```
Math 584    HW 7 Spring 2021, due Thursday, March 18.
Exam 2 is Tuesday, March 23. Final: Tuesday, May 4, 2:45-4:45.
```

1) Assume that $Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{p-1} x_{i,p-1} + \epsilon_i$ for $i = 1, ..., n$ where the ϵ_i are iid $N(0, \sigma^2)$. Hence $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$. Assume that the $n \times p$ matrix $\boldsymbol{X}$ has full rank p. The most important test for multiple linear regression (the partial F test) tests whether a subset of the predictors can be used instead of all p predictors. Let $I = \{i_0, i_1, ..., i_{p-q-1}\}$ be the index set for $p - q$ terms in the candidate subset and let $O = \{i_{p-q}, ..., i_{p-1}\}$ be the index set of the q terms left out of the model. Let $\boldsymbol{X}_I = [\boldsymbol{x}_{i0}, ..., \boldsymbol{x}_{i,p-q-1}]$ and let $\boldsymbol{X}_O = [\boldsymbol{x}_{i,p-q}, ..., \boldsymbol{x}_{i,p-1}]$. Define $\boldsymbol{\beta}_I$ and $\boldsymbol{\beta}_O$ similarly. Then the test is $H : \beta_{i,p-q} = \cdots = \beta_{i,p-1} = 0$. Equivalently, $H : \boldsymbol{\beta}_O = \boldsymbol{0}$. The usual ANOVA F table corresponds to $q = p-1$ and $I = \{0\}$ or $H : \beta_1 = \cdots = \beta_{p-1} = 0$. The p–values for individual coefficients correspond to $q = 1$ and $O = \{i\}$ or $H : \beta_i = 0$. The test statistic is given by Theorem 4.1(iv) and Theorem 4.3, but theory using results on quadratic forms is simpler.

Consider testing $\boldsymbol{\beta}_0 = \boldsymbol{0}$. Then $RSS = \boldsymbol{Y}'(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{Y}$ and $RSS_H = \boldsymbol{Y}'(\boldsymbol{I} - \boldsymbol{P}_H)\boldsymbol{Y}$ where $\boldsymbol{P}_H$ is the projection matrix for the regression of $\boldsymbol{Y}$ on $\boldsymbol{X}_I$.

(This notation is following Seber. So $\boldsymbol{X}_I = \boldsymbol{X}_1, \boldsymbol{X}_O = \boldsymbol{X}_2, \boldsymbol{\beta}_O = \boldsymbol{\beta}_2$, and $\boldsymbol{P}_H = \boldsymbol{P}_1$.)

a) Show that $RSS_H - RSS = \boldsymbol{Y}'\boldsymbol{C}\boldsymbol{Y}$.

b) Show that $\boldsymbol{C}$ is idempotent.

c) Find rank($\boldsymbol{C}$). (Hint: use A.6.2: If $\boldsymbol{P}$ is a projection matrix, then rank($\boldsymbol{P}$) = tr($\boldsymbol{P}$). $\boldsymbol{C}$ is a projection matrix since $\boldsymbol{C}$ is symmetric and idempotent.)

d) Show $\boldsymbol{C}\boldsymbol{X}_I = \boldsymbol{0}$.

Theorem: If $\boldsymbol{C}$ is a projection matrix, if $\boldsymbol{C\mu} = \boldsymbol{0}$ and if $\boldsymbol{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{I})$, then $\boldsymbol{Y}'\boldsymbol{C}\boldsymbol{Y} \sim \chi^2(rank(\boldsymbol{C}))$.

e) If $H : \boldsymbol{\beta}_0 = \boldsymbol{0}$ is true, then $\boldsymbol{Y} \sim N_n(\boldsymbol{X}_I\boldsymbol{\beta}_I, \sigma^2\boldsymbol{I})$. Assuming H is true, use the above theorem to find the distribution of $\boldsymbol{Y}'\boldsymbol{C}\boldsymbol{Y}$.

f) Explain why RSS and $RSS_H - RSS$ are independent.

g) Assuming H is true, find the distribution of the partial F statistic

$$F_I = \frac{(RSS_H - RSS)/rank(\boldsymbol{C})}{RSS/rank(\boldsymbol{I} - \boldsymbol{P})}.$$

2) Let $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$ where $E(\boldsymbol{\epsilon}) = \boldsymbol{0}$, Cov($\boldsymbol{\epsilon}$) $= \sigma^2\boldsymbol{I}_n$, and $\boldsymbol{X}$ has full rank. Let $\boldsymbol{a}$ be a constant vector.

a) Find $E(\boldsymbol{a}'\hat{\boldsymbol{\beta}})$.

b) Is $\boldsymbol{a}'\boldsymbol{\beta}$ estimable? Explain briefly.

3) Let $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$ where $\boldsymbol{\epsilon} \sim N_n(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n)$. Assume $\boldsymbol{X}$ has full rank. Let $\boldsymbol{e}$ be the vector of residuals. Then the residual sum of squares RSS $= \boldsymbol{e}'\boldsymbol{e}$. The sum of squared fitted values is $\hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}}$. Prove that $\boldsymbol{e}'\boldsymbol{e}$ and $\hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}}$ independent (or dependent).
(Hint: write each term as a quadratic form.)