

# THE VECTORS OF MIND

MULTIPLE-FACTOR ANALYSIS  
FOR THE ISOLATION OF PRIMARY TRAITS

*By*

L. L. THURSTONE

*Professor of Psychology  
The University of Chicago*



1935

THE UNIVERSITY OF CHICAGO PRESS  
CHICAGO · ILLINOIS

# CONTENTS

## MATHEMATICAL INTRODUCTION

	PAGE
MATRICES . . . . .	1
DETERMINANTS . . . . .	3
MATRIX MULTIPLICATION . . . . .	13
DIAGONAL MATRICES . . . . .	21
THE INVERSE . . . . .	23
THE CHARACTERISTIC EQUATION . . . . .	26
THE SUMMATIONAL NOTATION . . . . .	28
LINEAR DEPENDENCE . . . . .	32
GEOMETRIC INTERPRETATIONS . . . . .	33
ORTHOGONAL TRANSFORMATIONS . . . . .	38
OBLIQUE TRANSFORMATIONS . . . . .	43

## CHAPTER I. THE FACTOR PROBLEM

ON THE NATURE OF SCIENCE . . . . .	44
PSYCHOLOGICAL POSTULATES AND DEFINITIONS . . . . .	48
MATRIX FORMULATION . . . . .	54
COMMUNALITY . . . . .	62
THE INTERCORRELATIONS . . . . .	64
THE RELIABILITY COEFFICIENT. . . . .	66
THE POPULATION SPACE . . . . .	68
THE COMMON-FACTOR SPACE . . . . .	69

## CHAPTER II. THE FUNDAMENTAL FACTOR THEOREM

THE CORRELATIONAL MATRICES . . . . .	70
THE NUMBER OF INDEPENDENT FACTORS . . . . .	72
ALGEBRAIC AND CONFIGURATIONAL UNIQUENESS . . . . .	73
THE CASE OF $n$ TESTS AND $n$ FACTORS . . . . .	77
A METHOD OF FACTORING ANY SYMMETRIC MATRIX . . . . .	78
THE RANK OF A MATRIX . . . . .	81
METHODS OF ESTIMATING COMMUNALITIES . . . . .	85
1. EXPANSION OF A MINOR OF ORDER $(r+1)$ . . . . .	86
2. GROUPING OF SIMILAR TESTS . . . . .	87
3. GROUPING OF THREE TESTS . . . . .	87
4. HIGHEST COEFFICIENT IN EACH COLUMN . . . . .	89
5. LINEAR DEPENDENCE OF ROWS OR COLUMNS . . . . .	89
6. SECTIONING OF THE MATRIX . . . . .	90
7. EXPANSION OF PRINCIPAL MINORS OF ORDER $(n-1)$ . . . . .	91
8. EXPANSION OF PRINCIPAL MINORS OF ORDER $(r+a)$ . . . . .	91

## CHAPTER III. THE CENTROID METHOD

	PAGE
PRINCIPLES OF THE METHOD . . . . .	92
EXAMPLE 1. UNITY IN THE DIAGONAL CELLS . . . . .	98
EXAMPLE 2. COMMUNALITIES IN THE DIAGONAL CELLS . . . . .	101
EXAMPLE 3. EACH DIAGONAL ENTRY GREATER THAN THE COMMUNALITY AND LESS THAN UNITY . . . . .	102
EXAMPLE 4. EACH DIAGONAL ENTRY LESS THAN THE COMMUNALITY . . . . .	102
EXAMPLE 5. A FICTITIOUS EIGHT-VARIABLE PROBLEM WITH KNOWN COM- MUNALITIES . . . . .	103
EXAMPLE 6. THE CENTROID METHOD WITH UNKNOWN DIAGONALS . . . . .	108
CORRECTION FOR UNIQUENESS . . . . .	118

## CHAPTER IV. THE PRINCIPAL AXES

A METHOD OF LOCATING THE PRINCIPAL AXES . . . . .	120
A NUMERICAL EXAMPLE OF THE METHOD OF PRINCIPAL AXES . . . . .	124
HOTELLING'S SPECIAL CASE . . . . .	129
THE PRINCIPAL AXES OF A BATTERY OF FIFTEEN PSYCHOLOGICAL TESTS . . . . .	132

## CHAPTER V. THE SPECIAL CASE OF RANK ONE

THE INTERCOLUMNAR CRITERION . . . . .	134
GRAPHICAL METHOD FOR RANK 1 . . . . .	136
THE TETRAD DIFFERENCE . . . . .	137
GRAPHICAL ANALYSIS OF TETRADS . . . . .	141
A SINGLE-FACTOR METHOD WITHOUT TETRADS . . . . .	144
NUMERICAL EXAMPLE OF SINGLE-FACTOR METHOD . . . . .	147

## CHAPTER VI. PRIMARY TRAITS

SIMPLE STRUCTURE . . . . .	150
OBLIQUE REFERENCE VECTORS . . . . .	154
UNIQUENESS OF SIMPLE STRUCTURE . . . . .	155
PRIMARY TRAIT VECTORS . . . . .	156
THE EQUATION OF AN OBLIQUE SIMPLE STRUCTURE . . . . .	162
FIVE METHODS OF ISOLATING PRIMARY TRAITS . . . . .	163
GRAPHICAL METHODS FOR LESS THAN FOUR DIMENSIONS . . . . .	164
THE PROBLEM OF NEGATIVE ABILITIES . . . . .	165
GRAPHICAL ANALYSIS OF FIFTEEN PSYCHOLOGICAL TESTS . . . . .	167

## CHAPTER VII. ISOLATION OF PRIMARY FACTORS

METHOD OF OBLIQUE AXES . . . . .	171
NUMERICAL EXAMPLE OF METHOD OF OBLIQUE AXES . . . . .	172
CONSTELLATIONS . . . . .	174
THE METHOD OF AVERAGES . . . . .	175
NUMERICAL EXAMPLE OF THE METHOD OF AVERAGES . . . . .	176
THE SPECIAL CASE OF RANK 2 . . . . .	177

# CONTENTS

xv

PAGE

PROJECTIONS OF UNIT TRAIT VECTORS INTO A HYPERELLIPSOID . . . . .	178
A METHOD OF MAXIMIZING THE NUMBER OF ZERO FACTOR LOADINGS (THE SINGLE HYPERPLANE METHOD) . . . . .	179
ANALYTICAL METHOD OF ISOLATING SIMPLE STRUCTURE . . . . .	185
A METHOD OF SUCCESSIVE APPROXIMATION FOR ISOLATING SIMPLE STRUCTURE	189
NUMERICAL EXAMPLE OF METHOD OF SUCCESSIVE APPROXIMATION . . . . .	194
COMPARISON OF METHODS . . . . .	197

## CHAPTER VIII. THE POSITIVE MANIFOLD

RESTRICTIONS ON THE FACTORIAL MATRIX . . . . .	199
THE ELIMINATION OF NEGATIVE FACTOR LOADINGS . . . . .	201
UNITARY FACTORS . . . . .	205

## CHAPTER IX. ORTHOGONAL TRANSFORMATIONS

ROTATION IN THREE DIMENSIONS . . . . .	213
ALTERNATIVE TRANSFORMATIONS . . . . .	220
ROTATION IN FOUR DIMENSIONS . . . . .	222

## CHAPTER X. THE APPRAISAL OF ABILITIES

THE REGRESSION $x$ ON $s$ . . . . .	226
THE REGRESSION $s$ ON $x$ . . . . .	228

## APPENDIXES

I. OUTLINE OF CALCULATIONS FOR THE CENTROID METHOD WITH UNKNOWN DIAGONALS . . . . .	232
II. A METHOD OF FINDING THE ROOTS OF A POLYNOMIAL . . . . .	251
III. A METHOD OF DETERMINING THE SQUARE ROOT ON THE CALCULATING MACHINE . . . . .	253
INDEX . . . . .	255

## PREFACE

Thirty years ago Professor Charles E. Spearman introduced the factor problem in psychology when he observed that the intercorrelations of a set of tests revealed an underlying order. He interpreted this order as the effect of a conspicuous factor that was common to all of the tests. There has been much controversy about different aspects of Spearman's single-factor hypothesis and about his single-common-factor methods of analyzing intercorrelations. His single-common-factor hypothesis is that the intercorrelations of a group of tests can be explained in terms of a single central intellectual factor which has been denoted "*g*," and that the variance of each test can be explained by the "*g*" factor and a factor that is specific and unique for each test. Since his hypothesis involves a factor which all of the tests have in common and a factor which is unique for each test, it is frequently called a "two-factor hypothesis." Spearman's single-factor methods are concerned with the isolation of a single common factor in each test battery which can be so analyzed.

Another interpretation of the mental abilities is that of Professor E. L. Thorndike, who has been a leader in this type of psychological research. It has been his judgment that the socially significant mental abilities are numerous and discrete.

The factor methods described in this volume are based on the assumption that a test score can be expressed, in first approximation, as a linear function of a number of factors. My previous papers on the multiple-factor problem are as follows:

"Multiple Factor Analysis," *Psychological Review*, XXXVIII, No. 5 (September, 1931), 406-27.

"A Multiple Factor Study of Vocational Interests," *Personnel Journal*, X, No. 3 (October, 1931), 198-205.

"Isolation of Blocs in a Legislative Body by the Voting Records of Its Members," *Journal of Social Psychology*, III, No. 4 (November, 1932), 425-33.

*Theory of Multiple Factors* (January, 1933). Pp. 65. University of Chicago Bookstore.

*Computing Diagrams for Tetrachoric Correlation Coefficients* (April, 1933). Pp. 57. University of Chicago Bookstore.

*A Simplified Multiple Factor Method* (May, 1933). Pp. 25. University of Chicago Bookstore.

"The Vectors of Mind," *Psychological Review*, XLI, No. 1 (January, 1934), 1-32.

"Unitary Abilities," *Journal of General Psychology*, XI, No. 1 (July, 1934), 126-32.

The fundamental equation in my first paper on factor theory is the same as the first equation in the present volume, but the development here presented is more formal and considerably extended. The fundamental assumptions and the corresponding theorems are given in *chapters i* and *ii*. The centroid method which was described in my first paper has been improved several times, and it is presented in *chapter iii*. The notation has been made more explicit and unambiguous. The fundamental factor theorem was first stated in *Theory of Multiple Factors*. In the present volume this theorem is the subject of *chapter ii*. The theorem states that the number of linearly independent common factors in a battery of tests is the rank of their reduced correlational matrix.

In Spearman's special case, where only one common factor is involved, the rank of the correlational matrix must therefore be one. Hence all second-order minors must vanish. The expansions of the second-order minors are, in fact, Spearman's tetrad differences. This case is discussed in *chapter v*. We are concerned here with the generalization of the factor problem to  $n$  dimensions.

The geometrical formulation of the factor problem which was described in my earlier papers has been reproduced in this volume. Each test may be regarded as a radial vector in a common-factor space of as many dimensions as there are common factors in a test battery. The correlation between any pair of tests is the scalar product of the test vectors. Since the scalar product of a pair of vectors is independent of the co-ordinate system, it follows that the intertest correlations define the test configuration in a common-factor space but that they do not define the co-ordinate system. But the co-ordinate axes are the scientific categories in terms of which the tests are to be comprehended. This is an interesting indeterminacy. One of the principal problems of factor analysis is to find a unique set of co-ordinate axes, either orthogonal or oblique, which shall represent scientifically meaningful categories in terms of which the tests may be comprehended. This problem has been solved in terms of what I have called "simple structure" of a trait configuration. This concept is developed in *chapters vi* and *vii*.

One of the important restrictions that must be satisfied by any acceptable solution to the factor problem is that the factorial description of a trait or test must be invariant when it is moved from one battery to another. No form of uniqueness can be scientifically meaningful which violates this principle. This is the reason why I have discarded one of my earlier solutions, namely, the principal axes of the configuration; and it is also the reason why Professor Harold Hotelling's special case of the principal axes solution must be discarded. His special case has been called a method of principal components. The principal axes are discussed in *chapter iv*.

In some applications of factor theory it seems appropriate to impose the

restriction that the direction cosines of each trait vector shall be positive or zero. This special case of the factor problem is developed in *chapter viii* on "The Positive Manifold." In some applications it may be appropriate to impose the restriction that the fundamental categories or reference traits shall be uncorrelated in the experimental population. The reference vectors are then orthogonal, and the determination of an orthogonal simple structure is then demanded. Orthogonal transformations of possible use for this problem are described in *chapter ix*.

Finally, when the common factors are known, it is of interest to appraise each individual member of the statistical population as regards each of the primary factors. The solution of this regression problem is given in *chapter x*. The other regression involves the prediction of a test performance in terms of the test coefficients and primary factors. The solution to this regression problem is also given in *chapter x*.

Since I was myself unfamiliar with matrix theory until very recently, I could hardly take this subject for granted in writing for other psychologists with limitations of training that are similar to mine. It was therefore imperative to supply students of factor analysis with a mathematical introduction to matrix theory and related topics. This seemed all the more necessary in view of the fact that the available textbooks on this subject are unsatisfactory. In the "Mathematical Introduction" I have attempted to present the essential mathematical ideas as clearly as may be possible in the scope of a single chapter. The introduction is written for students who have had the conventional undergraduate instruction in analytic geometry and in the calculus. It is explicitly limited to the real case, since complex numbers and imaginaries have not yet been introduced in factor analysis.

One of the turning-points in the development of multiple-factor analysis was the discovery in 1931 that the mathematics most adaptable to this problem was matrix theory. I once asked Professor Gilbert A. Bliss how to factor a correlation table, but I did not call it a "matrix." He suggested that matrix theory might be applicable to my problem, but I was entirely unfamiliar with this branch of mathematics. Since that time I have profited on numerous occasions by the generosity of the members of the Department of Mathematics at the University of Chicago. I appreciate especially the interest of Professor R. W. Barnard. He suggested the equation by which a simple structure can be represented. A "simple structure" may be regarded either as a combined configuration of test vectors and reference vectors or as the aggregate of co-ordinate hyperplanes. Professor Barnard has also made valuable suggestions in connection with the problem of determining the co-ordinate hyperplanes of a simple structure by successive approximation in the analytical method.

I owe a special acknowledgment to my tutor in mathematics, Mr. Patrick

Youtz, who assisted me in becoming familiar with the elements of matrix theory. Later I was fortunate when Mr. Youtz accepted a full-time assignment on the factor projects. He has examined, mathematically, each of the factor methods, and he has read and criticized in detail the manuscript for this volume. It is through him that I have become acquainted with some of the conventions of mathematical writing.

Special acknowledgment is due Miss Leone Chesire, who has been responsible for most of the computing on my factor studies during several years. I have relied constantly on Miss Chesire's competent work in testing the many leads that we have investigated in factor theory. Her careful criticism of the entire manuscript has been of great value, and she has prepared the appendix on the centroid method.

I am indebted to my colleagues, Professor Mortimer J. Adler, Professor A. C. Benjamin, and Professor C. W. Morris, for reading and criticizing the general sections of the first chapter.

The entire manuscript for this volume has been read and criticized in detail by four readers. Mr. Patrick Youtz and Miss Leone Chesire have read the manuscript and shared in the supervision of the computing. My wife, Thelma Gwinn Thurstone, has read and criticized the manuscript both for the mathematical and the psychological content. Mr. Joseph Novak, as a mathematician, has read and criticized the manuscript without previous familiarity with the factor problem. All of these readers have suggested many revisions that were intended to clarify the exposition, but I assume responsibility for all of the solutions, as well as for any errors that may be found. It cannot be hoped that this volume will be free from errors, since all of the chapters, except *chapter v*, cover new ground. I am indebted also to Mrs. Cypra Feinsot, who has supervised the work of preparing the manuscript for the publishers.

Three studies are now in progress which involve applications of the factorial methods. These will appear eventually in monograph form. They are (1) a factor analysis of sixty psychological tests that were taken by two hundred and forty college students who volunteered fifteen hours of testing, (2) a factor study of several hundred personality traits on which thirteen hundred adults were rated, and (3) a factor study of vocational interests of three thousand college students with respect to eighty professions.

In carrying out these theoretical investigations, as well as the practical applications, the Social Science Research Committee at the University of Chicago has been most generous. I wish to acknowledge especially the interest of Professor Donald Slesinger, chairman of the Committee. I am grateful for the financial assistance and for the physical facilities that this Committee has placed at my disposal during the past four years. I am grate-



ful for a grant by the Carnegie Corporation of New York by which it became possible to add several research assistants during the past year. This grant has considerably aided in the development of factor theory. I am also grateful to the Illinois Emergency Relief Commission for assigning relief funds to these studies. Twelve computers have been at work on these factor projects.

The future development of factor theory will probably reduce factor analysis to simpler computing methods. The linear approximations that are here used may eventually prove to be inadequate, but it is likely that much can be accomplished by these approximations in psychology and in other social sciences. The factor methods may be regarded as an intermediate stage in the development of science. No one would think of investigating the fundamental laws of classical mechanics by correlational methods or by factor methods, because the laws of classical mechanics are already well known. If nothing were known about the law of falling bodies, it would be sensible to analyze, factorially, a great many attributes of objects that are dropped or thrown from an elevated point. It would then be discovered that one factor is heavily loaded with the time of fall and with the distance fallen but that this factor has a zero loading in the weight of the object. The usefulness of the factor methods will be at the borderline of science.

No attempt has been made in this volume to integrate the present multiple-factor analysis with the previous work of Professor Sewell Wright on path coefficients and with the work of Professor Truman L. Kelley on multiple factors. While these several approaches to the problem seem to be quite different, it should be possible to unify them. As far as I am aware, my own work is not in conflict with the work of others on the multiple-factor problem or with that of Professor Spearman on the single-factor methods. The development of factor theory, as well as its applications in science, will be accelerated by the assistance of mathematicians; and it is gratifying that Professor E. B. Wilson has turned his attention to these problems in several papers. The future development of factor analysis in psychology will probably require more mathematical competence than we can supply in our own ranks.

L. L. THURSTONE

CHICAGO, ILLINOIS  
March, 1935

## MATHEMATICAL INTRODUCTION

The matrix theory which is used in the development of factor analysis is not generally available to students whose training in mathematics is limited to undergraduate courses in analytical geometry and in the calculus. This mathematical introduction reviews the elementary theory of matrices as well as the closely related theory of determinants. Summation that involves double subscript notation is included in this section, since it is used in factor theory and since it is unfamiliar to most students of statistics. In the geometrical interpretation of the factorial matrix, only non-homogeneous co-ordinates are used. For this reason, the introduction includes non-homogeneous co-ordinates and omits homogeneous co-ordinates which are conventional. Orthogonal and oblique transformations have been illustrated geometrically. No provocation has been found so far in factor theory to introduce imaginaries and complex numbers, but the future development of factor analysis may call for them. *This mathematical introduction is limited to the real case*, and all theorems have been written with this restriction in mind.

If this introduction is not self-sufficient, perhaps it may serve as a useful guide to the student of factor theory who seeks mathematical assistance on specified topics. If a student has the intention of attaining some competence in factor theory and in related statistical work, there is no short cut for formal courses in the mathematics that is involved.\*

### Matrices

Matrices and determinants involve rectangular arrangements of numbers. Any rectangular arrangement of numbers is called a *matrix*, irrespective of what the numbers mean. If the matrix has  $m$  rows and  $n$  columns, the matrix is said to be of *order*  $m \times n$ . In designating the order of a matrix, it is customary to refer to *rows first* and *columns second*. Thus a matrix of

\* The following references will be found useful:

W. F. Osgood and W. C. Graustein, *Plane and Solid Analytic Geometry* (New York: Macmillan Co., 1929).

L. E. Dickson, *Modern Algebraic Theories* (New York: B. H. Sanborn, 1926), chap. iii.  
Maxime Bôcher, *Introduction to Higher Algebra* (New York: Macmillan Co., 1931), chaps. i-vi, inc.

H. W. Turnbull and A. C. Aitken, *Canonical Matrices* (London: Blackie & Sons, 1932), chap. i.

V. Snyder and C. H. Sisam, *Analytical Geometry of Space* (New York: Henry Holt & Co., 1914).

order  $p \times q$  has  $p$  rows and  $q$  columns. *Tables 1a* and *1b* show a matrix of order  $3 \times 4$  and a matrix of order  $3 \times 3$ . A *row* is horizontal. A *column* is vertical. The general name for either a row or a column is an *array*. Each of the small squares into which a matrix is divided is called a *cell*, and the number in each cell is called a *cell entry* or *element*.

<i>Table 1a</i>	<i>Table 1b</i>
2 3 1 5	2 1 5
1 6 0 9	4 8 3
0 2 6 7	2 0 7

In order to designate a particular element, it is customary to use a double subscript, the first one for the row and the second one for the column. If a matrix is denoted  $A$ , then its elements may be denoted  $a_{ij}$ , where  $i$  shows the row and  $j$  shows the column at the intersection of which the element  $a_{ij}$  is found. Thus, in *Table 1a* the element  $a_{12}=3$  and  $a_{24}=9$ .

In developing the theory of matrices it is desirable to exhibit the elements as shown in *Table 2*. The elements in the first row are  $a_{11}, a_{12}, a_{13}, \dots, a_{1n}$ , showing that the table represents a matrix of  $n$  columns. The elements of the first column are  $a_{11}, a_{21}, a_{31}, \dots, a_{m1}$ , showing that the table represents  $m$  rows. The general element in this matrix  $A$  is  $a_{ij}$ , where  $i$  takes the successive values  $1, 2, 3, \dots, m$ , while  $j$  takes the successive values  $1, 2, 3, \dots, n$ . The first subscript refers to the row; the second subscript refers to the column.

<i>Table 2</i>				
$a_{11}$	$a_{12}$	$a_{13}$	$\dots$	$a_{1n}$
$a_{21}$	$a_{22}$	$a_{23}$	$\dots$	$a_{2n}$
$a_{31}$	$a_{32}$	$a_{33}$	$\dots$	$a_{3n}$
$\dots$	$\dots$	$\dots$	$a_{ij}$	$\dots$
$a_{m1}$	$a_{m2}$	$a_{m3}$	$\dots$	$a_{mn}$

The conventional representation of a matrix is shown in *Table 3*, where the rectangular arrangement of numbers is inclosed by double vertical lines on the left and on the right sides of the rectangle. It is also customary to denote specified matrices with letters. Thus, the matrix of *Table 3* might be conveniently designated  $A$  or any other letter. A matrix  $A$  might also be designated by its general element  $a_{ij}$ .

<i>Table 3</i>		
2	1	5
4	8	3
2	0	7

If the successive rows of matrix  $A$  are written as successive columns of a new matrix, the new matrix is called the *transpose* of  $A$ . It is denoted  $A'$ . *Table 4* shows a matrix  $A$  and its transpose  $A'$ .

*Table 4*

$\left\  \begin{array}{cccc} 2 & 3 & 1 & 5 \\ 1 & 6 & 0 & 9 \\ 0 & 2 & 6 & 7 \end{array} \right\ $	$\left\  \begin{array}{ccc} 2 & 1 & 0 \\ 3 & 6 & 2 \\ 1 & 0 & 6 \\ 5 & 9 & 7 \end{array} \right\ $
$A$	$A'$

**Determinants**

One particular interpretation of a square matrix is called a determinant. This interpretation of a square matrix probably had its origin in the practical work of solving simultaneous equations, and it is indicated by single vertical lines on the left and on the right sides of a square table. It is illustrated in *Table 5*. *Table 3* is called a matrix; while *Table 5*, which implies a particular interpretation, is called a determinant. A determinant is always square. Hence its order is  $n$ , in which  $n$  is the number of rows or the number of columns.

*Table 5*

$$\left| \begin{array}{ccc} 2 & 1 & 5 \\ 4 & 8 & 3 \\ 2 & 0 & 7 \end{array} \right|$$

The diagonal from the upper left corner to the lower right corner of a determinant is called the *principal diagonal*. In *Table 5* the principal diagonal contains the elements 2, 8, 7. The other diagonal from the lower left corner to the upper right corner is called the *secondary diagonal*.

In many problems it is convenient to assign a plus sign and a minus sign to alternate cells in a determinant. A convenient rule is to designate the upper left cell as positive and all other cells as alternately negative and positive, as the cells can be moved over by the castle in a chess game. This sign arrangement is illustrated in *Table 6* for a determinant of order 5.

*Table 6*

$$\left| \begin{array}{ccccc} + & - & + & - & + \\ - & + & - & + & - \\ + & - & + & - & + \\ - & + & - & + & - \\ + & - & + & - & + \end{array} \right|$$

Notice that the cells in the principal diagonal are all positive and that lines parallel to this diagonal are alternately negative and positive. The sign is positive for an even number of steps from the upper left cell, and it is negative when the number of steps is odd. The sign of a cell determined in this manner may be called the *position sign of the cell*. If the general element of the determinant is denoted  $a_{ij}$ , then the element with its position sign may be conveniently denoted  $(-1)^{i+j}a_{ij}$ . When the exponent  $(i+j)$  is odd, the sign of the cell is negative; and when  $(i+j)$  is even, the sign of the cell is positive.

The product of any  $n$  elements of a square matrix, selected with only one element from each of the  $n$  rows and only one element from each of the  $n$  columns is called a *term of the determinant* of the matrix. *Table 7* is a deter-

*Table 7*

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

minant of order 3. From this determinant six terms may be written. These are shown in *Table 8*, in which the elements of each term are arranged in the order of their columns.

*Table 8*

- 1)  $a_{11} a_{22} a_{33}$
- 2)  $a_{11} a_{32} a_{23}$
- 3)  $a_{21} a_{12} a_{33}$
- 4)  $a_{21} a_{32} a_{13}$
- 5)  $a_{31} a_{12} a_{23}$
- 6)  $a_{31} a_{22} a_{13}$

Each of these six terms is the product of three elements so selected that each term contains only one element from each row and only one element from each column. If a square matrix is of order  $n$ , the total number of terms in its determinant is  $n!$ . The term that contains all the elements of the principal diagonal is called the *leading term* of the determinant.

The sign of each of the  $n!$  terms of a determinant can be ascertained in the following manner. Let the  $n$  elements of each term be arranged in ascending order according to columns, as shown in *Table 8*. This can evidently be done without affecting the numerical value of the terms. Consider the fourth term as an example, and list the rows as follows:

Any interchange of two adjacent elements constitutes an *inversion*. It may be illustrated by interchanging 1 and 3. The resulting arrangement is

$$2 \quad 1 \quad 3.$$

If, now, the adjacent elements 1 and 2 are interchanged, the arrangement becomes

$$1 \quad 2 \quad 3,$$

in which the rows are in consecutive order.

The sign of a term of a determinant is positive if it represents an even number of inversions from the consecutive order of rows and columns. The sign of the term is negative if the number of inversions in the term is odd.

Applying this rule to the six terms of *Table 8*, we have the same terms with proper signs as shown in *Table 9*:

*Table 9:*

- 1)  $+a_{11} a_{22} a_{33}$
- 2)  $-a_{11} a_{22} a_{23}$
- 3)  $-a_{21} a_{12} a_{33}$
- 4)  $+a_{21} a_{32} a_{13}$
- 5)  $+a_{31} a_{12} a_{23}$
- 6)  $-a_{31} a_{22} a_{13}$

A complete definition of a determinant can now be given.

**Definition:** *If a square table is used as a symbol of the sum of  $n$  terms, each term being the product of  $n$  elements with only one element from each row and only one element from each column, the sign of each term taken positive or negative according as the term contains an even or an odd number of inversions, then the square table is called a **determinant**.*

Hence the sum of the six terms of *Table 9* is implied by the determinant of *Table 9*. The determinantal interpretation of a square matrix is denoted by single vertical lines on the left and on the right sides of the square table, as shown in *Table 10*.

If a *square matrix* is denoted by a letter such as  $A$ , then the *determinant* of the matrix is denoted  $|A|$ . If  $A$  represents a *number*, then  $|A|$  means the *absolute value*, ignoring the sign of the number  $A$ . It should be noted that a matrix is merely a rectangular table of numbers, and hence a *matrix has no numerical value*. But a determinant is, by definition, a sum of terms, and hence it has a numerical value. If a matrix is denoted  $a_{ij}$ , then its determinant is denoted  $|a_{ij}|$ .

Consider the second-order determinant

$$\begin{vmatrix} 1 & 5 \\ 8 & 3 \end{vmatrix}$$

and the  $2=2$  terms that it implies. These are  $1 \times 3$  and  $8 \times 5$ , in which the factors of each term are arranged in consecutive order by columns. The rows of the term  $1 \times 3$  are 1 and 2. Since these are in consecutive order, the sign of this term is positive. Its value is therefore  $+(1)(3) = +3$ . The rows of the term  $8 \times 5$  are 2 and 1. One inversion changes the order 2 and 1 into the consecutive order 1 and 2. Hence the sign of this term is negative. The determinant therefore has the numerical value  $+3 - 40 = -37$ . Any second-order determinant can be evaluated as follows:

$$\begin{vmatrix} a & d \\ c & b \end{vmatrix} = ab - cd.$$

An  $x$ -rowed minor of the matrix  $A$  is a determinant of order  $x$  which is formed by the intersections of any  $x$  rows and any  $x$  columns of the matrix  $A$ . If one or more columns of a determinant are eliminated and if the same number of rows are eliminated, the remaining cells constitute a *minor*. From the determinant of *Table 7* nine second-order minors may be drawn. A few of them are illustrated here:

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \quad \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}, \quad \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}.$$

If any two columns and any two rows are eliminated from the determinant of *Table 7*, there remains a 1-rowed minor which is a single element. In this sense each element can be regarded as a minor of the determinant.

If corresponding rows and columns are eliminated, the remaining minor is symmetrically placed with regard to the principal diagonal, and it is called a *principal minor*. In the determinant of *Table 7*, three second-order principal minors may be drawn. These are

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \quad \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}, \quad \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}.$$

There are three 1-rowed principal minors in this determinant, namely, the three elements in the principal diagonal.

If the row  $i$  and the column  $j$  which intersect in an element  $a_{ij}$  are eliminated from a determinant, the remaining  $(n-1)$ -rowed determinant is called the *first minor* of  $a_{ij}$ . This definition is illustrated with the determinant of *Table 5*. The second row and the first column intersect in the element 4. If these two arrays are eliminated from the determinant, the remaining 2-rowed determinant is

$$\begin{vmatrix} 1 & 5 \\ 0 & 7 \end{vmatrix} = 7 - 0 = +7.$$

This determinant, whose numerical value is  $+7$ , is the first minor of the element  $a_{21} = 4$  in the determinant of *Table 5*. Let the first minor of the element  $a_{ij}$  be denoted  $m_{ij}$ .

In some problems it is convenient to refer to the minor  $m_{ij}$  with the position sign of the element  $a_{ij}$ . This quantity is called the *cofactor* of  $a_{ij}$ . It is defined by the relation

$$(\text{cofactor of } a_{ij}) \equiv e_{ij} \equiv (-1)^{i+j} m_{ij}.$$

In *Table 5* the cofactor of the element 4 is

$$(-1)^{2+1} \begin{vmatrix} 1 & 5 \\ 0 & 7 \end{vmatrix} = - [7-0] = -7.$$

In the same table the cofactor of the element 3 is

$$(-1)^{2+3} \begin{vmatrix} 2 & 1 \\ 2 & 0 \end{vmatrix} = - [0-2] = +2.$$

Hence the absolute values of the first minor of  $a_{ij}$  and of its cofactor are identical. They differ only in the manner of determining the sign. If the position sign of the element  $a_{ij}$  is positive, the first minor and the cofactor have the same sign. If the position sign of  $a_{ij}$  is negative, they have opposite signs.

The numerical value of a determinant can be expressed conveniently for some problems in terms of the cofactors. For example,

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}e_{11} + a_{21}e_{21} + a_{31}e_{31}.$$



The numerical value of a determinant is the weighted sum of the elements in any array, each element being weighted by its cofactor. In the example, the determinant is expressed in terms of the elements of the first column.

As a numerical example, the value of the determinant of *Table 5* can be expressed as follows:

$$\begin{aligned} \begin{vmatrix} 2 & 1 & 5 \\ 4 & 8 & 3 \\ 2 & 0 & 7 \end{vmatrix} &= +2 \begin{vmatrix} 8 & 3 \\ 0 & 7 \end{vmatrix} - 4 \begin{vmatrix} 1 & 5 \\ 0 & 7 \end{vmatrix} + 2 \begin{vmatrix} 1 & 5 \\ 8 & 3 \end{vmatrix} \\ &= 2(56-0) - 4(7-0) + 2(3-40) \\ &= 112 - 28 - 74 = +10. \end{aligned}$$

The numerical value of a determinant can be expressed as the summation:

$$(1) \quad |a_{ij}| = \sum_{i=1}^n a_{ij}e_{ij} = \sum_{j=1}^n a_{ij}e_{ij},$$

where the weighted sum may be taken over any column or any row. The following is an example of a fourth-order determinant, evaluated by the method of (1).

$$|A| = \begin{vmatrix} 2 & 4 & 1 & 0 \\ 3 & 2 & 4 & 2 \\ 1 & 6 & 1 & 4 \\ 1 & 0 & 2 & 3 \end{vmatrix} = +2 \begin{vmatrix} 2 & 4 & 2 \\ 6 & 1 & 4 \\ 0 & 2 & 3 \end{vmatrix} - 3 \begin{vmatrix} 4 & 1 & 0 \\ 6 & 1 & 4 \\ 0 & 2 & 3 \end{vmatrix} + 1 \begin{vmatrix} 4 & 1 & 0 \\ 2 & 4 & 2 \\ 0 & 2 & 3 \end{vmatrix} - 1 \begin{vmatrix} 4 & 1 & 0 \\ 2 & 4 & 2 \\ 6 & 1 & 4 \end{vmatrix}.$$

$$\begin{aligned} \begin{vmatrix} 2 & 4 & 2 \\ 6 & 1 & 4 \\ 0 & 2 & 3 \end{vmatrix} &= +2 \begin{vmatrix} 1 & 4 \\ 2 & 3 \end{vmatrix} - 6 \begin{vmatrix} 4 & 2 \\ 2 & 3 \end{vmatrix} + 0 \begin{vmatrix} 4 & 2 \\ 1 & 4 \end{vmatrix} \\ &= 2(3-8) - 6(12-4) + 0(16-2) \\ &= -10 - 48 + 0 = -58. \end{aligned}$$

$$\begin{aligned} \begin{vmatrix} 4 & 1 & 0 \\ 6 & 1 & 4 \\ 0 & 2 & 3 \end{vmatrix} &= +4 \begin{vmatrix} 1 & 4 \\ 2 & 3 \end{vmatrix} - 6 \begin{vmatrix} 1 & 0 \\ 2 & 3 \end{vmatrix} + 0 \begin{vmatrix} 1 & 0 \\ 1 & 4 \end{vmatrix} \\ &= 4(3-8) - 6(3-0) + 0(4-0) \\ &= -20 - 18 + 0 = -38. \end{aligned}$$

$$\begin{vmatrix} 4 & 1 & 0 \\ 2 & 4 & 2 \\ 0 & 2 & 3 \end{vmatrix} = +4 \begin{vmatrix} 4 & 2 \\ 2 & 3 \end{vmatrix} - 2 \begin{vmatrix} 1 & 0 \\ 2 & 3 \end{vmatrix} + 0 \begin{vmatrix} 1 & 0 \\ 4 & 2 \end{vmatrix}$$

$$= 4(12-4) - 2(3-0) + 0(2-0)$$

$$= 32 - 6 + 0 = + 26 .$$

$$\begin{vmatrix} 4 & 1 & 0 \\ 2 & 4 & 2 \\ 6 & 1 & 4 \end{vmatrix} = +4 \begin{vmatrix} 4 & 2 \\ 1 & 4 \end{vmatrix} - 2 \begin{vmatrix} 1 & 0 \\ 1 & 4 \end{vmatrix} + 6 \begin{vmatrix} 1 & 0 \\ 4 & 2 \end{vmatrix}$$

$$= 4(16-2) - 2(4-0) + 6(2-0)$$

$$= + 56 - 8 + 12 = + 60 .$$

Hence

$$|A| = (+2)(-58) - (3)(-38) + (1)(+26) - (1)(+60) = - 36 .$$

For every element  $a_{ij}$  in the square matrix  $A$  there is a corresponding minor  $m_{ij}$  and a corresponding cofactor  $e_{ij}$ . Let  $M$  be the square matrix with elements  $m_{ij}$ ; let  $E$  be the square matrix with elements  $e_{ij}$ ; and let  $F$  be the transpose of  $E$ . Then the square matrix  $F$  is called the *adjoint* of  $A$ . Its elements may be denoted  $f_{ij} = e_{ji}$ .

These definitions are illustrated in the following numerical example:

$$\begin{vmatrix} +2 & -1 & +5 \\ -4 & +8 & -3 \\ +2 & -0 & +7 \end{vmatrix} = A \qquad \begin{vmatrix} 56 & -22 & -16 \\ -7 & 4 & 2 \\ -37 & 14 & 12 \end{vmatrix} = E$$

$$\begin{vmatrix} 56 & 22 & -16 \\ 7 & 4 & -2 \\ -37 & -14 & 12 \end{vmatrix} = M \qquad \begin{vmatrix} 56 & -7 & -37 \\ -22 & 4 & 14 \\ -16 & 2 & 12 \end{vmatrix} = F = \text{adjoint of } A .$$

A square matrix is said to be *symmetric* when  $a_{ij} = a_{ji}$ . It is symmetric about the principal diagonal.

$$\begin{vmatrix} 1 & 4 & 5 \\ 4 & 2 & 8 \\ 5 & 8 & 3 \end{vmatrix} = a \text{ symmetric matrix.}$$

If the matrix is symmetric except that the signs above the principal diagonal are opposite to the signs below the diagonal, then the matrix is said to be *skew symmetric*.

$$\begin{vmatrix} +2 & -3 & +4 \\ +3 & -5 & -5 \\ -4 & +5 & +6 \end{vmatrix} = a \text{ skew symmetric matrix.}$$

If all the principal minors of a matrix are greater than or equal to zero, then the matrix is said to be *positive-definite*. If, in addition, it is symmetric, it is a *Gramian* matrix.

$$\begin{vmatrix} 2 & 3 & -3 \\ 2 & 4 & 2 \\ 3 & 5 & 6 \end{vmatrix} = a \text{ positive definite matrix.}$$

$$\begin{vmatrix} 5 & 10 & 13 \\ 10 & 20 & 26 \\ 13 & 26 & 36 \end{vmatrix} = a \text{ Gramian matrix.}$$

In some problems it is important to know the highest order of the non-vanishing minors. The highest order of the non-vanishing minors is called the *rank* of a matrix. The rank of *Table 5* is equal to its order, namely, 3, because the determinant itself does not vanish. The determinant of *Table 10* does vanish, so that its rank must be less than 3. It contains second-order minors that do not vanish, and the rank of the determinant is therefore 2.

*Table 10*

$$\begin{vmatrix} 10 & 8 & 1 \\ 8 & 8 & 2 \\ 1 & 2 & 1 \end{vmatrix} = 0$$

If the determinant of a matrix is zero, the matrix is said to be *singular*. If the determinant does not vanish, the matrix is said to be *non-singular*.

If  $|a_{ij}| = 0$ , then  $a_{ij}$  is *singular*.

If  $|a_{ij}| \neq 0$ , then  $a_{ij}$  is *non-singular*.

Most of the theorems in the elementary theory of real determinants are concerned with the methods of ascertaining the numerical value of a determinant and with the operations that do, or do not, affect its numerical value. The following theorems are useful in dealing with determinants:

1) The value of a determinant is equal to that of its transpose.

$$\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix};$$

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}.$$

2) If any pair of parallel arrays of a determinant are interchanged, the absolute value of the determinant remains unaltered but the sign is reversed.

$$+ \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = - \begin{vmatrix} a_2 & b_2 & c_2 \\ a_1 & b_1 & c_1 \\ a_3 & b_3 & c_3 \end{vmatrix}.$$

3) If two parallel arrays of a determinant are proportional, the determinant vanishes.

$$\begin{vmatrix} a_1 & ka_1 \\ a_2 & ka_2 \end{vmatrix} = 0; \quad \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ kb_1 & kb_2 & kb_3 \end{vmatrix} = 0.$$

If  $k=1$ , then two arrays are identical and the determinant vanishes.

4) If a determinant has an array of ciphers, it is equal to zero.

$$\begin{vmatrix} a & b & c \\ 0 & 0 & 0 \\ d & e & f \end{vmatrix} = 0.$$

5) If each element of an array is multiplied by any factor, the value of the determinant is multiplied by that factor.

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ kc_1 & kc_2 & kc_3 \end{vmatrix} = k \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}.$$

This theorem is sometimes useful in reducing a determinant to simpler forms.

$$\begin{vmatrix} 6 & 9 & 8 \\ 12 & 18 & 4 \\ 24 & 27 & 2 \end{vmatrix} = 6 \times 9 \times 2 \begin{vmatrix} 1 & 1 & 4 \\ 2 & 2 & 2 \\ 4 & 3 & 1 \end{vmatrix} = 6 \times 9 \times 2 \times 2 \begin{vmatrix} 1 & 1 & 4 \\ 1 & 1 & 1 \\ 4 & 3 & 1 \end{vmatrix} = -648.$$

6) If each element in an array is reversed in sign, the value of the determinant is reversed in sign.

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = - \begin{vmatrix} -a_1 & b_1 & c_1 \\ -a_2 & b_2 & c_2 \\ -a_3 & b_3 & c_3 \end{vmatrix} = + \begin{vmatrix} -a_1 & b_1 & c_1 \\ +a_2 & -b_2 & -c_2 \\ -a_3 & b_3 & c_3 \end{vmatrix}.$$

7) If an array contains no zero elements, all the elements of the array may be made unity by means of multiplying factors.

$$\begin{vmatrix} 3 & 4 & 6 \\ 2 & 8 & 8 \\ 6 & 7 & 9 \end{vmatrix} = 3 \times 4 \times 6 \begin{vmatrix} 1 & 1 & 1 \\ \frac{2}{3} & 2 & \frac{8}{3} \\ 2 & \frac{7}{3} & \frac{8}{3} \end{vmatrix} = 3 \times 4 \begin{vmatrix} 1 & 1 & 1 \\ 4 & 12 & 8 \\ 2 & \frac{7}{3} & \frac{8}{3} \end{vmatrix} \\ = 3 \times 4 \times 4 \begin{vmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 2 & \frac{7}{3} & \frac{8}{3} \end{vmatrix} = 3 \times 4 \begin{vmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 8 & 7 & 6 \end{vmatrix} = -36.$$

8) Every determinant can be expressed as a sum of two determinants.

$$\begin{vmatrix} (a_1+p) & b_1 & c_1 \\ (a_2+q) & b_2 & c_2 \\ (a_3+r) & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} + \begin{vmatrix} p & b_1 & c_1 \\ q & b_2 & c_2 \\ r & b_3 & c_3 \end{vmatrix}$$

9) The value of a determinant remains unaltered if each element in any array is augmented by a multiple of the corresponding element in a parallel array.

$$\begin{vmatrix} (a_1+kb_1) & b_1 & c_1 \\ (a_2+kb_2) & b_2 & c_2 \\ (a_3+kb_3) & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} + k \begin{vmatrix} b_1 & b_1 & c_1 \\ b_2 & b_2 & c_2 \\ b_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.$$

The second determinant in the right member vanishes because two columns are identical.

10) If all the elements of a determinant on one side of the principal diagonal are ciphers, the determinant reduces to the leading term.

$$\begin{vmatrix} 5 & 1 & 2 \\ 6 & 3 & 3 \\ 11 & 3 & 8 \end{vmatrix} = \begin{vmatrix} 5 & 1 & 2 \\ 6 & 3 & 3 \\ 5 & 0 & 5 \end{vmatrix} = \begin{vmatrix} 3 & 1 & 2 \\ 3 & 3 & 3 \\ 0 & 0 & 5 \end{vmatrix} = \begin{vmatrix} 2 & 1 & 2 \\ 0 & 3 & 3 \\ 0 & 0 & 5 \end{vmatrix} = 2 \times 3 \times 5 = +30.$$

### Matrix multiplication

Consider the three simultaneous equations,

$$(2) \quad \begin{cases} y_1 + 2y_2 & = x_1, \\ 3y_2 + y_3 & = x_2, \\ 2y_1 + 2y_2 + 4y_3 & = x_3. \end{cases}$$

The equations (2) are written in the *expanded notation*. If the  $x$ 's are known, the equations may be solved numerically for the  $y$ 's. As an example, let  $x_1=1, x_2=3, x_3=2$ . Solving the equations simultaneously,

$$y_1 = -\frac{1}{4}, \quad y_2 = \frac{1}{4}, \quad y_3 = \frac{6}{4}.$$

In some problems it is convenient, as well as clarifying, to represent a set of simultaneous equations in the *rectangular notation* which is illustrated in *Table 11*. Here the coefficients of the three simultaneous equations have

Table 11

$$\begin{array}{c}
 \begin{array}{ccc}
 & j & \\
 & 1 & 2 & 0 \\
 i & \left\| \begin{array}{c} 1 & 2 & 0 \\ 0 & 3 & 1 \\ 2 & 2 & 4 \end{array} \right\| & \\
 & A & 
 \end{array}
 \times i & \begin{array}{ccc}
 & k & \\
 & 1 & \\
 & 2 & \\
 & 3 & \\
 & y & 
 \end{array}
 = i & \begin{array}{ccc}
 & k & \\
 & 1 & \\
 & 2 & \\
 & 3 & \\
 & x & 
 \end{array}
 \end{array}$$

been arranged in the form of a matrix that may be denoted  $A$ . The  $y$ 's have been arranged in a vertical column in a matrix denoted  $y$ . The  $x$ 's are also arranged in a vertical column matrix denoted  $x$ . The matrix  $A$  is of order  $3 \times 3$ , while the matrices  $x$  and  $y$  are both of order  $3 \times 1$ .

A matrix that consists of a single column will be called a *column vector*. A matrix that consists of a single row will be called a *row vector*. The vectorial terminology is probably due to the fact that the elements in any array may be regarded as the Cartesian co-ordinates of a point in a space of as many dimensions as there are elements in the array. This point, together with the origin, determines a direction in space. In this manner any array of a matrix can be given a vectorial interpretation.

The three matrices  $A$ ,  $y$ ,  $x$ , of *Table 11* may be regarded as symbolizing the simultaneous equations (2). This necessitates that a particular operation be implied by the adjacent matrices  $A$  and  $y$ . In order that these matrices shall symbolize the simultaneous equations, the following rule must be implied in *Table 11*:

The first equation in (2) can be produced from the matrices by writing

$$(3) \quad a_{11}y_{11} + a_{12}y_{21} + a_{13}y_{31} = \sum_{j=1}^3 a_{1j}y_{j1} = x_{11}.$$

In performing this operation with the matrices, a row of the first one is associated with a column of the second one. The first equation of (2) calls for the *cross products* of the corresponding elements of the first row of  $A$  and the first column of  $y$ . (In the present problem,  $y$  has only one column.) The cross product is illustrated by (3).

The second equation (2) is produced by performing the same *row-by-column* multiplication, using the second row of  $A$  and the first column of  $y$ . Then we have

$$(4) \quad a_{21}y_{11} + a_{22}y_{21} + a_{23}y_{31} = \sum_{j=1}^3 a_{2j}y_{j1} = x_{21}.$$

The third equation (2) is produced by a similar operation on the third row of  $A$  and the first column of  $y$ . The sum of the three products is recorded in the third row and first column of  $x$ . The equation is

$$(5) \quad a_{31}y_{11} + a_{32}y_{21} + a_{33}y_{31} = \sum_{j=1}^3 a_{3j}y_{j1} = x_{31} .$$

The three equations may be written in the more condensed form

$$(6) \quad \sum_{j=1}^3 a_{ij}y_{jk} = x_{ik} .$$

This interpretation of two adjacent matrices is called *matrix multiplication*. In the present problem  $k=1$ , because  $y$  has only one column. Since  $i$  can take three different values, namely,  $i=1, 2, 3$ , the equation (6) represents all three of the simultaneous equations in *summational notation*. *Table 11* represents the *rectangular* notation. The three equations (2) may also be represented conveniently in the still more condensed *matrix notation*, namely,

$$(7) \quad Ay = x ,$$

which is a *matrix equation*. The operation specified by this matrix equation is that if the matrix  $A$  is multiplied by the matrix  $y$ , row-by-column, the matrix product is another matrix, namely,  $x$ . This is an exceedingly powerful method of handling sets of equations, because many otherwise tedious numerical operations can be shunted, so that the calculations are performed only on a final set of matrices rather than on many intermediate steps. Still more important is the fact that significant relations in a problem are conspicuous in the matrix notation but they may be obscure when the problem is handled in expanded algebraic or numerical form.

Table 12

$$\begin{array}{c}
 \begin{array}{c} k \\ \parallel \\ y_1 \\ y_2 \\ y_3 \\ y' \end{array} \begin{array}{c} j \\ \hat{y}_2 \\ y_3 \\ j \end{array} \parallel \begin{array}{c} i \\ 1 \\ 2 \\ 3 \\ 0 \\ 1 \\ 4 \\ A' \end{array} \parallel = \begin{array}{c} k \\ \parallel \\ x_1 \\ x_2 \\ x_3 \\ x' \end{array} \begin{array}{c} i \\ x_1 \\ x_2 \\ x_3 \\ x' \end{array} \parallel
 \end{array}$$

The same set of simultaneous equations (2) may be represented by the matrix multiplication shown in *Table 12*. The multiplication of the first row of  $y'$  ( $y'$  has only one row) and the first column of  $A'$  reproduces the first



equation of (2). It should be noted that the matrices of *Table 12* are the transposes of the matrices of *Table 11*. The transpose of a column vector is a row vector with the same elements.

The summational notation for the matrix multiplication of *Table 12* is

$$(8) \quad \sum_{i=1}^3 y_{kj} a_{ji} = x_{ki}.$$

The general element of  $A$  in *Table 11* is  $a_{ij}$ . Hence the general element of  $A'$  in *Table 12* is  $a_{ji}$ . The general element of  $y$  is  $y_{jk}$ , so that the general element of  $y'$  is  $y_{kj}$ . The matrix equation for *Table 12* is

$$(9) \quad y'A' = x',$$

which represents the same set of equations as (7).

In order to multiply one matrix by another, the number of columns of the first one must be the same as the number of rows of the second. The columns of  $A$  in *Table 11* are represented by the subscript  $j$ , and this is also the subscript for the rows of  $y$ . If the subscripts for the first matrix are  $i$  and  $j$  and the subscripts for the second matrix are  $j$  and  $k$ , then the  $j$  subscript is eliminated from the matrix product which has the subscripts  $i$  and  $k$ . The same rule can be verified in the matrix multiplication of *Table 12*, where the subscripts of the first matrix are  $k$  and  $j$  and those of the second matrix are  $j$  and  $i$ . Eliminating the middle subscript  $j$ , which is common, the matrix product has the subscripts  $k$  and  $i$ .

The matrix equations (7) and (9) illustrate the following matrix theorem:

*Theorem: The transpose of any product of matrices is the product of their transposes in reverse order.*

Hence, if  $AB=C$ , it follows that  $B'A'=C'$ . Applying this theorem to the present example, we have, by *Table 11*,  $Ay=x$  and, by the theorem,  $y'A'=x'$ , which is the matrix equation for *Table 12*.

If the  $x$ 's are known in (2), then the  $y$ 's may be found. Let the  $y$ 's be expressed as linear functions of the  $z$ 's in (10).

$$(10) \quad \begin{cases} z_1 & + z_3 & = & y_1, \\ & 2z_2 & + z_3 & = & y_2, \\ z_1 & + 2z_2 & & = & y_3. \end{cases}$$

This set of simultaneous equations is represented in *Table 13*. If the three

*Table 13*

$$\begin{array}{c} \left\| \begin{array}{ccc} 1 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 2 & 0 \end{array} \right\| \cdot \left\| \begin{array}{c} z_1 \\ z_2 \\ z_3 \end{array} \right\| = \left\| \begin{array}{c} y_1 \\ y_2 \\ y_3 \end{array} \right\| \\ B \qquad \qquad \qquad z \qquad \qquad \qquad y \end{array}$$

matrices of *Table 13* are denoted  $B$ ,  $z$ , and  $y$ , we can represent the three equations (10) in the single matrix equation,

$$(11) \qquad \qquad \qquad Bz = y .$$

Since the  $y$ 's are known, the values of the  $z$ 's can be determined. Substituting the known values of the  $y$ 's in (10), we find that

$$z_1 = -\frac{3}{14}, \qquad z_2 = \frac{7}{14}, \qquad z_3 = -\frac{3}{14} .$$

Equation (7) shows that the  $x$ 's can be expressed linearly in terms of the  $y$ 's. Equation (11) shows that the  $y$ 's can be expressed linearly in terms of the  $z$ 's. It is desired now to express the  $x$ 's directly in terms of the  $z$ 's without the intermediate  $y$ 's. This can be done. From the equations

$$(7) \qquad \qquad \qquad Ay = x ,$$

$$(11) \qquad \qquad \qquad Bz = y ,$$

it follows that

$$A(Bz) = x ,$$

$$ABz = x .$$

Let  $AB=C$ . Then

$$(12) \qquad \qquad \qquad Cz = x .$$

In order to express the  $x$ 's in terms of the  $z$ 's, the matrix product  $AB=C$  must be determined numerically. This matrix product is shown graphically in *Table 14*. Consider the first row of  $A$  and the first column of  $B$ . The cross product is

$$(1)(1) + (2)(0) + (0)(1) = +1 .$$

This is therefore the element in the first row and the first column of the matrix product  $C$ . Consider, as another example, the second row of  $A$  and the third column of  $B$ . The cross product is

$$(0)(1) + (3)(1) + (1)(0) = +3.$$

This is the element  $c_{23}$  in the matrix  $C$ .

Table 14

$$\begin{array}{c} \left\| \begin{array}{ccc} 1 & 2 & 0 \\ 0 & 3 & 1 \\ 2 & 2 & 4 \end{array} \right\| \\ A \end{array} \cdot \begin{array}{c} \left\| \begin{array}{ccc} 1 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 2 & 0 \end{array} \right\| \\ B \end{array} = \begin{array}{c} \left\| \begin{array}{ccc} 1 & 4 & 3 \\ 1 & 8 & 3 \\ 6 & 12 & 4 \end{array} \right\| \\ C \end{array}$$

Since the numerical values of the  $x$ 's, the  $y$ 's, and the  $z$ 's are known, the matrix equation (12) may be tested graphically, as shown in *Table 15*. As a

Table 15

$$\begin{array}{c} \left\| \begin{array}{ccc} 1 & 4 & 3 \\ 1 & 8 & 3 \\ 6 & 12 & 4 \end{array} \right\| \\ C \end{array} \cdot \begin{array}{c} \left\| \begin{array}{c} -\frac{8}{14} \\ \frac{7}{14} \\ -\frac{9}{14} \end{array} \right\| \\ z \end{array} = \begin{array}{c} \left\| \begin{array}{c} 1 \\ 3 \\ 2 \end{array} \right\| \\ x \end{array}$$

sample check, consider the second row of  $C$  and the first column of  $z$ . It should reproduce the value  $x_{21}=3$ .

$$(1)(-\frac{8}{14}) + (8)(\frac{7}{14}) + (3)(-\frac{9}{14}) = +3.$$

*Table 14* shows the matrix product  $AB=C$ . If the order of the matrices  $A$  and  $B$  is interchanged in this multiplication, a different product is obtained. This is readily verified numerically in *Table 14*; and it illustrates the principle that if  $AB=C$ , then, in general,  $BA \neq C$ . Matrix multiplication is not commutative. In matrix algebra it is essential to note the order of the matrix factors because the order is not arbitrary, as in ordinary algebra, where  $ab=ba$ .

The following is an example of matrix algebra. If, instead of (7) and (11), the transposed forms of these equations were used, we should have

$$(9) \quad y'A' = x',$$

$$(13) \quad z'B' = y'.$$

Substituting (13) in (9),

$$(14) \quad z'B'A' = x'.$$

But

$$(15) \quad AB \equiv C.$$

Hence

$$(16) \quad B'A' = C'.$$

Substituting (16) in (14),

$$(17) \quad z'C' = x',$$

which could also be written directly as the transposed form of (12).

In order that there shall be a unique solution for the simultaneous equations (2), the matrix  $A$  of the coefficients must be non-singular, i.e.,  $|A| \neq 0$ . This may be tested by trying to solve a set of non-homogeneous simultaneous equations with coefficients whose determinant does vanish.

The multiplication of matrices is associative. This is illustrated as follows:

$$(AB)C = A(BC) = ABC.$$

The matrix product  $(AB)$  may be determined and then *postmultiplied* by  $C$ , or the matrix product  $(BC)$  may be determined and then *premultiplied* by  $A$ . The product is the same. This principle can be extended to any number of matrix factors. For example,

$$(ABC)D = (AB)(CD) = A(BCD) = ABCD.$$

Note that the order of the matrix factors is retained.

The sum, or difference, of two  $m \times n$  matrices is the  $m \times n$  matrix each of whose elements is the sum, or difference, of the corresponding elements in the given matrices.

$$\left\| \begin{array}{cc} 1 & 2 \\ 3 & 4 \end{array} \right\| + \left\| \begin{array}{cc} 2 & 3 \\ 4 & 5 \end{array} \right\| = \left\| \begin{array}{cc} 3 & 5 \\ 7 & 9 \end{array} \right\|.$$

The components may be written in any order.

$$A + B = B + A ;$$

$$(A + B) + C = A + (B + C) = A + B + C .$$

If  $k$  and  $m$  are scalars, then

$$kA + kB = k(A + B) ;$$

$$kA + mA = (k + m)A .$$

The multiplication of matrices is distributive.

$$A(B + C) = AB + AC ;$$

$$(B + C)A = BA + CA .$$

It can be shown that *the rank of a matrix product cannot exceed the lowest rank of any of the factors*. Thus, if the ranks of matrices  $A$ ,  $B$ , and  $C$  are 2, 4, and 3, respectively, then the rank of the matrix product  $ABC$  cannot exceed 2.

It is sometimes useful to know that *the determinant of the product of two square matrices is equal to the product of their determinants*. The following is an example:

$$\text{Let } |A| = \begin{vmatrix} 2 & 6 \\ 4 & 7 \end{vmatrix} = -10, \quad \text{and let } |B| = \begin{vmatrix} 2 & 1 \\ 3 & 4 \end{vmatrix} = +5 .$$

Then

$$|A| \cdot |B| = |AB| = \begin{vmatrix} 22 & 26 \\ 29 & 32 \end{vmatrix} = -50 = (-10)(+5) .$$

In the matrix product  $AB$ , the matrix  $B$  is said to be *premultiplied* by the matrix  $A$ , or the matrix  $A$  is said to be *postmultiplied* by the matrix  $B$ .

The operation of multiplying one matrix by another can be summarized for mnemonic purposes in the diagram of *Figure 1*. This diagram shows that rows of the first matrix are associated with columns of the second matrix and that the middle subscript is eliminated in the product. If the  $i$ th row of  $A$  is cross multiplied with the  $k$ th column of  $B$ , the cross product is recorded in the cell  $ik$  of  $C$ .

There is nothing magical or profound in the particular rules of matrix multiplication that have become conventional. The *row-by-column* rule is

entirely arbitrary. It would have been possible to set up a column-by-row rule provided that the matrices had been so arranged that the rule would have reproduced the original equations which the matrix notation represent-

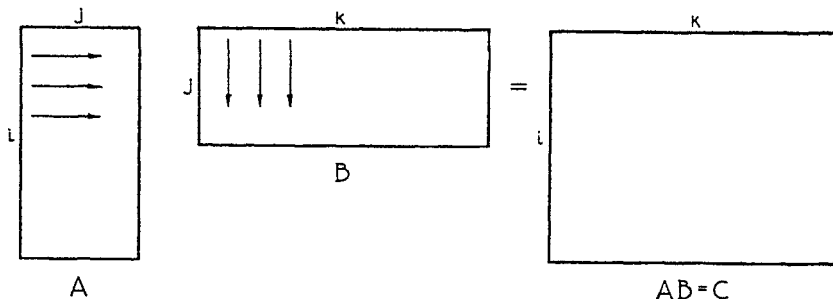


FIGURE 1

ed. It would also have been possible to have a notation which implied that one matrix was on top of another, but this would not have been so convenient for writing habits that go from left to right.

**Diagonal matrices**

In the manipulation of systems of equations there occurs frequently a type of matrix in which all of the elements are zero except the diagonal elements. A matrix in which only the elements of the principal diagonal are non-vanishing is called a *diagonal matrix*. The following is a diagonal matrix of order 4:

$$\begin{vmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & 0 & d \end{vmatrix} = D = \text{a diagonal matrix.}$$

It sometimes happens that all of the elements of a diagonal matrix are identical. Such a matrix is called a *scalar matrix*. The following is an example:

$$\begin{vmatrix} k & 0 & 0 & 0 \\ 0 & k & 0 & 0 \\ 0 & 0 & k & 0 \\ 0 & 0 & 0 & k \end{vmatrix} = K = \text{a scalar matrix.}$$

When a diagonal matrix has unity in each diagonal cell, it is called a *unit matrix* or the *identity matrix*. The following is an identity matrix of order 4:

$$\begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} = I = \text{the identity matrix.}$$

The properties of diagonal matrices are very useful in handling sets of linear equations:

1) Premultiplication  $DA$  with a diagonal matrix  $D$  multiplies each *row* of  $A$  by the corresponding element in  $D$ .

2) Postmultiplication  $AD$  with a diagonal matrix  $D$  multiplies each *column* of  $A$  by the corresponding element in  $D$ .

$$\begin{vmatrix} p & 0 \\ 0 & q \end{vmatrix} \begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} ap & bp \\ cq & dq \end{vmatrix};$$

$D \qquad A \qquad DA$

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} \begin{vmatrix} p & 0 \\ 0 & q \end{vmatrix} = \begin{vmatrix} ap & bq \\ cp & dq \end{vmatrix}.$$

$A \qquad D \qquad AD$

3) Premultiplication or postmultiplication with a scalar matrix  $K$  multiplies all elements of  $A$  by the constant element of  $K$ . This is a special case of the first two theorems. The reason why premultiplication with a scalar matrix has the same effect as postmultiplication is that if every row is multiplied by a constant  $p$ , the effect is the same as if every column is multiplied by the constant  $p$ . In either case every element of  $A$  is multiplied by  $p$ .

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} \begin{vmatrix} p & 0 \\ 0 & p \end{vmatrix} = \begin{vmatrix} ap & bp \\ cp & dp \end{vmatrix};$$

$A \qquad K \qquad AK$

$$\begin{vmatrix} p & 0 \\ 0 & p \end{vmatrix} \begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} ap & bp \\ cp & dp \end{vmatrix};$$

$K \qquad A \qquad KA$

$$\begin{vmatrix} p & 0 \\ 0 & p \end{vmatrix} \begin{vmatrix} a & b \\ c & d \end{vmatrix} = p \begin{vmatrix} a & b \\ c & d \end{vmatrix}.$$

$K \qquad A \qquad pA$

Since the effect of a scalar matrix is independent of its position before or after the other matrices in a matrix product, its constant element  $p$  can be used in the product instead of the scalar matrix  $K$  as shown in the third example. This illustrates the following theorem:

4) If  $K$  is a scalar matrix, then its constant element  $p$  may be substituted for the scalar matrix in a product.

$$AK = KA = pA = Ap .$$

A multiplier which is independent of the non-commutative rule of matrix algebra is called a *scalar*.

5) To multiply a matrix  $A$  by a scalar  $p$ , in either order,  $pA$  or  $Ap$ , is to multiply each element of  $A$  by  $p$ .

The identity matrix is a special case of the scalar matrix, and hence it is also independent of the non-commutative rule of matrix multiplication.

$$\begin{array}{c} \left\| \begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right\| \\ I \end{array} \cdot \begin{array}{c} \left\| \begin{array}{cc} a & b \\ c & d \end{array} \right\| \\ A \end{array} = \begin{array}{c} \left\| \begin{array}{cc} a & b \\ c & d \end{array} \right\| \\ IA = A \end{array} ;$$

$$\begin{array}{c} \left\| \begin{array}{cc} a & b \\ c & d \end{array} \right\| \\ A \end{array} \cdot \begin{array}{c} \left\| \begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right\| \\ I \end{array} = \begin{array}{c} \left\| \begin{array}{cc} a & b \\ c & d \end{array} \right\| \\ AI = A \end{array} .$$

6) To multiply a matrix  $A$  by the identity matrix, in either order,  $AI$  or  $IA$ , is to reproduce the matrix  $A$  unaltered.

$$AI = IA = A .$$

The identity matrix  $I$  in matrix algebra corresponds to unity in ordinary algebra. Hence the identity matrix is suppressed, just as unity is suppressed in ordinary algebra.

$$1 \times 5 = 5 ,$$

$$1 \times x = x ,$$

$$IA = A .$$

### The inverse

If a matrix  $A$  is non-singular, i.e.,  $|A| \neq 0$ , then there exists another unique matrix such that its multiplication by  $A$  produces the identity matrix. This other matrix is called the *inverse* of  $A$ , and it is denoted  $A^{-1}$ . Hence, if  $A$  is non-singular,

$$AA^{-1} = I .$$



The inverse of  $A^{-1}$  is  $A$ , so that

$$AA^{-1} = I = A^{-1}A .$$

Consider the ordinary algebraic equation,

$$ax = y .$$

If it is desired to state  $x$  explicitly, the equation is ordinarily written as

$$x = \frac{1}{a} y .$$

If the given matrix equation is

$$AB = C$$

and if it is desired to write it explicitly for  $B$ , this cannot be accomplished by ordinary division. A matrix is not a number but a rectangular table of numbers. There is an operation in matrix algebra which corresponds to division in ordinary algebra. If both members of the matrix equation  $AB = C$  are premultiplied by  $A^{-1}$ , we have

$$A^{-1}AB = A^{-1}C .$$

But

$$A^{-1}A = I .$$

Hence

$$IB = A^{-1}C ,$$

or

$$B = A^{-1}C .$$

This is the desired form. This example illustrates the operation in matrix algebra which corresponds to division in ordinary algebra. It consists in moving a premultiplying or postmultiplying factor from one member of the equation to the other member in the same relative position. This operation is illustrated in the following examples:

If

$$ABC = M ,$$

then

$$BC = A^{-1}M ,$$

$$C = B^{-1}A^{-1}M ,$$

$$CM^{-1} = B^{-1}A^{-1} ,$$

$$CM^{-1}A = B^{-1} ,$$

$$M^{-1}A = C^{-1}B^{-1} .$$

Since this operation is analogous to ordinary division, the inverse of  $A$  is sometimes called the *reciprocal* of  $A$ .

*The inverse of any product of matrices is the product of their inverses in reverse order.*

$$\begin{aligned} \text{Let} \quad ABC &= M, \\ BC &= A^{-1}M, \\ C &= B^{-1}A^{-1}M, \\ I &= C^{-1}B^{-1}A^{-1}M, \\ M^{-1} &= C^{-1}B^{-1}A^{-1}. \end{aligned}$$

But

$$(ABC)^{-1} = M^{-1}.$$

Hence

$$(ABC)^{-1} = C^{-1}B^{-1}A^{-1}.$$

A method of writing the inverse of a given matrix is as follows: Let the given matrix be  $A$  with elements  $a_{ij}$ .

1) Write the matrix  $M$  with elements  $m_{ij}$  which are the first minors of the elements  $a_{ij}$ ;

2) Reverse the signs of alternate elements of  $M$  so that it becomes the matrix  $E$  with elements  $e_{ij} = (-1)^{i+j}m_{ij}$ ;

3) Write the transpose of  $E$ , namely,  $E' = F$ , with elements  $f_{ij} = e_{ji}$ . The matrix  $F$  is the adjoint of  $A$ ;

4) Divide each element of  $F$  by the value of the determinant  $|A|$ . This is the inverse  $A^{-1}$  with elements

$$g_{ij} = \frac{f_{ji}}{|A|}.$$

The writing of an inverse will be illustrated by the numerical example of equations (2). The given matrix equation is

$$Ay = x.$$

It is desired to find the inverse of  $A$  so that the equation

$$y = A^{-1}x$$

may be written in numerical form.

$$\begin{vmatrix} 1 & 2 & 0 \\ 0 & 3 & 1 \\ 2 & 2 & 4 \end{vmatrix} = A ; \quad |A| = +14 .$$

$$\begin{vmatrix} 10 & -2 & -6 \\ 8 & 4 & -2 \\ 2 & 1 & 3 \end{vmatrix} = M ,$$

$$\begin{vmatrix} 10 & 2 & -6 \\ -8 & 4 & 2 \\ 2 & -1 & 3 \end{vmatrix} = E ,$$

$$\begin{vmatrix} 10 & -8 & 2 \\ 2 & 4 & -1 \\ -6 & 2 & 3 \end{vmatrix} = F ,$$

$$\begin{vmatrix} \frac{1}{14} & -\frac{3}{14} & \frac{2}{14} \\ \frac{2}{14} & \frac{4}{14} & -\frac{1}{14} \\ -\frac{6}{14} & \frac{2}{14} & \frac{3}{14} \end{vmatrix} = A^{-1} ; \quad |A^{-1}| = +\frac{1}{14}$$

It is of interest to verify numerically the matrix equation  $y = A^{-1}x$ . It is written in rectangular notation in *Table 16*.

*Table 16*

$$\begin{vmatrix} \frac{1}{14} & -\frac{3}{14} & \frac{2}{14} \\ \frac{2}{14} & \frac{4}{14} & -\frac{1}{14} \\ -\frac{6}{14} & \frac{2}{14} & \frac{3}{14} \end{vmatrix} \cdot \begin{vmatrix} 1 \\ 3 \\ 2 \end{vmatrix} = \begin{vmatrix} -\frac{10}{14} \\ \frac{12}{14} \\ \frac{6}{14} \end{vmatrix}$$

$A^{-1} \qquad \qquad \qquad x \qquad \qquad \qquad y$

### The characteristic equation

The characteristic equation is of considerable theoretical interest in factor analysis, and it appears in several of the fundamental factor problems. For this reason it is described in this introduction. In a more complete didactic presentation of this subject, the characteristic equation should be introduced with some geometric and other interpretation, so that the sig-

nificance of this equation might be apparent. The relation of the characteristic equation to the problems of factor theory will appear in later chapters.

If a constant  $\beta$  is added explicitly to each diagonal element of a square matrix  $A$ , the resulting matrix is called the *characteristic matrix* of  $A$ . It is illustrated as follows:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + \beta \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = \begin{vmatrix} (a_{11}+\beta) & a_{12} & a_{13} \\ a_{21} & (a_{22}+\beta) & a_{23} \\ a_{31} & a_{32} & (a_{33}+\beta) \end{vmatrix};$$

$$A + \beta I = (A + \beta I)$$

*Characteristic matrix of A.*

The determinant of the characteristic matrix is the *characteristic determinant* of  $A$ .

The expansion of a characteristic determinant of order  $r$  is a polynomial of degree  $r$ . When this polynomial is set equal to zero, the equation is called the *characteristic equation* of  $A$ . An example is the following equation:

$$\begin{vmatrix} (1+\beta) & 0 & 1 \\ 1 & (2+\beta) & 1 \\ 1 & 0 & (2+\beta) \end{vmatrix} = 0.$$

When the determinant is expanded, the characteristic equation becomes

$$\beta^3 + 5\beta^2 + 7\beta + 2 = 0.$$

The coefficients of the expansion of a characteristic determinant can be written in terms of the principal minors without expanding the whole determinant. Let the characteristic equation be as follows:

$$(18) \quad \beta^r + m_1\beta^{r-1} + m_2\beta^{r-2} + \dots + m_r = 0.$$

Then the coefficient  $m_x$  is the sum of all the  $x$ -rowed principal minors in  $A$ . The coefficient  $m_1$  is the sum of all the 1-rowed principal minors of  $A$ . These are 1, 2, 2, and the sum is +5. The coefficient  $m_2$  is the sum of all the 2-rowed principal minors of  $A$ . These are

$$\begin{vmatrix} 1 & 0 \\ 1 & 2 \end{vmatrix} = +2; \quad \begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix} = +1; \quad \begin{vmatrix} 2 & 1 \\ 0 & 2 \end{vmatrix} = +4.$$

The coefficient  $m_2 = 2 + 1 + 4 = +7$ . The coefficient  $m_3$  is the sum of all principal minors of order 3. This is the determinant  $|A|$  itself and its value is  $+2$ . The coefficient of the highest power of  $\beta$  is unity. These coefficients can be verified by expanding the determinant.

### The summational notation

If a set of  $n$  numbers is to be summed, the operation may be indicated in the *expanded* form,

$$(19) \quad x_1 + x_2 + x_3 + \cdots + x_n = y .$$

The same operation may be indicated in the more condensed summational form, namely,

$$(20) \quad \sum_{i=1}^n x_i = y ,$$

in which the subscript  $i$  takes the successive integral values from 1 to  $n$ , inclusive.

In statistical work it is pedantic to indicate the limits, because the summation is over the entire population except in rare cases, which can be specially indicated. It is acceptable practice in statistical work to write  $\Sigma x$  without subscripts when the usual form of summation over the population is implied. In factor analysis this simple and convenient notation becomes ambiguous because summation may be over the factors, over the population, or over the variables. It is therefore advisable to adopt the unambiguous double-subscript notation that is conventional in mathematics.

As an example, the sum of the elements in the first row of *Table 2* can be written in the form

$$a_{11} + a_{12} + a_{13} + \cdots + a_{1n} = \sum_{j=1}^n a_{1j} .$$

Here it is the second subscript  $j$ , representing the columns, which is found in the summation sign. This means that the  $a$ 's are to be summed for all values of  $j$  from 1 to  $n$ . The first subscript is fixed. The summation is thereby confined to one row.

The notation can be generalized to represent any row  $i$ . It then takes the form

$$a_{i1} + a_{i2} + a_{i3} + \cdots + a_{in} = \sum_{j=1}^n a_{ij} .$$

This notation means that the  $a$ 's are to be summed for a fixed value of  $i$ , since  $i$  does not occur in the summation. The  $a$ 's are to be summed in one row  $i$  for all the column values of  $j$  from 1 to  $n$ .

By analogy, the sum of all the elements in a column  $j$  of *Table 2* may be represented as follows:

$$a_{1j} + a_{2j} + a_{3j} + \dots + a_{mj} = \sum_{i=1}^m a_{ij} .$$

Here it is the column  $j$  that is fixed because it does not occur in the summation sign. The  $a$ 's are to be summed in some specified column  $j$  for all values of  $i$  from 1 to  $m$ .

If it is desired to designate the sum of all the elements in the matrix, each of the  $m$  row-sums must be summed. This involves a summation over both  $i$  and  $j$ . We then have

$$\text{Sum of all elements in } A = \sum_{i=1}^m \sum_{j=1}^n a_{ij} .$$

Since it does not matter whether the rows are summed before the columns, or vice versa, we have

$$\sum_{j=1}^n \sum_{i=1}^m a_{ij} = \sum_{i=1}^m \sum_{j=1}^n a_{ij} .$$

A matrix multiplication can also be designated by the summational notation. Consider the matrix multiplication  $AB=C$  of *Table 17*.

*Table 17*

$$\begin{array}{ccc}
 & \begin{array}{c} j(n) \\ \left\| \begin{array}{ccc} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{array} \right\| \\ i(m) \\ A \end{array} & \begin{array}{c} k(p) \\ \left\| \begin{array}{ccc} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{array} \right\| \\ j(n) \\ B \end{array} & = i(m) \begin{array}{c} k(p) \\ \left\| \begin{array}{ccc} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \\ c_{41} & c_{42} & c_{43} \end{array} \right\| \\ AB=C \end{array}
 \end{array}$$

Let  $A$  be a matrix of order  $m \times n$  with general element  $a_{ij}$ . Let  $B$  be a matrix of order  $n \times p$  with general element  $b_{jk}$ . Then the product  $AB=C$  must be a matrix of order  $m \times p$  with the general element  $c_{ik}$ . The middle subscript  $j$  for the general element disappears in the product, and so does the middle dimension  $n$  in the product  $(mn)(np) = (mp)$ . In the example,  $m=4$ ,  $n=3$ ,  $p=3$ .

The element  $c_{11}$  is obtained by the cross multiplication of the first row of  $A$  and the first column of  $B$ . In summational notation,

$$(21) \quad a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} = \sum_{j=1}^n a_{1j}b_{j1} = c_{11} .$$

The second row of  $A$  and the first column of  $B$ :

$$(22) \quad a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} = \sum_{j=1}^n a_{2j}b_{j1} = c_{21} .$$

The  $i$ th row of  $A$  and the first column of  $B$ :

$$(23) \quad a_{i1}b_{11} + a_{i2}b_{21} + a_{i3}b_{31} = \sum_{j=1}^n a_{ij}b_{j1} = c_{i1} .$$

The  $i$ th row of  $A$  and the  $k$ th column of  $B$ :

$$(24) \quad a_{i1}b_{1k} + a_{i2}b_{2k} + a_{i3}b_{3k} = \sum_{j=1}^n a_{ij}b_{jk} = c_{ik} .$$

The summation of (24) gives a single element in  $C$ . If it is desired to indicate the sum of all the elements in the  $i$ th row of  $C$ , the summation will be over  $k$  with a fixed value for  $i$ . We have then

$$(25) \quad \sum_{k=1}^p \sum_{j=1}^n a_{ij}b_{jk} = \sum_{k=1}^p c_{ik} .$$

Finally, if the sum of all the elements in  $C$  is to be indicated, the summation must also be made for all rows. Then

$$(26) \quad \sum_{i=1}^m \sum_{k=1}^p \sum_{j=1}^n a_{ij}b_{jk} = \sum_{i=1}^m \sum_{k=1}^p c_{ik} .$$

The summation of (24) represents the multiplication of two matrices,  $a_{ij}$  and  $b_{jk}$ , whose product is the matrix  $c_{ik}$ . It can be visualized in *Figure 1*.

As an example of manipulation with the summational notation, consider the first term of equation (12-i). It is

$$\frac{1}{N} \sum_{i=1}^N \sum_{m=1}^r a_{jm}^2 x_{mi}^2,$$

in which it is desired to substitute

$$\frac{1}{N} \sum_{i=1}^N x_{mi}^2 = 1,$$

in order to simplify the first term.

Since the order of summation is arbitrary, the order of the summations may be interchanged. Then the first term becomes

$$\frac{1}{N} \sum_{m=1}^r \sum_{i=1}^N a_{jm}^2 x_{mi}^2.$$

Since the subscript  $i$  does not occur in  $a_{jm}^2$ , this factor is a constant during the summation over  $i$ . Hence it may be placed in front of the summation over  $i$  without altering the value of the first term, which then becomes

$$\frac{1}{N} \sum_{m=1}^r a_{jm}^2 \sum_{i=1}^N x_{mi}^2.$$

But the reciprocal of  $N$  is a scalar, and so it can be placed anywhere in the summation. Changing its relative position, the first term becomes

$$\sum_{m=1}^r a_{jm}^2 \frac{1}{N} \sum_{i=1}^N x_{mi}^2,$$

and now the substitution can be made more clearly. Suppressing the part which is equal to unity, the first term simplifies into

$$\sum_{m=1}^r a_{jm}^2,$$

as it occurs in equation (18-i). These steps are more explicit than will ordinarily be found necessary, but they illustrate further the manner in which the summational notation can be handled.



### Linear dependence

A matrix of order  $m \times n$  may be regarded as  $m$  sets of numbers with  $n$  numbers in each set. Each row of numbers is then a set. *Table 18* is a matrix

*Table 18*

$$\left\| \begin{array}{cccccc} 2 & 3 & 5 & 1 & 4 & 2 \\ 4 & 6 & 10 & 2 & 8 & 4 \\ 1 & \frac{2}{3} & \frac{5}{3} & \frac{1}{3} & 2 & 1 \\ 6 & 9 & 15 & 3 & 12 & 6 \end{array} \right\|$$

of order  $4 \times 6$ . In this table every row can be expressed linearly in terms of the first row. By this is meant that for any row  $i$  there exists a constant  $c_i$ , such that

$$(27) \quad a_{ij} = c_i a_{1j} .$$

For the fourth row the constant  $c_i$  is 3, so that each element in the fourth row is three times the corresponding element in the first row. When any row can be so expressed in terms of the first row, the rows are *proportional*, and it can be shown that the columns are then also proportional. If two rows are not proportional, they are said to be *linearly independent*, for one of them cannot be expressed linearly in terms of the other. When each row can be expressed linearly in terms of one row, the rank of the matrix is 1. This means that all second-order minors vanish. This fact is readily verified in *Table 18*.

The idea of proportionality can be generalized to two or more dimensions. An example of rank 2 is shown in *Table 19*. In this matrix any row can be

*Table 19*

$$\left\| \begin{array}{cccccc} 7 & 5 & 4 & 6 & 7 & 2 \\ 2 & 2 & 0 & 4 & 2 & 4 \\ 5 & 4 & 2 & 6 & 5 & 4 \\ 9 & 7 & 4 & 10 & 9 & 6 \end{array} \right\|$$

expressed as a linear function of any two rows. In this particular matrix there are no two dependent rows. This requires that two constants  $c_1$  and  $c_2$  (not both zero) exist such that

$$(28) \quad a_{ij} = c_1 a_{1j} + c_2 a_{2j} ,$$

where the elements in the  $i$ th row are expressed linearly in terms of the first two rows.

In the following example the two constants for the third row of *Table 19* are determined. For the first two entries in the third row,

$$(29) \quad \begin{cases} 5 = 7c_1 + 2c_2, \\ 4 = 5c_1 + 2c_2. \end{cases}$$

Solving (29) simultaneously, we find that  $c_1 = 1/2$  and  $c_2 = 3/4$ . Testing this on the last column, as an example,

$$(2)\left(\frac{1}{2}\right) + (4)\left(\frac{3}{4}\right) = 4.$$

A different set of constants must be determined for each successive independent row.

Since each of the rows in *Table 19* can be expressed linearly in terms of the first two rows, it can be shown that the matrix of *Table 19* is of rank 2. This implies that all third-order minors vanish. As an example, the following third-order minor of *Table 19* vanishes.

$$\begin{vmatrix} 5 & 6 & 7 \\ 4 & 6 & 5 \\ 7 & 10 & 9 \end{vmatrix} = 0.$$

It can be shown that if the rank of a matrix is  $r$ , then there exists a set of  $r$  columns, or rows, in terms of which each column, or row, can be linearly expressed.

**Geometric interpretations**

The most frequent form of equation for a straight line in a plane is probably

$$(30) \quad y = mx + p,$$

in which  $x$  and  $y$  are the two variables while  $m$  and  $p$  are two independent parameters. This agrees with the well-known fact that any two points determine a straight line. The multiplying constant  $m$  is the slope, and the additive constant  $p$  is the  $y$ -intercept. In the present context it will be more useful to begin with the equation of a straight line in the more general form

$$(31) \quad a_1x_1 + a_2x_2 + k = 0.$$

This equation has two variables,  $x_1$  and  $x_2$ , and three parameters,  $a_1$ ,  $a_2$ ,  $k$ . Since only two points are needed to determine the line, it follows that the three parameters are not independent.

Equation (31) can evidently be multiplied by any arbitrary constant without affecting its geometrical representation in the plane. Then (31) becomes

$$(32) \quad ca_1x_1 + ca_2x_2 + ck = 0.$$

Let the multiplier  $c$  be so chosen that *the sum of the squares of the coefficients of the variables is equal to unity*. Then

$$(33) \quad (ca_1)^2 + (ca_2)^2 = 1,$$

or

$$(34) \quad c = \frac{1}{\sqrt{a_1^2 + a_2^2}}.$$

Let  $ca_1 = \lambda_1$ ;  $ca_2 = \lambda_2$ ;  $ck = d$ . Then

$$(35) \quad \lambda_1x_1 + \lambda_2x_2 + d = 0,$$

where

$$(36) \quad \lambda_1^2 + \lambda_2^2 = 1.$$

When the equation is written with this adjustment, it is said to be in *normal form*. This definition is applied not only to the equation of a line in a plane, and to the equation of a plane in a space of three dimensions, but also to the equation of a hyperplane of  $(n-1)$  dimensions in a space of  $n$  dimensions. The number of dimensions of the space defined by equation (35) is  $(n-1)$  where  $n$  is the number of variables. Hence equation (35) defines a space of one dimension, a line, in a space of  $n=2$  dimensions, a plane.

When a linear equation is in the normal form, the parameters have interesting meaning. The parameters  $\lambda_1$  and  $\lambda_2$  are the direction cosines of the normal to the linear space which is defined by equation (35); and the parameter  $d$  is the distance from the origin to the same linear space. The normal to the line makes  $\cos^{-1}\lambda_1$  with the  $x_1$ -axis and  $\cos^{-1}\lambda_2$  with the  $x_2$ -axis. *The direction cosines of a space are the cosines of the angles that its normal makes with the Cartesian co-ordinate axes*. In the present case the space is a one-dimensional space, namely, that which is defined by equation (35). In order to avoid ambiguity, the normal is taken positive on the side which contains the origin. Equation (35) may be interpreted geometrically as defining a space of one dimension whose normal has the direction cosines  $\lambda_1$  and  $\lambda_2$  and which is distant  $d$  from the origin.

If the parameter  $d$  vanishes, then the space which is defined by the equation contains the origin. In equation (35) the line contains the origin if  $d$  is zero.

Equation (35) locates a one-dimensional space (a line) in a two-dimensional space (a plane). If a new variable,  $x_3$ , is added, the equation takes the form

$$(37) \quad \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3 + d = 0 .$$

It defines a space of two dimensions (a plane) in a space of three dimensions with three orthogonal axes. Here, as before, if the equation is in normal form, then  $d$  is the distance of the plane from the origin, and the three coefficients  $\lambda_1, \lambda_2, \lambda_3$ , are the direction cosines of the normal to the plane. They are the cosines of the angles that the normal makes with intersecting lines that are parallel to the  $x_1, x_2$ , and  $x_3$  axes, respectively.

The direction cosines have the property that the sum of their squares is unity. In equation (35) the line is defined if the parameter  $d$  and one of the direction cosines are given. In equation (37) the plane is defined by its distance from the origin and any two of its direction cosines. The third direction cosine can be found from the fact that the sum of the squares of the direction cosines equals unity. If  $d$  vanishes in (37), the plane contains the origin. A plane through the origin is therefore defined by its direction cosines, which are the direction cosines of its normal.

An equation of the same form in  $n$  variables is

$$(38) \quad \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3 + \dots + \lambda_n x_n + d = 0 .$$

It defines a hyperplane of  $(n-1)$  dimensions in a space of  $n$  dimensions. If (38) is in normal form,  $d$  is the distance of the hyperplane from the origin. The  $\lambda$ 's are the direction cosines of the normal to the hyperplane and hence

$$(39) \quad \sum_{i=1}^n \lambda_i^2 = + 1 .$$

In the present factor analysis the hyperplanes of primary interest contain the origin, so that the parameter  $d$  vanishes. Then

$$(40) \quad \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3 + \dots + \lambda_n x_n = 0 .$$

The  $n$  values of  $\lambda_i$  are said to be the direction cosines of the hyperplane  $L$  which is defined by its normal  $\Lambda$ .

A matrix may be given a geometric interpretation. Let the matrix  $A$  be of order  $m \times n$ . Then the  $n$  elements of each row may be regarded as the Cartesian co-ordinates of a point in  $n$  dimensions. Since there are  $m$  rows, the matrix may be thought of as defining the positions of  $m$  points, one for each row, in a space of  $n$  dimensions. *Table 20* is of order  $6 \times 3$ . It can therefore be regarded as defining the positions of six points in a space of three dimensions.

*Table 20*

$$\begin{array}{|c|c|c|} \hline 1 & 2 & -4 \\ \hline -4 & 1 & -11 \\ \hline -2 & -3 & 5 \\ \hline 3 & 5 & -9 \\ \hline 4 & -2 & 14 \\ \hline 0 & 1 & -3 \\ \hline \end{array}$$

Let the rank of an  $m \times n$  matrix  $A$  be  $r$ . Then it can be shown that the  $m$  points are contained in a space of  $r$  dimensions which also contains the origin. The rank of the matrix of *Table 20* is 2. Hence the six points should lie in a plane which contains the origin. The equation of such a plane is

$$(41) \quad \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3 = 0,$$

in which the  $x$ 's are the three co-ordinates of each of the points in the plane and the  $\lambda$ 's are the direction cosines of the plane. The  $\lambda$ 's are not independent parameters because of the conditional equation (39). Hence any two  $\lambda$ 's define the plane. These may be found by any two of the six points which are not collinear with the origin. Since no two rows of *Table 20* are proportional, no two of the points are collinear with the origin. If two such points were found, then these two points would define a line through the origin and not a plane.

Substituting the  $x$ 's of the first two points of *Table 20* in (41),

$$(42) \quad \begin{cases} \lambda_1 + 2\lambda_2 - 4\lambda_3 = 0, \\ -4\lambda_1 + \lambda_2 - 11\lambda_3 = 0. \end{cases}$$

Solving for  $\lambda_1$  and  $\lambda_2$  in terms of  $\lambda_3$ , we have

$$(43) \quad \begin{cases} \lambda_1 = -2\lambda_3, \\ \lambda_2 = 3\lambda_3. \end{cases}$$

Substituting (43) in (39),

$$(44) \quad 4\lambda_3^2 + 9\lambda_3^2 + \lambda_3^2 = 1,$$

or

$$\lambda_3 = \frac{1}{\sqrt{14}},$$

and hence

$$\lambda_1 = -\frac{2}{\sqrt{14}}.$$

$$\lambda_2 = \frac{3}{\sqrt{14}}.$$

The equation of the plane in normal form is therefore

$$(45) \quad -\frac{2}{\sqrt{14}}x_1 + \frac{3}{\sqrt{14}}x_2 + \frac{1}{\sqrt{14}}x_3 = 0.$$

All of the four remaining points must lie in this plane, since the rank of *Table 20* is 2. The three coefficients are the direction cosines of the plane, i.e., the direction cosines of the normal to the plane.

The distance of a point from the origin is  $\sqrt{\Sigma x^2}$ , where the  $x$ 's are its co-ordinates. For example, the distance of the fourth point from the origin is  $\sqrt{3^2+5^2+(-9)^2}=10.72$ . If the sum of the squares of the co-ordinates of a point is equal to unity, then the point is at unit distance from the origin.

Each point may be interpreted as defining the terminus of a vector from the origin. The *scalar product* of any two vectors is  $h_1h_2 \cos \phi$ , where  $h_1$  and  $h_2$  are the lengths of the vectors and  $\phi$  is their angular separation. If the two vectors are of unit length, then the scalar product is the cosine of the angular separation. It can be shown that the scalar product of two vectors can be expressed in the form

$$\sum_{j=1}^n x_{ij}x_{lj},$$

where  $i$  and  $l$  refer to points (rows) and  $j$  refers to co-ordinates (columns). For example, the scalar product of the vectors defined by the second and fourth points of *Table 20* is

$$(-4)(3) + (1)(5) + (-11)(-9) = +92.$$

If the sum of the squares of the co-ordinates of each point is equal to unity, so that the points lie at unit distance from the origin, then the scalar product, or cross product, is the cosine of the angular separation of the vectors at the origin.

### Orthogonal transformations

The three simultaneous equations (2) may be regarded as representing the three co-ordinates of a point  $x$  ( $x_1, x_2, x_3$ ), the three co-ordinates of a point  $y$  ( $y_1, y_2, y_3$ ), and the law by which each point  $y$  is transformed into a corresponding point  $x$ . For every point  $y$ , there exists some other point  $x$  whose co-ordinates can be found by (2) when the co-ordinates of  $y$  are known. This relation is called a *linear transformation* by which the points  $y$  are moved to the corresponding points  $x$ . Every pair of the corresponding points are related by the linear transformation (2). The transformation is called *linear* when the equations by which the  $x$ 's can be found from the  $y$ 's are of the first degree, as is the case in (2).

If the transformation is of such a nature that the  $x$ 's can be obtained from the  $y$ 's by merely rotating the co-ordinate axes, then the transformation is an *orthogonal transformation* or *rotational transformation*. In order that a transformation shall be orthogonal, it is evidently necessary that the  $x$ 's be at the same distance from the origin as the corresponding  $y$ 's because the distance of a point from the origin remains invariant when the co-ordinate axes are rotated. It is also necessary that the angular separations, or scalar products, be invariant, because the configuration of the points is not altered by rotating the co-ordinate axes.

The matrix  $A$  of Table 11 is called the *matrix of the transformation* when it is regarded as the relation by which the points  $y$  are changed into the points  $x$ . A linear transformation is represented in the more general form by the square matrix of Table 21. It can be shown that a square matrix is

Table 21

$$\left\| \begin{array}{cccc} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \dots & \dots & \dots & a_{ij} & \dots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{array} \right\|$$

orthogonal, i.e., that it has the effect of rotating a set of points  $y$  into a set of points  $x$ , with the same configuration as the  $y$ 's, if it satisfies the following conditions:

1) The sum of the squares of the elements in each row is equal to unity, i.e.,

$$(46) \quad \sum_{j=1}^n a_{ij}^2 = +1,$$

and

2) The cross product of every pair of rows,  $i$  and  $l$ , is equal to zero, i.e.,

$$(47) \quad \sum_{j=1}^n a_{ij}a_{lj} = 0,$$

when  $i \neq l$ . It is immaterial whether the rotation is conceived as a rotation of the orthogonal co-ordinate axes in a fixed configuration of points or as a rotation of the configuration in a fixed reference frame of the co-ordinate axes. The result is the same.

The two conditions, (46) and (47), are of such frequent occurrence that it is sometimes convenient to combine them in a single statement. This can be done by writing the conditions in the more condensed form

$$(48) \quad \delta_{il} = \sum_{j=1}^n a_{ij}a_{lj},$$

where the symbol  $\delta_{il}$  is known as *Kronecker's delta*. It is defined as follows:

$$\delta_{il} = +1 \text{ when } i = l,$$

$$\delta_{il} = 0 \text{ when } i \neq l.$$

It can be seen that with this definition of  $\delta_{il}$ , the single statement (48) covers the two statements (46) and (47).

If the matrix of a transformation is square and if it satisfies (48), then the following conditions are also satisfied:

3) The sum of the squares of the elements in each column is unity, i.e.,

$$(49) \quad \sum_{i=1}^n a_{ij}^2 = +1.$$



4) The cross product of any pair of columns is zero, i.e.,

$$(50) \quad \sum_{i=1}^n a_{ij}a_{ik} = 0, \quad j \neq k,$$

where  $j$  and  $k$  refer to columns.

5) The determinant of the transformation is  $\pm 1$ , i.e.,

$$(51) \quad |A| = \pm 1.$$

If an orthogonal co-ordinate axis is reversed in direction, then the corresponding co-ordinate for each point is reversed in sign. If an odd number of orthogonal co-ordinate axes are reversed in direction in an orthogonal transformation, then the determinant of the transformation is equal to  $-1$ . If an even number of axes are reversed, the determinant is equal to  $+1$ . These two statements can be made with reference to the configuration. If the rotational transformation retains the configuration of the points, the determinant of the transformation is equal to  $+1$ . If the rotation involves a symmetric distortion of the configuration, the determinant of the orthogonal transformation is equal to  $-1$ .

Table 22

$$\begin{array}{c}
 \begin{array}{cc}
 x_1 & x_2 \\
 a \parallel \begin{array}{c} 1 \\ 2 \\ -1 \\ 5 \\ 4 \end{array} \parallel \\
 b \parallel \begin{array}{c} 2 \\ 3 \end{array} \parallel \\
 c \parallel \begin{array}{c} -1 \\ 4 \end{array} \parallel \\
 d \parallel \begin{array}{c} 5 \\ -2 \end{array} \parallel \\
 e \parallel \begin{array}{c} 4 \\ 1 \end{array} \parallel \\
 A
 \end{array}
 \cdot
 \begin{array}{cc}
 Y_1 & Y_2 \\
 \parallel \begin{array}{c} .866 \\ .500 \end{array} \parallel \\
 G
 \end{array}
 =
 \begin{array}{cc}
 y_1 & y_2 \\
 \parallel \begin{array}{c} -.134 \\ 3.232 \\ 1.134 \\ 3.330 \\ 3.964 \end{array} \parallel \\
 \parallel \begin{array}{c} -2.232 \\ 1.598 \\ 3.964 \\ -4.232 \\ -1.134 \end{array} \parallel \\
 B \\
 AG=B
 \end{array}
 \begin{array}{c}
 a \\
 b \\
 c \\
 d \\
 e
 \end{array}
 \end{array}$$

Each column of an orthogonal transformation shows the direction cosines of one of the new co-ordinate axes, referred to the given co-ordinate axes.

A rotation of the co-ordinate axes implies that the given configuration and the transformed configuration are contained in the same space. The rows of a transformation correspond to the dimensions of the given configuration, and the columns of a transformation correspond to the dimensions of the new configuration. If the transformation is merely a rotation of axes, it is evident that the matrix of an orthogonal transformation is necessarily square. If the matrix of a transformation is of order  $m \times n$  where  $m \neq n$ , then the transformation cannot be orthogonal, since the number of dimensions of the given configuration and the number of dimensions of the transformed configuration are not the same. However, such a matrix may satisfy

condition (47), and it is then said to be *orthogonal by rows*. If the condition is satisfied for columns, as in (50), instead of for rows, then the matrix is said to be *orthogonal by columns*.

Table 22 is a numerical example of the rotation of five points in a plane.  $A$  is a matrix of order  $5 \times 2$ . The orthogonal transformation  $G$  is of order  $2 \times 2$ . The matrix product  $AG = B$  is of order  $5 \times 2$ , and it shows the co-

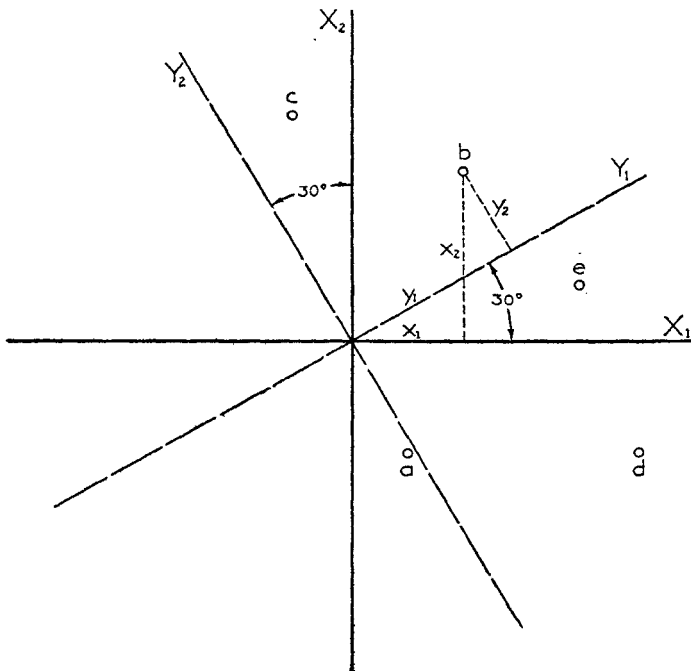


FIGURE 2

ordinates of the same five points with reference to the new rotated coordinate axes. In *Figure 2* the five points have been plotted for the given orthogonal co-ordinates axes,  $X_1$  and  $X_2$ , that are implied in  $A$ .

Let it be desired to rotate these axes through an angle  $\phi = 30^\circ$ . The usual formulas for rotation of axes\* can be written in the form of a  $2 \times 2$  transformation as follows:

$$\begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix}$$

\* W. F. Osgood and W. C. Graustein, *Plane and Solid Analytic Geometry* (New York: Macmillan Co., 1929), p. 220.

The first condition (46) gives

$$\cos^2 \phi + \sin^2 \phi = + 1 ,$$

and the second condition (47) gives

$$\cos \phi \sin \phi - \cos \phi \sin \phi = 0 .$$

Hence this is an orthogonal transformation in which the other properties may be readily verified.

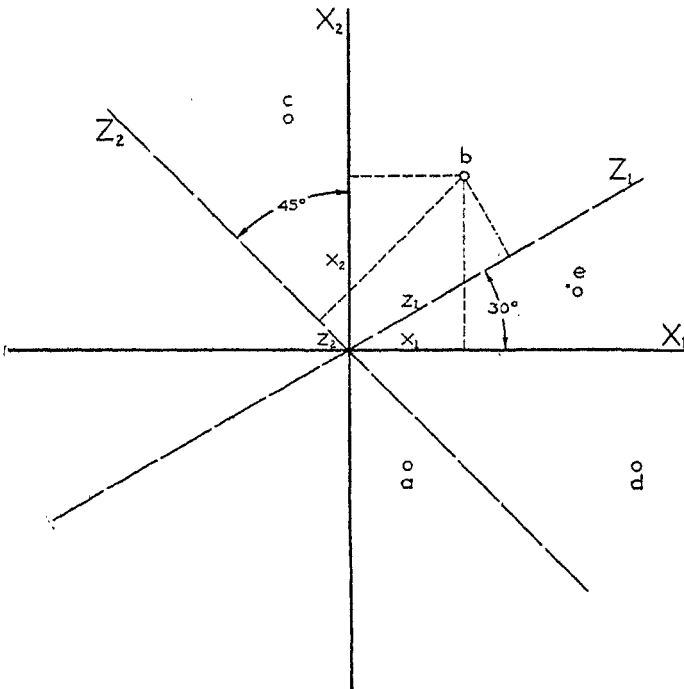


FIGURE 3

Substitution of  $\phi = 30^\circ$  in the transformation produces the matrix  $G$ , and the multiplication  $AG$  produces  $B$ . The numerical values in  $B$  may be checked in *Figure 2*, where  $Y_1$  and  $Y_2$  have been drawn so that  $Y_1OX_1 = \phi = 30^\circ$ . For example, the co-ordinates of the second point can be measured on the graph to be 3.23 and 1.60 for the two rotated axes, while they are 2 and 3 for the two original axes. This figure illustrates the geometric in-

terpretation of an orthogonal transformation. The two columns of the transformation  $G$  show the direction cosines of the new orthogonal reference vectors  $Y_1$  and  $Y_2$ .

**Oblique transformations**

In *Figure 2* the two co-ordinate axes  $Y_1$  and  $Y_2$  are orthogonal. If it is desired to define the  $n$  points with reference to oblique reference axes, the transformation is effected in a similar way. In *Figure 3* the same five points are plotted on the  $X_1$  and  $X_2$  axes which are implied in the given matrix  $A$  of *Table 22*. Here the new axes are oblique;  $Z_1$  is rotated  $30^\circ$  from  $X_1$ , and  $Z_2$  is rotated  $45^\circ$  from  $X_2$ .

In *Table 23* are shown the numerical values for the corresponding oblique transformation. The matrix  $A$  contains the co-ordinates of the given five

*Table 23*

	$x_1$	$x_2$		$Z_1$	$Z_2$		$z_1$	$z_2$
$a$	1	-2	·	+.866	-.707	=	-.134	-2.121
$b$	2	3		+.500	+.707		3.232	.707
$c$	-1	4					1.134	3.535
$d$	5	-2		$H$			3.330	-4.949
$e$	4	1					3.964	-2.121
	$A$						$AH=C$	

points in a space defined by the two orthogonal reference axes  $X_1$  and  $X_2$ . The matrix  $H$  is a square matrix of order  $2 \times 2$ . It is the matrix of the oblique transformation. Its columns show the direction cosines of the new oblique co-ordinate axes  $Z_1$  and  $Z_2$ . These direction cosines may be verified in *Figure 3*.

It should be noted that the sum of the squares of each column of  $H$  is equal to unity. Each column of  $H$  may be regarded as defining a unit vector in a space of two dimensions. The cross product of the columns of  $H$  is

$$\lambda_1\mu_1 + \lambda_2\mu_2,$$

which is the cosine of the angle between  $Z_1$  and  $Z_2$ .

The product  $AH=C$  shows the projection of each of the five points on each of the oblique axes  $Z_1$  and  $Z_2$ . This interpretation can be verified by actual measurement on *Figure 3*.

## CHAPTER I

### THE FACTOR PROBLEM

#### On the nature of science

This volume is concerned with methods of discovering and identifying significant categories in psychology and in other social sciences. It is therefore of interest to consider some phases of science in general that bear on the problem of finding a methodology for a psychological science.

It is the faith of all science that an unlimited number of phenomena can be comprehended in terms of a limited number of concepts or ideal constructs. Without this faith no science could ever have any motivation. To deny this faith is to affirm the primary chaos of nature and the consequent futility of scientific effort. The constructs in terms of which natural phenomena are comprehended are man-made inventions. To discover a scientific law is merely to discover that a man-made scheme serves to unify, and thereby to simplify, comprehension of a certain class of natural phenomena. A scientific law is not to be thought of as having an independent existence which some scientist is fortunate to stumble upon. A scientific law is not a part of nature. It is only a way of comprehending nature.

A simple example is the concept "force." No one has ever seen a force. Only the movement of objects is seen. The faith of science is that some schematic representation is possible by which complexities of movement can be conceptually unified into an order. The error of a literal interpretation of a force vector as the pictorial representation of a corresponding physical entity is seen in the resolution of forces. If a particle moves with uniform acceleration in a certain direction, it is, of course, possible to describe the movement by one force, or by two, or by three or more coplanar forces. This resolution of a movement into several simultaneous and superimposed movements is frequently done in order that a convenient and habitual reference frame may be retained. While the ideal constructs of science do not imply physical reality, they do not deny the possibility of some degree of correspondence with physical reality. But this is a philosophical problem that is quite outside the domain of science.

Consider, as another example, Coulomb's inverse-square law of electrical attraction. A postulated force is expressed as a function of the linear separation of the charges. Now, if the charges were to be personified, they would probably be much surprised that their actions were being described in terms

of their linear separations. No one assumes that there is a string between the charges, but Coulomb's law implies that the length of such a string is to be used in our simplified scheme of comprehending the postulated charges. It is more likely that the whole space surrounding the charges is involved in the phenomena of attraction and that Coulomb's law is a fortunate shortcut for representing approximately a part of the phenomena that are called charges and attractions. It is not unlikely that all of these entities will eventually vanish as such and become only aspects of an order more involved than Coulomb's law implies but not so chaotic as to individualize completely every moment of nature.

A science of psychology will deal with the activities of people as its central theme. A large class of human activity is that which differentiates individuals as regards their overt accomplishments. Just as it is convenient to postulate physical forces in describing the movements of physical objects, so it is also natural to postulate abilities and their absence as primary causes of the successful completion of a task by some individuals and of the failure of other individuals in the same task.

The criterion by which a new ideal construct in science is accepted or rejected is the degree to which it facilitates the comprehension of a class of phenomena which can be thought of as examples of a single construct rather than as individualized events. It is in this sense that the chief object of science is to minimize mental effort. But in order that this reduction shall be accepted as science, it must be demonstrated, either explicitly or by implication, that the number of degrees of freedom of the construct is smaller than the number of degrees of freedom of the phenomena that the reduction is expected to subsume. Consider, as an example, any situation in which a rational equation is proposed as the law governing the relation between two variables. If three observations have been made and if the proposed equation has three independent parameters, then the number of degrees of freedom of the phenomena is the same as the number of degrees of freedom of the equation, and hence the formulation remains undemonstrated. If, on the other hand, one hundred experimentally independent observations are subsumed by a rational equation with three parameters, then the demonstration can be of scientific interest. The convincingness of a hypothesis can be gauged inversely by the ratio of its number of degrees of freedom to that of the phenomena which it has demonstrably covered. It is in the nature of science that no scientific law can ever be proved to be right. It can only be shown to be plausible. The laws of science are not immutable. They are only human efforts toward parsimony in the comprehension of nature.

If abilities are to be postulated as primary causes of individual differences in overt accomplishment, then the widely different achievements of indi-

viduals must be demonstrable functions of a limited number of reference abilities. This implies that individuals will be described in terms of a limited number of faculties. This is contrary to the erroneous contention that since every person is different from every other person in the world, people must not be classified and labeled.

Each generalization in the scientific description of nature results in a loss in the extent to which the ideal constructs of science match the individual events of experience. This is illustrated by simple experiments with a pendulum in which the mass, the period, and the locus of the center of gravity with reference to a fulcrum are involved in the ideal construct that leads to experimental verification. But the construct matches only incompletely the corresponding experimental situation. The construct says nothing about the rusty set screw and other extraneous detail. From the viewpoint of immediate experience, scientific description is necessarily incomplete. The scientist always finds his constructs immersed in the irrelevancies of experience. It seems appropriate to acknowledge this characteristic of science in view of the fact that it is a rather common notion that the scientific description of a person is not valid unless the so-called "total situation" has been engulfed. A study of people does not become scientific because it attempts to be complete, nor is it invalid because it is restricted. The scientific description of a person will be as incomplete from the viewpoint of common sense as the description of other objects in scientific context.

The development of scientific analysis in a new class of phenomena usually meets with resistance. The faith of science that nature can be comprehended in terms of an order acknowledges no limitation whatever as regards classes of phenomena. But scientists are not free from prejudice against the extension of their faith to realms not habitually comprehended in the scientific order. Examples of this resistance are numerous. It is not infrequent for a competent physical scientist to declare his belief that the phenomena of living objects are, at least in some subtle way, beyond the reach of rigorous scientific order.

One of the forms in which this resistance appears is the assertion that, since a scientific construct does not cover all enumerable detail of a class of phenomena, it is therefore to be judged inapplicable. Since the analysis of cell growth by mathematical and physical principles does not cover everything that is known about cells, the biologist judges the analysis to be inapplicable. Since no mathematical analysis that can be conceived would cover all the subtle mysteries of personality, this realm is frequently judged to be outside the domain of rigorous science. But physical scientists accept rigorous scientific analyses about physical events that leave fully as much beyond the scientific constructs. Every explosion in the world has been different

from every other explosion, and no physicist can write equations to cover all of the detail of any explosive event. It is certain that no two thunderstorms have been exactly alike, and yet the constructs of physics are applied in comprehending thunder and lightning without any demand that the detail of the landscape be covered by the same scientific constructs.

The attitudes of people on a controversial social issue have been appraised by allocating each person to a point in a linear continuum as regards his favorable or unfavorable affect toward the psychological object. Some social scientists have objected because two individuals may have the same attitude score toward, say, pacifism, and yet be totally different in their backgrounds and in the causes of their similar social views. If such critics were consistent, they would object also to the statement that two men have identical incomes, for one of them earns when the other one steals. They should also object to the statement that two men are of the same height. The comparison should be held invalid because one of the men is fat and the other is thin. This is again the resistance against invading with the generalizing and simplifying constructs of science a realm which is habitually comprehended only in terms of innumerable and individualized detail. Every scientific construct limits itself to specified variables without any pretense to cover those aspects of a class of phenomena about which it has said nothing. As regards this characteristic of science, there is no difference between the scientific study of physical events and the scientific study of biological and psychological events. What is not generally understood, even by many scientists, is that no scientific law is ever intended to represent any event pictorially. The law is only an abstraction from the experimental situation. No experiment is ever completely repeated.

There is an unlimited number of ways in which nature can be comprehended in terms of fundamental scientific concepts. One of the simplest ways in which a class of phenomena can be comprehended in terms of a limited number of concepts is probably that in which a linear attribute of an event is expressed as a linear function of primary causes. Even when the relations are preferably non-linear and mathematically involved, it is frequently possible to use the simpler linear forms as first approximations. A well-known example of this type of relation is that in which the chroma of a spectral color is expressed as a linear function of two arbitrarily chosen primaries. If two spectral colors are chosen arbitrarily for use as primaries, it is possible to express any intermediate color as a linear function of the two arbitrarily chosen primaries. The coefficients of the two terms of this linear function represent the angular sizes of the two sectors into which a color rotator is divided. When the rotator is spun, the intermediate color is seen. But here, as elsewhere in science, although the chroma of the result-



ing color is expressed in terms of the linear function of the arbitrary primaries, it does not follow that the saturation or gray-value is expressed by the same law. There is still debate about which colors are to be considered primary. This question can be settled only by discovering that a certain set of primaries gives the most parsimonious comprehension of some phase of color vision. A parallel in the description of human traits is their description, in first approximation, as linear functions of a limited number of reference traits. The final choice of a set of primary reference traits or faculties must be made in terms of the discovery that a particular set of reference traits renders most parsimonious our comprehension of a great variety of human traits.

### Psychological postulates and definitions

The factorial methods have been developed primarily for the purpose of analyzing the relations of human traits. These are defined as follows:

Definition 1. *A trait is any attribute of an individual.*

The factorial methods are applicable also in the analysis of attributes of inanimate members of a group. The members of a statistical population may be moments in time or regions in space or any other entities, each of which has a set of attributes. This generalization will not be made explicitly, but it is implied in the following chapters. Since the methods have been developed primarily with psychological categories in mind, these will be explicitly discussed even though the same methods are applicable to problems which involve the attributes of inanimate members of a statistical group.

It is useful to distinguish between those traits which are descriptive of the individual as he appears to others and those traits which are exemplified primarily in the things that he can do. This distinction is involved in the definition of "ability."

Definition 2. *An ability is a trait which is defined by what an individual can do.*

This definition implies that there are as many abilities as there are enumerable things that individuals can do. Each ability is therefore objectively defined in terms of a specified task and of a specified method of appraising it.

Definition 3. *The task, together with the method of appraising it, which defines an ability is called a test.*

Definition 4. *The linear evaluation of a test performance is called a score.*

It is implied in these definitions that an index of ability is covariant with the score in the test which defines the ability, and that a true index of ability is covariant with the true score in the test.

Let there be  $N$  individuals in an experimental population, and let there

be  $n$  tests. Let  $S_{ji}$  be the raw score of individual  $i$  in test  $j$ , and let  $E_{ji}$  be the absolute variable error of the score. Then the true score  $T_{ji}$  of individual  $i$  in test  $j$  is defined by the relation

$$T_{ji} = S_{ji} - E_{ji}.$$

In psychological investigations it is sometimes desirable to postulate that the frequency distribution of the indices of a particular ability is normal in the experimental population, while in some investigations this is not a desirable restriction. Hence two indices of ability will be defined in accordance with these two cases. Both indices are so defined as to be covariant with the true score in the test which defines the ability, but they differ as regards the assumption of normality of the distribution of ability in the experimental population.

*Case 1, assuming that the distribution of ability is not necessarily Gaussian:*

Let  $v_{ji} = aT_{ji} + b$ , in which the parameter  $b$  and the positive parameter  $a$  are so chosen that the following conditions are satisfied:

$$\begin{aligned} 1) \quad & \sum_{i=1}^N v_{ji} = 0 ; \\ 2) \quad & \sum_{i=1}^N v_{ji}^2 = N . \end{aligned}$$

Then  $v_{ji}$  is an index of the ability  $j$  in individual  $i$  which is a linear function of the true score  $T_{ji}$  in the test  $j$ . Since this index is a linear function of the true score, it follows that the shape of the frequency distribution of true scores is retained in the frequency distribution of indices of the corresponding ability. This index will be called the *standard score in ability  $j$* .

*Case 2, assuming that the distribution of ability is Gaussian:*

Let  $\phi(T_{ji}) = t_{ji}$  be the monotonic increasing function which satisfies the three following conditions:

$$\begin{aligned} 1) \quad & \sum_{i=1}^N t_{ji} = 0 ; \\ 2) \quad & \sum_{i=1}^N t_{ji}^2 = N ; \end{aligned}$$

3) the frequency distribution of  $t_{ji}$  in  $N$  is Gaussian.

The index  $t_{ji}$  will be called the *normalized standard score in ability  $j$*  or simply the *standard score in ability  $j$* . It is assumed that in each investigation in which factorial methods are used, the statement will be explicitly made as

to whether a Gaussian distribution of ability has been assumed for the experimental population. The frequency distribution of raw scores in psychological tests is arbitrary, since the scores can be adjusted as regards skewness in any desired direction and to any desired extent by merely inserting or removing relatively easy or relatively difficult test items. Since this is a matter of judgment on the part of the person who assembles the psychological tests, no inference can be made concerning the skewness or normality of the distribution of a particular ability from the arbitrary skewness or artificial normality of the distribution of raw scores.

It is desirable to develop the factorial methods in such a manner that they are independent of the assumption of normality of ability in any particular experimental population. *In the present theoretical development of the factorial methods it will not be assumed that any of the distributions of ability are normal.*

The application of the factorial methods in science rests on a fundamental postulate.

Postulate. *The standard scores of all individuals in an unlimited number of abilities can be expressed, in first approximation, as linear functions of their standard scores in a limited number of abilities.*

The correlation between the true scores in two tests will be referred to as the correlation between the two abilities which are defined by the tests. In statistical work it is customary to refer to two variables as independent when their correlation is zero. The term *independence* will be used with three different meanings. They will be designated by appropriate adjectives unless the context makes the designation unnecessary.

Definition 5. *A set of  $n$  abilities are **linearly independent** if the rank of the matrix of their true intercorrelations is  $n$ .*

Definition 6. *Two abilities are **statistically independent** in a population if their correlation is zero in that population.*

Definition 7. *Two observations are **experimentally independent** if they are experimentally distinct, so that one is not derived from the other by a constraint either of the experimental situation or of the computations.*

In one sense, no two observations can ever be experimentally independent. The term can be used only with reference to the state of knowledge at the time the observations are made.

It is clarifying to interpret geometrically the relations of abilities. In such a context two abilities that are uncorrelated in a population will be called *orthogonal* in that population. Two abilities that are correlated in a population will be called *oblique* in that population.

There is special interest in the limited number of abilities in terms of which all other abilities can be defined, since these are the landmarks in terms of which all abilities can be comprehended.

Definition 8. *If the standard scores of  $N$  individuals in  $n$  abilities are expressed as linear functions of their scores in  $r$  linearly independent abilities, where  $r < n$ , then the  $r$  abilities will be called **reference abilities**.*

It will be shown that if a battery of tests can be described with reference to  $r$  orthogonal abilities, there exists an infinite number of sets of  $r$  orthogonal abilities in terms of which the description can be made with equal accuracy. An arbitrary set of  $r$  orthogonal abilities may be chosen for purposes of description. These are the statistically independent or orthogonal reference abilities. If a battery of tests can be described in terms of  $r$  orthogonal reference abilities, the tests can also be described by a set of  $r$  oblique reference abilities. It is not necessary that a reference ability be represented by a test in which it is involved exclusively. While each of the tests that are used in experimental work defines an ability, it may happen that the reference abilities in terms of which tests and individuals are described are not represented by actual tests but by linear combinations of several tests. A linear combination of tests may be thought of as a composite test.

The  $n \times r$  matrix of coefficients of the  $r$  reference abilities in terms of which the standard scores in each of the  $n$  abilities can be linearly expressed is not unique. The most parsimonious comprehension of the  $n$  abilities in terms of  $r$  reference abilities is obtained when the number of vanishing coefficients of the  $n$  linear functions is maximized.

Definition 9. *If the  $N$  standard scores in each of  $n$  diversified abilities can be expressed as linear functions of fewer than  $r$  of the  $r$  independent reference abilities, then that set of  $r$  reference abilities for which the number of vanishing coefficients is a maximum will be called **primary abilities**.*

If a large and diversified battery of tests can be described in terms of  $r$  reference abilities and if a particular set of  $r$  primary abilities can be found such that each test can be described in terms of less than  $r$  of these abilities, then the primary abilities have significance because of their identification with phenomena extraneous to the test scores and their intercorrelations even if the extraneous phenomena are unknown.

It is conceivable, and not improbable, that some reference abilities will be found to be sufficiently elemental that they can be declared to be either present or absent in each individual without intermediate gradations in amount or degree of presence.

Definition 10. *If only two numerical values occur in the population  $N$  for the standard scores in a primary ability, then the primary ability is a **unitary ability**.*

This is a genetic interpretation of factors.

The underlying idea from which the present factorial analysis originates is a very simple one. If there are  $N$  individuals in a random sample of the population and if each of these individuals has demonstrated his abilities by doing his best on  $n$  separate tests, then there will be  $nN$  test scores to be explained. At the present stage of development of psychology and of genetics there are no available ideal constructs for representing the mental abilities. The simplest possible formulation seems to be an analysis of the variance of each test into linear components.\* It is almost certain that this simple type of analysis will not be the ultimate one, but it is likely that the principal primary abilities will be discovered by factorial analysis of the variance of each test. As soon as some of the primary abilities have been isolated, detailed studies of inheritance should be undertaken.

The performance of an individual on a test is determined in part by the abilities that are called for by the test and in part by the degree to which the individual possesses these abilities. An individual's performance on a test may be regarded as a sum of the contributions of his primary abilities. His abilities are not called for to the same extent by the different tests, and it therefore seems natural to describe each of his test scores as a sum of weighted linear contributions of his different primary abilities. The weights are descriptive of the tests. This simple formulation of the problem is flexible enough to serve the descriptive purposes of psychology until more refined, and perhaps less obvious, constructs will be called for by future experimental inquiry and by the attainment of more accurate psychological measurement than now seems to be possible.

The assumption that a performance can be described approximately as a sum of weighted linear contributions of several independent factors can be represented in the following equation:†

$$(1) \quad a_{j1}x_{1i} + a_{j2}x_{2i} + a_{j3}x_{3i} + \dots + a_{jq}x_{qi} = s_{ji},$$

in which  $s_{ji}$  represents the standard score of individual  $i$  in test  $j$ . The  $x$ 's represent standard scores in the  $q$  statistically independent arbitrary reference abilities, while the  $a$ 's represent factor loadings in the tests. The  $x$ 's describe the individuals, and the  $a$ 's describe the tests. The first term represents the contribution of the arbitrary reference ability No. 1 to the test performance  $s_{ji}$ . It is determined by the amount of the first arbitrary reference ability that the subject possesses, namely,  $x_{1i}$ , and the extent to which the test calls for the first ability, namely,  $a_{j1}$ . Similar reasoning applies to the contribution of each other ability to the test performance. If the primary abilities are oblique, these orthogonal reference abilities may

\* The variance is the square of the standard deviation.

† "Multiple Factor Analysis," p. 409.

be regarded as arbitrary. They can be rotated or transformed into the primary abilities by methods that will be described.

There is no loss of generality in reducing all performances to standard scores. This reduction involves a translation of the origin of the raw test scores to the mean score of the distribution and a stretching of the scale so as to make the standard deviation unity. The shape of the distribution is not altered by reducing the raw scores to standard scores.

It should be noted that, even if each individual can be described in terms of a limited number of independent reference abilities, it is still possible for every person to be different from every other person in the world. Each person might be described in terms of his standard scores in a limited number of independent abilities. The number of permutations of these scores would probably be sufficient to guarantee the retention of individualities.

With a limited number of abilities this formulation not only allows that every person shall be different from every other person but it also allows the widest possible differences between several individuals who attain the same objective performance in a test. This may be readily seen by considering a hypothetical example. Assume that a test calls for two abilities, such as ability in abstraction and ability in the manipulation of numbers. Several individuals try the test and attain the same score. One of them may possess a high degree of ability in making the abstractions involved in the test, but he may be slow in numerical manipulation. Another may be slow in formulating the abstract part of the problem, but he may make up for this deficiency by superior numerical speed. The objective result might be the same. The purpose of factor analysis is to obtain a quantitative description of each primary ability in each individual by means of tasks that require these abilities in different amounts. Since every task is probably composite in the primary abilities required, it is necessary to make the appraisal of the abilities of individuals by analytical methods. This is exactly the object of the multiple-factor methods as applied to the problem of describing the abilities of people.

Factor analysis is reminiscent of faculty psychology. It is true that the object of factor analysis is to discover the mental faculties. The severe restrictions that are imposed by the logic of factor analysis make it an arduous task to isolate each new mental faculty, because it is necessary to prove that it is called for by the experimental observations. Factor analysis does not allow that a new faculty be added as soon as a new name can be found for the things that people can do. In order to prove that reasoning and abstraction are two different faculties, for example, it will be necessary to show that the tasks which call for such activities really do involve two factors, and not one.

There is an interesting difference between the logic of multiple correlation

and that of factor analysis. In multiple correlation it is necessary to designate one of the variables as dependent and all of the others as independent, and the problem is then to predict the one test score from all the rest. In factor analysis there is no problem of prediction of any test scores, and there is no distinction between independence and dependence among the given variables. The dependent variables are the primary abilities of the individual subjects which are to be estimated in terms of the given tests.

In the psychology of the future it may be found useful to postulate a different form of ideal construct for the description of mental endowment than the simple one that is implied in equation (1). The ideal constructs of the future may involve elements with location in a space frame with spatial, dynamic, and temporal constraints analogous to the ideal constructs of genetics. It would be unfortunate if some initial success with the analytical methods to be described here should lead us to commit ourselves to them with such force of habit as to retard the development of entirely different constructs that may be indicated by improvements in measurement and by inconsistencies between theory and experiment.

### Matrix formulation

Let  $N$  be the number of individuals in a random sample of the population, and let  $n$  be the number of tests from which the primary abilities are to be isolated. The raw data for factorial analysis consist of the entries in an  $n \times N$  table of standard scores in which each of the  $N$  subjects is represented by  $n$  test scores. This table will be referred to as an  $n \times N$  score matrix  $S$ .

Equation (1) implies that the matrix  $S$  is the product of two matrices, namely, one matrix with elements  $a$  which are descriptive of the tests and another matrix with elements  $x$  which are descriptive of the individuals. The former will be called the *factorial matrix*  $F_4$  and the latter the *population matrix*  $P_4$ .

In setting up these two matrices, an assumption will be made concerning the nature of the factors in the present psychological problem, namely, that there are at least three kinds of factors involved in the variance of each test. These factors are (a) the common factors or abilities, (b) the specific factors or abilities, and (c) the chance error factors. By *common factor* is meant any factor or ability which is called for by more than one of the  $n$  tests in a battery. By *specific factor* or ability is meant any factor or ability which is called for by only one of the  $n$  tests. By *error factor* is meant the variable chance error which is a part of the total variance of the test.

It is evident that *an ability which is a common factor for a test in one battery may become a part of the specific factor in the same test when it is placed in another test battery*. Whether any particular ability is common or specific depends on the battery as a whole.

In equation (1) test  $A$  is defined by the weights or test coefficients  $a_{j1}, a_{j2}, \dots, a_{jq}$ . These weights show the extent to which the test calls for each one of a set of reference abilities. The test coefficients therefore constitute a psychological description of the test. *It is a fundamental criterion for a valid method of isolating primary abilities that the weights of the primary abilities for a test must remain invariant when it is moved from one test battery to another test battery.* If this criterion is not fulfilled, the psychological description of a test will evidently be as variable as the arbitrarily chosen batteries into which the test may be placed. Under such conditions no stable identification of primary mental abilities can be expected. The factorial methods to be presented are consistent with this criterion, and stable identification of the primary abilities can therefore be expected. This criterion assumes that the several test batteries are given to the same population. The primary abilities that define a test in one population should be identical with the primary abilities which define it in a second population.

A test may call for two or more abilities that are unique for that test in a particular battery. Then the specific variance of the test should be divided into parts, one part for each of the several specific abilities. *In factorial analysis all of the abilities that are specific for a test combine into a single variance.*

Table 1 represents a *population matrix* in which the attributes of each member of the population  $N$  are recorded with reference to the common factors, the specific factors, and the error factors.

The notation is as follows:

Subscript  $i$  refers to a person;

Subscripts  $j$  and  $k$  refer to tests;

Subscripts  $m$  and  $M$  refer to common factors.

Let  $x$  refer to common factors;

$y$  refer to specific factors;

$z$  refer to error factors.

Let  $x_{mi}$  = the standard score of individual  $i$  in the common factor  $m$ ;

$y_{ji}$  = the standard score of individual  $i$  in the specific factor of test  $j$ ;

$z_{ji}$  =  $e_{ji}/\varepsilon_j$ , where  $e_{ji}$  is the absolute variable error in the standard score  $s_{ji}$ , and  $\varepsilon_j$  is the standard error of  $s_{ji}$ .

Let  $r$  = number of common factors;

$n$  = number of tests;

$N$  = number of individuals in a random sample of the population.

An interpretation of the cell entries of the population matrix  $P_4$  is then as follows: Each column is descriptive of one individual. The first  $r$  entries show his standard scores in the  $r$  common abilities. The next  $n$  entries show his standard scores in the  $n$  specific abilities. It is here assumed that every



psychological test in a finite test battery calls for some specific ability, although it may be of minor significance. This assumption fits the usual case. It is only in an unusual test battery (if such exists at all) that the number of specifics can be smaller than the number of tests. The present analysis would be essentially the same even for the ideal case in which specifics were

Table 1  
Population Matrix  $P_4$

$N$	
$r$ common factors	$x_{11}$ $x_{12}$ $x_{13}$ . . . $x_{1N}$
	$x_{21}$ $x_{22}$ $x_{23}$ . . . $x_{2N}$
	$x_{31}$ $x_{32}$ $x_{33}$ . . . $x_{3N}$
	. . . . . $x_{mi}$ . . .
	$x_{r1}$ $x_{r2}$ $x_{r3}$ . . . $x_{rN}$
$n$ specific factors	$y_{11}$ $y_{12}$ $y_{13}$ . . . $y_{1N}$
	$y_{21}$ $y_{22}$ $y_{23}$ . . . $y_{2N}$
	$y_{31}$ $y_{32}$ $y_{33}$ . . . $y_{3N}$
	. . . . . $y_{ji}$ . . .
	$y_{n1}$ $y_{n2}$ $y_{n3}$ . . . $y_{nN}$
$n$ error factors	$z_{11}$ $z_{12}$ $z_{13}$ . . . $z_{1N}$
	$z_{21}$ $z_{22}$ $z_{23}$ . . . $z_{2N}$
	$z_{31}$ $z_{32}$ $z_{33}$ . . . $z_{3N}$
	. . . . . $z_{ji}$ . . .
	$z_{n1}$ $z_{n2}$ $z_{n3}$ . . . $z_{nN}$

assumed to be absent. The last  $n$  cells of each column represent the variable errors in the  $n$  standard test scores of an individual.

The notation  $r$  for the number of common factors may be confusing on first sight, since the same letter is used for the coefficient of correlation; but the correlation will always be designated with a double subscript for the two variables. The notation  $r$  is retained for the number of common factors since it is a customary notation for the rank of a matrix, and it will be shown that the number of common factors is the rank of the correlational matrix.

Since equation (1) implies the product of two matrices, it is of interest to write both of them. The  $x$ 's in (1) refer to the population matrix  $P_4$  which is descriptive of the subjects. The  $a$ 's in (1) are descriptive of the tests. The description of the tests can be written in the form of a test matrix or *factorial matrix*, which is shown in *Table 2*.

Table 2  
Factorial Matrix  $F_4$

	$r$ common factors	$n$ specific factors	$n$ error factors
$n$ tests	$a_{11}$ $a_{12}$ $a_{13}$ . . . $a_{1r}$	$b_{11}$	$c_{11}$
	$a_{21}$ $a_{22}$ $a_{23}$ . . . $a_{2r}$	$b_{22}$	$c_{22}$
	$a_{31}$ $a_{32}$ $a_{33}$ . . . $a_{3r}$	$b_{33}$	$c_{33}$
	. . . . . $a_{jm}$ . . .	$b_{jj}$	$c_{jj}$
	$a_{n1}$ $a_{n2}$ $a_{n3}$ . . . $a_{nr}$	$b_{nn}$	$c_{nn}$

The additional notation is as follows:

- $a_{jm}$  ≡ loading of the common factor  $m$  in test  $j$ ,
- $b_{jj}$  ≡ loading of the specific factor of test  $j$  in test  $j$ ,
- $c_{jj}$  ≡ loading of the error factor of test  $j$  in test  $j$ .

Each row of the factorial matrix describes a test. The first  $r$  columns are filled, since each test may have a loading in each of the common abilities. A common-factor test loading is frequently zero; and this is, in fact, the situation that should be explicitly planned for in setting up factorial experiments. Since, by definition, there is only one specific factor in each test, the second section of  $F_4$  is necessarily a diagonal arrangement of the specific factor loadings  $b$ . The same is true for the error factor loadings  $c$ , which have a diagonal arrangement.

It will be assumed that the first  $r$  columns of  $F_4$  are linearly independent. This is a postulate concerning the test battery which is represented in  $F_4$ .

Postulate.

*The  $n$  tests which constitute the battery are so selected that the columns of the factorial matrix are linearly independent.*

It would be difficult to set up a battery which would violate this postulate. It would be a rare occurrence for the columns of  $F$  to be dependent when the number of tests is considerably in excess of the number of factors. If such a battery were to be assembled, the factorial solution would be a matrix  $F$  which reproduced  $R$  with less than the true number of reference abilities. This is probably a remote contingency. The geometrical re-







representation of such a solution would be a set of  $n$  radial vectors, one for each test, which would lie in a space of a number of dimensions less than the number of common factors.

Tables 1 and 2 show the two matrices whose row-by-column multiplication is implied in an equation of the type (1). These two matrices are the population matrix  $P_4$  and the test matrix or factorial matrix  $F_4$ . Rewriting (1) in matrix notation, we have

$$(2) \quad S = F_4 P_4 .$$

Inspection of the matrices  $F_4$  and  $P_4$  reveals that they may be written in several sections. The population matrix may be written in three sections, as shown in Table 3. Comparison of Tables 1 and 3 shows that the matrix  $P_4$

Table 3  
Three Components of the Population Matrix

	N	N	N
r com- mon factors	$x_{11} \ x_{12} \ x_{13} \ \dots \ x_{1N}$ $x_{21} \ x_{22} \ x_{23} \ \dots \ x_{2N}$ $x_{31} \ x_{32} \ x_{33} \ \dots \ x_{3N}$ $\dots \dots \dots x_{mi} \ \dots$ $x_{r1} \ x_{r2} \ x_{r3} \ \dots \ x_{rN}$		
n specific factors		$y_{11} \ y_{12} \ y_{13} \ \dots \ y_{1N}$ $y_{21} \ y_{22} \ y_{23} \ \dots \ y_{2N}$ $y_{31} \ y_{32} \ y_{33} \ \dots \ y_{3N}$ $\dots \dots \dots y_{ji} \ \dots$ $y_{n1} \ y_{n2} \ y_{n3} \ \dots \ y_{nN}$	
n error factors			$z_{11} \ z_{12} \ z_{13} \ \dots \ z_{1N}$ $z_{21} \ z_{22} \ z_{23} \ \dots \ z_{2N}$ $z_{31} \ z_{32} \ z_{33} \ \dots \ z_{3N}$ $\dots \dots \dots z_{ji} \ \dots$ $z_{n1} \ z_{n2} \ z_{n3} \ \dots \ z_{nN}$
	Matrix $P_1$ for common factors	Matrix $P_2$ for specific factors	Matrix $P_3$ for error factors

may be written as the sum of three matrices, namely,

matrix  $P_1$  for the common factors,  
 matrix  $P_2$  for the specific factors,  
 matrix  $P_3$  for the error factors,

so that

$$(3) \quad P_4 = P_1 + P_2 + P_3 .$$

The factorial matrix  $F_4$  may also be expressed as a sum of three parts in a similar manner. This is shown in *Table 4*.

*Table 4*  
*Three Components of the Factorial Matrix*

	$r$	$n$	$n$
$F_1 \equiv$	$a_{11} \ a_{12} \ a_{13} \ \dots \ a_{1r}$ $a_{21} \ a_{22} \ a_{23} \ \dots \ a_{2r}$ $a_{31} \ a_{32} \ a_{33} \ \dots \ a_{3r}$ $\dots \ \dots \ \dots \ a_{jm} \ \dots$ $a_{n1} \ a_{n2} \ a_{n3} \ \dots \ a_{nr}$	○	○
$D_1 \equiv$	○	$b_{11}$ $b_{22}$ $b_{33}$ $b_{jj}$ $b_{nn}$	○
$D_2 \equiv$	○	○	$c_{11}$ $c_{22}$ $c_{33}$ $c_{jj}$ $c_{nn}$

The three parts represent the common factors, the specific factors, and the error factors. The factorial matrix  $F_4$  may be written as a sum of these three parts, namely,

$$(4) \quad F_4 = F_1 + D_1 + D_2 .$$

Since, by definition, there is only one specific factor in each test, the middle section,  $D_1$ , is a diagonal matrix. By the same reasoning the third section,  $D_2$ , is a diagonal matrix in which each entry shows the error factor of a test.

Returning now to equation (2), we may express the standard scores in terms of the three kinds of factors. Substituting (3) and (4) in (2), we have

$$(5) \quad \begin{cases} S = F_4 P_4 , \\ S = (F_1 + D_1 + D_2)(P_1 + P_2 + P_3) , \\ S = F_1 P_1 + D_1 P_2 + D_2 P_3 . \end{cases}$$

A single element of the  $n \times N$  matrix  $S$  is the standard score of individual  $i$  in test  $j$ . It can be written as follows:

$$(6) \quad s_{ji} = \sum_{m=1}^r a_{jm} x_{mi} + b_{ji} y_{ji} + c_{ji} z_{ji} .$$

By definition, the sum of the standard scores,  $\sum_{i=1}^N s_{ji}$ , of the population  $N$  in test  $j$  is equal to zero. The sum may be written as follows:

$$(7) \quad \sum_{i=1}^N s_{ji} = \sum_{m=1}^r a_{jm} \sum_{i=1}^N x_{mi} + b_{ji} \sum_{i=1}^N y_{ji} + c_{ji} \sum_{i=1}^N z_{ji} = 0 .$$

By definition, the sum of the squares of the standard scores of the  $N$  subjects must equal  $N$ . Then

$$(8) \quad \sum_{i=1}^N s_{ji}^2 = N ,$$

and hence

$$(9) \quad \frac{1}{N} \sum_{i=1}^N s_{ji}^2 = 1 \equiv \text{total variance of test } j .$$

Since the factors are uncorrelated, we have

$$(10) \quad \sum_{i=1}^N x_{mi}x_{Mi} = \sum_{i=1}^N y_{ji}y_{ki} = \sum_{i=1}^N z_{ji}z_{ki} = 0,$$

where  $m$  and  $M$  refer to any pair of common abilities ( $m \neq M$ ), and where  $j$  and  $k$  refer to any pair of tests ( $j \neq k$ ). For the same reason the following cross products also vanish:

$$(11) \quad \sum_{i=1}^N x_{mi}y_{ji} = \sum_{i=1}^N x_{mi}z_{ji} = \sum_{i=1}^N y_{ji}z_{ki} = \sum_{i=1}^N y_{ji}z_{ji} = 0.$$

Substituting (6) in (9) and ignoring the vanishing cross products of (10) and (11), we have

$$(12) \quad \frac{1}{N} \sum_{i=1}^N \sum_{m=1}^r a_{jm}^2 x_{mi}^2 + \frac{1}{N} b_{jj}^2 \sum_{i=1}^N y_{ji}^2 + \frac{1}{N} c_{jj}^2 \sum_{i=1}^N z_{ji}^2 = 1.$$

Since  $x_{mi}$  and  $y_{ji}$  are standard scores,

$$(13) \quad \frac{1}{N} \sum_{i=1}^N x_{mi}^2 = \frac{1}{N} \sum_{i=1}^N y_{ji}^2 = 1.$$

Since

$$z_{ji} = \frac{e_{ji}}{\epsilon_j},$$

it follows that

$$(14) \quad z_{ji}\epsilon_j = e_{ji}.$$

Then

$$(15) \quad z_{ji}^2 \epsilon_j^2 = e_{ji}^2$$

and

$$(16) \quad \epsilon_j^2 \sum_{i=1}^N z_{ji}^2 = \sum_{i=1}^N e_{ji}^2 \equiv N \epsilon_j^2.$$

Hence

$$(17) \quad \frac{1}{N} \sum_{i=1}^N z_{ji}^2 = 1.$$

Substituting (13) and (17) in (12), we have

$$(18) \quad \sum_{m=1}^r a_{jm}^2 + b_{jj}^2 + c_{jj}^2 = 1,$$

in which the total variance of test  $j$  is expressed as a sum of three variances due to (a) the common factors, (b) the specific factors, and (c) the error factors. The  $(r+2)$  test coefficients in terms of which test  $j$  is described are the  $r$  values of  $a_{jm}$  and the values of  $b_{jj}$  and  $c_{jj}$ . Equation (18) can be restated as follows:

Theorem. *The sum of the squares of the test coefficients of a test is equal to unity.*

In fact,  $a_{jm}^2$ , the square of a test coefficient, is that part of the total variance of a test  $j$  which is attributable to the factor  $m$ . In the same manner  $b_{jj}^2$  is that part of the variance of the test  $j$  which is attributable to the specific factor in test  $j$ . Also,  $c_{jj}^2$  is the part of the variance of test  $j$  which is due to the variable chance errors in the scores of test  $j$ .

### Communality

It has been shown that the total variance of a test can be expressed as the sum of three variances which are due to (a) the abilities which are common to two or more of the tests, (b) the abilities which are unique in that they are called for by only one test, and (c) the variable errors. It will be convenient to name these three parts of the total variance. The following terminology\* will be used:

$$\begin{aligned} \sum_{m=1}^r a_{jm}^2 &\equiv h_j^2 \equiv \text{communality of test } j, \\ b_{jj}^2 &\equiv b_j^2 \equiv \text{specificity of test } j, \\ c_{jj}^2 &\equiv c_j^2 \equiv \text{error variance of test } j. \end{aligned}$$

The concept of communality is pivotal in factor analysis, and it will be necessary to refer to it frequently.

Definition 11. *The communality of a test is its common factor variance.*

\* "Theory of Multiple Factors," p. 8.

The object of factor analysis is to isolate the  $r$  values  $a_{jm}$  for each test  $j$  and the  $r$  values  $x_{mi}$  for each person  $i$ .

Since, in factor analysis, the specificity and the error variance combine into a single variance that is unique for each test, it will be convenient to combine them in the following manner:

$$b_j^2 + c_j^2 \equiv u_j^2 \equiv \text{uniqueness of test } j,$$

so that

$$(19) \quad h_j^2 + u_j^2 = 1.$$

Equation (19) shows that the variance of a test may be expressed as the sum of two parts, namely, the communality and the uniqueness.

Definition 12. *The uniqueness of a test is the complement of its communality.*

An interpretation of (19) is that the total variance of a test can be divided into two parts: namely, the communality, that part of its variance which is due to factors common to other tests in the battery; and the uniqueness, that part of its variance which is due to factors not common to other tests in the battery.

This distinction between communality and uniqueness is crucial in factorial analysis. If a test calls for several abilities which are unique in that they are not called for by any other tests in the battery, then these unique abilities combine with the error variance into a single specific variance for the test. The isolation of these unique abilities and the appraisal of individuals with reference to them cannot be effected by factorial methods until the test is inserted in a battery with other tests that do contain these abilities. Then the abilities which combine into a single specific factor in one battery become separate common factors in the new battery. They can then be isolated.

An object of psychological inquiry is to isolate an increasing number of abilities until the specific variance of each important test shall be reduced to a minimum. It is not likely that any single test will be completely described in terms of the factors which it has in common with those of one battery. In order to isolate all of the abilities that are called for by a test, it will probably be necessary to insert it in several test batteries in succession. The specific variance of a test should be regarded as a challenge; it is that part of the total variance of a test which is unique in a particular battery, and hence its factorial composition is unknown. In order to test a hypothesis concerning the abilities which are involved in the specific variance of a test, the test should be combined with others which involve the hypo-



tical abilities. If the specific variance is reduced the hypothesis is sustained.

For the next few years it will probably be more interesting to isolate new abilities than to reduce the specificity in particular tests. Increased knowledge of the primary mental abilities will facilitate the type of experiment by which the specificities of particular tests may be reduced. It will probably be found that a considerable fraction of the total variance of each test is attributable to factors of such limited social significance that the complete elimination of the specificity of each test will not be essential in the early stages of the scientific study of human abilities.

### The intercorrelations

Since  $s_{ji}$  are standard scores, the intercorrelation between two tests  $j$  and  $k$  can be written in the simple form

$$(20) \quad r_{jk} = \frac{1}{N} \sum_{i=1}^N s_{ji}s_{ki} .$$

This implies the multiplication of two matrices. The elements of a *moment matrix*  $M$  may be defined as follows:

$$(21) \quad m_{jk} = \sum_{i=1}^N s_{ji}s_{ki} = \sum_{i=1}^N s_{ji}s'_{ik} ,$$

so that we have, in matrix notation,

$$(22) \quad M = SS' .$$

Substituting (5) in (22),

$$(23) \quad M = (F_1P_1 + D_1P_2 + D_2P_3)(F_1P_1 + D_1P_2 + D_2P_3)' ,$$

$$(24) \quad = (F_1P_1 + D_1P_2 + D_2P_3)(P_1'F_1' + P_2'D_1' + P_3'D_2') .$$

Six of the terms of this product vanish. Hence

$$(25) \quad M = F_1P_1P_1'F_1' + D_1P_2P_2'D_1' + D_2P_3P_3'D_2' .$$

The population matrix  $P_1$  is orthogonal by rows, as can be seen by reference to (10) and (13). Hence

$$(26) \quad P_1 P_1' = D_3,$$

in which  $D_3$  is a diagonal matrix of order  $q \times q$  and where

$$q \equiv 2n + r \equiv \text{total number of factors.}$$

The matrix  $D_3$  has the constant element  $N$  in the  $r$  diagonal cells of the first  $r$  rows and columns, and zero in all other cells.

Similar reasoning applies to  $P_2$ . Hence

$$(27) \quad P_2 P_2' = D_4,$$

in which  $D_4$  is a diagonal matrix of order  $q \times q$  with constant element  $N$  in the  $n$  diagonal cells of the rows and columns  $(r+1)$  to  $(r+n)$  inclusive, and zero in all other cells.

By the same reasoning

$$(28) \quad P_3 P_3' = D_5,$$

in which  $D_5$  is a diagonal matrix of order  $q \times q$  with constant element  $N$  in the  $n$  diagonal cells of the rows and columns  $(r+n+1)$  to  $(r+2n)$ , inclusive, and zero in all other cells.

Substituting (26), (27), and (28) in (25),

$$(29) \quad M = F_1 D_3 F_1' + D_1 D_4 D_1' + D_2 D_5 D_2',$$

$$(30) \quad = N F_1 F_1' + N D_1 D_1' + N D_2 D_2',$$

$$(31) \quad = N (F_1 F_1' + D_1 D_1' + D_2 D_2').$$

By (20) and (21) we have

$$(32) \quad R_1 = \frac{1}{N} M,$$

where  $R_1$  is a square matrix of order  $n$ , the cells of which contain the true intercorrelations of the fallible tests. From (31) and (32) it follows that

$$(33) \quad R_1 = F_1 F_1' + D_1 D_1' + D_2 D_2',$$

or

$$(34)$$

The correlational entries of  $R_1$  are as follows:

$$(35) \quad r_{kk} = \sum_{m=1}^r a_{km}^2 + b_k^2 + c_k^2 = 1,$$

$$(36) \quad r_{kk} = h_k^2 + b_k^2 + c_k^2 = 1,$$

and

$$(37) \quad r_{jk} = \sum_{m=1}^r a_{jm}a_{km}, \quad \text{where } j \neq k.$$

Equations (35) and (37) show that the terms  $D_1^2$  and  $D_2^2$  of (34) affect only the diagonal entries of  $R_1$ . If a new matrix  $R$  is defined by the relation

$$(38) \quad R = F_1 F_1',$$

then  $R_1$  and  $R$  are identical except for the diagonal entries. By (35) the diagonal entry of  $R_1$  is 1. The diagonal entry of  $R$  in the column  $j$  and row  $j$  is the communality  $h_j^2$ . The matrix  $R$  will be called the *reduced matrix of the true correlations of fallible tests*. It will be referred to more briefly as a "reduced correlational matrix." The matrix  $R_1$  will be called the *complete matrix of correlations of fallible tests*. It will be referred to as a "complete correlational matrix" in the sense that the complete variance of each test is represented by the diagonal entries.

Let  $F$  be the matrix formed by the  $n$  rows and the first  $r$  columns of  $F_1$ . Then, by (38),

$$(39) \quad R = FF', \text{ \textit{Ward's method}}$$

in which the reduced correlational matrix is defined in terms of the common factors. The matrix  $F$  is an  $n \times r$ -rowed matrix which shows the weights of the  $r$  common factors in the  $n$  tests. This matrix will be called the "matrix of the common factors" or, more briefly, the "factorial matrix." Since in factorial analysis it is the common factors that are of principal interest, there is no confusion in referring to  $F$  as the factorial matrix without qualification for the common factors.

### The reliability coefficient

It is customary in psychological work to write the reliability coefficient in the diagonal cells of a correlation matrix. By the present analysis it is seen that the diagonal entries of  $R_1$  are unity, while the diagonal entries of  $R$  are the communalities  $h_j^2$ . The relation between the reliability and the com-

munity of a test may be shown by considering in detail the factorial matrix for a test  $j$ , a parallel test  $j'$ , another test  $k$ , and its parallel test  $k'$ . The factorial matrix for these four tests is shown in *Table 5*.

Let there be  $r$  common factors in the four tests. Let  $b_j^2$  be the specific variance in test  $j$ . Since  $j$  and  $j'$  are parallel tests, it is evident that they must require the same common abilities and the same specific ability. Hence  $b_j$  is recorded in the same column of  $F_4$  for both  $j$  and  $j'$ . For the same reason  $b_k$  must be common to tests  $k$  and  $k'$ , which are parallel. But the variable errors are uncorrelated by definition, even for parallel tests. Hence  $F_4$  of *Table 5* shows a separate error factor for each of the four tests.

*Table 5*  
Factorial Matrix  $F_4$  for Four Tests,  $j$  and Its Parallel  $j'$ , and  $k$  and Its Parallel  $k'$ .

$r$ common factors					Two specific factors		Four error factors			
$a_{j1}$	$a_{j2}$	$a_{j3}$	...	$a_{jr}$	$b_j$	○	$c_j$	○	○	○
$a_{j'1}$	$a_{j'2}$	$a_{j'3}$	...	$a_{j'r}$	$b_j$	○	○	$c_{j'}$	○	○
$a_{k1}$	$a_{k2}$	$a_{k3}$	...	$a_{kr}$	○	$b_k$	○	○	$c_k$	○
$a_{k'1}$	$a_{k'2}$	$a_{k'3}$	...	$a_{k'r}$	○	$b_k$	○	○	○	$c_{k'}$

The true correlation between the fallible parallel tests  $k$  and  $k'$  is the reliability of  $k$ . The complete correlational matrix is

$$(40) \quad R_1 = F_1 F_1' + D_1^2 + D_2^2 .$$

But  $D_2$  in *Table 5* does not contribute to the reliability coefficient  $r_{jj'}$  which is not a diagonal entry. The matrix  $D_1$  does contribute to the reliability coefficient because the specificity is an additional common factor in the special case of *Table 5*. Hence

$$(41) \quad r_{kk'} = \sum_{m=1}^r a_{km}^2 + b_k^2 ,$$

or

$$(42) \quad r_{kk'} = h_k^2 + b_k^2 = \text{reliability of test } k .$$

By (36) and (42),

$$(43) \quad r_{kk'} = 1 - c_k^2 .$$

Equation (43) merely states that *the reliability coefficient of a test is the complement of its error variance.*

Since

$$(44) \quad h_k^2 = 1 - u_k^2$$

and

$$(45) \quad u_k^2 = b_k^2 + c_k^2,$$

it follows that

$$(46) \quad h_k^2 \leq r_{kk'}.$$

*Theorem. The communality of a test is always smaller than the reliability except in the limiting case where the specific factor is absent, in which case the communality and the reliability are equal.*

It is of interest to note that the uniqueness cannot be separated into its two parts, the specificity and the error variance, by factorial methods. In order to estimate the specific variance of a test, it is necessary to estimate its reliability by experimental means. The uniqueness can be determined by factor methods. The specificity is then

$$(47) \quad b_k^2 = u_k^2 - c_k^2,$$

where

$$(48) \quad c_k^2 = 1 - r_{kk'}.$$

Since  $r_{kk'}$  can be estimated only more or less roughly by various experimental methods, it is clear that estimates of specific variance are necessarily equally uncertain.

The terminology for the different parts of the variance of a test is summarized as follows:

$$\begin{aligned} \text{Total variance} &= h_k^2 + b_k^2 + c_k^2 = 1. \\ \text{Reliability} &= h_k^2 + b_k^2 = 1 - c_k^2. \\ \text{Communality} &= h_k^2 = 1 - u_k^2. \\ \text{Specificity} &= b_k^2. \\ \text{Uniqueness} &= b_k^2 + c_k^2 = u_k^2. \\ \text{Error variance} &= c_k^2 = 1 - r_{kk'}. \end{aligned}$$

### The population space

The population matrix of *Table 1* may be regarded as exhibiting  $N$  coordinates for each of  $(r+2n)$  points in a population space of  $N$  dimensions. Each individual of an infinite population may be regarded as defining an

orthogonal reference vector. The sample of the population may be regarded as defining a set of orthogonal reference axes in as many dimensions as there are individuals in the sample. The factors may be regarded as vectors in the population space. By (13) and (17) it is seen that the  $(r+2n)$  factorial vectors are all at the distance  $\sqrt{N}$  from the origin in the population space. By (10) and (11) these vectors are orthogonal in the same space. Hence the entries of  $F_4$  may be regarded as  $\sqrt{N}$  times the direction cosines of  $(r+2n)$  orthogonal factorial vectors in the population space. Since the factors are represented by orthogonal vectors in the population space, it follows that the total factor space is a subspace of the population space.

The factorial matrix  $F_4$  may be regarded as the co-ordinates of  $n$  points in the same space. By (18) these points are at unit distance from the origin. The entries of  $F_4$  may be regarded as the direction cosines of  $n$  unit test vectors in the factor space.

The score matrix may be regarded as exhibiting the co-ordinates of each test in the population space of  $N$  dimensions. The cells of the moment matrix  $M$  show  $N$  times the scalar products of pairs of test vectors. The complete correlational matrix  $R_1$  shows the scalar products of pairs of test vectors in the population space, while the reduced correlational matrix  $R$  shows the scalar products of the projections of these test vectors in the common-factor subspace of the population space. Each test is represented by a unit vector in the population space. The square of the length of its projection in the common-factor subspace is its communality.

### The common-factor space

The geometrical representation of the factorial matrix is fundamental in factor analysis. The factor matrix of *Table 2* can be regarded as exhibiting the  $(r+2n)$  co-ordinates of  $n$  points in a total factor space of  $(r+2n)$  dimensions. The points may also be regarded as the termini of as many test vectors. Each test is then a unit vector in the total factor space. The scalar product of a pair of test vectors is the correlation between the two tests.

Since it is the common factors that are of primary interest in factor analysis, it is profitable to consider mainly the common-factor space. The *common-factor space* is defined by the first  $r$  columns of the factorial matrix  $F_4$ . It shows the  $r$  co-ordinates of each of  $n$  tests in a common-factor space of  $r$  dimensions. Here, again, the scalar product of a pair of test vectors is the correlation between the tests. The correlation is unaffected by the projections of the test vectors into the specific space and into the error space because these projections are orthogonal by definition. The length of each test vector in the common-factor space is the square root of its communality. The complement of the communality of each test is the square root of its projection in the unique factor space.

## CHAPTER II

### THE FUNDAMENTAL FACTOR THEOREM

#### The correlational matrices

The factor theorem which is basic for the present analysis is equation (39-i), namely,

$$FF' = R.$$

It states, in matrix notation:

Theorem 1. *The product of the factorial matrix and its transpose is the reduced correlational matrix.*

In the theoretical development of this theorem in the previous chapter the attributes of the individuals and of the tests were chosen as natural starting points, so that  $R$  could be written if  $F$  were known. The present scientific problem is the reverse. It is the intercorrelations,  $R_0$ , that are known. The object of the factorial analysis is to find  $F$ . The theory, as well as the statistical methods that are involved in factor analysis, is implied in this reversal, namely, that when  $R$  is given experimentally, the problem is to find  $F$ .

By the definition of a correlation coefficient in (22-i) and (32-i) it follows that

$$(1) \quad R_1 = \frac{1}{N} SS',$$

and hence the correlational matrix is symmetric and factorable. It can be shown that  $R_1$  is a positive-definite matrix. From this follows the fundamental factor theorem:

Theorem 2. *For any correlational matrix  $R$  there exists a corresponding factorial matrix  $F$  such that  $FF' = R$ .*

The bold-faced notation  $R$  refers to any correlational matrix. It may have any values in the diagonals which preserve the Gramian properties of the matrix. Hence  $R$  may contain unity or the reliabilities or the communalities in the diagonal cells. The bold-faced notation  $F$  refers to any factorial matrix which reproduces  $R$ .

To write the factorial description of the tests in the form of matrix  $F$  implies, of course, that an orthogonal co-ordinate system is given. In the reverse problem an interesting indeterminacy appears as regards the co-ordi-

nate system in that the co-ordinate reference axes are not defined by the correlational matrix  $R$ . It has been shown that the entries of  $R$  which are the true correlations of fallible tests can be regarded as the scalar products of pairs of vectors. Such a product is a function of the scalars of the two vectors and the angle of separation between the vectors. But all three of these quantities are independent of the location of the orthogonal axes of reference. Rotation of the orthogonal co-ordinates implies:

*Theorem 3. An infinite number of matrices  $F$  can be written which will reproduce a given correlational matrix  $R$ .*

In order that a unique solution of  $F$  may be found for any given matrix  $R$ , it will therefore be necessary to impose further restrictions on the solution. Such additional criteria are to be found in the psychological considerations that govern the problem.

Considerable psychological interest attaches to the signs of the co-ordinates which constitute the entries of  $F$ . If the variables are traits of people, it is usually possible to ascribe acceptable meanings to both positive and negative co-ordinates. If cheerfulness is one of the orthogonal axes, there is no difficulty in defining a personality trait as a vector with either positive or negative projection on the reference axis of cheerfulness. Thus, grouchiness might be a vector with a negative projection on cheerfulness.

The case seems to be different with those traits which concern the things that people can do. These are the traits which have been defined as abilities. An individual can, of course, be described as above or below the mean of a random sample of the population with regard to any specified ability; but, with current psychological concepts, it is preferable to avoid a formulation by which a task might have a negative projection on an axis of reference which defines an ability. One psychological interpretation would be that the performance of such a task is actually facilitated by some sort of ability which is less than totally absent!

Since the signs of the entries in  $F$  are of considerable interest, the following theorems will be found useful.

*Theorem 4. The signs of all the entries in a column of  $F$  may be changed without altering the correlational matrix  $R$ .*

This may be seen from the factor theorem 1. It may also be inferred from the geometrical consideration that the scalar products of  $R$  are independent, not only of the precise locations of the orthogonal co-ordinate axes, but also of reversal of their direction, as represented by a reversal of sign of the co-ordinates in a column of  $F$ . This geometrical fact has a psychological counterpart. The correlation between any two traits remains unaffected by the arbitrary decision to call one of the component reference traits "plus cheerfulness" or "minus gloominess." The theorem can be inferred algebraically



from (35-i) and (37-i), where it is seen that a change in sign of  $a_{jm}$  and  $a_{km}$  for a fixed column  $m$  does not alter the value of  $r_{jk}$ .

**Theorem 5.** *If all of the signs are reversed in a row of  $F$ , then all the signs are reversed in the corresponding row and the corresponding column of  $R$ .*

To change the signs in a row of  $F$  is to reverse the direction of a test vector. Its scalar remains the same, while its angular separation  $\phi$  from any other test vector is changed to the supplement of  $\phi$ . Hence the absolute values of the correlations of this test with the other tests remain unaltered, but their signs are reversed. The psychological interpretation can be shown by an example, namely, that if one variable correlates positively with "plus tactfulness," then it will correlate negatively with "minus tactfulness," which might be defined as "plus tactlessness."

This theorem can also be inferred algebraically from (37-i), where it is seen that a change in sign of  $a_{jm}$  for a test  $j$  alters the sign of  $r_{jk}$  where  $j \neq k$ . From (35-i) it is seen that when  $a_{km}$  is reversed in sign for a test  $k$ , the value of  $r_{jk}$  is not changed in sign for  $j = k$ . The self-correlation remains positive for all possible reversals of sign of tests and factors.

### The number of independent factors

One of the principal problems in factor analysis is to ascertain the number of linearly independent factors that must be postulated in order to describe the scores in the tests as linear combinations of the factors. The columns of  $F$  represent independent factors, so that the number of independent factors is the number of columns of  $F$ . But this is also the rank of  $F$ . It can be shown that the ranks of  $R$  and of  $F$  are always the same. Hence we have:\*

**Theorem 6.** *The number of linearly independent factors represented by the intercorrelations of  $n$  tests is equal to the rank of their correlational matrix  $R$ .*

Owing to sampling errors, the experimentally obtained correlation coefficients are not the true intercorrelations which are defined as the cell entries of  $R$ . The experimentally obtained correlations constitute a square matrix of order  $n$  which will be designated  $R_0$ . The distribution of discrepancies between the experimental values in  $R_0$  and the corresponding true correlations of the fallible tests in  $R$  should have a dispersion not excessively greater than that to be expected from the known standard errors of the experimental coefficients.

Since the sampling errors in  $R_0$  are fortuitous, it should be expected that the rank of  $R_0$  is equal to its order, namely  $n$ . The theorem concerning the number of factors shows that the number of common factors that are required in order to account exactly for the coefficients in  $R_0$  is equal to the

\* "Theory of Multiple Factors," p. 20.

number of tests. Such a solution is of no scientific interest. It corresponds to the more obvious situation in which the number of parameters in a hypothesis is equal to the number of observations. In a simple curve-fitting problem the analogous situation would be that in which a curve with  $r$  independent parameters is fitted to a set of  $r$  points. The significance of an equation so chosen is not convincing. One of the fundamental principles of science is that the convincingness of a scientific hypothesis varies with the degree to which it is overdetermined by the data. To postulate as many reference abilities as there are tests constitutes the absurdity of postulating as many categories as there are facts to be explained or described. To do so would be to acknowledge the defeat of scientific effort.

The problem of describing factorially the variables whose experimental intercorrelations are given in  $R_0$  is essentially that of finding another matrix  $R$  (a reduced correlational matrix) of lowest possible rank whose cell entries do not deviate from those of  $R_0$  by more than might be expected from the sampling errors in the experimental coefficients of  $R_0$ . If such a matrix  $R$  can be found, in which the rank  $r < n$ , a scientifically significant solution  $F$  may be possible. The converse is not necessarily valid, since the present reasoning is based on a set of postulates which by no means exhaust the possible ideal constructs in terms of which the variables may be described. But in any event the number of degrees of freedom of the construct must be considerably smaller than that of the experimental data that are to be unified.

In dealing with the experimentally obtained values in the correlational matrix  $R_0$ , it must be remembered that the diagonal entries are unknown. The communalities are numbers between 0 and +1. If the smallest number of factors in terms of which the scores can be linearly expressed is  $r$ , then the factorial matrix  $F$  will have  $r$  linearly independent columns. But the number of columns is then the rank of  $F$ . Since the rank of  $F$  and the rank of  $R$  are always the same, we have the following theorem:

*Theorem 7. The smallest number of independent factors that will account for the intercorrelations of  $n$  tests is the minimum rank of the correlational matrix with the diagonal entries treated as unknown positive values between 0 and +1.*

### Algebraic and configurational uniqueness

It has been shown that if a factorial matrix  $F$  has been found such that  $FF' = R$ , the solution  $F$  is not algebraically unique because the co-ordinate system of  $F$  may be rotated arbitrarily without affecting the reproduction of the correlations in  $R$ .<sup>\*</sup> Such rotation alters the numerical entries in  $F$ . It is in this sense that the matrix  $F$  is not algebraically unique.

<sup>\*</sup> "Theory of Multiple Factors," p. 10.

The entries in  $F$  may be regarded as the orthogonal co-ordinates of  $n$  points in the common-factor space. These points constitute the termini of the test vectors whose scalar products are shown in  $R$ . In this sense both  $F$  and  $R$  represent the same configuration. If there exists only one configuration that will satisfy  $R$ , then there is only one configuration that can be represented by  $F$ . Rotation of the co-ordinate system does not alter the configuration either in  $R$  or in  $F$ . It is in this sense that  $F$  may be a unique solution to the factor problem which is stated in  $R$ . If, on the other hand, the given matrix  $R$  with unknown diagonal entries does not define a unique configuration, then any corresponding matrix  $F$  cannot be unique.

Since the psychological problem consists in describing the abilities that are represented in the common-factor space, it seems evident that no psychologically meaningful solution can be expected unless the given matrix  $R$  defines a unique configuration in the common-factor space. It is therefore of considerable importance to ascertain the conditions under which a unique configuration is defined by the given intercorrelations.

This problem may be clarified by a very simple but extreme example of a correlational matrix which does not define a unique configuration. Consider a set of two tests. The correlational matrix is of order 2; and it contains only one intercorrelation in addition to the two communalities, which are unknown. If two abilities are involved, the rank of the correlational matrix must be 2. The two diagonals may be given any values between 0 and 1 by which the rank remains 2. Any pair of diagonal values defines the scalars of the two vectors. The angular separation is determined so that the scalar product is equal to the observed intercorrelation of the two tests. It is evident that for each pair of arbitrary diagonal values a different configuration will be obtained. Evidently, then, the two tests are not sufficient to define two common factors or abilities. The same type of reasoning can be extended to more tests and to higher dimensions.

The relation between the number of tests  $n$  and the number of independent factors  $r$  is subject to a limitation with regard to the present scientific problem. The number of reference abilities in  $n$  tests must satisfy one of the three following possibilities, namely,  $r > n$ ,  $r = n$ , or  $r < n$ . By the factor theorem (1) it is seen that  $n$  tests will produce a correlational matrix whose rank will not exceed its order  $n$ . If, then,  $r > n$ , the factors cannot be isolated by factorial methods. If more than  $n$  factors are involved, it is necessary to augment the test battery with additional tests before the reference factors can be isolated. If  $r = n$ , there are as many factors as there are tests. Such a solution is always possible, and it is therefore trivial as far as the scientific problem is concerned. The solution in which  $r = n$  violates the fundamental postulate of science that every valid hypothesis is overdetermined

by the data. This case is discussed further in *chapter iv* on “*The Principal Axes.*” The only allowable case is that in which  $r < n$ . This leads to the following postulate.

Postulate. *The number of reference abilities in a test battery is less than the number of tests.*

This condition must be satisfied, or the reference abilities cannot be isolated by factorial methods. In setting up a test battery for the purpose of discovering the primary abilities, the experimenter must so select the tests that the number of primary abilities is smaller than the number of tests in the battery. A more exact relation between  $r$  and  $n$  which must be satisfied in order that a unique solution shall exist will now be shown.

The number of intercorrelations in  $R$  which are to determine the configuration is

$$\frac{n(n-1)}{2}.$$

These intercorrelations constitute the observations. The number of parameters in  $F$  is  $nr$ , but this number can be reduced. If the first co-ordinate axis is passed through the first test, then

$$a_{12} = a_{13} = a_{14} = \dots = a_{1r} = 0.$$

The second orthogonal axis may be so placed that test 2 lies in the I-II plane. Then

$$a_{23} = a_{24} = \dots = a_{2r} = 0.$$

Table 1

$r$  common factors

	$a_{11}$	0	0	0	...	0
	$a_{21}$	$a_{22}$	0	0	...	0
	$a_{31}$	$a_{32}$	$a_{33}$	0	...	0
	...	...	...	...	...	...
	$a_{r1}$	$a_{r2}$	$a_{r3}$	$a_{r4}$	...	$a_{rr}$
	...	...	...	...	...	...
	$a_{n1}$	$a_{n2}$	$a_{n3}$	$a_{n4}$	...	$a_{nr}$

This process can be continued until there are one or more zero co-ordinates for each of the first  $(r-1)$  tests. The factorial matrix will then appear like *Table 1*, which has been arranged so as to represent  $n$  tests and  $r$  factors.

The number of parameters in  $F$  then becomes

$$nr - \frac{1}{2}r(r-1).$$

In order that there shall be a unique solution, the number of experimentally independent values in  $R_0$  must equal or exceed the number of linearly independent\* parameters in  $F$ . Hence

$$(2) \quad nr - \frac{r(r-1)}{2} \leq \frac{n(n-1)}{2}.$$

The condition for a maximum value of  $r$  for a given value of  $n$  is represented by substituting an equality sign for the inequality. The condition then becomes

$$(3) \quad nr - \frac{r(r-1)}{2} = \frac{n(n-1)}{2}$$

or

$$(4) \quad 2nr - r(r-1) - n(n-1) = 0.$$

Solving the quadratic in  $r$ , we have the following theorem.

*Theorem 8. In order that the correlational matrix  $R$  with unknown diagonals for  $n$  tests and  $r$  common factors shall represent a unique configuration, it is in general necessary that*

$$(5) \quad r \leq \frac{(2n+1) - \sqrt{8n+1}}{2}.$$

The suppression of the positive sign before the radical in (5) is justified by the postulate that  $r < n$ . When the equality sign is used in (5), the value of  $r$  becomes integral for certain values of  $n$ . Then the number of independent parameters of  $F$  is exactly equal to the number of experimentally independent coefficients in  $R_0$ . Such is the case when  $n=6$  and  $r=3$ .

\* In mathematical and scientific use the term *independence* has several different meanings. The context usually indicates clearly enough which of several meanings is implied. It may be useful to enumerate three of these meanings. *Linear independence* is here used in the sense in which the term is defined in current mathematical textbooks. The term *statistical independence* is here used to mean zero correlation, i.e., the case in which cross products of two variables vanish. Its geometrical representation is the orthogonality of a pair of vectors. Several values are here said to be *experimentally independent* if they have been separately determined in experimentation.

Since (4) is symmetric in  $n$  and  $r$ ,  $n$  can be expressed explicitly in terms of  $r$  by analogy from (5), so that

$$(6) \quad n \geq \frac{(2r+1) + \sqrt{8r + 1}}{2}.$$

This relation shows the minimum number of tests required for the determination of  $r$  factors. Formula (6) shows, for example, that there must be at least eight tests in order to determine four factors.

It is useful to have a table to show the smallest number of tests that will just determine a given number of factors or the largest number of factors that can just be determined by a given number of tests. This information is summarized for ten factors in *Table 2*.

*Table 2*

No. of Factors $r$	No. of Tests $n$	No. of Factors $r$	No. of Tests $n$
1.....	3*	6.....	10*
2.....	5	7.....	12
3.....	6*	8.....	13
4.....	8	9.....	14
5.....	9	10.....	15*

\* The asterisks refer to integral values of both  $r$  and  $n$  in (6).

**The case of  $n$  tests and  $n$  factors**

There is a simple solution which is satisfactory as long as the factor problem is regarded only in its mathematical aspects but which is fictitious as a solution to the present psychological problem. Since this simple solution with as many factors as there are tests is certain to occur to anyone who studies the factor problem, some discussion of its limitations is in order even though it can be shown to be psychologically trivial.

In this solution each test is represented by a radial unit test vector in space of  $n$  dimensions. Since the scalars are all unity, the angular separations between the vectors must be adjusted in order that the correlations shall represent scalar products of these vectors. In the correlational matrix there are  $n(n-1)/2$  experimentally independent correlation coefficients where unity is written in each diagonal cell. In the factorial matrix with as many factors as there are tests the number of independent parameters is also  $n(n-1)/2$ , since the factorial matrix is normalized by rows. Consequently, it may be expected that an exact solution exists in the form of a square matrix  $F$  of order  $n$  and rank  $n$  which reproduces exactly the experimentally obtained correlation coefficients in  $R_0$ .

The fallacious character of the solution in which there are as many factors as there are tests can be seen by considering the fact that it assumes as many degrees of freedom in the hypothesis  $F$  as there are independent experimental observations in  $R_0$ . This violates the postulate of science that a valid hypothesis is overdetermined by the data. Hence the solution is scientifically trivial even though it is mathematically valid.

That the description of  $n$  tests by as many factors is an erroneous solution can be seen as well from other considerations. If the number of postulated common factors is equal to the number of tests, then it is possible to account for the intercorrelations of  $R_0$  *exactly* by the  $n$  common factors. But the experimentally obtained correlations in  $R_0$  contain the effects of at least three sources of variance which are known to be unique for each test. These are (a) the variable chance errors in the scores of the  $N$  individuals, (b) the specific factors or abilities which are almost certain to be involved in each test of any finite battery, and (c) the sampling errors in the coefficients of  $R_0$ . All three of these sources of variance are unique for each test; and hence they must be accounted for by unique factors, i.e., factors which are, by definition, not common factors. But the solution in which  $n$  common factors account exactly for the  $n$  tests leaves no part of the variance to the unique factors that are known to exist. Hence such a solution can be discarded by psychological considerations apart from mathematical reasoning. The reason why these considerations are not immediately evident in dealing with the factor problem is that the existence of the three sources of unique variance in the  $n$  tests is a scientific fact quite extraneous to the correlational matrix  $R_0$ . In other words, more is known about the tests than is given in the correlational matrix. This additional information, which is not given by the intercorrelations as such, is our knowledge that each test is influenced by factors that are unique and not common. Although it seems evident from scientific, as well as psychological, considerations that the case of  $n$  common factors for  $n$  tests is trivial, there is some interest in knowing that such a solution can be written quite readily for any correlation table.

#### A method of factoring any symmetric matrix\*

The solution to be described is a simple general method of factoring any symmetric matrix. It will be called *the diagonal method*. Let *Table 3* represent a correlational matrix  $R$ , and let *Table 4* represent a factorial matrix  $F$  of order  $n \times r$  in which  $r$  is the rank of  $F$  and the rank of  $R$ . It will be assumed that  $F$  has been rotated as described in a preceding section (*Table 1*) so as to minimize the number of independent parameters.

By the factor theorem (1),

$$(7) \quad r_{11} = a_{11}^2.$$

\* "Theory of Multiple Factors," p. 13.

If the diagonal self-correlations are known, then  $a_{11}$  is known. If the self-correlations are unknown, then  $r_{11}$  may be set equal to unity, in which case  $a_{11}$  is also unity.

The correlation

$$(8) \quad r_{k1} = a_{11}a_{k1},$$

and hence

$$(9) \quad a_{k1} = \frac{r_{k1}}{a_{11}},$$

so that the entries in the first column of  $F$  can be determined.

Table 3

		$n$				
$n$	$r_{11}$	$r_{12}$	$r_{13}$	$\dots$	$r_{1n}$	
	$r_{12}$	$r_{22}$	$r_{23}$	$\dots$	$r_{2n}$	
	$r_{13}$	$r_{23}$	$r_{33}$	$\dots$	$r_{3n}$	
	$\dots$	$\dots$	$\dots$	$(r_{kj} = r_{jk})$	$\dots$	
	$r_{1n}$	$r_{2n}$	$r_{3n}$	$\dots$	$r_{nn}$	

Table 4

		$r$				
$n$	$a_{11}$	0	0	0	0	
	$a_{21}$	$a_{22}$	0	0	0	
	$a_{31}$	$a_{32}$	$a_{33}$	0	0	
	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	
	$a_{r1}$	$a_{r2}$	$a_{r3}$	$\dots$	$a_{rr}$	
	$\dots$	$\dots$	$\dots$	$a_{km}$	$\dots$	
$a_{n1}$	$a_{n2}$	$a_{n3}$	$\dots$	$a_{nr}$		

The correlation

$$(10) \quad r_{22} = a_{21}^2 + a_{22}^2,$$

so that

$$(11) \quad a_{22}^2 = r_{22} - a_{21}^2.$$



Here, as before, a given diagonal entry may be used; but if the diagonal entry is unknown, it can be given an arbitrary value of unity, which means that  $F$  shall represent the total variance of each test.

The correlation

$$(12) \quad r_{k2} = a_{21}a_{k1} + a_{22}a_{k2},$$

so that

$$(13) \quad a_{k2} = \frac{r_{k2} - a_{21}a_{k1}}{a_{22}},$$

and hence the second column of  $F$  can be determined.

The correlation

$$(14) \quad r_{33} = a_{31}^2 + a_{32}^2 + a_{33}^2,$$

so that

$$(15) \quad a_{33}^2 = r_{33} - a_{31}^2 - a_{32}^2.$$

The correlation

$$(16) \quad r_{k3} = a_{31}a_{k1} + a_{32}a_{k2} + a_{33}a_{k3},$$

so that

$$(17) \quad a_{k3} = \frac{r_{k3} - a_{31}a_{k1} - a_{32}a_{k2}}{a_{33}},$$

and hence the third column of  $F$  can be determined.

If  $R$  is of rank  $r$ , there will be  $r$  columns of  $F$ . If this procedure is continued to column  $(r+1)$ , it will be found that the entries in such a column all vanish. It will be seen by equations of the type (8), (12), (16), that each of the coefficients in  $R$  determines a parameter in  $F$  if  $r=n$ .

This method illustrates the following theorems.

**Theorem 9.** *Any symmetric matrix  $A$  of order  $n \times n$  and of rank  $r$  can be factored into the matrix  $B$  and its transpose  $B'$  where  $B$  is a matrix of order  $n \times r$  and of rank  $r$ .*

**Theorem 10.** *Any symmetric matrix  $A$  of order  $n \times n$  and of rank  $r$  in which all but  $r$  of its diagonal entries are unknown can be factored into an  $n \times r$  matrix  $B$  and its transpose  $B'$  where  $B$  is a matrix of order  $n \times r$  and of rank  $r$ .*

A simple numerical example is given in *Table 5* which shows the intercorrelations of four fictitious tests with unity in the diagonal cells. The rank is 4. The corresponding factorial matrix is shown in *Table 6*. In *Table 7* the same intercorrelations are reproduced with communalities in the diagonal cells by which the rank of the matrix is reduced to 2. Corresponding factorial matrices are shown in *Table 8*.

*Table 5*  
*Fictitious Correlational Matrix*

	I	II	III	IV
1	1.00	+ .56	+ .24	- .61
2	+ .56	1.00	- .12	- .63
3	+ .24	- .12	1.00	- .18
4	- .61	- .63	- .18	1.00

*Table 6*  
*Factorial Matrix Which Reproduces the Arbitrary Symmetric Matrix of Table 5*

	I	II	III	IV
1	+1.000 000	0	0	0
2	+ .560 000	+ .828 493	0	0
3	+ .240 000	- .307 064	+ .920 930	0
4	- .610 000	- .348 102	- .152 552	+ .695 308

### The rank of a matrix

Since the number of linearly independent factors has been shown to be the rank of the correlational matrix, it is of some interest to investigate the possible means for determining the rank of a matrix. The rank is defined as

the highest order of the non-vanishing minors, but to expand all of the minors even of a specified order is a prohibitive task when  $n$  is large. For the

*Table 7*

*The Same Correlational Matrix as in Table 5 except for Communalities in Diagonal Cells Which Reduce the Rank to 2*

	I	II	III	IV
1	+.58	+.56	+.24	-.61
2	+.56	+.74	-.12	-.63
3	+.24	-.12	+.72	-.18
4	-.61	-.63	-.18	+.65

scientific problem it is not of much value to have methods of determining the rank, because the rank of a correlational matrix  $R_0$  with experimentally obtained coefficients is known to be equal to its order. This is evident be-

*Table 8*

*Factorial Matrices Which Reproduce the Symmetric Matrix Table 7*

	I	II		I	II
1	+.70	+.30	1	+.761577	.000000
2	+.50	+.70	2	+.735316	+.446442
3	+.60	-.60	3	+.315135	-.787839
4	-.70	-.40	4	-.800969	-.091915

cause sampling errors and chance errors in the scores are fortuitous components in the coefficients.

The theorem to be described here is useful for estimating the rank of a matrix when the cell entries can be assumed to be free from experimental errors. It may be useful in estimating the rank of  $R_0$  containing experi-

mental coefficients when  $n$  is large, but it is not likely to be useful when  $n$  is as small as 10 or 15. The theorem is relevant to the factor problem, and some useful adaptation of it may be made to fallible data.

Theorem 11. *If any matrix of rank  $r$  is sectioned into a composite square matrix of order  $s$  where  $s > r$ , then the determinant of the composite matrix vanishes.*

Table 9

14	12	6	8	2
6	104	21	9	17
7	6	3	4	1
35	30	15	20	5

The matrix will be said to be sectioned when the columns have been divided into  $s$  groups, and when the rows have also been divided into the same number of groups. Let  $r=2$  as an example. Since  $s > r$ , we may let  $s=3$ . Then the  $n$  columns of  $R$  will be divided into three groups of  $p$ ,  $(q-p)$ , and  $(n-q)$  columns, respectively; while the  $n$  rows of  $R$  will be divided into three groups of  $t$ ,  $(u-t)$ , and  $(n-u)$  rows, respectively. The matrix  $R$  will then be sectioned.

Table 10

26	14	2
110	30	17
78	42	6

The composite matrix will be defined as the square matrix of order  $s$  in which the entries are the sums of the elements in the corresponding parts of the sectioned matrix. The example of *Table 9* illustrates the formation of a composite matrix. This  $4 \times 5$  matrix is of rank 2. It has been sectioned into a  $3 \times 3$  square matrix by arbitrarily dividing the columns into three groups of 2, 2, and 1 columns, respectively, and by arbitrarily dividing the rows into three groups of 1, 1, and 2 rows, respectively. The composite matrix is shown in *Table 10*. Its determinant vanishes.

The proof of the theorem will be written for rank 2, but it can readily be generalized for any rank. If  $R$  is of rank 2, it is possible to find two rows that

are linearly independent. Let these be the first and second rows. Then the elements of the  $j$ th row can be expressed as a linear function of the first two rows so that

$$(18) \quad \left\{ \begin{array}{l} r_{j1} = m_1 r_{11} + m_2 r_{21}, \\ r_{j2} = m_1 r_{12} + m_2 r_{22}, \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ r_{jk} = m_1 r_{1k} + m_2 r_{2k}. \end{array} \right.$$

It is evident that the sum of the first  $p$  entries of row  $j$  can also be expressed as the same linear function of the corresponding sums in the first two rows. We have then

$$(19) \quad \sum_{k=1}^p r_{jk} = m_1 \sum_{k=1}^p r_{1k} + m_2 \sum_{k=1}^p r_{2k}.$$

Similar summations may be written for the other two groups of columns so that

$$(20) \quad \sum_{k=(p+1)}^q r_{jk} = m_1 \sum_{k=(p+1)}^q r_{1k} + m_2 \sum_{k=(p+1)}^q r_{2k},$$

$$(21) \quad \sum_{k=(q+1)}^n r_{jk} = m_1 \sum_{k=(q+1)}^n r_{1k} + m_2 \sum_{k=(q+1)}^n r_{2k}.$$

These summations may be represented in an  $n \times 3$  matrix as shown in *Table 11*. Since each of the rows can be expressed as a linear function of the first two rows, it follows that the rank of this  $n \times 3$  matrix is also 2. The columns may be so arranged that the third column of this matrix may be expressed in terms of the first two columns. This reduction by columns is similar to the reduction by rows that has been described. This reduction by columns gives a  $3 \times 3$  composite matrix whose rank is 2, and hence its determinant vanishes. If the rank of  $R$  is equal to or greater than the order  $s$  of the square composite matrix, then the determinant of the composite does not necessarily vanish.

This theorem and other considerations about the rank of a matrix are of analytical interest because of the fact that the rank has been shown to be equal to the number of linearly independent common factors which are necessary to account for the intercorrelations. It does not seem to be feasible to apply this theorem directly to the determination of the communalities be-

cause of the sampling errors in the coefficients. It is possible that the theorem can be applied with profit to a large matrix whose rank is only a fraction of its order.

### Methods of estimating communalities

Before the correlational matrix  $R$  can be factored into the matrices  $F$  and  $F'$  which constitute the solution, it is necessary to compute or to estimate the communalities. If the cell entries of the correlational matrix are infallible, the computation of the communalities is a relatively simple matter;

Table 11

$\sum_1^p r_{1k} = b_{11}$	$\sum_{p+1}^q r_{1k} = b_{12}$	$\sum_{q+1}^n r_{1k} = b_{13}$
$\sum_1^p r_{2k} = b_{21}$	$\sum_{p+1}^q r_{2k} = b_{22}$	$\sum_{q+1}^n r_{2k} = b_{23}$
$\sum_1^p r_{3k} = b_{31}$	$\sum_{p+1}^q r_{3k} = b_{32}$	$\sum_{q+1}^n r_{3k} = b_{33}$
$\sum_1^p r_{nk} = b_{n1}$	$\sum_{p+1}^q r_{nk} = b_{n2}$	$\sum_{q+1}^n r_{nk} = b_{n3}$

but if the coefficients are experimentally obtained values, the communalities can be at best only estimated. Fortunately, the estimates of the communalities need not be at all close when the number of tests or variables is large. When the number of tests is as small as ten or twelve, it becomes essential to ascertain the communalities with some degree of exactness.

In this section several methods of computing or estimating the communalities will be described. Most of these methods are not suitable for purposes of computation, partly because of the limitation that experimental data are affected by sampling errors and partly because some of the methods are prohibitive in arithmetical labor. Those who study the factor problem analytically will find these methods of some interest.

One of the simplest of these methods is used as a first estimate for the

centroid method which is described in the next chapter. By successive approximations the communalities may be determined to any required degree of exactness.

### 1. Expansion of a minor of order $(r+1)$

If the correlation coefficients are infallible, a simple procedure for computing the communalities is as follows: In order to compute the communality of a test  $j$ , select any minor in the correlational matrix which contains the diagonal entry for test  $j$  but no other diagonal entries, and which is of order greater than the rank. By definition of the rank of a matrix this minor must vanish. Its expansion is a linear equation in one unknown by which the

Table 12

	1	2	3	4	5	6	7	8
1	...	.56	.16	.24	.72	.64	.40	.24
2	.56	...	.38	.49	.67	.72	.63	.53
3	.16	.38	...	.48	.24	.40	.52	.54
4	.24	.49	.48	...	.34	.52	.64	.65
5	.72	.67	.24	.34	...	.76	.52	.35
6	.64	.72	.40	.52	.76	...	.68	.56
7	.40	.63	.52	.64	.52	.68	...	.71
8	.24	.53	.54	.65	.35	.56	.71	...

communality may be computed. It is evident that this simple method is not applicable to fallible data, and consequently the method is not practically useful. It is possible that this method may be generalized into a useful summation formula.

In *Table 12* are reproduced the intercorrelations of eight hypothetical variables. The rank of the matrix is 2. *Table 13* shows a minor of order 3 with one unknown entry, namely, the communality for variable No. 1. In order that the expansion of the determinant of *Table 13* shall vanish, the unknown diagonal entry must be .64. If the rank is unknown and if it is assumed too high (say 3), it will be found that the coefficients of  $h_1^2$ , as well as the numerical terms, all vanish. This indeterminacy can be removed by assuming a lower rank. An exception is the case in which the minor of *Table 13* is of rank 2 when some other minor in *Table 12* of order 3 or higher does

not vanish. Such a situation would be discovered routinely by the centroid method, so that it is not necessary to evaluate all possible minors of order 3 in *Table 12*.

## 2. Grouping of similar tests

If the test battery is large enough so that each test belongs in a constellation of similar tests, then the tests in each constellation will be represented by vectors in the common-factor space with relatively small angular separations. The communality of a test is the square of the length of its vector. If the angular separations between several test vectors are relatively small, then the projection of a test vector on the centroid vector of the constellation will be nearly the same as the length of the vector. The square of the projection may be used as an estimate of the communality of the test with the knowledge that the estimate will be slightly too low. The projection of

*Table 13*

	1	2	3
1	$h_1^2$	.56	.16
4	.24	.49	.48
5	.72	.67	.24

each test vector on the best fitting single vector for the constellation is essentially the same as the loading of the test with the single common factor which best describes the intercorrelations of the tests in the constellation. Relatively simple methods for dealing with the special case of rank one are described in *chapter v*.

## 3. Grouping of three tests

A special case of the preceding method is that of using only three tests in a constellation. Since the intercorrelations of three tests can always be accounted for exactly by a single common factor, this method does not contain any check of internal consistency. To obtain such a check for a single common factor requires at least four tests. This is Spearman's problem, which is discussed in *chapter v*.

One procedure for estimating the communality of a test  $j$  is to select the two other tests which have the highest correlations with test  $j$ . Let these two additional tests be  $k$  and  $l$ . If the test battery is so constructed that each postulated ability is represented by several tests, it can be expected that the



three tests  $j$ ,  $k$ , and  $l$  will be represented by test vectors with relatively small angular separations. If this condition is satisfied, the three vectors can be represented approximately by their projections on a common centroid vector, so that the intercorrelations are nearly accounted for by a single factor common to the three tests. We have then

$$(22) \quad r_{jk} = a_{j1}a_{k1},$$

$$(23) \quad r_{jl} = a_{j1}a_{l1},$$

$$(24) \quad r_{kl} = a_{k1}a_{l1},$$

so that

$$(25) \quad \frac{r_{kl}}{r_{jl}} = \frac{a_{k1}}{a_{j1}},$$

or

$$(26) \quad a_{k1} = \frac{a_{j1}r_{kl}}{r_{jl}}.$$

But

$$(27) \quad r_{jk} = a_{j1}a_{k1},$$

and hence

$$(28) \quad r_{jk} = \frac{a_{j1}^2 r_{kl}}{r_{jl}}.$$

so that

$$(29) \quad a_{j1}^2 = \frac{r_{jk}r_{jl}}{r_{kl}};$$

where tests  $k$  and  $l$  are selected because of high correlations with  $j$ .

This formula is familiar. In fact, it is Spearman's\* formula for the correlation of a test with the common factor  $g$ , but it is here used under quite different circumstances and with different assumptions. Spearman uses this formula to ascertain the correlation of a test with the central intellectual factor under the assumption that only one principal factor is operative. Here two tests are selected because they correlate highest with test  $j$  under the assumption that the intercorrelations of these three tests may be described in terms of a single common factor, but it is also assumed that there are different common factors for different sets of three tests that may be selected in the battery. It is not assumed that the common factor is the

\* *The Abilities of Man* (Macmillan Co., 1927), Appendix, eq. (19).

same for all combinations of three tests. The formula is used here merely to estimate the communality of each test.

The diagonal entry for test  $j$  is then

$$(30) \quad h_j^2 \doteq \frac{r_{jk}r_{jl}}{r_{kl}},$$

where tests  $k$  and  $l$  are the two tests that correlate highest with  $j$ . This procedure is continued in estimating the diagonal entry for each of the  $n$  columns. In general, these values should be slightly too low.

If as many as four tests of each kind have been included in the battery, then an estimate of the communality of each of them may be taken as the average of four sets of three tests.

One useful circumstance is that the estimate of the communality is of significance only when the number of variables is relatively small—say eight or ten. When the number of variables is as large as thirty or forty, any value between 0 and +1 may be recorded in the diagonal cell of each column without affecting noticeably the resulting factor loadings as determined by the centroid method. The reason for this is that the diagonal entry has a very slight effect on the relative order of magnitude of the sum of a column.

In selecting the tests  $k$  and  $l$  which are to be used for estimating the communality of  $j$ , it is probably best first to correct for attenuation. Then the two highest correlations in each column indicate which two tests to select for each column. The communalities are determined by equation (30), in which raw coefficients are used. The correction for attenuation may be used only to ascertain which tests are to be selected in each column, although this refinement is probably not essential.

#### 4. Highest coefficient in each column

Inspection of equation (30) for estimating the communality of a test suggests a further simplification in the estimate. The numerator contains the product of the two highest correlations in the column for test  $j$ . The denominator is the intercorrelation of the two tests so selected, namely,  $k$  and  $l$ . If these coefficients are of the same order of magnitude, then the estimated communality of test  $j$  will be nearly equal to the highest intercorrelation in column  $j$ . This is the method that has been found in practice to give consistently better results than any of the many other much more elaborate methods that have been tried so far. This method is used as a first approximation in the centroid method of extracting the test coefficients.

#### 5. Linear dependence of rows or columns

If the rank of a correlational matrix is  $r$ , then any row may be expressed linearly in terms of any  $r$  independent rows. It may be possible to general-

ize this principle into a method of computing the communalities for fallible data.

Since the rank of the correlational matrix of *Table 12* is 2, any row can be expressed as a linear function of any two independent rows. *Table 13* shows a third-order minor of *Table 12* in which the first row can be expressed as a linear function of the second and third rows. The two multipliers may be determined from columns 2 and 3 by equations of the type (18). When these are known, the communality may be computed.

Another example will be shown with reference to *Table 12*. Assume that the second and third rows are independent. Consider a  $3 \times 6$  matrix consisting of the first three rows of *Table 12* and all of its columns except 2 and 3. This matrix may be assumed to be also of rank 2; and it contains only

*Table 14*

	1	2	3	4
1	$h_1^2$	.56	.16	.24
5	.72	.67	.24	.34
6	.64	.72	.40	.52
7	.40	.63	.52	.64
8	.24	.53	.54	.65

one unknown entry, namely, the communality of the first variable. If the first row is expressed linearly in terms of the second and third rows, the two multiplying coefficients may be determined from any pair of independent columns. When these are known, the unknown communality may be computed.

## 6. Sectioning of the matrix

A correlational matrix is square, and it may be divided into four quadrants in such a way that all of the unknown diagonal entries lie in the upper left and the lower right quadrants. All of the entries in the upper right and lower left quadrants are known. These two quadrants are symmetric about the diagonal. A part of the matrix of *Table 12* may be sectioned, as shown in *Table 14*. A composite matrix of rank 2 may be formed as shown in *Table 15*, in which the first row can be expressed as a linear function of the second and third rows. The two multiplying constants may be determined from the second and third columns of *Table 15*. When these multipliers are known, the communality of the first variable may be computed. The same

procedure can be repeated with the second and with each succeeding row of *Table 12*. In this manner all of the communalities in the upper left quadrant of *Table 12* may be determined. The same method can be used to determine the communalities in the lower right quadrant of *Table 12*. The reason these procedures have been investigated is the belief that if a communality is expressed as a function of a large number of fallible coefficients the determination is more stable than when the determination is made with a small number of fallible coefficients.

### 7. Expansion of principal minors of order $(n-1)$

It is possible to write  $n$  principal minors of order  $(n-1)$  in a square matrix of order  $n$ . If the expansion of each of these  $n$  principal minors of order  $(n-1)$  is set equal to zero, the rank of the matrix is assumed to be not

*Table 15*

	1	2	3, 4
1	$h_1^2$	.56	.40
5, 6	1.36	1.39	1.50
7, 8	.64	1.16	2.35

greater than  $(n-2)$ . This follows from the property of a Gramian matrix that if all of its principal minors of order  $m$  vanish, then the rank of the matrix does not exceed  $(m-1)$ . Since there are  $n$  principal minors of order  $(n-1)$ , their expansions give as many equations as there are unknown diagonal entries. A unique solution is obtained if the inequality (5) is satisfied. If this inequality is not satisfied, there should be no unique solution. In this method it is not necessary to know the rank. These considerations are of some analytical interest, but they do not seem to lend themselves to computing purposes.

### 8. Expansion of principal minors of order $(r+a)$

This should be a special case of the preceding method but less laborious. It is not necessary that the rank be known, but it is assumed that  $(r+a)$  is taken larger than the rank. The simplest case is that in which  $a=1$ . This method requires that the number of tests covered by the expanded principal minors is such as to satisfy inequality (5) even though all the tests in the correlational matrix are not utilized. The development of this type of analysis would be of interest, but it does not seem likely to yield practical computing methods.

## CHAPTER III

### THE CENTROID METHOD

#### Principles of the method

The centroid method is a general method of factoring a symmetric matrix with real elements.\* Its application to the factor problem involves finding  $F$  when  $R$  is known, so as to satisfy the fundamental factor theorem,  $FF' = R$ . The chief requirements of a method of factoring the correlational matrix are that it must be applicable even though the diagonal elements are unknown and that it must be applicable even though the intercorrelations are subject to sampling errors. These two requirements preclude the use of the diagonal method of *chapter ii*, which is very simple in application when the entries are infallible and the diagonals known.

The purpose of the centroid method in factor analysis is merely to factor the correlational matrix. Any other method would serve the purpose equally well provided that the minimum rank of  $R$  with unknown diagonals is not altered. When the correlational matrix has been factored into  $F$  and  $F'$ , the entries of  $F$  cannot be given scientific interpretation until  $F$  has been rotated so that the new reference axes represent primary factors.

Each correlation coefficient in  $R$  may be expressed in the form (37-i)

$$(1) \quad r_{jk} = a_{j1} a_{k1} + a_{j2} a_{k2} + \cdots + a_{jr} a_{kr},$$

in which there are as many terms in the right member as there are factors in  $R$ . The numerical values of  $a_{jm}$  are determined by the arbitrary locations of the orthogonal reference vectors, since  $a_{jm}$  is the projection of the trait vector  $j$  on the reference vector  $m$ . The subscript  $j$  defines a row of  $R$ , and the subscript  $k$  defines a column of  $R$ .

The traits are represented by a set of  $n$  trait vectors in a space of  $r$  dimensions, and the scalar product of each pair of vectors is the correlation between them. It has been shown that this configuration represents the intercorrelations and that these are independent of the locations of the orthogonal reference vectors that are implied in (1). Hence the reference vectors may be rotated without any effect on the intercorrelations.

\* The first form of the centroid method was described in "Multiple Factor Analysis." It was improved by the elimination of arbitrary subgroups in "A Simplified Multiple Factor Method." The method has been further improved as described in this chapter.

Let the co-ordinate system be rotated so that the centroid of the system lies in the first axis of reference. The case in which the centroid is at the origin will be discussed in a later paragraph. Then

$$(2) \quad r_{jk} = a'_{j1} a'_{k1} + a'_{j2} a'_{k2} + \cdots + a'_{jr} a'_{kr}.$$

Summing (2) for all traits  $j$  in column  $k$  of  $R$ , we have

$$(3) \quad \sum_{j=1}^n r_{jk} = a'_{k1} \sum_{j=1}^n a'_{j1} + a'_{k2} \sum_{j=1}^n a'_{j2} + \cdots + a'_{kr} \sum_{j=1}^n a'_{jr};$$

and summing for all columns  $k$  so as to include all entries in  $R$ , we have

$$(4) \quad \sum_{k=1}^n \sum_{j=1}^n r_{jk} = \sum_{k=1}^n a'_{k1} \sum_{j=1}^n a'_{j1} + \sum_{k=1}^n a'_{k2} \sum_{j=1}^n a'_{j2} + \cdots + \sum_{k=1}^n a'_{kr} \sum_{j=1}^n a'_{jr}.$$

But

$$(5) \quad \sum_{k=1}^n a'_{km} = \sum_{j=1}^n a'_{jm},$$

and hence

$$(6) \quad \sum_{k=1}^n \sum_{j=1}^n r_{jk} = \left[ \sum_{j=1}^n a'_{j1} \right]^2 + \left[ \sum_{j=1}^n a'_{j2} \right]^2 + \cdots + \left[ \sum_{j=1}^n a'_{jr} \right]^2.$$

The  $r$  co-ordinates of the centroid of the system of  $n$  points are

$$\frac{1}{n} \sum_{j=1}^n a'_{j1}, \quad \frac{1}{n} \sum_{j=1}^n a'_{j2}, \quad \dots, \quad \frac{1}{n} \sum_{j=1}^n a'_{jr}.$$

The co-ordinate axes have been so rotated that the centroid lies in the first axis of reference. The centroid therefore has zero projections on all the remaining  $(r-1)$  co-ordinate axes. Hence

$$(7) \quad \sum_{j=1}^n a'_{j2} = \sum_{j=1}^n a'_{j3} = \cdots = \sum_{j=1}^n a'_{jr} = 0,$$

so that the  $r$  co-ordinates of the centroid are

$$\frac{1}{n} \sum_{j=1}^n a'_{j1}, \quad 0, \quad 0, \dots, 0.$$

Substituting (7) in (6), we have

$$(8) \quad \sum_{k=1}^n \sum_{j=1}^n r_{jk} \equiv \left[ \sum_{j=1}^n a'_{j1} \right]^2 \equiv r_t,$$

in which  $r_t$  is defined as the sum of all the coefficients in  $R$  including the diagonal terms. The first co-ordinate of the centroid is also its distance from the origin, since the remaining  $(r-1)$  co-ordinates vanish. Hence the distance of the centroid from the origin is

$$(9) \quad d = \frac{1}{n} \sum_{j=1}^n a'_{j1},$$

or, by (8),

$$(10) \quad d = \frac{1}{n} \sqrt{\sum_{k=1}^n \sum_{j=1}^n r_{jk}} = \frac{1}{n} \sqrt{r_t}.$$

Substituting (7) in (3), we have

$$(11) \quad \sum_{j=1}^n r_{jk} = a'_{k1} \sum_{j=1}^n a'_{j1},$$

and from (8) it follows that

$$(12) \quad r_k = a'_{k1} \sqrt{r_t},$$

where  $r_k$  is the sum of all the coefficients in column  $k$  of  $R$ . If the sum of the coefficients in column  $k$  and the sum of all the coefficients in  $R$  are known, the projection of the vector  $k$  on the first axis of reference through the centroid is also known, namely,

$$(13) \quad a'_{k1} = \frac{r_k}{\sqrt{r_t}}.$$

By (13) the first co-ordinate of each trait may be found.

The numerical value of the first term in the right member of (2) is known by (13), and hence (2) may be transposed so as to show the first-factor residuals. Let the first-factor residuals from which the second co-ordinates are to be found be designated  $r_{2-jk}$  for  $j$  and  $k$ . We have then from (2)

$$(14) \quad r_{2-jk} = r_{jk} - a'_{j1} a'_{k1} = a'_{j2} a'_{k2} + a'_{j3} a'_{k3} + \dots + a'_{jr} a'_{kr}.$$

Summing for column  $k$ ,

$$(15) \quad \sum_{j=1}^n r_{2 \cdot jk} = \sum_{j=1}^n r_{jk} - a'_{k1} \sum_{j=1}^n a'_{j1} = a'_{k2} \sum_{j=1}^n a_{j2} + a'_{k3} \sum_{j=1}^n a'_{j3} + \cdots + a'_{kr} \sum_{j=1}^n a'_{jr}.$$

From (7) it follows that

$$(16) \quad \sum_{j=1}^n r_{2 \cdot jk} = 0.$$

The sum of the residuals is zero in each column.

The number of terms in the right member of (2) is the rank of  $R$ . The number of terms in the right member of (15) is  $(r-1)$ , and hence this is the rank of the table of first-factor residuals. The entries in the residual table may be regarded as the scalar products of all pairs of residual vectors in a space of  $(r-1)$  dimensions. From (7) it is seen that the  $(r-1)$  co-ordinates of the centroid of the residual vectors are zero, and hence the centroid is at the origin in the  $(r-1)$  subspace. This precludes the direct application of formulae of the type (13) in determining the second and subsequent co-ordinates of the  $n$  points.

In order to make the centroid method applicable in this situation, where the centroid of the system is at the origin, it is necessary to remove the centroid from the origin. In order to accomplish this purpose without destroying the identities of the traits, a new concept will be introduced. Every point represents a trait. The diametrically opposite point represents the diametrically opposite trait, which will be called the *reflection* or *image* of the given trait. If a trait  $+A$  is represented by the co-ordinates  $a_{11}, a_{12}, \dots, a_{1r}$ , then the co-ordinates of the reflected trait  $-A$  are  $-a_{11}, -a_{12}, \dots, -a_{1r}$ . Either the point or its reflection through the origin may be used to represent the trait as long as the proper sign is attached to it. In this sense the score on  $A$  may be replaced by the same score with negative sign to represent  $-A$ . Both scores represent the same trait except for sign. If  $A$  represents the trait "tactfulness," then  $-A$  represents "plus tactlessness," or "minus tactfulness." The identity of the trait is easily established with a simple reversal of sign. If some of the traits are reversed in sign, the centroid of the system will be removed from the origin without disturbing the identities of the traits. To reverse the signs in a row of  $F$  is to reflect the point through the origin, and it has been shown that this reversal of sign causes a reversal of sign in the corresponding row and in the corresponding column of  $R$ . The reflection of a trait is accomplished merely by changing the signs of the correlation coefficients in its row and in its column of  $R$  or



in the residual table. The correlation coefficients in  $R_0$  have the same relation to the common-factor space as the residual coefficients have to the residual subspace.

The next question is to decide which traits to reverse in sign. If all of them are reversed, it is clear that the correlational matrix, or the residual table, remains unaffected and the centroid remains at the origin. It is necessary, therefore, to reverse the signs of only some of the traits in the battery. It is desirable to account for as much as possible of the residual variance by each successive factor, and this should be a guiding consideration in deciding upon the traits which are to be reversed. If there is a clustering of traits in the  $(r-1)$  subspace which is balanced by a scattering of traits on the opposite side of the centroid, it is desirable to pass the second reference axis through the cluster. The second axis may be passed anywhere in the residual subspace because it is orthogonal to the first axis of reference, which has already been located through the centroid of the original system. Since the subspaces are frequently of order higher than the third, it is not feasible to use any direct graphical methods for finding the clusters. Rather simple considerations should make it possible to accomplish the same purpose analytically.

If a trait is in a cluster, its correlations will be high and positive with the remaining traits in the cluster; while if it is unique, in the sense that it is relatively remote from the other traits, its correlations with the other traits will be near zero. If a remote trait is reflected, its correlations will be reversed in sign, so that the majority of them are positive. The principle to be applied here is that every trait the majority of whose correlations are negative is to be reversed in sign. This will tend to bring all of them into a hemisphere, and the centroid will then be removed from the origin without destroying the identities of the traits.

When the sign reversals have been made so that the majority of the correlations for each trait are positive or zero, the centroid method may again be applied as before, by rotating the co-ordinate axes about the first centroid axis so that the centroid of the residual configuration lies in the second axis of reference. This axis will be orthogonal to the first axis of reference because the  $(r-1)$  subspace is orthogonal to the first axis of reference.

It will be found that the majority of the points in the subspace will have projections on the second centroid axis which are positive or zero. If it is desired, it is always possible to make a few additional sign changes so as to insure that the sum of every column in the residual table is positive or zero. This guarantees that the projection of every point in the subspace will have a positive projection on the new centroid axis. This additional adjustment is probably not ordinarily worth the additional computation, be-

cause it will not noticeably affect the location of the new centroid axis. The principle recommended for practical computations is to reverse the signs of one trait at a time until the number of negative coefficients in the residual table is less than  $n/2$ . It will be shown that this computation can be easily routinized.

After reflection, let (14) be written in the form

$$(17) \quad r'_{2jk} = a'_{j2}a'_{k2} + a'_{j3}a'_{k3} + \cdots + a'_{jr}a'_{kr}.$$

The value of  $a'_{j2}$  is equal to either  $+a'_{j2}$  or  $-a'_{j2}$ , depending on whether  $j$  has been reversed in sign. The correlation  $r'_{2 \cdot jk}$  is equal to the residual correlation  $r_{2 \cdot jk}$  if neither  $j$  nor  $k$  has been reflected or if both of them have been reflected. If only one of the traits  $j$  and  $k$  has been reflected, then  $r'_{2 \cdot jk} = -r_{2 \cdot jk}$ .

After the reflections, let the residual vectors in the residual subspace be rotated so that the new centroid lies in the second orthogonal reference axis. Then

$$(18) \quad r'_{2jk} = a'_{j2}''a'_{k2}'' + a'_{j3}''a'_{k3}'' + \cdots + a'_{jr}''a'_{kr}''.$$

Summing (18) in a manner similar to that shown in (3) and proceeding as in (3) to (12), inclusive, we have

$$(19) \quad r'_{2k} = a'_{k2}''' \sqrt{r'_{2i}},$$

where  $r'_{2k}$  is defined as the sum of the first-factor residual coefficients in column  $k$  after reflection, and  $r'_{2i}$  is defined as the sum of all the first-factor residual coefficients after reflection. From (19) the values of  $a'_{k2}'''$  may be found. If  $k$  has not been changed in sign, this is the second co-ordinate of  $k$ . If  $k$  has been changed in sign, then the second co-ordinate of  $k$  is  $-a'_{k2}'''$ .

The procedure for the remaining factors is the same. When the sign reversals have been made, the centroid method is used again by rotating the co-ordinate system about the axes that have been established, so that the centroid of the residual configuration lies in the next orthogonal reference axis. Each successive residual table is reduced in rank by 1. When  $r$  factors have been extracted, the  $r$ th-factor residuals all vanish if the rank of  $R$  is  $r$ .

A useful check on the arithmetical work is as follows: Summing (13) for all tests  $k$ ,

$$(20) \quad \sum_{k=1}^n a'_{k1} = \frac{1}{\sqrt{r_1}} \sum_{k=1}^n r_k.$$

But by (8), (11), and (12),

$$(21) \quad \sum_{k=1}^n r_k = r_t,$$

and hence

$$(22) \quad \sum_{k=1}^n a'_{k1} = \sqrt{r_t}.$$

The sum of the factor loadings is equal to the square root of the sum of all coefficients in the correlation table. This check is applicable for each of the successive factors.

### Example 1. Unity in the diagonal cells

A small correlation table will be used in four examples of the centroid method. The examples differ only in the diagonal entries. In *Table 1* are shown the intercorrelations of three fictitious variables with self-correla-

*Table 1*

	++1	+ -2	++3	
++1	++1.000000	-+ .480000	++ .560000	
+ -2	-+ .480000	++1.000000	-+ .420000	
++3	++ .560000	-+ .420000	++1.000000	
<i>A</i>	+ .080000	- .900000	+ .140000	
<i>B</i>	+1.040000	+ .900000	+ .980000	
<i>C</i>	+1.000000	+1.000000	+1.000000	
<i>D</i>	+2.040000	+1.900000	+1.980000	5.920000
<i>E</i>	+ .838435	+ .780895	+ .813775	2.433105
<i>K</i>	+ .838435	- .780895	+ .813775	.4109975

tions of unity. Before each entry there are two signs. The first one is the given sign. The given variables may be designated by number and sign, as +1, +2, +3. In order to displace the centroid from the origin, the signs may be reversed so that the sum of the coefficients in each column (omitting the diagonal entry) shall be positive. In row *A* these column sums are recorded. The second column has the largest negative sum. In this example there is only one column with negative sum. Hence, variable 2 is reversed in sign. These sign reversals in both column 2 and row 2 are shown by the second sign before each coefficient.

The new sums (omitting diagonals) are shown in row *B*. All of the sums are now positive. If negative sums remained, further sign reversals would be made, as shown in subsequent examples. In row *C* is shown the diagonal

entry for each column and in row *D* is shown the sum of all coefficients in each column.

The last entry in row *D* is the sum of all coefficients in the table. It is  $r_t$  in (8). The square root of this sum is also recorded; and immediately below this is recorded its reciprocal, as required in (13). In row *E* is shown the first-factor loading for each of the three variables with signs used in *Table 1*. In row *K* are shown the first-factor loadings with signs which correspond to the original positive signs of the variables.

In order to extract the second factor, *Table 2* is prepared with the residuals (14). In computing these residuals, the given variables are taken with the following signs, +1, -2, +3. These are the signs used in computing row *E* in *Table 1*. The factor loadings in row *E* then correspond to  $a'_{21}$  in (14).

Table 2

	++1	-+2	++3	
++1	++ .297027	-+ .174730	-- .122297	
-+2	-+ .174730	++ .390203	-+ .215473	
++3	-- .122297	-+ .215473	++ .337770	
$\Sigma_0$	.000000	.000000	.000000	
<i>A</i>	-- .297027	- .390203	- .337770	
<i>B</i>	+ .052433	+ .390203	+ .093176	
<i>C</i>	+ .297027	+ .390203	+ .337770	
<i>D</i>	+ .349460	+ .780406	+ .430946	1.560812
<i>E</i>	+ .279719	+ .624662	+ .344943	1.249325
<i>K</i>	+ .279719	+ .624662	+ .344943	.800432

In row  $\Sigma_0$  of *Table 2* are shown the sums of the columns. These all vanish, as proved by (16). Row *A* shows the sum of the coefficients in each column (omitting the diagonal entry). The second column has the largest negative sum. Hence, variable 2 is reversed in sign.

After reversing the signs of the residuals for variable 2, the new sums are recorded in row *B* (omitting diagonals). The diagonal entry for each column is recorded in row *C*. The sum of all coefficients in each column is shown in row *D*. The same procedure as before gives the second-factor loadings shown in row *E*. Row *K* is the same as row *E* because all three of the variables happen to be positive.

*Table 3* was prepared in order to extract the third-factor loadings. In it are recorded first the residuals from *Table 2*. The first sign is positive for each of the three variables because that is the sign arrangement which resulted after the sign reversal for the second factor. In row  $\Sigma_0$  is recorded the sum for each column which is zero. This is a check on the arithmetical work. Row *A* shows the sum of each column, omitting the diagonal entry. Since

both the first and the third columns have the same negative sum, it is immaterial which of them is reversed in sign for the extraction of the third-factor loadings. The first variable was here reversed in sign. In row *B* are recorded the new sums, omitting the diagonal entries. In row *C* are recorded the diagonal entries. In practice the repetition of row *C* is not needed, since the diagonal entries are available in the correlation table. Row *D* is the sum of rows *B* and *C*. The sum of all entries in row *D* is recorded at the right. It is the sum of all the coefficients in the residual table. Next below it is its square root; and next below that is recorded the reciprocal, as before. The multiplier is then applied to row *D*; and the result is row *E*, which contains the third-factor loadings with the signs of the variables after reversals for the third factor. In row *K* are recorded the third-factor loadings which correspond to the three variables taken with positive sign.

Table 3

	+ - 1	+ + 2	+ + 3	
+ - 1	+ + .218784	+ - .000000	- + .218784	
+ + 2	+ - .000000	+ + .000000	+ + .000000	
+ + 3	- + .218784	+ + .000000	+ + .218784	
$\Sigma_0$	.000000	.000000	.000000	
<i>A</i>	- .218784	.000000	- .218784	
<i>B</i>	+ .218784	.000000	+ .218784	
<i>C</i>	+ .218784	.000000	+ .218784	
<i>D</i>	+ .437568	.000000	+ .437568	.875136
<i>E</i>	+ .467744	.000000	+ .467744	.935487
<i>K</i>	- .467744	.000000	+ .467744	1.068962

If the same process is repeated in the attempt to extract a fourth factor from the residuals of *Table 3*, it will be found that all of the residuals vanish exactly. Hence the intercorrelations have been described in terms of as many factors as there are variables. The rank of the given *Table 1* is 3, and this is also the number of factors which will exactly account for the intercorrelations.

The three factor loadings for each of the three variables are summarized in the upper half of *Table 4*. These factors reproduce the given intercorrelations in *Table 1*. In the lower half of *Table 4* are recorded the factor loadings with the signs which correspond to those with which the factoring of *Table 1* was made. In that table the second variable was reversed in sign. When the factor loadings are taken with signs to correspond to the signs of the variables with which the factoring is made, the sum of each column after the first vanishes. This is shown at the bottom of *Table 4*.

**Example 2. Communalities in the diagonal cells**

In the first example unity was recorded in each diagonal cell. As a consequence, the rank of the correlational matrix became equal to its order, namely, 3. In the present example the communality is recorded in each

*Table 4*

	I	II	III
+1	+ .838435	+ .279719	-.467744
+2	-.780895	+.624662	.000000
+3	+.813775	+.344943	+.467744

	I	II	III
+1	+ .838435	+ .279719	-.467744
-2	+.780895	-.624662	.000000
+3	+.813775	+.344943	+.467744
Σ	+2.433105	.000000	.000000

diagonal, with the result that the rank of the matrix is reduced to 1. The procedure of extracting the factor loadings is here exactly the same as in the previous example.

*Table 5* shows the given intercorrelations, as well as the diagonal communalities which are assumed to be known in this example. The calcula-

*Table 5*

	++1	+ -2	++3	
++1	++ .640000	-+ .480000	++ .560000	
+ -2	-+ .480000	++ .360000	-+ .420000	
++3	++ .560000	-+ .420000	++ .490000	
A	+ .080000	- .900000	+ .140000	
B	+1.040000	+ .900000	+ .980000	
C	+ .640000	+ .360000	+ .490000	
D	+1.680000	+1.260000	+1.470000	4.410000
E	+ .800000	+ .600000	+ .700000	2.100000
K	+ .800000	- .600000	+ .700000	.4761905

tions are summarized in the several rows below the correlational matrix. Row A shows the sum of the coefficients in each column, omitting the diagonal entry. The largest negative sum is for the second column, and hence the second variable is reversed in sign. The resulting sums, omitting diagonals, are recorded in row B. The diagonal entries are repeated in row C.

Row *D* is the sum of rows *B* and *C*. Hence row *D* shows the sum of all the coefficients in each column. The total at the right of this row is the sum of all the coefficients in the table. Immediately below it is its square root, and below that is the reciprocal. This is the multiplier by which row *E* is obtained from row *D*. Row *E* shows the factor loading for each variable with the sign that was used for the factoring. Reversing the sign of the second variable, we have the factor loadings in row *K*, which represent the loadings when all of the variables are taken with positive sign.

If an attempt is made to obtain second-factor loadings for these three variables, it will be found that all of the residuals vanish. Hence one factor is sufficient to describe all of the intercorrelations in this table. The rank of the given correlation table is therefore 1, although its order is 3.

**Example 3. Each diagonal entry greater than the communality and less than unity**

In the first example it was shown that when unity is recorded in the diagonals of the correlational matrix, the intercorrelations can be described exactly in terms of as many factors as there are variables. In the second ex-

Table 6

	+1	+2	+3
+1	+.700000	-.480000	+.560000
+2	-.480000	+.500000	-.420000
+3	+.560000	-.420000	+.600000

ample it was shown that when the communalities are recorded in the diagonal cells, the rank of the matrix is reduced, so that a single factor is sufficient for the particular example here used. In the third example an arbitrary diagonal entry is recorded which is greater than the communality but less than unity. The resulting correlational matrix can be described in terms of as many factors as there are tests or variables.

Table 6 is such a matrix in which the diagonal entries exceed the communalities by arbitrary increments. The extraction of the factors is effected in exactly the same manner as has been described in the first two examples. The result is summarized in Table 7, which shows the three factor loadings for each of the three variables. The third-factor residuals vanish exactly.

**Example 4. Each diagonal entry less than the communality**

It is of some interest to know that the centroid method of factoring a symmetric matrix is applicable not only to those matrices whose factors are real but also to those symmetric matrices whose factors are imaginary.

When the diagonal entries are made less than the communalities, the Gramian properties of the correlational matrix are destroyed and the factors are then imaginary. The fourth example illustrates this case. *Table 8* contains

Table 7

	I	II	III
+1	+ .800900	+ .124013	-.207798
+2	-.644402	+.291112	.000000
+3	+.727254	+.167098	+.207798

Table 8

	+1	+2	+3
+1	+.600000	-.480000	+.560000
+2	-.480000	+.250000	-.420000
+3	+.560000	-.420000	+.400000

the same intercorrelations as those of *Table 1*, but the diagonal entries have been reduced below the communalities by arbitrary decrements. Application of the centroid method in exactly the same manner as for the previous examples gives the factor loadings shown in *Table 9*. The second and third columns of the factorial matrix of *Table 9* are imaginary. The co-ordinates of this table reproduce the intercorrelations exactly and the third-factor residuals all vanish.

Table 9

	I	II	III
+1	+.803111	+.106981 <i>i</i>	-.183145 <i>i</i>
+2	-.563157	+.259126 <i>i</i>	.000000
+3	+.675789	+.152146 <i>i</i>	+.183148 <i>i</i>

#### Example 5. A fictitious eight-variable problem with known communalities

The first four examples are intended to show the factoring of a symmetric matrix with four different conditions as regards the diagonal entries. The fifth example is intended to illustrate a variant procedure in selecting the variables which are to be reversed in sign. The sign-reversing method which is described here for the fifth example is recommended for most practical problems, since it is simpler in computation than the more complete method of the first four examples and since the simpler method gives results that are almost identical with those of the more elaborate sign-changing method.



*Table 10* shows the intercorrelations of eight variables whose self-correlations are known. All of the intercorrelations are here taken to be positive. At the bottom of each column is recorded the sum of the coefficients in the column, including the diagonal entry. The reason for taking the sum of all the entries in this example is that no sign reversals are necessary for the first factor when all of the given intercorrelations are positive. In the lower right corner of the table are shown the entries which are required to determine the multiplying factor. In the last row of the table are the first-factor loadings.

Table 10

	1	2	3	4	5	6	7	8	
1	.64	.56	.16	.24	.72	.64	.40	.24	
2	.56	.65	.38	.49	.67	.72	.63	.53	
3	.16	.38	.40	.48	.24	.40	.52	.54	
4	.24	.49	.48	.58	.34	.52	.64	.65	
5	.72	.67	.24	.34	.82	.76	.52	.35	
6	.64	.72	.40	.52	.76	.80	.68	.56	
7	.40	.63	.52	.64	.52	.68	.74	.71	
8	.24	.53	.54	.65	.35	.56	.71	.73	
<i>D</i>	3.60	4.63	3.12	3.94	4.42	5.08	4.84	4.31	33.94 = $\Sigma D$
<i>K</i>	.617940	.794740	.535548	.676301	.753693	.871983	830787	.739812	5.8258047 = $\sqrt{\Sigma D}$ .17165011 = $1/\sqrt{\Sigma D}$

The first-factor residuals are shown in *Table 11*. In front of each residual there are one, two, or three signs. Each sign is recorded in the first, the second, or the third position. The first sign is the sign of the residual as obtained from the given coefficients and the factor loadings of *Table 10*. In row  $\Sigma_0$  is shown the sum for all the coefficients in each column. These sums vanish, as shown in (16).

In order to select the variables which are to be reversed in sign so as to move the centroid of the system as far as possible from the origin, *Table 12* was prepared. It will be referred to as the *sign table*. This table illustrates a variant method of sign changing. At the top of the table are listed the variables from 1 to 8. In the first row is shown the number of negative entries in each column of *Table 11*. It so happens that in this example the number of negative entries is four for each column. Ordinarily these sums are not all the same.

The usual procedure is to select for sign reversal that variable whose column has the largest number of negative entries. Since this is four for each column, it is immaterial which of the variables is chosen for the first sign reversal. The first variable is so chosen. It is recorded at the right of the second row in *Table 12*. The fact that the first variable is to be reversed in sign is also indicated by the cross at the top of column 1.

There are eight entries in each column of *Table 11*; but since the diagonal



entry is always positive, there are only seven entries in each column that are subject to sign reversal. In the first column of *Table 11* there are four negative and three positive items, ignoring the diagonal entry. Hence, when the first variable is changed in sign, there will be three negative signs in the first column. This is the first entry in the second row of *Table 12*.

Each of the succeeding entries in the second row of *Table 12* is determined in the following manner. If the sign in the first row of *Table 11* is positive, then the entry in the first row of *Table 12* is augmented by 1. If the sign in the first row of *Table 11* is negative, then the entry in the first row of *Table 12* is reduced by 1. In this manner the remaining entries in the second row of *Table 12* are determined.

The procedure is summarized with the following notation:

$r_{ij}$   $\equiv$  given correlation or residual;

$N_j$   $\equiv$  number of negative signs in the  $j$ th column of the given table of correlations or residuals;

$n$   $\equiv$  number of variables;

$A_{ij}$   $\equiv$  the entry in the  $i$ th row and  $j$ th column in the sign table;

$B_{ij}$   $\equiv +1$  or  $-1$ . The sign of  $B_{ij}$  agrees with the sign of  $r_{ij}$ ;

$C_j$   $\equiv +1$  or  $-1$ . The sign is taken negative if variable  $j$  has been reversed in sign an odd number of times. Otherwise it is taken positive;

$k_i$   $\equiv$  variable which is reversed in sign in row  $i$ ;

$A_{ik} \equiv A_{ij}$  where  $j = k_i$ .  $A_{ik}$  is the largest value of  $A_{ij}$  in row  $i$ .

The successive steps in reflecting the variables are as follows:

- 1) The first row of the sign table contains  $N_j$  in column  $j$ ;
- 2) Select the highest value of  $N_j$ . Let it be column  $k$ . The variable  $k$  is to be reversed in sign;
- 3) Record  $k$  at the end of row 2;
- 4) Make a cross or check mark at the top of column  $k$ ;
- 5) Record  $A_{ij}$  in the next row where

$$(23) \quad A_{ij} = n - 1 - N_j \text{ when } j = k,$$

and

$$(24) \quad A_{ij} = A_{(i-1)j} + B_{ij}C_j \text{ when } j \neq k;$$

If a correlational entry is zero, count it as positive.

- 6) Find  $A_{ik}$ , the highest value in row  $i$ . Check the top of its column and record the number of the column at the end of row  $(i+1)$ ;
- 7) Record  $A_{ij}$ , as in step 5, for each row of the sign table until all entries  $A_{ij} < n/2$ .

The columns which are checked are to be reflected in the table of correlations or residuals.

A check on the arithmetical work of each row  $i$  is as follows:

$$(25) \quad \sum_{j=1}^n A_{ij} = \sum_{j=1}^n A_{(i-1)j} - 2[A_{(i-1)k} - A_{ik}] .$$

The sign table shows that variables 1, 2, 5, and 6 are to be reflected. The signs are reversed in the *second* position for these four *rows* in *Table 11*. Then the signs are reversed in the *third* position for the four *columns*. After making these sign reversals as shown in *Table 11*, each residual is to be taken with the sign that is next in front of it, irrespective of its position.

The rows and columns are designated by numbers. The sign reversals are also recorded in front of these numbers so as to show at a glance which of the variables have been reflected.

In row *D* is recorded the sum of each column after reflection. At the lower right corner of the table are shown the entries for the multiplier. In row *E* are shown the resulting second-factor loadings with signs to correspond to the reflected variables. In row *K* are shown the factor loadings for the original unreflected variables.

A check on the arithmetical work is that the sum of row *E* must equal  $\sqrt{\Sigma D}$ . This is the check described by (22).

A repetition of the same procedure for the second-factor residuals shows that they all vanish. Therefore the given coefficients in *Table 10* can be accounted for exactly by two factors.

*Table 13* gives a summary of the factor loadings for the eight variables. Two factor loadings are shown for each variable. The cross products in this table reproduce the correlations of the original unreflected variables.

The two methods of sign changing that have been described may be compared as follows: In the first and more complete method, that trait is reflected which has the largest absolute negative sum of coefficients in its column. After reflecting this trait, the sums are again determined, and the trait with the largest negative sum of coefficients is reflected. This procedure is continued until all of the sums of columns are positive, when the diagonal entries are ignored. In the second and shorter method, that trait is reversed in sign which has the largest *number* of negative coefficients in its

columns. After the reflection, the trait which has the largest number of negative coefficients is reflected. This process is continued until no trait remains for which the number of negative coefficients exceeds  $(n-1)/2$ . The two methods may be combined by using the shorter method first. When the number of negative signs has been minimized as described, there may still remain one or more small negative column sums, omitting the diagonal entries. The first method can then be used until all of the column sums are positive. This is the procedure which is illustrated in example 6, but in practice it is probably not worth the additional labor to make any refinements beyond the shorter procedure. The first method can be arranged with a check column in a manner similar to that of (25).

*Table 13*  
*Factor Loadings in Fictitious*  
*Eight-Variable Example*

	I	II
1	+ .617940	- .508084
2	+ .794740	- .135603
3	+ .535548	+ .336434
4	+ .676301	+ .350166
5	+ .758693	- .494353
6	+ .871983	- .199114
7	+ .830787	+ .223146
8	+ .739812	+ .427408
$\Sigma$	+5.825804	.000000

**Example 6. The centroid method with unknown diagonals** *This is what*

In the previous examples it has been assumed that the diagonal entries were known. The sixth example illustrates the application of the centroid method to an actual set of data. Since the communalities are unknown, the diagonal entries are also unknown. The diagonal entry will be estimated by method No. 4 in *chapter ii*. Fortunately, the diagonal entry may be given any value between zero and unity without affecting the results markedly, especially when the number of variables is as large as twenty or thirty or more. Hence even rough estimates of the diagonal entries are sufficient for reasonably accurate factor loadings by the centroid method.

*Table 14* contains the intercorrelations of fifteen psychological tests that were used by Professor Brigham in a recent experimental study. The tests will be identified by the same numbers that were used by Brigham.\* In each

\* Carl C. Brigham, *A Study of Error* (New York: College Entrance Examination Board, 1932), p. 275.

Table 14

## Computation of First-Factor Loadings

	10	2	5	3	4	1	8	7	9	6	15	14	17	11	18	Check
10	(.641)	.641	.564	.527	.424	.463	.242	.277	.265	.225	.273	.199	.303	.252	.217	5.513
2	.641	(.641)	.538	.496	.394	.468	.233	.240	.175	.203	.216	.122	.236	.224	.159	4.976
5	.564	.538	(.564)	.484	.428	.425	.203	.227	.252	.176	.189	.141	.245	.185	.200	4.821
3	.527	.496	.484	(.567)	.488	.567	.346	.388	.332	.356	.313	.219	.363	.393	.275	6.114
4	.424	.394	.428	.488	(.488)	.436	.297	.294	.342	.376	.287	.265	.314	.274	.332	5.439
1	.463	.458	.425	.567	.436	(.567)	.410	.424	.285	.356	.296	.223	.336	.355	.284	5.885
8	.242	.233	.203	.346	.297	.410	(.410)	.367	.328	.392	.245	.182	.312	.345	.232	4.544
7	.277	.240	.227	.388	.294	.424	.367	(.437)	.350	.437	.289	.209	.365	.337	.162	4.803
9	.265	.175	.252	.332	.342	.285	.328	.350	(.477)	.477	.202	.301	.367	.298	.234	4.685
6	.225	.203	.176	.356	.376	.356	.392	.437	.477	(.477)	.252	.303	.450	.345	.200	5.025
15	.273	.216	.189	.313	.287	.296	.245	.289	.202	.252	(.356)	.276	.356	.332	.179	4.081
14	.199	.122	.141	.219	.265	.223	.182	.209	.301	.303	.276	(.353)	.353	.332	.200	3.678
17	.303	.236	.245	.363	.314	.336	.312	.365	.367	.460	.356	.353	(.522)	.522	.272	5.316
11	.252	.224	.185	.393	.374	.355	.345	.337	.298	.345	.352	.332	.522	(.522)	.398	5.134
18	.217	.159	.200	.275	.332	.284	.232	.162	.234	.200	.179	.200	.272	.398	(.398)	3.742
D	5.513	4.976	4.821	6.114	5.439	5.885	4.544	4.803	4.685	5.025	4.081	3.678	5.316	5.134	3.742	73.756
E	.642	.579	.561	.712	.633	.685	.529	.559	.546	.585	.475	.428	.619	.598	.436	8.587
K	.642	.579	.561	.712	.633	.685	.529	.559	.546	.585	.475	.428	.619	.598	.436	8.587

Table 15  
 Computation of Second-Factor Loadings

	++10	++2	++5	++3	++4	++1	+-8	+-7	+-9	+-6	+-15	+-14	+-17	+-11	+-18	Check
++10	(.269)	.209	.204	.070	.018	.023	-.098	-.082	-.086	-.151	-.032	-.076	-.004	-.132	-.053	1.067
++2	.209	(.249)	.213	.084	.027	.081	-.073	-.084	-.141	-.136	-.059	-.122	-.004	-.122	-.093	1.870
++5	.204	.213	(.213)	.085	.073	.041	-.094	-.087	-.054	-.152	-.077	-.099	-.102	-.150	-.045	1.089
++3	.070	.084	.085	(.080)	.037	.079	-.031	-.010	-.057	-.061	-.025	-.086	-.073	-.033	-.035	.857
++4	.018	.027	.073	.037	(.136)	.032	-.038	-.009	-.064	-.000	-.014	-.006	-.078	-.105	-.050	.692
++1	.023	.081	.041	.079	.002	(.089)	-.048	-.041	-.089	-.046	-.029	-.070	-.088	-.055	-.015	.597
+-8	.098	.073	.094	.031	.038	.048	(.098)	-.071	-.039	-.083	-.006	-.044	-.015	-.029	-.001	.512
+-7	.082	.084	.087	.010	.060	.041	-.071	(.110)	-.045	-.110	-.023	-.050	-.019	-.003	-.082	.551
+-6	.086	.141	.054	.067	.004	.089	-.039	-.045	(.158)	-.158	-.057	-.067	-.029	-.029	-.004	.837
+-15	.151	.136	.152	.061	.096	.045	-.088	-.110	-.158	(.158)	-.026	-.063	-.088	-.045	-.056	1.103
+-14	.032	.050	.077	.025	.014	.029	-.006	-.023	-.057	-.026	-.077	-.073	-.062	-.008	-.028	.422
+-17	.070	.126	.099	.080	.066	.070	-.044	-.030	-.067	-.053	-.073	(.156)	-.088	-.076	-.013	.586
+-11	.094	.122	.102	.078	.078	.088	-.012	-.019	-.029	-.088	-.062	-.088	(.152)	-.092	-.013	1.159
+-18	.122	.122	.150	.093	.105	.055	-.020	-.003	-.029	-.025	-.068	-.076	-.152	-.137	-.137	1.180
+-18	.063	.093	.045	.055	.056	.016	-.001	-.082	-.004	-.055	-.028	-.013	-.002	-.137	(.157)	.310
Σ	.001	.004	.005	-.001	.001	.002	-.002	-.002	-.004	-.002	-.003	+.003	+.003	-.003	-.003	.015
D	1.067	1.870	1.089	+.857	+.505	+.507	-.512	+.551	+.857	1.103	+.422	+.855	1.139	+.180	+.316	14.169
E	.443	.499	.449	+.228	+.184	+.150	-.144	+.146	+.222	+.293	+.112	+.235	+.303	+.313	+.084	3.764
K	.443	.499	.449	+.228	+.184	+.159	-.144	-.146	-.222	-.293	-.112	-.235	-.303	-.313	-.084	.060

diagonal cell is recorded the highest correlation in the column. In row *D* is shown the sum of each column. In the lower right corner are recorded the sums required for the multiplier, and in row *E* are recorded the resulting first-factor loadings. These sums are checked by (22). Since all of the tests are positively intercorrelated, it is not necessary to reverse any of the signs in this table. The row *K* shows the factor loadings with the original signs of the tests. Since no sign changes are necessary, the last two rows are identical.

Table 15 shows the first-factor residuals. The diagonal entries are recorded first as residuals from the previous table. The sum of each column of residuals, including the diagonal, is recorded in row  $\Sigma_0$ . Each of these sums should be zero. Since the residuals are recorded to three decimals, the sum

Table 16

10	2	5	3	4	1	×	×	×	×	×	×	×	×	×	×	×	Check	<i>k</i> <sub>1</sub>
						8	7	9	6	15	14	17	11	18				
9	9	9	9	7	7	8	7	9	8	10	8	7	8	9	124	15		
8	8	8	8	6	6	7	8	8	7	4	9	8	9	8	112	14		
7	7	7	7	5	5	6	7	9	8	3	5	9	10	9	104	11		
6	6	6	6	4	4	7	8	8	7	2	4	10	4	10	92	18		
5	5	5	5	5	3	8	7	7	6	3	3	11	3	4	80	17		
4	4	4	4	4	2	7	8	8	7	2	2	3	2	3	64	7		
3	3	3	3	3	3	8	6	9	8	1	3	2	1	4	60	9		
2	2	2	2	2	2	9	5	5	9	2	2	1	2	5	52	6		
1	1	1	1	3	1	10	4	4	5	3	1	0	3	6	44	8		
0	0	0	0	2	2	4	3	3	4	4	2	1	2	5	32			

of the residuals in each column will be zero except for the discrepancies which are caused by rounding off the third decimal of each residual. The fact that these sums vanish within a small discrepancy in the last decimal place proves the arithmetical work.

Before the second-factor loadings can be extracted, some of the variables must be reflected. In order to ascertain which variables to reflect, Table 16 is prepared. A cross (×) at the top of a column indicates that the test of that column is to be reflected. The procedure in preparing this table is similar to that of the tables of sign changing which have already been described.

The sign changes are indicated in Table 15. A sign in the first position is the sign of the residual. The change of sign in each row is indicated in the second position. The change of sign in each column is shown in the third position. The sign next in front of the residual is the sign which is used in summing each column for the second-factor loading.



Table 17  
Computation of Third-Factor Loadings

	+ -10	+ -2	+ -5	+ +3	+ +4	+ +1	- +8	- +7	- +9	- +6	- -15	- -14	- -17	- -11	- -18	Check
+ -10	+(.048)	+.048	+.005	-.031	+.041	+.047	+.034	+.017	-.012	+.021	+.018	+.028	+.040	+.007	+.026	.347
+ -2	+.048	+(.051)	+.011	-.030	+.040	+.018	+.001	+.011	+.030	+.010	+.003	+.050	+.029	+.054	+.051	.208
+ -5	+.005	+.011	+(.046)	-.017	+.013	+.030	+.029	+.021	+.046	+.020	+.027	+.007	+.034	+.009	+.007	.096
+ +3	-.031	-.030	+.030	+(.043)	+.006	+.043	+.002	+.023	+.006	+.005	+.001	+.032	+.039	+.038	+.016	.213
+ +4	+.041	+.040	+.013	+.003	(.067)	+.019	+.019	+.040	+.026	+.045	+.001	+.025	+.037	+.063	+.067	.141
+ +1	+.047	+.018	+.030	+.043	+.019	(.071)	+.071	+.064	+.026	+.034	+.011	+.033	+.040	+.005	+.002	.364
- +8	+.034	+.001	+.020	-.002	-.019	+.071	(.078)	+.050	+.007	+.041	+.022	+.078	+.059	+.016	+.011	.480
- +7	+.017	+.011	+.021	+.023	+.040	+.064	+.050	(.094)	+.013	+.067	+.007	+.064	+.025	+.043	+.084	.539
- +9	+.012	+.050	+.046	+.006	+.026	+.054	+.007	+.013	+.098	+.003	+.082	+.015	+.038	+.098	+.023	.375
- +6	+.021	+.010	+.020	+.006	+.045	+.002	+.041	+.067	+.093	(.097)	+.059	+.010	+.001	+.097	+.080	.635
- -15	+.018	+.003	+.027	+.001	+.001	+.011	+.022	+.007	+.082	+.059	(.082)	+.047	+.028	+.063	+.037	.306
- -14	+.028	+.009	+.007	+.032	+.025	+.033	+.078	+.064	+.015	+.016	+.017	+.016	+.017	+.002	+.007	.346
- -17	+.040	+.029	+.034	+.009	+.037	+.040	+.059	+.025	+.038	+.001	+.028	+.017	(.059)	+.057	-.023	.450
- -11	+.007	+.034	+.009	+.038	+.033	+.005	+.016	+.043	+.008	+.097	+.033	+.002	+.057	(.111)	+.111	.680
- -18	+.026	+.051	+.007	+.016	+.007	+.002	+.011	+.004	+.023	+.050	+.037	+.007	+.023	+.111	(.111)	.230
$\Sigma$	.090	+.001	-.003	-.092	+.001	+.001	-.091	+.092	+.002	-.002	+.090	+.001	-.001	+.061	+.001	.001
D	.347	+.208	+.096	+.213	+.141	+.364	+.480	+.359	+.375	+.635	+.306	+.340	+.450	+.630	+.230	5.360
E	.150	+.090	+.041	+.092	+.061	+.157	+.207	+.233	+.162	+.274	+.132	+.149	+.194	+.272	+.099	2.313
K	.150	-.090	.041	+.092	+.061	.157	.207	.233	.162	.274	.132	.149	.194	.272	.099	.059

Each diagonal residual is erased. In each diagonal is recorded, instead, the largest residual of the column, irrespective of its sign. The diagonal entry is always recorded with positive sign.

The sum of each column is shown in row *D*. In row *E* is recorded the second-factor loading for each test after reflection. In row *K* is recorded the second-factor loading for each test taken with positive sign.

*Table 17* shows the second-factor residuals, and *Table 18* is the corresponding table of sign changes. The procedure is the same as for the preceding tables. The sum of each column, including the diagonal residual entry, is shown in row  $\Sigma_0$ . The fact that all of these sums vanish within a small discrepancy in the last decimal proves the arithmetical work. The diagonal residual entries are then erased, and the absolute maximum of each column is recorded in the diagonal cell. The sign changes indicated in *Table 18* are then made. The sum of each column, without the diagonal, is shown in the first row of *Table 19*. There are several negative entries in this row, namely, for columns 2 and 9. Variable 9 is changed in sign. The new sums, omitting diagonals, are recorded in the second row of *Table 19*. The second entry is still negative. The second variable is changed in sign, and the new sums are recorded in the third row. A negative sign appears in column 4. The fourth variable is then changed, and the sums are recorded in the last row. All sums are now positive. The entries in the last row of *Table 19* are added to the diagonal entries in *Table 17*. The sums are recorded in row *D* of *Table 17*. The factor loadings are recorded in row *E*. In row *K* are found the third-factor loadings for the original unreflected tests.

*Tables 20* and *21* are prepared in a similar way for determining the fourth-factor loadings.

*Tables 22*, *23*, and *24* are prepared for the fourth-factor residuals and the fifth-factor loadings. The residuals in *Table 22* are so small that they can be ignored. The standard deviation of discrepancies is .024.

The five factor loadings for each of Brigham's fifteen tests are summarized in *Table 25*. The contributions of the fifth factor to the correlations can be ignored. Each of the given intercorrelations can be reproduced from the first four factor loadings of this table within the discrepancies which are recorded in *Table 22*. It is an error, frequently made, to attempt a psychological interpretation of the factors in *Table 25*. It is not unlikely that each column of this table has psychological meaning, but there is no guaranty that such interpretation will be useful or fundamentally significant. The table represents merely the arbitrary centroid co-ordinates of a set of fifteen points in a space of five dimensions. The orthogonal reference axes which are obtained by the centroid method and which are represented by the five columns of *Table 25* must be rotated into a new set of orthogonal or oblique

Table 18

$\times$ 10	2	$\times$ 5	3	$\times$ 4	1	$\times$ 8	$\times$ 7	9	$\times$ 6	15	14	17	11	18	Check	$k_i$
8	7	6	8	8	7	7	6	7	9	7	7	8	7	8	110	6
9	6	7	7	7	6	8	7	8	5	6	6	6	6	7	102	10
5	7	8	6	6	5	9	8	7	4	5	5	5	8	94	8	
4	8	9	5	7	4	5	9	8	3	4	4	5	4	86	5	
3	7	5	4	8	3	4	10	7	2	5	3	4	5	78	7	
2	8	4	3	9	2	3	4	8	1	6	2	3	4	66	4	
3	7	3	4	5	1	2	3	7	2	5	1	4	5	58		

Table 19

Reflected Variables	10	$\times$ 2	5	3	$\times$ 4	1	8	7	$\times$ 9	6	15	14	17	11	18	$\Sigma$	$k_i$
$\Sigma(10, 5, 4, 8, 7, 6)$	.145	-.017	.190	.110	.058	.403	.424	.477	-.165	.282	.068	.366	.183	.129	.309	2.962	9
$\Sigma(10, 5, 4, 8, 7, 6, 9)$	.121	-.077	.098	.098	-.006	.295	.438	.503	.165	.468	.232	.336	.259	.325	.355	3.622	2
$\Sigma(10, 5, 4, 8, 7, 6, 9, 2)$	.217	.077	.076	.158	-.074	.331	.440	.525	.225	.448	.226	.318	.317	.393	.253	3.930	4
$\Sigma(10, 5, 4, 8, 7, 6, 9, 2, 4)$	.299	.157	.050	.170	.074	.293	.402	.445	.277	.538	.224	.268	.391	.519	.119	4.226	

Table 20  
Computation of Fourth-Factor Loadings

	--10	--2	--5	+13	+4	+11	+18	+17	+9	+6	+15	--14	--17	+11	+18	Check
--10	+.041	+.034	-.001	+.017	+.032	+.023	+.003	-.018	+.036	+.020	+.002	+.006	+.011	+.034	+.041	.217
--2	+.034	+.060	-.015	+.022	+.035	+.004	-.018	-.010	+.015	+.035	+.015	+.022	+.012	+.010	+.060	.117
--5	+.001	+.016	+.053	+.013	+.016	+.024	+.021	-.011	+.053	+.000	+.032	+.001	+.026	+.020	+.011	.256
+13	+.017	+.022	+.013	+.063	+.000	+.029	-.017	+.002	+.021	+.019	+.013	-.018	+.009	+.063	+.007	.297
+4	+.032	+.035	+.016	+.000	+.073	+.029	+.032	+.064	+.016	+.028	+.009	+.034	+.025	+.046	+.073	.154
+11	+.023	+.004	+.024	+.029	+.029	+.079	+.039	+.027	+.079	+.041	+.010	+.010	+.010	+.038	+.014	.456
+18	+.003	+.018	+.021	+.017	+.032	+.039	+.047	+.002	+.027	+.016	+.005	+.047	+.019	+.040	+.009	.272
+17	+.018	+.010	+.011	+.002	+.064	+.027	+.002	+.071	+.025	+.003	+.038	+.029	+.020	+.020	+.071	.157
+9	+.030	+.015	+.053	+.021	+.016	+.079	+.027	+.025	+.079	+.049	+.061	+.039	+.007	+.054	+.007	.524
+6	+.020	+.035	+.009	+.019	+.028	+.041	+.015	+.003	+.049	+.053	+.023	+.025	+.052	+.022	+.053	.454
+15	+.002	+.015	+.032	+.013	+.036	+.010	+.005	+.038	+.061	+.023	+.051	+.027	+.002	+.003	+.050	.160
--14	+.006	+.022	+.001	+.018	+.034	+.010	+.047	+.029	+.039	+.025	+.027	+.012	+.012	+.039	+.022	.256
--17	+.011	+.012	+.026	+.009	+.025	+.010	+.010	+.020	+.007	+.052	+.002	+.012	+.052	+.004	+.042	.145
--11	+.034	+.010	+.020	+.063	+.046	+.038	+.040	+.020	+.054	+.022	+.033	+.039	+.004	+.084	+.084	.557
--18	+.041	+.060	+.011	+.007	+.073	+.014	+.000	+.071	+.007	+.053	+.050	+.022	+.042	+.084	+.084	.226
Z <sub>2</sub>	.000	.000	+.001	+.001	+.002	+.001	+.002	.000	+.001	+.001	+.001	+.001	+.002	.000	+.001	.010
D	+.217	+.117	+.256	+.247	+.154	+.456	+.272	+.157	+.524	+.424	+.169	+.256	+.145	+.527	+.226	4.147
E	+.107	+.067	+.126	+.121	+.076	+.224	+.134	+.077	+.257	+.208	+.083	+.126	+.071	+.259	+.111	2.037
K	-.107	-.067	-.126	+.121	-.076	+.224	+.134	+.077	-.257	-.208	+.083	-.126	-.071	+.259	+.111	-.019

Table 21

10	2	5	3	× 4	1	8	7	× 9	× 6	× 15	14	17	× 11	× 18	Check	k <sub>i</sub>
7	7	7	6	7	6	8	7	7	7	10	7	5	8	9	108	15
6	6	6	5	6	5	7	6	6	6	8	4	8	6	7	8	9
5	7	5	4	7	4	6	5	6	9	3	7	7	8	9	92	6
4	6	6	3	8	3	5	6	5	5	2	6	6	9	10	84	18
3	5	5	4	7	2	4	7	4	4	3	5	5	10	4	72	11
2	6	4	3	8	1	3	6	3	3	4	4	6	4	3	60	4
3	7	3	4	6	0	2	5	2	2	5	3	7	3	4	50	

Table 22  
Computation of Fifth-Factor Loadings

	-+10	-+2	-+5	-+3	-+4	-+1	-+8	-+7	-+9	-+6	-+15	-+14	-+17	-+11	-+18	Check
-+10	+.040	+.028	+.014	+.004	+.040	+.001	+.011	+.026	+.009	+.002	+.007	+.007	+.003	+.006	+.029	.055
-+2	+.028	(.054)	+.022	+.015	+.039	+.009	+.026	+.014	+.030	+.023	+.010	+.029	+.008	+.025	+.054	.066
-+5	+.014	+.022	(.035)	+.002	+.006	+.004	+.004	+.001	+.021	+.035	+.022	+.016	+.017	+.013	+.003	.076
-+3	+.004	+.015	+.002	(.033)	+.006	+.002	+.039	+.007	+.010	+.006	+.003	+.003	+.018	+.032	+.020	.089
-+4	+.040	+.039	+.006	+.009	(.081)	+.012	+.022	+.048	+.004	+.012	+.015	+.024	+.030	+.026	+.081	.095
-+1	+.001	+.009	+.004	+.002	+.012	+.009	(.021)	+.009	+.010	+.021	+.006	+.018	+.006	+.029	+.011	.071
-+8	+.011	+.026	+.004	+.003	+.022	+.009	(.033)	+.008	+.007	+.012	+.006	+.030	+.009	+.005	+.006	.150
-+7	+.026	+.014	+.001	+.007	+.048	+.010	+.008	(.050)	+.005	+.010	+.010	+.032	+.019	+.025	+.000	.146
-+9	+.009	+.030	+.021	+.001	+.004	+.021	+.007	+.005	(.040)	+.004	+.040	+.007	+.025	+.013	+.022	.164
-+6	+.002	+.007	+.003	+.006	+.012	+.006	+.012	+.010	+.004	(.037)	+.006	+.001	+.037	+.032	+.030	.114
-+15	+.007	+.010	+.022	+.003	+.015	+.009	+.006	+.032	+.040	+.006	+.059	+.037	+.008	+.024	+.050	.203
-+14	+.007	+.029	+.003	+.024	+.018	+.030	+.030	+.019	+.007	+.001	+.037	(.037)	+.021	+.006	+.008	.180
-+17	+.003	+.008	+.017	+.018	+.030	+.006	+.009	+.025	+.025	+.037	+.038	+.021	(.037)	+.022	+.034	.098
-+11	+.006	+.025	+.013	+.032	+.026	+.020	+.005	+.005	+.013	+.032	+.024	+.006	+.022	+.055	+.055	.150
-+18	+.029	+.054	+.003	+.020	+.081	+.011	+.006	+.000	+.022	+.030	+.059	+.008	+.034	+.065	+.081	.233
$\Sigma$	+.001	+.001	+.000	+.002	+.001	+.001	+.001	+.001	+.001	+.001	+.002	+.000	+.000	+.002	+.000	.004
$D$	+.055	+.066	+.076	+.089	+.095	+.071	+.169	+.146	+.104	+.114	+.293	+.160	+.098	+.180	+.233	1.969
$E$	+.059	+.047	+.054	+.063	+.068	+.051	+.113	+.104	+.081	+.081	+.209	+.114	+.070	+.107	+.166	1.463
$K$	+.039	+.047	+.054	+.063	+.068	+.051	+.113	+.104	+.081	+.081	+.209	+.114	+.070	+.107	+.166	1.117

Table 23

$\times 10$	$\times 2$	5	$\times 3$	4	$\times 1$	8	7	$\times 9$	6	$\times 15$	14	17	11	18	Check	$\epsilon_i$
8	8	8	8	7	9	8	7	8	9	8	7	8	7	8	118	1
7	8	7	9	8	5	9	8	9	8	7	6	7	6	7	110	3
8	8	6	5	5	4	6	7	8	9	8	7	6	7	8	102	10
5	5	5	4	3	5	6	6	6	7	8	5	6	8	7	90	2
6	5	4	3	4	5	5	6	6	6	6	6	7	8	6	86	15
4	5	6	3	3	6	4	7	5	7	5	5	6	5	6	78	

Table 24

Reflected Tests	10	2	$\times$ 6	3	4	1	8	$\times$ 7	9	6	15	14	17	11	18	$\Sigma$	$k_i$
$\Sigma(1, 3, 10, 2, 9, 15)$	.085	.084	-.039	.074	.122	.038	.118	-.064	.072	-.031	.126	.131	.045	.069	-.014	.826	7
$\Sigma(1, 3, 10, 2, 9, 15, 7)$	.043	.056	-.041	.060	.026	.058	.134	.064	.082	.007	.190	.093	.095	.069	.146	1.082	5
$\Sigma(1, 3, 10, 2, 9, 15, 7, 5)$	.015	.012	.041	.056	.014	.050	.126	.066	.124	.077	.234	.123	.061	.095	.152	1.246	

Table 25

	I	II	III	IV	V
10	.642	.443	-.150	-.107	.039
2	.579	.499	-.090	-.057	.047
5	.561	.449	-.041	-.126	.054
3	.712	.228	.082	.121	-.063
4	.633	.134	.061	-.076	-.065
1	.685	.169	.157	.224	-.051
8	.529	-.144	.207	.134	.113
7	.559	-.146	.233	.077	-.104
9	.546	-.222	.162	-.257	.117
6	.585	-.293	.274	-.208	-.081
15	.475	-.112	-.132	.083	-.209
14	.428	-.235	-.149	-.126	-.114
17	.619	-.303	-.194	-.071	-.070
11	.598	-.313	-.272	.259	.107
18	.436	-.084	-.099	.111	.166

reference axes before psychological interpretation can be made with confidence. It has been shown that there exists an infinite set of orthogonal reference axes in terms of which the fifteen test vectors may be described as well as by those which are obtained by the centroid method. One of the principal problems in factor theory is to find a computationally feasible criterion by which this rotation can be effected uniquely and by which the reference axes so obtained have fundamental psychological meaning. The solution of this problem is described in several of the subsequent chapters. These solutions all begin with a given factorial matrix like that of *Table 25*. All of the solutions will be presented with the same set of illustrative data wherever feasible.

### Correction for uniqueness

It has been shown that the factorial matrix represents the co-ordinates of the termini of  $n$  trait vectors in a common-factor space of  $r$  dimensions. *Table 25* represents therefore the co-ordinates of fifteen points in five dimensions. The square of the length of each trait vector represents its communality. If the traits could be freed from the variable errors and from the specific factor, then the intercorrelations would be augmented in a manner analogous to the correction for attenuation. In correcting a coefficient for attenuation, the variable errors are removed. When a correlation between two traits is corrected not only for the variable errors but also for the specific factor in each test, the augmented correlation coefficient will be said to be "corrected for uniqueness." Hence the coefficients which are corrected for uniqueness are higher than those which are corrected only for attenuation.

The geometrical interpretation of the correction for uniqueness is of some interest. It has been shown that the correlation between two traits is the scalar product of the two trait vectors in the common-factor space. If each of the vectors is extended to meet the unit sphere so that each vector becomes a unit vector, then the scalar product of two such vectors is the cosine of their angular separation. The traits can then be represented as points on the surface of a hypersphere, and the angular separation between pairs of points represents the correlation after correction for uniqueness. For some problems in which the common-factor space can be reduced to three dimensions certain graphical methods are available in which each trait is represented as a point on the surface of a sphere. The plotting of the trait vectors on the surface of a sphere is facilitated by correcting the coefficients for uniqueness, because the augmented coefficient represents the cosine of the angular separation of a pair of vectors. The correction of the intercorrelations for uniqueness also facilitates the isolation of clusters of tests. These applications will be described in subsequent chapters.

The correlation coefficient can be regarded as the scalar product of a pair of test vectors. The lengths of the vectors are the square roots of their communalities. Hence

$$r_{jk} = h_j h_k \cos \phi_{jk},$$

so that

$$R_{jk} = \cos \phi_{jk} = \frac{r_{jk}}{h_j h_k},$$

in which  $R_{jk}$  is the correlation coefficient, corrected for uniqueness.

Table 26 shows the intercorrelations of Brigham's fifteen tests after correction for uniqueness for four factors.

Table 26

	10	2	5	3	4	1	8	7	9	6	15	14	17	11	18
10	1.000	.990	.988	.869	.886	.766	.478	.498	.480	.383	.648	.490	.520	.417	.655
2	.990	1.000	.993	.867	.856	.771	.448	.464	.401	.314	.572	.372	.412	.335	.590
5	.988	.993	.999	.866	.883	.764	.471	.499	.479	.392	.564	.408	.434	.308	.576
3	.869	.867	.866	.999	.956	.983	.826	.825	.636	.620	.794	.553	.628	.612	.827
4	.886	.856	.883	.956	.999	.906	.808	.833	.795	.751	.809	.697	.732	.594	.816
1	.766	.771	.764	.983	.906	1.000	.886	.873	.612	.628	.790	.515	.606	.644	.834
8	.478	.448	.471	.826	.808	.886	1.000	.994	.802	.856	.816	.673	.743	.732	.843
7	.498	.464	.499	.825	.833	.873	.994	.999	.853	.899	.800	.694	.751	.694	.821
9	.480	.401	.479	.636	.795	.612	.802	.853	.999	.983	.710	.844	.821	.557	.633
6	.383	.314	.392	.620	.751	.628	.856	.899	.983	.999	.679	.783	.775	.548	.662
15	.648	.572	.564	.794	.809	.790	.816	.800	.710	.679	1.000	.887	.943	.951	.995
14	.490	.372	.408	.553	.697	.515	.673	.694	.844	.783	.887	1.000	.988	.830	.841
17	.520	.412	.434	.628	.732	.606	.743	.751	.821	.775	.943	.988	.999	.898	.910
11	.417	.335	.308	.612	.594	.644	.732	.694	.557	.548	.951	.830	.898	.999	.947
18	.655	.590	.576	.827	.816	.834	.843	.821	.683	.662	.995	.841	.910	.947	.999



## CHAPTER IV

### THE PRINCIPAL AXES

#### A method of locating the principal axes\*

It has been shown that a set of traits may be regarded as  $n$  points in a common-factor space of  $r$  dimensions. It has also been shown that by rotation of  $F$  there exists an infinite number of factorial matrices which reproduce the correlations in  $R$ . It is natural to inquire whether a rotational criterion can be found by which a unique solution  $F$  may be obtained. One solution is to adopt the principal axes as the reference axes of  $F$ . The *principal axes* are defined as follows:

*Definition: If the sum of the squares of the projections of the test vectors on a radial axis is stationary, the axis is a principal axis.*

It can be shown that a set of vectors in a space of  $r$  dimensions has  $r$  principal axes and that these axes are orthogonal.

The attempted solution to the factor problem by which the trait vectors are described in terms of their projections on the principal axes is erroneous in spite of the fact that it is of considerable analytical interest. It will be described here with numerical examples partly because of its analytical interest but mainly because it will be shown in *chapter vii* to be psychologically meaningful when it is used in a modified form.

At the outset it may be stated that the method of principal axes does not give psychologically meaningful results. The matrix  $F$  which represents the principal axes of a battery has two serious limitations, namely, (a) the reference traits that are represented by the columns of  $F$  are a function of the *number* of traits of each kind that happen to be included in the battery, and (b) about half of the factor loadings beyond the first factor are necessarily negative. One of the fundamental requirements of a successful factorial method is that *the factorial description of a trait must remain invariant when the trait is moved from one battery to another which involves the same common factors or abilities*. When psychological tests are involved, a negative factor

\* The method of principal axes was first described in a paper which I presented at the Syracuse meeting of the American Association for the Advancement of Science in 1932. It was published in "The Theory of Multiple Factors," pp. 17-27. The method is given here in notation that is consistent with that of the previous chapters. Hotelling's special case of the method was described by him in "Analysis of a Complex of Statistical Variables into Principal Components," *Journal of Educational Psychology*, Vol. XXIV (September and October, 1933).

loading implies an ability whose possession is a detriment to the test performance. Such a situation can be comprehended for unusual situations, but it is not conceivable that half of the factor loadings in all special abilities should be negative. The reader may regard the method of principal axes as of analytical interest, but he should not expect to be able to give psychological meaning to the solution. A psychologically meaningful solution will be presented in *chapter vii*.

Each of the reference traits may be regarded as a unit vector in the same space of  $r$  dimensions in which the traits are represented by vectors whose scalars are less than unity. Let one such reference vector be  $\Delta_1$ , and let its direction cosines be  $\lambda_{11}, \lambda_{21}, \lambda_{31}, \dots, \lambda_{r1}$  in the common-factor space. The unit reference vector  $\Delta_1$  may be thought of as representing an imaginary pure trait. The correlation between a trait  $j$  and the reference vector  $\Delta_1$  will then be

$$(1) \quad r_{j\Delta_1} = a_{j1}\lambda_{11} + a_{j2}\lambda_{21} + \dots + a_{jr}\lambda_{r1},$$

or

$$(2) \quad r_{j\Delta_1} = \sum_{m=1}^r a_{jm}\lambda_{m1}.$$

This correlation is the projection of the vector  $j$  on the unit reference vector  $\Delta_1$ .

In order that the reference vector  $\Delta_1$  through the origin shall coincide with a principal axis of the system, it is necessary and sufficient that the sum of the squares of the projections of the trait vectors on the reference vector  $\Delta_1$  be stationary. We have then

$$(3) \quad r_{j\Delta_1}^2 = a_{j1}\lambda_{11} \sum_{m=1}^r a_{jm}\lambda_{m1} + a_{j2}\lambda_{21} \sum_{m=1}^r a_{jm}\lambda_{m1} + \dots + a_{jr}\lambda_{r1} \sum_{m=1}^r a_{jm}\lambda_{m1},$$

or

$$(4) \quad r_{j\Delta_1}^2 = \sum_{M=1}^r \sum_{m=1}^r a_{jm}a_{jM}\lambda_{m1}\lambda_{M1},$$

where the subscripts  $M$  and  $m$  refer to factors. Summing for  $n$  traits,

$$(5) \quad \sum_{j=1}^n r_{j\Delta_1}^2 = \sum_{j=1}^n \sum_{M=1}^r \sum_{m=1}^r a_{jm}a_{jM}\lambda_{m1}\lambda_{M1}.$$

For convenience, let

$$(6) \quad \sum_{j=1}^n r_{j\Delta_1}^2 \equiv u .$$

Then

$$(7) \quad \frac{\partial u}{\partial \lambda_{11}} = 2 \sum_{j=1}^n a_{j1}^2 \lambda_{11} + 2 \sum_{j=1}^n a_{j1} a_{j2} \lambda_{21} + \cdots + 2 \sum_{j=1}^n a_{j1} a_{jr} \lambda_{r1} ,$$

$$(8) \quad = 2 \sum_{m=1}^r \sum_{j=1}^n a_{j1} a_{jm} \lambda_{m1} ,$$

or

$$(9) \quad \frac{\partial u}{\partial \lambda_{M1}} = 2 \sum_{m=1}^r \sum_{j=1}^n a_{jM} a_{jm} \lambda_{m1} .$$

Since  $\lambda_{11}, \lambda_{21}, \dots, \lambda_{r1}$  are the direction cosines of the reference vector  $\Delta_1$  on the centroid axes, the solution is subject to the conditional equation,

$$(10) \quad v = \lambda_{11}^2 + \lambda_{21}^2 + \cdots + \lambda_{r1}^2 - 1 = 0 .$$

The constrained stationary values of  $u$  which satisfy the conditional equation (10) can be found by Lagrange's method of undetermined multipliers.\* We have then the following  $(r+1)$  simultaneous equations.

$$(11) \quad \begin{cases} \frac{\partial u}{\partial \lambda_{11}} + \beta \frac{\partial v}{\partial \lambda_{11}} = 0 , \\ \frac{\partial u}{\partial \lambda_{21}} + \beta \frac{\partial v}{\partial \lambda_{21}} = 0 , \\ \dots \dots \dots \\ \frac{\partial u}{\partial \lambda_{r1}} + \beta \frac{\partial v}{\partial \lambda_{r1}} = 0 , \end{cases}$$

$$(10) \quad v = \lambda_{11}^2 + \lambda_{21}^2 + \cdots + \lambda_{r1}^2 - 1 = 0 .$$

By means of these simultaneous equations the  $(r+1)$  unknowns  $\lambda_{11}, \lambda_{21}, \dots, \lambda_{r1}$  and  $\beta$  may be found. The partial derivatives of  $v$  are of the form

$$(12) \quad \frac{\partial v}{\partial \lambda_{m1}} = 2\lambda_{m1} .$$

\* William F. Osgood, *Advanced Calculus* (New York: Macmillan Co., 1928), p. 180.

Substituting (9) and (12) in (11), we have

(13)

$$\begin{aligned}
 &\lambda_{11} \left( \sum_{j=1}^n a_{j1}^2 + \beta \right) + \lambda_{21} \sum_{j=1}^n a_{j1}a_{j2} + \lambda_{31} \sum_{j=1}^n a_{j1}a_{j3} + \cdots + \lambda_{r1} \sum_{j=1}^n a_{j1}a_{jr} = 0, \\
 &\lambda_{11} \sum_{j=1}^n a_{j2}a_{j1} + \lambda_{21} \left( \sum_{j=1}^n a_{j2}^2 + \beta \right) + \lambda_{31} \sum_{j=1}^n a_{j2}a_{j3} + \cdots + \lambda_{r1} \sum_{j=1}^n a_{j2}a_{jr} = 0, \\
 &\lambda_{11} \sum_{j=1}^n a_{j3}a_{j1} + \lambda_{21} \sum_{j=1}^n a_{j3}a_{j2} + \lambda_{31} \left( \sum_{j=1}^n a_{j3}^2 + \beta \right) + \cdots + \lambda_{r1} \sum_{j=1}^n a_{j3}a_{jr} = 0, \\
 &\cdots \cdots \cdots \\
 &\lambda_{11} \sum_{j=1}^n a_{jr}a_{j1} + \lambda_{21} \sum_{j=1}^n a_{jr}a_{j2} + \lambda_{31} \sum_{j=1}^n a_{jr}a_{j3} + \cdots + \lambda_{r1} \left( \sum_{j=1}^n a_{jr}^2 + \beta \right) = 0.
 \end{aligned}$$

The multipliers  $\beta_p$  can be found from the fact that the determinant of the coefficients in (13) must vanish in order that solutions shall exist other than the trivial solution  $\lambda_{11}=\lambda_{21}=\cdots=\lambda_{r1}=0$ . The expansion of the determinant of the coefficients in (13) gives the *characteristic equation* of degree  $r$ , which may be written as follows:

(14)  $\beta^r + c_1\beta^{r-1} + c_2\beta^{r-2} + \cdots + c_r = 0.$

The coefficients  $c$  may be found by the following rules.

The numerical term  $c_r$  is the value of the  $r$ th order determinant of (13), ignoring  $\beta$ . The coefficient  $c_{r-1}$  is the sum of all the  $(r-1)$ -rowed principal minors in the same determinant. The coefficient  $c_{r-2}$  is the sum of all the  $(r-2)$ -rowed principal minors. The coefficient of  $\beta^r$  is always unity. The coefficient  $c_x$  in (14) is the sum of all the  $x$ -rowed principal minors. All of the roots of (14) are real and negative. They may be designated  $\beta_1, \beta_2, \dots, \beta_r$ .

Each of the roots  $\beta_p$  is substituted, in turn, in (13). Each of the  $r$  values of  $\beta_p$  gives a set of direction cosines for a principal axis. When the root  $\beta_p$  is substituted in (13), the solution gives the direction cosines of  $\Lambda_p$ , which are  $\lambda_{1p}, \lambda_{2p}, \dots, \lambda_{rp}$ .

The  $r$  principal axes  $\Lambda$  are orthogonal. In fact, their direction cosines may be arranged to form the matrix of the transformation from the given orthogonal co-ordinates in  $F$  to those of the principal axes. The matrix of the transformation is as given in *Table 1*.

It is of interest to note that each value  $-\beta_p$  is the sum of the squares of the projections of the trait vectors on the principal axis  $\Lambda_p$ . By inspection of the numerical values of the roots  $\beta_p$ , the major, mean, and minor axes of the system may be designated.

Table 1

$$\begin{vmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} & \dots & \lambda_{1r} \\ \lambda_{21} & \lambda_{22} & \lambda_{23} & \dots & \lambda_{2r} \\ \lambda_{31} & \lambda_{32} & \lambda_{33} & \dots & \lambda_{3r} \\ \dots & \dots & \dots & \dots & \dots \\ \lambda_{r1} & \lambda_{r2} & \lambda_{r3} & \dots & \lambda_{rr} \end{vmatrix}$$

**A numerical example of the method of principal axes**

The method of principal axes consists in the rotation of the co-ordinate system of  $F$  so that the principal axes constitute the orthogonal axes of reference. The principal axes may be defined as a set of orthogonal reference axes on each of which the sum of the squares of the projections of the trait vectors is stationary.

Let *Table 2* represent the factorial matrix  $F$  of a set of seven fictitious

Table 2

Tests	I	II	III	$h^2$
1	+ .5	- .2	+ .4	.45
2	+ .6	- .2	+ .6	.76
3	- .6	+ .5	+ .3	.70
4	- .3	+ .4	+ .6	.61
5	+ .2	+ .1	+ .7	.54
6	+ .6	- .4	0	.52
7	+ .7	- .3	+ .5	.83

tests in three factors. Geometrically, this table shows the three orthogonal co-ordinates of seven points. The communalities are listed in a column separate from  $F$ . They are all less than unity to correspond to the fact that every mental test may be assumed to have a specific factor in any finite battery of tests.

The sums required for the characteristic equation are as follows:

$$\sum_{j=1}^n a_{j1}^2 = + 1.95,$$

$$\sum_{j=1}^n a_{j1}a_{j2} = -1.07,$$

$$\sum_{j=1}^n a_{j1}a_{j3} = +.69,$$

$$\sum_{j=1}^n a_{j2}^2 = +.75,$$

$$\sum_{j=1}^n a_{j2}a_{j3} = +.11,$$

$$\sum_{j=1}^n a_{j3}^2 = +1.71.$$

These sums are substituted in the determinant of the characteristic equation as follows:

$$(15) \quad \Delta = \begin{vmatrix} (+1.95+\beta) & -1.07 & +.69 \\ -1.07 & (+.75+\beta) & +.11 \\ +.69 & +.11 & (+1.71+\beta) \end{vmatrix} = 0.$$

The values of  $\beta_p$  must be such as to make the determinant of the coefficients of the three homogeneous equations vanish in order that non-trivial solutions for  $\lambda_{1p}$ ,  $\lambda_{2p}$ ,  $\lambda_{3p}$  shall exist.

Expanding the determinant of (15), we get an equation of the form

$$(16) \quad \beta^3 + c_1\beta^2 + c_2\beta + c_3 = 0.$$

The numerical values of the coefficients of  $\beta_p$  are determined by the rule previously given. The numerical value of  $c_3$  is the value of the following determinant:

$$\begin{vmatrix} +1.95 & -1.07 & +.69 \\ -1.07 & +.75 & +.11 \\ +.69 & +.11 & +1.71 \end{vmatrix}.$$

It is found to be zero. Hence the rank of the determinant is less than 3. This proves that *Table 2* can be rotated so as to make at least one of its columns vanish.

The value of  $c_2$  is the sum of the three 2-rowed principal minors in (15), and  $c_1$  is the sum of the three 1-rowed principal minors, which is merely the sum of the three diagonal terms. The coefficient of  $\beta^3$  is unity. We then have the following values for the coefficients:

$$c_3 = 0, \quad c_2 = +4.4464, \quad c_1 = +4.41.$$

The expansion (16) can then be written as follows:

$$(17) \quad \beta^3 + 4.4100\beta^2 + 4.4464\beta + 0 = 0.$$

The roots of equation (17) are all real. One of the roots is zero. Dividing by  $\beta$ , we have the quadratic

$$(18) \quad \beta^2 + 4.4100\beta + 4.4464 = 0.$$

The two roots of this equation are  $-2.849690$  and  $-1.560310$ . Let these roots be designated by subscripts in the order of their numerical magnitude, namely,

$$(19) \quad \begin{cases} \beta_1 = -2.849690, \\ \beta_2 = -1.560310, \\ \beta_3 = 0. \end{cases}$$

Substituting  $\beta_1$  in the three simultaneous equations whose coefficients are shown in (13), we get three simultaneous equations,

$$(20) \quad \begin{cases} -.899690\lambda_{11} - 1.07\lambda_{21} + .69\lambda_{31} = 0, \\ -1.07\lambda_{11} - 2.099690\lambda_{21} + .11\lambda_{31} = 0, \\ +.69\lambda_{11} + .11\lambda_{21} - 1.139690\lambda_{31} = 0. \end{cases}$$

Solving for the ratios of  $\lambda_{11}$ ,  $\lambda_{21}$ ,  $\lambda_{31}$ , and normalizing them so that

$$(21) \quad \lambda_{11}^2 + \lambda_{21}^2 + \lambda_{31}^2 = 1,$$

we have

$$(22) \quad \begin{array}{l} \lambda_{11} = + .804972, \\ \lambda_{21} = - .386636, \\ \lambda_{31} = + .450036, \end{array}$$

$$\sum_{m=1}^3 \lambda_{m1}^2 = 1.000000.$$

These are the direction cosines of a unit reference vector which lies in the major principal axis of the system of seven points.

The second root,  $\beta_2$ , is then substituted in (15), and the same procedure gives the following values for the direction cosines of the unit reference vector  $\Delta_2$ , which lies in the mean principal axis of the system:

$$(23) \quad \begin{aligned} \lambda_{12} &= - .257498, \\ \lambda_{22} &= + .455692, \\ \lambda_{32} &= + .852080, \end{aligned}$$

$$\sum_{m=1}^3 \lambda_{m2}^2 = +1.000000.$$

The third root,  $\beta_3$ , is zero. The fact that the third root vanishes means that the sum of the squares of the projections of the seven test vectors on the minor principal axis is zero. Hence the projection of each of the seven tests on that axis is zero. Substituting  $\beta_3 = 0$  in (13), we obtain, by the same procedure as before, the values for the direction cosines of the unit reference vector  $\Delta_3$ , which lies in the minor principal axis of the system. These values are as follows:

$$(24) \quad \begin{aligned} \lambda_{13} &= - .534522, \\ \lambda_{23} &= - .801784, \\ \lambda_{33} &= + .267261, \end{aligned}$$

$$\sum_{m=1}^3 \lambda_{m3}^2 = +1.000000.$$

The direction cosines of the three principal axes of the system are arranged in *Table 3* to form the orthogonal transformation of the original

*Table 3*

	$\Delta_1$	$\Delta_2$	$\Delta_3$
1	+ .804972	- .257498	- .534522
2	- .386636	+ .455692	- .801784
3	+ .450036	+ .852080	+ .267261

co-ordinates of the matrix in *Table 2* to the co-ordinates that refer to the principal axes. This table corresponds to *Table 1* for the general case.

The fact that the new co-ordinates are orthogonal is verified by the fact



that the matrix in *Table 3* is orthogonal by columns. The correlations between the three principal axes can be expressed as follows:

$$(25) \quad \begin{cases} r_{A_1A_2} = \lambda_{11}\lambda_{12} + \lambda_{21}\lambda_{22} + \lambda_{31}\lambda_{32} , \\ r_{A_1A_3} = \lambda_{11}\lambda_{13} + \lambda_{21}\lambda_{23} + \lambda_{31}\lambda_{33} , \\ r_{A_2A_3} = \lambda_{12}\lambda_{13} + \lambda_{22}\lambda_{23} + \lambda_{32}\lambda_{33} . \end{cases}$$

Substituting the numerical values of *Table 3* in (25), it is seen that the three intercorrelations are zero. *Table 3* is the matrix of an orthogonal transformation. It must be orthogonal by rows and by columns; and its determinant must equal +1, since the matrix represents a rotation without reflection. These properties may be used as a check on the arithmetical work.

It is now possible to write the rotated form of  $F$  with the principal axes as co-ordinate axes. The new co-ordinates are shown in *Table 4*. The third column vanishes because one of the roots of the characteristic equation is zero. The communalities are listed in a separate column. They remain invariant under rotation.

*Table 4*

Test	$r_{jA_1}$	$r_{jA_2}$	$r_{jA_3}$	$h^2$
1	+ .659828	+ .120945	.00	.45
2	+ .830332	+ .265611	.00	.76
3	- .541290	+ .637969	.00	.70
4	- .126124	+ .770774	.00	.61
5	+ .437356	+ .590526	.00	.54
6	+ .637638	- .336776	.00	.52
7	+ .904489	+ .109084	.00	.83
$\sum_{j=1}^n r_{jA}^2$	2.849689	1.560312	.00	4.41

The three factor loadings of the first test in *Table 4* are obtained from the following equations:

$$(26) \quad \begin{cases} r_{1A_1} = a_{11}\lambda_{11} + a_{12}\lambda_{21} + a_{13}\lambda_{31} = + .659828 , \\ r_{1A_2} = a_{11}\lambda_{12} + a_{12}\lambda_{22} + a_{13}\lambda_{32} = + .120945 , \\ r_{1A_3} = a_{11}\lambda_{13} + a_{12}\lambda_{23} + a_{13}\lambda_{33} = .000000 . \end{cases}$$

The intertest correlations of the seven tests may be obtained either from the three factor loadings of *Table 2* or from the two factor loadings of *Table 4*. All of the intertest correlations are summarized in *Table 5*.

It is of interest to note that the sums of the squares of the factor loadings in the columns of *Table 4* are identical with the roots of the characteristic equation with reversed sign.

The present problem was set up so that one of the values of  $\beta$  would be zero. This was done by writing the loadings in *Table 2* so that the seven points were in the same plane. The points all satisfy the equation of an arbitrarily chosen plane, namely,  $2x+3y-z=0$ . In actual practice it is not likely that one of the values of  $\beta$  will be zero, but it may be very nearly

*Table 5*

	1	2	3	4	5	6	7
1	.....	+.58	-.28	+.01	+.36	+.38	+.61
2	+.58	.....	-.28	+.10	+.52	+.44	+.78
3	-.28	-.28	.....	+.56	+.14	-.56	-.42
4	+.01	+.10	+.56	.....	+.40	-.34	-.03
5	+.36	+.52	+.14	+.40	.....	+.08	+.46
6	+.38	+.44	-.56	-.34	+.08	.....	+.54
7	+.61	+.78	-.42	-.03	+.46	+.54	.....

zero. The fact that one or more of the roots of the characteristic equation are zero proves that the tests may be described in terms of less than  $r$  common factors. These common factors may be chosen to be statistically independent or dependent.

### Hotelling's special case

The method of principal axes is applicable for any diagonal values that preserve the Gramian properties of the correlation table. If reliabilities are recorded in the diagonal cells, or any other values greater than the communalities, the rank of the correlational matrix  $R$  will, in general, be equal to the number of tests. The centroid method will then give a factorial matrix  $F$  with as many columns as there are rows. This means that as many common factors are postulated as there are traits. Such a factorial matrix can be rotated by the method of principal axes so that the orthogonal reference vectors lie in the principal axes of the system of  $n$  points.

It has been shown in the previous chapter that in the special case where unity is recorded in the diagonals of the correlational matrix the centroid method gives a square factorial matrix  $F$  of  $n$  columns and  $n$  rows which re-

produces exactly all of the experimentally obtained correlation coefficients in  $R_0$ . The co-ordinate axes of this matrix may also be rotated into the principal axes of the system.

Hotelling has discussed this special case of the method of principal axes in which unity is recorded in the diagonals of  $R_0$ . He has called this special case the "method of principal components." The principal components are the projections of the trait vectors on the principal axes in the total factor space. He has described an ingenious iteration method by which the projections of the vectors on the principal axes in the total factor space may be found directly from the given coefficients in  $R_0$ . Unfortunately, this ingenious solution is not useful because it is subject not only to the fundamental limitations of the principal axes but also to additional limitations. The additional difficulties with Hotelling's case may be described as follows: To record unity in the diagonal cells of  $R_0$  implies that the total variance of each trait is to be described by *common* factors. It has been shown that the intercorrelations of  $n$  traits can always be accounted for exactly by  $n$  common factors. This can be done with the diagonal method described in *chapter ii*, by the centroid method of *chapter iii* with unity in the diagonals, or by the principal axes method with unity in the diagonals. Any solution in which the intercorrelations of  $n$  tests are accounted for exactly by  $n$  common factors must be an artifact as far as the psychological problem is concerned, because it is definitely known that each test has some *unique* variance. Three sources of unique variance may be listed, namely, (a) the chance errors in the test scores, (b) the specific factor in each test, and (c) the sampling errors in the correlation coefficients. Hotelling's case assumes that the tests are free from chance errors in the scores, that specific factors are absent, and that sampling errors are absent. This may be seen by considering the fact that his procedure gives a factorial matrix of  $n$  common factors which accounts for the coefficients exactly without any specific or unique variance whatever. As far as the psychological problem is concerned, such a solution is not acceptable.

In addition to these difficulties there must be considered the difficulties of the general principal axes solution which have been described in this chapter. These apply also to Hotelling's case. It is, of course, desirable that the axes of reference in terms of which the tests are described shall have psychological or genetic meaning. Consider any single test, such as a test of numerical manipulation. If this test is included in a battery which contains only a few number tests but many verbal tests, it is clear that the major principal axis will pass closer to the verbal tests than to the number tests. Now consider the same test when it is placed in a battery which contains only a few verbal tests but many number tests. The major principal

axis of this system will pass closer to the number tests. The factorial descriptions of the particular number test will be different in the two sets of computations, depending on the tests which are chosen arbitrarily for combination in a battery with the number test. It is not to be expected that such a factorial description should give psychologically meaningful axes

Table 6

	CENTROID CO-ORDINATES				$h_{IV}^2$
	I	II	III	IV	
10	.642	.443	-.150	-.107	.6424
2	.579	.499	-.090	-.057	.5956
5	.561	.449	-.041	-.126	.5339
3	.712	.228	.092	.121	.5820
4	.633	.134	.061	-.076	.4281
1	.685	.159	.157	.224	.5693
8	.529	-.144	.207	.134	.3614
7	.559	-.146	.233	.077	.3940
9	.546	-.222	.162	-.257	.4397
6	.585	-.293	.274	-.208	.5464
15	.475	-.112	-.132	.083	.2625
14	.428	-.235	-.149	-.126	.2765
17	.619	-.303	-.194	-.071	.5176
11	.598	-.313	-.272	.259	.5966
18	.436	-.084	-.099	.111	.2193
$\Sigma$	8.587	.060	.059	-.019	6.9653
$\Sigma k^2$	5.011317	1.183860	.428619	.341573	6.9654

Table 7

	1	2	3	4
1	$5.011317 + \beta$	.165747	.076365	.015854
2	.165747	$1.183860 + \beta$	-.053695	-.040502
3	.076365	-.053695	$.428619 + \beta$	-.044782
4	.015854	-.040502	-.044782	$.341573 + \beta$

of reference. This fundamental limitation is applicable also to the centroid method if any attempt is made to interpret the centroid coordinates directly without rotation. The purpose of the centroid method is merely to obtain a factorial matrix which accounts for the observed correlations within experimental errors and with the smallest possible number of common factors. The number of common factors is shown by the number of columns of  $F$  as found by the centroid method. Hotelling's iteration method might be used for the same purpose if it could be modified so as to use communalities instead of unity in the diagonals of  $R_0$ . As with the cen-

troid co-ordinates, a further rotation would be necessary in order to obtain a stable and fundamentally significant factorial description of the tests. No method is acceptable in this problem which distorts the rank of the correlational matrix in the common-factor subspace. These considerations make it necessary to discard the method of principal axes and also Hotelling's special case of this method as solutions to the psychological factor problem.

### The principal axes of a battery of fifteen psychological tests

In the previous chapter a battery of fifteen tests by Brigham was used as a numerical example of the centroid method of factoring the correlational

Table 8

	$\Delta_1$	$\Delta_2$	$\Delta_3$	$\Delta_4$
$\lambda_1$	.999	-.041	.015	-.012
$\lambda_2$	.043	.996	-.047	.070
$\lambda_3$	.016	-.072	-.909	.411
$\lambda_4$	.003	-.045	.414	.909

Table 9

	I	II	III	IV
10	.658	.431	.081	-.136
2	.598	.482	.043	-.061
5	.579	.433	-.028	-.107
3	.723	.186	-.034	.155
4	.639	.107	-.084	-.042
1	.694	.109	-.047	.271
8	.526	-.186	-.118	.190
7	.556	-.189	-.165	.149
9	.538	-.244	-.235	-.189
6	.576	-.326	-.313	-.104
15	.468	-.125	.167	.008
14	.415	-.235	.101	-.197
17	.602	-.310	.170	-.173
11	.580	-.328	.378	.095
18	.431	-.099	.146	.049

matrix. The same data will here be used as a numerical example of the principal axes. Table 6 contains the first four columns of Table (25-iii) and also the communalities for the first four centroid factors. The resulting coefficients of (13) are shown in Table 7. The expanded form (14) is as follows:

$$(27) \beta^4 + 6.965369\beta^3 + 10.810494\beta^2 + 5.407203\beta + .840052 = 0.$$

The four roots of this equation are

$$\beta_1 = -5.0197117$$

$$\beta_2 = -1.182688$$

$$\beta_3 = - .444971$$

$$\beta_4 = - .317999.$$

Substituting each of these four roots in (13) gives, after normalizing, the direction cosines of the four principal axes. These are listed in the four columns of *Table 8*. This table represents an orthogonal transformation  $L$  by which  $F$  is rotated into the principal axes. The factorial matrix  $FL$  is shown in *Table 9*. This matrix represents the same test configuration as the given *Table 6*. The only difference is that in *Table 9* the fifteen test vectors are described in terms of their projections on the principal axes, while in *Table 6* they are described in terms of their projections on the arbitrary orthogonal axes of the centroid method.

## CHAPTER V

### THE SPECIAL CASE OF RANK ONE

#### The intercolumnar criterion

The case of rank 1 is of special interest because it is the case to which Spearman and his students have given so much study. This is also the case which has been the subject of controversy during the past thirty years. Practically all of the scientific publications on the factor problem have been restricted to Spearman's special case of rank 1. It is only within the last few years that the more general case of the factor problem has been studied in which the rank exceeds 1 and in which any number of factors are treated analytically. Now that the factor problem has been generalized a step beyond the case of Spearman, it is of some interest to interpret a few of the old issues in a new light. The single-factor methods of Spearman may be interpreted in terms of the matrix formulation of the factor problem.

One of the earliest methods of Spearman was to ascertain the correlation between pairs of columns of  $R_0$ . Ideally, this correlation should be unity if the given correlations can be accounted for by a single common factor. It is a well-known property of determinants that if the rank is 1, then the columns are proportional and hence the intercolumnar correlations are unity. This property is stated in the following theorems.

*Theorem 1. If the correlational matrix is of rank 1, then all pairs of columns, or rows, are proportional.*

The converse of this theorem is also true, for if all pairs of columns are proportional, then all minors of second order or higher vanish, and hence the rank must be less than 2. The trivial case of rank 0 is here of no significance. The case in which the rank is 0 is, of course, identified by the fact that all the intercorrelations are 0. That is a case of no scientific interest. We have, therefore, the following converse theorem:

*Theorem 2. If all pairs of columns of the correlational matrix are proportional, then the rank of the matrix is 1 or 0.*

If a pair of columns are proportional, then the correlation between the columns is, of course,  $+1$ , so that we have the following theorem:

*Theorem 3. If the correlational matrix is of rank 1 then the correlation between any pair of columns is  $+1$  or  $-1$ .*

The converse of Theorem 3 is not necessarily true. A specific case which disproves the converse is as follows: Let the coefficients in a pair of columns be such that when one is plotted against the other, a linear plot is obtained

which does *not* pass through the origin. The correlation would be  $+1$ , but the coefficients in the two columns would not be proportional.

Spearman's former use of the intercolumnar criterion depended on the converse of Theorem 3, in that the high correlation between columns was the basis for the inference that a single factor was sufficient to describe the intercorrelations, i.e., that the rank of  $R_0$  was 1 within sampling errors. While the intercolumnar criterion is demonstrably fallible, it should be useful for rank 1, because it would be a rare situation in which a set of mental tests would satisfy the criterion when the rank was higher than 1.

Another type of difficulty appeared with the intercolumnar criterion. If all the coefficients in  $R_0$  are of the same order of magnitude and if these are overlaid with sampling errors, then the dispersion of a column may be comparable with the sampling errors, and the correlation between columns may be low because of the restricted range of the entries in the correlation table. The proportionality would still be maintained within sampling errors, but the points in the correlation table would be so restricted in range that the correlation coefficient would not reveal the proportionality. The *intercolumnar proportionality criterion* is therefore superior to the *intercolumnar correlation criterion*.

The limiting case of this effect is of some interest. If all of the coefficients in a correlation table are equal, then the proportionality criterion is satisfied but the correlation coefficient is indeterminate. The proportionality criterion would give the correct inference, namely, that the correlation matrix is of rank 1. We have then

$$(1) \quad \begin{cases} r_{jk} = a_{j1}a_{k1} , \\ r_{kl} = a_{k1}a_{l1} , \\ r_{jl} = a_{j1}a_{l1} . \end{cases}$$

But

$$(2) \quad r_{jk} = r_{jl} = r_{kl} ,$$

and hence

$$(3) \quad a_{j1}a_{k1} = a_{j1}a_{l1} = a_{l1}a_{k1} ,$$

$$(4) \quad a_{j1} = a_{k1} = a_{l1} .$$

It follows from (1) and (4) that

$$(5) \quad a_{j1} = a_{k1} = a_{l1} = \sqrt{r_{jk}} .$$



This limiting case is represented in the following theorem:

Theorem 4. *If all of the coefficients  $r_{jk}$  in a correlational matrix are equal, then the matrix is of rank 1 and each test has a single factor co-ordinate of  $\sqrt{r_{jk}}$ .*

If sampling errors are superimposed on this limiting case, the correlation between columns shows only the correlation between random errors. This correlation should be 0 or near 0. The intercolumnar proportionality criterion is still valid, and it would be only slightly affected by the sampling errors in a finite test battery.

Spearman's use of the correlational, rather than the proportionality, form of the intercolumnar criterion was determined, probably, by the fact that the standard error of a correlation coefficient can be determined, whereas the proportionality form of the criterion would require the development of an appropriate standard error formula. There does not seem to be any fundamental difficulty in doing so.

#### Graphical method for rank 1

The object of the factor problem is to find the factorial matrix which for rank 1 is a single column containing the one factor loading or co-ordinate for each test. The previous theorems suggest a simple graphical method of examining  $R_0$ . If the columns are proportional, then the plot of any column against any other column is linear through the origin. Let  $k$  and  $l$  designate any two columns, and let  $j$  designate any row of  $R_0$ . Then if  $r_{jk}$  is plotted against  $r_{jl}$ , the linear plot should be of the form

$$(6) \quad r_{jk} = c r_{jl},$$

where  $c$  is the slope constant. Substituting (1) in (6),

$$(7) \quad a_{j1}a_{k1} = c a_{j1}a_{l1},$$

or

$$(8) \quad c = \frac{a_{k1}}{a_{l1}}.$$

Since  $c$  is the slope of the plot, we have the following theorem:

Theorem 5. *If the rank of the correlational matrix is 1, and if any column  $k$  is plotted against any other column  $l$ , then the plot is linear through the origin with a slope which is the ratio of the single-factor loading of test  $k$  to that of  $l$ .*

Even if the tests are overlaid with sampling errors, this ratio is quite stable, and it may be determined by any suitable method of curve fitting. The

simplest method is probably the method of averages in which a line is drawn through the origin and through the centroid of the plot. If the coefficients deviate appreciably from 0, the slope of this line is not markedly affected by sampling errors. Hence this simple method should be useful in examining a table for rank 1. While this method is useful for examining a correlational matrix, it is not recommended for obtaining the single-factor loadings. A simpler and more direct method of solving the single-common-factor problem is described later in this chapter.

If the plot is not linear, or if the points scatter badly, the correlation table is of rank higher than 1, and Spearman's single-factor methods do not apply. If single-factor methods are to be used, the next step would be, no doubt, to try to find a subgroup which would give a linear plot through the origin. The intercorrelations of such a subgroup of tests could be accounted for by a single factor.

### The tetrad difference

Spearman's present method is to evaluate what are called "tetrad differences." The tetrad difference is of the form

$$(9) \quad r_{km}r_{ln} - r_{lm}r_{kn} = \rho,$$

where  $k$  and  $l$  refer to two rows, while  $m$  and  $n$  refer to two columns of  $R_0$ . The four subscripts refer to as many tests, and it is implied that four separate tests are involved in the tetrad-difference equation. Hence the tetrad difference is not written so as to include any diagonal terms of  $R_0$ . This is consistent with the fact that the communalities are unknown. Spearman has shown that if only one factor is involved, then all the tetrad differences in  $R_0$  vanish.

The tetrad differences have a very simple matrix interpretation. They are simply the expansions of second-order minors in the correlation table. If the rank of  $R_0$  is 1, then all second-order minors vanish. The converse is also true, for if the second-order minors vanish, then the rank must be 1, except for the trivial case when all entries are 0. The matrix interpretation of Spearman's tetrad-difference procedure is that rank 1 (i.e., a single common factor) is established by evaluating separately the second-order minors in the correlational matrix. One might speculate as to whether multiple-factor analysis would have developed earlier if this interpretation had been stated before. If the second-order minors must vanish in order to establish a single common factor, then must the third-order minors vanish in order to establish two common factors, and so on? To have put the matter in this way would have led to the matrix formulation of the problem.

To establish that a matrix is of any particular rank  $r$ , it is of course

necessary to prove that  $r$  is the highest order of the non-vanishing minors. Taken literally, this requires that all minors of order higher than  $r$  must be shown to be 0; but for computational purposes this is probably the most awkward way possible, especially for the single-common-factor case.

The tetrad-difference method of examining a correlation table cannot be recommended even for the restricted single-common-factor case to which it is theoretically applicable. The reason is that more effective methods are available for ascertaining whether one common factor is sufficient to account for the intercorrelations. If more than one factor is required, then the tetrad-difference criterion is not applicable. Some of the properties of the tetrad differences will be described here because of the fact that this way of ascertaining whether a correlation table is of rank 1 is in general use. There is considerable interest in the tetrads among students of factor theory.

Because of the great amount of labor that is involved in the computation of the tetrads for a large correlation table, it is convenient to know how many tetrads must be evaluated for  $n$  tests in order to cover the whole table. Since a tetrad difference is the value of a second-order minor which does not contain diagonal terms, there are as many tetrads as there are second-order minors which do not involve the diagonals. Each of these minors is defined by two rows and two columns. The number of pairs of rows that can be taken is the number of combinations of  $n$  things taken two at a time, or

$$(10) \quad C_2^n = \frac{n(n-1)}{2}.$$

The number of possible pairs that can be taken from the remaining columns, since diagonal elements are excluded, is then

$$(11) \quad C_2^{n-2} = \frac{(n-2)(n-3)}{2}.$$

Hence the total number of second-order minors in the correlation table, excluding diagonals, is

$$(12) \quad C_2^n C_2^{n-2} = \frac{n(n-1)(n-2)(n-3)}{4},$$

But since the correlational table is symmetric, it follows that every one of these minors is duplicated by a symmetric minor of the same value. Hence *the total number of different tetrads is*

$$(13) \quad \frac{1}{2} C_2^n C_2^{n-2} = \frac{n(n-1)(n-2)(n-3)}{8}.$$

Since every set of four variables gives three tetrads, it is possible to obtain the same result by considering the number of combinations of  $n$  things taken four at a time. Then the number of tetrads is

$$(14) \quad {}_3C_4^n = \frac{n(n-1)(n-2)(n-3)}{8}.$$

*Example:* If the number of tests is 20, then the correlation table contains 14,535 tetrads.

When the computation of all these tetrads has been made, the result is usually that the tetrads do not vanish. The inference must then be made that one common factor is insufficient to account for the intercorrelations of the tests. The question as to which of the tetrads will vanish and which of them will not vanish, and the question whether one common factor is sufficient, can be answered more easily by the other methods of this chapter.

If it is found that the tetrads do vanish within the sampling errors, then the next problem is to ascertain how much of the variance of each test is attributable to the single common factor. This can be done in terms of the correlation coefficients, as follows:

Consider the single-factor expression for the intercorrelations of any three tests  $j$ ,  $k$ , and  $l$ . Then

$$(15a) \quad r_{jk} = a_{j1}a_{k1},$$

$$(15b) \quad r_{jl} = a_{j1}a_{l1},$$

$$(15c) \quad r_{kl} = a_{k1}a_{l1}.$$

In order to find the loading  $a_{j1}$  of test  $j$  with the single common factor, divide (15b) by (15c). Then

$$(16) \quad \frac{r_{jl}}{r_{kl}} = \frac{a_{j1}}{a_{k1}},$$

so that

$$(17) \quad a_{k1} = \frac{a_{j1}r_{kl}}{r_{jl}}.$$

Substituting (17) in (15a),

$$(18) \quad r_{jk} = \frac{a_{j1}^2 r_{kl}}{r_{jl}},$$

from which we have Spearman's formula\* for the correlation of test  $j$  with the single common factor, namely,

$$(19) \quad a_{j1} = \sqrt{\frac{r_{jk}r_{jl}}{r_{kl}}}.$$

\* C. Spearman, *The Abilities of Man* (New York: Macmillan Co., 1927), Appendix, p. xvi, equation (19).

The value of (19) is subject to fluctuation with the sampling errors of the three coefficients in terms of which it is expressed. It is desirable to minimize this effect by taking an average value for  $a_{jl}$ , based on different pairs of tests  $k$  and  $l$  with which test  $j$  is combined. With  $n$  tests, the number of ways in which (19) can be written for test  $j$  is the number of pairs of tests that may be taken from the remaining  $(n-1)$  tests, excluding test  $j$ . Hence the number of ways in which (19) may be written is

$$\frac{1}{2}(n-1)(n-2).$$

Since there are  $n$  tests, the total number of formulae (19) for ascertaining the single-factor loading of all the tests is  $\frac{1}{2}n(n-1)(n-2)$ .

*Example:* In order to ascertain the correlation of each of 20 tests with the single common factor by all the determinations in  $R_0$ , formula (19) would be evaluated 3,420 times.

Spearman's procedure takes into consideration that the tetrad difference  $\rho$  in (9) does not quite vanish because of sampling errors in the four correlation coefficients. If a single common factor is fundamentally present and if the four coefficients have known standard errors, an expression for the standard error of  $\rho$  can be derived. This has been done by Wishart and by Holzinger.\* The experimentally observed deviations of  $\rho$  from zero should not exceed those which might be expected from the standard errors of  $\rho$ . This is the central idea in Spearman's single-common-factor method. The tetrads in a correlation table are first evaluated. A frequency distribution of these tetrad differences is then made and its standard deviation determined. If this dispersion is of an order of magnitude comparable with that which would be expected from the known standard errors of the tetrad differences, then Spearman draws the legitimate conclusion that a single common factor is sufficient to account for the observed intercorrelations. Applications of formulas of the type (19) give the loading of each test with the single common factor whose sufficiency has been established by the fact that the tetrads vanish within sampling errors.

It must be borne in mind that the vanishing of the tetrads, i.e., rank 1, does not prove the existence of a single common factor in the sense of a mental ability or in a genetic sense. This can be seen by considering the case where the  $r$  factor loadings are in the same proportion in all of the tests. All of the test vectors are then collinear in a common-factor space of

\* John Wishart, "Sampling Errors in the Theory of Two Factors," *British Journal of Psychology*, XIX (1928), 180-87.

Karl J. Holzinger, *Statistical Résumé of the Spearman Two-Factor Theory* (Chicago: University of Chicago Press, 1930), pp. 6-16.

$r$  dimensions, although their scalars may be different because of differences in the specific variances and in the error variances of the tests. This contingency is illustrated by the following fictitious factorial matrix of five tests and two factors.

If the matrix  $F$  of *Table 1* is multiplied by its transpose  $F'$ , it will be seen that the columns in  $R$  are proportional and that it is of rank 1. The tetrads vanish, and the intercorrelations of the tests can be described as well by one factor, as shown in the single-column matrix of *Table 2*. This example violates the postulate on page 57 in *chapter i*.

Table 1

	I	II
1	.60	.30
2	.40	.20
3	.80	.40
4	.20	.10
5	.30	.15

Table 2

	I
1	.6708
2	.4472
3	.8944
4	.2236
5	.3354

The reason why this result is obtained is that *Table 1* is of rank 1. It corresponds to the conceivable psychological situation in which each test of a battery calls for two primary mental abilities in the same ratio although they differ in specificity and reliability. In practice, it is possible to select from a large table of tests several groups whose intercorrelations are high when corrected for uniqueness. Each one of these groups of tests can be described in terms of one factor, but that factor is not necessarily psychologically significant. The tests may be composites, as illustrated in *Tables 1* and *2*. One way of avoiding this ambiguity is to work with several abilities or factors simultaneously, as is done in the multiple-factor methods, and to insure that a large number of zeros occur in each column of the table. This is one of the fundamental ideas developed in the subsequent chapters.

In order to reduce the labor of computing probable errors of the tetrads, Spearman and his students have developed several abbreviated procedures. These are all limited, however, to the single-common-factor case.

### Graphical analysis of tetrads

Although the tetrad method of examining a correlation table is not recommended, there is still so much interest in tetrads among students of factorial analysis that a simple graphical method of selecting the vanishing and the non-vanishing tetrads will be described. By Theorem 5, a plot of any column  $k$  against any other column  $l$  is linear if the rank of  $R_0$  is 1, i.e., if the intertest correlations can be described by a single common factor. If the test battery as a whole cannot be described by a single common factor, the

plot will not be linear but will scatter, as shown in *Figure 1*. This figure shows the plot of column 9 against column 1 from *Table II* of a recent factorial investigation by William Brown and William Stephenson.\* Column 9 represents a test of pattern perception, and column 1 represents a test of art

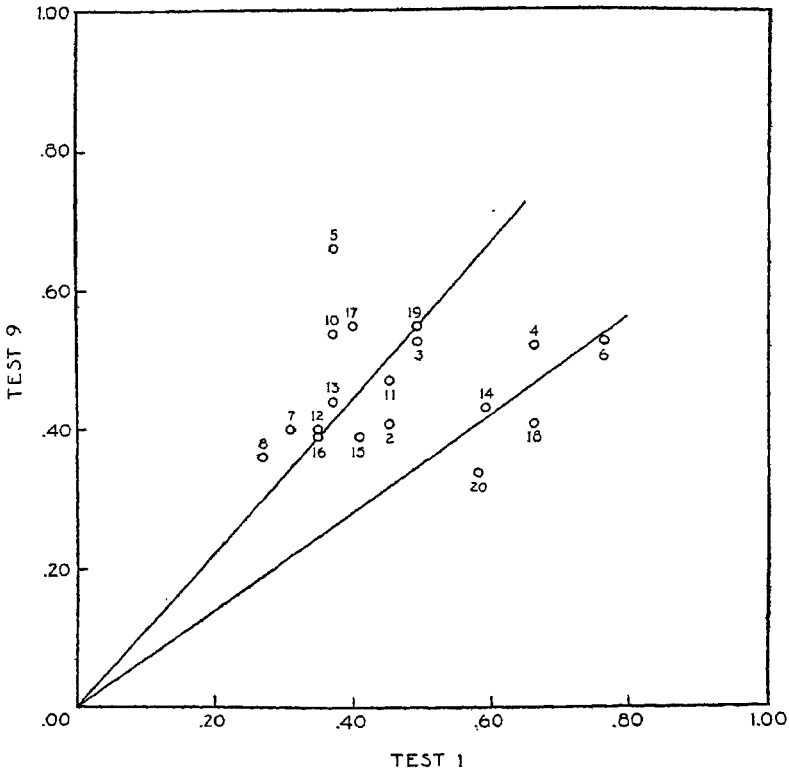


FIGURE 1

inventive synonyms. Although it is evident in *Figure 1* that a single factor is not sufficient to account for the intercorrelations and that, therefore, all of the tetrads do not vanish, it is still possible for smaller groups of tests to be of such a character that their intercorrelations can be described by a single common factor. Their tetrads should then vanish. The smallest group for which a tetrad can be written is four tests. Two tests are represented by the two columns that are plotted. Any pair of points on the diagram determines a tetrad. If they lie in a radial line through the origin, the corre-

\* "A Test of the Theory of Two Factors," *British Journal of Psychology (General Section)*, XXIII (1933), 352-70.

sponding tetrad vanishes. If they do not lie in a radial line through the origin, the corresponding tetrad does not vanish.

Let two points on the diagram represent tests  $m$  and  $n$ . If the two points are radial, then

$$(20) \quad \frac{r_{ln}}{r_{kn}} = \frac{r_{lm}}{r_{km}},$$

so that

$$(21) \quad r_{km}r_{ln} - r_{lm}r_{kn} = 0.$$

If the points  $m$  and  $n$  are not radial, the proportionality of (20) does not obtain and the tetrad (21) does not vanish.

A few numerical examples will be shown from *Figure 1*. A radial line can be drawn through the points 12 and 19. Hence the tetrad determined by these two points vanishes. The tetrad is as follows:

$$r_{12,17}r_{19,9} - r_{9,12}r_{19,1} = (.345)(.549) - (.401)(.489) = - .007.$$

A radial line cannot be drawn through the points 4 and 5, and hence the corresponding tetrad does not vanish. The tetrad is as follows:

$$r_{41}r_{59} - r_{49}r_{51} = (.656)(.655) - (.516)(.373) = + .237.$$

It may be of some interest to examine the tests, which are as follows:

- 1) Synonyms,
- 4) Disarranged sentences,
- 5) Fitting shapes,
- 9) Pattern perception.

This group of four tests probably has a common visual-perception factor in tests 5 and 9 which is not identical with the verbal factors in 1 and 4.

The previous group is as follows:

- 1) Synonyms,
- 9) Pattern perception,
- 12) Mutilated pictures,
- 19) Arithmetical equations.

Here there is either one common factor or several common factors whose factor loadings are in the same proportions in the four tests. It is probably significant that the battery contains only one arithmetical test. Any number factor in this test will remain specific in the battery as a whole, and hence it can have no effect on the vanishing of the tetrads.

The graphical method which has been described here can be used to indi-



cate, without numerical calculation, which of the tetrads will vanish almost exactly and which of them will have large residuals. It is possible that this graphical method could be extended so as to represent the probable errors of the tetrads; but since the tetrad method is not recommended even for the single-common-factor case, such a development would not be useful.

### A single-factor method without tetrads

If a correlation table is of rank 1 the single-common-factor loading of each test may be determined by a summation procedure. This method is simpler than the tetrad method for a single common factor, and it gives more information about the variables than the tetrad differences give.

If the correlation  $r_{jk}$  can be described by a single common factor, we have

$$(1) \quad r_{jk} = a_{j1}a_{k1} .$$

The sum of column  $k$  in the correlation table is

$$(22) \quad \sum_{j=1}^n r_{jk} = a_{k1} \sum_{j=1}^n a_{j1} .$$

Since the diagonal terms are unknown, the summation in the left member of (22) is unknown, and hence not suitable for computing purposes. Let  $(r)_k$  be the sum of column  $k$ , omitting the unknown diagonal entry. Then

$$(23) \quad (r)_k \equiv \sum_{j=1}^n r_{jk} - a_{k1}^2 ,$$

and hence

$$(24) \quad (r)_k = a_{k1} \sum_{j=1}^n a_{j1} - a_{k1}^2 .$$

Summing for all given coefficients in  $R_0$  and omitting the unknown diagonals, we have

$$(25) \quad (r)_t = \sum_{k=1}^n a_{k1} \sum_{j=1}^n a_{j1} - \sum_{k=1}^n a_{k1}^2 ,$$

or

$$(26) \quad (r)_t = \left[ \sum_{k=1}^n a_{k1} \right]^2 - \sum_{k=1}^n a_{k1}^2 ,$$

where  $(r)_t$  denotes the sum of all the coefficients in  $R_0$  except the diagonal ones.

Each of the two terms in the right member of (26) may be expressed in terms of summations of known coefficients, as follows:

$$(1) \quad r_{jk} = a_{j1}a_{k1},$$

and hence

$$(27) \quad \frac{r_{jk}}{a_{k1}} = a_{j1}.$$

Summing for column  $k$ ,

$$(28) \quad \frac{1}{a_{k1}} \sum_{j=1}^n r_{jk} = \sum_{j=1}^n a_{j1},$$

and from (23), it follows that

$$(29) \quad \frac{1}{a_{k1}} [(r)_k + a_{k1}^2] = \sum_{j=1}^n a_{j1},$$

or

$$(30) \quad \frac{(r)_k}{a_{k1}} + a_{k1} = \sum_{j=1}^n a_{j1} = \sum_{k=1}^n a_{k1}.$$

Hence the first term of the right member of (26) is

$$(31) \quad \left[ \sum_{k=1}^n a_{k1} \right]^2 = \frac{(r)_k^2}{a_{k1}^2} + 2(r)_k + a_{k1}^2.$$

The last term of (26) may be expressed in terms of summations of coefficients, as follows:

$$(1) \quad r_{jk} = a_{j1}a_{k1},$$

so that

$$(32) \quad r_{jk}^2 = a_{j1}^2 a_{k1}^2,$$

and hence

$$(33) \quad \frac{r_{jk}^2}{a_{k1}^2} = a_{j1}^2.$$

Summing for column  $k$ ,

$$(34) \quad \frac{1}{a_{k1}^2} \sum_{j=1}^n r_{jk}^2 = \sum_{j=1}^n a_{j1}^2 = \sum_{k=1}^n a_{k1}^2.$$

Summing for column  $k$ , except for the entry in row  $k$ ,

$$(35) \quad \frac{1}{a_{k1}^2} \left[ \sum_{j=1}^n r_{jk}^2 - r_{kk}^2 \right] = \sum_{j=1}^n a_{j1}^2 - a_{k1}^2.$$

Let the sum of the squares of the known coefficients in column  $k$  be denoted by  $(r^2)_k$ , so that

$$(36) \quad (r^2)_k \equiv \sum_{j=1}^n r_{jk}^2 - r_{kk}^2.$$

Then (35) can be written in the form

$$(37) \quad \frac{(r^2)_k}{a_{k1}^2} + a_{k1}^2 = \sum_{k=1}^n a_{k1}^2.$$

Substituting (31) and (37) in (26),

$$(38) \quad (r)_t = \frac{(r)_k^2}{a_{k1}^2} + 2(r)_k + a_{k1}^2 - \frac{(r^2)_k}{a_{k1}^2} - a_{k1}^2,$$

or

$$(39) \quad (r)_t - 2(r)_k = \frac{(r)_k^2}{a_{k1}^2} - \frac{(r^2)_k}{a_{k1}^2},$$

from which it follows that

$$(40) \quad a_{k1}^2 = \frac{(r)_k^2 - (r^2)_k}{(r)_t - 2(r)_k},$$

where

$(r)_k$   $\equiv$  the sum of the coefficients in column  $k$  omitting  
unknown diagonal entry,

$(r)_k^2$   $\equiv$  the square of  $(r)_k$ ,

$(r^2)_k$   $\equiv$  the sum of the squares of the known coefficients  
in column  $k$ ,

$(r)_t$   $\equiv$  the sum of all known coefficients in the correla-  
tion table.

Formula (40) gives the single-common-factor loading of each test in a correlation table of rank 1. This formula has been given, in different notation, by Spearman.\* In order to ascertain how well the single common factor accounts for the intercorrelations, a table of residuals should be computed. These residuals are defined as follows:

$$(41) \quad r_{jk} - a_{j1}a_{k1} = \rho_{jk},$$

where  $\rho_{jk}$  is the discrepancy between the given coefficient and that which is determined by the single-common-factor loadings of tests  $j$  and  $k$ . In order to ascertain which tests deviate most from the single-common-factor hypothesis for the whole battery, the mean absolute discrepancy for each test might be determined. It would be denoted

$$(42) \quad P_k = \frac{1}{n} \sum_{j=1}^n |\rho_{jk}|$$

for column  $k$ . This absolute mean discrepancy is a direct measure of the agreement between the given coefficients and the hypothesis of a single common factor. Hence it is superior to the tetrad differences, which constitute an indirect measure of the agreement.

A frequency distribution of the residuals may be made, and its dispersion may be compared with the standard error of the mean coefficient. A more formal treatment of the data would be to determine

$$(43) \quad E_{jk} = \frac{\rho_{jk}}{\epsilon_{jk}}$$

for each coefficient where  $E_{jk}$  is the ratio of the residual to the standard error of the given coefficient. A frequency distribution of these ratios should have a standard deviation not appreciably greater than unity in order to establish a single common factor as sufficient to account for the given correlation coefficients.

### Numerical example of single-factor method

An application of formula (40) has been made to a problem which has been used by several students of factor theory. *Table 3* contains the inter-

\* C. Spearman, *The Abilities of Man* (New York: Macmillan Co., 1927), Appendix, p. xvi, equation (21).

Table 3

	1	2	3	4	5	6	7	8
1	...	.58	.58	.51	.48	.43	.43	.46
2	.58	...	.47	.50	.53	.40	.44	.45
3	.58	.47	...	.51	.48	.41	.45	.38
4	.51	.50	.51	...	.34	.50	.35	.40
5	.48	.53	.48	.34	...	.33	.41	.39
6	.43	.40	.41	.50	.33	...	.41	.33
7	.43	.44	.45	.35	.41	.41	...	.31
8	.46	.45	.38	.40	.39	.33	.31	...
$(r)_k$	3.47	3.37	3.28	3.11	2.96	2.81	2.80	2.72
$(r)_k^2$	12.0409	11.3569	10.7584	9.6721	8.7616	7.8961	7.8400	7.3984
$(r^2)_k$	1.7447	1.6443	1.5628	1.4183	1.2864	1.1489	1.1358	1.0756
$a_{kl}^2$	.585677	.546265	.512004	.451027	.401892	.356995	.354345	.331384
$a_{kl}$	.765295	.739097	.715545	.671585	.633950	.597491	.595269	.575660

Table 4

First-Factor Residuals for Table 3

	1	2	3	4	5	6	7	8
1	.....	+.014	+.032	-.004	-.005	-.027	-.026	+.019
2	+.014	.....	-.059	+.004	+.061	-.042	.000	+.025
3	+.032	-.059	.....	+.029	+.026	-.018	+.024	-.032
4	-.004	+.004	+.029	.....	-.086	+.099	-.050	+.013
5	-.005	+.061	+.026	-.086	.....	-.049	+.033	+.025
6	-.027	-.042	-.018	+.099	-.049	.....	+.054	-.014
7	-.026	.000	+.024	-.050	+.033	+.054	.....	-.033
8	+.019	+.025	-.032	+.013	+.025	-.014	-.033	.....
$P_k$	.018	.029	.031	.041	.041	.043	.031	.023

correlations of eight tests which are reproduced from Holzinger.\* At the bottom of the table are recorded the three entries which are required by formula (40). The single-common-factor loading for each test is the last entry in each column. In *Table 4* are shown the residuals of (41). It will be seen that they are small. At the bottom of each column of *Table 4* is recorded the absolute mean discrepancy of (42). This example illustrates a single-factor method which does not require the computation of any tetrad differences. Each of the entries in *Table 4* might be divided by the standard error of the corresponding correlation coefficient, and a frequency distribution of these ratios might be prepared. Its standard deviation should not be much greater than unity.

\* Karl J. Holzinger, *Statistical Résumé of the Spearman Two-Factor Theory* (Chicago: University of Chicago Press, 1930), Table 6, p. 32.

## CHAPTER VI

### PRIMARY TRAITS

#### Simple structure

It has been shown that when the inequality (5—ii) is satisfied, there is a unique configuration of trait vectors that corresponds to the given correlational matrix. The correspondence between the configuration and the correlational matrix is independent of rotation of the orthogonal reference axes. The cell entries of  $F$  are altered within the range  $\pm h$ , under rotation of the reference axes. Since the rotation of the reference axes is arbitrary, it is clear that the numerical values in  $F$  can have no direct interpretation except in terms of some criterion which relates the configuration uniquely to the reference axes. If such a criterion can be found which satisfies the demands of the scientific problem, the reference axes become unique in relation to the configuration instead of remaining arbitrary under rotation.

The multiple-factor problem can be stated in two parts, namely:

- 1) What is the minimum number of factors that will account for the observed intercorrelations?
- 2) What is the minimum number of factors *for each trait* that will account for the intercorrelations?

The solution to the first of these two problems has been described in the previous chapters. The solution to the second problem will supply a unique factorial matrix.

The second problem is, in effect, to find the matrix representation of the simplest order among the traits that will account for the intercorrelations within the general restriction that the trait measures shall be linear functions of the several factors. If the traits involve  $r$  factors, the most complex order admissible in factor theory is that in which every one of the  $n$  traits involves every one of the  $r$  factors. The simplest possible order is that in which each trait can be described in terms of the smallest possible number of factors. The simplest order has been found when each row of the factorial matrix has at least one zero and when the number of zeros has been maximized.

It will be convenient to designate by special names some of the concepts that are involved in the problem of finding a unique factorial matrix of simplest possible order. The configuration of  $n$  trait vectors has  $r$  dimensions. The trait vectors are described in the matrix  $F$  by  $r$  orthogonal reference vectors.

**Definition:** The  $r$  orthogonal reference vectors that are implied in the factorial matrix  $F$  will be referred to as the **orthogonal reference vectors**.

**Definition:** When the factorial matrix  $F$  has been computed by the centroid method, its orthogonal reference vectors will be called the **centroid reference vectors**.

**Definition:** The unique configuration of trait vectors defined by the correlational matrix will be called the **correlational configuration** or the **trait configuration**.

Hence the factorial matrix  $F$  describes the trait configuration in terms of the orthogonal reference vectors.

**Definition:** The combined configuration of the  $n$  trait vectors and any set of  $r$  reference vectors will be called a **structure**.

Hence a structure is itself a configuration.

If the numerical entries in the factorial matrix shall have scientific interpretation, the reference vectors must have meaning beyond that of an arbitrary reference frame for the trait configuration. Each reference vector must be interpreted as a scientific category, so that the numerical entries of the factorial matrix have scientific meaning with reference to explanatory or descriptive categories. It is in this sense that the combined configuration of the trait vectors and the reference vectors constitutes a structure.

**Definition:** A structure in which each trait vector is contained in one or more of the  $r$  orthogonal co-ordinate hyperplanes will be called an **orthogonal simple structure**.

It follows from this definition that if a factorial matrix represents an orthogonal simple structure, each row of the matrix must have at least one zero. If  $r=3$ , each trait vector lies in at least one of the three orthogonal co-ordinate planes.

The nature of the second principal factor problem will be illustrated in three dimensions. Let *Table 1* represent a factorial matrix of rank 3. The crosses in the cells represent finite values of  $a_{jm}$ , while the other cells have zero entries. Some of the traits can be described in terms of only one factor, while the others have two factors each.

The corresponding configuration is shown in *Figure 1*, in which the three columns of the table are represented by the three reference vectors. The nine trait vectors are indicated by number. The trait configuration may be

Table 1

	I	II	III
1	○	○	×
2	○	×	○
3	×	○	○
4	×	×	○
5	×	○	×
6	○	×	×
7	×	○	×
8	○	×	×
9	×	×	○



described numerically by reference to any arbitrarily chosen set of co-ordinate axes. In general, an arbitrary set of axes is represented by a finite positive or negative value of  $a_{jm}$  in each cell of  $F$ . Orthogonal simple structure is shown if a set of axes can be found so that a large number of zeros appear in  $F$ , with at least one zero in each row. Since each one of the traits can be described in terms of fewer than three factors, simple structure can be found for the nine vectors of this example.

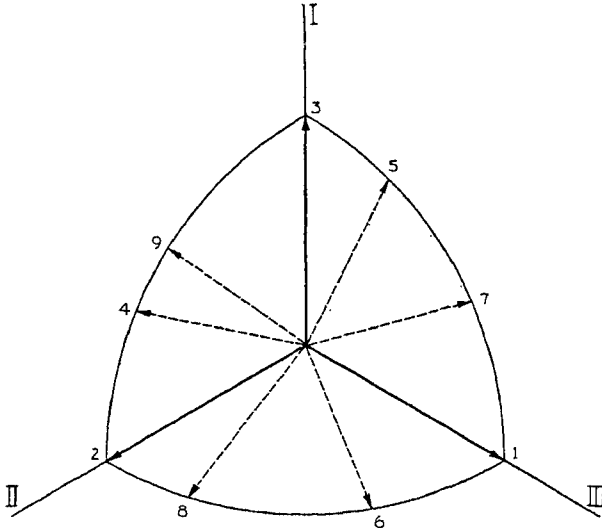


FIGURE 1

If an  $n \times r$  factorial matrix is set up with arbitrary entries in all cells, there is, in general, no transformation by which each of the  $n$  variables can be described in terms of fewer than  $r$  factors. It is assumed that  $n$  is large in comparison with  $r$ . Therefore the appearance of simple structure in a factorial matrix derived from observation commands attention. It is not a chance matter. When found in experimental data, it reveals order within the  $n$  variables in that  $r < n$  categories are required for describing them collectively and fewer than  $r$  categories are required for each one of them separately. When simple structure has been found by an orthogonal transformation  $L$  upon  $F$ , the reference vectors of  $FL$  represent fundamental categories which must be incorporated into the ideal constructs of the science.

*Definition: If an underlying physical order of the  $n$  traits is such that each of the traits can be described in terms of a smaller number of factors than are required for describing the traits collectively, then the underlying physical order will be called a **simple order**.*

It is here assumed that for the purposes of any particular scientific study a trait is completely described by that part of its total variance which is represented by the observed intercorrelations. Factors additional to those which are represented by the rank of the correlational matrix may be involved in the specific variance of each trait, but these are irrelevant in a factorial scientific study of the traits in which the correlational matrix is used as a datum.

Definition: *If a simple order exists for a set of  $n$  traits and if the  $r$  factors are statistically independent in the experimental population, then the corresponding physical order will be called an **orthogonal simple order**.*

Hence the configuration which represents an orthogonal simple order among the  $n$  traits is an orthogonal simple structure. An order among the traits involves, of course, not only the traits themselves but also the categories in terms of which they are described. These categories are themselves traits which may or may not be experimentally isolable.

It is useful to summarize the several fundamental concepts that are involved in this analysis. The concept *order* refers to the relation between the traits and the categories in terms of which the traits are to be described and comprehended. The *correlational matrix* describes merely the relations among the traits, independently of the descriptive categories. The *factorial matrix* describes the traits or variables in terms of an arbitrary set of descriptive categories. The *trait configuration* is a geometrical representation of the correlational matrix, and hence it is also independent of the descriptive categories. The *structure* is a configuration which represents not only the traits but also the arbitrary descriptive categories. The scientific problem is essentially a search for a set of descriptive categories in terms of which our conception of the traits or variables shall be the simplest possible. If an overdetermined and unique simplicity in our conception can be achieved, then the traits or variables will be said to reveal a *simple order*. The search for these categories has its direct analytical counterpart in the search for a set of *reference vectors* which shall reveal a *simple structure*. A simple structure is a configurational representation of a simple order. If the simplifying descriptive categories happen to be statistically independent in the experimental population, then the trait configuration can be so rotated in its arbitrary orthogonal frame that each trait vector is contained in one or more of the  $r$  orthogonal co-ordinate hyperplanes. The result is an *orthogonal simple structure*, and the reference vectors represent a set of statistically independent traits that serve the simplest possible comprehension of the given traits or variables.

### Oblique reference vectors

It is desirable to define the primary traits so as to describe the trait correlations in terms of the simplest possible order. Since the primary traits may not be orthogonal in the experimental population, the structure of the factorial matrix can be simplified by introducing oblique reference vectors. In the  $r$  dimensions of the common-factor space, each of  $r$  oblique reference vectors defines a co-ordinate hyperplane of  $(r-1)$  dimensions.

*Definition: The subspace of  $(r-1)$  dimensions which is orthogonal to the reference vector  $\Lambda_p$  will be called the **co-ordinate hyperplane**  $L_p$ .*

The oblique reference vectors will be referred to merely as reference vectors. Orthogonality will be explicitly designated.

The concept of orthogonal simple structure can be generalized to oblique reference axes.

*Definition: If a set of  $r$  hyperplanes of dimensionality  $(r-1)$  exists such that each trait vector is in one or more of the hyperplanes, then the combined configuration of the trait vectors and the reference vectors will be called a **simple structure** or an **oblique simple structure**.*

The factorial matrix which describes a simple structure of  $n$  traits in terms of  $r$  oblique reference vectors will have at least one zero in each row. Since its reference vectors are oblique, it cannot be obtained from  $F$  by an orthogonal transformation. A factorial matrix with oblique reference vectors will be denoted  $V$ .

The  $r$  reference vectors which are implied by the columns of  $F$  are orthogonal. Rotation of the system by the transformation  $L$  produces the matrix  $FL$ , which describes the same configuration. The  $r$  columns of  $FL$  also represent orthogonal reference vectors. Instead of subjecting  $F$  to an orthogonal transformation  $L$  in the attempt to find simple structure in  $FL$ , the new reference vectors will here be regarded as oblique. Then the factorial matrix  $F$  is subjected to some transformation  $G$ , not necessarily orthogonal, in the attempt to find simple structure in

$$(1) \quad FG \equiv V.$$

The only restriction upon the transformation  $G$  is that it shall be normalized by columns, i.e., that the sum of the squares of the elements in each column shall equal unity.

Each column of  $G$  represents the direction cosines of a reference vector. Since  $F$  is of rank  $r$ , it follows that  $G$  will be a square matrix of order  $r$ . Let  $\Lambda_p$  be the reference vector represented by column  $p$  in  $G$ , and let its direc-

tion cosines be  $\lambda_{1p}, \lambda_{2p}, \dots, \lambda_{rp}$ . The corresponding column in  $V$  contains the cell entries  $v_{jp}$ , where

$$(2) \quad v_{jp} = \sum_{m=1}^r a_{jm} \lambda_{mp}.$$

Each entry  $v_{jp}$  is the scalar product of the test vector  $j$  and the reference vector  $\Delta_p$ . Simple structure of  $V$  is shown if a transformation  $G$  can be found such that at least one of the entries  $v_{jp}$  vanishes in each row  $j$ . Each hyperplane  $L_p$  is determined by  $(r-1)$  points and the origin. Hence there must be at least  $r$  trait vectors in each hyperplane in order that it shall be overdetermined.

It will be useful to designate by a special name the number of reference vectors that are involved in the linear description of each trait.

Definition: *The number of reference vectors that are involved in the linear description of a trait will be called the **complexity** of the trait.*

It follows from this definition that in a simple structure every trait is of complexity less than  $r$ .

### Uniqueness of simple structure

When reference axes have been found which produce a simple structure, it is of considerable scientific interest to know whether the simple structure is unique. Consider a set of six trait vectors in three dimensions in which no three of the vectors are coplanar. There are fifteen different simple structures for this configuration. If there are seven trait vectors in three dimensions, no three of which are coplanar, then a simple structure is impossible because there exists no set of reference vectors, either orthogonal or oblique, by which a simple structure can be made.

The necessary and sufficient conditions for uniqueness of a simple structure need to be investigated. This is an important problem, because only in terms of its solution will it be possible to ascertain to what extent a particular simple structure is overdetermined by the experimental data on which it is based. In the absence of a complete solution to this problem three criteria will here be listed which are almost certain to constitute sufficient and more than necessary conditions for the uniqueness of a simple structure. The scientific interpretation of the cell entries in the oblique factorial matrix  $V$  should not be attempted except after reasonable assurance that the simple structure of  $V$  is unique. It is part of the faith of science that if several alternative simple structures can be found for the matrix  $V$  and if each of them can be given plausible descriptive categories, only one of the alternatives can eventually remain acceptable.

The three criteria by which the  $r$  reference vectors can be overdetermined are as follows:

- 1) Each row of  $V$  should have at least one zero,
- 2) Each column of  $V$  should have at least  $r$  zeros,
- 3) For every pair of columns of  $V$  there should be at least  $r$  traits whose entries  $v_{jp}$  vanish in one column but not in the other.

The first criterion demands that each trait should be describable in terms of fewer categories than are required by the whole set of  $n$  traits. It is conceivable that, in some experimental work, one or more of the traits will be so complex as to require description in terms of all of the factors that enter into the traits collectively. For the purpose of isolating the fundamental categories, these traits are not useful, and they should therefore be ignored. The criterion demands that the list of traits be long enough so that after elimination of several traits of complexity  $r$ , enough traits of complexity less than  $r$  remain to determine uniquely both the trait configuration and the simple structure. This principle may be illustrated with psychological tests. If one of the abilities to be isolated should be number sense, then this primary ability should not be required in all of the tests of a battery. The same restriction applies to each of the abilities that are to be isolated.

The second criterion seems to be essential for the following reason. Each column  $p$  of  $V$  is determined by a hyperplane  $L_p$ . A hyperplane through the origin is determined by  $(r-1)$  trait vectors. These trait vectors are contained in  $L_p$ , and therefore they have vanishing entries  $v_{jp}$  in column  $p$ . Therefore there must be at least  $(r-1)$  traits with vanishing entries in each column of  $V$  in order that the hyperplanes shall be determined. Since the hyperplanes should be overdetermined by the data, it follows that the number of vanishing entries in each column of  $V$  should equal or exceed  $r$ .

The third criterion is suggested by the fact that the  $r$  hyperplanes must be distinct. If two columns of  $V$  contain the same vanishing entries and if these exceed  $(r-2)$  in number, then the two corresponding hyperplanes are identical. The third criterion was written so as to insure both overdetermination and distinctness of the hyperplanes that define the columns of  $V$ .

### Primary trait vectors

If a set of  $r$  hyperplanes has been found such that each trait vector is contained in one or more of the hyperplanes, then the  $n$  traits can be comprehended as a simple order in which each trait is of complexity less than  $r$ . This implies that each row of the factorial matrix  $V$  has at least one zero. These hyperplanes will be specially designated as follows:

Definition: *The  $r$  hyperplanes whose normals produce a simple structure with a trait configuration will be called the **co-ordinate hyperplanes** for the trait configuration.*

Any other set of co-ordinate hyperplanes will be called arbitrary co-ordinate hyperplanes. The simple structure is defined by the trait configuration and the normals  $\Lambda_p$  to the co-ordinate hyperplanes  $L_p$ .

The corresponding co-ordinate axes are defined as follows:

Definition: *The intersection of any  $(r-1)$  co-ordinate hyperplanes defines a **co-ordinate axis** of the structure.*

The total number of sets of  $(r-1)$  hyperplanes is  $r$ , and consequently their intersections define  $r$  co-ordinate axes. These are of special scientific interest because they define the descriptive categories of the simple order.

Definition: *The unit vector defined by a co-ordinate axis will be called a **primary trait vector** or a **primary vector**.*

Definition: *The trait which corresponds to a primary vector will be called a **primary trait** or a **primary factor**.*

The object of a factorial analysis is to discover the primary traits and to describe them in terms of the traits that are experimentally observed.

A simple structure is represented diagrammatically in *Figure 2* for three dimensions. The hyperplanes of dimensionality  $(r-1)$  are planes in this special case. They are shown by the arcs  $L_p$ . The normals  $\Lambda_p$  are also shown. These are the reference vectors. Each primary vector  $T_p$  is the intersection of the  $(r-1)$  hyperplanes  $L_q$ , where  $q \neq p$ . In three dimensions there are three planes  $L_p$  which contain the origin. Their intersections determine the primary vectors  $T_p$ .

The intersection of all the co-ordinate hyperplanes, excepting  $L_p$ , defines a primary trait vector which will be denoted  $T_p$ . Hence  $T_p$  defines the linear subspace which is common to all the hyperplanes, excepting  $L_p$ . The trait vector  $T_p$  is not contained in the hyperplane  $L_p$ , but it is contained in all the other hyperplanes. It follows that the primary trait  $T_p$  is absent in all of the traits which have vanishing entries  $v_{jp}$  in column  $p$  of  $V$ . The primary trait  $T_p$  is present in all traits that have non-vanishing entries  $v_{jp}$  in column  $p$  of  $V$ .

Since the primary trait vector  $T_p$  is not contained in the hyperplane  $L_p$ , it might be inferred that it is identical with the normal to the hyperplane  $L_p$ . Such is not necessarily the case. If the primary traits are statistically independent in the experimental population, then the vectors  $T_p$  are orthogonal, and so are also the co-ordinate hyperplanes  $L_p$  and their normals  $\Lambda_p$ . In this case the reference vectors  $\Lambda_p$  are identical with the primary vectors  $T_p$ . However, if the primary traits  $T_p$  are not statistically independent in

the experimental population, then the hyperplanes  $L_p$  are oblique, and their normals  $\Lambda_p$  are oblique. The two sets of vectors  $\Lambda_p$  and  $T_p$  are then, in general, distinct.

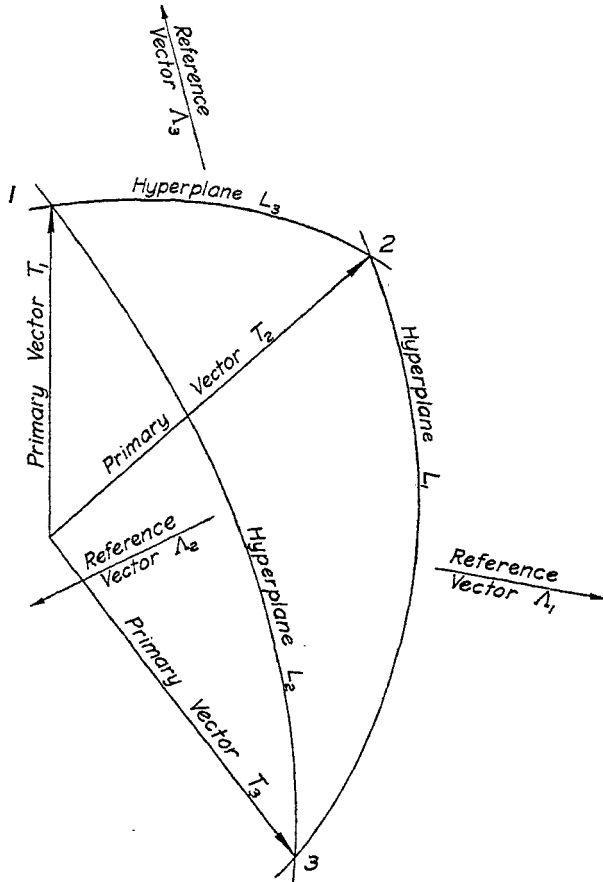


FIGURE 2

The geometrical interpretation of primary traits may be illustrated in three dimensions. Let the entries  $a_{jm}$  in each row of  $F$  be augmented by the multiplier  $1/h_j$ . The geometric representation of the augmented co-ordinates is that each trait vector is extended to unit length. The augmented co-ordinates are therefore the direction cosines of the trait vectors. The termini of the trait vectors can be represented as points on the surface of a hypersphere. If  $r=3$ , the trait configuration can be studied graphically on the surface of a ball.

Let *Figure 3* represent the trait configuration, and let the points represent the termini of the trait vectors on the surface of the sphere. Simple structure is shown by the fact that each point lies in one of the three arcs of great circles. All of the traits on the arc 1-2 can be described by two primary factors, since all of the corresponding trait vectors are coplanar. The whole set of traits can be described by three factors. Hence the same primary factor is absent in all of the traits along 1-2. The subspace 1-2 can be described by the direction cosines of the normal to the plane 1-2. Let this normal be denoted  $\Lambda_3$ . The subscript of  $\Lambda_3$  refers to the primary trait  $T_3$  that is absent in the subspace  $L_3$ . The vector  $\Lambda_3$  is the normal which defines the subspace  $L_3$ .

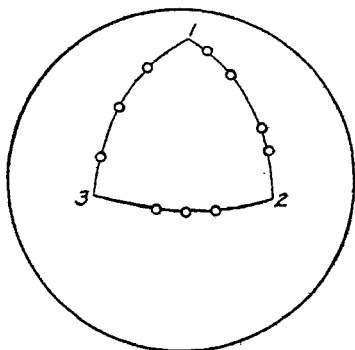


FIGURE 3

By analogy, the vector  $\Lambda_1$  is the normal to the plane 2-3 and  $\Lambda_2$  is the normal to the plane 1-3. If all trait vectors in the plane 1-2 represent two primary factors and if all trait vectors in the plane 1-3 represent two primary factors, it is clear that the vector which is determined by the intersection of these two planes represents the primary factor which the two planes have in common, namely, the primary factor No. 1. In the same manner the other primary factors, 2 and 3, are determined by intersections of planes.

The direction cosines of the reference vectors  $\Lambda_p$  constitute the columns of the matrix of transformation  $G$  by which simple structure is demonstrated in the traits of *Figure 3*.

The reasoning about *Figure 3* may be generalized as follows. The matrix of the transformation  $G$  is shown in (3) in which the  $r$  entries of column  $p$  show the direction cosines of the hyperplane  $L_p$ . The primary trait vector  $T_p$  is defined by the intersection of  $(r-1)$  hyperplanes, omitting  $L_p$ . Each column of (3) shows the  $r$  coefficients of the homogeneous linear equation which defines a hyperplane. If  $(r-1)$  of these equations are solved simul-



taneously, the ratios of the unknowns are the ratios of the direction cosines of the intersection. Normalizing these ratios, we have the direction cosines of the primary trait vector  $T_p$ :

$$(3) \quad \left\| \begin{array}{cccccc} \lambda_{11} & \lambda_{12} & \lambda_{13} & \dots & \lambda_{1r} \\ \lambda_{21} & \lambda_{22} & \lambda_{23} & \dots & \lambda_{2r} \\ \lambda_{31} & \lambda_{32} & \lambda_{33} & \dots & \lambda_{3r} \\ \dots & \dots & \dots & \lambda_{mp} & \dots \\ \lambda_{r1} & \lambda_{r2} & \lambda_{r3} & \dots & \lambda_{rr} \end{array} \right\| \equiv G.$$

The relation between the oblique reference vectors and the primary trait vectors can be generalized as follows. Let  $\mu_{mp}$  be the first minor of  $\lambda_{mp}$  in (3). Then it can be shown that the direction cosines of the primary trait vector  $T_p$  are proportional to the entries in column  $p$  in the matrix  $H$  in (4). Let the matrix  $T$  be produced by normalizing the columns of  $H$ . The columns of  $T$  are the direction cosines of the primary trait vectors  $T_p$ :

$$(4) \quad \left\| \begin{array}{cccccc} +\mu_{11} & +\mu_{12} & +\mu_{13} & \dots & +\mu_{1r} \\ -\mu_{21} & -\mu_{22} & -\mu_{23} & \dots & -\mu_{2r} \\ +\mu_{31} & +\mu_{32} & +\mu_{33} & \dots & +\mu_{3r} \\ \dots & \dots & \dots & (-1)^{m-1}\mu_{mp} & \dots \\ (-1)^{r-1}\mu_{r1} & (-1)^{r-1}\mu_{r2} & (-1)^{r-1}\mu_{r3} & \dots & (-1)^{r-1}\mu_{rr} \end{array} \right\| \equiv H.$$

The cosine of the angular separation of each pair of primary trait vectors is the correlation between the corresponding primary traits in the experimental population. It will probably be found that these correlations are positive. If the elements  $v_{ip}$  are taken positive or zero, then the angular separations of pairs of hyperplanes  $L_p$  exceed  $\pi/2$ . Hence it is to be expected that the scalar products, or correlations, of pairs of oblique vectors  $\Delta_p$  will be negative. This relation can be illustrated in the plane as follows.

Let  $I$  and  $II$  in *Figure 4* represent the orthogonal centroid axes of  $F$ . Let the small circles along the line 1 represent traits which all contain the same primary trait. In a similar way let the small circles along the line 2 represent traits that contain a second primary trait. Let the two primary traits be positively correlated in the experimental population. This is shown by the fact that the angular separation between 1 and 2 is less than a right angle.

The radial line 1 is a subspace of one dimension. It is defined by its normal  $\Lambda_2$ . The subscript of  $\Lambda_2$  shows the trait which is absent in the subspace 1. The correlations between  $\Lambda_2$  and the traits in 1 are all zero. The correlations between  $\Lambda_2$  and the traits in 2 are all positive. In a similar manner  $\Lambda_1$  is the normal to the subspace 2. All correlations between  $\Lambda_1$  and the traits in 2 are zero. The correlations between  $\Lambda_1$  and the traits in 1 are posi-

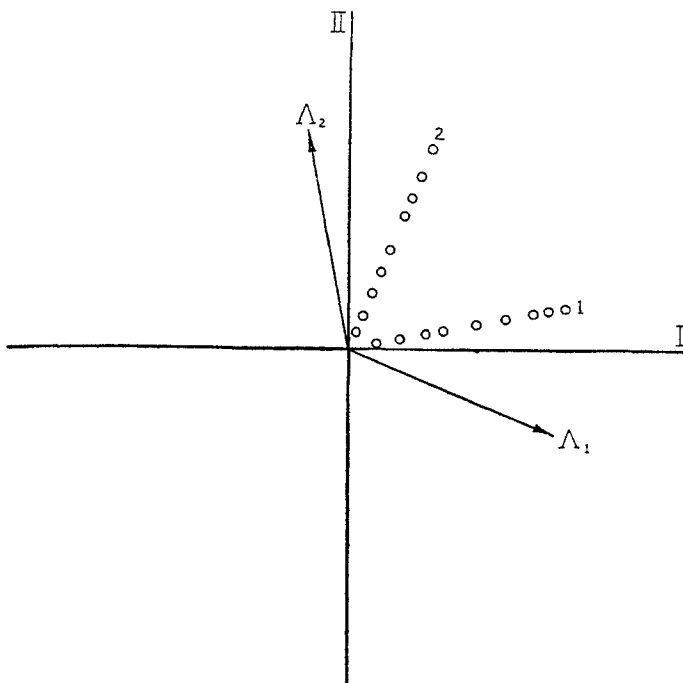


FIGURE 4

tive. While the angular separation between 1 and 2 is less than  $\pi/2$ , it is seen that the angular separation between  $\Lambda_1$  and  $\Lambda_2$  exceeds  $\pi/2$ . Hence its cosine is negative.

The convincingness of the primary traits that are isolated by the three criteria for overdetermined simple structure necessarily depends on the inventiveness of the scientist in formulating a plausible concept or descriptive category for each column of  $V$ . When that has been done, the further verification of these concepts demands that additional experiments be made with traits or variables in which the categories are represented in extreme form. For example, if facility in auditory imagery were postulated as a primary factor in dealing with certain verbal tests, the verification of this fac-

tor would require additional experimentation in which the same verbal tests are used in combination with non-verbal tests of auditory facility. If the latter tests retain even more conspicuous values in the auditory column than the auditory verbal tests, then the auditory factor is further experimentally affirmed.

### The equation of an oblique simple structure

If a set of  $r$  hyperplanes exists such that each of the trait vectors is in at least one of the hyperplanes, then the configuration is called a simple structure. The general case is that in which the hyperplanes are oblique. The special case in which the  $r$  hyperplanes are orthogonal is called an orthogonal simple structure. A simple structure is a set of  $r$  oblique hyperplanes, all of which contain the origin. This set of  $r$  hyperplanes may be regarded as a degenerate cone whose apex is at the origin and whose surface consists of the  $r$  hyperplanes. The equation of the hyperplane  $L_p$  is

$$(5) \quad \sum_{m=1}^r x_m \lambda_{mp} = 0.$$

The equation of a simple structure in  $r$  dimensions may be written by setting the product of  $r$  polynomials, like (5), for  $p=1, 2, 3, \dots, r$ , equal to zero. Then we have

$$(6) \quad \left[ \sum_{m=1}^r x_m \lambda_{m1} \right] \left[ \sum_{m=1}^r x_m \lambda_{m2} \right] \dots \left[ \sum_{m=1}^r x_m \lambda_{mr} \right] = 0.$$

This equation may be written in the more condensed form

$$(7) \quad \prod_{p=1}^r \left[ \sum_{m=1}^r x_m \lambda_{mp} \right] = 0.$$

Fitting (7) to a given trait configuration in which a simple structure is assumed, we have, for each test  $j$ ,

$$(8) \quad v_{j1} v_{j2} \dots v_{jr} = 0,$$

or

$$(9) \quad \prod_{p=1}^r v_{jp} = 0.$$

If the point  $j$  is in at least one of the  $r$  hyperplanes of (7), at least one of the  $r$  factors  $v_{jp}$  vanishes, and hence equation (9) is satisfied. Equation (7), or its equivalent (9), is therefore satisfied by all points in the  $r$  co-ordinate hyperplanes of a simple structure.

In order to determine the best-fitting degenerate cone for a given set of  $n$  points in a space of  $r$  dimensions, equation (7) may be written in the form of an observation equation, namely,

$$(10) \quad \prod_{p=1}^r \left[ \sum_{m=1}^r a_{jm} \lambda_{mp} \right] = \rho_j,$$

where  $\rho_j$  is the discrepancy for the point  $j$ . The best-fitting simple structure may be defined as that in which

$$\sum_{j=1}^n \rho_j^2$$

is minimized. Hence the criterion for a best-fitting simple structure is the minimizing of

$$(11) \quad \sum_{j=1}^n \prod_{p=1}^r \left[ \sum_{m=1}^r a_{jm} \lambda_{mp} \right]^2 = \sum_{j=1}^n \prod_{p=1}^r v_{jp}^2 = \phi.$$

The function  $\phi$  will be referred to as the criterion for the isolation of a simple structure.

### Five methods of isolating primary traits

When a factorial matrix  $F$  has been obtained from the correlational matrix  $R_0$  by any method, the second principal problem is to find the transformation  $G$  by which overdetermined simple structure may be discovered in the  $n$  traits or variables. Five methods will be described, namely:

- 1) Graphical method when  $r < 4$ ,
- 2) Method of oblique axes,
- 3) Method of averages,
- 4) Method of maximizing zero entries in each column of  $V$ ,
- 5) Analytical method.

When the rank of  $F$  is less than 4, the problem is quite simple, because the solution may be written by graphical methods. These will be described in later sections of this chapter.

When a psychological hypothesis in the form of a postulated primary trait is to be tested, the second and third methods are applicable. The

method of oblique axes is described in *chapter vii*. It is a variant of the method of principal axes which avoids the pitfalls of the principal axes. The method of averages is also described in *chapter vii*. It is an approximation to the method of oblique axes, and it does not require the determination of the roots of a characteristic equation. Either of these methods may be used for testing directly whether a postulated trait is primary. The fourth method isolates one hyperplane at a time in which the number of nearly vanishing entries is maximized.

The fifth method is entirely analytical in that it extracts the primary traits, if they exist, without presupposing any hypothesis regarding their nature. By means of the analytical method the primary traits may be found even if their nature is entirely unknown. The analytical method is described in *chapter vii*.

### Graphical methods for less than four dimensions

If the common-factor space has one dimension, the factor problem is solved directly by the methods of *chapter v*. If the rank is 2, the co-ordinates  $a_{jm}$  of  $F$  may be plotted on cross-section paper so that the configuration of  $F$  becomes visible in a plane. Simple structure is then revealed if all of the trait vectors are found to lie along two radial lines. The direction cosines of the normals to these lines constitute the columns of the transformation  $G$ , which can then be written by inspection. These direction cosines can also be thought of as defining two linear subspaces in the plane.

If the rank of  $F$  is 3, the graphical procedure implies a solid model. Two methods of handling this case will be described:

1) The trait vectors may be represented by old-fashioned hatpins that are stuck into a central spherical cork. The length of each hatpin  $j$  should be equal to  $h_j$ . The angular separation  $\phi_{jk}$  for each pair of hatpins,  $j$  and  $k$ , should be such that  $r_{jk} = h_j h_k \cos \phi_{jk}$ . Simple structure is demonstrated if each hatpin lies in one of three planes all of which contain the origin. The direction cosines of the normals to these planes constitute the columns of the transformation  $G$ .

2) The method of sticking hatpins in a cork has not been tried. A simpler method is to use the augmented co-ordinates  $A_{jm}$  of the trait vectors. These are also the direction cosines of unit trait vectors. They can be represented as points on the surface of a ball.

In plotting the trait configuration on a sphere, it will be found convenient to use the following method: Locate three orthogonal points on the surface and mark them I, II, III, to represent the three reference vectors for the direction cosines of the traits. Mark off on a narrow strip of paper the distance  $\pi D/4$ , where  $D$  is the diameter of the sphere. This is also the sur-

face distance between any two of the orthogonal reference points. Divide this distance into ninety parts in any convenient units to represent  $90^\circ$ . On the same strip of paper mark off the cosines of angles with any convenient unit such as .00, .05, .10, etc. In doing this, look up  $\cos^{-1} .05$ ,  $\cos^{-1} .10$ , and mark .05, .10 on the strip at the appropriate angles. The strip is then ready for use. In locating a point on the sphere, use the arithmetical check which is provided by the fact that the position of each point is determined by the angular separations from two reference points. The angular separation from the third reference point constitutes an arithmetical check.

Simple structure is demonstrated if each point lies in one of the sides of a spherical triangle. Each of the three sides determines a plane through the origin. The direction cosines of the normal to each plane constitute a column in the transformation  $G$ . The vertices of the triangle define the primary traits. The correlations between the primary traits in the experimental population are the cosines of the intervening sides of the spherical triangle. If the figure is a right spherical triangle, the primary traits are uncorrelated in the experimental population. If the figure is an oblique spherical triangle, the primary traits are correlated in the experimental population.

### The problem of negative abilities

All of the methods of factoring a correlational matrix that have been described give a factorial matrix with negative cell entries in the second and subsequent columns. The numerical values in each row of the factorial matrix  $F$  describe one of the traits in terms of arbitrary orthogonal reference axes. Since the axes are arbitrary, the psychological interpretation of the matrix  $F$ , as obtained by the centroid method or by any other equivalent method, is certain to lead to erroneous results unless the matrix is rotated so as to satisfy some additional criterion of the relation between the trait configuration and the reference axes.

If the variables in the correlational matrix represent personality traits other than abilities, then either positive or negative values of  $v_{jm}$  are admissible to psychological interpretation. For example, "tactfulness" and "tactlessness" are two traits which can be so defined that their co-ordinates are identical except for reversal of sign. Two disease symptoms might be in a similar inverse relation. The likes and dislikes of people might be related in the same manner. The trait "stability" probably has a negative projection on the reference vector of "emotionality."

If the traits in the correlational matrix represent abilities, it is not likely that the values of  $v_{jm}$  in  $V$  will be negative. By one interpretation, a negative value of  $v_{jm}$  would mean that performance in a psychological test  $j$  is

actually facilitated by the lack of some sort of ability  $m$ . Ideal constructs can be devised so as to allow some plausible interpretation for what might be called "negative abilities," but this does not seem to be necessary. This reasoning leads to the

Hypothesis: *When unique simple structure is found for a battery of psychological tests, then the non-vanishing entries in the factorial matrix are positive.*

It will be convenient to name the bounded space within which any radial vector has only positive direction cosines. This space will be defined as follows:

Definition: *The bounded space in which any radial vector has only positive direction cosines will be called the **positive region**.*

Definition: *If all the trait vectors that do not lie in a hyperplane are on the same side of it, the hyperplane will be called a **positive hyperplane** with reference to the trait configuration.*

Definition: *If a set of  $r$  positive hyperplanes exists such that each trait vector is contained in one or more of them, then the combined configuration of the trait vectors and the reference vectors will be called a **positive simple structure**.*

The geometrical interpretation of the restriction upon the numerical value of  $a_{im}$  in  $F$  in the case of mental ability tests is that all of the test vectors lie in the positive region of the common-factor space. When this condition is satisfied, all of the intertest correlations are positive or zero. It is a universally accepted fact that intertest correlations are positive.

The converse is not necessarily valid. The well-known fact that all intertest correlations are positive implies that all of the test vectors lie inside a cone with center at the origin and with a generating angle of  $\pi/4$ . Such a cone cannot be inscribed in the positive region except when the number of dimensions is as low as two.

The restriction that all of the test vectors shall be in the positive region of the common-factor space is not sufficient to determine  $F$  uniquely. In general, there exists an infinite number of orthogonal transformations by which all of the entries in  $F$  become positive or zero if the configuration of  $F$  can be inscribed in the positive region. Special cases may be set up in which one, and only one, orthogonal transformation will make the entries  $a_{jm}$  in  $F$  positive or zero. Such a case in three dimensions is that in which three test vectors are mutually orthogonal. These cases are not likely to be found in practice. Hence a unique matrix  $F$  is not to be expected with the single criterion that  $a_{jm} \geq 0$  in  $F$ .

## Graphical analysis of fifteen psychological tests

The graphical methods will be illustrated on the fifteen psychological tests of Brigham which were used for numerical examples in the third and fourth chapters. The fourth column in *Table (25-iii)* contains entries whose maximum contribution to any correlation coefficient is about .067. This is not large enough to justify serious consideration, and hence the test vectors can be represented in three dimensions with fair approximation. In *Table 2* are recorded the first three centroid factors and the communalities for the first three factors. The last three columns show the corresponding augmented co-ordinates.

Table 2

TESTS		CENTROID CO-ORDINATES			$h^2$	DIRECTION COSINES		
		I	II	III		I	II	III
10.	Opposites	.642	.443	-.150	.6309	.808	.558	-.189
2.	Opposites	.579	.499	-.090	.5923	.752	.648	-.117
5.	Opposites	.561	.449	-.041	.5180	.779	.624	-.057
3.	Analogies	.712	.228	.092	.5674	.945	.303	.122
4.	Artificial language	.633	.134	.061	.4224	.974	.206	.094
1.	Definitions	.685	.159	.157	.5192	.951	.221	.218
8.	Geometrical completion	.529	-.144	.207	.3434	.903	-.246	.353
7.	Arithmetical problems	.559	-.146	.233	.3881	.897	-.234	.374
9.	Arithmetical proportions	.546	-.222	.162	.3736	.893	-.363	.265
6.	Number series	.585	-.293	.274	.5032	.825	-.413	.386
15.	Card-turning	.475	-.112	-.132	.2556	.939	-.222	-.261
14.	Block construction	.428	-.235	-.149	.2606	.838	-.460	-.292
17.	Dice-counting	.619	-.303	-.194	.5126	.865	-.423	-.271
11.	Painted cubes	.598	-.313	-.272	.5296	.822	-.430	-.374
18.	Form-learning	.436	-.084	-.099	.2070	.958	-.185	-.218

The fifteen test vectors can be represented as points on a sphere. When this is done, the configuration of the test vectors can be inspected independently of the arbitrary centroid reference planes. In *Figure 5* the configuration is shown by plotting the augmented factor loadings *III* against *II*. The first factor is then perpendicular to the plane of the diagram.

Inspection of the sphere shows that three reference planes may readily be located so that each of the test vectors lies in at least one of the three planes or very near to one of them. The plane *AB* is determined by the centroids of the tests 2, 5, and 8, 7. The plane *AC* is determined by the centroids of the tests 6, 9, and 11, 14. The plane *BC* is determined by the centroids of the tests 10, 2, and 11, 14.

The direction cosines of the three planes are shown in *Table 3*. This table is the matrix *G* of the transformation of *F* into *V*. Each column shows the



direction cosines of an oblique reference vector  $\Lambda_p$ , and these are also the direction cosines of one of the three subspaces of dimensionality  $(r-1)$ . In the present example these subspaces are the planes  $AB$ ,  $AC$ , and  $BC$ .

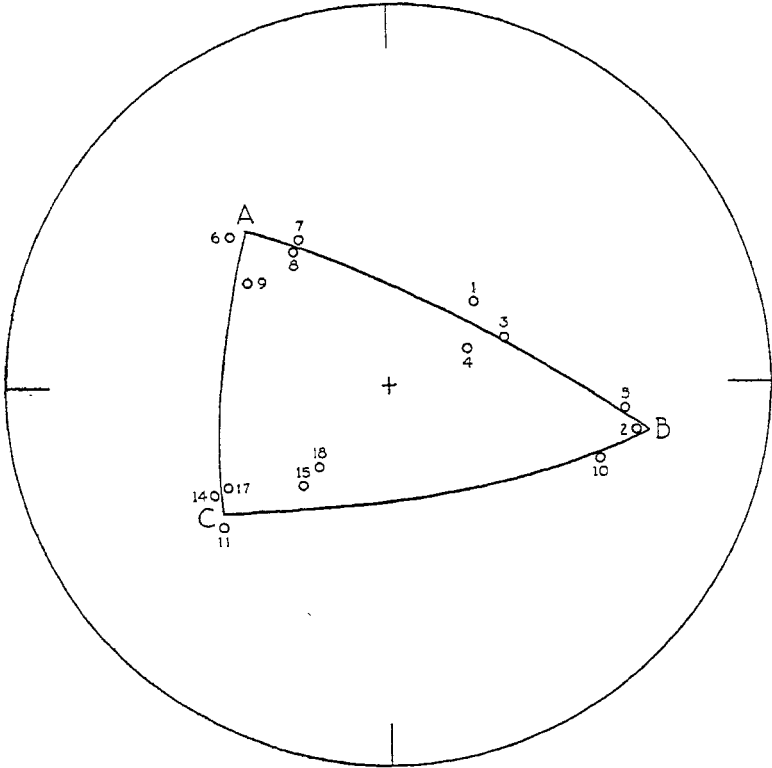


FIGURE 5

Table 3

	$\Lambda_A$	$\Lambda_B$	$\Lambda_C$
I	.304	.441	.244
II	-.154	.893	-.415
III	.940	-.088	-.876

Table 4 shows the matrix  $V$  for this particular problem. It can be seen that those test vectors which lie close to one of the three reference planes  $L_p$  in Figure 5 are also those which are represented with nearly vanishing entries in the corresponding column  $p$  of Table 4.

The three primary abilities are defined by the intersections of reference planes. These are shown at the points  $A$ ,  $B$ , and  $C$ . The direction cosines of the primary ability vectors  $T_p$  may be obtained either by the intersection of pairs of planes  $L_p$  or, more formally, by the matrix (4). The matrix  $T$  for the present example is shown in *Table 5*.

The matrix of *Table 4* shows simple structure, but it is not a unique configuration with the reference planes. The reason is that there are no test

Table 4

	$r_{j\Delta C}$	$r_{j\Delta B}$	$r_{j\Delta A}$
	I	II	III
10	.104	.692	-.014
2	.013	.709	.015
5	-.014	.652	.063
3	-.001	.510	.268
4	.045	.393	.229
1	-.036	.430	.331
8	.008	.086	.378
7	-.007	.096	.411
9	.083	.028	.352
6	.024	-.028	.481
15	.278	.121	.038
14	.332	-.008	.026
17	.447	.019	.052
11	.514	.008	-.026
18	.228	.126	.052

Table 5

	$T_A$	$T_B$	$T_C$
I	.834	.722	.829
II	-.372	.681	-.443
III	.408	-.122	-.341

vectors along the arc  $AC$  or along the arc  $BC$ . The five test vectors near  $C$  may be regarded as identical except for experimental errors. It is for this reason that the primary traits  $A$ ,  $B$ , and  $C$  cannot be inferred with certainty. Simple structure would be obtained as well by drawing a reference plane through the centroid of (1, 3, 4) and (15, 18) instead of the plane  $BC$ . The simple structure would involve this plane and the planes  $AC$  and  $AB$ . The psychological interpretation would be difficult because the tests at  $B$  would have negative factor loadings for the trait  $A$ . Hence the structure shown in *Figure 5* is the more probable one, though it cannot be demonstrated with certainty. The simple structure of *Figure 5* can be shown to be unique

only by a more extensive test battery which includes tests along  $AC$  and along  $BC$ .

With the reservations just written, it is of some interest to note the primary abilities  $A$ ,  $B$ , and  $C$ , and to consider tentatively their psychological nature. Since  $B$  consists of opposites tests, one might postulate a verbal factor. The tests at  $A$  are numerical, so that a number factor might also be postulated. Inspection of tests 1, 3, 4, shows them to lie in the plane  $AB$ . This seems reasonable since they are verbal in character; but they also contain some of the precision and restrictiveness of numerical work. This relation raises the psychological question whether the number factor is essentially concerned with number as such or with some kind of facility for logical or other restrictive thinking of which numerical work is only a good example. This is a question of fact which can be established by experimental inquiry with larger test batteries. The factor  $C$  is evidently concerned with visual imagery and perhaps with kinesthesia. The battery does not contain enough tests to establish their separation if they are separate abilities.

## CHAPTER VII

### ISOLATION OF PRIMARY FACTORS

#### Method of oblique axes

This method was devised for testing the hypothesis that a specified trait is primary. For example, if the hypothesis is entertained that a space factor is primary in the fifteen tests of Brigham, the method of oblique axes makes it possible to test this hypothesis directly. The method is general and in no sense limited to psychological tests which are used for illustrative purposes.

Let the trait which is postulated as primary be denoted  $T_p$ . If the hypothesis is clearly formulated, it should be possible to describe  $T_p$  in terms of other traits in which it is involved and in terms of still other traits in which the supposed primary component  $T_p$  is absent. If a battery of  $n$  traits has been found to involve  $r$  factors and if the postulated trait  $T_p$  is primary in this battery, then there should be only  $(r-1)$  factors in the residual battery which is obtained by merely eliminating those traits in which  $T_p$  can be involved. This idea can be illustrated with a postulated space factor as an example. If the fifteen tests of the battery contain several tests that involve space thinking and if the whole battery is well described by three factors, then the residual battery which is obtained by eliminating the space tests should be describable in terms of only two factors.

Each trait in the residual battery is described in terms of  $r$  factors in the factorial matrix  $F$ . It does not matter for the present problem that the reference axes of  $F$  are arbitrary.

After eliminating the traits that may conceivably involve the postulated primary trait  $T_p$ , let there remain  $n_0$  traits in which  $T_p$  is almost certainly absent. When these rows of  $F$  have been eliminated, there remains a factorial matrix  $F_0$  of  $n_0$  rows and  $r$  columns. The rank of the reduced factorial matrix  $F_0$  must be  $(r-1)$  if  $T_p$  is primary in  $F$ .

The trait configuration of the  $n_0$  traits involves, therefore,  $(r-1)$  dimensions; but each one of them is described in terms of  $r$  co-ordinates in  $F_0$ . If the  $r$  principal axes are determined for the matrix  $F_0$  by the methods of *chapter iv*, one of the roots  $\beta$  must vanish, because the extension of the trait configuration is vanishing in the  $r$ th dimension. This situation was anticipated in the first numerical example of *chapter iv*, where the system was intentionally devised so that one of the roots  $\beta$  did vanish. If  $n_1$  roots  $\beta$  vanish in the characteristic equation, then the trait configuration is of dimensionality  $(r-n_1)$ , and this is also the number of primary traits in the battery.

The verification of  $T_p$  as a primary trait is accomplished by the following procedures:

1) Eliminate from the factorial matrix  $F$  those traits which may conceivably involve the postulated primary trait  $T_p$ . This gives the reduced factorial matrix  $F_0$  of  $n_0$  rows and  $r$  columns.

2) Make sure that  $F_0$  satisfies the inequality (5-ii), so that the trait configuration is unique.

3) Determine the  $r$  roots  $\beta$  of the characteristic equation for  $F_0$ .

4) If only one root  $\beta$  vanishes, then one, and only one, primary factor was removed from  $F$  by its reduction to  $F_0$ . The trait  $T_p$  is therefore primary.

5) If  $n_1$  roots  $\beta$  vanish, then  $n_1$  primary factors were removed from  $F$  in its reduction to  $F_0$ , and the hypothesis is obscured. In supposedly removing one primary factor, several primary factors were removed. The traits must be re-examined in order to ascertain whether additional, but yet unformulated, primary factors were inadvertently removed from  $F$  together with  $T_p$ , or whether the trait  $T_p$  is itself a trait of multiple complexity  $n_1$ .

6) If none of the roots  $\beta$  vanish, then the hypothesis is disproved, because the residual factorial matrix  $F_0$  is of the same rank as the original matrix  $F$ . No primary factor has been removed from  $F$ . The problem then calls for another guess about the nature of the primary factors.

If this process is repeated for  $r$  successive postulated traits and if these are verified in the same manner, then the result will be a set of  $r$  primary trait vectors in terms of which the oblique factorial matrix  $V$  may be written. The present method is called a "method of oblique axes" because the co-ordinate axes of  $V$  are not necessarily orthogonal. The angular separations of the primary trait vectors of  $V$  are functions of the intercorrelations of the primary traits in the experimental population. These intercorrelations are affected by the fortuitous conditions that vary more or less uncontrollably from one sample population to another. It is one of the fundamental problems of factorial analysis to transcend these fortuitous conditions that characterize random samples. As long as the discovery of the fundamental categories of a science is markedly affected by the fortuitous elements of random sampling, the categories are not likely to be significant.

### Numerical example of method of oblique axes

The method will be illustrated on the battery of fifteen tests that were used in *chapter iii* as a numerical example. *Table (25-iii)* shows the factor loadings for the fifteen tests. Since the first three factors account for the intercorrelations within small residuals, only the first three columns of the factorial matrix will be used. This will facilitate a comparison of the results

of the present example with the graphical methods previously used for the same problem. Inspection of the tests suggests that a visual or space factor is present in some of them. The tests which are most conspicuously spatial in character are 11, 14, 15, 17, 18. After eliminating these five tests, there

Table 1

Tests		I	II	III
10	Opposites	.642	.443	-.150
2	Opposites	.579	.499	-.090
5	Opposites	.561	.449	-.041
3	Analogies	.712	.228	.092
4	Artificial language	.633	.134	.061
1	Definitions	.685	.159	.157
8	Geometrical completion	.529	-.144	.207
7	Arithmetic problems	.559	-.146	.233
9	Arithmetical proportions	.546	-.222	.162
6	Number series	.585	-.293	.274

remain ten tests for the residual factorial matrix  $F_0$ . This is reproduced for the first three centroid factors in *Table 1*. Although each of these ten tests is here described by three co-ordinates, the rank of the matrix should be only 2. Hence one of the roots  $\beta$  should vanish.

Table 2

	1	2	3
1	3.671647 + $\beta$	.730882	.528743
2	.730882	.919257 + $\beta$	-.255728
3	.528743	-.255728	.267573 + $\beta$

In *Table 2* the cross products are summarized. This table shows the coefficients of the simultaneous equations (13-iv). The three roots  $\beta$  of the characteristic equation are as follows:

$$\beta_1 = -3.910283, \quad \beta_2 = -.930210, \quad \beta_3 = -.017984.$$

It is seen that one of these roots is almost zero. Since this root is the sum of the squares of the projections of the test vectors on the minor principal axis, it is seen that the mean squared projection for the ten tests is .0018.

The direction cosines of the vector which is determined by substituting  $\beta_3$  in (13-iv) are as follows:

$$\lambda_{13} = +.211980, \quad \lambda_{23} = -.422010, \quad \lambda_{33} = -.881460.$$

In *Table 3* are listed the projections of the fifteen test vectors on the vector  $\Lambda_3$ ; and it is of special interest to note that the ten tests of the residual battery have nearly vanishing projections on  $\Lambda_3$ , whereas the five tests which were postulated to contain a space factor have marked positive projections on  $\Lambda_3$ . This result proves that a primary factor was removed from the test battery when the five space tests were eliminated.

*Table 3*

Tests		$r_{j\Lambda_3}$
10	Opposites	.081
2	Opposites	-.009
5	Opposites	-.034
3	Analogies	-.026
4	Artificial language	.024
1	Definitions	-.060
8	Geometrical completion	-.009
7	Arithmetic problems	-.025
9	Arithmetical proportions	.067
6	Number series	.006
15	Card turning	.264
14	Block construction	.321
17	Dice counting	.430
11	Painted cubes	.498
18	Form learning	.215

### Constellations

In formulating hypotheses concerning the nature of the primary traits, it is sometimes a considerable aid to know of constellations that may exist in the trait configuration. By a "constellation" is meant a grouping of trait vectors. It happens not infrequently that the trait configuration consists essentially in groups of trait vectors. The angular separations between the trait vectors within a constellation are relatively small, while the separations between constellations are marked.

When the dimensionality of the factorial matrix is less than four, the constellations may be inspected readily by graphical methods. When the dimensionality exceeds three, the graphical methods are not available, and it is then useful to have a routine by which the constellations may be isolated in the trait configuration. Since the constellations are to be used as an aid to intuition regarding the nature of the primary traits, it is not advisable to define a constellation rigorously as regards maximum angular separations or as regards the maximum generating angle of the cone which shall include a constellation. Such restrictions may be arbitrarily imposed by the investigator for each study in terms of the dimensionality of  $F$  and the mean order of magnitude of the communalities that are involved.

If an attempt is made to isolate constellations from a large battery of traits, say fifty or more, without some systematic procedure, it is usually found that the groupings become entangled in annoying complexity. If the constellations do not exist, the procedure must make this fact evident; but, on the other hand, constellations can be drawn for the purposes of studying the battery even though the traits arrange themselves more in the nature of chains than constellations. In three dimensions this situation is illustrated by a battery of traits whose configuration reveals a spherical triangle in which the sides of the triangle are pretty well defined by the trait vectors. If all of them lie in sides of a spherical triangle, then the isolation of constellations would be difficult, because there may be no sharp break between one constellation and the next. In three dimensions the graphical methods would, of course, be used because of their simplicity and directness; but in higher dimensions the groupings may be obtained by inspectional methods from the intercorrelations corrected for uniqueness.

One useful procedure is to ascertain first the average correlation in each column of the correlational matrix  $R_u$  where the given coefficients have been corrected for uniqueness. (An alternative is to count in each column the number of coefficients whose absolute values exceed, say, .80 or .90.) Select the trait  $T_x$  with highest mean coefficient. List all the traits whose correlations with  $T_x$  exceed .80 and complete the correlation table for the traits so selected. Eliminate from the table the trait which has the largest number of intercorrelations less than .80. Repeat the eliminating process until all the traits that remain in the table have intercorrelations that exceed .80. These traits constitute a constellation. Select the trait whose mean coefficient is next highest and which is not listed in the group just formed, and proceed with it in the same manner as with  $T_x$  until the majority of the traits have been assigned. These groupings are flexible, and they may be arranged to overlap. The arrangement of the traits in constellations should be regarded merely as a device for studying them and for formulating hypotheses concerning the underlying primary factors.

### The method of averages

This method is a modification of the method of oblique axes. Though not theoretically so satisfactory, it is useful, since it gives results which are closely similar to those of the method of oblique axes; and it is shorter in computational work, in that it does not require the determination of the roots of the characteristic equation.

When the traits which may contain the postulated primary factor have been eliminated from  $F$ , there remain  $n_0$  traits in the reduced factorial matrix  $F_0$ . These are arranged in ascending or descending order of magnitude, according to the co-ordinates in one of the columns of  $F_0$  which shows a con-



siderable range in numerical values. The second column is usually one of the best if  $F$  has been computed by the centroid method. The  $n_0$  traits of  $F_0$  are then divided into  $(r-1)$  groups. The co-ordinates of the centroid of each group are then determined. These define  $(r-1)$  points in the common-factor space. It is desired to find the hyperplane  $L$  which contains the  $n_0$  traits in  $F_0$ . In the method of averages the hyperplane  $L$  is taken to be that hyperplane which contains the vectors whose termini are the centroids of the  $(r-1)$  groups of traits. The projection of each of the  $n_0$  traits on the normal to the hyperplane  $L$  is then determined. These should all vanish. If  $\Lambda$  is a principal axis of the system, or if  $\Lambda$  is near one of the principal axes, then the sum of the squares of the projections of the  $n_0$  trait vectors on  $\Lambda$  will be nearly equal to that root  $\beta$  of the characteristic equation which is zero or nearly vanishing. On the other hand, the projections of the  $(n-n_0)$  eliminated traits should have appreciable projections on  $\Lambda$  in order to establish that a primary factor was removed in reducing  $F$  to  $F_0$ .

#### Numerical example of the method of averages

The same set of fifteen psychological tests will be used as an example. The ten tests in  $F_0$  may be divided according to the signs of the second column of  $F_0$  into  $(r-1)=2$  groups as follows:  $A = (10, 2, 5, 3, 4, 1)$  and  $B = (8, 7, 9, 6)$ . The centroids of these two groups of points are as follows:

	I	II	III
$A$	.6353	.3187	.0048
$B$	.5548	-.2012	.2190

Let the co-ordinates of the centroid of group  $A$  be  $A_1, A_2, A_3$  and let the corresponding coordinates for group  $B$  be  $B_1, B_2, B_3$ . Then if the two vectors  $A$  and  $B$  are to be contained in the plane  $L$ , it is necessary that both  $A$  and  $B$  have vanishing projections on the normal  $\Lambda$  to the plane  $L$ . Hence

$$A_1\lambda_1 + A_2\lambda_2 + A_3\lambda_3 = 0,$$

$$B_1\lambda_1 + B_2\lambda_2 + B_3\lambda_3 = 0.$$

Expressing all of the  $\lambda$ 's in terms of one of them, and normalizing, we have the values listed in *Table 4*. It is interesting to note that these values are nearly the same as those which were found by the method of oblique axes. These are shown in the first column of *Table 4*. In *Table 5* are shown the projections of the fifteen test vectors on  $\Lambda$  as defined in the second column of *Table 4*. Note that the projections of the ten tests of  $F_0$  almost vanish, while the projections of the five space tests which were removed from  $F$  are

appreciable in magnitude. These results could, of course, be predicted by inspection of the sphere on which the fifteen normalized test vectors were plotted.

Table 4

	Method of Oblique Axes	Method of Averages
$\lambda_1$	+ .212	+ .207
$\lambda_2$	- .422	- .400
$\lambda_3$	- .881	- .893

Table 5

	Tests	$r_{j\Delta_3}$
10	Opposites	.090
2	Opposites	.001
5	Opposites	-.027
3	Analogies	-.026
4	Artificial language	.023
1	Definitions	-.062
8	Geometrical completion	-.018
7	Arithmetic problems	-.034
9	Arithmetical proportions	.057
6	Number series	-.006
15	Card turning	.261
14	Block construction	.316
17	Dice counting	.423
11	Painted cubes	.492
18	Form learning	.212

### The special case of rank 2

When the matrix  $F$  has been obtained, the communalities are known, so that the matrix  $F_u$  can be written in which the cell entries show the direction cosines of the augmented or unit trait vectors. The cross products  $F_u F'_u = R_u$ , in which  $R_u$  contains the intercorrelations corrected for uniqueness.

Let some of the traits be of complexity 2, so that their intercorrelations may be described linearly in terms of two factors. Let a new matrix  $\Phi$  be formed whose entries  $\phi_{jk}$  are determined by the relation

$$(1) \quad \phi_{jk} = \cos^{-1} R_{jk},$$

where

$$R_{jk} = \frac{r_{jk}}{h_j h_k}.$$

Select any two columns of  $\Phi$ , say  $l$  and  $m$ . Then, if the other trait vectors lie in the plane of the two trait vectors  $l$  and  $m$ , we have

$$(2) \quad \phi_{jl} = \phi_{jm} + \phi_{ml}.$$

If the column  $l$  of  $\Phi$  is plotted against the column  $m$ , the plot should be linear with a slope of unity. The  $l$ -intercept is the angular separation  $\phi_{ml}$  between the trait vectors  $m$  and  $l$ . In this graphical procedure it becomes immediately evident which of the traits in  $R_0$  are coplanar, or nearly coplanar, with the pair of traits  $l$  and  $m$ . All the traits  $j$  which are represented in the graph on a line through the  $l$ -intercept of  $\phi_{lm} = \cos^{-1} R_{lm}$  with a slope of unity can be described linearly in terms of two factors. It is of interest to note that a constant can be added to each column of the matrix so that all columns become proportional. The rank is then reduced to 1.

### Projections of unit trait vectors into a hyperellipsoid

Let  $a_{jk}$  represent each element in a matrix  $Q$  of order  $n \times s$  of rank  $r$ , and let the matrix be normalized by rows so that

$$(3) \quad \sum_{k=1}^s a_{jk}^2 = 1.$$

Then each of the  $s$  cell entries  $a_{jk}$  in each row  $j$  can be expressed linearly in terms of  $r$  independent cell entries where  $r \leq s$ . There is no loss of generality in rearranging the rows and corresponding columns so that the  $r$  independent columns become the first  $r$  columns of the matrix. Then we have

$$(4) \quad a_{jk} = A_{1k}a_{j1} + A_{2k}a_{j2} + \dots + A_{rk}a_{jr}.$$

This linear description of  $a_{jk}$  can be condensed into the form

$$(5) \quad a_{jk} = \sum_{m=1}^r A_{mk}a_{jm}.$$

Squaring, we have

$$(6) \quad a_{jk}^2 = A_{1k}^2 a_{j1}^2 + A_{2k}^2 a_{j2}^2 + \dots + A_{rk}^2 a_{jr}^2 + 2A_{1k}A_{2k}a_{j1}a_{j2} \\ + \dots + 2A_{(r-1)k}A_{rk}a_{j(r-1)}a_{jr}.$$

This equation can be condensed as follows:

$$(7) \quad a_{jk}^2 = \sum_{p=1}^r \sum_{m=1}^r A_{mk}A_{pk}a_{jm}a_{jp}.$$

By (3) and (7) we have

$$(8) \quad \sum_{k=1}^s a_{jk}^2 = \sum_{k=1}^s \sum_{p=1}^r \sum_{m=1}^r A_{mk} A_{pk} a_{jm} a_{jp} = 1.$$

Equation (8) can be regarded as an equation in the  $r$  unknowns  $a_{jm}$  where  $m=1, 2, \dots, r$ . Equation (8) represents a hyperellipsoid in the space of  $r$  dimensions which is defined by the  $r$  columns of the given  $n \times s$  matrix. The hyperellipsoid of (8) contains the  $n$  points represented by the  $n$  rows of  $Q$ . Hence we have the following:

*Theorem: If a matrix  $Q$  of order  $n \times s$  is of rank  $r$  and if it has been normalized by rows, then any set of  $r$  independent columns of  $Q$  define the  $r$  Cartesian co-ordinates of each of  $n$  points which lie in the surface of a hyperellipsoid in a space of  $r$  dimensions with center at origin.*

Specializing this theorem to rank 2, we have

*Theorem: If a matrix  $Q$  of order  $n \times s$  is of rank 2 and if it has been normalized by rows, then any pair of independent columns determines the Cartesian co-ordinates of each of  $n$  points on an ellipse with center at origin.*

One example of this theorem will be shown for rank 2. *Table (10-iii)* shows the intercorrelations of eight fictitious variables whose correlation matrix is

Table 6

	1	2	3	4	5	6	7	8	$\Sigma^2$
1	.458079	.400819	.114520	.171780	.515338	.458079	.286299	.171780	1.000000
2	.336733	.390851	.228497	.294641	.402877	.432942	.378824	.318694	1.000000
3	.138219	.328270	.345547	.414657	.207328	.345547	.449212	.466489	1.000000
4	.166306	.339542	.332612	.401907	.235600	.360330	.443483	.450412	1.000000
5	.431563	.401594	.143854	.203794	.491503	.455539	.311685	.209788	1.000000
6	.349565	.393261	.218478	.284022	.415108	.436956	.371413	.305869	1.000000
7	.230121	.362440	.299157	.368193	.299157	.391205	.425723	.408464	1.000000
8	.151015	.333492	.339784	.408999	.220230	.352369	.446753	.459338	1.000000

of rank 2. When the table is normalized by rows, the cell entries take the values shown in *Table 6*. The first two columns are plotted in *Figure 1*, and it is seen that the points determine an ellipse with center at origin.

**A method of maximizing the number of zero factor loadings (the single hyperplane method)**

When the factorial matrix  $F$  has been obtained, it is desirable to be able to extract the primary abilities by methods that are not dependent on any hypotheses concerning their nature. The primary abilities have been de-

fined as those factors which reduce to a minimum the number of factors per test that will account for the intercorrelations. Since the primary abilities are likely to be positively correlated in all readily available experimental

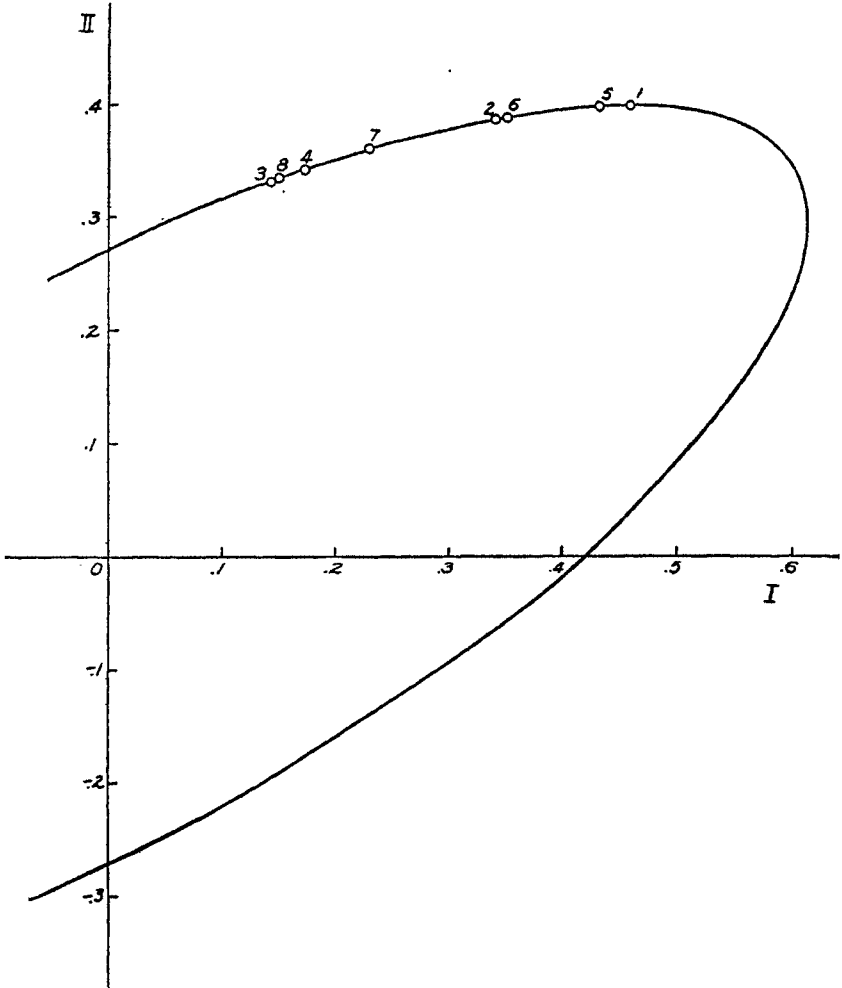


FIGURE 1

groups of subjects, the methods of isolating the abilities must be free from the restrictions of orthogonality. The simplest underlying structure is indicated by a transformation  $G$  of  $F$  by which the number of vanishing entries in  $V$  is maximized. A large number of zero entries in each column  $p$  of  $V$  constitutes assurance of an underlying order among the variables whereby

each one of them can be described by a number of scientific categories that is smaller than the rank of the correlational matrix. In general, such a transformation does not exist for a factorial matrix that is produced with arbitrary cell entries. Geometrically, this problem can be described as an attempt to discover a set of  $r$  hyperplanes so defined that each of the  $n$  test vectors lies in one or more of the hyperplanes. Since each hyperplane is of  $(r-1)$  dimensions, it is clear that one primary ability is absent from all test vectors in each hyperplane. If every test vector is so contained, it follows that there will be at least one vanishing entry in each row  $j$  of  $V$ .

In the extraction of primary abilities, each of the  $r$  hyperplanes will be sought in succession. This is feasible, since there are no conditions governing the relations between the hyperplanes or between the primary trait vectors except that they be distinct. The method to be described consists in finding a hyperplane that will contain as many as possible of the test vectors. The fact that a test is contained in a hyperplane can also be regarded as a zero correlation between the test and the normal to the hyperplane. This normal can be thought of as an imaginary test. It is desired, then, to find a vector  $\Lambda_p$  in the common-factor space with which the maximum number of tests have zero correlation or for which the number of zero correlations is larger than for any neighboring vector.

Since the correlation coefficient is a continuous function of the angular separation of the test vector and the reference vector  $\Lambda_p$ , it is desirable to maximize, not the absolute number of tests whose correlations with  $\Lambda_p$  vanish exactly, but rather some function of this correlation that has a large value when the correlation is near zero and which has insignificant values when the correlation becomes appreciable. This should be a function of the square of the correlation in order that the function be symmetric.

There is a very large number of functions of  $r_{j\Lambda}^2$  which can be used for the present purpose, but many of them must be eliminated for statistical or computational reasons. Let  $w$  represent the function of  $r_{j\Lambda}^2$ , and let

$$(9) \quad u \equiv \sum_{j=1}^n w_{jp}$$

be the function which is to be maximized in order to insure a large number of vanishing entries in a column  $p$  of  $V$ .

One of the simplest functions that satisfy the demands of this problem is

$$(10) \quad w'_{jp} = \frac{1}{r_{j\Lambda}^2};$$

but this function has the statistical limitation that when the correlation vanishes,  $w$  becomes infinite, so that it cannot be handled computationally.

This defect can be remedied by modifying the function to

$$(11) \quad w_{ip} \equiv \frac{1}{r_{iA}^2 + c},$$

where  $c$  is some small arbitrary constant such as  $+.01$ . The maximum value of  $w$  is then 100 when the correlation vanishes, and it is nearly unity when the correlation is unity.

It seems certain that better functions and simpler computational methods will be found than those which are to be described here. The excuse for presenting a method which is not yet the simplest is that it does locate each column of  $V$  in which the number of zero entries is maximized. While this method is useful, it cannot be applied automatically because hyperplanes may exist in a particular system which contain groups of traits but which do not define the most appropriate scientific categories.

If it is postulated that all of the entries in  $V$  shall be positive or zero and that negative factor loadings are to be excluded, then the precaution must be taken to carry the factorial matrix  $F$  to a sufficient number of factors. If this is not done, negative factor loadings will appear in  $V$  even though the tests can be described in the positive region of a common-factor space in higher dimensions than those which are assumed in  $F$ . This is largely a question of judgment as to when the residuals are small enough to be ignored. There should be no harm in carrying the factorial matrix to a number of columns larger than needed.

Since  $r_{iA} \equiv v_{ip}$ , we have

$$(12) \quad w_{ip} = \frac{1}{v_{ip}^2 + c}.$$

It is desired to find the vector  $\Lambda_p$  for which  $u$  is maximized. The vector  $\Lambda_p$  is defined in terms of its direction cosines, which may be denoted  $\lambda_{1p}$ ,  $\lambda_{2p}$ , . . . ,  $\lambda_{rp}$ . Then

$$(13) \quad r_{iA} = v_{ip} = a_{i1}\lambda_{1p} + a_{i2}\lambda_{2p} + \dots + a_{ir}\lambda_{rp}.$$

The unknown parameters are the direction cosines of  $\Lambda_p$ , while the values of  $a_{im}$  are given in the matrix  $F$ . The normal equations would be of the form

$$(14) \quad \frac{\partial u}{\partial \lambda_{mp}} + \beta_p \frac{\partial z_p}{\partial \lambda_{mp}} = 0$$

with the conditional equation

$$(15) \quad z_p = \sum_{m=1}^r \lambda_{mp}^2 - 1 = 0,$$

where  $\beta_p$  is a Lagrange multiplier.\* These  $r$  normal equations are evidently non-linear and awkward to solve.

Since the direct solution of the  $r$  simultaneous equations (14) is not feasible, the solution will be reached by starting with a trial vector. This vector will be adjusted until the normal equations are satisfied. Let  $\Lambda_p$  be an arbitrary trial vector. Then the  $n$  values of  $v_{jp}$  in (13) may be determined for the trial vector. It is desired to maximize  $u$ . One of the parameters may be expressed in terms of the remaining ones by the conditional equation (15), so that the solution may be found in terms of  $(r-1)$  independent parameters. Since the numbering of the orthogonal reference vectors is arbitrary, let  $\lambda_{1p}$  be expressed in terms of the remaining  $(r-1)$  direction cosines of  $\Lambda_p$ . By (15) we have

$$(16) \quad z_p = \lambda_{1p}^2 + \sum_{m=2}^r \lambda_{mp}^2 - 1 = 0,$$

so that

$$(17) \quad \lambda_{1p} = \left[ 1 - \sum_{m=2}^r \lambda_{mp}^2 \right]^{1/2}.$$

The first partial derivative of  $v_{jp}$  with respect to the independent parameter  $\lambda_{mp}$  is then

$$(18) \quad \frac{\partial v_{jp}}{\partial \lambda_{mp}} = -a_{j1}g_{mp} + a_{jm},$$

where

$$g_{mp} \equiv \frac{\lambda_{mp}}{\lambda_{1p}}.$$

Also,

$$(19) \quad \frac{\partial w_{jp}}{\partial \lambda_{mp}} = \frac{-2v_{jp} \frac{\partial v_{jp}}{\partial \lambda_{mp}}}{(v_{jp}^2 + c)^2}.$$

\* An alternative method of successive approximation is to treat (14) as  $r$  linear equations in  $\beta_p$  for  $r$  trial values of  $\lambda_{mp}$  with residuals  $\rho_{mp}$ . For each approximation  $\lambda_{mp}$  a new trial value  $\beta_p$  is determined so as to minimize the  $r$  residuals  $\rho_{mp}$ . These vanish for those values of  $\lambda_{mp}$  which maximize the function  $u$ .



From (18) and (19) we have

$$(20) \quad \frac{\partial w_{ip}}{\partial \lambda_{mp}} = \frac{-2v_{ip}(-a_{i1}g_{mp} + a_{im})}{(v^2_{ip} + c)^2}.$$

Let

$$(21) \quad y_{ip} \equiv \frac{2v_{ip}}{(v^2_{ip} + c)^2}.$$

Then

$$(22) \quad \frac{\partial w_{ip}}{\partial \lambda_{mp}} = -y_{ip}(-a_{i1}g_{mp} + a_{im}).$$

Since  $u$  is the function that is to be maximized, its partial derivatives with respect to the  $(r-1)$  independent parameters must be found. These derivatives are in the form

$$(23) \quad \frac{\partial u}{\partial \lambda_{mp}} = \frac{\partial}{\partial \lambda_{mp}} \sum_{j=1}^n w_{ip} = \sum_{j=1}^n \frac{\partial w_{ip}}{\partial \lambda_{mp}}.$$

Summing (22) for all tests  $j$ , we have, by (23),

$$(24) \quad \frac{\partial u}{\partial \lambda_{mp}} = g_{mp} \sum_{j=1}^n a_{j1} y_{ip} - \sum_{j=1}^n a_{jm} y_{ip}.$$

The numerical values of the  $(r-1)$  derivatives of (24) may be determined for the trial vector  $\Delta_p$ . Let these derivatives be denoted

$$(25) \quad \rho_{mp} \equiv \frac{\partial u}{\partial \lambda_{mp}}.$$

Since the derivatives (24) show the rates at which the function  $u$  is increasing at the point  $\Delta_p$  with respect to the  $(r-1)$  independent direction cosines of  $\Delta_p$ , it is clear that the small corrections to  $\lambda_{mp}$  should be proportional to  $\rho_{mp}$ . Let the corrections be denoted

$$(26) \quad \varepsilon_{mp} \equiv k\rho_{mp},$$

where  $k$  is a small arbitrary constant. Then

$$(27) \quad \mu_{mp} \equiv \lambda_{mp} + \varepsilon_{mp},$$

where  $\mu_{mp}$  are proportional to the  $(r-1)$  independent direction cosines of the new trial vector  $M_p$ . It is advisable to choose  $k$  so that none of the corrections  $\varepsilon_{mp}$  exceed .10 or .05.

When the  $(r-1)$  values of  $\mu_{m_p}$  have been computed by (27), the remaining value  $\mu_{1_p}$  is determined by the equation

$$(28) \quad \varepsilon_{1_p} = -\frac{1}{\lambda_{1_p}} \sum_{m=2}^r \lambda_{m_p} \varepsilon_{m_p},$$

which is obtained by taking differentials of (16), noting that  $d\lambda_{m_p} = \varepsilon_{m_p}$ . Normalizing the  $r$  values  $\mu_{m_p}$  gives the direction cosines of the new trial vector  $M_p$ .

This procedure is repeated until the  $(r-1)$  derivatives  $\rho_{m_p}$  all vanish. The vector  $\Delta_p$  for which all the derivatives  $\rho_{m_p}$  vanish gives a stationary point in the surface

$$(29) \quad u = \phi(\lambda_{1_p}, \lambda_{2_p}, \dots, \lambda_{r_p}).$$

It is advisable to choose for  $\lambda_{1_p}$  that direction cosine which has the highest absolute value.

The  $r$  direction cosines of the vector  $\Delta_p$  for which the function  $u$  is stationary constitute one of the columns in the transformation  $G$ . The order of the columns of  $G$  is arbitrary.

It will be found that the corresponding column of the matrix  $V$  shows a large number of vanishing entries if the variables of  $R$  have simple structure. In the case of psychological tests, this is in accordance with the hypothesis that each test in a diversified test battery does not require all of the primary abilities which are required by the battery as a whole. If the number of columns of  $F$  is smaller than the number of primary abilities in the  $n$  tests and if the primary abilities are involved only positively in the tests, then the insufficient number of columns of  $F$  will cause negative entries in  $V$ .

In evaluating the partial derivatives of (24) and in determining the value of  $u$  which is to be maximized, it is convenient to have facilitating tables for  $w$  and for  $y$ . *Table 7* shows the value of  $w$  for each given value of  $v$ . *Table 8* shows the value of  $y$  for each given value of  $v$ . The argument  $v$  is listed in these tables to two decimals.\*

### Analytical method of isolating simple structure.

The equation of a simple structure has been shown to be of the form (7-vi) or of the more condensed form (9-vi). The simple structure is defined by the  $r^2$  parameters  $\lambda_{m_p}$  of (7-vi). The best fitting simple structure for a particular factorial matrix  $F$  may be defined, by the usual statistical con-

\* Special data sheets have been prepared for determining the hyperplanes in which the function  $u$  is maximized. These are available at University of Chicago Bookstore.

Table 7  
Values of w

$c = .01$

$v$	0	1	2	3	4	5	6	7	8	9
.00	100	99	96	92	86	80	74	67	61	55
.10	50	45	41	37	34	31	28	26	24	22
.20	20	18	17	16	15	14	13	12	11	11
.30	10	9	9	8	8	8	7	7	6	6
.40	6	6	5	5	5	5	5	4	4	4
.50	4	4	4	3	3	3	3	3	3	3
.60	3	3	3	2	2	2	2	2	2	2
.70	2	2	2	2	2	2	2	2	2	2
.80	2	2	1	1	1	1	1	1	1	1
.90	1	1	1	1	1	1	1	1	1	1

$c = .10$

$v$	0	1	2	3	4	5	6	7	8	9
.00	10.00	9.99	9.96	9.91	9.84	9.76	9.65	9.53	9.40	9.25
.10	9.09	8.92	8.74	8.55	8.36	8.16	7.96	7.76	7.55	7.35
.20	7.14	6.94	6.74	6.54	6.35	6.15	5.97	5.78	5.61	5.43
.30	5.26	5.10	4.94	4.79	4.64	4.49	4.36	4.22	4.09	3.97
.40	3.85	3.73	3.62	3.51	3.41	3.31	3.21	3.12	3.03	2.94
.50	2.86	2.78	2.70	2.63	2.55	2.48	2.42	2.35	2.29	2.23
.60	2.17	2.12	2.06	2.01	1.96	1.91	1.87	1.82	1.78	1.74
.70	1.69	1.66	1.62	1.58	1.54	1.51	1.48	1.44	1.41	1.38
.80	1.35	1.32	1.29	1.27	1.24	1.22	1.19	1.17	1.14	1.12
.90	1.10	1.08	1.06	1.04	1.02	1.00	0.98	0.96	0.94	0.93

Table 8  
Values of y

$c = .01$

$v$	0	1	2	3	4	5	6	7	8	9
.00	0	196	370	505	595	640	649	631	595	549
.10	500	450	403	359	320	284	252	225	200	179
.20	160	144	129	116	105	95	86	79	72	66
.30	60	55	51	47	43	40	37	34	32	30
.40	28	26	24	23	21	20	19	18	17	16
.50	15	14	13	13	12	11	11	10	10	9
.60	9	8	8	8	7	7	7	6	6	6
.70	6	5	5	5	5	5	4	4	4	4
.80	4	4	4	3	3	3	3	3	3	3
.90	3	3	3	2	2	2	2	2	2	2

$c = .10$

$v$	0	1	2	3	4	5	6	7	8	9
.00	0.00	2.00	3.97	5.89	7.75	9.52	11.18	12.72	14.13	15.40
.10	16.53	17.51	18.34	19.03	19.57	19.99	20.28	20.46	20.54	20.51
.20	20.41	20.23	19.98	19.68	19.33	18.93	18.51	18.06	17.60	17.11
.30	16.62	16.12	15.62	15.12	14.63	14.14	13.66	13.19	12.72	12.27
.40	11.83	11.41	11.00	10.60	10.21	9.84	9.48	9.13	8.79	8.47
.50	8.16	7.87	7.58	7.31	7.04	6.79	6.55	6.31	6.09	5.88
.60	5.67	5.47	5.28	5.10	4.93	4.76	4.60	4.45	4.30	4.16
.70	4.02	3.89	3.77	3.64	3.53	3.42	3.31	3.21	3.11	3.01
.80	2.92	2.83	2.75	2.67	2.59	2.51	2.44	2.37	2.30	2.24
.90	2.17	2.11	2.05	2.00	1.94	1.89	1.84	1.79	1.74	1.70

ventions, as that set of  $r$  co-ordinate hyperplanes  $L_p$  with  $r^2$  parameters for which the function  $\phi$  in (11-vi) is minimized. For each trait  $j$ , let

$$(30) \quad w_j \equiv \prod_{p=1}^r v_{jp}^2,$$

so that

$$(31) \quad \phi = \sum_{j=1}^n w_j.$$

The elements of the oblique factorial matrix  $V$  are

$$(32) \quad v_{jp} = \sum_{m=1}^r a_{jpm} \lambda_{mp}.$$

The  $r$  parameters  $\lambda_{mp}$  for each co-ordinate hyperplane  $L_p$  are subject to the conditional equation

$$(33) \quad z_p = \sum_{m=1}^r \lambda_{mp}^2 - 1 = 0.$$

The form of the normal equations is as follows:

$$(34) \quad \frac{\partial \phi}{\partial \lambda_{mp}} + \beta_p \frac{\partial z_p}{\partial \lambda_{mp}} = 0,$$

where  $\beta_p$  is a Lagrange multiplier for each hyperplane  $L_p$ . The first term of (34) may be written in terms of  $\lambda_{mp}$ , namely,

$$(35) \quad \frac{\partial \phi}{\partial \lambda_{mp}} = \frac{\partial}{\partial \lambda_{mp}} \sum_{j=1}^n w_j = \sum_{j=1}^n \frac{\partial w_j}{\partial \lambda_{mp}}.$$

Substituting (32) in (30), the partial derivatives of (30) are

$$(36) \quad \frac{\partial w_j}{\partial \lambda_{mp}} = 2v_{jp} \frac{\partial v_{jp}}{\partial \lambda_{mp}} \prod_q v_{jq}^2,$$

where  $q$  takes all successive integral values from 1 to  $r$ , except  $p$ . For convenience, let

$$(37) \quad s_{jp} \equiv v_{jp} \prod_q v_{jq}^2.$$

Also,

$$(38) \quad \frac{\partial v_{jp}}{\partial \lambda_{mp}} = a_{jm}.$$

Substituting (37) and (38) in (36),

$$(39) \quad \frac{\partial w_j}{\partial \lambda_{mp}} = 2a_{jm}s_{jp}.$$

By (35) and (39),

$$(40) \quad \frac{\partial \phi}{\partial \lambda_{mp}} = 2 \sum_{j=1}^n a_{jm}s_{jp}.$$

By (33), the derivatives of the second term of (34) can also be written in terms of  $\lambda_{mp}$ , namely,

$$(41) \quad \frac{\partial z_p}{\partial \lambda_{mp}} = 2\lambda_{mp}.$$

Substituting (40) and (41) in (34),

$$(42) \quad 2 \sum_{j=1}^n a_{jm}s_{jp} + 2\beta_p \lambda_{mp} = 0.$$

Dividing by 2 and transposing, (42) becomes

$$(43) \quad \sum_{j=1}^n a_{jm}s_{jp} = -\beta_p \lambda_{mp}.$$

If the  $r^2$  parameters  $\lambda_{mp}$  have been correctly chosen, the  $r$  numerical values of the left member of (43) for each of the co-ordinate hyperplanes are proportional to the  $r$  values of  $\lambda_{mp}$ . If arbitrary trial values are chosen for the  $r^2$  parameters  $\lambda_{mp}$ , then the normalized left members of (43) define a new unit vector  $M_p$  with  $r$  direction cosines  $\mu_{mp}$ . The unit vector  $M_p$  has an angular separation from  $\Lambda_p$  of  $\theta_p$ . If it were desired to maximize the function  $\phi$ , then the  $r$  unit vectors  $M_p$  could probably be used as the new trial reference vectors  $\Lambda_p$ . But it is desired to minimize the function  $\phi$ . Hence each of the  $r$  reference vectors  $\Lambda_p$  may be adjusted in the plane of the angle  $\overline{\Lambda_p O M_p}$  by enlarging the angle  $\theta_p$ . Let the new trial reference vectors be  $N_p$  with direction cosines  $\nu_{mp}$ . These may be used instead of  $\Lambda_p$  in order to reduce the numerical value of the function  $\phi$ . Then

$$(44) \quad \overline{\Lambda_p O M_p} < \overline{N_p O M_p}.$$

By successive approximation, the function  $\phi$  may be reduced in numerical value toward its minimum value  $\phi_0$  at which the  $r^2$  parameters  $\lambda_{mp}$  define a simple structure. If the simple structure is perfect in the sense that each trait vector is contained in one or more of the co-ordinate hyperplanes, then each of the  $n$  values  $w_j$  in (30) vanishes, and hence  $\phi$  also vanishes. In this case the new vectors  $N_p$  become indeterminate, since the left members of (43) vanish.

#### A method of successive approximation for isolating simple structure

Equation (43) states the condition that is to be satisfied when the correct numerical values of the  $r^2$  parameters  $\lambda_{mp}$  have been found. A direct solution is not feasible for  $r^2$  parameters with as many non-linear normal equations. For computing purposes, a method of successive approximation will be described by which the minimum value of the function  $\phi$  may be approached with any required degree of accuracy.

The principle of the method is as follows: An arbitrary set of  $r$  reference vectors  $\Lambda_p$  is chosen for the first trial. It is convenient to choose the  $r$  orthogonal centroid axes for the first trial. Substituting their direction cosines in (43) gives the  $r^2$  initial numerical values of the left members of (43). There will be  $r$  such values for each of the  $r$  hyperplanes  $L_p$ . The  $r$  trial vectors  $\Lambda_p$  are to be adjusted so as to reduce the function  $\phi$  toward its minimum value.

Let the direction cosines of the  $r$  trial vectors  $\Lambda_p$  be denoted  $\lambda_{mp}$ , and let the corrections be denoted  $d\lambda_{mp}$ . The direction cosines of the resultant vectors  $N_p$  are proportional to

$$(45) \quad \nu_{mp} \equiv \lambda_{mp} + d\lambda_{mp} .$$

When the vectors  $N_p$  are normalized, they are reduced to unit vectors  $M_p$  with direction cosines

$$(46) \quad \mu_{mp} = m\nu_{mp} .$$

The vectors  $M_p$  are the new trial vectors.

It is desired to choose small corrections so that

$$(47) \quad \phi(\mu_{mp}) < \phi(\lambda_{mp}) .$$

By successive approximation the function  $\phi$  is to be minimized, subject to the conditional equation (33). The new trial vectors  $M_p$  have direction cosines

$$(48) \quad \mu_{mp} = m(\lambda_{mp} + d\lambda_{mp}) .$$

The inequality (47) will be satisfied if

$$(49) \quad d\phi(\lambda_{mp}) < 0 ,$$

subject to the conditional equation

$$(50) \quad dz_p = 0 .$$

By (31)

$$(51) \quad d\phi = \sum_{j=1}^n dw_j .$$

For convenience, let

$$(52) \quad t_{jp} \equiv \prod_q v_{jq}^2 ,$$

where  $q$  takes all integral values from 1 to  $r$ , except  $p$ . Then

$$(53) \quad dw_j = 2 \sum_{p=1}^r \sum_{m=1}^r v_{jp} t_{jp} a_{jm} d\lambda_{mp} .$$

But

$$(54) \quad s_{jp} = v_{jp} t_{jp} .$$

Hence

$$(55) \quad dw_j = 2 \sum_{p=1}^r \sum_{m=1}^r s_{jp} a_{jm} d\lambda_{mp} ,$$

and by (51)

$$(56) \quad d\phi = 2 \sum_{j=1}^n \sum_{p=1}^r \sum_{m=1}^r s_{jp} a_{jm} d\lambda_{mp} .$$

The conditional equation (50) can be expressed in terms of  $\lambda_{mp}$  by (33).

Then

$$(57) \quad dz_p = \sum_{m=1}^r \lambda_{mp} d\lambda_{mp} = 0 .$$

Let  $\lambda_{rp} \neq 0$ . Then

$$(58) \quad d\lambda_{rp} = -\frac{1}{\lambda_{rp}} \sum_{m=1}^{r-1} \lambda_{mp} d\lambda_{mp} .$$

Substituting (58) in (55),

$$(59) \quad \frac{1}{2}dw_j = \sum_{p=1}^r \sum_{m=1}^r \left[ a_{jm}S_{jp} - a_{jr}S_{jp} \frac{\lambda_{mp}}{\lambda_{rp}} \right] d\lambda_{mp},$$

and hence

$$(60) \quad \frac{1}{2}d\phi = \sum_{p=1}^r \sum_{m=1}^r \left[ \sum_{j=1}^n a_{jm}S_{jp} - \frac{\lambda_{mp}}{\lambda_{rp}} \sum_{j=1}^n a_{jr}S_{jp} \right] d\lambda_{mp}.$$

Here the differential  $d\phi$  is expressed as a sum of  $r^2$  terms, namely,  $r$  terms for each of the  $r$  hyperplanes. Each of these terms is of the form

$$\left[ \sum_{j=1}^n a_{jm}S_{jp} - \frac{\lambda_{mp}}{\lambda_{rp}} \sum_{j=1}^n a_{jr}S_{jp} \right] d\lambda_{mp}.$$

The  $r$  terms in which  $m=r$  vanish identically. The two conditions (49) and (50) are satisfied if each  $d\lambda_{mp}$  is so chosen that each of the non-vanishing terms of (60) is made negative. This can be accomplished if each of the corrections  $d\lambda_{mp}$  (excepting  $d\lambda_{rp}$ ) is taken with sign opposite to

$$\left[ \sum_{j=1}^n a_{jm}S_{jp} - \frac{\lambda_{mp}}{\lambda_{rp}} \sum_{j=1}^n a_{jr}S_{jp} \right].$$

When the corrections for  $(r-1)$  direction cosines of each of the  $r$  hyperplanes have been chosen (excepting  $d\lambda_{rp}$ ), the remaining  $r$ th correction for each hyperplane is determined by (58). In the method of successive approximation which is to be described, each correction  $d\lambda_{mp}$  will be taken proportional to the corresponding term in (56) with reversed sign.

The direction cosines of the new  $r$  trial vectors  $M_p$  may be determined in the following manner. Let

$$(61) \quad c_{mp} \equiv \frac{\lambda_{mp}}{\lambda_{rp}} \sum_{j=1}^n a_{jr}S_{jp} - \sum_{j=1}^n a_{jm}S_{jp},$$

where  $m \neq r$ . The value of  $c_{rp}$  is determined by the relation

$$(62) \quad c_{rp} = - \frac{1}{\lambda_{rp}} \sum_{m=1}^{r-1} \lambda_{mp} c_{mp}.$$

The corrections  $d\lambda_{mp}$  will be taken proportional to  $c_{mp}$ .



Let the  $r$  direction cosines of each of  $r$  vectors  $N_p$  be proportional to

$$(63) \quad v_{mp} = \lambda_{mp} + k_p c_{mp},$$

where  $k$  is so chosen that the maximum value of any one of the  $r^2$  terms  $k_p c_{mp}$  is equal to some assigned value such as .10 or .05. The new trial vectors  $M_p$  have the direction cosines

$$(64) \quad \mu_{mp} = m v_{mp},$$

where the constant  $m$  is so chosen that  $M_p$  is a unit vector. Hence

$$(65) \quad \sum_{m=1}^r \mu_{mp}^2 - 1 = 0.$$

Then by (64),

$$(66) \quad \sum_{m=1}^r \mu_{mp}^2 = m^2 \sum_{m=1}^r v_{mp}^2,$$

or

$$(67) \quad m = \frac{1}{\sqrt{\sum_{m=1}^r v_{mp}^2}}.$$

Instead of estimating the magnitudes of the corrections for the direction cosines of the vectors  $\Lambda_p$  by choosing a maximum value for  $k_p c_{mp}$ , it may be desirable to estimate the magnitude of the angular displacement between the given trial vectors  $\Lambda_p$  and the next trial vectors  $M_p$ . Let this angular displacement be  $\theta_p$ . The angles  $\theta_p$  may be determined as follows: The cosine of the angle  $\theta_p$  between the unit vectors  $\Lambda_p$  and  $M_p$  is

$$(68) \quad \cos \theta_p = \sum_{m=1}^r \mu_{mp} \lambda_{mp}.$$

By (48)

$$\mu_{mp} = m(\lambda_{mp} + d\lambda_{mp}).$$

Hence

$$(69) \quad \cos \theta_p = m \sum_{m=1}^r \lambda_{mp} (\lambda_{mp} + d\lambda_{mp})$$

or

$$(70) \quad \cos \theta_p = m \sum_{m=1}^r \lambda_{mp}^2 + m \sum_{m=1}^r \lambda_{mp} d\lambda_{mp}.$$

Since  $\Lambda_p$  are unit vectors,

$$(71) \quad \sum_{m=1}^r \lambda_{mp}^2 = 1 ,$$

and hence

$$(72) \quad \cos \theta_p = m + m \sum_{m=1}^r \lambda_{mp} d\lambda_{mp} .$$

But

$$(57) \quad dz_p = \sum_{m=1}^r \lambda_{mp} d\lambda_{mp} = 0 .$$

Hence

$$(73) \quad \cos \theta_p = m ,$$

or, by (67),

$$(74) \quad \cos \theta_p = \frac{1}{\sqrt{\sum_{m=1}^r v_{mp}^2}} .$$

By (45)

$$(75) \quad \sum_{m=1}^r v_{mp}^2 = \sum_{m=1}^r \lambda_{mp}^2 + 2 \sum_{m=1}^r \lambda_{mp} d\lambda_{mp} + \sum_{m=1}^r (d\lambda_{mp})^2 .$$

By (57) and (71)

$$(76) \quad \sum_{m=1}^r v_{mp}^2 = 1 + \sum_{m=1}^r (d\lambda_{mp})^2 ,$$

or

$$(77) \quad \cos \theta_p = \frac{1}{\sqrt{1 + \sum_{m=1}^r (d\lambda_{mp})^2}} ,$$

and hence

$$(78) \quad \frac{1}{\cos^2 \theta_p} - 1 = \sum_{m=1}^r (d\lambda_{mp})^2 ,$$

which can be written in the form

$$(79) \quad \tan^2 \theta_p = \sum_{m=1}^r (d\lambda_{mp})^2 .$$

By (45) and (63) it follows that

$$(80) \quad \sum_{m=1}^r (d\lambda_{mp})^2 = k_p^2 \sum_{m=1}^r c_{mp}^2 ,$$

so that

$$(81) \quad k_p^2 = \frac{\tan^2 \theta_p}{\sum_{m=1}^r c_{mp}^2} .$$

If a small angular displacement  $\theta_p$  is specified, the corresponding value of  $k_p$  is determined by (81).

It is probably best to choose  $\lambda_{rp}$  as that direction cosine  $\lambda_{mp}$  for each hyperplane  $L_p$  for which  $c_{mp}$  is the largest. The values of  $c_{mp}$  vanish for those trial values of  $\lambda_{mp}$  which minimize the function  $\phi(\lambda_{mp})$ . Their absolute values serve to indicate the rapidity with which the minimum value of  $\phi(\lambda_{mp})$  is being approached.

### Numerical example of method of successive approximation

The method of successive approximation for determining a simple structure will be illustrated by a numerical example of four points in a plane. The four points are shown in *Figure 2*. They were arranged in two groups of two points in each group. The two reference vectors  $\Lambda_p$  of the best-fitting simple structure will then necessarily be orthogonal to radial lines that pass through or near the two groups. Each trial consists of computations that are illustrated by *Tables 9* and *10* for the first trial. All of the trials are summarized in *Table 11*.

In *Table 9* the four points are numbered in column  $j$ . The two co-ordinates  $a_{j1}$  and  $a_{j2}$  for each of the four points are shown in the next two columns. The values of  $v_{jp}$  are shown in the next two columns. Since the first trial vectors  $\Lambda_p$  are taken as unit vectors along the two orthogonal reference axes, the initial values of  $a_{jm}$  and  $v_{jp}$  are identical. The values of  $a_{jm}$  and  $v_{jp}$  are different in all subsequent trials. The resulting values of  $w_j$  are shown in the next column. The sum of this column is the initial value of the function  $\phi(\lambda_{mp})$  which is to be minimized. The next two columns for  $s_{jp}$  facilitate computation of the values of  $c_{mp}$ . The last column is a check column.

Section  $A_{mp}$  of *Table 10* shows the numerical values of the left members of (43). They are obtained directly from *Table 9*. Section  $c_{mp}$  was computed

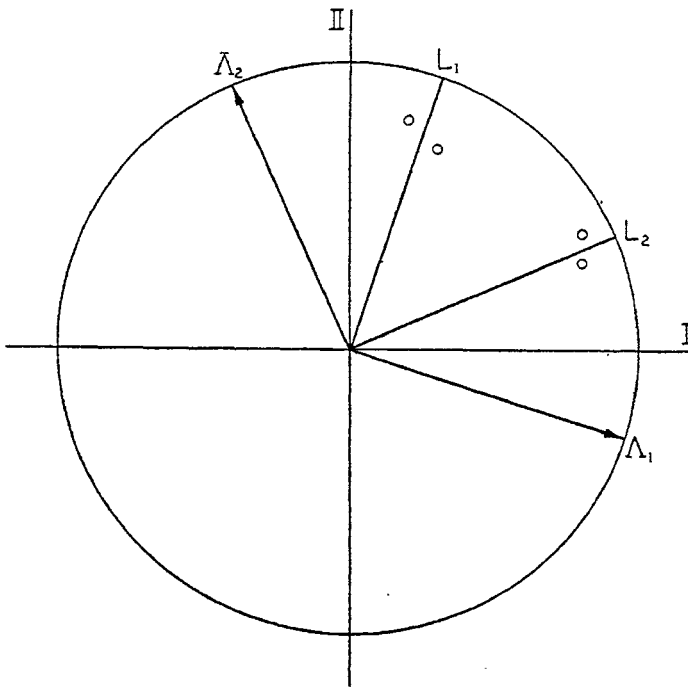


FIGURE 2

Table 9

$j$	$a_{j1}$	$a_{j2}$	$v_{j1}$	$v_{j2}$	$w_j$	$s_{j1}$	$s_{j2}$	$\sum_{m=1}^r a_{j m}$
1	.8	.3	.8	.3	.0576	.0720	.1920	1.1
2	.8	.4	.8	.4	.1024	.1280	.2560	1.2
3	.3	.7	.3	.7	.0441	.1470	.0630	1.0
4	.2	.8	.2	.8	.0256	.1280	.0320	1.0
$\Sigma$	2.1	2.2	2.1	2.2	.2297	.....	.....	4.3

by (61) and (62). The next section ( $d\lambda_{mp}$ ) was computed with such a multiplier  $k$  that the maximum correction was equal to an assigned value which was reduced for each trial. In the first trial this maximum value of the correction was .25, and it was denoted  $e$ . The next two sections are self-explana-

tory. The numerical values of the direction cosines of the two new trial vectors  $M_p$  are recorded in *Table 11*.

Table 10

	.22970000	.38370000	
$\sum_{j=1}^n a_{jms_{jp}} = A_{mp}$	.27810000	.22970000	
	.50780000	.61340000	
	.00000000	- .38370000	
$c_{mp}$	- .27810000	.00000000	
	.00000000	- .25000000	
$d\lambda_{mp} = k_p c_{mp}$	- .18119625	.00000000	$k_p = +.65155069$
	1.00000000	- .25000000	
$(\lambda_{mp} + d\lambda_{mp}) = \nu_{mp}$	- .18119625	1.00000000	
	.98397744	- .24253562	$m_1 = .98397744$
$\mu_{mp} = m\nu_{mp}$	- .17829302	.97014250	$m_2 = .97014250$

Table 11

Trial	$\phi$	$\epsilon$	$\lambda_{11}$	$\lambda_{21}$	$\lambda_{12}$	$\lambda_{22}$
1	.22970000	.25000000	1.00000000	.00000000	.00000000	1.00000000
2	.03658568	.14439991	.98397744	-.17829302	-.24253562	.97014250
3	.00534688	.03000000	.96391815	-.26619880	-.38271927	.92386469
4	.00450116	.01000000	.9544072	-.29518304	-.41250184	.91095677
5	.00442401	.00100000	.95450906	-.29818189	-.40247759	.91542984
6	.00442241	.00070000	.95419615	-.29918173	-.40314454	.91513632
7	.00442213	.....	.95397641	-.29988165	-.40320642	.91510906

The new trial reference vectors reduce the function  $\phi$ . In the second trial  $k$  is taken as unity, since the maximum value of  $c_{mp}$  is of the right order of magnitude. The maximum correction is then .144400. *Table 11* shows the

maximum correction  $e$  for the direction cosines of the trial reference vectors, the resulting reference vectors, and the corresponding reduced value of the function  $\phi$ . If the maximum correction  $e$  is taken too large, the fact will be known by the rise of the function.

This example has been carried to a higher degree of accuracy than will be expedient in most scientific problems. When the direction cosines of the last trial vectors are substituted in (43), it is found that they are proportional to the left members of (43) with a maximum discrepancy which is less than .002. This proves that the minimum value of  $\phi$  is reached. It is represented by the reference vectors  $\Lambda_p$  in *Figure 2*.

### Comparison of methods

All of the methods described in this chapter have been tried on actual psychological test data. It seems conclusive that the best method for most psychological problems is the method represented by equation (12) with *Tables 7* and *8* to facilitate the numerical work. Analytically, the method of (43) is the most interesting; but it is applicable with success only to a perfect simple structure which cannot be expected in any experimentally obtained data.

The analytical method of (43) can probably be modified so as to give as satisfactory results as that of (12) by a proper choice of the function  $w_i$  in (30). In that equation the second power of  $v_{ip}$  is used. It now seems certain that a requirement of the function  $w(v)$  is that the absolute value of its first derivative,  $dw/dv$ , must vary inversely with  $v$  except for small values of  $v$  in the range  $\pm .10$ , where the function should be more stable. This requirement is satisfied by equation (12). The requirement would also be satisfied by the analytical method of adjusting all of the co-ordinate hyperplanes simultaneously if it were modified so that

$$(82) \quad W_i = \prod_{p=1}^r v_{ip}^{2/n},$$

where  $n$  is an integer so chosen that the exponent is a small fraction. If it is desired to avoid having the derivatives become infinite when  $v_{ip}$  is zero, that can be accomplished by adding an arbitrary constant  $c$ , so that the function then becomes

$$(83) \quad W_i' = \prod_{p=1}^r (v_{ip} + c)^{2/n}.$$

The theoretical interest of the analytical method in which all of the hyperplanes are adjusted simultaneously should not be adequate reason for

accepting a solution which fails to maximize the number of vanishing entries in  $V$ . It seems reasonably certain that the method of equation (12) by which each hyperplane is separately adjusted does maximize the number of vanishing entries in  $V$ .

A consideration of primary importance in the determination of simple structure is the fact that its essential feature is a configuration. A requirement is that each trait vector shall lie close to one or more of the co-ordinate hyperplanes, but factorial analysis does not necessarily involve any assumption as to which hyperplane shall contain particular trait vectors. The statistical methods become applicable only when enough of the simple-structure configuration has been gleaned so that assumptions can be made regarding the groups of projections that are to be minimized.

One of the numerous methods that have been tried is to locate each hyperplane so as to minimize the sum of the absolute values of the projections  $|v_{ip}|$ . With slight modification this simple method is now being used successfully.

## CHAPTER VIII

### THE POSITIVE MANIFOLD

#### Restrictions on the factorial matrix

The scientific problems to which the factor methods are applied may require different restrictions on the elements of the factorial matrix. Several of these restrictions have already been discussed, and additional ones will be described in this chapter. Some of these restrictions may be considered under four cases, as follows:

1) The simplest case is that in which the factorial matrix  $F$  can be used as determined by the centroid method, or by any other equivalent method, without restrictions beyond those that are inherent in  $F$ . It is probably seldom that a scientific problem can be adequately solved without some restrictions on the elements of  $F$ .

2) One form of constraint that is of very general scientific interest as regards the factorial matrix is that of simple structure. It seems probable that this constraint will be almost universally imposed in order that the scientific interpretation of the factorial matrix shall be convincing.

3) If the scientific problem is such that negative cell entries in  $F$  are excluded, then we have the important case of a simple structure in which  $a_{jm} \geq 0$ . This is the assumption that underlies the application of factorial methods to the problem of isolating primary mental abilities; but the assumption is not absolutely necessary, since ideal constructs can be devised for a science of psychology which do not require that the cell entries of  $F$ , or those of the oblique factorial matrix  $V$ , be positive or zero.

4) A special case of the positive entries of  $F$  is the further restriction that each factor of  $F$ , or of  $V$ , be either completely present or completely absent in each test. This is a case of possible interest in genetics, but it is not likely that it will be directly applicable to scientific data without admitting a specific variance for each variable.

The first two cases have already been discussed. The last two cases will be described in this chapter. If all of the elements of  $V$  are positive or zero, then each column of  $V$  is defined by a positive hyperplane so located that all of the trait vectors which are not contained in it are on the same side of it. If it is assumed that all of the factors have positive or zero contributions to each variable, then all the trait vectors are in the positive region. The bounding planes of this region are then of special interest.



Definition: *The orthogonal hyperplanes which bound the positive region in  $r$  dimensions will be called the **orthogonal positive manifold**.*

Definition: *A set of  $r$  distinct and oblique positive hyperplanes for a trait configuration in  $r$  dimensions will be called an **oblique positive manifold**.*

An oblique positive manifold is not necessarily confined to the positive region.

Definition: *If the factorial matrix of the traits which are contained in a positive hyperplane is of rank  $(r-1)$ , then the hyperplane is a **bounding hyperplane** or a **positive co-ordinate hyperplane**.*

In the oblique factorial matrix  $V$  the elements of each column are the distances of the traits from a co-ordinate hyperplane. The factorial matrix  $V$ , or a corresponding matrix  $F$ , for those traits that are contained in one of the oblique co-ordinate hyperplanes is of rank  $(r-1)$ . A positive hyperplane can easily be located so that all of the trait vectors are either contained in it or on the same side of it, but it would not necessarily constitute a positive co-ordinate hyperplane. The trait vectors which are contained in it may be of rank less than  $(r-1)$ . A positive hyperplane may be determined so that only one trait vector lies in it, and evidently it would not be a unique co-ordinate hyperplane. However, if the rank of a factorial matrix for the trait vectors which are contained in such a hyperplane is  $(r-1)$ , then the hyperplane is quite likely to be scientifically significant as a positive co-ordinate hyperplane. If, in addition, the criteria of simple structure are satisfied, then the reference traits determined by the intersections of the  $r$  co-ordinate hyperplanes are almost certain to be scientifically significant categories of reference.

If a correlational matrix is of rank  $r_2$  and if the factorial matrix has been computed to  $r_1$  factors where  $r_1 < r_2$ , and if the trait configuration can be inscribed in the positive manifold in  $r_2$  dimensions, then it may be expected that the trait configuration in  $r_1$  dimensions cannot be inscribed in the positive manifold of  $r_1$  dimensions. In such a situation it is a matter of judgment whether the number of columns of  $F$  has been extended far enough. It is in the nature of most scientific problems in which the factor methods are likely to be called upon that all of the common factors of minor significance cannot be extracted. Those common factors which contribute only slightly to the variance of several traits cannot be differentiated with certainty from the variable errors. With only slight representation in the traits they cannot be identified and named with any degree of confidence. Hence it seems useless to carry the columns of  $F$  until the residuals are comparable with the known order of magnitude of the errors in the given correlation coefficients. Since the computations of successive factors must stop before the residuals reach

the order of magnitude of the variable errors in the correlations, it seems necessary to depend on judgment as to when the process should be discontinued. The practical criterion might be adopted that the factors should be extracted until they cease to be meaningful; but interpretation is not feasible until the factors have been rotated, even if an orthogonal system of reference traits is acceptable.

No complete analytical method is available for locating the positive manifold, but several methods of investigating it will be described. If a simple structure is found to exist in the trait configuration, then it may also happen that it can be inscribed in the positive manifold. If this situation is discovered, the simple structure in the positive manifold is especially convincing and the oblique factorial matrix is then almost certain to be scientifically meaningful. It may happen that the problem is of such a nature that simple structure is not to be expected but that the inscribing of the trait configuration in a positive manifold would be meaningful. Then the positive manifold is the means of locating a unique set of reference axes, either orthogonal or oblique, which may be scientifically significant.

### The elimination of negative factor loadings

The orthogonal transformation by which a factorial matrix  $F$  may be rotated into an orthogonal positive manifold  $F_p$  contains  $1/2 r(r-1)$  independent parameters, where  $r$  is the number of columns of  $F$ . Each one of these independent parameters may be thought of as determining an angle of rotation  $\phi$  for a pair of columns of  $F$ . There are as many independent parameters in an orthogonal transformation in  $r$  dimensions as there are pairs of columns. If any two columns are plotted on cross-section paper, the point that deviates farthest from the centroid may be brought into one of the two orthogonal axes by means of a rotation. This procedure may be continued with successive pairs of columns until all the elements of  $F_p$  are positive or zero if the trait configuration can be inscribed in the orthogonal positive manifold.

If the elements of  $F_p$  are theoretically positive or zero, it is to be expected that the variable errors will cause the theoretical zero elements to be higher or lower than zero. Small negative elements may therefore be expected in a factorial matrix which is theoretically positive.

The problem of finding an oblique positive manifold that will circumscribe the trait configuration can sometimes be solved approximately by a procedure not unlike that in which clusters are isolated. This procedure is based on the principle that those trait vectors which lie in or near the intersections of several of the bounding co-ordinate hyperplanes must have a maximum number of nearly vanishing correlations with the other traits.

The procedure is to select  $r$  such traits for which the number of low or nearly vanishing coefficients is maximized. Make a list of more than  $r$  traits with unusually large numbers of low coefficients. Arrange them in a square correlational matrix. Eliminate traits, as for the isolation of clusters, except that in this case it is a set of  $r$  trait vectors with the lowest possible intercorrelations that is sought. When this set of  $r$  trait vectors has been selected, a trial set of co-ordinate hyperplanes may be determined by taking  $r$  sets of  $(r-1)$  of these relatively uncorrelated extreme traits. These  $r$  co-ordinate hyperplanes may be adjusted by the method of maximizing the number of vanishing projections  $v_{ip}$  with due regard, in this case, also to the negative projections which are to be eliminated or reduced to values near zero. This method has been tried with some success, but its applicability can never be guaranteed unless it can be safely postulated that the trait configuration can be inscribed in a positive manifold.

It must be recalled that even if all of the original intercorrelations are positive or zero, it does not follow that the trait configuration can be inscribed in a positive orthogonal manifold. However, if the given correlation coefficients are all positive or zero, or if all the negative coefficients are near zero, then the existence of a positive manifold, as bounding planes for the configuration, is a plausible hypothesis. If the given correlational matrix contains negative coefficients that cannot be made positive by reflection, then the existence of a bounding positive manifold is definitely excluded.

One type of solution to the problem of locating an existing bounding positive manifold which has not yet been adequately investigated is based on the principle that the partial correlation coefficient is represented geometrically as the cosine of a dihedral angle. Consider a reference trait vector  $T$  and the two planes determined by the pair of vectors  $T$  and  $j$  and the pair of vectors  $T$  and  $k$ , where  $j$  and  $k$  are any two trait vectors. If the cosines of a large number of the dihedral angles  $jTk$  are near unity or near zero, then the reference vector  $T$  is likely to be the intersection of a set of  $(r-1)$  positive co-ordinate hyperplanes of a simple structure.

It is occasionally of some interest to determine one or more positive hyperplanes even though they may not be bounding hyperplanes. In *Figure 1* let  $\Delta$  be any trial unit vector in the common-factor space, and let  $u$  be the trait vector that has the largest negative projection  $c$  on  $\Delta$ . Let  $x$  be a vector which is coplanar with  $\Delta$  and with  $u$ . The direction cosines of  $\Delta$  are known, and it is desired to find the direction cosines of  $x$  which are orthogonal to  $u$  so that the projection of  $u$  on  $x$  is zero. The vector  $x$  will constitute the next trial vector.

Since  $x$  is coplanar with  $u$  and  $\Delta$ , the direction cosines of  $x$  can be expressed as linear functions of the direction cosines of  $u$  and of  $\Delta$ . Therefore

$$(1) \quad \begin{cases} x_1 = au_1 + b\lambda_1, \\ x_2 = au_2 + b\lambda_2, \\ \dots \dots \dots \\ x_r = au_r + b\lambda_r, \end{cases}$$

where  $a$  and  $b$  are the parameters to be determined, while  $u_m$  and  $\lambda_m$  are the direction cosines of  $u$  and  $\Delta$ . When the parameters  $a$  and  $b$  have been found,

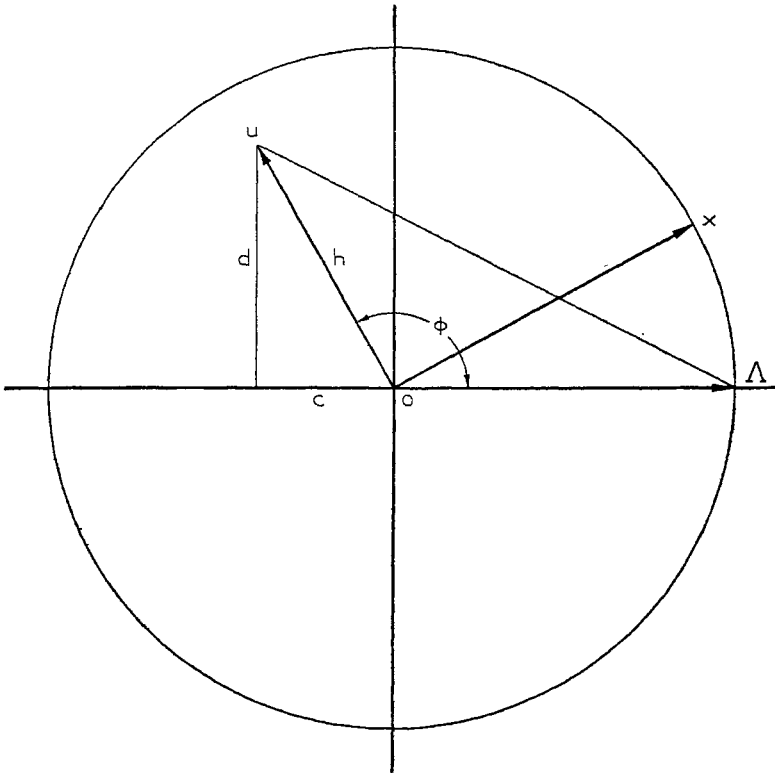


FIGURE 1

they can be used to determine the direction cosines of the vector  $x$  in the common-factor space of  $r$  dimensions which contains  $u$  and  $\Delta$ .

The determination of the vector  $x$  is subject to the conditional equation

$$(2) \quad \sum_{m=1}^r x_m u_m = 0,$$

since  $x$  and  $u$  are statistically independent. Substituting (1) in (2),

$$(3) \quad \sum_{m=1}^r u_m (a u_m + b \lambda_m) = 0,$$

which, after expanding and combining terms, becomes

$$(4) \quad a \sum_{m=1}^r u_m^2 + b \sum_{m=1}^r u_m \lambda_m = 0.$$

But

$$(5) \quad \sum_{m=1}^r u_m^2 = h^2,$$

where  $h^2$  is the communality of the trait  $u$ , and

$$(6) \quad \sum_{m=1}^r u_m \lambda_m = r_{u\lambda},$$

where  $r_{u\lambda}$  is the correlation between the trait  $u$  and the trial vector  $\Lambda$ .

The correlation  $r_{u\lambda}$  can also be written as a scalar product. Hence

$$(7) \quad r_{u\lambda} = h \cos \phi = c.$$

By (5), (6), and (7) equation (4) becomes

$$(8) \quad ah^2 + bc = 0,$$

and hence

$$(9) \quad b = -\frac{ah^2}{c}.$$

Since  $x_m$  are the direction cosines of the unit vector  $x$ ,

$$(10) \quad \sum_{m=1}^r x_m^2 = 1.$$

Substituting (1) in (10),

$$(11) \quad a^2 \sum_{m=1}^r u_m^2 + b^2 \sum_{m=1}^r \lambda_m^2 + 2ab \sum_{m=1}^r u_m \lambda_m = 1.$$

Since  $\Lambda$  is a unit vector, we have by (5), (6), and (7),

$$(12) \quad a^2 h^2 + b^2 + 2abc = 1.$$

Substituting (9) in (12),

$$(13) \quad a^2 h^4 - a^2 c^2 h^2 = c^2,$$

and hence

$$(14) \quad a = -\frac{c}{h\sqrt{h^2 - c^2}}.$$

Substituting (14) in (9),

$$(15) \quad b = \frac{h}{\sqrt{h^2 - c^2}}.$$

The parameters  $a$  and  $b$  are expressed in terms of the communality of  $u$  and the negative correlation  $r_{u\lambda}$ .

Applying these two parameters to the determination of the direction cosines  $x_m$  of  $x$ , we have

$$(16) \quad x_m = au_m + b\lambda_m,$$

by which the new trial vector  $x$  can be determined in the common-factor space of  $r$  dimensions. The projections of the traits in the battery on  $x$  are then found. If a significant negative projection exists, it is treated in the same manner as  $r_{u\lambda} = c$  until the direction cosines of a positive hyperplane have been reached. This method is quite simple in application.

### Unitary factors

A special case of the positive manifold is that in which each reference trait is either completely present or entirely absent in each member of the population  $N$ . Such reference traits may be called *unitary factors* in the sense that the raw scores in the dichotomous distribution of such a trait are either +1 or 0 for each member of the population. The corresponding standard scores in a unitary trait have only two numerical values in the population. These two values depend on the number  $N_s$  of individuals who possess the unitary trait and on the number  $M_s$  in whom the trait is absent.

It is a legitimate hypothesis that the intellectual and emotional traits of people can be reduced eventually to genetic origin. It seems likely that at least some human traits will be expressed in terms of unitary component elements which may be Mendelian in character. If that is to be the eventual outcome, then mental traits will be conceived as the resultant of a group of genetic unitary factors; and, in this context, the complexity of a trait will be the number of unitary factors that demonstrably contribute to the variance of the composite trait. Let  $n_j$  and  $n_k$  be the number of unitary factors that contribute to the variance of the composite traits  $j$  and  $k$ , respectively. It is not to be expected that these unitary factors or genes have equal importance in determining a composite trait. Hence some system of weighting each unitary factor seems essential in expressing the total variance of a trait  $j$  in terms of the  $n_j$  unitary factors that define  $j$ .

It also seems certain that here, as elsewhere in science, the primary causes do not combine in the manner of a weighted sum to produce the composite traits but that non-linear and discontinuous functions are involved. At the present time little is known about the unitary factors that combine to produce the observable human traits; little is known about the complexities,  $n_j$ , of these traits and the functions by which the unitary factors combine their effects to produce the composite traits.

It would seem unduly pessimistic to withdraw from the problem with a conviction that it cannot be solved. The present factorial methods are based on the hope that in some scientific problems, but not necessarily in all of them, a linear combination of factors may serve the purposes of a first approximation and that features of the problem will be revealed by these simple methods that would otherwise remain unnoticed for a long time. The correlation coefficient is itself a symbol of defeat in that its computation is an admission of ignorance about the underlying rational equation. Who would ever compute the correlation coefficient between the length of a pendulum and its period? It could be done by observing the period of each of one hundred pendulums of different lengths. But if the customary equations were unknown, the correlational method certainly would enable us to establish experimentally that there is an inverse relation between the length of a pendulum and its period and that there is no relation between the weight of the pendulum and its period within wide limits of weight. Such facts would be food for speculation concerning the non-linear functions that describe the phenomena more accurately. The factor methods are in a similar situation in that they will certainly be discarded eventually for each class of phenomena when they have served the purpose of revealing some of the significant relations.

In order to illustrate a type of factor analysis that may prove significant

in the future, the correlation between two composite traits will be considered as a function of the unitary elements, factors, or genes, which may be regarded as primary in this context. In order not to obscure the present purpose, the analysis will be made with simplifying assumptions that may account for experimental observations only as a first approximation.

Let the two composite traits be  $j$  and  $k$ , and let there be  $n_j$  unitary elements in  $j$  and  $n_k$  elements in  $k$ . Let there be  $N_e$  individuals in the population  $N$  who possess a particular element  $e$ , and  $M_e$  individuals in whom the element is absent. Let  $u_e$  be the standard score of every individual who possesses the unit factor, and let  $v_e$  be the standard score of each individual in whom the unit factor is absent. Then, by definition,

$$(17) \quad N_e + M_e = N .$$

Since  $u_e$  and  $v_e$  are standard scores,  $x_{ie}$ ,

$$(18) \quad \sum_{i=1}^N x_{ie} = N_e u_e + M_e v_e = 0 ,$$

so that

$$(19) \quad v_e = -\frac{N_e}{M_e} u_e .$$

For the same reason,

$$(20) \quad \sum_{i=1}^N x_{ie}^2 = N_e u_e^2 + M_e v_e^2 = N .$$

Substituting (19) in (20),

$$(21) \quad u_e^2 = \frac{NM_e}{N_e(M_e + N_e)} ,$$

and by (17),

$$(22) \quad u_e^2 = \frac{M_e}{N_e} .$$

Let

$$p_e = \frac{N_e}{N} \quad \text{and} \quad q_e = \frac{M_e}{N} .$$

Then

$$(23) \quad u_e^2 = \frac{q_e}{p_e} .$$



By (23) and (19),

$$(24) \quad v_e^2 = \frac{p_e}{q_e},$$

so that

$$(25) \quad u_e^2 v_e^2 = +1.$$

It is evident from (19) that  $u_e$  and  $v_e$  are of opposite sign. Then  $u_e$  may be taken positive and  $v_e$  negative, since  $u_e$  is the standard score of an individual possessing the unitary trait while  $v_e$  is the standard score for its absence.

The correlation between two composite traits  $j$  and  $k$  can be expressed as follows. Let  $d_{ji}$  be the raw deviation score of individual  $i$  in test  $j$ , and let it be assumed that  $d_{ji}$  is the sum of the standard scores of individual  $i$  in those unitary traits which are involved in  $j$ .

Let  $w_{ije} = u_{ie}$  if the unitary element  $e$  is in  $j$  and if the element  $e$  is present in individual  $i$ .

Let  $w_{ije} = v_{ie}$  if the unitary element  $e$  is in  $j$  and if the element  $e$  is absent in individual  $i$ .

Let  $w_{ije} = 0$  if the unitary element  $e$  is not in  $j$ .

Then

$$(26) \quad d_{ji} = \sum_e w_{ije}.$$

The standard score  $s_{ji}$  is related by a constant multiplier  $b$  to the deviation score  $d_{ji}$ . Hence

$$(27) \quad s_{ji} = b d_{ji} = b \sum_e w_{ije}.$$

To determine the multiplier  $b$ , the squares of the standard scores  $s_{ji}$  may be summed. Then

$$(28) \quad \sum_{i=1}^N s_{ji}^2 = N = b^2 \sum_{i=1}^N \left( \sum_e w_{ije} \right)^2.$$

But the elements are assumed to be uncorrelated in the population. Hence

$$(29) \quad N = b^2 \sum_e \sum_{i=1}^N w_{ije}^2.$$

If an element  $e$  is in  $j$ , then  $w_{ije} = x_{ie}$ , where  $x_{ie}$  is the standard score of individual  $i$  in the unitary trait  $e$ . Its numerical value is either  $u_{ie}$  or  $v_{ie}$ , depending on whether the element  $e$  is, or is not, present in individual  $i$ . Then

$$(30) \quad \sum_{i=1}^N w_{ije}^2 = \sum_{i=1}^N x_{ie}^2 = N.$$

Substituting (30) in (29),

$$(31) \quad N = b^2 \sum_e N.$$

The summation of the constant  $N$  over the elements  $e$  covers  $n_j$  elements. Hence

$$(32) \quad 1 = b^2 n_j,$$

or

$$(33) \quad b = \frac{1}{\sqrt{n_j}}.$$

Substituting (33) in (27),

$$(34) \quad s_{ji} = \frac{1}{\sqrt{n_j}} d_{ji},$$

and, by analogy,

$$(35) \quad s_{ki} = \frac{1}{\sqrt{n_k}} d_{ki}.$$

The correlation between the two composite traits  $j$  and  $k$  is

$$(36) \quad r_{jk} = \frac{1}{N} \sum_{i=1}^N s_{ji} s_{ki}.$$

Substituting (34) and (35) in (36),

$$(37) \quad r_{jk} = \frac{1}{N \sqrt{n_j n_k}} \sum_{i=1}^N d_{ji} d_{ki}.$$

Substituting (26) in (37), and ignoring vanishing cross products,

$$(38) \quad r_{jk} = \frac{1}{N \sqrt{n_j n_k}} \sum_e \sum_i w_{ije} w_{ike}.$$

The product  $w_{ije}w_{ike}$  is equal to  $u_{ie}^2$  if  $e$  is in both  $j$  and  $k$  and if  $e$  is present in  $i$ . It is equal to  $v_{ie}^2$  if  $e$  is in both  $j$  and  $k$  and if  $e$  is absent in  $i$ . It vanishes if  $e$  is absent in  $j$  or in  $k$  or in both. The cross products vanish because the elements are assumed to be uncorrelated. Since  $u_{ie}$  and  $v_{ie}$  are both standard scores,

$$(39) \quad \sum_{i=1}^N w_{ije}w_{ike} = N$$

if  $e$  is in both  $j$  and  $k$ . Then

$$(40) \quad r_{jk} = \frac{1}{N\sqrt{n_j n_k}} \sum_e N.$$

The summation of the constant  $N$  is here over the elements that are common to  $j$  and  $k$ . Hence

$$(41) \quad r_{jk} = \frac{n_{jk}}{\sqrt{n_j n_k}},$$

where  $n_{jk}$  is the number of elements that are common to  $j$  and  $k$ .

This well-known formula for the correlation coefficient expresses the correlation in terms of the number of unitary elements  $n_j$  that are involved in the composite trait  $j$ , the number of unitary elements  $n_k$  in  $k$ , and the number of unitary elements  $n_{jk}$  which are common to  $j$  and  $k$ .

In case the two composite traits  $j$  and  $k$  are of equal complexity as regards the unitary factors, so that  $n_j = n_k = n_i$ , then the formula reduces to the still simpler form

$$(42) \quad r_{jk} = \frac{n_{jk}}{n_i},$$

in which the correlation coefficient is interpreted directly as the ratio of common elements in  $j$  and  $k$ .

Equations (41) and (42) must be interpreted in the light of the simplifying assumptions that the unitary elements are equally weighted in their contributions to the variance of the composite traits and that they are statistically independent as regards their incidence in the population  $N$ .

A type of factor analysis may be developed from this conception of the correlation coefficient in that the three numerical values  $n_j$ ,  $n_k$ , and  $n_{jk}$  are all necessarily integral. It follows that, for a finite battery of traits with limited complexities, the frequency distribution of correlation coefficients

must show discontinuities, and even multimodality. It is conceivable that these multimodalities and discontinuities in the correlation coefficients may be used in an inverse process of reasoning whereby they become the experimental evidence for making inferences about the complexities of the composite traits and about the number of unitary elements that the traits have in common. This type of analysis will undoubtedly proceed by investigat-

Table 1

$$r_{jk} = \frac{n_{jk}}{\sqrt{n_j n_k}}$$

$n_j$	$n_k$	$n_{jk}$			
		0	1	2	3
1	1	.000	1.000		
1	2	.000	.707		
1	3	.000	.577		
1	4	.000	.500		
1	5	.000	.447		
1	6	.000	.408		
1	7	.000	.378		
1	8	.000	.354		
1	9	.000	.333		
1	10	.000	.316		
2	2	.000	.500	1.000	
2	3	.000	.408	.816	
2	4	.000	.354	.707	
2	5	.000	.316	.632	
2	6	.000	.289	.577	
2	7	.000	.267	.535	
2	8	.000	.250	.500	
2	9	.000	.236	.471	
2	10	.000	.224	.447	
3	3	.000	.333	.667	1.000
3	4	.000	.289	.577	.866
3	5	.000	.258	.516	.775
3	6	.000	.236	.471	.707
3	7	.000	.218	.436	.655
3	8	.000	.204	.408	.612
3	9	.000	.192	.385	.577
3	10	.000	.183	.365	.548

$n_j$	$n_k$	$n_{jk}$												
		0	1	2	3	4	5	6	7	8	9	10		
4	4	.000	.250	.500	.750	1.000								
4	5	.000	.224	.447	.671	.894								
4	6	.000	.204	.408	.612	.816								
4	7	.000	.189	.378	.567	.755								
4	8	.000	.177	.354	.530	.707								
4	9	.000	.167	.333	.500	.667								
4	10	.000	.158	.316	.474	.632								
5	5	.000	.200	.400	.600	.800	1.000							
5	6	.000	.183	.365	.548	.730	.913							
5	7	.000	.169	.338	.507	.676	.845							
5	8	.000	.158	.316	.474	.632	.791							
5	9	.000	.149	.298	.447	.596	.745							
5	10	.000	.141	.283	.424	.566	.707							
6	6	.000	.167	.333	.500	.667	.833	1.000						
6	7	.000	.154	.309	.463	.617	.772	.926						
6	8	.000	.144	.289	.433	.577	.722	.866						
6	9	.000	.136	.272	.408	.544	.680	.816						
6	10	.000	.129	.258	.387	.516	.645	.775						
7	7	.000	.143	.286	.429	.571	.714	.857	1.000					
7	8	.000	.134	.267	.401	.535	.668	.802	.935					
7	9	.000	.126	.252	.378	.504	.630	.756	.882					
7	10	.000	.120	.239	.359	.478	.598	.717	.837					
8	8	.000	.125	.250	.375	.500	.625	.750	.875	1.000				
8	9	.000	.118	.236	.354	.471	.589	.707	.825	.943				
8	10	.000	.112	.224	.335	.447	.559	.671	.783	.894				
9	9	.000	.111	.222	.333	.444	.556	.667	.778	.889	1.000			
9	10	.000	.105	.211	.316	.422	.527	.632	.738	.843	.949			
10	10	.000	.100	.200	.300	.400	.500	.600	.700	.800	.900	1.000		

ing the frequency distribution of coefficients separately for each column of a correlational matrix. These coefficients may be considered in their original form or after correcting them for attenuation or for uniqueness.

Table 1 has been prepared for the purpose of illustrating further the discreteness of the numerical values of the correlation coefficients that can be obtained under the assumptions of equation (41). The interpretation of the table can be illustrated by an example. Let two composite traits have complexities of 4 and 5, respectively, so that one of them is determined by four unitary elements and the other by five unitary elements. Then the only possible correlations between the two composite traits are .000, .224,

.447, .671, and .894, depending on whether they have 0, 1, 2, 3, or 4 unitary elements in common.

But while these unitary elements may be acknowledged to be a worthy objective, it must not be assumed that the larger and cruder categories will then vanish in significance. It is still useful to speak of a man's arms and legs even though much is known about the hierarchy of their parts and elements. Even if hundreds of unitary and elemental factors should eventually be discovered to be primary determiners of intellectual endowment, it might still be useful to retain such categories as verballity or visual imagery if they demonstrably simplify our comprehension of mental endowment.

## CHAPTER IX

### ORTHOGONAL TRANSFORMATIONS

#### Rotation in three dimensions

In the previous chapters the theory of multiple-factor analysis has been discussed, including the two cases of orthogonality and obliqueness of the co-ordinate axes. While it is probable that most scientific problems will require the more general oblique co-ordinate axes, it is always of interest to inquire whether the fundamental categories which are represented by the co-ordinate axes may be regarded as statistically independent. In this case the co-ordinate axes are orthogonal and the principal problem is then reduced to that of finding the orthogonal transformation by which the trait configuration of  $F$  can be rotated into a simple structure. In investigations where the co-ordinate axes may be expected to be orthogonal, it is convenient to deal with the rotational transformations in terms of the smallest possible number of parameters. A rotational transformation in a space of  $r$  dimensions is represented by a square matrix of order  $r$ , so that there are  $r^2$  parameters to be determined; but these are not all independent. In this chapter several methods will be described by which a rotational transformation of order  $r$  may be handled in terms of independent parameters. This considerably reduces their number, and it avoids the inconvenience of handling conditional equations. The principles will be described first for a rotational transformation in three dimensions and in four dimensions; but the methods are entirely general, so that they may be applied in a space of any number of dimensions.

Let the given co-ordinates of the points  $a$  be  $a_1, a_2, a_3$ , and let these points be subjected to a rotation. Let the new co-ordinates of the same points be  $A_1, A_2, A_3$ . The change from one set of co-ordinates to the other can be described by the orthogonal transformation

$$(1) \quad \begin{cases} A_1 = a_1x_{11} + a_2x_{21} + a_3x_{31} , \\ A_2 = a_1x_{12} + a_2x_{22} + a_3x_{32} , \\ A_3 = a_1x_{13} + a_2x_{23} + a_3x_{33} . \end{cases}$$

Here  $A_1, A_2, A_3$  represent the new co-ordinates of a point  $a$ , while the given co-ordinates of the same point are  $a_1, a_2, a_3$ . The nine  $x$  values constitute the nine parameters which define the orthogonal transformation.

The transformation may be represented in matrix form more briefly thus:

$$(2) \quad A = aX,$$

where  $X$  is an orthogonal matrix. The determinant of the third-order matrix must be  $+1$ , since the present problem concerns only rotation without reflection. Hence

$$(3) \quad \begin{vmatrix} x_{11} & x_{21} & x_{31} \\ x_{12} & x_{22} & x_{32} \\ x_{13} & x_{23} & x_{33} \end{vmatrix} = +1.$$

The nine parameters in this matrix are not independent. They must satisfy the following six conditional equations:

$$(4) \quad \begin{cases} x_{11}^2 + x_{12}^2 + x_{13}^2 = 1, \\ x_{21}^2 + x_{22}^2 + x_{23}^2 = 1, \\ x_{31}^2 + x_{32}^2 + x_{33}^2 = 1, \end{cases}$$

and

$$(5) \quad \begin{cases} x_{11}x_{21} + x_{12}x_{22} + x_{13}x_{23} = 0, \\ x_{11}x_{31} + x_{12}x_{32} + x_{13}x_{33} = 0, \\ x_{21}x_{31} + x_{22}x_{32} + x_{23}x_{33} = 0. \end{cases}$$

With nine parameters and six conditional equations there are only three independent parameters which determine a rotation in three dimensions.

In order to avoid the use of nine parameters with six conditional equations, the three Eulerian angles may be used as the three independent parameters. These are as follows:\*

$$(6) \quad \begin{cases} a_1 = A_1(\cos \phi \cos \psi - \sin \phi \sin \psi \cos \theta) \\ \quad - A_2(\cos \phi \sin \psi + \sin \phi \cos \psi \cos \theta) \\ \quad + A_3 \sin \phi \sin \theta, \end{cases}$$

$$(7) \quad \begin{cases} a_2 = A_1(\sin \phi \cos \psi + \cos \phi \sin \psi \cos \theta) \\ \quad - A_2(\sin \phi \sin \psi - \cos \phi \cos \psi \cos \theta) \\ \quad - A_3 \cos \phi \sin \theta, \end{cases}$$

$$(8) \quad a_3 = A_1 \sin \psi \sin \theta + A_2 \cos \psi \sin \theta + A_3 \cos \theta.$$

\* Virgil Snyder and C. H. Sisam, *Analytic Geometry of Space* (New York: Henry Holt & Co., 1914), chap. iii.

In this transformation there are only three parameters, namely, the three angles  $\phi$ ,  $\psi$ , and  $\theta$ ; but the transformation is nonsymmetric. In actual computation the three cosines might be regarded as independent parameters; but then the three sines are dependent parameters, so that a transformation by the Eulerian angles involves, in effect, six parameters with three conditional equations.

If this method is to be generalized to hyperspace, it is of interest to know the relation between the rank of the correlational matrix and the number of Eulerian angles, or other independent parameters, that will be needed to determine a rotation in more than three dimensions. The number of independent parameters for a rotation in  $r$  dimensions is  $\frac{1}{2}r(r-1)$ . A rotation in one plane can be effected by disturbing only two columns in the factorial matrix. The number of possible pairs of columns is  $\frac{1}{2}r(r-1)$ , and these rotations would seem to be independent. In the special case where  $r=3$  this gives three independent parameters such as the three Eulerian angles. In order to determine a rotation in four dimensions, we should have six independent parameters or Eulerian angles.

If the matrix  $F$  has rank 3, then its rotation will involve three dimensions. Since this can be effected by three independent parameters, it is desirable to have a transformation with not more than three parameters so as to avoid conditional equations. But there are other requirements that are more essential for convenience in computation. It is sometimes possible to effect a rotation of the factorial matrix on the basis of scientific hypotheses that can be tested. The fine adjustment of the rotation is in effect an infinitesimal rotation, and it will be convenient to have an orthogonal transformation in which the parameters become infinitesimal when the rotation is infinitesimal. In some situations it will also be convenient to start with trial values of the parameters and to solve for the corrections to these trial values. This can be done by means of linear simultaneous equations if second and higher powers of the corrections can be ignored. But that is feasible only if the parameters are themselves fractional—less than unity for any rotation. The most convenient form of orthogonal transformation seems to be one which satisfies the following requirements:

- 1) It should be possible to generalize the orthogonal transformation to any number of dimensions,
- 2) The parameters should become infinitesimal when the rotation is infinitesimal,
- 3) The parameters should be fractional for all rotations,
- 4) The number of parameters should be as small as possible so as to reduce to a minimum the number of conditional equations that are required for numerical work.

An orthogonal transformation will be described that satisfies all of these



requirements except that conditional equations for finite rotations are not eliminated. The entries of an  $n \times 3$  factorial matrix may be thought of as the three co-ordinates of each of  $n$  points. Let the three columns represent the three axes. If such a matrix is rotated about the first axis, it is clear that the first co-ordinate of each point remains unchanged while the second and third co-ordinates are changed. This rotation can be represented by an angle  $\alpha$  at the origin in the 2-3 plane. The transformation may be denoted  $X$ , and it is

$$(9) \quad X = \begin{vmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & -\sin \alpha \\ 0 & \sin \alpha & \cos \alpha \end{vmatrix}.$$

It will be convenient to adopt another notation for the trigonometric functions. Let

$$y_1 \equiv \cos \alpha ,$$

$$x_1 \equiv \sin \alpha .$$

Then the orthogonal transformation becomes

$$(10) \quad X = \begin{vmatrix} 1 & 0 & 0 \\ 0 & +y_1 & -x_1 \\ 0 & +x_1 & +y_1 \end{vmatrix}.$$

This rotation is represented by the matrix equation

$$(11) \quad b = \alpha X ,$$

by which the co-ordinates  $a$  are changed to the co-ordinates  $b$ .

The first rotation  $\alpha$  is in the 2-3 plane, while the first co-ordinate remains unchanged. The second rotation may be taken in the 1-3 plane, leaving the second co-ordinate of  $b$  unchanged. We then have, by analogy,

$$(12) \quad Y = \begin{vmatrix} y_2 & 0 & -x_2 \\ 0 & 1 & 0 \\ x_2 & 0 & y_2 \end{vmatrix}.$$

This rotation may be written in matrix notation as

$$(13) \quad c = bY ,$$

to represent the change in co-ordinates from  $b$  to  $c$ .

The third rotation is then in the 1-2 plane, which leaves the third co-ordinate in  $c$  unchanged. It is represented by the analogous transformation

$$(14) \quad Z = \left\| \begin{array}{ccc} y_3 & -x_3 & 0 \\ x_3 & y_3 & 0 \\ 0 & 0 & 1 \end{array} \right\| .$$

This rotation is shown in matrix notation by the transformation

$$(15) \quad A = cZ .$$

Summarizing the three rotations 11, 13, 15, we have

$$(11) \quad b = aX ,$$

$$(13) \quad c = bY ,$$

$$(15) \quad A = cZ ,$$

from which we have

$$(16) \quad \left\{ \begin{array}{l} A = cZ , \\ = bYZ , \\ = aXYZ . \end{array} \right.$$

Let  $u$  be the matrix product of the three transformations. Then

$$(17) \quad u \equiv XYZ ,$$

so that

$$(18) \quad A = au ,$$

where  $u$  is an orthogonal transformation which changes the co-ordinates of the  $n$  points from  $a$  to  $A$ .

Since the transformations  $X$ ,  $Y$ , and  $Z$  are orthogonal, their product is an orthogonal transformation. The criterion of orthogonality of a matrix  $X$  is that\*

$$(19) \quad X^{-1} = X', \quad \text{or} \quad XX' = I.$$

Then a second orthogonal matrix

$$Y^{-1} = Y', \quad \text{or} \quad YY' = I.$$

The product

$$D = XY.$$

Then

$$D' = Y'X',$$

and

$$DD' = XYY'X'.$$

But

$$YY' = I.$$

Hence

$$DD' = XX'.$$

But

$$XX' = I.$$

Hence

$$DD' = I.$$

By the same reasoning the matrix  $u$  can be shown to be orthogonal. With real parameters in  $X$ ,  $Y$ , and  $Z$ , it is clear that several successive rotations must give rotation as a product.

The row-by-column multiplication of the matrices  $XYZ = u$ , in that order, gives the transformation

$$(20) \quad u = \begin{vmatrix} y_2y_3 & -y_2x_3 & -x_2 \\ -x_1x_2y_3 + y_1x_3 & x_1x_2x_3 + y_1y_3 & -x_1y_2 \\ y_1y_3x_2 + x_1x_3 & -y_1x_2x_3 + x_1y_3 & y_1y_2 \end{vmatrix}.$$

\* H. W. Turnbull and A. C. Aitken, *An Introduction to the Theory of Canonical Matrices* (London and Glasgow: Blackie & Son, 1932), p. 33.

This orthogonal transformation satisfies the requirements in that it can readily be generalized to any number of dimensions. Its parameters are all fractional, since they represent sines and cosines of the successive angles of rotation. For infinitesimal rotations the sines become infinitesimal, so that second powers in the  $x$ -parameters can be neglected. If we suppress the terms of second degree in the  $x$ 's, we have

$$y = \sqrt{1 - x^2} \doteq 1 ,$$

so that the transformation takes the form

$$(21) \quad u \doteq v \equiv \begin{vmatrix} 1 & -x_3 & -x_2 \\ +x_3 & 1 & -x_1 \\ +x_2 & +x_1 & 1 \end{vmatrix}$$

for infinitesimal rotations. This is a skew-symmetric matrix, and it is of some interest to note that an infinitesimal orthogonal transformation seems always to take this form quite irrespective of the many alternative ways in which the finite rotation may be described. This generalization concerning infinitesimal orthogonal transformations seems also to hold for higher dimensions.

Finally, when the successive angles of rotation vanish, the respective sines vanish, the  $x$ -parameters vanish, and the transformation (20) reduces to the identity matrix. This is, of course, what one should expect.

For some purposes the skew-symmetric form (21) may be useful with a rotational criterion. When the  $x$ -parameters of (21) have been determined, they may be substituted in the orthogonal transformation (20) with assurance that the trait configuration will not be disturbed. The resulting factorial matrix can be subjected again to a rotation by the same criterion, estimating the parameters by (21) and rotating by (20). For some problems it may be best to retain all terms of second degree in the  $x$ -parameters of (20). In this case the third and higher powers may be ignored.

In the transformation (20) the  $y$ -parameters may be expressed in terms of the  $x$ -parameters. We have then

$$y = (1 - x^2)^{\frac{1}{2}} .$$

Expanding and ignoring terms of third and higher degree,

$$y \doteq 1 - \frac{x^2}{2} .$$

Then

$$u_{11} = y_2 y_3 \doteq \left(1 - \frac{x_2^2}{2}\right) \left(1 - \frac{x_3^2}{2}\right) \doteq 1 - \frac{x_2^2}{2} - \frac{x_3^2}{2},$$

$$u_{12} = -y_2 x_3 \doteq -x_3 \left(1 - \frac{x_2^2}{2}\right) \doteq -x_3,$$

$$u_{13} = -x_2.$$

Proceeding in the same manner for the other cells of (20), it takes the form

$$(22) \quad u \doteq W \equiv \begin{vmatrix} 1 - \frac{x_2^2}{2} - \frac{x_3^2}{2} & -x_3 & -x_2 \\ x_3 - x_1 x_2 & 1 - \frac{x_1^2}{2} - \frac{x_3^2}{2} & -x_1 \\ x_1 x_3 + x_2 & x_1 - x_2 x_3 & 1 - \frac{x_1^2}{2} - \frac{x_2^2}{2} \end{vmatrix}.$$

This transformation is obtained by ignoring terms of third and higher degree, while (21) is obtained by ignoring terms of second and higher degree. Forms like (21) and (22) may be used to estimate the numerical values of the parameters. The actual rotation can be effected by an orthogonal transformation (20) with the parameters so determined.

### Alternative transformations

The orthogonal transformation (20) is not unique. Other orthogonal transformations may be used, but the one that has been described may satisfy best the requirements that seem to be indicated for the factor problem. Among the various possible orthogonal transformations that have been investigated there may be mentioned the following, namely,

$$(23) \quad \frac{1}{(1+x_1^2)^{\frac{1}{2}}} \begin{vmatrix} (1+x_1^2)^{\frac{1}{2}} & 0 & 0 \\ 0 & 1 & +x_1 \\ 0 & -x_1 & 1 \end{vmatrix},$$

which can be generalized to hyperspace. The parameters are not necessarily fractional for finite rotations, and this would constitute a handicap in some forms of manipulation.

One type of orthogonal transformation of special interest has been mentioned by Professor E. B. Wilson.\* An interesting characteristic of this transformation is that the parameters are all rational and independent. It can be generalized to hyperspace. For a rotation in three dimensions it takes the following form:

$$(24) \quad \left\| \begin{array}{ccc} \frac{1 + p^2 - q^2 - r^2}{c} & \frac{2pq - 2r}{c} & \frac{2pr + 2q}{c} \\ \frac{2pq + 2r}{c} & \frac{1 - p^2 + q^2 - r^2}{c} & \frac{2qr - 2p}{c} \\ \frac{2pr - 2q}{c} & \frac{2qr + 2p}{c} & \frac{1 - p^2 - q^2 + r^2}{c} \end{array} \right\|$$

where  $c = 1 + p^2 + q^2 + r^2$ , and  $\tan^2 \theta/2 = p^2 + q^2 + r^2$ , while  $\theta$  is the angle of rotation about an axis  $l$ . The direction cosines of  $l$  are proportional to  $p, q, r$ . This would probably be the best form of orthogonal transformation for the factor problem except for the fact that the parameters are not necessarily fractional for finite rotations. (Consider for example  $\theta = \pi$  in (24).) Fractional parameters are convenient for some computing purposes in which second and higher powers of the parameters are to be ignored. Again, it may be desirable in some computations to start with trial values of the parameters and to solve for a small correction for each parameter. In order to be able to work with linear normal equations it is necessary to be able to ignore second and higher powers of the corrections. These considerations would lead one to prefer a transformation in which the parameters are fractional by definition. However, transformation (24) may be used with a suitable multiplier. This device could also be used on transformation (23), but such possibilities have not yet been investigated.

Ignoring the terms of second degree in (24), the transformation reduces to the form

$$(25) \quad \left\| \begin{array}{ccc} 1 & -2r & +2q \\ +2r & 1 & -2p \\ -2q & +2p & 1 \end{array} \right\|,$$

which is again a skew-symmetric matrix.

\* "On the Invariance of General Intelligence," *Proc. Nat. Acad. Sci.*, XIX, August, 1933, p. 771, n. 5.

**Rotation in four dimensions**

The procedure of writing any orthogonal transformation which has been described can be generalized to any number of dimensions. It will be extended here to four dimensions. Each independent rotation may be regarded as a disturbance of a pair of columns in the factorial matrix. Each of these independent rotations is determined by one of the independent  $x$ -parameters and its dependent  $y$ -parameter. The number of independent  $x$ -parameters required to determine a rotation in space of  $r$  dimensions is equal to the number of possible pairs of columns that may be taken in the  $n \times r$  matrix  $F$ . This is  $\frac{1}{2}r(r-1)$ , and consequently we should expect to have six independent parameters for a rotation in four dimensions.

If the four columns of  $F$  are numbered, then the six parameters may be associated with pairs of columns in  $F$ . These may be taken in the following arbitrary order: 1-2, 1-3, 1-4, 2-3, 2-4, 3-4. Let the corresponding independent parameters be  $x_1, x_2, x_3, x_4, x_5, x_6$ , and the corresponding dependent parameters  $y_1, y_2, y_3, y_4, y_5, y_6$ . These are related to the independent parameters as follows:

$$x_1 = \sin \alpha_1,$$

$$y_1 = \cos \alpha_1 = \sqrt{1 - x_1^2},$$

with analogous interpretation for each of the other five subscripts. Each pair of columns in  $F$  is represented by an orthogonal transformation. The matrix product of these six transformations is the matrix of the orthogonal transformation in four dimensions. The six independent rotations are as follows:

$$B_1 = \begin{vmatrix} y_1 & -x_1 & 0 & 0 \\ x_1 & y_1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix},$$

$$B_2 = \begin{vmatrix} y_2 & 0 & -x_2 & 0 \\ 0 & 1 & 0 & 0 \\ x_2 & 0 & y_2 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix},$$

$$B_3 = \begin{vmatrix} y_3 & 0 & 0 & -x_3 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ x_3 & 0 & 0 & y_3 \end{vmatrix},$$

$$B_4 = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & y_4 & -x_4 & 0 \\ 0 & x_4 & y_4 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix},$$

$$B_5 = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & y_5 & 0 & -x_5 \\ 0 & 0 & 1 & 0 \\ 0 & x_5 & 0 & y_5 \end{vmatrix},$$

$$B_6 = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & y_6 & -x_6 \\ 0 & 0 & x_6 & y_6 \end{vmatrix}.$$

Let the given co-ordinates be  $a$ , and let the final transformed co-ordinates be  $A$ . The points  $a$  are to be subjected to six independent and successive rotations which bring them to the co-ordinates  $A$ . Let the six independent rotations be represented as follows:

$$(26) \quad \begin{cases} c = aB_1, \\ d = cB_2, \\ e = dB_3, \\ f = eB_4, \\ g = fB_5, \\ A = gB_6. \end{cases}$$



Combining these six independent rotations, we get the single rotation

$$(27) \quad A = aB_1B_2B_3B_4B_5B_6 .$$

Let

$$v \equiv B_1B_2 \dots B_6 ,$$

Then

$$(28) \quad A = av ,$$

in which  $v$  is an orthogonal matrix of order 4 with six independent parameters and six dependent parameters.

After performing the matrix multiplication of (28) we have the following expressions for the cell entries of  $v$ :

$$v_{11} = y_1y_2y_3 ,$$

$$v_{21} = y_2y_3x_1 ,$$

$$v_{31} = y_3x_2 ,$$

$$v_{41} = x_3 ,$$

$$v_{12} = -y_4y_5x_1 - y_1y_5x_2x_4 - y_1y_2x_3x_5 ,$$

$$v_{22} = +y_1y_4y_5 - y_5x_1x_2x_4 - y_2x_1x_3x_5 ,$$

$$v_{32} = +y_2y_5x_4 - x_2x_3x_5 ,$$

$$v_{42} = +y_3x_5 ,$$

$$v_{13} = y_6x_1x_4 - y_1y_4y_6x_2 + y_4x_1x_5x_6 + y_1x_2x_4x_5x_6 - y_1y_2y_5x_3x_6 ,$$

$$v_{23} = -y_1y_6x_4 - y_4y_6x_1x_2 - y_1y_4x_5x_6 + x_1x_2x_4x_5x_6 - y_2y_5x_1x_3x_6 ,$$

$$v_{33} = y_2y_4y_6 - y_2x_4x_5x_6 - y_5x_2x_3x_6 ,$$

$$v_{43} = y_3y_5x_6 ,$$

$$v_{14} = -x_1x_4x_6 + y_1y_4x_2x_6 + y_4y_6x_1x_5 + y_1y_6x_2x_4x_5 - y_1y_2y_5y_6x_3 ,$$

$$v_{24} = +y_1x_4x_6 + y_4x_1x_2x_6 - y_1y_4y_6x_5 + y_6x_1x_2x_4x_5 - y_2y_5y_6x_1x_3 ,$$

$$v_{34} = -y_2y_4x_6 - y_2y_6x_4x_5 - y_5y_6x_2x_3 ,$$

$$v_{44} = y_3y_5y_6 .$$

If the terms in second and higher degree in the independent  $x$ -parameters are ignored, the matrix  $v$  reduces to

$$(29) \quad \begin{vmatrix} 1 & -x_1 & -x_2 & -x_3 \\ x_1 & 1 & -x_4 & -x_5 \\ x_2 & x_4 & 1 & -x_6 \\ x_3 & x_5 & x_6 & 1 \end{vmatrix}.$$

It will be seen that (29) is again a skew-symmetric matrix. It reduces to the identity matrix when the six rotational angles vanish.

The method which has just been described can be generalized to hyperspace. A rotation in five dimensions requires ten independent parameters. Six dimensions require fifteen independent parameters. If the number of primary factors is fairly large, it seems evident that the direct application of an orthogonal transformation to the factorial matrix  $F$  in the search for the primary factors is prohibitive in computational labor. The use of an orthogonal transformation on  $F$  presupposes the serious restriction that the primary factors are statistically independent in the experimental population. Since this is a condition that cannot be assumed in most factor problems, the rotational transformations must be subject to the same limitation.

## CHAPTER X

### THE APPRAISAL OF ABILITIES

#### The regression $x$ on $s$

The principal problem to which the previous chapters have been directed is that of isolating and identifying primary factors in a battery of traits. The psychological application of factor theory which is of most general current interest is the isolation of primary abilities. The present chapter is directed to the problem of appraising the several primary traits in each individual. The methods to be described are applicable not only to the psychological problem of describing the mental and physical traits of individuals, including native as well as acquired traits, but also to any situation in which it is desired to describe the individual members of a statistical group as regards the traits that may have been found to be primary.

Each individual member of the statistical population is described in terms of  $r$  abilities. Let the standard score of individual  $i$  in the primary ability  $p$  be denoted  $x_{pi}$ . It is desired to estimate  $x_{pi}$  in terms of the  $n$  tests which individual  $i$  has taken. The standard score of individual  $i$  on a test  $j$  has been denoted  $s_{ji}$ .

The regression  $x_{pi}$  on  $s_{ji}$  is as follows:

$$(1) \quad x_{pi} = \sum_{j=1}^n w_{pj}s_{ji} + \rho_{pi} ,$$

where the subscript  $p$  refers to primary abilities,  $j$  refers to the tests,  $i$  refers to the individuals,  $w_{pj}$  is the weight of the score  $s_{ji}$  in test  $j$  in the appraisal of the primary ability  $p$ ,  $\rho_{pi}$  is the residual or discrepancy between the true value of  $x_{pi}$  and the best value which can be obtained as a linear function of the test scores  $s_{ji}$ . Expressing this equation explicitly for the residual,

$$(2) \quad x_{pi} - \sum_{j=1}^n w_{pj}s_{ji} = \rho_{pi} .$$

It is desired to determine the values of  $w_{pj}$  which will minimize  $\rho_{pi}$ . Squaring (2),

$$(3) \quad x_{pi}^2 - 2x_{pi} \sum_{j=1}^n w_{pj}s_{ji} + \sum_{k=1}^n \sum_{j=1}^n w_{pj}w_{pk}s_{ji}s_{ki} = \rho_{pi}^2 .$$

Summing for the population, and dividing by  $N$ ,

$$(4) \quad \frac{1}{N} \sum_{i=1}^N x_{pi}^2 - \frac{2}{N} \sum_{i=1}^N x_{pi} \sum_{j=1}^n w_{pj} s_{ji} + \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^n \sum_{k=1}^n w_{pj} w_{pk} s_{ji} s_{ki} = \frac{1}{N} \sum_{i=1}^N \rho_{pi}^2 = u_p,$$

where  $u_p$  is the quantity to be minimized.

But

$$(5) \quad \frac{1}{N} \sum_{i=1}^N x_{pi}^2 = 1,$$

since  $x_{pi}$  is a standard score. Substituting (5) in (4) and rearranging,

$$(6) \quad 1 - 2 \sum_{j=1}^n w_{pj} \frac{1}{N} \sum_{i=1}^N s_{ji} x_{ip} + \sum_{k=1}^n \sum_{j=1}^n w_{pj} w_{pk} \frac{1}{N} \sum_{i=1}^N s_{ji} s_{ki} = u_p.$$

The summation

$$(7) \quad \frac{1}{N} \sum_{i=1}^N s_{ji} x_{ip} = r_{jp},$$

where  $r_{jp}$  is the correlation between the test  $j$  and the primary ability  $p$ . It is the scalar product of the test vector  $j$  and the primary vector  $T_p$ . It is here assumed that the primary abilities may be correlated in the experimental population  $N$ .

The summation

$$(8) \quad \frac{1}{N} \sum_{i=1}^N s_{ji} s_{ki} = R_{jk} = R_{kj},$$

where  $R_{jk}$  is the correlation between tests  $j$  and  $k$ . The correlation  $R_{jk}$  in (8) is equal to  $r_{jk}$  when  $j \neq k$ , but it is unity when  $j = k$ .

Substituting (7) and (8) in (6),

$$(9) \quad 1 - 2 \sum_{j=1}^n w_{pj} r_{jp} + \sum_{k=1}^n \sum_{j=1}^n w_{pj} w_{pk} R_{jk} = u_p.$$

The normal equations for determining  $w_{pj}$ , or  $w_{pk}$ , are in the form

$$(10) \quad \frac{\partial u_p}{\partial w_{pj}} = 0.$$

Taking partial derivatives in (9),

$$(11) \quad \frac{\partial u_p}{\partial w_{pj}} = -2r_{jp} + 2 \sum_{k=1}^n w_{pk} R_{jk}.$$

Setting the partial derivatives equal to zero, dividing the equation by 2, and transposing,

$$(12) \quad \sum_{k=1}^n w_{pk} R_{kj} = r_{jp} = r_{pj}.$$

Equation (12) represents a matrix multiplication which may be written in matrix notation,

$$(13) \quad w_{pk} R_{kj} = R_{pj},$$

where  $w_{pk}$  is a matrix of order  $r \times n$  and  $R_{kj}$  is a matrix of order  $n \times n$ . The latter is of rank  $n$  because the diagonal elements are unity, and hence specific factors and error factors are involved. The matrix  $R_{pj}$  is of order  $r \times n$ . Since  $R_{kj}$  is non-singular, the equation (13) may be written explicitly for  $w_{pk}$ . Then

$$(14) \quad w_{pk} = R_{pj} R_{kj}^{-1} = R_{pj} R_{jk}^{-1}.$$

Writing (14) in transposed form, and using  $w_{pk} = w_{pj}$ ,

$$(15) \quad w_{jp} = R_{kj}^{-1} R_{jp},$$

by which the numerical values of  $w_{jp} = w_{pj}$  in (1) can be determined.

### The regression $s$ on $x$

This regression implies that the primary abilities of an individual are known and that it is desired to estimate what his performance will be on a test with known factorial weightings. This is the reverse of the previous regression  $x$  on  $s$  in which it is assumed that an individual's scores are known and that his primary abilities are to be appraised.

The case in which  $s_{ji}$  is to be estimated by  $x_{pi}$  can be written in the form

$$(16) \quad s_{ji} = \sum_{p=1}^r w_{jp} x_{pi} + \rho_{ji},$$

where  $w_{jp}$  is the weight of the score  $x_{pi}$  in the estimate of the score  $s_{ji}$ , and  $\rho_{ji}$  is the discrepancy between the actual score  $s_{ji}$  and the estimated score

in test  $j$ . It should be noted that  $w_{ip}$  in (16) is not the transpose of  $w_{pj}$  in (1), since these are coefficients in two different regressions. Writing (16) explicitly for  $\rho_{ji}$ ,

$$(17) \quad s_{ji} - \sum_{p=1}^r w_{ip} x_{pi} = \rho_{ji}.$$

Squaring (17),

$$(18) \quad s_{ji}^2 - 2s_{ji} \sum_{p=1}^r w_{ip} x_{pi} + \sum_{q=1}^r \sum_{p=1}^r w_{ip} w_{iq} x_{pi} x_{qi} = \rho_{ji}^2.$$

Summing for the population and dividing by  $N$ ,

$$(19) \quad \frac{1}{N} \sum_{i=1}^N s_{ji}^2 - \frac{2}{N} \sum_{i=1}^N s_{ji} \sum_{p=1}^r w_{ip} x_{pi} + \frac{1}{N} \sum_{i=1}^N \sum_{q=1}^r \sum_{p=1}^r w_{ip} w_{iq} x_{pi} x_{qi} = \frac{1}{N} \sum_{i=1}^N \rho_{ji}^2 = u_j,$$

where  $u_j$  is the quantity to be minimized.

But, by definition,

$$(20) \quad \frac{1}{N} \sum_{i=1}^N s_{ji}^2 = 1.$$

Substituting (20) in (19), and rearranging,

$$(21) \quad 1 - 2 \sum_{p=1}^r w_{ip} \frac{1}{N} \sum_{i=1}^N s_{ji} x_{pi} + \sum_{q=1}^r \sum_{p=1}^r w_{ip} w_{iq} \frac{1}{N} \sum_{i=1}^N x_{pi} x_{qi} = u_j.$$

The summation

$$(22) \quad \frac{1}{N} \sum_{i=1}^N s_{ji} x_{pi} = r_{jp},$$

and the summation

$$(23) \quad \frac{1}{N} \sum_{i=1}^N x_{pi} x_{qi} = R_{pq},$$

where  $R_{pq}$  is the correlation between the primary abilities  $p$  and  $q$ . It can also be regarded as the cosine of the angular separation between the two primary unit vectors  $T_p$  and  $T_q$ .

Substituting (22) and (23) in (21),

$$(24) \quad 1 - 2 \sum_{p=1}^r w_{jp} r_{jp} + \sum_{q=1}^r \sum_{p=1}^r w_{jp} w_{jq} R_{pq} = u_j .$$

The normal equations are in the form

$$(25) \quad \frac{\partial u_j}{\partial w_{jp}} = 0 .$$

Taking partial derivatives in (24),

$$(26) \quad \frac{\partial u_j}{\partial w_{jp}} = -2r_{jp} + 2 \sum_{q=1}^r w_{jq} R_{qp} .$$

Setting the partial derivatives equal to zero, dividing the equation by 2, and transposing,

$$(27) \quad \sum_{q=1}^r w_{jq} R_{qp} = r_{jp} .$$

Writing equation (27) in matrix notation,

$$(28) \quad w_{jq} R_{qp} = R_{jp} .$$

Since the primary abilities are linearly independent, it follows that the rank of  $R_{qp}$  is  $r$ . Hence  $R_{qp}$  is non-singular. Equation (28) may therefore be written explicitly for  $w_{jq}$ ,

$$(29) \quad w_{jq} = R_{jp} R_{qp}^{-1} ,$$

by which the weights  $w_{jq} = w_{jp}$  in the regression equation (16) may be computed.

It is of interest to note the form which equation (29) takes in the special case where the primary abilities are orthogonal. Then

$$(30) \quad R_{qp}^{-1} = I , \text{ the identity matrix,}$$

and

$$R_{jp} = F_{jp} ,$$

where  $F_{jp}$  is a factorial matrix with orthogonal primary reference vectors, so that

$$(31) \quad w_{jp} = F_{jp} .$$

Substituting (31) in the regression equation (16), we have

$$(32) \quad s_{ji} = F_{ip}x_{pi} + \rho_{ji}.$$

In this equation  $\rho_{ji}$  is that part of the score  $s_{ji}$  which is not produced by the primary common factors. Hence  $\rho_{ji}$  is produced by specific and error factors. In the simplest case where all of the contributing factors are common, we have

$$(33) \quad s_{ji} = F_{ip}x_{pi},$$

which is the first equation of *chapter i*, as was to be expected.



## APPENDIX I

### OUTLINE OF CALCULATIONS FOR THE CENTROID METHOD WITH UNKNOWN DIAGONALS

An  $n \times r$  matrix  $F$  may be obtained from a given correlational matrix  $R$  with unknown diagonals by the following calculations. This is the method described in Example 6, *chapter iii*, except that the reflection of traits will not always be carried to the point where all column sums are positive when the diagonals are ignored. Only those traits will be reflected which minimize the number of negative signs in each column of  $R$ , as described in Example 5, *chapter iii*.

The method will be described in relation to the computations on a  $9 \times 9$  table of experimental correlations given by Professor Carl C. Brigham in his 1928 annual report to the College Entrance Examination Board. The data represent nine intelligence tests used by the College Entrance Board. The correlations are based on the records of 4,175 boys.

The calculations are recorded on data sheets\* devised for twenty variables or less. In working with more than twenty variables, the correlation table may be divided into  $20 \times 20$  sections with a data sheet for each section. The notation  $\Sigma$ ,  $B$ ,  $D$ ,  $E$ , and  $K$ , on the data sheet is the same as that used in the tables of *chapter iii*.

#### Steps in calculation

1. Record the table of intercorrelations as shown in *Table 1*. This may be any  $n \times n$  correlational matrix  $R$  with elements  $r_{jk}$  which satisfies the inequality (5—ii). In this example  $R_0$  is given,  $n=9$ , and the inequality is satisfied if the number of factors turns out to be 5 or less.

2. Record the signs of these correlations as indicated in the *upper part* of the narrow cells provided for the signs. This corresponds to the "first position" of the signs described in example 5 of *chapter iii*.

3. The diagonal cells of this table are blank, since the communalities are unknown. The cells for the entries  $r_{11}$ ,  $r_{22}$ , . . . ,  $r_{99}$  are the diagonal cells.

4. Since all of the entries of this table are positive, it is not necessary to reflect any of the tests. When any column of  $R$  has a majority of negative coefficients, traits are reflected at this stage of the procedure by the method described for *Table 2*.

\* These data sheets are available at the University of Chicago Bookstore, Chicago, Illinois.

5. Decide upon the estimate of the communalities to be used. A small number of variables demands a more accurate estimation of the communalities. When  $n$  is large, Method 4 of *chapter ii* is recommended. This method will be used here.

6. Pick the highest coefficient in each column, *disregarding sign*, and record it in the diagonal with *positive sign* placed in the upper half of the narrow sign cell.

EXAMPLE: The highest correlation in column 1 is .625. It is recorded in the diagonal cell of that column as +.625. If the highest coefficient in this column had been -.625, it would still have been recorded in the diagonal as +.625.

7. Add the entries in each column and record the sums in row  $D$  at the bottom of the data sheet. These are the sums  $\sum_{j=1}^n r_{jk} = r_k$  for each column  $k$  of equation (12-iii).

EXAMPLES: The sum of the nine entries in column 1 is 5.022. This is recorded in row  $D$ , column 1. It is the sum  $\sum_{j=1}^9 r_{j1} = r_1$ .

The sum of the nine entries in column 2 is 4.213. This is recorded in row  $D$ , column 2. It is the sum  $\sum_{j=1}^9 r_{j2} = r_2$ .

8. Add the entries in each row of *Table 1* and record the sums in column  $D$  at the extreme right of the data sheet. These are the sums  $\sum_{k=1}^n r_{jk} = r_j$  for each row  $j$ . These sums should agree with their corresponding column sums recorded in row  $D$  at the bottom of the data sheet.

9. Add all the column sums in row  $D$ . Record this value, 42.072, in row  $D$ , column  $D$ . This is the sum  $\sum_{k=1}^n \sum_{j=1}^n r_{jk} = \sum_{k=1}^n r_k = r_t$ , of equation (8-iii).

10. Add all the row sums of column  $D$ . This gives  $\sum_{j=1}^n \sum_{k=1}^n r_{jk} = \sum_{j=1}^n r_j = r_t = 42.072$ . Check to see that this sum agrees with the sum obtained in step 9.

11. Determine  $\sqrt{r_t}$ . In this example,  $\sqrt{r_t} = \sqrt{42.072} = 6.486293$ . Record this value in the space below  $r_t$ .

12. Compute the reciprocal,  $\frac{1}{\sqrt{r_t}}$ . For these data, the value is  $1/6.486293 = .154171$ . Record .154171 in the space below  $\sqrt{r_t}$ .

13. Multiply each sum in row *D* by the value  $\frac{1}{\sqrt{r_t}} = .154171$  obtained in step 12 and record the results in row *E* at the bottom of the data sheet. Each value in row *E* takes the sign of its corresponding sum in row *D*. These are the first-factor loadings,  $a'_{ki}$ , with the signs of the variables as used to obtain the sums in *D*.

EXAMPLES: Test 1:  $a'_{11} = .154171(+5.022) = +.774$ .

Test 2:  $a'_{21} = .154171(+4.213) = +.650$ .

. . . . .

Test 9:  $a'_{91} = .154171(+5.009) = +.772$ .

14. The product  $r_t \left( \frac{1}{\sqrt{r_t}} \right)$  should give  $\sqrt{r_t}$  recorded on the data sheet if the arithmetical work in determining the multiplier has been correct. In this example,  $42.072 (.154171) = 6.486282$ , which checks with the recorded value of  $\sqrt{r_t} = 6.486293$  to the fourth decimal place.

15. If the loadings in row *E* represent a centroid system, then  $\Sigma E$ , the sum of all the entries in row *E*, should equal  $\sqrt{r_t}$ . Record  $\Sigma E$  in the space  $\Sigma E$  in the lower right corner of the data sheet.

EXAMPLE:  $\Sigma E = 6.486$ .

$\sqrt{r_t} = 6.486$ .

16. Copy the values of row *E* in row *K* with the sign reversed for each test which has been reflected an odd number of times. Any test reflected an odd number of times will have the last recorded sign negative before its variable number. Since no tests were reflected in this table, the values in rows *E* and *K* are the same, all of the first-factor loadings are positive, and  $\Sigma K = \Sigma E = 6.486$ .

17. Record the values of row *K* in the first column of *Table 7*. *Table 7* will be the  $n \times r$  matrix *F* when *r* factors have been extracted.

18. Take a new data sheet and label it "First-Factor Residual Coefficients:  $r_{2 \cdot jk}$ " as shown in *Table 2*.

19. Insert the variable numbers with signs as given in *Table 1*, in the second row and second column provided for them in *Table 2*. Place the signs in the upper half of the narrow sign column. In this example, all of these signs are positive.

In data where traits are reflected in the first table, the signs transferred to *Table 2* are those of the traits on *Table 1* after reflection.

20. Copy the first-factor loadings from row  $E$  of *Table 1* in the first row and in the first column labeled  $E'$  in *Table 2*. This arrangement facilitates residual computations on a calculating machine.

EXAMPLES: The first-factor loading in test 1 from row  $E$  of *Table 1* is  $+.774$ . It is recorded in *Table 2* in the space in front of variable 1 in column  $E'$  and in the space above variable 1 in row  $E'$ .

Similarly, the first-factor loading in test 2 from row  $E$  of *Table 1* is  $+.650$ . It is recorded in *Table 2* in the space in front of variable 2 in column  $E'$  and in the space above variable 2 in row  $E'$ .

21. Check this transfer by adding the loadings for the nine rows of column  $E'$ . This gives  $\Sigma E' = \sum_{j=1}^9 a'_{j1} = 6.486$ , which is the value of  $\Sigma E$  on *Table*

1. Record this sum in the space marked  $\Sigma E'$  at the left of the data sheet.

Add the loadings for the nine columns of row  $E'$ . This gives  $\Sigma E' = \sum_{k=1}^9 a'_{k1} = 6.486$ , which is the value of  $\Sigma E$  on *Table 1*. Record this value of  $\Sigma E'$  in the space provided in the upper right corner of the data sheet.

22. Compute the first-factor residuals by formulae of the type (14-iii),

$$r_{2 \cdot jk} = r_{jk} - a'_{j1} a'_{k1},$$

and record in the  $j$ th row and  $k$ th column of *Table 2*.

23. For column 1 of *Table 2*, these residuals are

$$r_{2 \cdot j1} = r_{j1} - a'_{j1} a'_{11},$$

where  $k=1$  and  $j$  takes values from 1 to 9. The value  $r_{j1}$  is the entry in the  $j$ th row and first column of *Table 1*;  $a'_{j1}$  is the first-factor loading for test  $j$  recorded in row  $j$  of column  $E'$  in *Table 2* and  $a'_{11}$  is the first-factor loading in test 1 recorded at the top of column 1 in row  $E'$  of *Table 2*.

EXAMPLES:

$$r_{2 \cdot 11} = +.625 - .774(.774) = (+.026). \text{ Record above double line in column 1.}$$

$$r_{2 \cdot 21} = +.482 - .650(.774) = -.021. \text{ Record in row 2, column 1.}$$

$$r_{2 \cdot 31} = +.617 - .731(.774) = +.051. \text{ Record in row 3, column 1.}$$

$$r_{2 \cdot 41} = +.518 - .665(.774) = +.003. \text{ Record in row 4, column 1.}$$

$$r_{2 \cdot 51} = +.625 - .804(.774) = +.003. \text{ Record in row 5, column 1.}$$

$$r_{2.61} = +.422 - .593(.774) = -.037. \text{ Record in row 6, column 1.}$$

$$r_{2.71} = +.584 - .738(.774) = +.013. \text{ Record in row 7, column 1.}$$

$$r_{2.81} = +.563 - .759(.774) = -.024. \text{ Record in row 8, column 1.}$$

$$r_{2.91} = +.586 - .772(.774) = -.012. \text{ Record in row 9, column 1.}$$

The sign of each residual is recorded in the upper half of its narrow sign cell. The diagonal for this column and for all succeeding columns is recorded in the space just above the double line on the data sheet. This leaves the diagonal cell vacant in each column.

24. Add the entries in column 1, including the diagonal, and record in column 1 for the row marked "Actual  $\Sigma_0$ " at the bottom of the data sheet. This sum should be zero or nearly zero. It is  $+.002$ .

25. The expected value of this sum, designated "Check  $\Sigma_0$ ," on the data sheet may be calculated for each column  $k$  by the formula,

$$r_k - a'_{k1}\Sigma E',$$

where  $r_k$  is the sum in row  $D$ , column  $k$  of *Table 1*,  $a'_{k1}$  is the first-factor loading at the top of column  $k$  in *Table 2*, and  $\Sigma E'$  has the value already recorded on *Table 2*.

EXAMPLE: For column 1, this check is

$$r_1 - a'_{11}\Sigma E' = 5.022 - .774(6.486) = +.002.$$

This agrees with the "Actual  $\Sigma_0$ " value for column 1 indicated in step 24.

The "Actual  $\Sigma_0$ " and "Check  $\Sigma_0$ " values are not always exactly the same as in this case, but their difference very seldom exceeds .003 when three decimals are used in the calculations.

26. Since the residual tables are all symmetric about the diagonal, the calculated entries in column 1 may be copied in their corresponding cells in row 1, i.e.,

$$r_{2.21} = r_{2.12} = -.021. \text{ Record in row 1, column 2.}$$

$$r_{2.31} = r_{2.13} = +.051. \text{ Record in row 1, column 3.}$$

$$r_{2.41} = r_{2.14} = +.003. \text{ Record in row 1, column 4.}$$

$$r_{2.51} = r_{2.15} = +.003. \text{ Record in row 1, column 5.}$$

$$r_{2.61} = r_{2.16} = -.037. \text{ Record in row 1, column 6.}$$

$r_{2.71} = r_{2.17} = +.013$  . Record in row 1, column 7 .

$r_{2.81} = r_{2.18} = -.024$  . Record in row 1, column 8 .

$r_{2.91} = r_{2.19} = -.012$  . Record in row 1, column 9 .

When this step is completed, all cells in row 1 and column 1 of *Table 2* are filled except the diagonal.

27. Add all the entries in row 1, including the diagonal, and record the sum in the column labeled "Actual  $\Sigma_0$ " at the right of the data sheet. This sum should agree with the sum in step 24. It is  $+.002$ . This check is valuable in working with a large number of variables; it is not necessary when  $n$  is twenty or less.

28. Calculate the residuals *in the diagonal and below it* for each column  $k$  of *Table 2* in the manner described in steps 22 and 23.

EXAMPLES:

Column 2:  $r_{2.j2} = r_{j2} - a'_{j1}a'_{21}$  .

$$r_{2.22} = +.592 - .650(.650) = (+.170) .$$

$$r_{2.32} = +.397 - .731(.650) = -.078 .$$

$$r_{2.42} = +.397 - .665(.650) = -.035 .$$

Column 3:  $r_{2.j3} = r_{j3} - a'_{j1}a'_{31}$  .

$$r_{2.33} = +.626 - .731(.731) = (+.092) .$$

$$r_{2.43} = +.472 - .665(.731) = -.014 .$$

$$r_{2.53} = +.626 - .804(.731) = +.038 .$$

29. As soon as the residuals for column  $k$  are computed below the diagonal, fill in the entries in the row for that test above the diagonal by symmetry, as described in step 26.

EXAMPLES:

Row 2:  $r_{2.32} = r_{2.23} = -.078$  . Record in row 2, column 3 .

$r_{2.42} = r_{2.24} = -.035$  . Record in row 2, column 4 .

. . . . .

$r_{2.92} = r_{2.29} = -.062$  . Record in row 2, column 9 .

Row 3:  $r_{2.43} = r_{2.34} = -.014$  . Record in row 3, column 4 .

$r_{2.53} = r_{2.35} = +.038$  . Record in row 3, column 5 .

. . . . .

$r_{2.93} = r_{2.39} = -.007$  . Record in row 3, column 9 .

**30.** Check the accuracy of the residual calculations for each column  $k$  by the methods described in steps 24 and 25; make this check for each row  $j$  by the method of step 27.

**EXAMPLES:**

Test 2:  $r_2 - a'_{21}\Sigma E' = 4.213 - .650(6.486) = - .003$  .  
 Actual  $\Sigma_0 = - .002$  .

Test 3:  $r_3 - a'_{31}\Sigma E' = 4.744 - .731(6.486) = + .003$  .  
 Actual  $\Sigma_0 = +.004$  .

$n$  tests:  $r_i - \sum_{j=1}^n a'_{ji} \sum_{k=1}^n a'_{ki} = 42.072 - 6.486(6.486) = + .004$  .  
 Actual  $\Sigma_0 = +.006$  .

**31.** *Table 2* should now have every entry filled except the diagonals, and all of the sums,  $\Sigma_0$ , should be recorded.

**32.** Pick the highest coefficient in each column, disregarding sign, and record it in the diagonal with positive sign. This sign should be in the upper half of the narrow sign cell.

**33.** Prepare a table similar to *Table 3* with variable numbers 1 to  $n$  at the top of the columns. Add a check column and one labeled " $k_i$ ." This table will be used to minimize the number of negative signs in each column of *Table 2* in order to determine the tests to be reflected.

**34.** Count the number of negative signs in each column of *Table 2* and record in row 1 and in the proper column of *Table 3*. These are the values,  $N_j$ , described in Example 5, *chapter iii*.

**EXAMPLES:** The number of negative signs in column 1 of *Table 2* is four; hence 4 is the entry in the first row and first column of *Table 3*.

Similarly, there are six negative signs in column 2 of *Table 2*; consequently the entry in column 2 of the first row of *Table 3* is 6.

**35.** Check these values by counting the number of positive signs in each column excluding the diagonal. The sum of the positive and negative signs in each column must be  $(n-1)$ .

**36.** Add all the entries in the first row of *Table 3* and record in the check column. This sum is 46.

37. Pick the test having the highest number of negative signs to be reflected first. Tests 2, 6, and 9 have a maximum of six negative signs. The choice of one of these is arbitrary; test 2 is chosen here.

38. Record the variable number 2 in the column headed  $k_i$  and put an "×" above column 2 in *Table 3* to indicate that this test is to be reflected. An adjustment in the number of negative signs for each column will be made as if test 2 were reflected in *Table 2*; these results will be recorded in row 2 of *Table 3*.

39. For the trait being reflected, i.e., test 2 in this case, the entry in its column of row 2, *Table 3*, will be  $(n-1)$  minus the number of negative signs for its column in row 1 of *Table 3*. The value of  $(n-1)$  is the total number of entries in each column of *Table 2*, ignoring the diagonals.

In this example,  $(n-1)=8$ , and the entry in column 2, row 2, of *Table 3* becomes  $8-6=2$ .

40. Proceed to that row of *Table 2* for the test being reflected, i.e., row 2, and consider the sign of each entry there except the diagonal.

a) If that entry for a given trait *not previously reflected* is *positive*, increase by one the number of negative signs for that column recorded in row 1 of *Table 3*, and record the new value in its proper column of row 2, *Table 3*.

EXAMPLES: The entry for test 6 in row 2 of *Table 2* is *positive*. Test 6 has *not been previously reflected*. The number of negative signs for test 6 recorded in row 1 of *Table 3* is six. Consequently, 6 is increased one, giving 7 as the entry in row 2, column 6, of *Table 3*.

The entry for test 8 is also *positive* in row 2 of *Table 2*. Test 8 has *not been previously reflected*. In the same manner, its number of negative signs is increased one, giving 6 as the new value in row 2, column 8, of *Table 3*.

b) If the entry for a given test *not previously reflected* is *negative*, decrease the number of negative signs for that column by one and record the new value in its proper column of row 2, *Table 3*.

EXAMPLES: The entry for test 1 in row 2 of *Table 2* is *negative*. Test 1 has *not been previously reflected*. The number of negative signs for test 1 recorded in row 1 of *Table 3* is four. Hence, 4 is decreased one, giving 3 as the entry for test 1 in row 2 of *Table 3*.

The entries for tests 3, 4, 5, 7, and 9 are all negative in row 2 of *Table 2*. Since none of these tests has been previously reflected, the number of negative signs for each of them will be reduced one. This gives the entries 4, 4, 4, 3, and 5 for these respective tests in row 2 of *Table 3*.

All of the cells in row 2 of *Table 3* should now be filled.

41. Add all of the entries in row 2 of *Table 3*. This sum is 38.

42. If the sum of all the entries in row 1 of *Table 3* (from step 36) minus



the sum of all the entries in row 2 of *Table 3* (from step 41) is twice the difference between the number of negative signs for the reflected trait 2 in these two rows of *Table 3*, the arithmetical work in deriving row 2 of *Table 3* is checked.

EXAMPLE:  $46 - 38 = 2(6-2)$  .

$$8 = 8 .$$

43. Pick the test having the highest number of negative signs in row 2 of *Table 3* as the next test to be reflected. This is test 6 with a maximum of seven negative signs.

44. Record the variable number 6 in column  $k_i$  of *Table 3* and put an "X" above column 6 to indicate that this test is to be reflected.

45. The entry in column 6, row 3, of *Table 3* will be  $(n-1)$  minus the entry for test 6 in row 2 of *Table 3*, i.e.,  $8-7=1$ .

46. Proceed to row 6 of *Table 2* and consider the sign of each entry there except the diagonal in order to adjust the number of negative signs of row 2, *Table 3*, as if test 6 were reflected. These new values will become row 3 of *Table 3*.

a) For the tests which have *not been previously reflected* the same rules of adjustment of number of negative signs apply as in steps 40a and 40b, except that the adjustment is made with reference to row 2 instead of row 1 of *Table 3*.

EXAMPLES: The entry for trait 4 is *positive* in row 6 of *Table 2*. Trait 4 has *not been previously reflected*. Trait 4 has four negative signs in row 2 of *Table 3*. Hence, it will have  $4+1=5$  negative signs recorded in row 3 of *Table 3*.

The entries for tests 1, 3, 5, 7, 8, and 9 are *negative* in row 6 of *Table 2*. *None* of these tests has been *previously reflected*. Hence, the entries for these respective columns in row 2 of *Table 3* are each reduced one. This gives 2, 3, 3, 2, 5, and 4 as the entries for columns 1, 3, 5, 7, 8, and 9, respectively, in row 3 of *Table 3*.

b) When an entry in row 6, *Table 2*, is *positive* and the test has been *previously reflected* in this table, *decrease* by one the number of negative signs for that trait as recorded in row 2 of *Table 3*, and record the new value in its column of row 3, *Table 3*.

EXAMPLE: The entry for test 2 is *positive* in row 6, *Table 2*. Test 2 was *previously reflected*. The number of negative signs for test 2 in row 2 of *Table 3* is two. Hence, the entry for test 2 in row 3 of *Table 3* becomes  $2-1=1$ .

This gives the nine entries in row 3 of *Table 3*.

47. Add the entries in row 3 of *Table 3*. This sum is 26.

48. If the sum of the entries in row 2 of *Table 3* (in step 41) minus the sum of the entries in row 3 of *Table 3* (in step 47) is twice the difference between the entries in these rows for the test being reflected, i.e., test 6 in this case, the arithmetical work in deriving row 3 of *Table 3* is checked.

EXAMPLE:  $38 - 26 = 2(7-1)$ .

$$12 = 12.$$

49. Pick the trait having the highest number of negative signs in row 3 of *Table 3* as the next one to be reflected.

The maximum number of negative signs in row 3 of *Table 3* is five for tests 4 and 8. Either one may be reflected; test 4 is arbitrarily chosen here.

50. Write the variable number 4 in column  $k_i$  of row 3, *Table 3*, and put an "×" above column 4 to indicate that test 4 is to be reflected.

51. The entry for trait 4 in row 4 of *Table 3* will be  $(n-1)$  minus the number of negative signs for trait 4 in row 3 of *Table 3*.

EXAMPLE:  $8-5=3$ .

52. Proceed to row 4 of *Table 2* and consider the sign of each entry there except the diagonal, in order to adjust the number of negative signs of row 3, *Table 3*, as if test 4 were reflected. These new values will become row 4 of *Table 3*.

a) For the tests which have *not been previously reflected*, the same rules of adjustment of number of negative signs apply as in steps 40a and 40b, except that the adjustment is made with reference to row 3 instead of row 1 of *Table 3*.

EXAMPLES: The entries for the *unreflected* tests 1 and 7 in row 4 of *Table 2* are *positive*. Hence, their entries of row 3, *Table 3*, are each increased one and recorded in row 4 of *Table 3*. This gives the entry 3 for each of the columns 1 and 7 in row 4 of *Table 3*.

The entries for the *unreflected* tests 3, 5, 8, and 9 are *negative* in row 4 of *Table 2*. Hence, their entries in row 3 of *Table 3* are each reduced one, giving the entries 2, 2, 4, and 3, respectively, for these variables in row 4 of *Table 3*.

b) The entry for the *previously reflected* test 6 is *positive* in row 4 of *Table 2*. By the method of step 46b, its value in row 4 of *Table 3* becomes one less than its value in row 3 of *Table 3*.

EXAMPLE:  $1-1=0$ =entry for test 6 in row 4 of *Table 3*.

c) When the entry for a *previously reflected* test is *negative* in row 4 of *Table 2*, *increase* by one the number of negative signs for that test as recorded

Table I  
Matrix  $R_0$

$E'$	No.										$\Sigma E'$	Section	
		1	2	3	4	5	6	7	8	9			
.	1	.625	.482	.617	.518	.625	.422	.584	.563	.586		+ 5.022	
.	2	.482	(.592)	.397	.397	.410	.592	.398	.505	.440		+ 4.213	
.	3	.617	.397	(.626)	.472	.626	.356	.593	.500	.557		+ 4.744	
.	4	.518	.397	.472	(.541)	.520	.440	.541	.480	.409		+ 4.312	
.	5	.625	.410	.626	.520	(.763)	.322	.591	.594	.763		+ 5.214	
.	6	.422	.592	.356	.440	.322	(.592)	.363	.426	.332		+ 3.845	
.	7	.584	.398	.593	.541	.591	.363	(.593)	.592	.532		+ 4.787	
.	8	.563	.505	.500	.480	.594	.426	.592	(.633)	.633		+ 4.926	
.	9	.586	.440	.557	.403	.763	.332	.532	.633	(.763)		+ 5.009	
$\Sigma E'$													
Check $\Sigma_0$	B												
Actual $\Sigma_0$	D												
	D	+ 5.022	+ 4.213	+ 4.744	+ 4.312	+ 5.214	+ 3.845	+ 4.787	+ 4.926	+ 5.009			44.072
	E	+ .774	+ .650	+ .731	+ .665	+ .804	+ .593	+ .738	+ .759	+ .772			6.486
	K	+ .774	+ .650	+ .731	+ .665	+ .804	+ .593	+ .738	+ .759	+ .772			6.486
													$\sqrt{E}$
													$\sqrt{K}$

in row 3 of *Table 3* and record the new value in its column of row 4, *Table 3*.

EXAMPLE: The entry for the *previously reflected* test 2 is *negative* in row 4 of *Table 2*. The number of negative signs for test 2 in row 3 of *Table 3* is one. This value is increased one to give 2 as the entry for test 2 in row 4 of *Table 3*.

This gives the nine entries in row 4 of *Table 3*.

53. Add all the entries in row 4 of *Table 3* and record in the check column. This sum is 22.

54. If the sum of all the entries of row 3, *Table 3* (from step 47), minus the sum of all the entries for row 4, *Table 3* (from step 53), is twice the difference between the entries in these rows for the test being reflected, i.e., test 4, the arithmetical work in deriving row 4 is checked.

EXAMPLE:  $26 - 22 = 2(5-3)$ .

$$4 = 4 .$$

55. Zero entries sometimes appear in residual tables, such as *Table 2*. They are treated as of *positive* sign in making sign adjustments for the reflection of tests.

56. It sometimes happens that a test already reflected in a table may appear a second time (or any even number of times) as the test having the maximum number of negative signs. In this case, it is reflected back to its original position in the configuration by reversing each of the rules enumerated in steps 40a, 40b, 46b, and 52c.

EXAMPLE: For purposes of illustration only, reflect test 2 a second time *as if* it had had a maximum number of negative signs in row 4 of *Table 3*. The number of negative signs for each test would then be those of Row 5 in the following table:

	Tests									CHECK	$k_i$
	1	× × 2	3	× 4	5	× × 6	7	8	9		
Row 4.....	3	2	2	3	2	0	3	4	3	22	2
Row 5.....	4	6	3	2	3	1	4	3	4	30	6
Row 6.....	5	5	4	3	4	7	5	4	5	42	2
Row 7.....	4	3	3	4	3	8	4	5	4	38	6
Row 8.....	3	2	2	3	2	0	3	4	3	22	.....



Table 3

1	× 2	3	× 4	5	× 6	7	8	9	Check	$k_i$
4	6	5	5	5	6	4	5	6	46	2
3	2	4	4	4	7	3	6	5	38	6
2	1	3	5	3	1	2	5	4	26	4
3	2	2	3	2	0	3	4	3	22	.....

Again, for purposes of illustration, reflect test 6 a second time as if it had had a maximum number of negative signs in row 5 above. The sign adjustment recorded in row 6 is obtained in the same manner as for test 2 recorded in row 5, except to note that for the entry in row 6, column 2, of *Table 2*, the rule must be adjusted to take account of the fact that test 2 is now in its *unreflected* form, since it has been previously reflected twice. The sign of the entry in row 6, column 2, of *Table 2* is positive; hence rule 40a is the one to be reversed. This gives the entry 5 in row 6, column 2, of the foregoing table.

In case a test is reflected a third (or any odd number of times), the rules, 40a, 40b, 46b, and 52c apply directly by considering each test *previously reflected* an *even* number of times as an *unreflected* test, and each test *previously reflected* an *odd* number of times as a test *previously reflected once*.

**EXAMPLE:** Row 7 of the foregoing table gives the adjustment in number of negative signs of each column as though test 2 were reflected a third time. The rules apply directly except for the entry in column 6; test 6 has been *previously reflected twice*, so that it is considered as an *unreflected* test having a positive sign in row 2, *Table 2*; rule 40a then applies directly, and the entry for test 6 in row 7 becomes 8.

When test 6 is then reflected a third time, the rules apply directly, except for the entry for test 2, which has been *previously reflected three times*. Test 2 is then considered as a test *previously reflected once*, and rule 46b applies. The results are shown in row 8 of the foregoing sign table.

*The cases discussed in steps 55 and 56 do not occur in our present calculations.*

**57.** All entries in row 4 of *Table 3* are now equal to or are less than  $\frac{n-1}{2} = 4$ . Test 8 has four negative signs, which just balances the number of positive signs, ignoring the diagonal. All the other tests have a majority of positive signs. The tests listed in column  $k_i$  of *Table 3* are now ready to be reflected in *Table 2*.



Table 5

1	2	3	× 4	5	6	7	× 8	× 9	Check	$k_i$
5	4	5	6	5	3	3	4	5	40	4
4	3	4	2	4	2	2	5	6	32	9
3	2	3	1	5	1	1	6	2	24	8
2	1	2	0	4	2	2	2	1	16	.....

58. Reverse the signs of the tests indicated in column  $k_i$  of *Table 3* in their rows of *Table 2*. Record these signs in the lower half of the narrow sign cell for each entry involved.

EXAMPLE: Reverse all the signs of entries in rows 2, 4, and 6 of *Table 2*.

The entries in row 2 then have the signs +, -, +, +, +, -, +, -, +, in the lower half of the narrow sign cell. This corresponds to the "second position" of the signs described in Example 5, *chapter iii*.

59. Indicate that the signs in rows 2, 4, and 6 have been changed by reversing the signs before the variable numbers 2, 4, and 6 in the second column of *Table 2*. Record these new signs in the lower half of the narrow sign cell.

60. Reverse all of the signs in the columns for the tests being reflected and record these signs in the residual cells in front of each residual involved. Where two signs appear in the narrow sign cell for any entry, it is the sign in the "second position" that is reversed.

EXAMPLE: Reverse the signs of all entries in columns 2, 4, and 6. The entries in column 2 then become +, +, +, -, +, +, +, -, +.

61. Indicate that the signs in columns 2, 4, and 6 have been changed by reversing the signs before the variable numbers 2, 4, and 6 in the second row of *Table 2*. Record the new sign in the lower half of the narrow sign cell. These signs, after reflection for these variables, are the ones which will be transferred to the next residual table, i.e., *Table 4*.

62. Copy the last recorded sign for each entry in the columns representing the unreflected tests. Place these signs in the residual cells in front of each coefficient.

EXAMPLE: These are columns 1, 3, 5, 7, 8, and 9 in these data. By this procedure the signs of the residuals of column 1 are +, +, +, -, +, +, +, -, -, as indicated in *Table 2*.





Table 7

Test <i>j</i>	I	II	III
1	.774	.070	.094
2	.650	-.367	-.218
3	.731	.167	.184
4	.665	-.118	.261
5	.804	.306	-.011
6	.593	-.485	-.071
7	.738	.126	.167
8	.759	.039	-.163
9	.772	.291	-.255

63. Determine the sum of each column of *Table 2*, ignoring the diagonals, and record in row *B* at the bottom of the data sheet.

EXAMPLES: Test 1 = +.086 .

Test 2 = +.516 .

Test 9 = +.430 .

This step is useful only if it is desired to reflect tests until a maximum positive sum is secured. Steps 63 and 64 may be combined in cases where it is satisfactory to minimize the number of negative signs in each column of *R* without demanding a maximum positive sum.

64. Add the diagonal value for each column to the sum for that column in row *B*. These are the sums +.137, +.723, . . . , +.572 in row *D* at the bottom of *Table 2*.

65. Add all the entries in row *D*. This sum is 3.873.

66. Add all of the entries in each row of *Table 2*, including the diagonal, and record in column *D* at the extreme right of the data sheet. The row and column sums *D* for the same test should agree.

67. Add all the entries in column *D*. This sum is 3.873, which agrees with the sum in step 65. This is the value  $r'_{2i}$  of equation (19-iii).

68. The multiplier  $\frac{1}{\sqrt{r'_{2i}}}$  is obtained in the same manner as described in steps 11 and 12 for *Table 1*. Its value is .508132.

69. Multiply each sum in row *D* by .508132 and record in row *E*. These are the second-factor loadings,  $a''_{k2}$  of equation (19-iii). The signs of these loadings are those of the reflected variables.

EXAMPLES:  $a'_{12} = .508132(+.137) = +.070$  .

$a'_{22} = .508132(+.723) = +.367$  .

. . . . .

$a'_{92} = .508132(+.572) = +.291$  .

70. Check: Add all the entries in row  $E$  and compare with  $\sqrt{r'_{2i}}$ .

EXAMPLE:  $\Sigma E = 1.969$  .

$\sqrt{r'_{2i}} = 1.968$  .

71. Copy the values of row  $E$  in row  $K$ , with sign reversed for each test which has been reflected an odd number of times; i.e., for each test which has the last recorded sign negative before its variable number.

EXAMPLE: Tests 2, 4, and 6 have the last recorded sign negative before their variable numbers, since each test has been reflected once. Consequently, their loadings in row  $K$  take signs opposite to those in row  $E$ .

72. The sum of the loadings in row  $K$  should be approximately zero if a centroid system has been obtained. This sum for *Table 2* is  $+ .029$ .

73. Copy the values of row  $K$  as the second column of *Table 7*. These are the second-factor loadings of the unreflected tests.

74. Take a new data sheet and label it "Second-Factor Residual Coefficients:  $r_{3,jk}$ " as shown in *Table 4*.

75. Proceed, as in steps 18 through 73, to determine the second-factor residual coefficients,  $r_{3,jk}$ , and the third-factor loadings. These calculations are shown in *Table 4*.

Tests 4, 8, and 9 were reflected in *Table 4*. The sign table is shown in *Table 5*.

76. The third-factor residual coefficients,  $r_{4,jk}$ , shown in *Table 6*, are sufficiently small to ignore. Consequently a fourth factor was not determined.

77. *Table 7* shows the projections of the nine tests of this example on the three centroid axes obtained. This is the  $n \times r$  matrix  $F$  of the fundamental factor theorem  $FF' = R$ .

## APPENDIX II

### A METHOD OF FINDING THE ROOTS OF A POLYNOMIAL

Consider a polynomial of the type (14-iv),

$$f(\beta) = c_0\beta^r + c_1\beta^{r-1} + c_2\beta^{r-2} + \dots + c_{r-1}\beta + c_r = 0,$$

where  $r$  is a positive integer,  $c_0 \neq 0$ , and  $c_0, c_1, \dots, c_r$  are real coefficients. The  $r$  roots,  $\beta$ , of this equation are desired. Determine the upper and lower limits\* of the roots of this equation. Let a trial value of  $\beta$  within these limits be  $\beta'$ . If  $\beta'$  is a root of the polynomial  $f(\beta)$ , then by the Remainder Theorem,†  $f(\beta') = 0$ . The numerical value of  $f(\beta')$  may be determined on an electric calculating machine by computing  $f(\beta)/(\beta - \beta')$  by the process of synthetic division.‡ Consider the sign of the numerical value of  $f(\beta')$  and select a second trial root, designated  $\beta''$ , which will give  $f(\beta'')$  opposite in sign to that of  $f(\beta')$ . When two such trial values of  $\beta$  are found, there is at least one root\*\* between them. Determine a third trial value,  $\beta'''$ , by linear interpolation between  $f(\beta')$  and  $f(\beta'')$ . If the value  $f(\beta''') = 0$ , then  $\beta'''$  is one of the  $r$  roots of the polynomial. If  $f(\beta''') \neq 0$ , interpolate for successive trial values until that value of  $\beta$  is found for which the remainder,  $f(\beta)$ , is zero to as many decimals as required.

Repeat this process for each of the  $r$  roots of the polynomial by taking trial values in other regions between the upper and lower limits of the roots. In the method of principal axes of *chapter iv*, a very useful first approximation for each root,  $\beta_m$ , of the characteristic equation is  $-\sum_{j=1}^n a_{jm}^2$  for each column  $m$  of  $F$ , when  $c_0, c_1, \dots, c_r$  are all positive.

*Table 2* shows the application of the method of synthetic division in calculating one of the roots of equation (27-iv),

$$\beta^4 + 6.965369\beta^3 + 10.810494\beta^2 + 5.407203\beta + .840052 = 0.$$

The upper limit of the roots of this equation is any *positive* number; the lower limit is  $-7.965369$ . Since the equation for this example is from the

\* L. E. Dickson, *First Course in the Theory of Equations* (New York: John Wiley & Sons, 1922), pp. 21-23.

† *Ibid.*, p. 12.

‡ *Ibid.*, pp. 13-15.

\*\* *Ibid.*, p. 67.

method of principal axes, the first trial value  $\beta' = -5.011317 = -\sum_{j=1}^n a_{j1}^2$

from *Table (6-iv)*. Since  $f(\beta')$  is negative,  $\beta'' = -5.020000$  was chosen arbitrarily to secure a positive value of  $f(\beta'')$ . Linear interpolation between  $f(\beta')$  and  $f(\beta'')$  gave the third trial value,  $\beta''' = -5.019710$ . Linear interpolation between  $f(\beta''')$  and  $f(\beta'')$  gave the fourth trial value,  $\beta'''' = -5.019712$ , for which the value of the polynomial is  $+ .000018$ . Hence, one root of this equation is  $-5.019712$ .

The values of  $f(\beta)$  in each row of *Table 2* were determined by the equations in their corresponding rows of *Table 1*. In actual application, it is possible to carry out the  $r$  calculations in each row on a calculating machine without recording any of the values except the  $r$ th one in the column headed  $f(\beta)$ .

Table 1

Trial	$\beta^4$	$\beta^3$	$\beta^2$	$\beta$	$f(\beta)$	Trial $\beta$
	$c_0$	$c_1$	$c_2$	$c_3$	$c_4$	
1	.....	$c_1 + c_0\beta' = c_1'$	$c_2 + c_1\beta' = c_2'$	$c_3 + c_2\beta' = c_3'$	$c_4 + c_3\beta' = c_4'$	$\beta'$
2	.....	$c_1 + c_0\beta'' = c_1''$	$c_2 + c_1'\beta'' = c_2''$	$c_3 + c_2'\beta'' = c_3''$	$c_4 + c_3'\beta'' = c_4''$	$\beta''$
3	.....	$c_1 + c_0\beta''' = c_1'''$	$c_2 + c_1''\beta''' = c_2'''$	$c_3 + c_2''\beta''' = c_3'''$	$c_4 + c_3''\beta''' = c_4'''$	$\beta'''$
4	.....	$c_1 + c_0\beta'''' = c_1''''$	$c_2 + c_1''' \beta'''' = c_2''''$	$c_3 + c_2''' \beta'''' = c_3''''$	$c_4 + c_3''' \beta'''' = c_4''''$	$\beta''''$

Table 2

Trial	$\beta^4$	$\beta^3$	$\beta^2$	$\beta$	$f(\beta)$	Trial $\beta$
	1.0	+6.965369	+10.810494	+5.407203	.840052	
1	.....	1.954052	1.018120	.305081	-.688806	-5.011317
2	.....	1.945369	1.044742	.162598	+.023810	-5.020000
3	.....	1.945659	1.043850	.167379	-.000142	-5.019710
4	.....	1.945657	1.043856	.167347	+.000018	-5.019712

### APPENDIX III

#### A METHOD OF DETERMINING THE SQUARE ROOT ON THE CALCULATING MACHINE

Newton's\* iterative method of determining square roots may be used very advantageously on the calculating machine, using Barlow's Tables to determine the first trial value of the square root.

Let  $N$  be the number whose square root is desired, and let  $x_0$  be the first trial value for the  $\sqrt{N}$  derived from Barlow's Tables. Determine  $N/x_0$  on the calculating machine, and record it.

Compute a new trial value,  $x_1 = .5(x_0 + N/x_0)$  by determining the cumulative sum of the two products,  $(.5x_0)$  and  $[.5(N/x_0)]$ . Determine  $N/x_1$ . If the two values  $x_1$  and  $N/x_1$  are the same to as many decimals as desired, then  $\sqrt{N} = x_1 = N/x_1$ . If these two values differ, this process may be repeated for as many trials as are necessary to find that trial value  $x$  which agrees with  $N/x$  to the required number of decimals.

EXAMPLE		
$N$		= 42.072 .
$x_0$	= $\sqrt{42.07}$ in Barlow's Tables	= 6.4861391 .
$\frac{N}{x_0}$		= 6.48644738 .
$x_1$	= $.5x_0 + .5\left(\frac{N}{x_0}\right)$	= 6.48629324 .
$\frac{N}{x_1}$		= 6.48629324 .

\* E. T. Whittaker and G. Robinson, *The Calculus of Observations* (London: Blackie & Son, 1926), pp. 79-80.