Assignment 3

Rin Meng
Student ID: 51940633

December 2, 2024

Note: From this point onwards I will use tilde on top of a variable to denote that it is a vector.

1. Given the linear regression model in the form $\tilde{y} = \mathbf{X}\tilde{\beta} + \tilde{\epsilon}$ where $E[\tilde{\epsilon}] = 0$ and $E[\tilde{\epsilon}\tilde{\epsilon}^{\mathbf{T}}] = \sigma^2 I$, and $\tilde{\epsilon}$ is normally distributed, which implies that the least-squares estimator for $\tilde{\beta}$

$$\tilde{\beta} = (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\tilde{y}$$

Where we require the matrix $(\mathbf{X}^{\mathbf{T}}\mathbf{X})$ to be invertible.

a) Show that $\mathbf{X}^{\mathbf{T}}\mathbf{X} + \lambda\mathbf{I}$ is invertible, that is $\det(\mathbf{X}^{\mathbf{T}}\mathbf{X} + \lambda\mathbf{I}) \neq 0$ when $\lambda \neq 0$, and $\mathbf{I}$ is the identity matrix.

Proof. We know that $\mathbf{X}^{\mathbf{T}}\mathbf{X}$ is non-invertible, that is $\det(\mathbf{X}^{\mathbf{T}}\mathbf{X}) = 0$, and a and d are non-negative.

$$\begin{aligned}
\mathbf{X}^{\mathbf{T}}\mathbf{X} + \lambda\mathbf{I} &= \begin{bmatrix} a & b \\ c & d \end{bmatrix} + \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} a+\lambda & b \\ c & d+\lambda \end{bmatrix}
\end{aligned}$$

Using the formula of determinant of 2×2 matrix, we have

$$\begin{aligned}
\det(\mathbf{X}^{\mathbf{T}}\mathbf{X} + \lambda\mathbf{I}) &= \det\left(\begin{bmatrix} a+\lambda & b \\ c & d+\lambda \end{bmatrix}\right) \\
&= (a+\lambda)(d+\lambda) - bc \\
&= ad + a\lambda + d\lambda + \lambda^2 - bc \\
&= ad + \lambda(a+d) + \lambda^2 - bc
\end{aligned}$$

Given that $\det(\mathbf{X^TX}) = 0 \Rightarrow ad - bc = 0 \Leftrightarrow ad = bc$.

$$\begin{aligned}\det(\mathbf{X^TX} + \lambda\mathbf{I}) &= bc + \lambda(a+d) + \lambda^2 - bc \\ &= \lambda(a+d) + \lambda^2\end{aligned}$$

$$\begin{aligned}\because \lambda \neq 0 &\Rightarrow \lambda^2 > 0 \\ \because a, d \geq 0 &\Rightarrow a + d > 0 \\ &\Rightarrow \lambda(a+d) + \lambda^2 > 0\end{aligned}$$

$\therefore$ We have shown that $\det(\mathbf{X^TX} + \lambda\mathbf{I}) \neq 0$ when $\lambda \neq 0$. $\square$

b) Given the expression

$$(\tilde{y} - \mathbf{X}\tilde{\beta})^{\mathbf{T}}(\tilde{y} - \mathbf{X}\tilde{\beta}) + \lambda\tilde{\beta}^{\mathbf{T}}\tilde{\beta}$$

where $\lambda > 0$

Let us expand the expression algebraically, using what we learned about minimizing $(\tilde{y} - \mathbf{X}\tilde{\beta})^{\mathbf{T}}(\tilde{y} - \mathbf{X}\tilde{\beta})$ to obtain the estimator $\tilde{\beta}$ in terms of $\mathbf{X}$, $\tilde{y}$ and λ.

$$\begin{aligned}(\tilde{y} - \mathbf{X}\tilde{\beta})^{\mathbf{T}}(\tilde{y} - \mathbf{X}\tilde{\beta}) &= (\tilde{y}^{\mathbf{T}} - \tilde{\beta}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}})(\tilde{y} - \mathbf{X}\tilde{\beta}) \\ &= \tilde{y}^{\mathbf{T}}\tilde{y} - \tilde{y}^{\mathbf{T}}\mathbf{X}\tilde{\beta} - \tilde{\beta}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\tilde{y} + \tilde{\beta}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{X}\tilde{\beta} \\ &= \tilde{y}^{\mathbf{T}}\tilde{y} - 2\tilde{\beta}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\tilde{y} + \tilde{\beta}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{X}\tilde{\beta} \\ (\tilde{y} - \mathbf{X}\tilde{\beta})^{\mathbf{T}}(\tilde{y} - \mathbf{X}\tilde{\beta}) + \lambda\tilde{\beta}^{\mathbf{T}}\tilde{\beta} &= \tilde{y}^{\mathbf{T}}\tilde{y} - 2\tilde{\beta}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\tilde{y} + \tilde{\beta}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{X}\tilde{\beta} + \lambda\tilde{\beta}^{\mathbf{T}}\tilde{\beta} \\ &= \tilde{y}^{\mathbf{T}}\tilde{y} - 2\tilde{\beta}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\tilde{y} + \tilde{\beta}^{\mathbf{T}}(\mathbf{X}^{\mathbf{T}}\mathbf{X} + \lambda\mathbf{I})\tilde{\beta}\end{aligned}$$

Now let us minimize this expression my deriving it with respect to $\tilde{\beta}$ and setting it to zero.

$$\begin{aligned}\frac{\partial}{\partial\tilde{\beta}}(\tilde{y}^{\mathbf{T}}\tilde{y} + 2\tilde{\beta}^{\mathbf{T}}\mathbf{X}^{T}\tilde{y} + \tilde{\beta}^{\mathbf{T}}(\mathbf{X}^{\mathbf{T}}\mathbf{X} + \lambda\mathbf{I})\tilde{\beta}) &= 0 \\ -2\mathbf{X}^{\mathbf{T}}\tilde{y} + (\mathbf{X}^{\mathbf{T}}\mathbf{X} + \lambda\mathbf{I})\tilde{\beta} &= 0 \\ \mathbf{X}^{\mathbf{T}}\tilde{y} &= (\mathbf{X}^{\mathbf{T}}\mathbf{X} + \lambda\mathbf{I})\tilde{\beta} \\ \tilde{\beta} &= (\mathbf{X}^{\mathbf{T}}\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^{\mathbf{T}}\tilde{y}\end{aligned}$$

$\therefore$ The estimator $\tilde{\beta}$ in terms of $\mathbf{X}$, $\tilde{y}$ and λ is $(\mathbf{X}^{\mathbf{T}}\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^{\mathbf{T}}\tilde{y}$.

c) Show that the estimator $\tilde{\beta}$ is biased.

$$
\begin{aligned}
(\mathbf{X^TX}+\lambda\mathbf{I})^{-1}\mathbf{X^T}\tilde{y} &= (\mathbf{X^TX}+\lambda\mathbf{I})^{-1}\mathbf{X^T}(\mathbf{X}\beta+\epsilon) \\
&= (\mathbf{X^TX}+\lambda\mathbf{I})^{-1}\mathbf{X^TX}\beta+(\mathbf{X^TX}+\lambda\mathbf{I})^{-1}\mathbf{X^T}\epsilon \\
\text{Let } \mathbf{A} &= (\mathbf{X^TX}+\lambda\mathbf{I})^{-1}\mathbf{X^TX} \\
\mathrm{E}[\tilde{\beta}] &= \mathrm{E}[A\beta]+(\mathbf{X^TX}+\lambda\mathbf{I})^{-1}\mathbf{X^T}\mathrm{E}[\epsilon] \\
\mathrm{E}[\tilde{\beta}] &= \mathrm{E}[A\beta]
\end{aligned}
$$

For $\tilde{\beta}$ to be unbiased, we need $\mathrm{E}[\tilde{\beta}]=\beta \Leftrightarrow \mathrm{E}[\mathbf{A}\beta]=\beta \leftrightarrow \mathbf{A}=\mathbf{I}$, but we know that this is now true because $\mathbf{A}=(\mathbf{X^TX}+\lambda\mathbf{I})^{-1}\mathbf{X^TX}\neq\mathbf{I}$.
$\therefore$ The estimator $\tilde{\beta}$ is biased.

2. Given that $H_0:\beta_2=\beta_6=0$, and $H_0:\beta_2=\beta_6,\ \beta_3=\beta_4$.

a) A matrix T can rpresent the null hypothesis $H_0:\beta_2=\beta_6=0$ as

$$
T=\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}
$$

b) A matrix T can represent the null hypothesis $H_0:\beta_2=\beta_6,\ \beta_3=\beta_4$ as

$$
T=\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & -1 & 0 & 0 \end{bmatrix}
$$

Using the fact that $\beta_2=\beta_6 \Leftrightarrow \beta_2-\beta_6=0$ and $\beta_3=\beta_4 \Leftrightarrow \beta_3-\beta_4=0$.

c) Under $H_0:\mathbf{T}\tilde{\beta}=0$, show that $\mathrm{Var}(\mathbf{T}\tilde{\beta})=\sigma^2\mathbf{T}(\mathbf{X^TX})^{-1}\mathbf{T^T}$.

Proof.

$$
\begin{aligned}
\mathrm{Var}(\mathbf{T}\tilde{\beta}) &= \mathrm{Var}(\mathbf{T}(\mathbf{X^TX})^{-1}\mathbf{X^T}\tilde{\epsilon}) \\
&= \mathbf{T}(\mathbf{X^TX})^{-1}\mathbf{X^T}\cdot\mathrm{Var}(\tilde{\epsilon})\cdot\mathbf{X}(\mathbf{X^TX})^{-1}\mathbf{T^T} \\
&= \mathbf{T}(\mathbf{X^TX})^{-1}\mathbf{X^T}\cdot\sigma^2\mathbf{I}\cdot\mathbf{X}(\mathbf{X^TX})^{-1}\mathbf{T^T} \\
&= \sigma^2\mathbf{T}(\mathbf{X^TX})^{-1}\mathbf{X^TX}(\mathbf{X^TX})^{-1}\mathbf{T^T} \\
&= \sigma^2\mathbf{T}(\mathbf{X^TX})^{-1}\mathbf{T^T}
\end{aligned}
$$

$\therefore$ We have shown that $\mathrm{Var}(\mathbf{T}\tilde{\beta})=\sigma^2\mathbf{T}(\mathbf{X^TX})^{-1}\mathbf{T^T}$. □

d) Under $H_0:\mathbf{T}\tilde{\beta}=0$, show that $\mathbf{T}\tilde{\beta}=\mathbf{T}(\mathbf{X^TX})^{-1}\mathbf{X^T}\tilde{\epsilon}$.

Proof.

$$\begin{aligned}
\mathbf{T}\tilde{\beta} &= \mathbf{T}(\mathbf{X^TX})^{-1}\mathbf{X^T}\tilde{y} \\
&= \mathbf{T}(\mathbf{X^TX})^{-1}\mathbf{X^T}(\mathbf{X}\beta + \epsilon) \\
&= \mathbf{T}(\mathbf{X^TX})^{-1}\mathbf{X^TX}\beta + \mathbf{T}(\mathbf{X^TX})^{-1}\mathbf{X^T}\epsilon \\
&= \mathbf{T}\beta + \mathbf{T}(\mathbf{X^TX})^{-1}\mathbf{X^T}\epsilon \\
&= \mathbf{T}(\mathbf{X^TX})^{-1}\mathbf{X^T}\epsilon
\end{aligned}$$

$\therefore$ We have shown that $\mathbf{T}\tilde{\beta} = \mathbf{T}(\mathbf{X^TX})^{-1}\mathbf{X^T}\tilde{\epsilon}$. □

e) Under $H_0 : \mathbf{T}\tilde{\beta} = 0$, show that $\hat{\tilde{\beta}}^{\mathbf{T}}\mathbf{T^T}\Sigma^{-1}\mathbf{T}\hat{\tilde{\beta}} \sim \chi^2_{(r)}$, where $\Sigma^{-1} = \mathrm{Var}(\mathbf{T}\hat{\tilde{\beta}})^{-1}$.

Proof. Recall that the $\hat{\tilde{\beta}} = (\mathbf{X^TX})^{-1}\mathbf{X^T}\tilde{y}$, where $\tilde{y} = \mathbf{X}\beta + \tilde{\epsilon}$. The variance of $\mathbf{T}\hat{\tilde{\beta}}$ is $\mathrm{Var}(\mathbf{T}\hat{\tilde{\beta}}) = \sigma^2\mathbf{T}(\mathbf{X^TX})^{-1}\mathbf{T^T}$. Thus,

$$\begin{aligned}
\Sigma &= \sigma^2\mathbf{T}(\mathbf{X^TX})^{-1}\mathbf{T^T} \\
\Sigma^{-1} &= \frac{1}{\sigma^2}(\mathbf{T}(\mathbf{X^TX})^{-1}\mathbf{T^T})^{-1}
\end{aligned}$$

Now we have

$$\begin{aligned}
\hat{\tilde{\beta}}^{\mathbf{T}}\mathbf{T^T}\Sigma^{-1}\mathbf{T}\hat{\tilde{\beta}} &= \hat{\tilde{\beta}}^{\mathbf{T}}\mathbf{T^T}\left(\frac{1}{\sigma^2}(\mathbf{T}(\mathbf{X^TX})^{-1}\mathbf{T^T})^{-1}\right)\mathbf{T}\hat{\tilde{\beta}} \\
&= \frac{1}{\sigma^2}\hat{\tilde{\beta}}^{\mathbf{T}}\mathbf{T^T}(\mathbf{T}(\mathbf{X^TX})^{-1}\mathbf{T^T})^{-1}\mathbf{T}\hat{\tilde{\beta}}
\end{aligned}$$

Let us go back to our null hypothesis $H_0 : \mathbf{T}\tilde{\beta} = 0$. We know that the $\hat{\tilde{\beta}}$ is normally distributed with mean 0 and covariance matrix $\sigma^2(\mathbf{X^TX})^{-1}$, therefore so is $\mathbf{T}\hat{\tilde{\beta}}$. Then we have

$$\mathbf{T}\hat{\tilde{\beta}} \sim N(\mathbf{T}\beta, \sigma^2\mathbf{T}(\mathbf{X^TX})^{-1}\mathbf{T^T})$$

Under H_0 this becomes

$$\mathbf{T}\hat{\tilde{\beta}} \sim N(0, \sigma^2\mathbf{T}(\mathbf{X^TX})^{-1}\mathbf{T^T})$$

Now, let S be the statistic of our test, then

$$S = \frac{1}{\sigma^2}\hat{\hat{\beta}}^{\mathbf{T}}\mathbf{T}^{\mathbf{T}}(\mathbf{T}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{T}^{\mathbf{T}})^{-1}\mathbf{T}\hat{\hat{\beta}}$$

then S is definetly in a form of χ^2 distribution, we can show by letting

$$\mathbf{z} = \mathbf{T}\hat{\hat{\beta}}$$

$$\mathbf{A} = \frac{1}{\sigma^2}(\mathbf{T}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{T}^{\mathbf{T}})^{-1}$$

and that implies,

$$S = \mathbf{z}^{\mathbf{T}}A\mathbf{z}$$

where z $\sim N(0, \sigma^2\mathbf{T}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{T}^{\mathbf{T}})$, as shown above. The rank of $\mathbf{T}$ is $r = \mathbf{rank}(\mathbf{T})$, which represents the number of linearly independent rows in $\mathbf{T}$, where it ultimately implies that the degrees of freedom of χ^2 distribution is r.

$\therefore$ We have shown that $\hat{\hat{\beta}}^{\mathbf{T}}\mathbf{T}^{\mathbf{T}}\Sigma^{-1}\mathbf{T}\hat{\hat{\beta}} \sim \chi^2_{(\mathbf{rank(T)})} = \chi^2_{(r)}$. □

f) Show that

$$(\mathbf{I} - \mathbf{H})[\mathbf{X}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{T}^{\mathbf{T}}\mathbf{C}^{-1}\mathbf{T}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}] = 0$$

Proof. We know that $\mathbf{H} = \mathbf{X}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}$, and $\mathbf{C} = \mathbf{T}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{T}^{\mathbf{T}}$. The term $(\mathbf{I} - \mathbf{H})$ is the projection matrix that projects onto the orthogonal complement of the column space of $\mathbf{X}$. Thus we get:

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

$$\begin{gathered}(\mathbf{I} - \mathbf{H})[\mathbf{X}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{T}^{\mathbf{T}}\mathbf{C}^{-1}\mathbf{T}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}] \\ = \mathbf{X}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{T}^{\mathbf{T}}\mathbf{C}^{-1}\mathbf{T}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X} \\ -\mathbf{H}(\mathbf{X}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{T}^{\mathbf{T}}\mathbf{C}^{-1}\mathbf{T}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X})\end{gathered}$$

and since $\mathbf{HX} = \mathbf{X}$, we have

$$= \mathbf{X}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{T}^{\mathbf{T}}\mathbf{C}^{-1}\mathbf{T}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}$$

$$-\mathbf{X}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{T}^{\mathbf{T}}\mathbf{C}^{-1}\mathbf{T}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}$$

$$= 0$$

$\therefore$ We have shown that $(\mathbf{I} - \mathbf{H})[\mathbf{X}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{T}^{\mathbf{T}}\mathbf{C}^{-1}\mathbf{T}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}] = 0$. □

g) Under $H_0 : \mathbf{T}\tilde{\beta} = 0$, show that

$$F_0 = \frac{(\hat{\tilde{\beta}}^{\mathbf{T}}\mathbf{T}^{\mathbf{T}}\mathbf{C}^{-1}\mathbf{T}\hat{\tilde{\beta}})/r}{MSE} \sim F_{(r,n-p)}$$

where MSE is computed for the full model (with p parameters)

From earlier proof, we know that $\hat{\tilde{\beta}}^{\mathbf{T}}\mathbf{T}^{\mathbf{T}}\mathbf{C}^{-1}\mathbf{T}\hat{\tilde{\beta}} \sim \chi^2_{(r)}$. We also know that $MSE = \frac{SSE}{n-p} \Leftrightarrow MSE \sim \chi^2_{(n-p)}$. Then we have something like this

$$F_0 = \frac{(\hat{\tilde{\beta}}^{\mathbf{T}}\mathbf{T}^{\mathbf{T}}\mathbf{C}^{-1}\mathbf{T}\hat{\tilde{\beta}})/r}{SSE/(n-p)}$$

where we notice that this is a form of F distribution, where both numerator and denominator are some sort of χ^2 distribution now visually we can see that

$$= \frac{\sigma^2\chi^2_r}{\sigma^2\chi^2_{n-p}} = \frac{\chi^2_r}{\chi^2_{n-p}} \sim F_{(r,n-p)}$$

h) Find a matrix $\mathbf{T}$ that represents some hypothesis:

$$H_\gamma : \beta_0 = \beta_1 = \beta_2 = \cdots = \beta_k$$

Proof. Let us consider the hypothesis

$$H_\gamma : \beta_0 = \beta_1 = \beta_2 = \cdots = \beta_k = \beta$$

(here we let β can represent a common value for all the parameters). We can represent this hypothesis as

$$\beta_0 - \beta = 0, \beta_1 - \beta = 0, \beta_2 - \beta = 0, \ldots, \beta_k - \beta = 0$$

and our goal is to present this a matrix $\mathbf{T}$ such that

$$\mathbf{T}\beta = 0$$

where $\beta = [\beta_0, \beta_1, \ldots, \beta_k]^{\mathbf{T}}$.

Now lets start constructing the matrix $\mathbf{T}$, with these constraints:

i. We need the matrix $\mathbf{T}$ to be the size of $k_{\text{rows}} \times (k+1)_{\text{column}}$, where k is the number of parameters and the $+1$ is reserved for the intercept term.

ii. The first column of $\mathbf{T}$ represents constraints for β_0, the second column represents constraints for β_1, and so on.

iii. Each row of $\mathbf{T}$ represents the constraints of a parameter in some form $\beta_i - \beta = 0$. We can write this as $\mathbf{T}\beta = \mathbf{c}$.

$$\mathbf{T} = \begin{bmatrix} 1 & -1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

then the $\mathbf{T}\beta$ matrix will look like this

$$\mathbf{T}\beta = \begin{bmatrix} 1 & -1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} = \begin{bmatrix} \beta_0 - \beta_1 \\ \beta_1 - \beta_2 \\ \beta_2 - \beta_3 \\ \vdots \\ \beta_{k-1} - \beta_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

which implies that $\mathbf{c} = 0 \Rightarrow \mathbf{T}\beta = 0$. $\therefore$ We have shown that the matrix $\mathbf{T}$ that represents the hypothesis $H_\gamma : \beta_0 = \beta_1 = \beta_2 = \cdots = \beta_k$ is $\square$

3. Problem 3.25 on page 130, using "lm" function to answer the following questions. We are given that the linear regression model is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \epsilon$$

a. $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta \Leftrightarrow \beta_1 - \beta = 0, \beta_2 - \beta = 0, \beta_3 - \beta = 0, \beta_4 - \beta = 0$.

```
# Assuming that the data we are using is table.b1
# are the columns/predictors
# Load required library to use linearHypothesis()
library(car)
model = lm(y ~ x1 + x2 + x3 + x4, data = table.b1)

# testing the hypothesis
linear_hypothesis_test(model, "x1 = x2 = x3 = x4")
```

```
# function to perform hypothesis test using
# linearHypothesis()
linear_hypothesis_test <- function(model) {

# use matrix to specify the hypothesis and constraintss
# from the last proof we did in the previous question
hypothesis_matrix <- matrix(c(1, -1, 0, 0, 0,
                              0, 1, -1, 0, 0,
                              0, 0, 1, -1, 0,
                              0, 0, 0, 1, -1),
                nrow = 4, ncol = 5, byrow = TRUE)

# specify the hypothesis (all coefficients equal to 0)
hypothesis_values <- c(0, 0, 0, 0)

# perform the linear hypothesis test
linear_hypothesis_result <- linearHypothesis(model,
                                hypothesis_matrix,
                                hypothesis_values)
return(linear_hypothesis_result)}
```

b. $H_0 : \beta_1 = \beta_2, \beta_3 = \beta_4 \Leftrightarrow \beta_1 - \beta_2 = 0, \beta_3 - \beta_4 = 0.$

```
# Let us use the same technique as above

linear_hypothesis_test(model, "x1 = x2, x3 = x4")

linear_hypothesis_test <- function(model) {

hypothesis_matrix <- matrix(c(0, 1, -1, 0, 0,
                              0, 0, 0, 1, -1),
                nrow = 2, ncol = 5, byrow = TRUE)

hypothesis_values <- c(0, 0)

linear_hypothesis_result <- linearHypothesis(model,
                                hypothesis_matrix,
                                hypothesis_values)
```

```
return(linear_hypothesis_result)}
```

c. $H_0 : \beta_1 - 2\beta_2 = 4\beta_3,\ \beta_1 + 2\beta_2 = 0 \Leftrightarrow \beta_1 = -2\beta_2 + 4\beta_3, \beta_1 = -2\beta_2$.

```
# Again, we will use the same technique as above
# but redefine the hypothesis matrix and values
linear_hypothesis_test(model, "x1 - 2*x2 = 4*x3,
                         x1 + 2*x2 = 0")

linear_hypothesis_test <- function(model) {

hypothesis_matrix <- matrix(c(1, -2, 0, 4, 0,
                              1, 2, 0, 0, 0),
                nrow = 2, ncol = 5,
                byrow = TRUE)

hypothesis_values <- c(0, 0)

linear_hypothesis_result <- linearHypothesis(model,
                                hypothesis_matrix,
                                hypothesis_values)

return(linear_hypothesis_result)}
```

4. Problem 3.1 on page 125

   (a)
```
library(MPV)
model_3.1 <- lm(y ~ x2 + x7 + x8, data = table.b1)
#####################
Call:
lm(formula = y ~ x2 + x7 + x8, data = table.b1)

Residuals:
    Min      1Q  Median      3Q     Max
-3.0370 -0.7129 -0.2043  1.1101  3.7049
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.808372   7.900859  -0.229 0.820899
x2           0.003598   0.000695   5.177 2.66e-05 ***
x7           0.193960   0.088233   2.198 0.037815 *
x8          -0.004816   0.001277  -3.771 0.000938 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’
0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.706 on 24 degrees of freedom
Multiple R-squared:  0.7863, Adjusted R-squared:  0.7596
F-statistic: 29.44 on 3 and 24 DF,  p-value: 3.273e-08
```

(b)
```
anova_3.1 <- anova(model_3.1)
#####################
Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value    Pr(>F)
x2         1  76.193  76.193  26.172 3.100e-05 ***
x7         1 139.501 139.501  47.918 3.698e-07 ***
x8         1  41.400  41.400  14.221 0.0009378 ***
Residuals 24  69.870   2.911
---
Signif. codes:  0 ‘***’ 0.001 ‘**’
0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

(c)
```
t_stats <- summary_model_3.1$coefficients[, "t value"]
#####################
(Intercept)          x2          x7          x8
-0.228883    5.177090    2.198262   -3.771036
```

The conclusion we can draw about the roles the variables x_2, x_7, and x_8 play in predicting y is that x_2 and x_7 are significant predictors of y because their t-statistics are greater than 2 in absolute value, and their p-values are less than 0.05. x_8 is **highly** significant predictor of

y because its t-statistic is less than -2.

$\therefore$ We reject all null hypotheses that $\beta_2 = 0$, $\beta_7 = 0$, and $\beta_8 = 0$.

(d)
```
r_squared <- summary_model_3.1$r.squared
#####################
[1] 0.7863069
adj_r_squared <- summary_model_3.1$adj.r.squared
#####################
[1] 0.7595953
```

$\therefore$ The R^2 value is 0.7863 and the adjusted R^2 value is 0.7596.

(e)
```
model_reduced <- lm(y ~ x2 + x8, data = table.b1)
summary(model_reduced)
#####################
Call:
lm(formula = y ~ x2 + x8, data = table.b1)

Residuals:
    Min      1Q  Median      3Q     Max
-2.4280 -1.3744 -0.0177  1.0010  4.1240

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 14.7126750  2.6175266   5.621 7.55e-06 ***
x2           0.0031111  0.0007074   4.398 0.000178 ***
x8          -0.0068083  0.0009658  -7.049 2.18e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.832 on 25 degrees of freedom
Multiple R-squared:  0.7433, Adjusted R-squared:  0.7227
F-statistic: 36.19 on 2 and 25 DF,  p-value: 4.152e-08



model_full <- model_3.1

# Get the residual sum of squares (RSS) for both models
```

```
RSS_full <- sum(residuals(model_full)^2)
RSS_reduced <- sum(residuals(model_reduced)^2)

# Get the number of parameters
# (coefficients) in the full and reduced models
p_full <- length(coef(model_full))  # including the intercept
p_reduced <- length(coef(model_reduced))  # including the intercept

# Get the degrees of freedom for the full model
df_full <- df.residual(model_full)

MSR_full <- (RSS_full - RSS_reduced) / (p_full - p_reduced)
MSR_reduced <- RSS_reduced / df_full

# Calculate the partial F-statistic
F_statistic <-  MSR_full / MSR_reduced
F_statistic
#####################
[1] 4.832354
```

Using the adjusted R^2 value, we can see that the R^2 value for the full model is 0.7863, and the R^2 value for the reduced model is 0.7433. Which implies that the full model is better than the reduced model. The partial F-statistic is 4.832354, and we know that $F = t^2$, then we have $t = 2.198262$. This matches the t-statistic for x_7 in the full model. This implies that the F-statistic and the t-statistic for β_7 are directly related.

End of Assignment 3.