Factor Analysis

Christof Schuster
University of Notre Dame, Indiana, USA

Ke-Hai Yuan
University of Notre Dame, Indiana, USA

Glossary

communality Variance of a variable accounted for by the factors.
factor Latent variable that determines to a considerable extent the values of the observed variables.
factor loading Weight indicating the direct influence of a factor on a variable.
factor structure Matrix of covariances between factors and variables. If the variables are standardized, then the factor structure matrix contains correlations rather than covariances.
oblique rotation Factor transformation designed to achieve a simple interpretation of the factors.
orthogonal rotation Factor transformation designed to achieve a simple interpretation of mutually uncorrelated factors.
uniqueness Amount of variable variance unrelated to factors.

Factor analysis is a multivariate statistical technique for finding theoretical concepts that underlie the association between observed variables. To this end, the factor model introduces latent variables, commonly referred to as factors, and posits that the observed variables are determined, except for random error, by these factors. Factor analysis is particularly well-suited to psychology, wherein concepts such as "intelligence" can be observed only indirectly—for instance, by noticing which test problems an individual is able to solve correctly.

Introduction

Spearman noted in 1904 that measures of performance for certain cognitive tasks are often positively correlated. If an individual performs well on, say, an intelligence-related task, he or she also tends to do well on similar tasks. To explain this phenomenon, Spearman introduced a general ability factor, commonly referred to as the g-factor, that he claimed determines performance of intelligence-related tasks to a considerable extent, albeit not completely. Thus, positive correlations between intelligence-related tasks were explained by Spearman in terms of the g-factor's influence on each of these tasks. During the first half of the 20th century, factor analysis developed rapidly from a substantive theory on intelligence into a general statistical procedure. An early comprehensive account of factor analysis was given in 1947 by Thurstone.

The main idea of factor analysis is to "explain" correlations between observed variables in terms of a few unobservable variables, commonly called factors or latent variables. It will be convenient to refer to the observable and unobservable variables simply as variables and factors, respectively. More specifically, the model states that the factors determine the variables in such a way that when holding the factors constant, the residual variation of the variables is uncorrelated. If the number of factors needed to explain the correlations is "small," compared to the number of variables, then the factor model is appealing because the associations between the variables can be explained in a parsimonious manner.

Factor analysis is used if latent variables are assumed to underlie the association among the variables, but the conceptual nature of the factors is unclear before fitting the factor model to data. Because the purpose of factor analysis is to clarify the nature of the factors, one also speaks of exploratory factor analysis as opposed to confirmatory factor analysis, a closely related statistical technique that requires specific *a priori* assumptions about the factors and their relationships to the variables. Although in practice the factor model is almost always fitted to the sample correlation matrix, it is common to discuss the model in terms of the sample covariance matrix. This convention is followed here. Finally, note that factor analysis is different from principal component analysis. Although these two statistical techniques are frequently confused, they address different issues. While factor analysis explains observed correlations in terms of latent factors, principal component analysis is a data reduction technique that yields linear composites of the observed variables referred to as "components."

The Factor Model

Consider p variables, each having zero expectation, that is, $\mathrm{E}(x_i) = 0$, $i = 1, \ldots, p$. The factor model states that the conditional expectation of each variables is a linear function of q factors, $\xi_1, \ldots, \xi_q$. In other words, the factor model claims

$$\mathrm{E}(x_i \mid \xi_1 \cdots \xi_q) = \lambda_{i1}\xi_1 + \cdots + \lambda_{iq}\xi_q,$$

for $i = 1, \ldots, p$. The factor loadings, λ_{ij}, are regression coefficients that indicate the expected change in the ith variable that is related to a unit change in the jth factor, holding all other factors constant. Thus, the loadings measure the direct influence of a factor on the variables. Expressing the equations of all variables simultaneously in a single equation using matrix notation yields

$$\mathrm{E}(\boldsymbol{x} \mid \boldsymbol{\xi}) = \boldsymbol{\Lambda\xi}, \tag{1}$$

where $\boldsymbol{x}$ is a $(p \times 1)$ vector of random variables that have zero mean, $\boldsymbol{\xi}$ is a $(q \times 1)$ vector of random factors, and $\boldsymbol{\Lambda} = (\lambda_{ij})$ is a $(p \times q)$ matrix of factor loadings.

In addition to Eq. (1), the factor model makes two further assumptions. First, as has been mentioned, the variables are assumed to be uncorrelated when controlling for the factors. If the conditional covariance matrix is denoted as $\boldsymbol{\Psi}$, that is, $\boldsymbol{\Psi} = \mathrm{Var}(\boldsymbol{x} \mid \boldsymbol{\xi})$, then this assumption can be expressed as $\boldsymbol{\Psi} = \mathrm{diag}(\psi_{11}, \ldots, \psi_{pp})$. The ψ-parameters are referred to as uniquenesses. Second, the model assumes $\mathrm{E}(\xi_j) = 0$ and $\mathrm{Var}(\xi_j) = 1$, for $j = 1, \ldots, q$. This assumption removes the arbitrariness in the scales of the factors and does not restrict the generality of the factor model.

Defining the residual vector as $\boldsymbol{\epsilon} = \boldsymbol{x} - \mathrm{E}(\boldsymbol{x} \mid \boldsymbol{\xi})$, it is not difficult to see that $\boldsymbol{\Psi} = \mathrm{Var}(\boldsymbol{\epsilon})$. Both specific factors and measurement error may contribute to the residual term. However, because the factor model does not allow for the separation of these two sources of observed score variability, we do not distinguish between them.

From Eq. (1) and the preceding assumptions, it follows that the covariance matrix of the variables is

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda\Phi\Lambda}' + \boldsymbol{\Psi}, \tag{2}$$

where $\boldsymbol{\Phi} = (\phi_{ij})$ denotes the correlation matrix of the factors. Equation (2) shows that each variable's variance is the sum of two separate sources. One source encompasses the influences of specific factors and measurement error and is referred to as the variable's uniqueness, ψ_{ii}. The other source, commonly referred to as the variable's communality, h_i^2, is the influence the factors have on determining the variable values. More specifically, the communality is defined as $h_i^2 = \mathrm{Var}(\mathrm{E}(x_i \mid \boldsymbol{\xi}))$ and is given in terms of the model parameters by the ith diagonal element of $\boldsymbol{\Lambda\Phi\Lambda}'$.

The covariance between factors and variables,

$$\mathrm{Cov}(\mathbf{x}, \boldsymbol{\xi}) = \boldsymbol{\Lambda\Phi}, \tag{3}$$

is called the factor structure. Specializing this expression to a variable–factor pair yields $\mathrm{Cov}(x_i, \xi_j) = \lambda_{i1}\phi_{1j} + \cdots + \lambda_{iq}\phi_{qj}$. The covariances of the factor structure matrix do not control for other factors. This means that even if a particular variable has a zero loading on, say, the first factor, it still can be correlated with the first factor because of indirect effects. Specifically, if the variable loads on a factor that is also correlated with the first factor, the variable will be correlated with the first factor despite its zero loading on this factor. Also note that for standardized variables, the factor structure matrix gives the correlations rather than the covariances between the factors and the variables.

An important special case occurs if the factors are uncorrelated, that is, $\boldsymbol{\Phi} = \boldsymbol{I}$. In this case, the factors are said to be orthogonal and several of the preceding equations simplify. First, the communality of the ith variable reduces to $h_i^2 = \lambda_{i1}^2 + \cdots + \lambda_{iq}^2$. Second, the factor structure matrix [see Eq. (3)] is equal to the factor loading matrix, that is, $\mathrm{Cov}(\boldsymbol{x}, \boldsymbol{\xi}) = \boldsymbol{\Lambda}$. Thus, if the factors are orthogonal, a zero loading will imply a zero covariance between the corresponding variable–factor pair. Again, note that for standardized variables, the λ parameters represent correlations rather than covariances between variables–factor pairs.

Estimating the Model Parameters

Before discussing approaches to estimating the factor loadings, it should be noted that factor loadings are not uniquely defined. If $\boldsymbol{T}$ is a nonsingular $(q \times q)$ matrix such that $\mathrm{diag}(\boldsymbol{T\Phi T}') = \boldsymbol{I}$, then $\boldsymbol{\xi}^* = \boldsymbol{T\xi}$ will be an equally legitimate set of factors with $\mathbf{\Lambda}^* = \mathbf{\Lambda}\boldsymbol{T}^{-1}$. This follows because the assumptions of the factor model are also fulfilled by the new factors $\boldsymbol{\xi}^*$ and the fact that the conditional expectation of the variables is invariant to this transformation. This is easily seen by noting that $\mathrm{E}(\boldsymbol{x} \mid \boldsymbol{\xi}^*) = \mathbf{\Lambda}^*\boldsymbol{\xi}^* = (\mathbf{\Lambda}\boldsymbol{T}^{-1})(\boldsymbol{T}\boldsymbol{\xi}) = \mathbf{\Lambda}\boldsymbol{\xi} = \mathrm{E}(\boldsymbol{x} \mid \boldsymbol{\xi})$. When the factors are uncorrelated, the set of legitimate transformation matrices $\boldsymbol{T}$ is limited to orthogonal matrices, which fulfill the condition $\boldsymbol{T}\boldsymbol{T}' = \boldsymbol{I}$. In order to obtain a solution for the factor loadings, it is desirable to remove this indeterminacy. This can be achieved if the factors are uncorrelated, that is, $\mathbf{\Phi} = \boldsymbol{I}$, and if the so-called canonical constraint, which requires $\mathbf{\Lambda}'\mathbf{\Psi}^{-1}\mathbf{\Lambda}$ to be diagonal, is satisfied.

Having imposed sufficient model restrictions to define the parameters uniquely, the degrees of freedom (df) can be determined. The degrees of freedom are equal to the difference between the $p(p+1)/2$ freely varying elements in the unconstrained population covariance matrix $\mathbf{\Sigma}$ and the number of unrestricted model parameters. The degrees of freedom characterize the extent to which the factor model offers a simple explanation of the correlations among the variables. A necessary condition for the identification of the model parameters is $\mathrm{df} \geq 0$. Clearly, if the factors are uncorrelated, all model parameters are either loadings or uniquenesses and the total number of model parameters is $pq + p$. Because the canonical constraint introduces $q(q-1)/2$ restrictions on the model parameters, the degrees of freedom are

$$\begin{aligned}\mathrm{df} &= p(p+1)/2 - [pq + p - q(q-1)/2] \\ &= (1/2)\left[(p-q)^2 - (p+q)\right]. \end{aligned} \tag{4}$$

Estimation procedures may yield negative estimates for the ψ parameters that are outside the permissible range. Such inadmissible solutions are commonly referred to as Heywood cases. Heywood cases occur quite frequently in practice. A simple strategy of dealing with negative uniquenesses is to set them equal to zero. However, this strategy implies an unrealistic model characteristic, namely, that the factors perfectly explain the variation of the variables having zero uniquenesses. Finally, note that methods for estimating the factor loadings assume that the number of factors is known.

Maximum-Likelihood Factor Analysis

When estimating the factor loadings by maximum-likelihood, a multivariate normal distribution is assumed to underlie the variables. The maximum-likelihood estimates of loadings and uniquenesses are obtained from minimizing the discrepancy function

$$F(\mathbf{\Lambda}) = \log|\mathbf{\Sigma}| + \mathrm{trace}(\boldsymbol{S}\mathbf{\Sigma}^{-1}), \tag{5}$$

where $\boldsymbol{S}$ denotes the usual sample covariance matrix and $\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}$. Note that the expression for $\mathbf{\Sigma}$ differs from that in Eq. (2) because of the requirement $\mathbf{\Phi} = \boldsymbol{I}$ discussed previously. Minimization of Eq. (5) with respect to the λ and ψ parameters requires iterative numerical methods.

If the assumptions underlying maximum-likelihood estimation are met, then this parameter estimation method has several desirable properties. First, it provides a likelihood-ratio statistic for testing the hypothesis that a particular number of factors is sufficient to describe the sample covariance matrix adequately. Second, standard errors for the loadings can be derived that allow testing of whether the loadings are different from zero. Third, the solution is scale-free in the sense that the results obtained from analyzing the correlation matrix can be obtained by rescaling the results obtained from analyzing the covariance matrix, and vice versa. A drawback of maximum-likelihood factor analysis is that the sample covariance matrix has to be of full rank.

Principal-Axis Factor Analysis

Principal-axis factoring starts by considering the matrix $\boldsymbol{S}_r = \boldsymbol{S} - \tilde{\mathbf{\Psi}}$, where $\mathbf{\Psi}$ contains initial estimates of the uniquenesses. One popular method of obtaining these initial estimates is to calculate $\tilde{\psi}_{ii} = s_{ii}(1 - R_i^2)$, where s_{ii} is the ith variable's variance and R_i^2 is the squared multiple correlation coefficient obtained from a regression of x_i on the remaining variables.

Because $\boldsymbol{S}_r$ is symmetric, it is possible to write $\boldsymbol{S}_r = \mathbf{\Gamma}\mathbf{\Theta}\mathbf{\Gamma}'$, where the columns of $\mathbf{\Gamma}$ contain p eigenvectors of $\boldsymbol{S}_r$ and $\mathbf{\Theta} = \mathrm{diag}(\theta_1, \ldots, \theta_p)$ contains the corresponding eigenvalues θ_j, $j = 1, \ldots, p$. Without loss of generality, the eigenvalues can be assumed ordered such that $\theta_1 \geq \theta_2 \cdots \geq \theta_p$. Note that some of these eigenvalues may be negative. Let the number of positive eigenvalues be greater or equal to q, then $\mathbf{\Lambda}_q = \mathbf{\Gamma}_q\ \mathrm{diag}(\theta_1^{1/2}, \ldots, \theta_q^{1/2})$ can be defined, where $\mathbf{\Gamma}_q$ contains the first q columns of $\mathbf{\Gamma}$.

If one defines $\tilde{\mathbf{\Sigma}} = \mathbf{\Lambda}_q\mathbf{\Lambda}_q'$, then it can be shown that this matrix minimizes the least-squares discrepancy function

$$F(\mathbf{\Lambda}) = \mathrm{trace}[(\boldsymbol{S}_r - \tilde{\mathbf{\Sigma}})^2] \tag{6}$$

for fixed q. In other words, $\boldsymbol{S}_r$ can be optimally approximated in the least-squares sense by $\mathbf{\Lambda}_q\mathbf{\Lambda}_q'$, and therefore $\boldsymbol{S}$ is closely approximated by $\mathbf{\Lambda}_q\mathbf{\Lambda}_q' + \tilde{\mathbf{\Psi}}$. It is possible to iteratively apply this procedure. Having estimated $\mathbf{\Lambda}_q$ using the procedure just explained, the initial estimate of $\mathbf{\Psi}$ can be updated by calculating $\tilde{\mathbf{\Psi}} = \mathrm{diag}(\boldsymbol{S} - \mathbf{\Lambda}_q\mathbf{\Lambda}_q')$. An updated $\boldsymbol{S}_r$ matrix can then be calculated that leads to a new estimates of $\mathbf{\Lambda}_q$. The iteration is continued in this manner until the change in the factor loadings across successive iterations becomes negligible. Minimizing Eq. (6) has been recommended as more likely to find real but small factors when compared to the number of factors extracted from minimizing the maximum-likelihood discrepancy function [see Eq. (5)].

Determining the Number of Factors

Approaches to estimating the factor loadings require the number of factors to be known. However, this is hardly ever the case in practice. Therefore, starting with a one-factor model, models are fitted to the data, increasing the number of factors sequentially by one.

If the factor model has been fitted by maximum likelihood and the variables follow a multivariate normal distribution, a likelihood-ratio test can be used to test whether a specific number of factors is sufficient to explain the sample covariances. The null hypothesis states that at most q factors underlie the sample covariances. If H_0 is true, the test statistic follows asymptotically a chi-squared distribution, with degrees of freedom given by Eq. (4). A drawback of the likelihood-ratio test is that the number of significant factors will be overestimated if the factor model is only approximately true, especially when sample size is large. In addition, if several models are estimated by varying the number of factors, the test procedure is open to criticism because of an inflated type-one error rate due to multiple testing. Therefore, it is useful to consider two other commonly used rules of thumb for deciding on the number of factors that do not require specific distributional assumptions. However, both rules apply only if the sample correlation matrix $\boldsymbol{R}$ rather than the sample covariance matrix $\boldsymbol{S}$ is used to fit the factor model.

The first rule is based on the eigenvalues of $\boldsymbol{R}$. Because only a few of these eigenvalues will be larger than 1.0, this rule states that the number of factors should equal the number of eigenvalues greater than 1.0. The second rule for choosing the number of factors is a visual plotting procedure called the scree test, which plots the ordered eigenvalues against their rank. Ideally, the decreasing trend in the eigenvalues exhibited by this plot has a clifflike shape. Such a shape results if only the first few eigenvalues are "large" and the remaining eigenvalues exhibit a linear decreasing trend, representing the "scree." It has been recommended to retain as many factors as there are eigenvalues too large to be considered part of the scree. Although there is no mathematical rationale behind this procedure, the validity of the scree test has become accepted in standard factor analysis books. A drawback of the scree test is that it may be difficult sometimes to determine by visual inspection whether a particular eigenvalue should be considered large or small. To remedy the subjectivity involved in the scree test, there is a statistical test for the hypothesis, that the decrease in a set of eigenvalues follows a linear trend. Finally, note that the criteria discussed in this section may not agree when applied to a particular data set. In these cases, researchers may want to decide on the number of factors by taking substantive knowledge into consideration.

Rotating Factors to Simple Structure

Because the estimated factor loadings are based on arbitrary constraints used to define uniquely the model parameters, the initial solution may not be ideal for interpretation. Recall that any factor transformation $\boldsymbol{\xi}^* = \boldsymbol{T}\boldsymbol{\xi}$ for which $\mathrm{diag}(\boldsymbol{T}\boldsymbol{\Phi}\boldsymbol{T}') = \boldsymbol{I}$ is an equally legitimate solution to Eq. (1). To simplify interpretation, it is desirable to rotate the factor loading matrix to simple structure, which has been defined in terms of five criteria: (1) each row of $\boldsymbol{\Lambda}$ should have at least one zero; (2) each of the q columns of $\boldsymbol{\Lambda}$ should have at least q zeros; (3) for every pair of columns of $\boldsymbol{\Lambda}$, there should be several variables with a zero loading in one column but not in the other; (4) for every pair of columns of $\boldsymbol{\Lambda}$, there should be a considerable proportion of loadings that are zero in both columns if $q \geq 4$; and (5) for every pair of columns of $\boldsymbol{\Lambda}$, only few variables should have nonzero loadings in both columns. Rotation techniques differ according to their emphasis on particular simple structure criteria.

Generally, a distinction is made between orthogonal and oblique rotation techniques, which yield uncorrelated or correlated factors, respectively. If the substantive concepts identified by the factors are related, correlated factors are appealing because they allow for a more realistic representation of the concepts, as compared to orthogonal factors.

One of the most popular orthogonal rotation methods is the varimax approach. This approach aims at finding a loading pattern such that the variables have either large (positive or negative) loadings or loadings that are close to zero. The varimax approach tries to accomplish this loading pattern by maximizing the variance of the squared loadings for each factor. Another popular orthogonal rotation techniques is the quartimax rotation, which maximizes the variance of the squared factor loadings in each row.

One of the most common oblique rotation techniques is the promax approach. This approach improves the loading pattern obtained from an orthogonal rotation in the sense of further increasing large loadings and further decreasing small loadings. Varimax rotation is commonly used as prerotation to promax. The promax approach accomplishes this goal in two steps. First a "target" matrix is obtained from the normalized loading matrix by replacing each factor loading by its kth power. For even powers, the signs of the loading matrix elements carry over to the corresponing target matrix elements. Common values for k are 3 and 4. Second, the orthogonal factors are rotated such that the variable loadings are, in the least-squares sense, as close as possible to the corresponding elements of the target matrix.

Predicting the Factor Scores

Because the factor model assumes that the variables are determined to a considerable extent by a linear combination of the factors $\xi_{i1}, \ldots, \xi_{iq}$, it is often of interest to determine the factor scores for each individual. Two approaches to predicting factor scores from the variable raw scores are discussed here. Both approaches assume variables and factors to be jointly normally distributed.

The so-called regression factor scores are obtained as

$$\hat{\boldsymbol{\xi}} = \boldsymbol{\Phi}\boldsymbol{\Lambda}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}.$$

It can be shown that this predictor minimizes the average squared prediction error, $E\{\sum_{j=1}^{q}(\hat{\xi}_j - \xi_j)^2\}$, among all factor score predictors that are linear combinations of the variables.

The Bartlett factor score predictor is given by

$$\hat{\boldsymbol{\xi}} = \left(\boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}\right)^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}\boldsymbol{x}.$$

This expression also minimizes the average squared prediction error among all factor score predictors that are linear combinations of the variables. In addition, the Bartlett factor score predictor is conditionally unbiased, that is, $E(\tilde{\boldsymbol{\xi}} \mid \boldsymbol{\xi}) = \boldsymbol{\xi}$. Notice that when calculating the factor scores, the matrices $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$ are replaced by their estimates. In these cases, the optimum property under which the predictors have been derived may not apply. The formula for the Bartlett factor scores can be expressed equivalently when $\boldsymbol{\Psi}$ is replaced with $\boldsymbol{\Sigma}$. An advantage of this formulation is that the factor scores can be calculated even if $\boldsymbol{\Psi}$ is singular, which may occur if uniquenesses are zero (Heywood case).

The use of factor scores is problematic because the factor loadings and the factor intercorrelations cannot be defined uniquely. The predicted factor scores will depend on the selected rotation procedure; therefore, factor scores resulting from different rotation procedures may rank order individuals differently. In addition, factor scores are problematic to use as independent variables in regression models because their values differ from the true values, which typically leads to bias in the regression coefficients.

Example

Generally, the results from a factor analysis of a correlation matrix and the corresponding covariance matrix are not identical. When analyzing a covariance matrix, variables having large variance will influence the results of the analysis more than will variables having small variance. Because the variances of the variables are intrinsically linked to the measurement units, it is preferable to analyze standardized variables, which is equivalent to fitting the factor model based on the correlation matrix, if the variables have been measured using different units.

Factor analysis can be illustrated using the artificial data set given in Table I. The data set contains standardized performance scores of 10 individuals obtained from an algebra problem, a trigonometry problem, a logic puzzle, a crossword puzzle, a word recognition task, and a word completion task. The correlation matrix of

Table I Standardized Raw Scores of Six Performance Measures[a]

Observation	x_1	x_2	x_3	x_4	x_5	x_6
1	−0.697	−0.700	−1.268	−2.245	−1.973	−1.674
2	−1.787	−1.538	−2.018	0.486	−0.163	−0.065
3	0.206	−0.913	0.079	0.801	0.964	1.043
4	−0.191	−0.430	1.074	0.002	−0.071	−0.159
5	−0.606	−0.225	0.296	−0.602	−0.990	−1.174
6	0.171	−0.417	−0.620	−0.519	0.694	0.648
7	1.460	1.038	0.532	1.261	−0.364	0.848
8	−0.639	0.888	0.306	−0.372	−0.305	1.101
9	0.779	1.595	0.775	0.499	1.215	−1.055
10	1.304	0.702	0.844	0.688	0.992	0.488

[a] The performance measures (x_1–x_6) are scores obtained on six tests: an algebra problem, a trigonometry problem, a logic puzzle, a crossword puzzle, a word recognition task, and a word completion task.

the six performance measures calculated from the raw data of Table I is

$$R = \begin{pmatrix} 1.0 & 0.7 & 0.7 & 0.5 & 0.5 & 0.3 \\ 0.7 & 1.0 & 0.7 & 0.3 & 0.3 & 0.1 \\ 0.7 & 0.7 & 1.0 & 0.4 & 0.4 & 0.2 \\ 0.5 & 0.3 & 0.4 & 1.0 & 0.7 & 0.6 \\ 0.5 & 0.3 & 0.4 & 0.7 & 1.0 & 0.5 \\ 0.3 & 0.1 & 0.2 & 0.6 & 0.5 & 1.0 \end{pmatrix}.$$

It is hoped that subjecting this matrix to factor analysis will explain the correlations between the performance scores in terms of a small number of factors having easily interpretable relations to the tasks. First, the number of factors has to be determined. The eigenvalues of the sample correlation matrix are 3.34, 1.35, 0.47, 0.30, 0.28, and 0.26. Clearly, if the criterion to retain as many factors as there are eigenvalues greater than 1.0 is employed, two factors are retained. The scree plot in Fig. 1 confirms this conclusion. Clearly, the "small" eigenvalues appear to decrease linearly.

Finally, because the factor model has been fitted by maximum likelihood, the likelihood-ratio test can be used to evaluate the hypothesis that two factors are sufficient to explain the sample correlations. The test statistic yields a value of 0.0334 that can be compared against a suitable quantile of the chi-square distribution based on df = 4. If the nominal type-one error rate is set to 0.05, this test does not provide evidence against the null hypothesis. As a result, for this artificial data set, all three criteria for determining the number of factors agree.

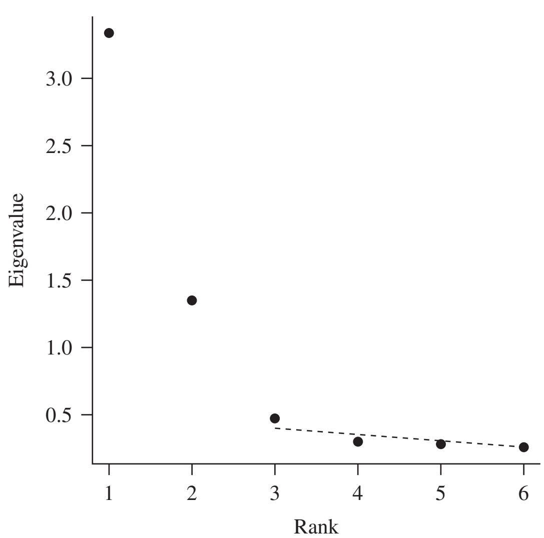

Figure 1 Scree plot depicting the ordered eigenvalues plotted against their rank. "Small" eigenvalues follow a linear decreasing trend.

Table II Maximum-Likelihood Estimates of Factor Loadings before Rotation and after Varimax and Promax Rotation

	Rotation method					
	Unrotated		*Varimax*		*Promax*	
Variable	ξ_1	ξ_2	ξ_1	ξ_2	ξ_1	ξ_2
x_1	0.828	−0.226	0.781	0.356	0.758	0.184
x_2	0.711	−0.489	0.860	0.079	0.920	−0.141
x_3	0.754	−0.335	0.794	0.225	0.807	0.037
x_4	0.746	0.486	0.263	0.851	0.055	0.864
x_5	0.693	0.373	0.294	0.730	0.123	0.722
x_6	0.482	0.485	0.061	0.681	−0.120	0.731

Next the focus is on the interpretation of the factors. Table II contains the factor loadings for the unrotated factors, the factor loadings after varimax rotation, and the factor loadings after promax rotation (the target matrix is created using $k = 3$ and varimax rotation). Because the maximum-likelihood estimates of the loadings on the unrotated factors satisfy the canonical constraint, the unrotated loadings are typically not of substantive interest. However, note that maximum-likelihood factor analysis together with the constraint $\boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}$ is equivalent to Rao's canonical factor analysis. Because the orthogonal varimax rotation and the oblique promax rotation attempt to provide loadings that fulfill Thurstone's simple structure criteria, these loadings are more useful for interpretation.

First, consider the factor loadings after varimax rotation. Because varimax is an orthogonal rotation, the loadings can be interpreted as correlations between the variables and the factors. Based on the pattern of these loadings, each of the six variables can be identified with one and only one of the factors. Variables x_1, x_2, and x_3 are considerably influenced by the first factor whereas variables x_4, x_5, and x_6 are mainly determined by the second factor. Because the first three variables involve cognitive processing of formal and abstract material, it may be desirable to label the first factor as "formal ability" or "mathematical ability." Similarly, because the last three variables, which involve cognitive processing of verbal material, load highly on the second factor, it may be convenient to label the second factor "verbal ability."

Based on the loadings from the oblique promax rotation, the conclusions about the direct influence of the factors on the variables are essentially the same as before. Overall, the comparison between the loadings from the orthogonal and the oblique rotation shows that the relationships between the factors and the variables have become clearer by allowing for correlated factors. Based on the promax rotation, the correlation between the factors is estimated to be moderately high—specifically, $\phi_{12} = 0.46$.

It is also interesting to consider the factor structure matrix, which for correlated factors gives the correlations between the variables and the factors. Using the factor

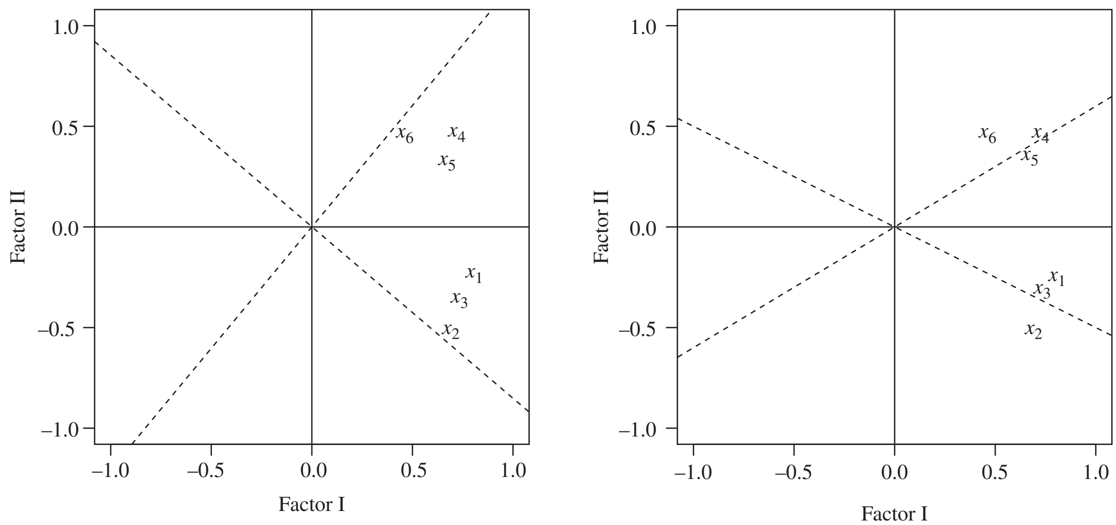

Figure 2 Loadings of variables on unrotated and rotated axes. The left-hand panel depicts the loadings with respect to the unrotated and varimax-rotated factors. The right-hand panel depicts the loadings with the respect to the unrotated and promax-rotated axes.

loadings from the promax rotation, the factor structure matrix is

$$\mathrm{Corr}(\boldsymbol{x}, \boldsymbol{\xi}) = \begin{pmatrix} 0.843 & 0.534 \\ 0.855 & 0.285 \\ 0.824 & 0.411 \\ 0.455 & 0.889 \\ 0.457 & 0.779 \\ 0.218 & 0.676 \end{pmatrix}.$$

This matrix shows that because of the factor intercorrelation, a small factor loading does not necessarily imply a small correlation between the corresponding variable–factor pair. In fact, the correlations in the factor structure matrix that correspond to small factor loadings have a moderately high value.

For the present example, factor rotation can be illustrated graphically. The left-hand panel of Fig. 2 depicts the loadings with respect to both the unrotated factors (solid lines) and the factors after varimax rotation (dashed lines). The right-hand panel of Fig. 2 depicts the loadings with respect to both the unrotated factors (solid lines) and the factors after promax rotation (dashed lines).

Finally, both the regression factor scores and the Bartlett factor scores are given in Table III. These scores are based on the results of the promax rotation. Because the factor scores are centered, positive and negative values can be interpreted as being above and below average, respectively. From Table III, the first individual appears to be considerably below average in terms of both math and verbal abilities. The second individual is considerably below average for math ability but about average for verbal ability. The factor scores of the other individuals may be interpreted similarly. Although there are small discrepancies between the regression factor scores and the Bartlett factor scores, overall they appear to agree well for this data set.

Table III Regression and Bartlett Factor Scores Calculated from the Result of a Maximum-Likelihood Factor Analysis after Promax Rotation

	Factor score			
	Regression		*Bartlett*	
Observation	ξ_1	ξ_2	ξ_1	ξ_2
1	−1.008	−2.122	−0.994	−2.463
2	−1.764	0.093	−2.076	0.300
3	−0.230	0.968	−0.356	1.197
4	0.070	−0.014	0.083	−0.024
5	−0.244	−0.846	−0.211	−0.998
6	−0.296	0.036	−0.351	0.075
7	1.065	0.862	1.173	0.931
8	0.187	−0.199	0.237	−0.261
9	1.214	0.389	1.389	0.342
10	1.007	0.833	1.107	0.902

See Also the Following Articles

Eysenck, Hans Jürgen • Guttman, Louis • Maximum Likelihood Estimation • Thurstone's Scales of Primary Abilities • Thurstone, L.L.

Further Reading

Bartholomew, D. J., and Knott, M. (1999). *Latent Variable Models and Factor Analysis,* 2nd Ed. Arnold, London.

Bartlett, M. S. (1937). The statistical conception of mental factors. *Br. J. Psychol.* **28,** 97–104.

Bentler, P. M., and Yuan, K.-H. (1997). Optimal conditionally unbiased equivariant factor score estimators. In *Latent Variable Modeling and Applications to Causality* (M. Berkane, ed.), pp. 259–281. Springer, New York.

Bentler, P. M., and Yuan, K.-H. (1998). Tests for linear trend in the smallest eigenvalues on the correlation matrix. *Pychometrika* **63**(2), 131–144.

Briggs, N. E., and MacCallum, R. C. (2003). Recovery of weak common factors by maximum likelihood and ordinary least squares estimation. *Multivar. Behav. Res.* **38,** 25–56.

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivar. Behav. Res.* **36,** 111–150.

Cattell, R. B. (1978). *The Scientific Use of Factor Analysis in Behavioral and Life Sciences.* Plenum, New York.

Chen, F., Bollen, K. A., Paxton, P., Curran, P. J., and Kirby, J. B. (2001). Improper solutions in structural equation models. *Sociol. Meth. Res.* **29**(4), 468–508.

Cudeck, R., and O'Dell, L. L. (1994). Applications of standard error estimates in unrestricted factor analysis: Significance tests for factor loadings and correlations. *Psychol. Bull.* **115,** 475–487.

Fuller, W. A. (1987). *Measurement Error Models.* Wiley, New York.

Gorsuch, R. L. (1983). *Factor Analysis,* 2nd Ed. Lawrence Erlbaum, Hillsdale, New Jersey.

Harman, H. H. (1976). *Modern Factor Analysis,* 3rd Ed. University of Chicago Press, Chicago.

Hendrickson, A. E., and White, P. O. (1964). Promax: A quick method for rotation to orthogonal oblique structure. *Br. J. Statist. Psychol.* **17,** 65–70.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika* **23,** 187–200.

Lawley, D. N., and Maxwell, A. E. (1971). *Factor Analysis as a Statistical Method,* 2nd Ed. Butterworths, London.

Rao, C. R. (1955). Estimation and tests of significance in factor analysis. *Psychometrika* **20,** 93–111.

Spearman, C. (1904). General intelligence, objectively determined and measured. *Am. J. Psychol.* **15,** 201–293.

Thurstone, L. L. (1947). *Multiple Factor Analysis.* University of Chicago Press, Chicago.

Yuan, K.-H., Marshall, L. L., and Bentler, P. M. (2002). A unified approach to exploratory factor analysis with missing data, nonnormal data, and in the presence of outliers. *Psychometrika* **67**(1), 95–122.