

# Homework 6: GLMMs and GEEs

your name

Due: April 19th at 11:59 PM

**Problem 1** [25 points]: The `toenail` data in the `faraway` package comes from a multicenter study comparing two oral treatments for toenail infection. Patients were evaluated for the degree of separation of the nail. Patients were randomized into two treatments and were followed over seven visits: four in the first year and yearly thereafter. The patients have not been treated prior to the first visit so this should be regarded as the baseline. Do the following:

- **part a** [5 points]: Calculate the proportion of patients with a normal or severe condition broken down by treatment and visit number. Plot these proportions and comment.
- **part b** [10 points]: Fit a GLMM for the outcome as a function of an interaction between the visit and the treatment. Since the two groups are selected at random, there should be no difference at the first visit. Does this model show a significant difference at this baseline (first visit)?
- **part c** [10 points]: Test for a significant treatment effect by fitting a model without treatment and comparing to the previous model.

Problems 2 and 3 below will ask you to compare GLMM and GEE fits. Your comparisons should include differences in interpretations of the  $\beta$ s obtained via GLMM and GEE, and the assumptions placed upon the data for each approach. [This paper](#) may be useful, but it is not required reading. Keep in mind that the hyperlinked reference that is included is critical of GLMM, perhaps overly critical.

**Problem 2** [25 points]: The National Youth Survey collected a sample of 11–17 year olds, 117 boys and 120 girls, asking questions about marijuana usage. The data is presented in the `potuse` data set in the `faraway` package. Do the following:

- **part a** [4 points]: Plot the total number of people falling into each usage category as it varies over time separately for each sex.
- **part b** [4 points]: Condense the levels of the response into whether the person did or did not use marijuana that year. Turn the year into a numerical variable. Fit a GLMM for the now binary response with an interaction between sex and year as a predictor using Gauss-Hermite quadrature. Comment on the effect of sex.
- **part c** [4 points]: Fit a reduced model without sex and use it to test for the significance of sex in the larger model.
- **part d** [4 points]: Fit a model with year as a factor. Should this model be preferred to the model with year as just a linear term? Interpret the estimated effects in the year as a factor version of the model.
- **part e** [4 points]: Fit your final model using PQL, Bayesian methods, and MCLA. Compare the results and discuss any inconsistencies if any arise.
- **part f** [5 points]: Fit GEE version of the model and compare it to the analogous GLMM fits.

**Problem 3** [25 points]: The `nitrofen` data in the `boot` package come from an experiment to measure the

reproductive toxicity of the pesticide **nitrofen** on a species of zooplankton called *Ceriodaphnia dubia*. Each animal produced three broods in which the number of live offspring was recorded. Fifty animals in total were used and divided into five batches. Each batch was treated in a solution with a different concentration of the pesticide. Do the following:

- **part a** [5 points]: Plot the total number of live offspring as they vary with concentration and comment. Now plot the numbers for each brood, taking care to distinguish the different broods. Is the trend within each brood the same?
- **part b** [10 points]: Fit a GLMM for the number of live offspring within each brood that varies with concentration and brood number (including an interaction). The model should take account of the dependence between observations from the same animal. Describe what the model says about how the number of live offspring change with concentration for the different broods. Provide reasons for choosing your selected GLMM implementation, or compare the GLMM implementations that were discussed in class.
- **part c** [10 points]: Fit an equivalent GEE model and compare it to the GLMM result(s).

**Problem 4** [25 points]: Show that a GLM in canonical form can be cast as a GEE. Explain your work in detail.