Types of moment matrices

Contents

1. Introduction

Some of the various types of moment matrices that the program can compute are defined in this extract and illustrated by means of a small data set.

The basis of analysis in PRELIS is a *data matrix* $\mathbf{Z}$ with N rows and k columns:

$$\mathbf{Z} = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1k} \\ z_{21} & z_{22} & \cdots & z_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ z_{N1} & z_{N2} & \cdots & z_{Nk} \end{bmatrix}$$

The columns represent *variables*. The rows represent statistical units (individuals, companies, regions, occasions, etc.) on which the variables have been observed or measured.

Here we shall refer to a row of the data matrix as a *case* on with the variables have been observed or measured. A case may be a *single observation* (as when the row characterizes an individual) or a *multiple case* (as when the row characterizes a whole group of individuals with identical responses to the variables). When a row of the data matrix represents a pattern of observations, the row carries a *weight* equal to the number of individuals having the same responses.

Each element $z_{\alpha i}$ is a numeric value. For continuous variables, these values represent observations or measurements on an interval scale or ratio scale. For ordinal variables, the values represent arbitrary score values, such as 1, 2, 3,4, a 5-category Likert scale. Still other values in the data matrix may represent missing observations.

An example of such a data matrix is shown below. It consists of 15 cases on four variables. (The sample size 25 is far too small to be useful in any LISREL model. Nevertheless, this small data set will be used here for illustrative purposes, as it is possible to check most of the computations by hand.)

Case	Var 1	Var 2	Var 3	Var 4
1	1	3	-0.7	-0.4
2	2	4	2.3	1.6
3	3	3	1.2	1.7
4	1	-9	-0.4	-0.3
5	3	2	-1.2	-0.7
6	2	1	-9.0	1.2
7	2	1	0.8	0.3
8	3	3	1.6	1.5
9	1	2	-0.9	-9.0
10	1	4	-0.8	-0.8
11	1	1	0.7	0.8
12	1	2	1.1	1.3
13	1	1	-9.0	0.8
14	2	2	0.7	0.3
15	3	3	1.8	1.7
16	1	2	-0.9	-0.9
17	2	4	-0.8	-0.7
18	2	1	1.1	1.2
19	3	1	1.2	1.7
20	2	2	1.6	1.8
21	2	4	2.3	1.6
22	3	3	1.2	1.7
23	3	2	-1.2	-0.7
24	2	1	-9.0	1.2
25	2	1	0.8	0.3

Variables 1 and 2 are assumed to be ordinal variables. The three entries of "-9" are specified by the user to represent missing observations. PRELIS can handle missing data using pairwise or listwise deletion or by imputation. Pairwise and listwise deletion are discussed here.

2. Pairwise deletion

To begin with, we shall pretend that all four variables are continuous. Let n_{ij} be the number of cases having real observations on both variables *i* and *j* (the effective sample sizes under pairwise deletion). The n_{ij} form a symmetric matrix $\mathbf{N}$ of order $k \times k$. For the data or our small illustrative example the matrix $\mathbf{N}$ is:

$$\mathbf{N} = \begin{bmatrix} 25 & & & \\ 24 & 24 & & \\ 22 & 21 & 22 & \\ 23 & 22 & 20 & 23 \end{bmatrix}$$

Some of the moment matrices can now be defined.

The moment matrix (MM) is defined as the symmetric matrix $\mathbf{M} = \left(m_{ij}\right)$ with elements

$$m_{ij} = (1/n_{ij})\sum_{\alpha} z_{\alpha i} z_{\alpha j}$$

where the summation is over all cases with real observations on both variables *i* and *j*. This definition applies when *i* = *j* as well. The elements of $\mathbf{M}$ represent moments about zero or mean squares and products. For the small data set:

$$\mathbf{M} = \begin{bmatrix} 4.440 & & & \\ 4.412 & 6.042 & & \\ 1.320 & 1.251 & 1.562 & \\ 1.664 & 1.562 & 1.246 & 1.395 \end{bmatrix}$$

The covariance matrix (CM) is defined as the symmetric matrix $\mathbf{S} = (s_{ij})$ with elements

$$s_{ij} = [1/(n_{ij} - 1)]\sum_{\alpha}(z_{\alpha i} - \bar{z}_i)(z_{\alpha j} - \bar{z}_j)$$

where

$$\bar{z}_i = (1/n_{ij})\sum_{\alpha} z_{\alpha i} \text{ and } \bar{z}_j = (1/n_{ij})\sum_{\alpha} z_{\alpha j}.$$

Note that the means use all univariate non-missing data whereas the covariances are based on all cases with non-missing observations on both variables *i* and *j*. For our small data matrix:

$$\mathbf{S} = \begin{bmatrix} 0.623 & & & \\ 0.097 & 1.216 & & \\ 0.310 & 0.101 & 1.350 & \\ 0.216 & -0.083 & 0.896 & 0.881 \end{bmatrix}$$

Here, for example, s_{21} = 0.097 is based on 24 cases, s_{11} = 0.623 is based on 25 cases, and s_{22} = 1.216 is based on 24 cases.

The correlation matrix (KM) is the matrix $\mathbf{R} = (r_{ij})$ with elements

$$r_{ij} = s_{ij} / d_i d_j$$

where

$$d_i^2 = [1/n_{ii} - 1)]\sum_{\alpha}(z_{\alpha i} - \bar{z}_i)^2$$

and

$$d_j^2 = [1/n_{jj} - 1)]\sum_{\alpha}(z_{\alpha j} - \bar{z}_j)^2 \ .$$

In the standard deviations d_i and d_j, the sums are over all real observations on each variable *i* and *j*, respectively. With these definitions it can technically happen that $r_{ij}^2 > 1$, although this is unlikely in large samples. For the small data set:

$$\mathbf{R} = \begin{bmatrix} 1.000 & & & \\ 0.100 & 1.000 & & \\ 0.337 & 0.079 & 1.000 & \\ 0.292 & -0.081 & 0.822 & 1.000 \end{bmatrix}$$

The three matrices **M**, **S**, and **R** were computed without distinguishing between ordinal and continuous variables. Data values on ordinal variables were treated as if they came from interval scales. However, by declaring the first two variables to be ordinal, other types of correlation matrices can be obtained.

When some of the variables are declared ordinal, the arbitrary score values of these variables can be replaced with their corresponding normal scores before **M**, **S**, and **R** is computed.

For variable 2 in our small data set, the computation of the normal scores is as follows:

Category	Marginal Frequency	Upper Threshold	Normal Score
1	8	-0.431	-1.118
2	7	0.319	-0.103
3	5	0.967	0.532
4	4	+∞	1.750

The resulting moment matrices are:

$$\mathbf{M} = \begin{bmatrix} 0.805 & & & \\ 0.051 & 0.989 & & \\ 0.345 & 0.052 & 1.562 & \\ 0.240 & -0.107 & 1.246 & 1.395 \end{bmatrix}$$

$$\mathbf{S} = \begin{bmatrix} 0.839 & & & \\ 0.054 & 1.032 & & \\ 0.361 & 0.055 & 1.350 & \\ 0.251 & -0.112 & 0.896 & 0.881 \end{bmatrix}$$

$$\mathbf{R} = \begin{bmatrix} 1.000 & & & \\ 0.058 & 1.000 & & \\ 0.339 & 0.046 & 1.000 & \\ 0.292 & -0.118 & 0.922 & 1.000 \end{bmatrix}$$

Besides the correlation matrix KM described above, there are four other types of correlation matrices, OM, PM, RM and TM, which can be used when some or all of the variables are ordinal. These correlation matrices consist of three different types of correlations. For each pair of variables on of the following three alternatives will occur:

1. When ***both variables are continuous*** (interval scaled). the product-moment correlation is computed from all complete pairs of observations. This correlation is the same in OM, PM, RM and TM.
2. When ***both variables are ordinal***, a contingency table is obtained from which the correlation is computed. Under OM, this correlation is the product-moment correlation of optimal scores or the canonical correlation (See Kendal & Stuart, 1961, pp. 568 – 573). Under PM, this correlation is the maximum-likelihood estimate of the *polychoric correlation*, where an underlying bivariate normal distribution is assumed. Under RM, this is Spearman's rank correlation and under TM, this is Kendall's tau-c correlation.

3. When ***one variable is ordinal and the other is continuous***, the program obtains the mean and variance of the continuous variable for each category of the ordinal variable and uses these summary statistics to compute the *polyserial correlation* (assuming again an underlying bivariate normal distribution). Under OM, RM and TM, a simple consistent estimator will be used, but under PM, a maximum-likelihood estimator will be used (see Jöreskog, 1986).

The end product of this procedure is a correlation matrix for all of the variables, where each correlation has been estimated separately. Although it is rare in practice, experience indicates that such a correlation matrix sometimes fails to be positive-definite. When a correlation matrix that is not positive-definite is to be used to estimate a LISREL model, the ML or GLS method cannot be used. The ULS, WLS or DWLS method must be used instead. Furthermore, even if the matrix of correlations is positive-definite, these correlations are unlikely to behave like ordinary sample moments, not even asymptotically. So, if one uses the ML or GLS methods for fitting the model, one should not rely on the normal theory standard errors and chi-square goodness-of-fit measures supplied by LISREL. Correct large sample standard errors and chi-square values can be obtained with WLS in LISREL.

When both variables are ordinal, information provided in the data may be represented as a contingency table. For the illustrative data, the contingency table for variables 1 and 2 is:

	VAR 2				
VAR 1	1	2	3	4	Marginal
1	2	3	1	1	7
2	5	2	0	3	10
3	1	2	4	0	7
Marginal	8	7	5	4	24

Let *x* and *y* be two ordinal variables with *p* and *q* categories, respectively. Let $n_{ij}\ (i=1,2,...,p,\ j=1,2,...,q)$ be the corresponding frequencies in the contingency table.

Optimal scores for *x* and *y* are defined as two sets of ordered score values that maximize the product-moment correlations, subject to the constraints that the means are 0 and the variances are 1 (see Kendall & Stuart, 1961, pp. 568 – 573). The product-moment correlation of these optimal scores, sometimes called *canonical correlation*, is obtained as the second largest eigenvalue of a symmetric matrix formed from the elements of the contingency table.

The *polychoric correlation* is not a correlation between a pair of score values. Rather it is an estimate of the correlation between two latent variables η and ξ underlying *y* and *x*, where η and ξ are assumed to have a bivariate normal distribution. For our illustrative data, the polychoric correlation between variables 1 and 2 is estimated as 0.098.

This latent correlation can be estimated by the maximum-likelihood method based on the multinomial distribution of the cell frequencies in the contingency table. The estimation procedure follows Olsson (1979), but the computational algorithm has been considerably improved.

Next, consider the third case, when one variable is ordinal and one variable is continuous. In our small data set, there will be four such pairs of variables: (3,1), (3,2), (4,1), and (4,2). In the illustration below, we use the pair (3,1). Let *x* be an ordinal variable with *p* categories, and let *y* be a continuous variable. As before, let n_i be the number of cases in category *i* of *x*. Corresponding to these cases, there will be n_i values on *y* denoted:

$$y_{i1}, y_{i2}, ..., y_{in_i}$$

Let $\bar{y}_i$ and s_i^2 be the mean and unbiased variance of these values. (If $n_i = 1$. the variance is zero and this category cannot be used in the computations. However, the required correlation can still be computed, provided there are at least two categories with $n_i > 1$.) For the pair (3,1), these summary statistics are:

Category	Number of observations	Mean	Standard Deviation
1	7	-0.271	0.826
2	8	1.100	1.006
3	7	0.657	1.290

The *polyserial correlation* is the correlation between the observed variable *y* and a latent variable ξ, where *y* and ξ are assumed to have a bivariate normal distribution. This can be estimated by the maximum-likelihood method as described by Jöreskog (1986). Under OM, Jöreskog's Method 1 is used, under PM, Jöreskog's Method 5 is used.

For our illustrative data, the correlation matrices obtained under the options OM and PM are:

$$\mathbf{O} = \begin{bmatrix} 1.000 & & & \\ 0.197 & 1.000 & & \\ 0.424 & 0.060 & 1.000 & \\ 0.313 & -0.111 & 0.822 & 1.000 \end{bmatrix}$$

$$\mathbf{P} = \begin{bmatrix} 1.000 & & & \\ 0.098 & 1.000 & & \\ 0.362 & 0.051 & 1.000 & \\ 0.340 & -0.112 & 0.822 & 1.000 \end{bmatrix}$$

3. Listwise deletion

So far, we have dealt with pairwise deletion. With *listwise deletion*, all cases with missing observations are deleted first so that the data matrix reduces effectively to a matrix without missing observations. All the definitions above still apply. The main difference is that under listwise deletion, all computations are based on the same cases. This will guarantee that all the matrices obtained under MM, CM, and KM are non-negative-definite. Correlation matrices obtained under OM and PM still cannot be guaranteed to be non-negative-definite, as they may consist of different types of correlations.

4. References

Jöreskog, K.G. (1986). *Estimation of the polyserial correlation from summary statistics*. Research Report 86_2. University of Uppsala, Department of Statistics.

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, **44**(**4**), 443–460.