DOCUMENT RESUME

ED 395 039 TM 025 059

AUTHOR Hester, Yvette
TITLE Mathematical Models Underlying Common Factor Analysis: An Introductory Primer.
PUB DATE 27 Jan 96
NOTE 22p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (New Orleans, LA, January 25-27, 1996).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Factor Analysis; Factor Structure; Heuristics; *Mathematical Models; *Matrices; Research Methodology
IDENTIFIERS *Data Reduction Methods; Statistical Analysis System; Statistical Package for the Social Sciences

ABSTRACT

Data reduction techniques seek to combine variables that account for patterns of variation in observed dependent variables in such a way that a simpler model is available for analysis. Factor analysis is a data reduction technique that attempts to model or explain a set of variables in terms of their associations. To understand why this technique yields an accurate analysis, an examination of the mathematical models underlying the procedure is necessary. Execution of factor analysis by the Statistical Analysis System and the Statistical Package for the Social Sciences will then not be a "black box." Mathematical models underlying true factor analysis and principal components analysis are presented and discussed. An explanation of the terms and basic differences is given in terms of the mathematical models. A small, heuristic example is included to illustrate the concepts and matrix algebra procedures involved in the factor analysis data reduction technique. An appendix presents commands for the MAPLE computer algebra system. (Contains 2 tables, 2 figures, and 10 references.) (Author/SLD)

Mathematical Models Underlying Common Factor Analysis: An Introductory Primer

Yvette Hester
Texas A&M University

Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans, LA, January 27, 1996

Abstract

Data reduction techniques seek to combine variables that account for patterns of variation in observed dependent variables in such a way that a simpler model is available for analysis. Factor analysis is a data reduction technique that attempts to model or explain a set of variables in terms of their associations. To understand why this technique yields an accurate analysis, an examination of the mathematical models underlying the procedure is necessary. Execution of factor analysis by SAS and SPSS will then not be a "black box". Mathematical models underlying true factor analysis and principal components analysis are presented and discussed. An explanation of terms and basic differences is given in terms of the mathematical models. A small, heuristic example to illustrate the concepts and matrix algebra procedures involved in the factor analysis data reduction technique is included.

Data Reduction Techniques

Data reduction techniques discussed in this paper are generalized regression-like techniques. In regression, the decision to keep regressors in the model is based on finding the "smallest-largest" subset of regressors--smallest in the sense that costs associated with a large number of variables should be minimized; largest in the sense that enough variables need to be retained for reliable predictions to maximize variance accounted for by the variables. There is no one statistical procedure to find this best subset of variables and personal judgment is required, as it is in all statistical analysis (Seber, 1977). To illustrate the magnitude of the number of possible regressions for a given situation, suppose that there are k possible regressors. Since each regressor is either in the equation or not, there are 2^k possible such regressions. If k is large, 2^k becomes extremely large, quickly, e.g., $2^{15} = 32,768$.

Methods for Selection of Subsets

The type of method used to select a regression subset or reduce the data varies based on the type of analysis performed. Specific methods discussed in the present paper are common (principal) factor analysis, principal components analysis and principal components factor analysis. Figure 1 shows the relationship between these methods and other common methods such as confirmatory factor analysis, exploratory factor analysis and maximum-likelihood factor analysis. Principal components factor analysis is a combination of the two primary methods, principal components analysis and common factor analysis. The other three methods, confirmatory factor analysis, exploratory factor analysis and maximum-likelihood factor analysis are considered types of true factor analysis. Maximum-likelihood analysis is frequently employed within both confirmatory and exploratory factor analysis.

Differences between principal components analysis and common factor analysis are illustrated here during a general explanation of factor analysis and a small, heuristic example of principal factor analysis is presented. Procedures involving the factor analysis

model as a basis have been more widely used and are generally better developed (Velicer & Jackson, 1990).

All generalized regression, data reduction techniques estimate parameters of regression-like linear models. Indeed, since canonical correlation analysis subsumes all parametric statistic methods (e.g., ANOVA, t-tests, discriminant analysis) as special cases (Knapp, 1978), and since canonical correlation analysis invokes a principal components analysis as part of its mathematics (Thompson, 1984), therefore all parametric methods implicitly invoke some kind of factor analytic logic.

Factor analytic techniques are multivariable, like the reality being modeled, and can be understood through the mathematics of matrix algebra. Each seeks a way to combine variables that accounts for patterns of variation in the observed dependent variables. This yields a simpler model, making further analysis less complicated. The following discussion is an introduction to factor analysis with similarities and differences to principal components analysis highlighted.

Principal (Common) Factor Analysis versus Principal Components Analysis

Factor analysis, like principal components analysis, regresses standardized observed variables on a set of unobserved factors. Factors are the underlying components or dimensions for which estimates of values are obtained. Factor analysis is a statistical model that includes unique, uncorrelated error terms, whereas principal components analysis is simply a mathematical transformation of data (Hamilton, 1992). Factor analysis attempts to model each of k standardized observed variables z_k as a linear combination of j unobserved factors F_j, where j < k, along with an error term for each observed variable, u_k. The factors F_j are common factors, since each of the observed variables z_k is written in terms of these factors. The error term u_k is called the unique factor, as each observed variable has it's own uniquely determined residual. In general, the linear function for each z_k has appearance

$$(1) \qquad z_k = l_{k1}F_1 + l_{k2}F_2 + \ldots + l_{kj}F_j + u_k .$$

The l_{kj} are equivalent to standardized regression coefficients and are called factor loadings. If there is only one factor or if the factors are all orthogonal (uncorrelated), the factor loadings are equivalent to the correlation between that factor and the observed variable. The matrix notation for this model is

$$(1)' \qquad Z = FL' + U,$$

where L' represents the transpose of the matrix containing the factor loadings.

Principal components analysis is a mathematical transformation of the data on the k observed variables represented by k principal components or factors. There is no unique factor or error term since k principal components will exactly explain all the variance of k observed variables (Hamilton, 1992). Principal components analysis is simpler mathematically than factor analysis and is a mathematical maximization procedure that uses uncorrelated linear functions. The linear function for principal components analysis is:

$$(2) \qquad z_k = l_{k1}F_1 + l_{k2}F_2 + \ldots + l_{kk}F_k .$$

Model (2) is similar to model (1) without the u_k term and j = k. The matrix equation (2)' will look like (1)' without the U matrix.

$$(2)' \qquad Z = FL' .$$

Principal components factor analysis is a combination of true factor analysis and principal components analysis in that if less than k factors explain a large amount of the variance of the observed variables, those factors will be used and an error term v_k will be introduced to represent the shared residual for each linear combination of factors. The difference between the v_k error terms and the u_k error terms for the true factor analysis

model is that the v_k are not unique, i.e., have nonzero correlation. Principal components factor analysis yields a principal components factor model that resembles the true factor analysis model and is

$$(3) \qquad z_k = l_{k1}F_1 + l_{k2}F_2 + \ldots + l_{kj}F_j + v_k,$$

where $v_k = l_{k,j+1}F_{j+1} + l_{k,j+2}F_{j+2} + \ldots + l_{k,k}F_k$. The v_k linear function shows that these residuals cannot be uncorrelated as in the true factor analysis model (Hamilton, 1992). For the remainder of this paper, the term principal components analysis will refer to model (3), since most researchers combine true factor analysis and principal components analysis into this model. The controversy regarding the similarities and differences between these three techniques is a lengthy issue. This paper will address only obvious differences in the representations of the mathematical models. For more extensive discussions of Component Analysis versus Common Factor Analysis see the January (1990) issue of the Journal of the Society of Multivariate Experimental Psychology, Multivariate Behavioral Research.

Factor analysis centers on attempting to explain a set of observed variables in terms of their correlations. Principal components analysis centers on attempting to explain a set of observed variables in terms of their variance. The decision as to which of the methods to use in an analysis in not clear-cut, especially as these methods produce similar results when applied to strongly correlated data. Confusion with terminology and computer packages further complicate the choices, as principal components is typically listed as an option within a factor analysis computer package. Principal components analysis is the default method for extraction in SPSS. If true factor analysis is desired, the researcher must indicate another method of extraction, e.g., principal axis factoring (Pedhazur & Smelkin, 1991). Component analysis will typically involve less computer processing time.

"Principal components appeals more to a 'data analysis' perspective, whereas factor analysis fits better with a 'model building' approach," as Hamilton (1992, p. 252) noted. The goal of both types of analyses is to find subsets of variables that are both highly correlated and weakly (or not at all) correlated with each other. Patterns for how the variables cluster are determined.

The goal of data reduction examines output in terms of which factors to retain in the model. Each factor will have an associated eigenvalue, denoted λ (lambda), to help in determining retention. Mathematically, eigenvalues are the roots of the characteristic polynomial associated with a given matrix. In the data reduction techniques, eigenvalues represent the variances of the original components. In principal components analysis, since k components explain k standardized variables, the sum of the eigenvalues will equal the number of variables. A component that has an eigenvalue of less than one will account for less than a single variables' variation since each standardized variable has variance of one. Thus, for principal components analysis, components with $\lambda \geq 1$ are retained in the model. For true factor analysis, eigenvalues are typically smaller and the eigenvalue greater than one criterion is inappropriate and not as useful (Pedhazur & Smelkin, 1991).

An analyst must bear in mind that these are simply recommendations and a large amount of subjectivity and thought are required when making these complex decisions. Substantive issues must be considered in the specific context of each particular research situation.

Screeplots can be helpful to get an overview of the data. A screeplot is a plot of eigenvalues in descending order plotted against the factor number. As the slope of the lines between points becomes less steep or smaller, a leveling off becomes apparent. A clear break in the slopes, i.e., where they begin to approach a horizontal line, will help a researcher determine a useful or natural cut-off for contributing factors.

Since true factor analysis centers on correlations between the original observed variables, a correlation matrix **R** must be obtained. The original observed variables are first standardized and placed in a data matrix **Z**. To form the original correlation matrix **R**, the vectors (columns) of **Z** must be normalized by taking the product **A** = $\frac{1}{\sqrt{n-1}}Z$. This scalar multiple takes out the $\sqrt{n-1}$ factor introduced by the standard deviation in the computation of the z-scores. The correlation matrix is the product of the matrix **A** and it's transpose A'. Thus, **R** = **A** '**A** and has ones on its major diagonal.

Factor analysis uses a modified correlation matrix, **R***, which has estimates of the proportion of a variables explained variance on the major diagonal instead of ones, like the original **R**. Principal components analysis does not involve this reduction of the variance on the diagonal elements. The reduced variance terms are referred to as communality and are denoted h_k^2. These values represent the proportion of variance explained by the extracted factors. The predictor is taken as the dependent variable and the factors are taken as the independent variables. One approach to estimating these values uses the coefficient of determination R_k^2. These values would appear on the major diagonal of **R*** as the initial estimates of h_k^2. Each R_i^2, $1 \le i \le k$, is determined by regressing the i th standardized observed variable z_i on the remaining standardized observed variables $z_1, z_2, \ldots, z_{i-1}, z_{i+1}, \ldots, z_k$. Matrix algebra allows this computation by first forming R^{-1}, the inverse of the correlation matrix **R**. Take the diagonal entries of R^{-1}, invert them and subtract them from 1. The resulting values become the initial estimates of the communalities h_k^2 and are placed on the diagonal of **R***. Thus, for a $k \times k$ correlation matrix **R**, the initial **R*** is given by

$$\mathbf{R}^* = R - I + \left[diag(R^{-1})\right]^{-1}, \text{ where } \mathbf{I} \text{ is the } k \times k \text{ identity matrix.}$$

The initial modified correlation matrix has the form

$$R^* = \begin{bmatrix} R_1^2 & r_{12} & \cdots & r_{1k} \\ r_{21} & R_2^2 & \cdots & r_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ r_{k1} & r_{k2} & \cdots & R_k^2 \end{bmatrix}.$$

It follows that $\mathbf{R} = \mathbf{R}^* + \mathbf{Q}$ where $\mathbf{Q}$ is the diagonal matrix containing the error terms u_n on the major diagonal and zeros everywhere else.

The initial factor loadings are derived by extracting the principal eigenvalues and forming the corresponding eigenvectors $e_1,\ldots,e_j$, $1 \le j \le k$, each having norm (length) one. Set $a_n = \sum_{i=1}^{j} e_{in}^2$, $n = 1,\ldots,k$. Then each a_n will be the sum of squares of the nth entries in each of the eigenvectors and $h_n^2 = 1 - u_n = a_n$.

Once the initial factor loadings are derived, the new $\mathbf{R}^*$ is given by

$$\mathbf{R}^* = R - I + \begin{pmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & a_k \end{pmatrix},$$

the new estimates for the communalities on the main diagonal. Iterations are performed until the communality estimates are stable.

Principal components first determines the factor loadings l_{ki} to compute the communalities directly. The residual is then found by $v_k = 1 - h_k^2$. Principal components analysis performs no iterations of any kind and does not begin with estimates. Some researchers are uncomfortable with the estimation involved in the true factor analysis procedure (Stevens, 1992), along with other objections with respect to multicollinearities (Hawkins, 1973).

Each factor can now be expressed in terms of the original variables. During the process of attempting to combine variables, composites called factor scores are formed. These scores derive from the coefficients found by the regression of factors on the observed variables. Factor scores are estimates of the factors and can be found by first computing the eigenvalues and corresponding eigenvectors for the last **R***. The first factor F_1 will be a linear combination of the original variables that explains the most variance. This will be the factor that has the largest eigenvalue. The second factor F_2 will have the second largest eigenvalue, and so on. Each factor can be expressed as

$$(4) \qquad F_n = e_{n1}z_1 + \ldots + e_{nk}z_k,\ n = 1, \ldots, j,$$

where $\sum_{j=1}^{k} e_{ji}^2 = 1$. In matrix notation $F_n = Ze_n,\ 1 \le n \le j$ where $e_n' e_n = 1$ and $e_j' e_m = 0$, for all $j < m$, since each component is uncorrelated with every other component (Hamilton, 1992).

Factor scores replace the original observed scores and can be analyzed or interpreted like any other variable through regression, etc., as a subsequent analysis. If the same number of factors are retained from both factor analysis and principal components analysis, and when the factors are well-defined, highly similar results are expected from the two methods. Velicer and Jackson (1990) report a correlation of .99 or better between alternative types of scores in this situation. Even when loadings were low and factors were poorly defined with few variables per factor, correlations were .9 or more. "Improvements in the quality of the data increased the degree of similarity" (Velicer & Jackson, 1990, p. 6). Some of the observed differences between the two methods are thought to be the result of overextraction of the number of components by the Kaiser rule; the default in many computer programs that employ principal components analysis. Maximum-likelihood factor analysis done with large sample sizes can also cause problems

with overextraction (Tucker & Lewis, 1973). This test assumes a multivariate normal distribution. Zwick and Velicer (1986) provide more on this topic.

A rotation of the factor loadings is sometimes required to simplify the factor structure and make the factors more interpretable. If only one factor is retained in the model, then rotation is ignored. Mathematically, rotation is a transformation or a rotating of the axes represented by the factors about the origin that enables variables to load more strongly or polarize on a single factor. Orthogonal rotation holds the factor axes perpendicular, i.e., keeps them uncorrelated during rotation. The factor loading matrix **L**, having columns $\sqrt{\lambda_1}\, l_1, \ldots, \sqrt{\lambda_j}\, l_j$, is multiplied by an orthogonal transformation matrix **M**, to obtain a new factor loading matrix **L***, where $\mathbf{L^* = LM}$. Then a least squares-like procedure is invoked. See Gorsuch (1983) for a discussion of various rotations.

Oblique rotation permits acute angles (correlation) between the factor axes. This type of rotation permits further polarization and involves a nonorthogonal matrix transformation represented by the matrix equation $\mathbf{L^{**} = L^*P}$, where $\mathbf{L^{**}}$ is the matrix of new factor loadings and **P** is the nonorthogonal transformation matrix. Oblique rotation is more complex than orthogonal rotation and somewhat arbitrary, but since the loadings are further polarized, it provides easier interpretation. An analyst should use different rotation methods and examine the results. If different methods reach the same results, conclusions can be considered stable (Hamilton, 1992). The two types of rotation "reflect different frames of reference in viewing phenomena" (Pedhazur & Smelkin, p. 615). Communalities are not affected by rotation or type of rotation.

Example

Suppose a survey of 5 questions concerning treatment by peers was given to 10 lecturers in a certain department at a large university. The questions are listed in Table 1.

Insert Table 1 about here

The responses are recorded in a raw data matrix X, where a negative response is coded as 0 and a positive response is coded as 1.

$$X := \begin{bmatrix} 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

The matrix is entered into a MAPLE session (a computer algebra system), after loading the linear algebra package and setting the digits to 6. The statistics package is also loaded. The exact MAPLE commands for this example are listed in appendix A.

A matrix of z-scores needs to be computed. The mean and standard deviation of each column of X is listed in Table 2. Exact arithmetic was used throughout all computations and then converted to 6 decimal places as each of the matrices needed to be examined. Only the decimal representations of the matrices are given in this paper, but the MAPLE commands for both the exact arithmetic matrices and the decimal representations are listed in appendix A.

Insert Table 2 about here

The matrix Z of standardized variables is

$$Z = \begin{bmatrix} -.474342 & -.948684 & .621059 & -1.44914 & .316228 \\ -.474342 & -.948684 & .621059 & .621059 & .316228 \\ -.474342 & .948684 & -1.44914 & -1.44914 & -2.84605 \\ 1.89737 & .948684 & .621059 & .621059 & .316228 \\ -.474342 & -.948684 & .621059 & -1.44914 & .316228 \\ -.474342 & .948684 & -1.44914 & .621059 & .316228 \\ 1.89737 & .948684 & -1.44914 & .621059 & .316228 \\ -.474342 & .948684 & .621059 & .621059 & .316228 \\ -.474342 & -.948684 & .621059 & .621059 & .316228 \\ -.474342 & -.948684 & .621059 & .621059 & .316228 \end{bmatrix}$$

Each column in **Z** must be normalized (given length 1) so that the correlation matrix **R** can be computed. This normalized data is listed in a matrix **A**.

$$A := \begin{bmatrix} -.158114 & -.316228 & .207020 & -.483046 & .105409 \\ -.158114 & -.316228 & .207020 & .207020 & .105409 \\ -.158114 & .316228 & -.483046 & -.483046 & -.948684 \\ .632456 & .316228 & .207020 & .207020 & .105409 \\ -.158114 & -.316228 & .207020 & -.483046 & .105409 \\ -.158114 & .316228 & -.483046 & .207020 & .105409 \\ .632456 & .316228 & -.483046 & .207020 & .105409 \\ -.158114 & .316228 & .207020 & .207020 & .105409 \\ -.158114 & -.316228 & .207020 & .207020 & .105409 \\ -.158114 & -.316228 & .207020 & .207020 & .105409 \end{bmatrix}$$

To obtain **R**, find A' (**A** transpose) and form the matrix product $\mathbf{R} = A'A$.

$$R := \begin{bmatrix} 1.00000 & .500000 & -.218219 & .327326 & .166666 \\ .500000 & 1.00000 & -.654658 & .218221 & -.333334 \\ -.218219 & -.654658 & .999999 & .0476189 & .509178 \\ .327326 & .218221 & .0476189 & .999998 & .509177 \\ .166666 & -.333334 & .509178 & .509177 & 1.00000 \end{bmatrix}$$

To obtain the initial estimates of the communalities or shared variances, R^{-1} (the inverse of R) is computed.

$$Rinv := \begin{bmatrix} 1.60000 & -1. & 0 & 0 & -.600000 \\ -1. & 2.70833 & 1.14565 & -.763767 & .875000 \\ 0 & 1.14565 & 2.10000 & 0 & -.687389 \\ 0 & -.763767 & 0 & 1.75000 & -1.14565 \\ -.600000 & .875000 & -.687389 & -1.14565 & 2.32500 \end{bmatrix}$$

The diagonal entries are inverted and subtracted from 1. The resulting values become the new entries along the diagonal of **R***.

$$Rstar := \begin{bmatrix} .375000 & .500000 & -.218219 & .327326 & .166666 \\ .500000 & .630769 & -.654658 & .218221 & -.333334 \\ -.218219 & -.654658 & .523809 & .0476189 & .509178 \\ .327326 & .218221 & .0476189 & .428569 & .509177 \\ .166666 & -.333334 & .509178 & .509177 & .569892 \end{bmatrix}$$

To start the iterative process, the eigenvalues of **R*** are found.

$$-.237107, -.138579, -.427933\ 10^{-6}, 1.21591, 1.68782$$

Two of the eigenvalues are positive and three are negative. The eigenvectors corresponding to the positive eigenvalues are found. MAPLE will return these eigenvectors already scaled to have norm 1.

$$e1 := [.477755 \quad .238664 \quad .0775400 \quad .630349 \quad .558054]$$

$$e2 := [-.282451 \quad -.629440 \quad .590471 \quad .009117 \quad .418665]$$

The next $\mathbf{R}^*$ is formed by taking the sums of squares of the corresponding entries in these two eigenvectors and placing them on the diagonal as new estimates of the communalities.

$$newRstar := \begin{bmatrix} .308029 & .500000 & -.218219 & .327326 & .166666 \\ .500000 & .453156 & -.654658 & .218221 & -.333334 \\ -.218219 & -.654658 & .354667 & .0476189 & .509178 \\ .327326 & .218221 & .0476189 & .397421 & .509177 \\ .166666 & -.333334 & .509178 & .509177 & .486704 \end{bmatrix}$$

This process will be repeated until the $\mathbf{R}^*$ matrix converges. Convergence can be checked by looking at the difference in the last two consecutive $\mathbf{R}^*$ matrices.

The next iteration yields the eigenvalues

-.352971, -.275860, -.0627888, 1.15194, 1.53967

and new $\mathbf{R}^*$ matrix

$$\begin{bmatrix} .316082 & .500000 & -.218219 & .327326 & .166666 \\ .500000 & .435714 & -.654658 & .218221 & -.333334 \\ -.218219 & -.654658 & .344568 & .0476189 & .509178 \\ .327326 & .218221 & .0476189 & .417134 & .509177 \\ .166666 & -.333334 & .509178 & .509177 & .486483 \end{bmatrix}$$

Four more iterations are given

$$\begin{bmatrix} .317330 & .500000 & -.218219 & .327326 & .166666 \\ .500000 & .431027 & -.654658 & .218221 & -.333334 \\ -.218219 & -.654658 & .343960 & .0476189 & .509178 \\ .327326 & .218221 & .0476189 & .423766 & .509177 \\ .166666 & -.333334 & .509178 & .509177 & .483915 \end{bmatrix}$$

$$\begin{bmatrix} .317135 & .500000 & -.218219 & .327326 & .166666 \\ .500000 & .429600 & -.654658 & .218221 & -.333334 \\ -.218219 & -.654658 & .344579 & .0476189 & .509178 \\ .327326 & .218221 & .0476189 & .426671 & .509177 \\ .166666 & -.333334 & .509178 & .509177 & .482011 \end{bmatrix}$$

$$\begin{bmatrix} .316738 & .500000 & -.218219 & .327326 & .166666 \\ .500000 & .429128 & -.654658 & .218221 & -.333334 \\ -.218219 & -.654658 & .345089 & .0476189 & .509178 \\ .327326 & .218221 & .0476189 & .428192 & .509177 \\ .166666 & -.333334 & .509178 & .509177 & .480855 \end{bmatrix}$$

$$\begin{bmatrix} .316402 & .500000 & -.218219 & .327326 & .166666 \\ .500000 & .428960 & -.654658 & .218221 & -.333334 \\ -.218219 & -.654658 & .345384 & .0476189 & .509178 \\ .327326 & .218221 & .0476189 & .429057 & .509177 \\ .166666 & -.333334 & .509178 & .509177 & .480177 \end{bmatrix}$$

and then a check for convergence:

$$\begin{bmatrix} -.000336 & 0 & 0 & 0 & 0 \\ 0 & -.000168 & 0 & 0 & 0 \\ 0 & 0 & .000295 & 0 & 0 \\ 0 & 0 & 0 & .000865 & 0 \\ 0 & 0 & 0 & 0 & -.000678 \end{bmatrix}$$

Two more iterations

$$\begin{bmatrix} .316173 & .500000 & -.218219 & .327326 & .166666 \\ .500000 & .428907 & -.654658 & .218221 & -.333334 \\ -.218219 & -.654658 & .345551 & .0476189 & .509178 \\ .327326 & .218221 & .0476189 & .429576 & .509177 \\ .166666 & -.333334 & .509178 & .509177 & .479794 \end{bmatrix}$$

$$\begin{bmatrix} .316016 & .500000 & -.218219 & .327326 & .166666 \\ .500000 & .428888 & -.654658 & .218221 & -.333334 \\ -.218219 & -.654658 & .345646 & .0476189 & .509178 \\ .327326 & .218221 & .0476189 & .429888 & .509177 \\ .166666 & -.333334 & .509178 & .509177 & .479573 \end{bmatrix}$$

and a check for convergence:

$$\begin{bmatrix} -.000157 & 0 & 0 & 0 & 0 \\ 0 & -.000019 & 0 & 0 & 0 \\ 0 & 0 & .000095 & 0 & 0 \\ 0 & 0 & 0 & .000312 & 0 \\ 0 & 0 & 0 & 0 & -.000221 \end{bmatrix}$$

Two more iterations

$$\begin{bmatrix} .315913 & .500000 & -.218219 & .327326 & .166666 \\ .500000 & .428879 & -.654658 & .218221 & -.333334 \\ -.218219 & -.654658 & .345685 & .0476189 & .509178 \\ .327326 & .218221 & .0476189 & .430062 & .509177 \\ .166666 & -.333334 & .509178 & .509177 & .479427 \end{bmatrix}$$

$$\begin{bmatrix} .315847 & .500000 & -.218219 & .327326 & .166666 \\ .500000 & .428876 & -.654658 & .218221 & -.333334 \\ -.218219 & -.654658 & .345713 & .0476189 & .509178 \\ .327326 & .218221 & .0476189 & .430176 & .509177 \\ .166666 & -.333334 & .509178 & .509177 & .479350 \end{bmatrix}$$

and another check for convergence:

$$\begin{bmatrix} -.000066 & 0 & 0 & 0 & 0 \\ 0 & -.3\,10^{-5} & 0 & 0 & 0 \\ 0 & 0 & .000028 & 0 & 0 \\ 0 & 0 & 0 & .000114 & 0 \\ 0 & 0 & 0 & 0 & -.000077 \end{bmatrix}$$

That is close enough.

To obtain estimates of the principal factors, find the eigenvalues and eigenvectors of the last **R***. The principal factors are the product of **Z** and these eigenvectors, and are given by

$$F1 := [-1.19815 \quad .158920 \quad -2.54522 \quad 1.74197 \quad -1.19815 \quad .490422 \quad 1.63098 \quad .601408 \quad .158920 \quad .158920]$$

$$F2 := [1.18905 \quad 1.23712 \quad -2.58979 \quad -.613639 \quad 1.18905 \quad -1.13620 \quad -1.82583 \quad .0759950 \quad 1.23712 \quad 1.23712]$$

A screeplot of the eigenvalues from the last iteration is given in figure 2.

Insert Figure 2 about here

Conclusion

Factor analysis is a regression-like data reduction technique that involves a generalized least squares procedure. As with all data reduction techniques, factor analysis seeks to combine variables into common, underlying factors that can be further analyzed. Matrix algebra helps illustrate the dynamic involved in the procedure. A computer algebra system such as MAPLE makes the matrix algebra bearable. An examination of factor analysis in this manner makes clear the processes that SAS and SPSS execute and do not allow them to be a black box.

References

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum. Hamilton, L. C. (1992). Regression with graphics: A second course in applied statistics. Pacific Grove, CA: Brooks/Cole.

Hawkins, D. M. (1973). On the investigations of alternative regressions by principal component analysis. *Appl. Stat.*, 22, 275-286.

Knapp, T.R. (1978). Canonical correlation analysis: A general parametric significance testing system. Psychological Bulletin, 85, 410 - 416.

Pedhazur, E. J., & Smelkin, L. P. (1991). Measurement, design and analysis: An integrated approach. Hillsdale, NJ: Erlbaum.

Seber, G. A. F. (1977). Linear regression analysis. New York: Wiley & sons.

Stevens, J. (1992). Applied multivariate statistics for the social sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Thompson, B. (1984). Canonical correlation analysis: Uses and interpretation. Newbury Park, CA: SAGE.

Tucker, L.R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38, 1-10.

Velicer, W. F. & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research*, 25, 1-28.

Zwick, W.R., & Velicer, W.F. (1986). A comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99, 432-442.

Table 1

Five questions asked to lecturers

X1: Do you feel you have input in departmental decisions?

X2: Do you feel the professional faculty consider you as an integral part of the program?

X3: Are there procedures or events that cause you to feel unnecessarily separated from the rest of the faculty?

X4: Does your Department Head or a designated supervisor discuss your evaluation with you each year?

X5: Are you interested in long term employment at this university?

Table 2

Mean and standard deviation of the columns of **X**

	Mean	Standard Deviation
X1	.2	.421636
X2	.5	.527048
X3	.7	.483046
X4	.7	.483046
X5	.9	.316228

Figure 1

The relationship among some data reduction techniques

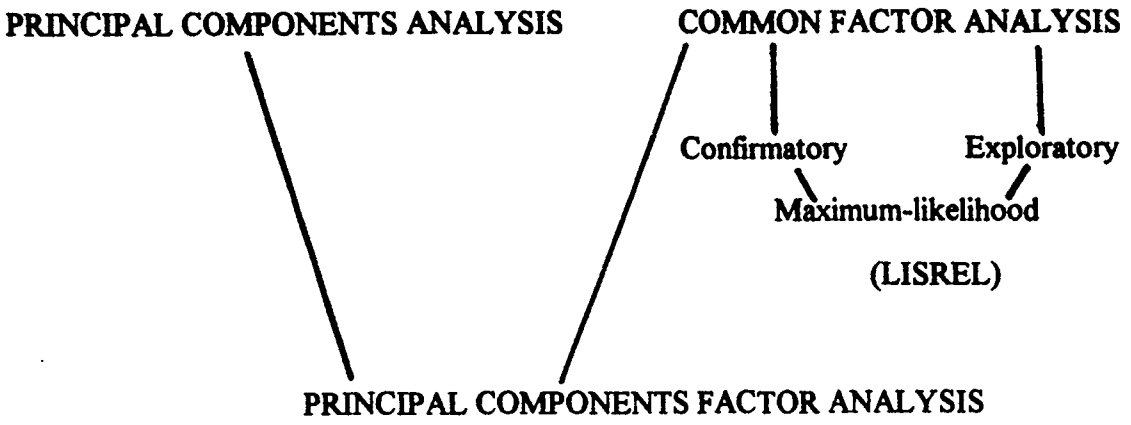
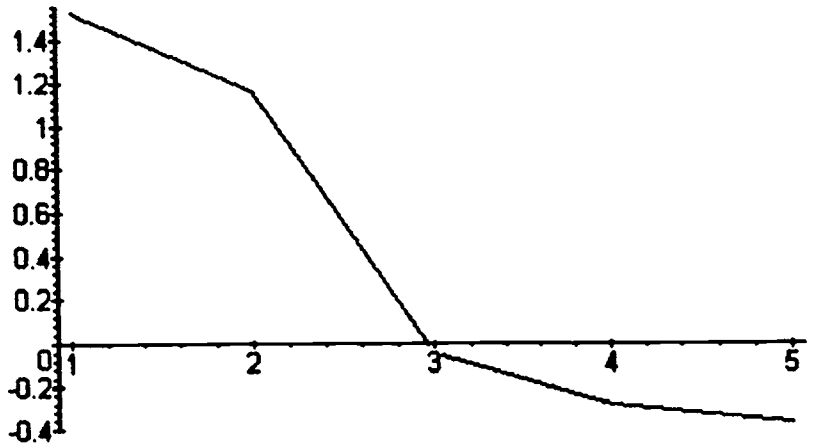


Figure 2

Screeplot of eigenvalues



Appendix A

Maple Commands -- Note: Calls to eig_info and f below are session dependent.

```
> Digits:=6:
> with(linalg):
> with(stats);
> with(describe);
> with(transform);
> X:=matrix(10,5,[0,0,1,0,1,0,0,1,1,1,0,1,0,0,0,1,1,1,1,1,0,0,1,0,1,0,1,0,1,1,1,1,0,1,1,0,1,
> barlist:=NULL: for i from 1 to 5 do barlist:=barlist,mean(convert(col(X,i),list)): od:
>  Xbar:=vector(5,[barlist]);
> devlist:=NULL: for i from 1 to 5 do
> devlist:=devlist,standarddeviation[1](convert(col(X,i),list)): od:
> Xdev:=vector(5,[devlist]);
> for i from 1 to 5 do evalf(Xdev[i]); od;
> matlist:=NULL: for i from 1 to 5 do
> matlist:=matlist,standardscore[1](convert(col(X,i),list)): od:
> Zstar:=transpose(matrix([matlist]));
> Z:=map(evalf,Zstar);
> Astar:=evalm(Zstar^1/3);
> A:=map(evalf,Astar);
> R1:=multiply(transpose(Astar),Astar);
> R:=multiply(transpose(A),A);
> R1inv:=inverse(R1);
> Rinv:=map(evalf,R1inv);
> diaglist:=NULL: for i from 1 to 5 do diaglist:=diaglist,1/Rinv[i,i]: od:
> Rstar:=evalm(R-diag(diaglist));
> eigenvals(Rstar);
> eig_info:=eigenvects(Rstar);
> e1:=eig_info[1][3][1]; e2:=eig_info[4][3][1];
> comlist:=NULL: for i from 1 to 5 do comlist:=comlist,e1[i]^2+e2[i]^2: od:
> newRstar:=evalm(R-diag(1,1,1,1,1)+diag(comlist));
> eigenvals(newRstar);
> eig_info:=eigenvects(newRstar);

> f:=proc(m,n)
>     local comlist, i, v1, v2:
>     global newRstar:
>     v1:=eig_info[m][3][1]: v2:=eig_info[n][3][1]:
>     comlist:=NULL:
>     for i from 1 to 5 do
>         comlist:=comlist,v1[i]^2+v2[i]^2:
>     od:
```

```
>     newRstar:=evalm(R-diag(1,1,1,1,1)+diag(comlist));
> end;

> f(3,5);
> eig_info:=eigenvects(newRstar);
> f(3,5);
> eig_info:=eigenvects(newRstar);
> f(1,3);
> eig_info:=eigenvects(newRstar);
> f(1,2);
> eig_info:=eigenvects(newRstar);
> f(1,2);
> evalm("-""");
> eig_info:=eigenvects(newRstar);
> f(2,3);
> eig_info:=eigenvects(newRstar);
> f(3,5);
> evalm("-""");
> eig_info:=eigenvects(newRstar);
> f(2,3);
> eig_info:=eigenvects(newRstar);
> f(1,3);
> evalm("-""");
> eig_info:=eigenvects(newRstar);
> e1:=eig_info[1][3][1]; e2:=eig_info[3][3][1];
> F1:=multiply(Z,e1); F2:=multiply(Z,e2);
> orderlist:=[3,1,4,5,2]: plist:=NULL: for i from 1 to 5 do
> plist:=plist,i,eig_info[orderlist[i]][1]: od: plist;
> plot([plist],style=line);
```