**BIOS6643 HW4 Due Friday, October 14th**

### Trent Hawkins

### Question 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 Vi if a random intercept for subjects will be included, plus an AR(1) structure for the error covariance matrix (Ri). 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.***

The structure of will depend on on R and G in this manner:

Where:

Finally:

### Question 2:

***First, use group and time as class variables, plus group\*time, and determine the best Kronecker Product structure to use for the error covariance structure. (Note: there are 3 options in SAS, and you may be limited by what will work.) Highlight results.***

Each of the models fit to answer this question take the form of:

Where is a complex, Kronecker-Product, covariance structure. The results for each of these Kronecker-Product structures are presented below. It should be noted that the LOCAL statement was added to the UN@AR(1) structure in SAS in order to add residual variance down the diagonal of the matrix to help with convergence. Models were fit using Maximum Likelihood (ML) estimation to maintain comparability across AIC scores.

| **Metric** | **UN (Time) @ CS(Trt)** | **UN @ UN** | **UN (Trt) @ AR(1) (Time)** |
| --- | --- | --- | --- |
| **-2 Log Likelihood** | 255.3 | 254.1 | 266.9 |
| **AIC (Smaller is Better)** | 317.3 | 324.1 | 312.9 |
| **AICC (Smaller is Better)** | 351.5 | 370.8 | 329.6 |
| **BIC (Smaller is Better)** | 310.8 | 316.9 | 308.1 |

The results indicate that the UN@AR(1) structure, with UN structure for Trt and AR(1) for time, achieves the lowest AIC-score. This model will be re-fit for the final results using Restricted Maximum Likelihood estimation.

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

| **Estimated R Matrix for id 1** | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Row** | **Col1** | **Col2** | **Col3** | **Col4** | **Col5** | **Col6** | **Col7** | **Col8** | **Col9** | **Col10** | **Col11** | **Col12** | **Col13** | **Col14** | **Col15** |
| **1** | 106.88 | 106.17 | 105.92 | 105.68 | 105.43 | 33.4631 | 33.3844 | 33.3059 | 33.2276 | 33.1495 | 56.7590 | 56.6255 | 56.4924 | 56.3596 | 56.2271 |
| **2** | 106.17 | 106.88 | 106.17 | 105.92 | 105.68 | 33.3844 | 33.4631 | 33.3844 | 33.3059 | 33.2276 | 56.6255 | 56.7590 | 56.6255 | 56.4924 | 56.3596 |
| **3** | 105.92 | 106.17 | 106.88 | 106.17 | 105.92 | 33.3059 | 33.3844 | 33.4631 | 33.3844 | 33.3059 | 56.4924 | 56.6255 | 56.7590 | 56.6255 | 56.4924 |
| **4** | 105.68 | 105.92 | 106.17 | 106.88 | 106.17 | 33.2276 | 33.3059 | 33.3844 | 33.4631 | 33.3844 | 56.3596 | 56.4924 | 56.6255 | 56.7590 | 56.6255 |
| **5** | 105.43 | 105.68 | 105.92 | 106.17 | 106.88 | 33.1495 | 33.2276 | 33.3059 | 33.3844 | 33.4631 | 56.2271 | 56.3596 | 56.4924 | 56.6255 | 56.7590 |
| **6** | 33.4631 | 33.3844 | 33.3059 | 33.2276 | 33.1495 | 12.0114 | 11.5247 | 11.4976 | 11.4706 | 11.4436 | 19.5082 | 19.4623 | 19.4166 | 19.3710 | 19.3254 |
| **7** | 33.3844 | 33.4631 | 33.3844 | 33.3059 | 33.2276 | 11.5247 | 12.0114 | 11.5247 | 11.4976 | 11.4706 | 19.4623 | 19.5082 | 19.4623 | 19.4166 | 19.3710 |
| **8** | 33.3059 | 33.3844 | 33.4631 | 33.3844 | 33.3059 | 11.4976 | 11.5247 | 12.0114 | 11.5247 | 11.4976 | 19.4166 | 19.4623 | 19.5082 | 19.4623 | 19.4166 |
| **9** | 33.2276 | 33.3059 | 33.3844 | 33.4631 | 33.3844 | 11.4706 | 11.4976 | 11.5247 | 12.0114 | 11.5247 | 19.3710 | 19.4166 | 19.4623 | 19.5082 | 19.4623 |
| **10** | 33.1495 | 33.2276 | 33.3059 | 33.3844 | 33.4631 | 11.4436 | 11.4706 | 11.4976 | 11.5247 | 12.0114 | 19.3254 | 19.3710 | 19.4166 | 19.4623 | 19.5082 |
| **11** | 56.7590 | 56.6255 | 56.4924 | 56.3596 | 56.2271 | 19.5082 | 19.4623 | 19.4166 | 19.3710 | 19.3254 | 35.8928 | 35.3500 | 35.2669 | 35.1840 | 35.1013 |
| **12** | 56.6255 | 56.7590 | 56.6255 | 56.4924 | 56.3596 | 19.4623 | 19.5082 | 19.4623 | 19.4166 | 19.3710 | 35.3500 | 35.8928 | 35.3500 | 35.2669 | 35.1840 |
| **13** | 56.4924 | 56.6255 | 56.7590 | 56.6255 | 56.4924 | 19.4166 | 19.4623 | 19.5082 | 19.4623 | 19.4166 | 35.2669 | 35.3500 | 35.8928 | 35.3500 | 35.2669 |
| **14** | 56.3596 | 56.4924 | 56.6255 | 56.7590 | 56.6255 | 19.3710 | 19.4166 | 19.4623 | 19.5082 | 19.4623 | 35.1840 | 35.2669 | 35.3500 | 35.8928 | 35.3500 |
| **15** | 56.2271 | 56.3596 | 56.4924 | 56.6255 | 56.7590 | 19.3254 | 19.3710 | 19.4166 | 19.4623 | 19.5082 | 35.1013 | 35.1840 | 35.2669 | 35.3500 | 35.8928 |

This 15X15 Covariance matrix above is the R matrix for Subject 1. The values highlighted in yellow is the variance for subject 1 in group ‘ch’ with the off-diagonal being the covariance at each time-point in each group. Since an UN covariance has been specified for treatment, there is a different covariance estimated for each pair of time-points. This pattern continues down the diagonal with red values being the variance in the ‘cl’ group and green being the variance in the ‘co’ group. The correlation matrix is similar, with decaying correlation on the off-diagonals for the AR(1) structure. Correlation decreases as time-points move further away, and obviously, as groups change.

***For the model in part a, write a contrast to test the null hypothesis that the means for the two treatment groups are equivalent for all time points and summarize the results.***

| **Contrast Results** | | | | |
| --- | --- | --- | --- | --- |
| **Label** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| Treatment Avg Differ at each time | 5 | 12.9 | 8.74 | 0.0008 |

The above contrast statement tests:

H0 : for all *i = j*

H­­A: At least one for all *i = j*

Using these results, we may reject H0, and conclude that at least one of the averages are not equal between time-points (p = 0.0008).

***Is there a time-as-continuous model that can ‘beat’ the time-as-class model in terms of AIC? (Remember that the highest degree of polynomial is one less than the number of time points.) Summarize your results.***

| **Metric** | **Linear**  **AR(1)** | **Quadratic**  **AR(1)** | **Cubic**  **AR(1)** | **Quartic**  **AR(1)** | **Time as Class**  **UN @ AR(1)** |
| --- | --- | --- | --- | --- | --- |
| **-2 Log Likelihood** | 453.5 | 426.4 | 398.5 | 394.7 | 266.9 |
| **AIC (Smaller is Better)** | 469.5 | 448.4 | 426.5 | 428.7 | 312.9 |
| **AICC (Smaller is Better)** | 471.3 | 451.8 | 432.1 | 437.2 | 329.6 |
| **BIC (Smaller is Better)** | 467.8 | 446.1 | 423.6 | 425.2 | 308.1 |

In this case, time is collected as a factor (0, 30, 60, 90, 120 days). Even though time points are evenly spaced, we would expect that the time-as-class model would provide the best fit to the data. Additionally, the more-complex Kronecker Product structure cannot be used for Time as a continuous variable. Therefore, the models that do not incorporate this structure cannot bring into account the within-group and between-group correlations and the Auto-regressive structure for time.

1. In a paragraph, describe your plan for your project. Include the data that you plan to use, and a couple of research questions of interest. (You can change these later if you need to.) If you plan to do an alternative project, describe that (e.g., special research topic; advanced data analysis with R).

I am hoping to complete an project that would give me a gentle introduction to geo-spatial time-series data analysis. I do not have a specific dataset yet, but I have several to choose from. I am waiting to meet with Dr. Strand until next Monday. Some questions of interest:

How have Colorado’s climate patterns determined snowpack over the past decades?

How do climate patterns effect overall storm severity in Colorado?

There are *Very* rough questions. I am hoping to work with Dr. Strand to identify more specific questions with easily obtainable data.