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

NAME:

NETID:

Instructions:

This exam has two questions with a total of 12 parts. 5 points will be awarded for each part, totaling 60 points.

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.

You may reference any result in the formulae by it's number; e.g. the Eigen-decomposition for a symmetric matrix is in 5.2a.

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

   First we note that since $H_j C H_j = H_j - \bar{J}$ (Formula sheet 5.2e) we have

   $$D_j = (H_j - \bar{J}) - (H_{j-1} - \bar{J}) = H_j - H_{j-1}$$

   Note: the Formula sheet now uses $D_j = H_j - H_{j-1}$ and the exam should have started from this.

   Since $H_j X_j = X_j$ (Formulae 5.2e) does not change the columns of X_j, $H_j X_{j-1} = X_{j-1}$ and

   $$D_j X_{j-1} = H_j X_{j-1} H_{j-1} X_{j-1} = X_{j-1} - X_{j-1} = \mathbf{0}.$$

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.

   We first observe that $(I - H)X = X - X = 0$ so that for

   $$\begin{aligned} SSE(\boldsymbol{y}) &= \boldsymbol{y}^T (I - H)\boldsymbol{y} \\ &= (X\boldsymbol{\beta} + \boldsymbol{e})^T (I - H)(X\boldsymbol{\beta} + \boldsymbol{e}) \\ &= (X\boldsymbol{\beta} + \boldsymbol{e})^T (I - H)\boldsymbol{e} \\ &= \boldsymbol{e}^T (I - H)\boldsymbol{e} \end{aligned}$$

   and by the same calculation $SSE(\boldsymbol{y}) = SSE(\boldsymbol{y}^*)$.

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 $\beta_0, \beta_1, \ldots, \beta_{j-1}$.

   Here we note that $\boldsymbol{y}^*$ is equivalent to use $\beta_k^* = \beta_k + \alpha_k$ for $k = 1, \ldots, j-1$, so if $\boldsymbol{y}^T D_j \boldsymbol{y} = \boldsymbol{y}^{*T} D_j \boldsymbol{y}^*$ this sum of squares doesn't change if we change $\beta_0, \beta_1, \ldots, \beta_{j-1}$. [Possibly this could have been phrased better.]

$$\begin{aligned}
\boldsymbol{y}^{*T} D_j \boldsymbol{y}^* &= (X\boldsymbol{\beta} + X_{j-1}\boldsymbol{\alpha} + \boldsymbol{e})^T D_j (X\boldsymbol{\beta} + X_{j-1}\boldsymbol{\alpha} + \boldsymbol{e}) \\
&= (X_{j-1}\boldsymbol{\alpha} + \boldsymbol{y})^T D_j (X_{j-1}\boldsymbol{\alpha} + \boldsymbol{y}) \\
&= \boldsymbol{y}^T D_j \boldsymbol{y} + \boldsymbol{\alpha}^T X_{j-1}^T D_j \boldsymbol{y} + \boldsymbol{y}^T D_j X_{j-1} \boldsymbol{\alpha} \\
&= \boldsymbol{y}^T D_j \boldsymbol{y}
\end{aligned}$$

   Since $D_j X_{j-1} = 0$ from Part 1.

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

4. Show that $HH_j = H_j$.

   First note that $HX_j = X_j$ then by the definition of H_j

$$HH_j = HX_j(X_j^T X_j)^{-1} X_j^T = X_j(X_j^T X_j)^{-1} X_j^T = H_j$$

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

   If $H_0 : (\beta_j, \ldots, \beta_p) = \mathbf{0}$, then we can write $X\beta = X_{j-1}\boldsymbol{\beta}_{j-1}$ for $\boldsymbol{\beta}_{j-1} = (\beta_0, \ldots, \beta_{j-1})$ and

$$\boldsymbol{y}^T D_j \boldsymbol{y} = (X_{j-1}\boldsymbol{\beta}_{j-1} + \boldsymbol{e})^T D_j (X_{j-1}\boldsymbol{\beta}_{j-1} + \boldsymbol{e}) = \boldsymbol{e}^T D_j \boldsymbol{e}.$$

   Taking the eigen-decomposition $D_j = U^T \tilde{D} U$ (Formulae 5.2a) and observing that $\boldsymbol{u} = U\boldsymbol{e} \sim N(\mathbf{0}, \sigma^2 I$ (Formulae 6.1a) and observing that $\tilde{D}$ has one non-zero diagonal element (Formulae 5.2c)

$$\boldsymbol{e}^T D_j \boldsymbol{e} = \boldsymbol{u}^T \tilde{D} \boldsymbol{u} = u_1^2 \sim \sigma^2 \chi_1^2$$

   with expectation σ^2.

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

   Writing $SSE = \boldsymbol{y}(I-H)^T(I-H)\boldsymbol{y} = \hat{\boldsymbol{e}}^t\hat{\boldsymbol{e}}$ *and* $\boldsymbol{y}^T D_j \boldsymbol{y} = \boldsymbol{y}^T D_j^T D_j \boldsymbol{y} = \boldsymbol{z}^T\boldsymbol{z}$ *by Formulae 6.1d it is enough to show*

   $$cov(\hat{\boldsymbol{e}}, \boldsymbol{z}) = \sigma^2(I-H)D_j = \sigma^2(I-H)(H_j - H_{j-1}) = 0$$

   since $HH_j = H_j$ *from Part 4 so* $(I-H_j) = 0$ *and similarly for* $(I-H)H_{j-1}$.

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

   $H_0 : (\beta_j, \ldots, \beta_p) = \mathbf{0}$ *we have*

   $$\frac{\boldsymbol{y}^T D_j \boldsymbol{y}}{MSE} = \frac{\sigma^2 X_1}{\sigma^2 X_2/(n-p-1)} = \frac{X_1}{X_2/(n-p-1)} \sim F^1_{n-p-1}$$

   where $X_1 = \boldsymbol{y}^T D_j \boldsymbol{y}/\sigma^2 \sim \chi^2_1$ *from Part 5 and* $X_2 = SSE/\sigma^2 \sim \chi^2_{n-p-1}$ *from Formulae 6.2a and these are independent by Part 6.*

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 $\boldsymbol{y}^T H_j C H_j \boldsymbol{y} = \sum_{k=1}^{j}(\beta_k^2 \boldsymbol{x}_k^T \boldsymbol{x}_k + 2\beta_k \boldsymbol{x}_k^T \boldsymbol{e}) + \boldsymbol{e}^T H_j C H_j \boldsymbol{e}$.

   For this question, we will write out $X\boldsymbol{\beta} = \sum_{j=1}^{p} \beta_j \boldsymbol{x}_j$, *we will also observe that since* $X_j \boldsymbol{x}_k = 0$ *if* $k > j$ *(the* $\boldsymbol{x}_k$ *are orthogonal to each other) then* $H\boldsymbol{x}_k = \boldsymbol{x}_k$ *if* $k \leq j$ *and* $\mathbf{0}$ *otherwise.*

   Also note that since the $\boldsymbol{x}_k$ *are centered,* $C\boldsymbol{x}_k = \boldsymbol{x}_k$ *and so* $H_j C H_j \boldsymbol{x}_k = \boldsymbol{x}_k$ *if* $k \leq j$ *and* $\mathbf{0}$ *otherwise. Thus using the notation in Part 6*

   $$H_j C H_j \boldsymbol{y} = H_j C H_j (X\boldsymbol{\beta} + \boldsymbol{e}) = X_j \boldsymbol{\beta}_j + H_j C H_j \boldsymbol{e}$$

   and

   $$\begin{aligned}
   \boldsymbol{y}^T H_j C H_j \boldsymbol{y} &= (X\boldsymbol{\beta} + \boldsymbol{e})^T (X_j \boldsymbol{\beta}_j + H_j C H_j \boldsymbol{e}) \\
   &= \boldsymbol{\beta}^T X^T X_j \boldsymbol{\beta}_j + \boldsymbol{e}^T X_j \boldsymbol{\beta}_j + \boldsymbol{\beta}^T X H_j C H_j \boldsymbol{e} + \boldsymbol{e}^T H_j C H_j \boldsymbol{e} \\
   &= \boldsymbol{\beta}_j^T X_j^T X_j \boldsymbol{\beta}_j + \boldsymbol{e}^T X_j \boldsymbol{\beta}_j + \boldsymbol{\beta}_j^T X_j^T \boldsymbol{e} + \boldsymbol{e}^T H_j C H_j \boldsymbol{e}
   \end{aligned}$$

because the columns of X *are orthogonal and* $X_j^T X = X_j^t [X_j\ \bar{X}_j] = [X_j^T X_j\ 0]$ *if* $\bar{X}_j = [\boldsymbol{x}_{j+1}, \ldots, \boldsymbol{x}_p]$. *Writing this equation out as summations gives us*

$$\begin{aligned}\boldsymbol{\beta}_j^T X_j^T X_j \boldsymbol{\beta}_j + 2\boldsymbol{e}^T X_j \boldsymbol{\beta}_j + \boldsymbol{e}^T H_j C H_j \boldsymbol{e} &= \sum_{k=1}^{j} (\sum_{l=1}^{j} \beta_k \beta_l \boldsymbol{x}_k^T \boldsymbol{x}_l + 2\boldsymbol{x}_k^T \boldsymbol{e}) + \boldsymbol{e}^T H_j C H_j \boldsymbol{e} \\ &= \sum_{k=1}^{j} (\beta_k^2 \boldsymbol{x}_k^T \boldsymbol{x}_k + 2\boldsymbol{x}_k^T \boldsymbol{e}) + \boldsymbol{e}^T H_j C H_j \boldsymbol{e}.\end{aligned}$$

Note that if we re-defined the problem as $\boldsymbol{y}^T H_j \boldsymbol{y}$ – appropriate for a simpler definition of D_j – then we would remove a couple of steps and get

$$\boldsymbol{y}^T H_j \boldsymbol{y} = n\beta_0^2 + \sum_{k=1}^{j} (\beta_k^2 \boldsymbol{x}_k^T \boldsymbol{x}_k + 2\boldsymbol{x}_k^t \boldsymbol{e}) + \boldsymbol{e}^T H_j C H_j \boldsymbol{e}.$$

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 $(\beta_{j+1}, \ldots, \beta_p)$.

   By subtracting

$$\begin{aligned}\boldsymbol{y}^T D_j \boldsymbol{y} &= \boldsymbol{y}^T H_j C H_j \boldsymbol{y} - \boldsymbol{y}^T H_{j-1} C H_{j-1} \boldsymbol{y} \\ &= \sum_{k=1}^{j} (\beta_k^2 \boldsymbol{x}_k^T \boldsymbol{x}_k + 2\boldsymbol{x}_k^T \boldsymbol{e}) + \boldsymbol{e}^T H_j C H_j \boldsymbol{e} \\ &\quad - \sum_{k=1}^{j-1} (\beta_k^2 \boldsymbol{x}_k^T \boldsymbol{x}_k + 2\boldsymbol{x}_k^T \boldsymbol{e}) - \boldsymbol{e}^T H_{j-1} C H_{j-1} \boldsymbol{e} \\ &= \beta_j^2 \boldsymbol{x}_j^T \boldsymbol{x}_j^T + \boldsymbol{e}^T (H_j C H_j - H_{j-1} C H_{j-1}) \boldsymbol{e}\end{aligned}$$

   Since this expression only involves β_j *and not* $(\beta_{j-1}, \ldots, \beta_p)$ *this sum of squares only depends on* β_j.

   Note that writing $D_j = H_j - H_{j-1}$ gives us exactly the same answer.

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

    In this case, the distribution of $\boldsymbol{y}^T D_j \boldsymbol{y} \sim \chi_1^2$ *(and hence our* $F \sim F_{n-p-1}^1$ *if* $\beta_j = 0$ *regardless of the values of* $(\beta_{j-1}, \ldots, \beta_p)$, *so this is only a test of the value of* β_j.

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.

    If $\beta_k \neq 0$ for some $k > j$ we have

    $$\boldsymbol{y}^T(I - H_j)\boldsymbol{y} = \boldsymbol{e}^T(I - H_j)\boldsymbol{e} + 2\boldsymbol{e}^T(I - H_j)\boldsymbol{x}_k\beta_k + \beta_k^2 \boldsymbol{x}_k^T(I - H_j)\boldsymbol{x}_k\beta_k$$

    and this is not distributed as χ^2_{n-j-1} – it will tend to be larger than this distribution, making the distribution of F smaller and costing power.

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.

    Here we will give

    - *The denominator degrees of freedom is 1, and the denominator variance is therefore 2, for MSE the variance is $2(n-p-1)/(n-p-1)^2 = 2/(n-p-1) < 1$ because var(SSE) is $2(n-p-1)$ (Formulae 6.2). This means our F statistic is more variable and we have less power.*
    - *If $\beta_{j+1} \neq 0$, and the $\boldsymbol{x}^k$ are not orthogonal, neither $\boldsymbol{y}^T D_j \boldsymbol{y}$ nor $\boldsymbol{y}^T D_{j+1}\boldsymbol{y}$ will necessarily be χ^2_1. Their distribution would depend on the unknown β_{j+2}.*

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

Without pushing through all the mathematical details, let us consider transforming the columns of X to

$$X^* = [X_j,\ (I - H_j)\bar{X}_j]$$

in the notation of Part 7, so that $(I - H_j)\bar{X}_j$ is orthogonal to X_j.

In this case, our F statistic only depends on β_j^* (the coefficient when we regress on X^*).

Loosely, we let X_j explain as much as possible about $\boldsymbol{y}$ and within that we ask how much does $\boldsymbol{x}_j$ add over what is already explained by X_{j-1}.