UNIVERSITY OF MASSACHUSETTS
Department of Mathematics and Statistics
ADVANCED EXAM - LINEAR MODELS
Friday, January 23, 2009

Work all problems. 75 points are required to pass.

1. (20 points) Consider data collected according to the simple linear regression model

$$Y_i = \alpha + \beta X_i + \epsilon_{ij}, i = 1, \ldots, n$$

where the ϵ_i are assumed i.i.d. $N(0, \sigma^2)$.

   (a) Write down the least squares estimators, $\hat{\alpha}$ and $\hat{\beta}$ for α and β. Give explicit expressions; do not leave in matrix form.

   (b) Write down an unbiased estimator, $\widehat{\sigma}^2$ for σ^2:

   (c) State what you know about the joint distribution of $\widehat{\alpha}$, $\widehat{\beta}$ and $\widehat{\sigma}^2$.

   (d) Use the previous part to **derive** a 95% confidence interval for $\alpha_1 + \beta_1 x_0$ for a specified x_0.

   (e) Consider x_0 fixed and a future random variable to be observed from $Y_0 = \alpha + \beta x_0 + \epsilon_0$, where ϵ_0 is independent of all of the other random quantities in the problem. **Derive** a 95% prediction interval for Y_0.

   (f) Consider the residuals $r_i = Y_i - (\hat{\alpha} + \widehat{\beta} X_i)$, set up in an $n \times 1$ 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.

2. (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, 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 (c) to 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 then use this to derive a $100(1-\alpha)\%$ confidence interval for $\sigma_A^2/(\sigma_A^2 + \sigma_\epsilon^2)$.

(f) Now special the previous part to get a test of $H_0 : \sigma_A^2 = 0$ and show how you would get the power function for this test. Your power calculation can be left in the form of an integral

3. (30 points) Consider the following three-treatment analysis of covariance model:

$$y_{ij} = \mu + \tau_i + \gamma_i x_{ij} + \varepsilon_{ij}$$

for $i = 1, \ldots, 3$, $j = 1, \ldots, n_i$ where we assume the ϵ_{ij} are i.i.d. $N(0, \sigma^2)$.

(a) Define estimability and explain why, without some restrictions the individual τ_i's are not estimable.

For the remainder of the problem assume $\tau_1 + \tau_2 + \tau_3 = 0$.

(b) Making use of the constraint in τ_i by representing τ_3 in terms of τ_1 and τ_2, write the model in the following matrix form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

by specifying $\mathbf{Y}$, $\mathbf{X}$, $\boldsymbol{\beta}$, and $\boldsymbol{\varepsilon}$.

We will assume that the x_{ij} are such that with this constraint incorporated $\mathbf{X}$ is of full column rank. There is no need in any of the following expressions to simplify $(\mathbf{X}'\mathbf{X})^{-1}$; you can leave it that way where needed.

(c) Write down the expression (in matrix form) for the L.S. estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ and state the distribution of $\hat{\boldsymbol{\beta}}$.

(d) Write down the expression for SSE (as some function of $\mathbf{Y}$, $\mathbf{X}$ and/or $\hat{\boldsymbol{\beta}}$) and specify the degrees of freedom of SSE.

(e) Set-up a 90% confidence interval for $\tau_1 - \tau_3$.

(f) Suppose we want to perform a statistical test with $\alpha = 0.05$ to see if the three groups have the same slope, i.e., to test the hypotheses

$$H_0: \quad \gamma_1 = \gamma_2 = \gamma_3 \text{ vs. } H_1: \quad \gamma_i \neq \gamma_{i'} \text{ for some } i \text{ and } i'$$

i. Specify $\mathbf{Y}$, $\mathbf{Z}$ and $\boldsymbol{\delta}$ to write the reduced model (under H_0) in matrix form

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\delta} + \varepsilon,$$

ii. Write down and expression for the error sum of squares, say SSE_1 for the reduced model, and specify the degrees of freedom of SSE_1.

iii. Give the likelihood ratio test in terms of SSE and SSE_1 (and other constants as needed), and provide the rejection rule for the test.

4. (20 points) Consider the two-way fixed effects model with $Y_{ijk} = \mu_{ij} + \epsilon_{ijk}$, for $i = 1, \ldots, I$, $j = 1, \ldots, J$ and $k = 1, \ldots, n_{ij}$ where the ϵ_{ijk} are assumed i.i.d. $N(0, \sigma^2)$.

(a) The model with μ_{ij} is called the cell means model. The effects model specifies $Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$, where the α's and β's are main effects and the γ's are interaction terms. Assuming marginal means and effects are defined using equal weights, explain how to get to the effects model from the cell means model, including specifying any restrictions.

(b) Working with the cell means model, give an estimator of $\theta = \sum_i \sum_j c_{ij}\mu_{ij}$, state the variance of the estimator and specify how to estimate the variance of the estimator.

(c) Define $\overline{\mu}_{i.} = \sum_j \mu_{ij}/J$. Provide one-at-a-time and simultaneous confidence intervals (based on Scheffe's method) for all contrasts in the $\overline{\mu}_{i.}$'s. In using Scheffe's method, state the general method and then apply it to this setting.