**BIOS6643 Fall 2019 HW 6 Due November 8**

1. 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.
   1. 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, what SAS procedure (or R package/function) would you suggest using to fit the data? Answer separately for each outcome.
   2. Related to a, talk about how you would set up the data and specify the REPEATED statement in SAS (or comparable code for R) for each outcome, so that the correlation between responses is accounted for properly, including gaps caused by no measurements on weekends. (NOTE: to deal with unequal spacing in GzLM/GEE, you need to include records for equally spaced time points and fill in with missing values as necessary, e.g., for weekends as described above; we will discuss this more soon.)
   3. Say 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 big users and on the other, very little users) than accounting for serial correlation (although the latter may still exist). What procedure would you suggest using if you wanted to account for between-subject variability of use, and also approximate the true likelihood in estimation? What are the drawbacks of this approach?
   4. For part c, 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?
2. In class we have discussed the albuterol data, which involves children who take rescue medication for their asthma. They take this ‘as needed’ (i.e., based on how they feel) but are also prescribed to take it (i.e., ‘pre-treats’). We would like to see how the daily albuterol use counts relate to air pollution measures, controlling for other covariates. The air pollution variable used here is ln(morning hourly maximum PM2.5). To help account for the pre-treats, the indicator variable Friday was included in the model (the one day they did not receive pre-treats since there is no gym class). Often meteorological variables are also controlled for in air pollution models; here we will include temperature, pressure and humidity. Complete the following.
   1. Run GEE and use the MODELSE option in the REPEATED statement to incorporate the scale parameter into the GEE process. Highlight the results. Does adding the scale parameter into the process modify the SE’s up or down? Use the AR(1) working covariance structure.
   2. Now fit the GzLMM using RSPL approach (the default method in PROC GLIMMIX). Include a spatial power structure to account for serial correlation (use ‘date’ as the indexing variable). Highlight the results. How does the scale parameter in GEE compare with the residual variance in the GzLMM PL approach? (Recall that the residual variance in GzLMM PL acts as the scale parameter; make sure to compare apples-to-apples, though; see the updated s13 slides.)
   3. Do the scale / residual variance estimates suggest over or under-dispersion in the data (considering Poisson distribution)?
   4. How do slope estimates and SE’s for ln(mmPM2.5) differ between the GEE and GzLMM PL approach? What about these SE’s compared to the empirical SE of GEE (which is also given in default output)?
   5. In a sentence, interpret the relationship between morning particulate matter (1 hour maximum) and children’s albuterol use. In order to make the slope more meaningful, interpret the effect per SD increase in the pollutant variable. Use the GzLMM PL estimate to do this.
3. Consider the exacerbation data fit using a GzLMM using pseudo-likelihood estimation, accounting for both serial correlation and subject heterogeneity by including a random intercept; results shown on the right side of slide 20 of the s13 GzLMM linearization slides.
   1. Manipulate and interpret the parameter estimates for B\_DAY and B\_WKEND for the layperson. Does your answer change if you rescale the day effect to a week or month? Do the two SAS approaches differ much?
   2. Do the slope estimates have subject-specific or population-averaged interpretations? Explain.
   3. If you were to fit the data using GzLM/GEE, how would you expect the beta estimates to change, relative to those using the GzLMM fits. Explain. (Think in terms of SS versus PA effects.)