UNIVERSITY OF MASSACHUSETTS  
Department of Mathematics and Statistics  
ADVANCED EXAM - LINEAR MODELS  
Friday, August 29, 2008  

Work all problems. 75 points are required to pass.  
- Read the questions carefully.  
- Where is says “state ...” you can state the result being asked for without proof.  

1. (30 PTS) Let $Y_1$ and $Y_2$ be independent random variables, $E(Y_i) = \mu + \alpha_i$ for $i = 1, 2$.  

(a) Let $\psi = c_1\alpha_1 + c_2\alpha_2$, where $c_1$ and $c_2$ are constants. Define what it means for $\psi$ to be estimable and then show what condition $c_1$ and $c_2$ must satisfy for $\psi$ to be estimable.  

(b) Use the previous part to show that $\alpha_2$ is not estimable and then to decide which, if either, of $\alpha_1 + \alpha_2$ and $\alpha_1 - \alpha_2$ is estimable.  

(c) With $\beta = (\mu, \alpha_1, \alpha_2)'$ and $Y' = (Y_1, Y_2)$, write $E(Y)$ as $X\beta$. Write out $X$, which is $2 \times 3$, explicitly with numbers.  

(d) Note that $X$ is not of full column rank (explain why!) and so the least squares estimate of $\beta$ (which solves $X'X\beta = X'Y$) is not unique. In order to avoid dealing with the singular matrix $X'X$ in computing a least squares estimator for $\beta$ one can impose some side conditions (also called constraints) on linear combinations of $\beta$. What type of side conditions and how many are needed in this problem to force a unique solution to $X'X\beta = X'Y$? Justify why these side conditions force a unique solution. Note: Part d) is needed to go on with the problem. If you can’t get it you can buy the answer (giving up 6 points) in order to proceed.  

(e) Using the previous part, select a side condition (or conditions) and rewrite the reparameterized model as $E(Y) = Z\gamma$. Be sure to specify $Z$ (which will be $2 \times 2$ of rank 2) and $\gamma$ explicitly.  

(f) Now compute the least squares estimate, $\hat{\gamma}$, of $\gamma$, explicitly (inverting and multiplying out the matrices) in terms of $Y_1$ and $Y_2$. Then write down a least squares estimate $\hat{\beta}$ of $\beta$.  

(g) Write down the BLUE for $\alpha_1 - \alpha_2$ explicitly in terms of $Y_1$ and $Y_2$. Which theorem have you employed to guarantee your answer being the BLUE? State the content of the theorem.
2. (30 PTS) Consider the linear model

\[ Y = X\beta + z\phi + \epsilon, \]  
(1)

with \( E(\epsilon) = 0 \) and \( \text{Cov}(\epsilon) = \sigma^2 I \). \( X \) is a known \( n \times p \) matrix of rank \( p < n \), \( z \) is a known \( n \times 1 \) vector and \( \beta \ (p \times 1) \), \( \phi \) (scalar) and \( \sigma^2 \) (scalar) are unknown parameters.

(a) Suppose the \( z\phi \) terms are ignored and the model assuming \( E(Y) = X\beta \) is fit; leading to \( \hat{\beta} = (X'X)^{-1}X'Y \), the ordinary least squares estimator using just \( X \). Let \( r_i = Y_i - \hat{Y}_i \) denote the resulting \( i \)th residual (that is using \( \hat{Y} = X\hat{\beta} \)) and let \( r \) be the \( n \) by 1 vector of residuals.

Derive \( E(r) \) and \( \text{Cov}(r) \) (assuming (1) holds).

(b) A plot of \( r_i \) versus \( z_i \) (the \( i \)th element of \( z \)) is often suggested as a way to assess if the variable \( z_i \) should be in the model. Explain (using your expression for \( E(r) \)) why this plot is often useful and when it might encounter some problems. Assume that if \( z_i \) enters into the model it does so via \( z_i\phi \).

(c) Consider a full least squares fit of the model in (1). Let \( M = X(X'X)^{-1}X' \). Show that

\[ \hat{\phi} = z'(I - M)Y \]
\[ z'(I - M)z. \]  
(2)

Do this by first rewriting (1) as \( Y = X\delta + (I - M)z\phi + \epsilon \) where \( \delta \) can involve both parameters and elements of \( X \) and/or \( z \).

(d) A client says ‘well, if the plot of \( r_i \) versus \( z_i \) represents the influence of \( z_i \) after accounting for the other variables, is it the case that if I ran a simple linear regression of \( r_i \) on \( Z_i \) that the slope I get will be the estimate \( \hat{\phi} \) in (2) from a full least squares fit of (1)?

Show that the answer to this question is no. Then show however that if we define \( w = (I - M)z \) (this is \( n \times 1 \)) and we regress \( r_i \) on \( w_i \) with NO intercept, then the estimated slope we obtain is exactly \( \hat{\phi} \) in (2) from the full least squares fit. (This result suggests that we plot \( r_i \) versus \( w_i \) rather than just \( z_i \) as this plot matches up with the estimate of \( \phi \) from the full least squares approach; this is known as an added variable plot.)
3. (40 PTS) Consider the one-factor fixed effects model: $Y_{ij} = \mu_i + \epsilon_{ij}, i = 1$ to $I$ and $j = 1$ to $n_i$, where $\mu_1, \ldots, \mu_I$, are fixed parameters (means) and the $\epsilon_{ij}$ are i.i.d. normal with mean 0 and variance $\sigma^2$.

(a) Write this out as a linear model, $Y = X\mu + \epsilon$, with $\mu' = (\mu_1, \ldots, \mu_I)$ and argue that the least squares estimator of $\mu$ has $\hat{\mu}_i = \bar{Y}_i = \sum_{j=1}^{n_i} Y_{ij}/n_i$. (You can just state the general form of the least squares estimator for a linear model and apply it here.)

(b) Show that $\hat{\sigma}^2 = \sum_i(n_i - 1)S_i^2 / (n - I)$ is an unbiased estimator of $\sigma^2$, where $S_i^2 = \sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_i)^2 / (n_i - 1)$ and $n = \sum_i n_i$. (Do this without using the normality assumption.) You can utilize the computing formula $\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_i)^2 = \sum_{j=1}^{n_i} Y_{ij}^2 - n_i\bar{Y}_i^2$.

(c) Using just the observations from “group” i, collected in $Y_i = (Y_{i1}, \ldots, Y_{in_i})'$, first write $(n_i - 1)S_i^2 / \sigma^2$ as a quadratic form in $Y_i$. Then state a general theorem on when a quadratic form is distributed chi-square and show how that result applies here to give the distribution of $(n_i - 1)S_i^2 / \sigma^2$.

(d) Use the previous part and whatever else you need to provide the distribution of $(n - I)\hat{\sigma}^2 / \sigma^2$.

For the rest of the problem you can use, without proof, that $\hat{\sigma}^2$ is independent of $\hat{\mu}$.

(e) State generally, Scheffe’s result for finding simultaneous confidence intervals for a collection of linear combinations of the coefficients in the general linear model. Then apply this result to give simultaneous confidence intervals for all pairwise differences of the $\mu_i$’s. In doing this last part a) define what a contrast is b) justify that the set of contrasts can be obtained by taking a linear combinations of a basis set consisting of $r$ linear combinations of the $\mu_i$’s, being sure to justify exactly what $r$ is.

(f) Now assume that all $n_i = n_1$. Argue that the the distribution of

$$Q = \frac{Max_i(Y_{i} - \mu_i) - Min_i(Y_{i} - \mu_i)}{\hat{\sigma}/n_1},$$

has a distribution that does not depend on any unknown parameters (but will depend on $d = n - I$ and $I$). Note: You do not have to find the density to do this.

(g) Use the previous part to derive simultaneous confidence intervals for all pairwise differences of the form $\mu_i - \mu_k$. In writing out your answer you can use $q_{\alpha,d,1}$ to denote the value for which $P(Q \leq q_{\alpha,d,1}) = 1 - \alpha$.

(h) In the general linear model with $Y = X\beta + \epsilon$ with with $E(\epsilon) = 0$ and $Cov(\epsilon) = \sigma^2I$ and $X$ being $n \times p$ of rank $p$, there are two ways to write out the F-statistic (which is the likelihood ratio test) for testing $H_0 : H\beta = h$, where $H$ is $q \times p$ of rank $q$. One is using a general matrix form, the other is using the full-reduced model approach.

i. For this problem, suppose $I = 2$ and consider testing $H_0 : \mu_1 = \mu_2$. Use each of the two methods to develop the F-test for this hypothesis (they will yield the same result). Give a final form that involves $\bar{Y}_1$, $\bar{Y}_2$, $\hat{\sigma}^2$ and the sample sizes $n_1$ and $n_2$.

ii. Set-up how you would compute the power of the test in the previous part (for testing $\mu_1 = \mu_2$ specifically). You can leave your answer in the form of an integral with integrand $f(x; d_1, d_2, \lambda)$ = density of a non-central F-distribution with $d_1$ and $d_2$ degrees of freedom and non-centrality parameter $\lambda$. You do not need to write out the density involved but be sure to specify the limits of integration along with $d_1$, $d_2$ and $\lambda$ for this particular problem.