UNIVERSITY OF MASSACHUSETTS
Department of Mathematics and Statistics
ADVANCED EXAM - LINEAR MODELS
August 31, 2009 (3 hours)

- Do all problems.
- Start each problem on a new page.
- A total of 70 points is needed to pass.

1. (*16 points*) Consider a general linear model

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon},$$

where $\underline{Y}$ is $n \times 1$; X is $n \times p$ with rank $r(\leq p)$; $\underline{\beta}$ is $p \times 1$, and $\underline{\varepsilon}$ is an n-dimensional random vector with $\mathrm{E}(\underline{\varepsilon}) = \underline{0}$.

   (a) Define what it means for $\underline{c}'\underline{\beta}$ to be estimable and show that $\underline{c}'\underline{\beta}$ is estimable if and only if $\underline{c}'$ is a vector in the linear space generated by the row vectors of X.

   (b) State and PROOF the Gauss-Markov theorem relating to an estimable linear function of $\underline{\beta}$, for $Cov(\underline{\varepsilon}) = \sigma^2 I$.

2. (*54 points*) Let $Y_{ij},\ j = 1, 2, \ldots, n_i,\ i = 1, 2, \ldots, m$ be observable r.v.'s such that

$$Y_{ij} = \alpha_0 + \alpha_i + \varepsilon_{ij},$$

where the α's are unknown parameters, the ε_{ij} are i.i.d. $N(0, \sigma^2)$ unobservable r.v.s, and $\sigma^2 > 0$ is unknown.

   (a) Let $\psi = \sum_{i=0}^{m} \ell_i \alpha_i$ with ℓ_i being specified constants. Show that ψ is estimable if and only if $\ell_0 = \sum_{i=1}^{m} \ell_i$.

   (b) Define $\mu_i = \alpha_0 + \alpha_i$, $i = 1, 2, \ldots, m$, which will be used for all the following questions as well. Write down the least squares estimates for the μ's. Are they UMVUE's? Why? (Explain briefly). Write down also the UMVUE of σ^2 (no need to prove the results).

   (c) Use the full-reduced model approach to construct the F test for testing H_0: all μ_i are equal versus H_1: at least two μ_i are different, using a test of size α. Write out the F-statistic explicitly and state the distribution of the test statistic, under both null and alternative hypotheses.

   (d) Here we are interested in simultaneous confidence intervals, with confidence $1 - \alpha$, for all pairwise differences in the means; e.g., differences $\mu_j - \mu_{j'}$, $j \neq j'$.

i. Give Bonferroni's inequality and explain how it is used to construct simultaneous confidence intervals. (No need to prove anything you can just write down the result.) Be complete.

ii. For $n_i = n$ $(i = 1, \ldots, m)$, define the studentized range distribution and use it to derive Tukey's method for the problem at hand.

iii. How would you decide which of the two methods is better here? Why is the Bonferroni method so popular in simultaneous confidence intervals?

(e) For this question, suppose there is a variable associated with each of the groups (e.g., a dose of some sort) with x_i denoting the value for group i, each $x_i > 0$ and all the x_i being different.

i. If it is known that $V(\epsilon_{ij}) = \sigma^2 x_i$, explain how you could transform the model to make optimal inferences on the μ's. Could you use a standard one-way analysis of variance routine found in the statistical packages or would you need a general regression package? Explain.

ii. Now suppose the x_i values are used to model the mean with the assumption that $\mu_i = \beta_0 + \beta_1 x_i$, but return to assuming that $V(\varepsilon_{ij}) = \sigma^2$.

A. Write the model as $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$. Write down the least squares estimate of $\underline{\beta}$ in terms of $\underline{Y}$ and X and the UMVUE for σ^2. No need to simplify these.

B. Use the full-reduced model approach to develop the F-test for the null hypothesis of $H_0 : \mu_i = \beta_0 + \beta_1 x_i$ versus H_A : the μ_i are not a linear function of the x_i, but change over i in some unspecified way - (This is just the model in part (b)). This is a test for "lack of fit". Lay out the general calculation and the degrees of freedom associated with the test, but there is no need to simply the expression.

3. (*30 points*) An experiment was run to compare three different primitive altimeters (an altimeter is a device which measures altitude). Each of three pilots used each of three altimeters and the response is the error in reading.

	Altimeter		
	1	2	3
Pilot 1	3	4	7
Pilot 2	6	5	8
Pilot 3	3	4	7

We assume first that the three pilots are FIXED factor levels, and consider it as a two-way fixed effects model with interactions.

(a) Plot these 9 observations to see any indication of Pilot-Altimeter interactions (no computation)? What should the graph look like if there is no interaction?

(b) Clearly with one observation per cell we cannot allow full interaction between pilots and altimeters, with interactions being denoted by γ_{ij}. There is a test for additivity assuming γ_{ij} to be of a special form. What is one form that can be used to test for interaction?

FROM NOW ON we consider that pilots (and hence pilot effects) are RANDOM, and that there is no interactions between pilots and altimeters. That is, we consider the model

$$Y_{ij} = \mu + A_i + \beta_j + \varepsilon_{ij},$$

where Y_{ij} is the observation for Pilot i and Altimeter j; A_1, A_2, and A_3 are i.i.d. $N(0, \sigma_A^2)$; $\sum_{j=1}^3 \beta_j = 0$, and ε_{ij} are assumed i.i.d. $N(0, \sigma^2)$ which are independent of the A's. Let $\underline{Y}' = [Y_{11}, Y_{12}, \cdots, Y_{33}]$.

(c) Find the covariance matrix of $\underline{Y}$.

(d) Let $\overline{Y}_{\cdot j} = \sum_{i=1}^3 Y_{ij}/3$. Find

$$\text{Cov} \begin{bmatrix} \overline{Y}_{\cdot 1} \\ \overline{Y}_{\cdot 2} \\ \overline{Y}_{\cdot 3} \end{bmatrix}$$

(e) Let $\text{SS}_T = \sum_{j=1}^3 3(\overline{Y}_{\cdot j} - \overline{Y}_{\cdot\cdot})^2$. Show that SS_T/K is distributed Chi-square with 2 degrees of freedom and noncentrality parameter λ, where K is a constant that will depend on parameters. Give K and λ explicitly. What is $E(SS_T)$?

− − − The End − − −