[1] Work supported by a fellowship from The Population Council and The Ford Foundation.

A CANONICAL REDUCTION OF THE FACTOR ANALYSIS MODEL

By

Germán Rodríguez[1]

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 992

MARCH 1975

A CANONICAL REDUCTION OF THE FACTOR ANALYSIS MODEL

Germán Rodríguez[1]

University of North Carolina at Chapel Hill

ABSTRACT

The canonical basis of the factor space proposed by Rao (1955) is reconsidered, and a complete canonical reduction of the unrestricted factor analysis model is given. Some results which do not appear to have been given explicitly in the literature are proved, and related to methods for estimating factor scores proposed by Bartlett (1937, 1938) and Thompson (1951).

[1]Work supported by a fellowship from The Population Council and The Ford Foundation.

1. INTRODUCTION

Consider the unrestricted factor analysis model

$$\underset{\sim}{x} = \underset{\sim}{\Lambda}\underset{\sim}{y} + \underset{\sim}{z}, \tag{1.1}$$

where $\underset{\sim}{x}: p \times 1$ is a stochastic vector of responses, $\underset{\sim}{\Lambda}: p \times q$ is a matrix of factor loadings of full column rank $q < p$, $\underset{\sim}{y}: q \times 1$ is a stochastic vector of factors with $E\underset{\sim}{y} = 0$ and $\text{var}(\underset{\sim}{y}) = \underset{\sim}{I}_q$, and $\underset{\sim}{z}: p \times 1$ is a stochastic error vector with $E\underset{\sim}{z} = 0$, $\text{var}(\underset{\sim}{z}) = \underset{\sim}{\Psi} = \text{diag}(\psi_1,\ldots,\psi_p)$ and $\text{cov}(\underset{\sim}{y},\underset{\sim}{z}) = \underset{\sim}{0}$. Under this model $E\underset{\sim}{x} = \underset{\sim}{0}$ and $\text{var}(\underset{\sim}{x}) = \underset{\sim}{V}$ where

$$\underset{\sim}{V} = \underset{\sim}{\Lambda}\underset{\sim}{\Lambda}' + \underset{\sim}{\Psi}. \tag{1.2}$$

It is well known that if $q > 1$ the model is not identified, for if $\underset{\sim}{M}$ is an orthogonal matrix of order q then $\underset{\sim}{\Lambda}$ may be replaced by $\underset{\sim}{\Lambda}\underset{\sim}{M}$ in (1.2). This corresponds to a rigid rotation of the factors $\underset{\sim}{y}$ to $\underset{\sim}{M}'\underset{\sim}{y}$. Since $\underset{\sim}{M}$ has $\frac{1}{2}q(q-1)$ free elements it is clear that $\frac{1}{2}q(q-1)$ independent constraints need be imposed upon the parameters.

A set of restrictions frequently used is that $\underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda}$ be diagonal and its elements be arranged in decreasing order of magnitude. These restrictions lead to factors which are the first q standardized principal components of $\underset{\sim}{x} - \underset{\sim}{z}$ if the responses are scaled so that their residual variances $\underset{\sim}{\Psi}$ are unity, and turn out to be quite convenient in maximum likelihood estimation. See Lawley and Maxwell (1971, Ch. 2 and Ch. 4).

Rao (1955) proposed to select as a basis for the common factor space the canonical variates $\underset{\sim}{y}^*$ of the factors $\underset{\sim}{y}$ with respect to the responses $\underset{\sim}{x}$. Then y_1^* is the linear function of the factors with the largest possible multiple correlation with $\underset{\sim}{x}$, and y_i^* $(i = 2,\ldots,q)$ has the next largest possible multiple correlation with $\underset{\sim}{x}$ subject to its being uncorrelated with $y_1^*,\ldots,y_{i-1}^*$. This has been called the canonical basis of the factor space. Rao (1955) also proposed a method of estimation based on canonical correlation analysis, called canonical factor analysis, and indicated that the method was equivalent to maximum likelihood estimation.

Several authors have indicated that the set of restrictions that $\underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda}$ be diagonal leads precisely to the canonical basis of the factor space. This result is implicit in Rao's work and is consistent with the equivalence of maximum likelihood and canonical estimation; see also McKeon (1964). No explicit proof, however, appears to have been given in the literature.

In this note we study this subject in some detail and derive a complete canonical reduction of the unrestricted factor analysis model. In this process we find the canonical variates $\underset{\sim}{x}^*$ of the responses with respect to the factors, and indicate how they may be used in analyzing a factor model. The case of canonical loadings is considered next, and the relationship between the set of restrictions that $\underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda}$ be diagonal and the canonical basis is clearly shown. Finally, the first q canonical variates of the responses with respect to the canonical factors are given explicitly, and related to methods for estimating factor scores proposed by Bartlett (1937, 1938) and Thompson (1951).

In our work we assume that the parameters of the factor model are known or have been estimated by the maximum likelihood method as described by Jöreskog (1967). The further computations required can be done using existing computer programs for canonical correlation analysis.

The following well-known identities will be useful. If $\underset{\sim}{V}$ satisfies (1.2) then

$$\underset{\sim}{V}^{-1} = \underset{\sim}{\Psi}^{-1} - \underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda}(\underset{\sim}{I} + \underset{\sim}{\Delta})^{-1}\underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1} \tag{1.3}$$

$$V^{-1}\Lambda = \Psi^{-1}\Lambda(I + \Delta)^{-1}, \text{ and} \tag{1.4}$$

$$\Lambda' V^{-1}\Lambda = \Delta(I + \Delta)^{-1}, \tag{1.5}$$

where $\Delta = \Lambda'\Psi^{-1}\Lambda$. (See for example Lawley and Maxwell (1971, p. 27).)

2. THE CANONICAL REDUCTION OF THE FACTOR MODEL

From (1.1), the joint variance-covariance matrix of x and y is

$$\text{var}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} V & \Lambda \\ \Lambda' & I_q \end{pmatrix} . \tag{2.1}$$

From canonical correlation theory (see for example Morrison (1967, Ch. 6)) we know that there exist linear transformations $x^* = L'x$ and $y^* = M'y$ such that

$$\text{var}\begin{pmatrix} x^* \\ y^* \end{pmatrix} = \begin{pmatrix} I_p & \Gamma \\ \Gamma' & I_q \end{pmatrix} , \tag{2.2}$$

where $\Gamma: p \times q = \begin{pmatrix} P \\ 0 \end{pmatrix}$ and $P = \text{diag}(\rho_1,\ldots,\rho_q)$. The ρ_i are the canonical correlations between the responses and factors, and x^*, y^* are the corresponding canonical variates.

Furthermore, ρ_i^2 is the i-th largest characteristic root of $\Lambda'V^{-1}\Lambda$, L is a matrix of eigenvectors of $V^{-1}\Lambda\Lambda'$ standardized so that $L'VL = I_p$, and M is a matrix of orthonormal eigenvectors of $\Lambda'V^{-1}\Lambda$. Note that y^* is a rigid rotation of y.

The canonical variates x^* and y^* have an interesting property. Since $x^* = L'x$, using (1.1) to write x in terms of y and z we have

$$x^* = L'\Lambda y + L'z;$$

but $y = My^*$, since $M^{-1} = M'$ by orthogonality, and hence

$$\underset{\sim}{x}^* = \underset{\sim}{L}'\underset{\sim}{\Lambda}\underset{\sim}{M}\underset{\sim}{y}^* + \underset{\sim}{L}'\underset{\sim}{z}.$$

Now $\underset{\sim}{L}'\underset{\sim}{\Lambda}\underset{\sim}{M}$ is $\text{cov}(\underset{\sim}{x}^*,\underset{\sim}{y}^*)$, which by (2.2) is $\underset{\sim}{\Gamma}$. Therefore

$$\underset{\sim}{x}^* = \underset{\sim}{\Gamma}\underset{\sim}{y}^* + \underset{\sim}{z}^*, \tag{2.3}$$

where $\underset{\sim}{z}^* = \underset{\sim}{L}'\underset{\sim}{z}$. Clearly $\text{cov}(\underset{\sim}{y}^*,\underset{\sim}{z}^*) = \underset{\sim}{0}$ and

$$\text{var}(\underset{\sim}{x}^*) = \underset{\sim}{\Gamma}\underset{\sim}{\Gamma}' + \underset{\sim}{L}'\underset{\sim}{\Psi}\underset{\sim}{L},$$

but from (2.2), $\underset{\sim}{\Gamma}\underset{\sim}{\Gamma}' = \text{diag}(\rho_1^2,\ldots,\rho_q^2,0,\ldots,0)$ and $\text{var}(\underset{\sim}{x}^*) = \underset{\sim}{I}_p$. Hence $\text{var}(\underset{\sim}{z}^*) = \underset{\sim}{L}'\underset{\sim}{\Psi}\underset{\sim}{L} = \underset{\sim}{\Psi}^*$, say, is a diagonal matrix,

$$\underset{\sim}{\Psi}^* = \text{diag}(1 - \rho_1^2,\ldots,1 - \rho_q^2,1,\ldots,1). \tag{2.4}$$

These results imply that (2.3) is a factor model, for the transformed responses $\underset{\sim}{x}^*$ are written as linear combinations of the factors $\underset{\sim}{y}^*$ plus uncorrelated random errors $\underset{\sim}{z}^*$. In view of the structure of the loadings $\underset{\sim}{\Gamma}$, this model has the property that x_i^* is loaded only on factor y_i^* for $i = 1,\ldots,q$, and is independent of the factors for $i = q + 1,\ldots,p$. Furthermore the loading of x_i^* on y_i^* for $i = 1,\ldots,q$ is the i-th largest canonical correlation between the responses and the factors. Thus, we have reduced the general model (1.1) to a particularly simple structure (2.3). This will be called the <u>canonical reduction of the factor model.</u>

It might be noted that the term "canonical" is used here both in the sense of a reduction of a model to a simple form, as in Anderson (1958, pp. 224-6), and in the sense of canonical correlation analysis. In this case the proposed canonical reduction is based on canonical variates.

Let $\underset{\sim}{x}^*$ be partitioned into vectors $\underset{\sim}{x}_1^* = (x_1^*,\ldots,x_q^*)'$ and $\underset{\sim}{x}_2^* = (x_{q+1}^*,\ldots,x_p^*)'$. Then $\underset{\sim}{x}_1^*$ represents those features of the responses that are explained by the factors. Let $\underset{\sim}{L}: p \times p$ be partitioned into matrices $\underset{\sim}{L}_1: q \times q$ and $\underset{\sim}{L}_2: p \times (p - q)$. Then $\underset{\sim}{x}_1^* = \underset{\sim}{L}_1'\underset{\sim}{x}$ and $\underset{\sim}{x}_2^* = \underset{\sim}{L}_2'\underset{\sim}{x}$. The matrix $\underset{\sim}{L}_1$ may be obtained from $\underset{\sim}{M}$ as

$$\underset{\sim}{L}_1 = \underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim}{M}\underset{\sim}{P}^{-1}. \tag{2.5}$$

To see this note that $\underset{\sim}{L}_1$ must satisfy the eigenvector-equation

$$\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim}{\Lambda}'\underset{\sim}{L}_1 = \underset{\sim}{L}_1\underset{\sim}{P}^2, \tag{2.6}$$

while $\underset{\sim}{M}$ in turn satisfies the eigenvector-equation

$$\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim}{M} = \underset{\sim}{M}\underset{\sim}{P}^2. \tag{2.7}$$

Substituting (2.5) for $\underset{\sim}{L}_1$ in the left-hand side of (2.6) we obtain

$$\begin{aligned}\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim}{\Lambda}'\underset{\sim}{L}_1 &= \underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim}{M}\underset{\sim}{P}^{-1} \\ &= \underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim}{M}\underset{\sim}{P} \quad \text{by (2.7)}, \\ &= \underset{\sim}{L}_1\underset{\sim}{P}^2 \text{ by (2.5);}\end{aligned}$$

hence (2.6) is satisfied by the proposed choice of $\underset{\sim}{L}_1$. We must also show that (2.5) gives standardized canonical variates $\underset{\sim}{x}_1^*$. Now (2.7) can also be written

$$\underset{\sim}{M}'\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim}{M} = \underset{\sim}{P}^2; \tag{2.8}$$

using (2.5) for $\underset{\sim}{L}_1$ we have

$$\begin{aligned}\underset{\sim}{L}_1'\underset{\sim}{V}\underset{\sim}{L}_1 &= \underset{\sim}{P}^{-1}\underset{\sim}{M}'\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim}{M}\underset{\sim}{P}^{-1} \\ &= \underset{\sim}{P}^{-1}\underset{\sim}{P}^2\underset{\sim}{P}^{-1} \text{ by (2.8)} \\ &= \underset{\sim}{I}_q .\end{aligned}$$

The canonical variates $\underset{\sim}{x}_2^*$ are also of interest, because they represent those features of the responses which are <u>not</u> explained by the factors, and thus may be useful in fitting and interpreting factor models.

For given $\underset{\sim}{\Lambda}$ and $\underset{\sim}{\Psi}$, the canonical variates $\underset{\sim}{x}^*$ and $\underset{\sim}{y}^*$ can easily be computed using any computer program for canonical correlation analysis with (2.1) as the basic input matrix.

3. THE FACTOR MODEL WITH CANONICAL LOADINGS

Suppose now that following Rao (1955), we define the factors $\underset{\sim}{y}$ as the canonical variates of the factor space with respect to the response space. In terms of our analysis in §2, this implies that $\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}$ must be diagonal with its elements arranged in decreasing order of magnitude, for then the diagonal matrix of eigenvalues $\underset{\sim}{P}^2$ is $\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}$ itself, and the matrix of orthonormal eigenvectors $\underset{\sim}{M}$ is $\underset{\sim}{I}_q$, indicating that no rotation of $\underset{\sim}{y}$ is required to obtain the canonical variates $\underset{\sim}{y}^*$ (i.e. $\underset{\sim}{y}^* = \underset{\sim}{y}$).

In view of (1.5), it is clear that a sufficient condition for $\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}$ to be diagonal is that $\underset{\sim}{\Delta} = \underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda}$ be diagonal. In this case

$$\underset{\sim}{P}^2 = \underset{\sim}{\Delta}(\underset{\sim}{I} + \underset{\sim}{\Delta})^{-1}, \text{ i.e. } \rho_i^2 = \frac{\delta_i}{1 + \delta_i} \quad (i = 1,\ldots,q) \qquad (3.1)$$

and if the δ_i are ordered so are the ρ_i^2. This proves that the usual set of restrictions that $\underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda}$ be diagonal and its elements be ordered does indeed lead to the canonical basis of the factor space.

Let us now consider the canonical variates for the $\underset{\sim}{x}$ set. As noted before, these can be obtained by conducting a canonical correlation analysis of matrix (2.1). An explicit expression for $\underset{\sim}{x}_1^*$, however, can be given. Using (2.5) with $\underset{\sim}{M} = \underset{\sim}{I}_q$ we find

$$\begin{aligned}\underset{\sim}{L}_1 &= \underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim}{P}^{-1} \\ &= \underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}(\underset{\sim}{I} + \underset{\sim}{\Delta})^{\frac{1}{2}}\underset{\sim}{\Delta}^{-\frac{1}{2}} \text{ by (3.1)}, \\ &= \underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda}(\underset{\sim}{I} + \underset{\sim}{\Delta})^{-\frac{1}{2}}\underset{\sim}{\Delta}^{-\frac{1}{2}} \text{ by (1.4)}.\end{aligned}$$

Thus,

$$x_1^* = \Delta^{-\frac{1}{2}}(I + \Delta)^{-\frac{1}{2}}\Lambda'\Psi^{-1}x. \tag{3.2}$$

If a battery of tests measuring q canonical factors is applied, the best standardized score on each factor is given by (3.2).

Thompson (1951) has considered the problem of estimating the factor scores given a sample of observations on x and has proposed the estimator

$$\hat{y}_1 = (I + \Delta)^{-1}\Lambda'\Psi^{-1}x, \tag{3.3}$$

which has the property of minimizing the variance of the residuals $y - \hat{y}_1$. Bartlett (1937, 1938), on the other hand, has proposed the estimator

$$\hat{y}_2 = \Delta^{-1}\Lambda'\Psi^{-1}x, \tag{3.4}$$

which has the property of minimizing the sum of squares of standardized residuals. For details see Lawley and Maxwell (1971, Ch. 8).

When Δ is diagonal, these estimators differ from the canonical variates x_1^* only by a scaling factor. Thus for all three methods the squared correlations between the estimators and the factor scores are the elements of $\Delta(I + \Delta)^{-1}$, which are also the squared canonical correlations between factors and responses.

The estimators $\hat{y}_1$ and $\hat{y}_2$ can be shown to be given by the following factor models:

$$\hat{y}_1 = P^2 y + z_1, \tag{3.5}$$

where $\text{var}(z_1) = \Psi_1 = \Delta(I + \Delta)^{-2}$ and is diagonal; and

$$\hat{y}_2 = y + z_2, \tag{3.6}$$

where $\text{var}(z_2) = \Psi_2 = \Delta^{-1}$ and is diagonal.

To obtain these results note that using (1.1) to write $\underset{\sim}{x}$ in terms of $\underset{\sim}{y}$ and $\underset{\sim}{z}$, we have

$$\hat{\underset{\sim}{y}}_1 = (\underset{\sim}{I} + \underset{\sim}{\Delta})^{-1}\underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda}\underset{\sim}{y} + (\underset{\sim}{I} + \underset{\sim}{\Delta})^{-1}\underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1}\underset{\sim}{z}$$

$$= (\underset{\sim}{I} + \underset{\sim}{\Delta})^{-1}\underset{\sim}{\Delta}\underset{\sim}{y} + \underset{\sim}{z}_1$$

$$= \underset{\sim}{P}^2\underset{\sim}{y} + \underset{\sim}{z}_1,$$

where

$$\text{var}(\underset{\sim}{z}_1) = (\underset{\sim}{I} + \underset{\sim}{\Delta})^{-1}\underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda}(\underset{\sim}{I} + \underset{\sim}{\Delta})^{-1}$$

$$= \underset{\sim}{\Delta}(\underset{\sim}{I} + \underset{\sim}{\Delta})^{-2}.$$

Similarly,

$$\hat{\underset{\sim}{y}}_2 = \underset{\sim}{\Delta}^{-1}\underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda}\underset{\sim}{y} + \underset{\sim}{\Delta}^{-1}\underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1}\underset{\sim}{z}$$

$$= \underset{\sim}{y} + \underset{\sim}{z}_1,$$

where

$$\text{var}(z_1) = \underset{\sim}{\Delta}^{-1}\underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda}\underset{\sim}{\Delta}^{-1} = \underset{\sim}{\Delta}^{-1} .$$

Both structures are pleasingly simple, particularly that of Bartlett's estimators. They do not provide, however, a complete canonical reduction of the factor model.

ACKNOWLEDGEMENT

The author would like to express his appreciation to Professors N. L. Johnson and D. Quade for their comments on the manuscript.

REFERENCES

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

Bartlett, M. S. (1937). The statistical conception of mental factors. British J. Psychology 28, 97-104.

_____________ (1938). Methods of estimating mental factors. Nature 141, 609-610.

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

Lawley, D. N. and Maxwell, A. E. (1971). Factor Analysis as a Statistical Method. New York: American Elsevier.

McKeon, J. J. (1964). Canonical Analysis: some relations between canonical correlation, factor analysis, discriminant function analysis and scaling theory. Psychometric Monographs No. 13.

Morrison, D. F. (1967). Multivariate Statistical Methods. New York: McGraw.

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