Statistics 550 Spring 2019 Exam 1 Take-home questions

1) For a one-way ANOVA model with two groups,

$$y_i = \mu + \gamma D_i + \varepsilon_i,\ i = 1, 2, ..., n,$$

where D_i is the dummy coded variable for group 1 and the sample sizes are n_1 and n_2, write out the $\mathbf{X}$ matrix and then calculate the least-squares estimator $\widehat{\beta}$ and the covariance matrix $V(\widehat{\beta})$. Write a sentence for each of the two estimators that gives an interpretation for them.

2) Data was collected on student GPAs (y_i) at a university from students in each of four different years (freshman, sophomore, junior, senior). Student high school GPAs were also collected (x_{1i}). An ANCOVA model is proposed for analyzing these data, where x_{2i}, x_{3i},and x_{4i} are dummy variables for the freshman, sophomore, and junior years, and x_{5i}, x_{6i},and x_{7i} are products of x_{1i} with x_{2i}, x_{3i},and x_{4i}, respectively:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i} + \beta_7 x_{7i} + \varepsilon_i$$

i) Write down a matrix $\mathbf{L}$ and a vector $\mathbf{c}$ so that the ANCOVA test of equal slopes can be written in the form $H_0 : \mathbf{L}\beta = \mathbf{c}$.

ii) Suppose that we want to test the hypothesis that the slope is equal for freshmen and sophomores. Write down a matrix $\mathbf{L}$ and a vector $\mathbf{c}$ so that this hypothesis can be written in the form $H_0 : \mathbf{L}\beta = \mathbf{c}$.

iii) Assuming that the slopes are equal, write down a matrix $\mathbf{L}$ and a vector $\mathbf{c}$ so that the ANCOVA test of equal group means can be written in the form $H_0 : \mathbf{L}\beta = \mathbf{c}$.

3) Some matrix problems. Suppose that $\boldsymbol{y} = (y_1, y_2)'$ has a multivariate normal distribution, so $\boldsymbol{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = (2, 3)'$ and

$$\boldsymbol{\Sigma} = \begin{bmatrix} 4 & 2 \\ 2 & 5 \end{bmatrix}.$$

Let $z_1 = 2y_1 + 3y_2$, and $z_2 = 2y_2 - y_1$.

i) Use vector and matrix calculations (from our matrix results in lecture 5) to find the means and variances of z_1 and z_2.

Let the matrix $\mathbf{X}$ be of size $n\ x\ (k + 1)$ and let $\mathbf{K} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

ii) Show that $\mathbf{K}$ is idempotent. (show that $\mathbf{KK} = \mathbf{K}$)

iii) Show that $\mathbf{KX} = \mathbf{0}$.

iv) Is $-\mathbf{K}$ idempotent? Show why or why not.

4) Consider the two-factor ANOVA data available in a separate file.

i) By fitting a series of regression models (using deviation coding), obtain sums of squares for all tests of interest (A, B, and the AB interaction) using all three Types of sums of squares (I, II, and III). For the Type I sums of squares fit in the order of A, B, AB. Do any of the three types of sums of squares satisfy

$$Total\ SS = SSA + SSB + SSAB + RSS?$$

ii) Do the results of the hypothesis tests differ based on the different sums of squares used? In general, which sum of squares type do you believe is best to use? Briefly explain your choice.