Statistics 550 Spring 2017 Exam 1 Take-home questions

1) Suppose you have a sample of n observations on a variable Y and another variable x, and are considering the model:

$$Y_i = \beta_0 + \beta_1 x_i^3 + \varepsilon_i$$

i) Use the general matrix-based formula for the least-squares estimator to give scalar expressions for the least-squares estimators of β_0 and β_1.

ii) Give the scalar expressions for the standard errors of the estimators of β_0 and β_1.

2) Data was collected on sales (Y_i) from different companies that are from one of four different sectors of the economy (the Hi Tech sector, the Energy sector, the Finance sector, or the Retail sector). The value of the assets for each company (x_{1i}) was also recorded. An ANCOVA model is proposed for analyzing these data, where x_{2i}, x_{3i},and x_{4i} are dummy variables for the Hi Tech, Energy, and Finance sectors, respectively. Also, 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) 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) In lecture 5, we encountered the matrix $\mathbf{M}_1 = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ along the way to finding the expected value of the residual sum of squares.

i) Show that $\mathbf{M}_1$ is idempotent, in other words that $\mathbf{M}_1\mathbf{M}_1 = \mathbf{M}_1$.

ii) Show that $\mathbf{M}_2 = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is also idempotent.

iii) Is the matrix $\mathbf{A}$ below idempotent? Show your work.

$$\mathbf{A} = \frac{1}{29}\begin{bmatrix} 4 & 10 \\ 10 & 25 \end{bmatrix}$$

4) Consider the two-factor ANOVA concrete data available in a separate file (adapted from Kuehl's nice Design of Experiments text). The first factor, called Aggregate has levels of "Basalt" and "Silicious", while the second

factor, called Compaction, has levels of "Static", "Regular", "Low", and "VeryLow". The response is the strength of a sample of concrete produced with those factor levels.

i) By fitting a series of regression models (using deviation coding), obtain sums of squares for all tests of interest (Aggregate and Compaction main effects, and the two-way interaction) using all three Types of sums of squares (I, II, and III). For the Type I SS, fit in the order of Aggregate, Compaction, and Aggregate x Compaction. (You can check your sums of squares using factors with SAS or R to be sure they are correct)

ii) Do any of the three types of sums of squares satisfy:

$$Total\ SS = SS\text{Aggregate} + SS\text{Compaction} + SS\text{Aggregate x Compaction} + RSS?$$

iii) For this set of data, which Type of SS should be used?