Exploratory factor analysis with ordinal 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 this example, we use the simplest form of ordinal data, i.e. binary data, to illustrate. The data represents the observed frequencies for 32 possible response patterns from a law school admissions test (LSAT). Data for 5 items is available for a subsample of 1000 students. This smaller data set is used in Bartholomew & Knott (1999), but originally reported on by Bock and Lieberman (1970). The observed frequencies are given in the table below.

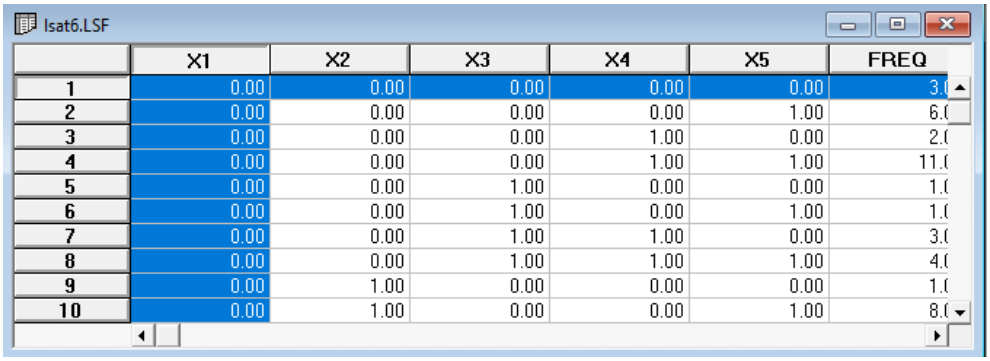
Index	Response pattern					
	1	2	3	4	5	Freq
1	0	0	0	0	0	3
2	0	0	0	0	1	6
3	0	0	0	1	0	2
4	0	0	0	1	1	11
5	0	0	1	0	0	1
6	0	0	1	0	1	1
7	0	0	1	1	0	3
8	0	0	1	1	1	4
9	0	1	0	0	0	1
10	0	1	0	0	1	8
11	0	1	0	1	0	0
12	0	1	0	1	1	16
13	0	1	1	0	0	0

14	0	1	1	0	1	3
15	0	1	1	1	0	2
16	0	1	1	1	1	15
17	1	0	0	0	0	10
18	1	0	0	0	1	29
19	1	0	0	1	0	14
20	1	0	0	1	1	81
21	1	0	1	0	0	3
22	1	0	1	0	1	28
23	1	0	1	1	0	15
24	1	0	1	1	1	80
25	1	1	0	0	0	16
26	1	1	0	0	1	56
27	1	1	0	1	0	21
28	1	1	0	1	1	173
29	1	1	1	0	0	11
30	1	1	1	0	1	61
31	1	1	1	1	0	28
32	1	1	1	1	1	298
Total						1000

The observed frequency for the first response pattern shows that 3 of the 1000 respondents failed to answer any of the 5 items correctly, while 298 respondents answered all items correctly. The frequency serves as a weight for each response pattern.

2. Exploratory factor analysis

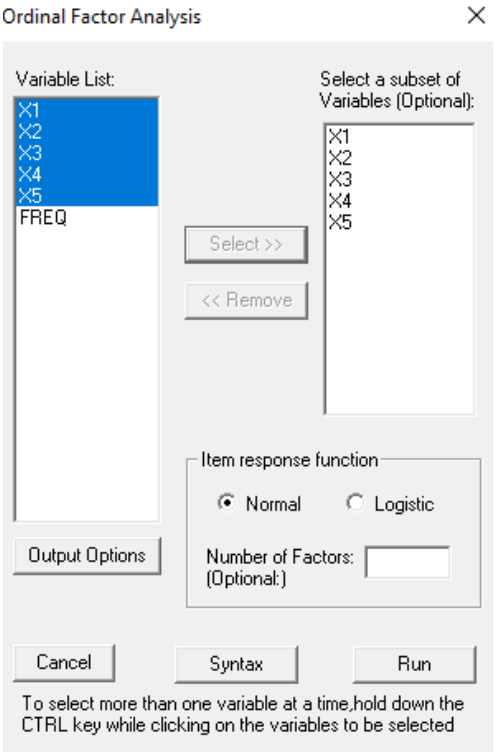
Start by opening the data file lsat6.lsf. The first few lines of this file are shown below.

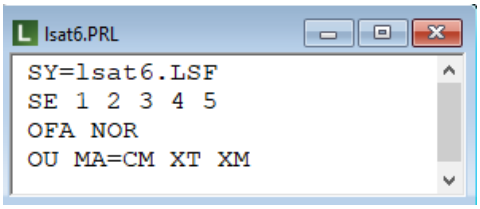


	X1	X2	X3	X4	X5	FREQ
1	0.00	0.00	0.00	0.00	0.00	3.(
2	0.00	0.00	0.00	0.00	1.00	6.(
3	0.00	0.00	0.00	1.00	0.00	2.(
4	0.00	0.00	0.00	1.00	1.00	11.(
5	0.00	0.00	1.00	0.00	0.00	1.(
6	0.00	0.00	1.00	0.00	1.00	1.(
7	0.00	0.00	1.00	1.00	0.00	3.(
8	0.00	0.00	1.00	1.00	1.00	4.(
9	0.00	1.00	0.00	0.00	0.00	1.(
10	0.00	1.00	0.00	0.00	1.00	8.(

As a first step, assign the variable FREQ as a weight using the **Data**, **Weight Cases** option from the main menu bar. After adding FREQ, click **OK** to return to the main window and save the LSF file.

Next, select the **Statistics**, **Ordinal Factor Analysis** option from the main menu bar to access the **Ordinal Factor Analysis** dialog box. Mark all five 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 leave the **Item response function** field at its default value of Normal. Click **Syntax** to generate a PRELIS syntax file.





The general form of the PRELIS command requesting factor analysis of ordinal variables is

```
OFA NF = k NOR
```

or

```
OFA NF = k POM.
```

In our example, we opted to not specify NF.

The first section of the output file gives univariate statistics for the 5 variables. We note that very few respondents answered the first question correctly, while the third question was not that hard for the sample of respondents considered here.

```
Total Sample Size(N) =   1000

 Univariate Marginal Parameters

 Variable      Mean St. Dev.   Thresholds
 --------      ---- --------   ----------
       X1    0.000    1.000  -1.433
       X2    0.000    1.000  -0.550
       X3    0.000    1.000  -0.133
       X4    0.000    1.000  -0.716
       X5    0.000    1.000  -1.126
```

```
Univariate Distributions for Ordinal Variables

X1              Frequency Percentage Bar Chart
     0      76         7.6    □□□□
     1     924        92.4    □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

X2              Frequency Percentage Bar Chart
     0     291        29.1    □□□□□□□□□□□□□□□□□□□
     1     709        70.9    □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

X3              Frequency Percentage Bar Chart
     0     447        44.7    □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
     1     553        55.3    □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

X4              Frequency Percentage Bar Chart
     0     237        23.7    □□□□□□□□□□□□□□□
     1     763        76.3    □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

X5              Frequency Percentage Bar Chart
     0     130        13.0    □□□□□□□
     1     870        87.0    □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
```

The univariate marginal parameters contain the mean, standard deviation and thresholds of the underlying variables used to estimate the tetrachoric correlations. For each item on the test, there was more correct than incorrect answers, so all thresholds are thus negative.

The output file also contains the unrotated factor loadings:

```
Unrotated Factor Loadings

            Factor 1 Unique Var
            -------- ----------
       X1      0.382      0.854
       X2      0.397      0.842
       X3      0.472      0.777
       X4      0.375      0.859
       X5      0.341      0.884
```

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 **LSAT6.NOR**, **BIVFITS.NOR** and **MULFITS.NOR**.