4/1/2010 36-402/608 ADA-II H. Seltman

Breakout #19 Results

These data come from Snijders and Bosker's book "Multilevel Analysis", chapter 5 (partly from: http://www.ats.ucla.edu/stat/sas/examples/mlm_ma_snijders/ch5.htm).

There are multiple students within each school, and school is identified with the "schoolnr" variable. The outcome is a score on a language test labeled "langpost".

The per-school explanatory variables are "GndC_size" (a centered variable indicating school size); "GrpMC_verb" (a centered average verbal IQ for each school); and "GrpMC_ses" (a centered variable indicating the mean socio-economic status for each school).

Different from the above is "GndC_verb" which is a (centered) student level verbal IQ variable indicating each student's IQ relative to the mean for his or her school.

Question 1: List some of the unmeasured school-level variables.

Question 2: Now ignoring those variables, very roughly sketch a plot of the means model whose fixed effects are defined by

MODEL langpost = GndC_size GrpMC_verb GrpMC_ses GndC_verb;

You should invent signs for the βs and not worry about the magnitudes. One effective technique is to plot the means model for the 1/3 and 2/3 quantile values of GndC_size, GrpMC_verb, GrpMC_ses.

If you used the "tercile" method suggested above, you will have 8 lines on a plot of GndC_verb vs. langpost, with each line representing a different kind of school (and other kinds of schools un-plotted but imagined between or beyond the plotted lines). To consider random effects, think about several schools all with the same values of the three school-level variables. You can do this by picking any one of the 8 lines on your plot from question 2. If there are no important student-level variables, note that for an "average" school all of the individual student points would fall on the line you picked. If there are important student-level variables, then the points randomly fall above or below that line with a spread based on the residual variance, σ^2.

Now consider what will happen if these several schools differ on important school level variables. Initially, to simplify things, again imagine that there are no important student-level variables. **Question 3: Add "prediction" lines for the several schools all with the same levels of the measured school-level variables.**

You should see that you drew lines with different intercepts and slopes, but with a mean intercept and slope that gives the original line.

This exercise shows how to think about which random effects are possible, and how they relate to the fixed effects models (and why random effects are defined to have mean zero).

Now relate what you learned here to the idea that a random effect can be represented as adding a mean-zero per-group (per-school, here), random variable to any beta value in the (fixed) means model, assuming that it makes sense for that quantity to vary from group to group.

Here is a simple model:

```
OPTIONS LINESIZE=66;
LIBNAME here ".";
TITLE 'Snijders and Bosker School Data, Chapter 5';
TITLE2 "Random slope by IQ model";
proc mixed data=schools2 covtest noitprint noclprint method=ml;
  class schoolnr;
  model langpost = GndC_verb GrpMC_verb / solution outpred=RSres;
  random intercept GndC_verb / subject=schoolnr type=un;
run;
```

Dimensions

Covariance Parameters	4
Columns in X	3
Columns in Z Per Subject	2
Subjects	131
Max Obs Per Subject	35
Number of Observations Used	2287

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	schoolNR	7.9177	1.3287	5.96	<.0001
UN(2,1)	schoolNR	-0.8198	0.2914	-2.81	0.0049
UN(2,2)	schoolNR	0.1994	0.1003	1.99	0.0234
Residual		41.3525	1.2902	32.05	<.0001

BIC (smaller is better) 15247.7

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	40.7498	0.2859	129	142.54	<.0001
GndC_verb	2.4588	0.08315	130	29.57	<.0001
GrpMC_verb	1.4052	0.3214	2025	4.37	<.0001

Question 4: What are there 131 of?

Question 5: Make a rough fixed effect plot based on the model fit here, plotting GndC_verb at -2 and +2 (roughly IQs 80 and 120). Ignoring the difficulty of making a plot that incorprates the negative correlation of the intercept and slope, pick the mean line representing one particular level of school mean IQ and draw several lines for separate schools consistent with the Covariance Parameter Estimates.

In practice we would do a lot of model selection and checking of residual plots at this stage. For the Breakout, we will just look at the best model, which also include the indicator variable for whether the school uses mixed-grade classrooms.

```
proc mixed data=schools2 covtest noclprint noitprint method=ml dfbw;
  class schoolnr;
  model langpost = GndC_verb GndC_ses GrpMC_verb mixedgra GndC_verb*mixedgra /
     solution outpred=RSparsRes;
  random intercept GndC_verb / subject=schoolnr type=un;
run;
```

Dimensions

Covariance Parameters	4
Columns in Z Per Subject	2

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	schoolNR	7.5574	1.2562	6.02	<.0001
UN(2,1)	schoolNR	-0.5890	0.2587	-2.28	0.0228
UN(2,2)	schoolNR	0.1277	0.08389	1.52	0.0640
Residual		39.3402	1.2253	32.11	<.0001

BIC (smaller is better) 15142.1

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	41.3213	0.3490	129	118.40	<.0001
GndC_verb	2.1134	0.09245	2152	22.86	<.0001
GndC_ses	0.1555	0.01464	2152	10.63	<.0001
GrpMC_verb	0.8754	0.3237	129	2.70	0.0078
mixedgra	-1.3961	0.5743	2152	-2.43	0.0151
GndC_verb*mixedgra	0.4472	0.1701	2152	2.63	0.0086

Question 6: Summarize what this model claims.