Biostatistics

Task3: Mixed models

1. The dataset `Cultivation` is part of the package `SASmixed`. It corresponds to an experiment in which 4 blocks were used, each block was divided in half, and two varieties of grass were assigned to each half. Each plot occupied by a variety was divided into three and each was inoculated with a different bacterium.

Fit the appropriate random effects model, test the fixed effects, and check the model assumptions.

2. The dataset `maths.txt` contains the following data:

- `math.8`: a math-test score when the student was eight years old.
- `math.11`: a current math-test score. `female`: a dummy variable coded 1 for girls and 0 for boys.
- `manual`: a dummy variable coded 1 if the student's parent (presumably the main wage earner) is in a manual occupation and 0 otherwise.
- `school`: a number indicating which school the student attends.

Add the following two variables to the data set:

– The mean age-8 math score in the student's school.

– The deviation between the student's age-8 math score and the mean score in her/his school (i.e., compute the school-centred age-8 math score).

i) Using the `lmList` function in the `nlme` package (this function will allow you to fit a model for each school at the same time), regress age-11 math scores on centred age-8 scores and the dummy variables for gender and class. Look at the within-schools coefficients. Why are some missing? Then plot each set of coefficients (i.e., starting with the intercepts) against the school mean age-8 math scores.

Do the coefficients appear to vary systematically by the school's mean age-8 scores?

ii) Fit linear mixed-effects models to the *Maths* data, proceeding as follows:

- Begin with a random-intercept model of age-11 math scores by schools.

  How much of the variation in age-11 scores is *between* schools?

- Fit a random-coefficients regression of age-11 math scores on the student's centred grade-8 scores, gender, and class. Initially include random effects for the intercept and all three explanatory variables. Test whether the random effects are needed and eliminate from the model those that are not. Check the significance of the fixed effects. Interpret the coefficients.

- Introduce the mean school age-8 math score as a level-2 explanatory variable, but only for the level-1 coefficients that were found to vary significantly among schools in part (*ii*).

  Test whether the random effects which are in the model, are still required now that there is a level-2 predictor in the model.

- Briefly summarize your findings.

3. The dataset `eating.txt` contains data on the exercise histories of 138 teenage girls hospitalized for eating disorders and on a group of 93 control subjects.

The variables are:

- `subject`: an identification code; there are several observations for each subject, but because the girls were hospitalized at different ages, the number of observations and the age at the last observation vary.
- `age`: the subject's age in years at the time of observation. All but the last observations for each subject were collected retrospectively at intervals of two years, starting at age 8.
- `log.exercise`: A transformation of the amount of exercise in which the subject engaged, expressed as estimated hours per week. The transformation consisted in taking logs (and using logs to the base 2 for interpretability), but because there are some 0 values of exercise, five minutes (5/60 of an hour) were added to each value of exercise before taking logs.
- `group`: a factor indicating whether the subject is a *patient* or a *control*.

i) Fit regressions of `log.exercise` on `age` for each subject and look at the coefficients as in the previous exercise.

ii) Fit a model for this longitudinal dataset considering the appropriate fixed and random effects.

Test the effects and give the appropriate conclusions.

4. Take two examples from Chapter 3, one from the *Multilevel Models* section and the other from the *Longitudinal data and repeated measurements* section. Apply and justify a Bayesian approach and explain the obtained results.

See, for example as references, these websites:

`https://mc-stan.org/users/documentation/case-studies/tutorial_rstanarm.html`
`https://m-clark.github.io/mixed-models-with-R/bayesian.html`
`https://cran.r-project.org/web/packages/brms/index.html`