**BIOS6643 GLM (Page 1) and LMM (Page 2) Review Questions Fall 2018**

**NOTE: To turn in: 3, 8, 9, 11, 12 (a and b only), 14. The others are for practice and discussion.**

**General comment: It would benefit you to also look at these solutions in depth, even for questions that you received full credit. In some cases there are alternative approaches that are ‘correct’. Also, in some rare cases I may have missed an error in your HW, especially for questions that give you flexibility such as #14.**

**On writing statistical models, make sure to include enough indicators and then make sure they correspond between left and right sides.**

1. The distribution of is given on page 62 of the GLM course notes, which was derived using the linear form result. Derive the distribution of  if  form some vector **a**. Note that the result can also be completed using the linear form result. The proof is short.



Let  (yes, the lower and upper case “a’s” are different)

Apply the linear form result to obtain 

This can be further reduced to  or 

1. For the Myostatin data, note that the population mean for the myostatin group at 48 hours is  for the one-way effects model (see notes). Write the population means for the following. NOTE: these are not numerical estimates, but parameters or combinations of parameters.
   1. Myostatin group at 48 hours; means model. **Answer =**
   2. Myostatin group at 48 hours; two-way effects model. **Answer = Note: we assume that we still have the interaction term in the model.**
   3. Myostatin group, difference between 48 and 72 hours, one-way effects model. **Answer = [or ]**
   4. Myostatin group, difference between 48 and 72 hours, two-way effects model. **To start, note that M group, 48 hours is , and the M group, 72 hours is . Thus, the difference is  [or ].**
2. Show that  satisfies the normal equations. (Here, tilde indicates that the beta estimate may not be unique.) **There are at least two approaches to show this. Here is one:**

** Start with normal equations.**

** Substitute in formula for beta tilde.**

** **

** Apply transpose property.**

** Projection matrix is symmetric.**

** X is in column space of X.**

** QED.**

*LMM models*

1. Review Section 3.6.3 in the GLM course notes. (Note: this should be in the LMM chapter since it is discussing models indexed by subject and time, i.e., applicable for repeated measures.)
   1. Write full-rank and less-than-full-rank models if there is a group variable with 4 levels (i.e., 4 groups), a time variable that is treated as a continuous variable (linear term only), plus group\*time interaction. How many columns are in **X** for each approach?

Full rank: , where *x*1*i*, *x*2*i* and *x*3*i* are indicator variables for Groups 1,2 and 3, respectively (group 4 is the reference). There are 8 beta coefficients including the intercept, so 8 columns in X.

LTFR: . Some simplification might occur, such as removing index ‘i’ from the x variable if time points are common to all subjects. There are 10 fixed-effect parameters, so 10 columns in X.

* 1. If time points are unequally spaced then would it be appropriate to treat time as a class variable? Explain.

1. Consider a study where subjects in 3 groups (e.g., race or treatment) are observed over 3 times and some health outcome, y, is measured. Unless otherwise mentioned, include a random intercept for subjects to account for the repeated measures. For simplicity, use 2 subjects per group.
   1. Consider modeling group and time as class variables, plus interaction. Write statistical models and the X matrix for the following cases.

**Some notes: you can use whatever type of model you wish, as long as you are consistent. I am using 2-way effects models. Some people used one-way effects models. Remember to include a random intercept for subjects, and use indices as appropriate on the statistical models.**

* + 1. No restriction placed on the model. I.e., write the less-than-full-rank statistical model.

, *i*=1,…,6 for subjects, *h*=1,2,3 for groups, *j*=1,2,3 for measures; , .



* + 1. A set-to-0 restriction is placed on the parameters associated with highest levels.

So let x1i=1 if subject i is in Group A, x2i=1 if subject i is in Group B, and 0 otherwise (Group C is reference); let x3=1 for Time 1 and 0 otherwise, x4=1 for Time 2 and 0 otherwise (Time 3 is reference).

, *i*=for subjects 1,…,6; *h*=1,2,3 for groups, *j*=1,2,3 for measures; , . The X matrix is below.

* + 1. A sum-to-0 restriction is placed on the parameters associated with highest levels.

**NOTE: the ‘associated with highest levels’ is not relevant here. The restrictions are , ,  for each fixed h, and for each fixed j. You could just write the model from i and state these restrictions. But you can rewrite the model with these constraints. First note that , , , **

**To illustrate how to incorporate this into the model, considering the group-effects part of the model that uses indices: , where x1, x2 and x3 are indicators for membership in Groups A, B and C, respectively. Using the sum-to-zero constraint, this can be rewritten as**

**. You can continue this for the other terms. It gets messy pretty quickly, so it’s actually more compact to just show the LTFR model and state the additional constraints.**

X matrix for ii (set to 0) X matrix for iii (sum-to-0)

 

1. Show that the linear trend for one group compared to another (say Group A versus B) is estimable by showing that **L**=**LH**, where the Moore-Penrose inverse is used in calculating **H**. First you need to construct **L**. (As a check, repeat using SAS’s g-inverse in calculating **H**, but you don’t need to turn that in.

**You can use SAS PROC IML or R to construct H. Note that ‘ginv’ is the function in both that uses the MP inverse. So, for example, you can use h=ginv(t(x)\*x)\*t(x)\*x; in SAS PROC IML. Just use the x from ai. Note that L=(0 0 0 0 0 0 0 -1 0 1 1 0 -1 0 0 0) and you will see that LH comes out to be the same. It is possible that there will be some really small numbers that should be 0, but this is just rounding error (in SAS). Below is H. Note that if you multiple this matrix by 16, you will get nice integers. So you can write H=(1/16)\*“Nice H”.**

| **H** | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **COL1** | **COL2** | **COL3** | **COL4** | **COL5** | **COL6** | **COL7** | **COL8** | **COL9** | **COL10** | **COL11** | **COL12** | **COL13** | **COL14** | **COL15** | **COL16** |
| **ROW1** | 0.5625 | 0.1875 | 0.1875 | 0.1875 | 0.1875 | 0.1875 | 0.1875 | 0.0625 | 0.0625 | 0.0625 | 0.0625 | 0.0625 | 0.0625 | 0.0625 | 0.0625 | 0.0625 |
| **ROW2** | 0.1875 | 0.5625 | -0.1875 | -0.1875 | 0.0625 | 0.0625 | 0.0625 | 0.1875 | 0.1875 | 0.1875 | -0.0625 | -0.0625 | -0.0625 | -0.0625 | -0.0625 | -0.0625 |
| **ROW3** | 0.1875 | -0.1875 | 0.5625 | -0.1875 | 0.0625 | 0.0625 | 0.0625 | -0.0625 | -0.0625 | -0.0625 | 0.1875 | 0.1875 | 0.1875 | -0.0625 | -0.0625 | -0.0625 |
| **ROW4** | 0.1875 | -0.1875 | -0.1875 | 0.5625 | 0.0625 | 0.0625 | 0.0625 | -0.0625 | -0.0625 | -0.0625 | -0.0625 | -0.0625 | -0.0625 | 0.1875 | 0.1875 | 0.1875 |
| **ROW5** | 0.1875 | 0.0625 | 0.0625 | 0.0625 | 0.5625 | -0.1875 | -0.1875 | 0.1875 | -0.0625 | -0.0625 | 0.1875 | -0.0625 | -0.0625 | 0.1875 | -0.0625 | -0.0625 |
| **ROW6** | 0.1875 | 0.0625 | 0.0625 | 0.0625 | -0.1875 | 0.5625 | -0.1875 | -0.0625 | 0.1875 | -0.0625 | -0.0625 | 0.1875 | -0.0625 | -0.0625 | 0.1875 | -0.0625 |
| **ROW7** | 0.1875 | 0.0625 | 0.0625 | 0.0625 | -0.1875 | -0.1875 | 0.5625 | -0.0625 | -0.0625 | 0.1875 | -0.0625 | -0.0625 | 0.1875 | -0.0625 | -0.0625 | 0.1875 |
| **ROW8** | 0.0625 | 0.1875 | -0.0625 | -0.0625 | 0.1875 | -0.0625 | -0.0625 | 0.5625 | -0.1875 | -0.1875 | -0.1875 | 0.0625 | 0.0625 | -0.1875 | 0.0625 | 0.0625 |
| **ROW9** | 0.0625 | 0.1875 | -0.0625 | -0.0625 | -0.0625 | 0.1875 | -0.0625 | -0.1875 | 0.5625 | -0.1875 | 0.0625 | -0.1875 | 0.0625 | 0.0625 | -0.1875 | 0.0625 |
| **ROW10** | 0.0625 | 0.1875 | -0.0625 | -0.0625 | -0.0625 | -0.0625 | 0.1875 | -0.1875 | -0.1875 | 0.5625 | 0.0625 | 0.0625 | -0.1875 | 0.0625 | 0.0625 | -0.1875 |
| **ROW11** | 0.0625 | -0.0625 | 0.1875 | -0.0625 | 0.1875 | -0.0625 | -0.0625 | -0.1875 | 0.0625 | 0.0625 | 0.5625 | -0.1875 | -0.1875 | -0.1875 | 0.0625 | 0.0625 |
| **ROW12** | 0.0625 | -0.0625 | 0.1875 | -0.0625 | -0.0625 | 0.1875 | -0.0625 | 0.0625 | -0.1875 | 0.0625 | -0.1875 | 0.5625 | -0.1875 | 0.0625 | -0.1875 | 0.0625 |
| **ROW13** | 0.0625 | -0.0625 | 0.1875 | -0.0625 | -0.0625 | -0.0625 | 0.1875 | 0.0625 | 0.0625 | -0.1875 | -0.1875 | -0.1875 | 0.5625 | 0.0625 | 0.0625 | -0.1875 |
| **ROW14** | 0.0625 | -0.0625 | -0.0625 | 0.1875 | 0.1875 | -0.0625 | -0.0625 | -0.1875 | 0.0625 | 0.0625 | -0.1875 | 0.0625 | 0.0625 | 0.5625 | -0.1875 | -0.1875 |
| **ROW15** | 0.0625 | -0.0625 | -0.0625 | 0.1875 | -0.0625 | 0.1875 | -0.0625 | 0.0625 | -0.1875 | 0.0625 | 0.0625 | -0.1875 | 0.0625 | -0.1875 | 0.5625 | -0.1875 |
| **ROW16** | 0.0625 | -0.0625 | -0.0625 | 0.1875 | -0.0625 | -0.0625 | 0.1875 | 0.0625 | 0.0625 | -0.1875 | 0.0625 | 0.0625 | -0.1875 | -0.1875 | -0.1875 | 0.5625 |

1. For the either the Dog data or Beta Carotene data, design and compute 2 contrasts and 2 estimates (other than those done in class or previously). Create your tests and estimates based on what you think is interesting. With the output, write up your results in a few sentences.

**Since you have freedom to choose the estimate or test, this will be graded on an individual basis. Note that I was not able to verify complete correctness of your estimates and contrasts since they varied by student, so if you have questions about whether you wrote it correctly, let me know. I won’t count off at this point if I did not already.**