Homework6

BIOS6643 Fall 2021

Due Wed 11/24/2021 at midnight

Question 1

We have learned that Generalized Linear Mixed Models (GLMM)s are basically a combination of LMMs and Generalized Linear Models. One of the complexities of GLMMs is that maximizing the (true) likelihood involves using a numerical technique like Gaussian quadrature. In 2-3 sentences, explain what makes the GLMM likelihood more complicated than the LMM likelihood, in terms of maximization.

Question 2

A study is planned where data will be collected on asthmatic subjects on every weekday for one month. There are two outcome measures of interest, (i) medication use counts and (ii) FEV1. You are the statistician and the PI is looking for your suggestions about models to use.

a. If it is anticipated that responses within subjects over time are serially correlated (but with some decay the further measurements are apart) for both outcomes, which R function (and package) would you suggest using to fit the data? Answer separately for each outcome.

b. Suppose that we now consider an indicator of whether subjects used medication or not on a given day (no $use = 0$, at least 1 $use = 1$). In this case, the researcher is more concerned about accounting for general differences between subjects in the model (e.g., on one extreme there may be heavy users and on the other, very light users) than accounting for serial correlation (although the latter may still exist). What procedure would you suggest using if you were very interested in accounting for between-subject variability of use? What are the drawbacks of this approach?

c. For **part b**, suggest a procedure you might use if you wanted to include both a random intercept for subjects in the model, as well as account for potential serial correlation of repeated measures. What are the drawbacks of this approach?

Question 3

A randomized double blinded trial of 294 patients was conducted to compare 2 oral treatments, corresponding to Itraconazole (treat=0) Terbinafine (treat=1)), for toenail infection. The response variable was the binary indicatro of presence of onycholysis (separation of the nail plate from the nail bed). Patients were evaluated for onycholysis at baseline (week 0) and at weeks 4, 8, 12, 24, 36 and 48. Suppose the interest is in finding the effect of treatment on changes in an individual's risk of onycholysis over time. The data are available in the *toenail-data.txt* file; response is a binary indicator for moderate or severe (response=1) versus none or mild (response=0) onycholysis.

a. Fit a GLMM model with month and the interaction between treatment and month, and random intercepts. Assume that there is not difference between treatments at baseline. Assume the scale parameter $\phi = 1$ and the random intercepts follow a normal distribution. Interpret the results of the model, including the coefficient of month and that of the interaction.

b. Fit a GEE model with the same mean structure as in a. Interpret the regression coefficients in the model and comment on the differences with the regression coefficients in a.

Question 4

Recall the epilepsy data that we have used in class (Thall 1988 Biometrics). Use the code provided to fit a GLMM and GEE model in SAS. The RANDOM statement functions just as in PROC MIXED, so there are options G and GCORR to print out the estimated among-individual covariance matrix G. See the proc glimmix documentation for more information on syntax, required statements, and options.

Recall we use the epilepsy data in lab 13. Here is the R code we used to read the data in, which shows the name of the variables.

```
dat.sz <- read.table("/Users/juarezce/Documents/OneDrive - The University of Colorado De
colnames(dat.sz) <- c("subj","seize","visit","trt","base","age")
## trt=0 corresponds to placebo
## seize= number of seizures

#  Create other covariates
dat.sz$o <- 8*(dat.sz$visit==0)+2*(dat.sz$visit>0)
dat.sz$logo <- log(dat.sz$o)
dat.sz$vm0 <- as.numeric(dat.sz$visit>0)
```

SAS code

```
/*Note vm0= 1 if visit>0; otherwise vm0=0*/
title "RANDOM INTERCEPT ONLY, ADAPTIVE QUADATURE";
```

```
proc glimmix data=seizure method=quad(qpoints=25);
  class subj;
  model seize = age vm0 trt trt*vm0 / solution link=log
        dist=poisson offset=logo;
  random int / subject=subj type=un;
run;

title "PROC GENMOD, UNSTRUCTURED CORRELATION";
proc genmod data=seizure;
  class subj visit;
  model seize = age vm0 trt trt*vm0 /  dist = poisson link = log offset=logo;
  repeated subject=subj / within=visit type=un corrw covb modelse;
run;
```

a. Fit a GLMM using the glimmix code above and interpret the results.

b. Fit a marginal model using the genmod code above to fit a marginal model. Interpret the results.

Note PROC GLIMMIX can also be used without random effects to fit population-averaged models. The R structure of a GLMM may be specified through definition of the structure of the errors using *random* residual*;* (this option is similar to the REPEATED statement in PROC MIXED).