# BTRY/STSCI 4030 - Linear Models with Matrices - Fall 2017
# Midterm - Wednesday, October 12

**NAME**:

**NETID**:

**Instructions**:

It is not necessary to complete numerical calculations (using a calculator) if you clearly show how the answer can be obtained, and if the exact answer is not required in subsequent parts. For example, if you are asked to calculate an F-statistic, then an answer in the form

$$F = \frac{45.2/6}{22.9/15}$$

would be acceptable.

A set of formulae and notes is provided with the exam; other outside material is not allowed. You may directly use any result on the notes without proving it.

**Problem 1: Sequential ANOVA**

In class we have seen that the whole-model ANOVA table can be used to test the hypothesis $H_0 : \beta_1 = \ldots = \beta_p = 0$ in Multiple Linear Regression. When this is rejected we may be interested in testing the individual coefficients.

Here we will develop the test that $\beta_j = 0$ using the sequential ANOVA table (see sheet of notes).

First we will establish that some of our quantities aren't affected by "true" parameters of the model:

1. Recall that the $j$th term in the sequential ANOVA is $D_j = H_j C H_j - H_{j-1} C H_{j-1}$. Show that $D_j X_{j-1} = 0$.

2. By comparing the sum of squared errors calculated from $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{e}$ to the same quantity calculated from $\boldsymbol{y}^* = X\boldsymbol{\alpha} + \boldsymbol{e}$ (keeping $\boldsymbol{e}$ the same), show that SSE does not change with the coefficient in the model.

3. Similarly, by changing just the first $j-1$ coefficients (ie using $\boldsymbol{y}^* = X\boldsymbol{\beta} + X_{j-1}\boldsymbol{\alpha} + \boldsymbol{e}$) show that the sum of squares for $x_j | X_{j-1}$ does not change with $\beta_0, \beta_1, \ldots, \beta_{j-1}$.

Now let's form a test statistic and show that it works

4. Show that $HH_j = H_j$.

5. Under the hypothesis $H_0 : (\beta_j, \ldots, \beta_p) = \boldsymbol{0}$, what is the distribution of $\boldsymbol{y}^T D_j \boldsymbol{y}$? What is its expectation?

6. Show that SSE and $\boldsymbol{y}^T D_j \boldsymbol{y}$ are independent.

7. Obtain and $F$ statistic to test the hypothesis that $(\beta_j, \ldots, \beta_p) = \boldsymbol{0}$ and give its distribution.

The next few questions will show that this hypothesis is stronger than we need.

8. Assuming that $X^T X$ is diagonal (all covariates are orthogonal and centered), show that $\boldsymbol{y}^T H_j C H_j \boldsymbol{y} = \sum_{k=1}^{j} (\beta_k^2 \boldsymbol{x}_k^T \boldsymbol{x}_k + 2\beta_k \boldsymbol{x}_k^T \boldsymbol{e}) + \boldsymbol{e}^T H_j C H_j \boldsymbol{e}$.

9. Hence show that $\boldsymbol{y}^T D_j \boldsymbol{y} = \beta_j^2 \boldsymbol{x}_j^T \boldsymbol{x}_j + 2\beta_j \boldsymbol{x}_j^T \boldsymbol{e} + \boldsymbol{e}^T D_j \boldsymbol{e}$ and that this sum of squares is unaffected by the values of $(\beta_{j+1}, \ldots, \beta_p)$.

10. Why does this mean that the $F$ statistic you derived early can be used to test $H_0 : \beta_j = 0$?

And a few extensions; do not assume $X^T X$ is diagonal.

11. Give a reason to use MSE for the full model in the $F$ statistic instead of $\boldsymbol{y}^T(I - H_j)\boldsymbol{y}$, the MSE for the model based on $X_j$.

12. A researcher observes that under $H_0 : (\beta_{j+1}, \ldots, \beta_p) = \mathbf{0}$, both $\boldsymbol{y}^T D_j \boldsymbol{y}$ and $\boldsymbol{y}^T D_{j+1} \boldsymbol{y}$ have the same expectation. They therefore suggest an alternative test based on their ratio $(\boldsymbol{y}^T D_j \boldsymbol{y})/(\boldsymbol{y}^T D_{j+1} \boldsymbol{y})$. Give two reasons this would be a bad idea.

bonus How can you interpret the test if $X^T X$ is not diagonal?