Statistical Methods: Multilevel Analysis

February 24, 2012

Please submit your reports (with Declaration of Authorship attached) to Reception in SPR 1, before 10 am Monday Week 7 (February 27th 2012). Since this is a group project, in case there was a distribution of tasks the report should contain a brief note about who did what. But note that the whole report is joint responsibility of all group members.

The file pupils_8.txt (made available in zipped form) contains data of eighth grade pupils in a state in the US, with the following variables.

dis_id	District Identifier
sch_id	School Identifier
stud_nm	Student Identifier
sex	Coded 1 if Female and 2 if Male
ethnicity	Coded 1 if White, 2 if Black, 3 if Hispanic, 4 if Asian, 5 if Other.
reading	California Test of Basic Skills Reading Score
math	California Test of Basic Skills Math Score
sch_size	School-Level Size (centered Natural Log)
sch_ses	School-Level SES (social-economic status) variable (centered)

Analyze how the reading score depends on available pupil and school characteristics. The math score should be left out of the analysis, as it would not be a very meaningful independent variable here. Unfortunately there is no individual-level SES variable available.

There are some missing values in the data set. You may assume that these are missing at random, and, for the purpose of this practical, simply omit the cases with any missing values from the multilevel analysis.

This is a three-level data set, because of the districts. Analyze it first as a two-level data set (pupils nested in schools); in a second stage analyze it as a three-level data set; the first (two-level) part is more important than the second (three-level) part. It is recommended that you complete the analysis and report of the two-level model before proceeding to the three-level model.

Your report should not have more than 10 pages per group (but fewer pages is fine!!!), and should pay attention to descriptives (including descriptives for the missing data pattern); analysis plan; results; assumption checks; and interpretation, the latter in terms understandable for a non-technical audience. The R code used should be given as an appendix (not included in the count of maximally 10 pages).