BIOS6643 HW3 Due Friday, September 28, 2018, 5pm

1. Consider a basic science experiment conducted where cell counts are measured at 4 time points for samples taken from individual subjects or animals. A linear mixed model will be fit for the data (perhaps after log transformation), and fixed effects will be included for time, and possibly treatment group as well as their interaction. (To answer this question we do not need to know the specific form of **Xβ**.) Determine the structure for **V***i* if a random intercept for subjects will be included, plus an AR(1) structure for the error covariance matrix (**R***i*). What does the combination of non-simple **R** and **G** allow you to do in modeling covariances that using only one cannot do? Discuss in a few sentences.

**Solution**: Recall that . Here, **Z***i* is just a column of 1’s since there is only a random intercept, so . Also,  based on the AR(1), so we have 

**The advantage of this structure is that you get some decay between covariances as the gap between time points is increased, but it does not necessarily decay to 0. This is helpful for correlated longitudinal data where some correlation between responses is expected regardless of gap between responses, but it is not as simple as the CS structure, either. Of course, we might be able to find covariates that explain the subject heterogeneity, which might then eliminate the need for the random intercept component.**

1. One model we used for the Mt. Kilimanjaro data included random effects for subject, up to the quadratic term (plus covariances between random effects), along with a simple **R** structure. (We did find at least one model with a better AIC, but let’s focus on this one for now.) We talked about how including multiple random effects can induce a covariance structure that is time sensitive (or in this case, altitude sensitive). Show this by considering a simple data set and model. In particular, let times be *t*=0, 1, 2, and consider a model that includes a random intercept and slope for time by subject, plus covariance between them (i.e., UN structure in **G**). Show that it is possible to obtain Cov(*Yi*1,*Yi*2) > Cov(*Yi*1,*Yi*3) < Cov(*Yi*2,*Yi*3), i.e., decaying covariance as distance between time points is increased. For what covariance parameter values will these hold?

**Solution**: , so 

**To get the ‘decay’ pattern we need . Note that since variances are changing over time, we could also consider finding constraints on the correlations, which might be even more meaningful. However, this is an algebraic nightmare and would be aided by some algebra software.**

1. Fit the Beta Carotene data using a continuous model for time, including group and group\*time in the model. (For a description of the data, see the file in the HW folder.) Determine the degree of polynomials for time that is important and sufficient for the model. For covariance structure, define the UN structure for **R**.
   1. Write your final model, fit it and compare it to the model that used group, time and group\*time as class variables. Which would you go with in a final report? Explain. NOTE: in comparing model AIC’s use method=ML for a more apples-to-apples comparison, particularly when changes are being made to the fixed effects.

**We have 5 time points. If you include up to cubic terms for time as well as the group-by-time interaction, you actually get a lower AIC (1243.1) than the model that treats time as class (1245.3). But given they are so close, you could probably go with one or the other depending on other criteria, such as whether you want to do inference between time points, or whether you want complete flexibility in estimating group-by-time combinations.**

**LTFR model: . We could further trim off the *i* subscript on *x* if all subjects have the same time points. If they actually had slightly different time points, keep it, though.**

* 1. Write an estimate or contrast statement for your continuous model based on what you think is interesting. The custom estimate and/or test could involve a subset of the data (e.g., comparing 2 specific groups), or the whole data.

**Below is the SAS code for the curve comparison for the 2 BASF groups. Note that this looks a little different than the curve comparison in the notes (Mt. K application) since I am treating prepar as a class variable. When I say ‘curve comparison’, I mean anything that will differentiate the functions, including intercept differences. Thus, this is a very general test. If you have any questions on this, let me know.**

\*LTFR model;

**proc** **mixed** order=data data=univar method=ml;

class id prepar;

model y= prepar time time\*time time\*time\*time prepar\*time prepar\*time\*time

prepar\*time\*time\*time / solution;

repeated / subject=id(prepar) type=un;

estimate 'test' prepar **0** **0** **1** **0**;

contrast 'BASF curve comparison' prepar **0** **0** **1** -**1**, prepar\*time **0** **0** **1** -**1**,

prepar\*time\*time **0** **0** **1** -**1**, prepar\*time\*time\*time **0** **0** **1** -**1**; **run**;

| **Contrasts** | | | | |
| --- | --- | --- | --- | --- |
| **Label** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **BASF curve comparison** | 4 | 19 | 4.56 | 0.0095 |

1. Consider a study where children are sampled from schools, and then measured over time. We will include a random intercept for schools and for subjects within schools (but simple **R**). Determine **V***h*, the covariance matrix for school *h*, if there are 3 children sampled from this school, where the first two kids have 3 measures and the last has 2. You might find it helpful start by writing the model for outcome *Yhij* and determining the design matrix for the random effects. (You can just write something generic for the fixed-effect part of the model.) For thought, not to turn in: how would **V***h* change if we had more measures for subjects and employed the AR(1) structure for **R***i(h)* (the error covariance structure for subject *i* within school *h*)?

, ,; . Combining these, we get

