Ordinal variables without missing values

Contents

1.   Ordinal Variables

Observations on an ordinal variable represent responses to a set of ordered categories, such as a five-category Likert scale. It is only assumed that a person who selected one category has more of a characteristic than if he/she had chosen a lower category, but we do not know how much more. Ordinal variables are not continuous variables and should not be treated as if they are. It is common practice to treat scores 1, 2, 3, ... assigned to categories as if they have metric properties but this is wrong. Ordinal variables do not have origins or units of measurements. Means, variances, and covariances of ordinal variables have no meaning. The only information we have are counts of cases in response vector. To use ordinal variables in structural equation models requires other techniques than those that are traditionally employed with continuous variables.

There are many methods available for estimating structural equation models with ordinal variables, see e.g., Jöreskog & Moustaki (2001), Yang-Wallentin, Jöreskog, & Luo (2010) or Forero, Maydeu-Olivares, & Gallardo-Pujol (2011) and references given there. There are essentially two types of methods. One is a full information maximum likelihood method (FIML) using a probit, logit, or other link function. The other method fits the model to a matrix of polychoric correlation or covariance matrix using some fit function like ULS or DWLS. The focus is here on the two methods: FIML and DWLS. These methods are illustrated using exploratory and confirmatory factor analysis models, but any LISREL model can be used, even those involving a mean structure.

2.   Example: Attitudes Toward Science and Technology (SCITECH)

In the Eurobarometer Survey 1992, citizens of Great Britain were asked questions about Science and Technology. The questions are given below.

1. Science and technology are making our lives healthier, easier and more comfortable [COMFORT].
2. Scientific and technological research cannot play an important role in protecting the environment and repairing it [ENVIRON].
3. The application of science and new technology will make work more interesting [WORK].

4. Thanks to science and technology, there will be more opportunities for the future generations [FUTURE].
5. New technology does not depend on basic scientific research [TECHNOL].
6. Scientific and technological research do not play an important role in industrial development [INDUSTRY].
7. The benefits of science are greater than any harmful effects it may have [BENEFIT].

Note that the items COMFORT, WORK, FUTURE, and BENEFIT have a positive question wording, whereas the items ENVIRON, TECHNOL, and INDUSTRY have a negative question wording.

The response alternatives were strongly disagree, disagree to some extent, agree to some extent, and, strongly agree. These were coded as 1, 2, 3, 4, respectively. The data file is **scitech.lsf**. Missing values have been deleted beforehand, see Batholomew et.al. (2002). This data set is used here to illustrate the case of no missing values. The sample size is 392. Exploratory factor analysis is illustrated first.

3.   Exploratory Factor Analysis

To perform an exploratory factor analysis of ordinal variables by the FIML method, open the file **scitech.lsf** (from the **SIMPLIS examples** folder) and select **EFA of Ordinal Variables** from the **Statistics** Menu. This shows the following dialog box
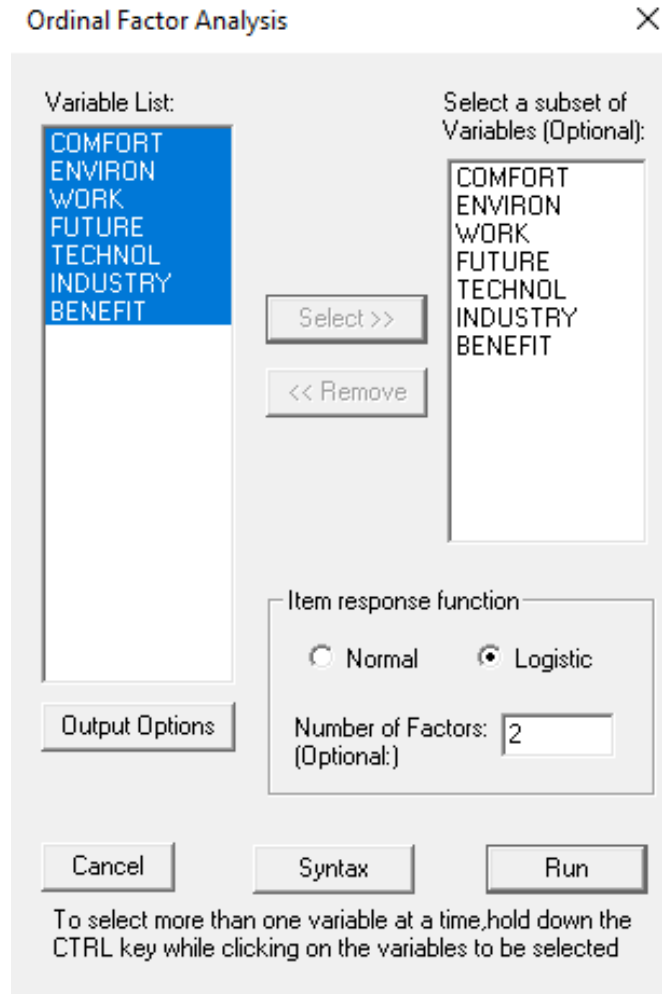
Select the variables to be analyzed, check the box **Logistic**, for example, and insert 2 for **Number of Factors**. Then click **Run**.

The output file **scitech.out** gives the following unrotated and rotated factor loadings:

```
Unrotated Factor Loadings

              Factor 1 Factor 2 Unique Var
              -------- -------- ----------
 COMFORT      0.764 0.000 0.416
 ENVIRON      0.237 0.817 0.277
 WORK         0.669 -0.4040.388
 FUTURE       0.847 -0.3230.179
 TECHNOL      0.234 0.832 0.252
 INDUSTRY     0.462 0.716 0.274
 BENEFIT      0.713 -0.2250.441

Varimax-Rotated Factor Loadings
              Factor 1 Factor 2 Unique Var
              -------- -------- ----------
 COMFORT      0.719 0.259 0.416
 ENVIRON      -0.0540.849 0.277
 WORK         0.767 -0.1530.388
 FUTURE       0.906 -0.0160.179
 TECHNOL      -0.0620.862 0.252
 INDUSTRY     0.192 0.830 0.274
 BENEFIT      0.747 0.030 0.441

Promax-Rotated Factor Loadings
              Factor 1 Factor 2 Unique Var
              -------- -------- ----------
 COMFORT      0.710 0.240 0.416
 ENVIRON      -0.0860.852 0.277
 WORK         0.774 -0.1750.388
 FUTURE       0.908 -0.0410.179
 TECHNOL      -0.0940.866 0.252
 INDUSTRY     0.161 0.826 0.274
 BENEFIT      0.747 0.010 0.441

Factor Correlations
Factor 1 Factor 2
-------- --------
Factor 1    1.000
Factor 2    0.064 1.000

Reference Variables Factor Loadings
              Factor 1 Factor 2 Unique Var
              -------- -------- ----------
 COMFORT      0.738 0.273 0.416
 ENVIRON      0.006 0.851 0.277
 WORK         0.757 -0.1400.388
 FUTURE       0.906 0.000 0.179
 TECHNOL      0.000 0.865 0.252
 INDUSTRY     0.251 0.837 0.274
 BENEFIT      0.750 0.043 0.441
```

```
Factor Correlations
Factor 1 Factor 2
-------- --------
Factor 1    1.000
Factor 2    -0.0891.000
```

The first solution is the standardized unrotated solution which is a transformation of the standardized solution obtained by the FIML procedure. The second solution is the varimax solution of Kaiser (1958). Both of these are orthogonal solutions, i.e., the factors are uncorrelated. The third solution is the promax solution of Hendrickson & White (1964). This is an oblique solution, i.e., the factors are correlated. The varimax and the promax solutions are transformations of the standardized unrotated solution and as such they are still maximum likelihood solutions. The fourth solution is the TSLS solution obtained in reference variables form. The reference variables are chosen as those variables in the promax solution that have the largest factor loadings in each column. In this case the reference variables are FUTURE and TECHNOL.

The Promax-Rotated Factor Loadings and the Reference Variables Factor Loadings suggest that there are two nearly uncorrelated factors and that Factor 1 is a Positive factor, having large loadings on the positively worded items COMFORT, WORK, FUTURE, and BENEFIT and Factor 2 is a Negative factor having large loadings on the negatively worded items ENVIRON, TECHNOL, and INDUSTRY. Three other files are also produced in this analysis: **SCITECH.POM**, **BIVFITS.POM** and **MULFITS.POM**. The first of these gives estimates of unstandardized factor loadings and thresholds and their standard errors. The other two give detailed information about the fit of the model.

The file **SCITECH.POM** give the following unstandardized parameters and standard errors

```
Unstandardized Thresholds Alpha^(i)_a
  COMFORT   -5.015    -2.744    1.535
  ENVIRON   -3.429    -1.243    1.002
     WORK   -2.941    -0.905    2.279
   FUTURE   -4.998    -2.125    1.885
  TECHNOL   -4.164    -1.482    1.086
 INDUSTRY   -4.695    -2.507    0.461
  BENEFIT   -3.379    -1.006    1.705

 Unstandardized Factor Loadings Beta_ij
  COMFORT    1.184     0.000
  ENVIRON    0.450     1.552
     WORK    1.074    -0.649
   FUTURE    2.002    -0.763
  TECHNOL    0.467     1.656
 INDUSTRY    0.882     1.366
  BENEFIT    1.073    -0.339

 Standard Errors for Unstandardized Thresholds Alpha^(i)_a
  COMFORT    0.589     0.240     0.173
  ENVIRON    0.304     0.173     0.172
     WORK    0.251     0.147     0.227
   FUTURE    0.600     0.331     0.307
  TECHNOL    0.376     0.195     0.195
 INDUSTRY    0.444     0.243     0.168
  BENEFIT    0.290     0.143     0.179

 Standard Errors for Unstandardized Factor Loadings Beta_ij
  COMFORT    0.187     0.000
  ENVIRON    0.222     0.228
     WORK    0.192     0.191
   FUTURE    0.434     0.298
```

```
 TECHNOL    0.253    0.272
INDUSTRY    0.252    0.233
 BENEFIT    0.172    0.188
```

4.   Estimation Using Adaptive Quadrature

The above FIML approach to analyze ordinal variables is only intended for exploratory factor analysis with one or two factors. For other cases one can use the following SIMPLIS syntax file instead, see file **scitech2a.spl**:

```
Raw Data from file SCITECH.lsf
 $ADAPQ(8) LOGIT
 Latent Variables Factor1 Factor2
 Relationships
 COMFORT - BENEFIT = Factor1
 ENVIRON - BENEFIT = Factor2
 Set the covariance between Factor1 and Factor2 to 0
 Path Diagram
 End of Problem
```

This uses adaptive quadrature to evaluate the integrals involved. The line

```
$ADAPQ(8) LOGIT
```

specifies that the adaptive quadrature procedure is to be used. One can specify the number of quadrature points to be used in the adaptive quadrature procedure and the link function. Follow the following guidelines to specify the number of quadrature points:

- For models with one latent variable, use 8 quadrature points

- For models with two or three latent variables, use 5-10 quadrature points
- For models with four or five quadrature points, use 5-6 quadrature points

- For models with six to ten latent variables use 3-4 quadrature points

The following link functions are available

- LOGIT

- PROBIT

- LOGLOG

- CLL (Complimentary log-log)

The adaptive quadrature gives a more accurate solution than **scitech.out**. But the unstandardized parameters are essentially the same in this case.

This output file also gives the following information:

```
    Number of quadrature points =     8
    Number of free parameters =      34
```

```
Number of iterations used =       8


-2lnL (deviance statistic) =      5822.09128
Akaike Information Criterion      5890.09128
Schwarz Criterion                 6025.11419
```

5.   Confirmatory Factor Analysis

This FIML approach will now be used to illustrate a confirmatory factor analysis. Use the file **scitech3a.spl**:

```
Raw Data from file scitech.lsf
$ADAPQ(8) LOGIT GR(5)
Latent Variables Positive Negative
Relationships
COMFORT WORK FUTURE BENEFIT = Positive
ENVIRON TECHNOL INDUSTRY    = Negative
Path Diagram
End of Problem
```

This gives

```
Number of quadrature points =  8
Number of free parameters =    29
Number of iterations used =    24
-2lnL (deviance statistic) =   5841.79954
Akaike Information Criterion   5899.79954
Schwarz Criterion  6014.96613
```

The difference between the two deviance statistics is $5841.800 - 5822.091 = 18.29$. If this is used as a chi-square with 5 degrees of freedom, the confirmatory factor model is rejected in favor of the exploratory two-factor model. This means that not all of the seven specified zero factor loadings are zero. One can relax the loading of COMFORT on Negative, see file **scitech4a.spl**. This gives a good two-factor solution with a small but significant loading of COMFORT on Negative. The deviance for this model is $-2\ln L = 5833.559$, corresponding to a chi-square of 11.468 with 4 degrees of freedom which is significant at the 1% level. This result can be interpreted as follows. Some fraction of people who generally have a positive attitude to science and technology will respond in the Strongly Disagree or Disagree to Some Extent categories to the COMFORT item.