Exploratory factor analysis of polytomous data

Contents

1. Introduction

The analysis of ordinal data requires other techniques than those used in the analysis of continuous data. The observations on an ordinal variable represent responses to a set of ordered categories. While it may be tempting to treat an ordinal variable as a continuous variable, this should be avoided. Ordinal variables do not have origins or units of measurements. In LISREL, exploratory factor analysis of ordinal variables is based on the methods of Jöreskog & Moustaki (2001). The likelihood function used for estimating an ordinal factor analysis model is different from that used for continuous variables, and is the sum of individual log-likelihoods for each case in the data. Each such individual log-likelihood is an integral, to be evaluated numerically. For details, the reader is referred to Section 16.2.1 of the *Multivariate Analysis with* LISREL text.

In our previous example, we used the simplest form of ordinal data, *i.e*. binary data, to illustrate. In this example we take a look at polytomous data, typically encountered as items in surveys. We will show how LISREL can be used to perform both exploratory and confirmatory factor analysis using such variables.

The Eurobarometer Survey from 1192 asked respondents questions about science and technology. Here we look at 7 questions, with responses from UK citizens. The sample size is 392, and there are no missing values.

The response alternatives for the questions considered here were:

- strongly disagree (coded 1)
- disagree to some extent (coded 2)
- agree to some extent (coded 3)
- strongly agree (coded 4)

The questions we look at in this example are:

- COMFORT: science and technology are making our lives healthier, easier and more comfortable

- ENVIRON: scientific and technological research cannot play an important role in protecting the environment and repairing it
- WORK: the application of science and new technology will make work more interesting
- FUTURE: thanks to science and technology there will be more opportunities for the future generations
- TECHNOL: new technology does not depend on basic scientific research
- INDUSTRY: scientific and technological research do not play an important role in industrial development
- BENEFIT: the benefits of science are greater than any harmful effects it may have.

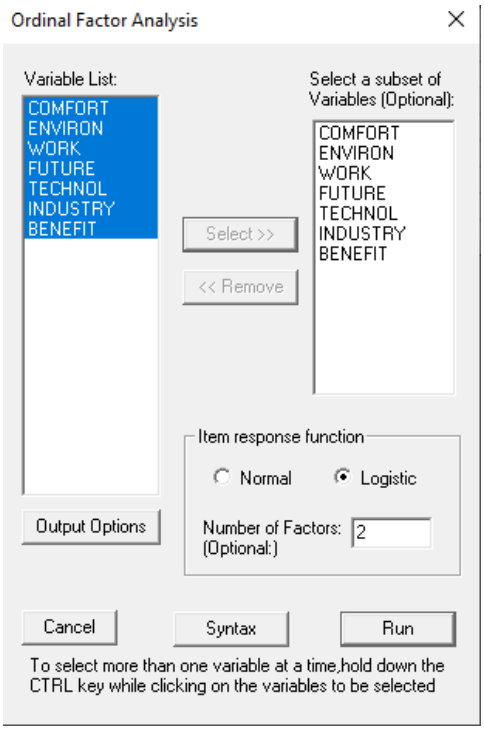
In the previous example we used ordinary data, weighting by the frequency of each of the possible response patterns. In this case, that is not feasible given that we have 4^7 possible response patterns for these 7 questions. Here we use the individual data from the file **scitech.lsf**.

SCITECH.LSF

	COMFORT	ENVIRON	WORK	FUTURE	TECHNOL	INDUSTRY	BENEFIT
1	4.00	4.00	4.00	3.00	4.00	3.00	2.00
2	3.00	4.00	3.00	3.00	3.00	3.00	3.00
3	3.00	2.00	2.00	2.00	4.00	4.00	3.00
4	3.00	3.00	2.00	2.00	4.00	4.00	3.00
5	3.00	1.00	4.00	4.00	2.00	3.00	1.00
6	4.00	3.00	4.00	3.00	3.00	4.00	3.00
7	3.00	2.00	2.00	3.00	4.00	4.00	4.00
8	3.00	2.00	2.00	3.00	3.00	4.00	4.00
9	3.00	3.00	3.00	4.00	4.00	4.00	2.00
10	4.00	3.00	3.00	3.00	3.00	3.00	3.00

2. Exploratory factor analysis

Request exploratory factor analysis by selecting the **Statistics**, **Ordinal Factor Analysis** option from the main menu bar to access the **Ordinal Factor Analysis** dialog box. Mark all seven variables in the **Variable List** field and click **Select** to move them to the **Select a subset of variables** box. Note that this dialog box can also be used to specify the number of factors and whether a normal or logistic item response function is requested. We opt to set the **Item response function** field to the value of Logistic. We also set the **Number of Factors** field to 2. Click **Syntax** to generate a PRELIS syntax file.



```
SCITECH.PRL
!PRELIS SYNTAX: Can be edited
SY=SCITECH.LSF

SE 1 2 3 4 5 6 7
OFA POM NF=2
OU MA=CM XT XM
```

The analysis is performed by clicking the **Run Prelis** button on the main menu bar.

The first section of the output file gives univariate statistics for the 7 variables. The program reports that only 298 actual response patterns were observed, and lists the 20 most common patterns.

```
Total Sample Size(N) =    392

 Univariate Marginal Parameters

 Variable     Mean St. Dev.   Thresholds
 --------     ---- --------   ----------
  COMFORT    0.000    1.000  -2.234  -1.314   0.749
  ENVIRON    0.000    1.000  -1.447  -0.514   0.450
     WORK    0.000    1.000  -1.377  -0.428   1.079
   FUTURE    0.000    1.000  -1.803  -0.774   0.691
  TECHNOL    0.000    1.000  -1.686  -0.589   0.464
 INDUSTRY    0.000    1.000  -1.951  -1.056   0.219
  BENEFIT    0.000    1.000  -1.611  -0.500   0.845

 Univariate Distributions for Ordinal Variables

 COMFORT           Frequency Percentage Bar Chart
      1       5          1.3   □
      2      32          8.2   □□□□□□
      3     266         67.9   □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
      4      89         22.7   □□□□□□□□□□□□□□□□

 ENVIRON           Frequency Percentage Bar Chart
      1      29          7.4   □□□□□□□□□
      2      90         23.0   □□□□□□□□□□□□□□□□□□□□□□□□□□□□□
      3     145         37.0   □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
      4     128         32.7   □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

 WORK              Frequency Percentage Bar Chart
      1      33          8.4   □□□□□□□□
      2      98         25.0   □□□□□□□□□□□□□□□□□□□□□□
      3     206         52.6   □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
      4      55         14.0   □□□□□□□□□□□□□

 FUTURE            Frequency Percentage Bar Chart
      1      14          3.6   □□□
      2      72         18.4   □□□□□□□□□□□□□□□□
      3     210         53.6   □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
      4      96         24.5   □□□□□□□□□□□□□□□□□□□□□

 TECHNOL           Frequency Percentage Bar Chart
      1      18          4.6   □□□□□
```

```
     2      91        23.2     ▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒
     3     157        40.1     ▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒
     4     126        32.1     ▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒

INDUSTRY         Frequency Percentage Bar Chart
     1      10         2.6     ▒▒▒▒
     2      47        12.0     ▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒
     3     173        44.1     ▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒
     4     162        41.3     ▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒

BENEFIT          Frequency Percentage Bar Chart
     1      21         5.4     ▒▒▒▒▒▒
     2     100        25.5     ▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒
     3     193        49.2     ▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒
     4      78        19.9     ▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒

There are 298 distinct response patterns, see FREQ-file.
The 20 most common patterns are :
     11         3    3    3    3    3    3    3
      8         3    3    2    3    3    3    3
      7         4    4    3    3    4    4    3
      6         3    4    3    3    4    4    3
      5         3    4    3    4    4    4    4
      5         3    2    3    3    3    3    3
      4         3    3    2    3    3    3    2
      4         3    3    2    2    3    3    2
      4         3    2    3    3    2    2    2
      4         3    2    3    3    2    3    3
      3         3    3    3    3    4    3    3
      3         3    4    3    3    3    4    2
      3         4    1    4    4    2    3    4
      3         3    4    3    2    4    4    2
      3         3    4    1    3    4    4    2
      3         3    4    1    3    4    4    4
      3         3    3    3    3    2    2    3
      3         3    4    3    3    3    3    3
      3         4    4    4    4    4    4    4
      3         3    3    3    4    3    3    3
```

This is followed by three sets of factor loadings results. The unrotated factor loadings are followed by the variamax-rotated loadings. This solution is the varimax solution of Kaiser (1958). Next, we have the promax solution of Hendrickson and White (1964) in which the factors are correlation. Both these solutions are transformations of the unrotated solution and s such they are still maximum likelihood solutions.

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	-0.404	0.388
FUTURE	0.847	-0.323	0.179
TECHNOL	0.234	0.832	0.252
INDUSTRY	0.462	0.716	0.274
BENEFIT	0.713	-0.225	0.441

Varimax-Rotated Factor Loadings

	Factor 1	Factor 2	Unique Var
COMFORT	0.719	0.259	0.416
ENVIRON	-0.054	0.849	0.277
WORK	0.767	-0.153	0.388
FUTURE	0.906	-0.016	0.179
TECHNOL	-0.062	0.862	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	-0.086	0.852	0.277
WORK	0.774	-0.175	0.388
FUTURE	0.908	-0.041	0.179
TECHNOL	-0.094	0.866	0.252
INDUSTRY	0.161	0.826	0.274
BENEFIT	0.747	0.010	0.441

Finally, we have the reference variables factor loadings. The reference variables are chosen as those variables in the promax solution that have the largest factor loadings in each column. The main advantage of the reference variable solution is that standard errors can be obtained for all the variables except the reference variables. This enables the researcher to decide which loadings are statistically significant or not. Here the reference variables are FUTURE (loading of 0.906) and TECHNOL (loading is 0.865).

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	-0.140	0.388
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	-0.089	1.000

Additional output automatically produced during the analysis provide estimates of unstandardized factor loadings, threshold and their standard errors, and model fit information. For this analysis, the default names of these files are **SCITECH.POM**, **BIVFITS.POM** and **MULFITS.POM**.