

FACTOR ANALYSIS AND RELATED STATISTICAL TECHNIQUES

By

Masil Bryan Danford

Institute of Statistics
Mimeo. Series No. 70
For Limited Distribution

June, 1953.

Factor Analysis and Related Statistical Techniques

by

MASIL BRYAN DANFORD

A thesis submitted to the Faculty of the North Carolina State College in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Experimental Statistics.

Raleigh

1953

Approved by

.....

Adviser

MASIL BRYAN DANFORD. Factor Analysis and Related Statistical Techniques.
(under the direction of H. FAIRFIELD SMITH).

The purpose of this thesis is two-fold: 1) To establish a rapport between psychologists and statisticians by defining terms and phrases used by psychologists in factor analysis in standard statistical language, and 2) To compare and contrast factor analysis as a statistical technique with related statistical procedures. The first three chapters deal with the first problem. Terms and symbols that are to be used are defined. The postulated factor equation is given and discussed. Two of the more common solutions are given, one being considered an approximate solution to the other. The psychologist's rationale for "factoring" a correlation matrix is sketched with discussion of Spearman's and Thurstone's theories. One chapter is devoted to the clarification of various kinds of correlation.

In Chapter IV, factor analysis is compared to the analysis of variance. Classification and identification problems are discussed in Chapter V. It was concluded that factor analysis is most closely allied to canonical correlation analysis, and that the problem set by the psychologists may be solved by this technique or by analogous methods.

Under specified moderately mild conditions, it is shown, in Chapter VI, how factor loadings may be predicted for a test not included in an original factored battery of tests without re-factoring the correlation matrix with the new test added.

Chapter VII includes a discussion of tests of significance in factor analysis and the estimation of factor loadings by maximum likelihood as given by Lawley.

A summary of the conclusions reached is given in Chapter VIII.

Biography

- Born:** May 22, 1923 at Nineveh, Leon County, Texas
- Parents:** George Frederick and Cordelia Ruth Scott Danford
- Married:** June 14, 1952 to Inez Mary Sherman
- Degrees:** B. A. (1947) with honors in mathematics from
The University of Texas, Austin, Texas.
M. A. (1948) in mathematics from
The University of Texas, Austin, Texas.
- Societies:** Phi Beta Kappa, Phi Delta Kappa, Phi Kappa Phi
- Service:** An officer in the U. S. Navy, 1943-1946.
- Fellowships:** Teaching Fellow in mathematics, University of Texas,
Austin, Texas 1948-1949.
Graduate Assistant with the Institute of Statistics,
North Carolina State College, Raleigh, 1950-1952.
National Science Foundation Fellowship, 1952-1953.
- Employment:** United States Air Forces' School of Aviation Medicine,
Biometrics Department, Randolph Field, Texas, June 1949-
August 1950, June 1951-September 1951.

ACKNOWLEDGMENT

This is to express appreciation to the members of my thesis committee:

To Professors R. L. Anderson, D. J. Moffie, and Wasily Hoeffding for their helpful criticisms and assistance,

To Professor George E. Nicholson for suggesting the thesis topic and for continued aid and encouragement,

And to the chairman of my committee, Professor H. Fairfield Smith.

Particular thanks are extended for his critical reading of the manuscript, for his invaluable assistance, and for his many suggestions for clarity of exposition.

The research for and writing of the thesis was made possible by a National Science Foundation Fellowship for the academic year 1952-1953.

TABLE OF CONTENTS

	<u>Page</u>
BIOGRAPHY	ii
ACKNOWLEDGEMENT	iii
LIST OF TABLES	vi
LIST OF FIGURES	vii
 <u>Chapter</u>	
I. INTRODUCTION	1
1.1 Statement of the problem	1
1.2 Historical background	1
II. GENERAL FACTOR ANALYSIS EQUATION AND SOLUTION . . .	11
2.1 Definitions and geometrical representations .	11
2.2 Centroid and principal axis solutions	26
2.3 A critique of Spearman's one-factor theory and Thurstone's multiple factor theory	38
III. CORRELATION, RELIABILITY, AND COMPONENTS OF VARIANCE	55
3.1 Purpose of Chapter	55
3.2 Intra-class and inter-class correlations and reliability coefficients	55
3.3 What to factor: common parts and steady parts of raw scores	65
3.4 Reliability and communality	75
3.5 Correlations corrected for attenuation	78
3.6 Reliability and length of test	80
IV. FACTOR ANALYSIS AND ANALYSIS OF VARIANCE	81
4.1 Univariate analysis of variance	81
4.2 Multivariate analysis of variance and canonical correlation analysis	96
V. DISCRIMINATION AND IDENTIFICATION PROBLEMS	113
VI. PREDICTION OF FACTOR LOADINGS	128

VII. ESTIMATION OF FACTOR LOADINGS AND TESTS OF SIGNIFICANCE	141
VIII. SUMMARY AND CONCLUSIONS	153
BIBLIOGRAPHY	159
APPENDIX A: GLOSSARY OF TERMS AND SYMBOLS	162

LIST OF TABLES

<u>Table</u>	<u>Page</u>
4.1. Correlations Among 4 Tests	81
4.2. Principal Axis Factor Loadings	82
4.4. Analysis of Variance of Unitary Standardized Test Scores	84
4.5. Analysis of Variance of Unitary Standardized Test Scores In Terms of Sample Correlation Coefficients	87

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
2.1. Geometrical Representation of N Points in $n = 2$ Space . .	12
2.2. Geometrical Representation of $n = 6$ Points in $N = 2$ Space	13
2.3. n Tests Configuration in $N = 3$ Space	15
2.4. Factor Loading as Projection on Common Factor Axis . . .	41
2.5. Factor Loadings as Projection on 2 Common Factor Axes . .	51

Chapter I

INTRODUCTION

1.1 Statement of the problem

The statistical technique of factor analysis has been developed for the most part by psychologists. This thesis is an attempt to establish a rapport or rapprochement between statisticians and psychologists. The terms and concepts that psychologists use in factor analysis will be defined in statistical language. Then analogies will be drawn between factor analysis as a statistical technique and other related statistical analyses.

Terminology and notation being rather extensive, a glossary with definitions of terms as used in this thesis is presented in Appendix A to serve as an aide-memoire.

To establish perspective, a historical sketch will be given.

1.2 Historical background.

It is not easy to pinpoint an idea, but it is probably safe to say that factor analysis sprang from the works of Galton and Pearson.

Galton (1888) is credited with the concept of concomitant variation between two "variable organs". This concept was eventually defined as a calculable quantity and was regarded as the "consequence of variations of the two organs being partly due to common causes". In recent years correlational analysis has fallen somewhat into disrepute among statisticians because of misinterpretations; however, due credit should be given these early workers. The average, of one form or another, has long been

considered as a statistic that in some way describes, or at least summarizes, an important part of an aggregate of values of a single variable. Such description was amplified by a measure of variation or dispersion. The next logical development was to provide a measure of concomitant variation among two or more variables. These descriptive statistics are all attempts at parsimony of expression as an aid to reasoning.

If there are N measurements on n variable organs, the degree of association of corresponding observations on each individual may be summarized by $\frac{1}{2}n(n - 1)$ correlation coefficients. The information provided by these statistics consists of a number of figures which form only $(n - 1)/2N$ th of the original nN observations. But if n is large, the simultaneous consideration of $\frac{1}{2}n(n - 1)$ correlation coefficients remains virtually impossible. By the turn of the century, Pearson was talking of the need for "multivariate correlation" techniques for studying the inter-relationship of several organs. Then multiple and partial correlation coefficients were evolved. However, even with these at hand, comprehensive description of the data may remain a hopeless tangle. Still more powerful methods were needed for a joint or multivariate analysis of several variables.

These early biometric workers were concerned with factors of evolution. In an attempt to bring some order out of chaos, they hit upon the idea of classification of types of individuals. The classical problem that arose at this time was to classify and identify criminals recorded by Scotland Yard, utilizing anthropometric measurements. It was noticed that many of the traits were correlated. Those criminals who were tall tended

to have long legs; if they were large around the chest, they were likely to be large around the neck. In order to reduce the number of necessary measurements, use was to be made of these high inter-correlations. It was suggested that perhaps all traits or characters were superficially related by some common causes or factors. Edgeworth proposed that the n correlated measures be reduced to n uncorrelated measures. He considered first the bivariate case. If the two correlated traits are normally distributed, the resulting equal frequency contours are ellipses. He suggested that the principal axes of this ellipse be calculated by well-known rules. The principle could be extended to any number of traits. Back of all these considerations was the implied hope that the number of uncorrelated variates needed to explain in some sense (or merely to describe) the n correlated variates would be fewer than n . Pearson (1901), following up these various suggestions, gave the first general formulation of the problem. He would pass through the swarm of points the line of closest fit. (See figure 2.1, Chapter II). Since there were no dependent and independent variates in the usual sense, distances perpendicular to the fitted line would be squared and the sum minimized. The problem of transforming n correlated variates to n uncorrelated variates was made determinate by agreeing to choose the longest axis of the ellipsoid of points at each successive stage. For the two tests i and j of figure 2.1, the first axis would be approximately in the position of the dashed line PA_1 ; the second axis, perpendicular to the first, would be in the approximate position of PA_2 . That is, the line of best fit was passed through the swarm of joints along the line of longest variation. Today, factor analysts speak of "explaining away" the maximum amount of variation at each

successive factoring. This solution of the problem was given explicitly by Hotelling in 1933 and has come to be known as the principal component or principal axis solution of the factor problem. Pearson was not able to solve the Scotland Yard problem because of the terrific amount of computation involved. In later years (Burt (1949)) it has been factored and re-factored. The question remains as to what these uncorrelated factors are, or whether any meaning can be ascribed to them other than as descriptive statistics; but more important than that, have we come any nearer sufficiently describing the universe with economy?

Data for seven bodily measurements were finally compiled by the criminologists: head length, head breadth, face, foot, forearm, height, and finger. A recent re-factorization by Burt (1949) shows the first factor to be a general body factor. The second factor is a bi-polar factor, head versus body-limb measurement. The other factors are not so easily interpreted. Together, the first two factors account for about 75% of the total variation, and the first 2 or 3 factors account for nearly all the inter-correlations. The remaining factors must be a factorization of the errors of the tests. How does this help in the classification of criminals? There is no reduction in labor of taking the 7 observations on each individual, but if only 2 indices of body types are recorded instead of the 7 correlated measures, there is a real saving. And if the researchers seek to correlate other attributes with body types, they have to correlate only with 2 indices instead of 7. It might be concluded, then, that the factor analysis technique has some merit when used for classification. That is not to say that problems should be handled in this fashion, however.

But back to the historical development. With this background in

biometric research, the psychologists enter the scene. Here, it is believed, are tools, techniques, and concepts that may be used in a study of human behavior. There had been attempts to describe the mind or mental functions in terms of faculties. But these ideas had gone out of fashion, most psychologists considering it impossible to describe the mind in terms of a disjoint set of faculties.

But by then mental testing had become popular. It had, further, been observed that the results of various mental tests were positively correlated. Correlational analysis seemed to be indicated. Here was a tool or technique made to order. So we enter the phase of psychology which attempted to summarize everything by correlation coefficients. In some quarters this persists even today. But some researchers finally asked; what is being measured? What are the causal effects? True, the test scores are related, but why and by what mechanism? If it could be determined through these instruments or tests what the causal relations are, then perhaps it would be possible to approximately describe mental processes after all.

Spearman championed the idea of a general intellectual factor which he called "g". The test scores were conceived as consisting of two parts, the part due to "g" and a unique part peculiar to the test. (In figure 2.4, Chapter II the part due to the common factor g is the length of the projection of the test vector onto the common factor axis, X_g). A common factor is an ability, faculty, or trait common to two or more tests. Spearman noticed that if the results from n tests were correlated and set forth in a correlation matrix, the elements in different columns (or rows) of this matrix seemed to be approximately proportional. That is, deter-

minants of minors of order 2 x 2 were approximately zero. This is the same as saying that the "effective" rank of the correlation matrix was one. Thus if the correlation coefficients could be measured without error, under the hypothesis of one common factor, the rank of the correlation matrix would be exactly one. Spearman and his workers observed that certain tests upset the proportionality. After much re-testing, if this persisted on the part of a given test, it was discarded. This was considered permissible since the primary objective of his inquiry was to construct a battery of tests that would measure pure g, and the non-proportionality of, say, the jth column was taken to mean that test j was measuring some factor already included as the specific part in other tests. Detailed studies of the tests were made to see what the unique parts of the tests were and why they were not measuring the common factor g. Some psychologists allowed these tests that upset the proportionality to remain in the battery. The hypothesis was altered to allow for "group factors", i.e. factors common only to certain sub-sets of tests.

Thurstone, in this country, developed what is known as multiple factor analysis. Instead of postulating in advance the number of factors to be evaluated, he proposed to assess the number of common factors needed to describe a battery of tests by ascertaining the matrix of minimum rank $r \leq n$, say, which could be fitted to the observed correlation matrix within the limits of random sampling.

For the common "g" theory as well as for the multiple factor theory, the diagonal elements of the correlation matrix are to be determined so as to preserve proportionality or rank r. The off-diagonal elements would be used to determine the rank and the diagonal elements adjusted accordingly.

The substitution of reduced correlations for the diagonals has not been universally accepted. Hotelling (1942) considers it nonsensical. The correlation of a test with itself is one. For an $n \times n$ sample correlation matrix, the rank is n and no less. But this criticism has not stopped the practice of substituting reduced correlations in the diagonal. A critique of Spearman's and Thurstone's theories will be discussed in the third section of chapter II.

The question might be asked: If one is given a battery of n tests with all the inter-correlations, how can one decide how many common factors are present? There is no completely satisfactory answer. Bartlett (1950) has derived some large sample tests of significance. These tests of significance promise to be helpful in answering the question. Bartlett hastens to add that statistical or numerical significance does not imply the existence of real or meaningful factors; on the other hand, factors not statistically significant may yet be psychologically real. Analogous ambiguity, of course, exists in interpreting all significance tests. One treatment may give higher yields than another, but economically the difference may not be large enough to matter. On the other hand, a real difference between two treatments may exist which is extremely important, but the sample size may not be large enough to detect it.

Probably the most important question of all in factor analysis is whether or not the factors extracted are meaningful and interpretable. The answer to this is not a decided yes or no. It seems to depend on the problem to be solved. For the Scotland Yard problem, there seemed to be two body indices that could be used for classification, a real saving. How about the results of a factor analysis applied to mental tests aimed at

measuring intelligence? The extraction of several uncorrelated common factors seems to help understand the structure of mental processes. Factors like numeric ability, verbal ability, and depth perception emerge. The psychologists now speak not just of intelligence but of the factors that go to make up intelligence. Two people may be rated as equally intelligent, but they very probably wouldn't have equal numeric and verbal abilities, say. Does this help the psychologists in mapping the mind, as it were? The psychologists say yes. If so, then it is an acceptable scientific technique.

It is interesting to note the change of emphasis during the development of these techniques. Galton and Pearson were concerned with classification. Spearman came along, adapted these concepts to a study of mental processes and tried to define or measure "g". In the multiple factor type of analysis, the problem is that of description. There is little or no attempt to classify the people who take the tests, nor do the factor analysts try to measure the factors. The factor analysts today want to locate and isolate what they hope are basic parameters or common factors that give rise to a matrix of correlation coefficients. This shift in thinking has been gradual and subtle. Or is there a difference at all in the three problems? Surely there is a common thread running through them.

In this thesis, factor analysis will be compared with other types of analysis to see whether or not the problem to be solved may be attacked by other means. To do this, however, general terminology will need to be defined and a general factor analysis equation will need to be set-up.

This will be done in section 1 of chapter II. In section 3, the two most common types of factor analysis will be discussed, the centroid method of factoring and the principal component method. In section 3 of the same chapter, a critique of the work of Spearman and Thurstone will be given.

In Chapter III, a detailed analysis of various types of correlation will be given. In certain cases, it will be shown how the correlations may be estimated by variance components. The chapter has a dual purpose, to clarify existing confusion in psychological literature and to try to ascertain what should be "factored", assuming the psychologists' general factor equation.

With this background material, factor analysis will be compared with the analysis of variance in chapter IV. The first part will deal with univariate analysis of variance and the second part, multivariate analysis and canonical correlation.

Chapter V has to do with the general subject of discriminatory analysis. It is conjectured that the factor problem may be approached through this technique.

Factor analysts frequently want to know how a new test will behave when placed in a factored battery of tests, without re-factoring the battery with the new test added. In chapter VI it is shown how the factor loadings for the new test may be estimated.

The estimation of factor loadings by maximum likelihood is discussed in chapter VII, along with tests of significance in factor analysis.

General conclusions are given in chapter VIII.

In a discussion of factor analysis as a statistical technique, Kendall

Chapter II

DEFINITIONS AND GENERAL FACTOR ANALYSIS SOLUTIONS

2.1 Definitions and geometrical representations.

In this thesis, "test" and "test score" are generic terms meaning any kind of measurement, whether mental, biological, or physical made on an individual; "individual" refers to a person, skull, or any object on which all test scores are measured. N shall refer to the number of individuals in a sample and n to the number of tests. "Raw" test scores (i.e. as observed) will be represented by $Y_{i\alpha}$, where the first subscript refers to the test, the Greek subscript to the individual; $i = 1, 2, \dots, n$; $\alpha = 1, 2, \dots, N$; that is, n tests are scored on each of a sample of N individuals. Usually test scores are recorded as deviations from the sample mean for each test, $Y_{i\alpha} - \bar{Y}_i$, or "standardized" as

$$s_{i\alpha} = (Y_{i\alpha} - \bar{Y}_i) / \sqrt{\hat{\sigma}_{ii}}$$

where \bar{Y}_i is the sample mean for test i ; i.e.,

$$\sum_{\alpha=1}^N Y_{i\alpha} / N,$$

and $\hat{\sigma}_{ii}$ is the sample variance defined as $\sum_{\alpha} (Y_{i\alpha} - \bar{Y}_i)^2 / (N - 1)$.

(Psychologists more commonly evaluate variance as the sample second moment, $m_2 = \sum_{\alpha} (Y_{i\alpha} - \bar{Y}_i)^2 / N$; but the definition used here will be more suitable when comparisons are made to the analysis of variance). Hence $\sum_{\alpha} s_{i\alpha}^2 = N - 1$. Note that "standardized" is defined differently from statistical usage, the usual definition being $(Y_{i\alpha} - \eta_i) / \sqrt{\sigma_{ii}}$ where $\eta_i = E(\bar{Y}_i)$. Written without the circumflex, σ_{ii} will, as usual in

statistics, refer to the population variance; σ_{ij} , $\hat{\sigma}_{ij}$ denote respectively the population and sample covariances between tests i and j with analogous definitions. For some discussions, the statistic $s_{i\alpha}^* = s_{i\alpha} / \sqrt{N-1}$ $= (Y_{i\alpha} - \bar{Y}_i) / \sqrt{(N-1) \hat{\sigma}_{ii}}$ will be used. The $s_{i\alpha}^*$ will be referred to as being of "unitary standard measure", since $\sum_{\alpha} (s_{i\alpha}^*)^2 = 1$. Except when otherwise qualified, "variance" will denote the sample variance of a single test as defined above. "Vector", as usual, shall mean a line in space with indicated magnitude and direction.

One geometrical picture of the nN observations is obtained by imagining n orthogonal test axes. The test scores made by the individuals may then be plotted as N points in n -space. For two tests, the geometrical representation is shown in Figure 2.1, the familiar scatter-diagram, where each of the N points represents an individual's score on tests i and j .

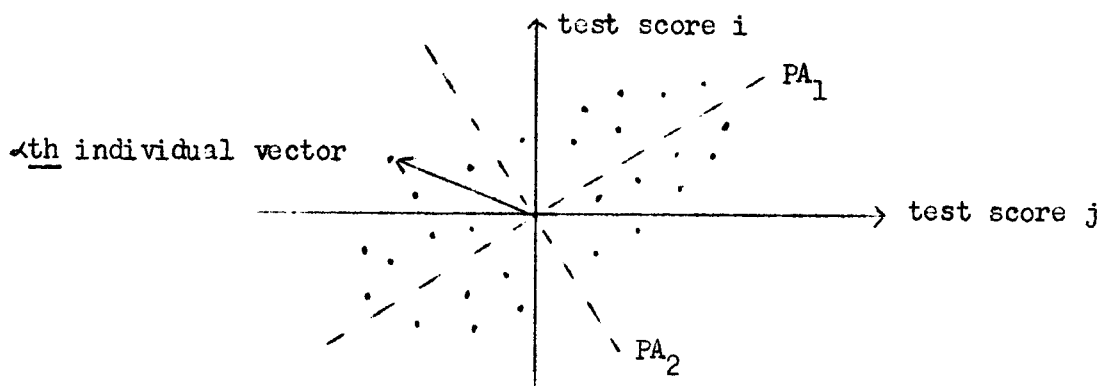


Figure 2.1

When the test scores are measured from their sample means, the origin will be at the mean of the two tests. If the two tests are approximately normally distributed and correlated, the swarm of points has an elliptical shape. For the general picture, there are n tests and thus n dimensions. The swarm of points is in general ellipsoidal. In figure 2.1,

a vector may be drawn from the origin to each point. The two elements of the vector for the α th individual are the coordinates of the point in reference to the two test axes. For the n-dimensional picture, there are N individual vectors, each having n elements or coordinates.

Another geometrical picture arises when the individuals are used as the reference orthogonal axes. Since there are N individuals, this diagram will have N dimensions and n points or vectors. As for figure 2.1, test scores will usually be measured from their sample means, and since $\sum_2 (Y_{i\alpha} - \bar{Y}_i) = 0$, this implies that the n points are restricted to lie on an N-1 dimensional hyperplane passing through the origin and so inclined that no point can have all coordinates positive or all negative. Figure 2.2 shows a diagram for N = 2, n = 6, where the restriction implies that all points lie on a line with slope -1.

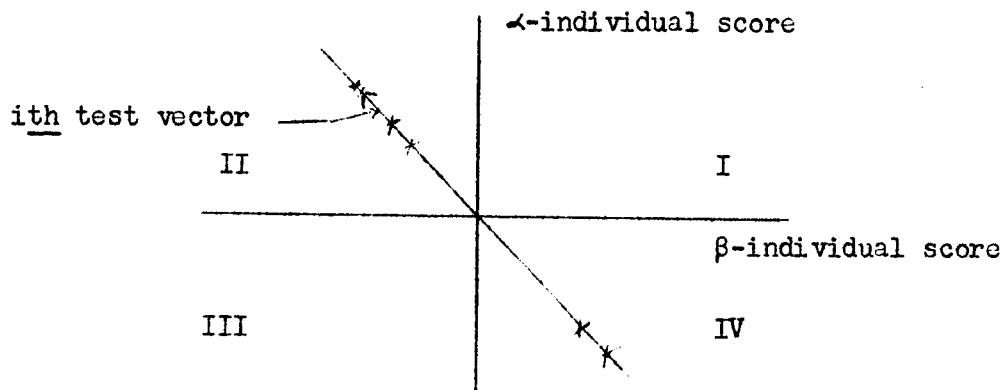


Figure 2.2

In this diagram, the points will not usually be symmetrically distributed about the origin; e.g. if in figure 2.2 individual α is superior to β on all tests, then all the points will lie in quadrant II. The length of test vector i (for the N-dimensional picture) is $\sqrt{(N-1) \hat{\sigma}_{ii}}$ (see any text on solid analytic geometry). If in figure 2.2, the $s_{i\alpha}$ are plotted, then the points all fall on the intersection of a circle with radius $\sqrt{N-1}$ and the

straight line with slope -1. Although this figure is initially conceived as in N dimensions, since invariably n will be less than N , all points will lie on an $n (< N-1)$ dimension hyperplane, which, owing to the above restriction, passes through the origin. If unitary standardized scores $s_{i\alpha}^*$ are plotted, then every vector has unit length and the n points lie on the intersection of that hyperplane with an $(N-1)$ dimensional hypersphere with center at the origin.

Since the direction cosines of any vector i with reference to the individual axes are $(Y_{i\alpha} - \bar{Y}_i) / \sqrt{\sum (Y_{i\alpha} - \bar{Y}_i)^2} = s_{i\alpha}^*$, it follows that the angle θ_{ij} between any two test vectors i and j is such that

$$\cos \theta_{ij} = \frac{\sum (Y_{i\alpha} - \bar{Y}_i)(Y_{j\alpha} - \bar{Y}_j) / \sqrt{\sum (Y_{i\alpha} - \bar{Y}_i)^2 \sum (Y_{j\alpha} - \bar{Y}_j)^2}}{\sqrt{\sum (Y_{i\alpha} - \bar{Y}_i)^2 \sum (Y_{j\alpha} - \bar{Y}_j)^2}} = \sum s_{i\alpha}^* s_{j\alpha}^* = r_{ij},$$

where r_{ij} is the sample correlation coefficient between tests i and j .

Note that altering the scale of any test alters all N coordinates for that test point proportionally and so affects only the length of the vector and not its direction. Hence the angles between vectors are invariant under change of scale. "Test vector" will refer to a vector as given in figure 2.2 on either standardized or unstandardized scales. The group of n test-vectors so drawn (i.e. with the angles between them representing the sample correlation coefficients of pairs of tests), and taken without reference to any particular set of axes is termed the "configuration" of the sample. Later, the configuration will be referred to other axes in varying numbers of dimensions (figures 2.4 and 2.5, section 3 of this chapter).

If individuals tend to make similar relative scores on a group of two or more tests (i.e., if there are correlations between the tests of the group), the θ_{ij} are small and the result is a group of vectors with

similar direction; the test points tend to lie in the same segment of the diagram, or for similarly scaled tests or standardized scores, to form a "cluster" of points on the surface of the hyper-sphere. Such clustering was in fact observed for experimental data. This led to the idea of trying to represent n correlated tests by $r < n$ composite variates, and then, by a rotation and change of origin, by a number of uncorrelated variates. For example if $N = 3$, $n = 5$, the points lie on a circle and the result may be like figure 2.3, drawn on the plane $\sum_{i=1}^3 s_{i\alpha} = 0$.

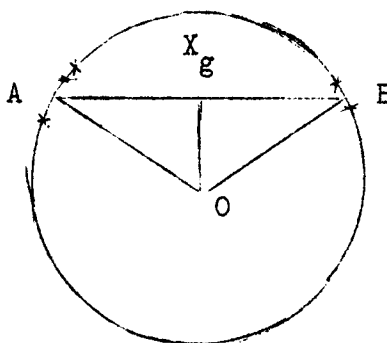


Figure 2.3

This picture results from plotting the sample points on the 3 individual axes analogous to figure 2.2. The points lie on a circle: The intersection of a sphere with radius $\sqrt{N-1}$ and the plane $s_{i1} + s_{i2} + \dots + s_{iN} = 0$ for any i . If the variation of the points within each cluster may be ascribed to random errors of sampling, one can postulate that each cluster represents the same characteristic, and insofar as the two characteristics indicated are correlated (as represented by the angle between the mean vectors of each cluster). They can be represented by a common part \overline{OX}_g , and two specific parts \overline{AX}_g and \overline{X}_gB .

In general, of course, the picture is more complex since $N > n$ and the test points lie in n dimensions. But the general problem of factor

analysis can be stated in terms of figure 2.3: Can one find r clusters or factors, $r < n$, such that the n set of test scores can be represented by only r vectors and such that most of the variation about these may be considered as ascribable to random fluctuations of the observations. And secondly, may it be convenient to further split these average clusters into orthogonal components (analogous to resolution of forces)?

A "battery" of tests is a given group of n tests. A test score is postulated to consist of several parts. There is the "common" part which gives rise to the correlation between tests. The "specific" part of a test score is the part peculiar to a given test which is not common to the other tests in a given battery. That is, the common and specific parts of a test are relative. For example, one spelling test in a battery of verbal tests would have a specific part peculiar to spelling ability; but if another spelling test were included with the first in the same battery, then the specific part would be common to the specific part of the second spelling test. The third part of a test score is a random error term. The sum of the specific part and the random error term is referred to as the "unique" part of the test score.

Spearman's one factor theory postulates that every test score is made up of two parts; i.e., $s_{i\alpha} = g_{\alpha} + u_{i\alpha}$, and that $u_{i\alpha} = (s_{i\alpha} - g_{\alpha})$ and g_{α} are mutually uncorrelated for all pairs of tests. g_{α} is the effect of the common factor and $u_{i\alpha}$ is the unique part of the test score. If several common factors are postulated, then one can assume the model

$$s_{i\alpha} = g_{\alpha} + A_{1\alpha} + A_{2\alpha} + \dots + A_{r-1,\alpha} + u_{i\alpha}'$$

where the $A_{p\alpha}$ represent the parts common to two or more of the $u_{i\alpha}$ above.

This equation is analogous to the usual models for experimental designs. The comparison will be examined more carefully below. Thurstone noted that for such a model, the rank of the hypothetical matrix of correlations, apart from the diagonal, would be r .

Assume there are r common factors in a battery of tests and $M-r$ specific parts. At least one specific part will be assumed to exist in each test score and no specific part exists in more than one test, otherwise it would become a common factor. Then the standardized test score may be written as

$$\begin{aligned}
 s_{i\alpha} &= \sum_{p=1}^r \alpha_{ip} x_{p\alpha} + \sum_{p=r+1}^M \alpha_{ip} x_{p\alpha} + \epsilon_{i\alpha} \\
 &= \sum_{p=1}^r a_{ip} X_{p\alpha} + \sum_{p=r+1}^M a_{ip} X_{p\alpha} + e_{i\alpha}. \quad (2.1)
 \end{aligned}$$

The α_{ip} ($p \leq r$) measures the contribution of common factor p to the scores on test i ; α_{ip} ($p > r$) measures the contribution of the p th specific part to the test i ; $\alpha_{i,r+1}$, $\alpha_{i,r+2}$, ... will be zero for all but one value of i . (note that α is used for both the factor "loading", and as a subscript denoting person). $x_{p\alpha}$ measures the ability of the α th individual to indicate the existence of factor p ; $X_{p\alpha}$ is the value estimated from the sample and is coded such that $\sum_{\alpha=1}^N X_{p\alpha} = 0$. Since α_{ip} for $p > r$ is zero except for one value of i , it seems reasonable to set

$$\sum_{p=r+1}^M \alpha_{ip} x_{p\alpha} = \pi_{i\alpha}$$

where $\pi_{i\alpha}$ is the contribution of the specific part to the test score; the

sample estimate is denoted by $f_{i\alpha}$. $\epsilon_{i\alpha}$ measures the random error and is not distinguishable from $\pi_{i\alpha}$ unless the i th test is administered more than once to the α th individual. The notation $\psi_{i\alpha} = \pi_{i\alpha} + \epsilon_{i\alpha}$ will be used to indicate the unique part, the sum of the specific part and the error term. In (2.1), $e_{i\alpha}$ represents the deviation of the fitted or regression value from the observed value. For the sample estimates, denote $u_{i\alpha} = f_{i\alpha} + e_{i\alpha}$. Equation (2.1) may then be written as

$$\begin{aligned} s_{i\alpha} &= \sum_{p=1}^r \alpha_{ip} X_{p\alpha} + \pi_{i\alpha} + \epsilon_{i\alpha} \\ &= \sum_{p=1}^r a_{ip} X_{p\alpha} + f_{i\alpha} + e_{i\alpha} \end{aligned} \tag{2.2}$$

or

$$\begin{aligned} s_{i\alpha} &= \sum_{p=1}^r \alpha_{ip} X_{p\alpha} + \psi_{i\alpha} \\ &= \sum_{p=1}^r a_{ip} X_{p\alpha} + u_{i\alpha}. \end{aligned} \tag{2.3}$$

In psychological literature, α_{ip} of (2.2) and (2.3) is called the factor "loading" of test i on factor p or the amount of "saturation" test i has on factor p . In section 3 of this chapter, it will be shown that a_{ip} is the projection of test vector i onto the common factor axis, X_p . In the same way as X is sometimes spoken of as a random variable varying along an axis, or as a specific value measuring the position on the axis of a particular observation, so the term "common factor" will be used ambivalently to denote either the axis, X_p , or the position on the axis (= amount

of a factor possessed by a particular person), $\chi_{p\alpha}$.

The equations in (2.2) and (2.3) appear to be simple regression equations. If the $\chi_{p\alpha}$ were known, the α_{ip} could be estimated as regression coefficients in the usual least squares fashion. Also, if the α_{ip} were known, the person constants $\chi_{p\alpha}$ could be estimated by regression. But neither the α_{ip} nor the $\chi_{p\alpha}$ are known. The problem is to estimate both of them. The $s_{i\alpha}$ may be considered as "yields" in a two-way classification, n tests by N persons. The usual experimental model postulates an additive model: for the i th test on the α th person, the test score or "yield" would consist of a test effect, τ_i , say, a person effect, ρ_α , and an interaction, $(\tau\rho)_{i\alpha}$ and an error term, $\epsilon_{i\alpha}$. Thus $s_{i\alpha}$ would be written as

$$s_{i\alpha} = \tau_i + \rho_\alpha + (\tau\rho)_{i\alpha} + \epsilon_{i\alpha}.$$

If the experiment is not replicated, the $(\tau\rho)_{i\alpha}$ and $\epsilon_{i\alpha}$ cannot be separated. But as will be shown in Chapter III, the person "effect" ρ_α is not constant from test to test. The point is that the test score or "yield" is not composed of additive effects; instead, it is rather like an "interaction" between person and tests. As the model (2.2)(or(2.3)) implies, the test score will tend to be high if both the saturation of the test in factor p is large (α_{ip} is large) and if the person component $\chi_{p\alpha}$ is large. This will be discussed more in Chapter III.

Let $F_a = [\bar{a}_{ip}]$ be the $n \times r$ matrix of factor loadings; $X = [\bar{\chi}_{p\alpha}]$, the $r \times N$ matrix of factors, $U = [\bar{u}_{i\alpha}]$, the $n \times N$ matrix of unique parts, and $S = [\bar{s}_{i\alpha}]$, the $n \times N$ matrix of standardized test scores. Equations

(2.3) may be written as

$$S = F_a X + U \quad (2.4)$$

Recall that the test scores are standardized so that $SS' = (N-1)R$, where R is the usual correlation matrix. The factors are also coded so that $\sum X_{p\alpha} = 0$, $\sum X_{p\alpha}^2 = (N-1)$. The common factors are assumed uncorrelated in the population and the common factors are so chosen that for the sample

$$\sum X_{p\alpha} X_{m\alpha} = \begin{cases} N-1, & m=p \\ 0, & m \neq p \end{cases}.$$

This will be proved in section 2 of this chapter. It is also assumed that in the population, the common factors are uncorrelated with the unique parts. Also, for the sample, the common factor axes are not only chosen orthogonal to each other, but orthogonal to the unique axes, $\sum (X_{p\alpha}, u_{i\alpha}) = 0$. If one multiplies (2.4) by its transpose

$$\begin{aligned} SS' &= (F_a X + U) (F_a X + U)' \\ &= F_a XX' F_a' + UU'. \end{aligned}$$

Since

$$XX' = (N-1) I,$$

where I is the identity matrix,

$$SS' = (N-1) R = (N-1) F_a F_a' + UU',$$

or

$$R = F_a F_a' + \frac{1}{N-1} UU'.$$

The sample correlation coefficient is a "consistent" estimate; that is

$$r_{ij} \xrightarrow[N \rightarrow \infty]{} \rho_{ij},$$

where ρ_{ij} is the population correlation between tests i and j . Thus, as $N \rightarrow \infty$, $R \rightarrow R^*$, where

$$R^* = \begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1n} \\ \rho_{21} & 1 & \dots & \rho_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \rho_{n1} & \rho_{n2} & \dots & 1 \end{bmatrix}$$

(Note: r_{ij} is a biased estimate of ρ_{ij} for $E(r_{ij}) \neq \rho_{ij}$. This applies also to the standardized variates $s_{i\alpha}$. Thus, expected values will not be taken. Instead, the concept of consistency will be used).

Let

$$R_a^* = \begin{bmatrix} H_1^2 & \rho_{12} & \dots & \rho_{1n} \\ \rho_{21} & H_2^2 & \dots & \rho_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \rho_{n1} & \rho_{n2} & \dots & H_n^2 \end{bmatrix}$$

and

$$R_u^* = \begin{bmatrix} \sigma^2(u_1) & & & 0 \\ & \sigma^2(u_2) & & \\ & 0 & \cdot & \\ & & \cdot & \\ & & & \sigma^2(u_n) \end{bmatrix},$$

so that

$$R^* = R_a^* + R_u^*$$

H_i^2 is called the population "communality" of test i ; it is the population variance of

$$\left(\sum_{p=1}^r a_{ip} x_{p\alpha} \right),$$

under the assumption of r common factors. $\sigma^2(u_i)$ is the population unique variance of test i ; i.e., $\sigma^2(u_i) = \text{pop. var. of } (y_{i\alpha})$. Thus, $H_i^2 + \sigma^2(u_i) = 1$.

Let

$$R_a = \begin{bmatrix} h_1^2 & r_{12} & \dots & r_{1n} \\ r_{21} & h_2^2 & \dots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \dots & h_n^2 \end{bmatrix}, \quad R_u = \begin{bmatrix} \sigma^2(u_1) & & & \\ & \widehat{\sigma^2(u_2)} & & 0 \\ & & \ddots & \\ 0 & & & \ddots \\ & & & & \widehat{\sigma^2(u_n)} \end{bmatrix}$$

so that $R = R_a + R_u$. h_i^2 is the sample estimate of H_i^2 and $\widehat{\sigma^2(u_i)}$ is the sample estimate of $\sigma^2(u_i)$. It is postulated that to a first degree of approximation

$$F_a F_a' \simeq R_a, \quad \frac{1}{N-1} U U' \simeq R_u, \quad (2.5)$$

where " \simeq " stands for "is approximately equal to". The factor analysts extract $r < n$ common factors until, say $r_{ij}^* = r_{ij} - \sum_{p=1}^r a_{ip} a_{jp}$ is "small". This is one of the main problems in factor analysis: When are the residual correlations, r_{ij}^* , negligible? There have been various empirical rules given: One might stop factoring when all $|r_{ij}^*| < .03$, say. Or one

may use the criterion of significant factors: After r common factors are extracted, if

$$\sum_{i=1}^n a_{i, r+1}^2$$

is "small", then the $(r+1)^{\text{st}}$ factor is not significant. This criterion seems to be adaptable to rigorous tests of significance. (See Chapter VII for a discussion of tests of significance).

Since

$$(N-1) h_i^2 = \sum_{\alpha} \left(\sum_{p=1}^r a_{ip} X_{p\alpha} \right)^2 = \sum_{p=1}^r a_{ip}^2 \sum_{\alpha} X_{p\alpha} X_{p\alpha} = (N-1) \sum_{ip} a_{ip}^2$$

and

$$\begin{aligned} \sigma^2(u_i) &= \sum_{\alpha} u_{i\alpha}^2 / (N-1), \\ h_i^2 + \sigma^2(u_i) &= 1. \end{aligned} \quad (2.6)$$

And from (2.5)

$$r_{ij} = \sum_{p=1}^r a_{ip} a_{jp}, \quad i \neq j. \quad (2.7)$$

If it is assumed that the specific part is uncorrelated with the error term, then

$$\begin{aligned} \text{Population variance of } (f_i) &= \text{Pop. var. of } (n_{i\alpha} + \epsilon_{i\alpha}) \\ &= \text{Pop. var. of } (n_{i\alpha}) + \text{Pop. var. of } (\epsilon_{i\alpha}) \\ \text{or} & \\ \sigma^2(u_i) &= \sigma^2(f_i) + \sigma^2(e_i) \end{aligned} \quad (2.8)$$

where

$$\sigma^2(f_i) = \text{Pop. var. of } (\pi_{i\alpha})$$

and $\sigma^2(e_i) = \text{Pop. var. of } (\epsilon_{i\alpha}).$

The estimates, $\widehat{\sigma^2}(f_i)$ and $\widehat{\sigma^2}(e_i)$ are obtainable only by replicating the same test on the same individual.

It will be shown in section 3 how the h_i^2 are obtained. Using the calculated quantity h_i^2 , $\widehat{\sigma^2}(u_i)$ may be obtained by subtraction, using relationship (2.6.).

If equation (2.3) is written as if all parts of the test are common parts (i.e. without the unique parts), then

$$s_{i\alpha} = \sum_{p=1}^n \beta_{ip} X_{p\alpha} = \sum_{p=1}^n b_{ip} X_{p\alpha}. \quad (2.9)$$

Again, the common factors are chosen to be uncorrelated over the sample, and using $S = FX$, where $F = [\bar{b}_{ip}]$, an $n \times n$ matrix of factor loadings, then

$$SS' = FXX'F'$$

or

$$(N-1)R = (N-1)FF'.$$

Thus for n common factors, FF' will equal the sample correlation matrix exactly. The last few factors are obviously a factorization of the error term, since for the sample, the $e_{i\alpha}$ are correlated.

The factor analysis problem may be stated, then, as finding a matrix of factor loadings, F_a , such that $F_a F_a' \simeq R_a$ or a matrix of factor loadings,

F , such that $FF' = R$, depending on whether one postulates the model of (2.3) or the model of (2.9). There are an infinity of solutions for F_a and F . The problem is made determinate by agreeing to choose for the j th factor axis the largest axis of the ellipsoid of the swarm of points; i.e., maximum variation is accounted for by each successive factor. The principal component solution, which is to be discussed in the next section, is unique. Various approximate solutions, e.g. the centroid solutions are not. As will be mentioned, the factors extracted depend on how the tests are "reflected". However, many psychologists extract all "significant" factors, then "rotate" the new factor axes to "meaningful psychological structure". The problem of rotation will not be discussed in this thesis, for it is essentially a psychological, not a statistical problem. Many psychologists maintain that for certain purposes, the correlation matrix to be factored is R_a . This is called the "reduced" correlation matrix; that is, the diagonal values are reduced from unity to h_i^2 . Finally, the "steady part" of a test score should be defined. That part of the test score which remains steady when the test is given to the same person time and again is called the steady part. Since the sample of individuals is usually assumed to be random, the "steady part" is a random variable, for it differs from one individual to the next. In terms of equation (2.2), the "steady part" is

$$\omega_{i\alpha} = \sum_{p=1}^r \lambda_{ip} x_{p\alpha} + \pi_{i\alpha}$$

and the sample estimate is

$$c_{i\alpha} = \sum_{p=1}^r a_{ip} X_{p\alpha} + f_{i\alpha}.$$

Denote the population variance of $\omega_{i\alpha}$ by $\sigma^2(c_i)$, and the population variance of $\varepsilon_{i\alpha}$ by $\sigma^2(e_i)$. And since the error is assumed to be uncorrelated with the common and specific parts, the population variance of test i is $\sigma^2(c_i) + \sigma^2(e_i)$. The reliability coefficient of test i , ρ_{ii} , is defined to be the ratio of the variance of the steady part to the total variance of the test; thus

$$\rho_{ii} = \frac{\sigma^2(c_i)}{\sigma^2(c_i) + \sigma^2(e_i)} \quad (2.11)$$

Other terms and phrases needed in the discussion will be defined as they are used.

2.2 Centroid and Principal Axis Solutions.

The centroid method of factoring is one of many approximate methods. Since it is one of the most popular, it will be sketched first. The centroid method will be followed by the principal axis solution, which will be treated in detail, since nearly all other approximate methods of factoring are, admittedly, approximations to this solution.

The centroid method of factoring is a simple summation technique which involves summing the columns of the sample correlation matrix. Recall that if it is postulated that a test score consists only of common parts, the test score may be written as

$$s_{i\alpha} = \sum_{p=1}^n \beta_{ip} x_{p\alpha}, \quad = \sum_{p=1}^n b_{ip} x_{p\alpha},$$

where the summation runs from 1 to n instead of 1 to r. Then

$$\begin{aligned} r_{ij} &= \sum_{\alpha=1}^N s_{i\alpha} s_{j\alpha} / (N-1) \\ &= \sum_{p=1}^n b_{ip} b_{jp}. \end{aligned}$$

(Note: This procedure works equally well for $r < n$ common factors and with communalities in the diagonal, if it is assumed that $r_{ij} = \sum_{p=1}^r a_{ip} a_{jp}$ exactly. For this exposition, however, all parts are assumed common factors so that $r = n$). Recall that $F = \sqrt{b_{ip}}$ and that $FF' = R$, exactly. Thus r_{ij} is the inner product of the ith row of F with the jth row. Summing r_{ij} over i and j ,

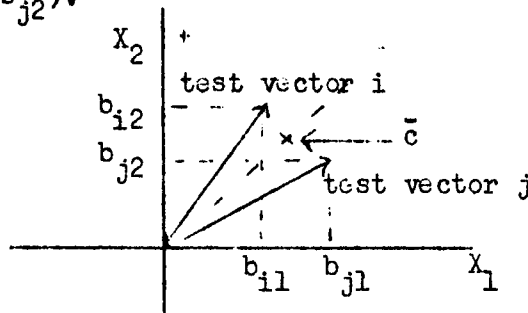
$$\begin{aligned} \sum_i \sum_j r_{ij} &= \sum_i \sum_j \left[\sum_p b_{ip} b_{jp} \right] \\ &= \sum_i b_{i1} \sum_j b_{j1} + \sum_i b_{i2} \sum_j b_{j2} + \dots + \sum_i b_{in} \sum_j b_{jn}. \end{aligned} \tag{2.13}$$

And since

$$\begin{aligned} \sum_i b_{ip} &= \sum_j b_{jp} \\ \sum_i \sum_j r_{ij} &= \left(\sum_i b_{i1} \right)^2 + \left(\sum_i b_{i2} \right)^2 + \dots + \left(\sum_i b_{in} \right)^2. \end{aligned} \tag{2.14}$$

In the test configuration (figure 2.2 and discussion) the "common factor" axes are placed such that the b_{ip} are the projections of the test vectors i onto the common factor axes p (See figures 2.4 and 2.5 in section 3 of this chapter). The common factor axis for the first factor X_1 , is placed such that the "centroid" lies on the first axis. The "centroid" is the center of gravity for the n -dimensional test configuration. For two tests and two common factors, the centroid, \bar{c} , is the point

$$\frac{1}{2} (b_{i1} + b_{j1}), \quad \frac{1}{2} (b_{i2} + b_{j2}).$$



The first common factor axis is so chosen that the coordinates of the centroid on this axis are

$$\frac{1}{n} \sum b_{i1}, \quad \frac{1}{n} \sum b_{i2} = 0, \quad \dots \quad \frac{1}{n} \sum b_{in} = 0,$$

i.e., all coordinates except one are zero.

Using this result in (2.14)

$$\sum_i \sum_j r_{ij} = \left(\sum_i b_{i1} \right)^2,$$

so that

$$\sum_i b_{i1} = \sqrt{\sum_i \sum_j r_{ij}}.$$

But

$$\begin{aligned} \sum_j r_{ij} &= b_{i1} \sum_j b_{j1} + \dots + b_{in} \sum_j b_{jn} \\ &= b_{i1} \sum_j b_{j1} \\ &= b_{i1} \sum \sum r_{ij}. \end{aligned}$$

Thus the "loading" of the ith test for the first factor is

$$b_{i1} = \sum r_{ij} / \sqrt{\sum \sum r_{ij}}. \quad (2.15)$$

By summing the columns of the correlation matrix, the loadings for all the tests on the first factor may be obtained. The "effect" of the first factor is eliminated by subtracting from r_{ij} the product of the loadings for tests i and j . The residual correlation is $r_{ij}^* = r_{ij} - b_{i1} b_{j1}$. It is easily shown that $\sum \sum r_{ij}^* = 0$. This is due to the restriction that the centroid has zero projections on the remaining $n - 1$ orthogonal axes. The second and successive factor loadings are obtained in the same way as the first factor loadings, but since $\sum \sum r_{ij}^* = 0$, expressions like (2.15) for the b_{i2} would not work. To circumvent this difficulty, the signs of certain tests are changed by a process called "reflection". One chooses the tests to reflect according to the comparisons he wishes to make. If there is no outside criterion, there are various rules laid down. A generally accepted rule is to reflect as many rows (and columns) of the residual correlation matrix as needed so as to make the total sum

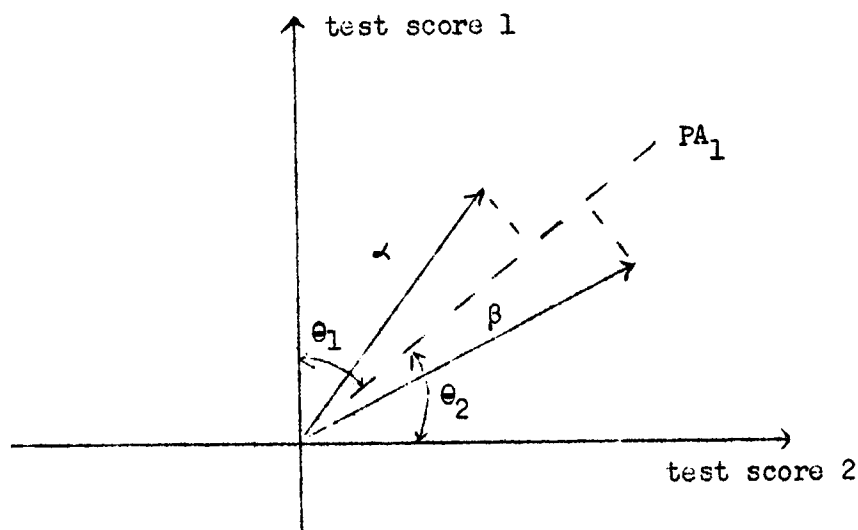
of the residual matrix a maximum. Then one proceeds as before by summing columns. After the loadings b_{i2} are obtained, the loadings corresponding to the tests that were reflected are made negative to account for this reflection. It is admittedly arbitrary as to which tests should be reflected. This will not be discussed here. For further details of this technique, see Thurstone (1948, Ch. VIII) or Thompson (1948, Ch. V.).

The principal axis solution was expanded and demonstrated by Hotelling in 1933. It is referred to as Hotelling's principal axis solution in this country, and in England, simply as the principal axis solution, which is usually credited to Pearson.

Consider the unitary standardized test scores $s_{i\alpha}^*$ as plotted in figure (2.1). The scores for the α th individual for the n tests are

$$s_{1\alpha}^*, s_{2\alpha}^*, \dots, s_{n\alpha}^*$$

Let l_{ip} be the direction cosines for the p th principal axis in relation to the test axis. A new axis is to be determined such that the sum of squares of the projections of the N n -coordinate test vectors onto this axis will be a maximum. (Or, equivalently, that the sum of squares of directions of the test points from the line shall be a minimum.) For two persons, α and β , say, and for two tests, the geometrical picture for the first principal axis is as follows, with $l_{11} = \cos \theta_1$, $l_{21} = \cos \theta_2$.



The squared length of the projection of individual vector α onto the line of best fit is $(l_{11} s_{12}^* + l_{21} s_{2\alpha}^*)^2$, or in general for n tests,

$$\left(\sum_{i=1}^n l_{i1} s_{i\alpha}^* \right)^2.$$

If the necessary restriction $\sum l_{i1}^2 = 1$ is imposed, the problem is to maximize, for all N individuals

$$\phi = S_{\alpha} \left(\sum_i l_{i1} s_{i\alpha}^* \right)^2 - \lambda \sum_i l_{i1}^2.$$

Then

$$\begin{aligned} \frac{\partial \phi}{\partial l_{i1}} &= 2 S_{\alpha} \left(\sum_j l_{j1} s_{j\alpha}^* \right) s_{i\alpha}^* - 2\lambda l_{i1} \\ &= 2 \sum_j l_{j1} S_{\alpha} s_{j\alpha}^* s_{i\alpha}^* - 2\lambda l_{i1} \end{aligned}$$

$$= 2 \sum_j l_{j1} r_{ij} - 2\lambda l_{i1}.$$

When all the partials of ϕ with respect to the l_{i1} are set equal to zero, there results the set of homogeneous equations

$$\begin{aligned} l_{11} (1-\lambda) + l_{21} r_{12} + \dots + l_{n1} r_{1n} &= 0 \\ l_{11} r_{12} + l_{21} (1-\lambda) + \dots + l_{n1} r_{2n} &= 0 \\ &\vdots \\ l_{11} r_{1n} + l_{21} r_{2n} + \dots + l_{n1} (1-\lambda) &= 0, \end{aligned} \tag{2.16}$$

or in matrix notation,

$$(R - \lambda I)\ell = 0 \tag{2.17}$$

where R is the usual correlation matrix with unity in the diagonal and ℓ is a column vector of direction cosines. Equation (2.17) has a non-trivial solution when

$$|R - \lambda I| = 0. \tag{2.18}$$

This is the general determinantal equation to be solved in factor analysis. In general there are n distinct non-zero roots of (2.18), λ_p , $p = 1, 2, \dots, n$. For each λ_p substituted back in (2.17), there results a set of direction cosines l_{ip} , one set for each test axis, $i = 1, 2, \dots, n$. The matrix of direction cosines is

$$L = \begin{bmatrix} l_{11} & l_{12} & \dots & l_{1n} \\ l_{21} & l_{22} & \dots & l_{2n} \\ \cdot & & & \\ \cdot & & & \\ l_{n1} & l_{n2} & \dots & l_{nn} \end{bmatrix} .$$

The elements of the pth column is the solution of (2.17) for the root λ_p .

Since the principal axes are chosen orthogonal and since

$$\sum_{i=1}^n l_{ip}^2 = 1,$$

$L'L = I$, the identity matrix. What relation exists between the l_{ip} and the b_{ip} ? Recall the problem in factor analysis is to find a matrix such that when multiplied by its transpose gives the correlation matrix R .

The question is, then, does $LL' = R$? The answer is no. But the l_{ip} and b_{ip} are proportional, the relationship being $b_{ip} = \sqrt{\lambda_p} l_{ip}$. Let the largest root of (2.18) be λ_1 , the next largest root λ_2 , etc. Recall the Pearsonian formulation of the problem was to take as the first principal axis the longest axis of the ellipsoid. Thus, if we let

$$F = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \cdot & & & \\ \cdot & & & \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{bmatrix} = \begin{bmatrix} l_{11}\sqrt{\lambda_1} & l_{12}\sqrt{\lambda_2} & \dots & l_{1n}\sqrt{\lambda_n} \\ l_{21}\sqrt{\lambda_1} & l_{22}\sqrt{\lambda_2} & \dots & l_{2n}\sqrt{\lambda_n} \\ \cdot & & & \\ \cdot & & & \\ l_{n1}\sqrt{\lambda_1} & l_{n2}\sqrt{\lambda_2} & \dots & l_{nn}\sqrt{\lambda_n} \end{bmatrix} \quad (2.19)$$

Then we have found the desired matrix of factor loadings such that $FF' = R$. (Note: The b_{ip} obtained by the centroid method do not equal the b_{ip} obtained by the principal axis solution. The principal axis solution is unique, while the centroid method gives the same answer to two different computers only if they agree to follow the same scheme in "reflecting". The centroid method places through the test configuration an "average" common factor axis, figure 2.2. The principal axis solution minimizes squared perpendicular distances of points to the line of best fit, figure 2.1).

The principal axis solution has many optimal properties. Consider the new variates $Y_{p\alpha}$ related to the $s_{i\alpha}$ by

$$Y_{p\alpha} = \sum_i l_{ip} s_{i\alpha} = \frac{1}{\sqrt{\lambda_p}} \sum_i b_{ip} s_{i\alpha}.$$

(These $Y_{p\alpha} \neq Y_{i\alpha}$, the raw scores discussed earlier). The new axes (Y_p and Y_m) are orthogonal. Also the variates $Y_{p\alpha}$ and $Y_{m\alpha}$ are statistically uncorrelated:

$$\begin{aligned} \sum_{\alpha} Y_{p\alpha} &= 0 \\ \sum_{\alpha} (Y_{p\alpha}, Y_{m\alpha}) &= \begin{cases} \lambda_p & \text{if } m = p \\ 0 & \text{if } m \neq p. \end{cases} \end{aligned} \quad (2.20)$$

This is seen to be true when it is realized that $Y_{p\alpha}$ is the (p, α) th element of the matrix $L'S$; recall that $\frac{1}{N-1} SS' = R = FF'$ so that (using matrix notation)

$$\begin{aligned}
YY' &= (L'S)(L'S)' = L'SS'L \\
&= (N-1) L'RL \\
&= (N-1) L'FF'L \\
&= (N-1)(L'F)(L'F)' \\
&= (N-1)(\Lambda)(\Lambda) \\
&= (N-1)\Lambda^2,
\end{aligned}$$

where Λ is a diagonal matrix with the $(p,m)^{\text{th}}$ element equal to

$$\begin{aligned}
\sum_i l_{ip} b_{im} &= \sum_i l_{ip} \sqrt{\lambda_m} l_{im} \\
&= \sqrt{\lambda_m} \sum_i l_{ip} l_{im} \\
&= \begin{cases} \lambda_p & \text{if } m = p \\ 0 & \text{if } m \neq p. \end{cases}
\end{aligned}$$

Thus

$$YY' = (N-1)\Lambda^2 = (N-1) \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & 0 \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}. \quad (2.21)$$

Obviously, then

$$\text{Var}(Y_{p\alpha}) = \lambda_p. \quad (2.22)$$

Also, the variates

$$Z_{p\alpha} = \sum_i b_{ip} s_{i\alpha} = \sqrt{\lambda_p} \sum_i l_{ip} s_{i\alpha}$$

have mean values zero and variances λ_p^2 .

That is,

$$\begin{aligned} \sum_{\alpha} Z_{p\alpha} &= 0 \\ \text{Var}(Z_{p\alpha}) &= \lambda_p^2. \end{aligned} \quad (2.23)$$

This result will be used in Chapter VI.

Finally, the variates

$$X_{p\alpha} = Y_{p\alpha} / \sqrt{\lambda_p}$$

have mean values zero and variances 1. These are the $X_{p\alpha}$ of the general factor analysis model (equations 2.2, 2.3, 2.4 of this chapter).

Recall that $FF' = R$; that is

$$\sum_{i=1}^n b_{ip}^2 = 1, \quad \sum_{p=1}^n b_{ip} b_{jp} = r_{ij}.$$

Using the definition of F of equation (2.19)

$$F'F = \Lambda^2 = \begin{bmatrix} \lambda_1 & & & & \\ & \lambda_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \lambda_n \end{bmatrix}. \quad (2.24)$$

F' may be viewed as a transformation matrix:

$$\begin{aligned}
F'RF &= F'(FF')F \\
&= (F'F)(F'F) \\
&= (F'F)(F'F)' \\
&= \Lambda^4 \\
&= \begin{bmatrix} \lambda_1^2 & & & 0 \\ & \lambda_2^2 & & \\ & & \ddots & \\ 0 & & & \lambda_n^2 \end{bmatrix}, \tag{2.25}
\end{aligned}$$

using (2.24).

Similarly for the L matrix of direction cosines,

$$\begin{aligned}
L'RL &= (L'F)F'L \\
&= (L'F)(L'F)' \\
&= \Lambda^2. \tag{2.26}
\end{aligned}$$

These results will be used in a discussion of factor analysis with the analysis of variance (Ch. IV).

If the h_i^2 are inserted in the diagonal and the resulting reduced correlation matrix R_a is factored, the approximate rank of R_a is r , the number of significant factors, and the solution of

$$|R_a - \lambda I| = 0 \tag{2.27}$$

again gives n roots λ_{ap} , but the last $n - r$ are approximately zero (i.e., they test non-significant), then the matrix of factor loadings, F_a is such that

$$F_a F_a' \simeq R_a$$

and

$$F_a' F_a = \Lambda_a^2 = \begin{bmatrix} \lambda_{a1} & & & 0 \\ & \lambda_{a2} & & \\ & & \ddots & \\ 0 & & & \lambda_{ar} \end{bmatrix} .$$

The roots λ_{ap} of (2.27) do not equal the roots λ_p of (2.18).

If the λ_p and λ_{ap} are numbered downward from the largest, then it can be proved that the pth component corresponding to λ_p (or λ_{ap}) accounts for the maximum remaining variation after the removal of the effect of the first (p-1) factors (i.e., after obtaining the projections of the test vectors onto the first (p-1) lines of best fit, corresponding to the largest (p-1) axis of the ellipsoid of points). Also,

$$\sum_{p=1}^n \lambda_p = n \quad (2.28)$$

and

$$\sum_{p=1}^r \lambda_{ap} \approx \sum_{i=1}^n h_i^2 .$$

2.3 A Critique of Spearman's One Factor Theory and Thurstone's Multiple Factor Theory.

To obtain a better understanding of factor analysis and what the psychologists are attempting to do, it is important to understand Spearman's one factor theory and how this was modified, primarily by Thurstone, to fit "multiple factor theory". The purpose of this section is to give a thumb-nail sketch of the rationale behind these two theories.

Figure (2.2), the test configuration, will be considered in this section.

In the introduction it was mentioned that Spearman postulated that the test score $s_{i\alpha}$ consisted of two parts, that part due to the effect of a common intellectual factor which he called "g", and a part due to a unique part of the test, consisting of the effect of a specific part and an error part. In the notation of section 1 of this Chapter, we may write

$$\begin{aligned} s_{i\alpha} &= \alpha_{ig} X_{g\alpha} + \psi_{i\alpha} \\ &= a_{ig} X_{g\alpha} + u_{i\alpha}, \end{aligned} \quad (2.29)$$

where α_{ig} is the amount of factor g the ith test possesses, $X_{g\alpha}$ (so scaled that

$$\left. \begin{aligned} \sum_{\alpha=1}^N X_{g\alpha} &= 0, & \sum_{\alpha=1}^N X_{g\alpha}^2 &= N - 1 \end{aligned} \right)$$

is the amount of factor p the \alphath individual possesses, and $\psi_{i\alpha}$ is the unique part of the test as measured on the \alphath individual. It is assumed in the population that the unique part of the test score is uncorrelated with the common factor part g. Also, for the sample, $\sum_{\alpha} X_{g\alpha} u_{i\alpha} = 0$. This is seen to be so by considering

$$u_{i\alpha} = s_{i\alpha} - a_{ig} X_{g\alpha}$$

so that

$$\begin{aligned} \sum_{\alpha} X_{g\alpha} u_{i\alpha} &= \sum_{\alpha} X_{g\alpha} (s_{i\alpha} - a_{ig} X_{g\alpha}) \\ &= \sum_{\alpha} X_{g\alpha} s_{i\alpha} - a_{ig} \sum_{\alpha} X_{g\alpha}^2. \end{aligned}$$

but

$$\begin{aligned} \sum_{\alpha} X_{g\alpha} s_{i\alpha} &= (N-1) \text{Cov}(\text{test } i, \text{factor } g) \\ &= (N-1) a_{ig} \end{aligned}$$

and since $\sum_{\alpha} X_{g\alpha}^2 = N-1$,

$$\sum_{\alpha} X_{g\alpha} u_{i\alpha} = (N-1) a_{ig} - (N-1) a_{ig} = 0.$$

$$\sum_{\alpha} s_{i\alpha} = a_{ig} \sum_{\alpha} X_{g\alpha} + \sum_{\alpha} u_{i\alpha}$$

and since $\sum_{\alpha} s_{i\alpha} = 0$ and $\sum_{\alpha} X_{g\alpha} = 0$, then $\sum_{\alpha} u_{i\alpha} = 0$ by necessity. And

$$\sum_{\alpha} s_{i\alpha}^2 = a_{ig}^2 \sum_{\alpha} X_{g\alpha}^2 + \sum_{\alpha} u_{i\alpha}^2,$$

or, dividing through by $(N-1)$

$$\underline{1} = a_{ig}^2 + \hat{\sigma}^2(u_i) \quad (2.30)$$

where $\hat{\sigma}^2(u_i) = \frac{1}{N-1} \sum_{\alpha} u_{i\alpha}^2$.

It has been mentioned that a_{ig} is called the factor loading or saturation. It may also be viewed as the length of the projection of test vector i onto the common factor axis X_g . Recall that in the discussion of figure (2.2), it was pointed out that the test vector i for the raw scores $Y_{i\alpha}$ has length $\sqrt{(N-1) \hat{\sigma}_{11}^2}$; while for the standard scores, the length is $\sqrt{N-1}$, and is 1 for the unitary standardized scores since $\sum_{\alpha} (s_{i\alpha}^*)^2 = 1$. Represent the common g factor axis by X_g and the axis of

the unique part of the test by u_i . Then the test vector i may be viewed as the sum of two components as in figure (2.4).

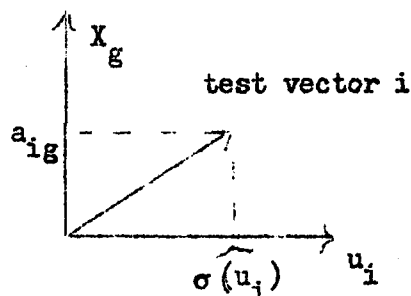


Figure 2.4

Test vector i lies in the plane spanned by X_g and u_i . For an n test "configuration" (test vectors so drawn that the angle between pairs of test vectors reflects the correlation between the two tests), there are $n + 1$ axes corresponding to figure 2.4: The common factor axis X_g and the n axes corresponding to the unique parts of the tests. In the sample, the unique axes are not mutually orthogonal, but are assumed to be in the population; $\sum_{\alpha} u_{i\alpha} u_{j\alpha} \approx 0$, so that u_i and u_j are not orthogonal.

This is seen to be so if $u_{i\alpha}$ and $u_{j\alpha}$ are written as the differences between the observed score and the calculated values, thus

$$\begin{aligned} \sum_{\alpha} u_{i\alpha} u_{j\alpha} &= \sum_{\alpha} (s_{i\alpha} - a_{ig} X_{g\alpha})(s_{j\alpha} - a_{jg} X_{g\alpha}) \\ &= \sum_{\alpha} s_{i\alpha} s_{j\alpha} - a_{jg} \sum_{\alpha} s_{i\alpha} X_{g\alpha} - a_{ig} \sum_{\alpha} s_{j\alpha} X_{g\alpha} \\ &\quad + a_{ig} a_{jg} \sum_{\alpha} X_{g\alpha}^2 \\ &= (N-1) r_{ij} - (N-1) a_{jg} a_{ig} - (N-1) a_{ig} a_{jg} \\ &\quad + (N-1) a_{ig} a_{jg} \end{aligned}$$

$$= (N-1) r_{ij} - (N-1) a_{ig} a_{jg} ;$$

but

$$a_{ig} a_{jg} \approx r_{ij}$$

so that

$$\sum_{i \neq j} u_{i\alpha} u_{j\alpha} \approx 0.$$

Each test vector i would lie in the plane spanned by the axes X_g and u_i . The projection of test vector i onto X_g is a_{ig} , the factor loading for the i th test on factor g . The correlation between the score on test i and the common factor g is

$$a_{ig} = \cos \theta_{ig} = r_{ig},$$

where r_{ig} is the correlation coefficient between the score on test i and the hypothetical g factor.

This general discussion of the common factor g is to show how Spearman's theory fits into the general formulation of section 1. How did Spearman approach the problem? He considered the "saturation" as the correlation of the test score with the hypothetical g . That is, he thought in terms of r_{ig} instead of a_{ig} . As mentioned earlier, Spearman noticed that if the score for n tests were correlated and set forth in a correlation matrix, the elements in different columns (or rows) of this matrix seemed to be approximately proportional.

Consider the correlation matrix of population correlation coefficients ρ_{ij} resulting from correlating all the test scores with each other and in addition with the hypothetical factor g .

	g	1	2	3	...	n
g	1	ρ_{1g}	ρ_{2g}	ρ_{3g}	...	ρ_{ng}
1		(H_1^2)	ρ_{12}	ρ_{13}	...	ρ_{1n}
2			(H_2^2)	ρ_{23}	...	ρ_{2n}
3				(H_3^2)	...	ρ_{3n}
.				.	.	.
.					.	.
.					.	.
n					.	(H_n^2)

(The H_i^2 will be discussed later).

If the proportionality in fact exists, then

$$\frac{\rho_{1g}}{1} = \frac{\rho_{12}}{\rho_{2g}} = \frac{\rho_{13}}{\rho_{3g}} = \dots$$

Using the first inequality,

$$\rho_{12} = \rho_{1g} \rho_{2g}$$

In general

$$\rho_{ij} = \rho_{ig} \rho_{jg} \tag{2.32}$$

Then, using the main body of the matrix, omitting g , expressions like

$$\rho_{ij} \rho_{kl} - \rho_{il} \rho_{jk} = 0. \tag{2.33}$$

This is seen to be so by substituting for each ρ_{ij} its value from (2.32).

The proportionality is maintained if one chooses

$$H_i^2 = \rho_{i2} \rho_{i3} / \rho_{23}$$

or, in general,

$$H_i^2 = \rho_{ij} \rho_{ik} / \rho_{jk} \quad (2.34)$$

Using (2.32),

$$H_i^2 = \rho_{ig} \rho_{jg} \rho_{ig} \rho_{kj} / \rho_{jg} \rho_{kg} = \rho_{ig}^2 \quad (2.34a)$$

The diagonal element H_i^2 is, then, the square of the correlation of test score i with the common factor g .

In experimental data, the proportionality would not be perfect (assuming one common factor g) because of the sampling errors in the correlation coefficients. If the sample estimates are substituted in (2.33), the resulting difference would not be zero, but would fluctuate around zero. ($r_{ij} r_{pq} - r_{iq} r_{jp}$ is called a "tetrad"). And if sample values are substituted in (2.34) for the corresponding parameters, the resulting estimate h_i^2 of H_i^2 varies according to the choice of j and k . For only two tests ($n = 2$), there are an infinite number of solutions of h_1^2 and h_2^2 such that

$$\frac{h_1^2}{r_{12}} = \frac{r_{12}}{h_2^2} .$$

For $n = 3$, the h_i^2 may be determined exactly:

$$h_1^2 = \frac{r_{12} r_{13}}{r_{23}} , \quad h_2^2 = \frac{r_{12} r_{23}}{r_{13}} , \quad h_3^2 = \frac{r_{13} r_{23}}{r_{12}} .$$

For $n > 3$, no exact solution is possible. Rather than compute $h_i^2 = r_{ij} r_{ik} / r_{jk}$ for all $j \neq k \neq i$, Spearman suggested the use of

$$h_i^2 = \frac{\sum_{j \neq k \neq i} r_{ij} r_{ik}}{\sum_{j \neq k \neq i} r_{jk}} \quad (2.35)$$

Comparing (2.35) with (2.34) and (2.34a), h_i^2 is seen to be the logical estimate of H_i^2 .

Thus, the factor loadings a_{ig} may be found by taking $a_{ig} = h_i$ where h_i^2 is computed from (2.35). For the one-common factor theory,

$$\begin{aligned} h_i^2 &= r_{ig}^2 = (\text{correlation of test score } i \text{ with factor } g)^2 \\ &= a_{ig}^2 = (\text{factor loading of test } i \text{ on factor } g)^2 \\ &= \text{Var} \quad (\text{part of test } i \text{ common to all other tests in a} \\ &\quad \text{given battery}). \end{aligned}$$

In the $[\bar{\rho}_{ij}]$ matrix above, if the H_i^2 are determined from (2.34) and substituted in the diagonal, the result, R_a^* , is a matrix of rank 1, under the assumption of one common factor. The rank of an $n \times n$ sample correlation matrix is exactly n , and the substitution of h_i^2 (determined according to (2.35), say) still gives a matrix R_a of rank n . It is often stated, incorrectly, that R_a has rank 1. What is meant is that R_a^* has rank 1, and that R_a has rank 1 aside from sampling fluctuations, since R_a is the estimate of R_a^* . One might say, then, that R_a has "effective" rank 1 or it has 1 "significant" factor. (See Chapter VII for tests of significance). Hotelling (1942) argues that for N points plotted in $n = 3$ space, the typical ellipsoid of constant probability may have any

shape from a sphere to a pancake or a needle. The "pancake" case may be thought of as approximating a distribution in 2 dimensions, and the "needle", a distribution of one dimension. But the points in fact all lie in 3 dimensions, in 2, or in one. And if the sample points lie in fewer than 3 dimensions, then all the values in the population lie in fewer than 3 dimensions, for the probability of the sample lying in 2 dimensions when the population is in 3 dimensions is zero. He states that "The problem therefore has been propounded more than once of finding a sampling distribution for testing by means of a sample the hypothesis that the distribution is really of fewer dimensions than the number of variates measured. Such efforts are completely futile In order to explain the manifest fact that the dimensionality of the set of observations is actually the number of variates and no less, they have resorted to highly unsatisfactory 'communalities' associated with 'specific factors' whose existence does not seem particularly clear on a priori grounds, and which is not tested by the statistical methods used." It is true that the problem is stated perhaps incorrectly in this manner. But what the psychologists really want to do is find the number of "significant" or psychologically important factors which will explain "satisfactorially" or to a certain degree of approximation, the observations.

Spearman and his workers observed that some tests upset the apparent proportionality. The tetrads formed with these tests and other tests did not fluctuate around zero. Consider the section of the correlation matrix

	i	j	k
p	ρ_{ip}	ρ_{jp}	ρ_{kp}
q	ρ_{iq}	ρ_{jq}	ρ_{kq}
r	ρ_{ir}	ρ_{jr}	ρ_{kr}

Form all tetrads using ρ_{ip} as the starting or pivot coefficients:

(td = tetrad)

$$td_1 = \rho_{ip} \rho_{jq} - \rho_{iq} \rho_{jp}$$

$$td_2 = \rho_{ip} \rho_{kq} - \rho_{iq} \rho_{kp}$$

$$td_3 = \rho_{ip} \rho_{jr} - \rho_{ir} \rho_{jp}$$

$$td_4 = \rho_{ip} \rho_{kr} - \rho_{ir} \rho_{kp}$$

(Let \hat{td}_i be the estimate of td_i , where the ρ_{ij} are replaced by their estimates, r_{ij}). Spearman noticed that a "tetrad of tetrads" fluctuated about zero for certain sub-sets of his tests. That is

$$\hat{td}_1 \hat{td}_4 - \hat{td}_2 \hat{td}_3 \approx 0.$$

This is called a "second order tetrad" (td_1 , e.g., is a "first order tetrad"). It can be shown that

$$td_1 td_4 - td_2 td_3 = \rho_{ip} \cdot \begin{vmatrix} \rho_{ip} & \rho_{jp} & \rho_{kp} \\ \rho_{iq} & \rho_{jq} & \rho_{kq} \\ \rho_{ir} & \rho_{jr} & \rho_{kr} \end{vmatrix}$$

(The bars " $\begin{vmatrix} \end{vmatrix}$ " stand for "determinant of"). Under the assumption of two common orthogonal factors among the 6 tests, the above determinant is exactly zero. But all of the first order tetrads, are zero, since the

two common factors act as if there is only one common factor. But assume, for definiteness, that test i has two common factors with tests p and q and that all the other tests have only one common factor, then td_3 , for example would not be zero; however, the above determinant would be exactly zero. And if the H_i^2 are so chosen to maintain this proportionality, the resulting matrix R_a^* would have exact rank 2. The sample matrix R_a corresponding to this R_a^* would have rank n , but "effective" rank 2. Thus we can state the following rule: if the rank of a correlation matrix R_a^* is r , then there are r common factors present. Postulating two or more common factors takes us into the realm of "multiple factor" analysis.

Thurstone proposed to view the problem in this light. Use the off diagonal terms of the sample correlation matrix to determine the number of significant factors present, then adjust the diagonal elements accordingly. The general problem of multiple factor analysis, then, is to find the effective rank of the sample correlation matrix R , aside from the diagonal elements. Exact rank $r < n$ will never be attained in experimental data, but it is postulated that the rank of the correlation matrix is r within sampling fluctuations of the correlation coefficients. The number of significant factors present, then, is the significant rank of the correlation matrix.

The factor analysis equation for two common factors may be written as

$$\begin{aligned}
 s_{i\alpha} &= \alpha_{i1} x_{1\alpha} + \alpha_{i2} x_{2\alpha} + \eta_{i\alpha}' \\
 &= a_{i1} x_{1\alpha} + a_{i2} x_{2\alpha} + u_{i\alpha}'
 \end{aligned}
 \tag{2.36}$$

The unique part of $\psi_{i\alpha} = s_{i\alpha} - \alpha_{ig} \chi_{g\alpha}$ of the one-common factor model has now been broken down into another common factor and the unique part $\psi'_{i\alpha}$. (For simplicity, the prime on ψ will be omitted). It should be mentioned, however, that the multiple factor analysts do not postulate other common factors over and above "g". If the "domain" or area of study is intelligence, the two common factors in (2.36) may be factors like numeric ability and verbal ability, both of which would be considered parts of the general intellectual factor "g". In effect, "g" is considered as consisting of many factors. (It should be observed that the multiple factor analysts study other domains of human behavior such as personality and vision as well as intelligence).

It is assumed that the common factors are uncorrelated with each other as well as with the unique part; and for the sample, the common factors are chosen orthogonal to each other as well as to the unique part. That is

$$\begin{aligned} \sum_{\alpha} X_{p\alpha} u_{i\alpha} &= \sum_{\alpha} X_{p\alpha} (s_{i\alpha} - \sum_{m=1}^r a_{im} X_{m\alpha}) \\ &= \sum_{\alpha} X_{p\alpha} s_{i\alpha} - \sum_{m=1}^r a_{im} \sum_{\alpha} X_{p\alpha} X_{m\alpha} \\ &= (N-1) a_{ip} - (N-1) a_{ip} \\ &= 0 \end{aligned}$$

since

$$\sum_{\alpha} X_{p\alpha} X_{m\alpha} \begin{cases} = N - 1, & \text{if } m = p \\ = 0 & , \text{if } m \neq p. \end{cases}$$

If $r = 2$,

$$0 = \sum_{\alpha} s_{i\alpha} = a_{i1} \sum_{\alpha} X_{1\alpha} + a_{i2} \sum_{\alpha} X_{2\alpha} + \sum_{\alpha} u_{i\alpha}.$$

The $X_{p\alpha}$ are scaled so that $\sum_{\alpha} X_{p\alpha} = 0$, and since $\sum_{\alpha} s_{i\alpha} = 0$ and $\sum_{\alpha} X_{p\alpha} = 0$, then

$$\begin{aligned} (N-1) &= \sum_{\alpha} s_{i\alpha}^2 = a_{i1}^2 \sum_{\alpha} X_{1\alpha}^2 + a_{i2}^2 \sum_{\alpha} X_{2\alpha}^2 \\ &\quad + 2a_{i1} a_{i2} \sum_{\alpha} X_{1\alpha} X_{2\alpha} \\ &\quad + 2a_{i1} \sum_{\alpha} X_{1\alpha} u_{i\alpha} + 2a_{i2} \sum_{\alpha} X_{2\alpha} u_{i\alpha} \\ &= (N-1) [a_{i1}^2 + a_{i2}^2 + \widehat{\sigma^2(u_i)}], \end{aligned}$$

where $\widehat{\sigma^2(u_i)} = \sum_{\alpha} u_{i\alpha}^2 / (N-1)$. Thus

$$a_{i1}^2 + a_{i2}^2 + \widehat{\sigma^2(u_i)} = 1. \quad (2.37)$$

h_i^2 is defined to be that part of the variance of test i attributable to common factors, thus

$$h_i^2 = a_{i1}^2 + a_{i2}^2.$$

In general, for $r \leq n$ common factors,

$$h_i^2 = \sum_{p=1}^r a_{ip}^2 \quad (2.38)$$

and

$$h_i^2 + \widehat{\sigma^2(u_i)} = 1. \quad (2.39)$$

The a_{ip} are, as usual, the saturations of test i on factor p . And as for the one-factor theory, the a_{ip} may also be viewed as the projections of test vector i onto common factor axes. Corresponding to figure 2.4, for two common factors, test vector i has three components as shown in figure 2.5.

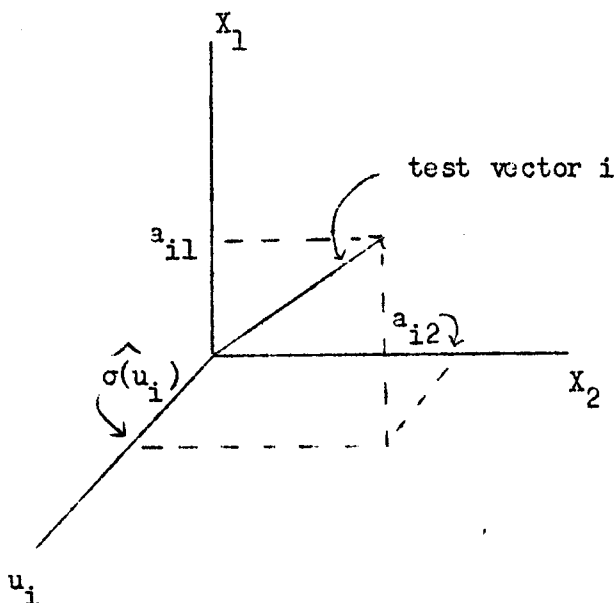


Figure 2.5

Thus, the length of test vector i (plotting $\sum_{\alpha} (s_{i\alpha}^*)^2 = 1$) is $\sqrt{a_{i1}^2 + a_{i2}^2 + \widehat{\sigma}^2(u_i)}$. For the n test-vector configuration and r common factors, there would be $r + n$ axes, corresponding to figure 2.5. Test vector i would lie in the hyper-plane spanned by the r common factor axes plus the unique axis u_i .

As for the one-common factor theory, the u_i and u_j are not orthogonal, since

$$\begin{aligned}
\sum_{\alpha} u_{i\alpha} u_{j\alpha} &= \sum_{\alpha} \left(s_{i\alpha} - \sum_{p=1}^r a_{ip} X_{p\alpha} \right) \left(s_{j\alpha} - \sum_{p=1}^r a_{jp} X_{p\alpha} \right) \\
&= \sum_{\alpha} s_{i\alpha} s_{j\alpha} - \sum_{p=1}^r a_{ip} \sum_{\alpha} X_{p\alpha} s_{j\alpha} - \sum_{p=1}^r a_{jp} \sum_{\alpha} X_{p\alpha} s_{i\alpha} \\
&\quad + \sum_{p=1}^r a_{ip} \sum_{p=1}^r a_{jm} \sum_{\alpha} X_{p\alpha} X_{m\alpha} \\
&= (N-1) r_{ij} - (N-1) \sum_{p=1}^r a_{ip} a_{jp} \\
&= (N-1) \left[\bar{r}_{ij} - \sum_{p=1}^r a_{ip} a_{jp} \right]
\end{aligned}$$

and $\sum_{p=1}^r a_{ip} a_{jp} \neq r_{ij}$ (unless $r = n$),

i.e., $\sum_{p=1}^r a_{ip} a_{jp} \approx r_{ij}$, for $r < n$,

Thus

$$\sum_{\alpha} u_{i\alpha} u_{j\alpha} \approx 0.$$

There are several procedures for estimating the h_i^2 for multiple factor analysis, but none which is wholly satisfactory. One could, for example, use the estimate r_{ii}' of the reliability coefficient, ρ_{ii}' , for h_i^2 . However, it will be shown in Chapter III that r_{ii}' is an overestimate of h_i^2 . For large batteries of 30 or 40 tests, the computer is usually instructed to substitute for h_i^2 the highest correlation occurring

in the *i*th row or column. This is considered to be sufficiently accurate since a battery of tests of that size would probably be factored by the centroid method. And the factor loading for the *i*th test on factor 1 is $\frac{\sum_j r_{ij}}{\sqrt{\sum_j \sum_j r_{ij}}}$, which would be a little affected by the value used for the diagonal element. Another procedure is to start factoring, using a guessed value usually less than 1, factor the matrix until the $(r+1)$ th factor tests non-significant, and obtain factor loadings a_{ip}^* . Then use

$$(h_i^*)^2 = \sum_{p=1}^r (a_{ip}^*)^2$$

to re-factor the correlation matrix. This iterative procedure is recommended by Burt and others, but is considered too laborious for large *n*. Too, the estimation of the h_i^2 becomes confounded with the number of significant factors present, a not too satisfactory situation.

Consider the parts of the test vectors that are in the common factor space; i.e., the space spanned by the *r* orthogonal common factor axes X_1, X_2, \dots, X_r . The correlation between two test vectors is the inner-product of their common factor coordinates, thus

$$r_{ij} \approx \sum_{p=1}^r a_{ip} a_{jp}$$

The sum of the inner-product does not exactly equal the sample correlation coefficient because the two unique axes u_i and u_j are not exactly orthogonal; i.e., for a sample of size *N*, $\sum_{\alpha} u_{i\alpha} u_{j\alpha}$ is only approximately zero. The variance of test score *i* common to all the other tests in the battery is h_i^2 . Thus

$$h_i^2 = \sum_{p=1}^r a_{ip}^2,$$

which is an estimate of H_i^2 , the population value. (Note: The a_{ip} are used to denote factor loadings when R_a , the reduced correlation matrix is factored, and the b_{ip} are used when R is factored. In general $a_{ip} \neq b_{ip}$, as will be demonstrated in Chapter VII).

The common factor axes are chosen orthogonal (they are postulated to be orthogonal in the population and are made to be orthogonal in the sample), thus

$$\sum_{i=1}^n a_{ip} a_{im} = 0, \quad m \neq p. \quad (2.40)$$

For the first factor, there are n factor loadings to be evaluated, for the second, only $(n-1)$ independent loadings because of the orthogonality of the axes (equation 2.40), and for the r th factor, $(n-r+1)$ independent loadings. Hence the total number of independent loadings for $r < n$ common factors is

$$n + (n-1) + \dots + (n-r+1) = nr - \frac{1}{2} r (r-1).$$

So as to be able to estimate these loadings from the $\frac{1}{2} n(n-1)$ different correlation coefficients, it is required at least

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

Determinateness results when the equality holds. A few such values of n and r are:

n(tests)	3	5	6	12	15	18
r(common factors)	1	2	3	7	10	12

Chapter III

CORRELATION, RELIABILITY, AND COMPONENTS OF VARIANCE

3.1 Purpose of chapter

The purpose of this chapter is: 1) to clarify the relationship between various types of correlations, 2) to write the raw scores $Y_{i\lambda}$ in the usual experimental model and see what relationship it has, if any, to the general factor analysis model, and 3) to accept the plausible statement that correlations are due to "common causes" or "common factors" and then study the test scores to see what part is due to common factors. The three objectives are not mutually exclusive and will not be dealt with necessarily in this order.

3.2 Intra-class and inter-class correlations and reliability coefficients

Consider the results of administering the same test to the same N individuals at two or more different times:

$$\begin{array}{l} \text{time 1: } Y_{11} \quad Y_{12} \quad \dots \quad Y_{1N} \\ \text{time 2: } Y_{21} \quad Y_{22} \quad \dots \quad Y_{2N}, \end{array}$$

where the $Y_{i\lambda}$ are the raw scores and the first subscript refers to the time rather than the test. It will be recalled that the $Y_{i\lambda}$ are related to the $s_{i\lambda}$ by

$$Y_{i\lambda} = \bar{Y}_i + s_{i\lambda} \sqrt{\sigma_{ii}}.$$

Suppose the test score $Y_{i\lambda}$ has a mean effect μ , a time effect τ_i , a person

effect ρ_{α} and an error component $\varepsilon_{i\alpha}$. (The ρ_{α} should cause no confusion. The population correlation coefficient ρ_{ij} has two subscripts instead of one, neither of which is Greek). If it is further assumed that

$$E(Y_{i\alpha}) = \mu + \tau_i = \eta_i,$$

$$E(\rho_{\alpha}) = 0 = E(\varepsilon_{i\alpha})$$

$$E(\rho_{\alpha}, \varepsilon_{i\alpha}) = 0 = E(\varepsilon_{i\alpha}, \varepsilon_{j\beta}), \quad i \neq j \text{ and/or } \alpha \neq \beta.$$

$$E(\rho_{\alpha})^2 = \sigma_p^2, \quad E(\varepsilon_{i\alpha})^2 = \sigma_{ei}^2 = \sigma_e^2.$$

(The σ_{ei}^2 are assumed equal for two different administrations of the same test). Then

$$E(Y_{i\alpha} - \mu - \tau_i)^2 = E(Y_{i\alpha} - \eta_i)^2 = \sigma_p^2 + \sigma_e^2.$$

The corresponding sample estimates of the parameters are

$$\eta_i = \mu + \tau_i \doteq \bar{Y}_i$$

$$\rho_{\alpha} \doteq p_{\alpha}$$

$$\tau_i \doteq t_i$$

The person effect ρ_{α} is a random variable. The sample of N individuals is a random sample from an infinite population of individuals. For the present discussion, assume the time effects are "fixed" (i.e., non-random), and that $\sum t_i = 0$. The usual analysis of variance has the following form:

<u>Source</u>	<u>df.</u>	<u>S. Sq.</u>	<u>M.Sq.</u>	<u>E(M.Sq.)</u>
Times	1	$\frac{1}{2} N (\bar{Y}_1 - \bar{Y}_2)^2$	V_1	$\sigma_e^2 + \theta(\tau)$
Persons	$N - 1$	$\frac{1}{2} \sum_{\lambda} (Y_{1\lambda} + Y_{2\lambda} - \bar{Y}_1 - \bar{Y}_2)^2$	V_2	$\sigma_e^2 + 2\sigma_p^2$
<u>Times x persons</u>	<u>$N - 1$</u>	$\frac{1}{2} \sum_{\lambda} (Y_{1\lambda} - Y_{2\lambda} - \bar{Y}_1 + \bar{Y}_2)^2$	V_3	σ_e^2
Total	$2N - 1$			

In the above analysis of variance table, d.f. = degrees of freedom, S. Sq. = sum of squares, M. Sq. = mean square = S. Sq./d.f., E(M. Sq.) = expected value of the mean square, and $\theta(\tau)$ is a function only of the time effects, τ_i . If it is assumed that the time effects are zero, $\tau_1 = \tau_2 = 0$, then the mean square for times is also an estimate of σ_e^2 , and hence, the one degree of freedom for the source "Times" would be added to the error sum of squares to give N degrees of freedom in the estimate $V_c = \frac{(N-1)V_3 + V_1}{N}$ of the error component σ_e^2 .

The intra-class correlation is defined to be

$$\rho_c = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_e^2}, \quad (3.1)$$

and the estimate of the correlation, assuming zero time effects, is

$$r_c = \frac{\frac{1}{2} (V_2 - V_e)}{\frac{1}{2} (V_2 + V_e)} = \frac{V_2 - V_e}{V_2 + V_e}. \quad (3.2)$$

If in the above set-up, the time effects are important (i.e., $\tau_1, \tau_2 \neq 0$),

or if the set-up is changed to one administration of parallel forms*, the one degree of freedom for differences in times or differences in parallel forms should be isolated. In this case, the best estimate of the intra-class correlation is

$$r'_c = \frac{\frac{1}{2}(V_2 - V_3)}{\frac{1}{2}(V_2 + V_3)} = \frac{V_2 - V_3}{V_2 + V_3}. \quad (3.3)$$

It will be observed that ρ_c , defined in (3.1), is the definition given in Chapter II, equation 2.11. That is, it is the ratio of the "steady" part of the test to the total variance. If the same test is administered twice, call the times i and i' . Then $\rho_{ii'}$ represents the fraction of the population variance of each accounted for by the steady part ($1 - \rho_{ii'}$) the fraction attributable to random "error". Psychologists call $r_{ii'}$, the estimate of $\rho_{ii'}$, the reliability coefficient. They, apparently, have never considered any form of statistic for this purpose other than the product-moment, alias "inter-class", correlation coefficient. But there is no basic reason for assessing reliability on only duplicate tests. Any number of replications may be made (at least theoretically; ignore here the effect of learning which may be a practical difficulty in some cases). It becomes apparent that the problem is that of "intra-class" correlation.

Historically the intra-class correlation coefficient was defined only for the case where there was no cross-classification (in our present example, if we knew for each individual only his two (or more) scores

*See glossary for definition.

without knowing which was made at which time (or on what form), or else deemed such information irrelevant to our purpose). In these cases, the correlation was (originally) computed, and thus implicitly defined, as the product-moment correlation of a "symmetric table"; i.e., one in which every pair of observations was entered twice in symmetric positions in the table. All this was finally brought to clear cut definition by Fisher in 1925 (Chapter VII, which may be consulted for fuller exposition of the above points) who showed that the analysis of variance approach was more fundamental and usually more meaningful.

Thus, r'_c shall be used as the only estimator for the reliability. The information thrown away by $(n - 1)$ degrees of freedom between times, if in fact they are equal, will be rather trivial if N is large.

Fix attention again on the definition of ρ_c ($\equiv \rho_{ii'}$) as given by (3.1). Obviously, the larger the σ_p^2 , the larger will be ρ_c . The estimate of σ_p^2 depends on the individuals in the sample. The more heterogeneous the individuals, the larger the estimate of σ_p^2 . The reliability of a test is, then, relative. If our population of individuals is very heterogeneous, σ_p^2 will be large, and for this group, ρ_c will be large. However, for another population of homogeneous individuals, σ_p^2 will be small, and ρ_c will be small. Thus, to establish the reliability of a test, we must specify the population of individuals from which our sample of N is presumed to be a random sample. Some psychologists tend to take the talk about random samples lightly; just any old group of people will do. For some studies, perhaps that is true, but where reliabilities are being evaluated for new tests, they should draw their sample very carefully. Ordinary product-moment correlations are affected in the same manner. Both are directly

influenced by the range of variables (or of abilities) under study. This is emphasized here since it is not appreciated by many research workers. It can be shown that

$$\begin{aligned} E(\text{numerator of } r'_c) &= E\left\{ \frac{1}{2}(V_2 - V_3) \right\} \\ &= \sigma_p^2 \\ &= E\left\{ \frac{S(Y_{1\alpha} - \bar{Y}_1)(Y_{2\alpha} - \bar{Y}_2)}{(N-1)} \right\} \end{aligned}$$

and

$$\begin{aligned} E(\text{denominator of } r'_c) &= E\left\{ \frac{1}{2}(V_2 + V_3) \right\} \\ &= \sigma_p^2 + \sigma_e^2 \\ &= E\left\{ \frac{S(Y_{1\alpha} - \bar{Y}_1)^2}{(N-1)} \right\} \\ &= E\left\{ \frac{S(Y_{2\alpha} - \bar{Y}_2)^2}{(N-1)} \right\}, \end{aligned}$$

thus r'_c may be written as

$$r'_c = \frac{S(Y_{1\alpha} - \bar{Y}_1)(Y_{2\alpha} - \bar{Y}_2)}{\frac{1}{2}\left\{ S(Y_{1\alpha} - \bar{Y}_1)^2 + S(Y_{2\alpha} - \bar{Y}_2)^2 \right\}} \quad (3.4)$$

The usual estimate of the association parameter, ρ_{12} , for the bivariate normal distribution is the product-moment correlation coefficient. It is estimated by

$$r_{12} = r_m = \frac{S(Y_{1\alpha} - \bar{Y}_1)(Y_{2\alpha} - \bar{Y}_2)}{\left\{ S(Y_{1\alpha} - \bar{Y}_1)^2 S(Y_{2\alpha} - \bar{Y}_2)^2 \right\}^{1/2}} \quad (3.5)$$

The product-moment correlation is sometimes referred to as the inter-class correlation. From the definitions in (3.4) and (3.5) it is apparent that only if

$$S(Y_{1\alpha} - \bar{Y}_1)^2 = S(Y_{2\alpha} - \bar{Y}_2)^2$$

are the statistics r'_c and r_m identical. Thus $|r'_c| \leq |r_m|$ since the geometric mean is less than or equal to the arithmetic mean.

To emphasize the point, clearly r'_c of (3.3) or (3.4) should be used as the estimate of the reliability coefficient ρ_{ii} , rather than r_m , the product-moment correlation.

One further comparison between intra- and inter-class correlation is of interest. Consider the regression of Y_1 on Y_2 . Let

$$Sy_1y_2 = S(Y_{1\alpha} - \bar{Y}_1)(Y_{2\alpha} - \bar{Y}_2)$$

$$Sy_i^2 = S(Y_{i\alpha} - \bar{Y}_i)^2.$$

The regression sum of squares is

$$(Sy_1y_2)^2 / Sy_2^2.$$

Since the total sum of squares is Sy_1^2 , the fraction of the sum of squares accounted for by regression is

$$r_{12}^2 = \frac{Sy_1y_2)^2 / Sy_2^2}{Sy_1^2} = \frac{\text{S. Sq. (Regression)}}{\text{Total S.Sq.}}$$

The definition of ρ_{12}^2 for the population is

$$(1 - \rho_{12}^2) = \frac{\sigma^2(\text{about regression})}{\text{Total } \sigma^2}$$

The intra-class correlation is

$$\rho_c = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_e^2} = \frac{\sigma^2(\text{steady part of test})}{\text{Total } \sigma^2}$$

or

$$(1 - \rho_c) = \frac{\sigma_e^2}{\sigma_T^2} = \frac{\sigma^2(\text{error of test})}{\text{Total } \sigma^2}$$

Note that ρ_{12} is squared while ρ_c is not. But the point is that ρ_{12} and ρ_c are not measures of the same thing.

There will be a digression here to consider the variance of a variance component. It is not often recognized that the usual unbiased estimate of σ_p^2 , viz. $\frac{1}{2}(V_2 - V_3)$, when $n = 2$, is identically equal to $\text{Cov}(Y_1, Y_2)$, i.e., $Sy_1y_2 / (N - 1) = S(Y_{1\alpha} - \bar{Y}_1)(Y_{2\alpha} - \bar{Y}_2) = \frac{1}{2} \{V_2(Y_1, Y_2) - V_3(Y_1, Y_2)\}$.

The variance of the estimate $\hat{\sigma}_p^2$ is usually found by taking the variance of the linear form $\frac{1}{2}(V_2 - V_3)$, thus

$$\begin{aligned} \text{Var}(\hat{\sigma}_p^2) &= \text{Var} \left\{ \frac{1}{2}(V_2 - V_3) \right\} \\ &= \frac{1}{4} \left\{ \frac{2V_2^2}{(N+1)} + \frac{2V_3^2}{(N+1)} \right\} \\ &= \frac{V_2^2 + V_3^2}{2(N+1)}, \end{aligned}$$

the degrees of freedom plus 2 being used in the estimator to adjust for bias. But the variance of $Sy_1y_2/(N-1)$ may also be obtained from the Wishart distribution, the distribution of variances and covariances. The moment generating function for the Wishart distribution of sums of squares and cross-products (not variances and covariances) as given by Wilks (1947, Chapter XI, p. 118) is

$$\phi(\theta_{ij}) = \left| A_{ij} \right|^{-\frac{N-1}{2}} \left| A_{ij} - 2\theta_{ij} \right|^{-\frac{N-1}{2}}$$

where $[A_{ij}]$ is the inverse of the population variance-covariance matrix. Wilks also introduces the factor 2 for the cross-product term in the moment generating function so that

$$\frac{\partial \phi}{\partial \theta_{12}} \Big|_{(\theta_{ij} = 0)} = 2(N-1) \rho_{12} \sigma_1 \sigma_2 = E(2Sy_1y_2),$$

where ρ_{12} is the association parameter in the bivariate normal distribution. Similarly, it can be shown that

$$\begin{aligned} \frac{\partial^2 \phi}{\partial \rho_{12}^2} \Big|_{(\rho_{ij} = 0)} &= 4(N-1)\sigma_1^2\sigma_2^2(1 + N\rho_{12}^2) \\ &= E(2Sy_1y_2)^2. \end{aligned}$$

Thus

$$\begin{aligned} \text{Var}(\hat{\sigma}_p^2) &= \text{Var}\left\{ \frac{Sy_1y_2}{N-1} \right\} = \frac{\sigma_1^2\sigma_2^2(1 + \rho_{12}^2)}{N-1} \\ &= \frac{1}{N-1} (\sigma_e^4 + 2\sigma_p^4 + 2\sigma_p^2\sigma_e^2) \\ &= \frac{V_2^2 + V_3^2}{2(N+1)}. \end{aligned}$$

Back to the intra-class correlation. Next consider the administration of the same test n times, or if there is a learning effect present, the administration of n parallel forms of the same test. The analysis of variance take the form:

<u>Source</u>	<u>df.</u>	<u>M.Sq.</u>	<u>E(M.Sq.)</u>
Tests (or times)	$n - 1$	V_1	$\sigma_e^2 + \theta(\tau)$
Persons	$N - 1$	V_2	$\sigma_e^2 + n\sigma_p^2$
Times x persons	$(n - 1)(N - 1)$	V_3	σ_e^2

It might be desirable to consider the time or test effect as random. In that case, $\theta(\tau) = N\sigma_t^2$. If there are test differences or differences due to times, then the best estimate of the intra-class correlation coefficient is

$$r'_c(n) = \frac{\frac{1}{n}(v_2 - v_3)}{\frac{1}{n}\{v_2 + (n-1)v_3\}} = \frac{v_2 - v_3}{v_2 + (n-1)v_3} \quad (3.6)$$

One might compute the average of all possible product moment correlation coefficients, \bar{r}_{ii} , which is also an estimate of the reliability of the test, but \bar{r}_{ii} is far more laborious to compute. (Fisher, 1925, Chapter VII). Since $r'_c(n)$ is an estimate of $\sigma_p^2/(\sigma_p^2 + \sigma_e^2)$, then for n replications of the same test, this statistic should be used to estimate reliability rather than \bar{r}_{ii} .

3.3 What to factor: common parts and steady parts of raw scores.

Recall that for standardized scores, the factor analysis model was written as

$$\begin{aligned} s_{i\alpha} &= \sum_{p=1}^r \lambda_{ip} X_{p\alpha} + \pi_{i\alpha} + \epsilon_{i\alpha} \\ &= \sum_{p=1}^r a_{ip} X_{p\alpha} + f_{i\alpha} + e_{i\alpha}, \end{aligned} \quad (3.7)$$

where $\sum_p \lambda_{ip} X_{p\alpha}$ = the effect of the common parts of the test score i on the α th individual, $\pi_{i\alpha}$ = the effect of the specific part, and $\epsilon_{i\alpha}$ = the random error term.

For the purpose of this argument, assume the raw score on two different tests X and Y can be written as the sum of three parts: (The tests are

given to the same group of persons)

$$\begin{aligned} Y_{\alpha} &= \gamma_y + \phi_{\alpha} + \varepsilon_{\alpha} \\ X_{\alpha} &= \gamma_x + \theta_{\alpha} + \delta_{\alpha}, \end{aligned} \tag{3.8}$$

where γ_y and γ_x are the mean effects, ϕ_{α} and θ_{α} are person effects, and ε_{α} , δ_{α} random error terms. Now what parts of (3.8) correspond to those of (3.7)? The standardized score has no mean effect ($\sum_{\alpha} s_{i\alpha} \equiv 0$) so that γ_y and γ_x have no counterparts in (3.7). It seems reasonable to assume that ε_{α} and δ_{α} correspond to the random error term $\varepsilon_{i\alpha}$, so that what we call the person effect, ϕ_{α} and θ_{α} , must correspond to the common part plus the specific part. Thus, they are the "steady" parts of the raw test scores. This agrees with our above conclusions since the steady part of the standardized score was defined in Chapter II to be

$$\omega_{i\alpha} = \sum_{p=1}^r \alpha_{ip} x_{p\alpha} + \pi_{i\alpha}.$$

But the steady parts of any two test scores are not necessarily equal.

The person effects, then, are rather like an "interaction" (used in the dictionary sense). Assume test Y is concerned with numerical problems and say that individual α has considerable numeric ability, then ϕ_{α} is large. If test X is a verbal test and individual α is relatively weak along such lines, θ_{α} will be small. Thus, for two different tests, ϕ_{α} and θ_{α} will be different; this is where the model of (3.8) differs from the usual experimental model.

Assume that

$$E(Y) = \eta_y, \quad E(X) = \eta_x$$

$$E(\phi_{\alpha}) = E(\theta_{\alpha}) = E(\varepsilon_{\alpha}) = E(\delta_{\alpha}) = 0$$

$$E(\phi_{\alpha})^2 = \sigma_{\phi}^2; \quad E(\theta_{\alpha})^2 = \sigma_{\theta}^2$$

$$E(\varepsilon_{\alpha})^2 = \sigma_{\varepsilon}^2; \quad E(\delta_{\alpha})^2 = \sigma_{\delta}^2$$

$$E(\phi_{\alpha}, \varepsilon_{\alpha}) = E(\theta_{\alpha}, \delta_{\alpha}) = E(\varepsilon_{\alpha}, \delta_{\alpha}) = 0.$$

The steady parts (i.e. the person effects), even though different and measured on different scales have a common part, γ_{α} ; thus

$$\phi_{\alpha} = \gamma_{\alpha} + \xi_{\alpha}$$

$$\varepsilon_{\alpha} = \gamma_{\alpha} + \zeta_{\alpha}$$

γ_{α} is the effect due to the common factor of X and Y, ξ_{α} is the specific part of Y, and ζ_{α} is the specific part of X. This is perhaps better seen by an illustration: Let

$$\sigma_x^2 = 12 = \text{Total variance of X in the population}$$

(units are unimportant)

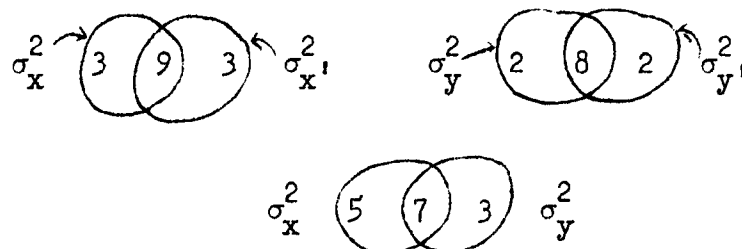
$$\rho_{xx'} = .75 = \text{True reliability of test X}$$

$$\rho_{xy} = 7/\sqrt{120} = \text{True correlation between X and Y}$$

$$\sigma_y^2 = 10 = \text{Total variance of Y in the population}$$

$$\rho_{yy'} = .8 = \text{True reliability of test Y}$$

Graphically:



Variance of the steady part of X = 9 = $\sigma^2(\epsilon_x)$

Variance of the steady part of Y = 8 = $\sigma^2(\delta_y)$

Variance of part of X common to Y = 7 = Variance of part of Y
common to X

Amount of variance of X explained by regression of X onto Y

$$= \rho_{xy} \sigma_x^2 = 84 / \sqrt{120}$$

Amount of variance of Y explained by regression of Y onto X

$$= \rho_{yx} \sigma_y^2 = 70 / \sqrt{120}$$

Unless Y is replicated, the variance of ξ_x cannot be separated from the variance of ϵ_x . If one desires a pseudo-psychological explanation, it might be reasoned that there are 7 (or 70 or 70×10^5) common "bonds" or synapses employed by the individuals for the two tests, X and Y.

(Psychologists do not postulate a finite or infinite set of bonds or synapses in reality; the terms and ideas are employed only as an aid to reasoning).

Assume, as for the standardized scores, that the specific factors are uncorrelated with each other as well as with the common part. Then if the

elements are assumed to be normal with zero means and variances σ^2 ,

$$E(\phi_{\alpha})^2 = \sigma^2(\phi_{\alpha}) = \sigma_{\gamma}^2 + \sigma_{\xi}^2$$

$$E(\theta_{\alpha})^2 = \sigma^2(\theta_{\alpha}) = \sigma_{\gamma}^2 + \sigma_{\eta}^2$$

$$E(\phi_{\alpha}, \theta_{\alpha}) = \sigma(\phi_{\alpha}, \theta_{\alpha}) = \sigma_{\gamma}^2$$

$$\sigma^2(Y) = \sigma_{\gamma}^2 + \sigma_{\xi}^2 + \sigma_{\epsilon}^2$$

$$\sigma^2(X) = \sigma_{\gamma}^2 + \sigma_{\eta}^2 + \sigma_{\delta}^2.$$

Then if the two tests are replicated n times (or if n parallel forms of each is administered) then the analysis of variance is:

<u>Source</u>	<u>df.</u>	<u>M.Sq.</u>	<u>E(M.Sq.)</u>
Tests (or times)	$n - 1$	$V_1(x,y)$	$\sigma_0^2 + \theta(\tau)$
Persons	$N - 1$	$V_2(x,y)$	$\sigma_0^2 + n\sigma^2$
Tests x persons	$(n - 1)(N - 1)$	$V_3(x,y)$	σ_0^2

($V(x,y)$ is the mean square for the joint analysis of two different tests on the same group of persons); where $\theta(\tau)$ is the fixed component due to differences between tests or times ($= N\sigma_t^2$ if tests are considered random), σ_{ξ}^2 is the variance of the common part, and σ_0^2 is the average unique (specific plus error) variance; that is

$$\sigma_0^2 = \frac{1}{2}(\sigma_{\xi}^2 + \sigma_{\epsilon}^2 + \sigma_{\eta}^2 + \sigma_{\delta}^2).$$

In general, $\sigma_x^2 \neq \sigma_y^2$ and $\sigma_\varepsilon^2 \neq \sigma_\delta^2$. If it is assumed that $\sigma_x^2 = \sigma_y^2$ and $\sigma_\varepsilon^2 = \sigma_\delta^2$, then $\sigma_x^2 + \sigma_\varepsilon^2 = \sigma_y^2 + \sigma_\delta^2$ and then V_3 would be an estimate of the unique variance of the two tests (now assumed to be equal) so that a test of the correlation between X and Y would be possible using $F = V_2(x,y)/V_3(x,y)$ (The null hypothesis to be tested is that $\sigma_\gamma^2 = 0$; i.e., there is nothing common between tests X and Y). The difference in "errors" is what makes interpretation of the analysis of variance more complex in psychological studies for testing purposes. However, the analysis of variance scheme may be used for estimation; $\frac{1}{n}[V_2(x,y) - V_3(x,y)]$ is an unbiased estimate of σ_γ^2 , the common factor variance. Let $V = [v_{ij}]$ be a matrix of variance components with elements defined as follows:

$$v_{ii} = \text{estimate of the variance component } \sigma_{pi}^2 \text{ (variance of the steady part of test } i \text{) obtained from the administration of the test } n \text{ times or from administration of } n \text{ parallel forms of test } i$$

$$= \frac{1}{n} [V_2 - V_3], \quad n \geq 2$$

$$v_{ij} = \text{estimate of the variance component } \sigma_\gamma^2 \text{ (variance of the part of tests } i \text{ and } j \text{ that is common to both)}$$

$$= \text{covariance of test } i \text{ and } j = S y_i y_j / (N - 1);$$

if the two tests are replicated n times, then from the analysis of variance table

$$= \frac{1}{n} [V_2(y_i, y_j) - V_3(y_i, y_j)], \quad n \geq 2.$$

Tukey, (1951) suggests that these component mean squares be used for a factor analysis. If it is desired to factor variances and covariances

(or correlations) that contain only common factors, then surely the v_{ij} ($i \neq j$) are the proper estimates. But $v_{ii}' = \hat{\sigma}_{pi}^2$ includes the specific part of the test; i.e., v_{ii}' is an estimate, say, of $\sigma^2(\phi_{\alpha}) = \sigma^2(\delta_{\alpha} + \xi_{\alpha})$. Thus v_{ii}' includes not only the estimate of the variance of the common part but also the estimate of the variance of the specific part. If we insist that the diagonal elements include only estimates of the variance of the common parts, then v_{ii}' is an over-estimate of the desired diagonal element. Thus a factor analysis of V would be identical to a factor analysis of R_r , the usual correlation matrix with reliabilities in the diagonals. The matrix V could be standardized to R_r by dividing through the i th row by $\sqrt{\text{est. variance of test } j}$. The diagonal elements of the standardized matrix would be $r_{ii}' = Sy_i y_i' / \sqrt{Sy_i^2 \cdot Sy_i^2}$, the product-moment correlation between the two replications of the same test. Consider the off-diagonal elements $\hat{\sigma}_{\gamma}^2 = \text{Cov}(X, Y)$. This quantity is the numerator of the product-moment correlation coefficient r_{xy} between X and Y . $\text{Cov}(X, Y)$ is also the numerator of the statistic

$$r'_{(x,y)} = \frac{\frac{1}{2} [V_2(x,y) - V_3(x,y)]}{\frac{1}{2} [V_2(x,y) + V_3(x,y)]} .$$

But the denominators of r_{xy} and $r'_{(x,y)}$ are different. ($r'_{(x,y)}$ is analogous to r'_c of equations (3.3) and (3.6) which measures the intra-class correlation of a given test). True, the product-moment correlation coefficient is the generally accepted statistic used to estimate the association parameter in the bivariate normal distribution, but why should it be used rather than $r'_{(x,y)}$? The point that is to be made is that both r_{xy} and $r'_{(x,y)}$ have as their numerators $\hat{\sigma}_{\gamma}^2$ which is the statistic that is

of importance, and that standardization beclouds the issue.

The model for the raw score is

$$Y_{i\alpha} = \eta_Y + \gamma_{i\alpha} + \xi_{i\alpha} + \varepsilon_{i\alpha}.$$

If this is converted to standardized form, $(Y_{i\alpha} - \bar{Y})/\sqrt{\text{Var}(Y)}$ when \bar{Y} is the sample mean and $\text{Var}(Y)$ the sample variance, then the resulting random variables are

$$\gamma_{i\alpha}' = \gamma_{i\alpha}/\sqrt{\text{Var}(Y)}$$

$$\xi_{i\alpha}' = \xi_{i\alpha}/\sqrt{\text{Var}(Y)}$$

$$\varepsilon_{i\alpha}' = \varepsilon_{i\alpha}/\sqrt{\text{Var}(Y)}$$

which are now ratios of random variables. For the standardized scores

$$s_{i\alpha} = \sum_{p=1}^r \alpha_{ip} X_{p\alpha} + \pi_{i\alpha} + \varepsilon_{i\alpha}.$$

The correspondence, then, is

$$\gamma_{i\alpha}' = \sum_p \alpha_{ip} X_{p\alpha}$$

$$\xi_{i\alpha}' = \pi_{i\alpha}$$

$$\varepsilon_{i\alpha}' = \varepsilon_{i\alpha}$$

and the steady part of both is

$$\begin{aligned}
 \gamma'_{i\alpha} + \xi'_{i\alpha} &= \text{steady part of test} \\
 &= \omega_{i\alpha} \\
 &= \sum_{p=1}^r \alpha_{ip} X_{p\alpha} + \pi_{i\alpha}
 \end{aligned}$$

The raw scores are easier to handle and manipulate. Also, expected values are straight-forward (assuming a linear model) for the raw scores while they are not for the standardized scores since the expected value of ratio of these two random variables is not equal to the ratio of the two expected values. The "standardization" procedure has as its purpose the equalizing, in some sense, the "variances" of the test scores. The raw scores are assumed to be of the form

$$Y_{i\alpha} = \eta_y + \gamma_{i\alpha} + \xi_{i\alpha} + \epsilon_{i\alpha} = \bar{Y} + c_{i\alpha} + b_{i\alpha} + e_{y_{i\alpha}}$$

$$X_{i\alpha} = \eta_x + \gamma_{i\alpha} + \xi_{i\alpha} + \delta_{i\alpha} = \bar{X} + c_{i\alpha} + d_{i\alpha} + e_{x_{i\alpha}}$$

where \bar{Y} , \bar{X} , c , b , and d are constants fitted to the sample and the e 's are the deviations of the observed values from the fitted values.

The standardization procedure takes estimates of the total variance of Y and X , $\text{Var}(Y)$ and $\text{Var}(X)$, and compute, say,

$$s_{i\alpha} = \frac{Y_{i\alpha} - \bar{Y}_i}{\sqrt{\text{Var}(Y)}}$$

$$s_{j\alpha} = \frac{X_{j\alpha} - \bar{X}_j}{\sqrt{\text{Var}(X)}}$$

This is an attempt to standardize the "errors". True, $\text{Var}(s_{i\alpha}) = \text{Var}(s_{j\alpha}) = 1$, but what is needed for analysis of variance procedures is the equivalence of $\sigma^2(\epsilon_\alpha)$ and $\sigma^2(\delta_\alpha)$. The quantity $\sum_\alpha [c_{y\alpha} / \sqrt{\text{Var}(Y)}]^2$ does not in general equal $\sum_\alpha [c_{x\alpha} / \sqrt{\text{Var}(X)}]^2$. If one had estimates $\hat{\sigma}^2(\epsilon_\alpha)$ and $\hat{\sigma}^2(\delta_\alpha)$, then one could compute for $\alpha = 1, \dots, N$

$$c_{y\alpha} / \sqrt{\hat{\sigma}^2(\epsilon_\alpha)} \text{ and } c_{x\alpha} / \sqrt{\hat{\sigma}^2(\delta_\alpha)}.$$

Now one could expect the error variances of the two tests X and Y to be approximately equal so that the analyses of variance would be appropriate (i.e., the assumption concerning equal "variances" would be approximately satisfied). The equalizing of what "variances" is what obscures the problem. The covariance of two tests alone, using raw scores, is a perfectly good measure of the degree of association between the two tests and seems a more fundamental statistic to use than the correlation coefficient.

If V is to be used by most psychologists, the diagonal elements would need to be reduced from v_{ii} to $v_{ii}^!$, say, which included only the estimate of the variance of the part of test i common to the other tests in the battery. (This is exactly the same problem of reducing the unities in the usual correlation matrix to communalities, h_i^2). Call this reduced matrix V_a . Some people assert that the elements v_{ii} should not be reduced but instead replaced by $\hat{\sigma}_{ii}$, the sample variance of test i. $\hat{\sigma}_{ii}$ includes the error components as well as the common and specific parts of the test.

Let V_b be the resultant matrix with v_{ii} replaced by $\hat{\sigma}_{ii}$.

Then the factorization of R_a , the reduced correlation matrix with diagonal elements h_i^2 , and V_a would be "similar". The word "similar" is used since it is not known what effect the standardization has on the extracted factors. A factorization of V and R_r , the correlation matrix with reliabilities in the diagonal would be similar; and a factorization of V_b and R , the usual correlation matrix would be similar.

From the common factor criterion, V_a and R_a should be used, V_a being the more fundamental of the two.

3.4 Reliability and communality

It has been suggested that the reliability coefficient of test i be used for the communality of test i in the correlation matrix which is to be factored. It has been stated, however, that the reliability is an over-estimate of the communality.

The reliability coefficient was defined in Chapter II, equation (2.11) for the standardized scores as

$$\rho_{ii'} = \frac{\sigma^2(c_i)}{\sigma^2(c_i) + \sigma^2(e_i)}$$

where

$\sigma^2(e_i)$ = true error variance of the random error term ($\epsilon_{i\alpha}$) for standardized test i

$\sigma^2(c_i)$ = true variance of the steady part $\psi_{i\alpha}$ for the standardized test i

$$e_{i\alpha} = s_{i\alpha} - \omega_{i\alpha}, \text{ and}$$

$$\omega_{i\alpha} = \frac{\sum_p \alpha_{ip} x_{p\alpha}}{p} + \pi_{i\alpha}.$$

For the raw test scores, the reliability was defined as

$$\rho_{ii'} = \frac{\sigma^2 p_i}{\sigma^2 p_i + \sigma^2 e_i}.$$

Conceptually, they are the same, and the estimate $r_{ii'}$ of $\rho_{ii'}$, whether using raw or standardized scores, will be the same. It was recommended that $r_{ii'}$ be computed using equation (3.3) if there is only one replication and equation (3.6) if there are n replications.

If the common factors are uncorrelated with the specific part, then

$$\begin{aligned} \text{Population variance of } \omega_{i\alpha} &= \sigma^2(c_i) \\ &= \text{Population variance of} \\ &\quad \left(\frac{\sum_p \alpha_{ip} x_{p\alpha}}{p} + \pi_{i\alpha} \right) \\ &= H_i^2 + \sigma^2(f_i), \end{aligned}$$

where H_i^2 , it will be recalled, is the population variance of the part of test i that is common to the other tests in the battery, and $\sigma^2(f_i)$ is the population specific variance. Thus

$$\sigma^2(\omega_{i\alpha}) + \sigma^2(\pi_{i\alpha}) = H_i^2 + \sigma^2(f_i) + \sigma^2(e_i)$$

= population variance of the standardized test i .

Thus

$$\sigma^2(c_i) \geq H_i^2. \quad (3.10)$$

The equality holds when $\sigma^2(f_i) = 0$; i.e., when the specific part of the test $\pi_{i\alpha}$ is zero. If $\sigma^2(c_i) + \sigma^2(e_i) = 1$, then dividing through both sides of (3.10) by $\sigma^2(c_i) + \sigma^2(e_i)$,

$$\frac{\sigma^2(c_i)}{\sigma^2(c_i) + \sigma^2(e_i)} \geq \frac{H_i^2}{\sigma^2(c_i) + \sigma^2(e_i)},$$

but

$$\frac{\sigma^2(c_i)}{\sigma^2(c_i) + \sigma^2(e_i)} = \rho_{ii'}, \quad (3.11)$$

by definition, so that

$$\rho_{ii'} \geq H_i^2. \quad (3.12)$$

The estimate of the reliability coefficient is $r_{ii'}$, and the estimate of the communality is h_i^2 . Due to the relationship (3.12), it is then stated that the estimate of the reliability coefficient $r_{ii'}$ is an over-estimate of h_i^2 . The over-estimation is due to the variance of the specific part of the test score, that fraction of the steady part of the test score that is not common to other tests.

3.5 Correlations corrected for attenuation

Psychologists use still another measure of correlation. Because of the unreliabilities of the tests, it is reasoned that the "true correlation" of the steady parts of the test is decreased. Let

$$r_{ij}(t) = \frac{r_{ij}}{\sqrt{r_{ii}' r_{jj}'}} \quad (3.13)$$

which is the product moment correlation coefficient between tests i and j adjusted (or corrected) for "attenuation". r_{ii}' and r_{jj}' are the estimates of the reliability coefficients for tests i and j , respectively. $r_{ij}(t)$ is an estimate of the quantity

$$\rho_t = \frac{E(\text{steady part of } X, \text{ steady part of } Y)}{\sqrt{E(\text{steady part of } X)^2 E(\text{steady part of } Y)^2}} \quad (3.14)$$

Since

$$\sigma^2(Y) = \sigma_\gamma^2 + \sigma_\xi^2 + \sigma_\varepsilon^2 = E(Y - \eta_Y)^2$$

$$\sigma^2(X) = \sigma_\delta^2 + \sigma_\gamma^2 + \sigma_\delta^2 = E(X - \eta_X)^2$$

$$E(\text{steady part of } X, \text{ steady part of } Y) = \sigma_\gamma^2 = E[(Y - \eta_Y)(X - \eta_X)]$$

$$E(\text{steady part of } Y)^2 = \sigma^2(\phi_\gamma) = \sigma_\gamma^2 + \sigma_\xi^2$$

$$E(\text{steady part of } X)^2 = \sigma^2(\theta_\gamma) = \sigma_\gamma^2 + \sigma_\delta^2$$

and r_{ij} is an estimate of

$$\rho_{ij} = \frac{E[(Y - \bar{y})(X - \bar{x})]}{\sqrt{E(Y - \bar{y})^2 \cdot E(X - \bar{x})^2}}$$

$$= \frac{\sigma_Y^2}{\sqrt{\sigma_Y^2 + \sigma_\xi^2 + \sigma_\varepsilon^2} \sqrt{\sigma_Y^2 + \sigma_\eta^2 + \sigma_\delta^2}}$$

and r_{ii}' and r_{jj}' are, respectively, estimates of

$$\rho_{ii}' = \frac{\sigma_Y^2 + \sigma_\xi^2}{\sigma_Y^2 + \sigma_\xi^2 + \sigma_\varepsilon^2}, \quad \rho_{jj}' = \frac{\sigma_Y^2 + \sigma_\eta^2}{\sigma_Y^2 + \sigma_\eta^2 + \sigma_\delta^2}$$

then, when the quantities of which r_{ij} , r_{ii}' , and r_{jj}' are estimates are substituted in (3.13), the result is

$$\rho_{ij}(t) = \frac{\sigma_Y^2}{\sqrt{\sigma_Y^2 + \sigma_\xi^2} \sqrt{\sigma_Y^2 + \sigma_\eta^2}}, \quad (3.15)$$

which is just (3.14). Thus, the correlation coefficient corrected for attenuation, $r_{ij}(t)$, is a proper estimate of ρ_t ; i.e., $\rho_t = \rho_{ij}(t)$. Each of the quantities in (3.15) are estimable from an analysis of variance table:

$$\sigma_Y^2 = \frac{1}{2} [V_2(X, Y) - V_3(X, Y)], \text{ for one administration of each test to the same individuals at the same time.}$$

$$\sigma_Y^2 + \sigma_\xi^2 = \frac{1}{2} [V_2(Y) - V_3(Y)], \text{ for 2 parallel forms of Y given to the same individuals}$$

$\sigma_{\gamma}^2 + \sigma_{\xi}^2 = \frac{1}{2} [V_2(X) - V_3(X)]$, for 2 parallel forms of X given to the same individuals.

3.6 Reliability and length of test

If the variance of a test of "unit" length is $\sigma_p^2 + \sigma_e^2$, the reliability is defined to be

$$\rho_{11'} = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_e^2},$$

the ratio of the steady part to the total variance of the test. If a test is replicated t times (or t parallel forms are administered), then the variance of the test of length t is

$$t\sigma_p^2 + \sigma_e^2.$$

The "steady" part is $t\sigma_p^2$ so that for a test of length t , the reliability is

$$\begin{aligned} \rho_{tt'} &= \frac{t\sigma_p^2}{t\sigma_p^2 + \sigma_e^2} = \frac{\frac{t\sigma_p^2}{\sigma_p^2 + \sigma_e^2}}{1 + \frac{(t-1)\sigma_p^2}{\sigma_p^2 + \sigma_e^2}} \\ &= \frac{t\rho_{11'}}{1 + (t-1)\rho_{11'}} \end{aligned}$$

Thus, theoretically, at least, the reliability of the test increases with increase in length.

Chapter IV

FACTOR ANALYSIS AND ANALYSIS OF VARIANCE

4.1 Univariate Analysis of Variance.

Burt (1947) takes a numerical example and argues that one gets answers from factor analysis that are similar to those from the analysis of variance. The example he used will be discussed here. The data consist of scores on 6 individuals for 4 tests: arithmetic, English, drawing, and handwork. The first two tests are academic and the last two, technical. In addition, arithmetic and drawing were given on Monday, English and handwork on Tuesday. The inter-correlations of the 4 tests are:

Table 4.1

		Arith.	Eng.	Draw.	Handw.	
Mon.	Arith.	1.0000				
Tues.	Eng.	.8808	1.0000			
Mon.	Draw.	.7125	.3753	1.0000		
Tues.	Handw.	.8477	.5193	.8247	1.0000	
Col.	Sums	3.4410	2.7754	2.9125	3.1917	12.3206

The factor loadings by the principal axis solution are shown in Table 4.2. (See Hotelling, 1933, pp. 417-41 and 498-520; or Thompson, 1948, Ch. VII, pp. 108-119).

Table 4.2

Factor loadings by principal axis
(or weighted summation) solution

tests \ factors	I	II	III	IV	S.Sq.
Arith.	.9791	.1908	.0275	.0610	1.0000
Eng.	.7850	.6126	-.0869	-.0376	1.0000
Draw.	.8310	-.4897	-.2640	-.0085	1.0000
Hand.	.9148	-.2851	.2851	-.0252	1.0000
S. Sq.	3.1021	.7328	.1593	.0058	4.0000
% of contr. to total variance	77.55	18.32	3.98	0.15	100.00

The correlation matrix of Table 4.1 was also factored by the centroid method, a simple summation method developed primarily by Thurstone. (See section 2 of Chapter II for rationale, and Thurstone, 1948, Ch. VIII and/or Thompson, 1948, Ch. V for details of the technique). The loadings by this method are given in Table 4.3.

Table 4.3

Factor loadings by centroid
(or simple summation) method

tests \ factors	I	II	III	IV	S.Sq.
Arith.	.9803	.1829	.0548	.0501	1.0000
Eng.	.7907	.6077	-.0548	-.0501	1.0000
Draw.	.8298	-.4828	-.2755	.0501	1.0000
Hand.	.9093	-.3078	.2755	-.0501	1.0000
S. Sq.	3.1016	.7306	.1578	.0100	4.0000
% of contr. to total variance	77.54	18.26	3.94	0.26	100.00

The loadings obtained by the two methods are very similar (Tables 4.2 and 4.3). The sign pattern for the loadings is the same except for factor IV, but the last factor is usually considered an "error" factor. The factors may be interpreted as

- I. General ability
- II. Academic vs. technical subjects
- III. Quantitative vs. aesthetic subjects
- IV. Monday vs. Tuesday subjects, or simply error.

The centroid method employs equal weights for each test. The resultant loadings for factor I are Arith. = .9803, Eng. = .7907, Draw. = .8298, Handw. = .9093. Since these loadings are not equal, it is argued that these weights should be used to weight the correlations to find the "true" loadings for factor I. An iterative process of this sort

results in convergence to a set of loadings for factor I which are identical to those for the principal axis solution, Table 4.2. Thus, the principal axis solution is sometimes referred to as the method of weighted summation.

Burt considers the unitary standardized test scores, $s_{i\alpha}^* = \frac{Y_{i\alpha} - \bar{Y}_i}{\sqrt{(N-1)\hat{\sigma}_{ii}}}$, such that

$$\sum_{\alpha} s_{i\alpha}^* = 0, \quad \sum_{\alpha} (s_{i\alpha}^*)^2 = 1,$$

and performs the usual analysis of variance on the $6 \times 4 = 24$ observations. By making the test means zero, the day means are also made zero.

Table 4.4
Analysis of Variance

Source	d.f.	adj.d.f.*	"S.Sq."	"M.Sq."(using adj. d.f.)
A. Persons	5	4	3.0810	.7702
B. Type of subj.	1			
C. Days	1			
AB. Persons x type of subj.	5	4	.6257	.1564
AC. Persons x days	5	4	.1521	.0380
BC. Type of subj. x days	1			
ABC. Residual	5	4	.1403	.0351

* adjusted by Burt "to allow for standardization"

The residual mean square is to be used to test all of the other effects for significance.

Next Burt observes that the average first factor loadings of the simple summation method is (from Table 4.3):

$$\begin{aligned} \frac{1}{4} (.9803 + .7907 + .8298 + .9093) &= \frac{1}{4} (3.5101) \\ &= .8776 \\ &= \frac{1}{4} \sum \sum r_{ij}. \end{aligned}$$

This quantity squared is

$$(.8776)^2 = .7702 = \frac{1}{16} \sum \sum r_{ij}^2,$$

which is 1/4 of the sum of squares for persons in Table 4.4. Similarly, it is shown that the (average of the absolute value of the second factor loadings)² is equal to 1/4 of the sum of squares for (AB), person x type of subject:

$$\begin{aligned} \left[\frac{1}{4} (.1829 + .6077 + .4828 + .3078) \right]^2 &= \left(\frac{1.5816}{4} \right)^2 \\ &= .1564. \end{aligned}$$

Thence, Burt suggests that by subtracting 1 from the degrees of freedom between persons "to compensate for the standardization of the tests" one can obtain a mean square equal to the square of mean factor loadings. But as will be shown below the ratio of the sum of squares to (mean factor loading)² has nothing whatever to do with degrees of freedom.

The other factors do not seem to correspond to the other sources in the analysis of variance table. Why do only the first two factors

apparently correspond to the first two sources in the analysis of variance?

The answer is found when the results are put in algebraic terms.

Denote the unitary standardized marks in the 4×6 table by $s_{i\alpha}^*$, $i = 1, 2, 3, 4$; $\alpha = 1, 2, \dots, N$. Recall that the first factor loading for the i th test (using the centroid method) is

$$b_{i1} = \frac{\sum_j r_{ij}}{\sqrt{\sum_i \sum_j r_{ij}}}$$

and

$$\begin{aligned} \sum b_{i1} &= \sum_i \sum_j r_{ij} / \sqrt{\sum_i \sum_j r_{ij}} \\ &= \sum_i \sum_j r_{ij} \\ &= \sqrt{\sum_i \sum_j \sum_{\alpha} S (s_{i\alpha}^* s_{j\alpha}^*)}. \end{aligned}$$

Thus

$$\begin{aligned} (\sum b_{i1})^2 &= \sum_i \sum_j r_{ij} = S \sum_i s_{i\alpha}^* \sum_j s_{j\alpha}^* \\ &= S \left(\sum_i s_{i\alpha}^* \right)^2 \\ &= n \text{ (Person "S. Sq.")}. \end{aligned}$$

Therefore

$$\left\{ \frac{\sum b_{i1}}{n} \right\}^2 = \frac{1}{n} \text{ (Person S. Sq.)}$$

$$= \frac{1}{n} \sum \sum r_{ij} .$$

Thus the relation is that the average first factor loading is the standard deviation between the means of the n standardized test scores per person. The factor n (Burt's "adjusted degrees of freedom") is merely the divisor introduced conventionally into the analysis of variance so that, on the null hypothesis, every mean square is an estimate of the variance of unit observations. Alternatively, one can say that the (mean factor loading)² is equivalent to the mean square between persons of standardized scores or to the sum of squares of unitary standardized scores, the degrees of freedom between persons having gone into the scaling. In other words, the first component taken out by the centroid method is simply, as in the analysis of variance, the unweighted sum of all test scores per person and does not, like the principal axis solution, seek the weighted compound of the scores with maximum variance.

The other sums of squares are evaluated in terms of the correlations and are shown in Table 4.5.

Table 4.5

Analysis of Variance of Unitary Standardized
scores in terms of correlations

<u>Source</u>	<u>d.f.</u>	<u>S. Sq.</u>
A. Persons	5	$\sqrt{4 + 2(r_{12} + r_{13} + r_{14} + r_{23} + r_{24} + r_{34})^2} / 4$
AB. Persons x type of subj.	5	$\sqrt{4 + 2(r_{12} - r_{13} - r_{14} - r_{23} - r_{24} + r_{34})^2} / 4$
AC. Persons x days	5	$\sqrt{4 + 2(-r_{12} + r_{13} - r_{14} - r_{23} + r_{24} - r_{34})^2} / 4$
ABD. Residual	5	$\sqrt{4 + 2(-r_{12} - r_{13} + r_{14} + r_{23} - r_{24} - r_{34})^2} / 4$

The second factor loadings are obtained by "reflecting" the residual correlations for drawing and handwork. One wants the second factor to correspond to the source AB, Persons x type of subject interaction. It can be shown after much tedious algebra that $\frac{1}{4}$ (sum of the absolute value of the second factor loading)² is

$$\frac{1}{4} \left(\sum |b_{i2}| \right)^2 = \frac{1}{4} \sqrt{4 + 2(r_{12} - r_{13} - r_{14} - r_{23} - r_{24} + r_{34})^2} - \frac{(r_{12} - r_{34})^2}{\sqrt{\sum \sum r_{ij}}} \quad 7.$$

The sum of squares for persons x type of subject is (Table 4.5)

$$\frac{1}{4} \sqrt{4 + 2 (r_{12} - r_{13} - r_{14} - r_{23} - r_{24} + r_{34})^2}.$$

The two expressions are only approximately equal. In Burt's example, it happens that the difference $(r_{12} - r_{34})^2 / \sqrt{\sum \sum r_{ij}} \approx .0002$, and is not detectable in his numerical comparison.

For n tests, the general expression for $\frac{1}{n}$ (the sum of the absolute values of the second factor loadings)² is

$$\frac{1}{n} \left(\sum |b_{i2}| \right)^2 = \frac{1}{n} \left[n + 2 \left(\sum_q + \sum_{n-q} - \sum_{q,n-q} \right) - 4 \frac{(\sum_q - \sum_{n-q})^2}{n \sqrt{\sum \sum r_{ij}}} \right]$$

where

\sum_q = sum of all correlation coefficients between the first $q < n$ unreflected tests

\sum_{n-q} = sum of all correlation coefficients between the $(n-q)$ reflected tests

and

$\sum_{q,n-q}$ = sum of all correlation coefficients between the q unreflected tests and the $(n-q)$ reflected tests.

The corresponding expression for the mean square of AB, Persons x type of subjects, using standardized scores (or sum of squares using unitary standardized scores) is

$$\frac{1}{n} \left[n + 2 \left(\sum_q + \sum_{n-q} - \sum_{q,n-q} \right) \right].$$

Obviously the two are approximately equal when the difference

$$4 \left(\sum_q - \sum_{n-q} \right)^2 / n \sqrt{\sum \sum r_{ij}}$$

is small.

When the effect of the second factor is subtracted from the reflected first residual matrix, the second residuals also sum to zero. If the second and fourth test vectors are reflected to correspond to the source AC, Person x days interaction, then

$$\begin{aligned} \frac{1}{4} \left(\sum |b_{i3}| \right)^2 &= \frac{1}{4} \left[4 + 2(-r_{12} + r_{13} - r_{14} - r_{23} + r_{24} - r_{34}) \right. \\ &\quad \left. -k_1 (r_{13} - r_{24})^2 \right. \\ &\quad \left. -k_2 (r_{12} - r_{34})(r_{13} - r_{24}) \right]. \end{aligned}$$

The corresponding mean square of the standardized scores (sum of squares for the unitary standardized scores) is (Table 4.5):

$$\frac{1}{4} \left[4 + 2(-r_{12} + r_{13} - r_{14} - r_{23} + r_{24} - r_{34}) \right].$$

For the Burt example, the sum of $k_1 f(r_{13}, r_{24})$ and $k_2 f(r_{12}, r_{34}, r_{13}, r_{24})$, where k_1 and k_2 are expressions like

$$1 / \sqrt{\sum \sum r_{ij}},$$

is large enough to cause a difference between the two expressions in the second decimal place. Thus

$$\left\{ \frac{\sum_{i=1}^4 |b_{i3}|}{n} \right\}^2 = .0276,$$

and (Table 4.4),

$$\frac{1}{4} (\text{Person S. Sq.}) = .0380.$$

A general expression could be written for each $(\sum |b_{ip}|)^2$, but the point has been made that the square of the sum of the factor loadings correspond to the sources only to the first order of difference among the r_{ij} . The correspondence is exact for the first factor but not for the other factors. The centroid method of factoring is a simple summation of correlation coefficients. And since the analysis of variance involves equal weighting of the observations, the two methods give results that may appear to be equal. Equivalence or near equivalence results when certain correlation coefficients are approximately equal.

A particular expression could be obtained for the fourth factor, and a general expression for the pth extracted factor. But the algebra becomes almost intractable the larger the number of extracted factors. Too, the expressions would depend on the way the tests were reflected after each factoring, so they would have little usefulness. Burt's example is not typical of factor analysis problems. Usually the tests cannot be so sharply grouped into academic vs. technical subjects, quantitative vs. aesthetic subjects, and so on. That is, there seldom are well defined sources of linear comparisons for an analysis of variance on batteries of tests assembled for a factorial study.

There appears to be no explicit relationship between the roots of $|R - \lambda I| = 0$ and the sums of squares that Burt has isolated. Even if the factor loadings found by the centroid method could be explicitly related to sources in the analysis of variance, the picture would not be

clarified, since the centroid solution is only an approximation to the principal axis solution. For the small example under discussion, the factor loadings for the two methods are similar, but for large batteries, only the first 2 or 3 sets of factor loadings are similar. That is, as more and more factors are extracted, the divergence becomes greater.

If the scores are not in unitary standard measure but the subject or test means are still zero, it can be shown that the results of an analysis of variance will be similar to that of Table 4.5. The unit Sums of Squares

$$\sum_i (s_{i\cdot}^*)^2 = 1$$

are replaced by $\hat{\sigma}_{ii}$ and the correlation coefficients by $\hat{\sigma}_{ij}$, where $\hat{\sigma}_{ii}$ and $\hat{\sigma}_{ij}$ are the sample variances and covariances for the tests. For example, the general mean square for Persons is

$$\frac{1}{n} \left\{ \sum \hat{\sigma}_{ii} + 2 (\hat{\sigma}_{12} + \hat{\sigma}_{13} + \dots + \hat{\sigma}_{n-1,n}) \right\} = \frac{1}{n} \sum_i \sum_j \hat{\sigma}_{ij},$$

and the mean square for residuals is

$$\frac{1}{n} \left[\sum \hat{\sigma}_{ii} + 2 \sum_i \sum_j \pm \hat{\sigma}_{ij} \right],$$

where the \pm depends on the linear comparisons isolated.

Instead of talking of correlations or of variances and covariances, one might ask what the analysis of variance table would look like if the test scores are written in the linear factor model

$$s_{i\alpha} = \sum_{p=1}^r a_{ip} x_{p\alpha} + \pi_{i\alpha} + e_{i\alpha},$$

where it is assumed that $r < n$ common factors account for all the inter-correlations of the tests; i.e., $r_{ij} = \sum_{p=1}^r a_{ip} a_{jp}$. Recall that h_i^2 and $\widehat{\sigma^2}(u_i)$ are the common factor "variance" and the unique "variance" of test i , where $u_{i\alpha} = f_{i\alpha} + e_{i\alpha}$ and $\widehat{\sigma^2}(u_i) = S u_{i\alpha}^2 / (N-1)$. Also $h_i^2 + \widehat{\sigma^2}(u_i) = 1$. It is found that the analysis of variance has the same form as presented earlier. For the set-up used above:

Analysis of Variance

Source

A. Persons	$\frac{1}{4}$	$\left[\sum \widehat{\sigma^2}(u_i) + \sum h_i^2 + 2(r_{12} + r_{13} + r_{14} + r_{23} + r_{24} + r_{34}) \right]$
AB. Persons x type of subj.	$\frac{1}{4}$	$\left[\sum \widehat{\sigma^2}(u_i) + \sum h_i^2 + 2(r_{12} - r_{13} - r_{14} - r_{23} - r_{24} + r_{34}) \right]$
AC. Persons x days	$\frac{1}{4}$	$\left[\sum \widehat{\sigma^2}(u_i) + \sum h_i^2 + 2(-r_{12} + r_{13} - r_{14} - r_{23} + r_{24} - r_{34}) \right]$
ABC. Residual	$\frac{1}{4}$	$\left[\sum \widehat{\sigma^2}(u_i) + \sum h_i^2 + 2(-r_{12} - r_{13} + r_{14} + r_{23} - r_{24} - r_{34}) \right]$

One might ask why the analysis of variance is used in these problems. The errors in the tests are generally quite different. The ruse of standardization seems to be of no help. The main difference between analysis of variance and factor analysis seems to be that the "sources" are stipulated a priori for the analysis of variance, while the "factors" are found a posteriori; i.e., at least for the principal axis solution,

the "factors" are chosen so as to account for the maximum variation at each extraction.

Finally, there is one other similarity between factor analysis and analysis of variance that should be mentioned. Both involve transformation matrices such that if G is the transformation matrix and S the matrix of observation, then $GS(GS)' = P$, a diagonal matrix. The application of a transforming matrix such that the result is a diagonal matrix is defined to be the reduction of a matrix to canonical form.

Let S be the matrix of standardized test scores s_{ij} , and L the matrix of direction cosines. Recall that $LL' = I$. L is obtainable from the factor matrix F by dividing each element of the p th column by $\sqrt{\lambda_p}$. Then $L'S$ gives a matrix such that if each element in the p th row is squared and summed, the result is $(N-1)\lambda_p$; i.e.

$$\frac{1}{(N-1)} L'SS'L = L'RL = (L'F)(F'L) = \Lambda^2 = \begin{bmatrix} \lambda_1 & & & 0 \\ & \cdot & & \\ 0 & & \cdot & \\ & & & \lambda_n \end{bmatrix}.$$

(See equation (2.26), Chapter II).

For Burt's example, it is easily verified that $(F'$, and consequently L' , obtained from Table 4.2)

$$L'RL = \begin{bmatrix} 3.1021 & & & \\ & .7328 & & 0 \\ & & .1593 & \\ & 0 & & .0058 \end{bmatrix},$$

where $\lambda_1 = 3.1021$, $\lambda_2 = .7328$, etc. (See Table 4.2).

From Table 4.3 (centroid factor loadings), the pth column is divided through by S.Sq. factor loadings for factor p and the result is G, say. Then it is easily verified that

$$G'RG \approx \begin{bmatrix} 3.1016 & & & \\ & .7306 & & 0 \\ & 0 & .1578 & \\ & & & .0100 \end{bmatrix}.$$

The analysis of variance may be viewed similarly. Consider the matrix weights

$$C = \begin{bmatrix} .5 & .5 & .5 & .5 \\ .5 & .5 & -.5 & -.5 \\ .5 & -.5 & .5 & -.5 \\ .5 & -.5 & -.5 & .5 \end{bmatrix}$$

C is obviously orthogonal. Next consider the matrix $C'S$. If each element of the pth row is squared and summed, the result is the sum of squares for the pth source (numbering downwards) in the above analysis of variance for Burt's example. Thus

$$\frac{1}{N-1} C'SS'C = C'RG = \begin{bmatrix} 3.0810 & & & \\ & .6257 & & 0 \\ & & .1521 & \\ & 0 & & .1403 \end{bmatrix}.$$

where 3.0810 = sum of squares for persons, .6257 = sum of squares for Persons x type of subject, etc.

Since the centroid method is an approximation to the principal axis solution, the closeness of the figures in the diagonal of L'RL and G'RG is not surprising. And the similarity of the elements of G'RG and C'RC is due to the fact that the "factors" in the centroid method roughly correspond, at least for this Burt-example, to the "sources" in the analysis of variance. (Note: The leading element in G'RG is, from Table 4.3, $\sum b_{i1}^2 = (.9803)^2 + \dots + (.9093)^2; = 3.1016 \neq$ S.Sq. for Persons; the equivalence between the centroid method and factor analysis was shown for $(\sum b_{i1})^2 = n$ Person S. Sq. This explains why the (1,1) element of G'RG \neq the (1,1) element of C'RC).

4.2 Multivariate analysis of variance and canonical correlation analysis.

Canonical correlation analysis is a study of the relationship between two different sets of variates. This is sometimes called an external factor analysis. Factor analysis takes only one set of variates and studies the internal relationships of only this one set. This is called an internal factor analysis. (Bartlett, 1948). In this section, an attempt is made to show that the two types of analyses can be related. First, some new notation and terminology is necessary. The discussion will first concern unstandardized or raw test scores.

Let S_1 be a $p \times N$ observation matrix of p test scores on N individuals. This will be labeled the independent set of tests. Let S_2 be a $q \times N$ observation matrix of test scores, q tests and N individuals, the dependent

set. The experimenter wants to predict S_2 from S_1 . Let B_{21} be a $q \times p$ matrix of regression coefficients, and $S_{2.1}$ the residual part of S_2 not accounted for by the regression of S_2 on S_1 . Then one may write

$$S_2 = B_{21}S_1 + S_{2.1}, \quad (4.1)$$

where S_1 and $S_{2.1}$ are assumed to be mutually orthogonal.

Post-multiplying (4.1) by S_1' ,

$$\begin{aligned} S_2 S_1' &= B_{21} S_1 S_1' + S_{2.1} S_1' \\ &= B_{21} S_1 S_1', \end{aligned}$$

or

$$\begin{aligned} B_{21} &= (S_2 S_1') (S_1 S_1')^{-1} \\ &= C_{21} C_{11}^{-1}, \end{aligned}$$

where $C_{21} = S_2 S_1'$, $C_{11} = S_1 S_1'$, and C_{11}^{-1} is the inverse matrix of C_{11} such that $C_{11} C_{11}^{-1} = I$, the identity matrix.

Post-multiplying both sides of (4.1) by its own transpose

$$S_2 S_2' = B_{21} S_1 S_1' B_{21}' + S_{2.1} S_{2.1}',$$

or

$$C_{22} = C_{21} C_{11}^{-1} C_{21}' + C_{22.1}, \quad (4.2)$$

where

$$C_{22} = S_2 S_2', \quad C_{22.1} = S_{2.1} S_{2.1}'.$$

Recall that in Chapter II, the general factor equation was written as

$$S = F_a X + U,$$

where S is the $n \times N$ observation matrix of n standardized test scores, $F_a = [\bar{a}_{ik}]$, the $n \times r$ matrix of factor loadings, $r < n$ common factors, and U the $n \times N$ matrix of unique parts. It is assumed that the common factors are uncorrelated with the unique factors in the population and are made to be orthogonal in the sample. In this light, S is the dependent set of scores, X , the independent variates, and F_a , the matrix of factor loadings or regression coefficients. Formally,

$$S X' = F_a X X' + U X'.$$

But $\frac{1}{N-1} X X' = I$ and $U X' = 0$, so that

$$F_a = \frac{S X'}{N-1} \quad (4.3)$$

If the common factors X were known, the loadings could be calculated according to (4.3). However, the $X_{p \times}$ are found only after the correlation matrix is factored. Further

$$\frac{1}{N-1} S S' = R = \frac{1}{N-1} F_a X X' F_a' + \frac{1}{N-1} U U',$$

or

$$R = F_a F_a' + \frac{1}{(N-1)} U U',$$

where

$$F_a F_a' \simeq R_a, \quad \frac{1}{N-1} U U' \simeq R_u$$

(equation 2.5, Chapter II). R_a is the amount of R explained by $r < n$ common factors. In equation (4.2), $C_{21} C_{11}^{-1} C_{21}'$ is defined to be the amount of the total variation $C_{22} = S_2 S_2'$ accounted for by the regression of S_2 onto S_1 . R_u is the matrix of error variances and covariances similar to $C_{22.1}$.

Next consider an arbitrary row matrix, a' , with q elements. Premultiply (4.1) by the row matrix a' (a' is not a row of F_a):

$$a' S_2 = a' B_{21} S_1 + a' S_{2.1}.$$

Corresponding to (4.2), there is the equation

$$a' C_{22} a = a' C_{22} C_{11}^{-1} C_{21}' a + a' C_{22.1} a.$$

The problem is to find the vector \underline{a} such that the sum of squares due to the regression of $a' S_2$ onto $a' B_{21} S_1$ relative to $a' C_{22} a$ will be a maximum. This may be found by maximizing the ratio

$$a' C_{21} C_{11}^{-1} C_{12} a / a' C_{22} a.$$

The solution requires finding an \underline{a} such that

$$(C_{21} C_{11}^{-1} C_{12} - r^2 C_{22}) a = 0. \quad (4.4)$$

This has a non-trivial solution when

$$\left| C_{21} C_{11}^{-1} C_{12} - r^2 C_{22} \right| = 0 \quad (4.5)$$

Pre-multiplying (4.4) by a' and solving for r^2 ,

$$r^2 = a' C_{21} C_{11}^{-1} C_{12} a / a' C_{22} a,$$

the fraction of the total sum of squares which was to be maximized. Thus, choose r^2 to be the largest root of (4.5). r^2 is usually written as λ and the problem is to find all the roots of $|C_{21} C_{11}^{-1} C_{12} - \lambda C_{22}| = 0$. Number the roots $\lambda_1, \lambda_2, \dots$ from the largest downward. There will in general be the smaller of (p, q) distinct roots. For each root λ_m that is found and substituted back into (4.4), there corresponds a canonical vector a_m . Then $a_m' S_2$ and $a_m' B_{21} S_1 = a_m' S_2 S_1' (SS')^{-1} S_1$ are two new vectors which are called the mth canonical variate. The correlation of the two new variates is the mth canonical correlation coefficient, $\sqrt{\lambda_m}$. It can be shown that the new canonical variates are independent. (Hotelling, 1936). For definiteness suppose that $p < q$. If the correlations between the variates before transformation was

$$k = \begin{matrix} & \begin{matrix} 1 & 2 & \dots & q \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ \vdots \\ \vdots \\ p \end{matrix} & \begin{bmatrix} \rho_{11} & & \dots & \rho_{1q} \\ & & & \\ & & & \\ \rho_{p1} & & \dots & \rho_{pq} \end{bmatrix} \end{matrix} \quad (4.6)$$

Then the transformation results in the matrix

$$k^* = \begin{matrix} & \begin{matrix} 1 & 2 & \dots & q \end{matrix} \\ \begin{matrix} \rho_1 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{matrix} & \begin{bmatrix} \rho_1 & 0 & \dots & 0 \\ 0 & \rho_2 & & \\ \vdots & \vdots & & \\ \vdots & \vdots & & \\ 0 & 0 \dots \rho_p & 0 & 0 \dots 0 \end{bmatrix} \end{matrix}, \quad (4.7)$$

where the ρ_m are the canonical correlations; i.e., $\rho_m = \sqrt{\lambda_m}$.

Recall that for internal factor analysis, the problem was to solve ℓ for

$$(R - \lambda I) \ell = 0,$$

which has a non-trivial solution when

$$|R - \lambda I| = 0,$$

giving n distinct roots, λ_p . And the matrix of latent vectors $L = [\ell_{ip}]$, transforms R to a diagonal form

$$L'RL = \begin{bmatrix} \lambda_1 & & & 0 \\ & \cdot & & \\ & & \cdot & \\ 0 & & & \lambda_n \end{bmatrix}. \quad (4.8)$$

The rank of the matrix k^* in (4.7) is $r < p$, the number of non-vanishing correlations. That is, there are r common factors between the two sets of variates S_1 and S_2 . Similarly, the rank of the matrix $L'RL$ in (4.8) is the number of non-vanishing roots $r < n$; that is, the number of common factors in the one set of variates S .

These are interesting analogies between external and internal factor analysis. To be more specific, consider the results of a testing procedure that involves n tests, each given at t different times to N individuals. The multivariate analysis of variance table for the scores has the following form ("raw" or unstandardized scores are used):

<u>Source</u>	<u>d.f.</u>	<u>S.Sq. and S.C.P.</u>
Between persons	$N - 1$	A
Between times	$t - 1$	B
Persons x times	$(N-1)(t-1)$	C

where A, B, and C are symmetric $n \times n$ matrices of sums of squares (S.Sq.) and sums of cross-products (S.C.P.) of the n tests. The general equation in (4.5) becomes

$$| A - \lambda (A+C) | = 0. \quad (4.9)$$

The two sets of variates S_1 and S_2 are not as distinct in this set-up as they were in the set-up described earlier. One might consider the N persons or individuals as the independent set of variates and the n tests as the dependent set. The analysis of variance model is a regression equation and A represents the sum of squares accounted for by the regression of tests onto the dummy variates of $\underline{0}$ and $\underline{1}$ for the persons. A, then, corresponds to $C_{21} C_{11}^{-1} C_{12}$ and $A + C$ to C_{22} . The differences between times is isolated as a source of variation which is of no interest in this discussion. There are several variations of (4.9). Sometimes it is written as $| A - \mu C | = 0$ or $| C - \nu(A+C) | = 0$. But $\lambda = 1 - \nu$, $\nu = \frac{1}{1+\mu}$, and the canonical variates are the same for all determinations. (Rao, 1952).

Bartlett (1948) suggests various operations on (4.9) so as to make it look like $| R - \lambda I | = 0$. First, he substituted

$$\lambda = \frac{\mu}{m+\mu}$$

where $m = (N-1)(t-1)$, the degrees of freedom for C . Then (4.9) may be written as

$$\left| A - \mu \frac{C}{m} \right| = 0. \quad (4.10)$$

Let $\theta = (N-1) \mu$. Dividing the matrix A through by $(N-1)$ reduces it to a variance covariance matrix, D . If m is larger or E , the true matrix of variances and covariances of the residuals or errors, is known otherwise, then (4.10) may be written as

$$\left| D - \theta E \right| = 0. \quad (4.11)$$

The (i,i) th element of D is

$$\frac{1}{N-1} \left\{ \frac{1}{t} \sum_{\alpha} \left(\sum_t Y_{it\alpha} \right)^2 - \frac{1}{tN} \left(\sum_{\alpha} \sum_t Y_{it\alpha} \right)^2 \right\},$$

where $Y_{it\alpha}$ is the test score on the i th test at the t th time as made by the α th individual. Now $Y_{it\alpha}$ is postulated to be of the form

$$Y_{it\alpha} = \mu + \tau_i + \beta_t + \rho_{\alpha} + \varepsilon_{it\alpha},$$

where μ = mean effect, τ_i = effect of test i , β_t = effect of time t , ρ_{α} the α th person effect, and $\varepsilon_{it\alpha}$ the usual random error. If the person effects are random and the test and time effects fixed, then

$$E(Y_{it\alpha}) = \mu + \tau_i + \beta_t = \eta_i$$

$$E(Y_{it\alpha} - \eta_i)^2 = \sigma_{e_i}^2 + \sigma_{p_i}^2.$$

The expected value, then, of the $(i,i)^{\text{th}}$ element of D is

$$\sigma_{e_i}^2 + t \sigma_{p_i}^2$$

(This is the result of section 3.6, Chapter III). And the $(i,j)^{\text{th}}$ element is

$$\frac{1}{N-1} \left[\frac{1}{t} \sum_{\alpha} (\sum_t Y_{it\alpha} \cdot \sum_t Y_{jt\alpha}) - \frac{1}{tN} (\sum_{\alpha} \sum_t Y_{it\alpha}) (\sum_{\alpha} \sum_t Y_{jt\alpha}) \right],$$

whose expected value is $t\sigma_Y^2(i,j)$, the variance of the common parts of tests i and j . Ordinarily, the tests are not replicated ($t = 1$) and the variances and covariances are taken over all items. In D, the estimates of the variances and covariances are over the N persons with the means of the times eliminated.

Further assume that the errors are uncorrelated, so that E becomes a diagonal matrix, E^* . Next standardize the D matrix by dividing each element of the i^{th} column by $\sqrt{(i,i)^{\text{th}} \text{ element of } D} = \sqrt{\sigma_{e_i}^2 + \sigma_{p_i}^2}$ and divide each element of the j^{th} row by $\sqrt{(j,j)^{\text{th}} \text{ element of } D} = \sqrt{\sigma_{e_j}^2 + t\sigma_{p_j}^2}$. The result is a correlation matrix R_t . Then (4.11) becomes

$$| R_t - \theta E^{**} | = 0. \quad (4.12)$$

The $(i,i)^{\text{th}}$ element of E^{**} is $\frac{\sigma_{e_i}^2}{\sigma_{e_i}^2 + t\sigma_{p_i}^2}$, which is the unreliability

of test i which is now of length t . (The variance of a test of "unit" length is $\sigma_{e_i}^2 + \sigma_{p_i}^2$, and of length " t ", $\sigma_{e_i}^2 + t\sigma_{p_i}^2$, see Chapter III).

Assume that the diagonal elements of D are "fairly" stable so that

$$\frac{\sigma_{e_i}^2}{\widehat{\sigma_{e_i}^2} + t \widehat{\sigma_{p_i}^2}}$$

is "fairly" stable. Then assume that all tests have the same reliability, $\rho_{ii} \approx .95$ say. Then the diagonal elements of E^{**} are all equal to .05. Let $\lambda = .05 \theta$. Then (4.12) becomes

$$|R_t - \lambda I| = 0, \quad (4.13)$$

which is similar to the usual factor analysis determinantal equation. R_t differs from R in that for R_t , the tests are replicated t times, with time effects eliminated. The correlations in R are obtained from single replications of each test.

Consider the case $n = 2$, for simplicity. If the expected values are substituted in $|D - \theta E| = 0$, the determinant of the expected values has the form

$$\begin{vmatrix} t \sigma_{p1}^2 + \sigma_{e1}^2 - \theta \sigma_{e1}^2 & t \sigma_{\gamma}^2 \\ t \sigma_{\gamma}^2 & t \sigma_{p2}^2 + \sigma_{e2}^2 - \theta \sigma_{e2}^2 \end{vmatrix} = 0.$$

Let $(1-\theta) = \lambda t$ and factor out t from each row and column, the result is

$$\begin{vmatrix} \sigma_{p1}^2 - \lambda \sigma_{e1}^2 & \sigma_{\gamma}^2 \\ \sigma_{\gamma}^2 & \sigma_{p2}^2 - \lambda \sigma_{e2}^2 \end{vmatrix} = 0.$$

The part corresponding to D,

$$E(V) = \begin{vmatrix} \sigma_{p1}^2 & \sigma_{\gamma}^2 \\ \sigma_{\gamma}^2 & \sigma_{p2}^2 \end{vmatrix},$$

is just the expected value of the matrix V discussed in Chapter III, the matrix of variance components that Tukey suggests factoring. If the σ_{ei}^2 are "small" relative to σ_{pi}^2 and σ_{γ}^2 , then a factorization of V should give results very similar to those of $|A - \lambda(A+C)| = 0$.

For a sample, E of (4.11) will not be a diagonal matrix, for the errors are correlated. But the assumption of equal reliabilities is not too stringent. Most tests used in a battery have established high reliabilities $>.90$, say. A solution of $|D - \theta E| = 0$ (or $|A - \lambda(A+C)| = 0$) should give factors very similar to the factors found from $|R - \lambda I| = 0$.

Results similar to that of (4.13) may be developed by a different approach. The general factor analysis equation is (for standardized scores)

$$S = F_a X + U.$$

Consider the matrix S of test scores as the dependent set of variates, F_a , a matrix of regression coefficients, and X, the matrix of the "independent" set of variates (the hypothetical factors). Then

$$(N-1) R = SS' = F_a X X' F_a' + U U'$$

and

$$R = F_a F_a' + \frac{1}{N-1} U U'.$$

Assume that the $r < n$ common factors explain all the intercorrelations so that $F_a F_a' = R_a$, exactly and $\frac{1}{N-1} U U' = R_u$. Recall that

$$R = R_a + R_u.$$

R_a is the reduced correlation matrix with communalities h_i^2 in the diagonal. Since $F_a F_a' = R$ exactly, the U 's are uncorrelated so that R_u is a diagonal matrix with unique (specific plus error) variances in the diagonal. Then analogous to (4.9),

$$\left| F_a F_a' - \lambda R \right| = \left| R_a - \lambda R \right| = 0. \quad (4.14)$$

To say that $F_a X$ accounts for all the inter-correlations, which are due to common causes or factors, is similar to the requirement that the errors are uncorrelated in the development of (4.13). Write $R_a = R - R_u$ and let $1 - \lambda = \frac{1}{\nu}$, then (4.14) becomes

$$\left| R - \nu R_u \right| = 0. \quad (4.15)$$

This result is comparable to (4.12), $\left| R_t - \theta E^{**} \right| = 0$. However, the (i,i) th element of R_u , $\widehat{\sigma^2}(u_i)$, consists of the specific plus error variances, i.e., $\widehat{\sigma^2}(u_i) = \widehat{\sigma^2}(f_i + e_i)$. The (i,i) th element of R is h_i^2 , which is the "variance" due to common factors; but because of replication, the (i,i) th element of D includes the variance of the specific part of test i as well as the random error; i.e., it is the variance of test i with time effects eliminated, which is standardized to unity in R_t . The (i,i) th element of E^* is $\widehat{\sigma_{e_i}^2}$, the variance of the random error term only, and for the standardized E^{**} , the element is

$$\frac{\hat{\sigma}_{e_i}^2}{\hat{\sigma}_{p_i}^2 + \hat{\sigma}_{e_i}^2}$$

which, for the standardized scores is equal to $\hat{\sigma}^2(e_i)$. Thus E^{**} includes only the variance of the random error term while R_u includes the variance of the specific plus the variance of the random error term.

A slightly different result is possible if in (4.14) R is written as the sum of R_a and R_u , and $\theta/(1+\theta)$ is substituted for λ . The determinantal equation to be solved is

$$\left| R_a - \theta R_u \right| = 0. \quad (4.16)$$

If one assumes equal unique (not error) variances, then (4.15) and (4.16) reduce to

$$\left| R - \lambda I \right| = 0 \text{ and } \left| R_a - \lambda I \right| = 0,$$

respectively. This is a more restrictive assumption than requiring the reliabilities to be equal. Instead of requiring all ρ_{ii} to be equal, it is necessary to require that all H_i^2 be equal. Before the analysis, the experimenter usually has little idea as to the relative sizes of the communalities, but the reliability is usually established before it is used in a factor analysis.

Burt (1949) defines two matrices

$$R_d = R_u^{-\frac{1}{2}} R R_u^{-\frac{1}{2}} \quad (4.16)$$

and

$$R_e = R_u^{-\frac{1}{2}} R_a R_u^{-\frac{1}{2}}, \quad (4.17)$$

where

$$R_u^{-\frac{1}{2}} = \begin{bmatrix} \frac{1}{\sigma(\hat{u}_1)} & & & \\ & \frac{1}{\sigma(\hat{u}_2)} & & 0 \\ & & \ddots & \\ 0 & & & \frac{1}{\sigma(\hat{u}_n)} \end{bmatrix} \cdot \quad (4.18)$$

Then from (4.15) and (4.16), there results

$$|R_d - \nu I| = 0 \quad (4.19)$$

and

$$|R_e - \theta I| = 0. \quad (4.20)$$

The typical off-diagonal element of R_d and R_e is

$$\frac{r_{ij}}{\sigma(\hat{u}_i) \sigma(\hat{u}_j)} = \frac{r_{ij}}{\sqrt{1-h_i^2} \sqrt{1-h_j^2}} \cdot$$

Burt calls this "adjusting for uniqueness". This is perhaps a poor choice of words for the reliability coefficient adjusted for attenuation, $r_{ij}/\sqrt{r_{ii} r_{jj}}$, where r_{ii} and r_{jj} are the reliability coefficients of tests i and j , respectively, has an entirely different purpose in mind. Because of unreliabilities of the tests (errors of measurements), the correlation is decreased. Thus, the "true" correlation is actually higher than the sample correlation coefficient. The

"true" correlation is estimated by dividing through by the geometric mean of the two reliability coefficients. The adjusted coefficients in R_d and R_e do have some appeal, however, for a test with all common factors would get infinite weighting since its communality would be one, and tests with no common factors would get no boost from the divisor $1/\sqrt{1 - h_i^2}$, for then h_i^2 would be zero. The purpose of a factorial study is to isolate and identify basic parameters in the domain under study. The assembling of a battery of tests does not have as its purpose the measurement of all factors related to a certain job, e.g. a lathe operator. For that purpose, surely one would want tests with specific parts, if those parts help to measure and thus predict performance as a lathe operator. Some people criticize this aspect of factor analysis and ask why should a portmanteau test consisting of all common factors get an infinite weight and a test with a certain specific factor get no increase in weight? When it is realized what the aim of a factorial study is, this may be rationalized. However, this all seems unnecessary. The numerator of the product moment correlation is an estimate of σ_{xy}^2 , the covariance of the steady parts of two tests. If the two tests have many common factors, then the estimate of σ_{xy}^2 is large. If they have nothing in common, then the estimate is near zero.

There is one other point that should be made. Is the set of loadings, F , the best that can be obtained so that the correlation is a maximum between the independent set of variates (the common factors) and the dependent set of variates (the test scores)? This is the problem of canonical correlation analysis. Again treat the problem formally, and write

$$S = F_a X + U.$$

A factor matrix F_a has been found whose factors are orthogonal (as the transformed canonical variates are orthogonal), that diagonalizes the matrix of observations S . It might be asked, then, is the matrix F_a so determined that the "sum of squares" due to extracted factors is a maximum? This may be checked by considering the new vector $b'F_a'S$ as the new vector $a'S_2$ was considered for the usual canonical problem. Thus

$$b'F_a'S = b'F_a'F_a X + b'F_a'U,$$

and

$$b'F_a'SS'F_a b = b'F_a'F_a XX'F_a'F_a b + b'F_a'UU'F_a b.$$

The problem is to maximize the ratio

$$b'F_a'F_a XX'F_a'F_a b / b'F_a'SS'F_a b$$

with respect to b . The solution is obtainable from

$$(F_a'F_a XX'F_a'F_a - \lambda F_a'SS'F_a) a = 0,$$

which has a non-trivial solution when

$$\left| F_a'F_a XX'F_a'F_a - \lambda F_a'SS'F_a \right| = 0. \quad (4.21)$$

But $\frac{1}{(N-1)} XX' = I$, $\frac{1}{(N-1)} SS' = R$, $F_a'F_a = \mathcal{L}_a^2$,

$F_a'R F_a = \mathcal{L}_a^4$, so that (4.21) becomes

$$\left| \mathcal{L}_a^4 - \lambda \mathcal{L}_a^4 \right| = 0 \quad (4.22)$$

giving all r roots $\lambda_p = 1$; i.e., all correlations, $\sqrt{\lambda_p}$, are unity. Thus, the principal axis solution is the best that can be found for the maximum explanation of the total variation in the correlation matrix using $r < n$ factors. This fact can be more elegantly proved (Kendall, 1950), but the demonstration shows how, at least formally, the factor analysis problem can be treated as the general regression problem of one set of variates onto another hypothetical set; and, in turn, through the devices of canonical correlation analysis, shows that the matrix of loadings, F_a , is the "best" transformation possible, "best" in the sense of accounting for the maximum variation using $r < n$ common factors.

Chapter V

DISCRIMINATION AND IDENTIFICATION

It was mentioned in the introduction that Galton and Pearson were concerned with problems of classification. The psychologists adapted and developed the ideas to a study of psychology, resulting in a technique called factor analysis, which has as its purpose description and identification. Since Galton and Pearson, there have been many advances in multivariate techniques that are used in many areas of research. However, factor analysis has developed virtually independently of these other methods. In this chapter, several multivariate techniques will be discussed to see how they are basically related. Hotelling's most predictable criterion will be mentioned, followed by a demonstration of how the usual discriminant problem may be considered as a special case of the general canonical correlation problem. An example of the use of canonical correlation analysis will be given, and the parallel thinking involved in canonical correlation analysis (external factor analysis) and factor analysis (internal factor analysis) will be emphasized.

In Chapter IV, section 2, it was shown how the determinantal equation of canonical correlation analysis may be reduced to the determinantal equation of factor analysis by imposing several simplifying conditions. In this chapter, the thinking involved in the two methods will be discussed, over and above the mathematical conditions necessary to effect equivalence. First, two special cases of canonical correlation analysis will be given.

Recall that for two sets of test scores (S_1 , the matrix of

observations for the independent set, and S_2 , the matrix of observations for the dependent set), the solution of \underline{a} from the equation

$$(C_{21} C_{11}^{-1} C_{12} - \lambda C_{22}) \underline{a} = 0$$

gives a column vector \underline{a} of coefficients such that $\underline{a}'S_2$ can best be predicted from $\underline{a}'B_{21}S_1 = \underline{a}'C_{21}C_{11}^{-1}S_1$, where \underline{a} depends on λ which is the largest root of

$$|C_{21} C_{11}^{-1} C_{12} - \lambda C_{22}| = 0. \quad (5.1)$$

The linear combination of test scores in the dependent set that is best predicted by the linear combination of scores in the independent set is called Hotelling's "most predictable criterion". (Hotelling 1935).

The second special case of canonical correlation analysis results when there is only 1 degree of freedom in one of the sets of variates. This was observed by Bartlett in 1938. Recall that the discriminant problem may be viewed as a multiple regression problem. Formally reverse the roles and consider the tests or predictors as the dependent set, S_2 , and the vector which differentiates between the two populations as the independent "set", S_1 . Let S_1 be a row vector with $N = N_1 + N_2$ elements, with each element being N_2/N or $-N_1/N$ according as the individual observed and measured belongs to population 1 or 2, respectively. N_1 is the number of individuals in the sample from population 1 and N_2 , the number in the sample from population 2. Suppose that n tests are being used as predictors or discriminators, then if S_2 is an $n \times N$ matrix,

$$S_2 S_1 = \frac{N_1 N_2}{N} \begin{bmatrix} d_1 \\ d_2 \\ \cdot \\ \cdot \\ d_n \end{bmatrix} = \frac{N_1 N_2}{N} d$$

where $d_i = \bar{Y}_{i1} - \bar{Y}_{i2}$, the difference in mean values for the i th discriminator or test between the two populations. $\bar{Y}_{ij} = \sum_{\alpha=1}^{N_j} Y_{ij\alpha} / N_j$, $j = 1, 2$; $\alpha = 1, 2, \dots, N_j$; $i = 1, 2, \dots, n$, where $Y_{ij\alpha}$ is the measurement on the α th individual for the i th test in population j . Then

$$S_1 S_1' = N_1 \left(\frac{N_2}{N}\right)^2 + N_2 \left(-\frac{N_1}{N}\right)^2 = \frac{N_1 N_2}{N} = K.$$

Thus

$$C_{21} C_{11}^{-1} C_{12} = S_2 S_1' (S_1 S_1')^{-1} S_1 S_2' = (Kd) \left(\frac{1}{K}\right) (d'K) = \frac{1}{K} dd',$$

where d is the column matrix of mean differences. Obviously, dd' is a matrix of rank 1. In equation (5.1) let $\mu = K\lambda / (1 - \lambda)$ and $\frac{1}{K} dd' = C_{21} C_{11}^{-1} C_{12}$. Recall that $C_{22} = C_{21} C_{11}^{-1} C_{12} + C_{22.1}$, so that (5.1) may be written as

$$\left| dd' - \mu C_{22.1} \right| = \left| dd' C_{11}^{-1} C_{12} - \mu I \right| = 0, \quad (5.2)$$

thus $dd' C_{11}^{-1} C_{12}$ is also of rank 1. The expansion of an equation of the form $|Q - \mu I| = 0$ is

$$(-1)^n \mu^n + (-1)^{n-1} p_1 \mu^{n-1} + (-1)^{n-2} p_2 \mu^{n-2} + \dots + c = 0.$$

This equation has only one root if Q is of rank 1. The coefficient P_1 of μ^{n-1} is the sum of all the roots which is also equal to the trace of Q . By definition, the trace of Q is the sum of the diagonal elements of Q . Further, P_2 , the coefficient of μ^{n-2} , involves the product of the roots of Q taken two at a time. Since only one root is non-zero, P_2 is zero. And in general, the coefficients of μ^{n-r} , $r \geq 2$, involves the product of the roots taken r at a time. Thus all the coefficients are zero except for the first two terms. The solution of

$$(-1) \mu^n + (-1)^{n-1} P_1 \mu^{n-1} = 0$$

is

$$\begin{aligned} \mu &= \sum (\text{roots}) = \text{trace of } Q \\ &= \text{trace of } dd' C_{22.1}^{-1}. \end{aligned}$$

But trace of $dd' C_{22.1}^{-1} = \text{trace } d' C_{22.1}^{-1} d$. This is seen to be so when the individual terms are considered. Let $A = d$, $C = d' C_{22.1}^{-1}$. Then it is stated that the trace of $AC = \text{trace of } CA$. But

$$\begin{aligned} \text{trace of } AC &= \sum_{i=1}^r \sum_{j=1}^s a_{ij} c_{ji} \\ \text{trace of } CA &= \sum_{i=1}^s \sum_{j=1}^r c_{ij} a_{ji}. \end{aligned}$$

An expansion of these two sums will show them to be equivalent. Thus, the rank of (5.2) is

$$\mu = \text{trace of } d' C_{22.1}^{-1} d.$$

But $d' C_{22.1}^{-1}$ is a scalar, hence

$$\mu = d' C_{22.1}^{-1} d.$$

The problem is to solve a from

$$(C_{21} C_{11}^{-1} C_{12} - \lambda C_{22}) a = (dd' - \mu C_{22.1}) a = 0.$$

Substituting $\mu = d' C_{22.1}^{-1} d$, then

$$(dd' - d' C_{22.1}^{-1} d C_{22.1}) a = 0. \quad (5.3)$$

Let $a = C_{22.1}^{-1} d$, and (5.3) becomes

$$\begin{aligned} dd' C_{22.1}^{-1} d - d' C_{22.1}^{-1} d C_{22.1} C_{22.1}^{-1} d \\ = d(d' C_{22.1}^{-1} d) - (d' C_{22.1}^{-1} d) d = 0, \end{aligned}$$

which is identically zero for $d' C_{22.1}^{-1} d$ is a scalar. The solution for a must be, then

$$a = C_{22.1}^{-1} d, \quad (5.4)$$

where $C_{22.1}$ is the within or residual matrix and a, the column of regression coefficients. As mentioned earlier, Bartlett (1938) sketched this solution. It has been expanded here to show the details. The result in (5.4) is the same result that Fisher obtained more directly.

It has been shown that the most predictable criterion and the linear discriminant analysis are special cases of canonical correlation analysis. Next, the discussion will concern the use of canonical correlation analysis to find how many factors (or how many dimensions) there

are to a given "domain". A "domain" is any area of study, whether psychological, biological, or physical; e.g., "intelligence" is a "domain" often studied by psychologists.

It is maintained by some psychiatrists and psychologists that the only difference between normal people, neurotics, and psychotics is one of degree. That is, these groups are conjectured to lie on a line or continuum. In a recent article, Rao and Slater (1949) study this problem statistically. The domain under study may be referred to as that of "neuroticism". The dimensionality of this "domain" is studied via tests. Factor analysts study the correlation matrix of tests to see how many "factors" are present or to determine the "dimensionality" of the tests. It should be emphasized here, that the number of significant factors present in a given domain is then determined through the use of the tests. For example, if a battery of tests is assembled to study intelligence, then all factors thought to bear on intelligence would be included in the battery. If numeric ability is considered a "factor" or aspect of intelligence, then surely some numeric tests would be included in the battery. Thus, dimensionality of a domain can be measured only in so far as the tests measure the domain. This places the burden of choosing a "good" battery on the experimenter. It was stated that Rao and Slater studied statistically the dimensionality of neuroticism; but this was possible only in so far as the tests they chose measured all aspects of neuroticism. For their study, they chose a group of British World War II servicemen. The "tests" used were 13 "pointers", some of which were: hereditary predisposition, physical ill health, neurotic traits in child-

hood, emotional instability, and alcoholism. The pointers were grouped into 3 "tests" A, B, C. (See the paper for the details of the grouping. For the purposes of this discussion, three "tests" were given to certain servicemen.) The test scores were listed for each individual of the following 6 groups: 1) normal, 2) personality change, 3) anxiety state, 4) hysteria, 5) psychopathy, and 6) obsession. The number of individuals in each group was different, and the number in the normal group was not proportionally represented since data were more difficult to collect for this group. (The technique employed is under study; the results obtained are of less interest here.) The total sum of squares and cross-products of the three tests A, B, C was obtained over all individuals in the 6 groups. Denote by T the 3 x 3 total sums of squares and cross-products matrix. From this, the between groups product-sum matrix, M, was broken out. Let W = the within groups product-sum matrix, so that, in this notation $T = M + W$. The total degrees of freedom was 255 and since M had 5 degrees of freedom, W had 250 degrees of freedom.

First, the generalized distance was computed between each pair of groups, as e.g., normal versus personality change. This was preliminary work to see if there were in fact differences between the groups as reflected by the tests A, B, and C. The statistic used was

$$D^2 = \sum_{i=1}^3 \sum_{j=1}^3 W^{ij} d_i d_j,$$

where $[W^{ij}]$ is the inverse of the within matrix, $\frac{1}{250} W$, and d_i is the difference in the mean values of the two groups under study as scored on

the tests A, B, C. A test of significance for D^2 may be obtained by using

$$F = \frac{N_1 N_2}{N} \cdot \frac{1}{p} \cdot \frac{p - m + 1}{m} D^2,$$

with $(m, p - m + 1)$ degrees of freedom, where N_1 is the number of individuals in the first group (normal, say), N_2 the number in the second group (personality change, say), m the number of variates or tests (3 in this example), and $p = 250$, the degrees of freedom in the within dispersion matrix W . This test is due to Hotelling.

Rao and Slater next performed a canonical correlation analysis. Corresponding to the determinantal equation

$$\left| C_{21} C_{11}^{-1} C_{12} - \lambda C_{22} \right| = 0,$$

there is the equation

$$\left| M - \lambda T \right| = \left| M - \lambda(M + W) \right| = 0.$$

(Rao and Slater actually solved $\left| M - \mu W \right| = 0$, where μ is related to the square of the canonical correlation λ by $\mu = \lambda/(1 - \lambda)$. The product-sum matrix M is the sums of squares resulting from the regression of tests onto the group dummy variates of 0 or 1. Thus it corresponds to $C_{21} \times C_{11}^{-1} C_{12}$, the sums of squares accounted for by the regression of S_2 or S_1 .

Rao and Slater found that only one canonical correlation tested statistically different from zero, indicating only one significant dimension or factor in neuroticism in so far as the tests A, B, C were able to measure neuroticism. That is not to say, of course, that there is

only one dimension to neuroticism, but merely that only one dimensionality can be recognized in such aspects of neuroticism as measured by these tests.

Rao and Slater proceed to determine statistical criteria for deciding to which group an individual α belongs (normal, etc.), given his test scores $A_\alpha, B_\alpha, C_\alpha$.

Here, then, is a use of external factor analysis. The tests A, B, C are used as devices to study the dimensionality of the domain of neuroticism. Using the significant canonical root, μ , the vector \underline{a} is found such that $(M - \mu W) \underline{a} = 0$. The significant factor is the linear compound

$$L = a_1 A_\alpha + a_2 B_\alpha + a_3 C_\alpha,$$

where the a_i are the elements of the vector \underline{a} , and $A_\alpha, B_\alpha, C_\alpha$ are the test scores made by individual α on tests A, B, and C.

This is the same problem as in factor analysis: to determine the number of significant factors in a correlation (or covariance) matrix, and once found, to concoct factors which are linear compounds of the test scores

$$X_{p\alpha} = \frac{1}{\sqrt{\lambda_p}} \sum_i l_{ip} s_{i\alpha}.$$

(See section 3, Chapter II).

The main difference in the two techniques is:

- 1) In the special type of "external factor analysis" (Bartlett,

1948, and Rao and Slater, 1949), the analysis is run on raw or unstandardized scores, while for "internal factor analysis", the scores are usually standardized. This is not a basic difference.

2) In the "external analysis" in Bartlett's discussion, the tests are replicated t times, while for the Rao-Slater example, the individuals were grouped into 6 categories so that tests given to individuals in a certain group are "replications" of a sort. Use is made of the "within" matrix, or the matrix of variances and covariances of the "errors". Again, this is no basic difference.

The internal factor analysts do not replicate their tests nor do they group their individuals into homogeneous groups. But there is no reason why the individuals could not be grouped, if the experiments were properly designed.

Factor analysts maintain that statisticians should aid them in the statistics of their problem. The main question they want answered is "when shall we stop factoring"? That is, how many significant factors are there in a correlation matrix. Bartlett has contributed considerably to this aspect of the problem, and his work along these lines will be discussed in Chapter VII. For the most part, theoretical statisticians have ignored the particular problems of internal factor analysis and have concerned themselves with other multivariate techniques. The problem is this, then, can the psychologists use other existing multivariate techniques to answer their questions? From the previous discussion, it seems that there are at least three alternative techniques available:

A) .If replication of the test is practicable (due to learning effects it is often not practicable), then the canonical analysis suggested by Bartlett (1948, p. 75) and discussed in Chapter IV could be used. That is, find the roots and the associated canonical variates (or factors) from the determinantal equation

$$|A - \lambda(A + C)| = 0, \quad (5.5)$$

$$|A - \mu C| = 0, \quad (5.6)$$

where there are n tests on N persons replicated t times, and

$A = n \times n$ matrix between persons, $N - 1$ d.f.

$B = n \times n$ matrix between times, $t - 1$ d.f.

$C = n \times n$ matrix, Persons \times times interaction, $(N-1)(t-1)$ df.

It was shown that (5.5) might be "reduced" or "standardized" to $|R_t - \lambda I| = 0$ by assuming that the errors are uncorrelated, and by assuming that the tests all have equal reliabilities.

B) If replication is not practicable and if the reliabilities ρ_{ii} , of all tests are known, then one might factor

$$|R - \lambda P| = 0 \quad (5.7)$$

where P is the $n \times n$ diagonal matrix of reliability coefficients of the n tests. This procedure would seem to lie somewhere between a factorization of

$$|A - \mu C| = 0 \quad (5.8)$$

(or $|A^* - \mu E| = 0$, where E is the true variance - covariance matrix

of "errors" and A^* is the $n \times n$ variance-covariance matrix between persons)

and

$$| R - \lambda I | = 0. \quad (5.9)$$

C) Finally, the procedure of Rao and Slater seems to offer a third alternative. If the factor analyst is studying the domain of intelligence, say, then on a priori bases, choose various groups of people from very intelligent to very dull. The $n \times n$ matrix of variation between groups, M , say and the $n \times n$ matrix of variation within groups, W , say, give rise to the determinantal equation

$$| M - \lambda (M + W) | = 0 \quad (5.10)$$

or

$$| M - \mu W | = 0. \quad (5.11)$$

(B) and (C) would in general be preferable since no replications of the tests on the same individuals is involved. Tests of significance are available for (A) and (C), using large sample approximations. Some theoretical work would need to be done for (B) to study the distribution of the roots.

There are two general classes of problems in factor analysis work:

- 1) the problem of studying only the "common" factors in a new area or domain; for these problems, the psychologists insist on substituting communalities in the diagonal of the correlation matrix;
- 2) the problem of predicting how a new individual will behave or into what group or category he will be placed; for these problems, the specific parts or

"factors" of the tests are important, and thus unities are used in the diagonal of the correlation matrix. For the second class of problems, techniques (A), (B), and (C) would seem to be appropriate. For the first class, the psychologists would not want to use these techniques until the diagonal elements of A in (5.6), R of (5.7), and M of (5.11) are "reduced" since they include the variance of the specific part of the tests. One could use an iterative technique in these problems, also: Compute the number of significant roots, r , say, and calculate the diagonal elements, assuming r common or significant factors, recompute the λ_p , etc.

It would be interesting to see the results of an internal and of an external factor analysis on the same data. The results of the two analyses should be similar: factors found from a solution of $|R - \lambda I| = 0$ and $|M - \lambda W| = 0$ should not differ greatly. Of course, one would not expect a factorization of $|R_a - \lambda I| = 0$ to give results similar to $|M - \lambda W| = 0$ since $|R_a - \lambda I|$ in general gives different results from $|R - \lambda I| = 0$. R_a , recall, has communalities in the diagonal. Technique (C) could be used for only two groups, a "successful" and the corresponding "unsuccessful" group of any particular occupation. A discriminant analysis could also be run for this special case of two groups. Batteries of tests are frequently given only to the "successful" individuals of a given domain and a factorial study run on this selected group. Perhaps sharper distinctions and a clearer understanding of the factors could be obtained by also utilizing the information on the "unsuccessful" group.

Factor analysts have seemed to place emphasis on the selection of a "good" battery of tests. Surely one wants all aspects or factors to be represented in the battery, in so far as he is able to ascertain what the

factors are. But of equal importance is the selection of the sample of individuals. In Chapter III it was emphasized that the reliability (and the correlations) of the tests depends on the people in the sample. If the people are very homogeneous, the reliability and covariances (and thence correlations) are small. It is extremely important, then, that equal care be taken by the psychologists in choosing their sample of persons. (This is important, of course, whether one solves $|R - \lambda I| = 0$ or one of the determinantal equations above).

To summarize the general idea of this chapter:

Psychologists have developed and expanded a multivariate technique called factor analysis, largely outside the general field of statistical growth and development. They have focused their attention on the correlations among tests and have directed their attention to a study of the correlation matrix. Parallel with this development have been advances in other multivariate techniques which are "similar". The purpose of the chapter has been an attempt to see what the problems are to be solved and what techniques are now available to help solve them. The problem of factor analysis is one of dimensionality of a given domain. This "apparent" dimensionality is studied via instruments called "tests". This is also the problem of canonical correlation analysis. In general, canonical correlation analysis studies involve the kind and number of factors between two sets of variates, and result in a solution of $|C_{21}C_{11}^{-1}C_{12} - \lambda C_{22}| = 0$. Special cases of this general equation arise when: (1) tests are replicated on the same individuals $t > 1$ times, and (2) by a grouping of the individuals into homogeneous groups on some a priori basis. These are called "external" factor analyses to

distinguish them from "internal" factor analysis. The latter involves one set of variates. The "two" sets of variates of (1) and (2) above are the set of n test scores and the "pseudo" set of persons. It is conjectured that the results of external and internal factor analysis will be similar, and thus canonical correlation analysis theory and techniques are available to the psychologists for a study of their dimensionality problems.

Chapter VI

PREDICTION OF FACTOR LOADINGS

One of the questions frequently asked in factor analysis is how will a test behave when placed in a battery of tests that has already been factored, without re-factoring the battery with the new test added.

Let the first r factors of the principal axis solution of $|R_a - \lambda I| = 0$ for $(n - 1)$ tests be given by the factor matrix

$$F_a = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1r} \\ a_{21} & a_{22} & \dots & a_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n-1,1} & a_{n-1,2} & \dots & a_{n-1,r} \end{bmatrix}$$

Then

$$F_a F_a' \simeq \begin{bmatrix} h_1^2 & r_{12} & \dots & r_{1,n-1} \\ r_{21} & h_2^2 & \dots & r_{2,n-1} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n-1,1} & r_{n-1,2} & \dots & h_{n-1}^2 \end{bmatrix}$$

$F_a F_a' = R$ if the number of common factors $r = n$. Recall also that

$F_a F_a' = \Lambda_a^2$ and $\frac{F_a' S S' F_a}{N - 1} = F_a' R F_a = \Lambda_a^4$. Hence the estimate of factor p for

the i th individual is

$$x_{pi} = \frac{1}{\sqrt{\lambda_p}} \sum_j l_{ip} s_{ij}$$

When all the n factors are extracted, $F = [b_{ip}]$ is the factor matrix of order $n \times n$. It can be written as

$$F = L\Lambda,$$

where L is the matrix of direction cosines of test axes relative to principal component axes obtained by the principal axis solution and

$$\Lambda = \begin{bmatrix} \lambda_1 & & & 0 \\ & \cdot & & \\ & & \cdot & \\ & & & \cdot \\ 0 & & & \lambda_n \end{bmatrix}.$$

Thus, X can be found from

$$S = FX$$

by pre-multiplying by F' ,

$$\begin{aligned} F'S &= F'FX \\ &= (L\Lambda)'(L\Lambda)X \\ &= \Lambda'\Lambda X \\ &= \Lambda^2 X. \end{aligned}$$

Thus

$$X = (\Lambda^2)^{-1} F'S \quad (6.1)$$

where

$$(\mathcal{L}^2)^{-1} = \begin{bmatrix} \frac{1}{\lambda_1} & & & 0 \\ & \cdot & & \\ & & \cdot & \\ & & & \cdot \\ 0 & & & & \frac{1}{\lambda_n} \end{bmatrix}.$$

A typical element of the matrix equation (6.1) is

$$\begin{aligned} X_{p\alpha} &= \frac{1}{\lambda_p} \sum_i b_{ip} s_{i\alpha} \\ &= \frac{1}{\lambda_p} \sum_i t_{ip} \sqrt{\lambda_p} s_{i\alpha} \\ &= \frac{1}{\sqrt{\lambda_p}} \sum_i t_{ip} s_{i\alpha}. \end{aligned} \quad (6.2)$$

When only a part of R is factored so that $F_a = [a_{ip}]$ is of order $n \times r$, $r < n$, then X is found from

$$S = F_a X + U$$

in a similar fashion. Pre-multiply by F_a' to get

$$F_a' S = F_a' F_a X + F_a' U. \quad (6.3)$$

However $F_a' U = 0$, since a typical element of $F_a' U$ is

$$\begin{aligned} \sum_i a_{ip} u_{i\alpha} &= \sum_i a_{ip} (s_{i\alpha} - \sum_p a_{ip} X_{p\alpha}) \\ &= \sum_i a_{ip} s_{i\alpha} - \sum_i a_{ip} \sum_p a_{ip} X_{p\alpha} \end{aligned}$$

$$= \sum_i a_{ip} s_{i\alpha} = \sum_i a_{ip}^2 X_{p\alpha}.$$

But $\sum_{i=1}^n a_{ip} s_{i\alpha} = \sqrt{\lambda_p} \sum_i b_{ip} s_{i\alpha}$

$$= \lambda_p X_{p\alpha}$$

and $\sum_{i=1}^n a_{ip}^2 X_{p\alpha} = \lambda_p X_{p\alpha}$, since $\sum_{i=1}^n a_{ip}^2 = \lambda_p$.

Thus

$$\sum_i a_{ip} u_{i\alpha} = 0 \text{ for all } p \text{ and } \alpha.$$

Therefore (6.3) can be written as

$$\begin{aligned} F'_a S &= F'_a F_a X \\ &= \Lambda_a^2 X \end{aligned}$$

or

$$X = (\Lambda_a^2)^{-1} F'_a S, \quad (6.4)$$

where

$$(\Lambda_a^2)^{-1} = \begin{bmatrix} \frac{1}{\lambda_{a1}} & & & 0 \\ & \cdot & & \\ & & \cdot & \\ & & & \cdot \\ 0 & & & \frac{1}{\lambda_{ar}} \end{bmatrix}$$

Recall that the λ_p obtained from a factorization of \dot{R} are not the same as the λ_{ap} obtained from a factorization of R_a . A typical element of (6.4) is, then

$$\begin{aligned} X_{p\alpha} &= \frac{1}{\lambda_{ap}} \sum_{i=1}^n a_{ip} s_{i\alpha} \\ &= \frac{1}{\sqrt{\lambda_{ap}}} \sum_i l_{ip} s_{i\alpha} \end{aligned}$$

For this discussion, it is more convenient to use

$$Z_{p\alpha} = \sum_{i=1}^n a_{ip} s_{i\alpha}$$

since the a_{ip} are the quantities obtained in a factor analysis. Thus

$$Z_{p\alpha} = \sum_i a_{ip} s_{i\alpha} = \sqrt{\lambda_{ap}} \sum_i l_{ip} s_{i\alpha}$$

Recall that $Y_{p\alpha}$ was defined to be (section 2, Chapter II)

$$Y_{p\alpha} = \sum_i l_{ip} s_{i\alpha}$$

(Note: this $Y_{p\alpha}$ is not the raw test score $Y_{i\alpha}$.)

The amount of the pth factor possessed by the α th individual was defined to be

$$X_{p\alpha} = Y_{p\alpha} / \sqrt{\lambda_p} = Z_{p\alpha} / \lambda_p$$

Thus $Z_{p\alpha}$ has zero mean and variance λ_p^2 .

Let s_n be the nth test that is to be added to the battery. It is desired to know its factorial content. Only the factors of the nth test that are common to the factors of the first $n - 1$ test will appear. The mean of s_n is zero and its variance is unity: $\sum_{\alpha} s_{n\alpha} = 0$, $\frac{1}{N-1} \sum_{\alpha} s_{n\alpha}^2 = 1$.

THEOREM: If test s_n has factors that are common to some or all of the common factors of the first $n - 1$ tests, and if its specific part does not form common factors with the specific parts of the first $n - 1$ tests, and further if $r_{ij} = \sum_{p=1}^r a_{ip} a_{jp}$, then

$$\begin{aligned} \hat{a}_{np} &= \frac{\text{Cov}(s_n, Z_p) / (N - 1)}{\sqrt{\text{Var}(s_n) \cdot \text{Var}(Z_p)}} \\ &= \frac{1}{N - 1} \sum_{\alpha} s_{n\alpha} \left[\sum_{i=1}^{n-1} a_{ip} s_{i\alpha} \right] / \lambda_{ap} \\ &= \frac{1}{\lambda_{ap}} \sum_{i=1}^{n-1} a_{ip} r_{in}, \end{aligned} \quad (6.5)$$

where \hat{a}_{np} is the estimate of the factor loading of test s_n on factor p .

Proof:

$$\text{Since } r_{in} = \sum_{p=1}^r a_{ip} a_{np},$$

$$\frac{\text{Cov}(s_n, Z_p)}{\sqrt{\text{Var}(s_n) \cdot \text{Var}(Z_p)}} = \frac{\text{Cov}(s_n, Z_p)}{\lambda_{ap}} = \frac{1}{\lambda_{ap}} \sum_{i=1}^{n-1} a_{ip} r_{in}$$

$$\begin{aligned}
&= \frac{1}{\lambda_{ap}} \left[a_{1p}(a_{11}a_{n1} + a_{12}a_{n2} + \dots + a_{1r}a_{nr}) \right. \\
&\quad + a_{2p}(a_{21}a_{n1} + a_{22}a_{n2} + \dots + a_{2r}a_{nr}) \\
&\quad + \dots \\
&\quad \left. + a_{n-1,p}(a_{n-1,1}a_{n1} + a_{n-1,2}a_{n2} + \dots + a_{n-1,r}a_{nr}) \right]
\end{aligned}$$

Using the fact that

$$\sum_{i=1}^{n-1} a_{ip}a_{im} = \begin{cases} \lambda_{ap} & \text{if } m = p \\ 0 & \text{if } m \neq p \end{cases},$$

$$\begin{aligned}
\frac{\text{Cov}(s_n, Z_p)}{\lambda_{ap}} &= \frac{a_{np}}{\lambda_{ap}} \sum_{i=1}^{n-1} a_{ip}^2 \\
&= a_{np}.
\end{aligned}$$

Obviously, then,

$$\begin{aligned}
h_n^2 &= \sum_{p=1}^r \left\{ \frac{\text{Cov}(s_n, Z_p)}{\lambda_{ap}} \right\}^2 \\
&= \sum_{p=1}^r a_{np}^2,
\end{aligned}$$

the usual definition for the communality.

Denote by $F_a(n, r)$ the factor matrix obtained by the principal axis solution for the first r common factors on n tests. Let F_m be the matrix

$F_a(n-1, r)$ augmented by the estimates of a_{np} . Then F_m and $F_a(n, r)$ represent the same test configuration in the r -common factor space, for

$$F_m F_m' = R_a^{-1} F_a(n, r) F_a'(n, r)$$

and

$$F_m' F_m = \bigwedge_a^2 F_a'(n, r) F_a(n, r)$$

That is, the correlations between the test vectors in the common-factor space are the same, and since the length of the test vector \underline{i} is $\sqrt{h_i^2}$ in both test-vector configurations, the two test-vector configurations are the same. A "rotation" of either F_m or $F_a(n, r)$ could be made to effect equivalence.

Since the common factor structure (or test configuration in the common-factor space) is invariant under rotation, the estimation of the a_{np} could be done either for the factor matrix as first factored or for the rotated factor matrix.

The problem can be set up and solved with identical results by multivariate regression techniques. Let F_1 be the matrix $F_a(n-1, r)$ with each element of columns 1, 2, ..., r divided respectively by λ_{ap} , i.e.,

$$F_1 = \begin{bmatrix} \frac{a_{11}}{\lambda_{a1}} & \frac{a_{12}}{\lambda_{a2}} & \dots & \frac{a_{1r}}{\lambda_{ar}} \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ \frac{a_{n-1,1}}{\lambda_{a1}} & \frac{a_{n-1,2}}{\lambda_{a2}} & \dots & \frac{a_{n-1,r}}{\lambda_{ar}} \end{bmatrix}$$

The problem is to estimate the regression coefficients of the transformed variates $F_1' S_2$ on S_1 , where S_2 is the $(n - 1) \times N$ observation matrix for the first $(n - 1)$ tests and S_1 is the raw matrix, $1 \times N$, of observations for the nth test. (The test scores are considered standardized throughout this chapter). Thus B_{21} is to be estimated from the regression equation

$$F_1' S_2 = B_{21} S_1 + F_1' S_{2.1}. \quad (6.6)$$

Post-multiplying (6.6) by S_1' ,

$$F_1' S_2 S_1' = B_{21} S_1 S_1' + F_1' S_{2.1} S_1'. \quad (6.7)$$

If $F_1' S_{2.1} S_1' = 0$ (i.e., if the specific factor of the nth test does not form common factors with any of the specific factors of the first $(n - 1)$ tests), then

$$B_{21} = F_1' C_{21}, \quad (6.8)$$

where

$$C_{21} = S_2 S_1' = \begin{bmatrix} r_{1n} \\ r_{2n} \\ \vdots \\ r_{n-1,n} \end{bmatrix}, \quad S_1 S_1' = \text{Var}(s_n) = 1.$$

It will be observed that the result of (6.8) is the same as that of (6.5).

As an example, consider the artificial correlation matrix of 5 tests resulting from the operation of two given factors. Denote the

factor matrix of the given factors by F_g :

$$F_g = \begin{array}{c} \begin{array}{cc} \text{I} & \text{II} \\ \begin{bmatrix} .8 & .3 \\ .7 & -.2 \\ .9 & -.25 \\ .75 & .41 \\ .6 & .3041 \end{bmatrix} \end{array} \end{array}$$

$$\lambda'_{a1} = (.8)^2 + (.7)^2 + \dots + (.6)^2 = 2.8625$$

$$\lambda'_{a2} = (.3)^2 + (-.2)^2 + \dots + (.3041)^2 = .4531$$

$$\lambda'_{a1} + \lambda'_{a2} = 3.3156$$

$$F_g F'_g = R_a(5) = \begin{bmatrix} .73 & .50 & .645 & .723 & .5712 \\ .50 & .53 & .68 & .443 & .3592 \\ .645 & .68 & .8725 & .5725 & .4640 \\ .723 & .443 & .5725 & .7306 & .5740 \\ .5712 & .3592 & .4640 & .5747 & .4525 \end{bmatrix}$$

The solution of

$$|R_a(5) - \lambda I| = 0$$

gives

$$F_a(5) = \begin{array}{c} \begin{array}{cc} \text{I} & \text{II} \\ \begin{bmatrix} .8352 & .1800 \\ .6632 & -.3002 \\ .8537 & -.3791 \\ .8019 & .2960 \\ .6380 & .2133 \end{bmatrix} \end{array} \end{array}$$

$$\lambda_{a1}(5) = (.8352)^2 + \dots + (.6380)^2 = 2.9163$$

$$\lambda_{a2}(5) = (.1800)^2 + \dots + (.2133)^2 = .3993$$

$$\lambda_{a1}(5) + \lambda_{a2}(5) = 3.3156 .$$

It is easily verified that

$$h_1^2 = (.8)^2 + (.3)^2 = .73 = (.8352)^2 + (.1800)^2$$

$$r_{12} = (.8)(.7) + (.3)(-.2)$$

$$= .50$$

$$= (.8352)(.6632) + (.1800)(-.3002)$$

etc.

so that,

$$F_a(5) F_a'(5) = R_a(5).$$

The relationship is exact since there are exactly two common factors in this hypothetical example.

Next consider the correlation matrix $R_a(4)$, obtained by omitting test 5 from the battery. The solution of

$$|R_a(4) - \lambda I| = 0$$

gives the factor loadings on the first 4 tests as:

$$F_a(4) = \begin{array}{cc} & \begin{array}{c} \text{I} \\ \text{II} \end{array} \\ \begin{array}{c} .8225 \\ .6806 \\ .8757 \\ .7819 \end{array} & \begin{array}{c} .2313 \\ -.2585 \\ -.3250 \\ .3453 \end{array} \end{array}$$

$$\lambda_{a1}(4) = (.8225)^2 + \dots + (.7819)^2 = 2.5179$$

$$\lambda_{a2}(4) = (.2313)^2 + \dots + (.3453)^2 = .3452.$$

Using (6.5) or (6.8):

$$\begin{aligned} a_{51} &= \frac{(.8225)r_{15} + (.6806)r_{25} + (.8757)r_{35} + (.7819)r_{45}}{\lambda_{a1}(4)} \\ &= \frac{1.5700}{2.5179} \\ &= .6235. \end{aligned}$$

Similarly,

$$\begin{aligned} a_{52} &= \frac{.2313r_{15} - .2585r_{25} - .3250r_{35} + .3453r_{45}}{\lambda_{a2}(4)} \\ &= \frac{.08690}{.3452} \\ &= .2517 \end{aligned}$$

Thus,

$$h_5^2 = a_{51}^2 + a_{52}^2 = .4522,$$

which differs from .4525 due to rounding errors. Also note that

$$\left\{ \lambda_{a1}(4) + \hat{a}_{51}^2 \right\} + \left\{ \lambda_{a2}(4) + \hat{a}_{52}^2 \right\} = \lambda_{a1}(5) + \lambda_{a2}(5) = 3.3156 = \lambda'_{a1} + \lambda'_{a2}.$$

It can also be verified that F_m , the $F_a(4)$ matrix augmented by the estimates of a_{5p} , when multiplied by its transpose will give $R_a(5)$, thus:

$$\begin{bmatrix} .8225 & .2313 \\ .6806 & -.2585 \\ .8757 & -.3250 \\ .7819 & .3453 \\ \dots\dots\dots \\ .6235 & .2517 \end{bmatrix} \begin{bmatrix} .8225 & .6806 & .8757 & .7819 & | & .6235 \\ .2313 & -.2585 & -.3250 & .3453 & | & .2517 \end{bmatrix} \approx R_a(5).$$

The relationship is only approximate due to rounding errors.

Thus the structure represented by F_g , $F_a(5)$ and F_m , $[F_a(4)$ augmented $]_{}$ is the same. $F_a(5)$ may be obtained from F_m by an appropriate orthogonal transformation.

For this artificial example of two factors, there is no estimation in the true sense for the a_{5p} can be calculated exactly. However, in practice, there are likely to be new common factors due to overlapping of specific factors. Too, for a real problem, $r_{ij} \approx \sum_{p=1}^F a_{ip} a_{jp}$. But to a first approximation, this procedure should work quite well.

Chapter VII
ESTIMATION OF FACTOR LOADINGS
AND TESTS OF SIGNIFICANCE

Besides Hotelling's principal axis solution, two major contributions of mathematical statisticians to the theory of factor analysis have been Lawley's (1940, 1942) attempt to estimate factor loadings by maximum likelihood, and Bartlett's (1950) large sample tests of significance for the roots of the determinantal equation $|R - \lambda I| = 0$. These will be reviewed in this Chapter.

Lawley's "Method I" is based on the variance-covariance sampling distribution of test scores. He starts with Wishart's joint distribution of variances and covariances. Preliminary work will presumably indicate the approximate size of r , the number of common factors. The need for prior information about r is one of the main difficulties with the technique. Using the estimated value of r , the likelihood function is maximized with respect to the factor loadings. The solution is laborious, several iterations being necessary to extract each factor. In his "Method II", Lawley considers the test scores directly,

$$s_{i\alpha} = \sum_{p=1}^r \lambda_{ip} x_{p\alpha} + \psi_{i\alpha}$$

where $\psi_{i\alpha} = \pi_{i\alpha} + \epsilon_{i\alpha}$, the sum of the specific part and the error term. Recall that the population variance of $\psi_{i\alpha}$ was denoted by $\sigma^2(u_i)$. Lawley, then, maximized

$$L \propto [\sigma(u_1) \sigma(u_2) \dots \sigma(u_n)]^{-(N-1)} \exp \left[-\frac{1}{2} \sum_i \frac{1}{\sigma^2(u_i)} \sum_p (s_{ip} - \sum_p \lambda_{ip} x_{p\alpha})^2 \right]$$

with respect to the $\sigma^2(u_i)$, λ_{ip} , and $x_{p\alpha}$. Bartlett (1950) remarks that "At first sight the second method would appear to be more fundamental, but it is known to lead to difficulties the factor equation [including uniques, $s_{i\alpha} = \sum_p \lambda_{ip} x_{p\alpha} + \psi_{i\alpha}$ or $S = F_a X + U$] is equivalent to another set of structural equations (familiar in econometrics) for which it is definitely known that the standard maximum likelihood method breaks down". Lawley (1942) remarked that in his Method II, convergence to unacceptable solutions sometimes resulted, where the unique or "error" variances, $\sigma^2(u_i)$ vanished. Kendall (1950) also had doubts about the method and wondered "whether the maximum likelihood procedure is valid where the parameters are under-determined".

As for Lawley's "Method I" Bartlett (1950) said that he "was at first skeptical of the value of this method because of its obvious break-down in elementary cases". He cited the case of 2 tests and one common factor. There is only one correlation coefficient and 2 unknown factor loadings. In general, for r common factors, there are

$$n + (n-1) + \dots + (n-r+1) = r(2n+1-r) / 2$$

unknown independent factor loadings and $\frac{1}{2} n(n-1)$ correlation coefficients. If

$$r(2n+1-r)/2 < n(n-1)/2$$

the problem is said to be "over-determinate". Thus for large batteries, difficulties of the sort of finding factor loadings for one common factor

among 2 tests would probably not arise since \underline{r} is usually considerably less than \underline{n} . Bartlett (1950) states that the "Method I" solution is distinct from the principal axis solution, "but it is an attractive link between the two analyses that the total χ^2 corresponding to the significance of the unreduced correlation matrix is necessarily the same, and only because of the difference between the factors extracted in the two analyses does the analysis of the total χ^2 into its respective components differ".

Lawley gives an expression to test the residual variation in the variance-covariance matrix of the raw scores after r common factors have been extracted. Under the null hypothesis of r common factors,

$$N \sum \frac{(C_{ij} - \hat{\sigma}_{ij})^2}{s_i^2 \cdot s_j^2} \quad (7.1)$$

is approximately distributed as χ^2 with $\frac{1}{2} [(n-r)^2 - (n+r)]$ degrees of freedom. N is the number of individuals, n the number of tests, $\hat{\sigma}_{ij}$ is the sample covariance between tests i and j , C_{ij} the covariance calculated from the r extracted factors, and s_i^2 is the unique variance for test i . If A_i^2 is the fraction of the sample variance $\hat{\sigma}_{ii}$ of test i that is common to the other tests, then $s_i^2 = \hat{\sigma}_{ii} - A_i^2$. A_i^2 may be obtained by computing the sum of squares of the r factor loadings for test i , and thence one may calculate s_i^2 . A_i^2 and s_i^2 are, respectively, to the unstandardized test scores as h_i^2 and $\sigma^2(u_i)$ are to the standardized test scores.

If the correlation matrix is factored, the expression analogous to (7.1) is

$$\begin{aligned}
& N \sum_{i,j} \frac{(r_{ij} - \sum_{p=1}^r a_{ip} a_{jp})^2}{\sigma^2(u_i) \cdot \sigma^2(u_j)} \\
& = N \sum_{i,j} \frac{\bar{r}_{ij}^2}{(1-h_i^2)(1-h_j^2)} \tag{7.2}
\end{aligned}$$

h_i^2 is the communality for test i as estimated from the r common factors, and \bar{r}_{ij} is the residual correlation of test i with j after the removal of the effect of the r common factors.

Thompson (1948) gives a detailed example of the use of "Method I" in Chapter IX of his text-book, and Emmett (1949) demonstrates the procedure for 3 common factors in a battery of 9 tests.

The validity of the test was demonstrated by Henrysson (1950), who concocted an artificial example with 1 common factor among 9 variables. He considered 12 samples with 200 independent sets of 9 observations in each sample. For each of the 12 samples, the variance-covariance matrix among the 9 variables was calculated and then factored by Lawley's "Method I" technique. After the extraction of one factor, the expression in (7.2) was calculated. The 12 computed χ^2 's that he obtained ranged from 15.5 to 38.6, each with 27 d.f., giving probabilities ranging from .96 to .07. This range of .89 agrees well with the expected range of .85, assuming the values of P follow a rectangular distribution. Thus, Lawley's residual test is empirically verified, at least for a single factor in an example known to satisfy the basic postulates.

Bartlett (1950, 1951a, 1951b) derived tests for the λ_p resulting from the solution of (7.1). He approached the problem from two directions. First, he considered an internal factor analysis as a special case of external analysis: test 1 against test 2, test 3 against the two tests, 1 and 2; etc. The criterion to test the roots of a canonical analysis at each stage is

$$\chi^2 = - \left[(N-1) - \frac{1}{2} (p+q+1) \right] \sum_{p=1}^r \log_e (1-\lambda_p), \quad (7.3)$$

with r ($q-p+r$) degrees of freedom. N is the number of observations (individuals), p the number of independent variables (in the matrix S_1), q is the number of dependent variates (in the matrix S_2), and r is the number of extracted factors or roots. The canonical root λ_p is the square of the p th canonical correlation. Thus for test 1 against test 2;

$$\chi^2 = - \left[(N-1) - \frac{1}{2} (1+1+1) \right] \log_e (1-r_{12}^2), \quad 1 \text{ d.f.}$$

and for test 3 against tests 1 and 2:

$$\chi^2 = - \left[(N-1) - \frac{1}{2} (2+1+1) \right] \log_e (1-R_{3,12}^2), \quad 2 \text{ d.f.},$$

where r_{12} is the correlation between tests 1 and 2 and $R_{3,12}$ is the multiple correlation between test 3 and tests 1 and 2. Continuing this process, and choosing a mean multiplying factor so as to make the resulting test invariant of the path, he suggests the test

$$\chi^2 = - \left[(N-1) - \frac{1}{6} (2n+5) \right] \log_e |R| \quad (7.4)$$

with $\frac{1}{2}n(n-1)$ d.f., where

$$|R| = (1-r_{12}^2) (1-R_{3,12}^2) \dots (1-R_{n,12}^2 \dots n-1),$$

the determinant of the correlation matrix. The same approximate test is found by considering the moments of $-N \log |R|$, using indirectly Wilk's formula for the moments of $|R|$.

After r common factors have been extracted, the residual matrix may be tested by calculating

$$\chi^2 = - \left[(N-1) - \frac{1}{6} (2n+5) - \frac{2}{3} r \right] \log_e R_{n-r} \quad (7.5)$$

and using d.f. = $\frac{1}{2} (n-r)(n-r-1)$,

where

$$R_{n-r} = \lambda_1 \lambda_2 \dots \lambda_r \left\{ \frac{|R|}{n-r} \right\}^{n-r} \quad (7.6)$$

If N is larger relative to n , Bartlett would use the mean multiplier

$$(N-1) - \frac{1}{6} (2n+5) - \frac{2}{3} r.$$

After considering the effect of the standardization of the test scores on the "degrees of freedom", Bartlett (1951b) suggests that it is "safer" to use

$$\text{d.f.} = \frac{1}{2} (n-r-1) (n-r+2) \quad (7.7)$$

rather than

$$\text{d.f.} = \frac{1}{2} (n-r) (n-r-1). \quad (7.8)$$

That is, the limiting χ^2 approximating distribution for standardized scores has as its parameter the degrees of freedom in (7.7).

The test of significance given by χ^2 in (7.5) tests the residual variation in the correlation matrix after the extractions of r common factors. Bartlett (1950) states

"significance in an interval analysis can only be on the basis of the relative values of the roots, which must after each elimination be in effect re-scaled to unit mean variance. This explains the factor

$$(n-r) / \left(\sum_{p=r+1}^n \lambda_p \right)^{n-r} \text{ in } R_{n-r}''.$$

And since the remaining variation in the residual correlation matrix after the extraction of r roots involves $\lambda_{r+1}, \lambda_{r+2}, \dots, \lambda_n$, the test given in (7.5) is a test of λ_{r+1} , the next largest root. For if λ_{r+1} is non-significant, then $\lambda_{r+2}, \dots, \lambda_n$ are also non-significant since $\lambda_{r+1} > \lambda_{r+2} > \dots > \lambda_n$.

As an example of the use of this test, consider the problem of
Chapter IV:

$$\lambda_1 = 3.1021$$

$$\lambda_2 = .7328$$

$$\lambda_3 = .1593$$

$$\lambda_4 = .0058$$

$$N = 6, \quad n = 4$$

$$R_4 = |R| = \lambda_1 \lambda_2 \lambda_3 \lambda_4 = .002101$$

$$R_3 = (.002101)/(3.1021 \left[\frac{4-3.1021}{4-1} \right]^3)$$

$$= \frac{.002101}{3.1021} \left[\frac{3}{.8979} \right]^3$$

$$= .02526$$

$$R_2 = .1356$$

$$R_1 = 1.0000$$

The χ^2 test for residual variation is:

	d.f.	χ^2	$\chi^2_{.05}$
$\lambda_1 \lambda_2 \lambda_3 \lambda_4$	$\frac{1}{2}(n)(n-1) = 6$	$-2.883 \log_e (.002102) = 17.47$	12.59
$\lambda_2 \lambda_3 \lambda_4$	$\frac{1}{2}(n-1-1)(n-1+2) = 5$	$-2.167 \log_e (.02526) = 7.97$	11.07
$\lambda_3 \lambda_4$	$\frac{1}{2}(n-2-1)(n-2+2) = 2$	$-1.5 \log_e (.1356) = 3.00$	5.99
λ_4	$\frac{1}{2}(n-3-1)(n-3+2) = 0$		

Only the first root is adjudged significant. The test is not considered safe for $N < 10$, say. Here there were only 6 persons, an unrealistic example from this viewpoint. Observe that the degrees of freedom for λ_4 are zero, because significance of the roots is based on relative size, and the last root can never be tested, since there is nothing to compare it with.

If the estimates of the parameters in a distribution are "efficient", then χ^2 may be used as a test of goodness of fit by calculating

$$\chi^2 = \sum (\text{observed} - \text{expected})^2 / \text{expected}$$

with degrees of freedom equal to the total number of observations less the number of fitted parameters. Thus in fitting by maximum likelihood (Lawley's Method I) one gets deviations distributed approximately as χ^2 with $\frac{1}{2} [(n-r)^2 - (n+r)]$ degrees of freedom, but when "fitting" by the principal axis procedure, the approximating χ^2 distribution has its "degrees of freedom parameter" $\frac{1}{2} (n-r-1)(n-r+2)$, after the extraction of r common factors from a battery of n tests.

How do factor analysts view these developments? The solution of $|R - \lambda I| = 0$ gives results that have optimum properties. Bartlett furnished a test of significance for the λ_p and Lawley has worked out methods for obtaining estimates of the factor loadings by maximum likelihood. The trouble is that these methods are all too tedious and time consuming in practice, using existing computational facilities. Bartlett's test requires the value of all the λ_p or at least the value of the determinant of R . In most problems, there are 30-60 tests and the principal component solution is not at all practicable. There have been recent attempts to solve the problem using electronic computers (Wrigley and Nethaus, 1952), but few psychological laboratories have such facilities.

But, at least, there are rigorous techniques available. The psychologists are aware of these. But for expediency and to a first approximation, the correlation matrix is factored by some simple summation technique as

the centroid method. Empirical rules or tests have been worked out to tell when to stop factoring. It is felt that the results obtained by these approximate methods do not differ much from the more exact methods. An approximate test that seems reasonable is to compute the percent of the total variance n accounted for by each factor, λ_p/n . It might be agreed in advance that 80%, say, of the variation would be explained. When this percentage is reached, the factoring stops. Perhaps the statistician's complaint about arbitrariness has become a "smoke-screen", but practicing statisticians will admit that tests of significance are in many circumstances superfluous to an experienced worker. Probably the most important problem in statistics is estimation, and such is the case in factor analysis. The psychologists are trying to map a domain, to estimate planes. Too, as Bartlett and others emphasize, the tests of significance may or may not indicate meaningful factors. So to that extent, even the "exact" tests are arbitrary.

The problem of what to do about the communalities still remains. In the factorization of R , the psychologists still want to insert h_1^2 in the diagonals.

It has been suggested that

$$|R_a - \lambda I| = 0 \quad (7.9)$$

be solved by the principal component solution, where R_a is the "reduced" correlation matrix, the usual correlation matrix with communalities in the diagonal. This, presumably, would give r roots of appreciable size so that

$$F_a' F_a = \begin{bmatrix} \lambda_1 & & & & \\ & \lambda_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ 0 & & & & \lambda_r \\ & & & & & 0 \end{bmatrix}$$

This would be all right if some satisfactory estimates of the h_i^2 could be obtained. The best method seems to be an iterative approach as Burt suggests. But this leaves something to be desired, for significance of factors and the estimation of the h_i^2 become confounded. Care would be needed so that the estimates would not produce a negative definite form. Bartlett's test for the λ_p would not then be appropriate, but a similar approximation modifying the "degrees of freedom" and the multiplier could probably be worked out.

It should be stressed, however, that the factors isolated by (7.9) will be different from those of $|R - \lambda I| = 0$. Bartlett (1950) mentions the case of one common factor between two tests with a third test completely independent. Then, using $(R - \lambda I)$, the equation to be solved for λ if the correlation coefficients are known is:

$$\begin{vmatrix} 1 - \lambda & \rho & 0 \\ \rho & 1 - \lambda & 0 \\ 0 & 0 & 1 - \lambda \end{vmatrix} = 0.$$

The three roots are

$$\lambda_1 = 1 + \rho$$

$$\lambda_2 = 1$$

$$\lambda_3 = 1 - \rho.$$

Tests 1 and 2 are assumed to have a common part; and each test has its own specific and its own error part. λ_2 accounts for the variation of test 3, which has no part common to tests 1 and 2. If ρ is large, both λ_1 and λ_2 will probably test significant, indicating 2 significant factors. (Recall that there is no test for λ_3 since it is the residual variation not explained by the first 2 factors).

Next consider the one-common factor problem for the "reduced" matrix. Assume the communalities H_1^2 are known. Then the problem is to solve

$$\begin{vmatrix} H_1^2 - \lambda & \rho & 0 \\ \rho & H_2^2 - \lambda & 0 \\ 0 & 0 & H_3^2 \end{vmatrix} = 0$$

for the λ_p . Obviously $H_1^2 H_2^2 = \rho^2$, and thus the population correlation matrix is exactly of rank 1. The roots for this equation are

$$\lambda_1 = H_1^2 + H_2^2$$

$$\lambda_2 = H_3^2$$

$$\lambda_3 = 0.$$

But H_3^2 has a population value of zero, so that only one factor would be declared significant, rather than two. This demonstrates rather forcefully why the factor analysts prefer to factor R_a rather than R . The presence of the specific and error parts of the variance of the test in the diagonal of the correlation matrix will increase its "effective" rank. Thus when common factors are of interest, it would seem that the psychologists are justified in substituting communalities in the diagonal of the correlation matrix before factoring.

Chapter VIII

SUMMARY AND CONCLUSION

Following a statement of the problem in the introduction, the motivation for factor analysis was discussed and the history of its development was sketched from the beginning works of Galton and Pearson to the present. It was mentioned how the problem of classification and study of factors of types was modified to a study of the factors of the mind. First came an attempt to measure a hypothetical factor called g . This was altered, not to a measure of several factors, but to a definition or identification of multiple mental factors. Underlying causes or factors presumably gave rise to the hierarchy observed in correlation matrices. The important question of reducing the diagonal elements so that the proportionality will be preserved was discussed along with the difficulties of arriving at good estimates of these.

A detailed examination of two geometrical representations of the nN test scores was given in section 1 of Chapter II along with a definition of the terms and expressions used in factor analysis. The general factor analysis equation was given with explanations of the meaning of the various parts. The centroid method of factoring and the principal axis solution were discussed in section 2 of this Chapter. In section 3, a critique of Spearman's one-factor theory and Thurstone's multiple factor theory was given.

A need was felt to clear up the relationships between intra- and inter-correlations, reliability coefficients, and communalities. These statistics were discussed in Chapter III. A second purpose of the Chapter was

to study what the psychologists want to "factor", assuming their general factor analysis equation. It was concluded that a factor analysis of V_a , a variance-covariance matrix of variance components with the elements in the main diagonal somehow adjusted downwards, might be preferable to a factor analysis of the usual correlation matrix R . The rationale for factoring R_a , the correlation matrix with reduced diagonal elements, was given. That is, the psychologists want to factor only common factors. Correlations corrected for attenuation were defined, and it was observed that a factor analysis of these attenuated correlations might have some appeal.

The question is often asked as to whether or not the same answers might be obtained from other statistical techniques as from factor analysis. With this in mind, factor analysis was compared with the analysis of variance in Chapter IV. It had been observed that the first 2 factors obtained by the centroid method of factoring corresponded, in some fashion, to two of the sources in the analysis of variance performed on unitary standardized test scores. This problem was investigated algebraically. It was found that the square of the sum of the first factor loadings is equal to $n \times$ ("sum of squares" for Persons). But "sum of squares" for unitary standardized scores, $s_{i\alpha}^*$, is in reality the mean squares, since

$$s_{i\alpha}^* = \frac{Y_{i\alpha} - \bar{Y}_i}{\sqrt{(N-1) \hat{\sigma}_{ii}}},$$

which already involves the degrees of freedom in the denominator. It was shown that equivalence was not possible between other factor loadings and

other sources in the analysis of variance table. The relationships are only approximate. In both types of analysis, the observation matrices are transformed into orthogonal "factors". In the analysis of variance, the factors correspond to well-defined sources of variation which are definable before the analysis is performed, while in factor analysis, the orthogonal factors are taken successively to explain the maximum variation possible. The factors isolated cannot be defined or observed before the analysis is performed. And after extraction, there is the often difficult task of naming and identifying the factors so isolated. Thus, for the analysis of variance, the experimenter stipulates before the experiment what linear combinations of the test scores are to be taken, while for factor analysis, the linear combinations are taken after viewing the data so as to account for maximum variation with each successive factor.

In section 2 of Chapter IV, it was shown how a canonical analysis determinantal equation might be altered so as to resemble the factor analysis determinantal equation. And it was demonstrated how the factor equation $S = F_a X + U$ might be considered a special case of the general multivariate regression equation $S_2 = B_{21} S_1 + S_{2,1}$, if one considers the hypothetical factors X as the independent set of variates. Then by a formal manipulation of symbols, it was shown that the matrix of factor loadings, F_a , is the "best" set of "regression" coefficients that may be obtained. "Best", here, being interpreted to mean that the highest correlation possible between the dependent and independent sets of variates was effected.

The general problem of discriminatory analysis was discussed in Chapter V. It was shown how discriminant analysis may be considered a special case of canonical analysis. Then it was reasoned that the basic problems of classification and discrimination do not differ essentially from the problems of identification, the basic problem in factor analysis. Thus, the tools and techniques developed in discriminatory analysis might be used, with appropriate modifications, to answer the psychologists' problems. Particularly, canonical correlation analysis should be investigated since this technique also involves the isolation of factors. The factors isolated in canonical correlation analysis and factor analysis should be similar if care is taken in the design of the experiment.

In Chapter VI, with certain rather restrictive assumptions, it was shown that the factor loadings for a test not originally factored with a given battery may be estimated or predicted. No attempt was made to attach standard errors to the estimate. If certain assumptions are true, the calculated values are exactly what they would have been had the test been factored with the battery.

Finally, a discussion of Lawley's method of maximum likelihood estimation of factor loadings was discussed. Lawley's test for residual variation after the extraction of r common factors was indicated along with Bartlett's work on tests of significance of the roots of the determinantal equation $|R - \lambda I| = 0$. It was pointed out that these tests are impractical for large batteries of tests because of the prodigious amount of work involved in obtaining either a maximum likelihood or a principal component solution. It is useful, of course, to have these sound

techniques as a background for the approximate solutions necessarily used in practice. It was noted that the empirical tests of significance used by factor analysts probably give about the same results as the theoretically more proper tests by Lawley and Bartlett. Identification and isolation are the main problems facing factor analysts. Tests of significance may not be of too great importance. Usually, only the factors that contribute say, more than, 5% to the total variation are identifiable anyway. Until electronic or other high-speed computers become available, factor analysts will continue to use approximate methods of solution. This is one of the main difficulties with most multivariate problems. The practical worker feels that the amount of information obtained from a multivariate analysis is not worth the labor involved. Thus the suggestion that canonical correlation analysis be used is only of theoretical interest and, at this time, has little chance of being tried out on a large scale in practice.

From these investigations it can be concluded that the problem the factor analysts want to solve, the isolation and identification of basic parameters operating in a given domain, cannot be solved by other existing techniques with the possible exception of canonical correlation analysis. Psychologists are largely responsible for the development and present status of factor analysis. The problem they attempt to solve is surely laudable and of scientific value. Presumably, the technique could be applied in other fields. Factor analysts, when criticized, have replied that the statisticians have not produced tools and techniques that will solve their problem. Statisticians have countered that the psychologists

should modify their question and problems and use existing tools and techniques. Both parties are partly to blame. Surely the statistician has a responsibility to help the psychologists answer their questions. In this thesis an attempt has been made to compare, contrast, and discuss the factor analysis problem and technique with other possible statistical tools the psychologists might be able to use.

Bibliography

- Bartlett, M. S. 1934. The vector representation of a sample. Proc. Cam. Phil. Soc. 30: 327-340.
- Bartlett, M. S. 1938. Further aspects of the theory of multiple regression. *ibid.* 34: 33-40.
- Bartlett, M. S. 1947. Multivariate analysis. *ibid.* 9: 176-197.
- Bartlett, M. S. 1948. Internal and external factor analysis. Brit. Journ. Psych., Stat. Sec. 1: 73-81.
- Bartlett, M. S. 1949. Fitting a straight line when both variables are subject to error. Biometrics. 5: 207-212.
- Bartlett, M. S. 1950. Tests of significance in factor analysis. Brit. Journ. Psych., Stat. Sec. 3: 77-85.
- Bartlett, M. S. 1951a. A further note on tests of significance in factor analysis. *ibid.* 4: 1-2.
- Bartlett, M. S. 1951b. The effect of standardization on χ^2 approximation in factor analysis. Biometrika. 38: 337-344.
- Burt, Cyril. 1947. Factor analysis and analysis of variance. Brit. Journ. Psych., Stat. Sec. 1: 3-26.
- Burt, Cyril. 1949. Alternative methods of factor analysis. *ibid.* 2: 98-120.
- Burt, Cyril. 1952. Tests of significance in factor analysis. *ibid.* 5: 109-133.
- Cox, Gertrude M. 1939. The multiple factor theory in terms of common elements. Psychometrika. 4: 59-68.
- Emmett, W. G. 1949. Factor analysis by Lawley's method of maximum likelihood. Brit. Journ. Psych., Stat. Sec. 2: 90-97.
- Fisher, R. A. 1935. The Design of Experiments. Oliver and Boyd. London.
- Galton, F. 1888. On correlations and their measurement. Proc. Roy. Soc. 45: 135f.
- Grubbs, F. E. 1948. On estimating precision of measuring instruments and product variability. Journ. Am. Stat. Assn. 43: 243-264.

- Henrysson, Sten. 1950. The significance of factor loadings. Brit. Journ. Psych., Stat. Sec. 3: 159-165.
- Hoel, Paul G. 1939. A significance test for minimum rank in factor analysis. Psychometrika. 4: 245-253.
- Holzinger, K. J. and H. H. Harmon. 1941. Factor Analysis. Univ. of Chicago Press. Chicago.
- Hotelling, H. 1933. Analysis of a complex of statistical variables into principal components. Journ. Ed. Psych. 24: 417-441, 498-520.
- Hotelling, H. 1935. The most predictable criterion. *ibid.* 26: 139-142.
- Hotelling, H. 1936. Relations between two sets of variates. Biometrika. 28: 321-327.
- Hotelling, H. 1942. Rotations in psychology and the statistical revolution. Science. 95: 504-507.
- Kendall, M. G. and B. B. Smith. 1950. Factor analysis. Journ. Roy. Stat. Soc., B. 12: 60-94.
- Lawley, D. N. 1940. The estimation of factor loadings by the method of maximum likelihood. Proc. Roy. Soc. Edin. 60: 64-82.
- Lawley, D. N. 1942. Further investigations in factor estimation. *ibid.* 61: 176.
- Lawley, D. N. 1949. Problems in factor analysis. *ibid.* 62: Part IV No. 41.
- Pearson, Karl. 1901. On lines and planes of the closest fit to systems of points in space. Phil. Mag. 6 Sec. II: 559-572.
- Rao, C. R. and P. Slater. 1949. Multivariate analysis applied to difference between neurotic groups. Brit. Journ. Psych., Stat. Sec. 2: 17-29.
- Rao, C. R. 1952. Advanced Statistical Methods. John Wiley and Sons, Inc. New York.
- Smith, H. F. 1950. Estimating Precision of Measuring Instruments. Journ. Am. Stat. Assn. 45: 447-451.
- Spearman, C. 1927. Abilities of Man. Macmillan. London.
- Thompson, G. 1948. Factorial Analysis of Human Ability. Univ. of London Press. London.

- Thurstone, L. L. 1948. Multiple Factor Analysis. Univ. of Chicago Press. Chicago.
- Tukey, John W. 1951. Components in regression. *Biometrics*. 7: 33-69.
- Wald, Abraham. 1940. The fitting of a straight line if both variables are subject to error. *Ann. Math. Stat.* 11: 248-300.
- Wilks, S. S. 1947. *Mathematical Statistics*. Princeton Univ. Press. Princeton, New Jersey.
- Wrigley, C. and J. O. Neuhaus. 1952. A re-factorization of the Burt-Pearson matrix with the Ordvac electronic computer. *Brit. Journ. Psych., Stat. Sec.* 5: 105-108.

Appendix A

GLOSSARY OF TERMS AND SYMBOLS

(The page is listed where the term or symbol is first used or defined in the text).

Attenuation: a correlation coefficient is said to be corrected for attenuation when it is divided by the geometric mean of the reliability coefficients of the two tests. It is denoted by $r_{ij}(t)$. If r_{ij} is the correlation between two tests i and j and r_{ii} and r_{jj} are the estimated reliabilities for tests i and j , respectively, then $r_{ij}(t) = r_{ij} / \sqrt{r_{ii}r_{jj}}$. (Page 78)

Battery of tests: a group of n tests under factorial study. (Page 16)

Canonical form: a transformation of n correlated variates to n uncorrelated variates. The correlation (or variance - covariance matrix) of the transformed variates is a diagonal matrix. (Page 94)

Centroid method of factoring: a method of factoring a correlation matrix which involves summing the columns. The i th column sum is proportional to the factor loading of test i on the common factor. By this method one finds $r < n$ common orthogonal factors. (Page 26)

Common factor: a hypothetical factor that is common to two or more tests in a battery of n tests. The amount of factor p the α th individual possesses, is $X_{p\alpha}$, is estimated by $X_{p\alpha} =$ a linear combination of the test scores as made by the α th individual. The common factor may be

considered as an axis, X_p , or as a particular value taken on this axis, $X_{p\alpha}$. (Page 16)

Communality: ratio of the "common factor variance" to the total variance

of test i . If the scores are standardized, then $\sum_{\alpha} s_{i\alpha} = 0$, $\sum_{\alpha} s_{i\alpha}^2 = N - 1$ so that the i th test has unit variance. The standardized test score as made by the α th individual on the i th test is postulated to be of the form $s_{i\alpha} = \sum_{p=1}^r \alpha_{ip} X_{p\alpha} + \psi_{i\alpha}$. The "fitted" value is $s_{i\alpha} = \sum_{p=1}^r a_{ip} X_{p\alpha} + u_{i\alpha}$. Let

$$h_i^2 = S \left(\sum_{p=1}^r a_{ip} X_{p\alpha} \right)^2 = \sum_{p=1}^r a_{ip}^2.$$

The "common" part of $s_{i\alpha}$ is $\sum_{p=1}^r a_{ip} X_{p\alpha}$, and since the total variance of the test is unity, then h_i^2 is the sample estimate of the population communality $H_i^2 = \text{pop. variance of } \sum_{p=1}^r \alpha_{ip} X_{p\alpha}$.

(Page 22)

Configuration: the representation of the scores for N individuals on n tests (figure 2.2 of text) as n points or vectors in N -space. If the vectors are drawn from the point representing the sample means of each test, then the cosine of the angle between any two test vectors equals the sample correlation between the two tests.

(Page 14)

Correlation matrix: see R , R_a , R_u .

Diagonal matrix: a matrix with diagonal cell entries not necessarily equal, and with zeros elsewhere. (Page 94)

Individual: a generic term meaning person, skull, or any object that is being measured. (Page 11)

Loading (or saturation): the "correlation" between the test and the hypothetical factor p . When the standardized test score is written according as $s_{i\alpha} = \sum_{p=1}^r a_{ip} X_{p\alpha} + u_{i\alpha}$, a_{ip} is said to be the loading or saturation of test i on factor p . The loading is the amount of factor p the i th test includes. (Page 18)

Parallel forms: two sets of test questions, each of which involves questions of the same test material and of equal difficulty; i.e., the two tests are assumed to have the same population mean and variance. (Page 58)

Principal axis solution: the line of "best fit" obtained by minimizing the perpendicular distances of the plotted points of the N individuals to a line drawn through the swarm of points along the direction of greatest variation. (Figure 2.1 of text). Involves solution of the determinantal equation $|R - \lambda I| = 0$. (Page 30)

Raw test score: the test score as observed without any arbitrary alteration or "standardization".

Reliability coefficient: the correlation between the results of two administrations of the same test, or between two parallel forms of the same test. If it is postulated that the test score consists of a "steady" part plus a non-steady part, then the reliability coefficient is the ratio of the variance of the steady part to the total variance of the test. (Page 26)

Saturation: see loading.

Specific part of a test score: the steady part of the test score not common to the other tests in a given battery. Its effect is represented by $\pi_{i\alpha}$ in the factor analysis equation. The sample estimate of $\pi_{i\alpha}$ is denoted by $f_{i\alpha}$. (Pages 17, 18)

Standardized: if the raw test score for the α th individual on the i th tests is denoted by $Y_{i\alpha}$, and if μ_i and σ_{ii} are respectively, the population mean and variance of test i , the usual definition of a standardized variate is $(Y_{i\alpha} - \mu_i) / \sqrt{\sigma_{ii}}$. Since μ_i and σ_{ii} are seldom known, the corresponding sample values are used, so that, in this thesis, the quantity $s_{i\alpha} = (Y_{i\alpha} - \bar{Y}_i) / \sqrt{\hat{\sigma}_{ii}}$ is said to be standardized. It is sometimes convenient to use

$$s_{i\alpha}^* = s_{i\alpha} / \sqrt{N-1} = (Y_{i\alpha} - \bar{Y}_i) / \sqrt{(N-1) \hat{\sigma}_{ii}}$$

The starred quantities are said to be in unitary standard measure since $\sum_{\alpha} (s_{i\alpha}^*)^2 = 1$. (Page 12)

Steady part (of test score): that part of the test score that is constant when scored on the same individual repeatedly. In the notation of this paper, the steady part in the population is denoted by i_{α} and in the sample as $c_{i_{\alpha}}$, thus

$$i_{\alpha} = \sum_{p=1}^r a_{ip} X_{p\alpha} + \pi_{i_{\alpha}}; \quad c_{i_{\alpha}} = \sum_{p=1}^r a_{ip} X_{p\alpha} + f_{i_{\alpha}}.$$

Unique (part of test score): that part of the test score peculiar only to that test score. It is that part of the test score not common to the other tests in a given battery; thus, it is the sum of the specific and error parts of the test score. In the notation of this chapter, the population value is denoted by i_{α} and the sample value by $u_{i_{\alpha}}$, thus $i_{\alpha} = \pi_{i_{\alpha}} + \epsilon_{i_{\alpha}}; \quad u_{i_{\alpha}} = f_{i_{\alpha}} + e_{i_{\alpha}}$. (Page 18)

Unitary standard measure: see standardized. (Page 12)

Variance: unless otherwise specified, the variance (and covariance) shall refer to sums over individuals with the test fixed.
(Page 12)

a_{ig} (= r_{ig}): loading or saturation of test i on factor X_g . (Page 39)

a_{ip} : factor loading of the i th test on the p th factor. The (i,p) th element of the matrix F_a , where $F_a F_a' = R_a$ and $F_a' F_a = \frac{2}{a}$. Used to denote the factor loadings resulting from a factorization of R_a , the "reduced" correlation matrix. (Page 17)

b_{ip} : factor loading of ith test on the pth factor; the (i,p) th element of F , where $FF' = R$, $F'F = \Lambda^2$. Used to denote the factor loadings resulting from a factorization of R . In general $a_{ip} \neq b_{ip}$, since the a_{ip} are the factor loadings resulting from a factorization of R_a , the "reduced" correlation matrix. (Page 24)

$c_{i\alpha} = \sum_{p=1}^r a_{ip} X_{p\alpha} + f_{i\alpha}$. (See steady part of test score in glossary).
(Page 26).

$e_{i\alpha} = s_{i\alpha} - \sum_{p=1}^r a_{ip} X_{p\alpha} - f_{i\alpha}$. Deviation between test score and "fitted" value. Sample estimate of the random error term $\varepsilon_{i\alpha}$. (Page 17)

$f_{i\alpha} = \sum_{p=r+1}^M a_{ip} X_{p\alpha}$. Sample estimate of the specific part of the test score, $\pi_{i\alpha}$, of the standardized test score i on the α th person.
(Page 18).

F : $n \times n$ matrix of factor loadings such that $FF' = R$, and $F'F = \Lambda^2$.
(Page 24)

F_a : $n \times r$ matrix of factor loadings, $r < n$; $F_a F_a' = R_a$ and $F_a' F_a = \Lambda_a^2$.
(Page 19)

h_i^2 : sample estimate of H_i^2 , the communality of test i = variance of the "common" part of test i = $S \left[\sum_{p=1}^r a_{ip} X_{p\alpha} \right]^2 = \sum_{p=1}^r a_{ip}^2$.
Variance of that part of test i that is common to 2 or more tests in a given battery. (Page 22)

H_i^2 : population communality of test i = population variance of
 $(\sum_{p=1}^r \lambda_{ip}^2 X_p)$. Variance of that part of test i that is common to
 2 or more tests in a given battery. Estimated by h_i^2 . (Page 22)

i, j, k : subscripts pertaining to tests.

L : matrix of direction cosines. The elements of the p th column are the direction cosines of the n test axes in relation to the p th common factor; the elements of the i th row are the direction cosines of the i th test axis in reference to all common factor axes. Also, the p th column of L is the p th latent vector of R corresponding to the p th latent root, λ_p , where λ_p is the p th largest root of $|R - \lambda I| = 0$.
 $L L' = L' L = I$, the identity matrix. (Page 33).

n : number of tests in a given battery of tests.

N : number of individuals on which the tests are scored.

p : subscript referring to common factors; the range is from 1 to $r < n$.

r : number of "common" factors in a given battery of tests.

$$\rho_c = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_e^2} \quad (\text{Page 57})$$

r_c = estimate of $\frac{\sigma_p^2}{\sigma_p^2 + \sigma_e^2}$; intra-class correlation coefficient.
 (Page 57)

r'_c : estimate of $\frac{\sigma_p^2}{\sigma_p^2 + \sigma_e^2}$; intra-class correlation coefficient

obtained after eliminating a constant between replications or times. (Page 58)

r'_{ii} : estimate of the reliability coefficient ρ_{ii} . (Page 58)

r_{ij} : product moment correlation coefficient between tests i and j.

r_{ig} : correlation of test i with hypothetical factor X_g . (Page 42)

$r_{ij}(t) = r_{ij} / \sqrt{r'_{ii} r'_{jj}}$: correlation coefficient corrected for "attenuation". (Page 78)

$$R = \begin{bmatrix} 1 & r_{12} & \dots & r_{1n} \\ r_{21} & 1 & \dots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \dots & 1 \end{bmatrix} = \text{usual correlation matrix;} \\ r_{ij} = r_{ji}. \text{ (Page 20)}$$

$$R^* = \begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1n} \\ \rho_{21} & 1 & \dots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \dots & 1 \end{bmatrix}; \quad \rho_{ij} = \rho_{ji}. \text{ (Page 21)}$$

$$R_a = \begin{bmatrix} h_1^2 & r_{12} & \dots & r_{1n} \\ r_{21} & h_2^2 & \dots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \dots & h_n^2 \end{bmatrix};$$

"reduced" correlation matrix; $r_{ij} = r_{ji}$, $i \neq j$.
 h_i^2 = communality for test i . (Page 22)

$$R_a^* = \begin{bmatrix} H_1^2 & \rho_{12} & \dots & \rho_{1n} \\ \rho_{21} & H_2^2 & \dots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \dots & H_n^2 \end{bmatrix};$$

H_i^2 = population communality for test i ;
 $\rho_{ij} = \rho_{ji}$, $i \neq j$. (Page 21)

$$R_r = \begin{bmatrix} r_{11}' & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22}' & \dots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \dots & r_{nn}' \end{bmatrix};$$

$r_{ij} = r_{ji}$, $i \neq j$
 r_{ii}' = estimate of the reliability of test i .
 (Page 75)

$$R_u^* = \begin{bmatrix} \sigma^2(u_1) & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & \sigma^2(u_n) \end{bmatrix}$$

(Page 21)

$$R_u = \begin{bmatrix} \widehat{\sigma^2}(u_1) & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & \widehat{\sigma^2}(u_n) \end{bmatrix}$$

(Page 22)

- $s_{i\alpha} = (Y_{i\alpha} - \bar{Y}_i) / \sqrt{\hat{\sigma}_{ii}^2}$. (See standardized). (Page 12)
- $s_{i\alpha}^* = s_{i\alpha} / \sqrt{(N-1)}$. (Page 12).
- t_d : tetrad (Page 47)
- $u_{i\alpha}$: See unique part of test score. (Page 18)
- v_{ii} : estimate of the variance component σ_p^2 , the variance of the "steady" part of the test.

 $= \frac{1}{n} [V_2(\text{test } i) - V_3(\text{test } i)]$, $n \geq 2$

 $= \frac{1}{n} [\text{Mean square for persons (test } i) - \text{mean square for error (test } i)]$ (Page 70)
- v_{ij} : estimate of the variance component σ_{γ}^2

 $= \frac{1}{n} [V_2(X,Y) - V_3(X,Y)]$, $n \geq 2$

 $= \frac{1}{n} [\text{Mean square for persons on joint analysis of } X, Y - \text{Mean square for error on joint analysis of } X, Y]$. (Page 70)
- $V = [v_{ij}]$: matrix of variance components (Page 70)
- V_a : The matrix V with diagonal elements v_{ii} adjusted downward to include only variances of common factors. (Page 75)
- V_2 : Mean square for persons for 2 or more replications of the same test. (Chapter III)

- V_3 : Mean square for error for 2 or more replications of the same test. (Chapter III).
- $V_2(X,Y)$: Mean squares for persons for a joint analysis of two different tests X and Y given to the same persons. (Page 70).
- $V_3(X,Y)$: Mean square for error for a joint analysis of two different tests X and Y given to the same persons. (Page 70).
- $X_{i\alpha}, Y_{i\alpha}$: unstandardized or raw test scores.
- X_g : common factor "g". (See $X_{g\alpha}$).
- $X_{g\alpha}$: the ability of the α th individual to detect or respond to factor g. (Section 3, Chapter III).
- $X_{p\alpha}$: the ability of individual α to detect or respond to factor p. (Page 17).
- U : n x N matrix of unique parts of standardized scores. (Page 20).
- α, β : Greek subscripts referring to individuals.
- α_{ip} : The saturation of test i on factor p, estimated by a_{ip} . Used when model is written as $s_{i\alpha} = \sum_p \alpha_{ip} X_{p\alpha} + \psi_{i\alpha}$. (Page 17)

- β_{ip} : The saturation of test i on factor p , estimated by b_{ip} . Used when model is written as $s_{i\alpha} = \sum_p \beta_{ip} x_{p\alpha}$; i.e., with no specific or random error terms. (Page 24).
- γ : subscript referring to a common part between two different tests. Also, γ_{α} represents the "common part" of raw test scores. (Page 67).
- $\epsilon_{i\alpha}, \epsilon_{\alpha}, \delta_{\alpha}$: random error terms (Pages 17, 67).
- $\xi_{\alpha}, \rho_{\alpha}$: specific parts of unstandardized test scores. (Page 67).
- $\pi_{i\alpha}$: specific part of standardized test score, estimated by $f_{i\alpha}$. (Page 17).
- $\psi_{i\alpha} = \pi_{i\alpha} + \epsilon_{i\alpha}$: unique part of standardized test score, estimated by $u_{i\alpha}$. (Pages 18, 25).
- $x_{p\alpha}$: amount of factor p the α th individual possesses, estimated by $X_{p\alpha}$. (Page 17).
- $\omega_{i\alpha} = s_{i\alpha} - \epsilon_{i\alpha}$: The steady part of the standardized test score. (Page 25).
- η_i, η_y, η_x : expected values of raw scores. (Page 66).
- ρ_{ij} : population correlation coefficient between tests i and j , estimated by r_{ij} .

ρ_{α} : the person effect for the α th individual when $Y_{i\alpha}$ is written as $\eta_i + \tau_i + \rho_{\alpha} + (\tau\rho)_{i\alpha} + \varepsilon_{i\alpha}$. (Page 19).

$\rho_t = \frac{E(\text{steady part of X, steady part of Y})}{\sqrt{E(\text{steady part of X})^2(\text{steady part of Y})^2}}$
(Page 78).

$\theta_{\alpha}, \phi_{\alpha}$: the person effect of the α th individual when the raw score is written as $Y_{i\alpha} = \eta_i + \phi_{\alpha} + \varepsilon_{i\alpha}$; $\phi_{\alpha} = \gamma_{\alpha} + \xi_{\alpha}$ and $\theta_{\alpha} = \delta_{\alpha} + \int_{\alpha}$. Thus, the person effects are the steady parts of the raw test scores, consisting of the common part plus the specific part. (Page 66).

$\sigma^2(e_i)$: error variance of standardized test i, equal the population variance of $(\varepsilon_{i\alpha})$. (Page 23).

$\widehat{\sigma^2(e_i)}$: estimate of $\sigma^2(e_i)$. (Page 24).

$\sigma^2(f_i)$: specific variance of test i, equal the population variance of $(\pi_{i\alpha})$. (Page 23).

$\widehat{\sigma^2(f_i)}$: estimate of $\sigma^2(f_i)$. (Page 24).

$\sigma^2(u_i)$: unique variance of test i, equal the population variance of $(\psi_{i\alpha} = \pi_{i\alpha} + \varepsilon_{i\alpha})$. (Page 22).

$\widehat{\sigma^2(u_i)} = \sum_{\alpha} u_{i\alpha}^2 / (N - 1)$: estimate of $\sigma^2(u_i)$. (Page 23).

$\sigma^2(c_i)$: variance of the steady part of the test i , equal the population variance of $(i) = \sum_p \alpha_{ip} \lambda_{p\alpha} + \pi_{i\alpha}$. (Page 26).

λ_p : pth root (numbering downwards) of the determinantal equation $R - \lambda I = 0$. (Section 2, Chapter III).

λ_{ap} : pth root (numbering downwards) of the determinantal equation $R_a - \lambda I = 0$. (Section 2, Chapter III).

$${}^2 = \begin{matrix} \lambda_1 & & 0 \\ & \cdot & \\ & & \cdot \\ 0 & & \lambda_n \end{matrix} \quad (\text{Page 36}).$$

$${}^4 = \begin{matrix} \lambda_1^2 & & 0 \\ & \cdot & \\ & & \cdot \\ 0 & & \lambda_n^2 \end{matrix} \quad (\text{Page 37}).$$

$${}^2_a = \begin{matrix} \lambda_{a1} & & 0 \\ & \cdot & \\ & & \cdot \\ 0 & & \lambda_{ar} \end{matrix} \quad (\text{Page 38}).$$

$${}^4_a = \begin{matrix} \lambda_{a1}^2 & & 0 \\ & \cdot & \\ & & \cdot \\ 0 & & \lambda_{ar}^2 \end{matrix}$$

Notations used in Chapter IV.

$B_{21} = S_2 S_1' (S_1 S_1')^{-1} = C_{21} C_{11}^{-1}$: $q \times p$ matrix of regression coefficients.

$$S_2 = B_{21} S_1 + S_{2.1}$$

S_1 : $p \times N$ matrix of observations of p tests on N persons from an "independent" set.

S_2 : $q \times N$ matrix of observations of q tests on N persons of a "dependent" set.

$$C_{21} = S_2 S_1'$$

$$C_{12} = C_{21}' = S_1 S_2'$$

$$C_{11} = S_1 S_1' ; C_{11}^{-1} = (S_1 S_1')^{-1}$$

$$C_{22} = S_2 S_2' ; C_{22}^{-1} = (S_2 S_2')^{-1}$$

$S_{2.1}$ = $q \times N$ residual matrix of S_2 not explained by regression of S_2 on S_1 .

$C_{22.1} = S_{2.1} S_{2.1}'$ = within or variance - covariance matrix of "errors" or "residuals".

Other relationships.

$$r_{ii}' = 1 - \sigma^2(e_i). \quad (\text{Page 26})$$

$$h_i^2 = \sum_{p=1}^r a_{ip}^2 < 1, r < n.$$

$$h_i^2 + \widehat{\sigma^2(f_i)} = r_{ii}'$$

$$H_i^2 + \sigma^2(f_i) = \rho_{ii}'$$

$$\sigma^2(u_i) = \sigma^2(f_i) + \sigma^2(e_i).$$