

STATISTICAL ANALYSIS
for Students in
PSYCHOLOGY and EDUCATION

STATISTICAL ANALYSIS

for Students in

PSYCHOLOGY and EDUCATION

ALLEN L. EDWARDS

Associate Professor of Psychology

University of Washington

RINEHART & COMPANY, INC.

Publishers

New York

First Printing *May, 1946*
Second Printing *July, 1947*
Third Printing *November, 1948*
Fourth Printing *April, 1950*
Fifth Printing *January, 1951 .*

COPYRIGHT, 1946, BY ALLEN L. EDWARDS
PRINTED IN THE UNITED STATES OF AMERICA
BY VAIL-BALLOU PRESS, INC., BINGHAMTON, N.Y.
ALL RIGHTS RESERVED

PREFACE

Despite the recognition accorded statistical methods by active workers in psychology, education, and the social sciences, the average undergraduate major in these fields often regards the customary course in statistics as dull and uninteresting. And there is no denying the fact that his actual experience in the course may serve to bolster rather than to change his original opinions. This is apt to be true, I believe, when the instructor assumes a degree of mathematical training, however elementary, which the undergraduate major in the social sciences fails to have. Even when the student has had such training, its value may be largely nullified by the time interval which separates it from the course in statistics.

Another factor producing lack of interest, I believe, is the stress which is often placed upon long and involved problems which are essentially exercises in multiplication, subtraction, division, and addition. The student often regards these problems—and perhaps rightfully so—as so much “busy-work.” Such problems subtract from the time which the student feels could be more profitably spent in learning to appreciate the use and value of statistical techniques in his chosen field.

Under these conditions the student’s memories of statistics are memories of laborious computations and mysterious formulas, and it is these memories which he passes on to next semester’s class.

This text attempts to break the vicious circle. Little stress is placed upon calculative ability. I have tried in most cases to minimize the labor of computations by the use of illustrations and problems constructed with that end in view. In

addition, coding techniques for reducing the size of numbers are introduced early and stressed throughout. I would not change this emphasis in the *introductory* course even if a sufficient number of calculating machines were on hand to make one available to each student. In my opinion, it is not the function of this first course in statistics to train computers and machine operators. Let the beginning student get a picture of the use to which statistical techniques can be put in answering questions in his field of specialization. Let him see that statistical techniques are tools, instruments. Let him understand the simple formulas and the meaning of various statistics. He can then learn machine techniques of calculation if he ever needs to handle large masses of data.

To avoid assumptions concerning the student's mathematical training, a review of elementary principles has been included. An understanding of these should enable the student to follow subsequent developments. Stress on this section of the text will assist greatly in minimizing a major source of confusion for most students. Some theory and derivations are introduced throughout the text, but nothing is presented, I feel, which is beyond the comprehension of the nonmathematically trained student, if he is assisted by the instructor. In the last analysis, I am convinced that statistics, unlike many other college courses, must be taught, not lectured.

The content of the book follows what I hope is a pedagogical as well as a logical order. Correlation is introduced earlier than in most texts because it has been my experience that students follow this development quite easily and that interest and motivation are increased because they see in correlation a technique of practical use. A good case could be made for introducing correlation by way of regression. But since regression is most often used in psychology and education for prediction, it is reserved for a later chapter on this topic.

Tests of significance have been stressed. One chapter has been devoted to the "*t* test," two to the "*F* test," and one

to the " χ^2 test." It has been my experience again that students are able to relate these tests to problems in their own fields and are consequently interested in them. It is not uncommon to hear students exclaim with some degree of elation that for the first time they have some idea of what is meant by the frequent references to "critical ratios" and "significant differences" that they have encountered in their textbooks in other courses.

In discussing tests of significance, I have emphasized small sample theory since, whether the traditional attitude approves or not, more and more research as published in psychological and educational journals is being evaluated by small sample techniques. This does not mean, however, that there is a rigid division between large and small sample theory, but rather, as Kenney has said, that the "continuity between large and small sample theory is an essential part of the newer attitude" (53, II, 123).

There are omissions as well as additions in this text. The reader will look in vain for the customary treatments of partial and multiple correlation. They have been omitted because I have searched without much success to find many applications of these statistical techniques in the literature. That they have their uses as well as their limitations is fully recognized, but I feel that they may be developed more profitably, as far as the student is concerned, in advanced courses, where time spent on these topics will not be at the expense of statistical techniques which are more commonly in use. I have no excuse to offer for certain other omissions—for example, the customary chapter on the reliability and validity of tests, and the usual extended treatment of centiles and scaling techniques—other than the fact that I do not feel that they fall within the general orientation of the rest of the book and that these are problems which might well be taken up in a course in tests and measurements.

I have, as any student must, a desire to acknowledge my indebtedness to various individuals. Professor Harold Edgerton of Ohio State University first impressed upon me the

desirability of knowing more about statistical methods and theory. Professor Lloyd G. Humphreys of Northwestern University aroused my interest in small sample theory in 1939 and gave freely of his time in discussions of the subject. It is a pleasure also to acknowledge that I owe much to the various publications of Professor George W. Snedecor, Professor Helen M. Walker, Professor R. A. Fisher, and Professor C. C. Peters. In addition, Professor Walker and Professor Peters were kind enough to clarify, in personal communications, certain points of interpretation.

To Professors Herbert S. Conrad and Robert J. Wherry and Dr. Steuart H. Britt, who read a draft of the manuscript in its entirety, and to Dr. Edward E. Cureton, who read Chapters 10 and 11, and Professor Quinn McNemar, who read Chapter 14, a mere acknowledgment is hardly sufficient reward for the careful and painstaking service which they have rendered. I can only say that their comments proved invaluable in guiding me when I started to work on the revision of the original draft. The present text owes much to their efforts.

I am indebted to Professor R. A. Fisher, also to Messrs. Oliver & Boyd Ltd. of Edinburgh, for permission to reprint Tables C, D, and H, from their book *Statistical Methods for Research Workers*. Professors Peters and Van Voorhis and their publishers, The McGraw-Hill Book Company, kindly granted permission to reproduce Table F from their book *Statistical Procedures and Their Mathematical Bases*. Table E has been reproduced from Professor Snedecor's book *Statistical Methods* by permission of the author and his publisher, the Iowa State College Press. Additional values of t at the 1 and the 5 per cent levels were also taken from Professor Snedecor's book by permission. Table G was prepared by J. G. Peatman and R. Schafer and is reproduced by their permission and by the consent of Carl Murchison from the *Journal of Psychology*, where it first appeared. Other acknowledgments are made at various points throughout the text.

Preface

ix

Phyllis Covington and Jacqueline Charlton shared the major responsibility of typing the manuscript. Sidney S. Culbert assisted in the reading and proofing of the typed copy and Grace French in the checking of various computations. I am grateful to all of them for their assistance.

Finally, I owe a very special debt to Professor W. R. Clark of the University of Maryland, who encouraged me in my varied efforts to arouse student interest in statistical methods, and to my students both at the University of Maryland and at the University of Washington who responded to these efforts.

A. L. E.

Seattle, Washington
April, 1946

CONTENTS

	PAGE
PREFACE	v
CHAPTER	
1. INTRODUCTION	3
1. The text and the student	3
a. Previous mathematical training	3
b. Examples and problems	4
c. Use of tables	4
d. Symbols	5
e. Daily preparation	6
f. Empirical approach	6
2. Statistical terms and statements	7
a. Averages	7
b. Variability	8
c. Relationships	9
3. Functions of statistical methods	10
a. Precise description	10
b. Study of relationships	10
c. Formulation of experimental designs	11
d. Statistical inference	13
e. Prediction	13
2. SURVEY OF RULES AND PRINCIPLES	15
1. Fractions	15
2. Decimals	16
3. Proportions and per cents	17
4. Positive and negative numbers	18
5. Numbers in a series	21

CHAPTER	PAGE
6. Squares and square roots	21
a. Finding squares and square roots	21
b. Locating the decimal point	22
c. Squares and square roots of numbers less than 1	23
7. Summation	23
8. Simple equations	25
3. MEASURES OF CENTRAL TENDENCY AND VARIABILITY	30
1. An experiment on retention	31
2. The range as a measure of variation	31
3. The mean as a measure of concentration	34
4. Some basic symbols	34
5. The average deviation as a measure of varia- tion	36
6. The variance and standard deviation	38
7. The normal distribution curve	39
8. The median as a measure of central tendency	41
9. The semi-interquartile range	45
10. Centiles	45
11. Standard scores	47
12. Other measures of central tendency and vari- ability	49
13. Samples and statistics	50
4. SIMPLIFYING STATISTICAL COMPUTA- TIONS	54
1. The approximate nature of measurements	54
a. Significant figures	55
b. Common practice in reporting statistics	56
c. Rounding numbers	57
2. Coding by subtraction	58
a. The sum of scores	58
b. The sum of squares	60

Contents

xiii

CHAPTER	PAGE
3. Coding by division	63
a. The sum of scores	63
b. The sum of squares	65
4. Summary of "coding formulas"	66
5. Grouping measures into classes	67
a. The number of intervals	68
b. Size of the class interval	68
c. Limits of the intervals	69
d. Tallying the scores	70
e. Assumptions concerning grouped scores	70
f. Calculation of the mean and sum of squares	72
g. The "Charlier checks"	74
h. Calculation of the median	75
6. Summary of steps in coding	75
5. THE PRODUCT-MOMENT CORRELATION COEFFICIENT	79
1. The coefficient of correlation	80
a. A perfect positive correlation	80
b. A perfect negative correlation	83
c. A high positive correlation	84
d. A high negative correlation	85
e. A low correlation	86
2. Basic formulas for r	87
a. Standard deviations	87
b. Sum of squares method	88
c. Correlation using original measures	90
d. The difference method for Σxy	91
e. Correlation using coded scores	91
3. Correlation computed from a scatter diagram	94
a. Preparing the scatter diagram	94
b. The sum of scores and sum of squares	96
c. The sum of cross-products	97
4. Interpretation of correlation	99
a. The range of the correlation coefficient	99

CHAPTER		PAGE
	b. The coefficient of determination	99
	c. Common elements	100
	5. Purpose for which r is to be used	101
	6. Errors of measurement and correlation	102
6.	THE CORRELATION RATIO AND OTHER MEASURES OF ASSOCIATION	106
	1. The correlation ratio	107
	a. A simple method of computation	109
	b. Summary of steps in computing η^2	110
	2. Biserial correlation	112
	3. Tetrachoric correlation	116
	4. The phi coefficient	117
	a. The ϕ coefficient and true dichotomies	118
	b. The assumption of continuity	119
	c. The ϕ coefficient and tetrachoric r	121
	5. The contingency coefficient	122
	6. Rank-difference coefficient	123
	7. Multiple and partial correlation	125
	8. Summary of measures of association	127
7.	PROBABILITY AND FREQUENCY DISTRIBUTIONS	132
	1. Meaning of probability	133
	2. Combinations	136
	3. Binomial distribution	137
	4. The normal distribution curve	141
	5. The use of Table B	144
	6. Pragmatic considerations	146
	7. Skewed distributions	148
8.	SAMPLING DISTRIBUTIONS	151
	1. Samples and populations	151
	2. Sampling distributions	153
	3. Standard error of the mean	157

Contents

CHAPTER	PAGE
4. Large samples and the normal curve table	159
5. The concept of fiducial limits	162
6. Small samples and the table of t	166
7. Other standard error formulas	168
a. Standard error of the standard deviation	168
b. Standard error of a proportion	168
c. Standard error of a per cent or frequency	170
9. THE t TEST OF SIGNIFICANCE	172
1. An experiment involving paired observations	172
a. Standard error of a mean difference	174
b. Testing the null hypothesis	176
c. Establishing the fiducial limits	177
d. Another method for computing the standard error	178
2. Experiments involving matched groups	180
3. Experiments involving independent groups	181
4. The advantages of pairing observations	183
5. Testing the significance of a proportion	184
6. Testing the significance of r	185
a. The direct computation of t	187
b. The use of Table D	188
10. ANALYSIS OF VARIANCE: INDEPENDENT GROUPS	192
1. Nature of analysis of variance	192
a. The total sum of squares	194
b. The sum of squares within groups	195
c. The sum of squares between groups	195
d. Generalized formula for r groups	196
e. The variance ratio	198
2. A comparison of F and t in the case of two groups	199
3. The comparison of three groups	202
a. The total sum of squares	202

CHAPTER		PAGE
	b. The sum of squares between groups	203
	c. The sum of squares within groups	204
	d. The variance ratio	205
	4. A more complex analysis	207
11.	ANALYSIS OF VARIANCE: MATCHED GROUPS	217
	1. Analysis of variance of two matched groups	218
	2. Analysis of variance of several matched groups	225
	3. Correlation ratio and analysis of variance	232
	a. The correlation ratio without bias	232
	b. Tables of epsilon-square	233
	c. Epsilon-square and analysis of variance	234
	4. A test of rectilinear relationship	237
12.	THE χ^2 TEST OF SIGNIFICANCE	239
	1. Simple applications of χ^2	239
	a. Observed and expected frequencies	239
	b. Testing a 50 : 50 hypothesis	241
	c. Testing any a priori hypothesis	243
	d. χ^2 calculated from per cents	244
	e. χ^2 applied to more than two categories	245
	2. χ^2 applied to two samples	246
	3. χ^2 applied to more than two groups	250
	4. Testing "goodness of fit"	252
	5. χ^2 and small frequencies	253
	6. χ^2 and the ϕ coefficient	254
13.	PREDICTIONS AND THE EVALUATION OF PREDICTIONS	258
	1. Predicting simple characteristics	258
	2. Predicting measurements	261
	a. The regression line	262
	b. The regression coefficient	264
	c. The regression equation	265

Contents

xvii

CHAPTER	PAGE
d. Regression and correlation coefficients	265
e. The standard error of estimate	267
f. The index of forecasting efficiency	269
g. Predicting in the opposite direction	269
h. Coding and regression coefficients	270
3. Predicting from nonrectilinear relationships	270
14. RESEARCH AND EXPERIMENTATION	277
1. Interpretation of tests of significance	279
2. Samples and research	282
3. Size of the sample	287
4. Control groups	288
a. Control by random selection	289
b. Control by matching individuals	291
c. Control by matching groups	294
d. Single group serving as its own control	294
5. The <i>t</i> test and the assumption of homogeneity of variances	295
6. Additional problems in experimental design	299

BIBLIOGRAPHY

APPENDIX

TABLE A. SQUARES AND SQUARE ROOTS OF NUMBERS FROM 1 TO 1,000	307
TABLE B. AREAS AND ORDINATES OF THE NORMAL CURVE IN TERMS OF x/σ	320
TABLE C. VALUES OF <i>t</i> AT THE 5% AND 1% LEVELS OF SIGNIFICANCE	330
TABLE D. VALUES OF <i>r</i> AT THE 5% AND 1% LEVELS OF SIGNIFICANCE	331
TABLE E. VALUES OF <i>F</i> AT THE 5% AND 1% LEVELS OF SIGNIFICANCE	332
TABLE F. VALUES OF ϵ^2 AT THE 5% AND 1% LEVELS OF SIGNIFICANCE	336
TABLE G. TABLE OF RANDOM NUMBERS	340

<small>TABLE</small>	<small>PAGE</small>
TABLE H. TABLE OF χ^2	342
ANSWERS TO EXAMPLES	343
INDEX OF AUTHORS	353
INDEX OF SUBJECTS	355

STATISTICAL ANALYSIS
for Students in
PSYCHOLOGY and EDUCATION

CHAPTER 1

INTRODUCTION

1. THE TEXT AND THE STUDENT

Approached from the point of view that *statistical techniques are tools* to be used in experimentation and research, and in the discovery of new facts, the study of statistical methods can be a very interesting as well as valuable subject. As social scientists, are we interested in descriptions? Then statistical methods can assist us in making our descriptions more precise. Are we interested in differences between individuals and groups? Then statistical methods can assist us in describing and evaluating the reliability of observed differences. Are we interested in discovering whether there is any relationship between two traits, two abilities, or between information and attitude, or between juvenile delinquency rates and various areas of a city? Statistical methods again come to our assistance. These are applications of statistical methods to problems and there is no reason why such applications cannot be learned at the same time that the techniques are learned. That is the point of view stressed in this book.

a. **Previous mathematical training.** Not everyone who uses a stop watch is interested, or need be, in the detailed construction of the watch. The stop watch is a tool, an instrument, which can be used for measuring, describing, or evaluating time intervals. In a similar fashion statistical methods may be regarded as techniques for measuring, describing and evaluating data. To learn to apply elementary statistical techniques does not require any elaborate previous mathematical preparation. The field of mathematical statistics is so highly developed that not every worker in the field of psychology or education can be expected to be a specialist in both fields.

Automobile manufacturers publish two different sets of instructions to accompany the automobiles they produce; one book is intended for the driver of the car and the other is intended for the mechanic. Needless to say, the contents of the two books are not the same. The mechanic's book explains the working of the engine and other details. The driver's book tells him how to operate the car. The driver himself may never see the engine which makes his car go, but he takes it for granted that it is there and in good working order. Of course, if the car breaks down, then the driver must take it to the mechanic to get it repaired.

This text is more like the automobile book for drivers than like the one for mechanics. If while reading it you become interested in getting a better knowledge of the mathematical bases behind the techniques presented, then more advanced texts such as Peters and Van Voorhis (74) may be consulted.¹

b. Examples and problems. It is a generally recognized principle in psychology and education that one learns by doing. That is the purpose of the exercises and examples scattered throughout the text. Insofar as possible these examples have been selected for simplicity, but some are more complicated than others. Emphasis in the text is placed upon the procedures to be followed in making various computations and in interpreting the results of these computations. It is possible to learn to do this just as well with numbers that are small as with numbers that are large. In the few cases where large numbers have been used, you will find that the chapter on "simplifying computations" will enable you to "code" these numbers, i.e., to reduce their size, so that computations will be facilitated.

c. Use of tables. In the back of the book you will find a number of statistical tables which you will have occasion to refer to constantly. It is important that you know how to use these tables accurately. Each one will be explained

¹ Numbers in parentheses in boldface type refer to the bibliography at the end of the book.

in detail at the time at which it is first introduced in the discussion. Some of these tables are designed to simplify your work, such as the table of squares and square roots. This table will enable you to obtain square roots easily and will also give you the squares of numbers so that you may avoid unnecessary multiplication.

d. Symbols. A word or two should be said about the use of symbols. They are relatively few in number and each one has a specialized meaning. These symbols are in reality a form of shorthand, a simplified way of expressing something that would otherwise have to be written out in longhand. . . Some of these symbols stand for quantities and others stand for operations to be performed. You have used symbols before and they are nothing to be frightened about. See how much easier it is to write " $2 + 2 = 4$ " than it is to say, "The quantity two plus the quantity two gives the sum of four."

Here is a slightly different example and one which may be unfamiliar: $R = H - L$. If we were to have to put this into words we would say, "The range of measurements is equal to the highest measurement minus the lowest measurement." In the symbolic statement, $R = H - L$, R stands for range, H stands for the highest measurement, and L stands for the lowest measurement. Once having memorized the symbolic statement we can use it over and over again in place of the longer definition. In essence, then, symbols enable us to say a lot with little effort. Take them in stride, memorize each one as it is introduced, and you will find that they will give you little trouble.

What we have just said with respect to symbols applies also to formulas which are stated in terms of symbols. If you think of each formula as consisting of symbols which stand for quantities and operations to be performed, and that this is merely an abbreviated way of saying something, you will soon realize their value. The purpose of a formula, don't forget, is to simplify your work, not to make it more complicated.

e. Daily preparation. A book written about the subject of statistical techniques and a course in statistical techniques may not be quite like the usual texts and courses to which you are accustomed. Some courses do not require daily preparation, and many students get into the habit of waiting until just before an examination before getting down to work. By cramming they may succeed in absorbing a sufficient amount of knowledge, temporarily at least, to pass an objective or essay type of examination. But research upon the problem of retention of material learned in this fashion indicates that it is soon forgotten. Students may not consider this too great a handicap if they find that an understanding of later topics is not dependent on what has come before.

This is not the case with statistical methods. They cannot be successfully learned or mastered by cramming. Nor can the student, once having taken an examination, afford to forget the material studied and still expect to understand what is to come later. Statistical methods, as presented in this book, start from scratch; the assumption is that the student knows nothing at all about the subject. But there is a continuity of development, each new topic or section building upon the foundation established in the beginning. In certain respects this approach is like the construction of a house, in which the foundation, sides, and roof are built one upon the other. No good contractor attempts to put a roof on a house until he is sure of his foundation. The first few chapters in this book are the foundation of everything which appears later. Don't make the mistake of rushing through them because they may seem familiar or easy. The chances are very good that many of the questions you may ask about later developments have their answers in one of the earlier chapters.

f. Empirical approach. For practically every topic developed in this book there are several possible approaches. There is an algebraic development, a geometrical development, and an "empirical" or, as some might prefer to call

it, arithmetical development. By the empirical approach is meant the actual working through of a simple set of arithmetical computations to see that certain theorems or statements check as they should. More will be said about the empirical approach in the third chapter when we take up the subject of "averages and measures of variability." The empirical approach is stressed throughout the discussion so that the student without much previous knowledge of mathematics can follow the development of various topics. The interested or advanced student should realize that there is nothing to prevent him from deriving some of the formulas and proofs by other means. Some examples will be cited later.

2. STATISTICAL TERMS AND STATEMENTS

a. Averages. In our daily conversation we often use the term "average." We say that "John is better than average" when someone questions us about his golfing ability. Or that "Mary is slightly below average as a dancer" and "slightly above average in height." Some of our college courses we say we like "better than average." Some of the shoes we buy are "poorer than average." And, although we may not have defined the term in our own thinking as precisely as a statistician would, we have some general understanding of the concept. We may be vaguely aware that our statements concerning averages are based upon a series of observations or measurements and that each of these observations or measurements taken singly may not be the same as the average we have in mind. We perhaps have some scale in mind when we refer to John's ability as a golfer or Mary's height, and our average represents some middle position or value. The statements that "John is better than average" and that "Mary is slightly above average" indicate that we do not believe they represent this middle position.

We can find statements similar to these in books about psychology, education, and the social sciences, but they are

usually expressed more precisely than the statements we make about averages in our daily conversation.

"A group of 50 high school students, after viewing a motion picture which presented the Chinese in a very favorable light, showed an average shift toward the favorable end of a scale measuring attitude toward the Chinese of 2.5 scale points. A control group which had not seen the motion picture showed a shift of only 1.2 scale points."

"The average reading comprehension test score for 200 sixth-grade students was 82.3, while the average score on the same test for a group of seventh-grade students was 96.8."

"A group of subjects which had been given one hour of practice daily for five days in simple arithmetic computations made an average of 13.3 errors on a speed test. Another group with ten days of daily practice made an average of 8.4 errors on the same test."

All of these statements concerning averages were made possible by statistical methods.

b. Variability. We encounter another kind of statement which is made possible by statistical methods. In their simplest form they may appear as follows:

"The individual shifts in attitude scores for the group viewing the motion picture ranged from .8 to 7.3. For the group which did not see the motion picture the shifts ranged from .2 to 3.4 points."

"The range of scores on the reading comprehension test for the sixth-grade students was from 30 to 101; for the seventh-grade students the range was from 39 to 135."

"The number of errors for the group with one week of practice ranged from 2 to 21, while for the group with two weeks of practice the range was from 2 to 11."

These statements indicate something of the spread or differences among measures of individual performance. They tell us, taken in conjunction with statements about averages, that some of the measurements were above average and that others were below. These differences are as much

a matter of interest as are the averages, so much so to some psychologists that entire books have been devoted to the subject (2, 34). But we experience variability also outside our books in daily life. We note that not all incomes are the same but that some are very high and others very low; that the temperature is not the same but varies from hour to hour, from day to day, and from month to month. Not all synthetic tires have the same life span. Some give more mileage than others. Not all individuals are equally good at golf, dancing, and other skills.

c. Relationships. Sometimes we find statements which are not directly about averages or differences, but about relations between averages or differences. For example, in connection with the previous statements about reading comprehension scores for the 200 sixth-graders, we might find something like this:

"Those students who were above average on the reading comprehension test also tended to be above average in intelligence, as measured by an intelligence test, while those who were below average on one test also tended to be below average on the other. There was, in other words, a decided relationship between performance on the two tests, the *correlation coefficient* being .78."

We need not concern ourselves at this time with the meaning of "correlation coefficient" other than to note that it is a measure of relationship or association. Our interest here is in pointing out that relationships are also a subject of discussion in psychology and education. Statements concerning relationships probably appear as often in these fields as do statements concerning averages and differences. They too are made possible by statistical methods.

We also make constant reference to relationships in daily life, although these statements, like those about averages and differences, are not expressed as precisely as the statistician would like to make them. We note that a person's income may be related to the number of years of education he has; or that the amount of rainfall is related to the

season of the year; or that an individual's opinions on political questions may be related to the section of the country in which he lives. Or we might say about John's golf: "He's good. He practices a great deal." In this case we would indicate that we thought there was some relationship between his ability and the amount of practice he puts in on his golf.

3. FUNCTIONS OF STATISTICAL METHODS

a. Precise description. If you have followed the rather elementary discussion up to this point, then you are already familiar with some of the chief functions of statistical methods. In the behavioral or social sciences (and the examples in this book are selected largely from these fields) statistical methods enable us to study and to describe averages, differences, and relationships in a precise fashion. The problem of studying averages and differences may seem simple enough. If we are interested in the performance of college freshmen on a test of verbal facility, for example, we give a group of freshmen the test, find some measure of average performances and some measure of variability or individual differences. We shall have more to say about this problem later, but now let us see how we might investigate relationships.

b. Study of relationships. One obvious method of studying relationships is by making comparisons. We might compare the average performance of freshmen on our test with the average performance of college sophomores to determine if there is any relationship between year in college and performance. If we found that sophomores made a higher average score than freshmen, then we might assume that such a relationship does exist. We might feel even more confident of our assumption if we had also tested a group of juniors and a group of seniors and found that average performance increased from year to year. If we were so inclined, we might even carry our investigation on down through the various grades in high school. Note here, in

this example, that the problem of studying relationships is essentially the problem of studying differences; we observe differences in average performance for different year groups. We find out, in other words, whether there is any tendency for these differences to go together, to be associated.

On some occasions we may not find any basis upon which to classify individuals in order to get more than two groups. If we were interested in the relationship between sex and performance on our test of verbal facility we should have to be content with classifying our subjects as men or women and studying the average performance of each of these two groups on our test.

There is another method of approaching the problem of relationships. Instead of studying average differences between groups, we study the difference or relationship between paired measurements. Some examples with which you are probably already familiar are the relationship between point hour ratio in college and intelligence test scores, the relationship between height and weight, the relationship between motivation and learning. The problem here is similar to that discussed above, except that all of our subjects are considered as members of a single group. For each subject we have a pair of measurements and we determine the relationship between these pairs.

c. Formulation of experimental designs. It is sometimes possible for an investigator to control various factors in which he is interested and to manipulate experimentally others in order to study the relationships between them. Such a situation may be called an *experiment*. The example cited earlier concerning the influence of a motion picture on attitudes is a case in point. The factor introduced into the situation was the motion picture about the Chinese. By testing attitudes before and after the children had seen it, the influence of the picture on attitudes could be measured (75). Practice periods of different lengths may be given subjects in order to study the relationship between the amount of practice and performance. The behavior of

children may be observed under normal play conditions, and then factors designed to produce frustration in the children may be introduced into the situation in order to observe whether these factors result in any changes in play behavior (4).

Usually this approach to the study of differences and relationships involves an *experimental* and a *control* group, and the behavior or performance of the two groups is compared. The experimental group is the group for which some factor (practice, frustration) is varied while the control group does not experience the factor. The factor which is introduced into the experimental situation is ordinarily called the *experimental* or *independent* variable; the variable for which we observe changes is called the *dependent* variable.

There are various techniques for selecting and equating the experimental and control groups so that various factors which are pertinent to the problems under investigation may be controlled. If we had reason to believe that, in a particular investigation, age might be related to the behavior under study, then obviously we would want to have some assurance that this factor would not account for the results of our experiment. One way in which we might accomplish this would be by matching each individual in our experimental group with another individual of the same age in the control group.

Sometimes a particular experiment demands that our groups *already differ* with respect to a variable in which we are interested. This might be the case if we wished to study the effects of differing attitudes upon the learning and retention of different kinds of prose (14, 15). Will, for example, individuals who favor a given issue learn material which presents a favorable picture of the issue more readily than material which is opposed to it? Will the opposite tendency be present in individuals who are opposed to the issue? In this instance we might select for study groups which differ with respect to the attitude they hold on the

issue but which are matched with respect to some other variable, such as level of intelligence.

Statistical methods play a very important part in the planning of experiments as well as in the evaluation of the results of experiments. Setting up an experiment so that the most advantageous analysis of the results is possible is called a problem in *experimental design*. A sound experimental design is like a good blueprint; it gives confidence that the various parts are all going to fit together at the end.

d. **Statistical inference.** Having conducted an experiment or having made a series of observations and having described such things as averages, differences, and relationships, and having quantified these descriptions, we find that statistical methods enable us to make another step. We are often interested in knowing how reliable our descriptions are. If we repeated the experiment with other groups, to what extent would the new averages, measures of variation, and relationships differ from those we obtained the first time? Statistical methods enable us to determine the reliability of observed differences and relationships so that we may make generalizations with a given *degree of confidence*. The process by which we arrive at such generalizations is known as *statistical inference*.

e. **Prediction.** Suppose that we had studied a group of workmen operating a particular machine and that we had then constructed a test of some sort which we believed to be capable of measuring performance on the machine itself. Giving the test to a group of "good" workmen, we find that they make an average score of so many points and that a group of "poor" workmen make a much lower average score. Could we then predict from the scores of a *new* group of workmen how well they would probably perform on the machine in question? If we find the relationship between a scholastic aptitude test and college grades, then how accurately can we predict the average grades of other individuals, knowing only their scholastic aptitude test scores when they have not taken any college work?

If 70 per cent of a group of "maladjusted" individuals answer "Yes" to an item in a personality inventory and only 30 per cent of a group of "adjusted" individuals give a "Yes" response, then how effectively can we predict whether a person is adjusted or maladjusted merely from knowledge of his response to this item? Our prediction, of course, might be in reverse form; knowing that an individual is adjusted or maladjusted, how well can we predict his answer to the item? The problem of prediction and the accuracy of prediction is the final function of statistical methods with which we shall be concerned.

In summary, we now know something about the kinds of problems to which statistical methods can be applied. The chapters which follow simply discuss in greater detail the use of statistical methods: (1) in making precise descriptions of averages, differences, and relationships; (2) in the planning and design of experiments; (3) in determining the degree of confidence we may place in certain generalizations about our observations; and (4) in making predictions.

As a final note to this introduction and survey of what is to come, we might add that there are a number of statistical problems peculiar to test construction which are dealt with by various statistical techniques. But this is a field which has expanded so rapidly that it requires separate treatment. We shall touch upon such problems only indirectly; the student who desires additional information should consult Guilford (39) and Greene (37).

CHAPTER 2

SURVEY OF RULES AND PRINCIPLES

The rules and principles outlined in this chapter are extremely simple as well as extremely important. They deal with fractions, decimals, positive and negative numbers, squares and square roots, and simple equations. The material may be familiar to many students, but merely being able to work the examples is not sufficient. Working a problem when it is expressed in simple form is one thing, but unless you clearly understand the rule or principle involved which guided you in determining the answer, you may not be able to apply it to some of the formulas developed later.

1. FRACTIONS

A *fraction* is one method of stating that we are dealing with a sum which has been divided into a number of equal parts. The *numerator* of a fraction indicates the number of parts considered and the *denominator* indicates the equal parts. For example, $3/4$ indicates that a given sum or number has been divided into four equal parts and that we are dealing with three of these four parts.

Rule 1. The numerator and denominator of a fraction may be multiplied or divided by the same number or symbol without changing the value of the fraction. Thus starting with the fraction on the left and multiplying both the numerator and denominator by the same value we get the following identities:

$$\frac{4}{5} = \frac{8}{10} = \frac{24}{30} = \frac{48}{60} \quad \text{and} \quad \frac{2x}{3y} = \frac{4x}{6y} = \frac{12x}{18y} = \frac{24xy}{36y^2}$$

Observe, however, that adding or subtracting the same number or symbol from the numerator and denominator of

a fraction will, in general, change the value of the fraction. If we subtract 1 from both the numerator and the denominator of the fraction $\frac{4}{5}$ we get $\frac{3}{4}$, which is *not* the same value as the original fraction; and if we add 2 to the numerator and denominator of the fraction $\frac{2}{3}$ we get $\frac{4}{5}$, which is *not* the same value as our first fraction. An exception to the rule would occur when the numerator of the fraction is equal to the denominator. Thus subtracting 3 from the numerator and denominator of $\frac{9}{9}$ gives $\frac{6}{6}$, which does not change the value of the original; and adding 2 to both the numerator and denominator of the fraction $\frac{3}{3}$ gives $\frac{5}{5}$, which is also equal to the original value.

Rule 2. To add or subtract fractions they must first be reduced to a common denominator. We then add or subtract the numerators only; the denominator of the answer is the common denominator of the group of fractions added or subtracted. Thus

$$\frac{2}{6} + \frac{2}{3} + \frac{1}{2} = \frac{2}{6} + \frac{4}{6} + \frac{3}{6} = \frac{2+4+3}{6} = \frac{9}{6}$$

Rule 3. To multiply fractions merely multiply the numerators and multiply the denominators. This, in effect, serves to reduce them all to a common denominator. Thus

$$\frac{2}{3} \times \frac{3}{4} \times \frac{5}{6} \times \frac{4}{5} = \frac{2 \times 3 \times 5 \times 4}{3 \times 4 \times 6 \times 5} = \frac{120}{360} = \frac{1}{3}$$

Rule 4. To divide fractions, invert the divisor and multiply according to Rule 3 above. Thus

$$\frac{2}{3} \div \frac{1}{2} = \frac{2}{3} \times \frac{2}{1} = \frac{2 \times 2}{3 \times 1} = \frac{4}{3}$$

2. DECIMALS

Common fractions, as we have seen above, may have very different denominators. Decimals or decimal fractions, on the other hand, always have a denominator of 10 or some power of 10 such as 100, 1,000, 10,000, 100,000, and so on.

Thus .3 equals 3/10, .03 equals 3/100, .003 equals 3/1,000, and .0003 equals 3/10,000. Common fractions such as 1/2, 3/4, and 2/5 may be written as decimals by dividing the numerator by the denominator. Thus 1/2, 3/4, and 2/5 may also be written .5, .75, and .4, respectively.

Rule 1. When adding or subtracting decimals, keep the decimal points in a straight line and the decimal point in the answer should be directly under the decimal points of the figures subtracted or added. Thus

$$\begin{array}{r}
 .82 \\
 \underline{.90} \\
 1.72
 \end{array}
 \quad
 \begin{array}{r}
 .333 \\
 \underline{1.222} \\
 1.555
 \end{array}
 \quad
 \text{and}
 \quad
 \begin{array}{r}
 1.28 \\
 \underline{-.05} \\
 1.23
 \end{array}
 \quad
 \begin{array}{r}
 .83 \\
 \underline{-.11} \\
 .72
 \end{array}$$

Rule 2. In multiplying numbers involving decimals, point off as many decimal places in the product as there are decimal places in the multiplier and multiplicand together. The answer, in other words, will have as many decimal places as the sum of those in the two numbers multiplied. Thus

$$\begin{array}{r}
 .03 \\
 \underline{.09} \\
 .0027
 \end{array}
 \quad
 \begin{array}{r}
 .222 \\
 \underline{.10} \\
 .02220
 \end{array}
 \quad
 \begin{array}{r}
 2.20 \\
 \underline{.03} \\
 .0660
 \end{array}
 \quad
 \begin{array}{r}
 .0005 \\
 \underline{.2} \\
 .00010
 \end{array}$$

Rule 3. When dividing, place the decimal point in the quotient as many places to the right of the decimal point in the dividend as there are decimal places in the divisor. In other words, the number of decimal places in the dividend minus the number of decimal places in the divisor equals the number of decimal places in the answer. Thus

$$\begin{array}{r}
 2 \overline{) .004} \\
 \underline{.02}
 \end{array}
 \quad
 \begin{array}{r}
 .2 \overline{) 4.2} \\
 \underline{21}
 \end{array}
 \quad
 \begin{array}{r}
 .02 \overline{) .008} \\
 \underline{.4}
 \end{array}
 \quad
 \begin{array}{r}
 .03 \overline{) .90} \\
 \underline{30}
 \end{array}$$

3. PROPORTIONS AND PER CENTS

Rule 1. To find what proportion of a sum or total a given number is, divide the number by the sum or total. If, in a

class of 60 students, 15 students receive a grade of "C," and we wish to find the proportion receiving this grade, we divide 15 by 60 and our answer is .25. If, in an experiment, 35 subjects out of a total of 70 show a characteristic in which we are interested, and we wish to know the proportion showing the characteristic, we divide 35 by 70 and our answer is .50.

Rule 2. To translate a proportion into a per cent, multiply the proportion by 100. In the example above, the proportion of the subjects showing the characteristic is .5 and the per cent showing the characteristic is (.5) (100) or 50 per cent.¹ We see from this also that if we wish to translate a per cent into a proportion, we must divide the per cent by 100.

Rule 3. To find the number that a given proportion of a total equals, multiply the total by the proportion. If in a group of 40 students the proportion receiving a grade of "B" is .1, the number receiving this grade is (40) (.1) or 4. The same rule applies to a per cent, the per cent being written, of course, as a proportion or decimal.

Rule 4. Just as the sum of all per cents of a given total is equal to 100 per cent, so also the sum of all proportions of any given total is equal to 1.00. We shall see the importance of this later when we deal with the normal curve, which is tabled in terms of proportions.

4. POSITIVE AND NEGATIVE NUMBERS

Perhaps the simplest illustration of the meaning of a negative number can be given in terms of readings on a thermometer. Suppose that the temperature is now 20 degrees above zero and the weather man says that we can expect a drop of 25 degrees by nightfall. What temperature will it be then? On the thermometer we have numbers above and below zero, and if the weather man's prediction

¹ When numbers or symbols are enclosed in parentheses without any intervening signs, the operation of multiplication is indicated.

comes true, we would say that the temperature is 5° below zero, or -5° . Temperatures that are above zero are represented by a plus sign and those below zero by a minus sign. Ordinarily we omit the plus sign for numbers above zero, but whenever the number is below zero, we write a minus sign in front of it.

Just as minus and plus signs can be used to indicate temperatures above and below zero, they can also be used to indicate *directions* or *deviations* from some value other than zero. For example, knowing that the average height of a group of students is 67 inches, we could designate an individual with a height of 69 inches as being 2 inches above the average and an individual with a height of 65 inches as being 2 inches below the average. For these two values we could write 2 and -2 , respectively. And for all other values above the average we could write the values without any sign, the plus being understood but each value below the average would carry a minus sign.

Rule 1. To add numbers with the same sign we merely add and give the sum the common sign. Thus, adding the following, we get

$$2 + 3 + 4 + 6 + 8 + 10 + 1 = 34$$

$$(-2) + (-3) + (-4) + (-6) + (-8) + (-10) + (-1) = -34$$

Rule 2. To add two numbers with unlike signs, take the difference between the two numbers and attach the sign of the larger number. Thus, adding the following pairs, we get

$$\begin{array}{r} -2 \\ \underline{6} \\ 4 \end{array} \quad \begin{array}{r} 4 \\ \underline{-8} \\ -4 \end{array} \quad \begin{array}{r} -10 \\ \underline{9} \\ -1 \end{array} \quad \begin{array}{r} 8 \\ \underline{-9} \\ -1 \end{array} \quad \begin{array}{r} -5 \\ \underline{6} \\ 1 \end{array}$$

Rule 3. When adding a group of numbers with unlike signs, add the positive and add the negative numbers separately, following Rule 1, and then take the difference be-

tween the two sums and attach the sign of the larger quantity, following Rule 2. Thus

$$\begin{array}{r}
 2 \\
 -3 \\
 -7 \\
 -5 \\
 -1 \\
 \hline
 4 \\
 -10
 \end{array}
 \begin{array}{r}
 -20 \\
 -10 \\
 5 \\
 15 \\
 10 \\
 \hline
 -5 \\
 -5
 \end{array}
 \begin{array}{r}
 -4 \\
 -6 \\
 -5 \\
 10 \\
 5 \\
 \hline
 5 \\
 5
 \end{array}
 \begin{array}{r}
 7 \\
 -6 \\
 -4 \\
 3 \\
 5 \\
 \hline
 -10 \\
 -5
 \end{array}$$

Rule 4. To subtract one signed number from another, change the sign of the subtrahend and add according to the rules above. Thus, subtracting the following pairs, remembering that the sign of the number is written only when the number is negative, we get

$$\begin{array}{r}
 5 \\
 -3 \\
 \hline
 8
 \end{array}
 \begin{array}{r}
 4 \\
 -6 \\
 \hline
 10
 \end{array}
 \begin{array}{r}
 -4 \\
 -3 \\
 \hline
 -1
 \end{array}
 \begin{array}{r}
 -4 \\
 -8 \\
 \hline
 4
 \end{array}
 \begin{array}{r}
 -4 \\
 3 \\
 \hline
 -7
 \end{array}
 \begin{array}{r}
 -4 \\
 8 \\
 \hline
 -12
 \end{array}
 \begin{array}{r}
 4 \\
 5 \\
 \hline
 -1
 \end{array}
 \begin{array}{r}
 6 \\
 2 \\
 \hline
 4
 \end{array}$$

Rule 5. The multiplication of numbers with like signs gives a positive product; the multiplication of numbers with unlike signs gives a negative product. Thus, multiplying the following pairs, remembering that the sign of the number is written only when it is negative, we get

$$\begin{array}{r}
 6 \\
 -3 \\
 \hline
 -18
 \end{array}
 \begin{array}{r}
 -4 \\
 2 \\
 \hline
 -8
 \end{array}
 \begin{array}{r}
 -4 \\
 -2 \\
 \hline
 8
 \end{array}
 \begin{array}{r}
 4 \\
 2 \\
 \hline
 8
 \end{array}
 \begin{array}{r}
 3 \\
 -3 \\
 \hline
 -9
 \end{array}
 \begin{array}{r}
 -5 \\
 2 \\
 \hline
 -10
 \end{array}
 \begin{array}{r}
 -5 \\
 -5 \\
 \hline
 25
 \end{array}
 \begin{array}{r}
 4 \\
 -3 \\
 \hline
 -12
 \end{array}$$

Rule 6. The division of numbers with like signs gives a positive quotient; the division of numbers with unlike signs gives a negative quotient. Thus, dividing the following pairs, remembering that the sign of the number is written only when the number is negative, we get

$$\begin{array}{r}
 -3 \overline{) -6} \\
 \underline{2}
 \end{array}
 \quad
 \begin{array}{r}
 2 \overline{) -6} \\
 \underline{-3}
 \end{array}
 \quad
 \begin{array}{r}
 3 \overline{) 6} \\
 \underline{2}
 \end{array}
 \quad
 \begin{array}{r}
 -2 \overline{) 4} \\
 \underline{-2}
 \end{array}$$

5. NUMBERS IN A SERIES

Rule 1. Numbers in a series involving only the operation of multiplication or addition may be multiplied or added in any order without changing the answer. Thus

$$\frac{2 \times 3 \times 4 \times 5}{1 \times 2 \times 3 \times 4} = \frac{120}{24} \quad \text{and} \quad \frac{5 \times 3 \times 2 \times 4}{2 \times 1 \times 4 \times 3} = \frac{120}{24}$$

$$2 + 4 + 5 + 3 + 6 = 20 \quad \text{and} \quad 4 + 2 + 3 + 6 + 5 = 20$$

Rule 2. When the operations of division and multiplication are involved in numbers in a series along with the operations of subtraction and addition, the multiplication and division should be performed first. Thus

$$\begin{array}{lll} 2 + 3 \times 8 = 26 & 4 + 4 \div 2 = 6 & 4 + 8 \times 2 - 2 \times 1 = 18 \\ 3 \times 2 - 1 = 5 & 6 \div 2 - 1 = 2 & 3 + 2 + 2 \times 4 - 3 = 10 \end{array}$$

Rule 3. Terms within parentheses should be treated as a single number. Thus

$$(8 - 2)(6) = 36 \quad (4 + 2 - 1)(-2) = -10 \quad (3 \times 2) \div 6 + 2 = 3$$

Rule 4. A quantity cannot be divided by zero, and multiplication of a quantity by zero gives zero. Thus

$$8 \times 0 = 0 \quad 4 \times 2 \times 0 = 0$$

6. SQUARES AND SQUARE ROOTS

At the back of this book you will find a table of squares and square roots for numbers from 1 to 1,000 (Table A, p. 390). It is important that you know how to use this table correctly and how to locate approximate values for the square roots of numbers with over four figures. After you practice with a few examples, you will find this fairly easy to do.

a. Finding squares and square roots. There are three columns in the table. One column is headed N , the second is headed N^2 , and the third column is headed \sqrt{N} . To find the square root of any number from 1 to 1,000 find the number in the column headed N and read the square root

in the column headed \sqrt{N} . To find the square of any number from 1 to 1,000; find the number in the column headed N and read the answer in the column headed N^2 .

Suppose you wanted to find the square root of 49. By looking in the N column until you came to 49, you could then read the answer—which, of course, is 7—in the column \sqrt{N} . Now look in the N column at the number 7. Across the table in the N^2 column you find that 7 squared is 49. This should give you an indication of a second way of finding the square root of a number, a method that is particularly valuable when you have to find the square root of a number larger than any of those given in the N column or a number with four or more figures. If 7 squared is 49, then the square root of 49 is 7. Therefore, if you have a number larger than 1,000 or with four or more figures, look for the closest approximation of it that you can find in the N^2 column and read the square root in the N column. In this way you can find a good approximation of the square root of any number with as many as six figures.

b. Locating the decimal point. Before using the table of squares and square roots to find the square root of a number, always point off the number in pairs starting at the decimal point. Thus 30.8025 and 2,520.04, when pointed off, would be 30 .80 25 and 25 20 .04, respectively. When the number of figures to the right or left of the decimal point is odd, assume that a zero has been added. Thus 63,001, 2,294.4, 778.41, and 21.068, when pointed off, would be 06 30 01, 22 94 .40, 07 78 .41, and 21 .06 80, respectively.

For convenience, you may assume that the square root will have one figure for every pair in the number, the decimal point being located according to the number of pairs on each side of it in the number for which you are seeking the square root. Thus, for the figures given above:

$$\sqrt{30 \cdot 80 \ 25} = 5.55$$

Since there is one pair to the left and two pairs to the right of the decimal point.

$$\sqrt{25 \ 20 \ .04} = 50.2$$

Since there are two pairs to the left and one pair to the right of the decimal point.

$$\sqrt{06 \ 30 \ 01} = 251$$

Since there are three pairs to the left of the decimal point.

$$\sqrt{22 \ 94 \ .40} = 47.9$$

Since there are two pairs to the left and one pair to the right of the decimal point.

$$\sqrt{07 \ 78 \ .41} = 27.9$$

Since there are two pairs to the left and one pair to the right of the decimal point.

$$\sqrt{21 \ .06 \ 80} = 4.59$$

Since there is one pair to the left and two pairs to the right of the decimal point.

c. Squares and square roots of numbers less than 1. The square root of a number less than 1 is always greater than the number itself, and the square of a number less than 1 is always less than the number itself. Thus

$$\begin{aligned} \sqrt{.81} &= .9 & (.4)^2 &= .16 \\ \sqrt{.64} &= .8 & \text{and } (.3)^2 &= .09 \\ \sqrt{.0025} &= .05 & (.02)^2 &= .0004 \end{aligned}$$

7. SUMMATION

To summate means to add. When, for example, we summate a *variable* (a quantity which may assume a succession of values or simply that which varies) such as X for a given series of N measurements, we would merely add all of the N values of X in the series. This operation is indicated by Σ , the Greek capital letter *sigma*. Thus²

$$\Sigma X = X_1 + X_2 + X_3 + X_4 + X_5 + \dots + X_N$$

² A more precise method of indicating the summation in this instance would be to write it thus $\sum_{i=1}^N X_i$. These additional symbols above and below the summation sign would indicate the limits of the summation and would be necessary in order to avoid confusion when the summation might not extend over the entire series. However, since the summation in most elementary statistical problems is over the entire series of N cases, the limits will not be written but will be understood to be from 1 to N .

Rule 1. The summation of a *constant* (a value which does not change for a given series) is obtained by multiplying the constant by N , the number of times the constant appears in the series. For example, if we let k represent a constant, then Σk is equal to Nk . If k is equal to 3 and N equals 6, then

$$\begin{aligned}\Sigma k &= k_1 + k_2 + k_3 + k_4 + k_5 + k_6 = Nk \\ \Sigma k &= 3 + 3 + 3 + 3 + 3 + 3 = (6)(3)\end{aligned}$$

Rule 2. The summation of an algebraic sum of two or more terms is the same as the algebraic sum of the sums of these terms taken separately. What this rather complicated sounding rule means is that it is possible to write $\Sigma(x + y + z)$ as $\Sigma x + \Sigma y + \Sigma z$ and that the two are identical. We may illustrate this by letting N equal 3 and assigning numerical values to x , y , and z . Thus

$$\begin{aligned}x_1 &= 4; & y_1 &= 2; & z_1 &= -2; \\ x_2 &= 3; & y_2 &= -2; & z_2 &= 3; \\ x_3 &= -6; & y_3 &= 4; & z_3 &= 3;\end{aligned}$$

$$\begin{aligned}\Sigma(x + y + z) &= (4 + 2 - 2) + (3 - 2 + 3) + (-6 + 4 + 3) \\ &= 4 + 4 + 1 = 9\end{aligned}$$

$$\Sigma x + \Sigma y + \Sigma z = 1 + 4 + 4 = 9$$

Thus

$$\Sigma(x + y + z) = \Sigma x + \Sigma y + \Sigma z$$

Rule 3. The summation of a variable times a constant is equal to the constant times the summation of the variable. Thus if k is a constant and X is a variable, and if N equals 3, k equals 2, and $X_1 = 2$, $X_2 = 3$, and $X_3 = 4$, we may write

$$\begin{aligned}\Sigma(kX) &= (2)(2) + (2)(3) + (2)(4) = 4 + 6 + 8 = 18 \\ k\Sigma X &= (2)(2 + 3 + 4) = (2)(9) = 18\end{aligned}$$

Thus

$$\Sigma(kX) = k\Sigma X$$

Rule 4. The summation of a variable divided by a constant is equal to the summation of the variable, divided by the constant. Thus if k is a constant and X is a variable, and if N equals 3, k equals 2, and $X_1 = 2$, $X_2 = 4$, and $X_3 = 6$, we may write

$$\Sigma\left(\frac{X}{k}\right) = \frac{2}{2} + \frac{4}{2} + \frac{6}{2} = 1 + 2 + 3 = 6$$

$$\therefore \frac{\Sigma X}{k} = \frac{2 + 4 + 6}{2} = \frac{12}{2} = 6$$

Thus

$$\Sigma\left(\frac{X}{k}\right) = \frac{\Sigma X}{k}$$

8. SIMPLE EQUATIONS

For performing operations upon equations there is one simple rule: whatever is done to one side of the equation must also be done to the other side. If you multiply one side by a number or symbol you must multiply the other side by the same number or symbol. The same rule applies to division, addition, subtraction, squaring, and extracting the square root. If you have difficulty in seeing the relationships in the following examples, try substituting numerical values for the symbols. But it is important, very important, that you learn to do these operations with symbols other than numbers. Study the illustrations carefully.

1. Division

$$\Sigma X = NM \quad \text{and dividing both sides by } \frac{\Sigma X}{N} = M$$

N we get

$$\Sigma x^2 = Nt^2 \quad \text{and dividing both sides by } \frac{\Sigma x^2}{N} = t^2$$

N we get

2. Multiplication

$$\frac{x}{s} = t \quad \text{and multiplying both sides} \quad x = ts$$

by s we get

$$s^2 = \frac{\Sigma x^2}{N} \quad \text{and multiplying both sides} \quad Ns^2 = \Sigma x^2$$

by N we get

3. Subtraction

$$X = x + M \quad \text{and subtracting } x \text{ from both} \quad X - x = M$$

sides we get

4. Addition

$$x = X - M \quad \text{and adding } M \text{ to both sides} \quad x + M = X$$

we get

5. Extracting square root

$$s^2 = \frac{\Sigma x^2}{N} \quad \text{and extracting the square} \quad s = \sqrt{\frac{\Sigma x^2}{N}}$$

root of both sides we get

6. Squaring

$$x = (X - M) \quad \text{and squaring both sides we} \quad x^2 = (X - M)^2$$

get

There is another point to remember in dealing with equations. It is possible to substitute any identity for a term already present in the equation. Here is an example:

1. If $M = \frac{\Sigma X}{N}$ and multiplying both sides by N we get

2. $NM = \Sigma X$ and if

3. $\Sigma x^2 = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$ then we may substitute the identity NM for ΣX and get

4. $\Sigma x^2 = \Sigma X^2 - \frac{(NM)^2}{N}$

These principles are fundamental. If you understand them they will take away any of the mystery which might surround some of the formulas we shall develop later. We

shall also point out various identities as we go along. If you memorize them when they first appear you should be able to recognize them when they are substituted for one another later.

SURVEY OF RULES AND PRINCIPLES

Example 1.—Add each of the following:

(a)	(b)	(c)	(d)	(e)	(f)	(g)
-8	4	8	20	10	-6	-6
-3	-2	-9	-10	-8	2	3
(h)	(i)	(j)				
-9	0	-4				
1	-16	-2				

Example 2.—Subtract each of the following:

(a)	(b)	(c)	(d)	(e)	(f)	(g)
-8	4	8	20	10	-6	-6
-3	-2	-9	-10	-8	2	3
(h)	(i)	(j)				
-9	0	-4				
1	-16	-2				

Example 3.—Check each of the following by marking (1) if true or (2) if false:

(a) $\frac{(49 + 8)}{7} = 7 + 8$

(e) $\frac{(6 \times 2 \times 2)}{2} = (3)(2)(2)$

(b) $2(2 \times 5) = (4)(10)$

(f) $\frac{(6 \times 5)}{(2 \times 3)} = \left(\frac{6}{2}\right)\left(\frac{5}{3}\right)$

(c) $\left(\frac{1}{4}\right)(4 \times 6) = \left(\frac{4}{4}\right)\left(\frac{6}{4}\right)$

(g) $\frac{2x}{2y} = \frac{x}{y}$

(d) $\frac{(4 + 2)}{4} = 2$

(h) $\frac{4x}{3y} = \frac{12x}{9y}$

$$(i) \frac{\left(\frac{2}{6}\right)}{\left(\frac{1}{3}\right)} = \left(\frac{2}{18}\right)$$

$$(k) \frac{\left(\frac{2}{3}\right)}{\left(\frac{2}{3}\right)} = 1$$

$$(j) \frac{\left(\frac{3}{4}\right)}{\left(\frac{1}{4}\right)} = \left(\frac{12}{4}\right)$$

$$(l) \frac{(8-3)}{2} = 4-3$$

Example 4.—Perform the operations indicated: .

$$(a) (6 + 1)^2$$

$$(h) \frac{8}{-2}$$

$$(o) \frac{.04}{.02}$$

$$(b) (4 + 1 - 2)$$

$$(i) \frac{9}{-3}$$

$$(p) \frac{.04}{.002}$$

$$(c) (2 + 3)^2$$

$$(j) (-3)(-8)$$

$$(q) \frac{.04}{.2}$$

$$(d) (8 - 2)^2$$

$$(k) (2)(-5)$$

$$(r) \frac{.4}{.01}$$

$$(e) (4 - 1 - 2)^2$$

$$(l) (-3)(2)$$

$$(s) \frac{.3}{.5}$$

$$(f) (8 - 5 + 1)^2$$

$$(m) (-1)(-6)$$

$$(t) (.02)(.02)$$

$$(g) \frac{-8}{2}$$

$$(n) (-1)(0)$$

$$(u) (.1)(.1)$$

Example 5.—There are 60 students in a class and 15 receive a grade of B.

(a) What per cent received a grade of B?

(b) What proportion failed to get B?

Example 6.—Sixteen out of 64 students passed an item on a test.

(a) What per cent failed the item?

(b) What proportion succeeded in passing the item?

Example 7.—A student poll showed that in a sample of 200, 60 voted "No."

(a) What per cent voted "No"?

(b) What proportion failed to vote "No"?

Simple Equations

Example 8.—Find the square root of each of the following from Table A in the appendix:

- | | | | |
|-------------|------------|-------------|-------------|
| (a) .04 | (f) 20,736 | (k) 5.9536 | (p) 38,809 |
| (b) .81 | (g) 1,024 | (l) 10.0489 | (q) 98,596 |
| (c) .0016 | (h) 4,356 | (m) 51.6961 | (r) 157,609 |
| (d) .000025 | (i) 9,801 | (n) 99.6004 | (s) 30,276 |
| (e) .4624 | (j) 6,724 | (o) 37.21 | (t) 966,289 |

CHAPTER 3

MEASURES OF CENTRAL TENDENCY AND VARIABILITY

A simple and effective experimental design is to observe changes in performance or behavior of members of the same group under differing sets of conditions or before and after they have experienced some variable which the experimenter has introduced. When variables which might have influenced the results have been excluded or equated, any observed changes may be assumed to be the result of the differing conditions. In this way one might study the influence of motion pictures upon attitudes, the effect of a course in propoganda analysis upon ability to analyze propoganda, and, in general, the effect of any variable or set of conditions which it is possible for the experimenter to introduce upon behavior.

When it is not possible or feasible to study the behavior of the same individuals under differing conditions, the experimenter may resort to a matching procedure in order to select two comparable groups for observation. Individuals might be matched upon the basis of intelligence test scores, reading comprehension scores, attitudes, or some other variable which may be related to the variable under study.¹ We need not concern ourselves at this point with why this particular type of experimental design is efficient; we have mentioned the subject by way of introduction to the hypothetical experiment, the data of which we wish to discuss.

¹The reasons for this must await the development of correlational techniques and tests of significance.

1. AN EXPERIMENT ON RETENTION

Suppose that on some nights we read a sociology text just before going to bed and that on other occasions we do our reading in the morning. After a period of several weeks we have the impression that our memory of what we have read is much better when our period of study has been followed by sleep than when it has been followed by a period of waking activity. In order to investigate the problem further, we design a simple experiment to test retention under the two conditions.

We have as subjects for our experiment two groups. Each individual in one group has been matched with another individual in the second group on the basis of an academic aptitude test which we already have reason to believe is a variable related to retention and learning. Our experimental procedure is to have both groups of subjects learn a list of twenty words by the *method of paired associates*. In this method words are presented in pairs, and the subject is supposed to learn to respond with the second member of a pair when the first is presented. We have all of our subjects go through the list until they achieve one perfect trial, i.e., one trial with no errors. This learning period in the case of one of our groups is followed by eight hours of sleep and in the case of the other group is followed by eight hours of uncontrolled waking activity. At the end of the eight-hour period both groups are retested. The figures given in Table 1 show the number of correct responses on this second test.

2. THE RANGE AS A MEASURE OF VARIATION

In this idealized and hypothetical experiment, the superiority of the members of the "sleep" group is, as we see by the figures in Table 1, apparent in every case.² Every one of the members of this group makes a higher score than

² The data cited are hypothetical for purposes of illustration and simplicity, but see the study by Jenkins and Dallenbach (48).

TABLE 1.—RETENTION SCORES OF MATCHED INDIVIDUALS FOLLOWING EIGHT HOURS OF DIFFERING DEGREES OF ACTIVITY

PAIR	GROUP		DIFFERENCE BETWEEN PAIRS	DEVIATIONS AND SQUARED DEVIATIONS					
	Sleep	Wake		(5)	(6)	(7)	(8)	(9)	(10)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	X	Y	D	<i>x</i>	<i>x</i> ²	<i>y</i>	<i>y</i> ²	<i>d</i>	<i>d</i> ²
1	18	14	4	3	9	4	16	-1	1
2	12	8	4	-3	9	-2	4	-1	1
3	15	10	5	0	0	0	0	0	0
4	16	9	7	1	1	-1	1	2	4
5	14	8	6	-1	1	-2	4	1	1
6	15	10	5	0	0	0	0	0	0
7	15	9	6	0	0	-1	1	1	1
8	17	11	6	2	4	1	1	1	1
9	18	13	5	3	9	3	9	0	0
10	13	6	7	-2	4	-4	16	2	4
11	16	10	6	1	1	0	0	1	1
12	19	14	5	4	16	4	16	0	0
13	20	16	4	5	25	6	36	-1	1
14	17	8	9	2	4	-2	4	4	16
15	14	8	6	-1	1	-2	4	1	1
16	10	8	2	-5	25	-2	4	-3	9
17	14	9	5	-1	1	-1	1	0	0
18	15	10	5	0	0	0	0	0	0
19	13	11	2	-2	4	1	1	-3	9
20	9	8	1	-6	36	-2	4	-4	16
Σ	300	200	100	0	150	0	122	0	66

$$M_x = \frac{\Sigma X}{N} = \frac{300}{20} = 15$$

$$\sigma_x = \sqrt{\frac{\Sigma x^2}{N}} = \sqrt{\frac{150}{20}} = \sqrt{7.5} = 2.74$$

$$M_y = \frac{\Sigma Y}{N} = \frac{200}{20} = 10$$

$$\sigma_y = \sqrt{\frac{\Sigma y^2}{N}} = \sqrt{\frac{122}{20}} = \sqrt{6.1} = 2.47$$

$$M_d = \frac{\Sigma D}{N} = \frac{100}{20} = 5$$

$$\sigma_d = \sqrt{\frac{\Sigma d^2}{N}} = \sqrt{\frac{66}{20}} = \sqrt{3.3} = 1.82$$

his mate in the "wake" group. But observe the variation exhibited by the scores within each group. If there were no differences in the retention scores of the members within each group, then the differences in the retention scores

between pairs of matched individuals would all be the same. In fact, the differences would all be the same whether matching had taken place or not. In this case we would have no need of statistical methods nor would we have any need to observe more than one pair of individuals, since all additional pairs would show the same constant difference in retention. The difference in retention for a single pair would, under these circumstances, give us complete information. But the tendency for individual measurements to vary is a fundamental fact of nature. That is one reason why we need the assistance of statistical methods in evaluating data.

A simple measure of the variation present in each group would be the *range*, which we have already defined as being the difference between the highest and the lowest measurement. For the "sleep" group the highest score is 20 and the lowest score is 9 and the range is therefore 11. For the "wake" group the highest score is 16 and the lowest score is 6 and the range is 10. We could find a similar measure of spread or variation for the *differences* between pairs. The range of these differences is from 9 to 1 and the range is therefore 8. Symbolically, we define the range as

$$R = H - L \tag{1}^*$$

where R = the range

H = the highest measurement in the series

L = the lowest measurement in the series

There is another type of range which is similar to the range discussed above. It is known as the *inclusive range* and tells us the number of *possible* different measurements we may have in a series with a given range. The range for a simple set of measurements might be, for example, from 15 to 10 or, in other words, 5. The inclusive range is merely the range plus 1, or, in this instance, 6. This means

* Formulas are numbered at the time they are first introduced and then referred to later by these numbers.

34 *Measures of Central Tendency and Variability*

that there *could be* 6 different possible values or scores in our series: 15, 14, 13, 12, 11, and 10. Symbolically, then, the inclusive range is defined as

$$R_i = (H - L) + 1 \quad (2)$$

where R_i is the inclusive range and H and L have the same meaning as before.

3. *THE MEAN AS A MEASURE OF CONCENTRATION*

Note that despite the spread or variability of the scores within each group, there is also a tendency for the various scores to cluster around the middle values rather than at the extremes. A single score toward the middle of the range would be more representative of *all* of the scores than a value selected from either extreme. The statistics which we use to measure this concentration are known as *averages* or measures of central tendency. The statistician may not always mean by average; however, the measure you may have in mind. The measure of which you are thinking is probably the *mean*, which is found by adding all of the scores and dividing by the number of scores. The mean is only one among several possible kinds of averages.

Let us find the mean for the "sleep" group, for the "wake" group, and for the differences between pairs. The totals or sums of the scores for each series are given at the bottom of Table 1. For the "sleep" group the total is 300, and since this sum is based on 20 observations we divide 300 by 20 and find the mean score for the group to be 15. In a similar manner we determine that the mean for the "wake" group is 10 and that the mean of the differences is 5. Note that *the difference between the two means is equal to the mean of the differences*, a relationship that will always hold true when measurements have been paired.

4. *SOME BASIC SYMBOLS*

Let us see now how it is possible to indicate symbolically the computations involved in finding the mean. We shall

let N equal the number of scores in a given series and let X represent the scores in the series. Then the individual scores might be represented by $X_1, X_2, X_3, X_4, \dots, X_N$, where the subscripts, 1, 2, 3, . . . N , stand for the particular measures. In the example under consideration we may let X represent scores for the "sleep" group. In a similar manner we may let Y represent scores for the "wake" group, Y_1 corresponding to X_1 , Y_2 to X_2 , and so on for each matched pair. The differences between the paired values of X and Y may be represented by D , and particular values of D may be represented by D_1, D_2, D_3 , and so forth.

Since N is the same for the X , Y and D scores, we do not need to worry about a separate symbol for indicating the number of cases in each series. If this were not true, however, we could use N_x, N_y and N_d to represent the different N 's. We shall use the symbol M to represent the mean, and the mean of the X series would be M_x , the mean of the Y series M_y , and the mean of the differences, M_d . We need one more symbol, one that we shall use very frequently, Σ , which is the Greek capital *sigma*. The symbol is an operational as well as a descriptive symbol and means to sum. Thus ΣX would mean "to sum the variable X ," or simply "summation X ," or "sum of the X 's." ΣY would mean "to sum the variable Y ," or "summation Y ;" and ΣD would mean "to sum the variable D ."

In terms of the symbols we have just discussed, it would now be possible for us to represent the mean of the X series by the following formula

$$M_x = \frac{X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8 + \dots + X_N}{N}$$

But since we have the symbol Σ , meaning to sum, we may merely write, in abbreviated form,

$$M = \frac{\Sigma X}{N} \quad (3)$$

where M = the mean

Σ = the sum of

X = each of the individual measurements or scores

N = the number of measurements in the series

Formula (3) is the generalized formula for the mean. We would only need to substitute Y for X to apply it to the Y series or D for X if we wished to find the mean of the D series. We have already pointed out that symbols and formulas are a kind of shorthand. You may observe, in this instance, how much more quickly, and with how much less space, $M = \frac{\Sigma X}{N}$ can be written than the statement for which it stands: "The mean of a series is equal to the sum of the individual measures in the series divided by the number of measures in the series."

5. THE AVERAGE DEVIATION AS A MEASURE OF VARIATION

We are now ready for a new symbol. You will soon learn to recognize its meaning as quickly as you now recognize that green is a symbol for "go." The new symbol that we want is one that will represent the *deviation* of an observed measure from the mean of the series. We shall use the symbol x to designate a deviation of X from the mean of the X series. Thus

$$x = X - M \quad (4)$$

where x = a deviation from the mean

X = the original measurement

M = the mean

In a similar manner we could use y to represent the deviation of a Y score from the mean of the Y 's and d to represent the deviation of a difference score (D) from the mean of the differences.

If we were to subtract the mean of the X scores from each of the X scores and sum for the series, in other words,

$\Sigma(X - M)$ or Σx , as we have done in column (5) of Table 1, we should find that *the sum of the deviations from the mean equals zero*, i.e., $\Sigma x = 0$. This is a basic statistical theorem. You will find it holds true for any series of measurements and can easily be verified in the case of the *Y* and *D* distributions of scores. This is one reason why we cannot simply add the deviations from the mean and divide by *N* in order to get a measure of average deviation or spread of scores from the mean. The simple average deviation would always equal zero and consequently would be of no value as a measure of variability.

We could, however, ignore the *signs* of the deviations and find the sum of the *absolute* values and divide this by *N*. The resulting value is called the *average deviation*. Symbolically, we would write

$$AD = \frac{\Sigma|x|}{N} \tag{5}$$

where *AD* = the average deviation

|x| = the absolute value of *x*, i.e., without regard to algebraic sign

N = the number of measures in the series

The average deviation is one of the easiest measures of variability to understand and had great popularity at one time. It is still of value if one must describe variation to a group of statistically inexperienced individuals, but it has been found to be of limited utility in statistical theory. You may wonder, if the average deviation is of so little value, why have we bothered to mention it? Why not simply use the range as our measure of variability? The answer to the first question is that the average deviation provides an introduction to the *standard deviation*, the measure of variability that we shall use most often. The answer to the second question is that the range also has its disadvantages. It is determined by only two scores and fluctuates much more from one series to another than do the other measures of variation such as the average devia-

tion or standard deviation. If we were to repeat our experiment on the effect of sleeping and waking periods on retention, for example, the range for each group and for the differences between pairs might differ greatly from the values we got the first time.

6. THE VARIANCE AND STANDARD DEVIATION

The most valuable measure of variability is the *standard deviation*, which is computed from the squares of the deviations from the mean and is represented by the symbol σ . We have already pointed out that ignoring the signs of the deviations as we did in calculating the average deviation does not lead to the development of any very significant statistical techniques. Squaring is the next step in simplicity of operations and, incidentally, the squared deviations will all be positive. Squared deviations from the mean, as we shall see later, form the basis of much of statistical theory.

If we square each of the deviations from the mean, sum, and divide by N , we obtain a measure which is called the *mean square* or *variance* and which is symbolized by v . The standard deviation is simply the square root of the variance.

Thus, if the variance is equal to $\frac{\sum x^2}{N}$, then the standard deviation is equal to \sqrt{v} , or, as it is more commonly expressed

$$\sigma = \sqrt{\frac{\sum x^2}{N}} \quad (6)$$

where σ = the standard deviation

x^2 = a deviation from the mean squared

N = the number of cases

The calculation of the standard deviation may be summarized in the following steps:

1. Find the mean

$$M = \frac{\sum X}{N}$$

- | | |
|---|--|
| 2. Find the deviation of each score from the mean | $x = (X - M)$ |
| 3. Square each deviation | x^2 |
| 4. Find the sum of the squared deviations (sum of squares) | Σx^2 |
| 5. Divide the sum of squares by N to find the variance or mean square | $v = \frac{\Sigma x^2}{N}$ |
| 6. Extract the square root to find the standard deviation | $\sigma = \sqrt{\frac{\Sigma x^2}{N}}$ |

Extracting the square root (Step 6) returns us to our original unit of measurement. You may follow these steps in the calculation of the standard deviations of the X , Y , and D series of measurements in Table 1. For the D series, for example, column (4) gives the scores which we sum to find the mean. Column (9) gives the deviations of each of these scores from the mean, and column (10) gives the deviation squared. The sum of the squared deviations is 66, which, divided by $N = 20$, gives the variance, 3.3. The standard deviation is the square root of 3.3 and from Table A, page 311, we find this to be equal to approximately 1.82.

7. THE NORMAL DISTRIBUTION CURVE

You may already be familiar with the concept of a *normal distribution* from other sources. A normal distribution is represented by a bell-shaped, symmetrical frequency curve, with very few measurements at the extremes and more and more as you move in toward the middle. It may look something like the curve shown in Figure 1.

Suppose that this distribution curve represented measurements of differences in retention for 10,000 pairs of subjects. That is, suppose that instead of merely 20 pairs as we had in the experiment mentioned earlier we had 10,000. We would not expect all of the differences in retention to be the same for these 10,000 pairs any more than they were for our 20 pairs. If we had 10,000 pairs we might sometimes get a difference of zero; sometimes, also, differences might

be in reverse, i.e., in some of the pairs the "waking" member might show a higher retention score than the "sleeping" member. But in terms of what we have already observed, we would expect most of these 10,000 differences to tend toward the middle or mean of the distribution.

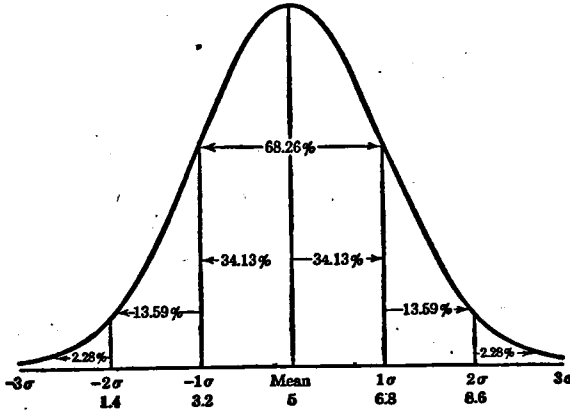


FIG. 1.—Normal distribution curve with mean equal to 5 and standard deviation equal to 1.8.

If the mean and standard deviation of this new distribution were the same as the mean and standard deviation of our 20 observations, then between the mean (5) plus and minus one standard deviation (1.8) would fall approximately 68.26 per cent of these 10,000 differences. In other words, between 5 ± 1.8 or between 3.2 and 6.8 would fall 68.26 per cent of the cases, and outside these limits would lie approximately 31.74 per cent of the differences. About 15.87 per cent of the differences would be greater than 6.8 and about 15.87 per cent would be smaller than 3.2. These statements are made possible by the fact that the equation for the normal curve is known, and tables have been prepared which enable us to find the proportion or per cent of cases between the mean and any given distance from the mean expressed in terms of standard deviation units. These tables are discussed in detail in a later chapter.

8. THE MEDIAN AS A MEASURE OF CENTRAL TENDENCY

In general, if a distribution is approximately normal, the mean is the appropriate measure to use to describe the central tendency of the group. If the distribution departs very much from the normal form so that scores are piled up at one end or the other of the scale, then another measure of central tendency may be used to supplement the description provided by the mean. This measure of central tendency is called the *median* and is defined as that point in a distribution of measurement above which and below which 50 per cent of the measurements lie. The median would also be the appropriate measure of central tendency to use if a distribution is truncated, i.e., cut off at one end so that we have no knowledge of the exact values of the measures at this end, as, for example, in a distribution of incomes where we might have at one end 7 cases which are simply recorded as \$15,000 *and over*. In a perfectly normal distribution the mean and median coincide, have the same value.

To illustrate the calculation of the median, let us suppose that we have a number of ratings on a 5-point scale and wish to find the median. Instead of writing out the value of each rating, we shall simply list the five possible values under the heading "Ratings" and then under *f* list the *frequency* or number of times each value occurs, as in Table 2. The rating "5," for example, occurs 4 times, the rating "4," occurs 3 times, and so on. Measurements arranged in the manner of Table 2 are called *frequency distributions*.

Since we have defined the median as a point, we shall have to pause for a moment to consider whether a score or a rating can be considered a precise point or not. It is customary in statistical work to think of a measurement, regardless of the instruments used in making it, as representing an *interval* ranging from half a unit below to half a unit above the given value. A height reported in terms of inches, for example, may be considered as representing

an interval ranging one-half inch below to one-half inch above the reported value. A height of 61 inches, in other words, may indicate a value ranging from 60.5 to 61.5. Even if the height were reported to the nearest 1/10 inch, 61.8 inches, for example, it might still represent an interval ranging from 61.75 to 61.85. This is because there are

TABLE 2.—FREQUENCY OF RATINGS ON
A 5-POINT SCALE

RATING	LIMIT	<i>f</i>
5	4.5-5.5	4
4	3.5-4.5	3
3	2.5-3.5	2
2	1.5-2.5	1
1	.5-1.5	1

limits to the accuracy of any measuring instrument. Regardless of how fine we may make our units of measurement, i.e., how many decimal places may be used in reporting them, we still do not know the *precise value* of the final number. Considered in this fashion, then, a rating of 5 may mean a value from 4.5 to 5.5 and a rating of 1 may mean from .5 to 1.5.

To find the median we must first find out how many ratings we have under consideration. This we do by adding the frequencies, 4, 3, 2, 1, and 1. *N*, then, is 11, and we wish to find the point above which and below which exactly 50 per cent or 5.5 of these 11 cases will fall. If we start counting upward from the lowest rating, we find that 1 + 1 + 2 will give us 4 of the needed 5.5 cases. This carries us through the rating 3, the upper limit of which is 3.5. We have moved up the scale, in other words, to the point 3.5 and have found 4 cases below here. But this is not sufficient; we need 5.5 cases or 1.5 more than the 4 we have so far. The rating 4 occupies the interval from 3.5 to 4.5, and there are 3 cases located within this interval. We do not know

how these 3 cases are distributed in the interval 3.5 to 4.5, but for convenience we *assume that they are distributed evenly throughout the interval*. We must move up into this interval until we have 1.5 more cases. We need, in other words, 1.5 of the 3 cases or $1.5/3$, which is equal to .5. We add this value (.5) to the lower limit (3.5) of the interval in which we know the median falls and this gives us the value of the median, 4.0.

We may, if we wish, check this value by counting down from the highest rating. We have 4 cases for the rating 5 which extends *down* to 4.5. We still need 1.5 more cases in order to get our 50 per cent. We need to go down into the interval 4.5 to 3.5 far enough to include 1.5 of the 3 cases which we assume to be distributed evenly throughout the interval. And $1.5/3$ gives us .5 which we now *subtract* (we are moving downward) from the *upper limit* (4.5) of the interval in which we know the median falls and arrive at the same value as before, 4.0, for the median.

Sometimes in computing the median we may find that 50 per cent of the measurements or scores take us exactly through a given score but that there is a gap between the upper limit of this score and the next score. For example, suppose we had the following measurements: 18, 17, 16, 14, 10, 8, 7, 5. N is equal to 8 and 50 per cent or $1/2 N$ is equal to 4. We need to find the point above which and below which 4 scores will fall. Counting up from the bottom or lowest score we find that 4 scores take us through 10, the upper limit of which is 10.5. It is true that 50 per cent of the scores do fall below the point 10.5, and that 50 per cent fall above this point. But it is also true that 50 per cent fall above and below any other point we might choose to select between 10.5 and 13.5. Under these circumstances *we assume that the value which best represents the median is the middle of the gap*. The range of the gap is equal to $13.5 - 10.5$ or 3. One half of 3 is equal to 1.5 and 1.5 added to the upper limit of 10.5 gives us a value of 12 for the median. You may check these calculations by

44. *Measures of Central Tendency and Variability*

counting down from the top, only in this instance, since you are moving downward, you would have to subtract 1.5 from the lower limit of the score 14. The value of the median remains the same, regardless of whether we calculate it by counting up or down.

If, in the distribution above, there had been no gap, i.e., if 10 had been followed by 11 rather than by 14, then the median would become the *dividing point between these two scores*. Since the upper limit of 10 is 10.5 and the lower limit of 11 is 10.5, the value arrived at for the median would be 10.5.

The following formula for computing the median will handle all situations except when the median falls in a gap in the distribution of measurements.

$$Mdn = l + \left(\frac{\frac{N}{2} - \Sigma f_o}{f_w} \right) i \quad (7)$$

where Mdn = the median

l = the lower limit of the interval containing the median

N = the total number of scores

Σf_o = the sum of the frequencies or number of scores up to the interval containing the median

f_w = the frequency or number of scores within the interval containing the median

i = the size or range of the interval (in the illustrations considered, since i has always equaled 1, it may be ignored—we include it here because this is a more generalized formula which can be used later)

The value of the median obtained with formula (7) may be checked, in the manner indicated earlier, by working from the top interval down. The formula in this case becomes

$$Mdn = u - \left(\frac{\frac{N}{2} - \Sigma f_o}{f_w} \right) i \quad (8)$$

where u = the upper limit of the interval containing the median, and Σf_o = the sum of the frequencies or number of scores down to the interval containing the median.

9. THE SEMI-INTERQUARTILE RANGE

The measure of variation which is generally used in connection with the median is the *semi-interquartile range* or Q . To find the value of Q , two other values must be computed, Q_1 , the first quartile, and Q_3 , the third quartile. These two values are also points on a scale, Q_1 being the point below which 25 per cent of the measurements fall and above which 75 per cent fall, and Q_3 being the point below which 75 per cent fall and above which 25 per cent fall. Q_1 and Q_3 are found in the same way that the median is found, i.e., by means of formula (7), the only difference being that for $\frac{N}{2}$ we substitute $\frac{N}{4}$ for Q_1 and for Q_3 we substitute $\frac{3N}{4}$.

The interval $Q_3 - Q_1$ contains the middle 50 per cent of the measurements and is known as the *interquartile range*. The semi-interquartile range is one half the range of the middle 50 per cent of the cases and is given by the following formula

$$Q = \frac{Q_3 - Q_1}{2} \quad (9)$$

where Q = the semi-interquartile range

Q_3 = the third quartile

Q_1 = the first quartile

10. CENTILES

Just as we used formula (7) to find the median or point above which and below which 50 per cent of the cases fall, and to find Q_1 and Q_3 , so also it can be used to find the point in a distribution above which and below which any given per cent of the cases fall. Such points are commonly called

centiles. Since the median marks the point above which and below which 50 per cent of the cases fall, it is also the 50th centile. The 25th centile is the same as Q_1 and the 75th centile is the same as Q_3 . The points dividing the distribution into tenths are also given special names; they are called *deciles*. Thus the 10th centile is also the 1st decile, the 20th centile is also the 2nd decile, and so forth.

If we wished to find a given centile, we would only need to substitute that per cent of the total scores or measurements for $\frac{N}{2}$ in formula (7). Thus, if we wished to find the 80th centile, which would be the point below which 80 per cent of the cases fall, $\frac{N}{2}$ would be replaced by $\left(\frac{80}{100}\right)(N)$ or by $\frac{4N}{5}$. To find the 33rd centile we would substitute $\left(\frac{33}{100}\right)(N)$ or $\frac{33N}{100}$. The 50th centile, the median, would be, of course, $\frac{50N}{100}$, which, simplified, is $\frac{N}{2}$.

Centiles are often used to describe an individual's relative position in a group with respect to some variable. For example, if we were told that John's score on a reading test was 49, and this was all that we were told, we would know no more about his ability than if we had not been told his score. If we knew that the mean score for college freshmen on the test was 40, we would at least know that he performed better than the average freshman. But if we were told that his score corresponded to the 75th centile, we would know that he does better than 75 per cent of the students who take the test.

One major difficulty with centiles as a means of expressing relative position is that, when distributions are fairly normal, individual differences relatively near the center of the distribution are exaggerated in comparison with the extremes. The actual measured differences represented by the centile

range 40 to 60, for example, are not as great as the actual measured differences represented by the centile range 1 to 21 and 79 to 99. This is because, as we know from our earlier discussion of the normal curve, frequencies are greater in the center of the distribution than at the extremes.

11. STANDARD SCORES

When a distribution of measures is approximately normal we may overcome the difficulty mentioned above in connection with centiles by expressing scores in terms of *relative deviates* or *standard scores*, symbolized by z . Standard scores derived from one distribution may also be compared directly with standard scores derived from another distribution, when both distributions are normal in form.

To illustrate the use of z scores, let us suppose we wish to find an average of an individual's score on a history test and on an English test. The history test is scored in terms of the number of right answers and shows a spread of scores from 10 to 190 with a mean of 95. The English test, however, is scored in terms of the number of right answers minus the number of wrong, and the range of scores is from 50 to 70 with a mean of 59. Obviously, we cannot compare directly the standing of our subject on one test with his standing on the other. We could not find his average standing on both tests by adding his score on the history examination with his score on the English examination and dividing by 2. This average would have no meaning, for we would be combining different units from different scales. It is as though we added together an individual's height, as measured in terms of inches, and his weight, as measured in terms of pounds, and divided by 2 to get an average. Suppose that we were foolish enough to do so and found that this average was 110. This would be 110, but 110 what? Inches? Pounds? Surely not either of these, nor would such an average have any other meaning.

If we wish to compare measurements from various dis-

tributions, we must first reduce the measurements of each distribution to a common scale. To do this for the distribution of scores on the history test, which we may designate as the X variable, we would subtract the mean of the distribution from each score in the distribution to reduce the measurements to deviation scores. Then we divide each deviation score by the standard deviation of the distribution to arrive at z scores. Thus

$$z_x = \frac{x}{\sigma_x} \quad (10)$$

where z = a standard score

x = a deviation from the mean of the series

σ_x = the standard deviation of the series

In the same manner we would transform each of the scores on the English examination, which we may designate as the Y variable, to z scores by subtracting the mean of the distribution from each score and dividing the resulting deviation scores by the standard deviation of the distribution. We may then add the z score of our subject on the history examination with the z score he obtained on the English examination and divide the sum of these by 2 in order to get an average. If we were interested in comparing individual scores on various other tests with those on the history and English examinations, these scores could also be reduced to z scores. As long as the various distributions are approximately normal in form, z scores derived from one distribution may be compared with z scores derived from other distributions. By following the procedure outlined we have, in effect, reduced the scores of each distribution to a common scale. Some of the z scores will, of course, carry negative signs, since some of the scores will be smaller than the mean. In general, a distribution of z scores will range in size from about plus 3 to minus 3. The mean of this distribution will be equal to zero and the standard deviation will be equal to 1. That the mean will be equal to zero, we know

because we have already found that the sum of deviations from the mean is equal to zero. The fact that the standard deviation of a full set of z scores equals 1 can be shown as follows:⁴

$$\sigma_{z_z} = \sqrt{\frac{\sum \left(\frac{x}{\sigma_x}\right)^2}{N}} = \sqrt{\frac{\sum \frac{x^2}{\sigma_x^2}}{N}} = \sqrt{\frac{\sum x^2}{N\sigma_x^2}} = \sqrt{\frac{\sigma_x^2}{\sigma_x^2}} = \sqrt{1} = 1$$

12. OTHER MEASURES OF CENTRAL TENDENCY AND VARIABILITY

There are other kinds of averages than those we have mentioned. One is the *mode*, or measure which occurs most frequently in a distribution of measurements. Another is the *geometric mean* which is the n th root of the product of the N values in a series. The geometric mean of 3 and 12, for example, would be $\sqrt{(3)(12)} = \sqrt{36} = 6$, whereas the arithmetic mean would be 7.5. We shall have occasion to refer again briefly to the geometric mean in connection with measures of relationships. Another measure of central tendency is the *harmonic mean*, which is defined as the reciprocal⁵ of the arithmetic mean of the reciprocals of the values. The harmonic mean is used in problems involving the averaging of rates, but we shall have no need to refer to it again in this text.

There are also other measures of variability in addition to those which we have described. One such is the *middle 80 per cent range* or the spread of scores between the 10th and 90th centiles. Another is the *probable deviation* or *probable error* which was widely used in the past, but which is practically never used now to describe variability. The probable deviation is approximately 2/3 the size (more

⁴ Let it be emphasized again that if this development is not perfectly clear, you should go back and study the rules of Chapter 2. Nothing is involved here except the application of these rules.

⁵ A reciprocal of a given value is 1 divided by that value.

precisely, .6745) of the standard deviation and is found by multiplying the standard deviation by that value. In a normal distribution the interval established by the mean plus and minus one probable deviation contains the middle 50 per cent of the measures and is therefore equivalent to $Q_3 - Q_1$. The probable deviation has no advantages over the standard deviation; rather, the disadvantage lies in the fact that in order to find it we must first calculate the standard deviation.

The measures of central tendency and variability which we have treated briefly in this section are used very infrequently in psychology and education and, with the exception of the geometric mean, have little bearing upon the statistical methods developed later. We shall consequently say no more about them. Our basic measure of central tendency will be the mean and our basic measure of variability will be the standard deviation. We shall refer to these measures constantly. Be sure that you thoroughly understand their calculation.

13. *SAMPLES AND STATISTICS*

We have more or less avoided the use of the term "sample" up to this point but to continue to do so would prove awkward. In your own experience you have "sampled" foods and then made judgments or based future reactions on your experience with these samples, i.e., you may ask for more or you may refuse more because you assume that the remainder of the food will be very much like the sample you experienced. An observer standing by would probably note that you do two things when you sample: (1) you deal only with a part or a portion of some whole, and (2) you assume that this part or portion is in some way representative of the whole. This is very similar to the meaning of a sample in statistics.

The statistical sample consists of the particular group of observations which have been selected for investigation or study (88, p. 129) and, generally, the sample under study is

assumed to be representative of some larger group from which the sample was selected. The larger group is called a *population* or *universe*. A measure derived from a sample, such as the mean or standard deviation, is called a *statistic*. The corresponding mean or standard deviation which would be obtained if the population instead of the sample had been studied is called a *parameter*. Parameters, since they are based upon all the existing cases, have fixed, single values. Statistics, on the other hand, since they are based upon only a part of the total population, may vary from sample to sample.

Statistics, in the absence of any other information, are the best *estimates* of the population parameters we have. The two statistics which we have demonstrated in this chapter, the mean and standard deviation are, as we have emphasized previously, basic. To find them you need compute but two sums: the sum of scores (ΣX) and the sum of squares (Σx^2). The sum of scores is necessary for the mean and the sum of squares for the standard deviation. Later we shall find that there are easier ways of computing these statistics when we have to deal with either a large number of observations or when the measures as such have large numerical values.

MEASURES OF CENTRAL TENDENCY AND VARIABILITY

Example 1.—A class in applied psychology made the following scores on a weekly quiz. (a) Find the mean of the scores. (b) Now subtract the mean from each score and sum these deviations. Are you surprised at your finding? Could you use this device to verify the calculation of a mean?

30	28	26	25	23	21	20
29	28	26	24	23	21	20
29	27	26	24	22	21	19
29	27	25	24	22	20	19
28	26	25	24	21	20	18

(g) v

(j) NM

(m) x

(h) $\frac{\sum x^2}{N}$

(k) $N\sigma^2$

(i) $\frac{\sum(X - M)^2}{N}$

(l) $\sqrt{\frac{\sum x^2}{N}}$

Example 8.—Show, algebraically, that the sum of the deviations from the mean is equal to zero.

CHAPTER 4

SIMPLIFYING STATISTICAL COMPUTATIONS

The computation of the mean and standard deviation is quite simple, as long as we are dealing with relatively few measurements or when the numerical size of the measurements is small. But when we have a great many scores and when the values of these are large, as may often be the case, then we need some method for simplifying our work. This is achieved through *coding*, a means of reducing scores or measurements.

1. THE APPROXIMATE NATURE OF MEASUREMENTS

You may recall that in the last chapter we touched briefly upon the meaning of a measurement or score when we considered the calculation of the median. At that time we pointed out that measurements are made and reported to the nearest unit, whatever that unit happens to be. Height, for example, may be reported to the nearest inch despite the fact that there is not a jump from one inch to the next, but a *theoretically* infinite gradation of units between each. The distance between 61 inches and 62 inches, for example, might be divided into tenths and reported 61.1, or divided into hundredths and reported 61.01, or thousandths and reported 61.001, and so on. A height, then, reported simply as 61 inches is not the precise value upon close examination that it might at first seem to be. But then neither would a reported value of 61.001 inches be an exact figure, for, regardless of the units of measurement, theoretically an instrument might be constructed which would measure with a greater degree of precision.

This is true of all measurements. Time may be measured in terms of years, months, weeks, days, hours, minutes, seconds, milliseconds, and so on, each succeeding unit being

more precise than the one before, but even milliseconds are not exact values but only approximate. What we have said about time applies also to other measurements with which you may be familiar: temperature, weight, brightness, intensity of sound, and so forth.

Because of the approximate nature of measurements, we customarily, in statistics, regard a height reported in terms of the nearest inch, such as 61 inches, as representing an interval ranging from 60.5 to 61.5, i.e., half a unit above and half a unit below the value reported. We regard psychological test scores and other measurements in the same manner. An intelligence test score of 82 is taken to mean from 81.5 to 82.5; an attitude test score of 23 is considered as representing an interval from 22.5 to 23.5. It is conceivable, in each instance, that if our units of measurement on these scales had been more refined, then the obtained values might have been somewhat higher or somewhat lower than the scores, 82 and 23, indicate. If this disturbs your previous beliefs about the accuracy of figures, then you might take comfort in the thought that most of our units of measurement are precise enough for the situations in which we are interested.

a. **Significant figures.** A frequent question coming from students is: How many decimal places shall I carry in my computations? There is no exact answer to this question as it is phrased. More properly, the question should be: How many *significant figures* should I carry? But even here there is no exact answer; there are only "good" or "established" practice and "poor" or not common practice—like "good" and "bad" usage in English. In view of what we have said concerning the approximate nature of measurements, the figures 28, 280, and 2,800 each contain but two significant figures. That is because the zeros used in the second and third numbers are merely used to locate decimal points, they are "fillers." The first value, 28, represents a range from 27.5 to 28.5; the second, 280, a range from 275 to 285; and the third, a range from 2,750 to 2,850. How-

ever, if 280 and 2,800 had been written 280. and 2,800., with a decimal point, then the zeros would have been considered significant figures and the range would be 279.5 to 280.5 and 2,799.5 to 2,800.5, respectively. In the measurements used throughout this book, we shall follow the fairly common practice of *not* writing the decimal point after figures such as 70 or 60 or 210 but *assume that it is understood*. When a score is written as 60, for example, it will be assumed that this represents a range from 59.5 to 60.5.

There are "rules" governing the number of significant figures in the answers to problems involving multiplication, division, addition, and subtraction, but, as Snedecor (86) has pointed out, they would have to be discarded when an involved series of operations must be performed. Following rigidly any single set of rules would involve "exaggerations of inaccuracies" (86, p. 87). The best single principle to follow is to carry along more figures in various computations than you intend to retain in the final answer, and then to round back to a reasonable number of places in reporting your answer. Let's consider first what we mean by a "reasonable" number of places in an answer before turning to the technique of "rounding."

b. Common practice in reporting statistics. An examination of the research literature in a given field will indicate current practice. In psychology, education, and the social sciences, since many or most of our measures are concerned with scores, usually measured in terms of whole numbers and seldom in terms of decimals or fractions, the following is common practice:

1. The mean is usually reported to one or two decimal places.
2. The median is usually reported to one or two decimal places.
3. The standard deviation is usually reported to one or two decimal places.
4. Standard errors, which we have not discussed as yet,

are usually reported to two and ordinarily not more than to three decimal places.

5. Correlation coefficients are usually reported to two and sometimes to three places.

6. Per cents, written as decimal fractions, are seldom reported to more than four places, and usually to two.

7. Proportions are usually reported to two or sometimes to three decimal places.

8. Ratios, used in tests of significance, which we shall take up later, are usually reported to two or sometimes to three decimal places.

When the number of observations with which we are dealing is very large, we might report the statistics listed above to another decimal place, but when the number of observations is small, say less than 100, such "professed accuracy" is apt to be looked upon as misleading. Remember that if you are going to report the mean of a sample to two decimal places, then you should carry the division $\Sigma X/N$ to three places and round back to two. This practice should be followed in computing all other statistics also: carry along two or three extra figures in making your computations and then round back in your final answer.

c. Rounding numbers. In rounding numbers to the nearest whole number, we would proceed as follows: 8.4 becomes 8; 7.1 becomes 7; 3.2 becomes 3; 7.6 becomes 8; 7.8 becomes 8; and 6.6 becomes 7. What is the rule we have followed? If the decimal fraction was less than .5, we dropped it and let the number stand; if the decimal fraction was over .5, we raised the number by one. If we round to one decimal we follow the same rule: 8.46 becomes 8.5; 7.32 becomes 7.3; 6.11 becomes 6.1; and 4.654 becomes 4.7.

Difficulties in rounding are apt to arise when we are asked to round numbers such as these: 5.5 and 4.5 to the nearest whole number; 8.550 and 4.650 to one decimal place. The answers may surprise you: 5.5 becomes 6; 4.5 remains 4;

8.550 becomes 8.6; and 4.65 remains 4.6. All of these numbers involve the dropping of a 5, which is right on the border line. The rule, by common practice, is this: if the number preceding the 5 which is to be dropped is an even number then we do not change it, but if the number preceding the 5 is odd, then it is raised by one. This is an arbitrary rule, to be sure, and it could just as well be the other way around. Either one would work and would tend to balance out errors that might be present in rounding if we had a long series to work with.

2. CODING BY SUBTRACTION

a. The sum of scores. We are now ready to consider some of the techniques of coding measurements. Consider the simple set of scores on a Thurstone attitude scale listed in Table 3. If you subtracted 5 from each of these scores

TABLE 3.—A SIMPLE SET OF SCORES
ON A THURSTONE ATTITUDE SCALE

11
8
5
2
4
7
1
2
5
9
7
1
4
5
4
Σ 75

and added the resulting deviations you would find that the sum would equal zero. What would this indicate? Do you recall a theorem introduced earlier, that the sum of deviations from the mean is equal to zero? This should

tell you, then, that the value we have subtracted from each of these scores is really the mean.

Now try subtracting 4 from each of the scores. The sum of the deviations is now no longer zero but 15. If you were to divide this value by N , which is equal to 15, the result would be 1, which is just the amount you need to add to the value 4, which you subtracted from each score, in order to obtain the mean. Try subtracting 3 from each score and you will now find that the sum of the deviations is equal to 30, and 30 divided by N gives 2, which is just the amount you need to add to 3, the value subtracted from each score, in order to obtain the mean.

As a matter of fact, any value at all could be subtracted from these scores and you could still find the mean by summing the deviations from the value subtracted. If you subtract a constant value from each score, then that value must be subtracted N times (once for each measurement in the series). Then N times the value subtracted, added to the sum of deviations from this value, will give the sum of scores (ΣX). For example, when 3 was subtracted from each score in the above series, we found the sum of deviations to be 30. And $(15)(3) + 30 = 75$, or the value that would be found by summing the original scores.

We are going to have to resort to some more symbols. The deviations we have just used may be symbolized by X' . This means that the deviation is not from the actual mean (M) of the series, but from some other point of arbitrary origin, symbolized by M' . Now for some algebraic manipulations, but manipulations which involve nothing more complicated than the application of the rules introduced in the second chapter. If the discussion is not clear, then you should go back and study the rules.

We can arrive at the equation $\Sigma X = (N)(M') + \Sigma X'$, empirically, that is, by actually working out several problems, subtracting different values, but we can also arrive at it in terms of our rules for equations and summation. Thus if we start with

$$\begin{aligned} X - M' &= X' \\ \Sigma(X - M') &= \Sigma X' \end{aligned}$$

By definition

Summating, and since the summation of a constant (M') is equal to N times the constant, we get

$$\Sigma X - (N)(M') = \Sigma X'$$

And adding $(N)(M')$ to both sides we get

$$\Sigma X = (N)(M') + \Sigma X'$$

Now having arrived at the above equation, we can readily see how it can be used to find the mean. If we divide both sides of the equation by N , we get $\frac{\Sigma X}{N} = M' + \frac{\Sigma X'}{N}$, and since $\frac{\Sigma X}{N}$ is equal to the mean, we may substitute in the left side to arrive at

$$M = M' + \frac{\Sigma X'}{N} \quad (11)^1$$

where M = the mean

M' = some constant which is subtracted from each score

$\Sigma X'$ = the sum of deviations from M'

N = the number of scores in the series

The value $\left(\frac{\Sigma X'}{N}\right)$ is called the *correction term for the mean* when deviations have been taken from some value other than the actual mean of the series.

b. The sum of squares. Perhaps you are wondering whether the X' values can be squared, summed, and then corrected in some fashion to arrive at the sum of squares (Σx^2). The answer is "Yes." All that we need to do to obtain the sum of squares is to subtract $\frac{(\Sigma X')^2}{N}$ from $\Sigma X'^2$.

¹ If M' were equal to the mean, then $\frac{\Sigma X'}{N}$ would, of course, be zero, since the "sum of deviations from the mean is equal to zero."

In other words,

$$\Sigma x^2 = \Sigma X'^2 - \frac{(\Sigma X')^2}{N} \quad (12)$$

where Σx^2 = the sum of squared deviations from the mean of the series

$\Sigma X'^2$ = the sum of squared deviations from some value (M') other than the mean

$(\Sigma X')^2$ = the sum of the deviations from M' , squared

N = the number of cases

The correction term $\frac{(\Sigma X')^2}{N}$ is not a correction term for subtraction as such, but for failure to take the deviations from the actual mean. Measures of variation such as the standard deviation and range are uninfluenced by subtraction or addition of a constant from every member in the series; the variation or spread of scores remains the same. For example, if the lowest score in a set was 20 and the highest was 40, the range would be 20. If a constant such as 10 was subtracted from every score in the series, the lowest score would become 10, the highest would become 30, and the range would remain 20. If 10 were added to each score, the lowest score would become 30, the highest 50, and the range would be the same as before. The standard deviation would also remain the same, regardless of the constant which is subtracted or added.

We may illustrate formula (12) with the series of Thurstone attitude scale scores we used before. Column (2) of Table 4 gives the deviations of each of the scores listed in column (1) from the actual mean. Working with these deviations from the mean, we can readily see from column (3) that the sum of squares is equal to 122. Now let us try the formula involving a correction term, when we work with deviations from some value other than the mean. In column (4) we have the values of X' when 4 has been subtracted, the sum of the deviations being 15. In column (5)

TABLE 4.—DEVIATIONS AND SQUARED DEVIATIONS
FROM VARIOUS POINTS OF ORIGIN

$$(X' = X - 4)$$

(1) X	(2) x	(3) x ²	(4) X'	(5) X' ²	(6) X ²
11	6	36	7	49	121
8	3	9	4	16	64
5	0	0	1	1	25
2	-3	9	-2	4	4
4	-1	1	0	0	16
7	2	4	3	9	49
1	-4	16	-3	9	1
2	-3	9	-2	4	4
5	0	0	1	1	25
9	4	16	5	25	81
7	2	4	3	9	49
1	-4	16	-3	9	1
4	-1	1	0	0	16
5	0	0	1	1	25
4	-1	1	0	0	16
Σ 75	0	122	15	137	497

we have these deviations squared, the sum being 137. Applying formula (12), we have

$$\begin{aligned}\Sigma x^2 &= \Sigma X'^2 - \frac{(\Sigma X')^2}{N} \\ &= 137 - \frac{(15)^2}{15} \\ &= 137 - 15 \\ &= 122\end{aligned}$$

which is precisely the value we obtained when we worked directly with the deviations from the actual mean of the distribution.

In column (6) we have squared the original scores. We may sum these, apply formula (12), and obtain the sum of squares as we did before. In this instance we are merely subtracting zero from each score, and our correction term

becomes the square of the sum of the original scores, divided by N . Thus we may write

$$\Sigma x^2 = \Sigma X^2 - \frac{(\Sigma X)^2}{N} \quad (13)$$

Substituting the appropriate values in the above formula, we get

$$\begin{aligned} \Sigma x^2 &= 497 - \frac{(75)^2}{15} \\ &= 497 - \frac{5,625}{15} \\ &= 497 - 375 \\ &= 122 \end{aligned}$$

We now have several different ways of finding the sum of squares: we may work with deviations from the actual mean; we may subtract some value other than the mean and apply a correction term to the resulting sum of squared deviations; or we may work with the measurements as they stand. This latter method is particularly valuable if you have a calculating machine to assist you in your computations.

3. CODING BY DIVISION

a. **The sum of scores.** We have just seen how we may subtract any constant from a series of scores, thus reducing the numerical size of the scores. We found also that we could work with these reduced or "coded" scores and, by applying a correction term, arrive at the same value for the *sum of scores* and the *sum of squares* that we would have obtained working with the original measures. We shall now see how division, too, can be used to reduce the size of scores.

In Table 5, column (1), we have a set of original measurements, the sum of which is 100. Since N is equal to 10, the mean of these scores is $100/10$ or 10. Column (2) gives the deviation of each score from the mean and the sum of this column is zero, as it should be. Column (3) gives the

deviations squared, and the sum of squares is equal to 96. In column (4) we have divided each X by 2 and we shall symbolize this "coded" score by x' . We shall let i represent the value by which we divided the scores. Column (5) contains the squares of the coded scores $\left(\frac{X}{2}\right)^2$.

TABLE 5.—CODING SCORES BY DIVISION

$$(x' = X/2)$$

(1) X	(2) x	(3) x^2	(4) x'	(5) x'^2
12	2	4	6	36
10	0	0	5	25
8	-2	4	4	16
10	0	0	5	25
14	4	16	7	49
6	-4	16	3	9
8	-2	4	4	16
16	6	36	8	64
6	-4	16	3	9
10	0	0	5	25
Σ 100	0	96	50	274

Note that the $\Sigma x'$ needs to be multiplied by 2 (the value by which each X was divided) in order to equal the sum of X . In other words, $\Sigma X = (\Sigma x')(i)$. If we divide both sides of this equation by N , we get $\frac{\Sigma X}{N} = \left(\frac{\Sigma x'}{N}\right)(i)$, and since $\frac{\Sigma X}{N}$ is equal to the mean, we may substitute in the left-hand side of the equation to get another basic formula:

$$M = \left(\frac{\Sigma x'}{N}\right)(i) \quad (14)$$

Thus we see that if we have reduced scores by dividing each one by the same constant, we may sum these coded scores,

divide by N , and multiply the result by i , the value by which we divided each score, to arrive at the mean. Substituting the appropriate numerical values in formula (14), we get

$$\begin{aligned} M &= \left(\frac{50}{10}\right) (2) \\ &= (5) (2) \\ &= 10 \end{aligned}$$

which is the value we obtained by working with the original measures.

b. The sum of squares. The formula for the sum of squares now requires a *correction term for coding* as well as one for failure to take the deviations from the mean of the series. Measures of variation, although uninfluenced by subtraction or addition, are changed by multiplication or division. Note, for example, that the range of the scores in column (4) is no longer the same as that of the original measurements in column (1). The formula we need is

$$\Sigma x^2 = \left[\Sigma x'^2 - \frac{(\Sigma x')^2}{N} \right] i^2 \quad (15)$$

where Σx^2 = the sum of squares

$\Sigma x'^2$ = the coded sum of squares from some point other than the mean

$\frac{(\Sigma x')^2}{N}$ = the correction term for origin

i^2 = the correction term for coding, that is division

Substituting in formula (15), we get

$$\begin{aligned} \Sigma x^2 &= \left[274 - \frac{(50)^2}{10} \right] (2)^2 \\ &= \left[274 - \frac{2,500}{10} \right] 4 \\ &= (274 - 250) 4 \\ &= (24) (4) \\ &= 96 \end{aligned}$$

which is precisely the value we obtained when we worked with deviations from the mean of the distribution.

You may not quite grasp, at this time, the value of the coding techniques we have described. That is because the problems and data we have had to work with up to now have been selected for simplicity and ease of computation. In each illustration the mean has been a whole number and the figures have been small rather than large. But suppose that the mean for a distribution of over 100 scores turned out to be 152.67. If you tried to compute the standard deviation by working with deviations from this mean, the computations would involve squaring four- or five-place figures. Coding the series by subtracting some even value and reducing them even more by dividing by a constant would simplify your computations.

4. SUMMARY OF "CODING FORMULAS"

We might summarize the discussion so far before we turn to coding measures by grouping. It is possible, we have seen, to code measurements by subtraction and division, if we remember to return our coded measurements to the units of our original scores for our final answers.²

1. When scores have been reduced by subtraction only ($X - M'$) then

$$M = M' + \frac{\Sigma X'}{N} \quad \text{and} \quad \Sigma x^2 = \Sigma X'^2 - \frac{(\Sigma X')^2}{N}$$

² It is also possible to code by multiplication and addition, but we seldom have need for these coding techniques in handling the data of the social sciences. The rules are these: The mean is influenced by every operation; the standard deviation only by multiplication and division. When more than one operation has been performed, for example, subtraction and then division, the coded results must be decoded with the inverse operation (the inverse operation of subtraction is addition, of division it is multiplication) and in reverse order. If we have subtracted 5 and then divided each measure by 2, we must decode the resulting mean by first multiplying by 2 and then adding 5. The sum of squares, being influenced only by the one operation, division, must be multiplied by the square of the value by which each measure was divided.

2. When scores have been reduced by means of division only $\left(\frac{X}{i}\right)$ then

$$M = \left(\frac{\Sigma x'}{N}\right) i \quad \text{and} \quad \Sigma x^2 = \left[\Sigma x'^2 - \frac{(\Sigma x')^2}{N}\right] i^2$$

3. When scores have been reduced first by subtraction and then by division, then

$$M = M' + \left(\frac{\Sigma x'}{N}\right) i \quad \text{and} \quad \Sigma x^2 = \left[\Sigma x'^2 - \frac{(\Sigma x')^2}{N}\right] i^2$$

4. When scores are treated in terms of the original measurements, then

$$M = \frac{\Sigma X}{N} \quad \text{and} \quad \Sigma x^2 = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$$

The formulas given above are basic. Memorize them and make sure that you know what every term means and what every term does.

5. GROUPING MEASURES INTO CLASSES

The most common method of coding scores is by "grouping" data or measurements into "classes" to form a *frequency distribution*. You may recall that earlier in this chapter we discussed "precision of measurement." Grouping may be thought of as the equivalent of using a less precise measuring unit and is most valuable when we have a large number of measurements. Instead of treating each measurement separately, we group them into a number of equal intervals, classes, or steps. We then code the classes and work with these coded values in order to simplify our computations.

Examine the scores in Table 6. They are hypothetical but we shall assume that they were made by a class in psychology on an objective examination. These scores, as they stand, do not give a very concise description of the

performance of the group—and one of the purposes of statistics is to summarize and describe. Nor are these scores, as they stand, very convenient to use in computation.

a. **The number of intervals.** The first thing we need to do in making a frequency distribution is to determine how we shall group the scores. We could group them in terms

TABLE 6.—HYPOTHETICAL SCORES MADE BY STUDENTS
ON AN OBJECTIVE TYPE OF EXAMINATION

87	76	73	70	67	66	64	63	61	60
85	75	72	69	67	65	64	62	61	60
82	74	71	69	67	65	63	62	61	60
78	74	71	68	66	65	63	62	61	60
77	74	70	68	66	64	63	62	61	60
60	59	58	57	56	54	52	50	46	43
60	59	58	57	55	54	52	49	46	42
60	59	58	57	55	53	51	49	46	38
60	59	58	56	55	53	51	48	45	35
60	59	57	56	54	53	50	47	44	33

of a *class interval* of 1 by placing numbers ranging from 87 to 33 at the left-hand side and then making a tally mark (/) each time one of these numbers was found in the distribution. This, however, would still leave the scores spread out; the class range would be from 87 to 33 or 54. Fortunately, experience has shown that quite accurate results can be obtained in statistics when, for purposes of computation, we work with a much smaller class range, say, from 10 to 20 classes. Our first rule for grouping scores, then, will be that we shall group them so as to have from 10 to 20 classes or groups. The larger the number of intervals, the more precise will be the computations, but also the more complicated the computations. Consequently, the number of class intervals we decide to work with will be dictated by our desire for accuracy and also by our desire for convenience.

b. **Size of the class interval.** One method which might be used to determine the appropriate size of the class interval

to use in grouping scores is first to find the range and then to divide this value by the contemplated number of class intervals with which you wish to work. In the present problem, if we wished to work with the minimum number, 10, we would divide the range, 54, by 10 and the quotient would be 5.4. This quotient rounded off to the nearest integer would be 5, which suggests the size of the interval to use. We shall let the symbol i represent the value of the class interval.

If we wished to work with approximately 15 class intervals we would divide the range by 15 and the quotient would be 3.6, which, when rounded, is 4. In this instance, however, instead of using 4 for the size of the class interval, we might prefer to use 3. The reason for this is that, in general, an odd number for the size of the class interval is easier to work with than an even number. We shall later make the assumption that the midpoint of the class interval best represents all of the scores located within the interval. If the interval has an odd number of units, then the midpoint will be an integer. If the interval has an even number of units, then the midpoint will be a fraction. It is to avoid the latter situation that intervals with an odd number of units are to be preferred.³ However, if the quotient obtained by dividing the range by the contemplated number of intervals is 9 or 11, then most workers would probably select 10 for the size of the interval. This is partly because classifying things by 10's is common in everyday life and also because computations involving 10's are particularly easy.

c. **Limits of the intervals.** It is customary in psychology and education to start class intervals so that the lowest score of the interval is some multiple of the size of the class interval. For example, when the size of the interval is 3,

³This rule, like most rules based upon common practice, has exceptions. If we had a range of 25, we would be forced to decide between working with the measures as they stand, or using an i of 3, which would give us less than the 10 intervals we desire, or using an i of 2. You would probably find advocates of each procedure.

intervals are started with some multiple of 3 such as 6, 9, 12, or 15, and so forth.⁴ Although it is customary to write these intervals as 6-8, 9-11, 12-14, and so forth, for a class interval of 3; and 10-14, 15-19, 20-24, and so forth, for a class interval of 5; we must remember what we have previously said about the meaning of a score, i.e., that it represents a range extending .5 of a unit above and below the recorded value. The same reasoning applies to class intervals; the *theoretical limits* of the interval 10-14 are 9.5-14.5, that is, .5 of a unit below and .5 of a unit above the *recorded limits*.

d. Tallying the scores. The next step in making a frequency distribution, after the size of the class interval has been determined, is to tally the scores. The various class intervals are listed as in Table 7 according to the accepted practice of placing the highest interval at the top. As the scores are taken one at a time, a tally mark is placed opposite the interval in which each score falls. When four tally marks (////) have been made in a given interval, the fifth is made as a cross tally, thus $\diagup \diagdown$. The sum of the tally marks for each interval gives the *frequency* of the interval, and the sum of all of the frequencies gives the total N .

e. Assumptions concerning grouped scores. What assumptions can we make concerning these scores as they are now grouped? We might assume that *the scores within each interval are evenly distributed throughout that interval*. This is the assumption we shall have to make in order to find the median which is a point. A second assumption we might make, and one that we use in computing the mean and standard deviation, is that *the best single value to represent all of the scores within a given interval is the midpoint of that interval*. This, of course, will not always be true, but the errors introduced tend to be small and the errors in one direction tend to counterbalance errors in the other direction.

⁴ This again is an arbitrary practice, and there is much to be said for starting intervals in such a way that the *midpoint* of the interval is an even multiple of the size of the interval. When 5, for example, is the size of the interval, the intervals may be started with 8, 13, 18, or 23, and so forth.

We shall find that the mean and standard deviation based upon this assumption will not be seriously in error.

We could, if we wished, now compute the mean by locating the midpoint of each class interval and multiplying this

TABLE 7.—FREQUENCY DISTRIBUTION OF SCORES GIVEN IN TABLE 6

(1) SCORES	(2) TALLY MARKS	(3) <i>f</i>
85-89	//	2
80-84		1
75-79		4
70-74		9
65-69		13
60-64		26
55-59		19
50-54		12
45-49		8
40-44		3
35-39	//	2
30-34		1

value by *f*, the number of scores within the interval; we could sum these values and divide by *N* and this would give us the mean. Locating the midpoint of an interval is an easy process. The midpoint of the interval is halfway between the lower limit and the upper limit of the interval. The lower limit of the interval 30-34 is 29.5 and the upper limit is 34.5, a range of 5. Half of 5 is 2.5 and this value added to the lower limit of the interval gives the midpoint, 32. The midpoint of any class interval, in other words, is the lower limit of the interval plus $i/2$. It is important not to forget that the lower limit of any interval extends .5 of a unit below the recorded value and the upper limit, .5 of a unit above.

f. **Calculation of the mean and sum of squares.** You may wonder why we are not now going ahead to find the mean of the distribution of scores in Table 8, by multiplying the midpoints of the class interval as given in column (2) by the corresponding frequencies as listed in column (4). We could calculate the mean in this manner, but we may simplify our computations even more by coding the values of the midpoint. To do this we (1) subtract the midpoint of the lowest interval from the midpoint of all of the other intervals and then (2) divide these values by i , the size of the interval. This will not change the values of the mean and standard deviation, except for the slight errors already introduced by grouping, if we take into consideration the proper corrections for origin and coding. The formulas are similar to those we used before except that we now use fx' to indicate that each coded score or midpoint has been multiplied by its frequency. Thus

$$M = M' + \left(\frac{\sum fx'}{N}\right) i \quad (16)$$

$$\sum x^2 = \left[\sum fx'^2 - \frac{(\sum fx')^2}{N} \right] i^2 \quad (17)$$

The essential steps in the application of formula (16) and formula (17) are illustrated in Table 8. You may note several things from this table. It would not have been necessary to find the midpoint of the lowest interval and then to subtract this value from the midpoint of every other interval, dividing the remainders by i . We could have simply coded the lowest interval 0, the next 1, the next 2, and so on. This will be true of all distributions, regardless of the size of the class interval or the number of intervals or any other considerations. We might also have subtracted the midpoint of some interval other than the lowest. We could have started, for example, with the midpoint of some interval toward the center of the distribution and numbered this 0. Then intervals above this would be numbered 1, 2, 3, and so on, and those below would be numbered -1, -2,

TABLE 8.—CALCULATION OF THE MEAN, MEDIAN, AND STANDARD DEVIATION FROM SCORES CODED BY GROUPING

(1) SCORES	(2) MID- POINT	(3) MID- POINT - 32	(4) <i>f</i>	(5) <i>x'</i>	(6) <i>fx'</i>	(7) <i>fx'²</i>
85-89	87	55	2	11	22	242
80-84	82	50	1	10	10	100
75-79	77	45	4	9	36	324
70-74	72	40	9	8	72	576
65-69	67	35	13	7	91	637
60-64	62	30	26	6	156	936
55-59	57	25	19	5	95	475
50-54	52	20	12	4	48	192
45-49	47	15	8	3	24	72
40-44	42	10	3	2	6	12
35-39	37	5	2	1	2	2
30-34	32	0	1	0	0	0
Σ			100		562	3,568

$$\begin{aligned}
 M &= M' + \left(\frac{\Sigma fx'}{N}\right) i \\
 &= 32 + \left(\frac{562}{100}\right) 5 \\
 &= 32 + (5.62)(5) \\
 &= 32 + 28.1 \\
 &= 60.1 \\
 \\
 Mdn &= l + \left(\frac{\frac{N}{2} - \Sigma f_c}{f_s}\right) i \\
 &= 59.5 + \left(\frac{50 - 45}{26}\right) 5 \\
 &= 59.5 + \left(\frac{5}{26}\right) 5 \\
 &= 59.5 + .96 \\
 &= 60.46
 \end{aligned}$$

$$\begin{aligned}
 \Sigma x^2 &= \left[\Sigma fx'^2 - \frac{(\Sigma fx')^2}{N}\right] i^2 \\
 &= \left[3,568 - \frac{(562)^2}{100}\right] 25 \\
 &= \left[3,568 - \frac{(315,844)}{100}\right] 25 \\
 &= (3,568 - 3,158.44) 25 \\
 &= (409.56)(25) \\
 &= 10,239 \\
 \\
 \sigma &= \frac{\Sigma x^2}{N} \\
 &= \frac{10,239}{100} \\
 &= 102.39 \\
 \\
 \sigma &= \sqrt{\sigma} \text{ or } \sqrt{\frac{\Sigma x^2}{N}} \\
 &= \sqrt{102.39} \\
 &= 10.1
 \end{aligned}$$

-3, and so on. This would give us slightly smaller figures to deal with, but would have introduced some negative values into our computations. As general practice, it is better to start the lowest interval with 0 and number up from there. There is perhaps less chance of making a mistake if you follow this practice.

g. The "Charlier checks." There are checks on the accuracy of your computations. They are known as the "Charlier checks" and in the present problem may be effected by raising each coded interval by one point. The 0 interval in the frequency distribution becomes 1, the 1 interval becomes 2, and so on. We may designate these new coded values for the class intervals as x'' . Now find the $\Sigma fx''$ and the $\Sigma fx''^2$ as before. If the computations in the first and second instance have both been correctly made, then the following relations will hold:

$$\Sigma fx'' = \Sigma fx' + N \quad (18)$$

$$\Sigma fx''^2 = \Sigma fx'^2 + (2)(\Sigma fx') + N \quad (19)$$

As an illustration of these checks we may examine the computations in Table 9.

TABLE 9.—ILLUSTRATION OF THE "CHARLIER CHECKS"

(1) SCORES	(2) f	(3) x'	(4) fx'	(5) fx'^2	(6) f	(7) x''	(8) fx''	(9) fx''^2
30-32	1	5	5	25	1	6	6	36
27-29	2	4	8	32	2	5	10	50
24-26	5	3	15	45	5	4	20	80
21-23	7	2	14	28	7	3	21	63
18-20	3	1	3	3	3	2	6	12
15-17	2	0	0	0	2	1	2	2
Σ	20		45	133			65	243

$$\begin{aligned} \Sigma fx'' &= \Sigma fx' + N \\ 65 &= 45 + 20 \\ 65 &= 65 \end{aligned}$$

$$\begin{aligned} \Sigma fx''^2 &= \Sigma fx'^2 + (2)(\Sigma fx') + N \\ 243 &= 133 + (2)(45) + 20 \\ 243 &= 133 + 90 + 20 \\ 243 &= 243 \end{aligned}$$

h. Calculation of the median. The median, Q_1 , Q_3 , and centiles can also be found from a frequency distribution. Formula (7) given earlier will work without any change. But if we have our scores grouped in intervals greater than one, as will usually be the case, the value within the parentheses must be multiplied by the size of the interval, i . Thus

$$Mdn = l + \left(\frac{\frac{N}{2} - \Sigma f_o}{f_w} \right) i$$

where Mdn = the median

l = the lower limit of the interval containing the median

Σf_o = the sum of frequencies up to the interval containing the median

f_w = the frequency within the interval containing the median

N = the total number of cases in the distribution

i = the size of the class interval

The application of formula (7) is illustrated in Table 8. The value of the median obtained by means of this formula may be checked by calculating the median using formula (8).

6. SUMMARY OF STEPS IN CODING

Here is a summary of the steps in coding measurements by first grouping them in a frequency distribution.

1. Determine the range: $(H - L)$.
2. Divide the range by the number of intervals you wish to work with (10 to 20). This figure gives the approximate size of the class interval i .
3. Begin the lowest interval with some multiple of the size of the interval.
4. Code the lowest interval 0, the next 1, the next 2, and so forth until the highest interval has been coded.
5. Apply formula (16) for the mean and formula (17) for the sum of squares.

If you are working with a calculating machine, you may not want to record the scores in a frequency distribution but may still wish to code them. This is easily accomplished. Follow the procedure above through the third step. Then take the lower limit (recorded limit) of the first interval and divide this by i (the size of the interval). This will be a whole number since the lower recorded limit is a multiple of i , and may be designated as k . Now divide each measurement by i , discarding any remainder. Subtract the value k and this will give you the coded value of the score, which is identical with the value you would have obtained if you had grouped the scores into a frequency distribution.

Suppose we take a few of the values from Table 6 to illustrate how they might have been coded without making the frequency distribution. The lower limit (recorded) of the first interval is 30, and this divided by the size of the interval, 5, gives the value of k , which is 6. Check the coded values listed in Table 10 against those of Table 8.

TABLE 10.—CODING SCORES WITHOUT
GROUPING INTO INTERVALS

($i = 5$; $k = 6$)

X	$\frac{X}{i}$	$(\frac{X}{i}) - k$
33	6	0
35	7	1
38	7	1
42	8	2
43	8	2
50	10	4
68	13	7
76	15	9
85	17	11

SIMPLIFYING COMPUTATIONS

Example 1.—Here is an easy practice series for coding.

29 28 27 25 24 22 20

- (a) Find the mean and sum of squares by the usual method.
- (b) Subtract 22 from each score and find the mean and sum of squares.
- (c) Find the sum of squares assuming M' to be zero.

Example 2.—Given the following highest and lowest scores in a number of different distributions:

These

(1) 87-22 ~~20~~ (2) 95-50 (3) 185-~~82~~ 60 (4) 66-~~42~~ 50 (5) 52-18 0 (6) 110-~~54~~ 14

- (a) Find the range.
- (b) Find the value of i .
- (c) Find the limits of the lowest interval.
- (d) Find the midpoint of the lowest interval.

Example 3.—Code the following scores without making a frequency distribution. The range is from 84 to 33.

(a)	(b)	(c)	(d)	(e)	(f)	(g)
80	45	53	39	42	56	43
33	62	64	66	47	83	58
84	68	59	55	35	38	60
76	71	37	84	45	61	77
42	36	78	72	41	59	52

Example 4.—Demonstrate the “Charlier check” with the following frequency distribution. Check both (a) the sum of scores and (b) the sum of squares.

SCORES	f	x'
38-31	1	
25	1	5
37-21	2	4
24	2	4
24-23	5	3
22-22	7	2
20-20	7	2
21	3	1
18-19	2	0
20	2	0
16-17	2	0

Example 5.—By making a frequency distribution, code the following "error" scores made by a class in general psychology on an objective examination: Let $i = 3$. Find the mean and standard deviation of the distribution. Check your computations by means of the "Charlier checks."

44	40	35	34	32	31	30	29	27
43	40	35	34	31	31	30	29	27
42	37	35	33	31	30	29	29	27
40	36	34	33	31	30	29	28	26
40	35	34	32	31	30	29	28	26
26	25	24	24	23	23	22	22	22
26	25	24	23	23	23	22	22	22
26	25	24	23	23	23	22	22	22
25	25	24	23	23	23	22	22	22
25	25	24	23	23	22	22	22	22
22	21	20	20	20	19	18	18	18
22	21	20	20	19	18	18	18	17
21	21	20	20	19	18	18	18	17
21	21	20	20	19	18	18	18	17
21	20	20	20	19	18	18	18	17
17	17	16	15	14	14	13	12	9
17	17	16	15	14	14	13	12	9
17	16	16	15	14	14	13	12	9
17	16	16	15	14	14	13	11	8
17	16	15	15	14	13	12	11	7

Example 6.—Find the mean, median, and standard deviation of the following distribution:

SCORES	f	x'
60-62	1	8
57-59	3	7
54-56	2	6
51-53	7	5
48-50	11	4
45-47	10	3
42-44	9	2
39-41	5	1
36-38	5	0

Example 7.—Show, algebraically, that $\Sigma x^2 = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$.

Example 8.—Show, algebraically, that (a) $\Sigma fx'' = \Sigma fx' + N$. and that (b) $\Sigma fx''^2 = \Sigma fx'^2 + (2)(\Sigma fx') + N$.

CHAPTER 5

THE PRODUCT-MOMENT CORRELATION COEFFICIENT

The statistical techniques discussed so far are useful for describing single variables. We are now ready to consider statistical techniques which will permit us to study two variables and to describe the relationships between them. The problem in which we are interested is the extent to which two variables are associated. As values of one increase from small to large can we expect corresponding changes in the second? Is it true, for example, that tall men tend to marry tall women and short men tend to marry short women? If we studied this problem and found that the statements were true, then we would say that these variables, height of husband and height of wife, are *positively* related.

There are occasions, however, when we observe not positive but negative relationships. Such a relationship has been reported between an index of economic prosperity and the number of lynchings occurring during a given year (9). A negative relationship between these two variables means that as values of one increased, values of the other tended to decrease; the higher the index of economic prosperity, the fewer the lynchings; the lower the index of economic prosperity, the greater the number of lynchings. We describe this relationship by saying that the two variables are *negatively* related.

We must take care in studying relationships that we do not confuse the concepts of "correlation" and "causation." When two things are related it does not necessarily follow that one is the cause of the other. We might find that there is a positive relationship between scores on a test of aggressiveness and yearly incomes for a sample of several hundred men, but we cannot legitimately say that one is the

cause of the other. The assumption that changes in one variable are the cause of changes in the second may or may not be valid, but this must be determined by considerations other than the mere fact that the two variables are related. The changes in each variable, for example, might possibly be the common result of some third variable. This is perhaps the most probable explanation of the positive relationship that has been found between scores on an academic aptitude test and subsequent grades in college. No one would assume that the scores on the test determine or cause the student to get good grades, or that the good grades cause the student to get a high score. Whatever it is that causes the student to get a high score may also be the cause of his better than average grades.

1. THE COEFFICIENT OF CORRELATION

One of the statistical techniques for describing relationships, both positive and negative, is the *product-moment correlation coefficient*. This coefficient measures the degree to which two variables are associated and is symbolized by r . In terms of absolute size, r may vary from $+1.00$, through zero, to -1.00 . A correlation coefficient of $+1.00$ indicates a perfect positive relationship between two variables; a zero coefficient indicates no relationship; and -1.00 indicates a perfect negative relationship. It is very seldom, if at all, that perfect relationships are found in the behavioral sciences, in part because of the limitations of our measuring instruments and also because of the difficulties of controlling all possible factors which may influence the two variables being studied. But for purposes of illustration and understanding let us see what a perfect relationship would mean.

a. A perfect positive correlation. If we are going to study relationships between variables, then we must take measurements in pairs and we must have more than one pair of measurements. We might take an individual's score on one test and pair it with his score on a second test, and

continue to do this for a number of different individuals. Or we might study the relationship between scores of brothers and sisters on the same test, pairing the score of each boy with that of his sister. If we were interested in studying a problem in motivation or drive in the psychological laboratory we might take the number of hours that an animal has been without food and pair this value with the number of crossings it will make over an electrically charged grid in order to reach food. Suppose, for reasons of simplicity, we assume that we have 10 individuals and that each one has taken two psychological tests. We shall call one of the tests *X* and the other *Y*. The paired scores for each individual are given in Table 11.

Observe that each individual's score on the *Y* test is exactly 1 point higher than his score on the *X* test. There are no exceptions and we would find that the correlation coefficient is positive and perfect and expressed by $+1.00$. Perhaps the best way to visualize what this means is in terms of a simple graph.

Figure 2 indicates the customary method of plotting a graph for two variables. The base line is called the *X-axis* or *axis of abscissas*. The left vertical line is called the *Y-axis* or *axis of ordinates*. The point at which the two

lines intersect is called the *point of origin* or *O*. The scale on the *X-axis* always runs from left to right and on the *Y-axis* from the bottom up, as shown in the diagram. A vertical distance from the *X-axis* to a given point is called the *ordinate* of that point and a horizontal distance from the *Y-axis* to a given point is called the *abscissa* of that point. Thus to plot a point for a given value of *X* and *Y*,

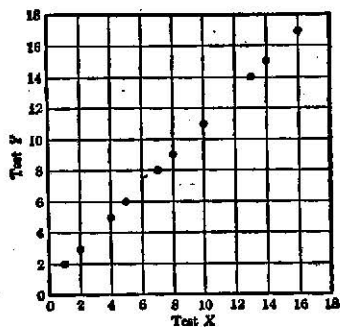


FIG. 2.—Correlation chart for scores on Test *X* and Test *Y* given in Table 11. Correlation coefficient equal to 1.00.

we would move out the X -axis away from O until we come to that value of X , and then move up the Y -axis until we come to the value for Y . The point would be plotted at the intersection of the horizontal line drawn from the value of

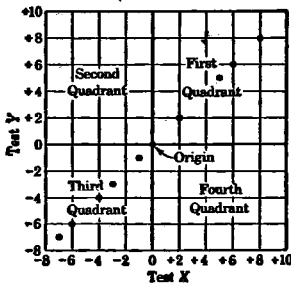
TABLE 11.—SCORES ON TESTS X AND Y FOR 10 INDIVIDUALS $(r = 1.00)$

TEST	INDIVIDUALS									
	1	2	3	4	5	6	7	8	9	10
X	16	14	13	10	8	7	5	4	2	1
Y	17	15	14	11	9	8	6	5	3	2

Y and the vertical line drawn from the value of X . Together the ordinate and abscissa of the point would be called the *coordinates* of the point.¹

In plotting the test scores of Table 11, values of the X test are plotted along the abscissa and along the ordinate values of the Y test are plotted. This is similar to plotting or making a frequency distribution with a single set of test scores, except that here we move the Y scores out to the right so that they will be tallied (plotted) also in terms of the corresponding X values.

¹ It is sometimes necessary to deal with negative values of X and Y , as might be the case if we were plotting the points in terms of deviation scores. In this instance we would take the origin of the X and Y axes at the means of X and Y . The two axes would be extended as shown. The upper right-hand section takes care of positive values of both X and Y , the upper left-hand section takes care of negative values of X and positive values of Y , the lower left-hand section takes care of negative values of both X and Y , and the lower right-hand section takes care of positive values of X and negative values of Y . These sections are called the first, second, third, and fourth *quadrants*, respectively.



We can readily see from Figure 2 that the plotted points fall along a straight line and that, therefore, the relationship between Test X and Test Y is *rectilinear*. The line which might be drawn through the plotted points to represent their trend is called a *regression line* and is described more fully on pages 262-264. When the relationship between the two variables is not perfect, then we have two regression lines, one showing the correlation or regression of Y on X and the other showing the correlation or regression of X on Y . The two regression lines always cross at the point which would be designated by M_x and M_y in the graph. The larger the absolute value of r , the closer together the two lines tend to lie. When r equals 1.00 the lines coincide. When r equals zero, on the other hand, the two regression lines are perpendicular to each other. The correlation coefficient, r , however, has the same value, regardless of whether we are speaking of the regression of X on Y or the regression of Y on X . That is to say, r_{xy} equals r_{yx} .

In some instances, the trend of the plotted points may be more accurately described by a curved regression line, that is, the relationship between X and Y may be *curvilinear* rather than *rectilinear*. To describe the degree of association between two variables related in this fashion, we shall need to compute another correlation coefficient, known as the *correlation ratio*. All that we have said concerning the product-moment correlation coefficient and all that follows assumes that the relationship between the variables under consideration can best be described in terms of a straight line, i.e., that changes in one variable are accompanied by a uniform change in the second.²

b. A perfect negative correlation. Suppose that our 10 individuals had made the scores given in Table 12. We shall let their X scores remain the same, but we shall change

² In a later section, where X and Y are grouped into classes in order to compute the correlation coefficient, it is assumed that the *average* value of Y , from class to class, changes uniformly with changes in X , and vice versa. For any given class the individual values may, of course, vary about the class mean.

TABLE 12.—SCORES ON TESTS X AND Y FOR 10 INDIVIDUALS
($r = -1.00$)

TEST	INDIVIDUALS									
	1	2	3	4	5	6	7	8	9	10
X	16	14	13	10	8	7	5	4	2	1
Y	2	4	5	8	10	11	13	14	16	17

the value of the Y scores. If we now plot these scores as we have done in Figure 3 we may note that the plotted values again fall exactly on a straight line but that the slant or direction of the line is different. High values of Y now tend to be associated with low values of X. In this instance

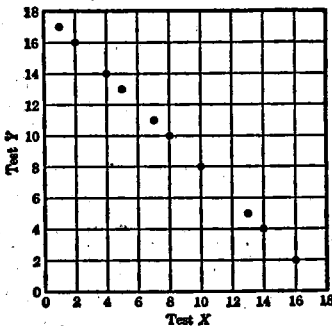


FIG. 3.—Correlation chart illustrating correlation between scores on Test X and Test Y given in Table 12. Correlation coefficient equal to -1.00 .

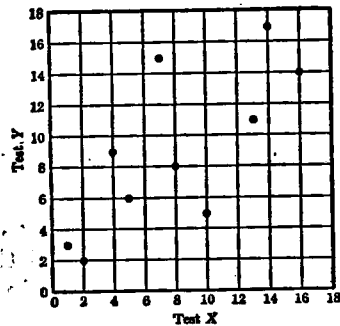


FIG. 4.—Correlation chart for scores on Test X and Test Y given in Table 13. Correlation coefficient equal to $.74$.

the correlation coefficient is -1.00 and we say that the relationship is negative.

c. **A high positive correlation.** Consider another illustration with the same X scores as before for our 10 individuals

but with the values of *Y*, as given in Table 13. Observe now that although there is a tendency for large values of *X*

TABLE 13.—SCORES ON TESTS *X* AND *Y* FOR 10 INDIVIDUALS
($r = .74$)

TEST	INDIVIDUALS									
	1	2	3	4	5	6	7	8	9	10
<i>X</i>	16	14	13	10	8	7	5	4	2	1
<i>Y</i>	14	17	11	5	8	15	6	9	2	3

and *Y* to be associated, the plotted points in Figure 4 deviate somewhat from any straight line that might be drawn among them to describe the trend. The correlation coefficient between *X* and *Y* is now .74.

d. A high negative correlation. A coefficient similar to that obtained from the data in Table 13 but with a negative relationship between *X* and *Y* would be indicated by the set of scores in Table 14. The scatter of the plotted points

TABLE 14.—SCORES ON TESTS *X* AND *Y* FOR 10 INDIVIDUALS
($r = -.73$)

TEST	INDIVIDUALS									
	1	2	3	4	5	6	7	8	9	10
<i>X</i>	16	14	13	10	8	7	5	4	2	1
<i>Y</i>	5	3	2	8	9	17	6	15	11	14

in Figure 5 is about the same but the trend is different. The correlation coefficient is $-.73$.

e. A low correlation. Perhaps you are wondering what a set of scores yielding a very small correlation coefficient would look like when plotted. The set of scores in Table 15

TABLE 15.—SCORES ON TESTS X AND Y FOR 10 INDIVIDUALS
($r = -.12$)

TEST	INDIVIDUALS									
	1	2	3	4	5	6	7	8	9	10
X	16	14	13	10	8	7	5	4	2	1
Y	6	3	17	11	15	2	5	8	9	14

yields a coefficient of $-.12$ and you may observe from Figure 6 that the plotted points tend to scatter all over.

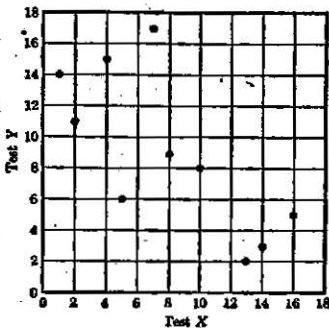


FIG. 5.—Correlation chart for scores on Test X and Test Y given in Table 14. Correlation coefficient equal to $-.73$.

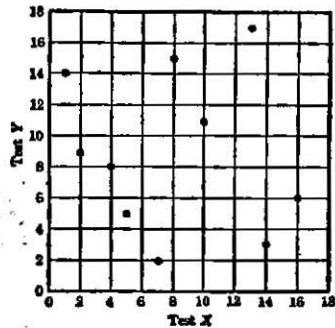


FIG. 6.—Correlation chart for scores on Test X and Test Y given in Table 15. Correlation coefficient equal to $-.12$.

There is, in other words, no very apparent tendency for values of X to increase or decrease in any consistent fashion with increases or decreases in Y.

2. BASIC FORMULAS FOR r

a. Standard deviations. The formula for the coefficient of correlation is

$$r_{xy} = \frac{\Sigma xy}{N\sigma_x\sigma_y} \quad (20)$$

where r_{xy} = the correlation coefficient of X on Y ³

Σxy = the sum of the cross-products or $(x)(y)$

σ_x = the standard deviation of the X variable

σ_y = the standard deviation of the Y variable

N = the number of pairs of measurements

The only new term involved in formula (20) is the Σxy . This term, when divided by N , is known as the *covariance* and is similar to the variance $\left(\frac{\Sigma x^2}{N}\right)$ of individual sets of measurements. The deviation x , as we already know, is $X - M_x$ and y is $Y - M_y$. The Σxy is the sum of the products of all of the paired deviations.

We can now see what makes the correlation negative or positive in sign. If an individual is above the mean on both the X and Y variable, then xy is positive in sign and this will be true of all such cases. On the other hand, if an individual is below the mean on both variables, then xy will still be positive in sign since $(-x)(-y)$ gives a positive product. The sum of xy is positive in sign and at a maximum when the largest value of x is paired with the largest value of y , and other values of x and y are paired accordingly.

A negative relationship between two variables means that values below the mean of one variable tend to be associated with values above the mean of the second variable. The x and y deviations will thus tend to be associated in the manner $(x)(-y)$ and $(-x)(y)$, and consequently the products will be negative in sign. The sum of xy will be at

³ The correlation coefficient of Y on X would take the same formula. The two are identical, as pointed out before.

its maximum negative value when the individual farthest above the mean on the X variable is also the farthest below the mean on the Y variable, and other x and y deviations are paired accordingly.

By having the values σ_x and σ_y in the denominator of formula (20), the deviations of x and y are reduced to a common basis. The values of $\frac{x}{\sigma_x}$ and $\frac{y}{\sigma_y}$ are *standard scores*, which we discussed earlier, and are comparable regardless of the unit of measurement involved in X and Y . Thus

$$\left(\frac{x}{\sigma_x}\right)\left(\frac{y}{\sigma_y}\right) = \frac{xy}{\sigma_x\sigma_y} \quad \text{and summing for the series and dividing by } N \text{ to get an average we arrive at}$$

$$\frac{\Sigma\left(\frac{x}{\sigma_x}\right)\left(\frac{y}{\sigma_y}\right)}{N} = \frac{\Sigma xy}{N\sigma_x\sigma_y} \quad \text{which is the formula for } r.$$

The coefficient of correlation, in other words, is the mean of the products of paired standard scores.

The steps involved in the calculation of the correlation coefficient by means of formula (20) are illustrated in Table 16. The scores are the same as those presented earlier in Table 13.

b. Sum of squares method. If we are not interested in the standard deviations of the X and Y variables as such, but have as our main objective the determination of the correlation coefficient, then formula (20) may be simplified so that

$$r_{xy} = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}} \quad (21)$$

where r_{xy} = the correlation coefficient of X on Y

Σxy = the sum of the cross-products

Σx^2 = the sum of squares for X

Σy^2 = the sum of squares for Y

The coefficient obtained by formula (21) would be identical with that obtained by formula (20). You may observe

TABLE 16.—CALCULATION OF THE CORRELATION COEFFICIENT FOR THE DATA OF TABLE 13

(1)	(2)	(3)	(4)	(5)	(6)	(7)
X	Y	x	y	x^2	y^2	xy
16	14	8	5	64	25	40
14	17	6	8	36	64	48
13	11	5	2	25	4	10
10	5	2	-4	4	16	-8
8	8	0	-1	0	1	0
7	15	-1	6	1	36	-6
5	6	-3	-3	9	9	9
4	9	-4	0	16	0	0
2	2	-6	-7	36	49	42
1	3	-7	-6	49	36	42
Σ 80	90	0	0	240	240	177

$$\sigma_x = \sqrt{\frac{\Sigma x^2}{N}} = \sqrt{\frac{240}{10}} = \sqrt{24} = 4.9$$

$$\sigma_y = \sqrt{\frac{\Sigma y^2}{N}} = \sqrt{\frac{240}{10}} = \sqrt{24} = 4.9$$

$$r = \frac{\Sigma xy}{N\sigma_x\sigma_y} = \frac{177}{(10)(4.9)(4.9)} = \frac{177}{240.1} = .74$$

that in the formula just given r is the ratio between two averages of variance. If both the numerator and the denominator are divided by N , then the numerator becomes the covariance $\left(\frac{\Sigma xy}{N}\right)$, and the denominator becomes the geometric mean of the variance $\left(\frac{\Sigma x^2}{N}\right)$ of X and the variance $\left(\frac{\Sigma y^2}{N}\right)$ of Y . The geometric mean of two numbers, you may recall from an earlier discussion, is the square root of their product.

It can easily be demonstrated that the denominators of

formulas (20) and (21) are identical. Thus by substituting identities for σ_x and σ_y , we get

$$\begin{aligned} N\sigma_x\sigma_y &= N\sqrt{\left(\frac{\Sigma x^2}{N}\right)\left(\frac{\Sigma y^2}{N}\right)} && \text{and multiplying we get} \\ &= N\sqrt{\frac{\Sigma x^2\Sigma y^2}{N^2}} && \text{and taking the square root of the} \\ & && \text{denominator we arrive at} \\ &= \frac{N}{N}\sqrt{\Sigma x^2\Sigma y^2} && \text{and since } \frac{N}{N} \text{ is equal to 1, then} \\ &= \sqrt{\Sigma x^2\Sigma y^2} \end{aligned}$$

Empirically we can determine the identity by substituting the values for Σx^2 and Σy^2 and solving for r .

$$r_{xy} = \frac{\Sigma xy}{\sqrt{\Sigma x^2\Sigma y^2}} = \frac{177}{\sqrt{(240)(240)}} = \frac{177}{240} = .74$$

c. **Correlation using original measures.** The calculation of the coefficient of correlation from original measures, without subtracting the mean of the Y distribution from each Y score and the mean of the X distribution from each X score, is usually a more convenient method of finding r than either of the two methods described above. In this instance, we do as we did before and assume that the means in each case are equal to zero. We square each of the scores, sum, and apply a correction term for origin to obtain the sum of squares. We already know that

$$\Sigma x^2 = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$$

$$\Sigma y^2 = \Sigma Y^2 - \frac{(\Sigma Y)^2}{N}$$

All that we need to do to get the sum of products (Σxy) is to apply a similar correction term for point of origin. Thus

$$\Sigma xy = \Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{N} \quad (22)$$

And since we now have identities for Σx^2 , Σy^2 , and Σxy , we may substitute in formula (21) to obtain

$$r_{xy} = \frac{\Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{N}}{\sqrt{\left[\Sigma X^2 - \frac{(\Sigma X)^2}{N}\right] \left[\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}\right]}} \quad (23)^4$$

The application of formula (23) is illustrated in Table 17.

d. The difference method for Σxy . The difference formula for obtaining the sum of cross-products (Σxy) is sometimes valuable. It is obtained as follows

$$\begin{aligned} (X - Y)^2 &= X^2 - 2XY + Y^2 \\ \Sigma(X - Y)^2 &= \Sigma X^2 - 2\Sigma XY + \Sigma Y^2 \\ 2\Sigma XY &= \Sigma X^2 + \Sigma Y^2 - \Sigma(X - Y)^2 \\ \Sigma XY &= \frac{\Sigma X^2 + \Sigma Y^2 - \Sigma(X - Y)^2}{2} \end{aligned} \quad (24)$$

Since we already have ΣX^2 and ΣY^2 , we merely need to find the value of $\Sigma(X - Y)^2$ and then to substitute in formula (24) above to solve for ΣXY . The steps are illustrated in Table 18.

e. Correlation using coded scores. The techniques of simplifying computations by coding apply especially well to the calculation of r . That is because the coded results of our computations do not need to be decoded as they do when we find the mean and standard deviation. Consequently, if we are not interested in these statistics, but only in r , we may work directly with the coded values. If we code X and Y by division,⁵ then we obtain the following formula for r :

$$r_{xy} = \frac{\left[\Sigma x'y' - \frac{(\Sigma x')(\Sigma y')}{N}\right] i_x i_y}{\sqrt{\left(\left[\Sigma x'^2 - \frac{(\Sigma x')^2}{N}\right] i_x^2\right) \left(\left[\Sigma y'^2 - \frac{(\Sigma y')^2}{N}\right] i_y^2\right)}} \quad (25)$$

⁴ An alternative formula may be obtained by multiplying both the numerator and the denominator of formula (23) by N . Thus

$$r_{xy} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]}}$$

⁵ The sum of cross-products, like the sum of squares, is uninfluenced by addition or subtraction of a constant.

The Product-Moment Correlation Coefficient

The prime sign indicates that we are dealing with values of X and Y which have been coded, and i_x and i_y are the constants by which X and Y scores, respectively, were

TABLE 17.—CALCULATION OF THE CORRELATION COEFFICIENT FROM ORIGINAL MEASUREMENTS

(1) X	(2) Y	(3) X^2	(4) Y^2	(5) XY
16	14	256	196	224
14	17	196	289	238
13	11	169	121	143
10	5	100	25	50
8	8	64	64	64
7	15	49	225	105
5	6	25	36	30
4	9	16	81	36
2	2	4	4	4
1	3	1	9	3
Σ 80	90	880	1,050	897

$$\begin{aligned}
 r &= \frac{\Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{N}}{\sqrt{\left[\Sigma X^2 - \frac{(\Sigma X)^2}{N}\right]\left[\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}\right]}} \\
 &= \frac{897 - \frac{(80)(90)}{10}}{\sqrt{\left[880 - \frac{(80)^2}{10}\right]\left[1,050 - \frac{(90)^2}{10}\right]}} \\
 &= \frac{897 - \frac{(7,200)}{10}}{\sqrt{\left[880 - \frac{(6,400)}{10}\right]\left[1,050 - \frac{(8,100)}{10}\right]}} \\
 &= \frac{897 - 720}{\sqrt{(880 - 640)(1,050 - 810)}} \\
 &= \frac{177}{\sqrt{(240)(240)}} \\
 &= \frac{177}{240} \\
 &= .74
 \end{aligned}$$

TABLE 18.—CALCULATION OF ΣXY BY THE DIFFERENCE METHOD

(1) X	(2) Y	(3) (X - Y)	(4) (X - Y) ²
16	14	2	4
14	17	-3	9
13	11	2	4
10	5	5	25
8	8	0	0
7	15	-8	64
5	6	-1	1
4	9	-5	25
2	2	0	0
1	3	-2	4
Σ			136

$$\begin{aligned}\Sigma XY &= \frac{\Sigma X^2 + \Sigma Y^2 - \Sigma(X - Y)^2}{2} \\ &= \frac{880 + 1,050 - 136}{2} \\ &= \frac{1,794}{2} \\ &= 897\end{aligned}$$

divided. The numerator of formula (25) is equal to the decoded or "true" value of Σxy , and the denominator contains the decoded or "true" values for Σx^2 and Σy^2 . There is no need to find these "true" values, however, if we are not going to compute the standard deviations of the X and Y distributions. Since the term, $i_x i_y$, appearing in the numerator of formula (25) is canceled by i_x^2 and i_y^2 appearing under the radical sign in the denominator, we may write

$$r_{xy} = \frac{\Sigma x'y' - \frac{(\Sigma x')(\Sigma y')}{N}}{\sqrt{\left[\Sigma x'^2 - \frac{(\Sigma x')^2}{N}\right] \left[\Sigma y'^2 - \frac{(\Sigma y')^2}{N}\right]}} \quad (26)$$

3. CORRELATION COMPUTED FROM A SCATTER DIAGRAM

The technique of calculating a correlation coefficient from a scatter diagram involves principles already familiar. The paired X and Y scores are tallied in a two-way frequency distribution which groups them into class intervals. To illustrate the steps involved we shall make use of the data of Curtis (8) on the relationship between a measure of susceptibility to hypnosis and a measure of intelligence. These data are given in Table 19.

TABLE 19.—SCORES ON A MEASURE OF SUSCEPTIBILITY TO HYPNOSIS AND ON A MEASURE OF INTELLIGENCE FOR 32 SUBJECTS*

SUBJECT	HYP. SUS. SCALE	STANFORD-BINET	SUBJECT	HYP. SUS. SCALE	STANFORD-BINET
MJ	22	136	TF	0	101
DJR	6	106	AEH	22	128
HIR	20	116	RR	16	122
SRB	8	139	JM	13	111
IC	0	103	SN	7	129
JDC	17	126	WP	10	117
MC	21	131	FW	6	116
JLF	13	137	SR	16	129
BHH	14	144	HF	13	109
MEG	5	130	CEF	0	103
DC	6	133	MM	0	104
SS	4	123	HMD	0	111
GG	9	134	JMD	12	131
FES	8	132	GA	4	112
MNS	6	117	GH	12	134
MLC	0	128	TF	0	101

* Data from Curtis (8).

a. Preparing the scatter diagram. Our first step is to make the *scatter diagram*, which is, in fact, a simple two-way frequency distribution or double-entry table. On the left in Table 20 we group scores on the hypnotic scale (Y variable) in terms of an interval of 2 with code numbers and frequencies at the right. At the top we give the classes

TABLE 20.—ILLUSTRATION OF THE COMPUTATION OF THE PRODUCT-MOMENT CORRELATION COEFFICIENT FROM A CORRELATION CHART DATA TAKEN FROM TABLE 19

X = Scores on Stanford-Binet

	X = Scores on Stanford-Binet																(1)	(2)	(3)	(4)	(5)	(6)
	99-101	102-104	105-107	108-110	111-113	114-116	117-119	120-122	123-125	126-128	129-131	132-134	135-137	138-140	141-143	144-146	f	y'	f y'	f y' ²	Σx'y'	Σx' ² y'
22-23									/				/				2	11	22	242	21	231
20-21					/					/							2	10	20	200	15	150
18-19																	0	9	0	0	0	0
16-17								/	/	/							3	8	24	192	26	208
14-15														/			1	7	7	49	15	105
12-13			/	/						/	/						5	6	30	180	40	240
10-11						/					/	/					1	5	5	25	6	30
8-9											//		/				3	4	12	48	35	140
6-7		/			/	/				/	/						5	3	15	45	34	102
4-5					/			/		/							3	2	6	12	22	44
2-3																	0	1	0	0	0	0
0-1	//	///			/				/								7	0	0	0	16	0
(1)	f	2	2	1	1	3	2	2	1	1	3	5	4	2	1	0	1	32	141	903	230	1250
(2)	x'	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15					
(3)	f x'	0	2	4	3	12	10	12	7	8	27	50	44	24	13	0	15	230				
(4)	f x' ²	0	1	4	9	16	25	36	49	64	81	100	121	144	169	196	225	2208				
(5)	Σy'x'	0	0	3	6	8	13	8	8	2	19	29	17	17	4	0	7	141				
(6)	ΣΣy'x'	0	0	6	18	32	65	48	50	16	171	290	187	204	82	0	105	1250				

Y = Scores on Hypnotic Susceptibility Scale

$$\begin{aligned} \Sigma xy &= \left[\Sigma x'y' - \frac{(\Sigma x')(\Sigma y')}{N} \right] / n \\ &= \left[1,250 - \frac{(230)(141)}{32} \right] / (3)(2) \\ &= \left(1,250 - \frac{32,430}{32} \right) / (6) \\ &= (1,250 - 1,013.44) / (6) \\ &= (236.56) / (6) \\ &= 1,419.36 \end{aligned}$$

$$\begin{aligned} \Sigma x^2 &= \left[\Sigma x'^2 - \frac{(\Sigma x')^2}{N} \right] / n \\ &= \left[2,208 - \frac{(230)^2}{32} \right] / (6) \\ &= \left(2,208 - \frac{52,900}{32} \right) / (6) \\ &= (2,208 - 1,653.12) / (6) \\ &= (554.88) / (6) \\ &= 4,993.92 \end{aligned}$$

$$r_{xy} = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}} = \frac{1,419.36}{\sqrt{(4,993.92)(1,486.88)}} = \frac{1,419.36}{2,724.92} = .521$$

$$\begin{aligned} \Sigma y^2 &= \left[\Sigma y'^2 - \frac{(\Sigma y')^2}{N} \right] / n \\ &= \left[993 - \frac{(141)^2}{32} \right] / (6) \\ &= \left(993 - \frac{19,881}{32} \right) / (6) \\ &= (993 - 621.28) / (6) \\ &= (371.72) / (6) \\ &= 1,486.88 \end{aligned}$$

Correlation Computed from a Scatter Diagram

for the measure of intelligence (X variable) in terms of an interval of 3, with the code numbers and frequencies at the bottom.⁶ On each individual we have two measurements, the score on the hypnotic scale and the score on the intelligence test. We make a tally mark in the proper cell in the table for each individual, taking both measurements into consideration. For example, the first subject, MJ, makes a score of 22 on the hypnotic scale and a score of 136 on the intelligence test. To find the cell in which to place the tally, we run up the left-hand intervals until we come to class 22-23 where, in terms of hypnotic scores, this particular one falls. We now move to the right until we come to the class interval 135-137 on the intelligence test. MJ's score on this test is 136 so that it belongs in this interval. Consequently, we place a tally mark in the twelfth cell from the bottom and the thirteenth cell from the left. In this cell we would make a tally for every other subject who has a hypnotic score ranging from 22 to 23 and, at the same time, an intelligence test score of 135-137. In the bottom left-hand corner cell, for example, we find 2 tallies. That is because there are two subjects who have made hypnotic scores from 0 to 1 and intelligence test scores from 99 to 101.

In a similar manner we make a tally for each pair of scores. When we have finished, we could enter numbers to take the place of the individual tallies and in this form, the table is often called a *correlation chart*. We have not entered the numbers in our table because of the small number of cells with more than one tally.

b. The sum of scores and sum of squares. Let us look now at the various entries in the columns at the right of Table 20 and the rows at the bottom. The first four columns numbered (1), (2), (3), and (4) are already familiar. Column (1) is the sum of the tallies for each interval in the Y distribution. It is the f column we used when we worked with a single frequency distribution to find the standard deviation.

⁶ The Y variable is always placed on the vertical scale and the X variable on the horizontal scale in a scatter diagram, as in a graph.

Column (2) is the coded value (y') for the various intervals, column (3) is the frequency for each interval times the coded value (fy') for that interval, and column (4) is the coded value squared times the frequency (fy'^2). All of these we have already used in our work with a single frequency distribution. The first four rows at the bottom of the table are the similar entries for the X variable. We could easily decode the sums at the bottom of the columns and at the end of the rows to find the means and standard deviations of the X and Y distributions if we were interested in them.

c. **The sum of cross-products.** Columns (5) and (6) and rows (5) and (6) are new. They are used to find the sum of cross-products (Σxy) needed for r . Let us see how we get these entries. Column (5) is the sum of x' values for all individuals having a common y' value. For example, there are three individuals with a y' value of 8. One of these has an x' value of 7, another an x' value of 9, and the third an x' value of 10. The sum of all of these x' values is 26 and that is recorded opposite the coded y' value of 8, in the column headed $\Sigma x' \cdot y'$.

To take another case: there are seven individuals with y' values of 0. What is the sum of their x' values? Looking at the table we see that two of these individuals have an x' value of 0, three have an x' value of 1, one has an x' value of 4, and another has an x' value of 9. Summing these x' values gives us 16, and that figure is recorded opposite the coded y' interval of 0.

The other entries are found in a similar fashion. The entries in row (5) at the bottom of the table give us the sum of y' values for all individuals with a common x' value. For example, we find that in the interval 111-113 there are three individuals. What is the sum of their y' values? Running down the table, we find that one of these individuals has a y' value of 6, one has a y' value of 2, and the third has a y' value of 0. Their sum is 8 and that figure is recorded in the $\Sigma y' \cdot x'$ row opposite the coded x' interval of 4.

Column (6) and row (6) are simply the products of the entries in column (2) and column (5) or (y') ($\Sigma x' \cdot y'$) and row (2) and (5) or (x') ($\Sigma y' \cdot x'$), respectively. The total sum of column (6), as indicated by the heading at the *bottom* of the column, is the sum of the cross-products which we need in the numerator of our formula for r . The total sum of row (6) should check exactly with the total sum of column (6). Note also that other checks are provided. Arrows have been drawn to indicate the values that should be precisely the same if computations have been correctly made. Once the column and row totals have been found, all that we need to do is to substitute in the formulas as shown at the bottom of Table 20 and solve for r .⁷

You now have at your disposal a number of different methods of finding a product-moment correlation coefficient. Which method you will want to use depends upon the type of problem you may be called upon to work and upon whether or not you have available a calculating machine. The major advantage of using a scatter diagram is that you can get a picture of the trend of the paired values which unplotted scores will not give. This provides a visual test of whether the relationship is rectilinear—which is always assumed in computing a product-moment correlation coefficient.⁸ But there are many opportunities for errors in making the entries, and there is no check upon this part of the process except to tally the scores a second time. Even then, if you find a discrepancy, you have no way of knowing whether an error was made in the first or second plotting or in both.⁹

Many workers express a preference for using original

⁷ Although it would not have been necessary, we have decoded $\Sigma x'y'$, $\Sigma x'^2$, and $\Sigma y'^2$ before computing r . We did so because we shall want to refer to the decoded values later on when we discuss "predictions."

⁸ A precise test of whether or not the relationship is rectilinear is described on page 237.

⁹ It should be pointed out also that the formula for r is based upon measurements taken by pairs. The calculation of r from a correlation table results in a slight loss in precision. This, however, is negligible if there are 12 or more class intervals and if N is approximately 50 or greater.

scores as they stand, but when N is large and the measurements are large, the calculations are much too laborious without mechanical equipment. In such instances, however, it is possible to code the scores and then to work directly with the coded values. Even when calculating machines are used, it is often timesaving to code the scores before beginning calculations.

4. INTERPRETATION OF CORRELATION

a. **The range of the correlation coefficient.** We already have some general ideas about the size of r . We know that it can range from $+1.00$, through zero, to -1.00 . And we know that a coefficient of $+1.00$ indicates a perfect positive relationship, -1.00 indicates a perfect negative relationship, and $.00$ indicates complete independence. Our problem now is to find some basis for interpreting or evaluating correlation coefficients between $.00$ and 1.00 . We must remember that the correlation coefficient is not expressed in the units of measurement from which it is obtained as are the mean and standard deviation. An r of $.60$, for example, does not indicate twice the relationship that an r of $.30$ does.

b. **The coefficient of determination.** One very useful way of looking at r is in terms of the *coefficient of determination* and the *coefficient of nondetermination*. The coefficient of determination is r^2 and when multiplied by 100 it gives the percentage of variance in Y which is associated with the variance in X or the other way around. If r equals $.80$, then r^2 would be $.64$, and this would indicate the proportion or percentage (the proportion times 100) of the variance in X which is associated with the variance in Y and vice versa. This is, in other words, the amount of variance in one variable that can be accounted for by the variance in the other. The coefficient of nondetermination equals $1 - r^2$ and indicates the amount of variance in one variable or the other which is independent of changes in the second variable. As an approximate standard, then, ignoring other factors such as the size of the sample upon which the r is based, it is

possible to view the size of a coefficient of correlation, regardless of sign, as follows:

r	VARIANCE EXPLAINED	QUALITATIVE EVALUATION
.90-1.00	.81-1.00	very high
.78- .89	.61- .80	high
.64- .77	.41- .60	moderate
.46- .63	.21- .40	low
.00- .45	.00- .20	very low

c. **Common elements.** Another way of looking at r is in terms of the theory of common elements. Out of a box in which we had placed several hundred discs numbered variously, we might draw three, and let these three numbers constitute common factors in a single X and a single Y score. Then if we drew a fourth number out of the box and added this to the first three to get the total value of the X score and drew a fifth number out of the box to add to the three to get the total value of the corresponding Y score, we would have a pair of X and Y values, each member of which was made up of four numbers or elements. Three of the elements or numbers making up the Y score would also be common to the X score; the pair would differ in only one element. We could draw a whole series of such pairs, each member having three elements or numbers in common and differing in only one number or element. If we computed the correlation coefficient for such paired X and Y values, then we would assume that

$$r = \frac{N_{xy}}{\sqrt{(N_x)(N_y)}}$$

where r = the correlation coefficient

N_{xy} = the number of elements common to both X and Y

N_x = the total number of elements in X

N_y = the total number of elements in Y

Since, in the case cited above, N_{xy} equals 3 and N_x equals 4 and N_y equals 4, we would assume the resulting correlation coefficient would be equal to

$$r = \frac{3}{\sqrt{(4)(4)}} = \frac{3}{4} = .75$$

It is doubtful, however, whether there are many correlations obtained from data in the social sciences which can be advantageously interpreted in terms of the theory of common elements. We are never sure of the number of factors determining such things as traits, abilities, aptitudes, motivation, and so forth, which constitute our X and Y variables. If we could assume that X and Y have the same number of total factors, then the obtained r might be interpreted as indicating the proportion of common factors, but in most cases this assumption is probably not justified. Furthermore, we should also have to assume that the contribution of each element is a simple additive function, i.e., if one factor should double or triple the effects of other factors, the interpretation would be in error (86, p. 131).

5. PURPOSE FOR WHICH r IS TO BE USED

It is well to keep in mind, when dealing with correlation coefficients, the purpose for which the r was originally computed. If we are merely interested in determining whether *any* relationship at all exists between two variables, then we have a means for testing whether a given r differs significantly from zero, as we shall see later. By the appropriate "test of significance" we can determine whether an obtained r is sufficiently high to indicate that the hypothesis of no relationship is untenable. If, on the other hand, prediction is our interest, that is, if we desire to predict individual scores on Y from corresponding scores on X , or vice versa, then we also have a means of determining the extent to which our efficiency of prediction is increased by the presence of some relationship. This problem will be taken up later in the chapter on "prediction."

6. ERRORS OF MEASUREMENT AND CORRELATION

Every set of measurements is subject to errors of observation. If, for example, we had several hundred objects of varying lengths and we measured the length of each one twice, we would not expect the pairs of measurements to be *precisely* the same. Slight errors of observation are apt to be present, despite efforts to reduce these to a minimum. Sometimes the second reading might be slightly less than the first, sometimes it might be slightly more, and in other cases we might have exactly the same recorded value for both readings. If these errors of observation are chance errors, then they are just as likely to be positive as negative and would, consequently, have a negligible influence upon the mean. But measures of variability are increased by such errors and measures of relationship are decreased. This means, of course, that the obtained correlation coefficient between two variables is probably always somewhat lower than the correlation which might be obtained between "true" measures of these same variables.

A formula is available for correcting for this state of affairs and it is known as the *correction for attenuation*. By correcting for attenuation we may estimate the correlation which might be obtained between "true" measures of our variables rather than those made with our fallible measuring instruments.¹⁰ Thus

$$r_{\infty\infty} = \frac{r_{xy}}{\sqrt{r_{xx} r_{yy}}}$$

where $r_{\infty\infty}$ = the correlation between "true" measures of X and Y

r_{xx} = the correlation between two sets of measurements of the X variable made with the same measuring instrument

r_{yy} = the correlation between two sets of measurements of the Y variable made with the same measuring instrument

¹⁰ We shall not, in this text, directly concern ourselves with the problems of reliability and validity of educational and psychological measurements. These are problems which can be treated in context in courses in "test construction" or "tests and measurements." Students who wish further information on these problems should consult Greene (37).

THE PRODUCT-MOMENT CORRELATION COEFFICIENT

Example 1—Here are 10 paired measurements for easy practice. Find the coefficient of correlation by the following formula.

$$r = \frac{\Sigma xy}{N\sigma_x\sigma_y}$$

X Distribution	2	5	4	3	6	3	6	4	5	2	7, 4
Y Distribution	1	5	2	3	4	2	5	3	4	1	2, 4

Example 2—Using the data of Example 1, show that the same value of r is obtained when the following formula is used.

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$$

Example 3—Now use the raw score formula to find r . Are you clear as to what the values $\frac{(\Sigma X)^2}{N}$, $\frac{(\Sigma Y)^2}{N}$ and $\frac{(\Sigma X)(\Sigma Y)}{N}$ do in the formula,

$$r = \frac{\Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{N}}{\sqrt{\left[\Sigma X^2 - \frac{(\Sigma X)^2}{N}\right] \left[\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}\right]}}$$

Example 4.—As an algebraic exercise you might try showing that $\Sigma xy = \Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{N}$. Start with $x = (X - M_x)$ and $y = (Y - M_y)$. Then $xy = (X - M_x)(Y - M_y)$. Carry out the multiplication of the right-hand term, summate, and try substituting identities.

Example 5.—A group of women students at a university filled out a Likert-type attitude scale (61). The students also had their mothers fill out the questionnaire. The scores are given below. Is there any correspondence between mothers' and daughters' scores? Without making a frequency distribution, code the scores

in the manner described in the chapter by letting $i = 5$; k will then equal 6.

MOTHER	DAUGHTER	MOTHER	DAUGHTER
31	33	43	43
85	74	58	51
61	60	66	64
91	84	56	52
63	45	71	56
77	60	59	62
81	58	60	59
80	57	70	64
84	64	48	51
58	45	54	56

Example 6.—A class in applied psychology was given Shaffer's S-Scale and C-Scale (82). Shaffer states that there is little relationship between scores on these two scales. Would you be inclined to agree on the basis of the data below? Code the scores without making a frequency distribution by subtracting 10 from each one. After you have found r , ask yourself whether it would have been simpler to have subtracted a smaller number from each score so that negative deviations would not have occurred.

C	S	C	S	C	S	C	S	C	S
5	10	15	7	14	10	9	10	18	9
19	9	11	6	13	7	6	19	14	14
17	10	18	11	19	8	8	8	13	6
14	6	13	11	11	11	18	9	18	8
13	10	14	4	18	8	16	6	18	7
7	12	13	8	5	6	5	7	8	4
13	14	13	6	18	13	14	8	19	12
8	10	4	7	17	6	15	7	22	17
6	17	8	9	23	18	18	12		

Example 7.—The data below are scores on two tests given to an introductory class in general psychology. One test was designed to measure the students' general understanding of the subject matter of the course. We shall call this variable X . The Y variable consists of scores on a vocabulary test of psychological terms. Construct a scatter diagram, letting $i = 5$ on both variables, and

CHAPTER 6

THE CORRELATION RATIO AND OTHER MEASURES OF ASSOCIATION

The Pearson product-moment correlation coefficient discussed in the last chapter assumes that the relationship between the two variables under consideration is rectilinear, that is, that a straight line best describes the increase or decrease in Y with changes in X or vice versa. Sometimes, however, we find variables which do not seem to be related in this fashion; the trend of the plotted points in the scatter diagram is not as accurately described by a straight line as it would be by a curved line. In such cases we speak of a *curvilinear* relationship between the variables and we need a new measure of association to describe it.

Think for a moment of a correlation chart in which the means of each Y column are the same. A line drawn through these means, from left to right, would be a straight line across the correlation table at the level of the mean of the entire Y distribution. Hence there would be no change in Y with change in X ; the average Y score for all individuals with a given high average X score would be the same as the average Y score for individuals with a given low average value of X . The relationship between X and Y would be zero. If the means of the Y columns increased with corresponding increases in X , then the relationship between X and Y would be positive. If the means of Y decreased with increases in X , then the relationship would be negative.

Let us suppose, however, that the means of the Y columns increase at first with increases in X , then level off and begin to drop with values of X beyond a certain point. This situation is illustrated in Figure 7, where we have plotted the paired values of X and Y given in Table 21. Obviously, any straight line drawn through these points would not

accurately describe the trend of the relationship between X and Y. There is, however, a very definite relationship between the two variables.

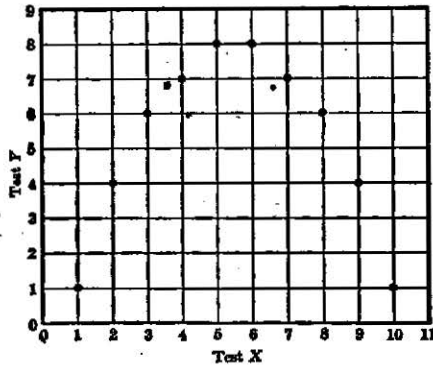


FIG. 7.—A correlation chart showing the relationship between the scores on Test X and Test Y given in Table 21.

TABLE 21.—PAIRED SCORES ON TESTS X AND Y FOR 10 INDIVIDUALS

TEST	INDIVIDUALS									
	1	2	3	4	5	6	7	8	9	10
X	1	2	3	4	5	6	7	8	9	10
Y	1	4	6	7	8	8	7	6	4	1

1. THE CORRELATION RATIO

The *correlation ratio* or *eta* (η) is used to determine the degree of relationship between two variables when the assumption of rectilinearity is not warranted.¹ Eta will be equal to r if rectilinearity does prevail, but if it does not; then η will be greater than the r that would have been obtained from the same set of data. Eta has a maximum

¹ Page 237 provides a precise test of whether the relationship departs sufficiently from rectilinearity to make the assumption invalid.

value of 1.00 and a minimum value of .00 and does not ever carry a negative sign.

Eta differs from r in still other respects. As we shall see later, it is possible, knowing the value of the correlation coefficient, to write an equation which may be used for predicting Y from X or X from Y , but a similar equation cannot be written in terms of η .² And unlike the correlation coefficient which has the same value for r_{xy} and r_{yx} , the values of η_{yx} and η_{xy} may be, and usually are, different. We shall therefore have two formulas for the correlation ratio: one for Y on X and the other for X on Y .

In computing η , measurements of the X and Y variables must first be grouped into classes in a correlation table. As is customary, we place the class intervals for X along the horizontal scale and the class intervals for Y along the vertical scale. We may then find for each class interval of X the mean Y value of all X 's within the class. These values we shall call the *means of the columns* or m_{cy} . In a similar manner, we may find for each class interval of Y the mean X value of all Y 's within the class. These values we shall call the *means of the rows* of the correlation table or m_{rx} . We are now ready for the formulas for the correlation ratio:

$$\eta_{yx} = \frac{\sigma_{m_{cy}}}{\sigma_y} \quad \text{and} \quad \eta_{xy} = \frac{\sigma_{m_{rx}}}{\sigma_x} \quad (27)$$

where η_{yx} = the correlation ratio for Y on X , and η_{xy} = the correlation ratio of X on Y
 $\sigma_{m_{cy}}$ = the standard deviation of the means of the columns, and $\sigma_{m_{rx}}$ = the standard deviation of the means of the rows
 σ_y = the standard deviation of the entire Y distribution, and σ_x = the standard deviation of the entire X distribution

² If the relationship is curvilinear, a curve may be fitted to the means of the columns (or rows, as the case may be) of the correlation table, and the equation of this curve may serve as a basis for making predictions.

Since the computation of η_{yx} and η_{xy} is basically similar, we shall describe only the calculation of η_{yx} . We proceed in finding η_{yx} as in computing r from a correlation table. The standard deviation of the Y distribution is calculated from the values given at the right side of the table (see Table 22), and is already a familiar process. The only new calculation is finding the standard deviation of the means of the columns. One possible way of doing this would be to find first the mean of the complete Y distribution from the values at the right. We would then find the mean of each column and subtract the mean of the entire Y distribution from each of the column means. These values would represent column mean deviations from the mean of the total. The next step would be to square each of the deviations and to *weight* each squared deviation by multiplying by the number of cases in the column upon which it is based. The sum of these squared, weighted deviations, when divided by the total number of cases (N), is equal to the *variance* of the means of the columns. To find the standard deviation of the means, we need only to extract the square root. If the relationship between Y and X were perfect, then $\frac{\sigma_{mcy}}{\sigma_y}$ would equal 1.00 and the standard deviation of the means of the columns would have to be as great as the standard deviation of the entire Y distribution. When σ_{mcy} equals zero, then the correlation ratio also equals zero.

a. A simple method of computation. There is a simpler method for computing the correlation ratio which eliminates the calculation of the mean and standard deviation of the Y distribution as such, and the mean Y value of each column.

This method deals with the *sums* of the columns rather than the means, but the correlation ratio obtained with this method will be identical with that computed according to the procedure outlined above. All that we need to do is to find the *sum of squares for the columns* and the *sum of squares for the complete Y distribution*. In terms of a formula,

which may appear complicated but in terms of which the calculations involved are really simple, the *correlation ratio squared* is

$$\eta^2_{yz} = \frac{\frac{(\sum y'_0)^2}{n_0} + \frac{(\sum y'_1)^2}{n_1} + \frac{(\sum y'_2)^2}{n_2} + \dots + \frac{(\sum y'_k)^2}{n_k} - \frac{(\sum fy')^2}{N}}{\sum fy'^2 - \frac{(\sum fy')^2}{N}} \quad (28)^3$$

where η^2_{yz} = the squared correlation ratio of Y on \bar{X}

$\frac{(\sum y'_0)^2}{n_0}$ = the squared sum of the coded y' values for the first column, divided by the number of cases within the column, and the other similar terms equal the corresponding values for successive columns

$\frac{(\sum fy')^2}{N}$ = the square of the sum of all the y' values divided by the N of the complete Y distribution

$\sum fy'^2 - \frac{(\sum fy')^2}{N}$ = the sum of squares for the complete Y distribution and which is found in the usual way

b. Summary of steps in computing η_{yz} . The method of calculating η_{yz} just described is illustrated in Table 22 where a problem has been worked out in detail in order that you may see where each value in formula (28) is obtained. The steps involved in the calculations may be summarized as follows:

1. Make a correlation table as for the correlation coefficient.
2. Find the $\sum fy'$ and the $\sum fy'^2$ for the entire distribution in the customary fashion.
3. Find the sum of scores for each of the k columns: $(\sum y'_0)$, $(\sum y'_1)$, $(\sum y'_2)$, . . . $(\sum y'_k)$.

³ Note that formula (28) is expressed in terms of eta squared and that in order to find eta we need to extract the square root of the value obtained with the formula.

TABLE 22.—A TABLE TO ILLUSTRATE THE COMPUTATION OF THE CORRELATION RATIO IN TERMS OF FORMULA (27) OR FORMULA (28): SCALE VALUES (X) AND Q VALUES (Y) OF 129 ATTITUDE TEST ITEMS

Q VALUES (Y)	SCALE VALUES (X)								f	f'	fy'	fy ²	
	0	1	2	3	4	5	6	7					8
4.0-4.4			1	1	1	2				5	8	40	320
3.5-3.9		1	1	1	4	5	4	2		18	7	126	882
3.0-3.4		4	3	1	6	3	2	2		27	6	162	972
2.5-2.9		3	2		2	2	1	3		13	5	65	325
2.0-2.4	4	3	2					3		15	4	60	240
1.5-1.9	8	6			1			5		20	3	60	180
1.0-1.4	6							9		15	2	30	60
.5-.9	3							11		14	1	14	14
.0-.4	1							1		2	0	0	0
n	22	17	9	3	14	12	7	14	31	129		557	2,993
Σf'	0	1	2	3	4	5	6	7	8				
Σfy'	55	76	51	21	85	79	45	77	68	557			

COL.	n	Σy'	(Σy') ²	$\frac{(\Sigma y')^2}{n}$
0	22	55	3,025	137.50
1	17	76	5,776	339.76
2	9	51	2,601	289.00
3	3	21	441	147.00
4	14	85	7,225	516.07
5	12	79	6,241	520.08
6	7	45	2,025	289.29
7	14	77	5,929	423.50
8	31	68	4,624	149.16
Σ	2,811.36

Formula (28):

$$r_{yz}^2 = \frac{2,811.36 - \frac{(557)^2}{129}}{2,993 - \frac{(557)^2}{129}}$$

$$= \frac{406.33}{587.97}$$

$$= .6911; r_{yz} = .83$$

Formula (27):

$$r_{yz} = \frac{.89}{1.07} = .83$$

$$\Sigma y'^2 = \left[\frac{\Sigma fy'^2 - \frac{(\Sigma fy')^2}{N}}{N} \right] i_y^2$$

$$= \left[2,993 - \frac{(557)^2}{129} \right] (.5)^2$$

$$= \left[2,993 - \frac{310,249}{129} \right] (.25)$$

$$= (2,993 - 2,405.03) (.25)$$

$$= (587.97) (.25)$$

$$= 146.9925$$

$$\sigma_y = \sqrt{\frac{146.9925}{129}}$$

$$= \sqrt{1.1395}$$

$$= 1.07$$

$$\sigma_{m_{yz}} = \sqrt{\frac{\frac{\Sigma (\Sigma y')^2}{n} - \frac{(\Sigma fy')^2}{N}}{N}} i_y^2$$

$$= \sqrt{\frac{2,811.36 - \frac{(557)^2}{129}}{129}} (.5)^2$$

$$= \sqrt{\frac{(406.33) (.25)}{129}}$$

$$= \sqrt{\frac{101.5825}{129}}$$

$$= \sqrt{.7875}$$

$$= .89$$

4. Square the sums found in Step 3.
5. Divide each of the squared sums of Step 4 by its corresponding number of cases within the column: $n_0, n_1, n_2, \dots, n_x$.
6. Sum the values found in Step 5.
7. Subtract the correction term for origin $\frac{(\sum fy')^2}{N}$ from the value obtained in Step 6.
8. Divide the value obtained in Step 7 by the sum of squares, corrected for origin, for the entire Y distribution:

$$\sum fy'^2 - \frac{(\sum fy')^2}{N}$$
9. Extract the square root of the value obtained in Step 8 to find η_{yx} .

To find η_{xy} it is merely necessary to remember that you are dealing with *rows* instead of columns. Consequently, all that you need to do is to substitute the word *row* for the word *column* and substitute X for Y and x' for y' in the steps outlined above and solve for η_{xy} .

2. BISERIAL CORRELATION

Sometimes an investigator is faced with a situation in which he desires to find the relationship between two variables, but the data for one variable are expressed in terms of a dichotomy (only two categories) or else have been reduced to a dichotomy. We might, for example, be interested in the relationship between the classification of a group of employees as "satisfactory" and "unsatisfactory" and the scores of the group on some test. Or we might have individuals classified as "radicals" and "conservatives" and wish to determine the relationship between scores on a personality inventory and the dichotomy "radical-conservative." Another problem of this nature is the deter-

mination of the relationship between response to a single item on a test and total scores on the same test. Many other problems similar to these arise in psychological and educational research.

If N is at least 50 and preferably larger and if we can make certain assumptions about the data at hand, then there is a statistical technique for determining the relationship between a variable expressed in terms of only two categories and one for which we have a series of measurements. This technique is called the *biserial coefficient of correlation* and is symbolized as r_{bs} . The assumptions which are involved in computing biserial r are that the dichotomized variable is really continuous and, furthermore, that it is normally distributed. These assumptions are usually defended by the argument that if our measurement of the dichotomized variable were sufficiently refined, we would find not just two categories as we have at hand but an infinite gradation along a scale ranging from one extreme to the other. And applying this refined measuring instrument, we would find not just two categories with individuals piling up in one or the other, but instead a normal distribution such as was described earlier.

There is another condition which must be met before we can legitimately compute biserial r . Our dichotomous variable must not constitute merely the two extremes of a larger group, but must include the entire group. We could not, for example, give a test to a large group and then select only the top 25 per cent and the bottom 25 per cent as the members of our dichotomy. If we attempted to compute biserial r with only these two extreme groups, the assumptions concerning continuity and normality of the dichotomous variable would indeed be difficult to justify. In this situation, we would need to make use of a special formula for the biserial coefficient of correlation computed from widespread classes. This formula is not given here, but is developed in Peters and Van Voorhis (74, pp. 384-389).

The formula for biserial r is

$$r_{bis} = \left(\frac{M_p - M_t}{\sigma_t} \right) \left(\frac{p}{y} \right) \quad (29)^4$$

where M_p = the mean score on the continuous variable of the individuals in the category with the higher mean
 M_t = the mean score on the continuous variable of the entire distribution (both categories combined)
 σ_t = the standard deviation of the continuous variable for the entire distribution

good performance = the proportion of the total N in the category with the higher mean on the continuous variable
~~_____~~
 y = the ordinate or height of the normal curve at the point of division between the two groups

This formula for finding biserial r involves only one new value with which we are not familiar, y , the ordinate of the normal curve. This we find from Table B on page 320. Determine first the proportion in the group with the higher mean by dividing the number in this group by the total N . Then look down column B or column C in Table B until you find the value most closely approximating this proportion. Then read the corresponding value of y from the last column of the table. To illustrate the calculation of biserial r we have worked out a problem in detail in Table 23.

If the dichotomous variable cannot be assumed to be continuous and normally distributed, as would surely be the case if this variable consisted of the two categories,

⁴You may also encounter the following formula for biserial r in other sources:

$$\left(\frac{M_p - M_q}{\sigma_t} \right) \left(\frac{pq}{y} \right)$$

where M_q = the mean score on the continuous variable of the group in the category with the lower mean

q = the proportion of the total N in the category with the lower mean and the other values are the same as in formula (29). Formula (29) is to be recommended, however, since as long as the standard deviation must be found in either case, we may as well find the mean of the total distribution while we are at it.

TABLE 23.—ILLUSTRATION OF THE COMPUTATION OF BISERIAL r BETWEEN RESPONSE TO A SINGLE ITEM ON A TEST AND TOTAL SCORES ON THE TEST

SCORES	TOTAL SAMPLE OR COMBINED GROUPS				GROUP MAKING CORRECT RESPONSE	
	f_i	x'	fx'_i	fx'^2_i	f_p	fx'_p
85-89	2	8	16	128	2	16
80-84	3	7	21	147	2	14
75-79	5	6	30	180	3	18
70-74	10	5	50	250	7	35
65-69	15	4	60	240	11	44
60-64	20	3	60	180	8	24
55-59	20	2	40	80	10	20
50-54	4	1	4	4	1	1
45-49	1	0	0	0	1	0
Σ	80		281	1,209	45	172

Midpoint of total

$$M_i = M' + \left(\frac{\Sigma fx'_i}{N_i} \right) i = 47 + \left(\frac{281}{80} \right) 5 = 47 + 17.56 = 64.56$$

$$\sigma_i = \sqrt{\frac{\Sigma fx'^2_i - \frac{(\Sigma fx'_i)^2}{N_i}}{N_i}} = \sqrt{\frac{1,209 - \frac{(281)^2}{80}}{80}} = 8.33$$

$$M_p = M' + \left(\frac{\Sigma fx'_p}{N_p} \right) i = 47 + \left(\frac{172}{45} \right) 5 = 47 + 19.11 = 66.11$$

$$r_{bp} = \left(\frac{M_p - M_i}{\sigma_i} \right) \left(\frac{p}{y} \right) \quad p = \frac{45}{80} = .5625$$

$$= \left(\frac{66.11 - 64.56}{8.33} \right) \left(\frac{.5625}{.394} \right) \quad y = .394$$

$$= \left(\frac{1.55}{8.33} \right) \left(\frac{.5625}{.394} \right)$$

$$= \frac{.8719}{3.282}$$

$$= .27$$

men and women, then Richardson and Stalnaker⁵ suggest that the following formula be used:

$$r_{p-bis} = \left(\frac{M_p - M_q}{\sigma_i} \right) (\sqrt{pq}) \quad (30)$$

where

- r_{p-bis} = the point biserial coefficient of correlation
- M_p = the mean score on the continuous variable of the group in the category with the higher mean
- M_q = the mean score on the continuous variable of the group in the dichotomy with the lower mean
- σ_i = the standard deviation of the continuous variable for the entire distribution
- p = the proportion of the total N in the category with the higher mean
- q = the proportion of the total N in the category with the lower mean

3. TETRACHORIC CORRELATION

Another special case of association is where the data for both variables are in terms of dichotomies or have been reduced to dichotomies. We might, for example, be interested in analyzing responses to items in a questionnaire and we might wish to determine whether, for any given pair of items, a "Yes" response to one is associated with a "Yes" response to the other, while a "No" response to one tends to be associated with a "No" response to the other. Or we might wish to determine the association between a dichotomy such as "normal-abnormal" and another dichotomy such as "Yes" and "No" responses to an item in a test. Or the relationship between being classified as "satisfactory" or "unsatisfactory" as an employee and being "above the mean" or "below the mean" on a given test.

Again, if we are willing to make the assumption that both

⁵ Richardson, M. W., & Stalnaker, J. M. A note on the use of biserial r in test research. *J. gener. Psychol.*, 1933, 8, 463-465.

of the dichotomous variables are really continuous and normally distributed, we can determine the relationship between the two by means of another technique of measuring association, *tetrachoric correlation*. The assumption of continuity of response to a single item demanding a "Yes" or "No" answer, for example, might be justified by stating that all who respond "Yes" do not do so with equal intensity. And likewise all who respond "No" probably do not do so with equal feeling. Some of the "Yes" responses indicate a very emphatic "YES" and others a very weak "Yes," and likewise for the "No" responses. Assuming such a continuous scale, then the additional assumption of normality of distribution along this scale might also be justified.

But even when these assumptions are valid, the labor of computing tetrachoric r is excessive, if all of the terms that properly belong in the formula are included in the solution. Fortunately, a set of computing diagrams, prepared by Chesire, Saffir, and Thurstone (7), simplifies the labor of calculating tetrachoric r tremendously. For this reason the formula for tetrachoric r is not given here. If tetrachoric r 's are to be computed, the diagrams mentioned should be consulted.

4. THE PHI COEFFICIENT

We may, however, inquire as to whether there is not a simpler coefficient which will serve our purpose in determining the degree of association between two dichotomous variables. The *phi coefficient* (ϕ) seems to be suited to this task for, in the first instance, it is applicable to truly dichotomous distributions, that is, where we have discrete categories such as men-women, married-single, alive-dead, and so forth.⁶ In the second instance, the ϕ coefficient may be

⁶ Some might wish to take issue with the statement that the examples cited are true dichotomies, arguing that they are really continuous. A case might be made for this point of view, but in either instance we are dealing with *assumptions*. If we assume continuity, then, as we shall see, the ϕ coefficient may be adapted to the assumption.

adapted to the assumption of continuity, in which case we may also derive an estimate of the corresponding r .

a. **The ϕ coefficient and true dichotomies.** Let us examine first the application of the ϕ coefficient to a problem where we do not feel justified in assuming that our variables are continuous. We shall assume, in other words, that each of our variables represents a true dichotomy.

Suppose we had interviewed a sample of 200 college students, of which 75 were men and 125 were women. Suppose also that one of the questions we had asked was whether they were employed part of the time or not and we found that 45 of the men said "Yes" and 45 of the women said "Yes." We give these data in Table 24, where the cell

TABLE 24.—RELATIONSHIP BETWEEN SEX AND EMPLOYMENT STATUS IN A SAMPLE OF 200 COLLEGE STUDENTS

	WOMEN	MEN	TOTALS
Employed part time.....	45 (a)	45 (b)	90 (a + b)
Not employed part time.....	80 (c)	30 (d)	110 (c + d)
Totals.....	125 (a + c)	75 (b + d)	200 (a + b + c + d)

entries are also represented symbolically by the letters a , b , c , and d . Using these letters, the formula for the ϕ coefficient may be written

$$\phi = \frac{(bc) - (ad)}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} \quad (31)$$

where

(bc) = the number in cell (b) times the number in cell (c)

(ad) = the number in cell (a) times the number in cell (d)

$(a + b)$, $(c + d)$,

$(a + c)$, and $(b + d)$ = the respective border totals

If the product of cells (*b*) and (*c*) is greater than the product of cells (*a*) and (*d*), then this will indicate that "being employed" and "being a male" are positively associated, while if the product of cells (*a*) and (*d*) is greater than the product of cells (*b*) and (*c*), the association will be negative. Substituting the values of Table 24 in the formula, we get

$$\phi = \frac{(45)(80) - (45)(30)}{\sqrt{(90)(110)(125)(75)}} = .23$$

The obtained value of ϕ , .23, indicates that there is a slight tendency for sex classification to be associated with employment status. College men, in other words, are more apt to be employed than college women.

b. The assumption of continuity. Now let us consider a case where we feel justified in making the assumption that our dichotomized variables are really continuous, as in the case where tetrachoric r would customarily be applied to determine the relationship between the variables. Suppose that we had interviewed a sample of 300 college students and found that 200 of them were under 19 years of age and that 100 were 19 years of age or over. We shall call these two groups the "younger group" and "older group," respectively. Of the older group, 20 answered "Yes" to the question: "Do you believe that three years of English should be required of all college students?" In the younger group, 140 answered "Yes" to the same question. Is the relationship between age and a "Yes" response to the question positive or negative?

We give the essential data in Table 25. Note that the data are so arranged as to be consistent with the pattern of a correlation table (page 82). A concentration in the first and third quadrants (cells *b* and *c*) represents a positive relationship and a concentration of the data in the second

TABLE 25.—RELATIONSHIP BETWEEN "YES" AND "NO" RESPONSE AND AGE STATUS

	No	YES	TOTALS
Older students	80	20	100
Younger students	60	140	200
Totals	140	160	300

and fourth quadrants (cells *a* and *d*) represents a negative relationship. Solving for ϕ , we get

$$\phi = \frac{(20)(60) - (80)(140)}{\sqrt{(100)(200)(140)(160)}} = -.472$$

The ϕ coefficient of $-.472$ is based upon the assumption that our variables are "true" dichotomies, whereas, in this instance, we might feel justified in assuming that the dichotomies are really artificial, that we have imposed them upon the data. If we had allowed for varying degrees of response to the question, such as "Strongly Agree," "Agree," "Undecided," "Disagree," and "Strongly Disagree," we would have had the beginning of a continuum. And likewise we might have recorded age as a continuous variable. If these assumptions are valid and if we wish to estimate the corresponding *r*, then we must first find the proportion of cases in the category (*a* + *b*), (*c* + *d*), (*a* + *c*), or (*b* + *d*), whichever is the largest. Then by reference to Table 26,⁷ we find the value of *k*, the constant by which the ϕ coefficient is to be divided in order to estimate the corresponding Pearson *r*. In this problem, the category (*c* + *d*) has

⁷ The values in Table 26 are taken from Camp, B. H. *The mathematical part of elementary statistics*. Boston: Heath, 1931, p. 309. Camp points out that the use of the constants in the table is limited to those cases where no one of the frequencies in the 2x2 table is less than 1 per cent of *N*; the proportion in no one of the border categories is greater than .9; and *r* is not greater than .80 (pp. 309-311). Since these assumptions will be met in most cases where tetrachoric *r* is applicable, the ϕ coefficient divided by the proper constant from Table 26 may be used profitably to provide a quick approximation of *r*.

the largest frequency and the proportion of cases in this category is equal to $200/300 = .67$. According to Table 26, when $p = .67$, then k is approximately equal to .625. Thus

$$r_{\phi} = \frac{-.472}{.625} = -.76$$

We would say, in this instance, that there is a very decided negative association between age and a "Yes" response to the question.

TABLE 26.—SHOWING THE VALUE OF k , THE CONSTANT BY WHICH THE ϕ COEFFICIENT IS TO BE DIVIDED, IN TERMS OF p , THE PROPORTION IN CELL $(a + c)$, $(c + d)$, $(a + b)$, OR $(b + d)$, WHICHEVER IS THE LARGEST, ON THE ASSUMPTION OF CONTINUITY OF BOTH VARIABLES*

p	.5	.6	.7	.8	.9
k	.637	.63	.62	.60	.56

* From Camp, B. H. The mathematical part of elementary statistics. Boston: Heath, 1931, p. 309.

c. The ϕ coefficient and tetrachoric r . To illustrate the correspondence between the estimate of the Pearson r derived by means of the ϕ coefficient and the corresponding r derived by means of tetrachoric correlation, we cite a specific case. Garrett (31, p. 372) reports a tetrachoric r of .53 for the data of Table 27. Using these data and solving for ϕ , we get

$$\phi = \frac{(35)(30) - (25)(10)}{\sqrt{(60)(40)(55)(45)}} = .328$$

Estimating the corresponding Pearson r , we get $.328/.63$, which equals .52, as compared with a tetrachoric r of .53. Some other illustrations are cited in the examples at the end of the chapter. You will find that, in general, ϕ divided by the proper value of k will correspond very closely to tetrachoric r when the conditions mentioned previously are

satisfied. If you are without copies of the computing diagrams for tetrachoric r , ϕ , in terms of simplicity, might well be the most appropriate measure of association to use with 2×2 tables.

TABLE 27.—RELATIONSHIP BETWEEN SUCCESS AS A SALESMAN AND SOCIAL ADJUSTMENT*

	UNSUCCESSFUL SALESMEN	SUCCESSFUL SALESMEN	TOTALS
Socially adjusted.....	25	35	60
Socially maladjusted.....	30	10	40
Totals.....	55	45	100

* Data from Garrett (31).

5. THE CONTINGENCY COEFFICIENT

The *contingency coefficient* (C) is another measure of association which may be applied to data arranged in an $n \times n$ table. This coefficient may be used when both variables can be classified in two or more categories, but when the categories themselves are not quantitative. Such might be the case if we wish to determine whether there is any relationship between the eye color of fathers and their sons. Sons may be classified according to whether the color of their eyes is brown, blue, hazel, or gray, and so may fathers. The problem is to find out whether brown-eyed fathers tend to have brown-eyed sons, blue-eyed fathers tend to have blue-eyed sons, and so forth.

The contingency coefficient varies between 0 and 1, but its sign must be determined by inspection of the table from which it is computed. The contingency coefficient can reach its maximum value only when the number of classes or categories is large. When the data are arranged in a 4×4 table, for example, as would be the case for the hypothetical problem mentioned above, C cannot exceed .866, and for a

10 × 10 table, the maximum value of C would be .949 (102).⁸

The formula for the contingency coefficient may be stated in terms of another statistic, *chi-square* (χ^2). The method of computing χ^2 is described in Chapter 12 and we shall not go into the details here. Once χ^2 has been found, then we may substitute in the following formula to obtain the contingency coefficient.

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}} \quad (32)$$

where C = the contingency coefficient
 χ^2 = the value of chi-square (obtained in the manner described in Chapter 12)
 N = the total number of cases in the table

6. RANK-DIFFERENCE COEFFICIENT

When we have a small number of pairs of observations which have been made in terms of ranks rather than in terms of some measurement, there is another method of measuring the association between the two sets. Ranking differs from measuring in that it merely arranges things in serial order. We might rank individuals in terms of height without making any actual measurements of height in terms of inches. We could simply line up a group of individuals, put the tallest one at the head of the line, followed by the next tallest, and so on until we had arranged all individuals in order. We could then assign rank 1 to the tallest, rank 2 to the next, and so on. Ranks, as contrasted to measurements, do not tell us how much taller the individual ranked 1 is as compared to the individual with rank 2, or rank 3, and so forth.

The method of measuring association between two sets of ranks is known as the "rank-difference method" and the

⁸ For a further discussion of this coefficient and the conditions under which various corrections may be applied, see Kelley (52, pp. 265-271) and Guilford (39, pp. 357-360).

resulting coefficient is called *rho*. The computations are fairly straightforward. In terms of a formula

$$\rho = 1 - \frac{6\sum D^2}{N(N^2 - 1)} \quad (33)$$

where ρ = the rank-difference correlation coefficient
 D^2 = the difference squared between each pair of ranks
 N = the number of pairs of ranks

The computations are illustrated in Table 28, which gives the rank order of importance assigned to various "morale" items by two different groups, a group of employers and a group of employees (28). We wish to determine whether there is any association or relationship between the ranks assigned to the items by the two groups.

TABLE 28.—RANKS ASSIGNED TO VARIOUS MORALE ITEMS BY EMPLOYERS AND EMPLOYEES*

ITEM	EMPLOYER RANKING	EMPLOYEE RANKING	DIFFERENCE	DIFFERENCE SQUARED
1. Credit for work done.	1	7	-6	36
2. Interesting work.....	2	3	-1	1
3. Fair pay.....	3	1	2	4
4. Understanding and appreciation.....	4	5	-1	1
5. Counsel on personal problems.....	5	8	-3	9
6. Promotion on merit..	6	4	2	4
7. Good physical working conditions.....	7	6	1	1
8. Job security.....	8	2	6	36
				92

* Data from Fosdick (28).

$$\begin{aligned} \rho &= 1 - \frac{6(92)}{8(8^2 - 1)} \\ &= 1 - \frac{552}{504} \\ &= 1 - 1.095 \\ &= -.10 \end{aligned}$$

Because formula (33) seems easier to apply than the Pearson product-moment formulas, some individuals would go so far as to transfer a set of measurements into ranks in order to use ρ as a measure of association. For example, when observations were originally made in terms of measurements, these are then converted into ranks by assigning the largest measurement in each series rank 1, the next largest rank 2, and so on. When two or more measurements in the same series are identical, that is, tied for a given rank, the practice is to give each one the average of the tied ranks. Thus two individuals with the same score, tied for, let us say, rank 8, would be given the average of ranks 8 and 9 or 8.5. If three individuals had been tied for rank 8, then each would be given the average of the ranks they would ordinarily occupy, ranks 8, 9, and 10, or the average rank of 9.

Frankly, however, there is little reason for converting measurements into ranks in order to determine the degree of association between the two series. The problem of converting, dealing with ties, finding differences, and squaring these differences, even for a short series, may require more time than the application of one of the formulas for the Pearson product-moment r discussed earlier.

7. MULTIPLE AND PARTIAL CORRELATION

We have discussed under measures of association only those techniques which enable us to describe the relationship between two variables. A product-moment correlation coefficient, for example, describes the relationship between one variable, which we call X , and a second, which we call Y . Technically, correlation coefficients of this nature are called *zero-order* correlations. It is possible, however, through the use of multiple correlation techniques to study the relationship between one variable and several others considered simultaneously. We might, for example, be interested in the relationship between high school marks and intelligence tests scores considered together and later

success in college as measured by point hour ratio. The maximum correlation that we could obtain between these first two variables and point hour ratio can be described by the *coefficient of multiple correlation*. Multiple correlation analysis is not limited to studying the relationship between two variables considered jointly and a third, but can be extended to determine the relationship between a combination of several factors and some other one. But since the multiple correlation coefficient is not the simple sum of the zero-order correlations of each of the combined factors with the variable under study, but takes into consideration the intercorrelations between the several variables in the combination, the law of diminishing returns is soon reached. This is to say that the increase in the size of the multiple correlation coefficient as a result of adding new variables is quite slow when the variables themselves are correlated with one another.

Partial correlation, in contrast to multiple correlation where we combine variables, enables us to hold constant certain variables while we study the relationship between two others. In the problem described above, we might be interested in the correlation we would find between intelligence test scores and college marks, if high school grades were held constant. If we hold one variable constant while we study the relationship between two others, the resulting correlation coefficient is called a *first-order* partial correlation. If two variables are held constant at the same time, we have a *second-order* partial correlation. There is some basis, however, for believing that the application of partial-correlation methods much beyond the first-order stage is inadvisable. We cannot go into the reasons for this belief here,⁹ nor can we give a detailed treatment of the applications and limitations of partial and multiple correlation. For this the interested student is referred to Peters and Van Voorhis (74).

⁹ See, for example (38, pp. 270-271), (74, pp. 244-245), (64, pp. 251-252).

8. SUMMARY OF MEASURES OF ASSOCIATION

By way of summary of the methods of measuring association, let us go back over the various techniques mentioned.

a. Pearson product-moment r : for measuring relationships between two variables when both are continuous and the relationship is rectilinear. The coefficient of correlation is most reliable when based upon a large number of pairs of observations. An r based upon 15 pairs of observations would have to be at least .64 to indicate that the correlation in the population from which the sample was drawn was not zero, for example, whereas the same inference might be made for an r of .18 if the sample consisted of 200 cases.¹⁰ The coefficient of correlation if computed from a correlation table with a small number of classes is likely to be less than the r which would be obtained if calculated directly from ungrouped measures. A table of correction factors has been worked out by Peters and Van Voorhis (74) for r 's computed from tables with varying numbers of classes and should be consulted for more precise estimates of association when r is computed from tables with less than 10 classes.

b. The correlation ratio: for measuring relationships between two variables which are related in a curvilinear fashion. The correlation ratio, η_{yx} , unlike the correlation coefficient, is overestimated when the number of class intervals of X is large so that but a few cases are found in each class. Obviously, if only a single case were present in each column, then the variance of the means of the columns would be as great as the total variance of Y , and the correlation ratio would be 1.00. However, if N is sufficiently large and the grouping of X is in terms of 8 to 10 intervals, each interval is apt to have a sufficient number of cases in it to make the obtained correlation ratio approximately accurate.

c. Biserial r : for measuring relationships when one variable is recorded in terms of a dichotomy and the other is

¹⁰ See pages 185-189 for the method of arriving at this statement.

continuous. Biserial r assumes that the individuals in each of the two categories represent a complete distribution (i.e., not just the two extremes), that the dichotomized variable is really continuous and normally distributed, and that the relationship between the two variables is rectilinear.

d. **Point biserial r :** for measuring the relationship between a truly dichotomous variable and a continuous variable.

e. **Phi coefficient:** for measuring the relationship between two variables that are truly dichotomous.

f. **Pearson r estimated by ϕ and tetrachoric r :** for measuring the relationship between two variables, when each one is recorded in terms of a dichotomy. It is assumed that both variables are essentially continuous and normally distributed and that the measures in each of the categories represent a complete distribution, and that the relationship is rectilinear.

~~g. **Contingency coefficient:** for measuring the relationship between two variables which can be classified in two or more categories, but when the categories themselves are not quantitative.~~

h. **Rank difference coefficient:** for measuring the relationship between two variables, each of which is arranged in terms of rank order.

i. **Multiple correlation coefficient:** for measuring the maximum relationship that may be obtained between a combination of several variables and some other one variable.

j. **Partial correlation coefficient:** for measuring the relationship between two variables with the effects of a third (or several others) held constant.

OTHER MEASURES OF ASSOCIATION

✓ **Example 1.**—Here is a simple set of data for practice. Assume that the X variable represents chronological age and that the Y variable represents scores on a vocabulary test. Compute η_{YX} to determine the correlation ratio of Y on X .

VOCABULARY TEST SCORES	CHRONOLOGICAL AGE								
	15	16	17	18	19	20	21	22	23
150-159					3	1	2	1	2
140-149				2	2	4	4	3	1
130-139			3	1	2	5	5	2	1
120-129		1	2	2	1	4	1	3	4
110-119		3	1	3	3	2			2
100-109	1	1	1						
90-99		2		1					
80-89	1								
70-79	2	2							
60-69	4								

Example 2.—Hay and Blakemore (41) report the following distributions of scores for “inexperienced” and “experienced” workers on the Minnesota Vocational Test for Clerical Workers. Use biserial r to determine whether there is any relationship between classification as an “experienced” or an “inexperienced” worker and scores on the test.

SCORES	INEXPERIENCED GROUP	EXPERIENCED GROUP
190-199		1
180-189	2	4
170-179	3	5
160-169	5	8
150-159	9	30
140-149	25	26
130-139	27	26
120-129	33	42
110-119	48	39
100-109	38	28
90-99	17	20
80-89	14	7
70-79	3	5
60-69	4	
50-59	1	

Example 3.—The following data were obtained from a class in social psychology on a final examination. Using the formula for

the point biserial coefficient of correlation, determine whether there is any relationship between response to Item 22 on the examination and total scores.

TOTAL SCORES	RESPONSE TO ITEM 22	
	Incorrect	Correct
80-84		2
75-79		3
70-74	1	5
65-69	4	7
60-64	3	4
55-59	8	10
50-54	5	9
45-49		1
40-44	2	
35-39	1	1

Example 4.—A group of 100 men and 100 women were polled to determine whether they liked or disliked a particular radio commentator. Of the men, 55 liked the commentator and 45 did not, while 40 of the women said they liked him and 60 said they did not. Find the ϕ coefficient to determine whether there is any relationship between response to the question and sex classification.

Example 5.—Peters and Van Voorhis (74) report a tetrachoric r of .569 for the following data. What is the value of r estimated by way of the ϕ coefficient?

NUMBER OF HOURS OF PEDAGOGY	UNSUCCESSFUL TEACHERS	SUCCESSFUL TEACHERS	TOTAL
Six hours or more.....	20	80	100
Less than 6 hours.....	70	55	125
Totals.....	90	135	225

Example 6.—Lindquist (64) reports a tetrachoric r of .35 for the following data on responses of 150 students to two test items. What value of r is obtained when estimated by means of the ϕ coefficient?

RESPONSE TO ITEM 2	RESPONSE TO ITEM 1		TOTAL
	Wrong	Right	
Right.....	24	56	80
Wrong.....	36	34	70
Totals.....	60	90	150

Example 7.—Assign ranks to the scores listed below and find the rank difference coefficient of correlation.

X Variable	8	13	13	18	14	19	8	4	17	15	22	6	18	8	12
Y Variable	4	14	6	13	8	12	10	7	6	7	17	17	9	9	4

Example 8.—Lo (65) had Chinese men and women and also boys and girls rank the “vices” listed below in order of seriousness. Are the average ranks assigned to these “vices” by boys more closely related to the ranks assigned by girls than to the ranks assigned by men?

“VICES”	MEN	WOMEN	BOYS	GIRLS
Snobbishness.....	1	2	2	2
Cheating.....	2	3	1	4
Sex irregularity.....	5	1	3	1
Stealing.....	4	4	4	3
Selfishness.....	3	8	5	7
Lying.....	6	5	7	6
Gambling.....	8	9	6	8
Laziness.....	7	6	8	11
Gossip.....	9	7	10	5
Extravagance.....	10	13.5	9	10
Vulgar talk.....	12	10	14	9
Swearing.....	11	13.5	13	12
Smoking.....	15	11	11	13
Drinking.....	14	12	12	14
Dancing.....	13	15	15	15

CHAPTER 7

PROBABILITY AND FREQUENCY DISTRIBUTIONS

We have already observed how individual members of a group vary from one another and we know how to measure this variation in terms of the standard deviation of the distribution of measurements. Individuals, however, not only vary from one another; they also differ from themselves if measured at different times. Height, for example, is said to be different in the morning upon arising and at night before retiring. Surely one's weight increases with a heavy meal. Individuals tend to perform better on achievement tests when not fatigued, and so on. Now since measurements on the same individual made at different times may vary, and since measurements of different individuals at the same time may vary, we may expect statistics derived from samples of individual measurements to vary also.

The mean achievement score of a group of college freshmen tested in the morning may not be precisely the same mean score that would have been obtained if the same group had been tested in the afternoon. Nor would we necessarily expect another sample of college freshmen, drawn from the same larger group or population as the first sample and in the same manner, to have precisely the same mean score as the first sample.

If we found the mean intelligence test score of a group of freshmen at a given college to be 115, we might expect the mean intelligence test score of another sample of freshmen to differ from this value. If the difference was only 1 point we might be inclined to say that this is just a "chance" difference. But would we also be willing to attribute a difference of 3 points between the two means as being due

to "chance"? If so, then what about a difference as great as 10 points? How much would the two means have to differ, in other words, before we would be willing to give up the hypothesis that the difference is due to chance?

To take another case: suppose that we were given a box and were told that it contained 50 white discs and 50 red discs. Suppose that we shook the box and drew out a sample of 30 discs. We would expect to have close to 15 red and 15 white discs, but we would probably not be too surprised if our sample had 16 of one color and 14 of the other. What if our sample had 20 of one color and 10 of the other? How far would our sample have to depart from the expected 50-50 division in order for us to suspect that we had been misinformed concerning the contents of the box?

These questions bring us to our next problem in statistical methods: the problem of how much confidence we can place in means, proportions, and other statistics derived from samples. The statistical methods used in investigating this problem are known as "tests of significance," and they enable us to determine, among other things, whether observed differences in sample statistics may be assumed to be the result of chance factors or whether we may reject this hypothesis. But in order to understand the use of these statistical techniques we shall have to consider first something of the general nature of probability and chance, and some of the properties of known frequency distributions.

1. MEANING OF PROBABILITY

The probability of an event may be defined as a ratio, the numerator of which gives the number of times that the event is expected to occur or the number of outcomes favorable to the event, and the denominator of which gives the total number of possible outcomes. The probability, for example, of the head of a coin coming up is $1/2$, the 1 in the numerator representing the outcome favorable to the event.

and the 2 in the denominator representing the total number of possible outcomes—the head or the tail of the coin may come up. The probability of an event occurring plus the probability that it will not occur equals unity (1.00) assuming a dichotomy of “occur” vs. “not occur.” The probability that the head of a coin will come up is $1/2$ and the probability that it will not is $1/2$. The sum of these two is equal to 1.00. It is customary to let p equal the probability that an event will happen, and $1 - p$, which is represented by q , the probability that the event will not happen.

A distinction is made between *a priori* and *empirical* probability (53, II, pp. 2-3). *A priori* probability refers to assumed probability. It is sometimes called rational or mathematical probability. An example would be finding the probability of getting exactly 7 heads and 3 tails in a single toss of 10 coins, on the assumption that each coin is as likely to come up heads as it is tails. *Empirical* probability refers to statements of probability based upon the actual observation of the number of times that an event has occurred in a given number of trials. The ratio of the number of times the event has occurred to the total number of trials is called the *relative frequency of success*. Thus in the coin problem we could toss 10 coins up in the air, say 10,000 times, and note the number of heads and the number of tails occurring on each toss. Then the number of times that 7 heads and 3 tails come up, divided by the total number of tosses, would give the probability of this event occurring. This statement of probability would be derived from empirical observation.

Let us examine first a few cases of a priori probability. Assume that if we toss a single coin, the probability of its head coming up is $1/2$. What is the probability of getting two heads when two coins are tossed or when a single coin is tossed twice? The possible outcomes are HH, HT, TH, and TT. Of these four possible outcomes, only one is favorable to getting two heads. The probability then is $1/4$. If we are not interested in which particular coin comes up “heads”

and which one "tails," then the probability of getting one head and one tail is $2/4$. The probability of getting two tails is the same as that of two heads, $1/4$. The sum of these probability ratios equals 1.00. We have, then, an answer to the question we raised: the probability of getting two heads with two tosses of a single coin is $1/4$. We also have in this simple illustration the basis of a general rule or principle: *the probability that all of a set of independent events will occur is the product of the separate probabilities of each event.* When a single coin is tossed twice, the probability of getting a head on the first toss is $1/2$ and the probability of getting a head on the second toss is $1/2$; the probability of getting two heads—the two tosses are independent, i.e., regardless of how the first toss comes out it will not influence the second toss—is therefore $(1/2)(1/2)$ or $1/4$. In a similar fashion we could determine that the probability of getting three heads from tossing a single coin three times or three coins once would be $(1/2)(1/2)(1/2)$ or $1/8$.¹

Suppose we think for a moment of a single item on a "true-false" test, and let us suppose that a student is going to answer this item by flipping a coin, as students sometimes do in answering true-false questions. The chances of getting a correct answer are $1/2$, since there are two possible outcomes, "correct" and "incorrect," and only one of them is favorable. If this test consisted of 10 items, and if the student answered each item by flipping a coin, then what is the probability that he will get a score of 10 correct on the basis of chance alone? Since each response is an independent event, the chances of getting all correct would be the product of the separate probabilities according to the rule above. Thus $\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}$ or $(\frac{1}{2})^{10}$ would give the probability of this happening. What are the chances that he will get all 10 items

¹ There are a number of possible ways of stating the probabilities just described. We sometimes say that the chances of getting a head on a single toss are even; that the chances are 50-50 of getting a head; that the proportion of heads expected if ten coins were tossed is .5; or that 50 per cent of the coins are expected to be heads.

wrong? Since, again on the basis of chance, this is also $1/2$ for each individual item, then the probability of getting a score of zero would also be $(\frac{1}{2})^{10}$.

2. COMBINATIONS

The cases just discussed are simple enough. But suppose that we had asked what the probability was of the student getting precisely 7 correct answers and therefore 3 wrong ones? Note that we are not here specifying which particular 7 answers need to be correct, but only that 7 be correct. In order to answer this question we need to know *how many combinations of 10 things taken 7 at a time are possible*. This can be determined by the formula for combinations of independent events

$${}_n C_r = \frac{n!}{(n-r)! (r)!} \quad (34)$$

where ${}_n C_r$ = the number of combinations of n things taken r at a time

$n!$ = factorial n or the product of all the integers from n to 1

$(n-r)!$ = the product of all the integers from $(n-r)$ to 1

$(r)!$ = the product of all the integers from r to 1

In the present problem, we want to know the possible number of combinations of 10 things (items) taken 7 (7 correct) at a time. Substituting in formula (34) we get

$$\begin{aligned} {}_{10} C_7 &= \frac{10!}{(10-7)! (7)!} \\ &= \frac{10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{(3 \times 2 \times 1) (7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1)} \\ &= \frac{10 \times 9 \times 8}{3 \times 2 \times 1} \\ &= \frac{720}{6} \\ &= 120 \end{aligned}$$

Thus we find that there are 120 ways in which a student might get precisely 7 items correct and 3 incorrect on a 10 item test, but we still do not know how frequently these particular combinations will turn up. Our complete formula for the probability of getting 7 items correct and 3 incorrect on the test should read

$${}_nC_r p^r q^{n-r} = \frac{n!}{(n-r)! (r)!} p^r q^{n-r} \quad (35)$$

In this formula, p is the probability of getting a correct answer to a single item considered alone, and the exponent of p indicates the total number of correct items in which we are interested. The value of q is equal to $1 - p$, and the exponent of q indicates the number of incorrect items. Substituting in the formula we get

$${}_{10}C_7 \left(\frac{1}{2}\right)^7 \left(\frac{1}{2}\right)^3 = \frac{10!}{(10-7)! (7)!} \left(\frac{1}{2}\right)^7 \left(\frac{1}{2}\right)^3 = \frac{120}{1,024}$$

In a similar fashion we could use formula (35) to determine the probability of the student getting any particular score ranging from 10 to zero.*

3. BINOMIAL DISTRIBUTION

If you have had algebra, then you may have noticed that the value of ${}_nC_r$ gives the coefficient of the $(n - r + 1)$ term in the binomial expansion $(p + q)^n$. That is, ${}_{10}C_7$, for example, gives the coefficient of the $(10 - 7 + 1)$ or fourth term of $(p + q)^{10}$. Expanding,[‡] we would get

$$(p + q)^{10} = p^{10} + 10p^9q + 45p^8q^2 + 120p^7q^3 + 210p^6q^4 + 252p^5q^5 + 210p^4q^6 + 120p^3q^7 + 45p^2q^8 + 10pq^9 + q^{10}$$

and the fourth term is $120p^7q^3$, the coefficient being the number 120. The exponent of p in each of the terms of the

* It is customary to let $0! = 1$.

‡ The rules for expanding the binomial $(p + q)^n$ are summarized below:

1. Each term in the binomial consists of the product of a numerical coefficient and a power of p and a power of q .

2. The first term always has a numerical coefficient of 1 which is understood

binomial expansion, as in formula (35), indicates the number of items correct (successes) and that of q indicates the number of items incorrect (failures), and the coefficients represent the number of ways in which each of these combinations of successes and failures may occur.

Thus to interpret the expansion above in terms of the true-false test on which each of the 10 answers is determined by chance: the probability of getting a score of 10 correct is $1/1,024$; the probability of getting a score of precisely 9 is $10/1,024$; the probability of getting a score of precisely 8 is $45/1,024$, and so forth. The advantage of the binomial expansion is that from it we can readily determine the probability of obtaining a score as large as or larger than any given score. The probability of getting a score of 7 or above, for example, is the sum of the probabilities for the

and therefore is not written; the power of p is always n and the power of q is zero and therefore does not appear; thus the first term is always p^n .

3. In each succeeding term, the power of p decreases by 1 in regular order, while the power of q increases by 1 in regular order, until the final term, q^n , is reached.

4. The product of the numerical coefficient and the power of p in any given term, divided by 1 plus the power of q in that term, will give the numerical coefficient of the term which follows.

If you have difficulty in remembering these rules, then you may always rely upon *Pascal's triangle* shown in Table 29 to find the binomial coefficients.

TABLE 29.—PASCAL'S TRIANGLE FOR FINDING THE BINOMIAL COEFFICIENTS

n	BINOMIAL COEFFICIENTS											
1						1	1					
2					1	2	1					
3				1	3	3	1					
4				1	4	6	4	1				
5				1	5	10	10	5	1			
6			1	6	15	20	15	6	1			
7			1	7	21	35	35	21	7	1		
8			1	8	28	56	70	56	28	8	1	
9			1	9	36	84	126	126	84	36	9	1
10	1	10	45	120	210	252	210	120	45	10	1	

scores, 7, 8, 9, and 10, or $\frac{120 + 45 + 10 + 1}{1,024}$, which is equal to 176/1,024. Thus about 17 times in 100 we would expect a score of 7 or higher to occur by chance alone.

We may now ask another question about our student. How high a score must he make before we would begin to suspect the hypothesis that his answers were determined by chance alone? A score of 8 or above would occur by chance just slightly more than 5 per cent of the time, and a score of 9 or higher would occur by chance just slightly more than 1 per cent of the time. Although the limits are arbitrary, it is customary in most statistical work to refer to the occurrence of an event that would happen by chance alone 5 per cent of the time as representing a *significant* departure from chance expectations, and the occurrence of an event that would happen by chance alone 1 per cent of the time is regarded as representing a *very significant* departure. If we accept these standards, then we would surely regard a student who can *consistently*; i.e., repeatedly, make scores of 9 or higher as answering each item on some basis other than chance.

Coming back to our true-false test of 10 items, suppose that we gave it to 1,024 students and that each student answered each item by flipping a coin, that is, by chance. The probabilities that we worked out earlier could now be considered as the frequencies of each score and we could make a frequency distribution of the 1,024 scores as we have done in Table 30.

Note the symmetry of the distribution of scores in Table 30 and observe that the distribution begins to approximate the bell-shaped, normal distribution to which we have had occasion to refer before. We can perhaps see this more clearly when the frequencies are plotted as in Figure 8. Now if the number of items on our test were increased and scores were rescaled so that the resulting frequency distribution conformed to the base line of Figure 8, the ordinates (the perpendicular straight lines representing frequencies)

of the distribution would be crowded closer and closer together, the steps would become smaller and smaller, and the distribution would approach a continuous curve, normal in shape.

TABLE 30.—DISTRIBUTION OF SCORES ON
A 10-ITEM TRUE-FALSE TEST AS DE-
TERMINED BY CHANCE

SCORE	<i>f</i>
10	1
9	10
8	45
7	120
6	210
5	252
4	210
3	120
2	45
1	10
0	1
Σ	1,024

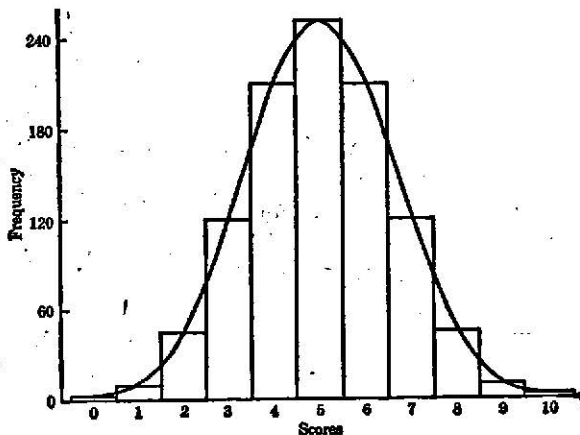


FIG. 8.—Theoretical distribution of scores of 1,024 students on a true-false test of 10 items when responses to each item are determined by chance.

4. THE NORMAL DISTRIBUTION CURVE

The equation for the normal curve is

$$y = \frac{N}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \quad (36)$$

where y = the frequency or height of the curve at any given point along the base line
 N = the number of cases involved in the distribution
 σ = the standard deviation of the distribution
 π = 3.1416, the ratio of the circumference of a circle to its diameter
 e = 2.718, the base of the Napierian system of logarithms
 x = the deviation of a measurement from the mean of the series

Since the values for e and π are known, and for any given distribution we would know the values of N and σ , we could solve for any given frequency (y) corresponding to any given value of x within our series of measurements.

We could, in this way, determine for any given distribution the best fitting normal distribution with the same mean and standard deviation as those actually computed from the data at hand. We shall do this for the true-false test of 10 items and compare the frequencies we obtained by the binomial expansion with those to be expected if the distribution were normal and had the same mean and standard deviation.

Fortunately the labor which would be involved in the calculations based upon formula (36) is greatly simplified by the fact that we have available tabled values of the *unit normal curve*, where N is assumed to be equal to 1 and the standard deviation is assumed to be 1. Table B on page 320 is a table of the unit normal curve. The first column of Table B headed z or $\frac{x}{\sigma}$ gives the standard score value,

$\frac{(X-M)}{\sigma}$. The second column gives the proportion of cases or area between the mean and the standard score value $\frac{x}{\sigma}$. The third column gives the area or proportion below the value $\frac{x}{\sigma}$ or the area in the larger portion of the curve, and the fourth column gives the area above $\frac{x}{\sigma}$ or the area in the smaller portion of the curve. The ordinate y , erected at the point $\frac{x}{\sigma}$, is tabled in the fifth column. Since the normal curve is symmetrical, the tabled values are given for only one half of the curve. Deviations below the mean would be the same except that we would think of them as having a negative sign attached. Thus we see from the third and fourth columns of Table B that exactly .5 of the total area of the curve is above the mean and .5 below the mean.

If we are to fit a normal distribution to the chance distribution of scores on the true-false test of 10 items, then we need to find the frequencies which would be expected, assuming a normal distribution with a mean of 5, standard deviation of 1.58, and N of 1,024. The essential calculations are given in Table 31. Column (1) gives the scores, column (2) the deviations of each score from the mean of the series, and column (3) the deviations divided by the standard deviation of the distribution, that is, $\frac{x}{\sigma}$. The values of y given in column (4) are obtained from Table B. To get the value of y corresponding to a z of 3.16, for example, we enter column (1) of Table B and run down the tabled values of z until we arrive at 3.16. We then read the corresponding value of y from column (5), which, for a z of 3.16, is equal to .0027. Since these ordinates are for the unit normal curve, where N equals 1 and the standard deviation equals 1,

TABLE 31.—FITTING A NORMAL DISTRIBUTION TO THE DATA OF TABLE 30

(1) SCORE	(2) $(X - M)$ z	(3) $\left(\frac{x}{\sigma}\right)$ z	(4) TABLE B y	(5) EXPECTED f_e	(6) OBSERVED f_o
10	5	3.16	.0027	1.7	1
9	4	2.53	.0162	10.5	10
8	3	1.90	.0656	42.5	45
7	2	1.27	.1781	115.4	120
6	1	.63	.3271	212.0	210
5	0	.00	.3989	258.5	252
4	-1	-.63	.3271	212.0	210
3	-2	-1.27	.1781	115.4	120
2	-3	-1.90	.0656	42.5	45
1	-4	-2.53	.0162	10.5	10
0	-5	-3.16	.0027	1.7	1
Total				1,022.7	1,024

$M = 5$
 $\sigma = 1.58$

we must adjust the tabled values of y for the N of 1,024 and the standard deviation of 1.58 of our distribution. We need to compute

$$k = \frac{iN}{\sigma} \tag{37}$$

where k = a constant by which each value of y taken from the table must be multiplied

i = the size of the interval in which the scores are grouped (in this instance $i = 1$, and can therefore be ignored)

N = the number of cases in the distribution

σ = the standard deviation of the distribution

Substituting in formula (37) we obtain

$$k = \frac{1,024}{1.58} = 648.1$$

and multiplying each of the values of y in column (4) by 648.1, we obtain the expected frequencies given in column

(5). In column (6) the frequencies which would result from the expansion of the binomial are given. You may note that there is a slight discrepancy between the sums of columns (5) and (6). The sum of frequencies determined by the normal curve will ordinarily be somewhat less than the N of the data to which we are fitting the normal distribution. This is because the normal curve extends beyond the range of the data. But the discrepancy is very small. By inspection we can see that the departure of the "chance" frequencies from those expected in terms of a normal distribution is not very great.

5. THE USE OF TABLE B

Although we shall not often be concerned with fitting a normal distribution to a given set of measurements, the discussion of how this can be accomplished has served to introduce us to the tabled values of the unit normal curve. We are now ready to consider some of the ways in which this table may be used.

The tabled values are in terms of proportions but we can, if we wish, readily think of them in terms of per cents by multiplying each by 100. Thus the proportion of scores below the mean and above the mean in a normal distribution is .5 and the percentage above and below the mean is 50 per cent. What percentage of the cases in a normal distribution will fall between the mean and plus 1 standard deviation, that is when $\frac{x}{\sigma}$ equals 1? From Table B, column (2), we find that the proportion is .3413 or 34.13 per cent. Since a similar per cent will lie on the other side of the curve between the mean and minus 1 standard deviation, we now see the basis for the statement made earlier that in a normal distribution 68.26 per cent of the scores will fall within the limits set by plus and minus 1 standard deviation from the mean.

Even more important than the per cent falling between the mean plus and minus 1 standard deviation is the per

cent falling between the mean and plus and minus 1.96 standard deviation units. On each side of the curve between the mean and 1.96 standard deviation units would lie 47.5 per cent. These two points will, in a normal distribution, define the zone which embraces the middle 95 per cent of the measures. In a similar fashion we may determine that between the mean plus and minus 2.58 standard deviation units will lie the middle 99 per cent of the measures. Since the entire area under the curve is tabled as 1.00, it follows that the area between ± 1.96 or between ± 2.58 standard deviation units can be expressed as a proportion of the total area, as can the area outside these limits. This is the essence of a probability ratio, as we have already discovered.

Suppose we had a normal distribution of scores with a mean of 60 and a standard deviation of 10. Then, if we wished to determine the limits which would include the middle 95 per cent of the scores, we would solve

$$\frac{x}{\sigma} = \pm 1.96$$

$$\frac{x}{10} = \pm 1.96$$

$$x = \pm 19.6$$

Then between 60 ± 19.6 or between 40.4 and 79.6 we would expect 95 per cent of the scores to fall. And we would expect scores as high as 79.6 or above and as low as 40.4 or below to occur only 5 per cent of the time. Suppose we wrote the value of each score in the distribution on a disc, placed all of the discs in a box, and mixed them thoroughly. We then draw out one disc at a time, record the value appearing on it, and put it back in the box. We do this N times. We would expect, in the long run, that scores of 79.6 and above and scores of 40.4 and below would account for only 5 per cent of N , i.e., of the total number of recorded values. They would occur, in other words, with a frequency of about 5 in 100. The significance of these statements will become

clearer when we take up the problem of sampling in the next chapter.

It should be obvious that using the same procedure that we have just described, we could determine from Table B the answers to the following questions about any given normal frequency distribution:

1. What per cent of the cases would be expected to fall between the mean and any given value of $\frac{x}{\sigma}$?
2. What per cent of the cases would be expected to fall above a given value of $\frac{x}{\sigma}$?
3. What per cent of the cases would be expected to fall below a given value of $\frac{x}{\sigma}$?
4. What per cent of the cases would be expected to fall between any two given values of $\frac{x}{\sigma}$?

6. PRAGMATIC CONSIDERATIONS

If you are about to place some magical belief in the normal probability curve, let us hasten to point out that it is, in nature, a mathematical ideal. Seldom, if ever, do distributions of actual measurements conform precisely to the expected frequencies of the normal curve. Deviations of actual from expected frequencies occur in the most precise set of observations made under the most ideal conditions and where the expectance of a normal distribution is logically justified, for example, in the actual tossing of a set of n coins several thousand times. Distributions of measurements in psychology and education likewise seldom, if ever, give a precisely normal distribution. But there are statistical tests to determine whether or not the departure of observed frequencies from those expected in terms of a normal distribution is sufficient to cause us to reject the

hypothesis of a normally distributed population. As an ideal pattern, the normal distribution curve is not very different from such physical ideals as Boyle's law, to which gases are expected to conform in fact but do so only in theory.

Since many of the measurements in the social sciences are concerned with traits, aptitudes, abilities, achievements, attitudes, and so forth, and since these measurements are usually made by means of a test or scale, we might point out that one of the reasons why we so often find approximately normal distributions with these instruments is because of the manner in which the instruments are reconstructed. If we make up an achievement test for a particular grade level and then find that the test we originally constructed fails to give us a normal distribution when we administer it to students at that particular grade level, we proceed to alter the test. We may make it more difficult if there are too many high scores; or if there are too many low scores we may make it a bit easier. Or we may lengthen the test so that there will be a greater spread of scores. Or we might change some of the items or replace some of the items with others. We might alter the method of scoring the test or do something else to change the test until we do get a normal distribution. There is nothing wrong with this practice. As a matter of fact if the normal curve is to be used in interpreting scores on a test, then the test should be altered or reconstructed to yield a normal distribution of scores.

A question which is often asked is: How do you know whether or not the particular trait or ability in which you are interested is actually distributed normally in the population or not? The answer is that we do not. The question as it stands is operationally meaningless since our measurements are confined to samples and we can only estimate the distribution in the population from what we obtain from our samples. If we intentionally construct a measuring instrument to yield a normal distribution with a sample,

then, if our sample is representative of some larger group which we call the population, we assume that a similar distribution would result if the population were measured with the same instrument.

7. SKEWED DISTRIBUTIONS

When a distribution of measurements departs from normality by yielding more measurements at one end of the scale than at the other, we say that the distribution is *skewed*. In a normal distribution of measurements the mean and median and mode coincide, and there is symmetry or balance between the areas on each side of these measures of central tendency. When measurements are massed at one end or the other of the distribution, the mean, since it is the center of balance, tends to be displaced toward the pointed or tail end of the curve. The mean, as you recall, is influenced by the numerical size of the measurements; the sum of deviations above the mean equals the sum of deviations below the mean. The median, on the other hand, is not influenced by the size of the scores; it is merely the point on each side of which there is an equal number of scores. Consequently, when a distribution is skewed negatively, the median will be larger than the mean as in Figure 9, where the horizontal scale, as usual, increases

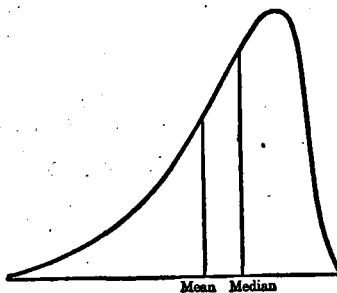


FIG. 9.—Relative position of the mean and median in a negatively skewed distribution.

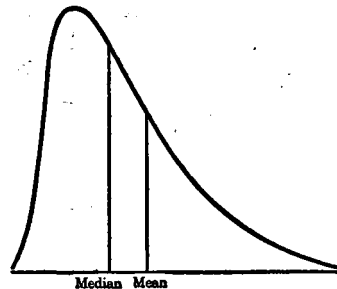


FIG. 10.—Relative position of the mean and median in a positively skewed distribution.

from left to right. When a distribution is skewed positively, the median will be smaller in value than the mean, as in Figure 10.

Another term which is used to describe distributions is *kurtosis*, which refers to the relative peakedness or flatness of a distribution in the neighborhood of the mode. A distribution which is flatter than a normal curve is called *platykurtic* and a distribution which has a higher peak than a normal curve is called *leptokurtic*. There are measures of skewness and kurtosis, but we shall have little need of them and they are not included here.⁴

PROBABILITY AND FREQUENCY DISTRIBUTIONS

Example 1.—Suppose a student is taking a true-false test consisting of 8 items and answers each question by flipping a coin, i.e., by chance.

- (a) What is the probability of his getting a score of 8 correct?
- (b) What is the probability of his getting a score of 6 or higher?

Example 2.—What is the probability of a student getting *at least* one correct answer by chance on two true-false questions?

Example 3.—If you flipped four coins in the air, what is the probability that you will get exactly three heads?

Example 4.—On a multiple-choice test of 4 items, each item with 4 alternatives, what is the probability of getting a score of precisely 3 on the basis of chance?

Example 5.—Given the following normal distribution of a large sample of measurements with mean of 80 and standard deviation of 12, what per cent of the measurements would you expect to find:

- | | |
|--------------|------------------------|
| (a) Above 98 | (g) Below 110 |
| (b) Above 86 | (h) Below 92 |
| (c) Above 65 | (i) Between 68 and 104 |
| (d) Above 50 | (j) Between 68 and 92 |
| (e) Below 68 | (k) Between 56 and 98 |
| (f) Below 86 | (l) Between 68 and 98 |

⁴ See Tippett (90), Peters and Van Voorhis (74), or Kenney (53).

Example (6.) Given the following means and standard deviations of normal distributions of large samples, within what limits would you expect to find the middle 95 and 99 per cent of the measurements?

(a) $M = 25 \quad \sigma = 5$

(b) $M = 30 \quad \sigma = 15$

(c) $M = 50 \quad \sigma = 88$

(d) $M = 42 \quad \sigma = 459.5$

(e) $M = 18.5 \quad \sigma = 224.4$

(f) $M = 22.4 \quad \sigma = 442.2$

(g) $M = 47.0 \quad \sigma = 552.1$

(h) $M = 23.1 \quad \sigma = 215.5$

CHAPTER 8

SAMPLING DISTRIBUTIONS

You now have at your command a number of statistical methods which are capable of giving a good description of samples. Central tendency can be described in terms of the mean or median, and variation in terms of the standard deviation. Raw scores can be translated into z scores to determine how far above or below the mean a given score lies, and to compare standing on one variable with relative position on another. In addition, you are familiar with various methods of measuring the degree of association between two variables. You have also at your disposal a means of determining by inspection how closely a set of obtained sample frequencies corresponds to the frequencies to be expected in a normal distribution with the same mean, standard deviation, and number of cases.

1. SAMPLES AND POPULATIONS

Samples, however, are not often studied for themselves but in order to generalize beyond the samples to the populations from which they were drawn. Why, if our interest is in the population, do we not study it instead of a sample drawn from it? Some of the more obvious reasons for not doing so are that it would be too time-consuming, expensive, impractical, and, in many instances, impossible. If you stop to consider the labor, expense, and time involved in conducting the United States Census once every ten years, this should be clear.¹ An investigator who wishes to study the state of public opinion on a particular issue could not afford to conduct a census, that is, a study of every member of a defined population. Yet, the scientific polling organi-

¹ Stock (88, p. 127) estimates that the 1940 census cost approximately \$50,000,000.

zations can determine quite accurately the state of public opinion by studying a sample drawn from the population (30). This is accomplished with relatively little expenditure of time, money, and labor as compared to the census.

The factors mentioned above are not the only reasons that deter us from studying populations. We may experiment and control conditions and variables much more readily in a small sample than we could in a large population. If we were interested in studying the relative effectiveness of two different methods of teaching arithmetic at the sixth-grade level, we might select as subjects in our experiment two small matched groups, samples, of sixth-grade students. For these small groups we can obviously control the experimental situation more effectively than if our subjects exhausted the population of sixth-grade students. One group we would teach by one method and the other by the second method, comparing the performance of the two groups on some measure of achievement at the end of the experimental period. If our samples are representative of the population from which they were drawn, then we assume that the method of instruction which is found most effective in our experiment will also prove to be most effective for similarly matched groups in the population at large.

You are probably most familiar with "samples" from reading about the opinion polls and, although the samples used in these polls are not samples typical of those we shall deal with in this chapter and later ones, they may serve as an introduction to the subject of "reliability of statistics." The opinion polls, by studying a relatively small sample of the voting population, make generalizations from the sample data about the population which, experience has shown, correspond very closely to the actual outcomes of elections. The polls report their data in terms of per cents, and these per cents, being measures derived from samples, are statistics in the same sense that a mean or standard deviation derived from a sample is a statistic. Statistics, you recall, are estimates of corresponding measures in the population

which are called parameters. The per cent favoring a given issue as derived from an opinion-poll sample would be a statistic; the per cent favoring the issue if it were possible to interview the population from which the sample has been drawn would be the corresponding parameter. The mean intelligence test score derived from a sample of grade school students in New York City is a statistic and is an estimate of the corresponding mean or parameter in the population which the sample is supposed to represent.

Reports of opinion-polling organizations such as the American Institute of Public Opinion usually include a warning statement with pre-election reports of public opinion. A poll may state, for example, that a sample of the voting public shows that Candidate A is favored by 52 per cent of the voters with a *margin of error* of 4 per cent. This means that the per cent in the population favoring Candidate A can quite confidently be assumed to be between 48 and 56 per cent. The 4 per cent margin of error, in other words, gives some indication of the *reliability* of the sample per cent or statistic. The theory underlying the determination of measures of reliability is known as *sampling theory*, and sampling theory is based upon frequency distributions and probability (53, II, 98).

2. SAMPLING DISTRIBUTIONS

We know from the discussion in the last chapter that if we had a large normal distribution with a given N , mean, and standard deviation, we could easily determine what proportion of scores to expect at given distances above or below the mean if we drew scores at random from the distribution. That is to say, if we put each score on a disc, mixed the discs in a box, and drew them forth one at a time, we could make a probability statement concerning the frequency with which we expect to obtain scores at or above a given point, or between two given points.

Let us assume that the distribution of scores in Table 32 was obtained by giving an objective type of examination

to a psychology class of 100 students. If we think of these scores as making up a population, we could compute the mean and standard deviation of the population, which we would find to be 60 and 10, respectively. If our distribution of 100 scores is a population, then the mean and standard

TABLE 32.—HYPOTHETICAL SCORES OF 100 STUDENTS ON AN OBJECTIVE TYPE OF EXAMINATION

87	76	73	70	67	66	64	63	61	60
85	75	72	69	67	65	64	62	61	60
82	74	71	69	67	65	63	62	61	60
78	74	71	68	66	65	63	62	61	60
77	74	70	68	66	64	63	62	61	60
60	59	58	57	56	54	52	50	46	43
60	59	58	57	55	54	52	49	46	42
60	59	58	57	55	53	51	49	46	38
60	59	58	56	55	53	51	48	45	35
60	59	57	56	54	53	50	47	44	33

deviation may be considered as fixed values or parameters of this population.

Suppose now that we placed each of the numbers in Table 32 on a disc and mixed them up in a box and drew *samples* of 1 case each out of the box, replacing the disc after each drawing. The mean of each sample would be equal to $\frac{\Sigma X}{N}$

and since we have but a single X and since N equals 1, the mean of each sample would be $\frac{X}{1}$ or the score itself. If we

drew a large number of samples of 1 case each, we could plot the means of these samples in a frequency distribution. If we then found the standard deviation of this distribution of sample means, it would be approximately the same as the population standard deviation, 10. The reason for this is simply that each sample mean would deviate from the population mean in the same way that each score does from the mean. But if we increased the size of our sample to 10 cases, and drew a large number of samples of this size from

the box, we would find that the sample means now show less variation about the population mean. And with samples of 40 cases each, the variation of sample means would be still less.

The relationship between the variation of sample means and the size of the sample is illustrated by an actual sampling experiment. Figure 11 is a distribution of means of 820 samples of 10 cases each. These samples were drawn by

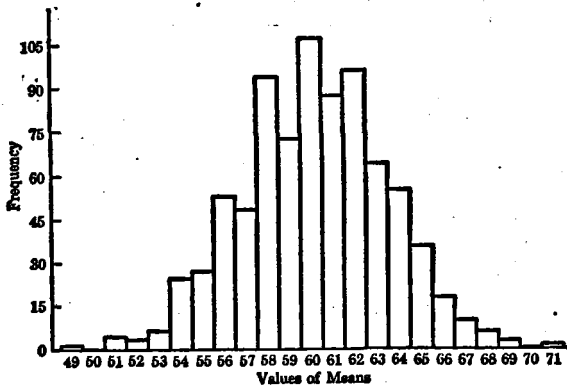


Fig. 11.—Distribution of 820 means of samples of 10 cases each drawn from the scores of Table 32.

students in statistics classes at the University of Maryland and the University of Washington from the scores of Table 32. Note that the lowest mean is 49 and that the highest mean is 71, the range being 22. Observe also the concentration of the sample means around the population value of 60, and the approximately normal shape of the distribution. If we combine the means of four samples, each sample with 10 cases, and find the mean of these combined samples, it would be the same as finding the means of samples of 40 cases each. This we have done, and the distribution of the 205 sample means is shown in Figure 12. You may observe that the range of means is now less than it was when each sample consisted of only 10 cases. The lowest mean is

now 54 and the highest is 66. The range, 12, is only about half that for the samples of 10 cases each.

Frequency distributions of statistics, such as the mean, derived from a large number of samples of a given size are known as *sampling distributions* and the standard deviation of a sampling distribution is called a *standard error* in order to differentiate it from the similar measure, the standard deviation, derived from a single sample. You should see

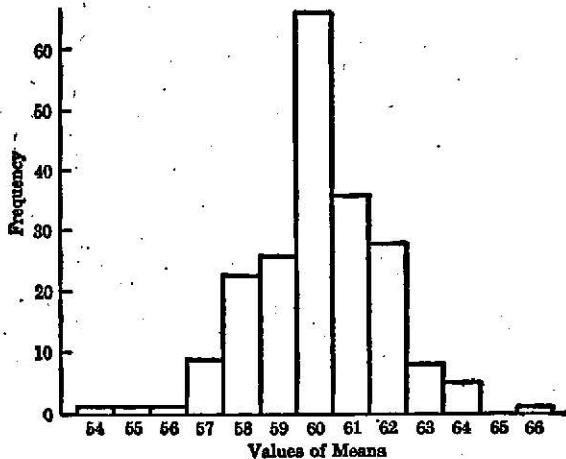


FIG. 12.—Distribution of 205 means of samples of 40 cases each drawn from the scores of Table 32.

from Figure 11 and Figure 12 that the standard error of a distribution of sample means is related to the size of the sample. As more individuals are included in the sample, the less the means will scatter or vary around the population mean. The standard error of a sampling distribution is related also to the amount of variability present in the population from which the samples are drawn. If the standard deviation of the 100 scores making up the population described above had been greater than 10, then the variability of means of samples drawn from this population would be greater also.

3. STANDARD ERROR OF THE MEAN

In the sampling experiments just described we have been dealing with samples drawn from a population where the parameters are known values. Under these circumstances, the formula for the standard deviation of the sampling distribution of means drawn from the population is

$$\sigma_m = \frac{\tilde{\sigma}}{\sqrt{N}} \quad (38)$$

where σ_m = the standard error of the mean

$\tilde{\sigma}$ = the population standard deviation (the tilde is placed over the standard deviation to indicate that it is the population value, not the sample value)

N = the number of cases in the sample

In most research problems in the past, it has been the practice to consider the standard deviation of the sample as an estimate of the population standard deviation, since the latter is seldom, if ever, known. Thus σ has usually been substituted for $\tilde{\sigma}$ in formula (38). It can be shown, however, that a more precise *estimate* of the unknown population standard deviation is obtained by the following formula:²

² The proof of this is not developed here, but can be found in Lindquist (64, pp. 48-50) or Peters and Van Voorhis (74, p. 70). We can, however, give some indication of why we divide the sample sum of squares by $N-1$ in estimating the population standard deviation. The best estimate of the population standard deviation is $\sqrt{\Sigma(X - M)^2/N}$, when M is the *population mean* and N is the number of cases in the sample. But the population mean is not known, and deviations must be taken from the sample mean. The result is that the estimate of the population standard deviation derived from the sample, particularly when the sample is small, is underestimated. The reason for this is that the sum of squares which we calculate, taking deviations from the sample mean, is at a minimum, i.e., less than it would be from any other value. Only in the unusual case where the sample mean happened to be identical with the population mean would the sum of squares based upon the sample mean be as large as the sum of squared deviations from the population mean. Regardless of how slightly the sample mean varies from the population mean, any variation at all would give us a smaller sum of squared deviations if the deviations are taken from the sample mean than would be found if the deviations were taken from the population mean. Division by N would thus give us an estimate of the population standard deviation which is biased, an estimate which is too small. It can be demonstrated algebraically that this bias can be corrected for by dividing by $N-1$ instead of N . The proof in Lindquist is not beyond the comprehension of the student with an elementary knowledge of algebra.

$$\begin{aligned}
 \bar{\sigma} &= \sigma \sqrt{\frac{N}{N-1}} \\
 &= \sqrt{\frac{\Sigma x^2}{N}} \sqrt{\frac{N}{N-1}} \\
 &= \sqrt{\frac{\Sigma x^2 N}{N(N-1)}} \\
 &= \sqrt{\frac{\Sigma x^2}{N-1}} \qquad (39)
 \end{aligned}$$

from which we see that if we want an estimate of the population standard deviation rather than the standard deviation of the sample, we simply divide the sum of squares by $N - 1$ instead of by N . And having computed this estimate of the population standard deviation we could substitute in formula (38) for the standard error of the mean and get

$$\begin{aligned}
 \sigma_m &= \frac{\bar{\sigma}}{\sqrt{N}} \\
 &= \frac{\sqrt{\frac{\Sigma x^2}{N-1}}}{\sqrt{N}} \\
 &= \sqrt{\frac{\Sigma x^2}{N(N-1)}} \\
 &= \frac{\sqrt{\frac{\Sigma x^2}{N}}}{\sqrt{\frac{N(N-1)}{N}}} \\
 &= \frac{\sqrt{\frac{\Sigma x^2}{N}}}{\sqrt{N-1}} \\
 &= \frac{\sigma}{\sqrt{N-1}} \qquad (40)
 \end{aligned}$$

Thus you can see that the standard error of the mean could also be obtained directly from the sample standard

deviation by means of formula (40). It makes no difference which procedure you use. You can divide the sample sum of squares by $N - 1$ and then use the square root of N as the denominator in the formula for the standard error of the mean, or you can divide the sample sum of squares by N and use $N - 1$ as the denominator in the formula for the standard error of the mean.

4. LARGE SAMPLES AND THE NORMAL CURVE TABLE

How shall we interpret the standard error of the mean, once we have computed it? Let us take a concrete case. Suppose that we have given an achievement test to a random sample of 400 students at a given university where the total enrollment is 10,000. The mean of our sample is 200 and the estimate of the population standard deviation is 40. The standard error of the mean is therefore 2. Suppose that our real interest is not in the mean score of our sample of 400 students but rather in the mean score that would have been obtained if we had given the test to every student in the university. We may ask how reliable an estimate of the population mean is our obtained sample mean of 200. We might even wish to ask what the probability is that the population mean is the same as that derived from our sample. Unfortunately, if we insisted upon asking the question in this way, we would be in for a disappointment. For the manner in which the question is phrased eliminates any possibility of an answer.

But, you may ask, didn't we say before that the statistic derived from a sample is an estimate of the population parameter? Aren't we justified, therefore, in saying that the best estimate of the population mean is 200? True enough, but note that this is but another way of stating that the best "hypothesis" we can make about the population mean with the data at hand is 200. Another sample of 400 cases drawn from the same population might have a mean of 201; a third sample might have a mean of 203. Without actually drawing a second and third sample, we might make the

hypothesis that the population mean is actually 201 or 203 or some other value, and that our obtained sample mean of 200 simply represents a *chance* deviation from this value. Obviously, whether we care to accept or reject the various hypotheses that might be set up concerning the population mean will depend upon the relative frequency with which observed sample means of 400 cases would deviate from these assumed or hypothetical values as a result of sampling variation.

Recall that in a normal distribution we may find the ratio $\frac{X - M}{\sigma} = \frac{x}{\sigma} = z$, and that we may then enter the table of the normal curve with any given value of z , in order to determine the relative frequency with which deviations as large as or larger than the given x occur. Now, since the distribution of means of random samples is also normal,³ and since these means will cluster around the population mean at the center of the distribution, it is also possible to write

$$z = \frac{M_o - M_h}{\sigma_m} \quad \downarrow (41)$$

Thus formula (41) tells us that we may set up some hypothetical value of the population mean, find the extent to which our sample mean deviates from this value of the population mean, and then, by reference to the table of the normal curve, determine how frequently such deviations or larger may be expected to occur by chance *if the hypothesis is true*. If deviations as large as the one we have obtained or larger would occur quite frequently as a result of sampling variation, then we would have very little confidence in rejecting the hypothesis that the population mean is the value that we have set up. On the other hand, if a deviation from the hypothetical value of the population mean as large as or larger than the one we have obtained would occur

³ This is true even when the population from which the samples were drawn departs considerably from normality. For evidence bearing upon the problem, see Kenney (53).

quite infrequently as a result of sampling variation, then we might reject the hypothesis concerning the population mean with a greater degree of confidence.

Let us test the hypothesis that the population mean is 199, assuming that our sample mean of 200 represents a deviation from this value. Substituting in formula (41), we get

$$z = \frac{200 - 199}{2} = .5$$

Entering Table B of the normal curve, we find that 31 per cent of the cases in a normal distribution may be expected to deviate from the mean by plus .5 standard deviation units or more. On the assumption, then, of random sampling from a population with a mean of 199, sample means of 200 or larger would occur in the long run 31 per cent of the time. We must admit, that if this is the case, then we would have very little confidence in rejecting the hypothesis that the population mean is 199.

In a similar manner we could test the hypothesis that the population mean is 194. The deviation of our observed mean in terms of standard deviation units would be $6/2$, or 3, and we would find from the table of the normal curve that z values of plus 3 or larger may be expected to occur by chance much less than 1 per cent of the time. Consequently, if the population mean is 194, then sample means of 200 or larger could be expected to occur by chance less than 1 per cent of the time. In this instance we would have much more confidence in rejecting the hypothesis that the true mean is 194.⁴

From these two examples you may see that the degree of confidence we may have in rejecting or accepting a given hypothesis about the population mean depends, as we have said before, upon the relative frequency with which devia-

⁴ Note that in both of the examples cited, we have *not* made the assumption that our sample mean is at the center of the distribution of sample means. It is M_1 which is assumed to be at the center of this distribution and M_0 represents a deviation from M_1 .

tions as great as our sample mean or greater might be expected to occur from the hypothetical value as a result of sampling variation. In other words, assuming a given hypothesis to be true, we test it by finding the relative frequency with which deviations from it as large as or larger than our sample deviation might be expected to occur by chance. If such deviations would occur very frequently by chance, then we cannot reject the hypothesis about the population mean with much confidence. On the other hand, if such deviations would occur very infrequently by chance, then we may reject the hypothesis with a high degree of confidence.

5. THE CONCEPT OF FIDUCIAL LIMITS

The discussion of the previous section, let us hope, has provided a basis for understanding the method now to be described. Instead of testing one hypothesis after another, as we might possibly do, it is customary to determine the *interval* within which any hypothesis might be considered tenable and outside which any hypothesis might be considered untenable. This interval is known as a *confidence interval* and the limits defining it are called *fiducial limits*. Statements of probability made in terms of the fiducial limits are called statements of *fiducial probability*.

It may be observed from Table B, page 320, that absolute values of z of 1.96 or greater will occur, by chance, 5 per cent of the time. It may also be observed that absolute values of z of 2.58 or greater will occur, by chance, 1 per cent of the time. It has recently become common among statistical workers to agree, arbitrarily, to reject a hypothesis about the population mean such that our sample mean deviates from it to the extent that the resulting value of z or greater would occur by chance 5 per cent or less of the time. If we set up a hypothesis concerning the population mean value and found that our sample mean deviated from this hypothetical value to the extent that we obtained an absolute value of z equal to 1.96, we would say that we reject this

hypothesis at the 5 per cent level of confidence. In a similar manner, if we obtained an absolute value of z equal to 2.58, we would say that the hypothesis is rejected at the 1 per cent level of confidence. A z of 1.96, in other words, indicates a *significant* deviation and a z of 2.58 indicates a *very significant* deviation. If we agree upon these standards, then we may determine for a given sample mean the line dividing hypotheses that would be acceptable from those that would be rejected at these levels of confidence. Let us do this for the problem discussed earlier where the sample mean was 200 and the σ_m was 2. The formula we require is the following:

$$\frac{x}{\sigma_m} = \pm z$$

where x = a deviation of the sample mean from a hypothetical value of the population mean
 σ_m = the standard error of the mean
 z = an absolute value of z that would cause us to reject a hypothesis at the 5 per cent level (1.96) or at the 1 per cent level (2.58)

Since σ_m and z are known values, we may substitute in the formula and solve for x . Let us do so for a value of z at the 5 per cent level of confidence.

5% Level

$$\frac{x}{\sigma_m} = \pm z$$

$$\frac{x}{2} = \pm 1.96$$

$$x = (2)(\pm 1.96)$$

$$x = \pm 3.92$$

Having found the value of x (± 3.92), we may note (Figure 13) that if the population mean is as low as 196.08, then our sample mean deviates from this value to the extent that a *plus* value of z of 1.96 is obtained. By reference to Table B we find that plus values of z of 1.96 or larger would occur

.025 or $2\frac{1}{2}$ per cent of the time by chance. Similarly, if the population mean is as large as 203.92, then our sample mean deviates from this value to the extent that a *minus* value of z of 1.96 is obtained (Figure 13). And by reference to Table B, we find that minus values of z of 1.96 or larger would occur by chance .025 or $2\frac{1}{2}$ per cent of the time. Putting these two figures together, we may observe that

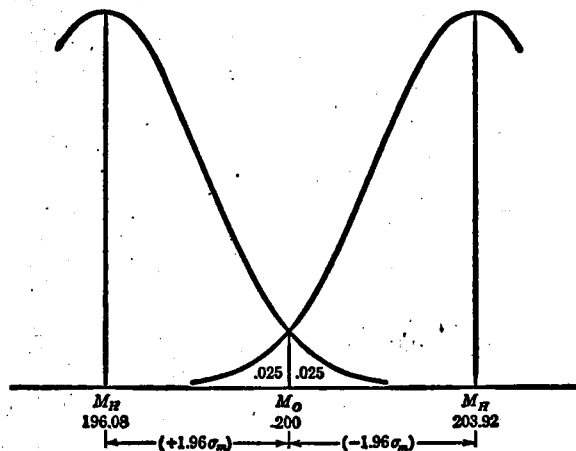


FIG. 13.—The fiducial limits at the 5 per cent level as determined from the tables of the normal probability curve.

absolute values of z (regardless of sign) of 1.96 or greater would arise 5 per cent of the time by chance.

Hence, any hypothesis that the population mean is as low as 196.08 or lower, or as high as 203.92 or higher, will, in terms of the sample mean we have obtained, yield a value of z which would occur 5 per cent of the time or less by chance. The sample mean would be said, therefore, to differ significantly from either of these two hypothetical values of the population mean (or any values outside these two), and any such hypothesis concerning the population mean would be rejected according to the standards we have

agreed upon. We also know that any hypothesis that the population mean is greater than 196.08 but less than 203.92 will be in accord with the value of the sample mean we have obtained, i.e., the sample mean will not differ significantly from any of these hypothetical values.

The limits set by the interval described above have been termed by Fisher (26) the fiducial limits of the parameter at the 5 per cent level. Just as we saw that the fiducial probability that the population mean was 203.92 or greater was $2\frac{1}{2}$ per cent and that the fiducial probability that the population mean was 196.08 or less was $2\frac{1}{2}$ per cent, so also we may say that the fiducial probability that the population mean lies within the fiducial limits is 95 per cent (26, pp. 190-191). That is to say, in the long run, we shall be correct 95 times in 100 in inferring that a population mean lies within the 5 per cent fiducial limits.

If we desire a higher degree of confidence before rejecting a hypothesis concerning the population mean, then we would, of course, work with the fiducial limits of the parameter at the 1 per cent level. In this case the fiducial probability that the population mean is 205.16 or greater is .005 ($z = -2.58$) and the fiducial probability that the population mean is 194.84 or less is .005 ($z = +2.58$). The fiducial probability, therefore, that the population mean lies within these limits is 99 per cent.

If we consistently follow the rule that a hypothesis is acceptable if it falls within the limits set by one of the above levels, then we may be in error because (1) we reject a true hypothesis, or (2) we accept a false one. "Errors of the first kind" may be minimized by arbitrarily increasing the level of significance we demand before accepting the hypothesis and, therefore, making the probability low that if the hypothesis is true we would obtain the value we have obtained.⁵ By adopting the 1 per cent level, for example,

⁵ But, as Tippett points out, by choosing a level that is too severe, "the proportion of mistaken inferences of the second kind may be too great, and advance of knowledge may be unjustifiably impeded. Too much scepticism may be obstructive" (90, p. 75).

"errors of the first kind" would occur, in the long run, only once in every hundred experiments; that is, we would reject 1 per cent of all of the true hypotheses tested. If we work at the 5 per cent level, then we would reject, in the long run, 5 per cent of the true hypotheses tested. As with any rule-of-thumb procedure, caution must be exercised in critical cases. Under some circumstances an "error of the first kind" may be more serious and, under other circumstances an "error of the second" may have more serious consequences.⁶ It may be said, by way of conclusion, that the 1 and 5 per cent levels that we have cited are the most commonly used.

6. SMALL SAMPLES AND THE TABLE OF t

As long as the samples with which we are working are quite large, the ratio, $\frac{\bar{x}}{\sigma_m}$, may be interpreted by reference to the tables of the normal curve. But with small samples (considered by some statisticians as samples under 30 cases and by others as samples under 100 cases) the distribution of the ratio, z , is not precisely normal. This means, of course, that the use of the table of the normal curve in interpreting the ratio is not justified. We must make use of the tabled values of t .⁷

The t ratio is the same ratio that we have just been discussing under the designation of z , and, for large samples, the distribution of t is equivalent to z . The sampling distribution of t depends, however, upon the number of cases in the sample or, more precisely, upon the number of *degrees of freedom* involved. The concept of degrees of freedom, symbolized by n or df , refers to the number of observations that

⁶ See, for example, the discussion by Walker (96, pp. 288-292) and Jackson (47, pp. 13-15).

⁷ The problem of satisfactorily dealing with small samples was recognized and investigated by W. S. Gosset, who published under the pseudonym of "Student." Professor R. A. Fisher has done much to develop the theory and technique of treating small samples and the table of t which we have included (page 330) is taken from his book (25) by permission of the author and his publishers.

are free to vary after certain restrictions have been placed upon the data.⁸ In determining the fiducial limits for the mean, the number of degrees of freedom is one less than the number of cases in the sample or $N - 1$.

Table C, page 330, gives the 1 and 5 per cent points for t for samples with varying degrees of freedom. Table C is to be used in the same way that we used the table of the normal curve, except that the value of t , instead of being 1.96 at the 5 per cent level and 2.58 at the 1 per cent level, will be somewhat larger, the exact value depending upon the number of degrees of freedom involved. Thus for a sample of 20 cases, the number of degrees of freedom is 19 and the value of t at the 5 per cent level is 2.09. For a sample of 10 cases, the number of degrees of freedom is 9, and t at the 5 per cent level is 2.26.

In the case which we have already dealt with in terms of z , we find that for a sample of 200 cases, degrees of freedom = 199, t would be 1.97 at the 5 per cent level and 2.60 at the 1 per cent level.⁹ Using these values we may find the fiducial limits. Thus

1% Level	5% Level
$\frac{x}{\sigma_m} = \pm t$	$\frac{x}{\sigma_m} = \pm t$
$\frac{x}{2} = \pm 2.60$	$\frac{x}{2} = \pm 1.97$
$x = (2)(\pm 2.60)$	$x = (2)(\pm 1.97)$
$x = \pm 5.20$	$x = \pm 3.94$

⁸ An adequate treatment of the concept of degrees of freedom would take us beyond the scope of this elementary text. It can be said, in general, however, that whenever a sample of N cases is used for the purpose of estimating a population value, 1 degree of freedom is lost for every statistic calculated from the sample and used in making the estimate. In the case of estimating the standard deviation of the population, the statistic calculated from the sample and used in making this estimate is the mean of the sample. Since this is the only statistic used in making the estimate, only 1 degree of freedom ($N - 1$) is lost. An excellent discussion of this problem, in terms which are not beyond the comprehension of the beginning student, can be found in Goulden (35, pp. 33-34). See also the more technical article by Walker (94).

⁹ There is no entry in Table C for degrees of freedom 199, but we may use the entry for 200 degrees of freedom, the discrepancy being very slight.

For a sample as large as this ($N = 200$) the fiducial limits would be little changed by using the table of t instead of the table of the normal curve. The greatest discrepancies between the fiducial limits established by z and by t will occur when N drops below 100.

7. OTHER STANDARD ERROR FORMULAS

a. **Standard error of the standard deviation.** The same interpretation that we have applied to the mean of a sample and its standard error also applies to the standard error formulas for other statistics. The sampling distribution of the variance for small samples is, however, slightly skewed, as is the sampling distribution of the standard deviation, but both approximate a normal distribution when N is approximately 50 or greater. We shall not include a formula for the standard error of the variance, but you may find one in Peters and Van Voorhis (74) or Tippett (90) if you should ever have need of it. The standard error of a standard deviation may be estimated by the formula

$$\sigma_s = \frac{\sigma}{\sqrt{2(N-1)}} \quad (42)$$

when the standard deviation has been computed by dividing the sum of squares by N . If the sum of squares has been divided by $N - 1$ in computing the standard deviation, then formula (42) becomes

$$\sigma_s = \frac{\sigma}{\sqrt{2N}} \quad (43)$$

In finding the fiducial limits for the standard deviation, we would proceed as we did in the case of the mean. If we use the table of t , then we must enter it with the appropriate number of degrees of freedom, $N - 1$, or one less than the number of cases in our sample.

b. **Standard error of a proportion.** If, in a given sample, our data are expressed in proportions, frequencies, or per cents, we may wish to compute standard errors for these sta-

tistics, and to establish the fiducial limits. If we interviewed a sample of 100 students and found that 60 ($p = .6$) of them favored a given proposition and 40 ($q = .4$) were opposed, we might be interested in finding the limits within which hypotheses concerning the population proportion would be tenable.

On the assumption that the obtained sample value of p is the most probable population value, the standard error of a proportion is

$$\sigma_p = \sqrt{\frac{pq}{N}} \quad (44)$$

where p = the proportion in one category (in this instance the proportion favoring the proposition)
 $q = 1 - p$ or the proportion in the second category
 N = the number of cases in the sample.

The standard error of the proportion in the case cited above would be, according to formula (44),

$$\begin{aligned} \sigma_p &= \sqrt{\frac{(.6)(.4)}{100}} \\ &= \sqrt{\frac{.24}{100}} \\ &= \sqrt{.0024} \\ &= .05 \end{aligned}$$

To establish the fiducial limits we would enter the table of t with degrees of freedom equal to $N - 1$. The value of t at the 5 per cent level is 1.98 and the fiducial limits would be (.05) (± 1.98) or $\pm .1$ (rounded). Thus we would consider any hypothesis concerning the proportion in the population which fell within the limits .5 and .7 as tenable, i.e., our sample would offer no evidence against the hypothesis. On the other hand, any hypothesis that the population proportion was equal to or outside these limits would be considered untenable and rejected.

c. Standard error of a per cent or frequency. Since we know that a per cent is simply a proportion multiplied by 100 and that a frequency in a given category is equal to the proportion in the category multiplied by the total number of cases involved, we may write the formula for a standard error of a per cent and a standard error of a frequency as follows:

$$\sigma_P = 100 \sqrt{\frac{pq}{N}} \quad (45)$$

$$\sigma_f = N \sqrt{\frac{pq}{N}} \quad (46)$$

A word of caution should be expressed concerning the calculation of standard errors for proportions, per cents, and frequencies, and using these values in establishing the fiducial limits. The procedure described will give approximately the same results as those obtained with more exact methods as long as the product Np (or Nq if q is less than p) equals at least 20. If this relationship does not hold, however, then a more exact treatment is necessary.¹⁰

SAMPLING DISTRIBUTIONS

Example 1.—Place the scores of Table 32 on discs or beans. Assume that the 100 scores make up a *population* with known *parameters*. From this population each member of the class will draw 10 *samples* of 10 cases each. The technique to be used in drawing the samples is this: place the numbered discs in a box with a hole cut in one end; shake the box and draw out one disc; record the number and put the disc back in the box; shake it, draw out another disc and so on until 10 numbers have been recorded. These numbers will make up Sample No. 1. Repeat the process until you have drawn 10 samples.

- (a) Find the mean of each of your 10 samples. Do not worry about the decimal place; round the number.

¹⁰ See, for example: C. J. Clopper & E. S. Pearson. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 1934, 26, 404-413.

(b) To get some idea of the sampling distribution of means, make a frequency distribution of all of the sample means drawn by the members of your class.

(c) What would you expect to happen to the range of means if the sample size had been larger than 10? Why?

(d) To get some idea of the sampling distribution of t you can find the standard deviation and standard error of the mean for each of your samples. Then using the formula

$$t = \frac{M_o - M_h}{\sigma_m}$$

and substituting the mean of the population (60), since it is known, for M_h , solve for t . Some of the values will carry a minus sign and others will be positive.

(e) Combine all of the t 's from the class into a frequency distribution, in terms of absolute values.

(f) What absolute value of t would you expect to find exceeded 1 per cent of the time? 5 per cent of the time?

Example 2.—Given the following means, N 's, and sample standard deviations, find the fiducial limits of the parameter at the 5 per cent and 1 per cent levels, using the table of t .

	N	M	σ		N	M	σ
(a)	65	25	5	(e)	26.17	25	10
(b)	17	30	7	(f)	17.26	50	12
(c)	101	35	10	(g)	50.40	50	7.5
(d)	10	25	5	(h)	200.30	22.5	4.25

Example 3.—Assume that the scores given below are "time" scores on a maze test. Using the table of t , find the standard error of the mean and the fiducial limits at the 1 and the 5 per cent levels.

RL	26	TH	17	LW	21
SS	20	BM	20	AM	17
EG	19	AE	23	VB	22
RG	15	WW	25	MG	19
EE	21	SC	16	NK	19
GH	19	PC	21	AB	20

CHAPTER 9

THE t TEST OF SIGNIFICANCE

In studying the reliability of various sample statistics in the last chapter, we discussed most of the basic essentials of the t test of significance applied to differences between sample statistics. In experimental and research work the determination of whether an observed difference is of such magnitude that it cannot be attributed to chance factors or sampling variation is often our major interest. We may find, for example, that a group working under one set of experimental conditions has a higher mean output than a comparable group working under a different set of experimental conditions. Is the observed difference between the means one that might occur frequently by chance, i.e., as a result of sampling variation? If not, then we might infer that the difference is a product of the experimental variables.

1. AN EXPERIMENT INVOLVING PAIRED OBSERVATIONS

Let us suppose that we are interested in the problem of whether attitudes toward working conditions are important determiners of output. We have 20 subjects who have been practicing adding numbers under quiet conditions. On the basis of their performance during these practice periods we divide the 20 subjects into 10 pairs, so that each subject is matched with another individual of comparable level of performance. We assign one member of each pair at random to one of two groups. Thus we have two groups or 10 pairs of subjects, each pair, and consequently each group, being fairly equal in ability to add. One of these groups we shall designate as "Group A" and the other as "Group B."

On the critical trials we tell the members of Group A that they are to be subjects in an experiment on distraction which

is to be a check on previous experiments which have been done. It has previously been found, we add, that working under conditions of noise results in an increase in production, that is, that most individuals find that noise tends to facilitate adding. The members of Group B are also told that they are to be subjects in an experiment on distraction, but they are told that previous experiments have shown that working under conditions of noise tends to decrease production. Each group is then put to work adding problems under noisy conditions, and performance is measured in terms of the number of problems correctly added. The scores of Group A and Group B on the critical trials are given in Table 33.¹ We see that the mean score of Group A

TABLE 33.—MEASURES OF PERFORMANCE OF MATCHED PAIRS WORKING UNDER CONDITIONS OF NOISE BUT WITH DIFFERING ATTITUDES TOWARD INFLUENCE OF NOISE—GROUP A UNDER ATTITUDE THAT NOISE INCREASES OUTPUT AND GROUP B UNDER ATTITUDE THAT NOISE DECREASES OUTPUT

GROUP	MATCHED PAIRS										SUM	MEAN
	1	2	3	4	5	6	7	8	9	10		
A	22	25	24	23	26	23	26	24	25	22	240	24
B	21	25	22	23	24	22	25	23	24	21	230	23

is higher than the mean score of Group B. Is the difference between the two means significant?

If we took an infinite number of samples of 10 paired observations each, the differences between the means, i.e., the mean differences, would constitute a sampling distribution in the same sense that means of single samples give us a sampling distribution.² We could plot these mean differences in a frequency distribution and they would tend to

¹ The data are hypothetical for the sake of simplicity, but see the experiment by Baker (3).

² It has been demonstrated earlier that the difference between the means is equal to the mean of the paired differences.

cluster around the mean of this distribution, which would be the population mean difference. The standard deviation of this distribution would indicate the amount of variability in mean differences which is to be expected when the samples consist of only 10 pairs of observations. How large a mean difference will we have to observe before we are willing to reject the hypothesis that the population mean difference is zero? It should be obvious from earlier discussions that we would need to know the standard deviation (standard error) of the distribution of mean differences based upon samples of 10 paired observations. As in the case of the standard error of a single mean, we must use our available statistics to estimate the standard error of the mean difference.

a. **Standard error of a mean difference.** In the case of matched pairs, the technique of finding the standard error of the mean difference is quite simple. We find the difference between each pair of observations and then the mean of this distribution of 10 differences. We then find the sum of squares and the standard deviation of the distribution of 10 differences; using this standard deviation as the numerator and the square root of the number of pairs minus 1 as the denominator, we may estimate the standard error of the mean difference. By formula this is

$$\sigma_{m_d} = \frac{\sigma_d}{\sqrt{N - 1}} \quad (47)^3$$

where σ_{m_d} = the standard error of the mean difference (or difference between the means)
 σ_d = the standard deviation of the distribution of differences between pairs
 N = the number of pairs

The essential calculations for the problem at hand are given in Table 34. The scores for both groups have been

³ If the numerator is the estimate of the population variance, i.e., if the sum of squares has been divided by $N - 1$, then the denominator of formula (47) becomes the square root of N .

reduced by subtracting 20 from each one. Subtraction, you may recall, will not change the resulting sums of squared deviations from the mean. And since the same constant has been subtracted from each series, the difference between

TABLE 34.—PERFORMANCE SCORES OF MATCHED PAIRS GIVEN IN TABLE 33 CODED BY SUBTRACTING 20 AND ARRANGED FOR COMPUTATIONAL PURPOSES

(1) GROUP A X	(2) GROUP B Y	(3) D D	(4) D d	(5) d ²	(6) x	(7) y	(8) xy	(9) x ²	(10) y ²
2	1	1	0	0	-2	-2	4	4	4
5	5	0	-1	1	1	2	2	1	4
4	2	2	-1	1	0	-1	0	0	1
3	3	0	-1	1	-1	0	0	1	0
6	4	2	1	1	2	1	2	4	1
3	2	1	0	0	-1	-1	1	1	1
6	5	1	0	0	2	2	4	4	4
4	3	1	0	0	0	0	0	0	0
5	4	1	0	0	1	1	1	1	1
2	1	1	0	0	-2	-2	4	4	4
Σ 40	30	10	0	4	0	0	18	20	20

$$M_d = \frac{\Sigma D}{N} = \frac{10}{10} = 1$$

$$\sigma_x = \sqrt{\frac{\Sigma x^2}{N}} = \sqrt{\frac{20}{10}} = \sqrt{2} = 1.41$$

$$\sigma_{m_x} = \frac{\sigma_x}{\sqrt{N-1}} = \frac{1.41}{3} = .47$$

$$\sigma_y = \sqrt{\frac{\Sigma y^2}{N}} = \sqrt{\frac{20}{10}} = \sqrt{2} = 1.41$$

$$\sigma_{m_y} = \frac{\sigma_y}{\sqrt{N-1}} = \frac{1.41}{3} = .47$$

$$\sigma_d = \sqrt{\frac{\Sigma d^2}{N}} = \sqrt{\frac{4}{10}} = \sqrt{.4} = .63$$

$$\sigma_{m_d} = \frac{\sigma_d}{\sqrt{N-1}} = \frac{.63}{3} = .21$$

$$r_{xy} = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}} = \frac{18}{\sqrt{(20)(20)}} = \frac{18}{20} = .90$$

the pairs and the means will be unchanged. Consequently, we may treat these coded scores as original measures, without taking into consideration any corrections for coding.

We find, from Table 34, that the mean difference between Groups A and B is 1, with a standard error of .21. How may we evaluate this observed difference between the means of the two groups? Is the difference so small that it might simply be the result of sampling variation? One way in which we might approach the problem is to set up some hypothesis concerning the population mean difference and then see whether our sample difference departs significantly from this hypothetical value. The deviation of our sample difference from the hypothetical population mean difference when divided by the standard error of the difference would give us the familiar t ratio. Assuming the hypothesis to be true, we could, by reference to the table of t , determine how frequently absolute values of t this size or larger would occur by chance. According to the standards we have agreed upon, if the value of t is such that it would be expected to occur less than 5 per cent of the time by chance we could say that t was significant. We might, therefore, conclude that the hypothesis concerning the population mean difference is not likely and reject it as untenable. Suppose, however, that we found just the opposite, that t was not significant at the 5 per cent level. What might we then conclude? We would have no basis for rejecting the hypothesis, but would this mean that the hypothesis was true? The answer is definitely no. The hypothetical value we tested is but one among many values that might result in a non-significant value of t .

b. Testing the null hypothesis. There is another approach to the evaluation of our observed mean difference that is more satisfactory if we are merely interested in finding out whether the difference is significant. The hypothesis that we may set up to test is the *null hypothesis*.⁴ We

⁴In this section and in those that follow, the term "null hypothesis" is often used to designate the hypothesis that a given population parameter is zero. It should be pointed out, however, that the term covers any hypothesis which is set up to be tested for possible rejection. For a more detailed discussion of this point, see the footnote on page 281.

assume that the population mean difference is zero and that, therefore, any observed difference such as the one we have obtained is merely due to chance or sampling fluctuations. This, in many respects, is the most logical hypothesis to test, since the major point we wish to determine is whether our observed difference between the means of the groups is significant. If we assume that the population mean difference is zero, then

$$t = \frac{M_o - M_k}{\sigma_{\bar{d}}} = \frac{1 - 0}{.21} = 4.76$$

or the observed mean difference divided by the standard error of the difference, since M_k is assumed to be zero.

We may now evaluate this t by entering Table C (page 330) with the appropriate number of degrees of freedom. In the present problem the number of degrees of freedom is equal to $N - 1$, where N is the number of pairs of observations. According to Table C, if the null hypothesis is true, then for 9 degrees of freedom we would expect to get an absolute value of t of 2.26 or larger 5 per cent of the time, and an absolute value of t of 3.25 will be exceeded only an average of once in 100 times. As before, we agree that if our obtained value of t is such that it would be exceeded 5 per cent of the time or less, then we shall reject the hypothesis being tested and infer that our observed mean difference is significant. In the present experiment, since t exceeds the 1 per cent point, the null hypothesis must be abandoned. We still do not know, however, what the population mean difference is; we only know that the hypothesis that it is zero is not tenable at the level of significance we have agreed upon.

c. Establishing the fiducial limits. We may go a step further in our development. In the manner of the last chapter we may determine the fiducial limits; that is, find out something about the limits within which we would be

willing to accept hypotheses about the population mean difference. Thus

1% Level	5% Level
$\frac{x}{\sigma_{m_d}} = \pm t$	$\frac{x}{\sigma_{m_d}} = \pm t$
$\frac{x}{.21} = \pm 3.25$	$\frac{x}{.21} = \pm 2.26$
$x = (.21)(\pm 3.25)$	$x = (.21)(\pm 2.26)$
$x = \pm .68$	$x = \pm .47$

Any hypothesis that the population mean difference is within the limits .53 to 1.47 would have to be accepted as tenable. On the other hand, any hypothesis that the population mean difference is as large as 1.47 or larger and as small as .53 or smaller would be considered untenable at the 5 per cent level. At the 1 per cent level the fiducial limits would be, of course, .32 and 1.68. Observe that in this instance also, we do not know what the population mean difference is, and that we cannot make any satisfactory statement of probability about its value. We are only confident, at a defined level, that it does not equal or exceed, in either direction, the fiducial limits.⁵

d. Another method for computing the standard error. The method of evaluation just described is for testing the significance of the difference between means of samples where members of one sample have been paired against the members of the second sample, as in the experiment described. The method would apply also when we give the same group a pre- and an end-test designed to measure the same factor, or if we have observations on the same variable for the same group before and after some intervening experimental situation. Under any of these conditions, we might, if we so desired, work directly with the differences between pairs in finding the standard error of the mean difference.

⁵In this connection, the student should reread the section of the previous chapter concerning "errors of the first, and second kinds."

Another method which might be used, however, would be to compute the correlation coefficient between the pairs of measurements and then to apply the following formula

$$\sigma_{m_d} = \sqrt{\sigma_{m_1}^2 + \sigma_{m_2}^2 - 2r_{12} \sigma_{m_1} \sigma_{m_2}} \quad (48)$$

where σ_{m_d} = the standard error of the mean difference
 σ_{m_1} = the standard error of the mean of one group
 σ_{m_2} = the standard error of the mean of the second group
 r_{12} = the correlation coefficient between the pairs of measurements on the experimental variable (the variable on which we are comparing the difference between the means)

We have all the necessary data in Table 34 to use formula (48). Substituting we find

$$\begin{aligned} \sigma_{m_d} &= \sqrt{(.47)^2 + (.47)^2 - (2)(.90)(.47)(.47)} \\ &= \sqrt{.4418 - .3976} \\ &= \sqrt{.0442} \\ &= .21 \end{aligned}$$

which is the same value that we obtained by working directly with the differences between pairs. In evaluating the t based upon formula (48) we would again have 9 degrees of freedom available or $N - 1$, where N is the number of pairs.⁶ The application of formula (48), however, is more laborious than working directly with the differences. In addition to finding the correlation coefficient, we must com-

⁶ Guilford (38) suggests that the number of degrees of freedom in this instance would be equal to $N - 2$, where N is the number of pairs. If this were the case, however, then the evaluation of t based upon the difference formula (45) and that based upon the correlation formula (46) might differ significantly in critical cases, despite the fact that the two formulas would lead to equivalent standard errors.

Professor Helen M. Walker, in a personal communication, has explained the rationale for using $N - 1$ with both formulas: "One arrives at the value of $N - 1$ (where N is the number of pairs) by either of two legitimate courses of reasoning: (1) We may think of the N differences as measurements on a random sample of N cases restricted only by the fact that deviations are measured from the sample mean. (2) We may think of the $2N$ cases as pre-

pute the standard deviations of the scores for each group and the standard error of the means. The difference formula takes the correlation element into consideration and simplifies the work involved. That is why it is preferred.

2. EXPERIMENTS INVOLVING MATCHED GROUPS

Sometimes in an experiment another method of matching groups is used. We do not pair off the individuals in one group with the individuals in the other, but arrange our two groups so that the means, standard deviations, and distributions of the two groups are comparable on some variable, without regard to individuals as such. The basis on which we match the groups may be preliminary practice or performance on the variable on which we intend to measure final outcomes in the experiment, or it may be on the basis of some other variable which we have reason to believe will be correlated with the variable on which we are going to measure outcomes. The reason for matching our groups on the basis of a variable which will be correlated with the experimental variable is that the standard error of the difference will be reduced if the two are positively correlated. The formula for the standard error of the mean difference is as follows:⁷

$$\sigma_{m_d} = \sqrt{(\sigma_{m_{x_1}}^2 + \sigma_{m_{x_2}}^2)(1 - r_{xy}^2)} \quad (49)$$

where σ_{m_d} = the standard error of the mean difference
 $\sigma_{m_{x_1}}$ = the standard error of the mean for one group on the X variable on which we are testing the difference.

senting an analysis of variance problem, in which the total number of degrees of freedom is divided as follows:

Among pairs	$N - 1$
Between groups	1
Interaction	$N - 1$
Total	$2N - 1$

The $N - 1$ degrees of freedom among pairs are the same $N - 1$ obtained in considering this as a sample of N pairs" (97). This second line of reasoning is phrased in terms of "analysis of variance," a technique which you will encounter later in Chapters 10 and 11.

⁷ See Wilks (99) and McNemar (68).

σ_{m_2} = the standard error of the mean for the second group on the X variable

r_{xy} = the correlation coefficient between the X variable on which we are testing the difference and the Y variable on which the groups were equated. The r is based upon the total N , that is, the N of the combined samples.

The difference between the means is then divided by the standard error to arrive at t . The number of degrees of freedom available for evaluating t when the two groups have been matched on only one variable is equal to $N_1 + N_2 - 3$ (96).⁸

3. EXPERIMENTS INVOLVING INDEPENDENT GROUPS

If we have two groups in which the subjects are not matched or paired upon any basis so that there is no reason to believe the means of the samples are correlated,⁹ then the formula for the standard error of the mean difference becomes

$$\sigma_{m_d} = \sqrt{\sigma_{m_1}^2 + \sigma_{m_2}^2} \quad (50)$$

⁸ It is possible, of course, to match the groups upon more than one variable. Professor Walker (96) points out that if " r is a multiple correlation of x against k other traits, there would be $(N_1 - 1) + (N_2 - 1) - k = N_1 + N_2 - (k + 2)$ degrees of freedom. Where r is a correlation of zero order, of course this becomes $N_1 + N_2 - 3$."

Whether much is to be gained, statistically, from the use of several variables for matching purposes, however, has been questioned by McNemar, who makes the point that "the efficacy of using additional controls is somewhat limited by the well-known fact that the increase in the multiple correlation coefficient resulting from adding more variables is usually slow. That this phenomenon of diminishing returns, associated with the problem of multiple correlation should be operative here has probably not been suspected by experimentalists" (68, p. 357).

⁹ The correlation between means of pairs of samples is, of course, unknown, since we ordinarily have but one pair of means and a correlation coefficient cannot be computed for one pair of observations. Fortunately, the correlation between means can be estimated by the correlation between the two sets of individual measurements in the two samples. It is the latter coefficient which we have used in the standard error formula.

To illustrate the application of formula (50) let us suppose that we are testing the difference between the means of a random sample of Democrats and a random sample of Republicans on an attitude test. The essential data are

TABLE 35.—SUMMARY OF ATTITUDE TEST DATA FOR 50 DEMOCRATS AND 50 REPUBLICANS

	DEMOCRATS	REPUBLICANS
Number of cases.....	50	50
Mean of distribution.....	8.7	6.2
Standard deviation.....	1.4	2.1
Standard error of mean.....	.2	.3

given in Table 35. Substituting the necessary values, we obtain

$$\sigma_{m_d} = \sqrt{(.2)^2 + (.3)^2} = \sqrt{.04 + .09} = \sqrt{.13} = .36$$

$$t = \frac{M_1 - M_2}{\sigma_{m_d}} = \frac{8.7 - 6.2}{.36} = \frac{2.5}{.36} = 6.9$$

When formula (50) is used to compute the standard error of the mean difference, the number of degrees of freedom available for evaluating t becomes $(N_1 - 1) + (N_2 - 1)$, or $N_1 + N_2 - 2$. According to Table C, for 98 degrees of freedom a t of 2.63 is significant at the 1 per cent level. Our observed t of 6.9 is, therefore, highly significant according to the standards agreed upon. Can we make any inference concerning the limits within which the population mean difference may lie? We may determine the fiducial limits of the parameter at the 1 per cent level in the same manner as before. Thus:

1% Level

$$\frac{x}{\sigma_{m_d}} = \pm t$$

$$\frac{x}{.36} = \pm 2.63$$

$$x = (.36)(\pm 2.63)$$

$$x = \pm .95$$

We may infer, then, at the 1 per cent level of confidence, that the population mean difference is within the limits 1.55 and 3.45.

4. THE ADVANTAGES OF PAIRING OBSERVATIONS

For the experiment cited earlier on the influence of attitudes toward conditions of work as a factor influencing output, let us find the standard error of the mean difference, assuming no correlation, and see what happens to both the standard error of the mean difference and the test of significance based upon it. Substituting in formula (50), we get

$$\sigma_{m_d} = \sqrt{(.47)^2 + (.47)^2} = \sqrt{.4418} = .66$$

Note that this value (.66) is larger than the value (.21) we obtained by use of the proper formula which takes the correlation into consideration. Computing the t ratio, we find that our observed mean difference divided by the standard error of the difference is now $1/.66$ or 1.52. This value of t would not meet the requirements of significance regardless of the number of degrees of freedom available.

The degrees of freedom for evaluating this t are, however, greater than in the case where we dealt with pairs. Taking cognizance of the pairing or the correlation, the number of degrees of freedom is the number of pairs minus 1 or, in this experiment, 9. For 9 degrees of freedom, t must be at least 2.26 to be significant at the 5 per cent level. When we have not paired our subjects, then the number of degrees of freedom is the number of subjects in one group minus 1, plus the number of subjects in the second group minus 1 or, in this experiment, 18. For 18 degrees of freedom a t of 2.10 is significant at the 5 per cent level. We can now see that in order for a statistical advantage to result from pairing or matching when we are dealing with two groups of 10 cases each, assuming the mean difference remains the same, the reduction in the standard error must be at least enough to raise t from 2.10 to 2.26.

If we had matched our groups and the resulting correlation coefficient had not been large, to take the extreme case, zero, then the standard error of the difference would not have been reduced at all, thus

$$\sigma_{m_d} = \sqrt{\sigma_{m_1}^2 + \sigma_{m_2}^2 - 2r_{12} \sigma_{m_1} \sigma_{m_2}} = \sqrt{\sigma_{m_1}^2 + \sigma_{m_2}^2 - 0}$$

In effect then, our experiment would really* have lost in precision because we would have only 9 degrees of freedom for evaluating the obtained t instead of 18, and the greater the number of degrees of freedom we have available, the smaller the required value of t to be significant at the 5 or the 1 per cent level. On the other hand, if we had matched our groups so well that the correlation coefficient had been quite high, then the standard error of the difference would have been reduced sufficiently to offset the loss in degrees of freedom.

In the present experiment, if we had not taken the correlation element into consideration, we could not have rejected the null hypothesis. But let us hasten to emphasize that this would only mean that the value of t computed offered no basis for rejecting the null hypothesis and that this is entirely different from concluding that the population mean difference must, therefore, be zero. You must keep clearly in mind the nature of the hypothesis that is tested and also the nature of the inference it is possible to make upon the basis of the test of significance.

5. TESTING THE SIGNIFICANCE OF A PROPORTION

We can apply the t test to observed sample proportions or per cents also, in order to determine whether these statistics depart significantly from some given hypothesis about the population parameters. Suppose that in a poll of student opinion we found that a sample of 100 was divided 65 to 35, with 65 individuals favoring the issue on which they were polled and 35 opposing the issue. What is the null hypothesis applied to this problem? We might assume that there is no difference between the number favoring

and the number disapproving, i.e., that the division in the population is 50:50. Assuming this hypothesis to be true, we can then find the standard error of the proportion by formula (44), where $\sigma_p = \sqrt{\frac{pq}{N}}$. Thus if the hypothesis we set up is true, the standard error of the proportion would be $\sqrt{\frac{(.5)(.5)}{100}}$ or .05. To translate this into the standard error of a per cent we need only multiply by 100 to get 5 per cent. Our observed sample departs from the hypothetical value we set by 15 per cent. This is the difference which we divide by the standard error of the hypothetical per cent to find t . Carrying through our computations, we find t is equal to 15/5 or 3. According to Table C, a t of 2.63 is required at the 1 per cent level of confidence and our value of 3 is, therefore, very significant. We may reject the hypothesis of evenly divided opinion which we set up as the population division and infer that a majority does favor the issue.

We shall see later that the chi-square test of significance is particularly useful in dealing with problems such as the one just described. In order to apply the standard error formulas for proportions and per cents, the data must first be translated into proportions or per cents, whereas the chi-square test enables us to work directly with the data in their original form.

6. TESTING THE SIGNIFICANCE OF r

When N is large and the population value of the correlation coefficient is not excessively high, then the standard error of r is given by the following formula:

$$\sigma_r = \frac{1 - \bar{r}^2}{\sqrt{N - 1}} \quad (51)$$

where \bar{r} = the population value of the correlation coefficient
 N = the number of pairs of observations

Since the population value of the correlation coefficient is unknown, it has been the practice of some investigators to substitute the sample value of r for \bar{r} in the formula. The standard error of r thus derived has then been given an interpretation similar to that which we have given other standard errors. This practice, however, is not recommended. The reason is that the sampling distribution of r is normal only under the conditions stated above: when N is large and when the population correlation is not very high.

The sampling distribution of r based upon a small number of observations drawn from a population having an absolute value of r of .80, for example, is markedly skew. One reason for this is that we have placed a limitation on one end of the sampling distribution. If the population r is .80, then samples could vary from 1.00 to -1.00 , but they could exceed the population value by not more than .20 at one end of the distribution, whereas in the opposite direction they could deviate by as much as 1.80 from the value.

If, however, the number of pairs in the samples upon which the r 's are based were increased to, let us say, 300 pairs, then the restriction of unity at one end of the scale would no longer be an important determining factor in the sampling distribution. Samples of 300 pairs of observations, even when the population r is as high as .80, would not tend to range more than .05 on each side of the population value (86, pp. 131-132). But if the population value were .96 or higher, then the restriction would again be a factor to consider.

Even when the population r is zero, however, the sampling distribution of r for small samples departs from the normal form.¹⁰ Figure 14 shows the curves for samples of 8 pairs

¹⁰ Fisher (25) has introduced a transformation of r into another statistic which is known as z (not to be confused with the z mentioned earlier), the distribution of which approximates the normal distribution and which remains nearly constant in form despite changes in the population parameter. He has also developed a formula for the standard error of z which permits the testing of any hypothesis concerning the population value or which may be used to test the difference between two r 's obtained from independent samples. A table has been prepared by Lindquist (64) which facilitates the computation of z for a given value of r . These sources should be consulted for testing hypotheses of the nature just described.

of observations which were drawn from a population where the correlation was zero and from a population where the correlation was .80.

a. The direct computation of t . The hypothesis which we are probably most interested in testing, once we have obtained a given value of r , is the hypothesis that the true,

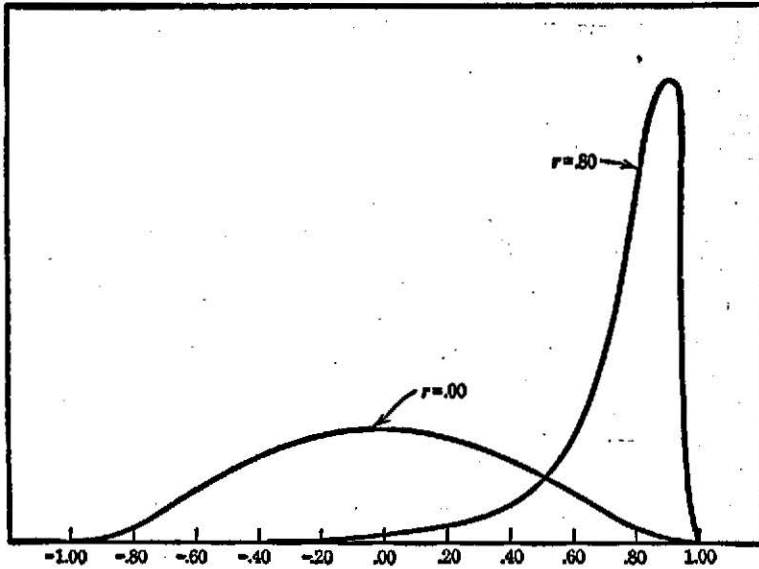


FIG. 14.—Sampling distribution of correlation coefficients for samples of eight pairs drawn from two populations having the indicated values of r . (Reproduced by permission from Fig. 7.4 in [86] with slight modifications.)

i.e., the population, r equals zero. If we set up this hypothesis for testing, assuming that our sample value is the result of sampling variation or chance, then the formula for t is

$$t = \left(\frac{r}{\sqrt{1-r^2}} \right) (\sqrt{N-2}) \quad (52)$$

where r = the observed sample value of the correlation coefficient

N = the number of pairs of observations in the sample

According to Fisher (25), the t calculated from formula (52) is distributed in accordance with the tabled values of t for degrees of freedom equal to $N - 2$. In other words, once we have obtained the value of t from the formula above, we may enter Table C, page 330, with degrees of freedom equal to the number of pairs of observations minus 2, to determine whether the obtained value is significant at the 5 or 1 per cent levels. Let us suppose, for example, that we obtained an r of .60 with 11 pairs of measurements. Substituting in formula (52) we get

$$\begin{aligned} t &= \left(\frac{.60}{\sqrt{1 - (.60)^2}} \right) (\sqrt{11 - 2}) \\ &= \left(\frac{.60}{\sqrt{.64}} \right) (\sqrt{9}) \\ &= \left(\frac{.60}{.80} \right) (3) \\ &= 2.25 \end{aligned}$$

Entering Table C with 9 degrees of freedom, we find that a t of 2.26 is required in order for us to reject the hypothesis tested at the 5 per cent level of confidence. Our obtained value of 2.25 is, therefore, not quite significant and, if we abide by the standards we have agreed upon, we would have to conclude that the hypothesis that the population r is zero is tenable.

b. The use of Table D. There is a much simpler method for finding out whether an observed value of r is sufficiently large to cause us to reject the hypothesis of zero correlation. Table D, page 331, gives the values of r which would be needed to meet the requirements of significance at the 5 and the 1 per cent levels for samples of various sizes. Table D is entered with degrees of freedom equal to $N - 2$, where N is the number of pairs. If we enter Table D with the 9 degrees of freedom available from the example described previously, we find that our r would need to be .602 to be

significant at the 5 per cent level. Our sample value of .60 is just slightly below the value required for significance. This is precisely the same conclusion we arrived at by using formula (52) and entering Table C to evaluate the obtained value of t . Obviously, the use of Table D is an easier method for testing the significance of an obtained r .

It should be evident from Table D that small r 's may be significant when they are based on a large number of pairs of observations, whereas large values of r may not be significant when based on a small number of observations. An r of .55 based upon 10 pairs of observations, for example, may be expected to occur quite frequently as a result of sampling variation, even when there is no correlation in the population from which the sample was drawn. The larger the value of N , on the other hand, then the less the value of the observed sample r need be in order to consider the hypothesis of zero correlation untenable.

THE t TEST OF SIGNIFICANCE

Example 1—Watson (98) has reported the following data concerning the performance of eighth-grade and tenth-grade students on a test of musical meanings. Is the difference between the means significant?

GROUP	N	MEAN	STANDARD DEVIATION
Eighth-grade students.....	200	90.76	19.32
Tenth-grade students.....	200	99.32	18.36

Example 2—The following measurements have been made on an experimental variable for individuals who have been matched on some criterion prior to the experiment proper.

GROUP	MATCHED PAIRS OF SUBJECTS									
	10	5	6	7	10	6	7	8	6	5
A	7	3	5	7	8	4	5	6	3	2
B	7	3	5	7	8	4	5	6	3	2

(a) Assuming that the groups are independent, test the difference between the means.

(b) Taking cognizance of the fact that the measurements have been paired, test the difference between the means. Does the new value of t change the conclusions you might have drawn from the test of significance in (a)?

Example 3.—Klineberg (54) has reported the mean scores of children in several European cities on a performance test of ability. Do these means differ significantly? To facilitate computations, assume that the sums of squares have been divided by $N - 1$ in computing the standard deviations.

CITY	N	MEAN	STANDARD DEVIATION
Paris.....	100	219.0	46.2
Hamburg.....	100	216.4	45.6
Rome.....	100	211.8	42.6

Example 4.—In a sample of 100 college students, 63 answered “yes” to the question: Would you make use of the library facilities if the building were open on Sunday afternoons? Do you have any confidence in the hypothesis that opinion is really evenly divided on the question?

Example 5.—Out of 200 adults polled, 110 announced that they approved of “singing commercials” on the radio. Only 90 disapproved of “singing commercials.” Is the hypothesis of evenly divided opinion tenable?

Example 6.—A random sample of 50 students showed that 35 of them were going to vote for Candidate A in a student election and 15 were going to vote for Candidate B. Are you confident that Candidate A will win the election?

Example 7.—In drawing a sample of 30 balls from a ballot box containing both black and white balls, a student obtained 22 black balls and 8 white. Is the hypothesis that the box contains an equal number of black and white balls tenable?

Example 8.—From another box containing white and black balls, the same student drew a sample of 50 and found 32 white and 18 black balls. Is the hypothesis that the box contains an equal number of black and white balls tenable?

Example 9.—An investigator reports an r of .88 for 10 pairs of measurements. Is the hypothesis that the population r is zero tenable?

Example 10.—Would a value of r of .33 for 10 pairs of measurements cause you to reject the null hypothesis? Why?

Example 11.—What value of r would you want to obtain before abandoning the null hypothesis for a sample of 50 pairs of measurements?

Example 12.—An investigator reports an r of .25 for a series of paired observations.

- (a) How large would his sample have to be before you would be willing to reject the null hypothesis?
- (b) What if he had reported an r of .55?

Example 13.—Lewis and Franklin (60) found that “task-oriented” subjects recalled a total of 118 interrupted and completed tasks at the end of an experimental session. The total was divided as follows: 75 interrupted tasks recalled and 43 completed tasks recalled. Someone proposes the hypothesis that this is just the result of sampling variation; that actually interrupted tasks are not recalled any more frequently than completed tasks. Could you offer any evidence to the contrary?

CHAPTER 10

ANALYSIS OF VARIANCE: INDEPENDENT GROUPS

The test of significance, t , developed in the last chapter is adequate for any experiment which involves only two groups and consequently a test of a single mean difference. But suppose that we had an experimental design involving three variables, for example, the performance of three groups, A, B, and C, under three differing sets of conditions. We could still use t to evaluate the differences between the means, by comparing A and B, B and C, and A and C. This seems a relatively simple procedure and it is, as long as there are not too many groups in our experiment. But if we had five groups, the number of comparisons we would have to make would be 10. And if we had ten groups, then the number of comparisons would be 45.¹ And we would have no assurance before going through all of the calculations involved that any single mean difference would be significant. Obviously, if we had some method of testing differences among all of the means at the same time, it would prove very valuable. Analysis of variance, and the corresponding test of significance, F , permits us to do just this.

1. NATURE OF ANALYSIS OF VARIANCE

Analysis of variance, as the name indicates, deals with variances rather than with standard deviations and standard errors. The variance of a sample, you may recall, is the standard deviation squared or $\frac{\sum x^2}{N}$, and the estimate of the

population variance is $\frac{\sum x^2}{N - 1}$. The rationale of analysis of

¹ The formula ${}_nC_r = \frac{n!}{(n-r)!r!}$ when r equals 2, simplifies to $\frac{n(n-1)}{2}$.

variance is that the total sum of squares of a set of measurements composed of several groups can be analyzed or broken down into specific parts, each part identifiable with a given source of variation. In the simplest case, the total sum of squares is broken down into two parts, a sum of squares based upon variation *within* the several groups and a sum of squares based upon variation *between* the group means. Then from these two sums of squares, *independent* estimates of the population variance are computed. On the assumption that the groups making up the total series of measurements are random samples from a homogeneous population, the two estimates of the population variance may be expected to differ only within the limits of chance fluctuations. This is the null hypothesis and it is tested by dividing the larger variance by the smaller variance to get the ratio of the variances. The 5 per cent and 1 per cent points of the variance ratio, which has been designated as F , have been tabled by Snedecor (86) and are reproduced in Table E, page 332. If the value of F exceeds the value at the level of significance agreed upon, then the null hypothesis—namely that there is no difference among the populations from which the samples have been drawn—is considered untenable. If we reject the null hypothesis, the populations from which the samples have been drawn may differ in terms of either means or variances or both. If the variances are approximately the same, then it is the means which differ. Since we are ordinarily concerned with differences in means, it is fortunate that, while possible, it is “unlikely in experimental data, that it is the variances which differ” (86, p. 188).²

This basically, then, is analysis of variance. Our first step will be to show that the total sum of squares for a series of measurements composed of several groups can be analyzed

² Just as the agricultural and biological data Snedecor has in mind may at times prove to be exceptions to this general principle, so may educational and psychological data. E. E. Cureton, for example, in a personal communication, reports that he has found such an exception in an analysis of test scores of people at different salary levels in certain occupations. Cf. also Lindquist (64, p. 99).

into the two parts mentioned, one part associated with variation within groups and the other with variation between group means. Let us take the data of Table 36. Assume

TABLE 36.—SCORES (X) AND SQUARES OF SCORES (X^2)
ON AN ACHIEVEMENT TEST FOR SUBJECTS TAUGHT
BY THE LECTURE AND THE PROJECT METHODS

LECTURE GROUP		PROJECT GROUP	
X	X^2	X	X^2
7	49	2	4
10	100	2	4
10	100	3	9
11	121	7	49
12	144	6	36
Σ 50	514	20	102

that the values given are scores on an achievement test for a group taught by the lecture method and another group taught by the project method.³

a. **The total sum of squares.** We first determine the total sum of squares by combining the scores of the two groups and treating them as one set of measurements. We could find the mean, which is 7, of the combined distribution, subtract this value from each of the scores, square the deviations and sum, to get the total sum of squares. Since the scores are rather small and few in number, however we shall apply the formula for the sum of squares, using the measures as they stand. Thus

$$\begin{aligned}\Sigma x^2 &= \Sigma X^2 - \frac{(\Sigma X)^2}{N} \\ &= 616 - \frac{(70)^2}{10}\end{aligned}$$

³ It is possible to make use of coding techniques before computing the various sums of squares needed in analysis of variance. The same coding constants, however, should be applied to every measure in the combined series. If this is done, then no corrections for coding are necessary.

$$\begin{aligned}
 &= 616 - \frac{4,900}{10} \\
 &= 616 - 490 \\
 &= 126
 \end{aligned}$$

b. The sum of squares within groups. Now let us find the sum of squares *within* each group. That is, considering each group separately, we find the mean of each group and the sum of squared deviations within each group from its own mean. Again we shall use the formula for the scores as they stand.

Lecture Group	Project Group
$\Sigma x_1^2 = \Sigma X_1^2 - \frac{(\Sigma X_1)^2}{n_1}$	$\Sigma x_2^2 = \Sigma X_2^2 - \frac{(\Sigma X_2)^2}{n_2}$
$= 514 - \frac{(50)^2}{5}$	$= 102 - \frac{(20)^2}{5}$
$= 514 - \frac{2,500}{5}$	$= 102 - \frac{400}{5}$
$= 514 - 500$	$= 102 - 80$
$= 14$	$= 22$

The sum of these two sums of squares ($14 + 22 = 36$) is called the sum of squares *within* groups. It obviously does not equal the total sum of squares (126). The reason is that for the total sum of squares the deviations were taken from the mean of the combined groups which was 7, whereas the sum of squares for the lecture group was computed from its own mean of 10, and the sum of squares for the project group was computed from its own mean, which was 4. If the means of the two groups had been equal, then the sum of squares computed within the lecture group plus the sum of squares computed within the project group would have equaled the total sum of squares.

c. The sum of squares between groups. Since the two means differ we may compute a second sum of squares based upon the variation of the group means. We find the mean

of the total which is 7 and the *deviation* of each of the group means from this value. We shall let d represent the deviation of a group mean from the mean of the total. Then $d_1^2 = (M_1 - M)^2$ and $d_2^2 = (M_2 - M)^2$. But since each of these squared deviations is based upon 5 cases, each one must be weighted or multiplied by n , the number of subjects in each group, in order to put them on a per individual measure basis. Thus the sum of squares based upon variation of group means of r groups will be equal to

$$n_1 d_1^2 + n_2 d_2^2 + n_3 d_3^2 + \dots + n_r d_r^2 \quad (53)$$

In the case at hand we have only two means, each based upon 5 cases. The deviation of the lecture mean from the total mean is 3, and the deviation of the project mean from the total is -3 . Thus the sum of squares is

$$(5)(3)^2 + (5)(-3)^2 = (5)(9) + (5)(9) = 90$$

The sum of squares between means (90) plus the sum of squares within groups (36) is now equal to the total sum of squares (126).

d. Generalized formula for r groups. In symbolic form, for the case at hand, we may now write

$$\Sigma x^2 = (\Sigma x_1^2 + \Sigma x_2^2) + (n_1 d_1^2 + n_2 d_2^2) \quad (54)$$

where Σx^2 = the total sum of squares for the combined distributions

$\Sigma x_1^2 + \Sigma x_2^2$ = the sum of squares within groups

$n_1 d_1^2 + n_2 d_2^2$ = the sum of squares between group means

Now, if we recall that the sum of squares for any given set of measurements can be stated in terms of $N\sigma^2$, since

$$\sigma = \sqrt{\frac{\Sigma x^2}{N}} \quad \text{by definition}$$

$$\sigma^2 = \frac{\Sigma x^2}{N} \quad \text{squaring both sides}$$

$$N\sigma^2 = \Sigma x^2 \quad \text{multiplying both sides by } N$$

then we may generalize formula (54) for more than two

groups and at the same time put it into an easily remembered form. We merely substitute the appropriate symbols, $n\sigma^2$, for the corresponding sums of squares within groups. Thus, if we have r groups

$$\Sigma x^2 = (n_1\sigma_1^2 + n_2\sigma_2^2 + n_3\sigma_3^2 + \dots + n_r\sigma_r^2) + (n_1d_1^2 + n_2d_2^2 + n_3d_3^2 + \dots + n_rd_r^2) \quad (55)$$

where Σx^2 = the total sum of squares based upon deviations from the mean of the combined measurements
 n = the number of cases within a given group
 σ = the standard deviation of a given group
 d = the deviation of a given group mean from the mean of the entire distribution

Thus from formula (55) above we see that the total sum of squares is equal to

$$\Sigma x^2 = \Sigma n\sigma^2 + \Sigma nd^2 \quad (56)$$

where Σx^2 = the total sum of squares
 $\Sigma n\sigma^2$ = the sum of squares within groups
 Σnd^2 = the sum of squares between group means

Formulas (55) and (56) make apparent what we have said before: That the total sum of squares can be analyzed into two parts: the $\Sigma n\sigma^2$ which is called the sum of squares within groups, and the Σnd^2 which is called the sum of squares between group means. Each of these sums of squares when divided by the appropriate number of degrees of freedom provides an *independent* estimate of the population variance. The number of degrees of freedom for the total sum of squares, we have already seen, is equal to $N - 1$ where N is the total number of cases in the combined groups. The number of degrees of freedom within each group is equal to $n - 1$ where n is the number of cases within each group. But since we have two groups, in this instance, the number of degrees of freedom is equal to $r(n - 1)$, where r is the number of groups. The number of degrees of freedom between groups is equal to $r - 1$, where r equals the number of groups. We may see these relations in Table 37.

TABLE 37.—ANALYSIS OF VARIANCE OF ACHIEVEMENT SCORES OF GROUPS TAUGHT BY THE LECTURE AND THE PROJECT METHODS

SOURCE OF VARIATION	SUM OF SQUARES	<i>df</i>	ESTIMATE OF VARIANCE
Between groups.....	90	1	90
Within groups.....	36	8	4.5
Total.....	126	9	

Degrees of Freedom

Between groups.....	$r - 1$
Within groups.....	$r(n - 1)$ or $N - r$
Total.....	$rn - 1$ or $N - 1$

e. The variance ratio. F , as you recall, is the ratio of the two estimates of the population variance, or

$$F = \frac{\text{larger variance}}{\text{smaller variance}} \quad (57)$$

In the present instance F is equal to $90/4.5$ or 20. To determine whether this F is significant at the 5 per cent or 1 per cent level, we enter the column of Table E (page 332) with the degrees of freedom of the larger variance ($df = 1$) and follow down to the row entry corresponding to the degrees of freedom of the smaller variance ($df = 8$). The values of F at the 5 per cent point are given in lightface type; the boldface type is the value at the 1 per cent point. These values for $df = 1$ and $=8$ are 5.32 at the 5 per cent level and 11.26 at the 1 per cent level. Our value of F , which is 20, is therefore highly significant, since 11.26 is the value of F which would be exceeded only 1 per cent of the time as a result of sampling variation if the null hypothesis were true.

Now since (i) the mean for the lecture group is higher than that for the project group, and since (ii) the difference between the means (between-groups variance) leads to a larger estimate of the population variance (is greater than

A Comparison of F and t in the Case of Two Groups 199

the within-groups variance) in (iii) a reliable fashion (the F ratio exceeds the 1 per cent level), therefore, (iv) we may be quite confident that the difference in achievement between the group taught by the lecture method and the group taught by the project method is indicative of a real difference, or, in other words, that achievement is reliably greater under the lecture method of instruction.

2. A COMPARISON OF F AND t IN THE CASE OF TWO GROUPS

You may ask whether we could not have obtained a similar test of the null hypothesis by the method described in the last chapter. The answer would be "Yes," and the results obtained with the t test would be identical with those we arrive at by means of the F test. But consider first another method of finding the standard error of the mean and the standard error of the mean difference.

You may recall that we said that the population standard deviation may be estimated from the sample standard deviation by $\sigma \sqrt{\frac{N}{N-1}}$ and, therefore, the population variance may be estimated by

$$\sigma^2 \frac{N}{N-1} = \left(\frac{\sum x^2}{N}\right) \left(\frac{N}{N-1}\right) = \frac{\sum x^2}{N-1}, \text{ in which case } \sigma_m^2 = \frac{\sum x^2}{N}$$

But if we have more than one group we may combine our data to arrive at another estimate of the population variance on the assumption that our two samples are random samples from the same population, or from populations having a common standard deviation as well as a common mean.⁴ We arrive at our estimate by pooling the sums of squares

⁴ The assumption that the populations are normally distributed is also involved. Tippett points out, however, that experience seems to indicate that "the tests are not very sensitive to moderate departures from normality nor to small differences in standard deviations" (90, p. 115). See also the discussion in Chapter 14, pp. 295-299.

computed within each of the groups along with the associated degrees of freedom. Thus for two groups

$$\sigma^2 = \frac{\Sigma x_1^2 + \Sigma x_2^2}{(N_1 - 1) + (N_2 - 1)} = \frac{\Sigma x_1^2 + \Sigma x_2^2}{N_1 + N_2 - 2} \quad (58)$$

$$\sigma_{m_1}^2 = \frac{\Sigma x_1^2 + \Sigma x_2^2}{N_1} \quad \text{and} \quad \sigma_{m_2}^2 = \frac{\Sigma x_1^2 + \Sigma x_2^2}{N_2} \quad (59)$$

$$\sigma_{m_d}^2 = \frac{\Sigma x_1^2 + \Sigma x_2^2}{N_1} + \frac{\Sigma x_1^2 + \Sigma x_2^2}{N_2} \quad (60)$$

Since the numerators of the two values in (60) are the same, we rewrite the formula for the standard error of the difference squared as

$$\sigma_{m_d}^2 = \left(\frac{\Sigma x_1^2 + \Sigma x_2^2}{N_1 + N_2 - 2} \right) \left(\frac{1}{N_1} + \frac{1}{N_2} \right) \quad (61)$$

If we have the same number of subjects in each group, then the standard error of each sample will be the same, and formula (61) may be simplified to

$$\sigma_{m_d} = \sqrt{2 \left(\frac{\Sigma x_1^2 + \Sigma x_2^2}{N_1 + N_2 - 2} \right) \left(\frac{1}{N_1} \right)} \quad (62)$$

Thus the standard error of the mean difference for the lecture group and the project group could be found directly from the sum of squares within each group, without first finding the standard deviations and the standard errors of the means. We merely need to substitute in formula (61) or formula (62) and get

$$\begin{aligned} \sigma_{m_d} &= \sqrt{2 \left(\frac{14 + 22}{5 + 5 - 2} \right) \left(\frac{1}{5} \right)} \\ &= \sqrt{2 \left(\frac{36}{8} \right) \left(\frac{1}{5} \right)} \\ &= \sqrt{2 \left(\frac{36}{40} \right)} \\ &= \sqrt{1.8} \\ &= 1.34 \end{aligned}$$

A Comparison of F and t in the Case of Two Groups 201

The difference between the means is 6, and t would be equal to $6/1.34$ or 4.48. The number of degrees of freedom is $(N_1 - 1) + (N_2 - 1)$, or 8, and for 8 degrees of freedom a t of 3.355 would be significant at the 1 per cent level. We would consider the null hypothesis, that there is no difference between the means, as untenable and reach, therefore, the very same conclusion that we reached by means of the F test. As a matter of fact, when only two groups are involved, F is equal to t^2 . That is to say, t which is 4.48 (rounded), when squared is 20 (rounded), which is the value of F we obtained. Thus you can see that there is no advantage in using F rather than t when testing the difference between only two means. It is only when we have a number of groups and consequently a number of means involved that F can be used to advantage. F provides us with an over-all test of significance among a number of different means. If F meets the level of significance we have adopted, we may then make specific comparisons with the t test.

What if the value of F which is obtained fails to be significant? May we then use the t test to test the differences between the pairs of means? Let us suppose that we had 10 sample means and found that F was not significant at the 5 per cent level. Now let us suppose that we tested, by means of t , the difference between the *largest* and the *smallest* mean in the group of 10. The value of t thus obtained may greatly exceed the tabled value at the 5 or the 1 per cent points. Could we conclude that the F test and the t test are inconsistent: that the former shows no significant differences, yet the latter does? When we examine the hypothesis tested by the t test we see the fallacy involved in this comparison. The hypothesis tested by t is that the two sample means have been drawn *at random* from the same population. We have *selected* the largest and the smallest for the comparison. The difference we are comparing is but 1 out of 45 possible comparisons which might be made.⁵ Fisher (26), although warning that com-

⁵ Determined by the formula for combinations, $n(n - 1)/2$.

parisons suggested *after* the data are in are open to suspicion, proposes that under these circumstances instead of demanding that the probability be 1 in 20 (5 per cent level) that it be 1 in $(45)(20) = 1$ in 900. In other words, t , in this particular example, would have to be equal to a value that could be expected to occur as a result of sampling variation but 1 time in 900 rather than 1 time in 20. Fisher contends, however, that it would be better to regard such unforeseen comparisons "only as suggestions for future experimentation, in which they can be deliberately tested" (26, p. 57).

3. THE COMPARISON OF THREE GROUPS

Let us now introduce a third group, taught by the discussion method, into our experiment. The scores on the achievement test for this group are given in Table 38, where,

TABLE 38.—SCORES (X) AND SQUARES OF SCORES (X^2) ON AN ACHIEVEMENT TEST FOR SUBJECTS TAUGHT BY THE LECTURE, DISCUSSION, AND PROJECT METHODS

LECTURE GROUP		DISCUSSION GROUP		PROJECT GROUP	
X	X^2	X	X^2	X	X^2
7	49	4	16	2	4
10	100	6	36	2	4
10	100	7	49	3	9
11	121	9	81	7	49
12	144	9	81	6	36
Σ 50	514	35	263	20	102

for the sake of comparison and easy examination, we repeat the scores of the other two groups.

a. The total sum of squares. The total sum of squares—we are dealing now with all of the groups combined—could be found by finding the mean of the total distribution and subtracting this value from each of the 15 scores and squar-

ing the resulting deviations. We shall, however, use the following formula for treating the measures as they stand

$$\begin{aligned}\Sigma x^2 &= \Sigma X^2 - \frac{(\Sigma X)^2}{N} \\ &= 879 - \frac{(105)^2}{N} \\ &= 879 - \frac{11,025}{15} \\ &= 879 - 735 \\ &= 144\end{aligned}$$

b. The sum of squares between groups. To find the sum of squares between group means we must first find the mean of the entire distribution. This will be equal to $\frac{\Sigma X}{N}$. Thus

$$M = \frac{50 + 35 + 20}{15} = \frac{105}{15} = 7$$

We then subtract 7 from each of the group means to get d , square each of these deviations, and weight each squared deviation by n , the number of cases within each group. Thus

$$\begin{aligned}\Sigma nd^2 &= n_1 d_1^2 + n_2 d_2^2 + n_3 d_3^2 \\ &= (5)(3)^2 + (5)(0)^2 + (5)(-3)^2 \\ &= 45 + 0 + 45 \\ &= 90\end{aligned}$$

Another method which is convenient for finding the sum of squares between groups is to work directly with the *sum of scores* for each group. If we square the sum of scores for each group and divide each of these values by the number of cases on which the sum is based, then we need only to apply a correction term for origin to get the sum of squares. Thus

$$\Sigma nd^2 = \frac{(\Sigma X_1)^2}{n_1} + \frac{(\Sigma X_2)^2}{n_2} + \frac{(\Sigma X_3)^2}{n_3} + \dots + \frac{(\Sigma X_r)^2}{n_r} - \frac{(\Sigma X)^2}{N} \quad (63)$$

where

Σnd^2 = the sum of squares between groups
 $\Sigma X_1, \Sigma X_2, \Sigma X_3, \Sigma X_r$ = the sum of scores or measurements for each of the r groups

n_1, n_2, n_3, n_r = the corresponding number of cases in the various groups

$\frac{(\Sigma X)^2}{N}$ = the correction term for origin or the total sum of scores for all of the groups, squared, and divided by the total N .

Substituting in formula (63), we may obtain the sum of squares between means for the case at hand. Thus

$$\begin{aligned}\Sigma nd^2 &= \frac{(50)^2}{5} + \frac{(35)^2}{5} + \frac{(20)^2}{5} - \frac{(105)^2}{15} \\ &= \frac{2,500 + 1,225 + 400}{5} - \frac{11,025}{15} \\ &= \frac{4,125}{5} - 735 \\ &= 825 - 735 \\ &= 90\end{aligned}$$

which is the same value that we obtained by the other method of computation. Note also that the "correction term," $\frac{(\Sigma X)^2}{N}$, is the same correction for origin that is used to find the total sum of squares.

c. The sum of squares within groups. We still have to obtain the sum of squares within groups. This we do by adding the sum of squares computed for each of the three groups when considered separately. We already have the sum of squares for the lecture and the project groups and we can find the sum of squares for the discussion group in the customary fashion. Thus

$$\begin{aligned}\Sigma x_3^2 &= \Sigma X_3^2 - \frac{(\Sigma X_3)^2}{n_3} \\ &= 263 - \frac{(35)^2}{5}\end{aligned}$$

$$\begin{aligned}
 &= 263 - \frac{1,225}{5} \\
 &= 263 - 245 \\
 &= 18
 \end{aligned}$$

The sum of squares within groups is thus equal to $14 + 18 + 22$ or 54 . This is a good place to point out that the direct computation of the sum of squares within groups is not necessary. We have merely done so up to this point in order to illustrate the source of this sum of squares. Actually, if our other computations are correct, the sum of squares within groups can be obtained by subtraction. The reason for this is that the sum of squares within groups plus the sum of squares between groups must equal the total sum of squares. Consequently, if we have the sum of squares between groups, we can subtract this value from the total sum of squares in order to get the sum of squares within groups. For example, the total sum of squares is equal to 144 , and between groups the sum of squares is 90 , and $144 - 90$ equals 54 or the sum of squares within groups.

d. The variance ratio. The results of our various computations are summarized in Table 39, where the two in-

TABLE 39.—ANALYSIS OF VARIANCE OF ACHIEVEMENT SCORES OF GROUPS TAUGHT BY THE LECTURE, DISCUSSION, AND PROJECT METHODS

SOURCE OF VARIATION	SUM OF SQUARES	df	ESTIMATE OF VARIANCE
Between groups.....	90	2	45
Within groups.....	54	12	4.5
Total.....	144	14	

Degrees of Freedom

Between groups.....	$r - 1$
Within groups.....	$r(n - 1)$ or $N - r$
Total.....	$rn - 1$ or $N - 1$

dependent estimates of the population variance have been made by dividing the sum of squares between groups and the sum of squares within groups by the appropriate degrees of freedom. F is the larger variance divided by the smaller variance or $\frac{45}{4.5} = 10$. We enter the column of Table E (page 332) with the 2 degrees of freedom of the larger variance and run down the column until we come to the row entry corresponding to the 12 degrees of freedom of the smaller variance. The value of F at the 1 per cent point is 6.93. If the null hypothesis, that the groups are random samples from the same population, is true, then values of F as large as 6.93 or larger would be expected to occur as a result of sampling variation less than 1 per cent of the time. Since our obtained value of F , 10, greatly exceeds the 1 per cent point, we may reject the null hypothesis with a great deal of confidence. Consequently, we may infer that the differences in achievement between the three groups taught by different methods of instruction are indicative of real differences.

We should note that the F test, although permitting us to infer that there are significant differences between the groups, does not specify that each group differs significantly from each of the others. It may be that only the difference between the lecture group ($M = 10$) and the project group ($M = 4$) is significant and that the discussion group ($M = 7$) does not differ significantly from either the lecture or the project group. The F test is an over-all test, as we have pointed out before. To determine whether any particular mean difference is significant or not, we would need to test the null hypothesis applied to the mean difference by the t test.

We summarize the computations needed for a simple case of analysis of variance in Table 40. The necessary formulas and methods of determining the appropriate degrees of freedom are included also for convenient reference.

TABLE 40.—SUMMARY OF COMPUTATIONS IN ANALYSIS OF VARIANCE FOR r GROUPS WITH n INDEPENDENT SUBJECTS IN EACH GROUP—TOTAL SUM OF SQUARES ANALYZED INTO TWO PARTS

INDIVIDUAL	MEASUREMENTS			
	Group 1	Group 2	Group 3	Group r
a	X_{1a}	X_{2a}	X_{3a}	X_{ra}
b	X_{1b}	X_{2b}	X_{3b}	X_{rb}
c	X_{1c}	X_{2c}	X_{3c}	X_{rc}
\vdots	\vdots	\vdots	\vdots	\vdots
n	X_{1n}	X_{2n}	X_{3n}	X_{rn}
Sum of columns....	ΣX_1	ΣX_2	ΣX_3	ΣX_r

Computations:

1. Total sum of scores = $\Sigma X_1 + \Sigma X_2 + \Sigma X_3 + \dots + \Sigma X_r$
2. Correction for origin = $\frac{(\Sigma X)^2}{N}$
3. Total sum of squares = $\Sigma X^2 - \frac{(\Sigma X)^2}{N}$
4. Sum of squares between groups

$$= \frac{(\Sigma X_1)^2 + (\Sigma X_2)^2 + (\Sigma X_3)^2 + \dots + (\Sigma X_r)^2}{n} - \frac{(\Sigma X)^2}{N}$$
5. Sum of squares within groups = Total - Between groups

Degrees of freedom:

1. Between groups = $r - 1$
2. Within groups = $r(n - 1)$ or $N - r$
3. Total = $rn - 1$ or $N - 1$

4. A MORE COMPLEX ANALYSIS

Let us now consider a somewhat more complicated application of analysis of variance. Suppose that we wished to study simultaneously the interaction of two or more variables, each varying in several ways. Specifically, we might be interested in the differential effects of three methods of

instruction (the lecture method, the discussion method, and the project method) upon three different types of achievement as measured by three different but comparable tests (a test of factual information, a test of understanding of general principles, and a test of ability to make applications). The questions which we might be interested in answering by experiment might be these: which of the three methods of instruction will result in the greatest over-all achievement, that is to say, on the combined tests? Will achievement be greater in the area of facts, applications, or principles? Is achievement in each area independent of method of instruction or will achievement in the various areas be dependent upon the type of instruction?

For purposes of illustration, let us assume that we have 45 subjects and that each subject is assigned at random to one of the nine experimental conditions of Table 41.

TABLE 41.—EXPERIMENTAL DESIGN FOR STUDYING THE INFLUENCE OF THREE DIFFERENT METHODS OF INSTRUCTION UPON THREE DIFFERENT KINDS OF ACHIEVEMENT

TYPE OF ACHIEVEMENT	METHOD OF INSTRUCTION			TOTAL
	Lecture	Discussion	Project	
Facts.....	<i>a</i>	<i>b</i>	<i>c</i>	$(a + b + c)$
Principles....	<i>d</i>	<i>e</i>	<i>f</i>	$(d + e + f)$
Applications..	<i>g</i>	<i>h</i>	<i>i</i>	$(g + h + i)$
Total....	$(a + d + g)$	$(b + e + h)$	$(c + f + i)$	$(a + b + c + d + e + f + g + h + i)$

With the border totals alone, we would have three comparisons to make for achievement and three comparisons to make for methods of instruction. If we compared every cell in the table, i.e., every experimental condition with every other experimental condition, we would have 36 additional comparisons to make. Since we do not know whether any of these differences are significant or not, we

shall make over-all comparisons first by means of the *F* test. We may then make the specific comparisons if *F* is significant. The results of the outcomes on the various achieve-

TABLE 42.—SCORES ON THREE DIFFERENT MEASURES OF ACHIEVEMENT FOR GROUPS TAUGHT BY THREE DIFFERENT METHODS OF INSTRUCTION

TYPE OF ACHIEVEMENT	IND.	METHOD			SUM AND MEAN FOR ACHIEVEMENT
		Lecture	Discussion	Project	
Facts	1	7	4	2	
	2	10	6	2	
	3	10	7	3	
	4	11	9	7	
	5	12	9	6	
	Σ Mean		50 10	35 7	20 4
Principles	1	6	10	5	
	2	5	10	4	
	3	8	11	7	
	4	9	11	8	
	5	12	13	11	
	Σ Mean		40 8	55 11	35 7
Applications	1	3	4	7	
	2	3	6	9	
	3	4	7	9	
	4	8	8	10	
	5	7	10	10	
	Σ Mean		25 5	35 7	45 9
Sum for Method.....		115	125	100	340
Mean for Method.....		7.67	8.33	6.67	7.56

ment tests for each subject are given in Table 42, We proceed with the calculation of the sums of squares in the manner already familiar.

Sum of squares:

$$\begin{aligned} 1. \text{ Total} &= (7)^2 + (10)^2 + (10)^2 + \dots + (10)^2 - \frac{(340)^2}{45} \\ &= 2,938 - 2,568.89 \\ &= 369.11 \end{aligned}$$

$$\begin{aligned} 2. \text{ Between groups} &= \frac{(50)^2}{5} + \frac{(40)^2}{5} + \dots + \frac{(45)^2}{5} - \frac{(340)^2}{45} \\ &= 2,770 - 2,568.89 \\ &= 201.11 \end{aligned}$$

$$\begin{aligned} 3. \text{ Within groups} &= \text{Total} - \text{Between groups} \\ &= 369.11 - 201.11 \\ &= 168 \end{aligned}$$

Before analyzing further the sum of squares between groups, let us test for the significance of the differences between groups (cells). Table 43 summarizes the data.

TABLE 43.—ANALYSIS OF VARIANCE OF SCORES ON THREE DIFFERENT MEASURES OF ACHIEVEMENT FOR GROUPS TAUGHT BY THREE DIFFERENT METHODS OF INSTRUCTION

SOURCE OF VARIATION	SUM OF SQUARES	df	ESTIMATE OF VARIANCE
Between groups.....	201.11	8	25.14
Within groups.....	168.00	36	4.67
Total.....	369.11	44	

Degrees of Freedom

$$\begin{aligned} \text{Between groups.....} & r - 1 \\ \text{Within groups.....} & r(n - 1) \text{ or } N - r \\ \text{Total.....} & rn - 1 \text{ or } N - 1 \end{aligned}$$

F is equal to the larger variance divided by the smaller variance or $\frac{25.14}{4.67} = 5.38$. We enter the column of the table of

F with the 8 degrees of freedom of the larger variance and find the row entry corresponding to the 36 degrees of freedom of the smaller variance and find that an F of 3.04 will be significant at the 1 per cent level. According to the

standards agreed upon, our obtained F of 5.38 is highly significant. We find the null hypothesis untenable, since if there were no differences in the populations the divergence between our estimates of the variance would occur as a result of sampling variation less than 1 per cent of the time. Hence we infer that the observed differences between our groups are not the result of chance.

But the information we have at the present time is not entirely satisfactory. We are pretty confident that there are differences between the nine experimental groups, but what about differences in type of achievement? And are the methods of instruction equally effective as far as total achievement is concerned? Or is one method more effective with one type of achievement while another method of instruction is more effective with another kind of achievement? Let us analyze the sum of squares between groups to see if we can get any additional information which would assist us in answering these questions.

We may compute a sum of squares for achievement, by squaring the sum of scores for each type of achievement, dividing each of these values by the number of cases on which it is based, and then subtracting the correction term for origin. In a similar manner we may compute a sum of squares for methods by squaring the sum of scores for each method, dividing each of these squares by the number of cases on which the sum is based, and then subtracting the correction term for origin. Thus

Sum of squares:

$$\begin{aligned} 1. \text{ Achievement} &= \frac{(105)^2}{15} + \frac{(130)^2}{15} + \frac{(105)^2}{15} - \frac{(340)^2}{45} \\ &= 2,596.67 - 2,568.89 \\ &= 27.78 \end{aligned}$$

$$\begin{aligned} 2 \text{ Methods} &= \frac{(115)^2}{15} + \frac{(125)^2}{15} + \frac{(100)^2}{15} - \frac{(340)^2}{45} \\ &= 2,590 - 2,568.89 \\ &= 21.11 \end{aligned}$$

The sum of these two sums of squares ($27.78 + 21.11$) which is equal to 48.89 does not equal the sum of squares between groups; we have a remainder or residual. We shall call this residual the sum of squares for *interaction*. It is found by subtraction. Thus

$$\begin{aligned} 3. \text{ Interaction} &= \text{Between groups} - (\text{Methods} + \text{Achievement}) \\ &= 201.11 - (21.11 + 27.78) \\ &= 152.22 \end{aligned}$$

Let us see what we have accomplished. First we analyzed the total sum of squares into two parts, one part associated with variation between each of the cells or groups of Table 42, the second part associated with variation within each of the groups. We then proceeded to analyze further the sum of squares between groups. One part can be traced to variation between methods of instruction, another to variation between types of achievement. The third, or remainder, is called interaction, since it is the result of the joint effect of a particular method of instruction and a particular kind of achievement.

We summarize the results of our analysis in Table 44, showing what has happened to the total sum of squares and how the total number of degrees of freedom has been partitioned. Note that we have 9 experimental groups with 5 subjects in each group. Consequently, we have 4 degrees of freedom within each of these groups or $(9)(4) = 36$ degrees of freedom within groups. In the previous analysis we had 8 degrees of freedom available for the sum of squares based upon differences between the 9 experimental groups. This made up our total of 44 degrees of freedom ($N - 1$). But we have further analyzed the sum of squares between groups into an achievement sum of squares, a methods sum of squares, and a residual or interaction sum of squares. And the 8 degrees of freedom must also be divided among these sums of squares. The methods sum of squares and the achievement sum of squares are based upon 3 groups each and consequently each of these

sums of squares will have 2 degrees of freedom. Thus, if 2 of the 8 degrees of freedom are allotted to methods and 2 to achievement, then we have a remainder of 4 degrees of freedom for the residual or interaction sum of squares. The degrees of freedom for interaction may also be obtained by multiplying the number of degrees of freedom allotted to methods by the number of degrees of freedom allotted to achievement as shown in Table 44.

TABLE 44.—FURTHER ANALYSIS OF VARIANCE OF SCORES ON THREE DIFFERENT MEASURES OF ACHIEVEMENT FOR GROUPS TAUGHT BY THREE DIFFERENT METHODS OF INSTRUCTION

SOURCE OF VARIATION	SUM OF SQUARES	df	ESTIMATE OF VARIANCE
Type of achievement....	27.78	2	13.89
Method of instruction...	21.11	2	10.56
Interaction.....	152.22	4	38.06
Within groups.....	168.00	36	4.67
Total.....	369.11	44	

Degrees of Freedom*

Achievement groups.....	$r_A - 1$
Method groups.....	$r_M - 1$
Interaction.....	$(r_A - 1)(r_M - 1)$
Within groups.....	$r_T(n - 1)$ or $N - r_T$
Total.....	$r_T n - 1$ or $N - 1$

- * r_T = the total number of experimental groups or 9
- r_A = the number of achievement groups or 3
- r_M = the number of methods groups or 3
- n = the number of subjects in each group or 5

If we divide the achievement, method, and interaction estimates of the population variance by the variance within groups, we get the following values of F :

Achievement:	$\frac{13.89}{4.67} = 2.97$	F at 1% point for df 2 and 36 = 5.25
Method:	$\frac{10.56}{4.67} = 2.26$	F at 1% point for df 2 and 36 = 5.25
Interaction:	$\frac{38.06}{4.67} = 8.15$	F at 1% point for df 4 and 36 = 3.89

Each of the above values of F must be evaluated according to the number of degrees of freedom involved in computing it. For achievement and method of instruction the degrees of freedom are the same, 2 and 36, but for interaction we have 4 and 36 degrees of freedom. The values of F at the 1 per cent point listed above were found by entering the column of Table E with the degrees of freedom of the larger variance and finding the row entry corresponding to the degrees of freedom for the smaller variance. The F ratios for achievement and method of instruction fail to meet the value of F at the 5 per cent level (3.26) also. Consequently, if we abide by our standards, we must regard the null hypothesis as tenable in each of these cases.

What conclusions can we now draw from the analysis? The failure of the F ratio for methods to meet the requirements of significance indicates that differences in total achievement of groups taught by different methods of instruction are not significant. Likewise we find that the F ratio for achievement is not significant, and hence we cannot say that our subjects tend to learn facts better than principles or applications. It is the highly significant F ratio for interaction that is of primary interest. How may we interpret this?

The interaction variance, as we have said before, is a product of the joint effect of method of instruction and type of achievement. The fact that it is significant indicates that the effectiveness of a particular method of instruction depends upon the kind of achievement we are interested in measuring. One method of instruction is, in other words, more effective with one kind of achievement than with another. Note that again the F test does not tell us specifically which method is most effective with which kind of achievement. To complete our analysis we should now proceed to try the t test with the various specific comparisons.

ANALYSIS OF VARIANCE: INDEPENDENT GROUPS

Example 1.—The following data consist of measurements of outcomes of an experiment involving a “control” group and an “experimental” group. The two groups were not matched in any manner.

Control group.....	9	10	20	14	18	5	8	11	12	13
Experimental group.....	21	19	18	13	15	20	22	25	17	10

- (a) Test the difference between the means, using the pooled sum of squares in computing the standard error of the mean difference.
- (b) Compute F for the same data.

Example 2.—~~The following data consist of samples selected from the population we used earlier to study the sampling distribution of means.~~ Assume that each value represents a score made by an individual assigned at random to one of five different experimental groups.

I	II	III	IV	V
58	9	4	7	1
55	9	3	5	9
60	1	4	5	0
67	0	2	4	9
60	1	2	9	1

- (a) Find the total sum of squares, the sum of squares between means, and the sum of squares within groups. ~~Work the data through first with actual deviation scores.~~ Is the null hypothesis tenable? *state*

(b) Code the scores by subtracting 60 from each one. Does this influence the values for the sums of squares? Is the value of F changed?

(B) *Is the null hypothesis tenable*
Example 3.—Assume that we are interested in studying differences in retention between groups which have been presented with material by different methods. We are also interested in studying the relative effectiveness of the method of presentation, as far as retention is concerned, at three different age levels. We have

30 subjects at each age level. Within each age level subjects have been assigned at random to one of the three methods groups. The hypothetical outcomes of our experiment are listed below.

AGE GROUPS	METHODS		
	I	II	III
I	8	9	5
	5	4	3
	6	8	7
	8	4	5
	9	3	3
	10	6	1
	9	7	5
	7	6	4
	8	7	3
	10	6	4
II	6	6	9
	1	5	8
	1	9	9
	6	7	11
	5	5	8
	4	6	11
	3	6	10
	6	5	7
4	4	8	
4	7	9	
III	7	7	5
	3	5	9
	6	8	8
	2	5	7
	3	5	7
	5	8	8
	3	6	6
	3	6	8
	4	6	5
	4	4	7

Find the total sum of squares, the sum of squares within groups, the sum of squares between methods groups, the sum of squares between age groups, and the sum of squares attributed to interaction. Make the various tests of significance and interpret your results.

CHAPTER 11

ANALYSIS OF VARIANCE: MATCHED GROUPS

You may recall that when we discussed the t test applied to groups that had been matched or in which the individuals had been paired, we were forced to modify the simple formula for the standard error of the mean difference to take into account the correlation. In the case of analysis of variance, where groups have been matched or individuals have been paired on some basis, we must do something very similar. As our illustration of the procedure to be followed we shall use the case of two matched groups, from the data of Table 34 in Chapter 9. Although we would not ordinarily use F to evaluate a single mean difference, this example will serve as an introduction to analysis of variance applied to several matched groups.

Group A, you recall, worked under the suggestion that noise facilitated performance, and Group B worked under the suggestion that noise resulted in a decrease in output. Individuals in the two groups had been matched on the basis of earlier performance. We found that the value of t , when we failed to take cognizance of the fact that the subjects had been paired, was only 1.52, or not significant at the 5 per cent level of confidence. On the other hand, when we took the pairing into consideration, we obtained a value of t equal to 4.76, a highly significant value. In the former case, we would have had to accept the null hypothesis as tenable, whereas, with the proper evaluation of the mean difference, the null hypothesis was rejected. The proper evaluation of the mean difference, in other words, determined the conclusion we could draw from our experiment.

We shall now see that if we were to apply the methods of analysis of variance described in the last chapter to this same set of data, we would find that F is not significant,

but is equal to the *nonsignificant* value of t^2 . That is, the F ratio obtained by dividing the estimate of the population variance derived from the sum of squares between the two groups by the estimate of the population variance based upon the sum of squares within groups should be equal to $(1.52)^2$. On the other hand, if we analyze the total sum of squares in the manner to be described in this chapter, the F ratio will be significant and equal to the *significant* value of t^2 or $(4.76)^2$.

1. ANALYSIS OF VARIANCE OF TWO MATCHED GROUPS

The necessary data are given in Table 45. We could, if we so desired, work with deviations from the actual means of the distributions. Since the scores have been coded by subtraction and are small in size, we shall work directly

TABLE 45.—COMPUTATION OF SUM OF SQUARES BASED UPON MEANS OF PAIRS AND SUM OF SQUARES BASED UPON MEANS OF GROUPS

PAIR	(1) GROUP A	(2) GROUP B	(3) SUM	(4) MEAN OF PAIR	(5) $(M_p - M_c)$ d	(6) d^2	(7) nd^2
1	2	1	3	1.5	-2.0	4.00	8.00
2	5	5	10	5.0	1.5	2.25	4.50
3	4	2	6	3.0	-.5	.25	.50
4	3	3	6	3.0	-.5	.25	.50
5	6	4	10	5.0	1.5	2.25	4.50
6	3	2	5	2.5	-1.0	1.00	2.00
7	6	5	11	5.5	2.0	4.00	8.00
8	4	3	7	3.5	.0	.00	.00
9	5	4	9	4.5	1.0	1.00	2.00
10	2	1	3	1.5	-2.0	4.00	8.00
Σ	40	30	70	38.00
Group Mean	4.0	3.0		3.5			
d	.5	-.5					
d^2	.25	.25					
nd^2	2.50	2.50					

with them as they stand, merely applying a correction term for origin.

Sum of squares:

$$1. \text{ Total} = (2)^2 + (5)^2 + (4)^2 + \dots + (1)^2 - \frac{(70)^2}{20}$$

$$= 290 - \frac{4,900}{20}$$

$$= 290 - 245$$

$$= 45$$

$$2. \text{ Between groups} = \frac{(40)^2}{10} + \frac{(30)^2}{10} - \frac{(70)^2}{20}$$

$$= \frac{1,600 + 900}{10} - \frac{4,900}{20}$$

$$= \frac{2,500}{10} - 245$$

$$= 250 - 245$$

$$= 5$$

$$3. \text{ Within} = \text{Total} - \text{Between groups}$$

$$= 45 - 5$$

$$= 40$$

We see from Table 46, where the results of our computations have been summarized, that using the methods of the

TABLE 46.—ANALYSIS OF VARIANCE OF SCORES OF GROUP A AND GROUP B INTO TWO PARTS

SOURCE OF VARIATION	SUM OF SQUARES	df	ESTIMATE OF VARIANCE
Between groups.....	5	1	5
Within groups.....	40	18	2.22
Total.....	45	19	

Degrees of Freedom

Between groups..... $r - 1$

Within groups..... $r(n - 1)$ or $N - r$

Total..... $rn - 1$ or $N - 1$

$$F = \frac{\text{larger variance}}{\text{smaller variance}} = \frac{5}{2.22} = 2.25$$

last chapter, analyzing the total sum of squares into only two parts, results in an F of 2.25. This value, as a result of errors of rounding, is not quite equal to $(1.52)^2$, the value of t^2 obtained by ignoring the correlation.

Note in Table 45 that if one member of a pair, assigned to one of the experimental groups, tends to have a high score, the corresponding member in the second group also tends to have a high score ($r = .90$). We anticipated that the previous level of performance of our subjects might be a factor influencing their performance under the experimental conditions and we attempted to control this variable by pairing the individuals in the two groups on the basis of their previous performance. In analyzing the total sum of squares into two parts, however, we have included all of the variation due to this factor in the within-groups sum of squares. As a result the value of F which we derived was underestimated in the same manner that t was underestimated when we failed to take into account the matching. We should take cognizance of this matching by calculating a sum of squares based upon differences *between pairs* which may then be subtracted from the within-groups sum of squares.

Since the scores in Table 45 may be classified by rows (pairs) as well as by columns (groups), the sum of squares *between pairs* may be computed in the same manner that we use to find the sum of squares between groups. We could find the sum of each pair and divide by 2 to find the mean of each pair. We could then subtract the mean of all 20 scores from each of the means for pairs, square the deviations, weight each squared deviation by multiplying by 2, the number of cases on which it is based, and then sum. This would give us a sum of squares based upon the variation of the means of pairs and the procedure we have used in computing it is simply the application of formula (53). The computations are clearly indicated in Table 45.

A simpler method for finding the sum of squares between

pairs, however, would be to work with the sum of scores for each pair and apply formula (63). Thus

$$\begin{aligned}
 \text{Between pairs} &= \frac{(3)^2}{2} + \frac{(10)^2}{2} + \frac{(6)^2}{2} + \dots + \frac{(3)^2}{2} - \frac{(70)^2}{20} \\
 &= \frac{9 + 100 + 36 + \dots + 9}{2} - \frac{4,900}{20} \\
 &= \frac{566}{2} - 245 \\
 &= 283 - 245 \\
 &= 38
 \end{aligned}$$

The number of degrees of freedom for this sum of squares is equal to the number of pairs minus 1.

If we now subtract the sum of squares between pairs from the sum of squares within groups, we are left with a remainder, or the *residual sum of squares*. Thus

$$\begin{aligned}
 \text{Residual} &= \text{Within groups} - \text{Between pairs} \\
 &= 40 - 38 \\
 &= 2
 \end{aligned}$$

This residual sum of squares is based upon variation remaining in the data that cannot be accounted for in terms of variation of the column means (between groups) and the row means (between pairs). Note that if we add the sum of squares between groups and the sum of squares between pairs and then subtract this value from the total sum of squares, we are left with the residual sum of squares. Thus

$$\begin{aligned}
 \text{Residual} &= \text{Total} - (\text{Between groups} + \text{Between pairs}) \\
 &= 45 - (5 + 38) \\
 &= 45 - 43 \\
 &= 2
 \end{aligned}$$

from which you may see that it would not have been necessary to calculate the within-groups sum of squares in order to find the residual sum of squares. This sum of squares is always found most easily, as shown above, by subtraction.

It is possible to calculate the residual sum of squares

directly, that is, by working with deviations rather than by subtraction, but this involves additional work. The residual sum of squares, as Snedecor (86) has pointed out, may be thought of as the result of the differences between a set of expected values and the actual observed values.

If we subtract 3.5, the mean of all of the scores, from each group mean and from the mean of each pair, we may represent these deviations by d . The expected value for each subject then becomes the mean of all the scores (3.5) plus the deviation of the mean of the group and the mean of the pair of which he is a member. The deviation of the mean of Group A = + .5, and the deviation of the mean of Group B = - .5. The deviation of the mean of the first pair = - 2.00, the deviation of the mean of the second pair = +1.5, the deviation of the mean of the third pair = - .5, and so on. The expected value for the first subject in the A group is thus equal to $3.5 + .5 - 2.0 = 2$. For the second subject in the A group, the expected value is equal to $3.5 + .5 + 1.5 = 5.5$. The expected values for the other subjects in the A group are found in the same manner. For the first subject in Group B, the expected value is equal to $3.5 - .5 - 2.0 = 1.00$. For the second subject in Group B, the expected value is $3.5 - .5 + 1.5 = 4.5$. In a similar manner, the expected values for the other subjects are found and have been entered in Table 47.

If we now take the difference between each of the observed scores and the corresponding expected values, square these differences, and sum them, we would get

$$(2.0 - 2.0)^2 + (5.0 - 5.5)^2 + \dots + (1.0 - 1.0)^2 = 2.00$$

which is the residual sum of squares. You will probably never want to use this method of calculation, however, since the residual sum of squares can be found so easily by subtraction.

The residual sum of squares, when divided by the appropriate number of degrees of freedom, provides us with an estimate of the population variance. The number of

TABLE 47.—OBSERVED AND EXPECTED SCORES FOR SUBJECTS IN GROUP A AND IN GROUP B ARRANGED TO ILLUSTRATE THE DIRECT COMPUTATION OF THE RESIDUAL SUM OF SQUARES

PAIR	GROUP A		GROUP B		SUM	MEAN	<i>d</i>
	Observed	Expected	Observed	Expected			
1	2.0	2.0	1.0	1.0	3	1.5	-2.0
2	5.0	5.5	5.0	4.5	10	5.0	1.5
3	4.0	3.5	2.0	2.5	6	3.0	-.5
4	3.0	3.5	3.0	2.5	6	3.0	-.5
5	6.0	5.5	4.0	4.5	10	5.0	1.5
6	3.0	3.0	2.0	2.0	5	2.5	-1.0
7	6.0	6.0	5.0	5.0	11	5.5	2.0
8	4.0	4.0	3.0	3.0	7	3.5	.0
9	5.0	5.0	4.0	4.0	9	4.5	1.0
10	2.0	2.0	1.0	1.0	3	1.5	-2.0
Σ	40.0	30.0	70		
Mean	4.0	3.0		3.5	
<i>d</i>	.5	-.5			

degrees of freedom for this sum of squares is equal to $(r - 1)(n - 1)$ where r is the number of matched groups and n equals the number of cases in each group.¹ In the present problem, then, the number of degrees of freedom is equal to $(2 - 1)(10 - 1)$ or 9. The residual sum of squares, when divided by 9, the number of degrees of freedom, gives the estimate of the population variance which we use in the denominator of the formula for the F ratio to test the between-groups variance.²

The results of our computations are summarized in Table

¹The number of degrees of freedom can also be obtained by subtraction from the total or from the number allotted to within groups. Thus the number of degrees of freedom for the residual sum of squares is

$$df_{res} = \text{Total} - (\text{Between groups} + \text{Between pairs})$$

$$= 19 - (1 + 9)$$

$$= 9$$

and also

$$df_{res} = \text{Within groups} - \text{Between pairs}$$

$$= 18 - 9$$

$$= 9$$

²On the basis of the assumptions noted in Chapter 10.

48. Let us see what we have accomplished. In this analysis the total sum of squares has been broken down into three parts with accompanying degrees of freedom. The variation between group means was, of course, our major interest. We wanted to know whether the difference was significant or not in mean performance score between a group working under the suggestion that noise was a hindrance and a

TABLE 48.—ANALYSIS OF VARIANCE OF SCORES OF GROUP A AND GROUP B INTO THREE PARTS

SOURCE OF VARIATION	SUM OF SQUARES	df	ESTIMATE OF VARIANCE
Between groups.....	5	1	5
Between pairs.....	38	9	4.22
Residual.....	2	9	.22
Total.....	45	19	

Degrees of Freedom

Between groups.....	$r - 1$
Between pairs.....	$n - 1$
Residual.....	$(r - 1)(n - 1)$
Total.....	$rn - 1$ or $N - 1$

group working under the suggestion that noise facilitated performance. But we also realized that the performance of our subjects under the experimental conditions might be related to performance prior to the experiment proper. We took this into account by pairing our subjects on the basis of previous performance and then "eliminating" the sum of squares attributable to this source of variation from the within-groups sum of squares. That is to say, the variation of the means of the groups and the means of the pairs has been taken into account and the residual variance is based upon whatever variation remains in the data.

The value of F which we obtain, using the residual variance in the denominator and the variance between groups in the numerator, is now equal to 22.73. This value corre-

sponds, within errors of rounding, to the value of t^2 , $(4.76)^2$, which we obtained when we took cognizance of the fact that the groups had been matched. Entering the column of Table E with the 1 degree of freedom corresponding to the variance estimated from the between-groups sum of squares and running down to the row entry for the 9 degrees of freedom of the residual variance, we find that F need be only 10.56 to be significant at the 1 per cent level. Since the value we have obtained is 22.52, we would reject the null hypothesis and infer that the difference between the two groups could not be the result of sampling variation.³

2. ANALYSIS OF VARIANCE OF SEVERAL MATCHED GROUPS

We are now ready to extend the method of analysis just described to the case of several matched groups. Let us suppose that we are interested in the effectiveness upon attitudes of various methods of presentation of propaganda material. We might, for example, be concerned with whether the same piece of propaganda presented by radio, by a face-to-face speech, by reading, and so forth, would be equally effective in modifying attitudes. Let us suppose that we have five different methods of presentation and that we have available twenty-five subjects, or five for each of the experimental groups. We could, of course, merely assign the subjects at random to one of the five experimental groups, but suppose that the degree of change in attitude as a result of the propaganda is related to the intensity of the attitude one has before being subjected to the propaganda. It is questionable as to whether an individual who has a very intense attitude, either positive or negative, will be influenced as much as an individual who has a relatively neutral attitude. If there is correlation between attitude prior to the experiment and attitude after being subjected

³ We could make a test of significance *between pairs* also, if we were interested. The F ratio, in this instance, would become the variance between pairs divided by the residual variance.

to the propaganda material, then we should take this into consideration in assigning our subjects to the various experimental groups. If we can control this factor, then we can reduce the sum of squares within groups as we did in the case of the two matched groups discussed previously.

If we have available scores on the attitude test for the twenty-five subjects prior to the experiment, then we may use these as a basis for equating our groups. We first divide our twenty-five subjects into five groups of five subjects each, assigning to each of the five groups subjects with approximately the same attitude test scores. These five groups we shall designate as the "attitude groups." To each experimental group we now assign, at random, one subject from each attitude group. This scheme will tend to balance the experimental groups so that each will have individuals with similar "intensities" of attitudes. That is to say, we may have some assurance that no one of the experimental groups has all of the subjects with high or low attitude test scores, but that each group has approximately the same range and mean.⁴

The fictitious attitude test scores of the subjects after the presentation of the propaganda material are given in Table 49. If we ignored the fact that our experimental groups had been matched, we would proceed with the customary analysis of the total sum of squares. We would find the part associated with differences between the means of the experimental groups (columns) and the part associated with differences within the experimental groups. Thus

Sum of squares:

$$\begin{aligned} 1. \text{ Total} &= (8)^2 + (7)^2 + (6)^2 + \dots + (8)^2 + (6)^2 - \frac{(200)^2}{25} \\ &= 1,678 - 1,600 \\ &= 78 \end{aligned}$$

⁴ That adequate matching would be virtually impossible with so few subjects and so many groups must be recognized. The experiment is hypothetical and is for purposes of illustration. It is the procedure that is important.

$$\begin{aligned}
 2. \text{ Between groups} &= \frac{(33)^2}{5} + \frac{(38)^2}{5} + \dots + \frac{(45)^2}{5} - \frac{(200)^2}{25} \\
 &= 1,620.8 - 1,600 \\
 &= 20.8
 \end{aligned}$$

$$\begin{aligned}
 3. \text{ Within groups} &= \text{Total} - \text{Between groups} \\
 &= 78 - 20.8 \\
 &= 57.2
 \end{aligned}$$

TABLE 49.—ATTITUDE TEST SCORES OF GROUPS SUBJECTED TO DIFFERENT METHODS OF PRESENTATION OF PROPAGANDA. INDIVIDUALS IN EACH GROUP MATCHED ACCORDING TO INTENSITY OF ATTITUDE PRIOR TO THE PRESENTATION

ATTITUDE LEVEL	EXPERIMENTAL GROUPS					SUM	MEAN
	1	2	3	4	5		
1	8	10	10	11	11	50	10
2	7	9	9	10	10	45	9
3	6	7	8	9	10	40	8
4	6	6	7	8	8	35	7
5	6	6	5	7	6	30	6
Σ	33	38	39	45	45	200	
Mean	6.6	7.6	7.8	9	9	8

The results of our computations, which are summarized in Table 50, show that F is equal to 1.82. According to Table E, the value of F at the 5 per cent level for 4 and 20 degrees of freedom is 2.87. We must assume, therefore, that the F we obtained could easily have resulted from sampling variation and that the null hypothesis is tenable.

In a similar manner we might test the differences between attitude levels, that is, between rows. The new sum of squares which we would need is

$$\begin{aligned}
 \text{Between rows} &= \frac{(50)^2}{5} + \frac{(45)^2}{5} + \dots + \frac{(30)^2}{5} - \frac{(200)^2}{25} \\
 &= 1,650 - 1,600 \\
 &= 50
 \end{aligned}$$

and the sum of squares within rows may be obtained by subtraction. Thus

$$\begin{aligned} \text{Within rows} &= \text{Total} - \text{Between rows} \\ &= 78 - 50 \\ &= 28 \end{aligned}$$

TABLE 50.—ANALYSIS OF VARIANCE INTO TWO PARTS OF ATTITUDE TEST SCORES OF EXPERIMENTAL GROUPS SUBJECTED TO DIFFERENT METHODS OF PRESENTATION OF PROPAGANDA

SOURCE OF VARIATION	SUM OF SQUARES	df	ESTIMATE OF VARIANCE
Between groups.....	20.8	4	5.2
Within groups.....	57.2	20	2.86
Total.....	78	24	

Degrees of Freedom

$$\begin{aligned} \text{Between groups.....} & r - 1 \\ \text{Within groups.....} & r(n - 1) \text{ or } N - r \\ \text{Total.....} & rn - 1 \text{ or } N - 1 \end{aligned}$$

$$F = \frac{\text{larger variance}}{\text{smaller variance}} = \frac{5.2}{2.86} = 1.82$$

TABLE 51.—ANALYSIS OF VARIANCE OF ATTITUDE TEST SCORES CLASSIFIED ACCORDING TO ROWS OF TABLE 49

SOURCE OF VARIATION	SUM OF SQUARES	df	ESTIMATE OF VARIANCE
Between rows.....	50	4	12.5
Within rows.....	28	20	1.4
Total.....	78	24	

Degrees of Freedom

$$\begin{aligned} \text{Between rows.....} & n - 1 \\ \text{Within rows.....} & n(r - 1) \text{ or } N - n \\ \text{Total.....} & nr - 1 \text{ or } N - 1 \end{aligned}$$

$$F = \frac{\text{larger variance}}{\text{smaller variance}} = \frac{12.5}{1.4} = 8.93$$

In Table 51 we have made the estimates of the population variance based upon these computations. F is 8.93, a highly significant value. We expected that it would be if our matching of the various groups were effective and if intensity of attitude prior to the experiment were related to subsequent scores.

The sum of squares within groups in this and in the previous analysis, however, includes a source of variation which we can control. We shall compute the residual sum of squares, which will enable us to hold the variation between rows constant while testing the columns, and to hold the columns constant while testing the rows. This new residual sum of squares is the same for both rows and columns and is found by subtraction. Thus

$$\begin{aligned} \text{Residual} &= \text{Total} - (\text{Between columns} + \text{Between rows}) \\ &= 78 - (20.8 + 50) \\ &= 7.2 \end{aligned}$$

We may now set up Table 52 and test the columns and rows, using the new estimate of the population variance based upon the residual sum of squares as the denominator in the formula for the F ratio. According to Table E, F

TABLE 52.—ANALYSIS OF VARIANCE INTO THREE PARTS OF ATTITUDE TEST SCORES OF TABLE 49

SOURCE OF VARIATION	SUM OF SQUARES	df	ESTIMATE OF VARIANCE
Between columns.....	20.8	4	5.2
Between rows.....	50	4	12.5
Residual.....	7.2	16	.45
Total.....	78	24	

Degrees of Freedom

Between columns.....	$r - 1$
Between rows.....	$n - 1$
Residual.....	$(r - 1)(n - 1)$
Total.....	$N - 1$

at the 1 per cent level for 4 and 16 degrees of freedom is 4.77. The F ratio for columns is $5.2/.45 = 11.56$; hence we may infer that the differences between the columns (methods of presentation) are indicative of real differences. Although we are not primarily interested in the differences between rows (attitude levels), we may test these for significance also. The F ratio for rows is $12.5/.45 = 27.78$, a highly significant value for 4 and 16 degrees of freedom, and we must conclude that the null hypothesis applied to rows is not tenable.

Perhaps you are wondering about the similarity between the sum of squares which we have called the "residual" in this chapter and the sum of squares which we called "interaction" in the last chapter. If so, then you may note that without replication, i.e., without more than one subject, within each of the "attitude levels" for each of the "experimental groups," we cannot evaluate the residual variance in the manner in which we evaluated the interaction variance in the last chapter. You may see this more clearly if you examine more closely the design of our experiment in Table 49.

If we had five subjects instead of one at each of the attitude levels for each of the experimental conditions, so that we would have twenty-five groups of five subjects each, we could take the deviations within each of these groups from the mean of the group to get a sum of squares within each of the twenty-five groups. The sum of these sums of squares, based upon deviations within each group from the group mean, would correspond to the sum of squares within groups which we used in the methods-achievement experiment, described in the last chapter, to evaluate the interaction variance. The residual variance in the present experiment, in other words, could be evaluated only if a number of subjects were included in each of the attitude levels for each of the experimental groups.

In Table 53 we have summarized the computations for analysis of variance into three parts. The necessary formulas and methods of determining the number of degrees

TABLE 53.—SUMMARY OF COMPUTATIONS IN ANALYSIS OF VARIANCE FOR *r* GROUPS WITH *n* MATCHED SUBJECTS IN EACH GROUP—TOTAL SUM OF SQUARES ANALYZED INTO THREE PARTS

INDIVIDUAL	MEASUREMENTS				SUM OF ROWS
	Group 1	Group 2	Group 3	Group <i>r</i>	
<i>a</i>	X_{1a}	X_{2a}	X_{3a}	X_{ra}	ΣX_a
<i>b</i>	X_{1b}	X_{2b}	X_{3b}	X_{rb}	ΣX_b
<i>c</i>	X_{1c}	X_{2c}	X_{3c}	X_{rc}	ΣX_c
⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮
<i>n</i>	X_{1n}	X_{2n}	X_{3n}	X_{rn}	ΣX_n
Sum of columns..	ΣX_1	ΣX_2	ΣX_3	ΣX_r	ΣX

Computations:

1. Total sum of scores = $\Sigma X_1 + \Sigma X_2 + \Sigma X_3 + \dots + \Sigma X_r$
2. Correction for origin = $\frac{(\Sigma X)^2}{N}$
3. Total sum of squares = $\Sigma X^2 - \frac{(\Sigma X)^2}{N}$
4. Sum of squares between columns

$$= \frac{(\Sigma X_1)^2 + (\Sigma X_2)^2 + (\Sigma X_3)^2 + \dots + (\Sigma X_r)^2}{n} - \frac{(\Sigma X)^2}{N}$$
5. Sum of squares between rows

$$= \frac{(\Sigma X_a)^2 + (\Sigma X_b)^2 + (\Sigma X_c)^2 + \dots + (\Sigma X_n)^2}{r} - \frac{(\Sigma X)^2}{N}$$
6. Residual sum of squares = Total - (Between columns + Between rows)

Degrees of freedom:

1. Between columns = $r - 1$
2. Between rows = $n - 1$
3. Residual = $(r - 1)(n - 1)$
4. Total = $N - 1$

of freedom available for each of the sums of squares have been included for convenient reference.⁵

3. CORRELATION RATIO AND ANALYSIS OF VARIANCE

a. The correlation ratio without bias. Peters and Van Voorhis (74) make a strong case for the use of the correlation ratio in place of analysis of variance techniques. The correlation ratio, however, as we computed it earlier, is influenced by the size of the sample and by the number of classes into which the sample is divided. But this bias can be corrected for by substituting the two estimates of the population variance in the formula for η .⁶ If η^2 has already been computed, and we assume that it has in cases of interest, then the correlation ratio without bias, designated as *epsilon* (ϵ) by T. L. Kelley, who developed the formula, is the square root of *epsilon-square* which is given by formula (64).

$$\epsilon_{yx}^2 = \frac{\eta_{yx}^2 (N - 1) - (k - 1)}{N - k} \quad (64)$$

where ϵ_{yx}^2 = epsilon-square or the square of the correlation ratio without bias

η_{yx} = the value of the correlation ratio

N = the total number of cases in the distribution

k = the number of columns in the correlation table

⁵ Lindquist (64) provides many illustrations of the use of analysis of variance in the field of educational research. Additional illustrations in educational and psychological research may be found in Baxter (5), Dunlap (11), Garrett and Zubin (32), Jackson (47), and Shen (83). The critical discussions by Grant (36) and Peters (73) should be consulted also. In addition to these references, the student would do well to read Snedecor (86). He has achieved a simplicity of presentation which is notable and to be commended. Tippett (90) and Goulden (35) are well worth reading. Kenney (53) approaches analysis of variance from a mathematical viewpoint and his text should be of interest to advanced students, as should the two classical publications of Fisher (25, 26). The article by Helen Walker (94), although technical, is a valuable aid to an understanding of the concept of degrees of freedom.

⁶ The rationale of this is developed in Peters and Van Voorhis (74, pp. 319-322) from which formula (64) is taken. A further correction for broad categories is suggested if the classes are continuous and groupings are large. This will probably not be true of most cases where analysis of variance techniques might be applied. If needed, however, tables to facilitate the corrections can be found in the reference cited.

b. Tables of epsilon-square. Peters and Van Voorhis have tabled the values of ϵ^2 at the 1 per cent and 5 per cent points, assuming the true correlation to be zero. Assuming the null hypothesis, that the population correlation is zero, Table F, pp. 336-339, gives the values of ϵ^2 which can be expected to occur by chance 5 per cent and 1 per cent of the time. The column of Table F is to be entered with 1 less than the number of columns in the correlation table from which the correlation ratio was computed, $(k - 1)$, and the value of ϵ^2 at the 5 or 1 per cent point is to be found by running down this column to the row entry corresponding to the total number of cases less the number of columns, $(N - k)$.

Before turning to the relationship between the correlation ratio without bias and analysis of variance, we may illustrate the use of Table F by an example. Let us suppose that we had found a correlation ratio of .60 from a correlation table consisting of 10 columns, and with a total N of 90 cases. Substituting in formula (64), we get

$$\begin{aligned} \epsilon_{yz}^2 &= \frac{\eta_{yz}^2 (N - 1) - (k - 1)}{N - k} \\ &= \frac{(.60)^2 (90 - 1) - (10 - 1)}{90 - 10} \\ &= \frac{(.36) (89) - 9}{80} \\ &= \frac{32.04 - 9}{80} \\ &= \frac{23.04}{80} \\ &= .288 \\ \epsilon_{yz} &= .54 \end{aligned}$$

If we now assume the null hypothesis, that the population correlation is zero, we may test the significance of the observed value of ϵ^2 , .288, to find out whether it would occur

as a result of sampling variation less than 5 per cent of the time. If our observed value is larger than the tabled value of ϵ^2 at the 5 per cent point, then we may reject the null hypothesis and infer that the population correlation is not zero. We enter the column of Table F (page 336) with $(k - 1) = 9$, and run down to the row entry corresponding to $(N - k) = 80$. The value of ϵ^2 at the 5 per cent point is .091, and at the 1 per cent point it is .142. Since our observed value exceeds the value of ϵ^2 at the 1 per cent point, we may reject the hypothesis that the population correlation is zero with a high degree of confidence. If the hypothesis of zero correlation were true, then we would get such a large value of ϵ^2 much less than 1 per cent of the time. Thus, you see that the table of ϵ^2 is used in much the same fashion that the table of F is used.

c. **Epsilon-square and analysis of variance.** In the simple analysis of variance experiment, where the total sum of squares is analyzed into two parts, ϵ^2 is found readily. By formula it is

$$\epsilon_{yx}^2 = 1 - \frac{v_w}{v_t} \quad (65)$$

where ϵ_{yx}^2 = epsilon-square
 v_w = the population variance estimated from the within-groups sum of squares
 v_t = the population variance estimated from the total sum of squares (the sum of squares for total divided by $N - 1$)

To illustrate formula (65) we may take the data of Table 39, where we found that the estimate of the population variance based upon variation within groups was 4.5. The total sum of squares was equal to 144 and this divided by $N - 1$ or 14 gives us 10.29, the denominator for formula (65). Solving for ϵ^2 , we get

$$\epsilon_{yx}^2 = 1 - \frac{v_w}{v_t} = 1 - \frac{4.5}{10.29} = 1 - .44 = .56; \epsilon = \sqrt{.56} = .75$$

We may evaluate ϵ^2 by reference to Table F. In this instance $(k - 1) = 2$ and $(N - k) = 12$. Entering the table of ϵ^2 with these values, we find that an ϵ^2 of .291 is significant at the 5 per cent level and an ϵ^2 of .459 at the 1 per cent level. Our observed ϵ^2 of .56 is much greater than the value at the 1 per cent level, which means that if the null hypothesis were true, then as large a value of ϵ^2 as we obtained would occur much less than 1 per cent of the time. Thus you see that none of our conclusions is changed by the ϵ^2 test. But, in addition to getting a test of significance, we also get an indication of the extent of the relationship between achievement and method of instruction. This is expressed by the unbiased correlation ratio, .75, the square root of ϵ^2 (74, p. 325).

When the total sum of squares in an analysis of variance problem has been analyzed into more than two parts, formula (65) must be revised. If, for example, we have analyzed the total sum of squares into three parts, one associated with columns, another with rows, and a third which we have called the residual, then we must use a different method of obtaining ϵ^2 . Peters and Van Voorhis (74) give the following formula:

$$\epsilon_{yz}^2 = \frac{(v_{col})(k - 1) - (v_{res})(k - 1)}{(v_{col})(k - 1) + (v_{res})(df_{res})} \quad (66)$$

where ϵ_{yz}^2 = epsilon-square
 v_{col} = the population variance estimated from the sum of squares between columns
 $k - 1$ = the degrees of freedom associated with the between-column variance or the number of columns minus 1
 v_{res} = the population variance estimated from the residual sum of squares
 df_{res} = the number of degrees of freedom associated with the residual sum of squares

The value of ϵ^2 at the 5 and 1 per cent points, when formula (66) is used, is found by entering the column of Table E

with $(k - 1)$ and running down to the row entry corresponding to the number of degrees of freedom associated with the residual variance, that is, df_{res} . We may illustrate the use of formula (66) by applying it to the data of Table 52. The estimated population variance from the sum of squares between columns was 5.2 for 4 degrees of freedom, and for the residual sum of squares the estimated population variance was .45 for 16 degrees of freedom. Substituting in formula (66) and solving for ϵ^2 , we get

$$\begin{aligned}\epsilon_{yz}^2 &= \frac{(5.2)(4) - (.45)(4)}{(5.2)(4) + (.45)(16)} \\ &= \frac{19}{28} \\ &= .68\end{aligned}$$

We find from the table of ϵ^2 that a value of .287 is significant at the 5 per cent point and a value of .43 at the 1 per cent point for 4 and 16 degrees of freedom. Our observed value of .68 is, therefore, highly significant; and we must consider the null hypothesis untenable. Thus we arrive at the same conclusion using ϵ^2 that we did using F .

You may wonder at this point whether F or ϵ^2 should be used as a test of significance in a given problem. There is no definite answer to this question, and there are advocates of each method. The calculations involved are much the same for both F and ϵ^2 , and tables of the 1 per cent and 5 per cent points of both ϵ^2 and F are available and easily used. The results of the tests of significance by both methods are consistent. Peters and Van Voorhis, however, point out that the F test does not "directly indicate the *strength* of the relation that is present, only its reliability. Analysis of variance, that is, tells only the negative side of the story, limiting itself to confirming or refuting the null hypothesis. Epsilon, on the other hand, shows in language with a uniform meaning what is the strength of the relation that is present and at the same time permits an 'exact' test of its reliability" (74, p. 353).

4. A TEST OF RECTILINEAR RELATIONSHIP

Epsilon-square may also be used to provide a test of whether two variables, X and Y , are related in a rectilinear fashion (74, p. 329). The test involves the computation of both ϵ^2 and r . The statistic derived from the test may be designated as ϵ'^2 and is obtained from the following formula:

$$\epsilon'^2 = \frac{\epsilon^2 - r^2}{1 - r^2} \tag{67}$$

Since ϵ'^2 has the same form of distribution as ϵ^2 , we may use the same tables for its interpretation. The table of ϵ^2 is entered with degrees of freedom equal to $k - 1$ and $N - k$. If ϵ'^2 is such that it exceeds the 5 or 1 per cent points, we may infer that the departure from rectilinearity is significant. If it does not, then we may infer that the relationship between the two variables is essentially rectilinear.

ANALYSIS OF VARIANCE: MATCHED GROUPS

Example 1.—Here is the set of scores of Example 2 in Chapter 9 on “The t Test of Significance.”

Group I.....	10	5	6	7	10	6	7	8	6	5
Group II.....	7	3	5	7	8	4	5	6	3	2

(a) Compute the value of F based upon an analysis into two parts: the between and the within sum of squares. Is the value of F thus obtained equal to the value of t^2 obtained when the pooled sum of squares is used in calculating the standard error of the mean difference?

(b) Test the difference between the groups, taking cognizance of the fact that the measurements have been paired. Analyze the total sum of squares into three parts, using the estimate of the variance based upon the residual sum of squares as the denominator in the F ratio. Is the value of F now equal to the value of t^2 obtained when the correlation is taken into account?

(c) Set up a table and compute the “residual” sum of squares directly. Does this value check with that obtained by subtraction?

CHAPTER 12

THE χ^2 TEST OF SIGNIFICANCE

Chi-square is a statistic similar to t and F in that its sampling distribution is known and in that it is also used for testing hypotheses. It is particularly applicable to situations where we wish to test the departure of observed frequencies in a given sample from the frequencies we would expect to obtain on the basis of a given hypothesis. Chi-square is also useful in testing the hypothesis that two samples have been drawn from a homogeneous population. We shall see some additional uses to which χ^2 may be put later, but now let us examine a specific case.

1. SIMPLE APPLICATIONS OF χ^2

Suppose that you are in charge of selecting a title for a new college magazine. You have eliminated all of the titles except two, but you cannot decide which of the two would have more appeal. To get some basis for making your decision, you go out and interview a small random sample of 60 students, asking each one to state which of the two titles he prefers. Your final count shows that 36 prefer Title No. 1 and 24 prefer Title No. 2. In this instance, as in most others where you are working with samples, it is not the sample that is of primary interest, but the population from which the sample was drawn. What you really wish to know is, if you go ahead and select Title No. 1, will your choice meet with the approval of a majority of the entire student body?

a. **Observed and expected frequencies.** When we wanted to make an inference on the basis of a sample mean concerning the population mean, we found that we could approach the problem by setting up some hypothesis about the population mean. Then by finding the deviation of our observed

sample mean from the hypothetical population value and dividing this deviation by the standard error of the mean, we arrived at a statistic called t . And, since the sampling distribution of t was known, we were able to make a probability statement concerning the frequency with which values of t as large as, or larger than, the one we obtained would occur by chance, assuming the hypothesis to be true.

Similarly, in the case of χ^2 , we must set up some hypothesis concerning the population ratio. We can then determine the frequencies we would expect to get in our sample, assuming the hypothesis to be true. Our deviations now become the difference between these *expected* frequencies and those actually *observed* in our sample. We may then use χ^2 to test the hypothesis that our sample may have been drawn from a population with parameter equal to the theoretical ratio. We assume, in other words, that any difference between our observed sample frequencies and those to be expected on the basis of the hypothesis can be accounted for in terms of sampling variation. The computations are simple: we merely take the difference between each observed and expected number, square these discrepancies, divide each squared discrepancy by the corresponding expected number, and sum. By formula

$$\chi^2 = \sum \frac{(o - e)^2}{e} \quad (68)$$

where

χ^2 = chi-square

o = the observed frequency

e = the corresponding expected frequency in terms of the hypothesis

If the obtained value of χ^2 is such that it would occur 5 per cent of the time or less, then, according to the standards we have used before, we would have to assume that the hypothesis being tested is untenable. On the other hand, if χ^2 is such that a value as large as, or larger than, the obtained value would occur more than 5 per cent of the

time, then we would have to consider the hypothesis tenable. The interpretation of χ^2 corresponds to the interpretation we have learned to place upon t , F , and ϵ^2 .

b. Testing a 50:50 hypothesis. In the problem at hand, since we are really interested in knowing whether or not a majority of the students prefer Title No. 1, our working hypothesis would be that the population is evenly divided. We assume that if we had interviewed every member of the population we would find 50 per cent favoring Title No. 1 and 50 per cent favoring Title No. 2. If the hypothesis is true, then the calculation of χ^2 proceeds as in Table 54.

TABLE 54.—TESTING A 50:50 HYPOTHESIS BY MEANS OF χ^2

	APPROVE TITLE NO. 1	APPROVE TITLE NO. 2	TOTAL
Observed.....	36	24	60
Expected.....	30	30	60
$(o - e)$	6	-6	
$(o - e)^2$	36	36	
$\frac{(o - e)^2}{e}$	1.2	1.2	$\chi^2 = 2.4$

Note that the computed value of χ^2 would be zero if our observed frequencies were the same as those expected on the basis of our hypothesis. The greater the departure of the observed frequencies from the expected, the larger the resulting value of χ^2 will be. If our sample had divided 55:5, for example, χ^2 would be equal to 41.66. In this respect, χ^2 is like t , for t also becomes larger the greater the departure of an observed sample mean from the hypothetical value of the population mean being tested. But, whereas the distribution of t was dependent upon the size of the sample, the general form of the distribution of χ^2 is independent of sample size. For example, if we drew a large number of random samples of 20 cases each from a box in which we

had placed 50 red and 50 white discs, the expected frequencies would be 10 red and 10 white. Computing χ^2 for each of these samples, we could plot a frequency distribution of the obtained values. If we then repeated the sampling process, this time drawing samples of 30 cases each, the expected numbers would be 15 and 15. If we made a frequency distribution of the values of χ^2 obtained with samples of this size, the two frequency distributions would be quite similar, despite the fact that the values of χ^2 in one case were based upon samples of 20 and in the other upon samples of 30.

To evaluate the χ^2 of 2.4 which we obtained in our illustration, we must enter Table H (page 342) with the number of degrees of freedom that are related to our sample data. We stated earlier that the concept of degrees of freedom may be regarded as having to do with the number of observations that are free to vary once certain restrictions have been placed upon a set of data. In the present problem we have only 1 degree of freedom because once we have entered one of the frequencies in Table 54, the other can be determined from the border total by subtraction. If we enter Table H with 1 degree of freedom we find that a value of χ^2 of 2.4 or larger may be expected to occur by chance more than 10 per cent of the time. Thus we have little basis for rejecting the hypothesis that our sample was drawn from a population in which the ratio is 50:50. In order for us to reject the hypothesis of evenly divided opinion at the 5 per cent level, our computed value of χ^2 would have to be 3.841.

Perhaps you have noticed that whenever a 50:50 hypothesis is being tested, the $(o - e)^2$ values are the same, and of course the expected numbers are the same also. Thus there is a simplified formula for χ^2 which may be used whenever this hypothesis is being tested:

$$\chi^2 = \frac{2(o - e)^2}{e} \quad (69)$$

Substituting in the problem under consideration, we get

$$\chi^2 = \frac{2(36 - 30)^2}{30} = \frac{(2)(6)^2}{30} = \frac{2(36)}{30} = \frac{72}{30} = 2.4$$

c. **Testing any a priori hypothesis.** In the same manner that we tested the hypothesis that the population ratio was 50:50 in the last example, we might test *any* hypothesis concerning a population ratio. Suppose, for example, that on the basis of past experience we knew that about 75 per cent of the members of a general psychology class could be expected to pass an item on a test. We now have a new class consisting of 200 members. On the basis of our past experience we would predict that about 75 per cent or 150 of these students would pass the item and that about 25 per cent or 50 would fail. Of course, we would not expect to obtain exactly these numbers; we can assume that some variation will be present as a result of chance or sampling factors. But suppose now, after giving the test, we find that only 137 pass the item and 63 fail it. Is this departure from the expected frequencies too great to be attributed to chance? Can we assume that our sample was drawn from a population in which the ratio of those passing to those failing the item is 75:25? The hypothesis may be tested by calculating χ^2 .

The value of χ^2 based upon the data of Table 55 is 4.51 and with 1 degree of freedom, a value of 3.841 is significant

TABLE 55.—TESTING A 75:25 HYPOTHESIS BY MEANS OF χ^2

	FAILING THE ITEM	PASSING THE ITEM	TOTAL
Observed	63	137	200
Expected	50	150	200
$(o - e)$	13	-13	
$(o - e)^2$	169	169	
$\frac{(o - e)^2}{e}$	3.38	1.13	$\chi^2 = 4.51$

at the 5 per cent point. Hence, if our sample was drawn from a population in which the ratio of those passing to those failing the item is 75:25, something has happened to our sample that would occur on the basis of chance less than 5 per cent of the time. According to our standards, we would thus consider the hypothesis untenable.

d. χ^2 calculated from per cents. Sometimes data are reported in terms of per cents by other investigators; as readers of their reports, we may wish to test their findings against some hypothesis which we may have. The value of χ^2 cannot be found directly from per cents, but requires a correction term (86). If we work with per cents, then the resulting value of χ^2 must be multiplied by $\frac{N}{100}$, where N is

the observed sample total. In our title illustration we found observed frequencies of 36 and 24. In terms of per cents these would be 60 and 40. Our expected frequencies must be expressed in the same units and, on the basis of a hypothesis of evenly divided opinion, they would be 50 and 50. Using formula (69), we find that χ^2 is equal to $(2)(10)^2/50$ or 4. This is not the same value we found before, 2.4. But applying our correction term, we get $(4) \left(\frac{N}{100} \right) = (4) \left(\frac{60}{100} \right) =$

$$\frac{240}{100} = 2.4.$$

You may think that translating the title data into per cents has served to increase the sample size (from 60 to 100) and that this is the cause of the discrepancy between the two values of χ^2 . If so that would be inconsistent with the statement we made earlier that the distribution of χ^2 is independent of sample size. It is not the increase in the sample size, but rather the increase in the value of the deviations that has changed the value of χ^2 . While a deviation of $36 - 30$ or 6 is not uncommon in samples of 60 cases, a deviation of $60 - 50$ or 10 in samples of 100 drawn from the same population is uncommon. This may become

clearer if you think of tossing 10 coins into the air. According to the laws of probability the expected frequencies are 5 heads and 5 tails. But if you found 6 of one and 4 of the other you would not be too upset. Suppose now that you tossed 1,000 coins. Would you be surprised to get 600 heads and 400 tails?

e. χ^2 applied to more than two categories. The application of χ^2 is not limited to dichotomous distributions. It can be used when we have sample data divided into three or more categories. In the absence of any a priori hypothesis concerning the population, we may assume that the sample frequencies should be distributed according to chance, that is, with an equal number in each category. Suppose, for example, we polled 60 students and asked their opinions concerning a contemplated change in the hours during which the library is to be open. We allow for three categories of response: favorable, indifferent, and unfavorable. According to chance we would expect to find 20 students in each category. The data and the calculation of χ^2 are given in Table 56.

TABLE 56.—TESTING THE HYPOTHESIS OF A UNIFORM DISTRIBUTION

	UNFAVORABLE RESPONSE	INDIFFERENT RESPONSE	FAVORABLE RESPONSE	TOTAL
Observed.....	15	10	35	60
Expected.....	20	20	20	60
$(o - e)$	-5	-10	15	
$(o - e)^2$	25	100	225	
$\frac{(o - e)^2}{e}$	1.25	5	11.25	$\chi^2 = 17.5$

The value of χ^2 now equals 17.5. How many degrees of freedom are involved in our data? In general, in a problem of this nature, the number of degrees of freedom is equal to $r - 1$, where r equals the number of categories that we have.

Thus if we have three categories we have 2 degrees of freedom, if we have 4 categories we have 3 degrees of freedom, and so on. Another way of looking at the matter is this: once the frequencies for two of the categories have been entered, then the third can be determined by subtraction from the border total. If our sample total is to remain 60, and if we had 40 in the first category and 5 in the second, the third would have to be 15. If we had 20 frequencies in the second category and 10 in the third, then the first would have to have 30, if our sample total is to remain the same. According to Table H, a χ^2 of 17.5 with 2 degrees of freedom is significant beyond the 1 per cent point. Hence we may reject the hypothesis that our sample was drawn from a population in which the same proportion would be found in each of the categories. Our observed frequencies, in other words, differ significantly from those to be expected, on the basis of the hypothesis tested.

2. χ^2 APPLIED TO TWO SAMPLES

In all of the problems considered so far, we have had only a single sample and either we have had an a priori hypothesis to work with or else we have assumed a chance hypothesis, i.e., that the frequencies would be distributed uniformly in the various categories. But in many problems our interest is in comparing two or more samples, and in such situations it sometimes happens that we have no a priori hypothesis, and frequencies in each of our samples depart so very far from chance expectations that this does not provide us with a very good basis for comparing the groups. We might have, for example, two groups of 100 cases each, one dividing 70:30 and the other 90:10. In each sample the departure from chance or 50:50 is highly significant.

Let us suppose that the samples we have consist of two groups of eighth-grade students and that 90 is the number passing an item on a standardized test in one of the samples and 70 is the number passing in the other. Since we have

only these data and no a priori hypothesis against which to test the frequencies, how may we compare the two groups?

If we make the assumption that both of our samples are drawn from a homogeneous population, then we may combine the frequencies from each sample to get a new common estimate of the population and we can measure the deviations of each sample from it. If the resulting value of χ^2 is small, i.e., does not reach the 5 per cent point, then we may infer that the two samples are drawn from the same population and that the best estimate of the parameter of this population is given by the frequencies from the pooled samples. On the other hand, if the value of χ^2 exceeds the 5 per cent point, then we may reject the hypothesis that the samples are drawn from a homogeneous population and infer that the observed frequencies in each sample are significantly different.

The hypothesis in the present problem is based upon the combined samples in Table 57. Specifically, we test the

TABLE 57.—TESTING THE HYPOTHESIS THAT TWO SAMPLES HAVE BEEN DRAWN FROM A HOMOGENEOUS POPULATION

	FAILING THE ITEM	PASSING THE ITEM	TOTAL
Sample 1.....	10	90	100
Sample 2.....	30	70	100
Total.....	40	160	200

<i>o</i>	<i>e</i>	<i>o - e</i>	$(o - e)^2$	$\frac{(o - e)^2}{e}$
10	20	-10	100	5.00
90	80	10	100	1.25
30	20	10	100	5.00
70	80	-10	100	1.25
Total 200	200	$\chi^2 = 12.50$

hypothesis that the two samples have been drawn from a population in which the probability of passing the item is 160/200 or 80 per cent and the probability of failing the item is 40/200 or 20 per cent. On the basis of this hypothesis, we calculate the expected frequencies for Sample 1 and Sample 2 by simply multiplying the total number in each of the samples by the theoretical percentages 80 and 20. The calculation of χ^2 then proceeds as before.

The computed value of χ^2 , 12.50, must be evaluated in terms of the number of degrees of freedom involved in the set of data upon which it is based. If we place the restriction upon our data that the border totals of Table 57 must remain the same in each sampling, then only one cell in the table can be filled in independently. Once a single frequency has been entered in any one of the four cells, then the values for the remaining three cells can be determined by subtraction from the border totals. If 90 is entered in the upper right-hand cell, then $100 - 90$ or 10 must be entered in the upper left-hand cell. And if 10 is the value of the upper left-hand cell, then $40 - 10$ or 30 must be entered in the lower left-hand cell. The value for the lower right-hand cell can be determined in a similar manner. Consequently, we say that only one degree of freedom is involved in this problem.

According to Table H, for one degree of freedom a value of χ^2 as large as 12.5 or larger would occur much less than 1 per cent of the time, if the two samples were drawn from a homogeneous population. Consequently, we reject the hypothesis that we set out to test and say that the observed frequencies in the two samples differ significantly.

It is worth noting in a problem of this nature that only one of the expected frequencies needs to be calculated; the others can be obtained by subtraction from the border totals, since the sum of the expected frequencies in each row and each column must check with these values. There is also a very simple rule for determining the number of degrees of freedom involved when the hypothesis to be tested is

based upon the border totals, as in this problem. The rule is this: if you have a table with r rows and k columns, the number of degrees of freedom is equal to $(r - 1)(k - 1)$. In a 2×2 table, such as we have here, the number of degrees of freedom is equal to $(2 - 1)(2 - 1)$ or 1. In a 3×4 table, the number of degrees of freedom would be equal to $(3 - 1)(4 - 1)$ or 6.

The application of χ^2 to the type of problem we have just discussed is of much value in psychological and educational research. Suppose that we were interested in the effectiveness of two methods of psychotherapy. We try the two methods out with a sample of 100 patients divided at random into two groups of 50 subjects each, and observe the number in each group showing "improvement" and the number showing "no improvement." We pool the data of the two groups to get our hypothesis and then calculate χ^2 , assuming the hypothesis to be true. If χ^2 exceeds the 5 per cent point, we reject the hypothesis that the two groups were drawn from the same population and infer that the numbers showing improvement and no improvement in the two samples differ significantly, i.e., that one method is more effective than the other. A problem of this nature is illustrated by the data of Table 58.

If the hypothesis to be tested is true, then we would expect to find 70/100 or 70 per cent of the 50 individuals treated by Method 1 to show improvement, and 30/100 or 30 per cent to show no improvement. Since the number of cases treated by each of the two methods is the same, the expected numbers for each of the two groups will be the same. The value of χ^2 is found as before.

The computed value of χ^2 , 4.76, exceeds the 5 per cent point, 3.841, for 1 degree of freedom. Hence we reject the hypothesis that the two samples were drawn from the same population and conclude that the numbers showing improvement in the two groups differ significantly. Method 1 is more effective than Method 2, and the difference in the numbers showing improvement, we believe, cannot be the result of sampling variation.

TABLE 58.—TESTING THE EFFECTIVENESS OF TWO METHODS OF PSYCHOTHERAPY WITH TWO SAMPLES OF 50 CASES EACH

	NUMBER SHOWING NO IMPROVEMENT	NUMBER SHOWING IMPROVEMENT	TOTAL
Method 1.....	10	40	50
Method 2.....	20	30	50
Total.....	30	70	100

<i>o</i>	<i>e</i>	<i>o - e</i>	$(o - e)^2$	$\frac{(o - e)^2}{e}$
10	15	-5	25	1.67
40	35	5	25	.71
20	15	5	25	1.67
30	35	-5	25	.71
Total 100	100	$\chi^2 = 4.76$

3. χ^2 APPLIED TO MORE THAN TWO GROUPS

We are not limited to two groups nor need the number of subjects in each group be the same in order to apply the χ^2 test of significance to problems of the kind described. To illustrate the procedure to be used in a more complicated problem, we have the hypothetical data of Table 59, in which 250 subjects have been classified according to "letter grades" on a test of general information and according to "educational status." We see that 95 of the subjects are college graduates, 70 high school graduates, and 85 elementary school graduates. The subjects in each of these three groups are then classified according to letter grade received on the test.

If we assume that the three educational groups are samples from the same population, we may pool the data from each. On the basis of the pooled frequencies, we estimate the

TABLE 59.—TESTING THE SIGNIFICANCE OF DIFFERENCES BETWEEN GRADES ON A TEST OF GENERAL INFORMATION FOR THREE EDUCATIONAL LEVELS

	GRADE ON INFORMATION TEST			TOTAL
	C	B	A	
College graduates.....	10	35	50	95
High school graduates.....	20	40	10	70
Elementary school graduates.....	35	40	10	85
Total.....	65	115	70	250

<i>o</i>	<i>e</i>	<i>o - e</i>	$(o - e)^2$	$\frac{(o - e)^2}{e}$
10	24.7	-14.7	216.09	8.75
35	43.7	- 8.7	75.69	1.73
50	26.6	23.4	547.56	20.58
20	18.2	1.8	3.24	.18
40	32.2	7.8	60.84	1.89
10	19.6	-9.6	92.16	4.70
35	22.1	12.9	166.41	7.53
40	39.1	.9	.81	.02
10	23.8	-13.8	190.44	8.00
Total 250	250.0	$\chi^2 = 53.38$

proportion in the population to be found in each letter grade category. Thus, in the "A" category we have $70/250 = .28$; in the "B" category we have $115/250 = .46$; and in the "C" category we have $65/250 = .26$. Now if our samples have all been drawn from this population, then the expected numbers in each letter grade category for each sample will be these proportions of the sample totals. Thus the expected frequencies for the sample of 95 individuals who are college graduates will be: $(.28)(95) = 26.6$ in the "A" category; $(.46)(95) = 43.7$ in the "B" category; and $(.26)(95) = 24.7$ in the "C" category. We find the ex-

pected numbers for each of the other groups in the same manner.

The calculation of χ^2 is the same as before. We subtract each of the expected numbers from the corresponding observed numbers and square the deviations. Each squared deviation is then divided by the corresponding expected number. The sum of these values which, in the present instance, is 53.38, is equal to χ^2 . We have $(r - 1)(k - 1)$ or $(3 - 1)(3 - 1) = 4$ degrees of freedom and according to Table H a value of χ^2 of 13.277 is significant at the 1 per cent level. Since our obtained value of 53.38 greatly exceeds 13.277, we must reject the hypothesis that these samples were drawn from the same population and conclude that the observed differences are significant.

4. TESTING "GOODNESS OF FIT"

There is another valuable use of χ^2 . You may recall that when we discussed fitting a normal distribution to a set of observed frequencies we had no basis for comparing the "goodness of fit" other than inspection. It should be readily apparent, however, that χ^2 can be used to provide us with a more precise answer to the question of whether the observed set of frequencies and those to be expected on the basis of a normal distribution differ significantly or not. All that we need to do is to find the difference between each of the observed and expected frequencies for each class interval, square each of these deviations, and divide each by the corresponding expected number. The sum of these values is χ^2 .

How many degrees of freedom do we have in evaluating this χ^2 ? That depends, as Peters and Van Voorhis (74) point out, upon the question we pose. If the only restriction placed upon the set of frequencies is that their totals remain the same, then the number of degrees of freedom is equal to the number of class intervals minus 1. If we place the further restriction upon the data that the mean and standard deviation must remain the same, which we ordinarily do in

fitting a normal distribution, then the number of degrees of freedom is equal to the number of intervals minus 3.

If the computed value of χ^2 is small so that it may be expected to occur quite frequently as a result of sampling variation, then we may infer that the discrepancies between the observed frequencies and those to be expected on the basis of a normal distribution are not significant. For example, suppose we had a distribution consisting of 12 class intervals and we fitted a normal distribution to the observed frequencies in the manner described in an earlier chapter. We compute χ^2 and find it to be 8.5. By reference to Table H we find that for 9 degrees of freedom values of χ^2 as large as 8.5 may be expected to occur quite frequently ($P = .50$). Consequently, we may be quite confident that our distribution does not depart significantly from the normal form. To reject the hypothesis of normality we would have had to obtain a χ^2 of at least 16.919, the tabled value at the 5 per cent level for 9 degrees of freedom.

5. χ^2 AND SMALL FREQUENCIES

It seems to be generally agreed that the χ^2 test should not ordinarily be applied to tables in which any cell entry is less than 5.¹ If a given cell entry is less than this value, then a correction may be applied to 2×2 tables or other tests involving 1 *df*.² The correction consists of adding .5 to the smallest observed frequency of the table (77). Since the marginal totals must remain the same, this means that the other cell entries must be adjusted accordingly.³

¹ This limit is set by Kenney (53) and others (38, 64, 86), but Tippett (90) suggests a value of 10.

² In applying the correction to the test of a ratio, for example, a 50:50 hypothesis with observed frequencies of 18 and 12, we add .5 to the smaller frequency, 12, and subtract .5 from the larger frequency, 18. We thus reduce each deviation of observed from expected by .5, i.e., from 3 to 2.5 in this instance. For a more detailed discussion of this point and other limitations of the χ^2 test which are beyond the scope of this brief treatment, the student is referred to Goulden (35, pp. 101-104).

³ Snedecor, however, has recently said: "Accumulating evidence indicates that the inaccuracies which may be introduced by small expected numbers are not so serious as was formerly thought" (86, p. 169).

The limitation imposed by small frequencies would apply also in the fitting of a normal distribution to a set of observed frequencies. Since in this instance, it is apt to be the extreme class intervals that contain small frequencies, the limitation is usually circumvented by combining the frequencies in the lowest interval with those in the interval directly above it and by combining the frequencies in the highest interval with those in the interval directly below. Sometimes it may be necessary to combine the frequencies of several intervals, the number of degrees of freedom being reduced accordingly.

6. χ^2 AND THE ϕ COEFFICIENT

We discussed earlier the use of the ϕ coefficient as a measure of association or relationship when two variables had been reduced to a 2×2 table. Although we found that the ϕ coefficient could be used to give us some indication of the strength of the relationship present, we had no means of testing whether or not the relationship was significant, i.e., whether it might be the result of chance or not. In the case of r we had a test of the null hypothesis by reference to Table D or by the computation of t according to formula (52) and then by reference to the table of t . The calculation of χ^2 provides us with a similar test for the ϕ coefficient. The ϕ coefficient and χ^2 are related in the following way

$$\phi = \sqrt{\frac{\chi^2}{N}} \quad (70)$$

and
$$\chi^2 = N\phi^2 \quad (71)$$

Hence, if we have computed ϕ as a measure of association and wish to test the null hypothesis, we need merely square the obtained value and multiply by N , the number of cases upon which it is based, to arrive at χ^2 . The number of degrees of freedom involved in the computation of χ^2 is, of course, 1, since we have a 2×2 table. We evaluate χ^2 as we have before by reference to Table H.

If we take the data from Table 27 cited earlier in our discussion of ϕ , we have a computed value of ϕ equal to .328 with an N of 100. Substituting in formula (71) we get

$$\begin{aligned}\chi^2 &= (100)(.328)^2 \\ &= (100)(.107584) \\ &= 10.76\end{aligned}$$

By reference to Table H we find that for 1 degree of freedom a value of χ^2 of 6.635 is significant at the 1 per cent level of confidence. Since our obtained value is 10.76 we would reject the hypothesis of no relationship with a great deal of confidence.

We could, of course, in a 2×2 table, reverse the procedure and compute χ^2 first. This would tell us whether or not there was any association present and whether we could reject the null hypothesis with any degree of confidence. If we were then interested in getting some indication of the strength of this relationship we could substitute in formula (70) and solve for ϕ .

CHI-SQUARE

Example 1.—Previous experience with a particular achievement test indicated that for seventh-grade children the ratio of those receiving a passing mark to those failing was 3 to 1. We wished to test whether this hypothesis (3:1) would also hold for sixth-grade children. A sample of 100 students drawn from the sixth-grade revealed that 60 passed. Is the hypothesis tenable?

Example 2.—A poll of fraternity men on a university campus showed that the ratio of those on the honor list to those not on the list was 1:4. To find out whether this ratio would hold for sorority members a sample of 150 sorority members was drawn. Forty of the sorority members were on the honor list. Must we abandon the hypothesis of 1:4?

Example 3.—A chairman of a committee confronted with a choice between the use of two slogans decided to sample a number of individuals to determine which they preferred. In a sample of 80 he found that 50 approved Slogan No. 1 and 30 approved

Slogan No. 2. Can we assume that the two slogans are equally popular?

Example 4.—A sample of 30 schizoid cases in a mental institution gave the following responses to an item in a personality inventory: "Yes" = 18; "?" = 9; "No" = 3. Another sample of 30 manic cases gave the following responses to the same item: "Yes" = 6; "?" = 9; "No" = 15. Test the hypothesis that there is no difference between the responses of the two groups, i.e., that the two samples are drawn from the same population.

Example 5.—A group of men and a group of women distributed their responses to an item in an attitude test in the fashion listed below. Test the hypothesis that there is no difference between the distribution of responses of the two groups, i.e., that they are both samples from the same population.

	STRONGLY DISAGREE	DISAGREE	UNDECIDED	AGREE	STRONGLY AGREE	TOTAL
Men.....	5	5	12	18	60	100
Women...	25	25	20	20	10	100

Example 6.—Two methods of psychotherapy were tried with two different samples. The results are given below. Can we assume that one method is more effective than the other?

METHOD	NUMBER SHOWING NO IMPROVEMENT	NUMBER SHOWING IMPROVEMENT	TOTAL
1	10	42	52
2	58	60	118

Example 7.—Kuo (57) reared kittens under three different conditions: (1) one group of kittens was isolated from all contact with rats except on the experimental test; (2) the kittens in another group were reared with their mothers whom they saw kill a rat or mouse every 4 days outside the cage; (3) one group lived with a single rodent from age 6-8 days onward. The test situation consisted of putting a kitten together with a rat to determine whether or not the kitten would kill. The data are as follows:

GROUP	NO. KILLING	NO. NOT KILLING	TOTAL
(1) Isolated.....	9	11	20
(2) Reared with mother.....	18	3	21
(3) Reared with rodent.....	3	15	18

Are there any significant differences between the groups or can we assume that they are all samples drawn from the same population?

Example 8.—Rosenzweig (80) has studied the recall of subjects for finished and unfinished tasks when they worked on the tasks under differing sets of instructions. The “informal” group was told that the experimenter was interested in knowing something about the task, that the ability of the subject was not under investigation. The “formal” group, on the other hand, was under the impression that the tasks were an intelligence test. The data for recall are given below. Calculate χ^2 and interpret your results.

GROUP	NO. SUBJECTS RECALLING PREPONDERANCE OF FINISHED TASKS	NO. SUBJECTS RECALLING PREPONDERANCE OF UNFINISHED TASKS	NO. SUBJECTS WITH NO PREPONDERANT TENDENCY
Informal group..	7	19	4
Formal group..	17	8	5

Example 9.—Determine whether the following distributions came from the same population. Assume that Distribution II gives the frequencies expected according to some hypothesis held by an experimenter and Distribution I gives the observed frequencies.

SCORES	DISTRIBUTION I	DISTRIBUTION II
70-74	5	7
65-69	18	15
60-64	24	20
55-59	20	30
50-54	16	20
45-49	20	15
40-44	14	10

CHAPTER 13

PREDICTIONS AND THE EVALUATION OF PREDICTIONS

Once we have discovered and measured differences and relationships by means of the statistical techniques discussed so far, we are then often interested in making predictions based upon our discoveries. The simplest problem of prediction is one of predicting the presence or absence of a given characteristic from knowledge of some other characteristic, that is, predictions based upon data classified in categories. If we have found by the χ^2 test, for example, that there are significant differences between the responses of men and women to an opinion poll on a given issue, then we might be interested in determining how accurately we can predict the poll results on the basis of this knowledge. Or perhaps we have found that employees who are above the mean on a psychological test tend to be rated as "satisfactory" by their supervisors, while those who are below the mean tend to be rated as "unsatisfactory." How accurately can we predict the ratings of a group of employees on the basis of this knowledge?

1. PREDICTING SIMPLE CHARACTERISTICS

Suppose that we have polled 200 students to determine whether they approve or disapprove of a contemplated university ruling. Our final tabulations show that 120 approve and 80 disapprove. With only the knowledge that our first sample divides 120 to 80, our prediction of the responses of the members of another similar sample would follow the *principle of maximum likelihood* or *maximum probability* (43). This means that we would predict for every individual the most probable or most frequent response, which

happens to be "approve" since it was made 120 times as compared with a frequency of 80 for "disapprove." If we made this prediction and if the ratio 120:80 held true for the second sample, we would have made 120 correct predictions out of a total of 200. We would be correct, in other words, $120/200 = 60$ per cent of the time. And we would be wrong $80/200 = 40$ per cent of the time.

Now let us suppose that we have knowledge of another factor: the manner in which men and women voted on the issue. Let us assume that we had 100 men and 100 women in the first sample and that the women voted 40 "approve" and 60 "disapprove," whereas the men voted 80 "approve" and 20 "disapprove." With knowledge of this factor what will our predictions be? It is obvious that if knowledge of the division of opinion within each sex does not increase the number of correct predictions over the number that we could successfully predict without this knowledge, then the information contributes nothing.

We shall still follow the principle of maximum likelihood, but we shall now consider the two groups separately. The most frequent response for men is "approve" and consequently that will be the predicted response for all men. The most frequent response for women, on the other hand, is "disapprove" and that will be our predicted response for all women. How many correct predictions will we now make? In the two groups combined we would have 80 correct for the men and 60 correct for the women for a total of 140 correct predictions. This is better than the total of 120 correct predictions we would have made without knowledge of the division of opinion within each sex.

To determine how much better our predictions are with knowledge of the vote according to sex, we divide the excess correct predictions made with this knowledge by the number we would have had without the knowledge. The result, $20/120 = 16.7$ per cent. Our predictions made with knowledge of the division of opinion within each sex are, in other words, 16.7 per cent better than predictions made without

this knowledge. Table 60 illustrates the basic computations involved in this evaluation.

Let us turn to another example. Suppose that we have entrance examination scores on 800 college students and we

TABLE 60.—PREDICTIONS OF RESPONSE FROM KNOWLEDGE OF SEX CLASSIFICATION

GROUP	ACTUAL VOTE		TOTAL	PREDICTION	NUMBER CORRECTLY PREDICTED
	Disapprove	Approve			
Men.....	20	80	100	(Approve) (Disapprove)	80
Women.....	60	40	100		60
Total.....	80	120	200	140

Correct predictions without knowledge of division of opinion within each sex.....120 or $120/200 = 60$ per cent
 Correct predictions with knowledge of division of opinion within each sex.....140 or $140/200 = 70$ per cent
 Increase in correct predictions with knowledge of division of opinion within each sex.....20 or $20/120 = 16.7$ per cent

then divide the students into two groups: those with scores above the median and those with scores below. We then classify the subjects within each group according to whether they earned average grades of "C or better" or "below C" during their freshman year in college. The results of our classification are given in Table 61, where we have evaluated the accuracy of our predictions of "academic standing" made from knowledge of classification on the entrance examination, in terms of predictions made without such knowledge.

Perhaps you are wondering whether or not it is possible to make predictions in the opposite direction. The answer is "yes," though ordinarily we are interested in predicting in only one direction. As an exercise, you might try pre-

TABLE 61.—PREDICTIONS OF "ACADEMIC STANDING" FROM KNOWLEDGE OF CLASSIFICATION ON AN ENTRANCE EXAMINATION

GROUP	ACTUAL FREQUENCIES		TOTAL	PREDICTION	NUMBER CORRECTLY PREDICTED
	Below C	C or Above			
Above Mdn.,..	150	250	400	(C or Above)	250
Below Mdn.,..	225	175	400	(Below C)	225
Total.....	375	425	800	475

Correct predictions without knowledge of classification on entrance examination.....425 or $425/800 = 53$ per cent
 Correct predictions with knowledge of classification on entrance examination.....475 or $475/800 = 59$ per cent
 Increase in correct predictions with knowledge of classification on entrance examination..... 50 or $50/425 = 12$ per cent

dicting whether an individual will be above or below the mean on the entrance examination from knowledge of his grades.

2. PREDICTING MEASUREMENTS

We are now ready to consider another problem of prediction. This time our interest is in predicting measurements of one quantitative variable from knowledge of measurements of another quantitative variable. We might wish to predict, for example, the most probable scores on one test from knowledge of scores on a second test. Obviously, if there is no relationship between the two tests, then we cannot expect to use scores on one as a basis for predicting scores on the other. It is the presence of correlation or association between the two that makes prediction possible, and the efficiency or accuracy of such predictions is a function of the degree or strength of the relationship that exists. For purposes of illustration, let us take the data cited earlier concerning the relationship between scores on a test of intelligence and scores on a measure of susceptibility to

hypnosis. We plot these scores in a scatter diagram (Figure 15).

a. **The regression line.** If we can assume that the relationship between the scores on the two tests is rectilinear, then we can draw a straight line through the plotted values which will indicate the trend of the relationship. This line is called the *regression line*.¹ It indicates the average change to be expected in one variable with change in the other. The regression line can then be used for prediction purposes. Given an individual's score on the X variable, we could erect an ordinate at that point and follow it up to the point where it meets the regression line. The Y value corresponding to this point would be the individual's most probable Y score.

It is obvious, however, that a number of straight lines might be drawn through the plotted values, since they will not fall precisely on any single line. We could, perhaps, draw one by inspection which seemed to "best fit" the data, but we can fit this line much more accurately by another technique, the *method of least squares*. This simply means that we shall demand that the line drawn to describe the trend be the one from which the sum of the squared errors of estimate (errors of prediction) is at a minimum. Our errors of estimate are the deviations of the observed values from the line describing their trend. If we let Y' equal a predicted or estimated value (a point on the line corresponding to some value of X) and Y equal the actual value, then we must find the line for which the $\Sigma(Y - Y')^2$ is at a minimum.

The line just described is the regression line we desire to find and it has a number of properties. It will pass through the point where the mean of the X variable and the mean

¹ Francis Galton first used the term "regression" in studying the inheritance of stature. It was his observation that, on the average, the offspring of abnormally tall parents and abnormally short parents tend to move back toward the population mean. The line describing this trend was called a "regression line." The term is still used to describe the line drawn among a group of points to represent the trend present, but it no longer necessarily carries the original implications that Galton intended.

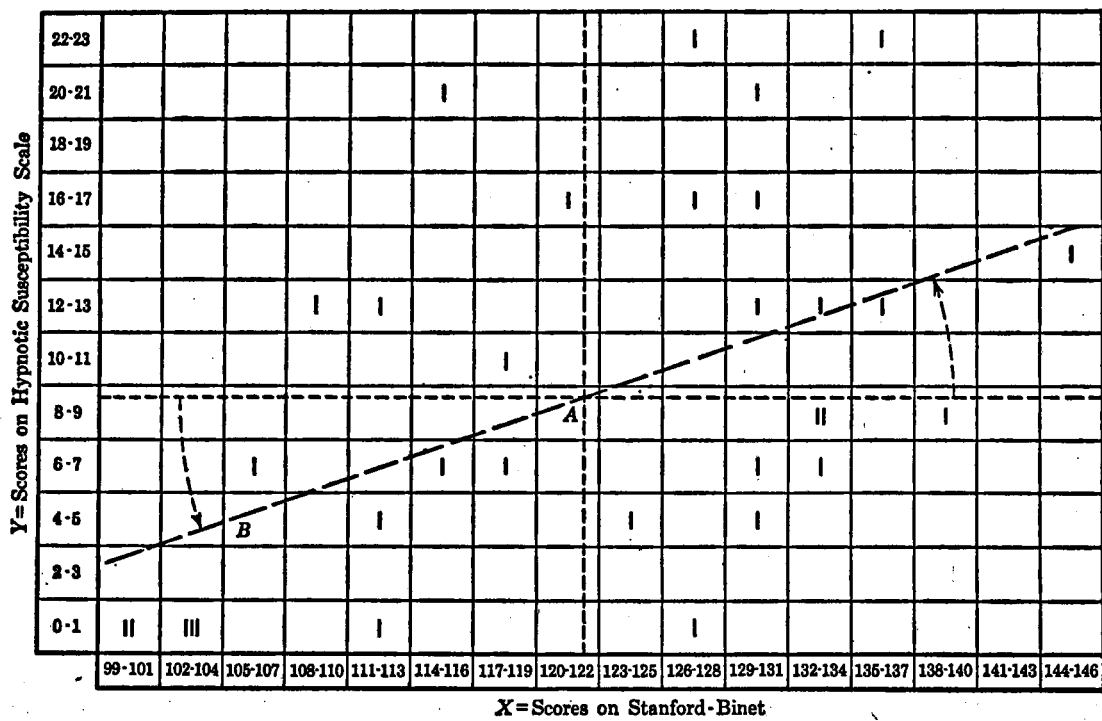


FIG. 15.—A scatter diagram of scores on a "hypnotic susceptibility scale" and scores on the Stanford-Binet. Dotted lines have been drawn through the means of X and Y. The regression line (Y on X) is represented by the line AB.

of the Y variable meet in the scatter diagram. The sum of the deviations, $\Sigma(Y - Y')$, will be equal to zero. And the sum of the squared deviations, $\Sigma(Y - Y')^2$, will be less than it would be from any other straight line.²

b. The regression coefficient. The slope of the regression line is given by the *regression coefficient*, b , which is simply the "covariance" divided by the variance of X . Thus

$$b_{yx} = \frac{\frac{\Sigma xy}{N}}{\frac{\Sigma x^2}{N}} \quad (72)$$

where b_{yx} = the regression coefficient of Y on X

$\frac{\Sigma xy}{N}$ = the covariance

$\frac{\Sigma x^2}{N}$ = the variance of the X distribution

You are already familiar with both of these values from the discussion of the correlation coefficient. The Σxy is the sum of the cross-products and is called, when divided by N (the number of paired measurements on which it is based), the covariance. We can simplify formula (72) by multiplying both the numerator and the denominator by N to get

$$b_{yx} = \frac{\Sigma xy}{\Sigma x^2} \quad (73)$$

Since we have the values needed from our earlier calculation of the correlation coefficient (Table 20) we may substitute in formula (73) and get

$$b_{yx} = \frac{1,419.36}{4,993.92} = .284$$

² The line described is for the regression of Y on X and is used in the prediction of Y from X . When predicting X from Y , as we shall see later, another line for the regression of X on Y is used.

c. **The regression equation.** Once we have computed the regression coefficient we may substitute in the formula for the *regression equation* to find the most probable value of Y for any given value of X . The formula for the regression equation is

$$Y' = b_{yx}(X - M_x) + M_y \quad (74)$$

where Y' = the predicted or estimated value of Y
 b_{yx} = the regression coefficient of Y on X
 X = the value of X for which we are predicting a value of Y
 M_x = the mean of the X distribution
 M_y = the mean of the Y distribution

In order to avoid subtracting M_x from X each time that a prediction is made, we may rewrite formula (74) so that

$$Y' = b_{yx}X + k \quad (75)$$

where $k = M_y - b_{yx}M_x$

To illustrate the use of the regression equation, let us suppose that we are given an X score of 132. What is our best prediction of the corresponding Y score? Substituting in the above formula with the necessary values calculated from Table 20, we solve for Y' .

$$\begin{aligned} Y' &= (.284)(132) + [9.31 - (.284)(121.56)] \\ &= 37.49 + (9.31 - 34.52) \\ &= 37.49 - 25.21 \\ &= 12.28 \end{aligned}$$

In a similar manner we could find the most probable Y value for any other given value of X .

d. **Regression and correlation coefficients.** You may have perceived already that the formula for the regression coefficient bears a very close resemblance to one of the formulas we used for the correlation coefficient. As a matter of fact, when we have already computed the correla-

tion coefficient, we may use it to find the regression coefficient. Thus if

$$\begin{aligned}
 b_{yx} &= \frac{\Sigma xy}{\Sigma x^2} && \text{then substituting an identity} \\
 &= \frac{\Sigma xy}{N\sigma_x^2} && \text{and multiplying both numerator and} \\
 &= \left(\frac{\Sigma xy}{N\sigma_x^2} \right) \left(\frac{\sigma_x\sigma_y}{\sigma_x\sigma_y} \right) && \text{rearranging} \\
 &= \left(\frac{\Sigma xy}{N\sigma_x\sigma_y} \right) \left(\frac{\sigma_x\sigma_y}{\sigma_x^2} \right) && \text{simplifying and} \\
 &= \left(\frac{\Sigma xy}{N\sigma_x\sigma_y} \right) \left(\frac{\sigma_y}{\sigma_x} \right) && \text{substituting an identity} \\
 &= r \left(\frac{\sigma_y}{\sigma_x} \right) && (76)
 \end{aligned}$$

If we substitute in formula (76) with the value of r that we had previously obtained for these test scores, we get

$$\begin{aligned}
 b_{yx} &= r \left(\frac{\sigma_y}{\sigma_x} \right) \\
 &= .52 \left(\frac{6.82}{12.49} \right) \\
 &= (.52)(.546) \\
 &= .284
 \end{aligned}$$

which equals the value we obtained by using formula (73).

How can we evaluate the reduction in our error of prediction when we use the regression coefficient and regression equation to predict values of Y ? In the absence of any knowledge of relationship between X and Y our best prediction for any given value of X would, of course, be the mean of the Y distribution and the extent of the errors of our predictions would be the standard deviation of the entire Y distribution. If you look for a moment at Figure

15, in which we have plotted the paired X and Y scores, you may be able to see more clearly just what influence correlation will have in reducing our errors of prediction. If we drew a horizontal line through the mean of the Y distribution, then the vertical deviations of each plotted point from this line (y deviations) would represent the actual deviations of $Y - M_y$. The sum of these deviations would be equal to zero, but the sum of these squared deviations would be greater than the sum of squared deviations from the regression line—if there is any relationship between X and Y . If the horizontal line through the mean of the Y distribution is rotated counterclockwise about the point A , where the mean of the X and the mean of the Y distribution fall, then the sum of squared deviations would become smaller and smaller until the line coincides with the regression line—line AB in Figure 15. It is the second variable, X , which makes this regression line meaningful. As long as the Y measures are considered alone, the best predicted value of Y for any single X measure would be the horizontal line, or mean of the Y distribution. But when there is regression of Y on X we find that different values of Y are associated with different values of X . It is these associated values that become our predictions when we have knowledge of the relationship between the two variables.

e. **The standard error of estimate.** The extent to which we have errors of prediction, i.e., the degree to which our predicted Y' values fail to correspond to the actual Y values, is indicated by the *standard error of estimate*. In terms of a formula

$$\sigma_{y.s} = \sqrt{\frac{\Sigma(Y - Y')^2}{N}} \quad (77)$$

where $\sigma_{y.s}$ = the standard error of estimate
 Y' = predicted value of Y corresponding to X
 Y = the corresponding actual value of Y
 N = the number of cases in the series

Formula (77), however, involves computing the Y' value for each X and finding the discrepancy between this value and the actual value, squaring, and summing for all predicted values. If we are not interested in individual errors of estimate, we can use the following formula to find $\Sigma(Y - Y')^2$, and then divide this by N and extract the square root to get $\sigma_{y \cdot x}$. Thus

$$\begin{aligned} \Sigma(Y - Y')^2 &= \Sigma y^2 - \frac{(\Sigma xy)^2}{\Sigma x^2} & (78) \\ &= 1,486.88 - \frac{(1,419.36)^2}{4,993.92} \\ &= 1,486.88 - 403.41 \\ &= 1,083.47 \end{aligned}$$

Then we find $\sigma_{y \cdot x}$ by formula (77). Thus

$$\begin{aligned} \sigma_{y \cdot x} &= \sqrt{\frac{1,083.47}{32}} \\ &= \sqrt{33.8584} \\ &= 5.82 \end{aligned}$$

An even simpler method of deriving the standard error of estimate when we have the correlation coefficient available is to use the following formula.

$$\begin{aligned} \sigma_{y \cdot x} &= \sigma_y \sqrt{1 - r^2} & (79) \\ &= 6.82 \sqrt{1 - (.52)^2} \\ &= 6.82 \sqrt{1 - .2704} \\ &= 6.82 \sqrt{.7296} \\ &= (6.82)(.85) \\ &= 5.80 \end{aligned}$$

The value 5.80 calculated by means of formula (79) differs slightly from the value 5.82, calculated by means of formula (77) as a result of the errors of rounding involved in the calculations. Formula (79) is valuable in that it enables us to see how we can evaluate our predictions. If there

had been no relationship at all between the two variables, then r would be zero, the value under the radical would be 1, and the standard error of estimate would therefore be equal to the standard deviation of the entire Y distribution. The other limiting case would arise if the relationship between X and Y were perfect. In this instance r^2 would equal 1 and the value under the radical would be zero; hence the standard error of estimate would be zero also. This is as it should be, for with a perfect correlation we could predict precisely the corresponding Y value for any given value of X with no errors; on the other hand, with no correlation present, our best prediction is the mean of the Y distribution and our errors would be measured by the standard deviation of the Y distribution.

f. **The index of forecasting efficiency.** In the present case we can get an indication of the reduction in errors of prediction by comparing the standard error of estimate to the standard deviation of the Y distribution. The standard error of estimate is 5.82. This is a reduction of 1.00 point compared to the standard deviation, which is 6.82. In terms of per cent reduction of error we have $1.00/6.82 = 15$ per cent. We have reduced the extent of our errors of prediction, in other words, from knowledge of the correlation between the two variables by 15 per cent.

You may note that the index of forecasting efficiency, in this instance 15 per cent, can be obtained with reference to the size of the correlation coefficient alone. The formula is

$$\begin{aligned}
 E &= 1 - \sqrt{1 - r^2} & (80) \\
 &= 1 - \sqrt{1 - (.52)^2} \\
 &= 1 - \sqrt{.7296} \\
 &= 1 - .85 \\
 &= .15 \\
 &= 15 \text{ per cent}
 \end{aligned}$$

g. **Predicting in the opposite direction.** Ordinarily we are interested in predicting in one direction only, but we

could, if we were so interested, make predictions of X values from values of Y . When $r = 1.00$, the two regression lines coincide, but when r is less than 1.00, the two regression lines will be different, and if we are to predict X values from given values of Y , then we must do so from the line showing the regression of X on Y . Our procedure would be the same as that which we have already outlined, except that the regression coefficient would now be

$$b_{xy} = \frac{\Sigma xy}{\Sigma y^2} \quad (81)$$

or

$$b_{xy} = r \left(\frac{\sigma_x}{\sigma_y} \right) \quad (82)$$

and the regression equation would be

$$X' = b_{xy}(Y - M_y) + M_x \quad (83)$$

or

$$X' = b_{xy}Y + (M_x - b_{xy}M_y) \quad (84)$$

and the standard error of estimate would be

$$\sigma_{x,y} = \sigma_x \sqrt{1 - r^2} \quad (85)$$

h. Coding and regression coefficients. What we have said before about coding applies to the computations we used in finding the regression coefficients. If X and Y have been divided by the same coding constant, then we do not need to worry about decoding, since $i_x i_y$ in the numerator will cancel i_x^2 or i_y^2 in the denominator of formulas (73) and (81). If, however, X and Y have been coded by dividing by different constants, then we must decode as follows

$$b_{yx} = \frac{\Sigma xy}{\Sigma x^2} = \frac{\left[\Sigma x'y' - \frac{(\Sigma x')(\Sigma y')}{N} \right] i_x i_y}{\left[\Sigma x'^2 - \frac{(\Sigma x')^2}{N} \right] i_x^2} \quad (86)$$

3. PREDICTING FROM NONRECTILINEAR RELATIONSHIPS

We have still another problem of prediction to consider, one that is something of a mixture of the two problems we

have already considered. The regression coefficient and regression equation we have just discussed are based, you recall, upon the assumption that the relationship between *X* and *Y* is rectilinear. What predictions can we make when this assumption is not warranted, i.e., when the relationship between *X* and *Y* is curvilinear? Our procedure can best be illustrated by a concrete problem. Suppose we had 129 items we contemplated using in a questionnaire. These items have been scaled by the Thurstone technique (89), the details of which are unimportant here. For each item, let us suppose that we also have a measure of ambiguity, *Q*, and a measure of scale value, *S*. Our problem is to determine the most probable *Q* value from knowledge of the scale value of the item. Our first step is to plot these paired values in a correlation table as in Table 62.

In the absence of any knowledge at all of the relationship between scale and *Q* values, our best prediction for each of the items would be the mean *Q* value of the entire dis-

TABLE 62.—PREDICTION OF *Y* (*Q* VALUES) FROM *X* (SCALE VALUES) WHEN THE RELATIONSHIP BETWEEN THE TWO VARIABLES IS CURVILINEAR

<i>Q</i> VALUES (<i>Y</i>)	SCALE VALUES (<i>X</i>)								
	0	1	2	3	4	5	6	7	8
4.0-4.4			1	1	1	2			
3.5-3.9		1	1	1	4	5	4	2	
3.0-3.4		4	3	1	6	3	2	6	2
2.5-2.9		3	2		2	2	1	3	
2.0-2.4	4	3	2					3	3
1.5-1.9	8	6			1				5
1.0-1.4	6								9
.5-.9	3								11
.0-.4	1								1
Means of columns	1.45	2.44	3.04	3.70	3.24	3.49	3.42	2.95	1.30

Mean of entire *Y* distribution = 2.36

tribution. This prediction follows from the principle of least squares which, in the present case, means that we wish to predict values which will give us the smallest sum of squared deviations, i.e., the smallest discrepancy between predicted and observed values. The standard deviation of the complete distribution of Q values would give us a measure of the extent of our errors of prediction in this instance.

But when we have the items classified according to scale and Q value, as in Table 62, we note that the means of the columns differ from the mean of the entire distribution. Could we not predict more accurately for items of a given range of scale values (items in a given column) if we predicted the mean Q value of the column rather than the mean of the entire distribution? If we predict as the most probable Q value for each item in column 0 the mean Q value of that column, then the sum of squared deviations of the items in the column from the mean would give us a measure of our errors of prediction for the single column. If we repeat this procedure for the items of column 1, column 2, and so on, and divide the total sum of squared deviations within groups by N , we could get an over-all measure of our errors of prediction. Thus

$$\sigma_{y \cdot z}^2 = \frac{\Sigma y_0^2 + \Sigma y_1^2 + \Sigma y_2^2 + \Sigma y_3^2 + \dots + \Sigma y_k^2}{N} \quad (87)$$

You may recognize $\Sigma y_0^2, \Sigma y_1^2, \Sigma y_2^2, \dots, \Sigma y_k^2$ as being the *sum of squares within groups* of analysis of variance. We did not apply analysis of variance to these scores when they were presented earlier in Table 22, page 111. If we had, we could get the sum of squares which we need in formula (87) by subtracting the sum of squares between groups (columns) from the total sum of squares. But the solution for the standard error of estimate is even simpler if we know the value of the correlation ratio, since

$$\sigma_{y \cdot z} = \sigma_y \sqrt{1 - r_{yz}^2} \quad (88)$$

Substituting in formula (88) with the value of σ_y and η^2_{yx} taken from Table 22, we get

$$\begin{aligned} \sigma_{y \cdot x} &= 1.07 \sqrt{1 - .6911} \\ &= 1.07 \sqrt{.3089} \\ &= (1.07)(.556) \\ &= .595 \end{aligned}$$

The value of $\sigma_{y \cdot x}$, .595, provides us with an estimate of our errors of prediction when we predict as the most probable Y value for any given value of X , the mean of the Y column in which the X value is classified. We have reduced the extent of the errors of prediction from the standard deviation of the entire Y distribution, 1.07, to the value given by $\sigma_{y \cdot x}$, .595. This represents a reduction of .475 points, or an improvement of $.475/1.07 = 47$ per cent.

PREDICTIONS AND THE EVALUATION OF PREDICTIONS

Example 1.—A study of 100 women who thought their marriage was a success and a study of 100 women who thought their marriage was unsuccessful revealed a differential in response to the question: Did you have a happy childhood? The data are as follows:

CHILDHOOD STATUS	MARITAL STATUS		TOTAL
	Unsuccessful Marriage	Successful Marriage	
Happy.....	40	70	110
Unhappy.....	60	30	90
Total.....	100	100	200

(a) What per cent of the predictions would be correct without a knowledge of the response to the question concerning childhood status? What per cent would be correct with a knowledge of the response to this question? To what extent is the efficiency of prediction increased by this knowledge?

(b) Reverse the table and determine the efficiency of prediction of response to the question concerning childhood status without knowledge of marital status and with knowledge of marital status. To what extent is the efficiency of prediction increased by this knowledge?

Example 2.—The data below, adapted from Newcomb (69), show student preference for presidential candidates in the 1936 election arranged according to college status. Determine the efficiency of prediction without a knowledge of student status and with this knowledge. To what extent is the efficiency of prediction increased by this knowledge?

COLLEGE STATUS	STUDENT PREFERENCE		TOTAL
	Against Roosevelt	For Roosevelt	
Juniors—Seniors.....	24	28	52
Sophomores.....	23	17	40
Freshmen.....	37	15	52
Total.....	84	60	144

Example 3.—Dorcus (10) had an industrial concern select two extreme groups of workers, a “satisfactory group” and an “unsatisfactory group.” Each member of both groups was then given the Humm-Wadsworth Scale and predictions were made on the basis of scores on the scale as to the group in which the individual belonged. The results were as follows:

HUMM-WADSWORTH SCALE	COMPANY RATINGS		TOTAL
	Unsatisfactory	Satisfactory	
Satisfactory.....	6	18	24
Unsatisfactory.....	16	8	24
Total.....	22	26	48

Determine the efficiency of prediction without a knowledge of Humm-Wadsworth score, but assuming you know the number of individuals placed by the company ratings in each of the two groups. Determine the efficiency of prediction with a knowledge of Humm-Wadsworth score and the increase in efficiency of prediction with this knowledge.

Example 4.—Kratwohl (56) reports the following data on relationship between classification on the ACE test of academic aptitude and grades in the social sciences. Determine the efficiency of prediction without and with knowledge of ACE classification. What is the increase in efficiency of predictions with knowledge of ACE classification?

ACE CLASSIFICATION	GRADES IN SOCIAL SCIENCE			TOTAL
	D and E	C	A and B	
More than 75.....	7	30	39	76
25-75.....	62	59	30	151
Less than 75.....	49	19	3	71
Total.....	118	108	72	298

Example 5.—In an earlier problem (Example 7, Chapter 5, page 104) you found the coefficient of correlation between scores on a test of subject matter and a test of vocabulary in psychology.

- (a) Find the regression coefficients: b_{vz} and b_{zv}
- (b) Using the regression equation, predict the most likely score on the test of vocabulary (Y) for the following scores on the test of subject matter (X):

If X	48	55	73	82	90
Then Y'					

- (c) Using the regression equation, predict the most likely score on the test of subject matter (X) for the following scores on the test of vocabulary (Y):

If Y	58	71	76	80	95
Then X'					

- (d) Find the standard errors of estimate, $\sigma_{y \cdot x}$ and $\sigma_{x \cdot y}$, and interpret them.

Example 6.—In an earlier problem (Example 1, Chapter 6, page 128) you found the correlation ratio of vocabulary test scores (Y) on chronological age (X). Find the standard error of estimate for the same data, i.e., $\sigma_{y \cdot x}$, and interpret it.

CHAPTER 14

RESEARCH AND EXPERIMENTATION

Research problems consist essentially of defining issues operationally, and collecting, analyzing, describing, and interpreting data as they bear upon the particular issue or question. Thus any question on which it is possible to collect data can be the basis of a research problem. To be sure, as Lynd (66) has stated, not all questions and consequently not all research problems are of equal significance. But the exact significance of a question is not always an easy matter to determine beforehand, as any trained researcher knows. Many problems, which may at first thought seem relatively insignificant, may later lead to very important contributions to knowledge.

The following questions illustrate something of the scope of research activities in the behavioral sciences. Will children, on the average, work harder when they are praised than when they are criticized (45)? Is there any relationship between grades earned in college and scores on a college entrance examination (33)? Does frustration result, on the average, in aggression or regression or both (4, 9)? Will one method of teaching mathematics result in greater average achievement upon the part of students than another method (6)? Do students, on the average, learn just as much from straight lectures as they do from discussion groups (100)? In terms of average achievement, are small classes to be preferred to large classes (100)? Are individuals who are honest in one situation likely to be honest in other situations (67)? Are personality traits related to color of the hair (71)? What is the greatest source of anxiety for college students (18)? Do we tend to suppress experiences which are unpleasant (17)? Is there any real difference between the results obtained by the counseling procedures used in

"nondirective" therapy and "directive" therapy (78)? To what extent can attitudes be changed as a result of viewing motion pictures (75, 79)? To what extent do "stereotypes" determine our responses to social issues (16)? How can children's fears be eliminated most effectively (49)? To what extent can children's intelligence test scores be modified by changes in the environment (84)? Do the attitudes we have toward various political concepts influence the meaning which these concepts have for us (12, 13)?

The *posing of a question*, similar to those just listed, is the first step in research. Questions, when properly phrased, become hypotheses which can be subjected to empirical test. Once a question has been formulated, the next step is planning the experimental design. This consists of determining what data will need to be collected, the manner in which they will be collected, and the methods by which they will be analyzed. The third step is actually carrying out the research and analysis. The final steps are interpreting the analyzed data and seeing that the results are then made available to other investigators.

Since it is the initial step which most often proves to be a stumbling block for the student called upon to undertake a research project as part of his academic training, we might examine briefly some of the sources of problems suitable for investigation. One of the most fruitful sources of hypotheses is to be found in the theories advanced by various writers in a given field. In psychology, for example, the learning theories of Hull (44), Tolman (91), and Guthrie (40) have been the starting point for many experiments. Allport's (1) theory of functional autonomy in the field of motivation has been another source of research problems (76). Sears (81) has summarized many experiments which had their origin in the psychoanalytic theories of Sigmund Freud. Likewise, *Gestalt* theory (55) and the variation of *Gestalt* theory advocated by Lewin (58, 59) have resulted in many important research projects.

Students who have been indoctrinated with critical atti-

tudes and who read challengingly the literature in their field of specialization will not fail to find innumerable projects demanding investigation. Students who read with blind acceptance of authority and the printed word, on the other hand, are apt to overlook assertions and statements made by writers which are not based upon research evidence. Such assertions and statements, when questioned, may very well become the basis of research problems for the critical reader. The proverb "The grass looks greener on the other side of the fence" and the many variations of this proverb, when questioned, became the basis of a series of ingenious experiments for Wright (101). And Irwin's (46) questioning attitude applied to certain aspects of Wright's research led to additional experimentation.

1. INTERPRETATION OF TESTS OF SIGNIFICANCE

Once a problem has been selected for investigation, if the investigation requires a test of significance—and this will usually be the case—it is important that we understand clearly the kind of hypothesis which can be tested by means of the χ^2 , t , ϵ^2 , or F tests. Despite the fact that the questions or hypotheses raised earlier were stated in such forms as "Is there any relationship between . . .," "To what extent can attitudes be changed by . . .," "Is there any difference between . . .," and so forth, these questions must be restated when it comes to the application of tests of significance. To be sure, it is in the forms just given that the investigator usually gets his hunch or hypothesis, i.e., it is usually the investigator's opinion that one method is more effective than another, or that attitudes can be changed by motion pictures, or that there is some relationship between two variables. But, as McNemar has pointed out in the case of a mean difference, "Regardless of the experimental hunch or hypothesis, the only workable statistical hypothesis is that no difference exists between the universe means" (68, pp. 336-337).

The hypothesis of no difference is the familiar *null hypothesis* and applies not only to differences between means. It is stated in a variety of ways depending upon the particular investigation.¹ If we are testing the difference between the means of achievement scores of children taught by one method and children taught by another method, our working hypothesis would be that there is no difference between the means of the populations of which our two groups are assumed to be samples. Sometimes we state our working hypothesis in this form: Two samples have been drawn from the same population; consequently there is no difference between them other than would be expected as a result of sampling variation. In testing the significance of an observed correlation coefficient, the hypothesis may be that the correlation in the population is zero, or that the sample value of r does not differ significantly from some other specified value of the population r . The same sort of hypothesis is applied to the means of several samples being tested by the F test. Our working hypothesis is that they have all been drawn from the same population or populations with a common mean and variance.² The calculation of χ^2 tests the same null hypothesis: That there is no difference between our observed and expected frequencies or that the observed sample data have been drawn from a population distributed according to the expected frequencies.

Our tests of significance are based on the assumption that

¹ A hypothesis which is set up with the possibility of being rejected at some designated level of significance is called a null hypothesis, the term "null" referring to our interest in possible rejection rather than to the fact that the hypothesis is stated negatively. Fisher has emphasized that "every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis" (26, p. 16).

In a broad sense, the term is used by some statisticians to refer to any eligible hypothesis, that is, any hypothesis which is exact and free from vagueness, so that it may form the basis for the sampling distribution used in making a test of significance. (See the note by T. A. Bancroft in *Biometrics Bulletin*, 1945, 1, p. 38.)

² More precisely, the hypothesis is that the samples have been drawn from the same or identical populations, since the assumption is that not only the means and the variances are the same, but that the populations are, in addition, normally distributed.

the null hypothesis is true. If we then obtain a value of t , F , e^2 , or χ^2 such that it would occur on the basis of sampling variation alone 5 per cent or less of the time, this may lead us to reject the null hypothesis at a defined level of significance.³ And if the null hypothesis is not tenable, this tells us that we would not expect to get the results we have obtained on the basis of sampling variation alone. If our results are not due to sampling errors, then this fact may, through logic and insight, lead us to infer that some difference between our experimental groups does exist and that this difference is attributable to one of our variables. But this inference must come from our experimental design and from analysis of our variables, not from tests of significance. Should the value of t , for example, be such that it could be expected to occur much more frequently than the 5 per cent of the time which we have agreed to regard as significant, then we must assume that the null hypothesis is tenable, i.e., that we have no evidence against it. But observe that the statement that we have no evidence against the hypothesis does not prove it to be true; it merely means that this particular sample offers insufficient evidence for rejecting it.

The insistence upon this form of statement of a statistical hypothesis, we agree with McNemar, is not "mere quibbling." If the null hypothesis is true, then successive repetitions of the research would give a sampling distribu-

³ It was pointed out previously, but it is worth repeating, that the null hypothesis which is tested is not set up necessarily because we believe it to be true, but so that it may be considered for possible rejection at a designated level of significance. On the assumption that the hypothesis is true, then the sampling distribution of the statistic under consideration may be used to determine the probability that random sampling from the population for which the hypothesis holds would yield statistics deviating from the parameter as much as the sample one does. Since the null hypothesis specifies the frequencies with which the different results of an experiment may occur, we may also divide these results into two classes, one of which shows a significant discrepancy or deviation from the null hypothesis, and the other no significant discrepancy or deviation from this hypothesis (26, pp. 16-17). "If these classes of results are chosen, such that the first will occur when the null hypothesis is true with a known degree of rarity in, for example, 5 per cent or 1 per cent of the trials, then we have a test by which to judge, at a known level of significance, whether or not the data contradict the hypothesis to be tested" (26, p. 182). Cf. also the discussion in Walker (96, pp. 290-292).

tion of differences which would have a mean of zero. The standard deviation of this sampling distribution is estimated from the standard error of the difference we have calculated. By reference to the table of t , we can then make a statement of the probability of obtaining a value of t as large as the one we have obtained. Under no circumstances, however, can we make any statement of probability that the true population difference is some specified value. Population parameters, though single, fixed values, usually remain unknown.

But despite the limitations imposed by the fact that population values are unknown, we sometimes lose sight of the fact that by establishing the fiducial limits we may infer something as to the *degree* of the difference between two means. That is, a simple test of significance may lead us to reject the hypothesis of no difference, but the fiducial limits permit us to determine, again with a given degree of confidence, whether we may consider as tenable a hypothesis that the population mean difference is within a certain range.

2. SAMPLES AND RESEARCH

Since most research deals with samples, we may expand our earlier discussion to take into consideration certain aspects of sampling which we neglected. When, as is generally the case, we are interested in making inferences from statistics derived from samples about the population or universe, certain assumptions are necessary. We must assume that our sample is representative of the population from which it was drawn. We may, however, define this population in various ways. We may start with a definition of a population of males as consisting of all males in the world. We may limit this definition by restricting our population to all males of a given age in a given country; all males of a given age in a given state; all males of a given age in a given college; and so on. If we have conducted an experiment with a random sample of male college students

selected from a given college, we can safely generalize from tests of significance about the population of college students in general only if we can make the assumption that our sample is representative of college students in general. Generalizations from samples must always, in other words, be considered as applying only to the population of which the samples are representative.

Samples used in experimentation and research, and consequently statistics derived from these samples, are subject to two kinds of errors: *constant errors* and *errors of random sampling*. Constant errors are errors which bias statistics in one direction. The Literary Digest poll of 1936, which predicted a Republican victory on election day, illustrates what happens when constant errors are present in a sample. One of the reasons why the Digest poll was in error was that much of Roosevelt's support lay in the lower income groups which were not represented on the lists of telephone subscribers and automobile owners from which the Digest largely selected its sample. The results of the poll consequently were biased. The sample was not representative of the voting population, and the inaccuracy of the generalization concerning the outcome of the election, based upon the sample, was evident to polling experts long before election day arrived (29). The important point in this connection is that the standard error formulas, upon which statistical tests of significance are based, do not provide any estimate of the direction or magnitude of constant errors. Constant errors can possibly be detected by a logical examination of the manner in which the sample was selected.

Only the errors of random sampling, the second source of errors to which samples are subject, can be estimated by standard error formulas. The standard error of a statistic estimates the variation to be expected of the statistic from sample to sample when the successive samples are of the same size and randomly selected from a given population. A random sample is customarily defined as a sample in

which each individual in the defined population must have an equal chance of appearing.⁴ Such a sample, if possible to obtain, would insure the fact that the sample is representative of the population. But strictly speaking, it is obvious that, according to this definition of randomness, the samples dealt with by psychologists, educators, sociologists, and others in the social sciences are never random. For all individuals in any large population never have an equal chance of appearing in the sample selected for investigation.

In an experiment, for example, involving a comparison of two teaching methods for sixth-grade children, the subjects being two classes in a given school, could we generalize from the sample results concerning the effectiveness of the two methods for sixth-grade students in general? If we could assume that our subjects constitute a random sample of the population of sixth-grade students, then we could generalize about this population. But our subjects could be considered a random sample only if every sixth-grade student had an equal opportunity to appear in the sample. Obviously, this is not the case.⁵

The problems of random sampling which are faced by the investigator in the social science fields, however, are not peculiar to these fields alone. In the same sense that individuals in a large population never have an equal chance of appearing in samples under investigation, the samples used by geologists, agronomists, biologists, physicists, chemists, and engineers are not, strictly speaking, random samples. A geologist may examine a sample of ore from a given area, but does every possible unit of ore have an equal

⁴ Snedecor contends, however, that for tests of significance one can specify that, after having decided upon the size of the sample to be used, the only requirement which needs to be imposed is that "every possible sample of n individuals shall have the same chance of being drawn" (86, p. 389). Note that this says nothing about individuals.

⁵ We could, of course, define the population about which we were generalizing as consisting not of individuals, but of intact sixth-grade classes in schools of a given size. We could then select by chance the schools and the classes in the schools in which we carried out our experiment and assume that our samples were representative of this defined population.

opportunity of appearing in the sample? An industrial engineer testing the tensile strength of thread or some other material in a plant, regardless of how the sample of thread is selected, does not have a random sample in which every unit in the population has an equal opportunity to appear. The sample of thread he studies can be a random sample only of the population of thread already manufactured. Thread which has not yet been produced does not have an equal opportunity to appear in the sample. Yet this does not prevent the industrial engineer from assuming that his sample is a random sample and that it is representative of all the thread that is currently coming.

The same is true of the agronomist who conducts research with a variety of corn grown in a given soil mixture. He may have a random sample of the population of corn grown this year, but he obviously does not have a random sample of the corn to be grown *next* year, for this corn does not have an equal opportunity to appear in the sample. But, again, this does not prevent the agronomist from dealing with his sample as if it were random, nor does it prevent him from assuming that his sample is representative of the corn to be grown in the future. From the sample he has he conducts tests of significance and generalizes on the basis of them about the corn which is yet to be grown.

The problem here, it seems, is one of an ideal. As an ideal, a random sample is to be regarded in much the same fashion as Boyle's law in physics, which gases are assumed in theory to obey, but which in practice they do not. The concept of the ideal is a useful and convenient way of looking at the behavior of gases. The failure of gases to satisfy the ideal does not prevent physicists and chemists from experimenting and making generalizations on the basis of it.

In the behavioral sciences we may recognize and acknowledge that we never have perfectly random samples of a large population. But, if we are willing to make the necessary assumptions, we find that in many, many instances the assumptions are justified in terms of practical considera-

tions. Much has been said against the samples used in behavioral research—and that many studies have violated even simple precautions to insure representativeness is not to be denied—but generalizations based upon samples in which care has been taken to eliminate bias and which are then treated *as if* they were random samples from larger populations have been found to be sound and useful. The fact is that no one has ever studied the correlation between college grades and tests of academic aptitude in a strictly random sample of college students from the population of all college students. Yet generalizations have held up that have been made on the basis of available correlation coefficients derived from samples and tests of significance have been applied to these coefficients as though they were derived from random samples.

McNemar, in the article to which we have referred previously, summarizes the problem of representativeness in this way:

In the absence of any rule-of-thumb method for checking representativeness in psychological research, the investigator must resort to logical considerations. If the sample has been drawn by some mechanical means or by stratifying the universe on the basis of pertinent facts, one can feel fairly sure that the sample is representative. In the absence of an obviously valid scheme for drawing the sample, the only thing one can do is to describe the sample as completely as possible with regard to known characteristics of the universe from which it was drawn. If the sample is typical of the universe in several variables which are related to the variate being studied, it is safe to assume that it is representative. This reasoning is, of course, posterior use of the principles of stratified sampling. The importance of fully describing the sample and how it was drawn cannot be overemphasized. Without such information it is impossible to evaluate a given research. (68, p. 384.)⁶

⁶ For a discussion of stratified samples, the article by McNemar (68) should be consulted as should Stock (88). The use of tables of random numbers for mechanically selecting samples will be discussed in a later section.

3. SIZE OF THE SAMPLE

It would not be difficult to find many articles stressing the need and value of large samples in psychological and educational research. Indeed, as McNemar points out:

Some psychologists frown upon the use of small samples, as, for example, N less than 25; a few use such small samples, but scorn the necessity of evaluating their results in terms of the mathematics of small samples . . . while others will rightfully argue that when small samples, properly evaluated, yield a difference which would arise by chance only once in a hundred times, the result is just as dependable as if the same chance figure had been found for large samples. It is assumed in either case—small or large sampling—that the sampling technique is such as to avoid bias. It is commonly and erroneously thought that some magic lies in large samples and that bias is less apt to be present. The larger the sample, the greater the precision so far as random errors are concerned, but it does not follow that bias is avoided by increasing the size of the sample. (68, p. 340.)

If, in comparing the means of two independent samples, each consisting of 15 cases, for example, we obtained a t of 2.76, which is significant at the 1 per cent level for 28 degrees of freedom, we could reject the hypothesis being tested with the same degree of confidence that we would have if the samples had consisted of 100 cases each and the value of t obtained was 2.60. Our confidence, in other words, is not increased in rejecting a hypothesis at the 1 or 5 per cent levels of significance when we are dealing with samples of 100 cases each, over what it would be if the same hypothesis were rejected at the 1 or 5 per cent levels for samples of 15 cases each.

Let us suppose, however, that in a given experiment we are testing the difference between the means of two samples of 25 cases each and that our computed value of t just fails to meet the required value at the 5 per cent level of significance. Now if we assume that the difference between the means will remain fairly constant, regardless of the size

of the samples, then we would, of course, expect to get a significant value of t if the size of the samples were increased. We could, in other words, repeat the experiment, if this were feasible, with samples sufficiently large to insure more or less that we would obtain a significant value of t .⁷

With respect to this persistent question of how many cases one should use, there is no one set answer. But we agree with McNemar "that the demonstration of a difference (or effect) which is large enough to possess any practical or social significance will not require large samples; certainly, a difference which is so small as to require 1000 cases in each sample to demonstrate it is apt to possess little psychological meaning" (68, p. 340).

4. CONTROL GROUPS

Suppose that an investigator reported the following experiment: At the beginning of the term in a Latin class, the members were given an English vocabulary test. At the end of the term the test was repeated and the difference between the means of the first and second tests is significant at the 1 per cent level. What conclusions may be drawn from this experiment? Can we assume that the study of Latin increases English vocabulary? This might be the case, but the experiment described does not establish the fact. All that we can conclude is that there has been a significant increase in the mean English vocabulary score, but we do not know whether this is largely due to the Latin course or not. The reason that we cannot draw the conclusion concerning the effects of the Latin course is that the acquisition of English vocabulary is a *growth* or achievement process that may have occurred, and probably did to some extent, in students who were not members of the Latin class. If the test had been given to an arithmetic class we would expect this group to show some increase in vocabulary also.

⁷ See, for example, the discussion in Peters and Van Voorhis (74, pp. 469-470).

In this and every other experiment where the changes in behavior we are measuring might possibly occur as a result of factors external to the experiment proper, a control group which does not experience the experimental variable is necessary for evaluating the changes observed in the experimental group. If we fail to take the precaution of having a control group, then we have no way of determining how much of the change in the experimental group can be attributed to the experimental variable, in this instance the Latin course, and how much to other unknown factors.

a. Control by random selection. Control groups are sometimes formed in research by dividing the total number of subjects available into two groups by some method of random selection. Each subject might be assigned a number, the numbers placed on discs, put into a box, and thoroughly mixed. The numbers are then drawn out one at a time, the first number being assigned to one group and the second number to the second, the third to the first, the fourth number to the second, and so on, until the numbers in the box are exhausted. Then by flipping a coin one of the two groups could be assigned to the experimental condition and the other could serve as a control.

A still more efficient method of random selection, however, is to make use of a *table of random numbers* (27, 72). These tables consist of numbers arranged at random in columns and rows. The tables can be used by entering at any point and by reading in any direction, down or up, right or left, or obliquely.

Let us suppose that we wished to divide 80 subjects into two groups of 40 each. We first number our subjects from 00 to 79. To determine where to enter the table, we might close our eyes and put our pencil on the table. Suppose we then open our eyes and find that the pencil is resting, let us say, on column (5) and row (26). It makes no difference, once the point of entry has been determined, in which direction we read. Let us assume that we are going to read downward. Since we need two place numbers, we obtain

these by combining the digits in adjoining columns. We read down the column until we have 40 unlike numbers below 80. We skip any number which is 80 or above and any number which is a repetition of a number previously read. Going down columns (5) and (6) of Table 63, which is a section of the table of random numbers, Table G, pp. 340-341, we would have 31, 14, 17, and 22. We would skip 99, since we have no subject who is assigned a number as large as this. If 99 were followed by 31, we would skip

TABLE 63.—A BLOCK OF RANDOM NUMBERS SELECTED FROM TABLE G

Row	COLUMNS								
	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
26	3	1	7	7	5	2	2	3	4
27	1	4	8	5	7	0	9	6	4
28	1	7	3	8	0	3	6	2	3
29	2	2	0	9	7	2	3	9	2
30	9	9	5	6	9	8	2	8	0

this number also since the subject assigned number 31 has already been selected. When we have reached the last row in the table we may continue to read numbers by going up or down the adjoining columns, for example, columns (7) and (8), until we have 40 unlike numbers below 80.

The individuals with the numbers corresponding to the first 40 numbers below 80 that we have read from the table of random numbers would constitute one of our groups and the remaining 40 individuals would constitute the second group. In a similar fashion we could divide a large group of subjects at random into any number of smaller groups. Tables of random numbers can also be used for selecting at random a single small group of subjects from a larger total, and for assigning groups at random to one of a number of given experimental conditions.

When individuals have been assigned to control and experimental groups in the manner just described or by some other method of random selection, then the proper formula for the standard error of the difference to use in evaluating the difference between the means of the experimental and control groups is formula (50) or formula (61).

b. **Control by matching individuals.** It is obvious, however, that many experiments will not permit the random selection of an experimental and control group and, in many instances, greater precision in terms of a reduced standard error of the difference can often be achieved by not making use of groups selected at random. In the experiment cited on the effect of Latin on English vocabulary, for example, it would not be possible to assign subjects to the Latin class and to a control group at random. We must deal with the Latin class as it stands. In this instance we could establish a control group by pairing with every member of the Latin class a non-Latin student with a similar English vocabulary score. At the end of the experiment we repeat the English vocabulary test with the experimental and control groups. We thus have an initial and final test score for each individual. We could now test the difference between the means of the initial and final test for the Latin group. But since correlation is involved between the means we are comparing, we must make use of the difference formula,

$$\sigma_{m_d} = \frac{\sigma_d}{\sqrt{N - 1}}$$
, to find the standard error of the mean difference. We could, of course, compute the standard error of the difference in terms of formula (48), if we first found the correlation coefficient between the initial and final scores for the Latin group, but working directly with the differences between scores involves less labor. In a similar manner we could determine whether the initial and final means of the control group differ significantly.

Suppose that we found a significant value of t in each comparison. This would mean that both the Latin group

and the non-Latin group show significant differences between the first and final tests. But we would, on the basis of the comparisons just made, have no way of knowing whether the study of Latin resulted in a greater mean increase in English vocabulary. It is in comparing the difference between the *mean gains* (or mean losses or simply mean changes, as might be the case in another experiment) that we would find the answer to this question. The standard error of the difference between the mean gains can be found quite simply when we have paired the members of our experimental and control groups.

We first find the difference between the initial and the final test for each member of the two groups. We thus have a difference score for each member of the two groups, and these scores are shown in columns (4) and (7) of Table 64. And since we have paired the members of each group,

TABLE 64.—THE COMPUTATION OF THE STANDARD ERROR OF THE MEAN DIFFERENCE IN GAINS FOR MATCHED INDIVIDUALS

PAIR (1)	LATIN (EXPERIMENTAL)			NON-LATIN (CONTROL)			DIFFERENCE IN GAIN (8)
	Initial (2)	Final (3)	Gain (4)	Initial (5)	Final (6)	Gain (7)	
1	35	42	7	35	36	1	6
2	42	46	4	43	48	5	-1
3	28	28	0	27	25	-2	2
4	32	31	-1	30	32	2	-3
5	37	40	3	38	37	-1	4

the difference scores are also paired. All that we need to do is to find the difference between each pair of difference scores. These scores, listed in column (8), will represent the differences in gains or losses for each pair of subjects. We then find the standard deviation of this distribution. This standard deviation divided by the square root of $N - 1$, where N equals the number of pairs, will give us the

standard error that we need for evaluating the difference in mean gain in English vocabulary for the two groups. Thus

$$\sigma_{m_d} = \frac{\sigma_{d_i}}{\sqrt{N-1}} \quad (89)$$

Perhaps you have noticed from Table 64 that the matched pairs do not have precisely the same scores on the initial test. Indeed, it is seldom that we can match the members of the two groups so that they have *exactly* the same initial scores. We must expect to have some discrepancies. Peters and Van Voorhis (74, pp. 448-449) suggest that differences between the pairs on the matching variable as large as 5 to 10 per cent of the range of scores are permissible as long as they are balanced between the two groups so as to keep the means approximately equal.

Sometimes it is not feasible to pair individuals in the control and experimental groups on the variable on which we are going to make a test of significance, but we may be able to pair our subjects on the basis of some variable which we have reason to believe is correlated with the variable on which the two groups are to be compared. We might be able to pair the subjects on the basis of age, intelligence test scores, reading speed, or some other factor that has a fair degree of correlation with the variable under study. If we match subjects in the experimental and control groups on this basis, we may still make use of the difference formula for finding the standard error of the mean difference.

If we succeed in pairing on some variable so that we introduce a correlation coefficient as high as .75, then this has the effect of reducing the standard error of the difference by about one half (68, p. 354). We already know that the standard error of the difference can be reduced by increasing the number of subjects in each group. But we would have to quadruple the number of subjects to achieve this same reduction in the standard error of the difference for groups selected at random. If we can introduce a correlation co-

efficient of .50, this would reduce the standard error of the difference as much as would doubling the number of subjects for groups selected at random (68, p. 354). Pairing individuals on the basis of some *relevant* variable, we see, is obviously advantageous statistically and also from the point of view of experimental design. But if the subjects are paired on the basis of a variable which fails to be correlated very highly with the experimental variable, then the reduction in the number of degrees of freedom available for evaluating the t obtained may offset the slight reduction in the standard error of the difference. Our experimental design, in this case, would really be less efficient than if we had used groups selected at random.

c. **Control by matching groups.** Even when circumstances do not permit the pairing of individuals, it may still be possible to match groups. If we can match individuals this insures the matching of groups. We can, however, match or equate entire groups, without regard for individuals, by seeing to it that the two groups have approximately the same mean and standard deviation on the initial test on the experimental variable or, if this is not possible, on some other *relevant* variable. If the variable on which the two groups are matched is correlated with the variable under study, we may still take advantage of this condition to reduce the standard error of the difference. When groups have been matched on this basis, the proper formula to use in finding the standard error of the mean difference is formula (49).

d. **Single group serving as its own control.** Earlier in our discussion we pointed out that a single group cannot serve as its own control whenever there is any basis for assuming that the changes in behavior which we are measuring might possibly be the result of factors external to the experimental situation. We should point out now that if we have no reason to assume that external factors are important, then, of course, the experimental group may serve as its own control. For example, if we measured the attitudes of sub-

jects *just before* and *immediately after* they had seen a motion picture designed to change attitudes, we would have no need of a control group. We could logically assume that whatever changes in attitude occurred were directly attributable to the influence of the motion picture, since it is difficult to see what possible external factors might be at work in this particular experiment. When a single group serves as its own control, the presence of correlation must be recognized and taken into account in computing the standard error of the difference between means.

5. THE *t* TEST AND THE ASSUMPTION OF HOMOGENEITY OF VARIANCES

The experimental designs and problems we have been discussing concern largely the test of significance of a mean difference based upon the sampling distribution of *t*. Testing the significance of the difference between two sample means, by dividing the difference by the standard error of the difference and referring the obtained value to the table of *t*, involves, as we have said before, the assumption that the variances of the populations from which the samples were drawn do not greatly differ. If the variances of the populations do differ, and if a significant value of *t* is obtained, then we can still reject the hypothesis that the two samples were drawn from the same or identical populations. But we do not know, however, whether the two populations differ only with respect to variances or perhaps with respect to both means *and* variances.

Since our interest is usually in the difference between means, it is fortunate that in most *experimental* work variances will not differ significantly. But a test of this hypothesis, in case of doubt, is proper and may be easily made. We may determine whether the assumption of homogeneity of variances is justified by calculating the ratio between the two estimates of the population variance derived from the samples. That is, we first find the sum of squares within one of our samples and divide this by $n_1 - 1$; then we find

the sum of squares within the second sample and divide this by $n_2 - 1$. We then calculate the ratio between these two variances, always placing the larger variance in the numerator. The value obtained is F , and we enter the column of the table of F with the number of degrees of freedom corresponding to the larger variance and find the row entry corresponding to the degrees of freedom of the smaller variance. Since the calculated value of F will always be larger than 1, so that only one end of the F distribution is involved, the tabled value at the 1 per cent level will indicate the value at the 2 per cent level.⁸ Let us apply this test to the variances of the "lecture" and "project" groups mentioned earlier in connection with analysis of variance.

From the data of Table 36, we find that the estimate of the population variance based upon the scores of the "lecture" group is equal to

$$\frac{\Sigma x^2}{n_1 - 1} = \frac{14}{5 - 1} = 3.5.$$

For the "project" group, we shall have

$$\frac{\Sigma x^2}{n_2 - 1} = \frac{22}{5 - 1} = 5.5.$$

F will be equal to $\frac{5.5}{3.5} = 1.571$. By reference to Table E,

we find that an F of 15.98 will be required in order for us to reject the hypothesis of homogeneity of variances at the 2 per cent level for 4 and 4 degrees of freedom. Thus there is no reason to suspect that the two variances differ significantly.

Let us take a case where the two estimates of the population variances do differ significantly, but where we have the same number of subjects in each group. When we tested

⁸ This discussion and the following one are based upon the new, fourth edition (1946) of Snedecor's book (86) which appeared too late to be included in the list of references.

the difference between the means of attitude test scores of 50 Republicans and 50 Democrats (Table 35), we found a value of *t* equal to 6.9 which, for 98 degrees of freedom, was highly significant. If we apply the test for homogeneity of variances, however, we find that $\frac{\sigma_1^2}{\sigma_2^2}$ is equal to 2.25. According to Table E, an *F* equal to 2.25 slightly exceeds the tabled value at the 2 per cent point for degrees of freedom equal to 49 and 49. Thus we would reject the hypothesis of homogeneity of variances. But can we test the hypothesis that the means of the population from which the samples were drawn are equal, irrespective of variances? As long as the number of cases in each sample is the same so that $n_1 - 1$ equals $n_2 - 1$, the solution is fairly simple. Snedecor recommends that we merely enter the *t* table with degrees of freedom equal to $n_1 - 1$ or just half the number ($n_1 + n_2 - 2$) we have when the variances do not differ significantly. Thus, entering the table of *t* with degrees of freedom equal to 49, we find that our conclusion concerning the difference between the means remains unchanged. The obtained value of *t*, 6.9, still exceeds the 1 per cent point for 49 degrees of freedom.

What is the solution, however, if n_1 and n_2 differ greatly as do also the variances? Let us suppose that we have one sample of 10 cases and another of 30, and that the two estimates of the population variances are 27.04 and 5.76, respectively. *F* would be equal to $\frac{27.04}{5.76}$ or 4.694, a highly significant value for degrees of freedom equal to 9 and 29, and the hypothesis of homogeneity of variances would be rejected. Can we, under these circumstances, test the hypothesis that the population means are equal without any hypothesis concerning the variances? Snedecor reports an unpublished paper by Cochran and Cox in which an approximate method is proposed for testing this hypothesis.

We calculate the sum of squares within each group and

find the estimate of the population variances by dividing the sum of squares for the first group by $n_1 - 1$ and that for the second group by $n_2 - 1$. These two values for the case at hand are equal to 27.04 and 5.76, respectively. The variances of the two sample means are found by dividing the estimates of the population variances by the corresponding number of cases in the samples. For the first group the variance of the mean would be equal to $\frac{27.04}{10} = 2.704$, and for the second group we would have $\frac{5.76}{30} = .192$.

Let us suppose that the difference between the means is equal to 4.2 and the obtained value of t will then be equal to

$$\frac{M_1 - M_2}{\sigma_{m_d}} = \frac{4.2}{\sqrt{2.704 + .192}} = \frac{4.2}{1.7} = 2.47$$

To find out whether this value is significant at the 5 per cent level, we find the tabled value of t at the 5 per cent level for both $n_1 - 1$ and $n_2 - 1$. For the first sample, with 9 degrees of freedom, t at the 5 per cent point is equal to 2.262; for the second sample, with 29 degrees of freedom, the value of t at the 5 per cent point is equal to 2.045. These two values, which for convenience may be called t_1 and t_2 , are substituted along with the corresponding variances of the means of the two samples in the formula given below to find the approximate value of t required at the 5 per cent level of significance. Thus

$$5\% \text{ level} = \frac{(\sigma_{m_1}^2)(t_1) + (\sigma_{m_2}^2)(t_2)}{\sigma_{m_1}^2 + \sigma_{m_2}^2} = t$$

Substituting in the above formula, we obtain

$$5\% \text{ level} = \frac{(2.704)(2.262) + (.192)(2.045)}{2.704 + .192} = 2.248$$

The hypothesis tested by comparing the value of t (2.47), obtained by dividing the difference between the means by

the standard error of the difference, with the value of t (2.248) approximated above is that the means of the populations from which the two samples were drawn are equal, with no hypothesis concerning the population variances involved. Since the calculated value of 2.47 is greater than the approximated value of 2.248, we may reject the hypothesis at the 5 per cent level. We could, of course, have used the tabled values of t at the 1 per cent point instead of the 5 per cent point in the approximation formula, if we had wanted to find the value of t required at the 1 per cent level of significance.

6. *ADDITIONAL PROBLEMS IN EXPERIMENTAL DESIGN*

Similar problems of experimental design involving ϵ^2 , F , and χ^2 , could be treated in much the same manner that problems involving t have been treated. For a discussion of these, however, the student is referred to Peters and Van Voorhis (74), Lindquist (64), Fisher (25, 26), Snedecor (86), Tippett (90), and Goulden (35). The articles mentioned previously in connection with analysis of variance should be consulted also.

BIBLIOGRAPHY

1. Allport, G. W. *Personality*. New York: Holt, 1937.
2. Anastasi, A. *Differential psychology*. New York: Macmillan, 1937.
3. Baker, K. H. Pre-experimental set in distraction experiments. *J. gen. Psychol.*, 1937, 16, 471-488.
4. Barker, R., Dembo, T., & Lewin, K. Frustration and regression: An experiment with young children. *Univ. Ia. Stud. Child Welf.*, 1941, 18, No. 1.
5. Baxter, B. Problems in the planning of psychological experiments. *Amer. J. Psychol.*, 1941, 54, 270-280.
6. Brownman, D. E. Measurable outcomes of two methods of teaching experimental geometry. *J. exp. Educ.*, 1938, 7, 31-34.
7. Cheshire, L., Saffir, M., & Thurstone, L. L. *Computing diagrams for the tetrachoric correlation coefficient*. Chicago: University of Chicago Bookstore, 1933.
8. Curtis, J. W. A study of the relationship between hypnotic susceptibility and intelligence. *J. exp. Psychol.*, 1943, 33, 337-339.
9. Dollard, J., Doob, L. W., Miller, N. E., Mowrer, O. H., Sears, R. R., *et al.* *Frustration and aggression*. New Haven: Yale University Press, 1939.
10. Doreus, R. M. A brief study of the Humm-Wadsworth Temperament Scale and the Guilford-Martin Personnel Inventory in an industrial situation. *J. appl. Psychol.*, 1944, 28, 302-307.
11. Dunlap, J. W. Applications of analysis of variance to educational problems. *J. educ. Res.*, 1940, 33, 434-442.
12. Edwards, A. L. Studies in stereotypes: I. The directionality and uniformity of responses to stereotypes. *J. soc. Psychol.*, 1940, 12, 357-366.
13. Edwards, A. L. Four dimensions in political stereotypes. *J. abnorm. soc. Psychol.*, 1940, 35, 566-572.
14. Edwards, A. L. Political frames of reference as a factor influencing recognition. *J. abnorm. soc. Psychol.*, 1941, 36, 34-50.

15. Edwards, A. L. Rationalization in recognition as a result of a political frame of reference. *J. abnorm. soc. Psychol.*, 1941, 36, 224-235.
16. Edwards, A. L. Unlabeled fascist attitudes. *J. abnorm. soc. Psychol.*, 1941, 36, 575-582.
17. Edwards, A. L. The retention of affective experiences—A criticism and restatement of the problem. *Psychol. Rev.*, 1942, 49, 43-53.
18. Edwards, A. L. The development of an anxiety scale. *J. appl. Psychol.*, 1942, 26, 187-196.
19. Edwards, A. L., & Kenney, K. C. A comparison of the Thurstone and Likert techniques of attitude scale construction. *J. appl. Psychol.*, 1946, 30, 72-83.
20. Edwards, A. L. A critique of "neutral" items appearing in attitude scales constructed by the method of equal appearing intervals. *Psychol. Rev.* (In press.)
21. Enlow, E. R. Statistics in education and psychology. New York: Prentice-Hall, 1937.
22. Ezekiel, M. Methods of correlation analysis. New York: Wiley, 1930.
23. Ezekiel, M. "Student's" method for measuring the significance of a difference between matched groups. *J. educ. Psychol.*, 1932, 23, 336-450.
24. Ezekiel, M. Reply to Dr. Lindquist's "further note" on matched groups. *J. educ. Psychol.*, 1933, 24, 306-309.
25. Fisher, R. A. Statistical methods for research workers. (6th ed.) Edinburgh: Oliver & Boyd, 1936.
26. Fisher, R. A. The design of experiments. (3rd ed.) Edinburgh: Oliver & Boyd, 1942.
27. Fisher, R. A., & Yates, F. Statistical tables for biological, agricultural and medical research. Edinburgh: Oliver & Boyd, 1938.
28. Fosdick, S. J. Report to the National Retail Dry Goods Association, 1939. Quoted in Hartmann, G. W., & Newcomb, T. (Eds.) *Industrial conflict*. New York: Cordon, 1939, p. 119.
29. Gallup, G. H. A guide to public opinion polls. Princeton: Princeton University Press, 1944.
30. Gallup, G. H., & Rae, S. F. The pulse of democracy. New York: Simon & Schuster, 1940.

31. Garrett, H. E. *Statistics in psychology and education.* (2nd ed.) New York: Longmans, Green, 1937.
32. Garrett, H. E., & Zubin, J. The analysis of variance in psychological research. *Psychol. Bull.*, 1943, 40, 233-267.
33. Garrett, W. S. The Ohio State Psychological Examination; An instrument for predicting success in college. *Occupations*, 1944, 22, 489-495.
34. Gilliland, A. R., & Clark, E. L. *Psychology of individual differences.* New York: Prentice-Hall, 1939.
35. Goulden, C. H. *Methods of statistical analysis.* New York: Wiley, 1939.
36. Grant, D. A. On "the analysis of variance in psychological research." *Psychol. Bull.*, 1944, 41, 158-166.
37. Greene, E. B. *Measurements of human behavior.* New York: Odyssey, 1941.
38. Guilford, J. P. *Fundamental statistics in psychology and education.* New York: McGraw-Hill, 1942.
39. Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.
40. Guthrie, E. R. *The psychology of learning.* New York: Harper, 1935.
41. Hay, E. N., & Blakemore, A. M. The relationship between clerical experience and scores on the Minnesota Vocational Test for Clerical Workers. *J. appl. Psychol.*, 1943, 27, 311-315.
42. Holman, L. J. *Simplified statistics.* London: Pitman, 1938.
43. Horst, P. (Ed.) *The prediction of personal adjustment.* New York: *Soc. Sci. Res. Coun. Bull.*, 1941, No. 48.
44. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
45. Hurlock, E. B. The value of praise and reproof as incentives for children. *Arch. Psychol., N. Y.*, 1924, 11, No. 71.
46. Irwin, F. W., Armit, F. M., & Simon, C. W. Studies in object-preferences. I. The effect of temporal proximity. *J. exp. Psychol.*, 1943, 33, 64-72.
47. Jackson, R. W. B. Application of the analysis of variance and covariance method to educational problems. Toronto: *Dept. Educ. Res. Bull., Univ. Toronto*, 1940, No. 11.

48. Jenkins, J. G., & Dallenbach, K. M. Obliviscence during sleep and waking. *Amer. J. Psychol.*, 1924, 35, 605-612.
49. Jones, M. C. The elimination of children's fears. *J. exp. Psychol.*, 1924, 7, 382-390.
50. Jones, M. C. A laboratory study of fear: The case of Peter. *Ped. Sem.*, 1924, 31, 308-315.
51. Kellar, B. The construction and validation of a scale for measuring attitude toward any home-making activity. In Remmers, H. H. (Ed.) *Studies in attitudes: Bull. Purdue Univ.*, 1934, 35, 47-63.
52. Kelley, T. L. *Statistical method*. New York: Macmillan, 1923.
53. Kenney, J. F. *Mathematics of statistics*. (2 vols.) New York: Van Nostrand, 1939.
54. Klineberg, O. A study of psychological differences between "racial" and national groups in Europe. *Arch. Psychol.*, N. Y., 1931, 20, No. 132.
55. Koffka, K. *Principles of Gestalt psychology*. New York: Harcourt, Brace, 1935.
56. Krathwohl, W. C. A 3 by 3 analysis of the predictive value of test scores. *J. appl. Psychol.*, 1944, 28, 318-322.
57. Kuo, Z. Y. The genesis of the cat's response to the rat. *J. comp. Psychol.*, 1930, 11, 1-30.
58. Lewin, K. *A dynamic theory of personality*. New York: McGraw-Hill, 1935.
59. Lewin, K. *Principles of topological psychology*. New York: McGraw-Hill, 1936.
60. Lewis, H. B., & Franklin, M. An experimental study of the role of the ego in work. II. The significance of task-orientation in work. *J. exp. Psychol.*, 1944, 34, 195-215.
61. Likert, R. A technique for the measurement of attitudes. *Arch. Psychol.*, N. Y., 1932, No. 140.
62. Lindquist, E. F. The significance of a difference between "matched" groups. *J. educ. Psychol.*, 1931, 22, 197-204.
63. Lindquist, E. F. A further note on the significance of a difference between the means of matched groups. *J. educ. Psychol.*, 1933, 24, 66-69.
64. Lindquist, E. F. *Statistical analysis in educational research*. Boston: Houghton Mifflin, 1940.

65. Lo, C. F. Moral judgments of Chinese students. *J. abnorm. soc. Psychol.*, 1942, 37, 264-269.
66. Lynd, R. S. Knowledge for what? Princeton: Princeton University Press, 1939.
67. MacKinnon, D. W. Violation of prohibitions. In Murray, H. A. (Ed.) *Explorations in personality*. New York: Oxford University Press, 1938, 491-501.
68. McNemar, Q. Sampling in psychological research. *Psychol. Bull.*, 1940, 37, 331-365.
69. Newcomb, T. M. Personality and social change. New York: Dryden, 1943.
70. Odell, C. W. Statistical method in education. New York: Appleton-Century, 1935.
71. Paterson, D. G., & Ludgate, K. E. Blonde and brunette traits: A quantitative study. *J. Personn. Research.*, 1922, 1, 122-127.
72. Peatman, J. G., & Schafer, R. A table of random numbers from Selective Service numbers. *J. Psychol.*, 1942, 14, 295-305.
73. Peters, C. C. Interaction in analysis of variance interpreted as intercorrelation. *Psychol. Bull.*, 1944, 41, 287-299.
74. Peters, C. C., & Van Voorhis, W. R. Statistical procedures and their mathematical bases. New York: McGraw-Hill, 1940.
75. Peterson, R. C., & Thurstone, L. L. Motion pictures and the social attitudes of children. New York: Macmillan, 1933.
76. Rethlingshafer, D. Experimental evidence for functional autonomy. *Psychol. Rev.*, 1943, 50, 397-407.
77. Rider, P. R. An introduction to modern statistical methods. New York: Wiley, 1939.
78. Rogers, C. Counseling and psychotherapy. Boston: Houghton Mifflin, 1942.
79. Rosenthal, S. P. Change of socio-economic attitudes under radical motion picture propaganda. *Arch. Psychol., N. Y.*, 1934, No. 166.
80. Rosenzweig, S. An experimental study of "repression" with special reference to need-persistent and ego-defensive reactions to frustration. *J. exp. Psychol.*, 1943, 32, 64-74.
81. Sears, R. R. Survey of objective studies of psychoanalytic

- concepts. New York: *Soc. Sci. Res. Coun. Bull.*, 1943, No. 51.
82. Shaffer, L. F. *The psychology of adjustment*. Boston: Houghton Mifflin, 1936.
 83. Shen, E. Experimental design and statistical treatment in educational research. *J. exp. Educ.*, 1940, 8, 346-353.
 84. Skeels, H. M., Upedegraff, R., Wellman, B. L., & Williams, H. M. A study of environmental stimulation: An orphanage preschool project. *Univ. Ia. Stud. Child Welf.*, 1938, 15, No. 4.
 85. Snedecor, G. W. *Calculation and interpretation of the analysis of variance and covariance*. Ames, Iowa: Collegiate Press, 1934.
 86. Snedecor, G. W. *Statistical methods*. (3rd ed.) Ames: Iowa Collegiate Press, 1940.
 87. Sorenson, H. *Statistics for students of psychology and education*. New York: McGraw-Hill, 1936.
 88. Stock, J. S. Some general principles of sampling. In Cantril, H. (Ed.) *Gauging public opinion*. Princeton: Princeton University Press, 1944, 127-142.
 89. Thurstone, L. L., & Chave, E. J. *The measurement of attitude*. Chicago: University of Chicago Press, 1929.
 90. Tippett, L. H. C. *The methods of statistics*. (3rd ed.) London: Williams & Norgate, 1941.
 91. Tolman, E. C. *Purposive behavior in animals and men*. New York: Appleton-Century, 1932.
 92. Treloar, A. E. *Elements of statistical reasoning*. New York: Wiley, 1939.
 93. Treloar, A. E. *Random sampling distributions*. Minneapolis, Minn.: Burgess, 1942.
 94. Walker, H. M. Degrees of freedom. *J. educ. Psychol.*, 1940, 31, 253-269.
 95. Walker, H. M. *Mathematics essential for elementary statistics*. New York: Holt, 1934.
 96. Walker, H. M. *Elementary statistical methods*. New York: Holt, 1943.
 97. Walker, H. M. Personal communication. Feb. 19, 1945.
 98. Watson, K. B. The nature and measurement of musical meanings. *Psychol. Monogr.*, 1942, 54, No. 224.

99. Wilks, S. S. The standard error of the means of "matched" samples. *J. educ. Psychol.*, 1931, 22, 205-208.
100. Wolfe, D. The first course in psychology. *Psychol. Bull.*, 1942, 39, 685-712.
101. Wright, H. F. The influence of barriers upon strength of motivation. *Contr. psychol. Theor.*, Duke: Duke University Press, 1937.
102. Yule, G. U., & Kendall, M. G. An introduction to the theory of statistics. (11th ed.) London: Griffin, 1937.

APPENDIX

TABLE A.—SQUARES AND SQUARE ROOTS OF NUMBERS FROM 1 TO 1,000

N	N^2	\sqrt{N}	N	N^2	\sqrt{N}
1	1	1.000	41	16 81	6.403
2	4	1.414	42	17 64	6.481
3	9	1.732	43	18 49	6.557
4	16	2.000	44	19 36	6.633
5	25	2.236	45	20 25	6.708
6	36	2.449	46	21 16	6.782
7	49	2.646	47	22 09	6.856
8	64	2.828	48	23 04	6.928
9	81	3.000	49	24 01	7.000
<u>10</u>	<u>100</u>	<u>3.162</u>	<u>50</u>	<u>25 00</u>	<u>7.071</u>
11	121	3.317	51	26 01	7.141
12	144	3.464	52	27 04	7.211
13	169	3.606	53	28 09	7.280
14	196	3.742	54	29 16	7.348
15	225	3.873	55	30 25	7.416
16	256	4.000	56	31 36	7.483
17	289	4.123	57	32 49	7.550
18	324	4.243	58	33 64	7.616
19	361	4.359	59	34 81	7.681
<u>20</u>	<u>400</u>	<u>4.472</u>	<u>60</u>	<u>36 00</u>	<u>7.746</u>
21	441	4.583	61	37 21	7.810
22	484	4.690	62	38 44	7.874
23	529	4.796	63	39 69	7.937
24	576	4.899	64	40 96	8.000
25	625	5.000	65	42 25	8.062
26	676	5.099	66	43 56	8.124
27	729	5.196	67	44 89	8.185
28	784	5.292	68	46 24	8.246
29	841	5.385	69	47 61	8.307
<u>30</u>	<u>900</u>	<u>5.477</u>	<u>70</u>	<u>49 00</u>	<u>8.367</u>
31	961	5.568	71	50 41	8.426
32	10 24	5.657	72	51 84	8.485
33	10 89	5.745	73	53 29	8.544
34	11 56	5.831	74	54 76	8.602
35	12 25	5.916	75	56 25	8.660
36	12 96	6.000	76	57 76	8.718
37	13 69	6.083	77	59 29	8.775
38	14 44	6.164	78	60 84	8.832
39	15 21	6.245	79	62 41	8.888
40	16 00	6.325	80	64 00	8.944

TABLE A.—SQUARES AND SQUARE ROOTS—Continued

N	N^2	\sqrt{N}	N	N^2	\sqrt{N}
81	65 61	9.000	121	1 46 41	11.000
82	67 24	9.055	122	1 48 84	11.045
83	68 89	9.110	123	1 51 29	11.091
84	70 56	9.165	124	1 53 76	11.136
85	72 25	9.220	125	1 56 25	11.180
86	73 96	9.274	126	1 58 76	11.225
87	75 69	9.327	127	1 61 29	11.269
88	77 44	9.381	128	1 63 84	11.314
89	79 21	9.434	129	1 66 41	11.358
90	81 00	9.487	130	1 69 00	11.402
91	82 81	9.539	131	1 71 61	11.446
92	84 64	9.592	132	1 74 24	11.489
93	86 49	9.644	133	1 76 89	11.533
94	88 36	9.695	134	1 79 56	11.576
95	90 25	9.747	135	1 82 25	11.619
96	92 16	9.798	136	1 84 96	11.662
97	94 09	9.849	137	1 87 69	11.705
98	96 04	9.899	138	1 90 44	11.747
99	98 01	9.950	139	1 93 21	11.790
100	1 00 00	10.000	140	1 96 00	11.832
101	1 02 01	10.050	141	1 98 81	11.874
102	1 04 04	10.100	142	2 01 64	11.916
103	1 06 09	10.149	143	2 04 49	11.958
104	1 08 16	10.198	144	2 07 36	12.000
105	1 10 25	10.247	145	2 10 25	12.042
106	1 12 36	10.296	146	2 13 16	12.083
107	1 14 49	10.344	147	2 16 09	12.124
108	1 16 64	10.392	148	2 19 04	12.166
109	1 18 81	10.440	149	2 22 01	12.207
110	1 21 00	10.488	150	2 25 00	12.247
111	1 23 21	10.536	151	2 28 01	12.288
112	1 25 44	10.583	152	2 31 04	12.329
113	1 27 69	10.630	153	2 34 09	12.369
114	1 29 96	10.677	154	2 37 16	12.410
115	1 32 25	10.724	155	2 40 25	12.450
116	1 34 56	10.770	156	2 43 36	12.490
117	1 36 89	10.817	157	2 46 49	12.530
118	1 39 24	10.863	158	2 49 64	12.570
119	1 41 61	10.909	159	2 52 81	12.610
120	1 44 00	10.954	160	2 56 00	12.649

TABLE A.—SQUARES AND SQUARE ROOTS—Continued

N	N^2	\sqrt{N}	N	N^2	\sqrt{N}
161	2 59 21	12.689	201	4 04 01	14.177
162	2 62 44	12.728	202	4 08 04	14.213
163	2 65 69	12.767	203	4 12 09	14.248
164	2 68 96	12.806	204	4 16 16	14.283
165	2 72 25	12.845	205	4 20 25	14.318
166	2 75 56	12.884	206	4 24 36	14.353
167	2 78 89	12.923	207	4 28 49	14.387
168	2 82 24	12.961	208	4 32 64	14.422
169	2 85 61	13.000	209	4 36 81	14.457
170	2 89 00	13.038	210	4 41 00	14.491
171	2 92 41	13.077	211	4 45 21	14.526
172	2 95 84	13.115	212	4 49 44	14.560
173	2 99 29	13.153	213	4 53 69	14.595
174	3 02 76	13.191	214	4 57 96	14.629
175	3 06 25	13.229	215	4 62 25	14.663
176	3 09 76	13.266	216	4 66 56	14.697
177	3 13 29	13.304	217	4 70 89	14.731
178	3 16 84	13.342	218	4 75 24	14.765
179	3 20 41	13.379	219	4 79 61	14.799
180	3 24 00	13.416	220	4 84 00	14.832
181	3 27 61	13.454	221	4 88 41	14.866
182	3 31 24	13.491	222	4 92 84	14.900
183	3 34 89	13.528	223	4 97 29	14.933
184	3 38 56	13.565	224	5 01 76	14.967
185	3 42 25	13.601	225	5 06 25	15.000
186	3 45 96	13.638	226	5 10 76	15.033
187	3 49 69	13.675	227	5 15 29	15.067
188	3 53 44	13.711	228	5 19 84	15.100
189	3 57 21	13.748	229	5 24 41	15.133
190	3 61 00	13.784	230	5 29 00	15.166
191	3 64 81	13.820	231	5 33 61	15.199
192	3 68 64	13.856	232	5 38 24	15.232
193	3 72 49	13.892	233	5 42 89	15.264
194	3 76 36	13.928	234	5 47 56	15.297
195	3 80 25	13.964	235	5 52 25	15.330
196	3 84 16	14.000	236	5 56 96	15.362
197	3 88 09	14.036	237	5 61 69	15.395
198	3 92 04	14.071	238	5 66 44	15.427
199	3 96 01	14.107	239	5 71 21	15.460
200	4 00 00	14.142	240	5 76 00	15.492

TABLE A.—SQUARES AND SQUARE ROOTS—Continued

N	N^2	\sqrt{N}	N	N^2	\sqrt{N}
241	5 80 81	15.524	281	7 89 61	16.763
242	5 85 64	15.556	282	7 95 24	16.793
243	5 90 49	15.588	283	8 00 89	16.823
244	5 95 36	15.620	284	8 06 56	16.852
245	6 00 25	15.652	285	8 12 25	16.882
246	6 05 16	15.684	286	8 17 96	16.912
247	6 10 09	15.716	287	8 23 69	16.941
248	6 15 04	15.748	288	8 29 44	16.971
249	6 20 01	15.780	289	8 35 21	17.000
250	6 25 00	15.811	290	8 41 00	17.029
251	6 30 01	15.843	291	8 46 81	17.059
252	6 35 04	15.875	292	8 52 64	17.088
253	6 40 09	15.906	293	8 58 49	17.117
254	6 45 16	15.937	294	8 64 36	17.146
255	6 50 25	15.969	295	8 70 25	17.176
256	6 55 36	16.000	296	8 76 16	17.205
257	6 60 49	16.031	297	8 82 09	17.234
258	6 65 64	16.062	298	8 88 04	17.263
259	6 70 81	16.093	299	8 94 01	17.292
260	6 76 00	16.125	300	9 00 00	17.321
261	6 81 21	16.155	301	9 06 01	17.349
262	6 86 44	16.186	302	9 12 04	17.378
263	6 91 69	16.217	303	9 18 09	17.407
264	6 96 96	16.248	304	9 24 16	17.436
265	7 02 25	16.279	305	9 30 25	17.464
266	7 07 56	16.310	306	9 36 36	17.493
267	7 12 89	16.340	307	9 42 49	17.521
268	7 18 24	16.371	308	9 48 64	17.550
269	7 23 61	16.401	309	9 54 81	17.578
270	7 29 00	16.432	310	9 61 00	17.607
271	7 34 41	16.462	311	9 67 21	17.635
272	7 39 84	16.492	312	9 73 44	17.664
273	7 45 29	16.523	313	9 79 69	17.692
274	7 50 76	16.553	314	9 85 96	17.720
275	7 56 25	16.583	315	9 92 25	17.748
276	7 61 76	16.613	316	9 98 56	17.776
277	7 67 29	16.643	317	10 04 89	17.804
278	7 72 84	16.673	318	10 11 24	17.833
279	7 78 41	16.703	319	10 17 61	17.861
280	7 84 00	16.733	320	10 24 00	17.889

TABLE A.—SQUARES AND SQUARE ROOTS—Continued

N	N^2	\sqrt{N}	N	N^2	\sqrt{N}
321	10 30 41	17.916	361	13 03 21	19.000
322	10 36 84	17.944	362	13 10 44	19.026
323	10 43 29	17.972	363	13 17 69	19.053
324	10 49 76	18.000	364	13 24 96	19.079
325	10 56 25	18.028	365	13 32 25	19.105
326	10 62 76	18.055	366	13 39 56	19.131
327	10 69 29	18.083	367	13 46 89	19.157
328	10 75 84	18.111	368	13 54 24	19.183
329	10 82 41	18.138	369	13 61 61	19.209
330	10 89 00	18.166	370	13 69 00	19.235
331	10 95 61	18.193	371	13 76 41	19.261
332	11 02 24	18.221	372	13 83 84	19.287
333	11 08 89	18.248	373	13 91 29	19.313
334	11 15 56	18.276	374	13 98 76	19.339
335	11 22 25	18.303	375	14 06 25	19.365
336	11 28 96	18.330	376	14 13 76	19.391
337	11 35 69	18.358	377	14 21 29	19.416
338	11 42 44	18.385	378	14 28 84	19.442
339	11 49 21	18.412	379	14 36 41	19.468
340	11 56 00	18.439	380	14 44 00	19.494
341	11 62 81	18.466	381	14 51 61	19.519
342	11 69 64	18.493	382	14 59 24	19.545
343	11 76 49	18.520	383	14 66 89	19.570
344	11 83 36	18.547	384	14 74 56	19.596
345	11 90 25	18.574	385	14 82 25	19.621
346	11 97 16	18.601	386	14 89 96	19.647
347	12 04 09	18.628	387	14 97 69	19.672
348	12 11 04	18.655	388	15 05 44	19.698
349	12 18 01	18.682	389	15 13 21	19.723
350	12 25 00	18.708	390	15 21 00	19.748
351	12 32 01	18.735	391	15 28 81	19.774
352	12 39 04	18.762	392	15 36 64	19.799
353	12 46 09	18.788	393	15 44 49	19.824
354	12 53 16	18.815	394	15 52 36	19.849
355	12 60 25	18.841	395	15 60 25	19.875
356	12 67 36	18.868	396	15 68 16	19.900
357	12 74 49	18.894	397	15 76 09	19.925
358	12 81 64	18.921	398	15 84 04	19.950
359	12 88 81	18.947	399	15 92 01	19.975
360	12 96 00	18.974	400	16 00 00	20.000

TABLE A.—SQUARES AND SQUARE ROOTS—Continued

N	N^2	\sqrt{N}	N	N^2	\sqrt{N}
401	16 08 01	20.025	441	19 44 81	21.000
402	16 16 04	20.050	442	19 53 64	21.024
403	16 24 09	20.075	443	19 62 49	21.048
404	16 32 16	20.100	444	19 71 36	21.071
405	16 40 25	20.125	445	19 80 25	21.095
406	16 48 36	20.149	446	19 89 16	21.119
407	16 56 49	20.174	447	19 98 09	21.142
408	16 64 64	20.199	448	20 07 04	21.166
409	16 72 81	20.224	449	20 16 01	21.190
410	16 81 00	20.248	450	20 25 00	21.213
411	16 89 21	20.273	451	20 34 01	21.237
412	16 97 44	20.298	452	20 43 04	21.260
413	17 05 69	20.322	453	20 52 09	21.284
414	17 13 96	20.347	454	20 61 16	21.307
415	17 22 25	20.372	455	20 70 25	21.331
416	17 30 56	20.396	456	20 79 36	21.354
417	17 38 89	20.421	457	20 88 49	21.378
418	17 47 24	20.445	458	20 97 64	21.401
419	17 55 61	20.469	459	21 06 81	21.424
420	17 64 00	20.494	460	21 16 00	21.448
421	17 72 41	20.518	461	21 25 21	21.471
422	17 80 84	20.543	462	21 34 44	21.494
423	17 89 29	20.567	463	21 43 69	21.517
424	17 97 76	20.591	464	21 52 96	21.541
425	18 06 25	20.616	465	21 62 25	21.564
426	18 14 76	20.640	466	21 71 56	21.587
427	18 23 29	20.664	467	21 80 89	21.610
428	18 31 84	20.688	468	21 90 24	21.633
429	18 40 41	20.712	469	21 99 61	21.656
430	18 49 00	20.736	470	22 09 00	21.679
431	18 57 61	20.761	471	22 18 41	21.703
432	18 66 24	20.785	472	22 27 84	21.726
433	18 74 89	20.809	473	22 37 29	21.749
434	18 83 56	20.833	474	22 46 76	21.772
435	18 92 25	20.857	475	22 56 25	21.794
436	19 00 96	20.881	476	22 65 76	21.817
437	19 09 69	20.905	477	22 75 29	21.840
438	19 18 44	20.928	478	22 84 84	21.863
439	19 27 21	20.952	479	22 94 41	21.886
440	19 36 00	20.976	480	23 04 00	21.909

TABLE A.—SQUARES AND SQUARE ROOTS—Continued

N	N^2	\sqrt{N}	N	N^2	\sqrt{N}
481	23 13 61	21.932	521	27 14 41	22.825
482	23 23 24	21.954	522	27 24 84	22.847
483	23 32 89	21.977	523	27 35 29	22.869
484	23 42 56	22.000	524	27 45 76	22.891
485	23 52 25	22.023	525	27 56 25	22.913
486	23 61 96	22.045	526	27 66 76	22.935
487	23 71 69	22.068	527	27 77 29	22.956
488	23 81 44	22.091	528	27 87 84	22.978
489	23 91 21	22.113	529	27 98 41	23.000
490	24 01 00	22.136	530	28 09 00	23.022
491	24 10 81	22.159	531	28 19 61	23.043
492	24 20 64	22.181	532	28 30 24	23.065
493	24 30 49	22.204	533	28 40 89	23.087
494	24 40 36	22.226	534	28 51 56	23.108
495	24 50 25	22.249	535	28 62 25	23.130
496	24 60 16	22.271	536	28 72 96	23.152
497	24 70 09	22.293	537	28 83 69	23.173
498	24 80 04	22.316	538	28 94 44	23.195
499	24 90 01	22.338	539	29 05 21	23.216
500	25 00 00	22.361	540	29 16 00	23.238
501	25 10 01	22.383	541	29 26 81	23.259
502	25 20 04	22.405	542	29 37 64	23.281
503	25 30 09	22.428	543	29 48 49	23.302
504	25 40 16	22.450	544	29 59 36	23.324
505	25 50 25	22.472	545	29 70 25	23.345
506	25 60 36	22.494	546	29 81 16	23.367
507	25 70 49	22.517	547	29 92 09	23.388
508	25 80 64	22.539	548	30 03 04	23.409
509	25 90 81	22.561	549	30 14 01	23.431
510	26 01 00	22.583	550	30 25 00	23.452
511	26 11 21	22.605	551	30 36 01	23.473
512	26 21 44	22.627	552	30 47 04	23.495
513	26 31 69	22.650	553	30 58 09	23.516
514	26 41 96	22.672	554	30 69 16	23.537
515	26 52 25	22.694	555	30 80 25	23.558
516	26 62 56	22.716	556	30 91 36	23.580
517	26 72 89	22.738	557	31 02 49	23.601
518	26 83 24	22.760	558	31 13 64	23.622
519	26 93 61	22.782	559	31 24 81	23.643
520	27 04 00	22.804	560	31 36 00	23.664

TABLE A.—SQUARES AND SQUARE ROOTS—Continued

N	N^2	\sqrt{N}	N	N^2	\sqrt{N}
561	31 47 21	23.685	601	36 12 01	24.515
562	31 58 44	23.707	602	36 24 04	24.536
563	31 69 69	23.728	603	36 36 09	24.556
564	31 80 96	23.749	604	36 48 16	24.576
565	31 92 25	23.770	605	36 60 25	24.597
566	32 03 56	23.791	606	36 72 36	24.617
567	32 14 89	23.812	607	36 84 49	24.637
568	32 26 24	23.833	608	36 96 64	24.658
569	32 37 61	23.854	609	37 08 81	24.678
570	32 49 00	23.875	610	37 21 00	24.698
571	32 60 41	23.896	611	37 33 21	24.718
572	32 71 84	23.917	612	37 45 44	24.739
573	32 83 29	23.937	613	37 57 69	24.759
574	32 94 76	23.958	614	37 69 96	24.779
575	33 06 25	23.979	615	37 82 25	24.799
576	33 17 76	24.000	616	37 94 56	24.819
577	33 29 29	24.021	617	38 06 89	24.839
578	33 40 84	24.042	618	38 19 24	24.860
579	33 52 41	24.062	619	38 31 61	24.880
580	33 64 00	24.083	620	38 44 00	24.900
581	33 75 61	24.104	621	38 56 41	24.920
582	33 87 24	24.125	622	38 68 84	24.940
583	33 98 89	24.145	623	38 81 29	24.960
584	34 10 56	24.166	624	38 93 76	24.980
585	34 22 25	24.187	625	39 06 25	25.000
586	34 33 96	24.207	626	39 18 76	25.020
587	34 45 69	24.228	627	39 31 29	25.040
588	34 57 44	24.249	628	39 43 84	25.060
589	34 69 21	24.269	629	39 56 41	25.080
590	34 81 00	24.290	630	39 69 00	25.100
591	34 92 81	24.310	631	39 81 61	25.120
592	35 04 64	24.331	632	39 94 24	25.140
593	35 16 49	24.352	633	40 06 89	25.159
594	35 28 36	24.372	634	40 19 56	25.179
595	35 40 25	24.393	635	40 32 25	25.199
596	35 52 16	24.413	636	40 44 96	25.219
597	35 64 09	24.434	637	40 57 69	25.239
598	35 76 04	24.454	638	40 70 44	25.259
599	35 88 01	24.474	639	40 83 21	25.278
600	36 00 00	24.495	640	40 96 00	25.298

TABLE A.—SQUARES AND SQUARE ROOTS—Continued

N	N^2	\sqrt{N}	N	N^2	\sqrt{N}
641	41 08 81	25.318	681	46 37 61	26.096
642	41 21 64	25.338	682	46 51 24	26.115
643	41 34 49	25.357	683	46 64 89	26.134
644	41 47 36	25.377	684	46 78 56	26.153
645	41 60 25	25.397	685	46 92 25	26.173
646	41 73 16	25.417	686	47 05 96	26.192
647	41 86 09	25.436	687	47 19 69	26.211
648	41 99 04	25.456	688	47 33 44	26.230
649	42 12 01	25.475	689	47 47 21	26.249
650	42 25 00	25.495	690	47 61 00	26.268
651	42 38 01	25.515	691	47 74 81	26.287
652	42 51 04	25.534	692	47 88 64	26.306
653	42 64 09	25.554	693	48 02 49	26.325
654	42 77 16	25.573	694	48 16 36	26.344
655	42 90 25	25.593	695	48 30 25	26.363
656	43 03 36	25.612	696	48 44 16	26.382
657	43 16 49	25.632	697	48 58 09	26.401
658	43 29 64	25.652	698	48 72 04	26.420
659	43 42 81	25.671	699	48 86 01	26.439
660	43 56 00	25.690	700	49 00 00	26.458
661	43 69 21	25.710	701	49 14 01	26.476
662	43 82 44	25.729	702	49 28 04	26.495
663	43 95 69	25.749	703	49 42 09	26.514
664	44 08 96	25.768	704	49 56 16	26.533
665	44 22 25	25.788	705	49 70 25	26.552
666	44 35 56	25.807	706	49 84 36	26.571
667	44 48 89	25.826	707	49 98 49	26.589
668	44 62 24	25.846	708	50 12 64	26.608
669	44 75 61	25.865	709	50 26 81	26.627
670	44 89 00	25.884	710	50 41 00	26.646
671	45 02 41	25.904	711	50 55 21	26.665
672	45 15 84	25.923	712	50 69 44	26.683
673	45 29 29	25.942	713	50 83 69	26.702
674	45 42 76	25.962	714	50 97 96	26.721
675	45 56 25	25.981	715	51 12 25	26.739
676	45 69 76	26.000	716	51 26 56	26.758
677	45 83 29	26.019	717	51 40 89	26.777
678	45 96 84	26.038	718	51 55 24	26.796
679	46 10 41	26.058	719	51 69 61	26.814
680	46 24 00	26.077	720	51 84 00	26.833

TABLE A.—SQUARES AND SQUARE ROOTS—Continued

N	N^2	\sqrt{N}	N	N^2	\sqrt{N}
721	51 98 41	26.851	761	57 91 21	27.586
722	52 12 84	26.870	762	58 06 44	27.604
723	52 27 29	26.889	763	58 21 69	27.622
724	52 41 76	26.907	764	58 36 96	27.641
725	52 56 25	26.926	765	58 52 25	27.659
726	52 70 76	26.944	766	58 67 56	27.677
727	52 85 29	26.963	767	58 82 89	27.695
728	52 99 84	26.981	768	58 98 24	27.713
729	53 14 41	27.000	769	59 13 61	27.731
730	53 29 00	27.019	770	59 29 00	27.749
731	53 43 61	27.037	771	59 44 41	27.767
732	53 58 24	27.055	772	59 59 84	27.785
733	53 72 89	27.074	773	59 75 29	27.803
734	53 87 56	27.092	774	59 90 76	27.821
735	54 02 25	27.111	775	60 06 25	27.839
736	54 16 96	27.129	776	60 21 76	27.857
737	54 31 69	27.148	777	60 37 29	27.875
738	54 46 44	27.166	778	60 52 84	27.893
739	54 61 21	27.185	779	60 68 41	27.911
740	54 76 00	27.203	780	60 84 00	27.928
741	54 90 81	27.221	781	60 99 61	27.946
742	55 05 64	27.240	782	61 15 24	27.964
743	55 20 49	27.258	783	61 30 89	27.982
744	55 35 36	27.276	784	61 46 56	28.000
745	55 50 25	27.295	785	61 62 25	28.018
746	55 65 16	27.313	786	61 77 96	28.036
747	55 80 09	27.331	787	61 93 69	28.054
748	55 95 04	27.350	788	62 09 44	28.071
749	56 10 01	27.368	789	62 25 21	28.089
750	56 25 00	27.386	790	62 41 00	28.107
751	56 40 01	27.404	791	62 56 81	28.125
752	56 55 04	27.423	792	62 72 64	28.142
753	56 70 09	27.441	793	62 88 49	28.160
754	56 85 16	27.459	794	63 04 36	28.178
755	57 00 25	27.477	795	63 20 25	28.196
756	57 15 36	27.495	796	63 36 16	28.213
757	57 30 49	27.514	797	63 52 09	28.231
758	57 45 64	27.532	798	63 68 04	28.249
759	57 60 81	27.550	799	63 84 01	28.267
760	57 76 00	27.568	800	64 00 00	28.284

TABLE A.—SQUARES AND SQUARE ROOTS—Continued

N	N^2	\sqrt{N}	N	N^2	\sqrt{N}
801	64 16 01	28.302	841	70 72 81	29.000
802	64 32 04	28.320	842	70 89 64	29.017
803	64 48 09	28.337	843	71 06 49	29.034
804	64 64 16	28.355	844	71 23 36	29.052
805	64 80 25	28.373	845	71 40 25	29.069
806	64 96 36	28.390	846	71 57 16	29.086
807	65 12 49	28.408	847	71 74 09	29.103
808	65 28 64	28.425	848	71 91 04	29.120
809	65 44 81	28.443	849	72 08 01	29.138
810	65 61 00	28.460	850	72 25 00	29.155
811	65 77 21	28.478	851	72 42 01	29.172
812	65 93 44	28.496	852	72 59 04	29.189
813	66 09 69	28.513	853	72 76 09	29.206
814	66 25 96	28.531	854	72 93 16	29.223
815	66 42 25	28.548	855	73 10 25	29.240
816	66 58 56	28.566	856	73 27 36	29.257
817	66 74 89	28.583	857	73 44 49	29.275
818	66 91 24	28.601	858	73 61 64	29.292
819	67 07 61	28.618	859	73 78 81	29.309
820	67 24 00	28.636	860	73 96 00	29.326
821	67 40 41	28.653	861	74 13 21	29.343
822	67 56 84	28.671	862	74 30 44	29.360
823	67 73 29	28.688	863	74 47 69	29.377
824	67 89 76	28.705	864	74 64 96	29.394
825	68 06 25	28.723	865	74 82 25	29.411
826	68 22 76	28.740	866	74 99 56	29.428
827	68 39 29	28.758	867	75 16 89	29.445
828	68 55 84	28.775	868	75 34 24	29.462
829	68 72 41	28.792	869	75 51 61	29.479
830	68 89 00	28.810	870	75 69 00	29.496
831	69 05 61	28.827	871	75 86 41	29.513
832	69 22 24	28.844	872	76 03 84	29.530
833	69 38 89	28.862	873	76 21 29	29.547
834	69 55 56	28.879	874	76 38 76	29.563
835	69 72 25	28.896	875	76 56 25	29.580
836	69 88 96	28.914	876	76 73 76	29.597
837	70 05 69	28.931	877	76 91 29	29.614
838	70 22 44	28.948	878	77 08 84	29.631
839	70 39 21	28.965	879	77 26 41	29.648
840	70 56 00	28.983	880	77 44 00	29.665

TABLE A.—SQUARES AND SQUARE ROOTS—Continued

N	N^2	\sqrt{N}	N	N^2	\sqrt{N}
881	77 61 61	29.682	921	84 82 41	30.348
882	77 79 24	29.698	922	85 00 84	30.364
883	77 96 89	29.715	923	85 19 29	30.381
884	78 14 56	29.732	924	85 37 76	30.397
885	78 32 25	29.749	925	85 56 25	30.414
886	78 49 96	29.766	926	85 74 76	30.430
887	78 67 69	29.783	927	85 93 29	30.447
888	78 85 44	29.799	928	86 11 84	30.463
889	79 03 21	29.816	929	86 30 41	30.480
890	79 21 00	29.833	930	86 49 00	30.496
891	79 38 81	29.850	931	86 67 61	30.512
892	79 56 64	29.866	932	86 86 24	30.529
893	79 74 49	29.883	933	87 04 89	30.545
894	79 92 36	29.900	934	87 23 56	30.561
895	80 10 25	29.916	935	87 42 25	30.578
896	80 28 16	29.933	936	87 60 96	30.594
897	80 46 09	29.950	937	87 79 69	30.610
898	80 64 04	29.967	938	87 98 44	30.627
899	80 82 01	29.983	939	88 17 21	30.643
900	81 00 00	30.000	940	88 36 00	30.659
901	81 18 01	30.017	941	88 54 81	30.676
902	81 36 04	30.033	942	88 73 64	30.692
903	81 54 09	30.050	943	88 92 49	30.708
904	81 72 16	30.067	944	89 11 36	30.725
905	81 90 25	30.083	945	89 30 25	30.741
906	82 08 36	30.100	946	89 49 16	30.757
907	82 26 49	30.116	947	89 68 09	30.773
908	82 44 64	30.133	948	89 87 04	30.790
909	82 62 81	30.150	949	90 06 01	30.806
910	82 81 00	30.166	950	90 25 00	30.822
911	82 99 21	30.183	951	90 44 01	30.838
912	83 17 44	30.199	952	90 63 04	30.854
913	83 35 69	30.216	953	90 82 09	30.871
914	83 53 96	30.232	954	91 01 16	30.887
915	83 72 25	30.249	955	91 20 25	30.903
916	83 90 56	30.265	956	91 39 36	30.919
917	84 08 89	30.282	957	91 58 49	30.935
918	84 27 24	30.299	958	91 77 64	30.952
919	84 45 61	30.315	959	91 96 81	30.968
920	84 64 00	30.332	960	92 16 00	30.984

TABLE A.—SQUARES AND SQUARE ROOTS—*Concluded*

N	N^2	\sqrt{N}	N	N^2	\sqrt{N}
961	92 35 21	31.000	981	96 23 61	31.321
962	92 54 44	31.016	982	96 43 24	31.337
963	92 73 69	31.032	983	96 62 89	31.353
964	92 92 96	31.048	984	96 82 56	31.369
965	93 12 25	31.064	985	97 02 25	31.385
966	93 31 56	31.081	986	97 21 96	31.401
967	93 50 89	31.097	987	97 41 69	31.417
968	93 70 24	31.113	988	97 61 44	31.432
969	93 89 61	31.129	989	97 81 21	31.448
970	94 09 00	31.145	990	98 01 00	31.464
971	94 28 41	31.161	991	98 20 81	31.480
972	94 47 84	31.177	992	98 40 64	31.496
973	94 67 29	31.193	993	98 60 49	31.512
974	94 86 76	31.209	994	98 80 36	31.528
975	95 06 25	31.225	995	99 00 25	31.544
976	95 25 76	31.241	996	99 20 16	31.559
977	95 45 29	31.257	997	99 40 09	31.575
978	95 64 84	31.273	998	99 60 04	31.591
979	95 84 41	31.289	999	99 80 01	31.607
980	96 04 00	31.305	1000	100 00 00	31.623

TABLE B.—AREAS AND ORDINATES OF THE NORMAL CURVE
IN TERMS OF $\frac{z}{\sigma}$

(1) z STANDARD SCORE $\left(\frac{z}{\sigma}\right)$	(2) A AREA FROM MEAN TO $\frac{z}{\sigma}$	(3) B AREA IN LARGER PORTION	(4) C AREA IN SMALLER PORTION	(5) y ORDINATE AT $\frac{z}{\sigma}$
0.00	.0000	.5000	.5000	.3989
0.01	.0040	.5040	.4960	.3989
0.02	.0080	.5080	.4920	.3989
0.03	.0120	.5120	.4880	.3988
0.04	.0160	.5160	.4840	.3986
0.05	.0199	.5199	.4801	.3984
0.06	.0239	.5239	.4761	.3982
0.07	.0279	.5279	.4721	.3980
0.08	.0319	.5319	.4681	.3977
0.09	.0359	.5359	.4641	.3973
0.10	.0398	.5398	.4602	.3970
0.11	.0438	.5438	.4562	.3965
0.12	.0478	.5478	.4522	.3961
0.13	.0517	.5517	.4483	.3956
0.14	.0557	.5557	.4443	.3951
0.15	.0596	.5596	.4404	.3945
0.16	.0636	.5636	.4364	.3939
0.17	.0675	.5675	.4325	.3932
0.18	.0714	.5714	.4286	.3925
0.19	.0753	.5753	.4247	.3918
0.20	.0793	.5793	.4207	.3910
0.21	.0832	.5832	.4168	.3902
0.22	.0871	.5871	.4129	.3894
0.23	.0910	.5910	.4090	.3885
0.24	.0948	.5948	.4052	.3876
0.25	.0987	.5987	.4013	.3867
0.26	.1026	.6026	.3974	.3857
0.27	.1064	.6064	.3936	.3847
0.28	.1103	.6103	.3897	.3836
0.29	.1141	.6141	.3859	.3825
0.30	.1179	.6179	.3821	.3814
0.31	.1217	.6217	.3783	.3802
0.32	.1255	.6255	.3745	.3790
0.33	.1293	.6293	.3707	.3778
0.34	.1331	.6331	.3669	.3765

TABLE B.—AREAS AND ORDINATES OF THE NORMAL CURVE

IN TERMS OF $\frac{x}{\sigma}$ —Continued

(1) z STANDARD SCORE $\left(\frac{x}{\sigma}\right)$	(2) A AREA FROM MEAN TO $\frac{x}{\sigma}$	(3) B AREA IN LARGER PORTION	(4) C AREA IN SMALLER PORTION	(5) y ORDINATE AT $\frac{x}{\sigma}$
0.35	.1368	.6368	.3632	.3752
0.36	.1406	.6406	.3594	.3739
0.37	.1443	.6443	.3557	.3725
0.38	.1480	.6480	.3520	.3712
0.39	.1517	.6517	.3483	.3697
0.40	.1554	.6554	.3446	.3683
0.41	.1591	.6591	.3409	.3668
0.42	.1628	.6628	.3372	.3653
0.43	.1664	.6664	.3336	.3637
0.44	.1700	.6700	.3300	.3621
0.45	.1736	.6736	.3264	.3605
0.46	.1772	.6772	.3228	.3589
0.47	.1808	.6808	.3192	.3572
0.48	.1844	.6844	.3156	.3555
0.49	.1879	.6879	.3121	.3538
0.50	.1915	.6915	.3085	.3521
0.51	.1950	.6950	.3050	.3503
0.52	.1985	.6985	.3015	.3485
0.53	.2019	.7019	.2981	.3467
0.54	.2054	.7054	.2946	.3448
0.55	.2088	.7088	.2912	.3429
0.56	.2123	.7123	.2877	.3410
0.57	.2157	.7157	.2843	.3391
0.58	.2190	.7190	.2810	.3372
0.59	.2224	.7224	.2776	.3352
0.60	.2257	.7257	.2743	.3332
0.61	.2291	.7291	.2709	.3312
0.62	.2324	.7324	.2676	.3292
0.63	.2357	.7357	.2643	.3271
0.64	.2389	.7389	.2611	.3251
0.65	.2422	.7422	.2578	.3230
0.66	.2454	.7454	.2546	.3209
0.67	.2486	.7486	.2514	.3187
0.68	.2517	.7517	.2483	.3166
0.69	.2549	.7549	.2451	.3144

TABLE B.—AREAS AND ORDINATES OF THE NORMAL CURVE
IN TERMS OF $\frac{x}{\sigma}$ —Continued

(1) z STANDARD SCORE $\left(\frac{x}{\sigma}\right)$	(2) A AREA FROM MEAN TO $\frac{x}{\sigma}$	(3) B AREA IN LARGER PORTION	(4) C AREA IN SMALLER PORTION	(5) y ORDINATE AT $\frac{x}{\sigma}$
0.70	.2580	.7580	.2420	.3123
0.71	.2611	.7611	.2389	.3101
0.72	.2642	.7642	.2358	.3079
0.73	.2673	.7673	.2327	.3056
0.74	.2704	.7704	.2296	.3034
0.75	.2734	.7734	.2266	.3011
0.76	.2764	.7764	.2236	.2989
0.77	.2794	.7794	.2206	.2966
0.78	.2823	.7823	.2177	.2943
0.79	.2852	.7852	.2148	.2920
0.80	.2881	.7881	.2119	.2897
0.81	.2910	.7910	.2090	.2874
0.82	.2939	.7939	.2061	.2850
0.83	.2967	.7967	.2033	.2827
0.84	.2995	.7995	.2005	.2803
0.85	.3023	.8023	.1977	.2780
0.86	.3051	.8051	.1949	.2756
0.87	.3078	.8078	.1922	.2732
0.88	.3106	.8106	.1894	.2709
0.89	.3133	.8133	.1867	.2685
0.90	.3159	.8159	.1841	.2661
0.91	.3186	.8186	.1814	.2637
0.92	.3212	.8212	.1788	.2613
0.93	.3238	.8238	.1762	.2589
0.94	.3264	.8264	.1736	.2565
0.95	.3289	.8289	.1711	.2541
0.96	.3315	.8315	.1685	.2516
0.97	.3340	.8340	.1660	.2492
0.98	.3365	.8365	.1635	.2468
0.99	.3389	.8389	.1611	.2444
1.00	.3413	.8413	.1587	.2420
1.01	.3438	.8438	.1562	.2396
1.02	.3461	.8461	.1539	.2371
1.03	.3485	.8485	.1515	.2347
1.04	.3508	.8508	.1492	.2323

TABLE B.—AREAS AND ORDINATES OF THE NORMAL CURVE
IN TERMS OF $\frac{x}{\sigma}$ —Continued

(1) z STANDARD SCORE $\left(\frac{x}{\sigma}\right)$	(2) A AREA FROM MEAN TO $\frac{x}{\sigma}$	(3) B AREA IN LARGER PORTION	(4) C AREA IN SMALLER PORTION	(5) y ORDINATE AT $\frac{x}{\sigma}$
1.05	.3531	.8531	.1469	.2299
1.06	.3554	.8554	.1446	.2275
1.07	.3577	.8577	.1423	.2251
1.08	.3599	.8599	.1401	.2227
1.09	.3621	.8621	.1379	.2203
1.10	.3643	.8643	.1357	.2179
1.11	.3665	.8665	.1335	.2155
1.12	.3686	.8686	.1314	.2131
1.13	.3708	.8708	.1292	.2107
1.14	.3729	.8729	.1271	.2083
1.15	.3749	.8749	.1251	.2059
1.16	.3770	.8770	.1230	.2036
1.17	.3790	.8790	.1210	.2012
1.18	.3810	.8810	.1190	.1989
1.19	.3830	.8830	.1170	.1965
1.20	.3849	.8849	.1151	.1942
1.21	.3869	.8869	.1131	.1919
1.22	.3888	.8888	.1112	.1895
1.23	.3907	.8907	.1093	.1872
1.24	.3925	.8925	.1075	.1849
1.25	.3944	.8944	.1056	.1826
1.26	.3962	.8962	.1038	.1804
1.27	.3980	.8980	.1020	.1781
1.28	.3997	.8997	.1003	.1758
1.29	.4015	.9015	.0985	.1736
1.30	.4032	.9032	.0968	.1714
1.31	.4049	.9049	.0951	.1691
1.32	.4066	.9066	.0934	.1669
1.33	.4082	.9082	.0918	.1647
1.34	.4099	.9099	.0901	.1626
1.35	.4115	.9115	.0885	.1604
1.36	.4131	.9131	.0869	.1582
1.37	.4147	.9147	.0853	.1561
1.38	.4162	.9162	.0838	.1539
1.39	.4177	.9177	.0823	.1518

TABLE B.—AREAS AND ORDINATES OF THE NORMAL CURVE
IN TERMS OF $\frac{x}{\sigma}$ —Continued

(1) z STANDARD SCORE $\left(\frac{x}{\sigma}\right)$	(2) A AREA FROM MEAN TO $\frac{x}{\sigma}$	(3) B AREA IN LARGER PORTION	(4) C AREA IN SMALLER PORTION	(5) y ORDINATE AT $\frac{x}{\sigma}$
1.40	.4192	.9192	.0808	.1497
1.41	.4207	.9207	.0793	.1476
1.42	.4222	.9222	.0778	.1456
1.43	.4236	.9236	.0764	.1435
1.44	.4251	.9251	.0749	.1415
1.45	.4265	.9265	.0735	.1394
1.46	.4279	.9279	.0721	.1374
1.47	.4292	.9292	.0708	.1354
1.48	.4306	.9306	.0694	.1334
1.49	.4319	.9319	.0681	.1315
1.50	.4332	.9332	.0668	.1295
1.51	.4345	.9345	.0655	.1276
1.52	.4357	.9357	.0643	.1257
1.53	.4370	.9370	.0630	.1238
1.54	.4382	.9382	.0618	.1219
1.55	.4394	.9394	.0606	.1200
1.56	.4406	.9406	.0594	.1182
1.57	.4418	.9418	.0582	.1163
1.58	.4429	.9429	.0571	.1145
1.59	.4441	.9441	.0559	.1127
1.60	.4452	.9452	.0548	.1109
1.61	.4463	.9463	.0537	.1092
1.62	.4474	.9474	.0526	.1074
1.63	.4484	.9484	.0516	.1057
1.64	.4495	.9495	.0505	.1040
1.65	.4505	.9505	.0495	.1023
1.66	.4515	.9515	.0485	.1006
1.67	.4525	.9525	.0475	.0989
1.68	.4535	.9535	.0465	.0973
1.69	.4545	.9545	.0455	.0957
1.70	.4554	.9554	.0446	.0940
1.71	.4564	.9564	.0436	.0925
1.72	.4573	.9573	.0427	.0909
1.73	.4582	.9582	.0418	.0893
1.74	.4591	.9591	.0409	.0878

TABLE B.—AREAS AND ORDINATES OF THE NORMAL CURVE
 IN TERMS OF $\frac{x}{\sigma}$ —Continued

(1) z STANDARD SCORE $\left(\frac{x}{\sigma}\right)$	(2) A AREA FROM MEAN TO $\frac{x}{\sigma}$	(3) B AREA IN LARGER PORTION	(4) C AREA IN SMALLER PORTION	(5) y ORDINATE AT $\frac{x}{\sigma}$
1.75	.4599	.9599	.0401	.0863
1.76	.4608	.9608	.0392	.0848
1.77	.4616	.9616	.0384	.0833
1.78	.4625	.9625	.0375	.0818
1.79	.4633	.9633	.0367	.0804
1.80	.4641	.9641	.0359	.0790
1.81	.4649	.9649	.0351	.0775
1.82	.4656	.9656	.0344	.0761
1.83	.4664	.9664	.0336	.0748
1.84	.4671	.9671	.0329	.0734
1.85	.4678	.9678	.0322	.0721
1.86	.4686	.9686	.0314	.0707
1.87	.4693	.9693	.0307	.0694
1.88	.4699	.9699	.0301	.0681
1.89	.4706	.9706	.0294	.0669
1.90	.4713	.9713	.0287	.0656
1.91	.4719	.9719	.0281	.0644
1.92	.4726	.9726	.0274	.0632
1.93	.4732	.9732	.0268	.0620
1.94	.4738	.9738	.0262	.0608
1.95	.4744	.9744	.0256	.0596
1.96	.4750	.9750	.0250	.0584
1.97	.4756	.9756	.0244	.0573
1.98	.4761	.9761	.0239	.0562
1.99	.4767	.9767	.0233	.0551
2.00	.4772	.9772	.0228	.0540
2.01	.4778	.9778	.0222	.0529
2.02	.4783	.9783	.0217	.0519
2.03	.4788	.9788	.0212	.0508
2.04	.4793	.9793	.0207	.0498
2.05	.4798	.9798	.0202	.0488
2.06	.4803	.9803	.0197	.0478
2.07	.4808	.9808	.0192	.0468
2.08	.4812	.9812	.0188	.0459
2.09	.4817	.9817	.0183	.0449

TABLE B.—AREAS AND ORDINATES OF THE NORMAL CURVE
 IN TERMS OF $\frac{x}{\sigma}$ —Continued

(1) z STANDARD SCORE $\left(\frac{x}{\sigma}\right)$	(2) A AREA FROM MEAN TO $\frac{x}{\sigma}$	(3) B AREA IN LARGER PORTION	(4) C AREA IN SMALLER PORTION	(5) y ORDINATE AT $\frac{x}{\sigma}$
2.10	.4821	.9821	.0179	.0440
2.11	.4826	.9826	.0174	.0431
2.12	.4830	.9830	.0170	.0422
2.13	.4834	.9834	.0166	.0413
2.14	.4838	.9838	.0162	.0404
2.15	.4842	.9842	.0158	.0396
2.16	.4846	.9846	.0154	.0387
2.17	.4850	.9850	.0150	.0379
2.18	.4854	.9854	.0146	.0371
2.19	.4857	.9857	.0143	.0363
2.20	.4861	.9861	.0139	.0355
2.21	.4864	.9864	.0136	.0347
2.22	.4868	.9868	.0132	.0339
2.23	.4871	.9871	.0129	.0332
2.24	.4875	.9875	.0125	.0325
2.25	.4878	.9878	.0122	.0317
2.26	.4881	.9881	.0119	.0310
2.27	.4884	.9884	.0116	.0303
2.28	.4887	.9887	.0113	.0297
2.29	.4890	.9890	.0110	.0290
2.30	.4893	.9893	.0107	.0283
2.31	.4896	.9896	.0104	.0277
2.32	.4898	.9898	.0102	.0270
2.33	.4901	.9901	.0099	.0264
2.34	.4904	.9904	.0096	.0258
2.35	.4906	.9906	.0094	.0252
2.36	.4909	.9909	.0091	.0246
2.37	.4911	.9911	.0089	.0241
2.38	.4913	.9913	.0087	.0235
2.39	.4916	.9916	.0084	.0229
2.40	.4918	.9918	.0082	.0224
2.41	.4920	.9920	.0080	.0219
2.42	.4922	.9922	.0078	.0213
2.43	.4925	.9925	.0075	.0208
2.44	.4927	.9927	.0073	.0203

TABLE B.—AREAS AND ORDINATES OF THE NORMAL CURVE
IN TERMS OF $\frac{x}{\sigma}$ —Continued

(1) z STANDARD SCORE $\left(\frac{x}{\sigma}\right)$	(2) A AREA FROM MEAN TO $\frac{x}{\sigma}$	(3) B AREA IN LARGER PORTION	(4) C AREA IN SMALLER PORTION	(5) y ORDINATE AT $\frac{x}{\sigma}$
2.45	.4929	.9929	.0071	.0198
2.46	.4931	.9931	.0069	.0194
2.47	.4932	.9932	.0068	.0189
2.48	.4934	.9934	.0066	.0184
2.49	.4936	.9936	.0064	.0180
2.50	.4938	.9938	.0062	.0175
2.51	.4940	.9940	.0060	.0171
2.52	.4941	.9941	.0059	.0167
2.53	.4943	.9943	.0057	.0163
2.54	.4945	.9945	.0055	.0158
2.55	.4946	.9946	.0054	.0154
2.56	.4948	.9948	.0052	.0151
2.57	.4949	.9949	.0051	.0147
2.58	.4951	.9951	.0049	.0143
2.59	.4952	.9952	.0048	.0139
2.60	.4953	.9953	.0047	.0136
2.61	.4955	.9955	.0045	.0132
2.62	.4956	.9956	.0044	.0129
2.63	.4957	.9957	.0043	.0126
2.64	.4959	.9959	.0041	.0122
2.65	.4960	.9960	.0040	.0119
2.66	.4961	.9961	.0039	.0116
2.67	.4962	.9962	.0038	.0113
2.68	.4963	.9963	.0037	.0110
2.69	.4964	.9964	.0036	.0107
2.70	.4965	.9965	.0035	.0104
2.71	.4966	.9966	.0034	.0101
2.72	.4967	.9967	.0033	.0099
2.73	.4968	.9968	.0032	.0096
2.74	.4969	.9969	.0031	.0093
2.75	.4970	.9970	.0030	.0091
2.76	.4971	.9971	.0029	.0088
2.77	.4972	.9972	.0028	.0086
2.78	.4973	.9973	.0027	.0084
2.79	.4974	.9974	.0026	.0081

TABLE B.—AREAS AND ORDINATES OF THE NORMAL CURVE
IN TERMS OF $\frac{x}{\sigma}$ —Continued

(1) z STANDARD SCORE $\left(\frac{x}{\sigma}\right)$	(2) A AREA FROM MEAN TO $\frac{x}{\sigma}$	(3) B AREA IN LARGER PORTION	(4) C AREA IN SMALLER PORTION	(5) y ORDINATE AT $\frac{x}{\sigma}$
2.80	.4974	.9974	.0026	.0079
2.81	.4975	.9975	.0025	.0077
2.82	.4976	.9976	.0024	.0075
2.83	.4977	.9977	.0023	.0073
2.84	.4977	.9977	.0023	.0071
2.85	.4978	.9978	.0022	.0069
2.86	.4979	.9979	.0021	.0067
2.87	.4979	.9979	.0021	.0065
2.88	.4980	.9980	.0020	.0063
2.89	.4981	.9981	.0019	.0061
2.90	.4981	.9981	.0019	.0060
2.91	.4982	.9982	.0018	.0058
2.92	.4982	.9982	.0018	.0056
2.93	.4983	.9983	.0017	.0055
2.94	.4984	.9984	.0016	.0053
2.95	.4984	.9984	.0016	.0051
2.96	.4985	.9985	.0015	.0050
2.97	.4985	.9985	.0015	.0048
2.98	.4986	.9986	.0014	.0047
2.99	.4986	.9986	.0014	.0046
3.00	.4987	.9987	.0013	.0044
3.01	.4987	.9987	.0013	.0043
3.02	.4987	.9987	.0013	.0042
3.03	.4988	.9988	.0012	.0040
3.04	.4988	.9988	.0012	.0039
3.05	.4989	.9989	.0011	.0038
3.06	.4989	.9989	.0011	.0037
3.07	.4989	.9989	.0011	.0036
3.08	.4990	.9990	.0010	.0035
3.09	.4990	.9990	.0010	.0034
3.10	.4990	.9990	.0010	.0033
3.11	.4991	.9991	.0009	.0032
3.12	.4991	.9991	.0009	.0031
3.13	.4991	.9991	.0009	.0030
3.14	.4992	.9992	.0008	.0029

TABLE B.—AREAS AND ORDINATES OF THE NORMAL CURVE
 IN TERMS OF $\frac{x}{\sigma}$ —Concluded

(1) z STANDARD SCORE $\left(\frac{x}{\sigma}\right)$	(2) A AREA FROM MEAN TO $\frac{x}{\sigma}$	(3) B AREA IN LARGER PORTION	(4) C AREA IN SMALLER PORTION	(5) y ORDINATE AT $\frac{x}{\sigma}$
3.15	.4992	.9992	.0008	.0028
3.16	.4992	.9992	.0008	.0027
3.17	.4992	.9992	.0008	.0026
3.18	.4993	.9993	.0007	.0025
3.19	.4993	.9993	.0007	.0025
3.20	.4993	.9993	.0007	.0024
3.21	.4993	.9993	.0007	.0023
3.22	.4994	.9994	.0006	.0022
3.23	.4994	.9994	.0006	.0022
3.24	.4994	.9994	.0006	.0021
3.30	.4995	.9995	.0005	.0017
3.40	.4997	.9997	.0003	.0012
3.50	.4998	.9998	.0002	.0009
3.60	.4998	.9998	.0002	.0006
3.70	.4999	.9999	.0001	.0004

TABLE C.—VALUES OF t AT THE 5% AND 1% LEVELS OF SIGNIFICANCE*

DEGREES OF FREEDOM	5%	1%	DEGREES OF FREEDOM	5%	1%
1	12.706	63.657	32	2.037	2.739
2	4.303	9.925	34	2.032	2.728
3	3.182	5.841	36	2.027	2.718
4	2.776	4.604	38	2.025	2.711
5	2.571	4.032	40	2.021	2.704
6	2.447	3.707	42	2.017	2.696
7	2.365	3.499	44	2.015	2.691
8	2.306	3.355	46	2.012	2.685
9	2.262	3.250	48	2.010	2.681
10	2.228	3.169	50	2.008	2.678
11	2.201	3.106	55	2.005	2.668
12	2.179	3.055	60	2.000	2.660
13	2.160	3.012	65	1.998	2.653
14	2.145	2.977	70	1.994	2.648
15	2.131	2.947	80	1.990	2.638
16	2.120	2.921	90	1.987	2.632
17	2.110	2.898	100	1.984	2.626
18	2.101	2.878	125	1.979	2.616
19	2.093	2.861	150	1.976	2.609
20	2.086	2.845	200	1.972	2.601
21	2.080	2.831	300	1.968	2.592
22	2.074	2.819	400	1.966	2.588
23	2.069	2.807	500	1.965	2.586
24	2.064	2.797	1000	1.962	2.581
25	2.060	2.787	∞	1.960	2.576
26	2.056	2.779			
27	2.052	2.771			
28	2.048	2.763			
29	2.045	2.756			
30	2.042	2.750			

*Table C is abridged from Table IV of Fisher: *Statistical Methods for Research Workers*. Oliver & Boyd, Ltd., Edinburgh, by permission of the Author and Publishers. Additional entries were taken from Snedecor: *Statistical Methods*. Ames, Iowa: Collegiate Press, by permission.

TABLE D.—VALUES OF t AT THE 5% AND 1% LEVELS OF SIGNIFICANCE*

DEGREES OF FREEDOM	5%	1%	DEGREES OF FREEDOM	5%	1%
1	.997	1.000	24	.388	.496
2	.950	.990	25	.381	.487
3	.878	.959	26	.374	.478
4	.811	.917	27	.367	.470
5	.754	.874	28	.361	.463
6	.707	.834	29	.355	.456
7	.666	.798	30	.349	.449
8	.632	.765	35	.325	.418
9	.602	.735	40	.304	.393
10	.576	.708	45	.288	.372
11	.553	.684	50	.273	.354
12	.532	.661	60	.250	.325
13	.514	.641	70	.232	.302
14	.497	.623	80	.217	.283
15	.482	.606	90	.205	.267
16	.468	.590	100	.195	.254
17	.456	.575	125	.174	.228
18	.444	.561	150	.159	.208
19	.433	.549	200	.138	.181
20	.423	.537	300	.113	.148
21	.413	.526	400	.098	.128
22	.404	.515	500	.088	.115
23	.396	.505	1000	.062	.081

* Table D is abridged from Table V.A. of Fisher: *Statistical Methods for Research Workers*. Oliver & Boyd, Ltd., Edinburgh, by permission of the Author and Publisher. Additional entries were taken from Snedecor: *Statistical Methods*. Ames, Iowa: Collegiate Press, by permission.

TABLE E.—VALUES OF *F* AT THE 5% (LIGHTFACE TYPE) AND THE 1% (BOLDFACE TYPE) LEVELS OF SIGNIFICANCE*

df	n, degrees of freedom (for greater mean square)																							
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
1	161 4,052	200 4,999	216 5,403	225 5,625	230 5,764	234 5,859	237 5,928	239 5,981	241 6,022	242 6,056	243 6,082	244 6,106	245 6,142	246 6,169	248 6,208	249 6,234	250 6,258	251 6,286	252 6,302	253 6,323	254 6,334	254 6,352	254 6,361	254 6,366
2	18.51 98.49	19.00 99.00	19.16 99.17	19.25 99.25	19.30 99.30	19.33 99.33	19.36 99.34	19.37 99.36	19.38 99.38	19.39 99.40	19.40 99.41	19.41 99.42	19.42 99.43	19.43 99.44	19.44 99.45	19.45 99.46	19.46 99.47	19.47 99.48	19.47 99.48	19.48 99.49	19.49 99.49	19.49 99.49	19.50 99.50	19.50 99.50
3	10.13 34.12	9.55 30.82	9.28 29.46	9.12 28.71	9.01 28.24	8.94 27.91	8.88 27.67	8.84 27.49	8.81 27.34	8.78 27.23	8.76 27.13	8.74 27.05	8.71 26.92	8.69 26.83	8.66 26.69	8.64 26.60	8.62 26.50	8.60 26.41	8.58 26.35	8.57 26.27	8.56 26.23	8.54 26.18	8.54 26.14	8.53 26.12
4	7.71 21.20	6.94 18.00	6.59 16.69	6.39 15.98	6.26 15.52	6.16 15.21	6.09 14.98	6.04 14.80	6.00 14.66	5.96 14.54	5.93 14.45	5.91 14.37	5.87 14.24	5.84 14.15	5.80 14.02	5.77 13.93	5.74 13.83	5.71 13.74	5.70 13.69	5.68 13.61	5.66 13.57	5.65 13.52	5.64 13.48	5.63 13.46
5	6.61 16.26	5.79 13.27	5.41 12.06	5.19 11.39	5.05 10.97	4.95 10.67	4.88 10.45	4.82 10.27	4.78 10.15	4.74 10.05	4.70 9.96	4.68 9.89	4.64 9.77	4.60 9.68	4.56 9.55	4.53 9.47	4.50 9.38	4.46 9.29	4.44 9.24	4.42 9.17	4.40 9.13	4.38 9.07	4.37 9.04	4.36 9.02
6	5.99 13.74	5.14 10.92	4.76 9.78	4.53 9.15	4.39 8.75	4.28 8.47	4.21 8.26	4.15 8.10	4.10 7.98	4.08 7.87	4.08 7.79	4.00 7.72	3.96 7.60	3.92 7.52	3.87 7.39	3.84 7.31	3.81 7.23	3.77 7.14	3.75 7.09	3.72 7.02	3.71 6.99	3.69 6.94	3.68 6.90	3.67 6.88
7	5.59 12.25	4.74 9.55	4.35 8.45	4.12 7.85	3.97 7.46	3.87 7.19	3.79 7.00	3.73 6.84	3.68 6.71	3.63 6.62	3.60 6.54	3.57 6.47	3.52 6.35	3.49 6.27	3.44 6.15	3.41 6.07	3.38 5.98	3.34 5.90	3.32 5.85	3.29 5.78	3.28 5.75	3.25 5.70	3.24 5.67	3.23 5.65
8	5.32 11.26	4.46 8.65	4.07 7.59	3.84 7.01	3.69 6.63	3.58 6.37	3.50 6.19	3.44 6.03	3.39 5.91	3.34 5.82	3.31 5.74	3.28 5.67	3.23 5.56	3.20 5.48	3.15 5.36	3.12 5.28	3.08 5.20	3.05 5.11	3.03 5.06	3.00 5.00	2.98 4.96	2.96 4.91	2.94 4.88	2.93 4.86
9	5.12 10.56	4.26 8.02	3.86 6.99	3.63 6.42	3.48 6.06	3.37 5.80	3.29 5.62	3.23 5.47	3.18 5.35	3.13 5.26	3.10 5.18	3.07 5.11	3.02 5.00	2.98 4.92	2.93 4.80	2.90 4.73	2.86 4.64	2.82 4.56	2.80 4.51	2.77 4.45	2.76 4.41	2.73 4.36	2.72 4.33	2.71 4.31
10	4.96 10.04	4.10 7.56	3.71 6.55	3.48 5.99	3.33 5.64	3.22 5.39	3.14 5.21	3.07 5.06	3.02 4.95	2.97 4.85	2.94 4.78	2.91 4.71	2.86 4.60	2.82 4.52	2.77 4.41	2.74 4.33	2.70 4.25	2.67 4.17	2.64 4.12	2.61 4.05	2.59 4.01	2.56 3.96	2.55 3.93	2.54 3.91
11	4.84 9.65	3.98 7.20	3.59 6.22	3.36 5.67	3.20 5.32	3.09 5.07	3.01 4.88	2.95 4.74	2.90 4.63	2.86 4.54	2.82 4.46	2.79 4.40	2.74 4.29	2.70 4.21	2.65 4.10	2.61 4.02	2.57 3.94	2.53 3.86	2.50 3.80	2.47 3.74	2.45 3.70	2.42 3.66	2.41 3.62	2.40 3.60
12	4.75 9.23	3.88 6.93	3.49 5.95	3.26 5.41	3.11 5.06	2.92 4.82	2.85 4.65	2.80 4.50	2.76 4.39	2.72 4.30	2.69 4.22	2.64 4.16	2.60 4.05	2.54 3.98	2.50 3.86	2.46 3.78	2.42 3.70	2.40 3.61	2.36 3.56	2.35 3.49	2.32 3.46	2.31 3.41	2.30 3.38	2.30 3.36
13	4.67 9.07	3.80 6.70	3.41 5.74	3.18 5.20	3.02 4.86	2.92 4.62	2.84 4.44	2.77 4.30	2.72 4.19	2.67 4.10	2.63 4.02	2.60 3.96	2.55 3.85	2.51 3.78	2.46 3.67	2.42 3.59	2.38 3.51	2.34 3.42	2.32 3.37	2.28 3.30	2.26 3.27	2.24 3.21	2.22 3.18	2.21 3.16

* Reproduced from G. W. Snedecor. *Statistical Methods*. Ames, Iowa: Collegiate Press, 1940, pp. 184-187, by permission of the author and the publisher.

TABLE E.—VALUES OF F AT THE 5% (LIGHTFACE TYPE) AND THE 1% (BOLDFACE TYPE) LEVELS OF SIGNIFICANCE—Continued

n_1	n_2 degrees of freedom (for greater mean square)																							
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
14	4.60 8.86	3.74 6.51	3.34 5.56	3.11 5.03	2.96 4.69	2.85 4.46	2.77 4.28	2.70 4.14	2.65 4.03	2.60 3.94	2.56 3.86	2.53 3.80	2.48 3.70	2.44 3.62	2.39 3.51	2.35 3.43	2.31 3.34	2.27 3.26	2.24 3.21	2.21 3.14	2.19 3.11	2.16 3.06	2.14 3.02	2.13 3.00
15	4.54 8.68	3.68 6.36	3.29 5.42	3.06 4.89	2.90 4.56	2.79 4.32	2.70 4.14	2.64 4.00	2.59 3.89	2.55 3.80	2.51 3.73	2.48 3.67	2.43 3.56	2.39 3.48	2.33 3.29	2.29 3.29	2.25 3.20	2.21 3.12	2.18 3.07	2.15 3.00	2.12 2.97	2.10 2.92	2.08 2.89	2.07 2.87
16	4.49 8.53	3.63 6.23	3.24 5.29	3.01 4.77	2.85 4.44	2.74 4.20	2.66 4.03	2.59 3.89	2.54 3.78	2.49 3.69	2.45 3.61	2.42 3.55	2.37 3.45	2.33 3.37	2.28 3.25	2.24 3.18	2.20 3.10	2.16 3.01	2.13 2.96	2.09 2.89	2.07 2.86	2.04 2.80	2.02 2.77	2.01 2.75
17	4.45 8.40	3.59 6.11	3.20 5.18	2.96 4.67	2.81 4.34	2.70 4.10	2.62 3.93	2.55 3.79	2.50 3.68	2.45 3.59	2.41 3.52	2.38 3.45	2.33 3.35	2.29 3.27	2.23 3.16	2.19 3.08	2.15 3.00	2.11 2.92	2.08 2.86	2.04 2.79	2.02 2.76	1.99 2.70	1.97 2.67	1.96 2.65
18	4.41 8.28	3.55 6.01	3.16 5.09	2.93 4.58	2.77 4.25	2.66 4.01	2.58 3.85	2.51 3.71	2.46 3.60	2.41 3.51	2.37 3.44	2.34 3.37	2.29 3.27	2.25 3.19	2.19 3.07	2.15 3.00	2.11 2.91	2.07 2.83	2.04 2.78	2.00 2.71	1.98 2.68	1.95 2.62	1.93 2.59	1.92 2.57
19	4.38 8.18	3.52 5.93	3.13 5.01	2.90 4.50	2.74 4.17	2.63 3.94	2.55 3.77	2.48 3.63	2.43 3.52	2.38 3.43	2.34 3.36	2.31 3.30	2.26 3.19	2.21 3.12	2.15 3.00	2.11 2.92	2.07 2.84	2.02 2.76	2.00 2.70	1.96 2.63	1.94 2.60	1.91 2.54	1.90 2.51	1.88 2.49
20	4.35 8.10	3.49 5.85	3.10 4.94	2.87 4.43	2.71 4.10	2.60 3.87	2.52 3.71	2.45 3.56	2.40 3.45	2.35 3.37	2.31 3.30	2.28 3.23	2.23 3.13	2.18 3.05	2.12 2.94	2.08 2.86	2.04 2.77	1.99 2.69	1.96 2.63	1.92 2.56	1.90 2.53	1.87 2.47	1.85 2.44	1.84 2.42
21	4.32 8.02	3.47 5.78	3.07 4.87	2.84 4.37	2.68 4.04	2.57 3.81	2.49 3.65	2.42 3.51	2.37 3.40	2.32 3.31	2.28 3.24	2.25 3.17	2.20 3.07	2.15 2.99	2.09 2.88	2.05 2.80	2.00 2.72	1.96 2.63	1.93 2.58	1.89 2.51	1.87 2.47	1.84 2.42	1.82 2.38	1.81 2.36
22	4.30 7.94	3.44 5.72	3.05 4.82	2.82 4.31	2.66 3.99	2.55 3.76	2.47 3.59	2.40 3.45	2.35 3.35	2.30 3.26	2.26 3.18	2.23 3.12	2.18 3.02	2.13 2.94	2.07 2.83	2.03 2.75	1.98 2.67	1.93 2.58	1.91 2.53	1.87 2.46	1.84 2.42	1.81 2.37	1.80 2.33	1.78 2.31
23	4.28 7.88	3.42 5.66	3.03 4.76	2.80 4.26	2.64 3.94	2.53 3.71	2.45 3.54	2.38 3.41	2.32 3.30	2.28 3.21	2.24 3.14	2.20 3.07	2.14 2.97	2.10 2.89	2.04 2.78	2.00 2.70	1.96 2.62	1.91 2.53	1.88 2.48	1.84 2.41	1.82 2.37	1.79 2.32	1.77 2.28	1.76 2.26
24	4.26 7.82	3.40 5.61	3.01 4.72	2.78 4.22	2.62 3.90	2.51 3.67	2.43 3.50	2.36 3.36	2.30 3.25	2.26 3.17	2.22 3.09	2.18 3.03	2.13 2.93	2.09 2.85	2.02 2.74	1.98 2.66	1.94 2.58	1.89 2.49	1.86 2.44	1.83 2.36	1.80 2.33	1.76 2.27	1.74 2.23	1.73 2.21
25	4.24 7.77	3.38 5.57	2.99 4.68	2.76 4.18	2.60 3.86	2.49 3.63	2.41 3.46	2.34 3.32	2.28 3.21	2.24 3.13	2.20 3.05	2.16 2.99	2.11 2.89	2.06 2.81	2.00 2.70	1.96 2.62	1.92 2.54	1.87 2.45	1.84 2.40	1.80 2.32	1.77 2.29	1.74 2.23	1.72 2.19	1.71 2.17
26	4.22 7.72	3.37 5.53	2.98 4.64	2.74 4.14	2.59 3.82	2.47 3.59	2.39 3.42	2.32 3.29	2.27 3.17	2.22 3.09	2.18 3.02	2.15 2.96	2.10 2.86	2.05 2.77	1.99 2.66	1.95 2.58	1.90 2.50	1.85 2.41	1.82 2.36	1.78 2.28	1.76 2.25	1.72 2.19	1.70 2.15	1.69 2.13

The function, $F = s$ with exponent $2s$, is computed in part from Fisher's table VI (7). Additional entries are by interpolation, mostly graphical.

TABLE E.—VALUES OF F AT THE 5% (LIGHTFACE TYPE) AND THE 1% (BOLDFACE TYPE) LEVELS OF SIGNIFICANCE—Continued

n_2	n_1 degrees of freedom (for greater mean square)																							
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
27	4.21	3.85	2.96	2.78	2.57	2.46	2.37	2.30	2.25	2.20	2.16	2.13	2.08	2.03	1.97	1.93	1.88	1.84	1.80	1.76	1.74	1.71	1.68	1.67
	7.68	5.49	4.60	4.11	3.79	3.56	3.39	3.26	3.14	3.06	2.98	2.93	2.83	2.74	2.63	2.55	2.47	2.38	2.33	2.25	2.21	2.16	2.12	2.10
28	4.20	3.84	2.95	2.71	2.56	2.44	2.36	2.29	2.24	2.19	2.15	2.12	2.06	2.02	1.96	1.91	1.87	1.81	1.78	1.75	1.72	1.69	1.67	1.65
	7.64	5.45	4.57	4.07	3.76	3.53	3.36	3.23	3.11	3.03	2.95	2.90	2.80	2.71	2.60	2.52	2.44	2.35	2.30	2.22	2.18	2.13	2.09	2.06
29	4.18	3.83	2.93	2.70	2.54	2.43	2.35	2.28	2.22	2.18	2.14	2.10	2.05	2.00	1.94	1.90	1.85	1.80	1.77	1.73	1.71	1.68	1.65	1.64
	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.08	3.00	2.92	2.87	2.77	2.68	2.57	2.49	2.41	2.32	2.27	2.19	2.15	2.10	2.06	2.03
30	4.17	3.82	2.92	2.69	2.53	2.42	2.34	2.27	2.21	2.16	2.12	2.09	2.04	1.99	1.93	1.89	1.84	1.79	1.76	1.72	1.69	1.66	1.64	1.62
	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.06	2.98	2.90	2.84	2.74	2.66	2.55	2.47	2.38	2.29	2.24	2.16	2.13	2.07	2.03	2.01
32	4.15	3.80	2.90	2.67	2.51	2.40	2.32	2.25	2.19	2.14	2.10	2.07	2.02	1.97	1.91	1.86	1.82	1.76	1.74	1.69	1.67	1.64	1.61	1.59
	7.50	5.34	4.46	3.97	3.66	3.42	3.25	3.12	3.01	2.94	2.86	2.80	2.70	2.62	2.51	2.42	2.34	2.25	2.20	2.12	2.08	2.02	1.98	1.96
34	4.13	3.78	2.88	2.65	2.49	2.38	2.30	2.23	2.17	2.12	2.08	2.05	2.00	1.95	1.89	1.84	1.80	1.74	1.71	1.67	1.64	1.61	1.59	1.57
	7.44	5.29	4.42	3.93	3.61	3.38	3.21	3.08	2.97	2.89	2.82	2.76	2.66	2.58	2.47	2.38	2.30	2.21	2.15	2.08	2.04	1.98	1.94	1.91
36	4.11	3.76	2.86	2.63	2.48	2.36	2.28	2.21	2.15	2.10	2.06	2.03	1.98	1.93	1.87	1.82	1.78	1.72	1.69	1.65	1.62	1.59	1.56	1.55
	7.39	5.25	4.38	3.89	3.58	3.35	3.18	3.04	2.94	2.86	2.78	2.72	2.62	2.54	2.43	2.35	2.26	2.17	2.12	2.04	2.00	1.94	1.90	1.87
38	4.10	3.75	2.85	2.62	2.46	2.35	2.28	2.19	2.14	2.09	2.05	2.02	1.96	1.92	1.85	1.80	1.76	1.71	1.67	1.63	1.60	1.57	1.54	1.53
	7.35	5.21	4.34	3.86	3.54	3.32	3.15	3.02	2.91	2.82	2.75	2.69	2.59	2.51	2.40	2.32	2.22	2.14	2.08	2.00	1.97	1.90	1.86	1.84
40	4.08	3.73	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.07	2.04	2.00	1.95	1.90	1.84	1.79	1.74	1.69	1.66	1.61	1.59	1.55	1.53	1.51
	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.88	2.80	2.73	2.66	2.56	2.49	2.37	2.29	2.20	2.11	2.05	1.97	1.94	1.88	1.84	1.81
42	4.07	3.72	2.83	2.60	2.44	2.32	2.24	2.17	2.11	2.06	2.02	1.99	1.94	1.89	1.82	1.78	1.73	1.68	1.64	1.60	1.57	1.54	1.51	1.49
	7.27	5.15	4.29	3.80	3.49	3.26	3.10	2.96	2.86	2.77	2.70	2.64	2.54	2.46	2.35	2.26	2.17	2.08	2.02	1.94	1.91	1.85	1.80	1.78
44	4.06	3.71	2.82	2.59	2.43	2.31	2.23	2.16	2.10	2.05	2.01	1.98	1.92	1.88	1.81	1.76	1.72	1.66	1.63	1.58	1.56	1.52	1.50	1.48
	7.24	5.12	4.26	3.78	3.46	3.24	3.07	2.94	2.84	2.75	2.68	2.62	2.52	2.44	2.32	2.24	2.15	2.06	2.00	1.92	1.88	1.82	1.78	1.75
46	4.05	3.70	2.81	2.57	2.42	2.30	2.22	2.14	2.09	2.04	2.00	1.97	1.91	1.87	1.80	1.75	1.71	1.65	1.62	1.57	1.54	1.51	1.48	1.46
	7.21	5.10	4.24	3.76	3.44	3.22	3.05	2.92	2.82	2.73	2.66	2.60	2.50	2.42	2.30	2.22	2.13	2.04	1.98	1.90	1.86	1.80	1.76	1.72
48	4.04	3.19	2.80	2.56	2.41	2.30	2.21	2.14	2.08	2.03	1.99	1.96	1.90	1.86	1.79	1.74	1.70	1.64	1.61	1.56	1.53	1.50	1.47	1.45
	7.19	5.08	4.22	3.74	3.42	3.20	3.04	2.90	2.80	2.71	2.64	2.58	2.48	2.40	2.28	2.20	2.11	2.02	1.96	1.88	1.84	1.78	1.73	1.70

TABLE E.—VALUES OF F AT THE 5% (LIGHTFACE TYPE) AND THE 1% (BOLDFACE TYPE) LEVELS OF SIGNIFICANCE—Continued

n_1	n_2 degrees of freedom (for greater mean square)																							
	1	2	3	4	5	6	7	8	9	10	11	12	14	15	20	24	30	40	50	75	100	200	500	∞
50	4.03 7.17	3.18 5.06	2.79 4.20	2.56 3.72	2.40 3.41	2.29 3.18	2.20 3.02	2.13 2.88	2.07 2.78	2.02 2.70	1.98 2.62	1.95 2.56	1.90 2.46	1.85 2.39	1.78 2.26	1.74 2.18	1.69 2.10	1.63 2.00	1.60 1.94	1.55 1.86	1.52 1.82	1.48 1.76	1.46 1.71	1.44 1.68
55	4.02 7.12	3.17 5.01	2.78 4.16	2.54 3.68	2.38 3.37	2.27 3.15	2.18 2.98	2.11 2.85	2.05 2.75	2.00 2.66	1.97 2.59	1.93 2.53	1.88 2.43	1.83 2.35	1.76 2.23	1.72 2.15	1.67 2.06	1.61 1.96	1.58 1.90	1.52 1.82	1.60 1.78	1.48 1.71	1.43 1.66	1.41 1.64
60	4.00 7.08	3.15 4.98	2.76 4.13	2.52 3.65	2.37 3.34	2.25 3.12	2.17 2.95	2.10 2.82	2.04 2.72	1.99 2.63	1.95 2.56	1.92 2.50	1.86 2.40	1.81 2.32	1.75 2.20	1.70 2.12	1.65 2.03	1.59 1.93	1.56 1.87	1.50 1.79	1.48 1.74	1.44 1.68	1.41 1.63	1.39 1.60
65	3.99 7.04	3.14 4.95	2.75 4.10	2.51 3.62	2.36 3.31	2.24 3.09	2.15 2.93	2.08 2.79	2.02 2.70	1.98 2.61	1.94 2.54	1.90 2.47	1.85 2.37	1.80 2.30	1.73 2.18	1.68 2.09	1.63 2.00	1.57 1.90	1.54 1.84	1.49 1.76	1.48 1.71	1.42 1.64	1.39 1.60	1.37 1.56
70	3.98 7.01	3.13 4.92	2.74 4.08	2.50 3.60	2.35 3.29	2.23 3.07	2.14 2.91	2.07 2.77	2.01 2.67	1.97 2.59	1.93 2.51	1.89 2.45	1.84 2.35	1.79 2.28	1.72 2.15	1.67 2.07	1.62 1.98	1.56 1.88	1.53 1.82	1.47 1.74	1.45 1.69	1.40 1.62	1.37 1.56	1.35 1.53
80	3.96 6.96	3.11 4.88	2.72 4.04	2.48 3.56	2.33 3.25	2.21 3.04	2.12 2.87	2.05 2.74	1.99 2.64	1.95 2.55	1.91 2.48	1.88 2.41	1.82 2.32	1.77 2.24	1.70 2.11	1.65 2.03	1.60 1.94	1.54 1.84	1.51 1.78	1.45 1.70	1.42 1.65	1.38 1.57	1.35 1.52	1.32 1.49
100	3.94 6.90	3.09 4.82	2.70 3.98	2.46 3.51	2.30 3.20	2.19 2.99	2.10 2.82	2.03 2.69	1.97 2.59	1.92 2.51	1.88 2.43	1.85 2.36	1.79 2.26	1.75 2.19	1.68 2.06	1.63 1.98	1.57 1.89	1.51 1.79	1.48 1.73	1.42 1.64	1.39 1.59	1.34 1.51	1.30 1.46	1.28 1.43
125	3.92 6.84	3.07 4.78	2.68 3.94	2.44 3.47	2.29 3.17	2.17 2.95	2.08 2.79	2.01 2.65	1.95 2.56	1.90 2.47	1.86 2.40	1.83 2.33	1.77 2.23	1.72 2.15	1.65 2.03	1.60 1.94	1.55 1.85	1.49 1.75	1.45 1.68	1.39 1.59	1.36 1.54	1.31 1.46	1.27 1.40	1.25 1.37
150	3.91 6.81	3.06 4.75	2.67 3.91	2.43 3.44	2.27 3.14	2.16 2.92	2.07 2.76	2.00 2.62	1.94 2.53	1.89 2.44	1.85 2.37	1.82 2.30	1.76 2.20	1.71 2.12	1.64 2.00	1.59 1.91	1.54 1.83	1.47 1.72	1.44 1.66	1.37 1.56	1.34 1.51	1.29 1.43	1.25 1.37	1.22 1.33
200	3.89 6.76	3.04 4.71	2.65 3.88	2.41 3.41	2.26 3.11	2.14 2.90	2.05 2.73	1.98 2.60	1.92 2.50	1.87 2.41	1.83 2.34	1.80 2.28	1.74 2.17	1.69 2.09	1.62 1.97	1.57 1.88	1.52 1.79	1.45 1.69	1.42 1.62	1.35 1.53	1.32 1.48	1.26 1.39	1.22 1.33	1.19 1.28
400	3.86 6.70	3.02 4.66	2.62 3.83	2.39 3.36	2.23 3.06	2.12 2.85	2.03 2.69	1.96 2.55	1.90 2.46	1.85 2.37	1.81 2.29	1.78 2.23	1.72 2.12	1.67 2.04	1.60 1.92	1.54 1.84	1.49 1.74	1.42 1.64	1.38 1.57	1.32 1.47	1.28 1.42	1.22 1.32	1.16 1.24	1.13 1.19
1000	3.85 6.66	3.00 4.62	2.61 3.80	2.38 3.34	2.22 3.04	2.10 2.82	2.02 2.66	1.95 2.53	1.89 2.43	1.84 2.34	1.80 2.26	1.76 2.20	1.70 2.09	1.65 2.01	1.58 1.89	1.53 1.81	1.47 1.71	1.41 1.61	1.36 1.54	1.30 1.44	1.26 1.38	1.19 1.28	1.13 1.19	1.08 1.11
∞	3.84 6.64	2.99 4.60	2.60 3.78	2.37 3.32	2.21 3.02	2.09 2.80	2.01 2.64	1.94 2.51	1.88 2.41	1.83 2.32	1.79 2.24	1.75 2.18	1.69 2.07	1.64 1.99	1.57 1.87	1.52 1.79	1.46 1.69	1.40 1.59	1.35 1.52	1.28 1.41	1.24 1.36	1.17 1.25	1.11 1.15	1.00 1.00

TABLE F.—VALUES OF r^2 AT THE 5% (ROMAN TYPE) AND 1% (ITALIC TYPE) LEVELS OF SIGNIFICANCE*

$k - 1$

$N-k$	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	
1	.988 <i>.9995</i>																			
2	.854 <i>.870</i>	.900 <i>.980</i>																		
3	.695 <i>.828</i>	.774 <i>.928</i>	.805 <i>.854</i>																	
4	.573 <i>.808</i>	.665 <i>.860</i>	.706 <i>.871</i>	.729 <i>.888</i>																
5	.483 <i>.718</i>	.578 <i>.808</i>	.623 <i>.828</i>	.651 <i>.833</i>	.689 <i>.833</i>															
6	.418 <i>.645</i>	.509 <i>.718</i>	.556 <i>.746</i>	.583 <i>.765</i>	.606 <i>.779</i>	.621 <i>.789</i>														
7	.365 <i>.584</i>	.454 <i>.665</i>	.501 <i>.691</i>	.532 <i>.714</i>	.553 <i>.739</i>	.570 <i>.741</i>	.582 <i>.760</i>													
8	.324 <i>.533</i>	.409 <i>.605</i>	.456 <i>.645</i>	.486 <i>.667</i>	.509 <i>.684</i>	.525 <i>.697</i>	.538 <i>.708</i>	.550 <i>.716</i>												
9	.292 <i>.489</i>	.372 <i>.561</i>	.417 <i>.600</i>	.447 <i>.625</i>	.470 <i>.644</i>	.487 <i>.658</i>	.500 <i>.669</i>	.512 <i>.678</i>	.522 <i>.686</i>											
10	.264 <i>.451</i>	.341 <i>.538</i>	.385 <i>.569</i>	.415 <i>.602</i>	.437 <i>.627</i>	.454 <i>.648</i>	.468 <i>.663</i>	.479 <i>.675</i>	.489 <i>.685</i>	.496 <i>.692</i>										
11	.242 <i>.418</i>	.314 <i>.488</i>	.357 <i>.538</i>	.386 <i>.555</i>	.407 <i>.574</i>	.425 <i>.590</i>	.439 <i>.601</i>	.451 <i>.612</i>	.461 <i>.620</i>	.470 <i>.628</i>	.476 <i>.634</i>									
12	.224 <i>.391</i>	.291 <i>.459</i>	.332 <i>.497</i>	.361 <i>.524</i>	.383 <i>.544</i>	.400 <i>.560</i>	.414 <i>.574</i>	.425 <i>.583</i>	.435 <i>.592</i>	.444 <i>.600</i>	.451 <i>.608</i>	.458 <i>.618</i>								
13	.208 <i>.366</i>	.272 <i>.438</i>	.311 <i>.471</i>	.339 <i>.497</i>	.359 <i>.517</i>	.377 <i>.533</i>	.392 <i>.548</i>	.403 <i>.557</i>	.413 <i>.566</i>	.421 <i>.574</i>	.428 <i>.581</i>	.434 <i>.587</i>								

*In the body of this table is given in standard type the maximum r^2 that could be expected 5 times in 100 and, in italics, the maximum r^2 that could be expected 1 time in 100, on the basis of chance fluctuation when the true correlation is zero, the values being given for the number of columns in the correlation table less 1 ($k - 1$) by columns and the total population of the sample less the number of columns ($N - k$) by rows. The reason the section of this table covered by this note is blank is because no valid meaning could attach to r^2 when the average number of items in each column is less than two.

This table was made by Charles H. Griffin by the formula derived on pp. 421-422, of Peters and Van Voorhis *Statistical Procedures and Their Mathematical Bases* and is reproduced from pp. 494-497 by permission of the authors and their publishers, McGraw-Hill Book Company.

TABLE F.—VALUES OF ϵ^2 AT THE 5% (ROMAN TYPE) AND 1% (ITALIC TYPE) LEVELS OF SIGNIFICANCE—Continued

$k - 1$

$N - k$	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50
14	.194	.255	.292	.319	.340	.357	.371	.382	.393	.400	.407	.414	.425						
15	<i>.344</i>	<i>.408</i>	<i>.446</i>	<i>.478</i>	<i>.493</i>	<i>.509</i>	<i>.523</i>	<i>.533</i>	<i>.543</i>	<i>.551</i>	<i>.557</i>	<i>.564</i>	<i>.574</i>						
16	.181	.240	.276	.302	.322	.338	.351	.363	.374	.383	.390	.397	.408	.418					
17	<i>.524</i>	<i>.587</i>	<i>.624</i>	<i>.650</i>	<i>.671</i>	<i>.687</i>	<i>.700</i>	<i>.711</i>	<i>.720</i>	<i>.728</i>	<i>.736</i>	<i>.743</i>	<i>.753</i>	<i>.761</i>					
18	.170	.228	.261	.287	.306	.322	.338	.348	.357	.364	.371	.378	.390	.399	.418				
19	<i>.307</i>	<i>.368</i>	<i>.404</i>	<i>.430</i>	<i>.450</i>	<i>.466</i>	<i>.480</i>	<i>.491</i>	<i>.500</i>	<i>.509</i>	<i>.516</i>	<i>.523</i>	<i>.533</i>	<i>.543</i>	<i>.558</i>				
20	.161	.214	.248	.272	.291	.307	.321	.332	.342	.349	.356	.363	.375	.385	.399				
21	<i>.591</i>	<i>.650</i>	<i>.686</i>	<i>.711</i>	<i>.728</i>	<i>.747</i>	<i>.761</i>	<i>.771</i>	<i>.781</i>	<i>.790</i>	<i>.797</i>	<i>.803</i>	<i>.815</i>	<i>.824</i>	<i>.839</i>				
22	.152	.203	.236	.260	.278	.293	.307	.317	.327	.335	.342	.349	.361	.370	.385				
23	<i>.377</i>	<i>.434</i>	<i>.469</i>	<i>.494</i>	<i>.514</i>	<i>.529</i>	<i>.544</i>	<i>.556</i>	<i>.564</i>	<i>.573</i>	<i>.581</i>	<i>.587</i>	<i>.598</i>	<i>.608</i>	<i>.621</i>				
24	.145	.193	.225	.248	.266	.281	.294	.305	.315	.322	.329	.336	.348	.356	.371				
25	<i>.564</i>	<i>.620</i>	<i>.654</i>	<i>.678</i>	<i>.698</i>	<i>.714</i>	<i>.727</i>	<i>.738</i>	<i>.748</i>	<i>.756</i>	<i>.764</i>	<i>.771</i>	<i>.783</i>	<i>.792</i>	<i>.808</i>				
26	.138	.185	.215	.238	.255	.270	.283	.293	.303	.310	.317	.324	.336	.344	.359	.371			
27	<i>.853</i>	<i>.908</i>	<i>.939</i>	<i>.964</i>	<i>.983</i>	<i>.998</i>	<i>.1.013</i>	<i>.1.022</i>	<i>.1.028</i>	<i>.1.034</i>	<i>.1.039</i>	<i>.1.044</i>	<i>.1.053</i>	<i>.1.061</i>	<i>.1.077</i>	<i>.1.092</i>	<i>.1.104</i>	<i>.1.117</i>	<i>.1.131</i>
28	.131	.177	.206	.227	.244	.259	.271	.281	.291	.299	.306	.313	.324	.332	.347	.359			
29	<i>.642</i>	<i>.704</i>	<i>.736</i>	<i>.760</i>	<i>.779</i>	<i>.794</i>	<i>.808</i>	<i>.820</i>	<i>.831</i>	<i>.841</i>	<i>.850</i>	<i>.857</i>	<i>.869</i>	<i>.878</i>	<i>.894</i>	<i>.908</i>	<i>.921</i>	<i>.934</i>	<i>.948</i>
30	.126	.169	.197	.219	.235	.249	.262	.272	.282	.289	.296	.303	.315	.322	.338	.350			
31	<i>.832</i>	<i>.888</i>	<i>.914</i>	<i>.937</i>	<i>.956</i>	<i>.972</i>	<i>.985</i>	<i>.996</i>	<i>.1.006</i>	<i>.1.014</i>	<i>.1.021</i>	<i>.1.028</i>	<i>.1.040</i>	<i>.1.050</i>	<i>.1.066</i>	<i>.1.081</i>	<i>.1.094</i>	<i>.1.107</i>	<i>.1.121</i>
32	.120	.162	.190	.211	.227	.240	.253	.263	.271	.279	.286	.291	.301	.311	.326	.338			
33	<i>.623</i>	<i>.679</i>	<i>.703</i>	<i>.726</i>	<i>.744</i>	<i>.759</i>	<i>.772</i>	<i>.783</i>	<i>.793</i>	<i>.801</i>	<i>.809</i>	<i>.816</i>	<i>.827</i>	<i>.837</i>	<i>.853</i>	<i>.866</i>	<i>.879</i>	<i>.892</i>	<i>.905</i>
34	.115	.156	.183	.203	.218	.232	.244	.254	.262	.270	.277	.282	.292	.304	.317	.329			
35	<i>.814</i>	<i>.868</i>	<i>.892</i>	<i>.916</i>	<i>.933</i>	<i>.948</i>	<i>.961</i>	<i>.971</i>	<i>.980</i>	<i>.989</i>	<i>.996</i>	<i>.1.004</i>	<i>.1.016</i>	<i>.1.025</i>	<i>.1.041</i>	<i>.1.054</i>	<i>.1.067</i>	<i>.1.080</i>	<i>.1.093</i>
36	.111	.150	.176	.195	.211	.224	.236	.245	.253	.262	.268	.273	.285	.293	.308	.320	.334		
37	<i>.707</i>	<i>.753</i>	<i>.773</i>	<i>.796</i>	<i>.813</i>	<i>.827</i>	<i>.840</i>	<i>.850</i>	<i>.859</i>	<i>.867</i>	<i>.875</i>	<i>.882</i>	<i>.894</i>	<i>.903</i>	<i>.919</i>	<i>.931</i>	<i>.944</i>	<i>.957</i>	<i>.970</i>
38	.107	.145	.170	.188	.204	.216	.228	.237	.246	.253	.260	.266	.278	.286	.301	.313	.325		
39	<i>.599</i>	<i>.644</i>	<i>.664</i>	<i>.686</i>	<i>.703</i>	<i>.718</i>	<i>.731</i>	<i>.742</i>	<i>.752</i>	<i>.761</i>	<i>.769</i>	<i>.776</i>	<i>.788</i>	<i>.797</i>	<i>.813</i>	<i>.825</i>	<i>.838</i>	<i>.851</i>	<i>.864</i>
40	.103	.139	.164	.182	.197	.210	.220	.229	.238	.245	.251	.258	.269	.277	.292	.304	.317		
41	<i>.583</i>	<i>.628</i>	<i>.648</i>	<i>.670</i>	<i>.687</i>	<i>.702</i>	<i>.715</i>	<i>.726</i>	<i>.736</i>	<i>.745</i>	<i>.753</i>	<i>.760</i>	<i>.772</i>	<i>.781</i>	<i>.797</i>	<i>.809</i>	<i>.822</i>	<i>.835</i>	<i>.848</i>

337

TABLE F.—VALUES OF r^2 AT THE 5% (ROMAN TYPE) AND 1% (ITALIC TYPE) LEVELS OF SIGNIFICANCE—Continued

E - 1

<i>N</i> - <i>k</i>	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50
28	.099	.135	.159	.176	.191	.203	.214	.223	.231	.238	.245	.251	.261	.271	.286	.296	.310		
	<i>.186</i>	<i>.229</i>	<i>.257</i>	<i>.277</i>	<i>.295</i>	<i>.309</i>	<i>.321</i>	<i>.331</i>	<i>.339</i>	<i>.345</i>	<i>.355</i>	<i>.363</i>	<i>.375</i>	<i>.383</i>	<i>.400</i>	<i>.419</i>	<i>.427</i>		
29	.096	.131	.153	.171	.185	.197	.208	.217	.224	.232	.239	.244	.255	.262	.277	.290	.302		
	<i>.180</i>	<i>.222</i>	<i>.249</i>	<i>.269</i>	<i>.286</i>	<i>.300</i>	<i>.312</i>	<i>.322</i>	<i>.330</i>	<i>.338</i>	<i>.346</i>	<i>.354</i>	<i>.366</i>	<i>.374</i>	<i>.391</i>	<i>.408</i>	<i>.418</i>		
30	.093	.127	.149	.166	.179	.191	.202	.211	.218	.225	.231	.237	.249	.256	.271	.283	.296		
	<i>.175</i>	<i>.215</i>	<i>.242</i>	<i>.262</i>	<i>.278</i>	<i>.292</i>	<i>.303</i>	<i>.314</i>	<i>.322</i>	<i>.331</i>	<i>.338</i>	<i>.345</i>	<i>.356</i>	<i>.366</i>	<i>.383</i>	<i>.395</i>	<i>.408</i>		
32	.087	.119	.140	.157	.169	.181	.192	.200	.207	.213	.220	.226	.237	.244	.259	.269	.284	.297	
	<i>.165</i>	<i>.203</i>	<i>.229</i>	<i>.248</i>	<i>.264</i>	<i>.276</i>	<i>.288</i>	<i>.298</i>	<i>.306</i>	<i>.316</i>	<i>.322</i>	<i>.329</i>	<i>.341</i>	<i>.351</i>	<i>.367</i>	<i>.378</i>	<i>.393</i>	<i>.410</i>	
34	.082	.112	.132	.148	.160	.171	.182	.190	.197	.203	.209	.215	.226	.233	.248	.258	.273	.286	
	<i>.155</i>	<i>.192</i>	<i>.217</i>	<i>.236</i>	<i>.251</i>	<i>.263</i>	<i>.274</i>	<i>.284</i>	<i>.292</i>	<i>.300</i>	<i>.308</i>	<i>.315</i>	<i>.326</i>	<i>.336</i>	<i>.353</i>	<i>.363</i>	<i>.378</i>	<i>.395</i>	
36	.078	.106	.125	.140	.152	.163	.172	.180	.187	.193	.199	.205	.215	.222	.237	.247	.262	.275	.286
	<i>.147</i>	<i>.183</i>	<i>.208</i>	<i>.224</i>	<i>.239</i>	<i>.251</i>	<i>.262</i>	<i>.271</i>	<i>.280</i>	<i>.288</i>	<i>.294</i>	<i>.301</i>	<i>.312</i>	<i>.322</i>	<i>.338</i>	<i>.351</i>	<i>.364</i>	<i>.381</i>	<i>.394</i>
38	.074	.101	.119	.134	.145	.155	.164	.171	.179	.185	.191	.197	.205	.214	.227	.236	.251	.267	.276
	<i>.140</i>	<i>.174</i>	<i>.198</i>	<i>.214</i>	<i>.228</i>	<i>.240</i>	<i>.251</i>	<i>.260</i>	<i>.268</i>	<i>.275</i>	<i>.282</i>	<i>.289</i>	<i>.300</i>	<i>.309</i>	<i>.326</i>	<i>.338</i>	<i>.350</i>	<i>.369</i>	<i>.380</i>
40	.069	.096	.114	.128	.139	.149	.157	.164	.171	.176	.183	.188	.198	.205	.219	.229	.241	.257	.268
	<i>.133</i>	<i>.168</i>	<i>.188</i>	<i>.205</i>	<i>.218</i>	<i>.230</i>	<i>.240</i>	<i>.249</i>	<i>.257</i>	<i>.265</i>	<i>.272</i>	<i>.278</i>	<i>.288</i>	<i>.299</i>	<i>.314</i>	<i>.322</i>	<i>.340</i>	<i>.357</i>	<i>.368</i>
42	.067	.092	.109	.121	.133	.142	.150	.158	.164	.169	.175	.180	.190	.197	.209	.221	.233	.249	.258
	<i>.127</i>	<i>.159</i>	<i>.180</i>	<i>.196</i>	<i>.209</i>	<i>.220</i>	<i>.231</i>	<i>.239</i>	<i>.247</i>	<i>.254</i>	<i>.261</i>	<i>.267</i>	<i>.278</i>	<i>.287</i>	<i>.303</i>	<i>.314</i>	<i>.328</i>	<i>.345</i>	<i>.357</i>
44	.064	.088	.104	.116	.127	.136	.144	.151	.157	.163	.168	.174	.182	.190	.202	.211	.226	.239	.251
	<i>.122</i>	<i>.152</i>	<i>.172</i>	<i>.188</i>	<i>.201</i>	<i>.212</i>	<i>.221</i>	<i>.230</i>	<i>.238</i>	<i>.245</i>	<i>.251</i>	<i>.258</i>	<i>.268</i>	<i>.277</i>	<i>.292</i>	<i>.304</i>	<i>.318</i>	<i>.335</i>	<i>.347</i>
46	.061	.084	.100	.112	.122	.130	.139	.144	.151	.157	.162	.167	.175	.183	.195	.204	.219	.232	.244
	<i>.117</i>	<i>.148</i>	<i>.168</i>	<i>.181</i>	<i>.193</i>	<i>.204</i>	<i>.213</i>	<i>.221</i>	<i>.229</i>	<i>.236</i>	<i>.243</i>	<i>.249</i>	<i>.259</i>	<i>.268</i>	<i>.283</i>	<i>.295</i>	<i>.308</i>	<i>.326</i>	<i>.338</i>
48	.058	.081	.097	.107	.117	.126	.133	.140	.146	.151	.156	.161	.169	.177	.189	.198	.212	.225	.237
	<i>.112</i>	<i>.140</i>	<i>.159</i>	<i>.174</i>	<i>.186</i>	<i>.196</i>	<i>.206</i>	<i>.213</i>	<i>.221</i>	<i>.228</i>	<i>.234</i>	<i>.240</i>	<i>.250</i>	<i>.259</i>	<i>.274</i>	<i>.286</i>	<i>.299</i>	<i>.317</i>	<i>.329</i>

338

TABLE F.—VALUES OF ϵ^2 AT THE 5% (ROMAN TYPE) AND 1% (ITALIC TYPE) LEVELS OF SIGNIFICANCE—Continued

$k - 1$

$N - k$	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50
50	.056	.077	.092	.104	.113	.121	.128	.135	.140	.145	.150	.155	.164	.171	.182	.194	.206	.219	.231
55	.108	.135	.153	.168	.180	.189	.199	.208	.214	.221	.226	.232	.242	.252	.265	.277	.292	.308	.320
60	.051	.071	.084	.095	.103	.111	.118	.124	.129	.133	.139	.143	.152	.158	.169	.179	.191	.204	.216
65	.099	.123	.140	.154	.165	.175	.183	.190	.197	.203	.209	.215	.225	.233	.247	.259	.272	.288	.300
70	.047	.065	.077	.087	.095	.102	.109	.115	.119	.124	.128	.133	.140	.146	.158	.167	.178	.191	.203
75	.091	.114	.130	.142	.153	.162	.169	.178	.183	.189	.195	.200	.209	.217	.231	.242	.256	.271	.283
80	.043	.060	.072	.080	.089	.095	.101	.106	.110	.116	.120	.123	.131	.136	.147	.155	.166	.178	.190
85	.084	.105	.120	.132	.142	.150	.158	.164	.171	.177	.182	.186	.195	.204	.217	.227	.240	.255	.268
90	.040	.056	.067	.075	.083	.089	.094	.099	.103	.108	.112	.115	.123	.128	.138	.146	.157	.169	.181
95	.078	.098	.112	.122	.132	.140	.148	.154	.160	.166	.170	.175	.184	.190	.204	.215	.227	.242	.255
100	.035	.049	.059	.066	.073	.078	.083	.087	.091	.095	.099	.103	.109	.114	.123	.130	.141	.153	.164
105	.069	.086	.099	.109	.117	.125	.131	.137	.142	.147	.152	.156	.164	.171	.182	.192	.204	.219	.231
110	.028	.039	.047	.053	.058	.063	.067	.071	.074	.077	.080	.083	.088	.094	.102	.109	.116	.127	.138
115	.055	.070	.080	.088	.095	.101	.106	.111	.116	.121	.124	.127	.134	.141	.150	.159	.170	.184	.196
120	.023	.032	.038	.043	.047	.051	.054	.057	.060	.063	.065	.068	.072	.076	.082	.088	.096	.106	.114
125	.044	.056	.064	.071	.077	.082	.087	.090	.095	.098	.102	.104	.110	.115	.124	.131	.141	.154	.163
130	.019	.026	.032	.036	.039	.043	.046	.048	.051	.053	.055	.057	.061	.064	.070	.075	.083	.090	.099
135	.037	.047	.054	.060	.065	.069	.073	.076	.080	.083	.086	.088	.093	.097	.105	.112	.122	.132	.142
140	.014	.020	.024	.027	.030	.032	.034	.036	.038	.040	.041	.043	.046	.049	.053	.058	.064	.070	.077
145	.022	.028	.031	.035	.039	.042	.045	.048	.051	.053	.055	.058	.062	.065	.071	.075	.081	.088	.093
150	.007	.010	.012	.014	.015	.016	.017	.018	.019	.020	.021	.022	.024	.025	.028	.030	.033	.037	.041
155	.014	.018	.021	.023	.025	.027	.028	.029	.031	.032	.033	.035	.038	.039	.042	.045	.049	.055	.060
160	.003	.004	.005	.005	.006	.007	.007	.008	.008	.008	.009	.009	.010	.010	.011	.012	.014	.016	.017
1,000	.006	.007	.008	.009	.010	.011	.011	.012	.012	.013	.013	.014	.014	.015	.016	.017	.019	.022	.025

TABLE G.—A TABLE OF RANDOM NUMBERS*

Row	COLUMN NUMBER															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	2	7	8	9	4	0	7	2	3	2	5	4	2	6	7	1
2	2	2	6	0	4	1	7	7	3	8	7	3	6	7	9	4
3	9	1	6	0	3	9	7	9	1	8	5	1	5	2	2	7
4	7	0	5	5	9	2	7	5	7	0	0	8	8	5	0	6
5	4	7	3	6	6	3	9	8	2	1	7	9	7	6	4	2
6	8	2	0	2	8	7	7	6	0	2	2	3	1	1	1	6
7	0	8	7	5	3	3	6	4	2	6	8	3	1	6	5	0
8	9	4	1	9	0	8	4	6	6	8	6	3	3	2	2	3
9	5	0	0	6	7	4	0	0	0	1	9	5	9	9	1	8
10	1	9	5	4	1	5	2	6	2	9	4	1	5	5	8	4
11	5	6	4	4	1	8	7	2	8	3	6	6	1	9	8	6
12	7	9	2	5	1	9	7	9	3	1	8	6	3	7	7	6
13	3	3	3	5	9	5	1	4	0	8	2	5	5	3	5	4
14	1	9	0	4	0	0	9	9	5	1	4	6	5	9	4	7
15	5	4	4	7	2	0	3	7	9	1	0	9	6	2	9	7
16	2	9	8	2	5	5	9	3	2	0	4	9	0	6	4	4
17	9	7	6	2	6	7	7	3	3	3	1	7	5	0	9	5
18	5	8	2	4	3	3	0	8	5	3	5	7	5	8	3	6
19	4	3	4	9	5	0	3	6	2	9	7	4	6	2	5	6
20	1	1	9	8	4	8	0	6	7	0	9	7	9	6	9	9
21	6	9	1	8	3	3	7	5	9	6	6	7	9	6	0	4
22	7	0	0	3	8	1	3	4	7	9	5	2	6	9	9	7
23	3	7	7	2	8	1	6	6	9	0	6	7	3	6	6	3
24	2	7	0	0	0	6	5	0	6	5	6	0	8	2	9	9
25	3	0	7	0	7	8	4	9	4	2	8	2	4	7	4	9
26	6	2	9	3	3	1	7	7	5	2	2	3	4	6	4	2
27	5	4	9	3	1	4	8	5	7	0	9	6	4	7	2	1
28	0	3	7	0	1	7	3	8	0	3	6	2	3	1	0	9
29	9	3	6	6	2	2	3	9	7	3	3	9	2	8	7	3
30	2	9	5	6	9	9	5	6	9	3	2	8	0	0	4	4
31	8	5	7	2	9	2	6	5	9	3	9	7	1	8	3	5
32	8	4	5	7	7	9	9	5	1	4	5	5	0	9	5	3
33	8	7	9	8	1	8	4	1	4	3	7	7	0	9	1	3
34	7	3	2	5	1	8	6	3	2	8	5	8	6	9	3	4
35	8	9	9	0	1	8	8	8	9	5	7	5	0	4	1	1
36	0	2	9	7	8	8	1	7	6	1	6	7	6	4	2	5
37	0	5	2	3	2	3	8	1	8	8	1	6	2	3	0	7
38	2	2	6	8	1	6	9	6	2	6	7	9	1	7	8	0
39	0	7	8	4	9	5	8	8	0	7	2	1	8	1	0	5
40	4	8	0	7	0	5	9	9	4	9	6	9	8	2	0	6
41	9	2	0	1	6	7	2	8	3	9	8	3	0	4	7	8
42	0	8	8	3	4	0	9	2	2	8	1	5	3	4	4	2
43	2	0	6	9	7	5	2	8	2	5	5	4	4	7	7	1
44	3	1	8	6	8	3	3	5	6	2	7	4	1	8	9	4
45	0	0	8	6	1	7	5	0	8	5	6	5	0	8	2	7
46	3	3	2	9	4	2	5	3	3	8	2	4	2	6	2	5
47	8	4	7	4	0	4	5	1	2	1	0	4	2	5	7	7
48	0	2	4	3	0	2	0	7	2	8	8	0	8	4	1	6
49	4	6	5	6	3	0	4	5	2	0	1	5	2	7	9	5
50	3	4	8	3	4	5	8	7	5	9	7	1	6	3	9	9

* From J. G. Peatman and R. Schafer. A table of random numbers from Selective Service numbers. *J. Psychol.*, 1942, 14, pp. 296-297, by permission of the editor and publisher.

TABLE G.—A TABLE OF RANDOM NUMBERS—Continued

COLUMN NUMBER													Row			
17	18	19	20	21	22	23	24	25	26	27	28	29		30	31	32
6	8	5	9	1	3	5	4	0	3	6	6	7	6	5	1	1
2	1	3	8	9	0	3	4	9	0	2	6	3	0	9	8	2
5	2	5	3	4	1	3	9	5	8	1	3	8	2	9	2	3
0	5	9	0	5	7	4	5	2	0	6	1	6	4	2	0	4
4	9	6	0	3	6	3	5	3	9	9	1	8	5	1	3	5
4	8	5	2	2	3	4	2	2	6	5	2	2	4	9	6	6
0	5	5	7	8	1	0	1	2	9	1	4	3	4	7	6	7
7	4	7	5	1	5	7	6	3	7	9	4	5	4	3	5	8
1	4	7	4	9	8	7	2	4	3	0	8	6	4	2	7	9
4	4	6	1	8	7	8	6	4	8	7	4	4	0	5	8	10
2	2	9	1	9	0	4	8	1	0	1	3	5	3	4	4	11
6	5	0	3	3	1	1	2	4	7	8	9	1	7	5	2	12
6	5	7	2	6	7	8	9	9	9	8	0	9	1	5	3	13
6	4	8	2	6	4	4	1	8	8	1	5	4	3	8	0	14
4	4	7	6	1	6	1	2	2	9	5	8	4	4	3	6	15
2	1	5	7	7	3	6	5	5	4	5	7	9	6	6	4	16
1	3	3	9	2	4	1	1	0	0	1	3	7	7	3	3	17
9	3	4	5	4	6	3	9	2	7	1	1	4	9	1	3	18
9	8	3	6	0	4	0	3	5	9	7	1	8	0	6	9	19
4	0	6	0	1	5	9	6	5	1	4	2	0	4	1	9	20
5	3	4	5	7	3	0	6	1	0	3	0	0	3	5	0	21
3	2	5	0	2	3	5	3	9	7	4	8	9	4	5	1	22
6	0	7	8	1	9	6	7	4	8	9	6	3	6	5	1	23
1	7	2	2	8	4	9	0	4	3	2	4	5	5	1	2	24
6	0	4	3	8	1	7	7	0	9	8	4	6	6	3	1	25
2	4	7	5	4	4	4	1	7	1	6	7	1	2	6	8	26
8	7	7	6	1	3	3	4	6	6	5	9	0	0	7	3	27
5	9	2	5	9	2	0	2	8	7	7	2	0	2	7	2	28
5	1	0	7	0	8	9	3	8	8	5	3	1	3	1	0	29
8	8	5	7	2	1	3	4	9	5	2	6	8	3	6	6	30
6	6	1	2	1	5	5	5	6	1	7	1	5	7	5	9	31
1	3	9	3	7	8	1	4	0	5	4	1	5	4	4	0	32
4	6	1	3	8	6	5	9	2	2	8	1	6	9	0	1	33
5	6	2	6	1	9	0	6	9	0	5	4	6	8	0	3	34
0	0	3	1	3	0	3	5	8	9	2	7	8	8	1	1	35
0	5	8	3	2	4	7	7	2	2	6	2	6	8	6	0	36
3	0	4	2	6	4	6	8	3	7	4	4	3	8	9	9	37
2	4	8	0	4	7	3	3	8	4	4	8	4	3	3	8	38
3	4	0	7	4	1	0	3	2	0	1	2	8	6	5	9	39
4	0	7	8	1	1	4	2	1	6	7	0	7	3	1	2	40
4	0	5	1	6	8	7	8	3	5	4	5	0	4	0	6	41
6	7	9	2	8	9	1	8	5	3	1	0	7	8	3	9	42
7	2	8	6	5	9	1	3	7	8	2	7	1	9	3	6	43
5	6	8	0	6	4	6	4	1	0	9	1	9	8	1	4	44
1	1	6	6	4	6	0	0	9	4	7	9	2	4	8	7	45
2	9	0	1	3	7	6	5	9	1	4	6	0	1	0	0	46
9	4	6	5	8	3	3	8	1	0	3	7	7	7	8	6	47
0	2	3	5	9	7	5	1	3	6	3	2	8	7	5	8	48
3	0	2	2	1	6	1	1	0	0	9	1	6	1	7	7	49
0	9	4	2	5	8	9	5	3	3	3	6	4	5	2	0	50

TABLE H.—TABLE OF χ^2 *

5% 1% 342

DEGREES OF FREEDOM n	P = .99	.98	.95	.90	.80	.70	.50	.30	.20	.10	.05	.02	.01
1	.000157	.000628	.00393	.0158	.0642	.148	.455	1.074	1.642	2.706	3.841	5.412	6.635
2	.0201	.0404	.103	.211	.446	.713	1.386	2.408	3.219	4.605	5.991	7.824	9.210
3	.115	.185	.352	.584	1.005	1.424	2.366	3.665	4.642	6.251	7.815	9.837	11.341
4	.297	.429	.711	1.064	1.649	2.195	3.357	4.878	5.989	7.779	9.488	11.668	13.277
5	.554	.752	1.145	1.610	2.343	3.000	4.351	6.064	7.289	9.236	11.070	13.388	15.086
6	.872	1.134	1.635	2.204	3.070	3.828	5.348	7.231	8.558	10.645	12.592	15.033	16.812
7	1.239	1.564	2.167	2.833	3.822	4.671	6.346	8.383	9.803	12.017	14.067	16.622	18.475
8	1.646	2.032	2.733	3.490	4.594	5.527	7.344	9.524	11.030	13.362	15.507	18.168	20.090
9	2.088	2.532	3.325	4.168	5.380	6.393	8.343	10.656	12.242	14.684	16.919	19.679	21.666
10	2.558	3.059	3.940	4.865	6.179	7.267	9.342	11.781	13.442	15.987	18.307	21.161	23.209
11	3.053	3.609	4.575	5.578	6.989	8.148	10.341	12.899	14.631	17.275	19.675	22.618	24.725
12	3.571	4.178	5.226	6.304	7.807	9.034	11.340	14.011	15.812	18.549	21.026	24.054	26.217
13	4.107	4.765	5.892	7.042	8.634	9.926	12.340	15.119	16.985	19.812	22.362	25.472	27.688
14	4.660	5.368	6.571	7.790	9.467	10.821	13.339	16.222	18.161	21.064	23.685	26.873	29.141
15	5.229	5.985	7.261	8.547	10.307	11.721	14.339	17.322	19.311	22.307	24.996	28.259	30.578
16	5.812	6.614	7.962	9.312	11.152	12.624	15.338	18.418	20.465	23.542	26.296	29.633	32.000
17	6.408	7.255	8.672	10.085	12.002	13.531	16.338	19.511	21.615	24.769	27.587	30.995	33.409
18	7.015	7.906	9.390	10.865	12.857	14.440	17.338	20.601	22.760	25.989	28.669	32.346	34.805
19	7.633	8.567	10.117	11.651	13.716	15.352	18.338	21.689	23.900	27.204	30.144	33.687	36.191
20	8.260	9.237	10.851	12.443	14.578	16.266	19.337	22.775	25.038	28.412	31.410	35.020	37.566
21	8.897	9.915	11.591	13.240	15.445	17.182	20.337	23.858	26.171	29.615	32.671	36.343	38.932
22	9.542	10.600	12.338	14.041	16.314	18.101	21.337	24.939	27.301	30.813	33.924	37.659	40.289
23	10.196	11.293	13.091	14.848	17.187	19.021	22.337	26.018	28.429	32.007	35.172	38.968	41.638
24	10.856	11.992	13.848	15.659	18.062	19.943	23.337	27.096	29.553	33.106	36.415	40.270	42.980
25	11.524	12.697	14.611	16.473	18.940	20.867	24.337	28.172	30.675	34.382	37.652	41.566	44.314
26	12.198	13.409	15.379	17.292	19.820	21.792	25.336	29.246	31.795	35.563	38.885	42.856	45.642
27	12.879	14.125	16.151	18.114	20.703	22.719	26.336	30.319	32.912	36.741	40.113	44.140	46.963
28	13.565	14.847	16.928	18.939	21.588	23.647	27.336	31.391	34.027	37.916	41.337	45.419	48.278
29	14.256	15.574	17.708	19.768	22.475	24.577	28.336	32.461	35.139	39.087	42.557	46.693	49.588
30	14.953	16.306	18.493	20.599	23.364	25.508	29.336	33.530	36.260	40.266	43.773	47.962	50.892

* Table H is reprinted from Table III of Fisher: *Statistical Methods for Research Workers*. Oliver & Boyd, Ltd., Edinburgh, by permission of the Author and Publishers. For larger values of n, the expression $\sqrt{2\chi^2} - \sqrt{2n - 1}$ may be used as a normal deviate with unit standard error.

Appendix

ANSWERS TO EXAMPLES

Chapter 2

- | | | | |
|------------|--|---|---|
| Example 1. | (a) -11
(b) 2
(c) -1
(d) 10
(e) 2 | (f) -4
(g) -3
(h) -8
(i) -16
(j) -6 | |
| Example 2. | (a) -5
(b) 6
(c) 17
(d) 30
(e) 18 | (f) -8
(g) -9
(h) -10
(i) 16
(j) -2 | |
| Example 3. | (a) 2
(b) 2
(c) 2
(d) 2
(e) 1
(f) 1 | (g) 1
(h) 1
(i) 2
(j) 1
(k) 1
(l) 2 | |
| Example 4. | (a) 49
(b) 3
(c) 25
(d) 36
(e) 1
(f) 16
(g) -4 | (h) -4
(i) -3
(j) 24
(k) -10
(l) -6
(m) 6
(n) 0 | (o) 2
(p) 20
(q) .2
(r) 40
(s) .6
(t) .0004
(u) .01 |
| Example 5. | (a) 25 per cent
(b) .75 | | |
| Example 6. | (a) 75 per cent
(b) .25 | | |
| Example 7. | (a) 30 per cent
(b) .70 | | |

Example 8.	(a) .2	(f) 144	(k) 2.44	(p) 197
	(b) .9	(g) 32	(l) 3.17	(q) 314
	(c) .04	(h) 66	(m) 7.19	(r) 397
	(d) .005	(i) 99	(n) 9.98	(s) 174
	(e) .68	(j) 82	(o) 6.1	(t) 983

Chapter 3

Example 1. (a) $M = 24$ (b) $\Sigma x = 0$

Example 2. (a) $Mdn = 16.9$ (d) $Mdn = 16.5$
 (b) $Mdn = 17.0$ (e) $Mdn = 31.5$
 (c) $Mdn = 9.0$

Example 3. Group 1: $M = 7$; $\Sigma x = 0$
 Group 2: $M = 5$; $\Sigma x = 0$
 Diff.: $M_d = 2$; $\Sigma d = 0$

Example 4. Group 1: $Mdn = 6.5$; $R = 5$; $v = 3.0$; $\sigma = 1.73$
 Group 2: $Mdn = 5.0$; $R = 6$; $v = 3.6$; $\sigma = 1.90$

Example 5. $M = 20$; $v = 10.1951$; $\sigma = 3.19$

Example 6. $Mdn = 20.12$; $Q_1 = 17.75$; $Q_3 = 22.42$

Example 7. (a) x
 (b) NM
 (c) $\Sigma(X - M)^2$; $N\sigma^2$
 (d) M
 (e) $\frac{\Sigma X}{N}$
 (f) x^2
 (g) $\frac{\Sigma x^2}{N}$; σ^2 ; $\frac{\Sigma(X - M)^2}{N}$
 (h) v ; σ^2 ; $\frac{\Sigma(X - M)^2}{N}$
 (i) $\frac{\Sigma x^2}{N}$; v ; σ^2
 (j) ΣX

- (k) Σx^2 ; $\Sigma(X - M)^2$
 (l) σ ; \sqrt{v} ; $\sqrt{\frac{\Sigma(X - M)^2}{N}}$
 (m) $X - M$

Example 8. $x = X - M$
 $\Sigma x = \Sigma(X - M)$
 $\Sigma x = \Sigma X - \Sigma M$
 $\Sigma x = \Sigma X - NM$
 $\Sigma x = \Sigma X - \Sigma X$
 $\Sigma x = 0$

Chapter 4

Example 1. (a) $\Sigma X = 175$; $M = 25$; $\Sigma x^2 = 64$

(b) $M = 22 + \left(\frac{21}{7}\right) = 25$

$\Sigma x^2 = 127 - \frac{(21)^2}{7} = 64$

(c) $\Sigma x^2 = 4,439 - \frac{(175)^2}{7} = 64$

Example 2.

	(a)	(b)	(c)	(d)
(1)	65	5	20-24	22
(2)	45	3 or 5	48-50 or 50-54	49 or 52
(3)	123	10	60-69	64.5
(4)	23	2	42-43	42.5
(5)	36	3	15-17	16
(6)	56	5	50-54	52

Example 3.

(a)	(b)	(c)	(d)	(e)	(f)	(g)
X'	X'	X'	X'	X'	X'	X'
10	3	4	1	2	5	2
0	6	6	7	3	10	5
10	7	5	5	1	1	6
9	8	1	10	3	6	9
2	1	9	8	2	5	4

Example 4. (a) $65 = 45 + 20$
 (b) $243 = 133 + (2)(45) + 20$

Example 5. $M = 22.17$; $\sigma = 7.41$
 Check: $\Sigma fx'' = 1,090 = 910 + 180$
 Check: $\Sigma fx''^2 = 7,700 = 5,700 + (2)(910) + 180$

Example 6. $M = 46.79$; $Mdn = 46.75$; $\sigma = 5.83$

Example 7. $x = X - M$
 $x^2 = (X - M)^2$
 $x^2 = X^2 - 2XM + M^2$
 $\Sigma x^2 = \Sigma X^2 - 2M\Sigma X + NM^2$
 $\Sigma x^2 = \Sigma X^2 - 2MN\bar{M} + NM^2$
 $\Sigma x^2 = \Sigma X^2 - 2NM\bar{M} + NM^2$
 $\Sigma x^2 = \Sigma X^2 - NM^2$
 $\Sigma x^2 = \Sigma X^2 - N\left(\frac{\Sigma X}{N}\right)\left(\frac{\Sigma X}{N}\right)$
 $\Sigma x^2 = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$

Example 8. (a) $x'' = x' + 1$
 $\Sigma x'' = \Sigma x' + N$
 (b) $x'' = x' + 1$
 $x''^2 = (x' + 1)^2$
 $x''^2 = x'^2 + 2x' + 1$
 $\Sigma x''^2 = \Sigma x'^2 + (2)(\Sigma x') + N$

Chapter 5

Example 1. $r = \frac{18}{(10)(1.414)(1.414)} = .90$

Example 2. $r = \frac{18}{\sqrt{(20)(20)}} = .90$

Example 3. $r = \frac{138 - \frac{(40)(30)}{10}}{\sqrt{\left[180 - \frac{(40)^2}{10}\right]\left[110 - \frac{(30)^2}{10}\right]}} = .90$

$$\begin{aligned} \text{Example 4. } xy &= (X - M_x)(Y - M_y) \\ xy &= XY - YM_x - XM_y + M_x M_y \\ \Sigma xy &= \Sigma XY - M_x \Sigma Y - M_y \Sigma X + NM_x M_y \\ \Sigma xy &= \Sigma XY - M_x NM_y - M_y NM_x + NM_x M_y \\ \Sigma xy &= \Sigma XY - 2NM_x M_y + NM_x M_y \\ \Sigma xy &= \Sigma XY - NM_x M_y \\ \Sigma xy &= \Sigma XY - N \left(\frac{\Sigma X}{N} \right) \left(\frac{\Sigma Y}{N} \right) \\ \Sigma xy &= \Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{N} \end{aligned}$$

Example 5. $r = .82$

Example 6. $r = .12$

Example 7. $r = .73$

Chapter 6

Example 1. $\eta_{y \cdot x} = .82$

Example 2. $r_{b \cdot i \cdot s} = .20$

Example 3. $r_{p \cdot b \cdot i \cdot s} = .24$

Example 4. $\phi = .15$

Example 5. $\phi = .37; \quad r_\phi = .59$

Example 6. $\phi = .22; \quad r_\phi = .35$

Example 7. $\rho = .27$

Example 8. $\rho_{BG} = .84; \quad \rho_{BM} = .91$

Chapter 7

Example 1. (a) $1/256$ (b) $37/256$

Example 2. $3/4$

Example 3. $1/4$

Example 4. 3/64

Example 5. (a) .0668 (g) .9938
 (b) .3085 (h) .8413
 (c) .8944 (i) .8185
 (d) .9938 (j) .6826
 (e) .1587 (k) .9104
 (f) .6915 (l) .7745

Example 6. Middle 95 per cent.

(a) 25 ± 9.80 (e) 18.5 ± 4.31
 (b) 30 ± 13.72 (f) 22.4 ± 8.62
 (c) 50 ± 11.76 (g) 47.0 ± 10.78
 (d) 42 ± 8.82 (h) 23.1 ± 4.12

Middle 99 per cent

(a) 25 ± 12.90 (e) 18.5 ± 5.68
 (b) 30 ± 18.06 (f) 22.4 ± 11.35
 (c) 50 ± 15.48 (g) 47.0 ± 14.19
 (d) 42 ± 11.61 (h) 23.1 ± 5.42

Chapter 8

Example 2.	Fiducial Limits at 5% Level	Fiducial Limits at 1% Level
(a)	23.75—26.25	(a) 23.34—26.66
(b)	26.29—33.71	(b) 24.89—35.11
(c)	33.02—36.98	(c) 32.37—37.63
(d)	21.16—28.84	(d) 19.48—30.52
(e)	20.88—29.12	(e) 19.42—30.58
(f)	43.64—56.36	(f) 41.24—58.76
(g)	47.85—52.15	(g) 47.13—52.87
(h)	22.01—22.99	(h) 21.85—23.15

Example 3. $M = 20$; $\sigma_m = .718$; $df = 16$
 Fiducial limits at 5% level: 18.48—21.52
 Fiducial limits at 1% level: 17.90—22.10

Chapter 9

Example 1. $t = 4.53$; $df = 398$

Example 2. (a) $t = 2.35$; $df = 18$
 (b) $t = 6.71$; $df = 9$

Example 3. Paris-Hamburg: $t = .40$; $df = 198$
 Paris-Rome: $t = 1.15$; $df = 198$
 Hamburg-Rome: $t = .74$; $df = 198$

Example 4. $t = 2.60$; $df = 99$

Example 5. $t = 1.43$; $df = 99$

Example 6. $t = 2.86$; $df = 49$

Example 7. $t = 2.56$; $df = 29$

Example 8. $t = 2.00$; $df = 49$

Example 9. $r = .765$ would be significant at 1% level for 8 df

Example 10. r must be at least .632 to be significant at 5% level
 for 8 df

Example 11. r must be at least .28 to be significant at 5% level
 for 48 df

Example 12. N must be at least 62 in order for $r = .25$ to be
 significant at 5% level
 N must be at least 13 in order for $r = .55$ to be
 significant at 5% level

Example 13. $t = 3.04$; $df = 117$

Chapter 10

Example 1. (a) $t = 2.99$; $df = 18$
 (b) $F = 8.95$; $df = 1$ and 18

Example 2. $F = 1.10$; $df = 4$ and 20

Example 3. Between the 9 groups:

$$F = 13.16; \quad df = 8 \text{ and } 81$$

Between the age groups:

$$F = 1.35; \quad df = 2 \text{ and } 81$$

Between the methods groups:

$$F = 5.40; \quad df = 2 \text{ and } 81$$

Interaction:

$$F = 22.94; \quad df = 4 \text{ and } 81$$

Chapter 11

- Example 1. (a) $t = 2.34; \quad df = 18$
 $F = 5.45; \quad df = 1 \text{ and } 18$
 (b) $t = 6.71; \quad df = 9$
 $F = 45.05; \quad df = 1 \text{ and } 9$

- Example 2. (a) $F = 1.98; \quad df = 2 \text{ and } 27$
 (b) $F = 22.52; \quad df = 2 \text{ and } 18$

- Example 3. $\eta_{xx}^2 = .7174; \quad \epsilon_{xx}^2 = .6856; \quad k - 1 = 9;$
 $N - k = 80$

Chapter 12

- Example 1. $\chi^2 = 12.00; \quad df = 1$
 Example 2. $\chi^2 = 4.16; \quad df = 1$
 Example 3. $\chi^2 = 5.00; \quad df = 1$
 Example 4. $\chi^2 = 14.00; \quad df = 2$
 Example 5. $\chi^2 = 64.50; \quad df = 4$
 Example 6. $\chi^2 = 13.47; \quad df = 1$
 Example 7. $\chi^2 = 18.99; \quad df = 2$
 Example 8. $\chi^2 = 8.76; \quad df = 2$
 Example 9. $\chi^2 = 9.20; \quad df = 6$

Chapter 13

- Example 1. (a) Per cent correct without knowledge of
 response to question..... 50
 Per cent correct with knowledge of re-
 sponse to question..... 65

	Per cent increase with knowledge of response to question.....	30
(b)	Per cent correct without knowledge of marital status.....	55
	Per cent correct with knowledge of marital status.....	65
	Per cent increase with knowledge of marital status.....	18
Example 2.	Per cent correct without knowledge of student status.....	58
	Per cent correct with knowledge of student status.....	61
	Per cent increase with knowledge of student status.....	5
Example 3.	Per cent correct without knowledge of Humm-Wadsworth score.....	54
	Per cent correct with knowledge of Humm-Wadsworth score.....	71
	Per cent increase with knowledge of Humm-Wadsworth score.....	31
Example 4.	Per cent correct without knowledge of ACE classification.....	40
	Per cent correct with knowledge of ACE classification.....	50
	Per cent increase with knowledge of ACE classification.....	27
Example 5.	$M_x = 65.25; \sigma_x = 12.77; r_{xy} = .73$ $M_y = 65.90; \sigma_y = 10.04; r_{yx} = .73$	
(a)	$b_{xy} = .929$ $b_{yx} = .574$	
(b)	$X = 48; Y' = 56.00$ $X = 55; Y' = 60.02$ $X = 73; Y' = 70.35$ $X = 82; Y' = 75.52$ $X = 90; Y' = 80.11$	

$$(c) Y = 58; X' = 57.91$$

$$Y = 71; X' = 69.99$$

$$Y = 76; X' = 74.63$$

$$Y = 80; X' = 78.35$$

$$Y = 95; X' = 92.29$$

$$(d) \sigma_{x,y} = 8.72; E = 32\%$$

$$\sigma_{y,x} = 6.86; E = 32\%$$

$$\text{Example 6. } \sigma_{y,x} = 13.01; E = 43\%$$

AUTHOR INDEX

- Allport, G. W., 278, 300
 Anastasi, A., 9, 300
 Armitt, F. M., 279, 302
- Baker, K. H., 173, 300
 Bancroft, T. A., 280
 Barker, R., 12, 277, 300
 Baxter, B., 232, 300
 Blakemore, A. M., 129, 302
 Brownman, D. E., 277, 300
- Camp, B. H., 120, 121
 Cantril, H., 305
 Chave, E. J., 271, 305
 Chesire, L., 117, 300
 Clark, E. L., 9, 302
 Clopper, C. J., 170
 Cochran, W. G., 297
 Cox, G. M., 297
 Cureton, E. E., 193
 Curtis, J. W., 94, 300
- Dallenbach, K. M., 31, 303
 Dembo, T., 12, 277, 300
 Dollard, J., 79, 277, 300
 Doob, L. W., 79, 277, 300
 Dorcus, R. M., 274, 300
 Dunlap, J. W., 232, 300
- Edwards, A. L., 12, 277, 278, 300, 301
 Enlow, E. R., 301
 Ezekiel, M., 301
- Fisher, R. A., 165, 166, 186, 188, 201,
 202, 232, 280, 281, 289, 299,
 301, 330, 331, 340
 Fosdick, S. J., 124, 301
 Franklin, M., 191, 303
- Gallup, G. H., 152, 283, 301
 Galton, F., 262
 Garrett, H. E., 121, 122, 232, 302
 Garrett, W. S., 277, 302
 Gilliland, A. R., 9, 302
- Gossett, W. S., 166
 Goulden, C. H., 167, 232, 253, 299,
 302
 Grant, D. A., 232, 302
 Greene, E. B., 14, 102, 302
 Guilford, J. P., 14, 123, 126, 179, 253,
 302
 Guthrie, E. R., 278, 302
- Hartmann, G. W., 301
 Hay, E. N., 129, 302
 Holman, L. J., 302
 Horst, P., 258, 302
 Hull, C. L., 278, 302
 Hurlock, E. B., 277, 302
- Irwin, F. W., 279, 302
- Jackson, R. W. B., 166, 232, 302
 Jenkins, J. G., 31, 303
 Jones, M. C., 278, 303
- Kellar, B., 238, 303
 Kelley, T. L., 123, 232, 303
 Kendall, M. G., 123, 306
 Kenney, J. F., 134, 149, 153, 160,
 232, 253, 303
 Kenney, K. C., 301
 Klineberg, O., 190, 303
 Koffka, K., 278, 303
 Krathwohl, W. C., 275, 303
 Kuo, Z. Y., 256, 303
- Lewin, K., 12, 277, 278, 300, 303
 Lewis, H. B., 191, 303
 Likert, R., 103, 303
 Lindquist, E. F., 126, 130, 157, 186,
 193, 232, 253, 299, 303
 Lo, C. F., 131, 304
 Ludgate, K. E., 277, 304
 Lynd, R. S., 277, 304
- MacKinnon, D. W., 277, 304
 McNemar, Q., 180, 181, 279, 281,
 286, 287, 288, 293, 304

- Miller, N. E., 79, 277, 300
Mowrer, O. H., 79, 277, 300
Murray, H. A., 304
- Newcomb, T. M., 274, 301, 304
- Odell, C. W., 304
- Paterson, D. G., 277, 304
Pearson, E. S., 170
Pearson, K., 106, 120, 125, 127, 128
Peatman, J. G., 289, 304, 341
Peters, C. C., 4, 113; 126, 127, 130,
149, 157, 168, 232, 235, 236, 237,
252, 288, 293, 299, 304, 336
Peterson, R. C., 11, 278, 304
- Rae, S. F., 152, 301
Remmers, H. H., 303
Rethlingshafer, D., 278, 304
Richardson, M. W., 116
Rider, P. R., 253, 304
Rogers, C., 278, 304
Rosenthal, S. P., 278, 304
Rosenzweig, S., 257, 304
- Saffir, M., 117, 300
Schafer, R., 289, 304, 341
Sears, R. R., 79, 277, 278, 300, 304
Shaffer, L. F., 104, 305
Shen, E., 232, 305
Simon, C. W., 279, 302
Skeels, H. M., 278, 305
- Snedecor, G. W., 56, 101, 186, 187,
193, 222, 232, 244, 253, 284, 296,
299, 305, 330, 331, 332
Sorenson, H., 305
Stalnaker, J. M., 116
Stock, J. S., 50, 51, 286, 305
- Thurstone, L. L., 11, 58, 61, 117, 271,
278, 300, 304, 305
Tippett, L. H. C., 149, 165, 168, 199,
232, 253, 297, 299, 305
Tolman, E. C., 278, 305
Treloar, A. E., 305
- Upedegraff, R., 278, 305
- Van Voorhis, W. R., 4, 113, 126, 127,
130, 149, 157, 168, 235, 236, 237,
252, 288, 293, 299, 304, 336
- Walker, H. M., 166, 167, 179, 180,
181, 232, 281, 305
Watson, K. B., 189, 305
Wellman, B. L., 278, 305
Wilks, S. S., 180, 306
Williams, H. M., 278, 305
Wolfe, D., 277, 306
Wright, H. F., 279, 306
- Yates, F., 289, 301
Yule, G. U., 123, 306
- Zubin, J., 232, 302

SUBJECT INDEX

- Abcissa**, 81
- Analysis of variance**, 192-231
of independent groups, 192-214
of matched groups, 217-231
nature of, 192-199
relation to epsilon-square, 232-236
relation to *t* test in case of two independent groups, 199-201
relation to *t* test in case of two matched groups, 224-225
of several independent groups, 202-207
of several matched groups, 225-231
- Area**, under normal curve, 40, 141-142
- Arithmetic mean**, 34
- Attenuation**, correction for, 102
- Attributes**, prediction of, 258-261
- Average deviation**, 36-38
- Averages**, 7-8, 34
correlation between, 181
of rates, 48
- Binomial**, rules for expanding, 137-138
- Binomial coefficients**, 137-138
- Binomial distribution**, 137-140
- Biserial coefficient of correlation**, 112-116, 127-128
assumptions involved in computing, 113
computed from widespread classes, 113
- Centiles**, 45-47
as measures of relative position, 46
- Central tendency**, measures of, 34, 41-45, 49-50
- Charlier checks**, 74
- Chi-square**, 239-255
applied to an a priori hypothesis, 241-246
- Chi-square—Continued**
applied to more than two categories, 245-246
applied to several samples, 250-252
applied to two samples, 246-250
calculated from per cents, 244
correction for small frequencies, 253-254
degrees of freedom for, 248
interpretation of, 240
relation to contingency coefficient, 123
relation to phi coefficient, 254-255
sampling distribution of, 241
as test of "goodness of fit," 252-253
- Class intervals**, 67-70
assumptions concerning scores within, 70
influence of number of on accuracy, 68
midpoint of, 71
recorded limits of, 69-70
size of, 68-69
theoretical limits of, 70
- Coding**, 58-67, 270
and calculation of correlation coefficient, 91-93
and calculation of regression coefficient, 270
corrections for, 60-67
by division, 63-66
formulas, 66-67
by subtraction, 58-63
- Coefficient**, biserial, 112-116
contingency, 122-123
of correlation, 9, 79-99
of determination, 99-100
of non-determination, 99-100
phi, 117-122
point-biserial, 115
rank-difference, 123-125
regression, 264, 266, 270
- Combinations**, 136-137
- Common elements**, theory of, 100-101

- Confidence intervals, 162-165
 Constant, definition of, 24
 Constant errors, 283
 Contingency coefficient, 122-123
 Control group, 12, 288-295
 formed by matching on experimental variable, 294
 formed by matching on related variable, 294
 formed by pairing individuals, 291-294
 formed by random selection, 289-291
 single group serving as own, 294
 Coordinates, 82
 Correction, for attenuation, 102
 for coding, 60-66
 for small frequencies in calculating chi-square, 253-259
 Correlation, accuracy of predictions based upon, 261
 and causation, 79
 chart, 96
 and common elements, 100-101
 as means of reducing standard error, 183
 multiple, 125-126, 181
 negative, 83-85
 partial, 125-126
 positive, 80, 84-85
 of ranks, 123-125
 and regression, 265-267
 and standard error of mean difference, 179-181
 tetrachoric, 116-117
 Correlation coefficient, 9, 79-99
 computed from coded scores, 91-93
 computed from original measures, 90-91
 computed from scatter diagram, 94-99
 computed from standard scores, 88
 estimated by phi, 120
 as mean of product of paired z scores, 88
 multiple, 125
 partial, 125
 purpose for which used, 101
 Correlation coefficient—*Continued*
 as ratio of two averages of variance, 89
 reduction in size from grouping, 127
 relation to regression coefficient, 265-266
 reliability of, 127
 significance of, 185-189
 size of, 80, 99
 zero order, 125
 Correlation ratio, 83, 107-112, 232-236
 without bias, 232-236
 relation of size of to number of class intervals, 127
 size of, 108
 summary of steps in computing, 110-111
 Covariance, 87, 89
 Cross-products, computation of, 87
 Curvilinear relationships, 83, 106-111
 Deciles, 46
 Decimals, 16-17
 Degree of confidence, 13, 161-162
 in rejecting hypotheses, 165
 Degrees of freedom, 166-167, 168, 170, 177, 179-180, 182-183, 188-189, 197-198, 201, 206, 210, 212-213, 221-223, 235-236, 242, 245-246, 248-249, 252-253, 296
 in calculating chi-square, 242, 249
 in determining fiducial limits of mean, 167
 formulas for in analysis of variance, 207, 231
 for mean difference between independent groups, 182
 for mean difference between matched groups, 177, 179, 180
 in testing "goodness of fit" of normal distribution, 252-253
 in testing significance of r , 188
 Dependent variable, 12
 Deviation, 36-37
 of observed from expected frequency, 240
 significant, 139, 163

- Deviation—Continued
 from some value other than mean, 59
- Diagrams, scatter, 94
 for computing tetrachoric correlation, 117
- Differences, sampling distribution of mean of, 281-282
- Distributions, binomial, 137-140
 frequency, 41
 leptokurtic, 149
 normal, 39, 41; 139
 platykurtic, 149
 reduced to common scales, 48
 of sample means, 154-156
 sampling, 153-156
 skewed, 148-149
 truncated, 41
- Epsilon, 232-236
 Epsilon-square, 232-236
 comparison with F , 234-236
 relation to analysis of variance, 234-236
 as test of rectilinear relationship, 237
 use of tables of, 233
- Equations, rules for performing operations upon, 25-27
- Errors, constant, 283
 of estimate, 262
 of first kind, 165
 of measurement, 102
 of random sampling, 281, 283
- Eta, 107
- Experimental design, 11, 13, 30, 192, 208, 278, 281, 299
- Experimental group, 12
- Experimental variable, 12
- Experiments, nature of, 11
- Fiducial limits, 162-166
 of mean difference, 177-178, 182
 of proportion, 169
 and small samples, 166-170
- Fiducial probability, 162
- Fractions, 15-16
- Frequency, standard error of, 170
- Frequency distributions, 41, 67
- Geometric mean, 49, 89
- "Goodness of fit," chi-square as test of, 252-253
- Graphs, 81-82
- Harmonic mean, 49
- Homogeneity of variance, test of, 296
- Hypotheses, testing of, 160-166, 176, 187-188, 201, 236-237, 240, 243, 247-248, 295-299
- Hypothesis, acceptance of false, 165
 based upon pooled frequencies, 247
 of homogeneity of variance, 295
 of no difference, 279-282
 null, 176-177, 184-185, 193, 280-281, 234
 rejection of true, 165
 and theory, 278-279
- Identities, substitution of in formulas, 26-27
- Independent groups, 181-183, 192-214
- Independent variable, 12
- Index of forecasting efficiency, 269
- Interaction, 208
- Interaction variance, 212-214
- Interval, 41
- Intervals, confidence, 162-165
- Kurtosis, 149
- Least squares, principle of, 272
- Matched groups, 12, 30, 152, 180-181
- Mean, arithmetic, 34
 calculated from grouped scores, 72-74
 geometric, 49
 harmonic, 49
 as a measure of concentration, 34-36
 of a set of z scores, 48
 standard error of, 157-159
- Mean difference, fiducial limits of, 177-178
 significance of when variances differ, 297-299

- Mean difference—*Continued*
 standard error of, 174-176
- Mean square, 33
- Means, variability of in correlation chart, 103-109
- Measurements, approximate nature of, 54-58
 errors of, 102
 paired, 11, 80
 precision of, 67
 predictions of, 261-273
 reduced to common scales, 47-48, 88
 reliability of, 102
 in social sciences, 147
 transformed into set of ranks, 125
- Median, 41-45, 75
- Method of least squares, 262
- Midpoint, of class interval, 71
- Mode, 49
- Multiple correlation coefficient, 125-126, 128, 181
- Normal curve, 39-40, 140-144
 and large samples, 159-162
 ordinates of, 114, 141-143
 and standard scores, 142
 use of tables of, 40, 144-146
 zone embracing middle 95 per cent of measures, 145
- Normal distribution, 39-40, 139-140
 fitted to sample set of measurements, 142-144
 as a mathematical ideal, 146
 relation of mode, median, and mean in, 148
 of test scores, 147
 testing "goodness of fit" of, 252-253
- Null hypothesis, 176-177, 184-185, 193, 234, 280-281
- Numbers, in a series, 21
 positive and negative, 18-20
 techniques of rounding, 57-58
- Opinion polls, 283
- Ordinate, 81
 of normal curve, 114
- Paired associates, method of, 31
- Paired observations, 11
 advantages of, 183-184
 reduction in degrees of freedom for, 183-184
- Parameters, 51, 153
- Partial correlation coefficient, 125-126, 128
- Per cents, 17-18
 calculation of chi-square from, 244
 standard errors of, 170
- Phi coefficient, 117-122, 128
 applied to true dichotomies, 118-119
 and assumption of continuity, 119-121
 estimate of r derived from, 120
 relation to chi-square, 254-255
 as substitute for tetrachoric r , 121-122
- Point biserial r , 114-116, 128
- Population, definition of, 51
 estimate of mean of, 159
 estimate of variance of, 192, 222
 ratio, 240
- Precision, of measurements, 67
- Predictions, 13
 accuracy of dependent upon correlation, 261
 of attributes, 258-261
 based upon means of columns of correlation table, 272
 errors of, 259, 262
 evaluation of efficiency of, 259, 269
 of measurements, 261-273
- Principle of least squares, 272
- Principle of maximum likelihood, 258
- Principle of maximum probability, 258
- Probability, a priori, 134
 empirical, 134
 fiducial, 162
 meaning of, 133-136
 principle of maximum, 258
- Probable deviation, 49
- Probable error, 49
- Proportions, 17-18
 standard error of, 168-169
 testing significance of, 184-185

- Public opinion polls, 152
margin of error in, 153
samples used in, 152
- Quadrant, 82
Quartile, 45
- Random numbers, tables of, 289-290
Range, 5
of correlation coefficient, 80, 99
inclusive, 33-34
as measure of variability, 8-9, 31-34
middle 80 per cent, 49
semi-interquartile, 45
Rank-difference coefficient, 123-125
Ranks, method of dealing with ties, 125
Rates, averaging of, 49
Rectilinear relationship, test of significance of departure from, 237
Regression coefficient, 263
coding of, 270
and correlation, 265-267
Regression equation, 265
Regression line, 83, 262-263
fitted by method of least squares, 263
Relationships, between averages and differences, 9
curvilinear, 83, 106-107
negative, 79, 85
positive, 79, 84
rectilinear, 83
study of, 9-10, 79-80
Relative deviates, 47
Reliability, 13, 153
Replication, 230
Research problems, 277-279
Residual sum of squares, 221-222, 229-230
Residual variance, 222, 224, 230
- Samples, definition of, 50
generalizations from, 283
large, 287
- Samples—Continued
random, 283-286
reasons for studying, 151-152
representativeness of, 282, 286
and research, 282-286
size of, 287-288
small, 166-168, 170, 287-288
used in opinion polls, 283
Sampling distribution, 153-156, 173
of chi-square, 241-242
of correlation coefficient, 186-187
of difference between correlated means, 181
of difference between means, 173
of epsilon-square, 233
of means, 157
of standard deviation, 168
of *t* ratio, 166-167
of variance ratio, 193, 296
Sampling, errors of random, 281, 283
Sampling theory, 153
Scatter diagram, 94
Significance, tests of, 57, 297-299
Significant figures, 55-56
Skewness, 148-149
Squares and square roots, 21-23
Squares, sum of, 39, 51, 60-63, 65, 157, 193-195, 202, 210, 219, 226
Standard deviation, 27-39
calculated from grouped scores, 73
of a distribution of *z* scores, 48
estimate of population, 157-158
of a sampling distribution, 156
standard error of, 168
Standard errors, 156-159, 168-170, 174-189, 200
based on pooled sum of squares, 200
of difference in mean gains, 293
of estimate, 267-269, 272-273
of frequencies, 170
of mean differences, 174, 179-180, 200
of mean differences of independent groups, 181, 200
of mean differences when correlation is present, 179-181
of means, 157-159
of per cents, 170, 184-185
of proportions, 168-169, 184-185

- Standard errors—*Continued*
 relation to population variability, 156
 relation to sample size, 156
 of standard deviation, 168
- Standard scores, 47-49, 88, 141
 mean of distribution of, 48
 mean of products of paired, 88
 as measures of relative position, 47-48
 standard deviation of distribution of, 48
- Statistical inference, 13
- Statistical methods, applications of, 3
 approaches to study of, 6
 continuity of development of, 6
 functions of, 10-14
 mathematical bases of, 4
- Statistics, 51, 152
- Sum of cross-products, 87
 computed from correlation chart, 97-98
 computed from original measures, 90
 correction for coding, 90
 difference method of computing, 91
- Sum of squares, 51, 60-63, 65, 157, 193-195, 202, 210, 219, 226
 computed from correlation chart, 96-97
 between groups, 193, 195-196, 203-204, 210, 219, 227
 within groups, 193, 195, 204-205, 210, 219, 227, 230
 interaction, 212, 230
 pooling of within groups, 199, 204-205, 210, 219, 227
 residual, 221-222, 229-230
- Summation, 23-25, 35
 of an algebraic sum of two or more terms, 24
 of a constant, 24
 of a variable, 24
 of a variable divided by a constant, 25
- Symbols, use of, 5, 34-36
- Test construction, 14
- Tests of significance, 133, 184
 chi-square, 239-255
 comparison of F and t in case of two groups, 199-202, 218-225
 epsilon-square, 232-237
 interpretation of, 279-282
 t ratio, 166, 187
 variance ratio, 193, 296
- Tetrachoric correlation, 116-117
 assumptions involved in computing, 117
 diagrams for computing, 117
 and phi coefficient, 121
- Two-way frequency table, 94
- Universe, definition of, 51
- Variability, 8-9
 in population as related to sample, 156
- Variable, definition of, 23
 dependent, 12
 dichotomous, 112-113
 experimental, 12
 independent, 12
- Variables, used for matching groups, 293-294
- Variance, analysis of, 192-231
 assumption of homogeneity of, 295-299
 definition of, 38
 estimate of population, 192, 197, 199
 interaction, 212-214
 of means of columns in correlation chart, 108-109
 in one variable associated with second, 99
 residual, 222, 224, 230
- Variance ratio, 193, 198, 296, 205-206, 213-214, 218, 225, 227, 296
- X-axis, 81-82
- Y-axis, 81-82