Statistics 206

Homework 5

NOT Due

1. Tell true or false of the following statements.

   (a) The multiple coefficient of determination R^2 is always larger/not-smaller for models with more X variables.

   False. When the X variables in a smaller model are not entirely included in the larger model, then it is possible that the R^2 of the smaller model is larger than the R^2 of the larger model.

   (b) If all the regression coefficients associated with the X variables are estimated to be zero, then $R^2 = 0$.

   True. The fitted regression surface is horizontal so $\widehat{Y}_i = \overline{Y}$ for all i and thus $SSR = 0$, then $R^2 = \frac{SSR}{SSTO} = 0$.

   (c) The adjusted multiple coefficient of determination R_a^2 may decrease when adding additional X variables into the model.

   True. $R_a^2 = 1 - \frac{n-1}{n-p}\frac{SSE}{SSTO}$, the decrease in SSE may be more than offset by the loss of degrees of freedom in the denominator $n - p$.

   (d) Models with larger R^2 is always preferred.

   False. Models with larger R^2 may have a lot of X variables that are unrelated to the response variable or are highly correlated with each other which just over-fit the data and make prediction and interpretation difficult.

   (e) If the response vector is a linear combination of the columns of the design matrix $\mathbf{X}$, then the coefficient of multiple determination $R^2 = 1$.

   TRUE. Since now $Y_i = \widehat{Y}_i$ for all $i = 1, \cdots, n$. Thus all residuals $e_i \equiv 0$ and then $SSE = 0$.

2. **Multiple linear regression by matrix algebra in R**. Consider the following data set with 5 cases, one response variable Y and two predictor variables X_1, X_2.

```
case     Y     X1       X2
1       -0.97  -0.63  -0.82
2       2.51  0.18   0.49
3     -0.19  -0.84  0.74
4       6.53  1.60   0.58
5       1.00  0.33  -0.31
```

   Consider the first-order model for the following questions.

(a) Write down the model equations and the coefficient vector $\boldsymbol{\beta}$. Write down the design matrix and the response vector.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i, \quad i = 1, \cdots, 5$$

$$\boldsymbol{\beta} = [\beta_0, \beta_1, \beta_2]'$$

$$\mathbf{Y} = [-0.97, 2.51, -0.19, 6.53, 1.00]'$$

$$\mathbf{X} = \begin{bmatrix} 1 & -0.63 & -0.82 \\ 1 & 0.18 & 0.49 \\ 1 & -0.84 & 0.74 \\ 1 & 1.60 & 0.58 \\ 1 & 0.33 & -0.31 \end{bmatrix}$$

(b) In R, create the design matrix $\mathbf{X}$ and the response vector $\mathbf{Y}$. Calculate $\mathbf{X}'\mathbf{X}$, $\mathbf{X}'\mathbf{Y}$ and $(\mathbf{X}'\mathbf{X})^{-1}$. Copy your results here.

```
> t(X)%*%X
             X1     X2
   5.00 0.6400 0.6800
X1 0.64 3.8038 0.8089
X2 0.68 0.8089 1.8926
> t(X)%*%Y
      [,1]
    8.8800
X1 12.0005
X2  5.3621
> solve(t(X)%*%X)
                          X1          X2
    0.21184719 -0.02140278 -0.06696786
X1 -0.02140278  0.29134054 -0.11682948
X2 -0.06696786 -0.11682948  0.60236791
```

(c) Obtain the least-squares estimators $\hat{\boldsymbol{\beta}}$. Copy your results here.

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = [1.265271, 2.679724, 1.233270]'$$

(d) Obtain the hat matrix $\mathbf{H}$ and copy it here. What are $rank(\mathbf{H})$ and $rank(\mathbf{I} - \mathbf{H})$? (Hint: You may use *rankMatrix()* in library *Matrix*)

```
> H = X%*%solve(t(X)%*%X)%*%t(X) # hat matrix
> H
            [,1]       [,2]        [,3]       [,4]        [,5]
[1,]  0.74859901 0.02181768  0.01132102 -0.1770289  0.39529119
[2,]  0.02181768 0.27197293  0.35049579  0.2534024  0.10231125
[3,]  0.01132102 0.35049579  0.82936038 -0.1072487 -0.08392853
[4,] -0.17702890 0.25340235 -0.10724866  0.7973084  0.23356681
[5,]  0.39529119 0.10231125 -0.08392853  0.2335668  0.35275928
```

$$rank(\mathbf{H}) = n - p = 5 - 2 = 3, rank(\mathbf{I} - \mathbf{H}) = p = 2$$

(e) Calculate the trace of $\mathbf{H}$ and compare it with $rank(\mathbf{H})$ from part (d). What do you find?

```
>sum(diag(H))
[1] 3
```

Thus $Tr(H) = 3$ which is equal to the $rank(\mathbf{H})$.

(f) Obtain the fitted values, the residuals, SSE and MSE. What should be the degrees of freedom of SSE? Copy your results here. You may use the following codes (with suitable modification) for SS:

```
> sum((Y-mean(Y))^2)
> sum((Y-Yhat)^2)
> sum((Yhat-mean(Y))^2)
```

```
> sum((Y-mean(Y))^2) # for SSTO
[1] 35.14712
> sum((Y-Yhat)^2)# for SSE
[1] 0.91145
> sum((Yhat-mean(Y))^2)# for SSR
[1] 34.23567
> Yhat
            [,1]
[1,] -1.43423719
[2,]  2.35192330
[3,] -0.07307774
[4,]  6.26812586
[5,]  1.76726576
> residuals
           [,1]
[1,]  0.4642372
[2,]  0.1580767
[3,] -0.1169223
[4,]  0.2618741
[5,] -0.7672658
> SSE
        [,1]
[1,] 0.91145
> MSE
         [,1]
[1,] 0.455725
```

Consider the nonadditive model with interaction between X_1 and X_2 for the following questions.

(f) Write down the model equations and the coefficient vector $\boldsymbol{\beta}$.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \epsilon_i, \quad i = 1, \cdots, 5$$

$$\boldsymbol{\beta} = [\beta_0, \beta_1, \beta_2, \beta_3]'$$

(g) Specify the design matrix and the response vector. Obtain the hat matrix $\mathbf{H}$. Find $rank(\mathbf{H})$ and $rank(\mathbf{I} - \mathbf{H})$. Compare the ranks with those from part (d), what do you observe?

$$\mathbf{Y} = [-0.97, 2.51, -0.19, 6.53, 1.00]'$$

$$\mathbf{X} = \begin{bmatrix} 1 & -0.63 & -0.82 & 0.5166 \\ 1 & 0.18 & 0.49 & 0.0882 \\ 1 & -0.84 & 0.74 & -0.6216 \\ 1 & 1.60 & 0.58 & 0.9280 \\ 1 & 0.33 & -0.31 & -0.1023 \end{bmatrix}$$

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

$rank(\mathbf{H}) = 4, rank(\mathbf{I} - \mathbf{H}) = 1$. $rank(\mathbf{H})$ is one more and $rank(\mathbf{I} - \mathbf{H})$ is one less, compared to part(d).

(h) Obtain the least-squares estimators $\hat{\boldsymbol{\beta}}$. Copy your results here.

$$\hat{\boldsymbol{\beta}} = [1.051738, 1.987286, 1.804233, 1.387774]'$$

(i) Obtain the fitted values, the residuals, SSE and MSE. What should be the degrees of freedom of SSE? Copy your results here.

$$\widehat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = [-0.963, 2.416, -0.145, 6.566, 1.006]'$$

$$\mathbf{e} = \mathbf{Y} - \widehat{\mathbf{Y}} = [0.007, 0.094, -0.045, 0.036, -0.006]'$$

$$SSE = \mathbf{e}'\mathbf{e} = 0.01223284$$

$$MSE = SSE/df = 0.01223284/1 = 0.01223284$$

(j) Which model appears to fit the data better?
The second model fits the data better since it has a much smaller SSE therefore much larger R^2 ($R^2 = 1 - SSE/SSTO$ and $SSTO$ is the same).

3. Under the general liner regression model, show that the residuals are uncorrelated with the fitted values and the estimated regression coefficients.

Proof.

$$e = (I - H)Y, \quad \hat{\beta} = (X'X)^{-1}X'Y$$

$$\mathrm{Cov}(e, \hat{\beta}) = (I - H)\mathrm{Cov}(Y)((X'X)^{-1}X')' = \sigma^2(I - H)X(X'X)^{-1} = 0$$

since $(I - H)X = X - X = 0$. Therefore $\hat{\beta}$ and the residuals e are uncorrelated.
Also, $\hat{Y} = X\hat{\beta}$.
Hence, $\mathrm{Cov}(\hat{Y}, e) = \mathrm{Cov}(X\hat{\beta}, e) = X\mathrm{Cov}(\hat{\beta}, e) = 0$.
Therefore $\hat{Y}$ and the residuals e are uncorrelated. $\square$

4. Under the multiple regression model (with X variables $X_1, \cdots, X_{p-1}$), show the following.

   (a) The LS estimator of the regression intercept is:

   $$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}_1 - \cdots - \hat{\beta}_{p-1} \overline{X}_{p-1},$$

   where $\hat{\beta}_k$ is the LS estimator of β_k, and $\overline{X}_k = \frac{1}{n} \sum_{i=1}^{n} X_{ik}$ $(k = 1, \cdots, p-1)$.
   (Hint: Plug in $\hat{\beta}_1, \cdots, \hat{\beta}_{p-1}$ to the least squares criterion function $Q(\cdot)$ and solve for b_0 that minimizes that function.)
   By the hint, taking partial derivative of $Q(\cdot)$ w.r.t. b_0 we have

   $$-\sum_{i=1}^{n} (Y_i - b_0 - \hat{\beta}_1 X_{i1} - \cdots - \hat{\beta}_{p-1} X_{i,p-1}) = 0$$

   $$\Rightarrow \hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}_1 - \cdots - \hat{\beta}_{p-1} \overline{X}_{p-1}.$$

   (b) SSE and the coefficient of multiple determination R^2 remain the same if we first center all the variables and then fit the regression model.
   (Hint: Use part (a) and the fact that SSE is the minimal value achieved by the least squares criterion function.)
   By part (a), if we center all the variables, the only change to the least squares criterion function is adding a constant term and taking out a constant term which is equal to $\hat{\beta}_0$, but now $\hat{\beta}_0^*$ for the new setup of centered data is 0 i.e. nothing really changed in the criterion function. Hence the minimal value achieved by the least squares criterion function SSE remains unchanged. And also the $SSTO$ is unchanged, therefore the coefficient of multiple determination would remain the same as well.

5. **Multiple regression.** The following data set has 30 cases, one response variable Y and two predictor variables X_1, X_2.

```
case Y     X1     X2
1    2.86  0.36  2.14
2   -0.50  0.66  0.74
3    3.24  0.66  1.91
4    0.44 -0.52 -0.41
5    0.04 -0.68  0.45
...  ...   ...   ...
29   2.60  0.84 -0.49
30   0.98 -0.11  2.41
```

   Consider fitting the nonadditive model with interaction between X_1 and X_2. (R output is given at the end.)

(a) Write down the first 4 rows of the design matrix $\mathbf{X}$.

$$\begin{vmatrix} 1 & 0.36 & 2.14 & 0.7704 \\ 1 & 0.66 & 0.74 & 0.4884 \\ 1 & 0.66 & 1.91 & 1.2606 \\ 1 & -0.52 & -0.41 & 0.2132 \\ ... & ... & ... & ... \end{vmatrix}$$

(b) What is error sum of squares of this model?
$SSE = 27.048$.

(c) We want to conduct prediction at $X_1 = 0, X_2 = 0$ and it is given that

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0.087 & -0.014 & -0.035 & -0.004 \\ -0.014 & 0.115 & -0.012 & -0.052 \\ -0.035 & -0.012 & 0.057 & -0.014 \\ -0.004 & -0.052 & -0.014 & 0.050 \end{bmatrix}.$$

What is the predicted value? What is the prediction standard error? Construct a 95% prediction interval.
The predicted value when $X_1 = 0, X_2 = 0$ is
$\hat{Y}_h = X_h'\hat{\boldsymbol{\beta}} = 0.9918 + 0 + 0 + 0 = 0.9918$ where

$$X_h' = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}, \quad \hat{\boldsymbol{\beta}}' = \begin{bmatrix} 0.9918 & 1.5424 & 0.5799 & -0.1491 \end{bmatrix}.$$

$s(pred) = \sqrt{MSE\left[1 + X_h'(X'X)^{-1}X_h\right]} = 1.02 \times \sqrt{(1 + 0.087)} = 1.063$.
A 95% confidence prediction interval is $0.9918 \pm t(0.975; 30-4) \times 1.063 = 0.9918 \pm 2.056 \times (1.063) = (-1.19, 3.18)$.

6. **(Optional Problem)** Under the Normal error model (with X variables $X_1, \cdots, X_{p-1}$), show that if $\beta_1 = \cdots = \beta_{p-1} = 0$, then $SSR \sim \sigma^2\chi^2_{(p-1)}$ and $SSTO \sim \sigma^2\chi^2_{(n-1)}$.

(Hint: use eigen- decomposition and the fact that $(\mathbf{H}_n - \frac{1}{n}\mathbf{J}_n)\mathbf{1}_n = \mathbf{0}$.*)*

When $\beta_1 = \cdots = \beta_{p-1} = 0$, $\boldsymbol{X\beta} = \beta_0\mathbf{1}_n$.

$$\begin{aligned} SSR &= Y^{'}(H - \frac{1}{n}J_n)Y \\ &= d'd \end{aligned}$$

Where $d = (H - \frac{1}{n}J_n)Y = (H - \frac{1}{n}J_n)(\beta_0 1_n + \epsilon) = (H - \frac{1}{n}J_n)\epsilon$, since $(\boldsymbol{H} - \frac{1}{n}\boldsymbol{J}_n)\mathbf{1}_n = \mathbf{0}_n$.
Thus,

$$SSR = \epsilon'(H - \frac{1}{n}J_n)\epsilon$$

Let $z = Q\epsilon$, then

$$SSR = \epsilon'\mathbf{Q}'\mathbf{\Lambda}\mathbf{Q}\epsilon = z'\mathbf{\Lambda}z = \sum_{1}^{p-1} z_i^2$$

$$E(z) = E(Q\epsilon) = 0_n, Var(z) = var(Q\epsilon) = Q'var(\epsilon)Q = \sigma^2 I_n$$

Under normal error model, z_i are iid $N(0, \sigma^2)$. Thus, $SSR \sim \sigma^2 \chi^2_{p-1}$

Similar for SSTO.

$$\begin{aligned}
SSTO &= \boldsymbol{Y}'(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n)\boldsymbol{Y} \\
&= (\beta_0 1_n + \epsilon)'(I_n - \frac{1}{n}J_n)(\beta_0 1_n + \epsilon) \\
&= \epsilon'(I_n - \frac{1}{n}J_n)\epsilon \\
&= \epsilon'\mathbf{Q}'\boldsymbol{\Lambda}\mathbf{Q}\epsilon \\
&= (Q\epsilon)'\boldsymbol{\Lambda}(Q\epsilon) \\
&= \sum_{1}^{n-1} z_i^2
\end{aligned}$$

Under normal error model, z_i are iid $N(0, \sigma^2)$. Thus, $SSTO \sim \sigma^2 \chi^2_{n-1}$

7. **(Optional Problem.) Expectation and covariance of quadratic forms.** Let $\mathbf{y}$ be a d-dimensional random vector with $E(\mathbf{y}) = \boldsymbol{\mu}$ and $Var(\mathbf{y}) = \boldsymbol{\Sigma}$. Let $\mathbf{A}$ and $\mathbf{B}$ be $d \times d$ symmetric matrices. Show the following:

   (a) $E(\mathbf{y}'\mathbf{A}\mathbf{y}) = tr(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}$.

   Proof.

$$\begin{aligned}
E(\mathbf{y}'\mathbf{A}\mathbf{y}) &= E(tr(\mathbf{y}'\mathbf{A}\mathbf{y})) \\
&= E(tr(\mathbf{A}\mathbf{y}\mathbf{y}')) \\
&= tr(E(\mathbf{A}\mathbf{y}\mathbf{y}')) \\
&= tr(\mathbf{A}E(\mathbf{y}\mathbf{y}')) \\
&= tr(\mathbf{A}\boldsymbol{\Sigma} + \mathbf{A}\boldsymbol{\mu}\boldsymbol{\mu}') \\
&= tr(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}
\end{aligned}$$

   □

   (b) Assume that $\mathbf{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then

$$Cov(\mathbf{y}'\mathbf{A}\mathbf{y}, \mathbf{y}'\mathbf{B}\mathbf{y}) = 2tr(\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\Sigma}) + 4\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\mu}.$$

   Specifically:

$$Var(\mathbf{y}'\mathbf{A}\mathbf{y}) = 2tr((\mathbf{A}\boldsymbol{\Sigma})^2) + 4\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\mu}.$$

   (Hint: (i) First show the above for $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}_d$; (ii) Use the fact: $X \sim N(0, \sigma^2)$, then $E(X^3) = 0$ and $E(X^4) = 3\sigma^4$.)

Proof. We can represent $\mathbf{y}$ as $\boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{z}$, where $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, then

$$\begin{aligned}\mathbf{y}'\mathbf{A}\mathbf{y} &= (\boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{z})'\mathbf{A}(\boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{z}) \\ &= (\boldsymbol{\mu}' + \mathbf{z}'\boldsymbol{\Sigma}^{1/2})\mathbf{A}(\boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{z}) \\ &= \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu} + 2\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\Sigma}^{1/2}\mathbf{z} + \mathbf{z}'\boldsymbol{\Sigma}^{1/2}\mathbf{A}\boldsymbol{\Sigma}^{1/2}\mathbf{z}.\end{aligned}$$

Thus,

$$\begin{aligned}&Cov(\mathbf{y}'\mathbf{A}\mathbf{y}, \mathbf{y}'\mathbf{B}\mathbf{y}) \\ &\quad= Cov(2\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\Sigma}^{1/2}\mathbf{z} + \mathbf{z}'\boldsymbol{\Sigma}^{1/2}\mathbf{A}\boldsymbol{\Sigma}^{1/2}\mathbf{z}, 2\boldsymbol{\mu}'\mathbf{B}\boldsymbol{\Sigma}^{1/2}\mathbf{z} + \mathbf{z}'\boldsymbol{\Sigma}^{1/2}\mathbf{B}\boldsymbol{\Sigma}^{1/2}\mathbf{z}) \\ &\quad= 4Cov(\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\Sigma}^{1/2}\mathbf{z}, \boldsymbol{\mu}'\mathbf{B}\boldsymbol{\Sigma}^{1/2}\mathbf{z}) + Cov(2\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\Sigma}^{1/2}\mathbf{z}, \mathbf{z}'\boldsymbol{\Sigma}^{1/2}\mathbf{B}\boldsymbol{\Sigma}^{1/2}\mathbf{z}) + \\ &\quad Cov(\mathbf{z}'\boldsymbol{\Sigma}^{1/2}\mathbf{A}\boldsymbol{\Sigma}^{1/2}\mathbf{z}, 2\boldsymbol{\mu}'\mathbf{B}\boldsymbol{\Sigma}^{1/2}\mathbf{z}) + Cov(\mathbf{z}'\boldsymbol{\Sigma}^{1/2}\mathbf{A}\boldsymbol{\Sigma}^{1/2}\mathbf{z}, \mathbf{z}'\boldsymbol{\Sigma}^{1/2}\mathbf{B}\boldsymbol{\Sigma}^{1/2}\mathbf{z}))\end{aligned}$$

Since $E(z_i^3) = 0$, we have

$$Cov(2\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\Sigma}^{1/2}\mathbf{z}, \mathbf{z}'\boldsymbol{\Sigma}^{1/2}\mathbf{B}\boldsymbol{\Sigma}^{1/2}\mathbf{z}) = Cov(\mathbf{z}'\boldsymbol{\Sigma}^{1/2}\mathbf{A}\boldsymbol{\Sigma}^{1/2}\mathbf{z}, 2\boldsymbol{\mu}'\mathbf{B}\boldsymbol{\Sigma}^{1/2}\mathbf{z}) = 0.$$

Besides, we have

$$4Cov(\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\Sigma}^{1/2}\mathbf{z}, \boldsymbol{\mu}'\mathbf{B}\boldsymbol{\Sigma}^{1/2}\mathbf{z}) = 4\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\mu}.$$

Let $\mathbf{z}'\boldsymbol{\Sigma}^{1/2}\mathbf{A}\boldsymbol{\Sigma}^{1/2}\mathbf{z} = \sum_{i,j} a_{i,j} z_i z_j$, and $\mathbf{z}'\boldsymbol{\Sigma}^{1/2}\mathbf{B}\boldsymbol{\Sigma}^{1/2}\mathbf{z} = \sum_{k,l} b_{k,l} z_k z_l$, then

$$\begin{aligned}&Cov(\mathbf{z}'\boldsymbol{\Sigma}^{1/2}\mathbf{A}\boldsymbol{\Sigma}^{1/2}\mathbf{z}, \mathbf{z}'\boldsymbol{\Sigma}^{1/2}\mathbf{B}\boldsymbol{\Sigma}^{1/2}\mathbf{z}) = \sum_{i,j,k,l} a_{i,j} b_{k,l} Cov(z_i z_j, z_k z_l) \\ &\quad= 2\sum_i a_{i,i} b_{i,i} + \sum_{i \neq j} a_{i,j} b_{i,j} + \sum_{i \neq j} a_{i,j} b_{j,i} = 2tr(\boldsymbol{\Sigma}^{1/2}\mathbf{A}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2}\mathbf{B}\boldsymbol{\Sigma}^{1/2}) \\ &\quad= 2tr(\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\Sigma}).\end{aligned}$$

□

(c) Use part (a) to derive $E(SSE)$ and $E(SSR)$.

Proof.

$$\begin{aligned}E(SSE) &= E(Y'(\mathbf{I} - \mathbf{H})Y) \\ &= tr(\mathbf{I} - \mathbf{H})\sigma^2 + \boldsymbol{\beta}'\mathbf{X}'(\mathbf{I} - \mathbf{H})\mathbf{X}\boldsymbol{\beta} \\ &= (n - p)\sigma^2\end{aligned}$$

$$\begin{aligned}E(SSR) &= E(Y'(\mathbf{H} - \frac{1}{n}\mathbf{J})Y) \\ &= tr(\mathbf{H} - \frac{1}{n}\mathbf{J})\sigma^2 + \boldsymbol{\beta}'\mathbf{X}'(\mathbf{H} - \frac{1}{n}\mathbf{J})\mathbf{X}\boldsymbol{\beta} \\ &= (p - 1)\sigma^2 + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{X}'\frac{1}{n}\mathbf{J}\mathbf{X}\boldsymbol{\beta}\end{aligned}$$

□