BIOS 667, Spring 2017 Midterm

1. Consider the model $E[Y] = X\beta$, $\text{cov}(Y) = \Sigma$, where Y is an $n \times 1$ random vector distribted as multivariate normal with mean $X\beta$ and covariance matrix Σ; β is $p \times 1$ and Σ depends on a $q \times 1$ vector θ that is functionally unrelated to β.

   (a) True or false: The REML estimator of the regression coefficients β are obtained by maximizing the REML likelihood with respect to β. Explain briefly (1–3 sentences).

   (b) True or false: The REML estimator of β is unbiased. Explain briefly.

   (c) True or false: The maximum-likelihood (full likelihood, not REML likelihood) estimator of β is unbiased. Explain briefly.

2. In the TLC study, in the "Active" group, suppose that, using i to index subjects, the otucome vector $(Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4})^{\top}$ is distributed as multivariate normal with mean $(25, 14, 16, 18)^{\top}$ in μg/dL, and covariance matrix

$$50 \begin{bmatrix} 1 & 0.5 & 0.5 & 0.4 \\ & 1 & 0.6 & 0.5 \\ & & 1 & 0.5 \\ & & & 1 \end{bmatrix}.$$

   Define $A_i = (Y_{i2} + Y_{i3} + Y_{i4})/3$.

   All numerical values must be simplified and reduced as much as possible.

   (a) Find the mean and variance of A_i.

   (b) Find $\text{cov}(Y_{i1}, A_i)$.

   (c) Find the conditional mean and variance of A_i given Y_{i1}.

   (d) What is the point of this question?

3. In the TLC study, one of the study statisticians asked for fitting 6 separate linear regression models: regression of Y_{ij} on Y_{i1} (with intercept), $j = 2, 3, 4$, for i in the Active group, and separately for i in the Placebo group. The estimated slopes (standard errors) were:

	A	P
j=2	0.613(0.202)	0.901(0.0877)
j=3	0.600(0.208)	0.961(0.0898)
j=4	0.912(0.231)	0.848(0.106)

   (a) The said statistician wanted to test the hypothesis that, at each occasion j, the corresponding true population slopes, say α_{1j} (A) and α_{0j} (P) are equal in the two groups. Why?

   (b) Perform the test for $j = 2$ only (the test can be either exact or approximate).

   (c) The tests above for $j = 2, 3, 4$ are correlated. How would you test all three hypotheses as a single hypothesis, $H_0 : \alpha_0 = \alpha_1$ (vectors) versus $H_0 : \alpha_0 \neq \alpha_1$? You can use the full original data set.

BIOS 667, Spring 2017, 2nd Midterm Exam

The notation for linear mixed models established in class will be used, $Y_{ij}, b_i, \nu_{ij}, \mu_{ij}, x_{ij}, Y_i, X_i, n_i, G, R_i,$ etc. Also, normality (uni- or multi-, as appropriate) of Y_i given b_i and normality of b_i are assumed throughout.

1. Suppose that in a longitudinal study there are 2 observations for each subject. Consider the model
$$\nu_{ij} = \mathrm{E}[Y_{ij}|b_i] = (\beta_1 + b_{i1}) + (\beta_2 + b_{i2})x_{ij}$$
where $x_{i1} = 0, x_{i2} = 1,\ \beta_1 = 1, \beta_2 = 1,\ g_{11} = g_{12} = 1, g_{22} = 4,\ R_i = I_{2\times 2}$, i=1, ..., K. The covariate x_{ij} varies over time, for example, a treatment indicator in a cross-over study, dose of a drug that is intentionally varied over time, etc.

   In the following, simplify the answers as much as possible.

   (a) Give two interpretations of β_2.

   (b) Compute $\mathrm{corr}(Y_{i1}, Y_{i2})$.

   (c) Find the marginal distribution of the response vector Y_i.

   (d) Find the best linear unbiased predictor (BLUP) of b_{i2} based on *only* Y_{i2} (not the vector Y_i).

   (e) Find the prediction mean squared error of the BLUP in the last part.

   (f) A random subject drawn from the same population has 0 observations. What is the BLUP of b_{i2} for that subject? What is the prediction mean squared error?

   (g) From this part on, assume that β is unknown (and possibly different from the numerical values given above).
   Define $D_i := Y_{i2} - Y_{i1}$. What is the distribution of D_i? Explain how β_2 can be estimated using only the D_i's. The variance of the best linear unbiased estimator of β_2 based on the D_i's is δ/K. Find the value of δ.

   (h) Suppose that β is estimated by weighted least squares (WLS), using the vectors $\{Y_i\}$. The variance of the WLS estimator of β_2 is γ/K. Find the value of γ.

2. Consider the model M_1:
$$\nu_{ij} = \mathrm{E}[Y_{ij}|b_i] = (\beta_1 + b_{i1}) + (\beta_2 + b_{i2})x_{ij}.$$
Using data from a large number of independent subjects, the model parameters were estimated by REML (not full ML) and the maximized -2 log-likelihood was 200.

   A model differing from M_1 only by not having b_{i2} in it was estimated by REML (using the same data) and the maximized -2 log-likelihood was 205.

   In the context of model M_1, and given the above information, is it possible to test the null hypothesis $g_{22} = 0$ against the alternative hypothesis $g_{22} > 0$. If yes, give the details. If no, explain why.