UNIVERSITY OF MASSACHUSETTS
Department of Mathematics and Statistics
ADVANCED EXAM - LINEAR MODELS
Friday, August 31, 2007

Work all problems. 75 points are required to pass with at least 2 out of problems 1, 3 and 4 essentially complete.

1. (25 points) Let $\mathbf{Y}$ be an $n \times 1$ random vector, $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$, where $\mathbf{X}$ is an $n \times p$ known matrix with non-full rank r $(1 \leq r < p)$, and $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters. We assume that elements in $\mathbf{Y}$ are uncorrelated. Let $\psi = \mathbf{a}'\beta$ for some p-vector $\mathbf{a}$.

   (a) Give the definition for ψ to be estimable, in terms of an expected value involving $\mathbf{Y}$.

   (b) Derive (with justifications) a necessary and sufficient condition for $\mathbf{a}$ so that $\mathbf{a}'\boldsymbol{\beta}$ is estimable in terms of the matrix $\mathbf{X}$ and how it relates to $\boldsymbol{a}$. You need to show both necessity and sufficiency.

   (c) Write down the normal equation for finding least squares estimator and then state carefully what constraints must be put on $\boldsymbol{\beta}$ in order to force a unique solution to the normal equations. (Recall that $\mathbf{X}$ is not of full column rank.) You don't need to prove that this yields a unique solution but state carefully what the conditions are.

   (d) Suppose that $\mathbf{a}'\boldsymbol{\beta}$ is an estimable function of $\boldsymbol{\beta}$. State the Gauss-Markov theorem concerning the "best estimator" of $\mathbf{a}'\boldsymbol{\beta}$ and, in light of the fact that there are many solutions to the normal equations, how you would compute the best estimator of $\boldsymbol{a}'\boldsymbol{\beta}$.

   (e) In the above discussions, does $\mathbf{Y}$ need to have a multivariate normal distribution? Use one or two sentences to explain your answer.

2. (15 points) Let $\mathbf{Y}$ be an $n \times 1$ multivariate normal random vector, with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ which is positive definite. Let $\mathbf{A}$ be a non-zero $n \times n$ symmetric matrix.

   (a) State a necessary and sufficient condition for the quadratic form $\mathbf{Y}'\mathbf{A}\mathbf{Y}$ to have a non-central chi-squared distribution with degrees of freedom r and non-centrality parameter λ. Specify what r and λ are.

   (b) Now, if
$$\mu = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1/5 & 0 \\ 0 & 1/5 \end{pmatrix}$$
   and
$$Q = y_1^2 + 4y_1y_2 + 4y_2^2.$$
   Show that Q has a non-central chi-squared distribution with a specified degrees of freedom and non-centrality parameter, where you give a numerical value for each.

3. (30 points) Consider the one-factor random effects model $Y_{ij} = \mu + A_i + \epsilon_{ij}$ where $i = 1, \ldots I > 2$, $j = 1, \ldots, J > 2$, μ is a fixed unknown parameter and the A_i and ϵ_{ij} are independent normal random variables with mean 0, $Var(A_i) = \sigma_A^2 > 0$ and $Var(\epsilon_{ij}) = \sigma_\epsilon^2 > 0$.

Define

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ . \\ \mathbf{Y}_I \end{bmatrix} \text{ with } \mathbf{Y}_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ . \\ Y_{iJ} \end{bmatrix}.$$

Let

$$F = \frac{S_1^2/(I-1)}{S_2^2/I(J-1)}$$

be the usual F-statistic for testing $H_0 : \sigma_A^2 = 0$, where $S_1^2 = J\sum_i(\bar{Y}_{i.} - \bar{Y}_{..})^2$ and $S_2^2 = \sum_i\sum_j(\bar{Y}_{ij} - \bar{Y}_{i.})^2$.

(a) Find $Cov(\mathbf{Y})$ which you can give by describing $Cov(\mathbf{Y}_i)$ for each i and $Cov(\mathbf{Y}_i, \mathbf{Y}_k)$ for $i \neq k$.

(b) Find the distribution of each of S_1^2 and S_2^2. You can do this by getting constants c_1 and c_2 so that $c_k S_k^2$ has a well known distribution (k = 1 or 2), being sure to state the parameters involved.

If for some reason you cannot get the above part you can "buy" the answer, that is get it from the exam monitor, but give up the points on that question. This will let you continue on.

(c) Argue that S_1^2 and S_2^2 are independent. (Hint: First prove that $\bar{Y}_{i.} - \bar{Y}_{..}$ has 0 covariance with $\bar{Y}_{ij} - \bar{Y}_{i.}$.)

(d) Use the previous two parts to find the distribution of F, which again you could describe by getting a constant c so that cF follows a known distribution (state the parameters involved).

(e) Use the result of (d) first derive a test of size α of $H_0 : \sigma_A^2/\sigma_\epsilon^2 = g$ versus $H_A : \sigma_A^2/\sigma_\epsilon^2 \neq g$ and show how you would get the power function for this test. Your power calculation can be left in the form of an integral involving a known distribution.

(f) Use the test in the previous part to derive a $100(1-\alpha)\%$ confidence interval for $\sigma_A^2/(\sigma_A^2 + \sigma_\epsilon^2)$.

4. (30 points) Consider data collected from different groups with a simple linear regression model for each group. That is, if Y_{ij} is the jth observation in the ith group, assume

$$Y_{ij} = \alpha_i + \beta_i X_{ij} + \epsilon_{ij}, \text{i} = 1 \text{ to I, j} = 1 \text{ to } n_i,$$

where the ϵ_{ij} are assumed i.i.d. $N(0, \sigma^2)$.

(a) Consider data from just group i. Write down the least squares estimators, $\hat{\alpha}_i$ and $\hat{\beta}_i$, and α_i and β_i. Give explicit expression; do not leave in matrix form.

(b) Considering all groups together, show that the least squares estimates of the coefficients are the same as given in part a).

(c) Assume that $\beta_1 = \ldots = \beta_I = \beta$. Show that the in this model the best estimator of β is $b = \sum_i w_i \hat{\beta}_i$ for appropriate weights $w_1, \ldots w_I$; specify what the weights are. (Hint: You can work directly from the normal equations and avoid ever finding the inverse of $\mathbf{X}'\mathbf{X}$)

(d) Consider testing the hypothesis $H_0 : \beta_1 = \ldots = \beta_I$. Show that the sum of squares due to the hypothesis is

$$\sum_{i=1}^{I} (\hat{\beta}_i - b)^2 a_{ii},$$

where $a_{ii} = \sum_j (X_{ij} - \overline{X}_{i.})^2$. What are the degrees of freedom associated with the F-test?

(e) Return to using the data for just a single group, say group 1 and consider the residuals $r_j = Y_{1j} - \hat{\alpha}_1 + \widehat{\beta}_1 X_{1j}$ set up in an $n_1 \times 1$ vector $\mathbf{r}$.

- Find $Cov(\mathbf{r})$ (best to do this using the matrix form of $\mathbf{r}$).

- Find a constant c_j (it might involve an unknown parameter) so that the variance of r_j/c_j is the same for all j. Give an explicit expression for c_i. Note that this constant motivates the use of standardized residuals for residual analysis.