Homework4

BIOS6643 Fall 2021

8/20/2021

Question 1 $\boldsymbol{R}$ and $\boldsymbol{G}$ matrix

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 $\boldsymbol{X\beta}$.)

Determine the structure for $\boldsymbol{V}_i$ if a random intercept for subjects will be included, plus an AR(1) structure for the error covariance matrix ($\boldsymbol{R}_i$). What does the combination of non-simple $\boldsymbol{R}$ and $\boldsymbol{G}$ allow you to do in modeling covariance that using only one cannot do? Discuss in a few sentences.

Question 2 Covariance for random effects

One model we used for the *Mt. Kilimanjaro data* included random effects for subject, up to the quadratic term (plus covariance between random effects), along with a simple $\boldsymbol{R}$ 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 $\boldsymbol{G}$). Show that it is possible to obtain $Cov[Y_{i1},\ Y_{i2}] > Cov[Y_{i1},\ Y_{i3}] < Cov[Y_{i2},\ Y_{i3}]$, i.e., decaying covariance as distance between time points is increased. For what covariance parameter values will these hold?

Question 3 Kronecker product

Consider the Dog Data (see course notes, p. 121). I have used this interesting data set for years but only recently someone pointed out to me that it involves ‘doubly repeated measures’. Previously, I just thought ID’s were nested within group, and that it was classical repeated measures over time. But the ID’s are unique, it’s just that we have repeated measures over treatment as well as time. Thus, the main purpose of this exercise is to fit the data using a *Kronecker Product* for the error covariance matrix that accounts for repeated measures over time as well as treatment.

a. First, use group and time as class variables, plus *group* $\times$ *time*, and determine a Kronecker Product structure that will work for these data. (Note: there are 3 options in SAS, and you may be limited by what will work.) Highlight results.

b. Add the R and RCORR options in the REPEATED statement. (Note that this is equivalent to the fitted $\boldsymbol{V}$ matrix since there are no random effects.) In 3-4 sentences, interpret the correlations and variances in the data.

c. Compare your model in **a** with another approach to modeling the correlated data. For example, you could try a random intercept for subjects, which is a simple but ‘better than nothing’ approach. But you can use another approach if you have another idea. Is model **a** better?

d. Try modeling time as continuous (modify model **a**). In order to get this to work, in SAS you need to define a second time variable that is identical to time (e.g., add t=time in the data step). Use ‘t’ in model statement, and ‘time’ in the class and repeated statements. Compare the AIC’s for the 1st, 2nd and 3rd degree time models (add polynomial for both time and *group timestime*). Note: in making comparisons, use method=ML, since REML does not have fixed effects in the likelihood. (If you go up to the 4th degree, note that you have saturated the fixed effects and should get the same AIC as model **a**.)

e. With your optimal model found in part **d**, go back to method=REML and conduct a custom test for interaction between the 2 treatments (ignoring the control treatment). Highlight results.

Question 4 Fixed and random effects

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 $\boldsymbol{R}$). Determine $\boldsymbol{V}_h$, 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 Y_{hij} 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 $\boldsymbol{V}_h$ change if we had more measures for subjects and employed the AR(1) structure for $\boldsymbol{R}_{i(h)}$ (the error covariance structure for subject i within school h)?