# STAT 510 Homework 6

**Due Date:** 11:00 A.M., Wednesday, February 28

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 non-negative definite if and only if all its eigenvalues are non-negative. (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 $i$, $\mu$ is an unknown real-valued parameter, $x_i$ is the $i$th 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 $\mu$. Simplify your answer as much as possible.

4. Consider the Gauss-Markov model with normal errors $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I})$. For any nonsingular $p \times p$ matrix $\boldsymbol{B}$, the model can be reparameterized by

$$\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}\boldsymbol{B}^{-1}\boldsymbol{B}\boldsymbol{\beta} = \boldsymbol{W}\boldsymbol{\alpha}, \text{ where } \boldsymbol{W} = \boldsymbol{X}\boldsymbol{B}^{-1} \text{ and } \boldsymbol{\alpha} = \boldsymbol{B}\boldsymbol{\beta}.$$

   From a previous homework problem, we know the column spaces of $\boldsymbol{X}$ and $\boldsymbol{W}$ are identical so that $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ and $\boldsymbol{y} = \boldsymbol{W}\boldsymbol{\alpha} + \boldsymbol{\epsilon}$ are the same models. Suppose $\boldsymbol{C}$ is a $q \times p$ matrix of rank $q < p$. Then there exists a $(p - q) \times p$ matrix $\boldsymbol{A}$ such that

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{C} \end{bmatrix} \text{ has rank } p \text{ and is therefore nonsingular.}$$

   Then we can write

$$\boldsymbol{B}\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{C} \end{bmatrix}\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{A}\boldsymbol{\beta} \\ \boldsymbol{C}\boldsymbol{\beta} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \end{bmatrix} = \boldsymbol{\alpha},$$

   where $\boldsymbol{\alpha}_1 = \boldsymbol{A}\boldsymbol{\beta}$ and $\boldsymbol{\alpha}_2 = \boldsymbol{C}\boldsymbol{\beta}$. If we let $\boldsymbol{W}_1$ be the matrix consisting of the first $p - q$ columns of $\boldsymbol{W} = \boldsymbol{X}\boldsymbol{B}^{-1}$ and $\boldsymbol{W}_2$ be the matrix consisting of the last $q$ columns of $\boldsymbol{W} = \boldsymbol{X}\boldsymbol{B}^{-1}$, then

$$\boldsymbol{W}\boldsymbol{\alpha} = [\boldsymbol{W}_1, \boldsymbol{W}_2]\begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \end{bmatrix} = \boldsymbol{W}_1\boldsymbol{\alpha}_1 + \boldsymbol{W}_2\boldsymbol{\alpha}_2.$$

   Now consider testing

$$H_0 : \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{0} \text{ vs. } H_A : \boldsymbol{C}\boldsymbol{\beta} \neq \boldsymbol{0}.$$

   (a) Rewrite these hypotheses in terms of the $\boldsymbol{\alpha}$ parameter vector.

   (b) If you wanted to fit a reduced model corresponding to the null hypothesis, what model matrix would you use?

(c) Consider the unbalanced experiment described in slide set 8. Assume the full model given on slide 6. Provide a matrix $\boldsymbol{C}$ for testing the main effect of time.

(d) Provide a matrix $\boldsymbol{A}$ so that

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{C} \end{bmatrix}$$

is a $4 \times 4$ matrix of rank 4.

(e) Provide a model matrix for a reduced model that corresponds to the null hypothesis of no time main effect.

(f) Find the error sum of squares for the reduced and full models.

(g) Find the degrees of freedom associated with the sums of squares in part (f).

(h) Compute the $F$-statistic for testing the null hypothesis of no time main effect using the sums of squares and degrees of freedom computed in parts (f) and (g).

5. 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 $\mu$ has the form $\boldsymbol{a}'\boldsymbol{y} = a_1y_1 + a_2y_2$ for some real-valued constants $a_1$ and $a_2$. What has to be true about $a_1$ and $a_2$ in order for $\boldsymbol{a}'\boldsymbol{y} = a_1y_1 + a_2y_2$ to be unbiased?

(b) Write a simplified expression for the variance of $\boldsymbol{a}'\boldsymbol{y} = a_1y_1 + a_2y_2$ 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 $\mu$.

(e) Use the result on slide 12 of slide set 10 to derive the BLUE of $\mu$.