

- (1) In 6612 you may have been told that for a class variable in a GLM or LMM, you need to pick one level as the reference, and that the estimates for other levels become a comparison to the reference. Using concepts of estimability in a less-than-full-rank model, prove that this is true. To make the argument, consider SAS's approach to determining a g-inverse, such as we did in class.

If we use SAS's g-inverse, the estimate for the highest level of a factor is 0, i.e., this level is treated as the reference level based on how the g-inverse is constructed. We also know that comparisons of levels of a factor are estimable. So, say, for a 4-level factor,  $\alpha_i - \alpha_4$  is estimable, for  $i=1,2,3$ . What this means is that estimate does not change based on the g-inverse you use, or what level you set as the reference (whichever way you want to think about it). So using SAS's g-inverse approach, we get  $\hat{\alpha}_i - \hat{\alpha}_4 = \hat{\alpha}_i - 0 = \hat{\alpha}_i$  as the unique estimator of  $\alpha_i - \alpha_4$ . Further notes: if we had used the M-P inverse, then the estimate of level  $i$  does not compare to a reference level; in that case alpha estimates will all be nonzero. Thus you could say  $\hat{\alpha}_i^{MP} - \hat{\alpha}_4^{MP} = \hat{\alpha}_i^{SAS}$  is the unique estimator of  $\alpha_i - \alpha_4$ .

- (2) Derive  $Var(\hat{\beta})$  in a full-rank linear mixed model, given the algebraic form of  $\hat{\beta}$  that is obtained via ML estimation. NOTE: there are two types of variance, model-based and empirical (or sandwich ☺). The difference is whether the middle 'V' is determined via the model or using squared residual quantities; derive the model-based form. To answer this question, work with the 'complete data' form of  $\hat{\beta}$ .

The ML estimator has form  $\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y$ , which is a linear form of  $Y$ . The linear form result says  $Var(AY) = AVar(Y)A'$ , so let  $A = (X'V^{-1}X)^{-1}X'V^{-1}$  and

$$\begin{aligned}
 Var(\hat{\beta}) &= AVar(Y)A' \\
 &= (X'V^{-1}X)^{-1}X'V^{-1}Var(Y)[(X'V^{-1}X)^{-1}X'V^{-1}]' \\
 &= (X'V^{-1}X)^{-1}X'V^{-1}V[(X'V^{-1}X)^{-1}X'V^{-1}]' && \text{For model-based variance, we let } Var(Y)=V \\
 &= (X'V^{-1}X)^{-1}X'V^{-1}V(V^{-1})'X[(X'V^{-1}X)^{-1}]' && \text{Applying transpose on right} \\
 &= (X'V^{-1}X)^{-1}X'V^{-1}VV^{-1}X(X'V^{-1}X)^{-1} && A=A' \text{ if } A \text{ is symmetric (terms on right)} \\
 &= (X'V^{-1}X)^{-1}X'V^{-1}X(X'V^{-1}X)^{-1} && \text{Applying regular inverse in middle} \\
 &= (X'V^{-1}X)^{-1} && \text{Applying regular inverse on left}
 \end{aligned}$$

- (3) For a linear mixed model that could either be full rank or less than full rank, for  $L\beta$  that is estimable, determine  $Var(L\hat{\beta})$ . A couple of tips: (a) the M-P inverse of a square symmetric matrix is also symmetric, (b) we know that  $P_X = X(X'X)^{-1}X'$  (in a GLM) is unique, i.e., not dependent on the g-inverse used; it is also true that  $X(X'V^{-1}X)^{-1}X'$  is unique in an LMM.

So say the model is less-than-full-rank (which is the general case). We have  $L\tilde{\beta} = a'X(X'V^{-1}X)^{-1}X'Y$ . First, note that the quantity is unique (does not depend on g-inverse used) since  $X(X'V^{-1}X)^{-1}X'$  is unique;  $L\tilde{\beta}$  is normally distributed with unique mean and variance. We can employ the linear form result, letting  $A = a'X(X'V^{-1}X)^{-1}X'$ . The linear form result says  $Var(AY) = AVar(Y)A'$ . Again using the model-based form,  $Var(Y)=V$ , so altogether we have  $AVA' = a'X(X'V^{-1}X)^{-1}X'VX(X'V^{-1}X)^{-1}X'a$ . Looks messy, but we know that every matrix has a M-P inverse, so  $X'VX$  has a M-P inverse such that

$(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}$ . Thus,  $\text{Var}(\mathbf{L}\hat{\boldsymbol{\beta}}) = \mathbf{a}'\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{a} = \mathbf{L}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{L}'$ . Given that it is a unique solution (again, we have  $\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'$  on the interior), we must arrive at the same solution even if we had used something other than the M-P inverse.

- (4) In a short paragraph, explain the difference between a general linear model (GLM; not a generalized linear model, which I denote with GzLM and which will be discussed more later) and a linear mixed model (LMM).

Basically, a general linear model (GLM) is for independent (e.g., cross-sectional) data, and a linear mixed model (LMM) accounts for correlated data. The GLM is a special case of the LMM when there are no random effects and the error covariance matrix is simple ( $\sigma^2 \mathbf{I}$ ). Both modeling approaches are regression-type models, where we are trying to understand the relationship between an outcome and several predictors. For the LMM, modeling the correlation (and covariance parameters in general) is usually a nuisance process (something we need to do but are not directly interested in). However, there are situations where we may be interested in random-effect estimates themselves, or even the other covariance parameter estimates.

- (5) For 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.

**This one will obviously vary by person. You can get a sense of whether the coefficients were written as long as you don't get a 'not estimable' statement.**

- (6) Consider a study where subjects in 3 groups (e.g., race or treatment) are observed over 3 equally spaced 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.

a. Consider modeling group and time as class variables, plus interaction. Write statistical models and the X matrix for the following cases.

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

$$Y_{hij} = \beta_0 + \alpha_h + \tau_j + \gamma_{hj} + b_i + \varepsilon_{hij},$$

$i=1,\dots,6$  for subjects,

$h=1,2,3$  for groups,

$j=1,2,3$  for measures,

$b_i \sim iid N(0, \sigma_b^2)$  and  $\varepsilon_{hij} \sim iid N(0, \sigma_\varepsilon^2)$ .

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

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

$$Y_{ij} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_{1i} x_3 + \beta_6 x_{2i} x_4 + \varepsilon_{ij}$$

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

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

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

**You can use SAS PROC IML or R to construct  $\mathbf{H}$ ; 'ginv' is the function in both that uses the MP inverse. So, for example, you can use  $\mathbf{h}=\text{ginv}(\mathbf{t}(\mathbf{x})*\mathbf{x})*\mathbf{t}(\mathbf{x})*\mathbf{x}$ ; in SAS PROC IML. Just use the  $\mathbf{x}$  from ai. Note that  $\mathbf{L}=(0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ -1 \ 0 \ 1 \ 1 \ 0 \ -1 \ 0 \ 0 \ 0)$  and you will see that  $\mathbf{LH}$  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  $\mathbf{H}$  (you did not need to turn this in). Note that if you multiple this matrix by 16, you will get nice integers. So you can write  $\mathbf{H}=(1/16)*\text{"Nice H"}$ .**

	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

- c. How would answers in a change an AR(1) structure for  $\mathbf{R}$  is included? (You do not need to rewrite entire models, just what changes).

Instead of  $\varepsilon_{hij} \sim iid N(0, \sigma_\varepsilon^2)$ , the errors have the AR(1) structure within subject  $i$ . You can write this as  $\varepsilon_i \sim N(\mathbf{0}, \mathbf{R}_i)$ , where  $\mathbf{R}_i$  has the AR(1) structure. This applies to both the FR and LTRF models.

- d. Say that Time is treated as continuous (i.e., not included in the CLASS statement in SAS or factor argument in R). Rewrite the models and X matrices in a. Say the linear term for Time is sufficient.

$Y_{hij} = \beta_0 + \alpha_h + \beta_1 x_j + \gamma_h x_j + b_i + \varepsilon_{hij}$ ,  $i=1, \dots, 6$  for subjects,  $h=1, 2, 3$  for groups,  $j=1, 2, 3$  for measures, where  $b_i \sim iid N(0, \sigma_b^2)$  and  $\varepsilon_{hij} \sim iid N(0, \sigma_\varepsilon^2)$ . NOTE: for  $\mathbf{X}$  below, I am using 1, 2 and 3 as the 3 times; you could make more general and say 'x1', 'x2' and 'x3'.

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 2 & 2 & 0 & 0 \\ 1 & 1 & 0 & 0 & 3 & 3 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 2 & 2 & 0 & 0 \\ 1 & 1 & 0 & 0 & 3 & 3 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 2 & 0 & 2 & 0 \\ 1 & 0 & 1 & 0 & 3 & 0 & 3 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 2 & 0 & 2 & 0 \\ 1 & 0 & 1 & 0 & 3 & 0 & 3 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 2 & 0 & 0 & 2 \\ 1 & 0 & 0 & 1 & 3 & 0 & 0 & 3 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 2 & 0 & 0 & 2 \\ 1 & 0 & 0 & 1 & 3 & 0 & 0 & 3 \end{pmatrix}$$

- e. Say that the times of observation were at 0, 1 and 6 months rather than equally spaced.
- i. Would it be appropriate to treat time as a class variable in this case? Explain.

There is no problem in using equally spaced or unequally spaced times for a class variable, since you are estimating levels separately. The unequal spacing does not impose any constraints metrically.

- ii. Suggest a structure for  $\mathbf{R}_i$  and write it out.

There are a couple of possibilities. Since there are only 3 times, it is not very expensive to use the UN structure, since it only adds 6 covariance parameters.

$$\mathbf{R}_i = \begin{pmatrix} \sigma_0^2 & \sigma_{01} & \sigma_{06} \\ \sigma_{01} & \sigma_1^2 & \sigma_{56} \\ \sigma_{06} & \sigma_{56} & \sigma_6^2 \end{pmatrix}$$

Another option would be the spatial power structure. It only adds 2 covariance parameters and handles the unequal spacing:

$$\mathbf{R}_i = \sigma_\varepsilon^2 \begin{pmatrix} 1 & \phi & \phi^6 \\ \phi & 1 & \phi^5 \\ \phi^6 & \phi^5 & 1 \end{pmatrix}$$