PSYCHOMETRIKA—VOL. 31, NO. 3
SEPTEMBER, 1966

SIMULTANEOUS FACTOR ANALYSIS OF SEVERAL GRAMIAN MATRICES

JOSEPH LEVIN
SZOLD INSTITUTE
JERUSALEM, ISRAEL

Given several Gramian matrices, a least-square fit to all the matrices by one factor matrix, with a predetermined number of factors, is shown to be the principal axes solution of the average of the matrices.

The problem of a simultaneous factor analysis of several matrices arises usually in the context of studies of factorial invariance. The issue of factorial invariance is relevant to correlation matrices based on different populations yet of similar nature, or to correlations based on presumably similar variables, so that a similarity of factorial structure is expected.

The prevalent solution to this problem is based on a separate factor analysis of each matrix and a rotation of the factor matrices to proportional profiles. This method was first advocated by Cattell [3] and has been re-examined from various points of view by Meredith and others [2, 5]. However, this method has two limitations. First, each rotation applies only to two matrices at a time; a simultaneous treatment of several matrices has not yet been developed. Second, the requirement of proportionality of the profiles is not a strict one. There are circumstances where it is desirable to fit the same factor matrix to several correlation matrices or covariance matrices simultaneously. A case in point would be correlations based on random samples of the same population. Though the most efficient approach would be to pool all the samples and compute one correlation matrix for the whole group, this procedure is not always feasible due to some technical obstacle, such as differences of scale among the tests (e.g., when alternate forms are employed, differing in mean and variance though essentially equivalent).

Multi-way classification matrices discussed by Abelson [1] may serve as another example. The latter case is discussed subsequently.

Simultaneous Least-Square Analysis with a Predetermined Number of Factors

As is well known, a least-square fit to a Gramian matrix when the number of factors is determined in advance is obtained from the principal axes solution. The following is a generalization of this theorem to a set of several matrices. It is shown that a least-square fit of the same factor matrix to several Gramian matrices yields the principal axes solution of the average of the matrices.

Notation

R_h is a correlation matrix (or more generally a Gramian matrix), $h = 1, \cdots, s$.
r_{hij} is an element of R_h, $i = 1, \cdots, n; j = 1, \cdots, n$.
X is a matrix of order n by m and rank m (where m is predetermined), also $m \leq n$.
x_{ik} is an element of X.
$\bar{r}_{ij}$ is the average correlation over h.
$\bar{R}$ is the matrix, the elements of which are $\bar{r}_{ij}$.

We want to solve for X such that XX' is a least-square approximation to all the matrices R_h. The sum of squares of the residuals is

$$t = \sum_h \sum_i \sum_j (r_{hij} - \sum_k x_{ik}x_{jk})^2. \tag{1}$$

To solve for x_{ik} which minimize t, the expression is differentiated with respect to each x_{ik} and the derivatives are set equal to zero.

$$\cancel{\frac{dt}{dx_{ik}} = 2 \sum_h \sum_j x_{jk}(r_{hij} -} \sum_k x_{ik}x_{jk}) = 0 \tag{2}$$

$$(i = 1, \cdots, n; k = 1, \cdots, m).$$

Opening the parentheses,

$$\sum_h \sum_j x_{hij}r_{hij} - s(\sum_j x_{jk'} \sum_k x_{ik}x_{jk}) = 0, \tag{3}$$

k' is introduced as an alternative subscript to k because the k outside the parentheses in expression (2) is independent of the k within the parentheses. Thus k' is a free subscript while k is eliminated by the summation.

Interchanging the order of summation of the left term in (3), dividing by s, and transferring the right term, we get

$$\sum_j x_{jk}\bar{r}_{ij} = \sum_j x_{jk'} \sum_k x_{ik}x_{jk}. \tag{4}$$

(4) can be formulated in matrix notation

$$\bar{R}X = XX'X. \tag{5}$$

It is easy to show that the solution of X can be rotated to the principal axes solution. Let L be the latent vector matrix and Λ the latent root diagonal matrix of $X'X$. Multiplying (5) by L we obtain

$$\bar{R}XL = X(X'X)L = XL\Lambda. \tag{6}$$

XL is the principal axes matrix of $\bar{R}$, but it is also a nonsingular rotation of X, and therefore serves as a solution of the original problem of a simultaneous factor analysis of a set of matrices.

Relation Between Factors of Input Matrices and Their Average

In order to clarify further the meaning of simultaneous factor analysis, it is necessary to relate $\bar{R}$ to the factors X_h of the matrices R_h (input matrices).

The simplest way to compare sets of data is to operate with their average. Our task, therefore, is to compare $\bar{X}_h$ with the factor matrix of $\bar{R}$. But in view of the complexity of the relation between the principal axes of a sum to the principal axes of the terms, a slightly different approach is adopted here. A formula comparing $\bar{R}$ with the correlations reconstructed from $\bar{X}_h$ gives us indirectly an idea of the relation between the factors of the input matrices and the factors of $\bar{R}$. This formula has also the merit of providing an index of similarity, or rather heterogeneity, of the input matrices.

At this point it is necessary to introduce an additional restrictive assumption, viz., that the number of significant factors, or rather the predetermined number of factors, is the same for all input matrices—say m factors. This assumption is a logical requirement, otherwise there would be no rationale for fitting the same factor matrix to all the input matrices. Let us introduce the following notation.

$X_{h(m)}$ the factor matrix of R_h extracted up to m factors.

$$R_{h(m)} = X_{h(m)}X'_{h(m)} .$$

$E_{h(m)} = R_h - R_{h(m)}$, this is a matrix of residuals.

$$\bar{R}_{h(m)} = \overline{X_{h(m)}X'_{h(m)}} = \bar{R}_h - \bar{E}_{h(m)} .$$

Employing this notation, the following equation is obtained by simple algebraic operations (note that the range of h is from 1 to s):

$$\begin{aligned} \bar{X}_{h(m)}\bar{X}'_{h(m)} &= \frac{1}{s}\sum_h X_{h(m)}\frac{1}{s}\sum_h X'_{h(m)} \\ &= \frac{1}{s^2}\sum_{i=1}^{s} X_{i(m)}X'_{i(m)} + \frac{1}{s^2}\sum_{i,j}^{s} X_{i(m)}X'_{j(m)} \\ &= \frac{1}{s}\overline{X_{i(m)}X'_{i(m)}} + \frac{s-1}{s}\overline{X_{i(m)}X'_{j(m)}}, \end{aligned} \tag{7}$$

i and j are alternative subscripts for h; $i \neq j$.

Multiplying by s and substituting $\bar{R}_{h(m)}$ for $\overline{X_{i(m)}X'_{i(m)}}$ we obtain (8),

$$s\bar{X}_{h(m)}\bar{X}'_{h(m)} = \bar{R}_{h(m)} - \overline{X_{i(m)}X'_{j(m)}} + \overline{sX_{i(m)}X'_{j(m)}} . \tag{8}$$

When the matrices $X_{h(m)}$ approach equality, the equation

$$\overline{X_{i(m)}X'_{j(m)}} = \overline{X_{h(m)}X'_{h(m)}} = \bar{R}_{h(m)}$$

is approximately correct. Substituting this relation in (8), we obtain

$$\bar{R}_{h(m)} = \overline{X_{h(m)}X'_{h(m)}} = \bar{X}_{h(m)}\bar{X}'_{h(m)} . \tag{9}$$

The term $\overline{X_{i(m)}X'_{i(m)}}$ is, therefore, an indicator of the resemblance (or heterogeneity) of the matrices $R_{h(m)}$.

The equations (7)–(9) assumed that m factors have been extracted. They relate $\bar{X}_{h(m)}$ to $\bar{R}_h - \bar{E}_{h(m)}$. Adopting a stricter assumption, that the full rank of all the matrices X_h is equal, (7)–(9) express the relations between $\bar{X}_h\bar{X}'_h$ and $\bar{R}_h$.

Rank Relation Between Input Matrices and Their Average

Mathematically the relation between the rank of a sum (or average) of Gramian matrices and the rank of the terms is determined by the following theorem:

$$\text{rank } (\bar{R}) \geq \text{rank } (R_h), \qquad h = 1, \cdots, s. \tag{10}$$

The proof follows directly from the formula employing the notation of sectioned matrices.

$$\begin{aligned} &[X_1 ; X_2 ; \cdots ; X_s][X_1 ; X_2 ; \cdots ; X_s]' \\ &\qquad = [X_1X'_1 + X_2X'_2 + \cdots + X_sX'_s] \end{aligned} \tag{11}$$

The rank of the product is equal to the rank of the supermatrix in the brackets. This supermatrix is a sequence of columns and its rank cannot be less than any of its submatrices X_h . Under special conditions of dependence among the submatrices, e.g., when they conform to the pattern of proportional profiles, the rank is retained. But in the general case the rank increases. In practical situations, even when simultaneous factor analysis seems justified, summation raises the rank, though it is assumed that for perfect data all the factor matrices are equal. This is due to the presence of error components in each matrix, and, as is well known, error inflates rank.

Practical procedures of factor analysis do not adhere strictly to the algebraic properties of the correlation matrices, and various estimation procedures like least-square approximations are employed in order to get around the statistical feature of error.

The mathematical relation of formula (10) should be applied here cautiously, if at all, because the exact ranks of the matrices are not utilized and practically the rank is defined by the number of "significant" factors extracted. The rank is determined by statistical decisions based on the magnitude of the latent roots, magnitude of residuals, and so on. There is, of course, a connection between mathematical rank and decisions concerning the number of factors, but the decisions in the context of our problem should take into account statistical considerations and the logical structure of the data.

Employing the term "estimated rank" to designate the number of significant factors, it is apparently even possible for the "estimated rank" of the average matrix to reduce.

The rank relations depend on the relations between the matrices which are averaged. Let us examine, for instance, the relation between semantic differential matrices, where each matrix represents the correlation of scales for a variety of concepts of one person. Even when there is a high degree of resemblance between the matrices, individual differences do exist. They cannot be considered as error components and there is no reason to assume that they will conform to the customary statistical assumption concerning random error. The differences between the factor matrices are a "genuine" source of variability and we may expect also the "estimated rank" of the average matrix to rise. When the individual differences are small, and irrelevant to the purpose of the investigation, a simultaneous factor analysis may be appropriate. However, the number of factors extracted should be even less than the "estimated rank" in order to minimize the effect of individual differences, i.e., only the major factors should be extracted.

On the other hand, when the differences among the matrices are attributed to sampling error of the correlation coefficients, the average of the error components tends to zero, as a limit, as the number of the matrices increases. Mathematically, the rank of the average matrix will still increase and approach its order. (As a matter of fact, for fallible data the rank is so inflated due to error that generally the rank equals the order of the matrix.) However, the residuals for a fixed number of factors may decrease, or, stated obversely, though the total number of factors increases, there may be less significant factors and a smaller number may be needed to account for the same residuals, i.e., the "estimated rank" may decrease.

Discussion

One important implication of the preceding theorem concerns permissible operation with the correlation coefficients. Some textbooks do not recommend averaging correlation coefficients and suggest a prior application of the z transformation [4]. This transformation normalizes the sampling distribution and is relevant to the application of the t test of significance, but is not related to the issue of additivity.

It seems that permissible operations cannot be defined in general but depend on the mathematical context defining the purpose and meaning of the operations. It should be noted, however, that the conclusion which dispenses with the transformation is based, indirectly, on the fact that the same weight is assigned to all the residuals in equation (1), irrespective of the magnitude of the correlations.

Simultaneous factor analysis applies to any set of matrices, independent as well as dependent ones, i.e., it can be employed, also, for the analysis of correlation matrices based on the same population as in multi-way classification tables. Indeed, let x_{ijk} be an element of a triple classification table. The table can be sectioned into a set of two-way tables, e.g., K tables of order J by I, each of which may serve for the computation of a correlation matrix. Thus, we obtain K matrices of order I or order J (depending which correlations are required by the design of the research).

This conclusion has interesting implications for the comparison of various methods of analysis of multi-way tables, especially Abelson's ANOVA model and Tucker's Three-Mode factor analysis model [1, 6]. It is easier to expound the main idea by a concrete illustration; semantic differential data consist of ratings of various objects, on scales, by persons. Since classical methods of factor analysis are not suitable for treatment of multi-way tables, the data have to be reduced to two-way tables. Within the framework of psycholinguistic studies individual differences are of secondary importance, as phrased by Abelson [1], they are "agents of discrimination." By this he means that scales are employed to discriminate objects, whereas the raters serve in capacity of an instrument. Since the variance due to the agents of discrimination is, in a sense, instrumental error, it is customary to reduce the data by averaging the scores of each object, on each scale, averaging over subjects.

On the other hand, Tucker's model of Three-Mode factor analysis [6] utilizes the principle of sectioning the data into two-way matrices, averaging the corresponding covariances (more precisely, the sum of products), and factoring of the resultant matrix. Thus, the semantic differential provides covariance matrices for each subject. The model requires analysis of the average of these matrices. The rationale of this procedure can be interpreted by the theorem of simultaneous factor analysis, for the underlying principle of the procedure consists of fitting a fixed factor matrix to all subjects. Obviously, the same reasoning applies to all the possible ways of sectioning the original table.

In order to avoid a possible misrepresentation, it should be emphasized that the preceding discussion refers only to the rationale of the technique of Three-Mode factor analysis; its fundamental idea is beyond the scope of our problem.

REFERENCES

[1] Abelson, R. P. Scales derived by considerations of variance components in multi-way tables. In H. Gulliksen and S. J. Messick (Eds.), *Psychological scaling: theory and applications*. New York: Wiley, 1960. Pp. 169–186.

[2] Ahmavaara, Y. The mathematical theory of factorial invariance under selection. *Psychometrika*, 1954, **19**, 27–38.

[3] Cattell, R. B. "Parallel proportional profiles" and other principles for determining the choice of factors by rotation. *Psychometrika*, 1944, **9**, 267–283.

[4] Guilford, J. P. *Fundamental statistics in psychology and education*. New York: McGraw-Hill, 1956.

[5] Meredith, W. Notes on factorial invariance. *Psychometrika,* 1964, **2**, 177–185.
[6] Tucker, L. R. The extension of factor analysis to three-dimensional matrices. In N. Frederiksen and H. Gulliksen (Eds.), *Contributions to mathematical psychology.* New York: Holt, Rinehart and Winston, 1964, 110–127.

Manuscript received 9/14/65
Revised manuscript received 10/28/65