PSYCHOMETRIKA—VOL. 20, NO. 2
JUNE, 1955

ESTIMATION AND TESTS OF SIGNIFICANCE IN FACTOR ANALYSIS

C. RADHAKRISHNA RAO

VISITING RESEARCH PROFESSOR, UNIVERSITY OF ILLINOIS

A distinction is drawn between the method of principal components developed by Hotelling and the common factor analysis discussed in psychological literature both from the point of view of stochastic models involved and problems of statistical inference. The appropriate statistical techniques are briefly reviewed in the first case and detailed in the second. A new method of analysis called the canonical factor analysis, explaining the correlations between rather than the variances of the measurements, is developed. This analysis furnishes one out of a number of possible solutions to the maximum likelihood equations of Lawley. It admits an iterative procedure for estimating the factor loadings and also for constructing the likelihood criterion useful in testing a specified hypothesis on the number of factors and in determining a lower confidence limit to the number of factors.

1. *Introduction*

Whatever may be the arguments for or against factor analysis as a tool in psychological research, the statistical problems it involves have been of considerable interest to the statistician mainly because of their complexity. Two important contributions on the statistical side are by Hotelling (**8**), who introduced the principal component analysis, and Lawley (**11**, **12**), who provided a test criterion for judging the significance of factors in addition to working out the maximum-likelihood equations of estimation. These two authors were, however, considering two different problems, both of which seem to have important application. They are sometimes considered as two possible formulations of the same problem providing the same answer. In theory it helps to make a distinction between the two. The term principal component analysis (PCA) should be used for Hotelling's formulation of the problem and its solution; the term factor analysis should be used for the specialized formulation considered in psychological literature and for the various solutions offered (see also **10**). Lawley was considering the latter problem under the assumption that the variables (test scores) are normally distributed.

Illustrations have appeared from time to time to show that PCA gives nearly the same relative magnitudes of factor loadings as any effective method of factor analysis. This is true only when what have been termed as communalities are very nearly equal for all the tests as shown in section 3.1 of this paper.

The PCA is sometimes modified (3, p. 114; 7) by the insertion of communalities in the diagonal of the correlation matrix. This method, called the principal factor analysis (PFA), seems to provide a valid approach to the problem of factor analysis; however, it carries with it the flavor of principal component analysis intended to explain the variations in the standardized scores. It will be seen that an alternative approach developed in section 3.2 explains most effectively the correlations between the test scores in a battery. This method may be called a canonical factor analysis (CFA). Formulas for estimation are detailed in section 4.

The tests of significance associated with component analysis and factor analysis also differ to some extent. In the former case interest chiefly lies in the magnitude of, or the relationship between, the latent roots of the hypothetical matrix of raw correlations or those corrected for attenuation. In factor analysis it is the decomposition of the correlation matrix as the sum of a diagonal matrix and a positive semi-definite matrix. The differences in nature of these tests are sometimes fundamental (section 2.2). The tests of component analysis are contained in Hotelling's paper (8), and an appropriate test for factor analysis is given by Lawley (11). An alternative form of Lawley's test yielding slightly more precise results is given in section 4.2. It is also shown (section 4.3) that the test criterion can be calculated during the process of estimation and used in obtaining a lower confidence limit to the number of factors.

Recently Bartlett (1, 2) proposed a test involving the latent roots of the correlation matrix intended to study "the correlation structure in relation to the variance of the measurements." The exact nature of the hypothesis for which Bartlett's test is applicable and the conditions under which it is valid are examined in section 2.2. It appears that this test does not provide a complete answer to either form of analysis under consideration.

For a full account of tests of significance in factor analysis developed up to 1952, the reader is referred to Burt (3).

The author of this article is not concerned here in examining which of the methods, component or factor analysis, is relevant in problems of psychological research or whether both methods provide rather similar numerical results (not identical in general) leading to the same psychological interpretation. The main emphasis is on the differences in statistical techniques needed in these two cases and a detailed examination of the methods for factor analysis.

2. *Problems of Factor and Component Analyses*

2.1 *Factor Analysis*

Factor analysis postulates an underlying structure of a set of measurements in terms of hypothetical variables (non-observable) depending on

what are called common and specific or individual factors. If $x_1, \cdots, x_p$ denote p different measurements on an individual, then x_i is written

$$x_i = z_i + s_i \qquad (i = 1, \cdots, p), \tag{2.1.1}$$

where z_i, the variables depending on common factors, and s_i, the variables depending on specific factors, satisfy the following conditions of zero covariance:

$$\text{cov}(z_i, s_i) = 0, \quad \text{cov}(z_i, s_j) = 0, \quad \text{cov}(s_i, s_j) = 0 \qquad (i \neq j). \tag{2.1.2}$$

Sometimes, another independent variable representing unreliability in the measurement x_i is added to $(z_i + s_i)$, but for purposes of factor analysis based on unrepeated test scores of individuals, this variable can be combined with s_i without loss of generality. If such repeated test scores are available, then a more comprehensive analysis of the common and specific factors is possible. This latter analysis is not considered here.

From the structural setup (2.1.1), (2.1.2) it follows

$$\begin{aligned} V(x_i) &= V(z_i) + V(s_i) \\ \sigma_{ii} &= \gamma_{ii} + \delta_{ii} . \end{aligned}$$

$$\begin{aligned} \text{cov}(x_i, x_j) &= \text{cov}(z_i, z_j) + \text{cov}(z_i, s_j) + \text{cov}(s_i, z_j) + \text{cov}(s_i, s_j) \\ &= \text{cov}(z_i, z_j) \\ \sigma_{ij} &= \gamma_{ij} \qquad (i \neq j). \end{aligned}$$

If Σ, Γ, Δ denote the dispersion matrices of the vector variables $\underline{x}$, $\underline{z}$, $\underline{s}$, then

$$\Sigma = \Gamma + \Delta,$$

where Δ is a diagonal matrix.

It is seen from the above analysis that any correlation between x_i, x_j is solely due to the correlation between z_i, z_j. What we can actually observe are the values of the variables $\underline{x}$ on a group of individuals but not $\underline{z}$, $\underline{s}$ which are not operationally defined, but whose hypothetical existence is postulated. We thus obtain an estimate of the matrix Σ. The subject of factor analysis is mainly concerned with the estimation of the matrix Γ starting with an estimate of Σ. The object is not to find any matrix Γ satisfying the condition $\Sigma = \Gamma + \Delta$ but the one which has the least complexity leading to a parsimonious description of the relationships between the observable variables $\underline{x}$. The complexity, when defined as the rank of the matrix Γ, has a special significance for the problems on which this technique is applied, as shown in the subsequent sections of this paper.

Some of the statistical problems of factor analysis are:

(a) to estimate the minimum rank of the dispersion matrix (variances and covariances) Γ of the variables $z_1, \cdots, z_p$ occurring in the structural equations (2.1.1, 2.1.2),

(b) to test any hypothesis specifying the minimum rank of Γ,

(c) to estimate *a basis of the common factor space* (defined below),

(d) to predict the value of any common factor from the observed set $x_1, \cdots, x_p$ for any individual.

The statement that the rank of Γ is $k < p$ implies that the variables $z_1, \cdots, z_p$ can be expressed as linear combinations of k independent variables only. To bring out the precise meaning of such a dependence let us consider the entire *vector space* of elements consisting of all linear combinations of the set $z_1, \cdots, z_p$ of variables introduced into the different tests of a battery with the restriction that any two variables differing by a constant represent the same element. We may call any element of this space a common factor variable or simply a factor variable unless otherwise specified. The vector product of any two elements f and g of this space is defined by $\text{cov}(f, g)$ and the square of the norm of f by variance of f, $V(f)$.

Vectors $f_1, f_2, \cdots, f_r$ of this space are said to be independent if no linear combination $a_1f_1 + a_2f_2 + \cdots + a_rf_r$ (all $a_i \neq 0$ simultaneously) has zero norm. A vector space is said to be finite dimensional if all its elements can be expressed as linear combinations of a finite number of elements. In such a case there is a minimal number of such elements called the rank of the space. A set of such elements necessarily independent (to be minimal) is called a basis. A basis of a vector space is not, however, unique but its rank is.

We can always choose a basis such that its elements are orthogonal (zero vector product), implying that the chosen factor variables are uncorrelated. A convenience provided by such a choice is that a basis can be simply represented by a set of correlations between the measurements and factors. If $Z_1, \cdots, Z_k$ is an orthogonal basis then each z_i can be expressed in terms of Z_j.

$$z_i = a_{i1}Z_1 + \cdots + a_{ik}Z_k . \tag{2.1.3}$$

The covariance of x_i with Z_j is a_{ij}, the coefficient of Z_j in the representation (2.1.3). This may be regarded as a correlation coefficient once x_i and Z_j are properly standardized. A basic set of factors or equations (2.1.3) can be represented by the matrix of correlations

$$\begin{bmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & & \vdots \\ a_{p1} & \cdots & a_{pk} \end{bmatrix}, \tag{2.1.4}$$

which is also called the factor loading matrix. Such a basis is not unique because the choice of Z_i, or the representation (2.1.3), is not unique. A basis is just meant to generate the entire space of factor variables which are linear combinations of some hypothetical variables introduced into the various

tests. From this point of view one basis is as good as any other. Of course, given one basis, orthogonal or oblique, the other can be derived by a linear transformation.

The choice of a suitable basis which is "psychologically meaningful" has largely rested with the psychologist, perhaps rightly so. But it is quite conceivable, once the psychological meaning is translated to mean some precisely stated restrictions on the basis, that its choice will turn out to be a problem of statistical estimation. In this sense the graphical methods of rotation of factor loadings advocated by Thurstone (**17**) and the quadrimax method of Neuhaus and Wrigley (**13**) are statistical methods of factor analysis, where the number of zero or small loadings is maximized. Such a restriction, or even the orthogonality of a basis, may not be the most helpful in leading to a suitable psychological interpretation or the discovery of "real entities." The choice of the restrictions to be imposed on the basis is perhaps a problem for psychological research. An objective method developed in this connection by Cattell (**4, 5**) seems to have interesting possibilities from a statistical point of view.

2.2. *Principal Components*

It was pointed out by Sir Cyril Burt that this method was originally put forward by Karl Pearson in 1901. But the statistical problems of estimation and testing connected with the principal components were first considered by Hotelling.

Hotelling (**8**) considers two types of problems:

First, without assuming a decomposition of the measurements as in (2.1.1), hypotheses are framed in terms of the latent roots of the correlation matrix with a view to studying the shape of the scatter of the standardized scores in the p-dimensional space or alternatively the relative importance of the different principal components in explaining the total variance. For instance, if some of the calculated roots are not significantly different then the components corresponding to them may be considered equally important.

Let us consider a specific hypothesis that the $(i + 1)$th to the $(i + r)$th roots of the population correlation matrix (denoted by ρ) are equal. The value of i may be 0, 1, $\cdots$ to $(p - r)$.

This hypothesis imposes a restriction on ρ, viz., that it admits the decomposition

$$\rho = \theta + \lambda I, \tag{2.2.1}$$

where θ is a matrix of rank $(p - r)$, λ is the common value of the roots from $(i + 1)$th to $(i + r)$th, and I is the unit matrix.

Any such hypothesis or a similar one based on the dispersion matrix Σ instead of ρ can be tested by a likelihood-ratio criterion Λ, provided sample size is moderately large. Exactly how large the sample size should be is a

matter for further investigation. The statistic $(-2 \log_e \Lambda)$ is distributed in large samples as χ^2 with degrees of freedom equal to the number of restrictions on the free parameters imposed by the hypothesis.

The number of restrictions in a hypothesis of the form (2.2.1) is equal to the number of restrictions on the symmetric matrix θ with rank $(p - r)$ minus one for the unknown λ. Only $(p - r)$ rows and columns in θ are independent; the rest of the elements depend on them. Therefore, the number of restrictions on the elements of θ is

$$(p - r)(p - r + 1)/2,$$

and with one less we have

$$(p - r - 1)(p - r + 2)/2 \tag{2.2.2}$$

degrees of freedom for the χ^2 approximation.

If A is the estimated dispersion matrix from observations on n individuals, then the test criterion for the hypothesis (2.2.1) with Σ instead of ρ is

$$-(n - 1) \log \frac{|A|}{|\hat{\Sigma}|}, \tag{2.2.3}$$

where $|\hat{\Sigma}|$ is the estimated dispersion matrix under the conditions of the hypothesis. The latent roots of $\hat{\Sigma}$

$$\mu_1, \mu_2, \cdots, \mu_i, \mu_{i+1}, \cdots, \mu_{i+r}, \mu_{i+r+1}, \cdots, \mu_p$$

are connected with the latent roots of A in the following way:

$$\lambda_j = \mu_j \qquad (j = 1, \cdots, i, i + r + 1, \cdots, p),$$

$$\mu_{i+1} = \frac{\lambda_{i+1} + \cdots + \lambda_{i+r}}{r}.$$

Since

$$|A| = \lambda_1 \cdots \lambda_p, \qquad |\hat{\Sigma}| = \mu_1 \cdots \mu_p,$$

the ratio $|A|/|\hat{\Sigma}|$ is

$$\frac{\lambda_{i+1} \cdots \lambda_{i+r}}{\left(\frac{\lambda_{i+1} + \cdots + \lambda_{i+r}}{r}\right)^r}, \tag{2.2.4}$$

which is a suitable power of the ratio of the geometric to the arithmetic mean of the $(i + 1)$th to $(i + r)$th roots of A. From this point of view it would appear, by choosing $i = (p - r)$ in (2.2.4), that Bartlett's (1) test using the dispersion matrix instead of the correlations is valid for judging the significance of equality of the least r roots.

Unfortunately the test does not seem to reduce to the form (2.2.4) in terms of the roots of the observed correlation matrix R when the hypothesis is as stated in (2.2.1) in terms of the population correlation matrix ρ. The effect of standardizing the variables by the *sample standard deviations* is not properly allowed for by a criterion of the form (2.2.4). This is also partly revealed by Bartlett's own evaluation of the degrees of freedom by the expectation method in a simple case. They depend on the unknown correlations and reach the value (2.2.2) only in a limiting case, while for a genuine likelihood ratio this is not expected. The exact evaluation of the test criterion depends on complicated equations which require further investigation.

Secondly, Hotelling considers the problem of "testing the variances of components against the variance to be expected on account of the inaccuracy of the tests as revealed by their self-correlations or reliability coefficients." For this purpose a test score is thought of as made up of two parts, a true score with variance unity and a random error. Thus

$$x_i = X_i + \epsilon_i \qquad (i = 1, \cdots, p), \tag{2.2.5}$$

with the conditions $\text{cov}(\epsilon_i, \epsilon_j) = 0$, $(i \neq j)$. The hypothesis stated above is interpreted to imply that the true scores X_i are linearly dependent, i.e., "the scatter diagram of the true scores will lie in a flat space of smaller dimensionality immersed in the p-dimensional space." If independent estimates of the variances of ϵ_i are available, either from an external source or by re-tests on individuals, there is no need to consider the true scores as random variables in order to test the above hypothesis. The general multivariate tests of dimensionality developed in more complicated situations are directly applicable for this problem. The non-stochastic model on the scores corrected for unreliabilities used in testing the second hypothesis provides a strong contrast to tests in factor analysis where, of necessity, all the variables (the common and specific factors) involved are considered to be stochastic, which makes the problem more complex.

3. *Special Characterizations of a Basis in Factor Analysis*

Using the vector notation

$$\underset{\sim}{x} = (x_1, \cdots, x_p), \qquad \underset{\sim}{z} = (z_1, \cdots, z_p), \qquad \underset{\sim}{s} = (s_1, \cdots, s_p).$$

The equation (2.1.1) can be written

$$\underset{\sim}{x} = \underset{\sim}{z} + \underset{\sim}{s}. \tag{3.1}$$

The dispersion matrix of $\underset{\sim}{x}$ (using D for dispersion) is

$$D(\underset{\sim}{x}) = D(\underset{\sim}{z}) + D(\underset{\sim}{s}),$$

or

$$\Sigma = \Gamma + \Delta, \tag{3.2}$$

where Σ, Γ and Δ are defined by equation (3.2). The covariance of $\underline{z}$ and $\underline{s}$ is zero because of conditions (2.1.2). The matrix Γ is positive semi-definite with rank $k < p$ and Δ is a positive-definite diagonal matrix. The equation (3.2) supplies the fundamental decomposition of the dispersion matrix Σ in terms of those of the hypothetical variables postulated by a factorial structure. If the rank of Γ is $k < p$, the space of common factors has a basis of k independent factors as shown in section 2.1. For a proper identification of the space and "an orderly selection of independent factors" there is a need to characterize a basis in a convenient way. A basis so characterized need not admit a psychological interpretation, for only mathematical and statistical convenience is being sought at this stage. A basis once obtained can always be transformed to meet other requirements. Two special characterizations are discussed here.

3.1 *First Characterization*

Let $\underline{l} = (l_1, \cdots, l_p)$ be a vector of arbitrary coefficients giving rise to a new factor variable

$$\underline{l}\underline{z}' = l_1 z_1 + \cdots + l_p z_p .$$

The variation in the variable x_i explained by the factor variable $\underline{l}\underline{z}'$, is

$$\frac{\operatorname{cov}^2(x_i, \underline{l}\underline{z}')}{V(\underline{l}\underline{z}')} = \frac{(l_1\gamma_{1i} + \cdots + l_p\gamma_{pi})^2}{\underline{l}\Gamma\underline{l}'}, \tag{3.1.1}$$

assuming that $\underline{l}\ \Gamma\ \underline{l}'$, the variance of $\underline{l}\underline{z}'$, is not zero. The total variation explained in all the variables is

$$\frac{\left\{\sum_1^p (l_1\gamma_{1i} + \cdots + l_p\gamma_{pi})^2\right\}}{\underline{l}\Gamma\underline{l}'} = \frac{\underline{l}\Gamma\Gamma\underline{l}'}{\underline{l}\Gamma\underline{l}'}. \tag{3.1.2}$$

Let us choose $\underline{l}$ such that (3.1.2) is a maximum. Differentiating with respect to the vector $\underline{l}$ (see **14**, p. 21), the equation leading to the optimum value λ of the ratio (3.1.2) and the vector $\underline{l}\ \Gamma$ is

$$\underline{l}\Gamma\Gamma - \lambda\underline{l}\Gamma = 0$$

or eliminating $\underline{l}\,\Gamma$

$$|\ \Gamma - \lambda I\ | = 0, \tag{3.1.3}$$

where I is the identity matrix. This shows that λ is the maximum latent root of Γ and $\underline{m} = \underline{l}\Gamma$ is the latent vector corresponding to it. Since the vector $\underline{m}$ satisfies the equation

$$\underline{m}\Gamma = \lambda\underline{m},$$

and

$$\underline{m} = \underline{l}\Gamma,$$

so that

$$\underline{m}\Gamma = \lambda \underline{l}\Gamma, \tag{3.1.4}$$

the vector $\underline{m}$ itself can be taken to be a solution of $\underline{l}$. We thus obtain the first factor variable as a linear combination of $z_1, \cdots, z_p$. From the theory of canonical roots and vectors (**14**, p. 24), it would then follow that the second factor variable, which explains the highest proportion of the residual variation independently of the first, is the linear combination corresponding to the second canonical vector. There are as many linear combinations as there are non-zero roots λ, which is equal to the rank of the matrix Γ. The linear combinations of $z_1, \cdots, z_p$ supplied by the canonical vectors of zero roots of λ vanish identically, indicating the dependence of the factor variables associated with the measurements $x_1, \cdots, x_p$.

The factor loading of the variable x_i on the first factor chosen above is the correlation between the two. The covariance is

$$\text{cov}\,(x_i, \underline{l}\underline{z}') = l_1\gamma_{1i} + \cdots + l_p\gamma_{pi} = \lambda_1 l_i,$$

and if the variables x_i are initially chosen to have unit variance the correlation is

$$\frac{\lambda_1 l_i}{\sqrt{\underline{l}\Gamma\underline{l}'}} = \frac{\lambda_1 l_i}{\sqrt{\lambda_1 \underline{l}\underline{l}'}} = \frac{\sqrt{\lambda_1}\, l_i}{\sqrt{l_1^2 + \cdots + l_p^2}}. \tag{3.1.5}$$

The factor loadings are then the elements of the first canonical vector suitably standardized. Similarly the factor loadings of any other factor are derived from the canonical vector defining the factor.

Even after exhausting all the independent factor variables, there still remains some variation left in $\underline{x}$ to be explained by the specific factors unless the number of independent common factors is equal to p. In the problem originally considered by Hotelling, the successive components explaining variation in $\underline{x}$ were not confined to the common factor portion $\underline{z}$ but were also functions of the specific factors $\underline{s}$, which then are equivalent to linear functions of $\underline{x}$. Hotelling's principal components are, therefore, important in problems where the total variation of a measurement vector $\underline{x}$ is sought to be accounted for, to the maximum amount possible, by a smaller number of linear functions of $\underline{x}$. The principal components of Hotelling are derived from the latent vectors of the matrix $\Sigma = \Gamma + \Delta$ instead of Γ alone as used above. It may be observed that when

$$\Delta = \delta^2 I,$$

i.e., when all the specific variables have the same variance δ^2, a latent vector $\underline{l}$ of $\Gamma + \Delta$ satisfying the equation

$$\underline{l}(\Gamma + \delta^2 I) = \mu \underline{l}$$

also satisfies the equation

$$\underline{l}\Gamma = (\mu - \delta^2)\underline{l} = \lambda\underline{l}$$

and is therefore a latent vector of Γ. The principal component analysis of Hotelling is thus a method of factor analysis with the factor loadings inflated keeping the same relative magnitudes, when all the specific variances are the same.

There is some arbitrariness in the above characterization of the basic set of factors because instead of maximizing the sum of the variations explained in $x_1, \cdots, x_p$ we could maximize a weighted sum and arrive at a different basis and consequently a different set of factor loadings. When the variables $x_1, \cdots, x_p$ are chosen to have unit variances the method adopted is equivalent to using reciprocals of total variances as weights.

The quantity δ_i^2, the residual variance of x_i unexplained by the factor variables, satisfies the equation

$$\sigma_{ii} = \frac{\lambda_1 l_i^2}{\underline{l}\underline{l}'} + \frac{\lambda_2 m_i^2}{\underline{m}\underline{m}'} + \cdots + \delta_i^2, \tag{3.1.6}$$

where $\underline{l}, \underline{m}, \cdots$ are the latent vectors defining the factors $\underline{l}\underline{z}', \underline{m}\underline{z}', \cdots$ and the subscript i relates to the ith element in the vectors.

The best formula for predicting a factor variable such as $\underline{l}\underline{z}'$ from the observed measurements $\underline{x}$ is obtained by the method of regression (16). If $\underline{k}\underline{x}'$ is the predicted value, then $\underline{k}$ satisfies the equation

$$\underline{k}\Sigma = \underline{l}\Gamma, \qquad \underline{k} = \underline{l}\Gamma\Sigma^{-1} = \lambda_1\underline{l}\Sigma^{-1}, \tag{3.1.7}$$

and similarly for other factors. The characterization of the basis considered here together with methods of estimation is known as the principal factor analysis (PFA) (7).

3.2 *Second Characterization*

Instead of asking for a factor variable which explains as much of variation as possible of $\underline{x}$, we may pose the problem in a different way. What is that factor variable which is predictable from $\underline{x}$ with the maximum possible precision? Or in other words, what is that factor variable which is maximally related to $\underline{x}$? The solution to this problem depends on a canonical correlation analysis of the hypothetical factor variables $\underline{z}$ with the measurable variables $\underline{x}$ of which $\underline{z}$ constitute a part.

If $\underline{l}\underline{z}'$ and $\underline{q}\underline{x}'$ represent two linear combinations of factor variables and test scores, then according to the theory of canonical correlations (8) the correlation (or its square) between the two linear functions

$$\frac{(\underline{l}\Gamma\underline{q}')^2}{(\underline{l}\Gamma\underline{l}')(\underline{q}\Sigma\underline{q}')} \tag{3.2.1}$$

has to be maximized. Using the algebra developed in a similar genetic problem (14), the optimum value of the correlation ν is found to be a root of the equation

$$| \Gamma - \nu^2 \Sigma | = 0, \tag{3.2.2}$$

or

$$| \Sigma - \lambda \Delta | = 0, \qquad \lambda = 1/(1 - \nu^2). \tag{3.2.3}$$

The vectors $\underline{l}$ and $\underline{q}$ are proportional and satisfy the same equation

$$\underline{l}(\Sigma - \lambda\Delta) = 0, \qquad \underline{q}(\Sigma - \lambda\Delta) = 0. \tag{3.2.4}$$

That factor variable which is highly correlated with $\underline{x}$ is $\underline{l}\underline{z}'$, where $\underline{l}$ is the latent vector corresponding to the largest root of the determinental equation (3.2.2). The second factor variable, uncorrelated with the first and possessing the highest correlation with $\underline{x}$ is $\underline{m}\underline{z}'$, where $\underline{m}$ is the latent vector corresponding to the second root of (3.2.2), and so on. We get as many factors as the number of non-zero values of ν^2 or values of λ greater than unity which is the same as the rank of Γ.

For any factor $\underline{l}\underline{z}'$ as determined above

$$\underline{l}\Gamma\underline{l}' = \nu^2 \underline{l}\Sigma\underline{l}' = \frac{\lambda - 1}{\lambda} \underline{l}\Sigma\underline{l}' = (\lambda - 1)\underline{l}\Delta\underline{l}'$$

$$\text{cov}\,(x_i\,,\ \underline{l}\underline{z}') = l_1\gamma_{1i} + \cdots + l_p\gamma_{pi} = (\lambda - 1)l_i\delta_i^2\,.$$

The correlation between x_i and $\underline{l}\underline{z}'$,

$$\frac{(\lambda - 1)l_i\delta_i^2}{\sqrt{(\lambda - 1)\underline{l}\Delta\underline{l}'}\,\sigma_{ii}}\,, \tag{3.2.5}$$

is the factor loading of x_i on the factor $\underline{l}\underline{z}'$. This is again an element of $\underline{l}$ multiplied by a constant. It can be shown that the same factor loadings are obtained if instead of $(x_1\,,\ \cdots\,,\ x_p)$ we consider $(c_1x_1\,,\ \cdots\,,\ c_px_p)$ with the variables arbitrarily scaled. In the previous case it is necessary to reduce the variables $(x_1\,,\ \cdots\,,\ x_p)$ to unit standard deviation before proceeding to derive factors in order to achieve uniqueness of factor loadings.

To predict the factor measurements we use the regression equation as in (3.1.7). In this case it turns out that $\underline{l}\underline{z}'$, $\underline{m}\underline{z}'$, $\cdots$, defined by the latent vectors of (3.2.3), can be best predicted by $\underline{l}\underline{x}'$, $\underline{m}\underline{x}'$, $\cdots$, avoiding the complication of multiplication by Σ^{-1} necessary in the case of factors defined in the earlier characterization of the basic set (3.1.7).

The residual variance δ_i^2 in x_i unexplained by the factor variables satisfies the equation

$$\sigma_{ii} = \frac{(\lambda_1 - 1)}{\underline{l}\Delta\underline{l}'}\, l_i^2\delta_i^4 + \frac{(\lambda_2 - 1)}{\underline{m}\Delta\underline{m}'}\, m_i^2\delta_i^4 + \cdots + \delta_i^2\,, \tag{3.2.6}$$

similar to the formula (3.1.6) in the earlier case. This second characterization of a basis together with methods of estimation may be called canonical factor analysis (CFA) to bring out its connection with the theory of canonical correlations.

Factor analysis thus fits in a general theory of canonical correlations involving two sets of variables: one set being observable and the other set, observable as in multiple regression; dummy as in multiple discrimination; or hypothetical as in problems of genetic selection.

3.3 *Which Is a Better Characterization?*

This question is meaningless if we are dealing with the true values of the dispersion elements satisfying the conditions of a given rank of the factor-variable space because one can be transformed into the other, and in fact they may be replaced by any other basic set and they all serve the same purpose.

But this is no longer true when we have only *estimates* of the dispersion elements and factors are estimated by formally substituting for Σ the estimated quantities and choosing Δ to satisfy the equation (3.1.6) in the first case (PFA) and (3.2.6) in the second case (CFA). Which then is a better estimate of a basis?

From the point of view of statistical estimation, PFA gives a least-squares estimate (16, p. 119) and CFA, a maximum-likelihood estimate, when normality of the distribution of the observations is assumed. At present there is not much to choose between the two methods except for the following reasons. The maximum-likelihood estimation leads in general to better results when the distribution of the variables is specified. No suitable test based on the least-squares estimates is available while there exists an easily computable test criterion on the basis of maximum-likelihood estimates. No further computations are needed to obtain the factor measurements if the factors are estimated by CFA; in fact, in this method, factors are deduced from a description of their measurement.

There is another logical argument which may have to be borne in mind in deciding the issue. A rigorous hypothesis concerning the number of independent factor variables is perhaps never true, and a test of this null hypothesis can detect its falsehood only when there is a serious departure. If then by following a rule of behavior (as determined by a test criterion) we decide to extract a certain number of factors, any method of estimation may be looked upon as providing only a summary of all the factors in terms of a few dominant ones having a definite existence with magnitudes bigger than standard errors calculable from the observations. It is then of interest to examine whether one method of estimation leads to a better summary than the other and at the same time has low errors of estimation.

From this viewpoint, PFA may be thought of as providing the best k

(given number not necessarily exhaustive) factors explaining the maximum possible variance in the measurements while CFA, the best k factors which have in some sense highest possible correlations with the measurements. This may mean that while the first set attempts to explain as much as possible of the variations in the individual measurements, the latter set focuses on the correlations. Perhaps the psychological interest chiefly lies in the latter set, which offers a better explanation of the correlations between the measurements.

4. *Estimation and Tests of Significance for Factors*

4.1 *Estimation of Factor Loadings*

Let $A = (a_{ij})$ denote the observed dispersion matrix of the vector variable $\underline{x}$. This is sufficient for the estimation of Γ and Δ, the two components of the population dispersion matrix Σ. Following the equations (3.2.3, 3.2.4) of the second characterization, we have on substituting A for Σ

$$| A - \lambda\Delta | = 0,$$

$$\underline{l}(A - \lambda\Delta) = 0, \tag{4.1.1}$$

where $\underline{l}$ is a latent vector corresponding to the latent root λ. From the point of view of mechanical computations it is convenient to solve for

$$\underline{b} = \underline{l}\Delta^{1/2}, \qquad \underline{b}\underline{b}' = 1, \tag{4.1.2}$$

in which case $\underline{b}$ is the latent vector of

$$| \Delta^{-1/2} A \Delta^{-1/2} - \lambda I | = 0. \tag{4.1.3}$$

Let us suppose, for the sake of illustration, that we are extracting two factors. If $\underline{b} = (b_1, \cdots, b_p)$ and $\underline{c} = (c_1, \cdots, c_p)$ are the first two latent vectors of (4.1.3) corresponding to the roots λ_1 and λ_2, then the equation (3.2.6) gives

$$a_{ii} = [(\lambda_1 - 1)b_i^2 + (\lambda_2 - 1)c_i^2 + 1]\delta_i^2, \tag{4.1.4}$$

or

$$\delta_i^2 = \frac{a_{ii}}{(\lambda_1 - 1)b_i^2 + (\lambda_2 - 1)c_i^2 + 1} = \frac{a_{ii}}{g_i^2}, \tag{4.1.5}$$

where g_i is defined by the last part of equation (4.1.5). The equation (4.1.3) can now be written in terms of the observed correlation matrix R instead of the dispersion matrix A

$$| GRG - \lambda I | = 0, \tag{4.1.6}$$

where the elements g_i of the diagonal matrix G satisfy

$$g_i = \sqrt{(\lambda_1 - 1)b_i^2 + (\lambda_2 - 1)c_i^2 + 1}. \tag{4.1.7}$$

The computational problem is then to solve for g_i's satisfying the equations (4.1.6) and (4.1.7), where λ_1 , λ_2 , are the latent roots of (4.1.6) and $\underline{b}$, $\underline{c}$, the latent vectors. A tentative method is to start with a trial matrix G and obtain successive approximations by solving (4.1.6) for λ_1 , λ_2 , and $\underline{b}$, $\underline{c}$, and substituting in (4.1.7). The process is repeated until the g_i converge.

A better approximation to g_i is obtained by using the formula

$$g_i = \sqrt{\left(\frac{\lambda_1}{\lambda_*} - 1\right)b_i^2 + \left(\frac{\lambda_2}{\lambda_*} - 1\right)c_i^2 + 1}, \tag{4.1.8}$$

where

$$\lambda_* = \frac{[\Sigma g_i^2]_0 - \lambda_1 - \lambda_2}{p - 2}, \tag{4.1.9}$$

the summation $[\Sigma g_i^2]_0$ refers to the g_i^2 at the previous stage used in equation (4.1.7) to obtain λ_1 , λ_2 .

The two formulas (4.1.7) and (4.1.8) should agree towards the final stages when convergence is expected to be slow. But in the initial stages (4.1.8) may accelerate convergence.

The estimated factor loadings on the first and second factors at any stage of approximation are

$$\sqrt{\lambda_1 - 1}\, bG^{-1}, \qquad \sqrt{\lambda_2 - 1}\, cG^{-1}.$$

The same method holds good for any number of factors. The estimates of factor loadings obtained from equations (4.1.6, 4.1.7) can be shown to satisfy the maximum-likelihood equations of Lawley (**11**, **12**) and thus constitute one out of a number of possible solutions. The equations (4.1.6, 4.1.7) are in a proper shape to admit an iterative procedure for solution. The use of equation (4.1.7) seems to avoid a difficulty which may occur in the iterative procedures. The iterative method given by Lawley (**16**, p. 130) may suffer a breakdown on the initial iteration due to an improperly chosen trial set of factor loadings leading to imaginary values of the quantities commonly designated by "h_1 , h_2 , $\cdots$." This may also occur in PFA with the guessed communalities at the first stage.

4.2 *Tests of Significance and Estimation of Number of Factors*

It is also necessary to lay down some rules for determining the number of factors to be estimated. This is partially answered by any reasonable test for a specified number of factors. We determine that number of factors for which the chosen test does not show significance, while for any smaller number the hypothesis is contradicted. If the level of significance is based on the 5 per cent level, then this method leads us to a lower confidence limit to the number of factors. That is, we can assert, with a risk of only 5 per cent, that the number of factors is at least as large as that discovered by the above

procedure. This is no doubt an objective rule for determining the lower limit to the number of factors, but in practice it may be better to extract one or two more factors, depending on the magnitude of the residual roots. If one or two such roots are sufficiently bigger than unity (though not significantly so) it may be worth while to extract the factor corresponding to them also.

The hypothesis we propose to test is that the population dispersion matrix admits the decomposition

$$\Sigma = \Gamma + \Delta, \tag{4.2.1}$$

where Δ is a diagonal matrix with positive terms and Γ is a positive semi-definite matrix of rank $k < p$.

The test criterion we use is derived by the principle of likelihood ratio, assuming that the observations are normally distributed.

The exact distribution of the test criterion is not known but in large samples (-2 log) of the likelihood ratio is distributed as χ^2 with degrees of freedom equal to the number of independent restrictions on the elements of Σ imposed by the hypothesis (4.2.1). This hypothesis specifies the rank of the matrix $\Sigma - \Delta$ for suitably chosen Δ. If its rank is k, then by fixing the first k rows and columns the rest of the elements can be computed, which implies $(p - k)\ (p - k + 1)/2$ restrictions. Allowing for p unknown values in Δ, the number of restrictions is equal to

$$\frac{(p-k)(p-k+1)}{2} - p = \frac{(p-k)^2 - p - k}{2}. \tag{4.2.2}$$

The test based on the likelihood-ratio criterion is

$$-(n-1)\log\frac{|A|}{|\hat{\Sigma}|}, \tag{4.2.3}$$

where $\hat{\Sigma}$ is the estimated dispersion matrix using the maximum-likelihood equations of section 4.1. The multiplying coefficient $(n - 1)$, where n is the sample size, may be replaced by the more appropriate value for the χ^2 approximation to hold when n is not large,

$$\left(n - 1 - \frac{2p+5}{6} - \frac{2k}{3}\right),$$

where p is the number of variables and k is the number of factors (1). Since the roots of the equation $|\Sigma - \lambda\Delta| = 0$ corresponding to k factors are estimated by $|A - \lambda\Delta| = 0$ or the equivalent forms (4.1.3), (4.1.6), it follows that the roots of $|\hat{\Sigma} - \lambda\Delta| = 0$ are

$$\lambda_1, \cdots, \lambda_k, 1, 1, \cdots, 1, \tag{4.2.4}$$

while the roots of $|A - \lambda\Delta|$ are, in descending order of magnitude,

$$\lambda_1, \cdots, \lambda_k, \lambda_{k+1}, \cdots, \lambda_p, \tag{4.2.5}$$

and, therefore,

$$\frac{|A|}{|\hat{\Sigma}|} = \lambda_{k+1} \cdots \lambda_p , \tag{4.2.6}$$

which is the product of the least $(p - k)$ roots, at the last stage of iteration of the equation (4.1.6), $|GRG - \lambda I| = 0$.

The χ^2 test is

$$-(n - 1) \log (\lambda_{k+1} \cdots \lambda_p) \tag{4.2.7}$$

with $[(p - k)^2 - p - k]/2$ degrees of freedom apart from the slight refinement in the multiplying coefficient.

4.3 *A Modified Criterion and Its Practical Use*

It may be recalled that the likelihood-ratio criterion is the ratio of the maximum likelihood under the restrictions of the hypothesis (4.2.1) to that without any restrictions on Σ. It is of interest to examine how the ratio (or its logarithm) of the likelihoods is converging to the maximum value (or the negative of log ratio to its minimum value) during the iterative process. Fortunately this can be expressed in terms of the roots λ_{k+1} , $\cdots$, λ_p at any stage of the iterative process

$$-(n - 1)[\log (\lambda_{k+1} \cdots \lambda_p) - (p - k) \log \lambda_a], \tag{4.3.1}$$

where λ_a of (4.1.9) is the arithmetic mean of λ_{k+1} , $\cdots$, λ_p . [Strangely the sequence (4.3.1) of statistics (likelihood ratios), which converge ultimately to the test criterion (maximum-liklihood ratio), resembles Bartlett's (1) ratio test but, of course, the roots λ_i are obtained differently and the ratios are used with different degrees of freedom. From this analysis it would appear that Bartlett's ratio is an initial approximation to the actual test criterion.] (4.3.1) converges to

$$-(n - 1) \log (\lambda_{k+1} \cdots \lambda_p)$$

at the final stage when

$$(p - k) = \lambda_{k+1} + \cdots + \lambda_p . \tag{4.3.2}$$

Suppose that (4.3.1) is not significant as χ^2 with $[(p - k)^2 - p - k]/2$ degrees of freedom, at any stage, then the same conclusion is reached even after completing the iterative process. If a test of significance is the only aim of analysis, then, sometimes, iteration can be stopped at some stage. Even if the result is significant, it is possible to terminate the computations provided the change in (4.3.1) at that stage is small from one cycle of operations to another.

The modified criterion is extremely useful in practice when the object

of the analysis is to estimate the number of factors (lower confidence value) as well as the factor loadings.

Before proceeding with the cycle of operations for estimation, let us fix some high value of k as the number of factors and calculate the roots after one or two iterations. At this stage, find that value of r for which Λ_r (with d.f. $[(p-r)^2 - p - r]/2$) is not significant, but Λ_{r-1} is. This shows that the number of factors is not greater than r. We may set the number of factors provisionally at r and continue the process of estimation. Each time we may calculate Λ_{r-1} and Λ_r to see whether Λ_{r-1} becomes not significant at any stage. If it is not significant, there is a case for switching over to $(r-1)$ factors instead of r.

5. *Summary*

The experimental situation and the nature of the data on which the technique of factor analysis can be successfully employed may be stated as follows. Each of the p measurements on an individual has a linear regression on a common set of a few hypothetical variables or factors. The deviations from regression for any two measurements are uncorrelated. The factor analysis seeks the smallest number of independent hypothetical variables necessary to explain the intercorrelations between the measurements.

If R is the observed correlation matrix, the computational problem of factor analysis depends on the solution of the diagonal matrix G satisfying the equations

$$| GRG - \lambda I | = 0, \tag{5.1}$$

$$g_i = [(\lambda_1 - 1)a_{1i}^2 + \cdots + (\lambda_k - 1)a_{ki}^2 + 1]^{1/2}, \tag{5.2}$$

where k is the number of factors assumed, $\lambda_1, \cdots, \lambda_k$, are the first k largest roots of (5.1) and $a_i = (a_{i1}, \cdots, a_{ip})$ is the latent vector corresponding to the root λ_i. Once G is found to satisfy the equations (5.1, 5.2), then the factor loadings are given by

$$(\lambda_j - 1)a_j G^{-1} \qquad (j = 1, \cdots, k),$$

and the test of the hypothesis that k factors are adequate to explain the intercorrelations is

$$\chi^2 = -(n-1) \log_e (\lambda_{k+1} \cdots \lambda_p)$$

with $[(p-k)^2 - p - k]/2$ degrees of freedom. The lower confidence limit to the number of factors is the smallest value of k for which χ^2 is not significant.

Some research remains to be done to find an elegant computational technique for solving the equations (5.1, 5.2). The method available at present is to guess suitable values of g_i, substitute in (5.1) and obtain better

approximations to g_i by using (5.2). This process is continued until convergence is secured. Unfortunately this appears to be a slow process unless the initial values of g_i are very near the true values. Even with a good set of trial values the problem can be best tackled only on an electronic computer when large numbers of variables are involved. A suitable program for Illiac is being written by Mr. Golub of the Digital Computer Laboratory at the University of Illinois. A numerical example solved on a tentative program is reported below. Full details will be presented soon.

First it may be noted that the relation between g_i and the communality h_i^2 for the ith variate is

$$g_i = 1/\sqrt{1 - h_i^2},$$

so that good trial values of g_i are available once the communalities are approximately determined by an initial factorization of the correlation matrix by a simpler method, such as the centroid. Another method suggested in the literature is to choose the squared multiple correlation as an estimate of the communality. In many cases it is sufficient to start with the initial approximation $g_i = 1/2$.

Second, although the test involves the product of the roots at the final stages of convergence, it is useful to compute at intermediate stages the statistic

$$\chi^2 = -(n - 1)[\log_e (\lambda_{k+1} \cdots \lambda_p) - (p - k) \log_e (\lambda_{k+1} + \cdots + \lambda_p)],$$

which, when not significant, implies the nonsignificance of the ultimate χ^2. We could stop at any stage after this, provided further iterations do not considerably alter the factor loadings.

The following correlation matrix was presented by Davis (6) in an attempt to study factors of comprehension in reading.

1.00								
.72	1.00							
.41	.34	1.00						
.28	.36	.16	1.00					
.52	.53	.34	.30	1.00				
.71	.71	.43	.36	.64	1.00			
.68	.68	.42	.35	.55	.76	1.00		
.51	.52	.28	.29	.45	.57	.59	1.00	
.68	.68	.41	.36	.55	.76	.68	.58	1.00

Assuming a single factor, the χ^2 was calculated and found to be significant. This indicated more than one factor. Under the hypothesis of two factors the value of χ^2

$$-(n - 1)[\log (\lambda_3 \cdots \lambda_9) - (9 - 2) \log (\lambda_3 + \cdots + \lambda_9)]$$

came down to 29.73 at an early stage of iteration. This being less than 30.1, the 5 per cent significance value of χ^2 with 19 degrees of freedom, the hypothesis of two factors stands unrejected. So the data admit an interpretation in terms of two significant factors only. A fairly stablilized set of factor loadings are

Factor 1	.845	.817	.477	.401	.669	.891	.834	.651	.833
Factor 2	−.309	−.084	.012	.153	.161	.145	.081	.122	.080

I wish to thank Dr. C. F. Wrigley, who read the manuscript and offered some helpful comments.

REFERENCES

1. Bartlett, M. S. Tests of significance in factor analysis. *Brit. J. Psychol., Statist. Sect.*, 1950, **3**, 77–85.
2. Bartlett, M. S. A further note on tests of significance in factor analysis. *Brit. J. Psychol., Statist. Sect.*, 1951, **4**, 1–2.
3. Burt, C. Tests of significance in factor analysis. *Brit. J. Psychol., Statist. Sect.*, 1952, **5**, 109–133.
4. Cattell, R. B. Parallel proportional profiles. *Psychometrika*, 1944, **9**, 267–283.
5. Cattell, R. B. The description and measurement of personality. Yonkers, New York: World Book Co., 1946.
6. Davis, F. B. Fundamental factors of comprehension in reading. *Psychometrika*, 1944, **9**, 185.
7. Holzinger, K. J., and Harman, H. H. Factor analysis. Chicago: Univ. Chicago Press, 1941.
8. Hotelling, H. Analysis of a complex of variables into principal components. *J. educ. Psychol.*, 1933, **24**, 417–441, 498–520.
9. Hotelling, H. Relations between two sets of variates. *Biometrika*, 1936, **28**, 321–377.
10. Kendall, M. G. Factor analysis. *J. roy. stat. Soc., Series B*, 1950, **12**, 60.
11. Lawley, D. N. The estimation of factor loadings by the method of maximum likelihood. *Proc. roy. Soc. Edin.*, 1940, **60**, 64–82.
12. Lawley, D. N. Further investigations in factor estimation. *Proc. roy. Soc. Edin.*, 1941, **61**, 176–185.
13. Neuhaus, J. O., and Wrigley, C. F. The quadrimax method: an analytic approach to orthogonal simple structure. Manuscript on file in the Univ. Illinois Library, 1953.
14. Rao, C. R. Advanced statistical methods in biometric research. New York: Wiley, 1952.
15. Rao, C. R. Discriminant functions for genetic differentiation and selection. Sankhyā, 1953, **12**, 229.
16. Thomson, G. H. The factorial analysis of human ability. London: Univ. London Press 5th ed., 1951.
17. Thurstone, L. L. A new rotational method in factor analysis. *Psychometrika*, 1938, **3**, 199–218.

Manuscript received 2/19/54

Revised manuscript received 5/11/54