**BIOS6643 Fall 2016 HW 8 (Solutions to be posted)**

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 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 for each outcome, so that the correlation between responses is accounted for properly, including gaps caused by no measurements on weekends. In doing this, recall all that was mentioned in class and in the notes regarding missing data.
   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. Consider the generalized linear mixed model , where , , and  is the jth row of **Z***i*, the covariate matrix for subject *i*, associated with random effects **b***i*. Write explicit forms of E(*Yij*| **x***ij*, **b**i=0) and E(*Yij*|**x***ij*) for outcome variables with the following distributions. (Sometimes conditioning on **x** is not stated but implied.)
   1. Normal
   2. Binomial
   3. Poisson
3. Consider the albuterol use data fit with PROC GENMOD with GEE, shown on slide 21 of Non-normal I slides.
   1. Often, slope estimates are expressed for a more common increase in the predictor, rather than ‘per unit’ increase, to improve their interpretability. Determine the relative increase in albuterol use for a 10 μg/m3 increase in mmaxpm25 (morning maximum PM2.5). Note that we did this in class so it is basically review.
   2. Another way to get meaningful slope estimates so that slopes of different predictors can be compared with each other is to standardize them per SD increase or IQR (interquartile range) increase. Determine the relative increase in doser use, per IQR increase in mmaxpm25, and compare it with that of temperature

(ͦ F) and relative humidity (%). The IQRs for mmaxpm25, temperature and humidity are 11, 16 and 29, respectively. Interpret the results. (Note that standardized estimates do not rely on original units.)

* 1. If there are random intercept differences for subjects in the population, will the estimates in parts a and b have subject-specific or population-averaged interpretations (or both)? Explain.

1. Consider the exacerbation data fit using a GzLMM using pseudo-likelihood estimation, results shown on the right-hand side of slide 15 of the Non-normal II slides.
   1. Manipulate and interpret the parameter estimates for B\_DAY and B\_WKEND for the layperson.
   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.
2. Name one advantage and one disadvantage of fitting the exacerbation discussed above using a GzLMM with Gaussian quadrature. What estimation approach is used?