BASIC IDEAS OF FACTOR AND COMPONENT ANALYSIS

Karl G. Jöreskog

University of Uppsala
Department of Statistics
P.O. Box 513
S-751 20 UPPSALA
SWEDEN

1. Introduction

Factor analysis is the common term for a number of statistical techniques for the resolution of a set of variables in terms of a small number of hypothetical variables, called factors. Within the statistical framework factor analysis belongs to the field of multivariate analysis. In fact, as we shall see, factor analysis can be formulated in terms of partial correlations.

Though the models and methods of factor analysis are of a statistical nature, factor analysis has been developed mainly by psychologists, particularly for analyzing the observed scores of many individuals on a number of psychological tests such as aptitude and achievements tests. The phenomenon continually observed in this situation is that the tests correlate (positively) with each other. Factor analysis attempts to "explain" these correlations by an analysis, which, when carried out successfully, yields underlying factors, smaller in number than the number of observed variables, that contain all the essential information about the linear inter-relationships among the test scores.

Factor analysis is most often employed in the behavioral sciences for studying the measurement properties (reliabilities and validities) of the instruments used in these sciences. But the techniques of factor analysis are not limited to behavioral sciences applications only. Factor analysis has been used in such diverse fields as meteorology, political science, medicine, geography and business. For a comprehensive

general description of the concepts, theories and techniques of factor analysis, see Harman (1967). For a statistical treatment of factor analsyis, see Lawley and Maxwell (1971).

This paper gives an elementary introduction to the basic ideas and concepts of factor analysis, developed by means of a simple example. To begin with we shall deal with population properties and ignore any sampling aspects that may arise. Thus we are concerned here with the explication of the meaning of the model rather than with the fitting of the model to empirical data.

Suppose six tests with test scores denoted by x_1, x_2, ..., x_6 have been administered to a large population of individuals. Let us assume that x_1, x_2 and x_6 are three different measures of verbal ability and that x_3, x_4 and x_5 are three different measures of numerical ability. For simplicity we assume that the tests are measured in standard scores so that each test has zero mean and unit standard deviation in the population. Furthermore suppose the following intercorrelations are obtained:

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	1.000	0.720	0.378	0.324	0.270	0.270
x_2	0.720	1.000	0.336	0.288	0.240	0.240
x_3	0.378	0.336	1.000	0.420	0.350	0.126
x_4	0.324	0.288	0.420	1.000	0.300	0.108
x_5	0.270	0.240	0.350	0.300	1.000	0.090
x_6	0.270	0.240	0.126	0.108	0.090	1.000

A typical correlation in this symmetric matrix is denoted by ρ_{ij}.

(The symbol ρ here indicates a population correlation as distinct from a sample correlation denoted by r .) For example $\rho_{32} = \rho_{23} = 0.336$ is the correlation between x_2 and x_3 .

2. Factor Analysis

Factor analysis sets out to explain these correlations by introducing underlying factors f_1, f_2, ... that account for the correlations. This is done in the following way. One first asks the question: Is there a factor f_1 such that if this is partialed out there remains no intercorrelations between the tests? If so the partial correlation between any pair of tests x_i and x_j after f_1 has been eliminated must vanish, i.e.,

$$\rho(x_i, x_j; f_1) = 0 , \quad i \neq j . \tag{1}$$

This is the same as to say that there exists numbers λ_1, λ_2, ..., λ_6 and residual variables e_1, e_2, ..., e_6 such that

$$\left.\begin{aligned} x_1 &= \lambda_1 f_1 + e_1 \\ x_2 &= \lambda_2 f_1 + e_2 \\ &\cdots\cdots\cdots \\ &\cdots\cdots\cdots \\ x_6 &= \lambda_6 f_1 + e_6 \end{aligned}\right\} \tag{2}$$

and such that the correlation between any pair of residuals e_i and e_j is zero, i.e.,

$$\rho(e_i, e_j) = 0 \ . \qquad (3)$$

Each equation in (2) represents the linear least squares regression of a test score on the factor f_1. Since all x-variables have been assumed to have zero means we omit the constant terms in the regressions. A well known property of least squares regression is that the residual e_i in the i:th equation is uncorrelated with f_1 [for a proof, see Cramér (1945), section 23.3]. Without loss of generality we may assume that f_1 is scaled so that its mean is zero and its variance is one. Using these results it is easily verified that

$$\rho_{ij} = \lambda_i \lambda_j \ , \quad i \neq j \ . \qquad (4)$$

Hence the correlations in any pair of rows or columns of the correlation matrix are proportional. This does not hold, however, for the given correlation matrix. For example, $\rho_{51}/\rho_{61} = 1$ but $\rho_{53}/\rho_{63} = 0.350/0.126 \neq 1$. We therefore conclude that no such factor f_1 exists.

It is then natural to try to explain the given correlations by means of two underlying factors instead of one. One then asks, Are there two factors f_1 and f_2 such that, when these are partialed out, there remains no intercorrelations between the tests? If so, the partial correlation between x_i and x_j, given f_1 and f_2 is zero:

$$\rho(x_i, x_j; f_1, f_2) = 0 \ , \quad i \neq j \qquad (5)$$

and the linear least squares regressions of the tests on these factors may be written as

$$\left.\begin{aligned} x_1 &= \lambda_{11}f_1 + \lambda_{12}f_2 + e_1 \\ x_2 &= \lambda_{21}f_1 + \lambda_{22}f_2 + e_2 \\ &\cdots\cdots\cdots\cdots \\ &\cdots\cdots\cdots\cdots \\ &\cdots\cdots\cdots\cdots \\ x_6 &= \lambda_{61}f_1 + \lambda_{62}f_2 + e_6 \end{aligned}\right\} , \qquad (6)$$

where the λ's are the regression coefficients and the e's are residuals in each equation that are uncorrelated with both f_1 and f_2. As before, we may, without loss of generality, assume that $E(f_1) = E(f_2) = 0$ and $E(f_1^2) = E(f_2^2) = 1$, where $E(x)$ is the expected value or mean of a random variable x. The condition of <u>partial linear independence</u> expressed in (5) is now equivalent to

$$\rho(e_i, e_j) = 0 , \quad i \neq j \quad . \qquad (7)$$

If f_1 and f_2 are uncorrelated we have, using these results,

$$\begin{aligned} \rho(x_i, x_j) &= E(x_i x_j) \\ &= E[(\lambda_{i1}f_1 + \lambda_{i2}f_2 + e_i)(\lambda_{j1}f_1 + \lambda_{j2}f_2 + e_j)] \\ &= \lambda_{i1}\lambda_{j1}E(f_1^2) + \lambda_{i2}\lambda_{j2}E(f_2^2) + \lambda_{i1}\lambda_{j2}E(f_1f_2) + \lambda_{i2}\lambda_{j1}E(f_ \\ &= \lambda_{i1}\lambda_{j1} + \lambda_{i2}\lambda_{j2} \cdot \qquad (8) \end{aligned}$$

In our example the following coefficient matrix $\underset{\sim}{\Lambda} = (\lambda_{im})$ satisfies this equation for all pairs of i and j :

$$\underset{\sim}{\Lambda} = \begin{bmatrix} 0.889 & -0.138 \\ 0.791 & -0.122 \\ 0.501 & 0.489 \\ 0.429 & 0.419 \\ 0.358 & 0.349 \\ 0.296 & -0.046 \end{bmatrix}$$

This matrix of regression coefficients has actually been obtained by the method of maximum likelihood, but this should not be of any concern now. At this point, it is sufficient to verify that this matrix does indeed reproduce the given correlations, as predicted in equation (8). For example, we may verify that

$$\begin{aligned} \lambda_{11}\lambda_{21} + \lambda_{12}\lambda_{22} &= 0.889 \times 0.791 + 0.138 \times 0.122 \\ &= 0.7032 + 0.0168 \\ &= 0.7200 \\ &= \rho_{12} \, . \end{aligned}$$

Thus we have established the fact that, for the particular correlation matrix at hand, there exists two uncorrelated factors f_1 and f_2 such that the representation

$$\left.\begin{aligned} x_1 &= 0.889\ f_1 - 0.138\ f_2 + e_1 \\ x_2 &= 0.791\ f_1 - 0.122\ f_2 + e_2 \\ x_3 &= 0.501\ f_1 + 0.489\ f_2 + e_3 \\ x_4 &= 0.429\ f_1 + 0.419\ f_2 + e_4 \\ x_5 &= 0.358\ f_1 + 0.349\ f_2 + e_5 \\ x_6 &= 0.296\ f_1 - 0.046\ f_2 + e_6 \end{aligned}\right\} \qquad (9)$$

holds with the condition (7) fulfilled for all pairs i and j. The factors f_1 and f_2 are not, however, the only factors that satisfy these requirements. For example, the following two factors also do it

$$\left.\begin{aligned} f_1^* &= 0.988\ f_1 - 0.153\ f_2 \\ f_2^* &= 0.153\ f_1 + 0.988\ f_2 \end{aligned}\right\} . \qquad (10)$$

It is easily verified that $E(f_1^*) = E(f_2^*) = 0$ and $\text{Var}\ (f_1^*) = \text{Var}\ (f_2^*) = 1$ and that f_1^* and f_2^* are uncorrelated if f_1 and f_2 have these properties. Solving (10) for f_1 and f_2 in terms of f_1^* and f_2^* we get

$$\left.\begin{aligned} f_1 &= 0.988\ f_1^* + 0.153\ f_2^* \\ f_2 &= -0.153\ f_1^* + 0.988\ f_2^* \end{aligned}\right\} . \qquad (11)$$

Substituting (11) into (9), we get the regression of each test on the factors f_1^* and f_2^*, as

$$\left.\begin{aligned} x_1 &= 0.90\ f_1^* + e_1 \\ x_2 &= 0.80\ f_1^* + e_2 \\ x_3 &= 0.42\ f_1^* + 0.56\ f_2^* + e_3 \\ x_4 &= 0.36\ f_1^* + 0.48\ f_2^* + e_4 \\ x_5 &= 0.30\ f_1^* + 0.40\ f_2^* + e_5 \\ x_6 &= 0.30\ f_1^* + e_6 \end{aligned}\right\} . \qquad (12)$$

The matrix of regression coefficients is now

$$\underset{\sim}{\Lambda}^{*} = \begin{bmatrix} 0.90 & 0 \\ 0.80 & 0 \\ 0.42 & 0.56 \\ 0.36 & 0.48 \\ 0.30 & 0.40 \\ 0.30 & 0 \end{bmatrix} .$$

The difference between $\underset{\sim}{\Lambda}^{*}$ and $\underset{\sim}{\Lambda}$ is that $\underset{\sim}{\Lambda}^{*}$ exhibits a kind of simplicity in the sense that some elements are zero. Factor analysts want to find factors such that the regression matrix $\underset{\sim}{\Lambda}$ is as simple as possible in the sense that it contains as many zero elements as possible. In general, even greater simplicity may be obtained by allowing the factors to become correlated. For example, the following factors

$$\left.\begin{aligned} f_1^{**} &= f_1^{*} \\ f_2^{**} &= 0.6\, f_1^{*} + 0.8\, f_2^{*} \end{aligned}\right\} , \tag{13}$$

with $\rho(f_1^{**}, f_2^{**}) = 0.6$, yields

$$\underset{\sim}{\Lambda}^{**} = \begin{bmatrix} 0.9 & 0 \\ 0.8 & 0 \\ 0 & 0.7 \\ 0 & 0.6 \\ 0 & 0.5 \\ 0.3 & 0 \end{bmatrix}$$

as is easily verified by solving (13) for f_1^{*} and f_2^{*} and

substituting these into (12). Note that $\underset{\sim}{\Lambda}^{**}$ has three zeroes in each column whereas $\underset{\sim}{\Lambda}^{*}$ has only three zeroes in one column.

Geometrically, the transformation from one set of factors to another, as in (10) or (11), corresponds to a rotation of coordinate axes in Euclidean space. The two elements in each row of the matrix $\underset{\sim}{\Lambda}$ may be regarded as the coordinates of a point in two-dimensional Euclidean space and these points may be plotted in this space as in Fig. 1. The points in the space are fixed but the coordinate axes are arbitrary. The correlation between the two factors is equal to the cosine of the angle between the two coordinate axes. If the axes are orthogonal the factors are uncorrelated. In Fig. 1, the axes labeled λ_{i1} and λ_{i2} correspond to the matrix $\underset{\sim}{\Lambda}$. If the axis λ_{i1} is rotated clockwise so that it passes through the points (1), (2) and (6) while maintaining the orthogonality of the axes, we get the coordinate system λ^{*}_{i1} and λ^{*}_{i2} corresponding to the matrix $\underset{\sim}{\Lambda}^{*}$. If the first axis is kept at λ^{*}_{i1} and the second axis rotated clockwise so that it passes through the points (3), (4) and (5), we obtain the coordinate system λ^{**}_{i1} and λ^{**}_{i2}, in which the cosine of the angle between the coordinate axes is 0.60. This corresponds to the matrix $\underset{\sim}{\Lambda}^{**}$ and the correlation $\rho(f_1^{**}, f_2^{**}) = 0.6$. Note that zero elements in $\underset{\sim}{\Lambda}^{*}$ or $\underset{\sim}{\Lambda}^{**}$ correspond to points in the space lying on one coordinate axis.

We have demonstrated the two basic principles of factor analysis. The first one is <u>the principle of conditional linear</u> independence. This principle expresses the idea that the factors shall account for all linear relationships among the variables.

Once the factors have been partialed out there shall remain no correlation between the variables. In this sense factor analysis is a method for classification of linear dependence. In general, for p variables $x_1, x_2, \ldots, x_p$, once it has been established that correlation exists, we may ask whether we can find one factor f_1 such that

$$\rho(x_i, x_j; f_1) = 0 , \quad i \neq j ,$$

for all pairs i and j. If the answer is yes we say that $x_1, x_2, \ldots, x_p$ has linear dependence of degree 1. If, on the other hand, the answer is no we ask for two factors f_1 and f_2 such that

$$\rho(x_i, x_j; f_1, f_2) = 0 , \quad i \neq j .$$

If such factors can be found we say that $x_1, x_2, \ldots, x_p$ have linear dependence of degree 2, etc. It may be shown that this process always ends after having found some $k < p$ factors $f_1, f_2, \ldots, f_k$ such that

$$\rho(x_i, x_j; f_1, f_2, \ldots, f_k) = 0 , \quad i \neq j , \qquad (14)$$

in which case we have linear dependence of degree k among the tests.

Equation (14) is equivalent to the linear factor analytic model

$$\left.\begin{aligned} x_1 &= \lambda_{11}f_1 + \lambda_{12}f_2 + \ldots + \lambda_{1k}f_k + e_1 \\ x_2 &= \lambda_{21}f_1 + \lambda_{22}f_2 + \ldots + \lambda_{2k}f_k + e_2 \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ x_p &= \lambda_{p1}f_1 + \lambda_{p2}f_2 + \ldots + \lambda_{pk}f_k + e_p \end{aligned}\right\} , \qquad (15)$$

where

$$\rho(e_i, e_j) = 0 , \quad i \neq j , \qquad (16)$$

for all pairs of i and j. Each equation in (15) represents the linear least squares regression of a variable on the factors $f_1, f_2, \ldots, f_k$. The coefficients λ_{im} are regression coefficients. In psychological terminology these are called factor loadings. The factors $f_1, f_2, \ldots, f_k$ are called common factors because they measure attributes that are common to two or more of the variables. The residuals $e_1, e_2, \ldots, e_p$ are called residuals or unique factors. They represent what is left over of each variable when the factors have been determined so as to account for all intercorrelations of the tests. A unique factor is usually regarded as a sum of an error of measurement and a specific factor, the specific factor having to do only with the the particular variable. None of these parts contribute to any correlation between variables. Each equation in (15) decompos the variable x_i into two uncorrelated parts,

$$x_i = c_i + e_i , \qquad (17)$$

where

$$c_i = \lambda_{i1} f_1 + \lambda_{i2} f_2 + \ldots + \lambda_{ik} f_k \qquad (18)$$

is called the common part of x_i .

Once a set of k factors has been found that account for the intercorrelations of the tests, as in (15), these may be transformed to another set of k factors that account equally well for the correlations. In fact, any non-singular linear transformation of the first set of factors yields a new set of factors with this property. For our particular data, (f_1, f_2) , (f_1^*, f_2^*) and (f_1^{**}, f_2^{**}) are three sets of factors, each one accounting for the correlations in the sense that (15) and (16) are fulfilled. The regression of the tests on the factors, as represented by the matrix of factor loadings, for two different sets of factors, may be quite different. Since there are infinitely many sets of factors this is a great indeterminacy in the model. However, Thurstone (1947) proposed giving attention only to factors for which the variables have a simple representation. The matrix of factor loadings shall have as many zero elements as possible. If $\lambda_{im} = 0$, the m:th factor does not enter into the i:th test. A variable should not depend on all common factors but only on a small portion of them. Also the same factor should be involved only in a small portion of the variables. Such a matrix is regarded as giving the simplest structure and presumably the one with the most meaningful psychological interpretation. This is the second basic principle of factor analysis, the principle

of simple structure. The factor analyst usually tries to meet these requirements when he chooses the tests to be included in the factorial study.

In our particular example, the matrix $\underset{\sim}{\Lambda}^{*}$ represents a simple structure with two uncorrelated factors and $\underset{\sim}{\Lambda}^{**}$ represents a simple structure with two correlated factors. Uncorrelated factors are also called orthogonal and correlated factors are called oblique. Usually simple structure is better achieved with oblique factors than with orthogonal. In our particular example the interpretation of the matrix $\underset{\sim}{\Lambda}^{*}$ might be that there is some general factor f_1^{*} that is required to perform well in all the six tests. In addition to this general factor f_1^{*} there is a second factor f_2^{*}, which has nothing to do with f_1^{*} but which is necessary to perform well on the numerical tests x_3, x_4 and x_5. This interpretation is not a very reasonable one. A better interpretation is based on $\underset{\sim}{\Lambda}^{**}$. Here the factors f_1^{**} and f_2^{**} are interpreted as a verbal and a numeric factor. These factors correlate 0.6 in the population of individuals examined. This means that individuals who perform well on verbal tests are likely to perform well on numerical tests too.

Consider the equation (17). Since the common part c_i and the unique part e_i are uncorrelated this equation also partitions the variance of x_i as

$$\text{Var}\ (x_i) = \text{Var}\ (c_i) + \text{Var}\ (e_i)\ . \qquad (19)$$

Var (c_i) is called the common variance or the communality of x_i and Var (e_i) is called the unique variance or the uniquene

of x_i. The communality of a variable is the portion of a variable's total variance that is accounted for by the common factors. The uniqueness is the portion left unexplained by the common factors.

To determine the communality of x_i we use equation (18) and get

$$E(c_i^2) = \lambda_{i1}^2 E(f_1^2) + \lambda_{i2}^2 E(f_2^2) + \ldots + \lambda_{ik}^2 E(f_k^2) +$$

$$+ 2\lambda_{i1}\lambda_{i2} E(f_1 f_2) + 2\lambda_{i1}\lambda_{i3} E(f_1 f_3) + \ldots +$$

$$+ 2\lambda_{i,k-1}\lambda_{ik} E(f_{k-1} f_k) \ .$$

If $f_1, f_2, \ldots, f_k$ all have unit variances this becomes

$$\text{Var } (c_i) = \lambda_{i1}^2 + \lambda_{i2}^2 + \ldots + \lambda_{ik}^2 +$$

$$+ 2\lambda_{i1}\lambda_{i2}\rho(f_1, f_2) + 2\lambda_{i1}\lambda_{i3}\rho(f_1, f_3) + \ldots +$$

$$+ 2\lambda_{i,k-1}\lambda_{jk}\rho(f_{k-1}, f_k) \ . \qquad (20)$$

If the factors are uncorrelated this reduces further to

$$\text{Var } (c_i) = \lambda_{i1}^2 + \lambda_{i2}^2 + \ldots + \lambda_{ik}^2 \ , \qquad (21)$$

which is simply the sum of squares of the elements in the i:th row of $\underset{\sim}{\Lambda}$.

Using (21) and the matrix $\underset{\sim}{\Lambda}^*$ in our example we get

Test	Communality	Uniqueness
1	0.81	0.19
2	0.64	0.36
3	0.49	0.51
4	0.36	0.64
5	0.25	0.75
6	0.09	0.91
Total	2.64	3.36

(22)

The same result is obtained if we apply the general formula (20) to the matrix $\underset{\sim}{\Lambda}^{**}$. This shows that communalities and uniquenesses are unaffected by linear transformation of factors such as (13). That this property is general is evident from the fact that the unique factors themselves $e_1, e_2, \ldots, e_p$ are unaffected by linear transformations of the common factors. Hence their variances, the uniquenesses, and their counterparts, the communalities, are invariant under such transformations.

3. Factor Analysis vs. Component Analysis

Factor analysis is often confused with principal component analysis. The two methods of analysis are similar to some exten but have entirely different aims. The distinction between facto and component analysis has been emphasized by Kendall and Lawley (1956), Lawley and Maxwell (1971) and others. It is advisable to keep this distinction clear.

The first principal component of $x_1, x_2, \ldots, x_p$ is defined as the linear combination of $x_1, x_2, \ldots, x_p$,

$$u_1 = \beta_{11}x_1 + \beta_{21}x_2 + \ldots + \beta_{p1}x_p$$

that has maximal variance, subject to the restriction that

$$\beta_{11}^2 + \beta_{21}^2 + \ldots + \beta_{p1}^2 = 1 \ .$$

The coefficients $\beta_{11}, \beta_{21}, \ldots, \beta_{p1}$ are determined as the elements of the latent vector of the covariance matrix of $x_1, x_2, \ldots, x_p$ which corresponds to the largest latent root θ_1 (see e.g. Anderson, 1958 , Chapter 11) . The variance of u_1 is θ_1 .

The second principal component of $x_1, x_2, \ldots, x_p$ is that linear combination

$$u_2 = \beta_{12}x_1 + \beta_{22}x_2 + \ldots + \beta_{p2}x_p \ ,$$

which have maximal variance and is uncorrelated with u_1 . The coefficients $\beta_{12}, \beta_{22}, \ldots, \beta_{p2}$ are the elements of the latent vector of the covariance matrix corresponding to the second largest latent root θ_2 . The variance of u_2 is θ_2 .

The third, fourth, ... principal components of $x_1, x_2, \ldots, x_p$ are similarly defined. There will be as many principal components as there are positive latent roots $\theta_1, \theta_2, \ldots$ of the variance-covariance matrix. Usually the covariance matrix is positive definite and then all p roots are positive, thus giving rise to p principal components

$$u_i = \beta_{1i}x_1 + \beta_{2i}x_2 + \ldots + \beta_{pi}x_p, \quad i = 1,2,\ldots,p \ . \quad (23$$

The matrix of coefficients $\underset{\sim}{B} = (\beta_{ij})$ form an orthogonal matrix (see e.g. Anderson, 1958 , Chapter 11) and equation (23) can therefore be inverted to give

$$x_i = \beta_{i1}u_1 + \beta_{i2}u_2 + \ldots + \beta_{ip}u_p \ , \quad i=1,2,\ldots,p. \quad (24)$$

The equation (24) expresses each variable x_i as a linear combination of p uncorrelated variables with descending variances.

In some applications of principal component analysis a few principal components account for a large portion of the total variance of all the variables. Equation (24) may then be cut off after a certain number of terms, $k < p$ say, so that

$$x_i = \beta_{i1}u_1 + \beta_{i2}u_2 + \ldots + \beta_{ik}u_k + v_i \ , \quad (25)$$

where

$$v_i = \beta_{i,k+1}u_{k+1} + \beta_{i,k+2}u_{k+2} + \ldots + \beta_{ip}u_p \quad (26)$$

is interpreted as a residual with small variance.

To bring out the formal similarity between factor and component analysis,we standardize the components $u_1, u_2,\ldots, u_k$ to unit variance by dividing them by $\sqrt{\theta_1}, \sqrt{\theta_2},\ldots, \sqrt{\theta_k}$, respectively. If we write

$$u_i^* = (1/\sqrt{\theta_i})u_i$$

and

$$\beta_{ij}^* = \beta_{ij}\sqrt{\theta_j} \ ,$$

equation (25) becomes

$$x_i = \beta_{i1}^*u_1^* + \beta_{i2}^*u_2^* + \ldots + \beta_{ik}^*u_k^* + v_i \ , \quad (27)$$

an equation that formally resembles the i:th equation in (15). However, the residuals $v_1, v_2,\ldots, v_p$ are not all uncorrelated

as in (16) . To show this, multiply (25) by β_{im} , sum over i from 1 to p and use (23) and the property that the matrix $\underset{\sim}{B}$ is orthogonal. We then obtain, for m=1,2,...,k,

$$\sum_{i=1}^{p} \beta_{im} v_i = \sum_{i=1}^{p} \beta_{im} x_i - \sum_{i=1}^{p} \sum_{n=1}^{k} \beta_{im} \beta_{in} u_n \qquad (28)$$

$$= u_m - u_m$$

$$= 0 \ .$$

The residuals $e_1, e_2, \ldots, e_p$, therefore, satisfy the k linear equations (28) and hence cannot all be uncorrelated.

Summing up we may say that factor analysis is correlation-oriented and principal component analysis is variance-oriented. Whereas factor analysis aims at reproducing the intercorrelations of the variables, principal component analysis aims at reproducing their total variance. Although, in principal component analysis, a few components may extract a large portion of the total variance, all components are required to reproduce the correlations exactly. In factor analysis, on the other hand, there are, by definition, a certain number of factors, fewer than the number of variables, that reproduce the intercorrelations exactly. These factors, however, do not account for as much variance as does the same number of principal components.

We may illustrate these remarks on the basis of the example considered in the previous section. The first two principal components (of unit variances) of $x_1, x_2, \ldots, x_6$

are

$$\left.\begin{aligned} u_1^* &= 0.32\,x_1 + 0.30\,x_2 + 0.26\,x_3 + 0.24\,x_4 + 0.21\,x_5 + 0.15\,x_6 \\ u_2^* &= -0.29x_1 - 0.32\,x_2 + 0.34\,x_3 + 0.38\,x_4 + 0.42\,x_5 - 0.60\,x_6 \end{aligned}\right\} \quad (29$$

The equations corresponding to (27) are

$$\left.\begin{aligned} x_1 &= 0.81\,u_1^* - 0.31\,u_2^* + v_1 \\ x_2 &= 0.78\,u_1^* - 0.34\,u_2^* + v_2 \\ x_3 &= 0.68\,u_1^* + 0.36\,u_2^* + v_3 \\ x_4 &= 0.62\,u_1^* + 0.40\,u_2^* + v_4 \\ x_5 &= 0.55\,u_1^* + 0.45\,u_2^* + v_5 \\ x_6 &= 0.39\,u_1^* - 0.60\,u_2^* + v_6 \end{aligned}\right\} \quad . \qquad (30)$$

It is readily verified that these representations do not reproduce the intercorrelations as do (9) . For example,

$$\begin{aligned} \beta_{i1}^*\beta_{j1}^* + \beta_{i2}^*\beta_{j2}^* &= 0.81 \times 0.78 + 0.31 \times 0.34 \\ &= 0.6318 + 0.1054 \\ &= 0.7372 \end{aligned}$$

which is not equal to ρ_{12} . The variance of each variable accounted for by the two principal components and the residual variances are

i	Var $(\beta^*_{i1}u^*_1 + \beta^*_{i2}u^*_2)$	Var (v_i)
1	0.75	0.25
2	0.72	0.28
3	0.59	0.41
4	0.55	0.45
5	0.51	0.49
6	0.51	0.49
Total	3.63	2.37

(31)

This is to be compared with the corresponding table (22). The residual variances in (31) are smaller than those of (22) except for x_1. The total variance accounted for by the two principal components is 3.63 compared with 2.64 for the two factors. This is a reflection of the specific factor included in each e_i but not in v_i.

In precisely the same way as in factor analysis the two principal components may be transformed linearly to another set of components that together account for as much variance as do the original principal components. In component analysis this may also be done to obtain some interesting interpretation. Prinicpal components do not usually have a meaningful interpretation except in terms of their variance and correlation properties. Another difference between factor analysis is that components are, by definition, linear combinations of the variables $x_1, x_2, \ldots, x_p$, whereas the common factors are not. Instead the factors are linear

combinations of the common parts $c_1, c_2, \ldots, c_p$ of the variables. For example, the factors f_1^{**} and f_2^{**} in (13) have the following representations

$$\left.\begin{aligned} f_1^{**} &= (0.9^2 + 0.8^2 + 0.3^2)^{-1}(0.9c_1 + 0.8c_2 + 0.3c_6) \\ f_2^{**} &= (0.7^2 + 0.6^2 + 0.5^2)^{-1}(0.7c_3 + 0.6c_4 + 0.5c_5) \end{aligned}\right\}, \qquad (32)$$

These are easily verified by substituting $c_1, c_2, \ldots, c_6$ from the equations corresponding to (18) into the right sides of (32) .

Principal component analysis is not a model in the usual sense. It is merely a descriptive method of analysis that can be used to analyze all kinds of quantitative variables. Factor analysis, on the other hand, postulates a certain model (15) , which is to be tested against empirical data. The equations (15) are not capable of direct verification, since the p variables x_i are expressed in terms of p+k other variables which are not observable, but the equations imply a hypothesis that can be tested, namely that the correlations ρ_{ij} of the x's are of the form

$$\rho_{ij} = \lambda_{i1}\lambda_{j1} + \lambda_{i2}\lambda_{j2} + \ldots + \lambda_{ik}\lambda_{jk} . \qquad (33)$$

If k is small, then (33) imposes restrictions on the ρ's . The smaller k is, the more restrictive the factor analysis model is.

4. Estimation

The basic model in factor analysis, as given by (15), can be written in matrix form as

$$\underset{\sim}{x} = \underset{\sim}{\Lambda}\underset{\sim}{f} + \underset{\sim}{e}, \tag{34}$$

where $\underset{\sim}{x}$ is a column vector of observations on p variables, $\underset{\sim}{f}$ is a column vector of k common factors, $\underset{\sim}{e}$ is a vector of p residuals, which represent the combined effects of specific factors and random measurement error, and $\underset{\sim}{\Lambda} = (\lambda_{im})$ is a pxk matrix of factor loadings.

The residuals in $\underset{\sim}{e}$ are assumed to be uncorrelated with each other and with the common factors $\underset{\sim}{f}$. All three vectors $\underset{\sim}{x}$, $\underset{\sim}{f}$ and $\underset{\sim}{e}$ are assumed to have zero mean vectors and their covariance matrices are denoted respectively by $\underset{\sim}{\Sigma}(pxp)$, $\underset{\sim}{\Phi}(kxk)$ and $\underset{\sim}{\Psi}^2(pxp)$. The matrix $\underset{\sim}{\Psi}^2$ is diagonal with elements ψ_{ii}^2, i=1,2,...,p, which are the residual or unique variances of the variables. Since each column of $\underset{\sim}{\Lambda}$ may be scaled arbitrarily, we may assume, without loss of generality, that the common factors have unit variances, so that the diagonal elements of $\underset{\sim}{\Phi}$ are unity. In addition, if for $k > 1$, the common factors are orthogonal or uncorrelated, then the off-diagonal elements of $\underset{\sim}{\Phi}$ are zeroes and the matrix $\underset{\sim}{\Phi}$ is an identity matrix. Otherwise, if the factors are correlated, $\underset{\sim}{\Phi}$ is the correlation matrix of the factors.

In view of equation (34) and the assumptions just made, the covariance matrix $\underset{\sim}{\Sigma}$, of the observed variables $\underset{\sim}{x}$, is

$$\begin{aligned}\underset{\sim}{\Sigma} &= E(\underset{\sim}{x}\underset{\sim}{x}') \\ &= E[(\underset{\sim}{\Lambda}\underset{\sim}{f} + \underset{\sim}{e})(\underset{\sim}{\Lambda}\underset{\sim}{f} + \underset{\sim}{e})'] \\ &= E(\underset{\sim}{\Lambda}\underset{\sim}{f}\underset{\sim}{f}'\underset{\sim}{\Lambda}) + E(\underset{\sim}{\Lambda}\underset{\sim}{f}\underset{\sim}{e}') + E(\underset{\sim}{e}\underset{\sim}{f}'\underset{\sim}{\Lambda}') + E(\underset{\sim}{e}\underset{\sim}{e}') \\ &= \underset{\sim}{\Lambda}\underset{\sim}{\Phi}\underset{\sim}{\Lambda}' + \underset{\sim}{\Psi}^2 , \end{aligned} \tag{35}$$

since the second and third terms are zero because $\underset{\sim}{f}$ and $\underset{\sim}{e}$ were assumed to be uncorrelated.

Equations (34) and (35) represent a model for a population of individuals. This population is characterized by the parameters $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}^2$. In practice, these parameters are unknown and must be estimated from data on N individuals. Let $x_{\alpha i}$ be the observed value of variable i for individual α. Then the available data may be written as a data matrix $\underset{\sim}{X}$ of order Nxp. From this we can compute the sample mean vector $\bar{\underset{\sim}{x}}' = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p)$ and the sample covariance matrix $\underset{\sim}{S} = (s_{ij})$, where

$$\bar{x}_i = (1/N) \sum_{\alpha=1}^{N} x_{\alpha i}, \tag{36}$$

$$s_{ij} = (1/n) \sum_{\alpha=1}^{N} (x_{\alpha i} - \bar{x}_i)(x_{\alpha j} - \bar{x}_j) , \tag{37}$$

with $n = N - 1$.

The information provided by $\underset{\sim}{S}$ may also be represented by a correlation matrix $\underset{\sim}{R} = (r_{ij})$ and a set of standard deviations $s_1, s_2, \ldots, s_p$ where $s_i = \sqrt{s_{ii}}$ and $r_{ij} = s_{ij}/s_i s$

In most applications both the origin and the unit of measurement in the observed variables $\underset{\sim}{x}$ are arbitrary or irrelevant, and then only the correlation matrix $\underset{\sim}{R}$ is of any

interest. In such cases one takes $\underset{\sim}{S}$ to be a correlation matrix $\underset{\sim}{R}$ in what follows. This is what we did in the previous sections.

In practical work with empirical data, the factor analysis model does not fit the data perfectly as it did in the artificial example used in the previous sections. The statistical problem then is how to fit the model matrix $\underset{\sim}{\Sigma}$ of the form (35) to a sample covariance matrix $\underset{\sim}{S}$. This assumes that the number of factors k is known or specified a priori. However, in most exploratory factor studies this is not the case. Instead the investigator wants to determine the smallest k for which the model fits the data. This is usually done by a sequential procedure testing increasing values of k until a sufficiently good fit has been obtained (see e.g. Lawley & Maxwell, 1971).

Several methods have been developed for estimating the parameters of the factor analysis model. Three different methods of fitting $\underset{\sim}{\Sigma}$ to $\underset{\sim}{S}$ will be considered here, namely the method of <u>unweighted least squares</u> (ULS), which minimizes the sum of squares of all the elements of $\underset{\sim}{S} - \underset{\sim}{\Sigma}$:

$$U = \mathrm{tr}(\underset{\sim}{S} - \underset{\sim}{\Sigma})^2 , \tag{38}$$

the method of <u>generalized least squares</u> (GLS), which minimizes the sum of squares of all the elements of $\underset{\sim}{I} - \underset{\sim}{S}^{-1}\underset{\sim}{\Sigma}$:

$$G = \mathrm{tr}(\underset{\sim}{I} - \underset{\sim}{S}^{-1}\underset{\sim}{\Sigma})^2 , \tag{39}$$

and the method of maximum likelihood (ML) , which minimizes

$$M = \text{tr}(\underset{\sim}{\Sigma}^{-1}\underset{\sim}{S}) - \log|\underset{\sim}{\Sigma}^{-1}\underset{\sim}{S}| - p \ , \qquad (40)$$

where $\text{tr}(\underset{\sim}{A})$ and $|\underset{\sim}{A}|$ denotes the trace and the determinant of $\underset{\sim}{A}$, respectively. The last method is equivalent to the maximization of the likelihood of the observations under multinormality of $\underset{\sim}{x}$, hence the name maximum likelihood method.

Each fitting function U, G and M is to be minimized with respect to $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$. Derivations and justifications of these methods are found in Anderson (1959), Jöreskog (1967), Lawley & Maxwell (1971) and Jöreskog & Goldberger (1972). All three functions may be minimized numerically by basically the same algorithm. For details of the minimization procedure, see Jöreskog (1977).

The GLS and ML methods are scale-free in the sense that analyses of the same variables in two different sets of scales are related by proper scale factors in the rows of $\underset{\sim}{\Lambda}$ and $\underset{\sim}{\Psi}$. This property does not hold for ULS. When $\underset{\sim}{x}$ has a multivariat normal distribution, both GLS and ML yield estimates that have good properties in large samples. Both GLS and ML require a positive definite matrix $\underset{\sim}{S}$, while ULS will work even on a matrix which is non-Gramian.

When $k > 1$, so that there is more than one common factor, it is necessary to remove an element of indeterminacy in the basic model before the procedure for minimizing the fitting function can be applied. As demonstrated in the previous section, this indeterminacy arises because there exists a non-singular linear transformation of the common factors which

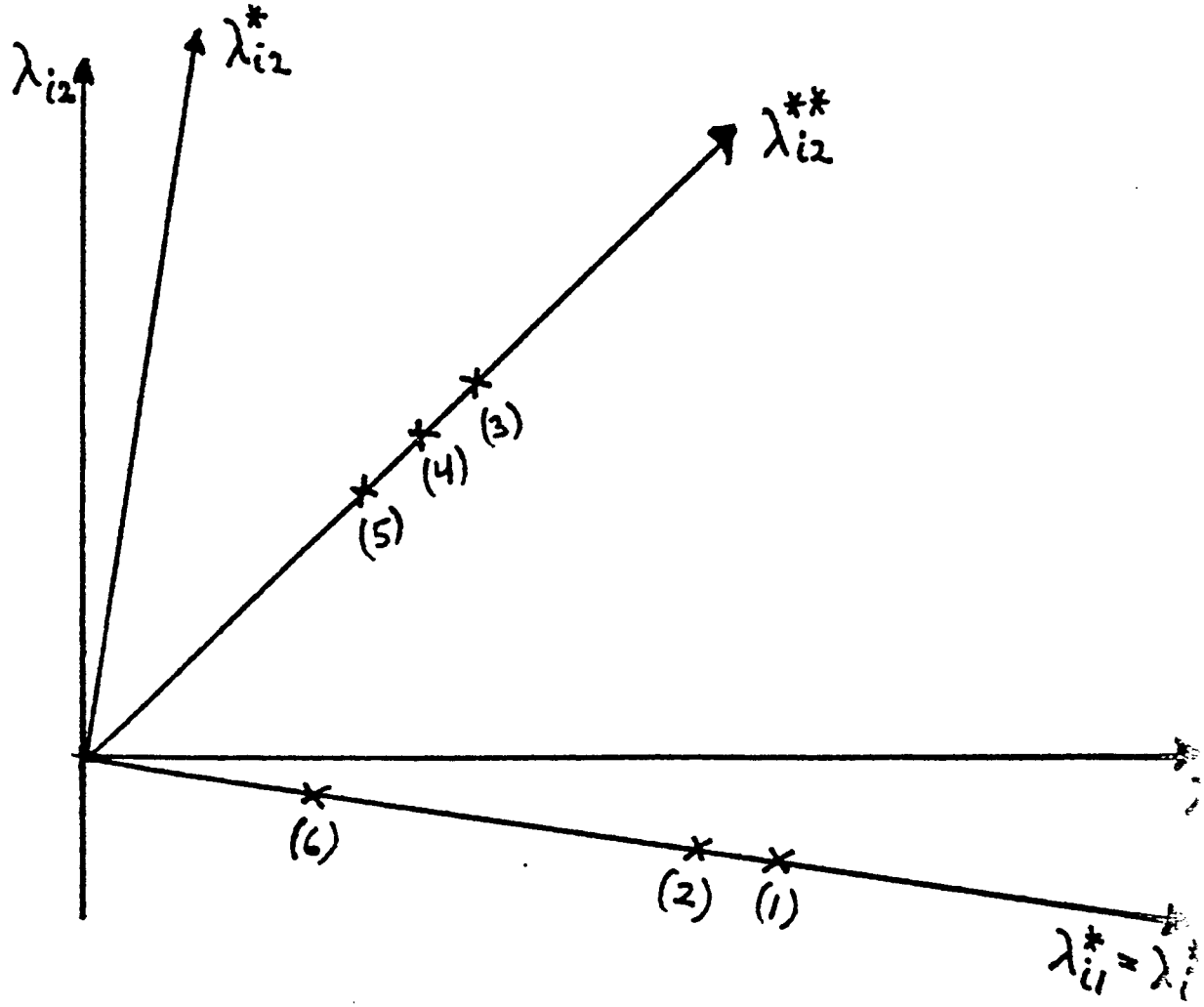

Fig. 1. Plots of factor 2 vs factor 1 in three coordinate systems

changes $\underset{\sim}{\Lambda}$ and in general also $\underset{\sim}{\Phi}$, but leaves $\underset{\sim}{\Sigma}$ and therefore the function unaltered. Hence, to obtain a unique set of parameters and a corresponding unique set of estimates, we must impose some additional conditions. These have the effect of selecting a particular set of factors and thus of defining the parameters uniquely.

The usual way to choose the conditions in exploratory factor analysis is to choose $\underset{\sim}{\Phi} = \underset{\sim}{I}$, $\underset{\sim}{\Lambda}'\underset{\sim}{\Lambda}$ to be diagonal in ULS and $\underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-2}\underset{\sim}{\Lambda}$ to be diagonal in GLS and ML and to estimate $\underset{\sim}{\Lambda}$ and $\underset{\sim}{\Psi}$ subject to these conditions. This leads to an arbitrary set of factors that may then be subjected to a rotation to another set of factors to facilitate a more meaningful interpretation. The rotation is guided by the principle of simple structure. Techniques for rotation of factors are given by Harman (1967).

REFERENCES

Anderson, T.W. (1958): An introduction to multivariate statistical analysis. New York, Wiley.

Anderson, T.W. (1959): Some scaling models and estimation procedures in the latent class model. Probability and Statistics, The Harald Cramér Volume, U. Grenander (ed.), New York, Wiley, 9-38.

Cramér, H. (1957): Mathematical methods of statistics. Princeton, Princeton University Press.

Harman, H.H. (1967): Modern Factor Analysis, Second Edition. Chicago, University of Chicago Press.

Jöreskog, K.G. (1967): Some contributions to maximum likelihood factor analysis. Psychometrika, 32, 443-482.

Jöreskog, K.G. (1977): Factor analysis by least squares and maximum likelihood methods. In K. Enslein, A. Ralst & H.S. Wilf (eds.): Statistical Methods for Digital Computers. New York, Wiley, 125-153.

Jöreskog, K.G. & Goldberger, A.S. (1972): Factor analysis by generalized least squares. Psychometrika, 37, 243-259.

Kendall, M.G. & Lawley, D.N. (1956): The principles of factor analysis. Journal of the Royal Statistical Society, Series A, 119, 83-84.

Lawley, D.N. & Maxwell, A.E. (1971): Factor analysis as a statistical method. Second Edition. London, Butterworth & Co publishers.

Thurstone, L.L. (1947): Multiple factor analysis. Chicago, University of Chicago Press.