Math 6010, Fall 2006
Take-home final exam
December 9, 2006

Turn in hard copy by 4 pm December 11 OR turn in a pdf file by email (use LaTeX or scan) by December 15.

Complete solutions to any four of the five problems. (There is no benefit to doing all five.) Work must be done independently.

1. Consider the following regression model for $n = 20$ pairs $(x_i, Y_i)$:

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \varepsilon_i$$

for $\varepsilon_1, \ldots, \varepsilon_n$ independent $N(0, \sigma^2)$ random variables. The model was fit, giving residual sum of squares $\text{RSS}_3 = 160$. When the cubic term was omitted, we got $\text{RSS}_2 = 180$. When both the quadratic and cubic terms were omitted, we got $\text{RSS}_1 = 200$. Total sum of squares, corrected for the mean, was $\text{RSS}_0 = 1000$.

(a) Give the sample multiple correlation coefficient for the cubic model above.

(b) For $\alpha = 0.05$ test $H_0$: true model is the simple linear regression $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$.

2. Let $x_0, x_1, \ldots, x_n$ be positive constants. Let $Y_i = \beta x_i + \varepsilon_i$ for $i = 0, \ldots, n$ for $\varepsilon_0, \ldots, \varepsilon_n$ independent $N(0, \sigma^2)$. The pairs $(x_i, Y_i)$ have been observed for $i = 1, \ldots, n$ and the value of $Y_0$ is to be predicted.

(a) Give a formula for a $100(1-\alpha)\%$ prediction interval for $Y_0$.

(b) Apply the formula for $n = 4$, $\alpha = 0.05$, and $(x_i, Y_i)$ pairs $(1, 2)$, $(2, 7)$, $(3, 10)$, $(4, 11)$, $x_0 = 5$. Repeat for the same pairs but with $x_0 = 10$.

(c) Repeat (a) and (b) for confidence intervals for $g(x_0) = \beta x_0$, rather than prediction intervals.

3. For pairs $(X_i, Y_i)$: $(0, 7), (1, 7), (2, 5), (3, 1)$ and the simple linear regression model sketch the 95% confidence ellipsoid for $\boldsymbol{\beta} = (\beta_0, \beta_1)'$. Suppose that you wished to test $H_0$: $\boldsymbol{\beta} = (7, -1)'$. Would you reject $H_0$ at level $\alpha = 0.05$?

4. (Three-way ANOVA) Consider the following $2 \times 2 \times 2$ array of means $\mu_{ijk}$: $\mu_{111} = 35$, $\mu_{112} = 25$, $\mu_{121} = 13$, $\mu_{122} = 7$, $\mu_{211} = 21$, $\mu_{212} = 11$, $\mu_{221} = 7$, and $\mu_{222} = 1$.

(a) Find the parameters $\mu$, $\alpha_i$, $\beta_j$, $\gamma_k$, $(\alpha\beta)_{ij}$, $\ldots$, $(\alpha\beta\gamma)_{ijk}$.

(b) Samples of size two were taken from each of the eight normal distributions with means $\mu_{ijk}$ as above and variances $\sigma^2 = 16$, then rounded to the nearest integer. The $Y_{ijkl}$ were: $Y_{1111} = 40$, $Y_{1112} = 31$, $Y_{1121} = 33$, $Y_{1122} = 21$, $Y_{1211} = 18$, $Y_{1212} = 21$, $Y_{1221} = 12$, $Y_{1222} = 14$, $Y_{2111} = 17$, $Y_{2112} = 5$, $Y_{2121} = 11$, $Y_{2122} = 6$, $Y_{2211} = 2$, $Y_{2212} = 10$, $Y_{2221} = 3$, $Y_{2222} = 1$. Estimate the parameters and fill out the ANOVA table.

(c) Since the eight subspaces $V_0, V_A, V_B, V_C, V_{AB}, V_{AC}, V_{BC}, V_{ABC}$ all have dimension 1, the correpsonding sums of squares may all be expressed in the form $\|p(\boldsymbol{Y} \mid \boldsymbol{x})\|^2 = (\boldsymbol{x}'\boldsymbol{Y})^2/\|\boldsymbol{x}\|^2$. Give a vector $\boldsymbol{x}$ for each of these subspaces.

5. The analysis of covariance model we have considered has assumed that the slope $\gamma$ is the same for every cell of the table. In the case of a $2 \times 3$ table with three observation per cell, suppose that $Y_{ijk} = \mu_{ij} + \gamma_{ij} x_{ij} + \varepsilon_{ijk}$.

(a) Describe how you could test the null hypothesis that the $\gamma_{ij}$ are the same for all $i$ and $j$.

(b) Could you carry out this test if there were only two observations per cell?

(c) Consider the model with $\gamma_{ij}$ replaced by $\gamma_i$. How could you test the null hypothesis that $\gamma_1 = \gamma_2$? Is there a corresponding $t$-test?