STAT 510 Homework 6

Due Date: 11:00 A.M., Wednesday, March 6

1. Consider the plant density example discussed in slide set 6.

   (a) For each of the tests in the ANOVA table on slide 38, provide a vector $\boldsymbol{c}$ so that a test of $H_0 : \boldsymbol{c}'\boldsymbol{\beta} = 0$ would yield the same statistic and p-value as the ANOVA test. (You can use R to help you with the computations like we did on slides 45 and 46 of slide set 6.) Label these vectors $\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3$, and $\boldsymbol{c}_4$ for the linear, quadratic, cubic, and quartic tests, respectively.

   (b) Are $\boldsymbol{c}_1'\boldsymbol{\beta}$, $\boldsymbol{c}_2'\boldsymbol{\beta}$, $\boldsymbol{c}_3'\boldsymbol{\beta}$, and $\boldsymbol{c}_4'\boldsymbol{\beta}$ contrasts? Explain.

   (c) Are $\boldsymbol{c}_1'\boldsymbol{\beta}$, $\boldsymbol{c}_2'\boldsymbol{\beta}$, $\boldsymbol{c}_3'\boldsymbol{\beta}$, and $\boldsymbol{c}_4'\boldsymbol{\beta}$ orthogonal? Explain.

2. Suppose $\boldsymbol{H}$ is a symmetric matrix. Prove that $\boldsymbol{H}$ is positive definite if and only if all its eigenvalues are positive. (If you wish, you may use the Spectral Decomposition Theorem in your proof.)

3. Consider the model
$$y_i = \mu + x_i \epsilon_i,$$
where for $i = 1, \ldots, n$, y_i is the response for observation i, μ is an unknown real-valued parameter, x_i is the ith known nonzero observation of an explanatory variable, $\epsilon_1, \ldots, \epsilon_n$ are independent and identically distributed as $N(0, \sigma^2)$, and $\sigma^2 > 0$ is an unknown variance component. Provide an expression for the best linear unbiased estimator of μ. Simplify your answer as much as possible.

4. Suppose
$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu \\ \mu \end{bmatrix}, \begin{bmatrix} 1/4 & 0 \\ 0 & 1 \end{bmatrix} \right).$$

   (a) A linear unbiased estimator of μ has the form $\boldsymbol{a}'\boldsymbol{y} = a_1 y_1 + a_2 y_2$ for some real-valued constants a_1 and a_2. What has to be true about a_1 and a_2 in order for $\boldsymbol{a}'\boldsymbol{y} = a_1 y_1 + a_2 y_2$ to be unbiased?

   (b) Write a simplified expression for the variance of $\boldsymbol{a}'\boldsymbol{y} = a_1 y_1 + a_2 y_2$ in terms of a_1 and a_2.

   (c) Use the result of part (a) to write the answer to part (b) in terms of a single variable.

   (d) Use parts (a) through (c) and a calculus-based argument to derive the BLUE of μ.

   (e) Use the result on slide 12 of slide set 10 to determine the BLUE of μ.

5. Consider the Aitken Model with normal errors described on slide 18 of slide set 10. As usual, assume that $\boldsymbol{y}$ is $n \times 1$ and that the rank of $\boldsymbol{X}$ is r. Suppose $\boldsymbol{c}'\boldsymbol{\beta}$ is estimable. Give an expression for a 95% confidence interval for $\boldsymbol{c}'\boldsymbol{\beta}$ in terms of $\boldsymbol{y}$, $\boldsymbol{X}$, $\boldsymbol{V}$, $\boldsymbol{c}$, n, and r. [Hint: Write the expression for the confidence interval in terms of the transformed model that involves $\boldsymbol{z}$ and $\boldsymbol{W}$, and then make substitutions.]

6. An experiment was conducted to study the effect of a fertilizer on corn yield. Four fertilizer amounts (0, 2, 4, and 10 pounds per acre) were assigned to 20 plots of land using a balanced and completely randomized design with 5 plots per fertilizer amount. Let $x_1 = 0$, $x_2 = 2$, $x_3 = 4$, and $x_4 = 10$. Let y_{ij} be the yield for the jth plot that received x_i pounds of fertilizer per acre ($i = 1, 2, 3, 4$, $j = 1, \ldots, 5$). The researchers fit to the data the following model:
$$y_{ij} = \beta_1 + \beta_2(x_i - \bar{x}_\cdot) + \epsilon_{ij}, \text{ where } \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2), \tag{1}$$

$\bar{x}. = 4$ (the average fertilizer amount), and $\beta_1, \beta_2 \in I\!R$ and $\sigma^2 \in I\!R^+$ are unknown parameters.

Suppose that, unknown to the researchers, the true model for the data is actually

$$y_{ij} \sim N(\mu_i, 36), \text{ where } \mu_1 = 160, \mu_2 = 180, \mu_3 = 200, \mu_4 = 252, \tag{2}$$

and all yields are independent. Note that this model (2) is a special case of the cell-means model discussed in class. The solutions to Exam 1 show that the estimator of $\theta \equiv E(y_{3j})$ obtained from the fit of Model (1) has lower mean squared error than the estimator of θ obtained from the fit of a cell-means model. With all other values of the parameters held fixed at the values specified in (2), for what values of μ_4 does the estimator of θ obtained from the fit of Model (1) have lower mean squared error than the estimator of θ obtained from the fit of a cell-means model?