73/85 B+ Chong Kim

BIOS6643 **Midterm** **2016**

(1) The fitted $\mathbf{R}_i$ matrix for one subject in the Beta Carotene data has the following form when using the UN structure ($\mathbf{R}_i$ matrix on left, $\mathbf{Rcorr}_i$ on right). This is the same as $\mathbf{V}_i$ as there are no random effects in this particular model. Recall that the 5 time points were at 0, 6, 8, 10 and 12 weeks. (We just used the second BL value in this analysis.)

a. Based on observing the numbers in the matrix, argue why you think the AR(1) either would work or would not work for these data; 2 to 3 sentences is sufficient.

Estimated R Matrix for Id(Prepar) 71 1

Col1	Col2	Col3	Col4	Col5
2825	4114	5334	3939	4304
4114	14162	13215	11600	12287
5334	13215	14514	11697	12450
3939	11600	11697	11396	11313
4304	12287	12450	11313	12552

Estimated R Correlation Matrix for Id(Prepar) 71 1

Row	Col1	Col2	Col3	Col4	Col5
1	1.0000	0.6505	0.8330	0.6943	0.7229
2	0.6505	1.0000	0.9218	0.9131	0.9216
3	0.8330	0.9218	1.0000	0.9095	0.9224
4	0.6943	0.9131	0.9095	1.0000	0.9459
5	0.7229	0.9216	0.9224	0.9459	1.0000

The correlation b/w time points do not decay over time (assuming $|\phi|<1$). Thus AR(1) would NOT be a good structure for R. (i.e. T_1T_1 T_1T_2 T_1T_3 1 → 0.65 → 0.83 dec inc?)

10/10

b. Say that we model the data in a different way; we include the baseline value as a covariate rather than as an outcome. How would you expect this to affect the selection of a correlation structure ($\mathbf{V}_i$) for the responses? Again, 2 to 3 sentences is sufficient.

If BAC baseline model is used, then the error matrix would reflect the systematic change in the difference from baseline to outcome. The selection of V matrix will depend on of these changes (based on baseline) have a appropriate structure (I don't think AR(1) worked for BAC model). Also less parameter for UN thus UN could work better.

5/10 What about the new Vi structure that is 4x4 instead of 5x5? What patterns do you expect?

$\beta_1 x_i$

$\mu+\gamma_i+\tau_j+(\gamma\tau)_{ij}$

c. Write an ESTIMATE statement (to be added to PROC MIXED code) to compare the linear trend for Preparation 1 versus linear trend for Preparation 3 (group, time as class variables and interaction between them included); indicate which type of model you are using (e.g., means, 2-way effects). Note that there are 4 total preparations and you now just have 4 times.

1x1

Estimate "1x1 for prep 1 vs 3"
group*time -2 -1 1 2 0 0 0 0 2 1 -1 -2 0 0 0 0;

9/10 Basic idea good. Just change "2" to "3".

using 2 way effect;

1 2 3 4 5 1 2 3 4 5 id(school)

4 schools

(2) Say we find 4 different schools to participate in 'health day', and within each school we find 5 children to take blood pressure measurements on.

5/5 a. Are schools and children nested or crossed factors? nested → children(school)

lvl 1 = $Y_{i0} = b_{i0} + b_{i1} + \varepsilon_{ij}$

$b_{i0} = \beta_0 + \beta$

b. Suppose that we build a model for blood pressure that includes age and gender as covariates. Write a complete observation-level statistical model for these data if we take just 1 measurement per subject, and either add terms or specify structures to address potential correlation between responses in your model.

highest "school" level → $Y_{hij} = b_{h0} + b_{h1}X_{1ij} + b_{h2}X_{2i} + \varepsilon_h$ $\varepsilon_h \sim N(0, R_h)$

lower level → $b_{h0} = \beta_0$; $b_{h1} = \beta_{1ij}$; $b_{h2} = \beta_{2i}$ $\underset{(5\times5)}{R_h} = R_{i(h)} \underset{5\times5}{I}$ for VC, $i = 1, 5$

7/10 A little unclear. Do you have random effects here?

$h = 1, \ldots, 4$ school, j = rep meas, β_1 age, where $R_{i(h)} = \sigma^2_\varepsilon$, $i = 1, 5$

$i = 1, \ldots, 5$ subject; β_2 gender

X_1 = age; X_2 gender (1 or 2); β_0 intercept

$$\begin{Bmatrix} R_{1(h)} & & & & \\ 0 & R_{2(h)} & & & \\ 0 & 0 & R_{3(h)} & & \\ 0 & 0 & 0 & R_{4(h)} & \\ 0 & 0 & 0 & 0 & R_{5(h)} \end{Bmatrix}$$ 5×5 sym

c. Repeat the previous part if we take 2 measurements per subject.

same as above but $j = 1, 2$ (repeated measures)

$\varepsilon_h \sim N(0, \underset{(10\times10)}{R_h})$

7/10 O.k. how do you account for correlation between responses within a asubject?

R_h is VC. and $R_{i(h)}$ is VC

$\underset{(10\times10)}{R_h} = R_{i(h)} \underset{5\times5}{I}$ where

$$\underset{(2\times2)}{R_{i(h)}} = \begin{Bmatrix} \sigma^2_\varepsilon & 0 \\ 0 & \sigma^2_\varepsilon \end{Bmatrix}$$ if VC

d. Determine the covariance between time measurements within one subject for this model. (You don't need to compute the entire covariance matrix unless you want to.)

$R_{i(h)} = VC$ $\sigma^2_\varepsilon \underset{2\times2}{I}$

$$R_{i(h)} = \begin{Bmatrix} \sigma^2_\varepsilon & 0 \\ 0 & \sigma^2_\varepsilon \end{Bmatrix}$$ covariance 0 if VC.

if $R_{i(h)}$ = UN then

$$R_{i(h)} = \begin{Bmatrix} \sigma^2_1 & \sigma_{12} \\ \sigma_{12} & \sigma^2_2 \end{Bmatrix}$$ covariance σ_{12} if UN.

10/10 O.k.

Short answer questions: Select 2 of the following questions to answer; 2 to 3 sentences and a simple illustration (if applicable) is sufficient. FIRST, CIRCLE THE TWO QUESTIONS THAT YOU WILL ANSWER, AND THEN WRITE YOUR ANSWERS IN THE EMPTY SPACE BELOW. IF YOU NEED MORE SPACE, USE AN ADDITIONAL BLANK PAGE RATHER THAN CONTINUING ON THE BACK.

(3) We fit a model for longitudinal data and use the UN structure for the R matrix. We then add a random intercept to this model but one of 3 things happens: (i) the final Hessian matrix was not positive definite, (ii) model did not converge, (iii) variance of the random intercept is estimated to be 0. (It might be any of the 3, it just depends on the data.) But then if we then change the structure for R to simple ($\sigma^2\mathbf{I}$) and keep the random intercept, the convergence criteria are met and the variance of the random intercept is sizeable. Explain what is happening. If you use an example, a simple matrix will do (e.g., 2 repeated measures per subject).

(4) A friend of yours fits data using a linear model. You then remind them that they actually have longitudinal data, and that they should use some kind of model that takes correlation between repeated measures into account. They then say, "Yeah, I tried that but the estimates did not change much, so I'm going to stay with the simpler model." At first you think, oh, o.k. But then you remember, "Ah, but…" Finish this statement (i.e., explain why you the longitudinal model is expected to be better despite the fact that the estimates themselves may not be much different between the approaches).

(5) In class we observed when we fit the Mt. Kilimanjaro data using PROC MIXED that the variance for the random intercept was 0 (and hence essentially that parameter was dropped from the model). However, we still obtained nonzero estimates using the EB method. These seem to conflict to each other. Describe why this may happen when fitting a mixed model and why it doesn't necessarily indicate a 'bad' fit.

3). For a simple 2 r.m. model if either of the 3 things above happening the parameter estimates (i.e. cov param α) are unidentifiable due to redundancy in the estimation of them thus trying to fit too many cov parameter when the model can only exhibit a few will lead to the above notes

10/10 O.k., but it's the other way around. Since some parameters are unidentifiable, one of these problems arises...

4). First, longitudinal is better b/c it allows for unobserved (random effects) factors, that the fixed effects don't capture, to be incorporated along with complex correlation structure that may better reflect the real world. Although the estimates may not change, the variance component will be different due to incorporating the complex error structures thus making inferences will be different as the end result.

10/10