**BIOS6643 Homework 2 (Pseudo review) Fall 2019**

1. In 6612 you may have been told that for a class variable in a GLM or LMM, you need to pick one level as the reference, and that the estimates for other levels become a comparison to the reference. Using concepts of estimability in a less-than-full-rank model, prove that this is true. To make the argument, consider SAS’s approach to determining a g-inverse, such as we did in class.
2. Derive  in a full-rank linear mixed model, given the algebraic form of  that is obtained via ML estimation. NOTE: there are two types of variance, model-based and empirical (or sandwich ☺). The difference is whether the middle ‘V’ is determined via the model or using squared residual quantities; derive the model-based form. To answer this question, work with the ‘complete data’ form of .
3. For a linear mixed model that could either be full rank or less than full rank, for  that is estimable, determine . A couple of tips: (a) the M-P inverse of a square symmetric matrix is also symmetric, (b) we know that  (in a GLM) is unique, i.e., not dependent on the g-inverse used; it is also true that  is unique in an LMM.
4. In a short paragraph, explain the difference between a general linear model (GLM; not a general*ized* linear model, which I denote with GzLM and which will be discussed more later) and a linear mixed model (LMM).
5. For the either the Dog data or Beta Carotene data, design and compute 2 contrasts and 2 estimates (other than those done in class or previously). Create your tests and estimates based on what you think is interesting. With the output, write up your results in a few sentences.
6. Consider a study where subjects in 3 groups (e.g., race or treatment) are observed over 3 equally spaced times and some health outcome, y, is measured. Unless otherwise mentioned, include a random intercept for subjects to account for the repeated measures. For simplicity, use 2 subjects per group.
   1. Consider modeling group and time as class variables, plus interaction. Write statistical models and the X matrix for the following cases.
      1. No restriction placed on the model. I.e., write the less-than-full-rank statistical model.
      2. A set-to-0 restriction is placed on the parameters associated with highest levels.
7. Show that the linear trend for one group compared to another (say Group A versus B) is estimable by showing that **L**=**LH**, where the Moore-Penrose inverse is used in calculating **H**. First you need to construct **L**. (As a check, you can repeat using SAS’s g-inverse in calculating **H**, but you don’t need to turn that in.)
8. How would answers in a change an AR(1) structure for **R** is included? (You do not need to rewrite entire models, just what changes).
9. Say that Time is treated as continuous (i.e., not included in the CLASS statement in SAS or factor argument in R). Rewrite the models and X matrices in a. Say the linear term for Time is sufficient. **SEP 23 UPDATE: you only need to do this for one model (less-than-full-rank or full rank).**
10. Say that the times of observation were at 0, 1 and 6 months rather than equally spaced.
    * 1. Would it be appropriate to treat time as a class variable in this case? Explain.
      2. Suggest a structure for **R***i* and write it out.