**BIOS6643 Final Quiz 2016**

You are allowed to use any course material in completing this quiz. Do not contact other students to discuss the exam, even for clarification. If you have any questions regarding this quiz, please contact me at 303-398-1862, or send me an e-mail: [strandm@njhealth.org](mailto:strandm@njhealth.org). Regarding question #3, I have also attached the GzLMM likelihood detail notes, in Word version so that you could use it as a template and not have to build all of the equations from scratch (if you use Word). If you want to use something else like LaTeX, that is fine, but are more on your own. You can space out the questions below according to your needs. If you write your answers and will not be using Word or LaTeX, just use separate pages and clearly write the question numbers with your work. Unless other arrangements are made, your work is due back by 1pm. Have fun! Try to write sufficient, succinct answers.

1. You conduct a longitudinal study involving subjects with and without a certain gene. It turns out that subjects with the gene drop out of the study more often than those without. In addition, the relationship between the outcome and time differs between those with and without the gene. If we have the genetic information for the subjects, do you expect there to be problems in estimating coefficients of interest in the longitudinal model (i.e., time-related coefficients) with respect to gene differences? Explain, and discuss what terms you would include in the model in order to carry out the analysis.
2. You have binary longitudinal data, and you fit a model to estimate the relationship between the health outcome [*y*: currently sick (0), currently healthy (1)] and several key predictors, including nutrition level (*x*). You consider fitting the data using 2 different approaches: (i) one that includes a random intercept for subjects to account for general differences in health, and (ii) the other that does not have random effects but takes into account serial correlation of the repeated measures (e.g,. using an AR(1) structure).
   1. Mention the model and method (real briefly) that you would use for these 2 approaches.
   2. For the approaches you suggested, can you compare model fits using goodness-of-fit statistics? Explain.
   3. It turns out that the variance for random intercept is relatively large. How would you expect this to affect the slope of *x*, and how would you interpret the slope for these 2 approaches (subject-specific or population-averaged)?
   4. Now say that you want account for both subject heterogeneity as well as serial correlation. What sort of approach would you take here?
3. Consider an analysis to be done on longitudinal count data using Poisson regression in a generalized linear mixed model (GzLMM); a random intercept will be included for subjects and a simple linear regression for time will be included (including intercept, *β*0 and slope, *β*1). The probability mass function of the Poisson is for *y*=0,1,2,…, where *λ* is mean and variance of *Y*. Write the likelihood function for the GzLMM in terms of *β*0 and *β*1. You will need to leave the integrands in your solution; as with the example given in class, use *i* to denote subjects and *j* for time.
4. Regarding the previous question, suppose that we find the Poisson distribution to be too stringent for the model (specifically, the requirement that the mean and the variance are equal). Mention one alterative way to model longitudinal count data that offers more flexibility in fitting the data than the true Poisson.
5. A study is conducted on subjects that come in for multiple hospital visits. Ideally they should come in once a year, but it turns out that they often come in at different times or even miss visits. A longitudinal model will be used to examine health outcomes as a function of time and other predictors.
   1. If the outcome can be modeled with a mixed model, what type of covariance structure would you suggest to account for repeated visits for subjects?
   2. If we’re now considering a binary outcome, how would you account for the repeated measures?