PSYCHOMETRIKA—VOL. 43, NO. 2.
JUNE, 1978
NOTES AND COMMENTS

THE PRINCIPAL COMPONENTS OF MIXED MEASUREMENT LEVEL MULTIVARIATE DATA: AN ALTERNATING LEAST SQUARES METHOD WITH OPTIMAL SCALING FEATURES

FORREST W. YOUNG

UNIVERSITY OF NORTH CAROLINA

YOSHIO TAKANE

MCGILL UNIVERSITY

AND

JAN DE LEEUW

UNIVERSITY OF LEIDEN

A method is discussed which extends principal components analysis to the situation where the variables may be measured at a variety of scale levels (nominal, ordinal or interval), and where they may be either continuous or discrete. There are no restrictions on the mix of measurement characteristics and there may be any pattern of missing observations. The method scales the observations on each variable within the restrictions imposed by the variable's measurement characteristics, so that the deviation from the principal components model for a specified number of components is minimized in the least squares sense. An alternating least squares algorithm is discussed. An illustrative example is given.

Key words: data analysis, factor analysis, qualitative data, nonmetric analysis.

Introduction

Principal components analysis [Hotelling, 1933] has been widely used in the social and behavioral sciences. The method postulates that an $m \times n$ matrix Z of m observations on n variables has a bilinear structure of the form,

(1) $$\hat{Z} = XF',$$

where X is an $m \times r$ matrix of m component scores on r components, and F is an $n \times r$ matrix of n loadings on the r components. For identification purposes it is conventional to restrict X and F such that $X'X/m = I$ and $F'F = D$ (diagonal). It is also conventional that Z is columnwise standardized. Hotelling's method finds X and F such that

(2) $$\theta = \mathrm{tr}(Z - \hat{Z})'(Z - \hat{Z})$$

is minimized for a prescribed number of components.

In this note we extend Hotelling's principal components analysis to the case where the variables have a variety of measurement characteristics. Some may be nominal, others ordinal and the rest interval. Furthermore, some may be discrete and others continuous, in the terminology of de Leeuw, Young & Takane [1976]. We place no restrictions on the pattern of measurement characteristics; any mix will do. No previous work [Roskam, 1968; Young, 1972; Kruskal & Shepard, 1974] has extended principal components analysis in this fashion. Since our procedure permits all of the variables to be nominal, all to be ordinal, or (for that matter) all to be interval, the previous developments are all considered as special cases, including Hotelling's original contribution.

Procedure

In the presence of nominal and/or ordinal variables the optimization criterion (2) is generalized to

(3) $$\theta^* = \mathrm{tr}(Z^* - \hat{Z})'(Z^* - \hat{Z})$$

Copies of this paper and of the associated PRINCIPALS program may be obtained by writing to Forrest W. Young, Psychometric Laboratory, Davie Hall 013-A, Chapel Hill, NC 27514.

where Z^* is an $m \times n$ matrix of optimally scaled observations. We employ the scaling convention that Z^* is columnwise centered and normalized; i.e.,

$$Z^{*\prime} l_m = o_n \quad \text{and diag} \left[\frac{Z^{*\prime} Z^*}{m} \right] = I_n \tag{4}$$

where l_m and o_n are vectors of ones and zeroes, and where their subscripts indicate their orders.

The procedure, which we refer to as PRINCIPALS (*Princip*al components analysis by *a*lternating *l*east *s*quares), optimizes θ^* defined in (3) under the normalization restriction on Z^* stated in (4). PRINCIPALS is based on the alternating least squares (ALS) principle, and is closely related to other ALS procedures developed previously [de Leeuw, Young & Takane, 1976; Young, de Leeuw & Takane, 1976; Takane, Young & de Leeuw, 1977]. It consists of two phases, a model estimation phase (optimization of θ^* with respect to the model parameters X and F) and an optimal scaling phase (optimization of θ^* with respect to the data parameters Z^*). In each phase conditional least squares estimates of a subset of the total set of parameters are obtained while the other parameters are held constant. The two phases are iteratively alternated until convergence is obtained. Such ALS procedures have been proven to be monotone-convergent by de Leeuw, Young & Takane [1976]. The philosophical, mathematical and algorithmic concepts of our approach have been presented in detail in the previous work cited above. Thus, we do not give a detailed explanation of the terms used in this section, but refer the interested reader to those articles.

PRINCIPALS consists of the following steps:

(Step 0.) Initialization: As in all of our previous work, the observed data Z are used for the initial Z^*. That is, $Z^* = Z$. For nominal variables random numbers may be assigned to categories, if initial category values are not externally provided. We then standardize Z^* to meet restriction (4), and proceed to Step 1.

(Step 1.) Model estimation: Let the Eckart-Young decomposition of Z^* be $PD^{1/2}Q'$. It is well known that the X and F are given by $X = P_r$ and $F = Q_r D_r^{1/2}$, where P_r is that portion of matrix P which is the r (normalized) eigenvectors of $Z^*Z^{*\prime}$ corresponding to the r dominant eigenvalues. Correspondingly, Q_r is a submatrix of Q with r (normalized) eigenvectors of $Z^{*\prime}Z^*$ corresponding to the r dominant eigenvalues, and D_r is the diagonal matrix of r dominant eigenvalues (of either $Z^*Z^{*\prime}$ or $Z^{*\prime}Z^*$). It is tacitly assumed that the $r + 1$ dominant eigenvalues are all distinct (in order to uniquely identify X and F). Computationally we first form $R^* = Z^{*\prime}Z^*/m$ (assuming that $n < m$), and obtain F by solving $R^*Q_r = Q_r D_r$. Then X can be obtained by $X = Z^*FD_r^{-1}$. (Several variants of this procedure, some of which may be more efficient, are conceivable within the alternating least squares framework. See Young, de Leeuw & Takane, 1976, for a discussion of some of these variants.)

(Step 2.) Termination: We evaluate θ^* at this point, and, if the improvement in fit from the previous iteration to the present iteration is negligible, stop.

(Step 3.) Data estimation (optimal scaling): From X and F we compute $\hat{Z}$ by (1). We then obtain the matrix of optimally scaled data Z^* which gives the minimum θ^* for fixed $\hat{Z}$ within each variable's measurement restrictions. The optimal scaling of data can be performed for each variable separately and independently, since the θ^* is separable with respect to the optimally scaled data for each variable. That is, we can rewrite (3) as a sum of independent problems, one for each variable:

$$\theta^* = \sum_{i=1}^{n} (z_i^* - \hat{z}_i)'(z_i^* - \hat{z}_i) = \sum_{i=1}^{n} \theta_i^*, \tag{7}$$

where z^*_i and $\hat{z}_i$ and i'th column vectors of Z^* and $\hat{Z}$, respectively. (A least squares loss function is said to be separable with respect to certain subsets of parameters when it can be decomposed into a sum of components each of which is a function only of the parameter subset.) Note that $\theta_i^* = (z_i^* - \hat{z}_i)'(z_i^* - \hat{z}_i)$, $(i = 1, \cdots, n)$ is a function of only z_i^*. The minimum of θ^* can be obtained by minimizing each θ_i^* separately with respect to each z_i^* $(i = 1, \cdots, n)$.

Each z_i^* may be obtained by the methods discussed by de Leeuw, Young & Takane [1976] in a fashion which permits each variable to be defined at the nominal, ordinal or interval (including ratio) level of measurement, and to be either discrete or continuous. These methods minimize θ^* for any selection of measurement characteristics.

The optimally scaled data are normalized before going back to Step 1. Steps 1 through 3 are iterated until convergence, which is assured, is obtained.

Example

Informal observations based on several sets of Monte Carlo data indicate that the PRINCIPALS procedure is capable of obtaining the desired results. Analysis of the artificial data used by Kruskal and Shepard [1974] in evaluating their procedure for nonmetric (ordinal) principal components analysis is also encouraging. The Kruskal and Shepard data resemble the "cylinder problem" originally proposed by Thurstone [1974, p. 117]. The data specify 12 physical characteristics of 30 cylinders, including such aspects as their height, volume, electrical resistance, moment of inertia, etc. Each of these 12 characteristics is related to the height and base area of each cylinder by a simple formula which is generally non-linear. (The entire list of variables and formulae are given by Kruskal and Shepard, 1974, p. 140.)

Since there is no error in these data, PRINCIPALS ought to find inverse transformations of each of the 12 variables which permit a perfectly fitting 2-component solution that is perfectly related to the two underlying variables of height and base area. This did in fact happen. After 30 iterations the STRESS (as defined by (3) and (4)) was .00790, the variance accounted for by each variable was 1.000, and the correlations between the two variables of height and base area and an orthogonal procrustes rotation [Cliff, 1966] of the derived principal components were .999263 and .999275. Thus, our solution is identical to the true underlying structure, lending credence to our procedure.

REFERENCES

Cliff, N. Orthogonal rotation to congruence. *Psychometrika,* 1966, *31,* 33–42.

Hotelling, H. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology,* 1933, *24,* 417–441.

Kruskal, J. B., & Shepard, R. N. A nonmetric variety of linear factor analysis. *Psychometrika,* 1974, *39,* 123–157.

de Leeuw, J., Young, F. W., & Takane, Y. Additive structure in qualitative data: An alternating least squares method with optimal scaling features. *Psychometrika,* 1976, *41,* 471–503.

Roskam, E. E. *Metric analysis of ordinal data in psychology.* Voorschoten, Holland: VAM Publishing Co. 1968.

Takane, Y., Young, F. W., & de Leeuw, J. Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika,* 1977, *42,* 7–67.

Thurstone, L. L. *Multiple factor analysis.* Chicago: University of Chicago Press, 1947.

Young, F. W. A model of polynomial conjoint analysis algorithms. In R. N. Shepard, et al. (Eds.) *Multidimensional scaling: Theory and applications in the behavioral sciences* (Vol. I). New York: Seminar Press, 1972.

Young, F. W., de Leeuw, J., & Takane, Y. Regression with qualitative and quantitative variables: An alternating least squares method with optimal scaling features. *Psychometrika,* 1976, *41,* 505–529.

Manuscript received 8/16/77
Final version received 11/11/77