**BIOS6643 HW4 2016**

1. Consider a study in which subjects’ blood pressures are observed over time (4 time points, equally spaced, no missing data). The model will have fixed effects for time, age (at start of experiment) and gender; a random intercept for subjects will also be included.
   1. Write out the mixed model, specifying all parts, if time is modeled as a continuous variable (linear term only) and the AR(1) structure is used to model the errors.

, where are fixed effects associated with time, age and gender, respectively and b is associated with the random effect (intercept) for subjects; i=1,…,n; j=1,…,4 (continuous).

* 1. How many fixed effect (β) parameters are in the model (considering full-rank parameterization)? How many covariance (α) parameters?

Fixed effects:

1 for time, 1 for age, 1 for gender, and 1 for the intercept thus 3 fixed effect parameters considering full-rank parameterization.

Covariance parameters:

There will be 1 random intercept (for subject) and since the model is AR(1) structure, there will be 1 for AR(1) and 1 for thus in total there will be 3 covariance parameters.

* 1. Write the form of Var(**Y***i*) in terms of the covariance parameters.
  2. Repeat parts a-c if time is modeled as a class variable, the UN structure is used for the covariance matrix of the errors and there is no random intercept term.

where are fixed effects associated with time, age, and gender, respectively; i=1,…,n; j=1,…,4; k=range of age at start of experiment to end of experiment; l=1 or 2.

Fixed effects:

3 parameter for time (4-1=3 considering full rank), 1 for age, and 1 for gender, thus 5 fixed effect parameters considering full-rank parameterization.

Covariance parameters:

There will be 0 random intercept (for subject) and since the model is UN structure, there will be 10 for UN.

* 1. **Not to turn in.** In the model, age at start of experiment (i.e., baseline) was used. How would estimates change if you used continuous age in the model instead? In order to answer the question, write out the statistical models for both approaches.

If age was considered to be continuous then the model would incorporate the parameter for time and age into one fixed parameter.

where are fixed effects associated with time, age, and gender, respectively; i=1,…,n; j=1,…,4; k=range of age at start of experiment to end of experiment; l=1 or 2.

Where now the age of individuals change with time but since they are a fixed effect, the variance due to the change in age will be the same as the variance due to time thus the sums of squares of error will not change but the parameter will change to reflect the change in age due to time.

1. Consider the linear mixed model (in subject form): **Y***i*=**X***i***β**+**Z***i***b***i*+**ε***i*, where **ε**i~*N*(**0**,**R***i*) independent of **b***i*~N(**0**,**G***i*), and subjects *i*=1,…,*n* are independent. [Here, **G***i* and **R***i* are the same for all subjects; the index is just used to indicate dimensions.] A longitudinal experiment is conducted where subjects are observed at 3 equally spaced time points – for simplicity you can let the time points be 0,1, and 2. The data will be modeled using **b***i*=(*bi*0,*bi*1)*t* defined for subjects (*bi*0 = random intercept; *bi*1 = random slope for time) and **R**=*σε*2**I**. There will be fixed-effect terms for time and possibly other variables, but their specification isn’t relevant to the questions below. Justify all responses / show work.
2. Determine **V**i=*Var*(**Y**i) (the model covariance matrix for subject *i*) if the unstructured covariance structure is used for **G**.
3. Write the SAS, PROC MIXED code for the analysis of the data, if we allow for a simple linear time trend for fixed effects. (You can use generic names, e.g., *dat* for the data set, *time* for time.)

PROC MIXED DATA=dat;

MODEL y = time ;

RANDOM intercept time/ type=un s;

RUN;

1. Often a realistic covariance structure for **V**i for longitudinal data is one where covariance between responses is positive but decays as time between responses increases. With the covariance structure you came up with in part a, is it possible for this structure to have covariance that decays as time between responses increases? Justify your response.

Yes it is possible to have covariance structure to decay over time. For example, the AR(1) covariance structure will have a decreasing covariance as time increases between responses increases. For the covariance structure depicted in part a, the covariance between random intercept and slope would need to be a negative value thus as time increases between responses the covariance between responses will decrease.

1. Consider the peak flow data fit with a simple random effect model, with SAS output presented on pages 12-14 of the *LMM inference* notes (pages 122-124 of the complete PDF).
   1. Interpret the tests for random effects, e.g., interpret the effect for subject 101S.

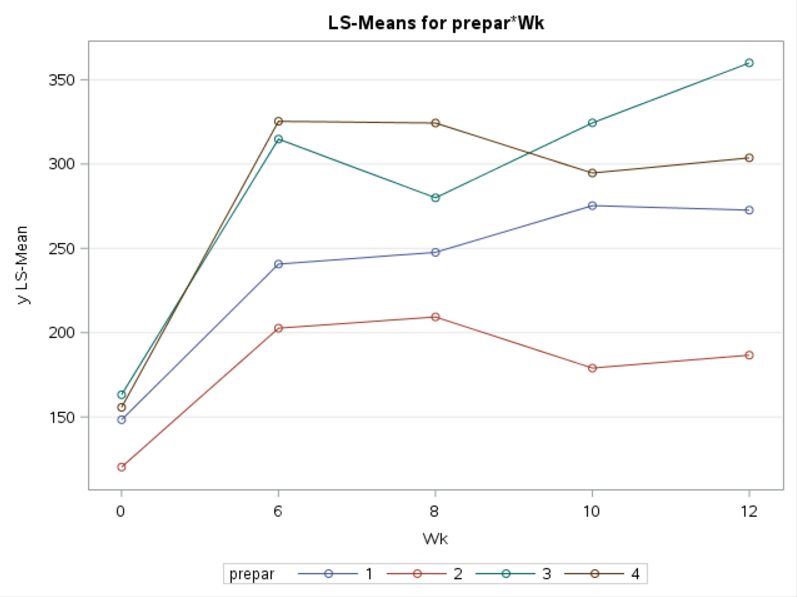
Based on the partial output in pg 122-124 of appended LMM: inference notes, the random intercept for subject 101S is significantly different compared to the overall expected random intercept (0) with p-value 0.0412. Subject 102S has a larger estimated random intercept (108.63) away from the null, thus a higher p-value (<0.0001), while subject 110S has a smaller difference away from null (-28.8484) thus a non-significant p-value (0.2562). Subjects 103S, 104S, and 105S all have random intercepts that are significantly different from the null (.

* 1. Based on the estimated covariance parameters, are you surprised that the majority of kids have p-values below 0.05? Explain.

No, I’m not surprised because the is so much larger than the which is indicative that the between subject variability is much larger than the within subject variability and thus the resulting p-value for the random intercept should be relatively significant for majority of the kids.

* 1. **Not to turn in.** For subject 101S, determine the following, replacing unknown parameters with their estimates. The data is posted on the course web site; you only need the January values for subject 101S. The quantities can be found in the slides and notes as mentioned above.
     1. The shrinkage factor,.
     2. The random effect estimate, .
     3. The predicted PEF, . Is more weight given to the subject’s data or the population average (estimate)? Does the result surprise you? Explain.

1. (From Rosner, 2006.) A clinical trial was planned comparing the incidence of cancer in a group taking beta-carotene in capsule form compared with a group taking beta-carotene placebo capsules. One issue in planning such a study is which preparation to use for the beta-carotene capsules. Four preparations were considered: (1) Solatene (30mg capsules), (2) Roche (60mg capsules), (3) BASF (30mg capsules), (4) BASF (60mg capsules). To test efficacy of the four agents in raising plasma-carotene levels, a small bioavailability study was conducted. After two consecutive-day fasting blood samples, 23 volunteers were randomized to one of the four preparations mentioned above, taking 1 pill every other day for 12 weeks. The primary endpoint was level of plasma carotene attained after moderately prolonged steady ingestion. For this purpose, blood samples were drawn at 6, 8, 10 and 12 weeks. In order to model the data, consider group and time as class variables. Other model specifications may depend on the question. **Just use the second baseline measure as the ‘time 0’ measure for parts a through f.**
   1. Create a graph for the data. You have some flexibility on what type of graph to include.



* 1. Consider the model with a random intercept for subjects (in addition to other terms mentioned above); let **R** have the independent structure (**R**=*σε*2**I**). Mention 2 approaches to fitting the model and conducting tests.
  2. For the model above, determine the ICC and interpret the quantity.
  3. Now consider the same model but include a random slope for continuous (linear) time for subjects in addition to the random intercept. (NOTE: since we’re modeling time as a class variable in the fixed-effect part of the model, you’ll need to define a new time variable so that you can treat it as a continuous variable for the random effect.) Try the UN structure for **G** and compare results using the VC structure. Which model is better? Interpret the covariance term between the intercept and slope.
  4. Now consider a model with no random effects, but specifying **R** to have the UN structure. Fit this model and compare with the previous ones.
  5. Write a paragraph about your findings. Include a couple of specific tests to support your story (with custom needs, if you’d like). You can also include a table with goodness-of-fit statistics for models fit in the previous parts. Out of all the models, which one would you use for your ‘final’ model? Why?
  6. **Not to turn in.** There are two baseline variables. There are different ways these variables could be used in an analysis. For example, you could simply use one or the other, or perhaps the average of the two. Suggest what you would use for a baseline measure. [Note: you receive the data not knowing for certain whether all the values are correct or not.]