Course 7: Understanding and Modelling Measurement Error in Social Surveys

Exercises Tuesday

I) Introduction to SQP: http://sqp.upf.edu [together]

1) In the questionnaire of the round 3 of the European Social Survey, look for the 4 different forms of the questions asking about if it is bad or good for the economy that people come to live here from other countries (should look at the main + the three supplementary questionnaires, one form in each).

2) What are the characteristics of the 4 different forms?

3) Which one do you expect to have the best quality? Justify.

4) Now, go to SQP. Find the reliability, validity and quality of each 4 forms for Great Britain when using the MTMM experiments and when using the prediction.

5) Do the same for France, Spain and the Netherlands.

6) Which kind of differences do we observe and what can be their effects for the analyses?

7) In each country, tell which form of the question should be used to maximize the quality.

8) Can we conclude something about characteristics of scales that should be avoided and characteristics that should be used?

9) SQP also gives the confidence intervals: what do they tell us?

II) Code a question in SQP to get the quality prediction [together]

1) In SQP, choose one question for which the authorized prediction is available (within ESS round 3). Write down the question, and its authorized quality prediction.

2) Go to the ESS questionnaire and look for the exact formulation of the question.

3) /!\ For this question, please use the DEMO version of SQP! Code this question in SQP until you obtain a quality prediction. Compare with the quality in the authorized prediction. Did you get the same? If not, compare the codes of the authorized prediction and yours to see where the differences come from.

4) Look for the suggestions of improvements. If there are some clear suggestions, try to implement them and see if the quality really improves.

III) Estimate simple measurement models in Mplus

Please, note that sometimes it is normal in the exercises to get an error message from Mplus. This is to show you situations that can happen when analysing data.

1) Use the dataset "ESSround3-trust-day2-n.dat". The dataset is in a fixed format with 12 variables of format 2. The 12 variables are the following:

- "cntry": 2 letters that correspond to the country
- A8 A9 A10 B4 B5 B6 B7 B8 B9 B10: correspond to the questions with the same names in the ESS questionnaire of round 3 (main questionnaire).
- "cntrynum": number from 1 to 25 corresponding to the country. The numbers have the following meaning:

1=Austria, 2=Belgium, 3=Bulgaria, 4=Switzerland, 5=Cyprus, 6= Germany, 7=Denmark, 8=Estonia, 9=Spain, 10=Finland, 11=France, 12=great Britain, 13=Hungary, 14=Ireland, 15=Latvia, 16=Netherlands, 17=Norway, 18=Poland, 19=Portugal, 20=Romania, 21=Russia, 22=Sweden, 23=Slovenia, 24=Slovakia, 25= Ukraine.

Missing values are coded 66, 77, 88 and 99.

Use the questions about trust in institutions.

a) **[together]** Specify the Mplus input for a factor model with 1 factor (trust in institutions) and 1 indicator (question B4 in the ESS round 3 questionnaire). Ask for the standardized estimates and for the modification indices.

Look at the path diagram and at the output.

What does the identification issue refer to?
What is the necessary condition for identification?
When is a model over-identified?
What can be done with an over-identified model that cannot be done with a just identified one?

b) Specify the Mplus input for a factor model with 1 factor (trust in institutions) and
- 2 indicators (B4-B5)
- 3 indicators (B4-B5-B6)
- 4 indicators (B4-B5-B6-B7)

Look at the path diagram and at the output.

Which estimator was used by default by Mplus?
How many observations do we have in the dataset?
How many parameters had to be estimated?
Where can you find the unstandardized estimates?
Where can you find the standardized ones?

2) Use again the dataset “ESSround3-trust-day2-n.dat”. Use again trust in institutions as one of the factor; and use social trust as a second factor.

a) Estimate using Mplus a factor model with these 2 correlated factors and
- 2 indicators each (B4-B5 and A8-A9)
- 3 indicators each (B4-B5-B6 and A8-A9-A10)

Do we have any identification problem here?

b) Repeat the analyses but selecting only the data for Germany. Then, do it only for UK.

NB: we should NOT look at the estimates and make conclusions before having tested the model, but we will speak about testing only in the next days. So for today’s exercise, let’s assume that the fit of our model is OK and that we can look at the estimates.

Then, compare the standardized estimates for the model with 3 indicators (UK versus Germany).

For each factor and country, tell which is the indicator that better measure it?

IV) Working on your own research

1) Go online and try to find data about the variables for which you formulated questions yesterday.

2) Compare these questions with the ones you made. Give pros and cons for their formulation and yours.

3) Go to SQP. Select one of the variables. Code in SQP your formulation and the one you found in existing data. Compare the quality of both.

4) Use SQP to improve the question. Decide of the final formulation you would use for this variable if you would have to collect your own data. Also tell what you will do with the “don’t know” and which mode you will use to collect the data with highest quality possible.