BIOS6643 **HW3 Solutions**

Note: For simplicity, you can write your answers for (1) and (2) for r_i=r=4. I.e., each answer should be a 4x4 matrix.

(1) Consider the mixed model $Y_{ij} = \mu + \tau_j + b_i + \varepsilon_{ij}$, where $b_i \sim iid\ N(0,\sigma_b^2)$ and $\varepsilon_{ij} \sim iid\ N(0,\sigma_\varepsilon^2)$, where b_i and ε_{ij} are independent; μ and τ are fixed effects; i denotes subject, i=1,…,n and j =1,…,r_i denotes time. In class we determined that $Var(\mathbf{Y}_i)$ had a compound symmetric structure. [This is actually the covariance matrix for $\mathbf{Y}_i$, or $Cov(\mathbf{Y}_i)$, but it is also often referred to as $Var(\mathbf{Y}_i)$.] Note that $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{ir_i})^t$.

a. Determine the correlation matrix, $Corr(\mathbf{Y}_i)$.

Solution: Note that $Corr(Y_i, Y_j) = Cov(Y_i, Y_j) / \sqrt{Var(Y_i)Var(Y_j)}$. But for the CS structure, the variances are equal, so this reduces to $Corr(Y_i, Y_j) = Cov(Y_i, Y_j) / Var(Y_i)$. Now $Var(Y_i) = \sigma_\varepsilon^2 + \sigma_b^2$. Also, we derived $Cov(Y_i, Y_j) = \sigma_b^2 \quad for\ i \neq j$ and $Cov(Y_i, Y_j) = \sigma_\varepsilon^2 + \sigma_b^2 \quad for\ i = j$ in class. Thus the correlation structure is

$$Corr(\mathbf{Y}_i) = \begin{pmatrix} 1 & \sigma_b^2/(\sigma_\varepsilon^2+\sigma_b^2) & \sigma_b^2/(\sigma_\varepsilon^2+\sigma_b^2) & \cdots & \sigma_b^2/(\sigma_\varepsilon^2+\sigma_b^2) \\ \sigma_b^2/(\sigma_\varepsilon^2+\sigma_b^2) & 1 & \sigma_b^2/(\sigma_\varepsilon^2+\sigma_b^2) & \cdots & \sigma_b^2/(\sigma_\varepsilon^2+\sigma_b^2) \\ \sigma_b^2/(\sigma_\varepsilon^2+\sigma_b^2) & \sigma_b^2/(\sigma_\varepsilon^2+\sigma_b^2) & 1 & \cdots & \sigma_b^2/(\sigma_\varepsilon^2+\sigma_b^2) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sigma_b^2/(\sigma_\varepsilon^2+\sigma_b^2) & \sigma_b^2/(\sigma_\varepsilon^2+\sigma_b^2) & \sigma_b^2/(\sigma_\varepsilon^2+\sigma_b^2) & \cdots & 1 \end{pmatrix}$$

We can simplify the notation by writing: $Corr(R_i) = [1/(\sigma_b^2 + \sigma_\varepsilon^2)](\sigma_b^2\mathbf{J} + \sigma_\varepsilon^2\mathbf{I})$. Note that the off-diagonal elements are the ICC!

b. Write the form of $Var(\boldsymbol{\varepsilon}_i) = \mathbf{R}_i$, where $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{ir_i})^t \sim N(\mathbf{0}, \mathbf{R}_i)$.

Solution: We can deduce that $\mathbf{R}_i = \sigma_\varepsilon^2\mathbf{I}$.

(2) Consider the model $Y_{ij} = \mu + \tau_j + \varepsilon_{ij}$, where $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \varepsilon_{i3}, \varepsilon_{i4})^t \sim N(\mathbf{0}, \mathbf{R}_i)$ (no random intercept).

(a) Write the form of $\mathbf{R}_i$ that yields the compound symmetric structure for $Var(\mathbf{Y}_i)$.

$$\mathbf{R}_i = \begin{pmatrix} \sigma_\varepsilon^2+\sigma_b^2 & \sigma_b^2 & \sigma_b^2 & \sigma_b^2 \\ \sigma_b^2 & \sigma_\varepsilon^2+\sigma_b^2 & \sigma_b^2 & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 & \sigma_\varepsilon^2+\sigma_b^2 & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 & \sigma_b^2 & \sigma_\varepsilon^2+\sigma_b^2 \end{pmatrix}$$

(b) Write the form of $\mathbf{R}_i$ that yields the AR(1) structure for $Var(\mathbf{Y}_i)$. (Hint: refer to HW1.)

$$\mathbf{R}_i = \sigma_\varepsilon^2 \begin{pmatrix} 1 & \phi & \phi^2 & \phi^3 \\ \phi & 1 & \phi & \phi^2 \\ \phi^2 & \phi & 1 & \phi \\ \phi^3 & \phi^2 & \phi & 1 \end{pmatrix}$$

(3) Re: slides 17 and 18 in the LMM II slides: on slide 18, tests for linear and quadratic trend are included in **C**. Show that the test H_0: $\mathbf{C\tau=0}$ versus H_A: H_0^C (considering rows simultaneously in the same test) is just the main effect test for time. I.e., it is the same test as when using the form of **C** as given on slide 17. Does it make sense that this would be right? Explain.

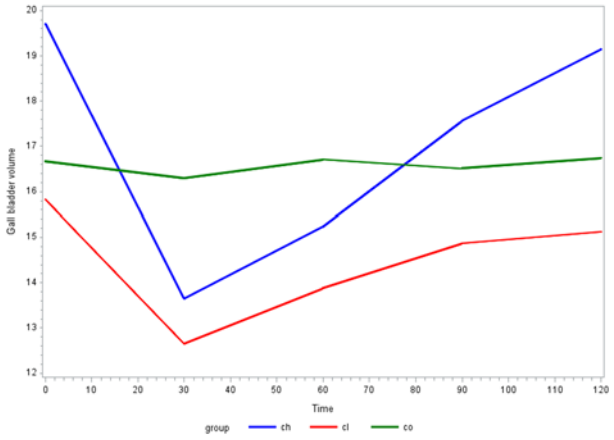
Solution: In order to get H_0: $\tau_1=\tau_2=\tau_3$, we can write $\mathbf{C} = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}$. However, note that for the orthogonal polynomial contrast matrix, $\mathbf{C} = \begin{pmatrix} -1 & 0 & 1 \\ 1 & -2 & 1 \end{pmatrix}$, the same constraints hold. To show this, note that for the first row, we have $\tau_1=\tau_3$. In the second row, we have $\tau_1+\tau_3=2\tau_2$, but since $\tau_1=\tau_3$, we can re-express the 2nd row as $\tau_3+\tau_3=2\tau_2$, or $\tau_2=\tau_3$; thus, $\tau_1=\tau_2$. What this means is that even though individual rows are polynomial tests, the test of H_0: $\mathbf{C\tau=0}$ (using the polynomial contrasts and considering the rows in the same test) is also the test of time effects. This does make sense since the d.f. for time is 2, and since we have 2 orthogonal contrasts, their sums of squares (thinking in terms of RM ANOVA) should add up to the 'time' sum of squares. Another way to think about it: if there are only 3 time points, then there can only be linear and/or quadratic effects. If there are neither, then there must be no time effects – i.e., the means must be equal across times. (Note that although a flat line is a straight line, it is not a 'linear effect'; a linear effect must have a nonzero slope.)

(4) With the dog data (see slide 4 of LMM II slides for description; data on website), do the following.

a. List 2 reasons why adding a random intercept for dogs (relative to the same model but without the random term) might help the model.

Solution: first, it captures variability between dogs into account based on factors such as dog size (e.g., Great Dane versus lap dog) and others not expressed in the fixed effects. Second, it helps us take correlation between responses within dogs into account.

b. Graph the mean response by time for each group.



c. Refer to class notes and slides for polynomial test output. Which polynomial trends for time and group*time most apparent in the graph? Which significant ones are not as apparent in the graph? (There is no correct or incorrect here, it's just based on effort.)

Solution: There appears to be quadratic interaction. Also, there may be an overall quadratic trend (averaged over groups) since 2 of the 3 groups have it. There may be a slight cubic trend with the flattening of the curves for the 2 drug groups; this may show up both as an overall cubic effect, and cubic interaction, since the control group is pretty flat. I do not notice linear or strong quartic effects here. After looking at the results, there is a marginally significant linear and quartic trend (averaged over groups).

d. Consider a CONTRAST to test for differences over time between the CH and CL groups. Using the means model, the test can be written as H_0: $\mu_{1j} - \mu_{1j'} = \mu_{2j} - \mu_{2j'}$ for all j, j', where subscript '1' denotes CH group, '2' denotes CL. Show that the same test can also be written as H_0: $(\gamma\tau)_{1j} - (\gamma\tau)_{1j'} = (\gamma\tau)_{2j} - (\gamma\tau)_{2j'}$ for all j, j' using the effects model. (See the LMM II slides 3 and 16 for how I've written the models.)
Solution: Based on the model I have in the notes, we can express $\mu_{hj} = \mu + \gamma_h + \tau_j + (\gamma\tau)_{hj}$. Using this form, you can use this to show directly that the above holds. (You can also call the interaction effects γ and the group effects something else, just be consistent.)

e. Write an ESTIMATE or CONTRAST statement for each question below and carry out the analysis. Summarize your results. (Note that these involve all 5 time points.)

Note I used the effects model for d and e..

i. Does the difference between the 2 drug groups change over time?

Solution:

```
contrast 'ch vs. cl int' group*time 1 -1 0 0 0 -1 1 0 0 0 0 0 0 0 0,
                         group*time 1 0 -1 0 0 -1 0 1 0 0 0 0 0 0 0,
                         group*time 1 0 0 -1 0 -1 0 0 1 0 0 0 0 0 0,
                         group*time 1 0 0 0 -1 -1 0 0 0 1 0 0 0 0 0;
```

This yields p<0.0001 (F=8.52).

ii. Is there a mean difference between drug groups for at least one time point?

Solution:

```
contrast 'ch vs. cl equal'
   group 1 -1 0 group*time 1 0 0 0 0 -1 0 0 0 0 0 0 0 0 0,
   group 1 -1 0 group*time 0 1 0 0 0 0 -1 0 0 0 0 0 0 0 0,
   group 1 -1 0 group*time 0 0 1 0 0 0 0 -1 0 0 0 0 0 0 0,
   group 1 -1 0 group*time 0 0 0 1 0 0 0 0 -1 0 0 0 0 0 0,
   group 1 -1 0 group*time 0 0 0 0 1 0 0 0 0 -1 0 0 0 0 0;
```

Results suggest that they differ: F=6.90, p<0.0001.

iii. Does control differ from the average of the CH and CL drug groups over time? (For the drug group means, you can use the straight average of drug groups means for each time.)

Solution: As mentioned in an e-mail, I was mainly interested in determining whether the difference between the average of CH and CL compared with CO depended on time:

```
contrast 'ch/cl ave vs. co int'
      group*time 0.5 -0.5 0 0 0 0.5 -0.5 0 0 0 -1 1 0 0 0,
      group*time 0.5 0 -0.5 0 0 0.5 0 -0.5 0 0 -1 0 1 0 0,
      group*time 0.5 0 0 -0.5 0 0.5 0 0 -0.5 0 -1 0 0 1 0,
      group*time 0.5 0 0 0 -0.5 0.5 0 0 0 -0.5 -1 0 0 0 1;
```

Significant: $p<0.0001$ with $F=18.23$. I will be a little flexible depending on how you interpreted the question.

f. Get estimates and 95% confidence intervals for each of the following.

i. The mean change in scores (from BL to 60 minutes after) for the CH group.

Solution:

```
estimate '60m-BL CH' time -1 0 1 0 0
   group*time -1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 / cl;
```

This yields a GBV difference estimate of -4.46 (60 minute time minus BL; 95% CI: -5.42 to -3.50).

ii. The difference in mean change of scores (BL to 60 min) for the CH group relative to the Control group.

Solution:

```
estimate '60-BL, CH vs. CO'
      group*time -1 0 1 0 0 0 0 0 0 1 0 -1 0 0 / cl;
```

The difference in 60m to BL between CH and CO groups is -4.50 (95% CI: -5.86 to -3.15). Note that the difference is similar to part i since control group means were very consistent over time.

g. Based on your own analysis and what is presented in the notes and slides, write a one paragraph (min. 5 sentences) summary of results and discussion for the dog and gallbladder volume data. Include what you believe is important and include statistical results in your write up.

Solution: Multiple approaches are 'correct' here, but in your paragraph, it is good practice to include inferential quantities (e.g., confidence limits and p-values) along with your descriptions. I think the polynomial trends (mainly quadratic) are interesting. The clearest pattern is that the drugs significantly reduced GBV, but then they increased again by 2 hours. The cubic trends for the drug groups, although initially hard to see by eye, indicate that the increase in gallbladder volume tapers off late. Differences between each drug group versus control were significant, and as we'd expect, control was pretty flat. (Note that this is not a linear trend, just flat.) Tests also indicated that the 2 drug groups differed from each other, both in terms of interaction and just comparing means over time ($p<0.0001$ for both). Both drug groups had quadratic patterns, but the quadratic trend was stronger trend for the CH group than the CL group ($p<0.0001$), suggesting that this drug is stronger than the CL drug. [To get this last test, use just the first row in the 'quadratic-by-quadratic contrast matrix' in the notes – the test has 1 degree of freedom.]

h. Say that your client requests that time be modeled as a continuous variable rather than class. In particular, they want to be able to estimate GBV values that might be in between the 30 minute intervals (e.g., 45 minutes) in addition to at the observed time points. Any cautions or things to consider when setting up the model? How would you proceed? Answer in 2 to 3 sentences.

Solution: Note that if you model time as class, you have the most flexibility possible if modeling changes over time. if you simply take 'time' out of the CLASS statement, then you are modeling time in a simple linear fashion. So these are the 2 extremes. A straight line might not fit the data (within group), so you would carefully need to examine what degree of polynomials will be sufficient. As far as interpolation, that should be fine as long as you have found a decent function for time. For this, we assume that the trend in between observed times is the same as what we observed. Again, this is probably a reasonable assumption. Extrapolation is where you might run into problems.