The Statistician, Vol. 29, No. 3 0039–0526/80/00110167 $02.00

Rotation in Factor Analysis

R. A. DARTON
Personal Social Services Research Unit, University of Kent at Canterbury, Canterbury, Kent CT2 7NF

The aim of factor rotation is to produce a solution having "simple structure"; here the main methods of rotation currently available are reviewed. Rotation may be for exploratory or confirmatory purposes. Exploratory studies typically involve two stages, the production of an initial solution, and then its rotation to a solution having simple structure. The initial solution is based upon an explanatory criterion. This paper includes an example of such a study. Confirmatory studies involve the examination of the fit of data to a model with specified parameters.

Introduction

Factor analysis, and particularly factor rotation, is a contentious topic. The purpose of this paper is to provide a review of the main methods of factor rotation currently available.

Factor analysis originated at the beginning of this century with the work of Spearman, Karl Pearson and Burt, amongst others. Spearman was responsible for the development of the Two-Factor Theory, in which each variable receives a contribution from two factors, a general factor common to all variables and a specific factor unique to itself. Pearson developed the method of principal axes, later to be extended by Hotelling to become the theory of principal components. Spearman's Two-Factor Theory was eventually superseded by multiple factor analysis in which several common or group factors are postulated and in which the specific factors are normally absorbed into the error term. From this standpoint Spearman's Two-Factor Theory may be regarded as a one-factor theory, that is one common or group factor. The concept of multiple factor analysis was introduced by Garnett (1919), but was mainly developed in the 1930s and 1940s, particularly by Thurstone, who introduced the term "multiple factor analysis" (Thurstone, 1931). Thurstone was also responsible for developing the concept, central to factor rotation, of simple structure.

A review of factor rotation must take account of the large number of methods developed over the last 40 years or so. In this treatment the methods of rotation have been classified according to the degree of subjectivity involved. Secondly, the final solution may consist of orthogonal or oblique, i.e. correlated, factors. Both types of solution are illustrated below.

Two further points should be made in connection with factor rotation. Firstly, most factor analyses involve two distinct steps, the calculation of an initial solution and its rotation to obtain the final solution. The first step is based on a statistical or mathematical criterion while the second is based upon some explanatory criterion.

Secondly, the development of methods of factor analysis, both for the initial solution and for the rotated solution, has been closely related to the availability of computers. Although the theory for more satisfactory solutions was known, computationally simple solutions were essential when calculations had to be performed by hand.

In the next section the common-factor model for multiple factor analysis is outlined. This is followed by a brief discussion of two of the main methods for obtaining an initial solution. The next sections are concerned with the problem of rotation from an initial solution. This is illustrated by the results obtained from some actual data. Finally, some alternatives and extensions to the two-stage procedure mentioned above are examined.

The Common Factor Model

The Model

The common factor model expresses each of p observable variables, denoted by x_i, $i=1,\ldots,p$, in terms of q factors, $f_1,\ldots,f_q$, and an error variable e_i. Following Lawley and Maxwell (1971), we may write

$$\boldsymbol{x}=\boldsymbol{\Lambda f}+\boldsymbol{e}, \tag{1}$$

where $\boldsymbol{x}$ is a $p\times 1$ vector of observable variables,
$\boldsymbol{\Lambda}$ is a $p\times q$ matrix of factor loadings (the pattern matrix),
$\boldsymbol{f}$ is a $q\times 1$ vector of factors, and
$\boldsymbol{e}$ is a $p\times 1$ vector of error variables.

All variates will be assumed to have zero means, i.e.

$$E(\boldsymbol{x})=E(\boldsymbol{e})=\boldsymbol{0}, \qquad E(\boldsymbol{f})=\boldsymbol{0}. \tag{2}$$

Then the $p\times p$ variance–covariance matrix of the observable variables may be written as

$$E(\boldsymbol{xx'})=\boldsymbol{\Sigma}. \tag{3}$$

In addition, we make the following assumptions:

(1) The error variables are uncorrelated. Thus

$$E(\boldsymbol{ee'})=\boldsymbol{\Psi}, \tag{4}$$

where $\boldsymbol{\Psi}$ is a diagonal matrix of order p. The diagonal elements of $\boldsymbol{\Psi}$, denoted by ψ_i, $i=1, \ldots, p$, are termed the residual or unique variances of the observable variables x_i, $i=1, \ldots, p$.

(2) The factors and error variables are uncorrelated, i.e.

$$E(\boldsymbol{fe'})=\boldsymbol{0}. \tag{5}$$

(3) The factors are scaled to have unit variances. Thus

$$E(\boldsymbol{ff'})=\boldsymbol{\Phi}, \tag{6}$$

where $\boldsymbol{\Phi}$ is a $q\times q$ matrix of correlations between the factors. In the case where the factors are orthogonal, $\boldsymbol{\Phi}$ is the identity matrix.

From equations (1), (3), (4)–(6) we have

$$E(\boldsymbol{xx'})=\boldsymbol{\Sigma}=\boldsymbol{\Lambda\Phi\Lambda'}+\boldsymbol{\Psi}. \tag{7}$$

This equation provides the normal starting-point for a factor analysis solution.

In addition to the pattern matrix $\boldsymbol{\Lambda}$, we may define a structure matrix $\boldsymbol{\varkappa}$ which contains the covariances between the observed variables and the factors, i.e.

$$\boldsymbol{\varkappa}=E(\boldsymbol{xf'})=\boldsymbol{\Lambda\Phi}. \tag{8}$$

If the factors are uncorrelated, $\boldsymbol{\varkappa}=\boldsymbol{\Lambda}$.

It should be emphasized that the factor model given in equation (1) is a linear model. Non-linear factor analysis, in which the model contains products of factors, has been discussed by McDonald (1962, 1967) but is rarely used.

Indeterminacy of the Model

The factor model (1) expresses p variables in terms of q factors and p error variables, and is consequently undetermined. The problems of identification have been examined in some detail by Anderson and Rubin

(1956) and for the purposes of this paper it is only necessary to examine the problem in terms of transformations, i.e. rotations.

Suppose M is a $q \times q$ non-singular matrix and that

$$f_0 = Mf, \tag{9}$$

$$\Lambda_0 = \Lambda M^{-1}, \tag{10}$$

$$\Phi_0 = M\Phi M'. \tag{11}$$

Equations (1) and (7) may then be written as follows:

$$x = \Lambda_0 f_0 + e, \tag{12}$$

$$\Sigma = \Lambda_0 \Phi_0 \Lambda_0' + \Psi. \tag{13}$$

This indeterminacy provides the scope for factor rotation. M is the matrix which transforms or rotates one set of factors to another, and would be chosen to give rotated factors with unit variances.

In the Introduction it was noted that factor analyses typically involve two stages, an initial solution and a final solution, the latter being obtained by rotating the initial solution. In order to obtain an initial solution it is necessary to apply some restrictions to resolve the problem of indeterminacy. In some methods, such as principal factor analysis, these constraints are implicit. In others, such as maximum likelihood factor analysis, the constraints have to be specified.

Using the Correlation Matrix

In the model developed above, Σ may represent the variance–covariance matrix or the correlation matrix. Normally the correlation matrix is used in factor analysis. However, not all factoring procedures are scale-invariant, that is when Σ is rescaled the estimates of the factor loadings do not differ simply by the scaling factor. In such cases an essentially arbitrary decision has to be made as to whether to use the correlation matrix or not. This problem is discussed in connection with principal component analysis by Anderson (1963).

Variance Components

Equation (7) shows that the variance of each x-variable may be expressed as the sum of two components. For variable x_i, $i = 1, \ldots, p$, the i, ith element of the matrix $\Lambda\Phi\Lambda'$ represents the common variance or communality and the i, ith element of Ψ represents the unique or residual variance. Thus the variance of variable x_i, $i = 1, \ldots, p$, may be written as

$$\sigma_i^2 = \sum_{j=1}^{q} \lambda_{ij}^2 + \sum_{j \neq k}^{q}\sum \lambda_{ij}\lambda_{ik}\phi_{jk} + \psi_i, \quad i = 1, \ldots, p, \tag{14}$$

where λ_{ij} is the i, jth element of $\mathbf{\Lambda}$,

ϕ_{jk} is the j, kth element of $\mathbf{\Phi}$, and

ψ_i is the i, ith element of $\mathbf{\Psi}$.

If the factors are orthogonal $\phi_{jk}=0, j \neq k$, and equation (14) reduces to

$$\sigma_i^2 = \sum_{j=1}^{q} \lambda_{ij}^2 + \psi_i, \quad i=1, \ldots, p. \tag{15}$$

In the psychological literature the composition of the variance is often developed further, for example see Harman (1976, Chapter 2). However, for the purposes of this paper the composition of the variance as given in equations (14) and (15) is sufficient.

Methods for Obtaining an Initial Solution

Introduction

Summaries of the main methods for obtaining initial solutions may be found in Harman (1976), and here we shall only outline the two most important methods, namely principal factor analysis and maximum likelihood factor analysis. As with most methods for obtaining an initial solution, both of these methods produce an orthogonal solution. Principal factor analysis is basically an application of principal component analysis and therefore automatically produces orthogonal factors. Maximum likelihood estimation of factor models with correlated factors is discussed below (Maximum Likelihood Estimation in Confirmatory Factor Analysis), but for initial, exploratory, solutions orthogonal factors are always extracted. A computer is necessary to obtain the solution for both methods.

Principal Factor Analysis

This method involves the application of a principal component analysis to the matrix $\mathbf{\Sigma}-\mathbf{\Psi}$. The factors obtained will therefore be orthogonal and equation (7) may be written as

$$\mathbf{\Sigma}-\mathbf{\Psi}=\mathbf{\Lambda}\mathbf{\Lambda}'. \tag{16}$$

Since the method is based on principal component analysis it is not scale-invariant. Provided that the model is correct, i.e. that there are q factors, the matrix $\mathbf{\Sigma}-\mathbf{\Psi}$ will be of rank q and have q non-zero eigenvalues. The q eigenvalues will be positive by virtue of equation (16) and for convenience will be assumed to be distinct. If the q non-zero eigenvalues are listed in descending order of size on the diagonal of a diagonal matrix $\mathbf{\Theta}_q$, then it is easy to show that the pattern matrix $\mathbf{\Lambda}$ will satisfy the equation

$$\mathbf{\Lambda}'\mathbf{\Lambda} = \mathbf{\Theta}_q. \tag{17}$$

This equation specifies $\mathbf{\Lambda}$, and therefore the factors, uniquely provided that the diagonal elements of $\mathbf{\Theta}_q$ are distinct, as assumed above. The total amount of common variance accounted for by factor j is given by

$$\theta_j = \sum_{=1}^{p} \lambda_{ij}^2, \quad i=1, \ldots, q, \tag{18}$$

where $\theta_1 > \theta_2 > \ldots > \theta_k > 0$.

In order to apply the method of principal factor analysis to a sample correlation or variance–covariance matrix two problems must be overcome. Firstly, the matrix $\mathbf{\Psi}$ must be estimated and, secondly, the number of factors to be extracted must be determined.

Harman (1976, Chapter 5) provides a summary of the main methods for estimating $\mathbf{\Psi}$. It has been shown that the squares of the multiple correlations (SMCs) between each x-variable and the remaining $p-1$ x-variables provide lower bounds for the communalities, i.e. the diagonal elements of the matrix $\mathbf{\Sigma}-\mathbf{\Psi}$ (Roff, 1936; Dwyer, 1939). In principal factor analysis initial estimates of the elements of $\mathbf{\Psi}$ are usually obtained by using the SMCs and an iterative procedure is employed to obtain the matrix $\mathbf{\Lambda}$ and a better estimate of $\mathbf{\Psi}$.

From equation (16) the matrix $\mathbf{\Sigma}-\mathbf{\Psi}$ was seen to be positive semi-definite. However, when $\mathbf{\Psi}$ is estimated this does not necessarily remain true. Taking the same number of factors as the number of positive eigenvalues would result in an over-factorization since the total common variance would be too large. In any case, a smaller number of factors is normally required. Guttman (1954) has provided lower bounds for the number of factors. If $\mathbf{\Sigma}$ represents the correlation matrix, Guttman's weakest lower bound for the number of factors is the number of eigenvalues of $\mathbf{\Sigma}$ which are greater than or equal to 1·0. This criterion is discussed by Rummel (1970, Chapter 15). Once a factor solution has been obtained it is possible to use a test developed by Rippe (1953) to assess whether sufficient factors have been extracted. This test, which assumes that the original variables have a multivariate normal distribution, is very similar to those employed for testing for the number of factors in maximum likelihood factor analysis (see below).

Maximum Likelihood Factor Analysis

The first attempts at maximum likelihood estimation of the elements of $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ were made by Lawley in 1940. Without a computer, however, the work involved in his iterative procedure was prohibitive. Later attempts at improving the procedure were made by various workers in the 1950s, but convergence presented a problem in all of these methods. In the 1960s Joreskög and Lawley produced a new method of estimation which, although iterative, converged very rapidly (see Joreskög, 1967).

Maximum likelihood estimation does not produce a unique solution without extra conditions being imposed. However, it is independent of scale. For further comparisons between principal factor analysis and maximum likelihood factor analysis, reference may be made to Gnanadesikan (1977, Table 1).

The method of Joreskög and Lawley will now be briefly described. Assuming that the factors and error variables have independent multivariate normal distributions, the x-variables will have a multivariate normal distribution. Suppose that a sample of n individuals is measured on each of the p variables, then the sample variance–covariance matrix is given by

$$\boldsymbol{S}=\frac{1}{n-1}\sum_{k=1}^{n}(\boldsymbol{x}_k-\bar{\boldsymbol{x}})(\boldsymbol{x}_k-\bar{\boldsymbol{x}})'. \tag{19}$$

Provided that $n>p$, $\boldsymbol{S}$ will have a Wishart distribution with $n-1$ degrees of freedom (Anderson, 1958, Chapter 7). The log-likelihood function will be

$$\log_e L=-\left(\frac{n-1}{2}\right)[\log_e|\boldsymbol{\Sigma}|+\operatorname{tr}(\boldsymbol{S}\boldsymbol{\Sigma}^{-1})]+K, \tag{20}$$

where K is a function independent of $\boldsymbol{\Sigma}$. As mentioned above, the initial solution obtained in maximum likelihood factor analysis is orthogonal. Hence the problem involves obtaining estimates $\hat{\boldsymbol{\Lambda}}$ and $\hat{\boldsymbol{\Psi}}$ which satisfy the equation

$$\boldsymbol{\Sigma}=\boldsymbol{\Lambda}\boldsymbol{\Lambda}'+\boldsymbol{\Psi}, \tag{21}$$

and which maximize $\log_e L$. A two-stage process is employed to minimize the function

$$F_q(\boldsymbol{\Lambda},\boldsymbol{\Psi})=\log_e|\boldsymbol{\Sigma}|+\operatorname{tr}(\boldsymbol{S}\boldsymbol{\Sigma}^{-1})-\log|\boldsymbol{S}|-p, \tag{22}$$

this being equivalent to maximizing $\log_e L$. In the first stage F_q is minimized for given $\boldsymbol{\Psi}$, which may be denoted by

$$F_q(\boldsymbol{\Psi})=\min_{\boldsymbol{\Lambda}} F_q(\boldsymbol{\Lambda},\boldsymbol{\Psi})=F_q(\hat{\boldsymbol{\Lambda}},\boldsymbol{\Psi}). \tag{23}$$

The result of this stage of the procedure is an equation of the following form:

$$(\boldsymbol{\Psi}^{-1/2}\boldsymbol{S}\boldsymbol{\Psi}^{-1/2})(\boldsymbol{\Psi}^{-1/2}\hat{\boldsymbol{\Lambda}})=(\boldsymbol{\Psi}^{-1/2}\hat{\boldsymbol{\Lambda}})(\boldsymbol{I}+\hat{\boldsymbol{\Lambda}}'\boldsymbol{\Psi}^{-1}\hat{\boldsymbol{\Lambda}}). \tag{24}$$

If $\hat{\boldsymbol{\Lambda}}'\boldsymbol{\Psi}^{-1}\hat{\boldsymbol{\Lambda}}$ is chosen to be diagonal, equation (24) defines q of the eigenvalues and eigenvectors of the matrix $\boldsymbol{\Psi}^{-1/2}\boldsymbol{S}\boldsymbol{\Psi}^{-1/2}$. The eigenvalues must be the q largest eigenvalues (Lawley and Maxwell, 1971, Chapter 4). Assuming that these eigenvalues are distinct and greater than unity, equation (24) may be written in the following form:

$$(\boldsymbol{\Psi}^{-1/2}\boldsymbol{S}\boldsymbol{\Psi}^{-1/2})\,\boldsymbol{\Omega}=\boldsymbol{\Omega}\boldsymbol{\Theta}, \tag{25}$$

where $\boldsymbol{\Theta}=\boldsymbol{I}+\hat{\boldsymbol{\Lambda}}'\boldsymbol{\Psi}^{-1}\hat{\boldsymbol{\Lambda}}$ and the columns of $\boldsymbol{\Omega}$ contain the normalized eigenvectors corresponding to the q largest eigenvalues, i.e.

$$\boldsymbol{\Omega}=\boldsymbol{\Psi}^{-1/2}\hat{\boldsymbol{\Lambda}}(\hat{\boldsymbol{\Lambda}}'\boldsymbol{\Psi}^{-1}\hat{\boldsymbol{\Lambda}})^{-1/2}=\boldsymbol{\Psi}^{-1/2}\hat{\boldsymbol{\Lambda}}(\boldsymbol{\Theta}-\boldsymbol{I})^{-1/2}. \tag{26}$$

Hence $\boldsymbol{\Lambda}$ may be estimated from the equation

$$\hat{\boldsymbol{\Lambda}}=\boldsymbol{\Psi}^{1/2}\boldsymbol{\Omega}(\boldsymbol{\Theta}-\boldsymbol{I})^{1/2}. \tag{27}$$

In the second stage of the process f_q is minimized, i.e.

$$\min_{\boldsymbol{\Psi}} f_q(\boldsymbol{\Psi})=\min_{\boldsymbol{\Lambda},\,\boldsymbol{\Psi}} F_q(\boldsymbol{\Lambda},\boldsymbol{\Psi}). \tag{28}$$

The result of this procedure is the following estimate for $\boldsymbol{\Psi}$:

$$\hat{\boldsymbol{\Psi}}=\operatorname{diag}(\boldsymbol{S}-\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Lambda}}'). \tag{29}$$

As in principal factor analysis an initial estimate of $\boldsymbol{\Psi}$ is required. Joreskög (1967) recommends the estimates

$$\hat{\psi}_i=\left(1-\frac{q}{2p}\right)\left(\frac{1}{s^{ii}}\right),\quad i=1,\ldots,p, \tag{30}$$

where s^{ii} is the i, ith element of $\boldsymbol{S}^{-1}$. The details of the iterative procedure are given by Lawley and Maxwell (1971, Chapter 4).

In the preceding discussion it was found to be convenient to take $\hat{\boldsymbol{\Lambda}}'\boldsymbol{\Psi}^{-1}\hat{\boldsymbol{\Lambda}}$ as a diagonal matrix. Hence $\hat{\boldsymbol{\Lambda}}'\hat{\boldsymbol{\Psi}}^{-1}\hat{\boldsymbol{\Lambda}}$ will be diagonal. This requirement introduces sufficient extra conditions to make the maximum likelihood solution unique, and can be regarded as the analogue of equation (17). In maximum likelihood factor analysis the basic model (equation (21)) is essentially rescaled so that each residual variance is unity, i.e.

$$\boldsymbol{\Psi}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Psi}^{-1/2}=\boldsymbol{\Psi}^{-1/2}\boldsymbol{\Lambda}\boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1/2}+\boldsymbol{I}. \tag{31}$$

The maximum likelihood solution involves finding eigenvalues and eigenvectors for this rescaled model. The total amount of common variance in the rescaled model accounted for by factor j is given by

$$\theta_j-1=\sum_{i=1}^{p}\frac{\lambda_{ij}^2}{\psi_i},\quad j=1,\ldots,q, \tag{32}$$

where $\theta_1 > \theta_2 > \ldots > \theta_q > 1$ are the diagonal elements of $\mathbf{\Theta}$. (Cf. equation (18).)

An important feature of maximum likelihood factoring is the provision of a likelihood-ratio test for the number of factors:

$$-2 \log_e l = (n-1)\,[\log_e |\hat{\mathbf{\Sigma}}| + \mathrm{tr}(S\hat{\mathbf{\Sigma}}^{-1}) - \log_e |S| - p], \tag{33}$$

where l is the likelihood-ratio and

$$\hat{\mathbf{\Sigma}} = \hat{\mathbf{\Lambda}}\hat{\mathbf{\Lambda}}' + \hat{\mathbf{\Psi}}. \tag{34}$$

(Cf. equation (22).) For large samples $-2 \log_e l$ is approximately distributed as χ^2 if sufficient factors have been extracted, the number of degrees of freedom being given by $\frac{1}{2}[(p-q)^2 - (p+q)]$. Lawley and Maxwell suggest a sequential procedure for determining the number of factors. The number of factors extracted is increased by one at each step and the process is terminated when the null hypothesis is accepted. It may be noted that work by Bartlett has suggested that the approximation to χ^2 may be improved if the multiplying factor $(n-1)$ is adjusted for p and q. (See Bartlett, 1950, 1951, 1954.)

Factor Rotation

Introduction

As has been shown above (equations (9)–(13)), a factor solution may be subject to a rotation by a non-singular matrix. In the two methods of factoring just discussed unique solutions were obtained, but in each case this was achieved by imposing conditions on the elements of the pattern matrix $\mathbf{\Lambda}$. For the given models, the (orthogonal) factors obtained using these methods accounted for the maximum amount of variance, the second highest amount of variance, and so on. The factors obtained may be said to be based upon a mathematical or statistical criterion. However, for the purposes of explanation, other criteria have been adopted. In particular, Thurstone developed the concept of simple structure which specifies certain patterns of loadings within the matrix $\mathbf{\Lambda}$. Furthermore, in order to achieve the patterns of loadings specified by the simple structure criteria, oblique or correlated factors are permitted. It should be noted that the simple structure criteria do not define the values of particular factor loadings. Methods for rotating an initial solution to approximate a specified set of loadings are considered below (Methods for Confirmatory Factor Analysis).

As may be seen above (the Common Factor Model), oblique factor solutions are specified by a pattern matrix $\mathbf{\Lambda}$, a structure matrix $\boldsymbol{\varkappa}$ and a correlation matrix $\mathbf{\Phi}$. Thus it is often more difficult to comprehend a solution consisting of oblique factors than a solution consisting of orthogonal factors.

(Note that in the rest of this paper it will be understood that factor rotations are applied to achieved factor solutions, i.e. estimates, but the model symbols are used for clarity.)

Simple Structure Criteria

Thurstone's simple structure criteria have been described as specifying certain patterns of loadings within the matrix $\mathbf{\Lambda}$. In fact, the list of the simple structure criteria given by Thurstone (1947, Chapter 14) defines the patterns of elements in the factor reference structure (see below). However, since there is a close relationship between this and the pattern matrix, as shown below, we may follow Harman (1976, Chapter 6) in listing the criteria in terms of the elements of $\mathbf{\Lambda}$:

(1) Each row of the matrix should have at least one zero.
(2) If there are q factors, each column of the matrix should have at least q zeros.
(3) For every pair of columns in the matrix there should be several variables whose entries vanish in one column but not in the other.
(4) For every pair of columns in the matrix, a large proportion of the variables should have vanishing entries in both columns when there are four or more factors.
(5) For every pair of columns in the matrix there should only be a small number of variables with non-vanishing entries in both columns.

Diagrammatically, rotation to simple structure takes the following form. The crosses relate to "large" loadings and the blanks to "near-zero" loadings.

Initial solution			*Rotated solution*		
f_1	f_2	f_3	f_1	f_2	f_3
×	×		×		
×	×	×	×		
×	×	×	×	×	
×		×		×	
×	×	×		×	
×	×			×	×
×		×			×

(See Rummel, 1970, Chapter 16.)

Variance Components in Rotated Solutions

Equations (9)–(13) demonstrate the effect of applying a non-singular rotation to the factor model (1). In particular, equation (13) shows that the variance explained by the factors is unchanged by applying a non-singular rotation.

Types of Factor Rotation

As was noted in the Introduction, methods of factor rotation may be classified according to the degree of subjectivity involved. The first, and most subjective, methods of rotation were graphical and required a large amount of labour and experience for their successful operation. The first attempts at making the procedure of rotation more objective occurred in the 1940s. These procedures are usually termed semi-analytical since they provide a compromise between complete subjectivity and complete objectivity. The development of completely objective, or analytical, methods of factor rotation began in the 1950s and has been aided by the increasing availability of computers since then. Each of these classes of rotational methods contain orthogonal and oblique methods of solution.

Graphical and semi-analytical methods of rotation are examined fairly briefly in the remainder of this section and analytical rotation is given a rather more detailed treatment below (Analytical Factor Rotation).

Graphical Rotation

The simplest method of graphical rotation involves selecting a pair of factors and subjecting them to an orthogonal or oblique rotation. One of the new factors generated by this procedure may then be rotated with a third factor. This procedure is repeated until the factors possess the structure required. Each rotation may be expressed in terms of a transformation matrix, the overall transformation matrix being obtained as a product of the individual transformation matrices. An example of the procedure is given by Lawley and Maxwell (1971, Chapter 6).

Using equations (8) and (9), the structure matrix for an oblique rotated solution obtained from an initial orthogonal solution may be written as

$$\varkappa_0 = E(xf_0') = E(xf'M') = \Lambda M'. \tag{35}$$

The columns of M' contain the direction cosines of the oblique factors with respect to the orthogonal factors (the orthogonal frame of reference). The factors have unit variances and hence unit length and so the elements of the columns of M' are also the projections of the termini of the oblique factors on the orthogonal frame of reference. Hence the correlations between the oblique factors are given by

$$\Phi_0 = E(f_0 f_0') = E(Mff'M') = MM'. \tag{36}$$

Using equation (8)

$$\varkappa_0 = \Lambda_0 \Phi_0, \tag{37}$$

where Λ_0 is the oblique factor pattern matrix. Hence

$$\varkappa_0 = \Lambda_0 M M'. \tag{38}$$

Thus from (35) and (38)

$$\Lambda_0 = \Lambda M^{-1}, \tag{39}$$

which is equation (12).

Graphical methods involve a substantial amount of work and Thurstone (1947) describes various procedures for obtaining oblique rotations by graphical methods more systematically. In particular, he made use of a system of reference axes. For each of the q factors f_j, $j=1, \ldots, q$, there is a coordinate hyperplane π_j of $q-1$ dimensions with an orthogonal reference axis γ_j. (The subscript "0" has been dropped for clarity.) Points within π_j have zero loadings on f_j and zero projections on γ_j. Figure 1 represents the case of two dimensions, in which the hyperplane π_1 is the same as factor f_2, and vice versa:

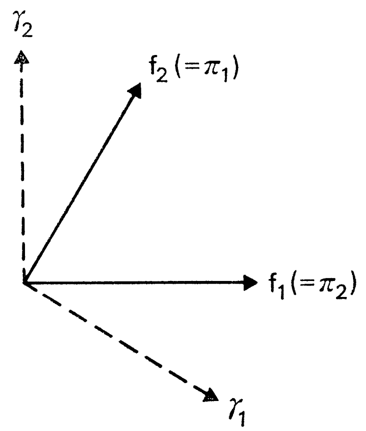

Figure 1

The oblique reference structure matrix $\varkappa_r$ may be written as

$$\varkappa_r = \Lambda T' \tag{40}$$

(cf. equation (35)). The columns of T' are the direction cosines of the oblique axes γ_j, $j=1, \ldots, q$, with respect to the original orthogonal factors. Thus the correlations between the oblique factors and the reference axes are given in the matrix MT'. However, the jth reference axis is orthogonal to all the factors contained in π_j and hence MT' is diagonal. Hence

$$T' = M^{-1}N, \tag{41}$$

where N is the diagonal matrix which normalizes M^{-1} (Harman, 1976, Chapter 12). From equations (39)–(41) we have

$$\varkappa_r = \Lambda M^{-1}N = \Lambda_0 N, \tag{42}$$

that is, the oblique reference structure matrix $\varkappa_r$ corresponds to the oblique factor pattern matrix Λ_0. Thus Thurstone's simple structure principles may be expressed in terms of either matrix.

Semi-analytical Rotations

As stated above, various attempts were made in the 1940s to make the process of factor rotation more objective. These methods involved a compromise between subjective and fully objective, or analytical procedures.

In common with analytical procedures, the first problem encountered is that of operationalizing the simple structure criteria. A procedure developed by Horst (1941), which avoids graphical methods completely, attempts this by maximizing, for each factor in turn, the ratio of the sum of squares of "significant" factor loadings to the sum of squares of all the factor loadings. The subjective aspect of this procedure involves the identification of variables having "significant" factor loadings. The use of Horst's criterion was extended by Tucker (1944) in a combination of analytical and graphical methods.

Analytical Factor Rotation

Introduction

The first completely objective method of factor rotation was produced by Carroll (1953), although a number of other solutions were produced soon afterwards which, in the orthogonal case, were found to be equivalent to Carroll's solution. The collective term for these equivalent solutions is the "quartimax method". Kaiser's varimax procedure (Kaiser, 1958) is a modification of the quartimax procedure and is probably the most commonly used analytic method of factor rotation. It may be noted that varimax and quartimax are special cases of the class of orthomax criteria for orthogonal rotations (Harman, 1976, Chapter 13).

A large number of methods of factor rotation has resulted from the extension of some of the orthogonal procedures to the more general, oblique, case. The first analytical oblique methods of rotation have been termed indirect methods since they involve the use of a reference structure, as outlined above (Factor Rotation). However, oblique rotation was simplified when Jennrich and Sampson (1966) produced a direct method not involving reference axes. Harman (1976, Chapter 14) summarizes the oblique methods and only Jennrich and Sampson's method is discussed here. Procrustes rotations, which also produce correlated factors, are discussed below (Methods for Confirmatory Factor Analysis).

Quartimax Rotation

The following people were responsible for this method: Carroll (1953), Saunders (1953a, quoted in Harman (1976), 1953b), Ferguson (1954) and Neuhaus and Wrigley (1954). Ferguson's approach developed from considerations of information theory while the other workers attempted to operationalize Thurstone's simple structure criteria. In this discussion the theory of the quartimax method is presented in general terms.

A non-singular rotation does not alter the amount of variance explained, as noted above (Variance Components in Rotated Solutions). In other words the total communality is unchanged. In addition, the communality of each variable is unchanged, as may be seen from equation (13). Thus the square of the communality is unchanged; that is, for an orthogonal rotation

$$\left(\sum_{j=1}^{q} \lambda_{ij}^2\right)^2 = \sum_{j=1}^{q} \lambda_{ij}^4 + \sum_{j\neq k}^{q}\sum \lambda_{ij}^2 \lambda_{ik}^2 = \text{constant}, \tag{43}$$

the subscript "0" having been omitted for clarity. Therefore, summing equation (43) over all p variables gives

$$\sum_{i=1}^{p} \sum_{j=1}^{q} \lambda_{ij}^4 + \sum_{i=1}^{p} \sum_{j\neq k}^{q}\sum \lambda_{ij}^2 \lambda_{ik}^2 = \text{constant}. \tag{44}$$

Carroll suggested minimizing the cross-product term while Saunders, Ferguson, and Neuhaus and Wrigley suggested maximizing the term of fourth powers. Equation (44) renders their criteria equivalent. While Ferguson did not offer a computational procedure, the computational procedures in the other three cases involved various methods of pairwise rotation of the factors.

An example of factor analysis is given below and indicates that the quartimax procedure tends to retain an important first factor. This is characteristic of this method of factor rotation and occurs because quartimax essentially attempts to simplify the rows of the pattern matrix via the minimization of the cross-product term in equation (44).

Varimax Rotation

The varimax method attempts to simplify the columns rather than the rows of the pattern matrix. It therefore precludes the retention of a fairly general first factor.

Kaiser (1958) defines the simplicity of a factor as the variance of its squared loadings. For factor j this may be denoted by $v_j{}^*$, where

$$v_j{}^* = \frac{1}{p}\left[\sum_{i=1}^{p} \lambda_{ij}{}^4 - \frac{1}{p}\left(\sum_{i=1}^{p} \lambda_{ij}{}^2\right)^2\right], \quad j=1,\ldots,q. \tag{45}$$

The varimax procedure involves maximizing the total simplicity, i.e. maximizing

$$V^* = \sum_{j=1}^{q} v_j{}^*. \tag{46}$$

In practice the factor loadings are usually normalized by their corresponding communalities which, as noted above, are unchanged by a non-singular rotation. Thus the "normal" varimax criterion is given by

$$V = \frac{1}{p}\sum_{j=1}^{q}\left[\sum_{i=1}^{p} \frac{\lambda_{ij}{}^4}{h_i{}^4} - \frac{1}{p}\left(\sum_{i=1}^{p} \frac{\lambda_{ij}{}^2}{h_i{}^2}\right)^2\right], \tag{47}$$

where

$$h_i{}^2 = \sum_{j=1}^{q} \lambda_{ij}{}^2, \quad i=1,\ldots,p, \tag{48}$$

is the communality of variable x_i.

As with quartimax rotation the computational procedure involves pairwise rotation of the factors.

The results of an application of the varimax procedure are given below. Examination of the relative proportions of variance explained by each factor indicates that varimax tends to spread the "large" loadings across the columns of the pattern matrix to a greater extent than does quartimax. This is due to the fact that, for V to be maximized, the second term in equation (47) must be small, i.e.

$$\sum_{i=1}^{p} \frac{\lambda_{ij}{}^2}{h_i{}^2}$$

must be fairly constant across factors. Since this is related to the variance accounted for by factor j (cf. equations (18) and (32)), it follows that varimax attempts to equalize the amount of variance explained by each factor.

Various extensions of the varimax procedure have been proposed by Horst (1965, Chapter 18). A more intelligible account of one of his methods is given by Lawley and Maxwell (1971, Chapter 6).

Indirect Methods for Oblique Rotation

These methods involve the use of a reference structure and have been largely superseded by the direct approach proposed by Jennrich and Sampson (1966). Consequently, they will only be dealt with very briefly here. Further details of these methods may be found in Harman (1976, Chapter 14).

As stated above, a number of methods of rotation proposed in the early 1950s turned out to be equivalent in the orthogonal case, forming the quartimax method. However, for those methods which have a generalization to the oblique case this equivalence is not maintained. For example, oblimax, the oblique version of Saunders' procedure, differs from quartimin, the oblique version of Carroll's procedure. On the other hand, the oblique version of varimax, termed covarimin, and quartimin are both members of the class of oblimin methods, just as varimax and quartimax are special cases of the class of orthomax rotations.

Direct Oblimin Rotation

Direct oblimin is the oblique procedure introduced by Jennrich and Sampson (1966) which does not involve the use of reference axes. The direct oblimin criterion takes the following form:

$$\sum_{j<k}^{q}\sum\left[\sum_{i=1}^{p}\lambda_{ij}{}^{2}\lambda_{ik}{}^{2}-\frac{\delta}{p}\sum_{i=1}^{p}\lambda_{ij}{}^{2}\sum_{i=1}^{p}\lambda_{ik}{}^{2}\right], \tag{49}$$

where, as before, the subscripts "0" have been omitted for clarity. δ is a parameter which controls the degree of correlation between the factors.

The rotational procedure effects a transformation of the form displayed in equation (39). Equation (36) places some conditions on the transformation matrix, namely

$$\operatorname{diag}(\boldsymbol{MM'})=\boldsymbol{I}_q. \tag{50}$$

Direct oblimin involves minimizing (49) subject to the constraints given in equation (50). As with quartimax and varimax the computational procedure involves an iterative process of pairwise rotation of the factors.

The value of the parameter δ is chosen before the rotation. If $0 < \delta = 1$ the factors are highly correlated. $\delta = 0$ produces a fairly correlated set of factors and $\delta < 0$ produces less correlated factors. An example of a solution obtained with $\delta = 0$ appears in the last part of this section.

Examples of Analytical Methods of Factor Rotation

A number of the methods of rotation discussed above are illustrated here. The data used were collected in the CPS 1972 American National Election Study (Inter-University Consortium for Political Research, 1975) and consist of measures of favourability towards Presidential candidates and some groups in American society. Respondents were required to state their position on a "feeling thermometer", running from 0 to 100 degrees, according to their degree of favourability towards each candidate and group. The analyses described here were performed at the University of Manchester Regional Computer Centre by means of the SPSS computer package (Nie *et al.*, 1975, Chapter 24).

Data were available at Manchester for 34 of the 36 feeling thermometer variables. Maximum likelihood factor analysis was judged inappropriate due to the bimodal nature of the frequency distributions of the variables and, instead, principal factor analysis was employed. The correlation matrix between the 34 variables was obtained using listwise deletion (Kim and Curry, 1977); full information was available for 1 439 of the original 2 191 cases.

Use of the "eigenvalue-one" criterion described above to determine the minimum number of factors suggested that six factors were required. However, since the sixth factor in a six factor solution accounted for only 1·5 per cent of the total variance five factors were judged to be sufficient. It may be noted that twelve factors would have been required to ensure that all communalities were greater than the estimated communalities (SMCs). However, twelve factors accounted for only 59·3 per cent of the total variance, as compared with the 50·2 per cent of the total variance accounted for by five factors.

The results are given in the tables below. In each case the figures have been rounded to two decimal places and rounded factor loadings with absolute values of less than 0·30 have been omitted for clarity. In each case the rotations have been obtained using Kaiser Normalization (see the section entitled Analytical Factor Rotation).

In the initial solution the first two factors account for four-fifths of the variance explained. Insofar as factors may be labelled, these factors may be regarded as representing dimensions of conservatism and radicalism respectively. The other three factors do not account for

Table 1
Initial factor pattern matrix

Variables	*Factors* 1	2	3	4	5	*Communality*
1. George Wallace	0·38	−0·38				0·36
2. Richard Nixon	0·45	−0·50	0·45			0·67
3. George McGovern		0·66			0·33	0·64
4. Spiro Agnew	0·41	−0·44	0·41			0·57
5. Sargent Shriver		0·61				0·54
6. Big business	0·33					0·22
7. Poor people	0·38	0·39				0·30
8. Liberals		0·62				0·41
9. Southerners	0·56					0·35
10. Intellectuals	0·32	0·34				0·25
11. Catholics	0·51	0·34				0·47
12. Radical students		0·59				0·53
13. Policemen	0·65					0·47
14. Protestants	0·70					0·59
15. Jews	0·55	0·35		−0·34		0·62
16. The Military	0·68					0·54
17. Whites	0·74					0·56
18. Democrats	0·33	0·53				0·54
19. Blacks	0·32	0·55				0·43
20. Republicans	0·62		0·44			0·62
21. Labour unions		0·31				0·22
22. Young people	0·48	0·37				0·39
23. Conservatives	0·56					0·43
24. Women's Lib. Movement		0·43				0·21
25. Marijuana users	−0·37	0·46	0·33			0·47
26. Black militants	−0·30	0·61	0·34			0·62
27. Urban rioters	−0·32	0·58	0·36			0·62
28. Civil Rights leaders		0·60				0·49
29. Suburban dwellers	0·67					0·53
30. Workingmen	0·74			0·40		0·81
31. Farmers	0·72			0·32		0·66
32. Ministers leading protest marches		0·67				0·61
33. Middle-class people	0·78					0·71
34. City dwellers	0·69	0·33				0·60
Variance component (eigenvalue)	8·08	5·78	1·61	0·80	0·78	
%of total variance	23·8	17·0	4·7	2·3	2·3	
Cum. % of total variance	23·8	40·8	45·5	47·9	50·2	
% of common variance	47·4	33·9	9·5	4·7	4·6	
Cum. % of common variance	47·4	81·3	90·7	95·4	100·0	

Table 2

Quartimax rotated factor pattern matrix

Variables	*Factors* 1	2	3	4	5
1. George Wallace		−0·39	0·38		
2. Richard Nixon		−0·31	0·74		
3. George McGovern		0·45	−0·34	0·55	
4. Spiro Agnew			0·68		
5. Sargent Shriver		0·39	−0·32	0·50	
6. Big business			0·41		
7. Poor people	0·51				
8. Liberals		0·47			
9. Southerners	0·49				
10. Intellectuals	0·42				
11. Catholics	0·60				0·32
12. Radical students		0·71			
13. Policemen	0·50	−0·31	0·33		
14. Protestants	0·71				
15. Jews	0·64				0·44
16. The Military	0·51	−0·32	0·37		
17. Whites	0·68				
18. Democrats	0·50			0·46	
19. Blacks	0·51	0·37			
20. Republicans	0·42		0·66		
21. Labour unions	0·34				
22. Young people	0·61				
23. Conservatives	0·38		0·49		
24. Women's Lib. Movement		0·42			
25. Marijuana users		0·65			
26. Black militants		0·77			
27. Urban rioters		0·77			
28. Civil Rights leaders		0·67			
29. Suburban dwellers	0·71				
30. Workingmen	0·81				−0·36
31. Farmers	0·73				−0·30
32. Ministers leading protest marches		0·75			
33. Middle-class people	0·82				
34. City dwellers	0·77				
Variance component	7·42	4·89	2·79	1·13	0·82
% of total variance	21·8	14·4	8·2	3·3	2·4
Cum % of total variance	21·8	37·2	44·4	47·7	50·2
% of common variance	43·5	28·7	16·4	6·6	4·8
Cum % of common variance	43·5	72·2	88·5	95·2	100·0

Table 3

Varimax rotated factor pattern matrix

Variables	*Factors*				
	1	*2*	*3*	*4*	*5*
1. George Wallace		−0·33	0·43		
2. Richard Nixon			0·75		
3. George McGovern		0·33	−0·30	0·66	
4. Spiro Agnew			0·70		
5. Sargent Shriver				0·61	
6. Big business			0·44		
7. Poor people	0·48				
8. Liberals		0·38		0·41	
9. Southerners	0·47		0·34		
10. Intellectuals	0·33				
11. Catholics	0·44				0·50
12. Radical students		0·68			
13. Policemen	0·39	−0·30	0·44		
14. Protestants	0·55				0·41
15. Jews	0·45				0·63
16. The Military	0·41	−0·30	0·49		
17. Whites	0·60		0·31		
18. Democrats	0·40			0·59	
19. Blacks	0·40	0·33			0·34
20. Republicans			0·71		
21. Labour unions				0·36	
22. Young people	0·58				
23. Conservatives	0·30		0·55		
24. Women's Lib. Movement		0·39			
25. Marijuana users		0·65			
26. Black militants		0·76			
27. Urban rioters		0·77			
28. Civil Rights leaders		0·63			
29. Suburban dwellers	0·63				
30. Workingmen	0·88				
31. Farmers	0·77				
32. Ministers leading protest marches		0·72			
33. Middle-class people	0·80				
34. City dwellers	0·72				
Variance component	5·87	4·17	3·53	1·99	1·50
% of total variance	17·3	12·3	10·4	5·8	4·4
Cum % of total variance	17·3	29·5	39·9	45·7	50·2
% of common variance	34·4	24·4	20·7	11·7	8·8
Cum % of common variance	34·4	58·9	79·6	91·2	100·0

Table 4a

Direct oblimin rotated factor pattern matrix ($\delta = 0 \cdot 0$)

Variables	*Factors*				
	1	*2*	*3*	*4*	*5*
1. George Wallace			0·40		
2. Richard Nixon			0·80		
3. George McGovern					0·70
4. Spiro Agnew			0·75		
5. Sargent Shriver					0·65
6. Big business			0·47		
7. Poor people	0·40				
8. Liberals					0·39
9. Southerners	0·42				
10. Intellectuals				−0·33	
11. Catholics				−0·57	
12. Radical students		0·68			
13. Policemen			0·40		
14. Protestants				−0·46	
15. Jews				−0·73	
16. The Military			0·46		
17. Whites	0·42				
18. Democrats					0·62
19. Blacks				−0·39	
20. Republicans			0·75		
21. Labour unions					0·39
22. Young people	0·51				
23. Conservatives			0·53		
24. Women's Lib. Movement		0·35			
25. Marijuana users		0·66			
26. Black militants		0·80			
27. Urban rioters		0·82			
28. Civil Rights leaders		0·57			
29. Suburban dwellers	0·47			−0·32	
30. Workingmen	0·98				
31. Farmers	0·82				
32. Ministers leading protest marches		0·67			
33. Middle-class people	0·74				
34. City dwellers	0·60				

sufficient variance to justify labelling and, in any case, factor 3 seems to lie between two different dimensions.

The first four factors in the quartimax and varimax solutions correspond fairly closely. Both solutions tend to increase the importance of the later factors at the expense of the first two, although this is more evident for varimax, as is to be expected on theoretical grounds. However, factors 1 and 2 of these rotated solutions do not differ greatly from the first two

Table 4b

Direct oblimin rotated factor structure matrix (δ=0·0)

Variables	*Factors*				
	1	*2*	*3*	*4*	*5*
1. George Wallace		−0·41	0·51		
2. Richard Nixon		−0·37	0·78		−0·33
3. George McGovern		0·44	−0·39		0·72
4. Spiro Agnew		−0·33	0·72		
5. Sargent Shriver		0·38	−0·34		0·68
6. Big business			0·46		
7. Poor people	0·48			−0·38	0·38
8. Liberals		0·45		−0·32	0·53
9. Southerners	0·53		0·45		
10. Intellectuals	0·33			−0·45	
11. Catholics	0·46			−0·66	0·30
12. Radical students		0·71			0·36
13. Policemen	0·49	−0·37	0·57		
14. Protestants	0·61		0·40	−0·62	0·34
15. Jews	0·46			−0·77	
16. The Military	0·53	−0·38	0·61		
17. Whites	0·67	−0·30	0·48	−0·42	
18. Democrats	0·42			−0·38	0·70
19. Blacks	0·38	0·32		−0·57	0·43
20. Republicans	0·41		0·76	−0·30	
21. Labour unions	0·31				0·44
22. Young people	0·57			−0·44	0·33
23. Conservatives	0·40		0·63		
24. Women's Lib. Movement		0·40			0·32
25. Marijuana users		0·67	−0·32		
26. Black militants		0·78			
27. Urban rioters		0·78			
28. Civil Rights leaders		0·67	−0·36		
29. Suburban dwellers	0·66		0·37	−0·55	
30. Workingmen	0·89		0·30	−0·34	0·31
31. Farmers	0·81		0·38	−0·30	
32. Ministers leading protest marches		0·76	−0·38		
33. Middle-class people	0·83		0·40	−0·49	
34. City dwellers	0·73		0·32	−0·57	0·32

factors of the initial solution; the main difference is the reduced importance of variables 1–5 in the rotated solutions. The third and fourth factors in the rotated solutions represent republicans and democrats respectively, having taken over the politicians from factors 1 and 2 of the initial solution. Factor 5 has a religious element in both of these rotated solutions, but otherwise the two solutions are not very similar on this factor.

Table 4c

Inter-factor correlations for direct oblimin solution ($\delta = 0 \cdot 0$)

Factors	*Factors* 1	2	3	4	5
1	1·00	−0·16	0·42	−0·47	0·37
2		1·00	−0·38	−0·18	0·31
3			1·00	−0·15	−0·13
4				1·00	−0·35
5					1·00

Oblique solutions were obtained for delta-values 0·5, 0·3, 0·1, 0, −0·1, −0·3, and −0·5. The solution obtained for $\delta = 0{\cdot}5$ is not tabulated here. This delta-value produced a highly correlated solution, each of the inter-correlations between the factors exceeding 0·6. The solutions obtained for each of the other delta-values employed were very similar and only that for $\delta = 0$ is presented here. This solution exhibits a greater degree of simple structure than the quartimax and varimax solutions although the factors are fairly similar to those obtained by using the orthogonal rotations, with the exception that factor 5 rather than factor 4 represents democrats.

Methods for Confirmatory Factor Analysis

Introduction

The methods discussed above all relate to the basic two-stage approach to factor analysis. In the first stage an initial, usually orthogonal, solution is obtained. Then, in the second stage, the initial solution is subjected to an orthogonal or oblique rotation. The methods of rotation described above attempt to transform the initial pattern matrix into a pattern matrix which conforms to Thurstone's simple structure criteria. Since the simple structure criteria merely specify an overall configuration of factor loadings in the rotated loading matrix this procedure may be regarded as essentially exploratory, particularly when analytical methods of rotation are employed. However, from such exploratory analyses it may be possible to develop a more specific model which can then be tested. This latter type of analysis is usually termed confirmatory factor analysis. Two types of approach to the problem of confirmatory factor analysis are briefly examined below. The first approach involves maximum likelihood estimation of a factor model which includes restrictions specifying hypotheses concerning the parameters of the model. In the second approach the values of certain factor loadings are again specified, but in this case rotational procedures are employed to transform the initial solution into a rotated solution with the required characteristics. This is sometimes termed target rotation (Rummel, 1970, Chapter 17).

Maximum Likelihood Estimation in Confirmatory Factor Analysis

Confirmatory factor analysis involves the specification of a factor model in which some of the parameters of the model have defined values. Thus a model may be specified in which certain elements of $\mathbf{\Lambda}$ and $\mathbf{\Phi}$, and more rarely $\mathbf{\Psi}$, take specified values. Since the simple structure criteria require that some elements of the pattern matrix be zero a model may be developed in which some of the elements of $\mathbf{\Lambda}$ are zero. For such a restricted model in which no parameters in $\mathbf{\Psi}$ are specified, a necessary condition for the model parameters to be unique is that the number of fixed parameters in $\mathbf{\Lambda}$ and $\mathbf{\Phi}$ is at least q^2 (Lawley and Maxwell, 1971, Chapter 7). It may be noted that q constraints are imposed on $\mathbf{\Phi}$ by requiring the factors to have unit variances.

The early attempts at using the method of maximum likelihood to estimate the parameters of a restricted model suffered from the same drawbacks as the early attempts at maximum likelihood estimation in the unrestricted case. Unfortunately, the two-stage procedure used for estimation in the unrestricted case is not appropriate for the restricted model and the likelihood function has to be maximized simultaneously with respect to all parameters (Lawley and Maxwell, 1971, Chapter 7). An improved method of estimation was obtained by Joreskög (1966) although the iterative procedures employed tended to be slow. Joreskög later developed a better iterative procedure (Joreskög, 1969) and extended the procedure to cover a wider range of models. Tests of hypotheses concerning restricted factor models are discussed by Joreskög (1966).

Target Rotation

Methods of target rotation involve the basic two-stage approach to factor analysis but, at the second stage, the choice of transformation matrix is determined by an hypothesized factor pattern. The term "procrustes rotations" has been applied by Hurley and Cattell (1962) to the methods designed to achieve this transformation.

Hurley and Cattell produced a method based on least-squares principles which starts from the relationship between the oblique reference structure matrix and the original orthogonal pattern matrix given in equation (40). The least squares estimate of the transformation matrix $\boldsymbol{T}'$ is

$$\hat{\boldsymbol{T}}' = (\mathbf{\Lambda}'\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'\boldsymbol{\varkappa}_r^*, \tag{51}$$

where $\boldsymbol{\varkappa}_r^*$ is the hypothesized reference structure matrix. Normalizing $\hat{\boldsymbol{T}}'$

by columns produces the procrustes transformation matrix. Equation (51) may be rewritten in terms of the oblique pattern matrix $\boldsymbol{\Lambda}_0$ by virtue of equation (42), as follows:

$$\hat{\boldsymbol{T}}' = (\boldsymbol{\Lambda}'\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Lambda}_0{}^*\boldsymbol{N}. \tag{52}$$

$\boldsymbol{N}$ may be omitted from this equation since it can be taken into account when $\hat{\boldsymbol{T}}'$ is normalized.

The promax method of Hendrickson and White (1964) is an extension of the procrustes method which, according to Lawley and Maxwell (1971, Chapter 6), seems to work well in practice. This method consists of three stages. In the first two stages an orthogonal solution and its varimax rotation are obtained. Then, in the third stage, a modification of the procrustes method is applied to the varimax pattern matrix. In this stage it is assumed that the varimax solution is reasonably close to the hypothesized oblique solution. By raising the elements of the varimax pattern matrix to a suitable integral power greater than unity and preserving the signs of the elements, variables with small loadings on a factor will be forced into that factor's hyperplane, thus improving the approximation to simple structure. Various studies by Hendrickson and White suggest that the power should normally be taken to be equal to 4. If the varimax pattern matrix is denoted by $\boldsymbol{\Lambda}$ and the matrix obtained by powering its elements is denoted by $\boldsymbol{\Lambda}^*$, the least-squares estimate of the transformation matrix using the promax procedure is given by

$$\hat{\boldsymbol{M}}^{-1} = (\boldsymbol{\Lambda}'\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Lambda}^*, \tag{53}$$

where $\hat{\boldsymbol{M}}^{-1}$ and $\hat{\boldsymbol{T}}'$ are related by equation (41).

Discussion

The aim of this paper is to provide a general review of factor rotation in factor analysis. As explained above, factors are rotated for the purpose of explanation, the aim being to find a solution which is parsimonious in the sense that each variable receives a substantial contribution, i.e. loading, from only one factor. However, as Guilford (1952) points out, variables subjected to factor analysis are often factorially complex and consequently this aim is rarely realized. The most common means of specifying parsimonious solutions involves the use of Thurstone's simple structure principles and one of the main problems in factor rotation concerns operationalizing these principles.

Apart from the labelling of factors in the example in the section, Analytical Factor Rotation, for the sake of convenience, and, secondly, any implicit labelling of factors in confirmatory factor analysis, this paper is not concerned with the problem of ascribing meanings to factors. The use of factor analysis to identify specific types of factors has been of great interest to psychologists, but in most cases the problems of sampling have been ignored. Saunders (1948) has identified three types of sampling problem, namely the selection of individuals, the selection of variables and the selection of scores, i.e. the methods of measuring the variables. That solutions depend upon the variables included in the analysis has been demonstrated by Guilford and Zimmerman (1963). In general, however, too little attention has been paid to the problem of variable-selection in theory construction. One exception to this is alpha factor analysis (Kaiser and Caffrey, 1965) which is concerned with generalizing from a sample to a universe of variables. It may be noted that even when "objective" procedures are employed to produce rotated factor solutions the uniqueness of the solution may be in doubt (Saunders, 1960). Much more attention must be paid to these problems if the use of factor analysis in theory-construction is to be justified.

Factor analyses normally involve two stages, the first producing an initial, usually orthogonal, solution and the second producing a rotated solution from the solution obtained at the first stage. Some attempts to combine these stages have been made, for example by Horst (1965, Chapter 19), but it seems unlikely that the two-stage procedure will be abandoned in the foreseeable future.

Acknowledgements

The data utilized in this paper were made available by the Inter-University Consortium for Political Research. The data for the CPS 1972 American National Election Study were originally collected by the Center for Political Studies of the Institute for Social Research, University of Michigan. Neither the original collectors of the data nor the Consortium bear any responsibility for the analyses or interpretations presented here. Miss M. R. Irvine of the University of Manchester kindly set up the data file for the analysis. I would also like to thank Dr R. W. Farebrother of the University of Manchester and an anonymous referee for their comments on previous versions of this paper.

REFERENCES

ANDERSON, T. W. (1958). *An Introduction to Multivariate Statistical Analysis.* Wiley, New York.

ANDERSON, T. W. (1963). Asymptotic theory for principal component analysis. *Annals of Mathematical Statistics*, **34**, 122–48.

ANDERSON, T. W. and RUBIN, H. (1956). Statistical inference in factor analysis. *Proc. 3rd Berkeley Symposium on Math. Statist. and Prob.*, **5**, 111–50. University of California Press.

BARTLETT, M. S. (1950). Tests of significance in factor analysis. *British Journal of Psychology* (*Statistics Section*), **3** (2), 77–85.

BARTLETT, M. (1951). The effect of standardization on a χ^2 approximation in factor analysis. *Biometrika*, **38,** 337–44.

BARTLETT, M. (1954). A note on the multiplying factors for various χ^2 approximations. *Journal of the Royal Statistical Society B*, **16** (2), 296–8.

CARROLL, J. B. (1953). An analytical solution for approximating simple structure in factor analysis. *Psychometrika*, **18** (1), 23–38.

DWYER, P. S. (1939). The contribution of an orthogonal multiple factor solution to multiple correlation. *Psychometrika*, **4** (2), 163–71.

FERGUSON, G. A. (1954). The concept of parsimony in factor analysis. *Psychometrika*, **19** (4), 281–90.

GARNETT, J. C. M. (1919). On certain independent factors in mental measurements. *Proceedings of the Royal Society, London, A*, **96,** 91–111.

GNANADESIKAN, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations.* Wiley, New York.

GUILFORD, J. P. (1952). When not to factor analyze. *Psychological Bulletin*, **49,** 26–37.

GUILFORD, J. P. and ZIMMERMAN, W. S. (1963). Some variable-sampling problems in the rotation of axes in factor analysis. *Psychological Bulletin*, **60** (3), 289–301.

GUTTMAN, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika*, **19** (2), 149–61.

HARMAN, H. H. (1976). *Modern Factor Analysis*, 3rd ed, revised. University of Chicago Press.

HENDRICKSON, A. E. and WHITE, P. O. (1964). Promax: a quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology*, **17** (1), 65–70.

HORST, P. (1941). A non-graphical method for transforming an arbitrary factor matrix into a simple structure factor matrix. *Psychometrika*, **6** (2), 79–99.

HORST, P. (1965). *Factor Analysis of Data Matrices.* Holt, Rinehart and Winston, New York.

HURLEY, J. R. and CATTELL, R. B. (1962). The procrustes program: producing direct rotation to test a hypothesized factor structure. *Behavioural Science*, **7,** 258–62.

INTER-UNIVERSITY CONSORTIUM FOR POLITICAL RESEARCH (1975). *The CPS* 1972 *American National Election Study*, Vols I–III. Inter-University Consortium for Political Research, Michigan.

JENNRICH, R. I. and SAMPSON, P. F. (1966). Rotation for simple loadings. *Psychometrika*, **31** (3), 313–23.

JORESKÖG, K. G. (1966). Testing a simple structure hypothesis in factor analysis. *Psychometrika*, **31** (2), 165–78.

JORESKÖG, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, **32** (4), 443–82.

JORESKÖG, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, **34** (2), 183–202.

KAISER, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, **23** (3), 187–200.

KAISER, H. F. and CAFFREY, J. (1965). Alpha factor analysis. *Psychometrika*, **30** (1), 1–14.

KIM, J.-O. and CURRY, J. (1977). The treatment of missing data in multivariate analysis. *Sociol. Meth. and Res.*, **6** (2), 215–40.

LAWLEY, D. N. (1940). The estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society, Edinburgh*, **60,** 64–82.

LAWLEY, D. N. and MAXWELL, A. E. (1971). *Factor Analysis as a Statistical Method*, 2nd edn. Butterworths, London.

MCDONALD, R. P. (1962). A general approach to nonlinear factor analysis. *Psychometrika*, **27** (4), 397–415.

MCDONALD, R. P. (1967). Factor interpretation in nonlinear factor analysis. *British Journal of Mathematical and Statistical Psychology*, **20** (2), 205–215.

NEUHAUS, J. O. and WRIGLEY, C. (1954). The quartimax method: an analytic approach to orthogonal simple structure. *British Journal of Statistical Psychology*, **7** (2), 81–91.

NIE, N. H., HULL, C. H., JENKINS, J. G., STEINBRENNER, K. and BENT, D. H. (1975). *Statistical Package for the Social Sciences*, 2nd edn. McGraw-Hill, New York.

RIPPE, D. D. (1953). Application of a large sampling criterion to some sampling problems in factor analysis. *Psychometrika*, **18** (3), 191–205.

ROFF, M. (1936). Some properties of the communality in multiple factor theory. *Psychometrika*, **1** (2), 1–6.

RUMMEL, R. J. (1970). *Applied Factor Analysis.* Northwestern University Press.

SAUNDERS, D. R. (1948). Factor analysis I: some effects of chance error. *Psychometrika*, **13** (4), 251–7.

SAUNDERS, D. R. (1953a). An analytic method for rotation to orthogonal simple structure. *Research Bulletin* 53-10. Educational Testing Service, Princeton.

SAUNDERS, D. R. (1953b). An analytical method for rotation to orthogonal simple structure. *American Psychologist*, **8,** 428.

SAUNDERS, D. R. (1960). A computer program to find the best-fitting orthogonal factors for a given hypothesis. *Psychometrika*, **25** (2), 199–205.

THURSTONE, L. L. (1931). Multiple factor analysis. *Psychological Review*, **38,** 406–27.

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

TUCKER, L. R. (1944). A semi-analytical method of factorial rotation to simple structure. *Psychometrika*, **9** (1), 43–68.