Applications of Multilevel Models to Research Policy and Higher Education Studies
Computer Lab Session
Lugano January 8th, 2015

The aim of this computer lab to is help you get acquainted with basic multilevel analysis using STATA.

Research Scenario

Suppose you are a social scientist who is interested in what determines people's perceptions of their country's education system. You suspect that a person's opinion about their education system is driven by both individual and contextual level factors. Using the existing theory to guide your analysis, you would like to understand both how a person's educational background AND the performance of their country's education system influence their opinion.

Data Background

The data comes from the latest round of the European Social Survey (ESS) that has been released in 2012. The ESS is a 36-country repeat cross-sectional survey that is released every two years since the year 2002. The survey population intends to cover people 15 years and older, with no upper age limit, who are resident in the country, regardless of nationality, citizenship or legal status. The sample is selected by strict random probability methods at every stage and respondents are interviewed face-to-face. In each round, the minimum effective sample sizes are at least 1,500 (or 800 where the population is less than two million inhabitants) in each participating country. The minimum target response rate set by ESS is 70 per cent.

The standardized ESS questionnaire includes both core models and rotating models and covers a range of social topics. The ESS is particularly suitable since it occurs frequently and has a nationally representative random sample for each European member state. The survey is designed to be comparable across countries so the questions are translated to have not only linguistic equivalence but also have functional equivalence.

Part A. Get to know the dataset

1. Identify the nested structure of the data and the number of level-2 unit (countries) and the number of level-1 units (individuals).

2. Become acquainted with the dependent variable "state of education nowadays" (*stfedu*) run a tabulation and generate descriptive statistics using the summarize command. Make note of the direction of the scale to assist in the interpretation of the results.

3. Do the same for the potential independent variables and the individual level: years in education (*eduyrs*) or education level (*edulvl);* at the country level*:* tertiary education enrollment (c_isced56_2011), secondary enrollment (*c_gerupsec_2010*) , Reading PISA score (*cnrtry_readPISA12),* Math PISA score (*cnrtry_mathPISA12),* GDP per capita *(cntry_GDP)*

4. Finally, explore potential individual control variables. For instance you could look at: *hinctnta*, *mainact*, *brncntr*, *gndr*, *agea*, amongst others that you are curious about in the dataset.

Part B. Partition the variance of the dependent variable

1. Build a null model (a multilevel model without an independent variable) first to partition the variance. Hint: command for a multilevel model is XTMIXED and details about the syntax have been distributed.

2. Calculate the interclass coefficient by hand. Note: In the future you can download a program called XTMRHO and it will calculate it for you after you have build a multilevel model.

3. Does the dependent variable have enough variance at level-2 (countries) to justify employing multi-level modeling?

Part C. Building a Random Intercept Model

To build our random-intercept model, we shall proceed in a stepwise fashion that is commonly used when presenting the results of multilevel modelling for publication. This means, we shall build our full model in blocks to ensure that adding additional variables into the equation improves the model.

1. Build a multilevel model with random intercepts. In this first iteration, only include the independent and dependent variable. Compare the goodness-of-fit statistics with your null model.

2. Next introduce your individual controls into the model. Note any improvement in the model fit.

3. Now introduce your contextual control variable(s) into the model in addition to the individual controls.

Part D. Cross Level Interactions

Let us now consider if the performance of the country's education system has heterogeneous effects on residents' opinions about the education system. For instance, we expect that individuals who are more educated are more likely to be satisfied by the education system when PISA scores are higher. Now we would like to test this.

1. Introduce the cross level interaction terms into the random intercept model that you have built. Some of these variables have been created for you in the data set (*eduyrsxPISAread eduyrsxPISAmath).* However if you would like to use other cross-level interaction variables, you may create them using the command GERNERATE.

2. What conclusions can you draw from the coefficients of these interaction terms?

Part E. Extending to a Random Coefficient Model

1. Next build a random coefficient model using the XTMIXED command. Remember to allow your independent variable of interest to vary randomly.

2. Calculate the covariance matrix between intercepts and slopes by using the post-command estat recovariance.

3. Let's calculate the intercepts and slopes for each country.
   You can use the following syntax:
   predict u*, reffects
   bysort country: generate groups=(_n==1)
   list country u2 u1 if country<=24 & groups

4. To make sure the added covariance term (age) is significant, let's run an LR test.
   You can do that with the following syntax:
   quietly [insert full xtmixed command with random intercept only]
   estimates store ri
   quietly [insert full xtmixed command with random coefficient]
   lrtest ri rc

5. With this information, what can you say about the random coefficient model? What does it tell us?

Part F. If you have extra time…

1. If you have extra time, you can try to explore other models that take into account additional interaction terms or control variables to help us better understand the how the performance of the country's education system is related to Europeans' opinions about education in their country.

2. If you would like to use the variables in the dataset to run a multi-level logistic model you may do so with the command XTLOGIT. Remember, in order to do so, you will need to transform your dependent variable into a dichotomous one.