On the Implementation of Maximum-likelihood Factor Analysis

Moon Sup Song*
Chi Hoon Choi*

ABSTRACT

The statistical theory of factor analysis is briefly reviewed with emphasis on the maximum-likelihood method. A modified version of Jöreskog (1975) is used for the implementation of the maximum-likelihood method. For the minimization of the conditional minimum function, an adaptive Newton-Raphson method is applied.

1. Introdution

Factor analysis is a branch of multivariate statistical analysis that is concerned with the internal relationships of a set of variables. It is one of the statistical techniques to resolve a set of variables in terms of a small number of hypothetical variables, called factors. Although factor analysis has mainly developed in the field of psychology, its application has now become very popular along with the use of digital computers.

The aim of this paper is to give a brief review of the theory of factor analysis and the method to implement the algorithm of maximum likelihood estimation. The main reason of choosing the maximum likelihood method is that it is scale-free and it gives statistics for the determination of the best

* Department of Computer Science & Stetistics, Seoul National University. This work was supported in part by the Research Fund of the Ministry of Education, Korean Government, in 1979. The assistance of Kyung-Mann Kim and Myung-Hoe Huh with computer programming is gratefully acknowledged.

number of factors to use. A modified version of the algorithm of Jŏreskog (1975) is used. The estimates are computed by the Newton-Raphson method. Most of the algebraic details of the derivation of the algorithm are omitted since they can be found in the references. An exclusive bibliography of factor analysis can be found in Harman (1976).

In Section 2 unrestricted basic factor analysis model is introduced with notations. The maximum likelihood method is reviewed in Section 3. In applying the Newton-Raphson method we used the adaptive rule suggested by Ramsay (1975). Two factor rotations, varimax and promax rotations, are reviewed in Section 4. In Section 5 a brief description of the computer program outputs is given with an example. All algorithms were programmed in Fortran IV. Computations were in double precision on IBM 370/125 at Seoul National University Computer Center.

2. Notations and Preliminaries

We consider the basic factor analysis model

$$x = \Lambda f + e, \qquad (2.1)$$

where x is a column vector of p responses, $\Lambda = (\lambda_{ij})$ is a pxk matrix of unknown factor loadings, f is a column vector of k common factors, and e is a column vector of p residuals. λ_{ij} denotes the loading for the i-th variable on the j-th factor.

The residuals e are assumed to be independent of each other and of the common factor f. The variance (or dispersion) matrices of x, f, and e are respectively denoted by $\Sigma = (\sigma_{ij})$, Φ, and Ψ^2. Ψ^2 is a diagonal matrix with diagonal elements $\Psi_i^2 (i=1, \ldots, p)$, which are called unique variances. Without loss of generality we assume that Φ is an identity matrix of order k, that is, the common factors are uncorrelated and have unit variances.

By the assumptions we have made, Σ is given by the equation

$$\Sigma = \Lambda\Lambda' + \Psi^2 \qquad (2.2)$$

The elements of Λ and Ψ are usually unknown parameters which have to be estimated from experimental data. Let S denote the $p \times p$ sample covariance matrix whose elements are usual unbiased estimates of the elements of Σ obtained from a random sample of size N. The estimation problem is then to fit the matrix Σ of the form (2. 2) to an observed sample covariance matrix S.

When $k>1$, there is an infinity of choice for Λ. This indeterminacy arises from the fact that a postmultiplication of Λ by an arbitrary $k \times k$ orthogonal matrix leaves Σ unaltered. Hence to obtain a unique set of parameters and a corresponding unique set of estimates, we must impose some restrictions upon the elements of Λ.

In the following section we assume that the number of factors k can be specified in advance. In practice we have to choose the smallest k for which the model fits the data. For example, we may use the sequential procedure suggested by Lawley and Maxwell (1971, Section 4. 4).

There are many different methods of fitting Σ to S. The traditional iterated principal facor method minimizes the unweighted sum of squares

$$U=tr(S-\Sigma)^2, \qquad (2.3)$$

which is equivalent to the unweighted least squares (ULS) method (Joreskog, 1975) and also to the minres (a contraction of "minimum residuals") method proposed by Harman (1976). A disadvantage of the ULS method is that it is not scale-free and is therefore usually applied to the correlation matrix R rathen than S.

Another method to fit Σ to S is the generalized least squares (GLS) method which minimizes

$$G=\frac{1}{2}\, tr(I_p-S^{-1}\Sigma)^2, \qquad (2.4)$$

where I_p is an identity matrix of order p. This method is proposed by Jöreskog and Goldberger (1972). The GLS method is scale-free and under the assumption of normality is asymptotically equivalent to the maximum likelihood (ML) method, which will be studied thoroughly in the following section.

3. Maximum Likelihood Estimation

In the basic factor analysis model (2. 1) we assume that f and e follow independent multivariate normal distributions with zero mean and with respective covariance matrices I_k and Ψ^2, where by the assumptions in Section 2 I_k is an identity matrix of order k and Ψ^2 is a diagonal matrix with diagonal elements Ψ_i^2, $i=1,\ldots, p$.

Suppose that a random sample of N observations of x is obtained. Then the elements of the sample covariance matrix S, or of nS, follow a Wishart distribution $W(\Sigma,n)$ with $n=N-1$ degrees of freedom. Hence the logarithm of the likelihood function L is given by

$$\log L=-\frac{1}{2}n\ [\log|\Sigma|+tr(\Sigma^{-1}S)]+\log C, \tag{3. 1}$$

where C is a function of observations. Just for convenience, instead of maximizing L we want to minimize

$$F(\Lambda,\Psi)=tr(\Sigma^{-1}S)-\log|\Sigma^{-1}S|-p. \tag{3. 2}$$

Then, according to Lawley and Maxwell (1971), a constant times the minimum value of F in (3. 2) can be used as a "goodness of fit" chi-square criterion.

Jöreskog and awley (1968), to find the minimum of F in (3. 2), introducec a two-stage procedure which is proved to be successful. We first find the conditional minimum $f(\Psi)$, for a given Ψ, such that

$$f(\Psi)=\min_{\Lambda}\ F(\Lambda,\Psi). \tag{3. 3}$$

Thus the problem of minimizing F with respect to Λ and Ψ has been transformed into that of minimizing f with respect to Ψ.

The minimization problem of f in (3. 3) may be accomplished by using a method of Fletcher and Powell (1963) or Jöreskog and Lawley (1968). Jöreskog (1975) has successfully applied the Newton-Raphson method which converges quadratically to solve the minimization problem. In this paper we adopt a modified version of the algorithm suggested by Jöreskog (1975).

As mentioned in Section 2 there is an indeterminacy in the choice of Λ. To eliminate this indeterminacy we impose that $\Lambda'\Psi^{-2}\Lambda$ be a diagonal matrix. If the vatiates are rescaled so that the residual variance of each variate is unity, the jth diagonal element of $\Lambda'\Psi^{-2}\Lambda$, $\Sigma_i(\lambda_{ij}/\Psi_i)^2$, is the total variance in x due to the jth factor. Thus our choice of factors is such that the first one makes a maximum contribution to the variance in x, the second one makes a maximum contribution subject to being uncorrelated with the first one, and so on.

Since the function values and the first and second derivatives of f in (3.3) may be represented in terms of eigenvalues and eigenvectors of a matrix, we introduce some notations in the following.

Let $\gamma_1 \leq \gamma_2 \leq \cdots \leq \gamma_p$ be the eigenvalues of $\Psi S^{-1}\Psi$ and let $\omega_1, \omega_2, \ldots, \omega_p$ be an orthonormal set of corresponding eigenvectors. Let $\Gamma_1 = \text{diag}(\gamma_1, \gamma_2, \ldots \gamma_k)$ and $\Gamma_2 = \text{diag}\ (\gamma_{k+1}, \gamma_{k+2}, \ldots, \gamma_p)$ and let $\Omega_1 = [\omega_1, \omega_2, \ldots, \omega_k]$ and $\Omega_2 = [\omega_{k+1}, \omega_{k+2}, \ldots, \omega_p]$. Then according to Equation (19) and (23) of Jöreskog (1975), the conditional solution $\tilde{\Lambda}$ and the conditional minimum f given Ψ are respectively given by

$$\tilde{\Lambda} = \Psi\Omega_1(\Gamma_1^{-1} - I_k)^{\frac{1}{2}} \tag{3.4}$$

$$f(\Psi) = \sum_{m=k+1}^{p} (\log\gamma_m + \frac{1}{\gamma_m} - 1). \tag{3.5}$$

Note that when one or more of the φ_i are close to zero, Equation (3.4) is not well defined in the sense that the corresponding rows of $\tilde{\Lambda}$ are diminished. Thus, we give seperate dicussion to the case when one or more of φ_i's are close to zero at the end of this section.

From Equation (41) and (42) of Jöreskog (1975) the first and second derivatives of f in (3.3) are respectively given by

$$\frac{\partial f}{\partial \varphi_i} = \frac{2}{\varphi_i} \sum_{m=k+1}^{p} \left(1 - \frac{1}{\gamma_m}\right) \omega_{im}{}^2 \tag{3.6}$$

$$\frac{\partial^2 f}{\partial \varphi_i \partial \varphi_j} = \frac{4}{\varphi_i \varphi_j} \sum_{m=k+1}^{p} \left\{ \frac{1}{\gamma_m} \omega_{im}{}^2 \omega_{jm}{}^2 + \left(1 - \frac{1}{\gamma_m}\right) \omega_{im}\ \omega_{jm} \right.$$

$$\left. \cdot \sum_{n \pm m} \frac{\lambda_m + \lambda_n}{\lambda_m - \lambda_n} \omega_{in}\ \omega_{jn} - \frac{1}{2} \delta_{ij} \left(1 - \frac{1}{\lambda_m}\right) \omega_{im} \omega_{im} \right\} \tag{3.7}$$

where ω_{im} is the ith element of the eigenvector ω_m. when one or more of the φ_i's are close to zero, the computation of the first and second derivatives are numerically unstable. To overcome this difficulty Jöreskog suggested the transformation from φ_i to v_i defined by

$$v_i = \log \varphi_i{}^2, \quad \varphi_i = \sqrt{e^{v_i}}. \tag{3.8}$$

Using this transformation the first and the second derivatives in (3. 6) and (3. 7) are respectively given by

$$\frac{\partial f}{\partial v_i} = \sum_{m=k+1}^{p} \left(1 - \frac{1}{\gamma_m}\right) \omega_{im}{}^2 \tag{3.9}$$

and

$$\frac{\partial^2 f}{\partial v_i \partial v_j} = -\delta_{ij} \frac{\partial f}{\partial v_i} + \sum_{m=k+1}^{p} \omega_{im} \omega_{jm}$$

$$\cdot \left(\sum_{n=1}^{k} \frac{\gamma_m + \gamma_n - 2}{\gamma_m - \gamma_n} - \omega_{in}\ \omega_{jn} + \delta_{ij} \right] \cdot \tag{3.10}$$

From the relation $\sum_{m=k+1}^{p} \omega_{im} \omega_{jm} = \delta_{ij} - \sum_{n=1}^{k} \omega_{in} \omega_{jn}$, when $\gamma_{k1}, \gamma_{k2}, \cdots, \gamma_p$ are all close to one, the second derivatives in (3. 10) can be approximated by

$$\frac{\partial^2 f}{\partial v_i \partial v_j} \approx \left(\sum_{m=k+1}^{p} \omega_{im} \omega_{jm} \right)^2. \tag{3.11}$$

Let v be the column vector of $v_i's$ and let h and H be the column vector and matrix of the corresponding first and second derivatives, respectively. Then the algorithm of the Newton-Raphson produre is given by

$$H^{(s)} \delta^{(s)} = h^{(s)} \tag{3.12}$$

$$v^{(s+1)} = v^{(s)} - \delta^{(s)}, \tag{3.13}$$

where the superscript s denotes the sth iteration and $\delta^{(s)}$ is a column vector of corrections deterined by (3. 12).

Ramsay (1975) proposed an adaptive rule to improve the rate of convergence of iterative procedures in solving implicit equations (see also Ramsay (1977)). Here we apply the Ramsay's adaptive rule to the Newton-Raphson procedure defined by (3. 12) and (3. 13).

Let $\hat{v}^{(s)}$ be the vector of estimates at sth iteration. Then the adaptive Newton-Raphson procedure is defined by

$$\hat{v}(^{(s+1)}=\theta\hat{v}^{(s)}+(1-\theta)v^{(s+1)}, \tag{3. 14}$$

where $v^{(s+1)}$ is the vector of estimates computed by Equation (3. 13) and θ is the acceleration parameter. As mentioned in Ramsay (1975), by an appropriate choice of θ, we can expect to

a) damp out oscillation of iterates by choosing

$$0<\theta<1$$

with more and more damping being induced as $\theta\rightarrow 1$,

b) accelerate slow convergence by choosing

$$\theta<0$$

with more and more acceleration being induced as $\theta\rightarrow-\infty$.

Note that if θ is very close to one, then the iteration is forced to terminate too soon before a satisfactory convergence is obtained. Thus in practice we set upper and lower limits of θ to be used.

Ramsay (1975) also developed a rule which permits the recalculation of θ after every third every third iteration as follows. Let θ_{old} be the value of θ used over the last three iterations. Then the updated θ_{new} is given by

$$\theta_{\text{new}}=1-(1-\theta_{\text{old}})\,\|\triangle\hat{v}^{(s-1)}\|\,/\,\|\triangle^{2}\hat{v}^{(s-1)}\|, \tag{3. 15}$$

where $\|\cdot\|$ is a measure of lengths of vectors and $\triangle\hat{v}^{(s-1)}$ and $\triangle^{2}\hat{v}^{(s-1)}$ are the first and the second difference vectors, respectively, defined by

$$\triangle\hat{v}^{(s-1)}=\hat{v}^{(s)}-\hat{v}^{(s-1)}, \tag{3. 16}$$

$$\triangle^{2}\hat{v}^{(s-1)}=\triangle\hat{v}^{(s)}-\triangle\hat{v}^{(s-1)}. \tag{3. 17}$$

For the convergence criterion we may use

$$\max_i \left| \hat{v}_i^{(s+1)} - \hat{v}_i^{(s)} \right| < \epsilon.$$

The initial value of v in the Newton-Raphson iteration (3. 12) and (3. 13) may be chosen as

$$v_i^{(1)} = \log[(1-k/2p=/s^{ii}], \qquad (3.\ 18)$$

where s^{ii} is the ith diagonal element of S^{-1}. This choice has been justified by Jöreskog (1963) and is widely used now.

In solving the system of equations (3. 12), we use the triagular factorization $H=TT'$ of symmetric positive definite matrices, where T is a lower triangular matrix. But, note that the exact matrix H of second derivatives given by (3. 10) may not be positive definite in the beginning. Therefore, we use the approximate matrix of second derivatives given by (3. 11), which is always positive definite, as long as the maximum correction is greater than ϵ_E(*e.g.*, $\epsilon_E=0.1$). After that, if H is positive definite, we use the exact second derivatives.

Because of model in appropirelty or small sampel size, the likelihood function (3. 1) may not have any true maximum for positive unique variances. In such case one or more of $\psi_i's$ tend to zero (or $v_i's$ tend to $-\infty$) in the course of iteration and would become negative if allowed to do so. This situation is usually referred to the Heywood case.

In Heywood case the system (3. 12) is unstable in the sense that $\partial f/\partial v_i \rightarrow 0$ and $\partial^2/\partial v_i \partial v_j \rightarrow 0$, $j=1, \ldots, p$. In this case, since the ith element of h and ith row and column of H are near zero, the system (3. 12) may produce a bad correction vector δ. Jöreskog (1975) suggested a simple and effective way to deal with this problem. We delete ith equation in the system (3. 12) and compute the corrections for all the other $v_j's$ from the reduced system. The correction term for v_i is then computed as

$$\delta_i = (\partial f/\partial v_i)/(\partial^2 f/\partial v_i^2).$$

In practice we may take v_i as a Heywood variable if $\partial^2 f/\partial v_i^2 < 0.01$.

In computing the numerical solution $\hat{v}_i$ of v_i, if v_i is a Heywood variable, $\hat{v}_i$ decreases very rapidly. Thus, if $\hat{v}_i$ is less than $\log(\epsilon)$, then we fix $\hat{v}_i$ at $\log(\epsilon)$ and the function f in (3. 5) is minimized with repect to the other variables. After the iteration is finished we set the Heywood variable ψ_i to zero and compute $\tilde{\Lambda}$ according to the following modified method.

When we have improper solutions, i.e., when one or more of the numerical solutions $\hat{\psi}_1, \hat{\psi}_2, \ldots, \hat{\psi}_p$ of $\psi_1, \psi_2, \ldots, \psi_p$ are zero or close to zero, we have to modify (3. 4) to compute the loading matrix $\tilde{\Lambda}$. To apply the Joreskog's (1975) modification we let S^{-1} be decomposed by $S^{-1}=TT'$, where T is a lower triangular matrix. We let $d_1 \geq d_2 \geq \ldots \geq d_p$ be the eigenvalues of $I_p - T'\Psi^2 T$ and let $u_1, u_2, \ldots, u_p$ be an orthonormal set of corresponding eigenvectors. We also let $D_1 = \mathrm{diag}(d_1, d_2, \ldots, dk)$ and let $U_1 = [U_1, U_2, \ldots, u_k]$. Then according to Equation (60) of Jöreskog (1975) the conditional solution $\tilde{\Lambda}$ is given by

$$\tilde{\Lambda} = T'^{-1} U_1 D_1^{1/2}. \tag{3. 19}$$

Note that when one or more of the ψ_i's are zero, the corresponding eigenvalues in D_1 are one and therefore $\tilde{\Lambda}$ in (3. 19) is well defined. Note also that the $\tilde{\Lambda}$ in (3. 4) and (3. 19) are actually equivalent.

In each iteration the computation of the first and second derivatives of f requires the eigenvalues and eigenvectors of asymmetric matrix. It is well known that the QR algorithm is the most efficient method available in the computation of eigenvalues. But it very often fails give orthogonal eigenvectors corresponding to eigenvalues which are equal or very close to zero. Thus, in the program, we would choose the Jacobi method which is known as "slow but safe" method. The Jacobi method usually requires about ten times as many operations as the QR method, but it is safe and sure in all situations and has the advantage of being capable of producing the eigenvectors along with eigenvalues.

4. Factor Rotations

As mentioned in Section 2 the factorization

$$\Sigma = \Lambda\Lambda' + \Psi^2$$

is not unique in the sense that a postmultiplication of Λ by an arbitrary $k \times k$ orthogonal matrix leaves Σ unaltered. Thus, once a factor matix is obtained, we may want to find an orthogonal transformation of the factor matrix which is more meaningful and can be interpreted more easily. Sometimes the pattern of loadings may be further simplified by transforming to oblique (or uncorrelated) factors.

There are numerous principles and procedures for factor rotation: analytical or topological methods and orthogonal or oblique transformations. For the analytic orthogonal rotation the varimax method of Kaiser (1958, 1959) is most widely used, and for the analytic oblique rotation the promax method of Hendrickson and White (1964) works well in practice. In this section we give a brief review on these two methods.

Given an unrotated factor matrix Λ, the varimax criterion requires that we make orthogonal rotations on this matrixsnch that

$$r = \sum_{j=1}^{k} \left\{ p \sum_{i=1}^{p} (\lambda_{ij}^2/h_i^2)^2 - \left[\sum_{i=1}^{p} (\lambda_{ij}^2/h_i^2) \right]^2 \right\} \tag{4.1}$$

is maximized, where h_i^2 is the communality of the ith variable defined by

$$h_i^2 = \sum_{j=1}^{k} \lambda_{ij}^2, \quad i = 1, \ldots, p. \tag{4.2}$$

The program for the varimax method can be found in the IBM Scientific Subroutine Package.

Starting with a matrix of factor loadings that has been rotated to orthogonal simple structure (e.g., varimax rotated factor loadings), the promax method transforms this matrix to an oblique simple structure. We let $\Lambda = (\lambda_{ij})$ be the varimax transformed matrix of factor loadings and define a

$p\times k$ matrix $Q=(r_{ij})$ by

$$q_{ij}=|\lambda_{ij}{}^{m-1}|\lambda_{ij}, \tag{4.3}$$

with $m>1$. Thus each element of Q is, except for sign which remains unchanged, the mth power of the corresponding element of Λ. Then the promax method seeks a transformation matrix U, not necessarily orthogonal, such that the columns of ΛU fits the columns of Q, generated by (4.3), in the least squares sense. It can be easily shown that the required matrix U is given by

$$U=(\Lambda'\Lambda)^{-1}\Lambda'Q. \tag{4.4}$$

We now normalize the columns of U so that the transformed factors have unit variances. Then the matrix of promax transformed loadings is given by

$$\Lambda^{*}=\Lambda(UD), \tag{4.5}$$

where

$$D^{2}=\mathrm{diag}[(U'U)^{-1}].$$

According to the report of Hendrickson and White (1964), the optimal value for m is 4 for the majority of cases; however, for the occasional factor analysis where the data are particularly "cleanly" structured, a lower power seems to provide the best solution.

5. A Numerical Example

In this section we describe what the program does. The input data may be raw data, covariance matrix, or correlation matrix. With a correlation matrix R of order $p\times p$, the performs a sequence of factor analysis for each number of factors

$$k=k_{L},\ k_{L}+1,\ldots k_{U}.$$

For each k, the main outputs are the unrotated factor matrix, the unique variances, eigenvalues of $\Psi S^{-1}\Psi$ at the minimum, residual correlation matrix, the varimax and/or promax rotated factor matrix, and the correlation matrix of factors after promax rotation. Various statics are produced to help the determination of the best number of factors to use.

To introduce the statisice we let $C_0=(N-1)-(2p+5)/6$ and let $f(\hat{\Psi})$ be the value of conditional minimum defined by (3.3). Then

$$\chi_k^2=(C_0-\frac{2}{3}k)\ f(\hat{\Psi}) \tag{5.1}$$

has approximately chi-square distribution with its number of degrees of freedom given by

$$d_k=\frac{1}{2}[(p-k)^2-(p+k)]. \tag{5.2}$$

(For the details, see Lawley and Maxwell (1971)). If the value of χ_k^2 in (5.1) is not significant, we accept the hypothesis H_k that, for specified k, there are k common factors. One of the main advantages of using the maximum likelihood method in factor analysis is the availability of statistics to determine the number of factors. But, the value of χ_k^2 is not always a good indicator of "goodness of fit" in pratice. Tucker and Lewis (1973) introduced a useful statistic, called relibility coefficient, to indicate the quality of representation of interrelations among attributes. The reliability coefficient ρ_k is defined by

$$\rho_k=(M_0-M_k)/(M_0-1) \tag{5.3}$$

where $M_0=C_0(-\log|R|)/[\frac{1}{2}p(p-1)]$ and $M_k=\chi_k^2/d_k$, χ_k^2 and d_k are defined by (5.1) and (5.2), respectively. The formula (5.3) is slightly different from Equation (11) of Tucker and Lewis (1973). We used the expression of Jöreskog (1975). The reliability coefficient ρ_k indicates how well a factor model with k common factors represents the covariances among variables. Lack of fit would indicate the relations among the variables are more complex than can be represented by k common factors.

An example is presented in Table 1, which is from Jöreskog (1975). Table 1 is the correlation matrix of intelligence tests administered to 286 senior high-school students at a high school, which was originally gathered and analyzed by Thurstone (1940). Nine tests are used in Jöreskog just for illustration.

The data were analyzed with $k=1,\ldots,$ 5 and $\epsilon=^{-3}$. For the convergence criterion $\epsilon=10^{-5}$ the test results which are not presented here show that usually one or two extra iterations are required. Table 2 shows the values of χ_k^2 and ρ_k. For $k=1$ or 2, χ_k^2 is too significant to be accepted and ρ_k is not large enough. For $k=3$, χ_k^2 is still significant but ρ_k is sufficiently large. For $k=4$, χ_k^2 is not significant at significance level 0.05 and ρ_k is large enough. For $k=5$, it is obvious that the data are overfitted. Thus $k=3$ and $k=4$ are candidates for the number of factors.

Table 1. Correlation Matrix for Intelligence Tests

Test*	1	2	3	4	5	6	7	8	9
1	1.000								
2	0.684	1.000							
3	0.284	0.368	1.000						
4	0.177	0.186	0.332	1.000					
5	0.072	0.091	0.358	0.727	1.000				
6	0.227	0.232	0.415	0.577	0.519	1.000			
7	0.288	0.421	0.096	0.099	0.052	0.240	1.000		
8	0.029	0.141	0.149	0.305	0.304	0.320	0.342	1.000	
9	0.321	0.352	0.120	0.306	0.178	0.322	0.500	0.401	1.000

*1: Additon, 2: Multiplication, 3: Arithmetic, 4: Figures, 5: Cards, 6: Squares, 7: Identical Numbers, 8: Identical Forms, 9: Repeated Letters

Table2. Test Values for Number of Factors

k	χ_k^2	d_k	ρ_k	significance probability of χ_k^2
1	414.00	27	0.409	$p<0.001$
2	135.99	19	0.746	$p<0.001$
3	32.83	12	0.928	$p<0.005$
4	10.74	6	0.967	$0.05<p<0.10$
5	3.30	1	0.905	$0.05<p<0.10$

The intermediate values of the maximum likelihood method are given in Table 3. Column 2 of Table 3 indicates what kind of second derivatives are used: 0 means approximate and 1 means exact. Column 3 gives the function value $f(\hat{\Psi})$. The maximum correction and maximum gradient appear in Column 4 anc Column 5, respectively. The last column is the value of θ defined in (3. 15).

Table 3. Intermediate Values of the Minimization

Iteration	Type	Function	Max. correction	Max.gradient	θ
0	—	0. 1904599	—	1.24×10^{-1}	—
1	0	0. 1210846	4.17×10^{-1}	2.86×10^{-2}	—
2	0	0. 1176815	1.45×10^{-1}	2.91×10^{-3}	—
3	1	0. 1176052	4.26×10^{-2}	1.00×10^{-3}	−0. 052
4	1	0. 1175988	1.67×10^{-2}	6.91×10^{-5}	−0. 052
5	1	0. 1175988	4.94×10^{-4}	3.56×10^{-6}	−0. 052

In Table 4 we give the unrotated factor loadings for $k=3$. The last column of Table 4 is the unique variances. Since all loadings on the first factor are positive and fairly large, it is a so-called *g* factor of general intelligence. But the interpretation of the second and third factors are not obvious. The varimax rotation and the promax rotation with multiplier $m=4$ are performed to the loadings of Table 4, and the results are shown in Table 5 (loadings after promax rotation appear within paranthesis).

Table 4. Unrotated Factor Loadings for $k=3$

i	λ_{i1}	λ_{i2}	λ_{i3}	ψ_i^2
1	0. 569	0. 476	−0. 197	0. 411
2	0. 674	0. 571	−0. 151	0. 198
3	0. 503	−0. 036	−0. 228	0. 693
4	0. 690	−0. 505	−0. 061	0. 263
5	0. 604	−0. 588	−0. 100	0. 279
6	0. 628	−0. 283	0. 056	0. 522
7	0. 448	0. 317	0. 482	0. 466
8	0. 405	−0. 146	0. 429	0. 631
9	0. 537	0. 131	0. 464	0. 478

The pattern of the varimax (or promax) rotated factor loadings in Table 5 is much simpler than that of the unrotated factor loadings in Table 4.

Table 5. **Varimax (Promax with $m=4$) Rotated Factor Loadings**

i	λ_{i1}	λ_{i2}	λ_{i3}
1	0.085 (−0.036)	0.747 (0.769)	0.151 (0.022)
2	0.079 (−0.077)	0.856 (0.866)	0.252 (0.118)
3	0.411 (0.396)	0.371 (0.358)	0.012 (−0.136)
4	0.837 (0.865)	0.091 (0.019)	0.167 (−0.003)
5	0.845 (0.903)	−0.001 (−0.102)	0.081 (−0.085)
6	0.613 (0.591)	0.158 (0.052)	0.279 (0.158)
7	−0.022 (−0.192)	0.283 (0.168)	0.673 (0.711)
8	0.289 (0.215)	−0.038 (−0.181)	0.533 (0.545)
9	0.174 (0.031)	0.216 (0.078)	0.667 (0.674)

The first factor has high loadings for tests 4-6, which may be interpreted as spatial factor. The second factor has high loadings for tests 1-2, which is a numerical factor. The third factor is concernec with tests 7-9, which may be interpreted as a perceptual speed factor. Note that test 3 (arithmetic) has almost equal loadings on the first and second factors. Thus four factor analysis is applied to the data on Table 1.

Table 6. **Promax (with $m=3$) Rotated Factor Loadings for $k=4$**

i	k_{i1}	γ_{i2}	λ_{i3}	λ_{i4}	ψ_i^2
1	−0.047	0.078	−0.900	0.044	0.225
2	0.113	−0.040	−0.677	−0.195	0.341
3	1.001	0.001	−0.028	0.047	0.000
4	−0.091	0.947	−0.085	0.049	0.184
5	0.037	0.830	0.055	0.049	0.184
5	0.037	0.830	0.055	0.065	0.332
6	0.116	0.511	−0.038	−0.180	0.516
7	−0.034	−0.197	−0.089	−0.776	0.431
8	0.038	0.154	0.224	−0.576	0.621
9	−0.097	0.094	−0.126	−0.630	0.501

When four factor analysis was applied to the data on Table 1, the third variate appeared to be a Heywood variable. The first factor has a loading of 1,000 for test 3. The promax (with $m=3$) rotated factor loadings are shown in Table 6. The first factor is obviously an arithmetic factor. It is interesting to note that the remaining factors can be intereted as the three factors in the three factor analysis.

REFERCNCES

[1] Fletcher, R. and Powell, M.J.D. (1963). "A rapidly convergent descent method for minimization," Comput. J., 2, 163-168.

[2] Harman, H.H. (1976). Modern Factor Analysis (3rd edition revised), Univ. of Chicago Press, Chicago.

3] Hendrickson, A.E. and White, P.O. (1964). "PROMAX: a quick method for rotation to oblique simple structure," Brit. J. Statist. Psychol., 17, 65-70.

[4] Jöreskog, K.G. (1963). Statistical Estimation in Factor Analysis, Almqvist and Wiksell, Stockholm.

[5] Jöreskog, K.G. (1975). "Factor analysis by least-squares and maximum likelihood methods," in *K*. Enslein A. Ralston, and H.S. Wilf (eds.): Statistical Methods for Digital Computers, John Wiley & Sons, Inc., New York.

[6] Jöreskog, K.G. and Goldberger, A.S. (1972). "Factor analysis by generalized least squares," Rsychometrika, 37, 243-259.

[7] Jöreskog, K.G. and Lawley, D.N. (1968). "New methods in maximum likelihood factor analysis," Brit. J. Math. Statist. Psychol., 21, 85-96.

[8] Kaiser, H.F. (1958). "The varimax criterion for analytic rotation in factor analysis," Psychometrika, 23, 187-200.

[9] Kaiser, H.F. (1959). "Computer program for varimax rotation in factor analysis," Educational and Psychological Measurement, 19, 413-420.

[10] Lawley, D.N. and Maxwell, A.E. (1971). Factor Analysis as a Statistical Method (2nd edition), Butterworth, London.

[11] Ramsay, J.O. (1975). "Solving implicit equations in psychometric data analysis," Psychometrika, 40, 337-360.

[12] Ramsay, J.O. (1977). "A comparative study of several robust estimates of slope, intercept, and scale in linear regression," J. Amer. Statist. Assoc., 72, 608-615.

[13] Thurstone, L.L. (1940). "Experimental study of simple structure," Psychometrika, 5, 153-168.

[14] Tucker, L.R. and Lewis, C. (1973). "A reliability coefficient for maximum likelihood factor analysis," Psychometrika, 38, 1-10.