SOC 30005:
Assignment 3: Two-Level Analysis of Covariance
with Multiple Imputation of Missing Data
Due May 18, 2019

Our aim in this assignment is the same as it was in Assignment 2: To study the impact of cognitive strategies on writing on writing (CSIW). But now we are going to confront and solve two key problems that were ignored in Assignment 2: a) the nested character of the data; b) missing data.

As before, our analysis will be restricted to the following variables, except that we add the school id:

CSIW (1=CSIW, 0=control)
Achievement Level (1=High, 2=Average, 3=Low, 4=Learning Disability)
Holistic pretest (pre-test on writing achievement)
Holistic posttest (posttest on writing achievement)
Grade (1=Grade 4, 2=Grade 5)
School (school identification number)

You should have already re-coded Achievement Level into 4 dummy variables and grade into 1 dummy variable. As always, whenever you write down a model, make sure all terms are defined and all assumptions stated. If you want to discuss a table or plot, paste it directly into the text. Do not include any appendices. Do not include any tables or figures that you do not discuss in the text. As before, groups of 3 or less can collaborate; the paper should be signed by all authors, each of whom will receive the same grade.

A. Nested Data

Run a cross tab in which the rows are the schools ("school") and rows are CSIW. Based on this, tell us at what level the treatment ("CSIW") varies.

B. Build the MDM file

Sort the cases by "school." The level-1 and level-2 files will be the same. Choose the relevant child level variables at level-1 and the relevant school variable at level 2.
B. Analysis of covariance (ANCOVA) model. Indicate that there are missing data, and tell the program to delete missing level-1 cases at run time.

C. Analysis of Covariance

1. Write down the level-1 model with relevant covariates (do not include quadratic terms or interactions at this time).
2. Write down the level-2 model.
3. Estimate the model and tell us the estimated treatment effect and its standard error (model based and robust).

4. For comparison, delete the random effect. How do the estimated treatment effect and standard error (model-based versus robust) compare with those obtained when you included the random effect. Which model is preferable and why?

D. Missing Data

1. What is the total sample size?
2. How many cases were used in the analysis?
3. Re-run the random effects model but now use “automatic imputation.”
4. Note how many cases were used in the “averaged results.”
5. How do the results compare with those in your original random effects model.

E. Interpretation

This assignment, unlike our others, has been pretty mechanical. Sorry about that! We’ll discuss the modeling framework, assumptions, and interpretations in detail in class!