Analysis of ordinal variables

In many cases, especially when data are collected through questionnaires, the variables are ordinal, i.e., responses are classified into different ordered categories.

An ordinal variable z (z may be either a y- or x-variable in LISREL sense) may be regarded as a crude measurement of an underlying unobserved or unobservable continuous variable z^*. For example, a four-point ordinal scale may be conceived as:

$$\text{if} \quad z^* \le \alpha_1, \quad z \text{ is scored } 1,$$
$$\text{if } \alpha_1 < z^* \le \alpha_2, \quad z \text{ is scored } 2,$$
$$\text{if } \alpha_2 < z^* \le \alpha_3, \quad z \text{ is scored } 3,$$
$$\text{if } \alpha_3 < z^*, \quad \quad z \text{ is scored } 4,$$

where $\alpha_1 < \alpha_2 < \alpha_3$ are threshold values for z^*. It is often assumed that z^* has a standard normal distribution, in which case the thresholds can be estimated from the inverse of the normal distribution function.

Suppose z_1 and z_2 are two ordinal variables with underlying continuous variables z_1^* and z_2^*, respectively. Assuming that z_1^* and z_2^* have a bivariate normal distribution, their correlation is called the *polychoric correlation coefficient*. A special case of this is the *tetrachoric correlation coefficient*, where both z_1 and z_2 are dichotomous. Now, suppose further that z_3 is a continuous variable measured on an interval scale. The correlation between z_1^* and z_3 is called the *polyserial correlation coefficient* assuming that z_1^* and z_3 have a bivariate normal distribution. A special case of this is the *biserial correlation* when z_1 is dichotomous.

An ordinal variable z does not have a metric scale. To use such a variable in a linear relationship we use the corresponding underlying variable z^* instead. The polychoric and polyserial correlations are not correlations computed from actual scores but are rather theoretical correlations of the underlying z^*-variables. These correlations are estimated from the observed pairwise contingency tables of the ordinal variables.

The weight matrix required for such an analysis is the inverse of the estimated asymptotic covariance matrix **W** of the polychoric and polyserial correlations. The asymptotic covariance matrix as well as the matrix of polychoric and polyserial correlations are obtained by PRELIS.

The steps involved in this analysis will be described in this example, and the one that follows. These examples involve only ordinal variables and polychoric correlations.

In this study of attitudes of morality and equality (Hasselrot & Lernberg (1980)), Swedish school children in grade 9 were asked questions about their attitudes regarding social issues in family, school, and society. Among the questions asked were the following eight items.

> For me, questions about……

> 1. human rights
> 2. equal conditions for all people
> 3. racial problems
> 4. equal value of all people
> 5. euthanasia
> 6. crime and punishment
> 7. conscientious objectors
> 8. guild and bad conscience

> are….

> unimportant   not important   important   very important

For the present example we use a subsample of 200 cases. Responses to the eight questions were scored 1, 2, 3, and 4, where 4 denotes very important. The data matrix consists of 200 rows and 8 columns and is stored in the file **EX71.RAW** in the **LISREL Examples** folder.

The PRELIS command file (**EX71.PRL**) is:

```
ATTITUDES OF MORALITY AND EQUALITY
DA NI=8 MI=0
LA
HUMRGHTS EQUALCON RACEPROB EQUALVAL EUTHANAS CRIMEPUN CONSCOBJ GUILT
RA=EX71.RAW
OR ALL
OU MA=PM PM=EX71.PML AC=EX71.ACP
```

The polychoric correlations are saved in the file **EX71.PML**, which will be read by LISREL. The asymptotic covariance matrix, saved in the file **EX71.ACP**, is not equal to the matrix given in the PRELIS output file, but is equal to N times this matrix, as required by LISREL, where N is the sample size.

The matrix of polychoric correlations as printed by PRELIS is:

Correlation Matrix

```
             HUMRGHTS   EQUALCON   RACEPROB   EQUALVAL   EUTHANAS   CRIMEPUN
             --------   --------   --------   --------   --------   --------
 HUMRGHTS      1.000
 EQUALCON      0.422      1.000
 RACEPROB      0.213      0.183      1.000
 EQUALVAL      0.370      0.636      0.262      1.000
 EUTHANAS      0.438      0.696      0.303      0.688      1.000
 CRIMEPUN      0.206      0.239      0.265      0.406      0.204      1.000
 CONSCOBJ      0.183      0.324      0.323      0.355      0.238      0.311
    GUILT      0.093      0.303      0.238      0.323      0.327      0.204
```

Correlation Matrix

```
             CONSCOBJ    GUILT
             --------   --------
 CONSCOBJ      1.000
    GUILT      0.209      1.000
```

There are $(1/2)\ 8 \times 7 = 28$ correlations in the polychoric correlation matrix. Hence, there should be $(1/2)\ 8 \times 29 = 406$ elements in the asymptotic covariance matrix. These are stored as 67 lines with six elements each plus one line with four elements.

The LISREL model to be estimated is a two-factor confirmatory factor analysis model, in which the variables HUMRGHTS, EQUALCON, EQUALVAL, and EUTHANAS are assumed to measure Equality and RACEPROB, CRIMEPUN, CONSOBJ, and GUILT are assumed to measure Morality.

The LISREL command file (**EX71.LIS**) is:

```
ATTITUDES OF MORALITY AND EQUALITY
DA NI=8 NO=195 MA=PM
LA
HUMRGHTS EQUALCON RACEPROB EQUALVAL EUTHANAS CRIMEPUN CONSCOBJ GUILT
PM FI=EX71.PML
AC FI=EX71.ACP
MO NX=8 NK=2
FR LX 1 1 LX 2 1 LX 4 1 LX 5 1 LX 3 2 LX 6 2 LX 7 2 LX 8 2
OU SE TV
```

The MA = PM on the DA command is essential. Otherwise, the program will think that a covariance matrix will be analyzed and, as a consequence, the program will think that the file **EX71.ACP** contains 666 elements instead of 406, as there are 666 asymptotic variances and covariances of the 36 variances and covariances in a covariance matrix for 8 variables.

The PM command tells the program to read the matrix of polychoric correlations and the AC command tells the program to read the asymptotic covariance matrix.

The AC command implies that the WLS method will be used as default method instead of the ML method. So it is not necessary to put ME = WLS on the OU command.

The fitted path diagram is shown below:
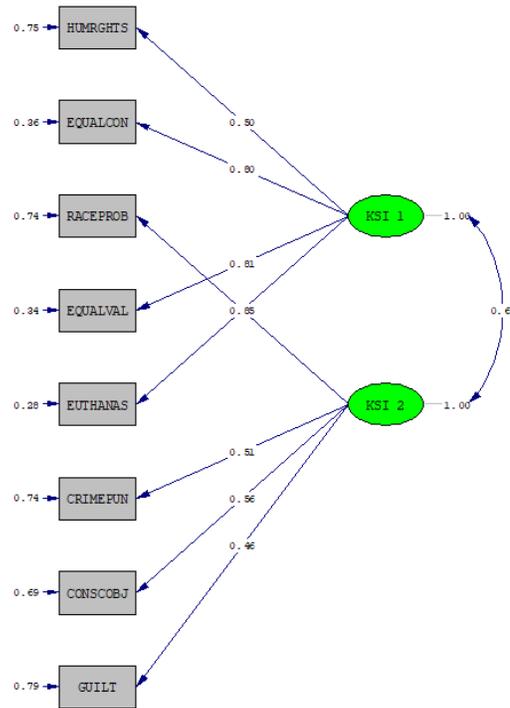
Table: Results using different types of correlations and different estimation methods (standard errors in parentheses)

Parameter	Normal theory GLS		Non-normal theory WLS	
	MA = KM	MA = PM	MA = KM	MA = PM
LX(1,1)	.430 (.074)	.506 (.070)	.431 (.049)	.501 (.084)
LX(2,1)	.733 (.066)	.790 (.062)	.763 (.029)	.803 (.047)
LX(3,2)	.470 (.087)	.534 (.082)	.429 (.056)	.514 (.087)
LX(4,1)	.769 (.066)	.814 (.061)	.804 (.029)	.810 (.050)
LX(5,1)	.801 (.065)	.878 (.059)	.876 (.030)	.851 (.053)
LX(6,2)	.450 (.086)	.479 (.081)	.456 (.059)	.509 (.095)
LX(7,2)	.469 (.086)	.494 (.081)	.485 (.057)	.559 (.085)
LX(8,2)	.439 (.087)	.488 (.082)	.496 (.062)	.463 (.087)
PH(2,1)	.690 (.083)	.703 (.070)	.708 (.060)	.675 (.086)
χ^2	20.06	37.15	37.27	41.48
GFI	.975	.953	.983	.952
AGFI	.952	.912	.969	.909
RMR	.042	.057	.050	.044

The results are shown in the last column of the table above. For comparison, the results obtained by normal theory GLS and the results from using the product-moment correlations based on normal scores, are also given in the table. In this table, MA = KM means "matrix analyzed = matrix of product-moment correlations based on normal score" and MA = PM means "matrix analyzed = matrix of polychoric correlations". Only the results of WLS with MA = PM are asymptotically correct. WLS with MA = KM is incorrect because the product-moment correlations based on normal scores are biased (inconsistent). The standard errors of the GLS estimates are wrong because they are based on the wrong formula.