UNIVERSITY OF MASSACHUSETTS
Department of Mathematics and Statistics
ADVANCED EXAM - LINEAR MODELS
January 15, 2010

Work all problems. 70 points are needed to pass.

1. (35 points) Suppose that we observe statistically independent observations with the same variance σ^2, Y_{ij} for $i = 1, 2$ and $j = 1, 2$, where $E(Y_{ij})$ is modeled as follows:

$$E(Y_{11}) = \mu, \quad E(Y_{12}) = \mu + \gamma, \quad E(Y_{21}) = \mu + \alpha, \quad E(Y_{22}) = \mu + \alpha + \gamma,$$

   (a) With $\boldsymbol{\beta} = (\mu, \alpha, \gamma)'$, present this model in a linear regression form $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, and specify $\mathbf{y}$, X, and $\boldsymbol{\varepsilon}$.

   (b) What is the rank of X?, and what will be the d.f. of SSE?

   (c) Are μ, α, and γ estimable? Give a reason.

   (d) Find a least squares (LS) estimate vector $\widehat{\boldsymbol{\beta}} = (\widehat{\mu}, \widehat{\alpha}, \widehat{\gamma})'$ for $\boldsymbol{\beta}$, and write each of $\widehat{\mu}$, $\widehat{\alpha}$ and $\widehat{\gamma}$ explicitely as a linear combination of Y_{ij}'s. You may use

$$\begin{pmatrix} 4 & 2 & 2 \\ 2 & 2 & 1 \\ 2 & 1 & 2 \end{pmatrix}^{-1} = \begin{pmatrix} 3/4 & -1/2 & -1/2 \\ -1/2 & 1 & 0 \\ -1/2 & 0 & 1 \end{pmatrix}$$

   (e) Are those LS estimators $\widehat{\mu}$, $\widehat{\alpha}$ and $\widehat{\gamma}$ unique LS estimators for μ, α and γ respectively? Further, are they the best linear unbiased estimators (BLUE) for μ, α and γ respectively? Based on what theorem you make such conclusions?

   (f) Find the BLUE for $2\mu + \gamma$ with explanations.

   (g) If Y_{11} is missing (not available), show that each of μ, α and γ is still estimable.

2. (15 points)

   (a) Let $\mathbf{Y}$ be a $N_p(\boldsymbol{\mu}, \Sigma)$ random vector, where Σ is positive definit, and A be a $p \times p$ symmetric matric with rank k ($k \leq p$). State a complete theorem concerning a necessary and sufficient condition for $\mathbf{Y}'A\mathbf{Y}$ to be a chi-square distribution.

   (b) Let $U_1, \cdots, U_n$ be i.i.d. random vectors, where $U_i = (x_i, y_i)'$ has a bivariate normal distribution with mean $\boldsymbol{\mu} = (\mu_1, \mu_2)'$, $\mathrm{Var}(x_i) = \mathrm{var}(y_i) = 1$, and $\mathrm{Cov}(x_i, y_i) = \rho$ with $\rho \neq -1$ or 1, for $i = 1, \ldots, n$. Define

$$q = n(\overline{x}^2 - 2\rho\overline{xy} + \overline{y}^2)/(1 - \rho^2)$$

   wnere $\overline{x} = \mathbf{1}'\mathbf{x}/n$ and $\overline{y} = \mathbf{1}'\mathbf{y}/n$ with $\mathbf{x} = (x_1, \cdots, x_n)'$ and $\mathbf{y} = (y_1, \cdots, y_n)'$. Show that q has a chi-square distribution, and specify its degrees of freedom and non-centrality parameter. [Hint: Cosider a quadratic form in

$$\mathbf{z} = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}. \;]$$

3. (10 points) Consider the simple regression model:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \ldots, n$$

where ε_i are independent with $E(\varepsilon_i) = 0$, and $\mathrm{Var}(\varepsilon_i) = \sigma^2 x_i^2$ (non-uniform variances) with $\sigma^2 > 0$. Find the BLUE for β_0 and β_1, and an unbiased estimator for σ^2. (Note: You may give answers in matrix form, no need to be simplified.)

4. (20 points) Suppose we want to compare two simple linear models:

$$\begin{aligned} \text{Model I: } Y_{1j} &= \alpha_1 + \beta_1 x_{1j} + \varepsilon_{1j}, \;\; j = 1,2,3 \\ \text{Model II: } Y_{2j} &= \alpha_2 + \beta_2 x_{2j} + \varepsilon_{2j}, \;\; j = 1,2,3,4 \end{aligned}$$

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

(a) As a full model, write the two models in one form $\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ and define $\mathbf{Y}$, X, $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$. Write down a LS $\widehat{\boldsymbol{\beta}}$ and SSE using these notations.

(b) Suppose we want to test the null hypothesis $H_0: \; \alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$ against the alternative $H_1: \; \alpha_1 \neq \alpha_2$ or $\beta_1 \neq \beta_2$. Explain why H_0 is a testable linear hypothesis if all matrices need to be inverted are non-singular.

(c) Write down the reduced model (under H_0) in the form $\mathbf{Y} = Z\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}_0$, and define $\mathbf{Y}$, Z, $\boldsymbol{\beta}_0$ and $\boldsymbol{\varepsilon}_0$. Then write down a LS $\widehat{\boldsymbol{\beta}}_0$ and SSE_0 using these notations.

(d) Write down the test statistic F and specify the criterion for rejecting H_0 at 5% level.

5. (10 points) Let $E_1, \cdots, E_k$ be k events, then we have the Bonferroni inequality:

$$\Pr\left[\bigcap_{i=1}^{k} E_i\right] \geq 1 - \sum_{i=1}^{k} \Pr[E_i^c]].$$

We want to apply this inequality to construct simultaneous confidence intervals for k contrasts of means. Consider independent Y_{ij} for $i = 1,2,3$ and $j = 1, 2, \ldots, n_i$, where $Y_{ij} \sim N(\mu_i, \sigma^2)$. Describe how to construct simultaneous confidence intervals for $\mu_2 - \mu_1$ and $\frac{1}{2}(\mu_2 + \mu_3) - \mu_1$ with confidence at least 0.95. (Note: you need to write down expicit formulae for your notations.)

6. (10 points) Suppose we consider two competing models: $Y = X\beta + \epsilon$ (reduced model) and $Y = X\beta + Z\gamma + \eta$ (full model), where the error vectors (ϵ and η respectively) follow a $N(0, \sigma^2\mathbf{I})$ distribution . Discuss how the estimated residuals from the reduced model can be used (through relevant plots) to assess model inadequacy. That is, what can the relevant residual plots reveal when the reduced model is correct and what can they reveal when the full model is correct. In, particular discuss what happens when the the reduced model is a simple linear regression and the full model is a quadratic regression of Y on X.