BIOS6643 **Homework 5** **2016**

<u>*To turn in:*</u>

(1) Consider data collected at three times, morning, noon and evening over 3 straight days. Thus, each subject has 9 measurements. We discussed this scenario in class.

a. *Review*: Say that the outcome measure is y=cortisol, which is related to stress levels. Explain why just using the AR(1) structure for repeated measures might not be the best model for such data.

<u>Solution</u>: the main problem is that you cannot allow for the possibility that correlation between days at the same time of day might be stronger than correlation within a day between times. We might expect cortisol to be higher in the morning, say, when people are getting ready for their day, rushing to work, etc. Thus, you could potentially see a stronger correlation between morning cortisol levels on two consecutive days than between cortisol levels from morning to noon of the same day. Applying one AR(1) structure to the 9 repeated measures would not allow for this to happen, because it forces a decay on correlation as the time between measurements increases.

b. If 'morning', 'noon' and 'evening' have consistent meanings across days, we can consider 'day' and 'time of day' as crossed factors that constitute two different types of repeated measures. Consider the AR(1)$\otimes$AR(1) Kronecker Product structure for these data that we discussed in class. What is the correlation between the morning measurement on one day with the evening measurement on the following day (in terms of parameters)?

<u>Solution</u>: if we let ϕ denote the correlation between consecutive days and ρ denote correlations between consecutive times within a day, then the correlation is $\phi\rho^2$.

c. Suppose that we find the AR(1) structure for repeated measures within a day too restrictive. Suggest another Kronecker Product structure for the data, and determine the correlation between the morning measurement on one day with the evening measurement on the following day for your choice.

<u>Solution</u>: there are many possibilities, but I would probably try the AR(1) structure for repeated measures over days and the UN structure for repeated measures within a day. This would also work well if we were to include more days in the study, since we won't have additional parameters for the AR(1) structure as we add days. The correlation of interest is $\phi\sigma_{13}/(\sigma_1\sigma_3)$. Of course the 'correct' answer depends upon the structure you use (this is for the AR(1)$\otimes$UN), and there were some other reasonable suggestions, such as AR(1)$\otimes$TOEP.

(2) Consider a study where families are recruited for a study where cholesterol levels of members within families are measured at two different times.

a. Identify the levels of this 3-level data.

<u>Solution</u>: Level 3 (big unit) is the family, level 2 (medium unit) is the subject, level 1 (small unit) is the time of measurement within subject.

b. If we include random *family* and *subject within family* terms and allow for UN structure for the error covariance matrix for subject i, write the $\mathbf{V}_i$ matrix for a family with 3 members involved in the study. (The dimension of the matrix should be 6×6.)

Solution:

$$\mathbf{V}_h = \mathbf{Z}_h \mathbf{G}_h \mathbf{Z}_h^t + \mathbf{R}_h\,,\ \mathbf{R}_h = \begin{pmatrix} \mathbf{R}_{1(h)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{2(h)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}_{3(h)} \end{pmatrix},\ \mathbf{R}_{i(h)} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix};\ \mathbf{Z}_h = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{G}_h = \begin{pmatrix} \sigma_F^2 & 0 & 0 & 0 \\ 0 & \sigma_S^2 & 0 & 0 \\ 0 & 0 & \sigma_S^2 & 0 \\ 0 & 0 & 0 & \sigma_S^2 \end{pmatrix}.$$ Combining these, we get

$$\mathbf{V}_h = \begin{pmatrix} \sigma_F^2+\sigma_S^2+\sigma_1^2 & \sigma_F^2+\sigma_S^2+\sigma_{12} & \sigma_F^2 & \sigma_F^2 & \sigma_F^2 & \sigma_F^2 \\ \sigma_F^2+\sigma_S^2+\sigma_{12} & \sigma_F^2+\sigma_S^2+\sigma_2^2 & \sigma_F^2 & \sigma_F^2 & \sigma_F^2 & \sigma_F^2 \\ \sigma_F^2 & \sigma_F^2 & \sigma_F^2+\sigma_S^2+\sigma_1^2 & \sigma_F^2+\sigma_S^2+\sigma_{12} & \sigma_F^2 & \sigma_F^2 \\ \sigma_F^2 & \sigma_F^2 & \sigma_F^2+\sigma_S^2+\sigma_{12} & \sigma_F^2+\sigma_S^2+\sigma_2^2 & \sigma_F^2 & \sigma_F^2 \\ \sigma_F^2 & \sigma_F^2 & \sigma_F^2 & \sigma_F^2 & \sigma_F^2+\sigma_S^2+\sigma_1^2 & \sigma_F^2+\sigma_S^2+\sigma_{12} \\ \sigma_F^2 & \sigma_F^2 & \sigma_F^2 & \sigma_F^2 & \sigma_F^2+\sigma_S^2+\sigma_{12} & \sigma_F^2+\sigma_S^2+\sigma_2^2 \end{pmatrix}$$

Note: are there identifiability issues with this structure? We'll discuss in class.