Homework Problems on Mixed-Effects Models

Soc. 761

Fall 2014

The file `Snijders.txt` (available on the course web site) contains data on 4106 grade-8 students (who are approximately 11 years old) in 216 primary schools in the Netherlands. The data are used for several examples, somewhat different from the analysis that we will pursue below, by Snijders and Boskers in *Multilevel Analysis, 2nd Edition* (Sage, 2012).

The data set includes the following variables:

- `school`: a (non-consecutive) ID number indicating which school the student attends.
- `iq`: the student's verbal IQ score, ranging from 4 to 18.5 (i.e., *not* traditionally scaled to a population mean of 100 and standard deviation of 15).
- `test`: the student's score on an end-of-year language test, with scores ranging from 8 to 58.
- `ses`: the socioeconomic status of the student's family, with scores ranging from 10 to 50.
- `class.size`: the number of students in the student's class, ranging from 10 to 42; this variable is constant within schools, apparently reflecting the fact that all of the students in each school were in the same class.
- `meanses`: the mean SES in the student's school, calculated from the data; the original data set included the school-mean SES, but this differed from the values that I computed directly from the data.
- `meaniq`: the mean IQ in the student's school, calculated (for the same reason) from the data.

Read the data into R in the usual manner. There are some missing data, and I suggest that you begin by using `na.omit` to filter out cases with missing data. How many students are lost when missing data are removed in this manner? Then add the following two variables to the data set:

1. school-centred SES, computed as the difference between each students' SES and the mean of his or her school; and
2. school-centred IQ.

Questions:

1. Using Trellis graphics (i.e., the R **lattice** package), examine scatterplots of students' test scores by centered SES and centred IQ for each of 20 randomly sampled schools. (See the script `Snijders.R` on the course website if you need some help drawing these graphs.) Do the relationships in the scatterplots seem reasonable linear? *Hint*: In interpreting these scatterplots, take into account the small number of students in each school, ranging from 4 to 34 in the full data set.
2. Using the `lmList` function in the **lme4** package, regress the students' test scores on centred SES and centred IQ within schools for the full data set. Then plot each set of coefficients (i.e., starting with the intercepts) against the schools' mean SES, mean IQ, and class size. Do the coefficients appear to vary systematically by the school's characteristics (that is, by the Level 2 explanatory variables centred SES, centred IQ, and class size)? (Once again, you'll find R commands for these computations and graphs in `Snijders.R`.)

3. Fit linear mixed-effects models to the Snijders and Boskers data, using `lmer` in the **lme4** package, and proceeding as follows:

   (a) Begin with a one-way random-effects ANOVA of test scores by schools. What proportion of the total variation in test scores among students is between schools (i.e., what is the intra-class correlation)?

   (b) Fit a random-coefficients regression of test scores on the students' centred SES and centred IQ. Initially include random effects for the intercept and both explanatory variables. Test whether each of these random effects is needed, and eliminate from the model those that are not (if any are not). How, if at all, are test scores related to the explanatory variables? *Hint:* In using the `anova` function to test variance and covariance components for the random effects, remember to set the argument `refit=FALSE` to retain the REML estimates. *Note*: You may obtain a convergence warning in fitting one or more of the null models that remove variance and covariance components; this warning will not prevent you from performing the likelihood-ratio test for the corresponding random effects.

   (c) Introduce mean school SES, mean school IQ, and class size as Level 2 explanatory variable, but only for the Level 1 coefficients that were found to vary significantly among schools in part (b). *Hint:* Recall that modeling variation in Level 1 coefficients by Level 2 explanatory variables implies the inclusion of cross-level interactions in the model; and don't forget that the intercepts are Level 1 coefficients that may depend on Level 2 explanatory variables.

       Test whether the random effects that you retained in part (b) are still required now that there are Level 2 predictors in the model. *Note*: Again, you may obtain a convergence warning.

   (d) Using the `Anova` function in the **car** package (with the argument `test="F"`), compute Type II tests of the various main effects and interactions in the model fit in part (c). Then simplify the model by removing any fixed-effects terms that are non-significant. Finally, interpret the results obtained for the simplified model. If your final model includes interactions, you may wish to use the `Effect` function in the **effects** package to visualize the interactions.