Directed Studies VHM 881 Course
Fall 2009, Atlantic Veterinary College, PEI
Henrik Stryhn

Home Assignment 2

The home assignment consists of 5 parts (below), which all should be answered. The home assignment is due on Wednesday, December 2, 2009. It is worth 15% of the course mark. All aids except personal assistance are allowed.

The topic of the home assignment is the impact of informative cluster sizes on the estimators in linear mixed models. The first two parts considers models without covariates, the next two parts deals with models with covariates, and the final applied part is about detecting informative cluster sizes in real data analysis.

Part 1
Exercises E 12.4 and 12.5 in the textbook (McCulloch, Searle & Neuhaus, 2008).

Part 2
Carry out a small simulation study to confirm the presence of the bias in the sample mean determined by the formulae proven in Part 1. Use a model of the type assumed for Equation (12.12), with values of $\gamma_0 \neq 0$ and $\gamma_1 \neq 0$, and a range of values for σ_b^2. Estimate also the correlation between the responses and cluster size, and graph the bias as a function of this correlation.

Part 3
Exercise E 12.6 in the textbook. Explain the difference between the ordinary and generalized least squares estimators, and determine also the limiting distribution of the latter.

Part 4
Extend the simulation study from Part 2 to include a covariate of similar type as used in the simulations for Table 12.1 of the textbook (note that you should still use a linear mixed model instead of a logistic mixed model). Summarize the results for the bias of estimators of β_0 and β_1 in a similar way as in Part 2, and draw conclusions.

Part 5
Consider one (e.g. the first) of your simulated datasets from Part 4. Carry out a statistical analysis to demonstrate the presence of informative cluster sizes in this dataset (without utilizing the knowledge you have about the true model for the data). Summarize your approach into a recommendation of how to detect instances of informative cluster sizes in datasets in general. As a demonstration of your approach, determine whether there seems to be informative cluster sizes in the `scc40_2level` dataset with a suitably chosen cluster variable and a suitably chosen predictor. The dataset is contained on the CD-ROM for the Epi on the Island 2009 summer course.