Multigroup model for ordinal data from 8 countries

Contents

1. Introduction 1
2. Exploratory data analysis 2
3. The model 4
4. Comparing 2 groups 5
5. Comparing all groups 9

1. Introduction

This example illustrates the analysis of the efficacy variables from the Political Action survey data, using data from eight countries. We use six political efficacy measurements as described in Aish & Jöreskog (1990).

The data are the responses to the following statements:

- NOSAY: "People like me have no say in what the government does."
- VOTING: "Voting is the only way that people like me can have any say about how the government runs things."
- COMPLEX: "Sometimes politics and government seem so complicated that a person like me cannot really understand what is going on."
- NOCARE: "I don't think that public officials care much about what people like me think."
- TOUCH: "Generally speaking, those who elect to Congress in Washington lose touch with the people pretty quickly."
- INTEREST: "Parties are only interested in people's votes but not in their opinions."

The ordered categories are:

- 1: agree strongly
- 2: agree
- 3: disagree
- 4: disagree strongly

Research questions of interest are:

- Do the efficacy items measure the same latent variable in all countries?
- If, are the factor loadings invariant over countries?
- Are the intercepts invariant over countries?

The latent variable of interest in this example is Efficacy, a component of political efficacy. Due to different languages used in the survey in the different countries, along with the problem that the same word may have different connotations in different countries, make complete factorial invariance over all 8 countries somewhat unlikely. It is also possible that cultural differences between the countries may influence either response style or response pattern.

2. Exploratory data analysis

For the first country, being the USA, use of the PRELIS syntax file

effitot1.prl

```
!Creating lsf file for USA
da ni=7
la
COUNTRY NOSAY VOTING COMPLEX NOCARE TOUCH INTEREST
ra = EFFITOT.RAW
cl COUNTRY 1=USA 2=GER 3=NET 4=AUS 5=BTN 6=ITY 7=SWI 8=FIN
cl NOSAY - INTEREST 1=AS 2=A 3=D 4=DS 6=DU 8=DK 9=NA
sd COUNTRY = 1
ou ra=effiusa.lsf
```

produces the following univariate statistics:

```
Univariate Distributions for Ordinal Variables

 NOSAY            Frequency Percentage Bar Chart
     AS     175         10.2     □□□□□□□□□□
      A     518         30.1     □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
      D     857         49.9     □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
     DS     130          7.6     □□□□□□□
     DK      29          1.7     □□
     NA      10          0.6     □

 VOTING           Frequency Percentage Bar Chart
     AS     283         16.5     □□□□□□□□□□□□□□□□□□□□
      A     710         41.3     □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
      D     609         35.4     □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
     DS      80          4.7     □□□□□
     DK      26          1.5     □□
     NA      11          0.6     □

 COMPLEX          Frequency Percentage Bar Chart
     AS     343         20.0     □□□□□□□□□□□□□□□□□□
      A     969         56.4     □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
      D     323         18.8     □□□□□□□□□□□□□□□□□
     DS      63          3.7     □□□
     DK       9          0.5
     NA      12          0.7     □
```

NOCARE		Frequency Percentage	Bar Chart
AS	250	14.5	
A	701	40.8	
D	674	39.2	
DS	57	3.3	
DK	20	1.2	
NA	17	1.0	

TOUCH		Frequency Percentage	Bar Chart
AS	273	15.9	
A	881	51.3	
D	462	26.9	
DS	26	1.5	
DK	60	3.5	
NA	17	1.0	

INTEREST		Frequency Percentage	Bar Chart
AS	264	15.4	
A	762	44.3	
D	581	33.8	
DS	31	1.8	
DK	62	3.6	
NA	19	1.1	

By replacing the country code in the second to last line of the syntax allows us to obtain similar results for any of the other countries. For example, the same statistics for Austria (country = 4) show that Austrians were more likely to agree strongly on the NOSAY question, for example, than did Americans where only 10.2% of responded in this category. Similarly, on the VOTING variable 41.3% of USA responded in the "Agree" category compared to 27.2 percent of Austrians, most of whom agreed strongly on this question. Turning to the variable INTEREST, 19.5% of Austrians disagreed compared to 33.8% of Americans.

NOSAY		Frequency Percentage	Bar Chart
AS	804	50.7	
A	432	27.3	
D	215	13.6	
DS	50	3.2	
DK	82	5.2	
NA	2	0.1	

VOTING		Frequency Percentage	Bar Chart
AS	903	57.0	
A	431	27.2	
D	135	8.5	
DS	28	1.8	
DK	85	5.4	
NA	3	0.2	

COMPLEX		Frequency Percentage	Bar Chart
AS	531	33.5	
A	557	35.1	
D	283	17.9	
DS	125	7.9	
DK	86	5.4	
NA	3	0.2	

NOCARE		Frequency	Percentage	Bar Chart
	AS	590	37.2	
	A	474	29.9	
	D	334	21.1	
	DS	86	5.4	
	DK	98	6.2	
	NA	3	0.2	

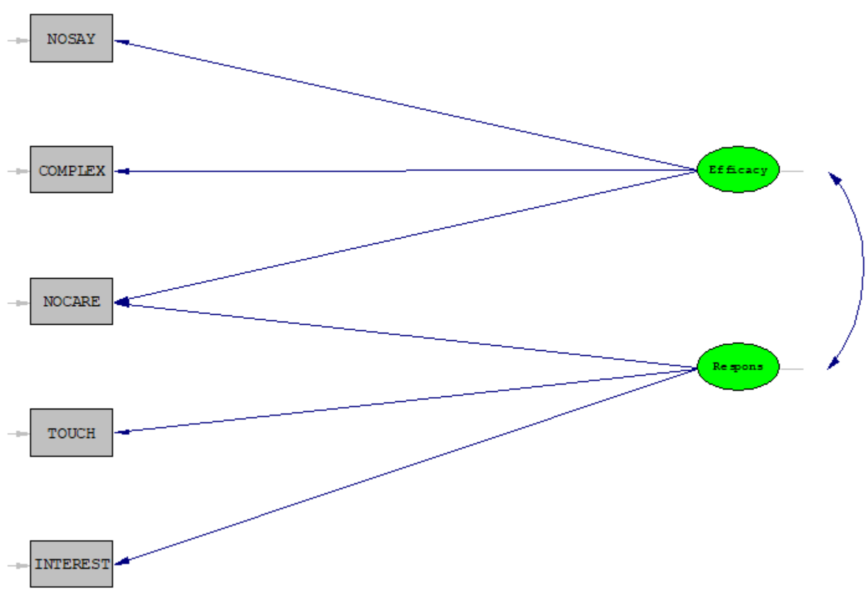
TOUCH		Frequency	Percentage	Bar Chart
	AS	602	38.0	
	A	533	33.6	
	D	257	16.2	
	DS	49	3.1	
	DK	140	8.8	
	NA	4	0.3	

INTEREST		Frequency	Percentage	Bar Chart
	AS	629	39.7	
	A	458	28.9	
	D	309	19.5	
	DS	76	4.8	
	DK	110	6.9	
	NA	3	0.2	

While results for the USA and the UK, both English speaking countries are similar to another, they are different from those of the German speaking countries (Germany and Austria). These differences may be indicative of differences in the means of the latent variables between countries. Note that in the examples to follow, we have opted not to include the variable VOTING.

3. The model

The conceptual diagram for the model of interest for each country is given below.



The model is defined by the cumulative response function

$$\Pr\left\{x_i \le c_i \mid \boldsymbol{\xi}\right\} = F\left(\alpha_{c_i}^{(i)} + \sum_{j=1}^{k} \beta_{ij}\xi_j\right) \tag{1}$$

with F representing a distribution function and $-\infty = \alpha_0^{(i)} < \alpha_1^{(i)} \ldots < \alpha_{m_i-1}^{(i)} < +\infty$. $\Pr\left\{x_i \le c_i \mid \boldsymbol{\xi}\right\}$ denotes the probability that an individual chooses a response in category c_i or lower on variable i. The $\alpha_s^{(i)}$ represent thresholds and the β_{ij} factor loadings. We will allow these to vary across groups and will test whether they are the same in all groups.

For this example, we opt to use the probit link function, defined as

$$F\left(u\right) = \Phi(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt.$$

Using standard LISREL notation, we can specify the model as

$$\mathbf{x}^{(g)} = \boldsymbol{\Lambda}_x^{(g)} \boldsymbol{\xi}^{(g)} \tag{2}$$

where $\mathbf{x}^{(g)}$ is a vector of observed variables of NOSAY, COMPLEX, NOCARE, TOUCH and INTEREST in group g. The vector of latent variables $\boldsymbol{\xi}^{(g)}$ is a scalar representing the latent variable Efficacy in group g with mean vector $\boldsymbol{\kappa}^{(g)}$, while $\boldsymbol{\Lambda}_x$ is a vector corresponding to the matrix $\left[\beta_{ij}\right]$ and is of order 5×1.

In the examples to follow, $\boldsymbol{\Lambda}_x$ is the vector

$$\boldsymbol{\Lambda}_x = \begin{bmatrix} 1 \\ \lambda_2^{(x)} \\ \lambda_3^{(x)} \\ \lambda_4^{(x)} \\ \lambda_5^{(x)} \end{bmatrix}.$$

Note that there are no intercept or error terms in equation (2). Instead, there are thresholds varying over variable, response categories and groups. We will assume the thresholds to be the same in all groups. This is necessary in order to estimate the means of the latent variable on the same scale for all 8 groups.

The parameter matrix $\boldsymbol{\Lambda}_x$ is considered an attribute of the observed items. It is assumed to be invariant over the groups. The first element, "1", is the unit of measurement in the latent variable. With these units the same across groups, we can compare the variances of the latent variable across the groups.

4. Comparing 2 groups

To start, we compare two English speaking countries, namely the USA and Britain. The data files contain some missing data, and as a result we will use full information maximum likelihood estimation (FIML). In order to estimate the model using the probit link function, we use adaptive quadrature.

effi1b.lis

```
Group: USA
!-----------------------------------------------------------------------
! Comparing USA and BTN
! Two-group Confirmatory Factor Analysis Model of Political Efficacy Items
!-----------------------------------------------------------------------
DA NG=2 NI=5
RA=effi_usa.lsf
$ADAPQ(10) PROBIT EQTH GR(5)
MO NX=5 NK=1 LX=FR KA=FR
LK
Efficacy
FI LX(1,1)
VA 1 LX(1,1)
OU

Group: BTN
DA
RA=effi_btn.lsf
MO LX=IN KA=FR
LK
Efficacy
PD
OU
```

The syntax file shown above is used to fit the model. In order to specify that the thresholds should be equal over groups, the EQTH keyword is invoked on the ADAPQ command.

The output file contains information on the effective sample size for both groups:

```
RA=effi_usa.lsf

          Note: there are     9 cases with missing values
          on all of the selected variables.
          These cases are excluded from the analysis.
          Effective sample size:     1710
RA=effi_btn.lsf

          Note: there are    27 cases with missing values
          on all of the selected variables.
          These cases are excluded from the analysis.
          Effective sample size:     1456
         LAMBDA-X
```

The lambda-X contain the estimated factor loadings common to both groups.

```
        LAMBDA-X

          Efficacy
           --------
    NOSAY     1.000

  COMPLEX     0.844
             (0.050)
              16.811

   NOCARE     2.217
             (0.121)
              18.394
```

```
    TOUCH      1.991
             (0.114)
              17.533

 INTEREST      2.397
             (0.145)
              16.489
```

The estimated variance and mean of the latent variable Efficacy for the two countries are given in the Phi and Kappa matrices.

USA:

```
         PHI

            Efficacy
            --------
               0.419
             (0.089)
               4.728
KAPPA

            Efficacy
            --------
               0.087
             (0.060)
               1.451
```

Britain:

```
         PHI

            Efficacy
            --------
               0.419
             (0.089)
               4.728

KAPPA

            Efficacy
            --------
               0.087
             (0.060)
               1.451
```

No χ^2 is available when a model is estimated using FIML for adaptive quadrature. However, the $-2\ln L$ may be used to compare models. The number of free parameters correspond to the sum of estimated parameters in $\mathbf{\Lambda}_x$ (4), $\mathbf{\Phi}$ (1 each per group), $\boldsymbol{\kappa}$ (1 each per group) and the 15 thresholds.

```
     Number of quadrature points =              10
     Number of free parameters =                23
     Number of iterations used =                 8

     -2lnL (deviance statistic) =      29615.52401
```

```
Akaike Information Criterion      29661.52401
Schwarz Criterion                 29800.90917
```

We have 5 ordinal variables in the model, each with 4 categories. So a total of 15 thresholds will be estimated. Results for these are given below. As these were constrained to be equal in all groups, these results are the same for the US and Britain.

```
Threshold estimates and standard deviations
-------------------------------------------
      Threshold             Estimates          S.E.      Est./S.E.

TH1_NOSAY                    -1.38974        0.06499     -21.38279
TH2_NOSAY                     0.05325        0.06183       0.86112
TH3_NOSAY                     1.91481        0.07332      26.11455
TH1_COMPLEX                  -0.92635        0.05473     -16.92484
TH2_COMPLEX                   0.83089        0.05572      14.91171
TH3_COMPLEX                   2.07541        0.06980      29.73301
TH1_NOCARE                   -1.86838        0.12670     -14.74684
TH2_NOCARE                    0.56501        0.13503       4.18427
TH3_NOCARE                    3.23096        0.17966      17.98417
TH1_TOUCH                    -1.50420        0.12339     -12.19055
TH2_TOUCH                     0.96193        0.11908       8.07783
TH3_TOUCH                     3.46062        0.15450      22.39939
TH1_INTEREST                 -1.71136        0.18047      -9.48272
TH2_INTEREST                  0.81913        0.12247       6.68836
TH3_INTEREST                  3.80682        0.00000       0.00000
```

To test the hypothesis of factorial invariance, we use the modified syntax file shown below. Note that LX = IN on the MO line is the second group has been replaced by LX = FR and that two additional lines

```
FI LX(1,1)
VA 1 LX(1,1)
```

have been added.

For this model, the deviance statistic is

```
Number of quadrature points =              10
 Number of free parameters =                27
 Number of iterations used =                10

 -2lnL (deviance statistic) =      29483.35782
 Akaike Information Criterion      29537.35782
 Schwarz Criterion                 29700.98388
```

The difference between the deviance statistics is 132.166, and 4 more parameters have been estimated here. The hypothesis of factorial invariance can thus be tested using a χ^2 of 132.166, with 4 degrees of freedom. It is clear that the hypothesis must be rejected.

When we repeat this analysis for the two German speaking countries (see **effi2b.lis**, **effi2b.spl**) we reach the same conclusion.

5. Comparing all groups

From the results in Section 4, it seems unlikely that the hypothesis of factorial invariance over groups will hold when the complete data set with 8 countries is used. However, this analysis will be briefly discussed here as a way to estimate the latent mean of Efficacy for each country on the same scale. The syntax file for the combined group is given below.

effi3b.lis

```
!-----------------------------------------------------------------------
DA NG=8 NI=5
RA=effi_usa.lsf
$ADAPQ(10) PROBIT EQTH GR(5)
MO NX=5 NK=1 LX=FR KA=FR
LK
Efficacy
FI LX(1,1)
VA 1 LX(1,1)
OU

Group: GER
DA NI=5
RA=effi_ger.lsf
MO LX=IN KA=FR
LK
Efficacy
OU

Group: NET
DA NI=5
RA=effi_net.lsf
MO LX=IN KA=FR
LK
Efficacy
OU

Group: AUS
DA NI=5
RA=effi_aus.lsf
MO LX=IN KA=FR
LK
Efficacy
OU

Group: BTN
DA NI=5
RA=effi_btn.lsf
MO LX=IN KA=FR
LK
Efficacy
OU

Group: ITY
DA NI=5
RA=effi_ity.lsf
MO LX=IN KA=FR
LK
Efficacy
OU

Group: SWI
DA NI=5
RA=effi_swi.lsf
MO LX=IN KA=FR
LK
Efficacy
OU

Group: FIN
DA NI=5
RA=effi_fin.lsf
MO LX=IN KA=FR
LK
Efficacy
OU
```

The order in which countries are given does not matter, but the results will depend on which variable is chose to define the scale of the latent variable Efficacy. Here we are using NOSAY.

The common estimated factor loadings are

	Efficacy
NOSAY	1.000
COMPLEX	0.896
	(0.025)
	36.223
NOCARE	1.927
	(0.052)
	37.246
TOUCH	1.742
	(0.050)
	34.876
INTEREST	1.863
	(0.056)
	33.459

while the estimated latent means by country are obtained as

USA	-0.145
	(0.028)
	-5.187
Germany	0.103
	(0.029)
	3.568
Netherlands	-0.139
	(0.030)
	-4.686
Austria	0.402
	(0.034)
	11.852
Britain	0.000
	(0.029)
	0.017
Italy	0.266
	(0.029)
	9.136
Switzerland	0.068
	(0.035)
	1.966
Finland	0.111
	(0.031)
	3.537

Estimated latent means vary between -0.145 for the USA and 0.402 for Austria. It is clear that the differences between countries first noted in the data screening carry through to the estimated latent variable means. Combining these results with the lack of evidence supporting factorial invariance across groups we conclude that the countries perception and evaluation of the terms in question are not the same.