Jin Kweon
SID: 3032235207
prof: Oscar Madrid
section: 102 (21047)

<HW 2>

#1.

I will follow the notation book uses.
Book says, $b = \hat{\beta}$.
So, $b = \begin{bmatrix} A \\ B \end{bmatrix}$.

$\hat{y} = A + Bx$ (simple regression)

It is given that $Var(b) = \sigma^2 (X^T X)^{-1}$.

$var(b) = cov(b) \Rightarrow$ get diagonals of the variance-covariance matrix.

Since $\hat{y} = Xb$, X will be $\begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$.

(So, $\hat{y}$ will be $n \times 1$ vector.)

$$X^T X = \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix}$$

$$(X^T X)^{-1} = \frac{1}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \begin{bmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{bmatrix}$$

So,

$$Var(b) = \sigma^2 (X^T X)^{-1} = \frac{\sigma^2}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \begin{bmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{bmatrix}$$

By the definition of variance matrix,

$$Var(A) = \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} = \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n\sum_{i=1}^n x_i^2 - 2\left(\sum_{i=1}^n x_i\right)^2 + \left(\sum_{i=1}^n x_i\right)^2}$$

$$= \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n\sum_{i=1}^n x_i^2 - 2n\bar{x}\sum_{i=1}^n x_i + \left(n \cdot \frac{1}{n}\sum_{i=1}^n x_i\right)^2} =$$

$$= \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n\sum_{i=1}^n x_i^2 - 2n\bar{x}\sum_{i=1}^n x_i + (n\bar{x})^2}$$

$$= \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n\left(\sum_{i=1}^n x_i^2 - 2\bar{x}\sum_{i=1}^n x_i + (n\bar{x})^2\right)}$$

$$= \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n\left(\sum_{i=1}^n x_i^2 - 2\bar{x}\sum_{i=1}^n x_i + \left(\sum_{i=1}^n \bar{x}\right)^2\right)}$$

$$= \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n\sum (x_i - \bar{x})^2}$$

And,

$$Var(B) = \frac{n\sigma^2}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}$$

$$= \frac{n\sigma^2}{n\sum_{i=1}^n (x_i - \bar{x})^2}$$, as I proved above.

$$= \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

<other way>

I know from the normal equation that $b=(X^TX)^{-1}X^Ty$, assumed that (X^TX) is invertible.

I already got $(X^TX)^{-1}$ before.

$$X^Ty=\begin{bmatrix}1 & \cdots & 1\\ x_1 & \cdots & x_n\end{bmatrix}\begin{bmatrix}y_1\\ \vdots\\ y_n\end{bmatrix}=\begin{bmatrix}\sum_{i=1}^n y_i\\ \sum_{i=1}^n x_iy_i\end{bmatrix}$$

I will write $\sum x_i$ instead of $\sum_{i=1}^n x_i$ for convenience.

$$b=(X^TX)^{-1}X^Ty=\frac{1}{n\sum x_i^2-(\sum x_i)^2}\begin{bmatrix}\sum x_i^2 & -\sum x_i\\ -\sum x_i & n\end{bmatrix}\begin{bmatrix}\sum y_i\\ \sum x_iy_i\end{bmatrix}$$

$$=\frac{1}{n\sum x_i^2-(\sum x_i)^2}\begin{bmatrix}\sum x_i^2\sum y_i-\sum x_i\sum x_iy_i\\ -\sum x_i\sum y_i+n\sum x_iy_i\end{bmatrix}$$

So, $A=\dfrac{\sum x_i^2\sum y_i-\sum x_i\sum x_iy_i}{n\sum x_i^2-(\sum x_i)^2}$

$$B=\frac{-\sum x_i\sum y_i+n\sum x_iy_i}{n\sum x_i^2-(\sum x_i)^2}=\frac{\sum x_iy_i-\frac{1}{n}\sum x_i\sum y_i}{\sum(x_i-\bar{x})^2}$$

$$=\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}=\frac{\sum(x_i-\bar{x})y_i}{\sum(x_i-\bar{x})^2}$$

And, $\mathrm{var}(B)=\mathrm{var}\left(\dfrac{\sum(x_i-\bar{x})(A+Bx_i+e_i)}{\sum(x_i-\bar{x})^2}\right)$

$$=\mathrm{var}\left(\frac{\sum(x_i-\bar{x})e_i}{\sum(x_i-\bar{x})^2}\right)=\frac{\sigma^2\sum(x_i-\bar{x})^2}{\left(\sum(x_i-\bar{x})^2\right)^2}$$

$$=\frac{\sigma^2}{\sum(x_i-\bar{x})^2}$$

And, $A=\dfrac{\sum x_i^2\sum y_i-\sum x_i\sum x_iy_i}{n\sum(x_i-\bar{x})^2}$

Since $\bar{y}=A+B\bar{x}$, $\mathrm{var}(A)=\mathrm{var}(\bar{y}-B\bar{x})$

$=\mathrm{var}(\bar{y})+\mathrm{var}(B\bar{x})-2\mathrm{cov}(\bar{y},B\bar{x})$

$=\mathrm{var}(\frac{1}{n}(y_1+\cdots+y_n))+\bar{x}^2\mathrm{var}(B)-2\mathrm{cov}(\bar{y},B\bar{x})$

Here, $-2\mathrm{cov}(\bar{y},B\bar{x})=-2\bar{x}\,\mathrm{cov}(\frac{1}{n}\sum y_i,B)$

$$=\frac{-2\bar{x}}{n}\sum_i\mathrm{cov}(y_i,B)=-\frac{2\bar{x}}{n}\sum_i\mathrm{cov}\left(y_i,\frac{\sum_j(x_j-\bar{x})y_j}{\sum_k(x_k-\bar{x})^2}\right)$$

$$=\frac{-2\bar{x}}{n\sum_k(x_k-\bar{x})^2}\sum_i\mathrm{cov}\left(y_i,\sum_j(x_j-\bar{x})(y_j-\bar{y})\right)$$

$$=\frac{-2\bar{x}}{n\sum_k(x_k-\bar{x})^2}\sum_i\sum_j(x_j-\bar{x})\,\mathrm{cov}(y_i,y_j-\bar{y})$$

as $x_j-\bar{x}$ is fixed

$$=\frac{-2\bar{x}}{n\sum_k(x_k-\bar{x})^2}\sum_i\sum_j(x_j-\bar{x})\left[\mathrm{cov}(y_i,y_j)-\mathrm{cov}(y_i,\bar{y})\right]$$

$$=\frac{-2\bar{x}}{n\sum_k(x_k-\bar{x})^2}\sum_i\left[\sigma^2(x_i-\bar{x})-\mathrm{cov}(y_i,\bar{y})\sum_j(x_j-\bar{x})\right]$$

as $y_i\perp\!\!\!\perp y_j$ and $\mathrm{var}(y_i,y_i)=\sigma^2$

$\hookrightarrow \begin{cases}y_i=\beta_0+\beta_1x_i+e_i\\ y_j=\beta_0+\beta_1x_j+e_j\end{cases}\rightarrow$ and $e_i\perp\!\!\!\perp e_j$

$$=\frac{-2\bar{x}}{n\sum_k(x_k-\bar{x})^2}\sum_i\sigma^2(x_i-\bar{x})$$

as $\sum_j(x_j-\bar{x})=0$

$=0$

as $\sum_i(x_i-\bar{x})=0$.

So, $\frac{1}{n^2}n\,\mathrm{var}(y_i)+\bar{x}^2\dfrac{\sigma^2}{\sum(x_i-\bar{x})^2}$

$$=\frac{\sigma^2\sum(x_i-\bar{x})^2+\sigma^2\cdot n\cdot\bar{x}^2}{n\sum(x_i-\bar{x})^2}=\frac{\sum x_i^2}{n}\cdot\frac{\sigma^2}{\sum(x_i-\bar{x})^2}$$

And, $\mathrm{cov}(\hat{\beta}_0,\hat{\beta}_1)=\mathrm{cov}(\bar{y}-\hat{\beta}_1\bar{x},\hat{\beta}_1)=-\mathrm{cov}(\hat{\beta}_1\bar{x},\hat{\beta}_1)=$

$=-\bar{x}\,\mathrm{var}(\hat{\beta}_1)$

$\boxed{\begin{array}{l}\star\text{Tip}\Rightarrow\begin{cases}\mathrm{cov}(ax+b,y)=a\,\mathrm{cov}(x,y),\ \text{if } a,b\in\mathbb{R}.\\ \mathrm{cov}(ax+b,y)=a\,\mathrm{cov}(x,y)\,a^T,\ \text{if } a \text{ is a vector.}\end{cases}\end{array}}$

Jin Kweon

#2

$H = X(X^TX)^{-1}X^T$ → Think it as orthogonal projection matrix.

ⓐ prove $H = H^T$. (Symmetric)

$H^T = [X(X^TX)^{-1}X^T]^T$

$= (X^T)^T[(X^TX)^{-1}]^T(X)^T$

$= X[(X^TX)^{-1}]^T X^T$

$= X(X^TX)^{-1}X^T$, as $(X^TX)^{-1}$ is a symmetric (inverse of a symmetric matrix is also symmetric, and we know X^TX is always symmetric).

So, $H^T = X(X^TX)^{-1}X^T = H$

So, H is symmetric

ⓑ prove $H = H^2$ (idempotent)

$H^2 = H \cdot H = [X(X^TX)^{-1}X^T][X(X^TX)^{-1}X^T]$

$= X(X^TX)^{-1}(X^TX)(X^TX)^{-1}X^T$

$= X(X^TX)^{-1}X^T$, as $(X^TX)^{-1}(X^TX) = I$.

So, H is idempotent.

∴ proof is finished...

∴ So, H is orthogonal projection as H is symmetric and idempotent

#3.

a) I will prove it in three different ways.

① I am proving

$\frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n} \hat{y}_i$.

So, $\frac{1}{n}(y_1 + \cdots + y_n) = \frac{1}{n}(\hat{y}_1 + \cdots + \hat{y}_n)$

So, $\frac{1}{n}[1 \cdots 1]\, y = \frac{1}{n}[1 \cdots 1]\, \hat{y}$

I know that $\hat{y}_i = X\hat{\beta}_i$, where X is a design matrix with an intercept such that

$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}$$

(In OLS, sum of y-values are equal to sum of $\hat{y}$-values when X have an intercept. otherwise, it is not true)

↳ Goal of our proof

I want to mimimize RSS, and I get a normal equation, $X^Ty = X^TX\hat{\beta}$, where $\hat{\beta}$ is the least square estimator. (solution of normal equation.)

So, $X^Ty = X^TX\hat{\beta}$.

$$\underbrace{\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ x_{1p} & \cdots & x_{np} \end{pmatrix}}_{(p+1)\times n} \underbrace{y}_{n\times 1} = \underbrace{\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ x_{1p} & \cdots & x_{np} \end{pmatrix}}_{(p+1)\times n} \underbrace{\begin{pmatrix} 1 & \cdots & x_{1p} \\ \vdots & & \vdots \\ 1 & \cdots & x_{np} \end{pmatrix}}_{n\times(p+1)} \underbrace{\hat{\beta}}_{(p+1)\times n}$$

*Intuition → Sum of residuals is zero, when there is an intercept

⇒ c.t.

$\Longrightarrow$ c.t.

So,

$$\begin{pmatrix} 1\cdot y_1 + \cdots + 1\cdot y_n \\ \vdots \\ x_{1p}\cdot y_1 + \cdots + x_{np}\cdot y_n \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ x_{1p} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} \hat\beta_0 + \cdots + \hat\beta_p x_{1p} \\ \vdots \\ \hat\beta_0 + \cdots + \hat\beta_p x_{np} \end{pmatrix}}$$

$$\| $$

$$\begin{pmatrix} 1\cdot(\hat\beta_0 + \cdots + \hat\beta_p x_{1p}) + \cdots + 1\cdot(\hat\beta_0 + \cdots + \hat\beta_p x_{np}) \\ \vdots \\ x_{1p}(\hat\beta_0 + \cdots + \hat\beta_p x_{1p}) + \cdots + x_{np}(\hat\beta_0 + \cdots + \hat\beta_p x_{np}) \end{pmatrix}$$

So, in linear algebra, matrix is associative, but since we do not know X^T is full rank (X^T is invertible), I need to prove $y = X\hat\beta$ algebraically

So,

$$\begin{pmatrix} (1 \cdots 1)\, y \\ \vdots \\ (x_{1p} \cdots x_{np})\, y \end{pmatrix} = \begin{pmatrix} (1 \cdots 1) \underbrace{\begin{pmatrix} \hat\beta_0 + \cdots + \hat\beta_p x_{1p} \\ \vdots \\ \hat\beta_0 + \cdots + \hat\beta_p x_{np} \end{pmatrix}}_{X\hat\beta} \\ \vdots \\ (x_{1p} \cdots x_{np}) \begin{pmatrix} \hat\beta_0 + \cdots + \hat\beta_p x_{1p} \\ \vdots \\ \hat\beta_0 + \cdots + \hat\beta_p x_{np} \end{pmatrix} \end{pmatrix}$$

This shows $(1 \cdots 1)\, y = (1 \cdots 1)\, \hat y$!!!

☆ <warning> Don't say $y = \begin{pmatrix} \hat\beta_0 + \cdots + \hat\beta_p x_{1p} \\ \vdots \\ \hat\beta_0 + \cdots + \hat\beta_p x_{np} \end{pmatrix}$!!! $\Rightarrow$ Above matrix does not mean $y = X\hat\beta = \hat y$!!!

② Not good one

Need to prove $\frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n} \hat y_i$,

or prove $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat y_i$.

$$\sum_{i=1}^{n} \hat y_i = \sum_{i=1}^{n} (\hat\beta_0 + \hat\beta_1 x_{i1} + \hat\beta_2 x_{i2} + \cdots + \hat\beta_p x_{ip})$$

$$= \sum_i \left[(\bar y - \hat\beta_1 \bar x_{i1} - \cdots - \hat\beta_p \bar x_{ip}) + \hat\beta_1 x_{i1} + \cdots + \hat\beta_p x_{ip} \right]$$

$$= \sum_i \bar y - \hat\beta_1 \sum_i \bar x_{i1} \cdots - \hat\beta_p \sum_i \bar x_{ip} + \hat\beta_1 \sum_i x_{i1} + \cdots + \hat\beta_p \sum_i x_{ip}$$

$$= n\bar y - \hat\beta_1 \cdot n \cdot \bar x_{i1} - \cdots - \hat\beta_p \cdot n \cdot \bar x_{ip} + \hat\beta_1 \cdot n \cdot \bar x_{i1} + \cdots + \hat\beta_p \cdot n \cdot \bar x_{ip}$$

$$= n\bar y = n \cdot \frac{1}{n}\sum_i y_i = \sum y_i$$

And, be aware from what I proved, if there is no intercept ($= \hat\beta_0$) we cannot solve it !!! So, we need an intercept.

☆ actually this claim is consequence of 3a.

$$\bar y = \bar{\hat y} = \frac{1}{n}\sum_i \hat y_i \Longrightarrow$$

$$= \frac{1}{n}\sum_i (\hat\beta_0 + \hat\beta_1 x_{i1} + \cdots + \hat\beta_p x_{ip})$$

$$= \hat\beta_0 + \hat\beta_1 \bar x_1 + \cdots + \hat\beta_p \bar x_p$$

Jin Kwon

$\Longrightarrow$ c.t. 3-a)

③ $\bar{\hat{y}} = \frac{1}{n}\sum_i \hat{y}_i = \frac{1}{n}\sum_i (X\hat{\beta})_i$

$= \frac{1}{n}\sum_i (\underbrace{X(X^TX)^{-1}X^T}_{=H} y)_i$

And, $Hy = \begin{bmatrix} \text{Symmetric and} \\ \text{idempotent} \\ \text{projection matrix} \\ \text{onto } Col(X) \end{bmatrix} y$

$= \begin{bmatrix} y_1 h_{11} + \cdots + y_n h_{1n} \\ \vdots \\ y_1 h_{n1} + \cdots + y_n h_{nn} \end{bmatrix}$,

where $H = \begin{bmatrix} h_{11} & \cdots & h_{1n} \\ \vdots & \ddots & \vdots \\ h_{n1} & \cdots & h_{nn} \end{bmatrix}$.

So, $\frac{1}{n}\sum_i (Hy)_i = \frac{1}{n}\left(y_1(h_{11}+\cdots+h_{n1}) + \cdots + y_n(h_{1n}+\cdots+h_{nn})\right)$.

Since $(1,\cdots,1)^T \in col(X)$,
$H(1,\cdots,1)^T = (1,\cdots,1)^T$ as $Hv = v$ for all $v \in col(X)$.
↳ we assume intercept exists in the model.

So, sum of every row is 1 $(H \cdot 1 = 1)$.

Since $H = H^T$, sum of every column should be 1, as well.

So, $\frac{1}{n}(y_1(h_{11}+\cdots+h_{n1}) + \cdots + y_n(h_{1n}+\cdots+h_{nn}))$

$= \frac{1}{n}(y_1(1) + \cdots + y_n(1)) = \frac{1}{n}\sum y_i$

$= \bar{y}$

$\boxed{\therefore \bar{\hat{y}} = \bar{y}}$

b)

we do not know X^TX is invertible.
Use normal equation: $X^TX\hat{\beta} = X^Ty$.
Although $\hat{\beta}$ is not unique (as X^TX is not invertible), $\hat{y} = X\hat{\beta}$ is unique

$\hat{y}^T\hat{y} = (X\hat{\beta})^T(X\hat{\beta}) = \hat{\beta}^TX^TX\hat{\beta} =$

$= \hat{\beta}^TX^Ty = \hat{\beta}^T(X^TX\hat{\beta}) = (\hat{\beta}^TX^TX)\hat{\beta}$

$= (X^TX\hat{\beta})^T\hat{\beta} = (X^Ty)^T\hat{\beta} = y^TX\hat{\beta}$

$= y^T\hat{y}$ ↳ ☆ or, $(y-\hat{y})^T\hat{y} = 0$

$\therefore$ proved.

c) we said $\widehat{cov}(x,y) = \frac{1}{n}\sum_i (x_i - \bar{x})(y_i - \bar{y})$

because this is random (empirical) and actually $cov(X,Y) = E[(X-E(X))(Y-E(Y))]$
$= E(XY) - E(X)E(Y)$.

$var(\hat{y}) = cov(\hat{y},\hat{y}) = \frac{1}{n}\sum_i (\hat{y}_i - \bar{\hat{y}})(\hat{y}_i - \bar{\hat{y}}) =$

$= \frac{1}{n}(\hat{y} - \bar{\hat{y}})^T(\hat{y} - \bar{\hat{y}}) = \frac{1}{n}(\hat{y} - \bar{y})^T(\hat{y} - \bar{y}) =$

$= \frac{1}{n}(\hat{y}^T - \bar{y}^T)(\hat{y} - \bar{y}) =$

$= \frac{1}{n}(\hat{y}^T\hat{y} - \hat{y}^T\bar{y} - \bar{y}^T\hat{y} + \bar{y}^T\bar{y})$.

Now, I will find $cov(y,\hat{y})$ and compare it with $var(\hat{y})$.

$cov(y,\hat{y}) = \frac{1}{n}\sum_i (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}}) =$

$= \frac{1}{n}\sum_i (y_i - \bar{y})(\hat{y}_i - \bar{y}) = \frac{1}{n}(y - \bar{y})^T(\hat{y} - \bar{y}) =$

$= \frac{1}{n}(y^T - \bar{y}^T)(\hat{y} - \bar{y}) = \frac{1}{n}(y^T\hat{y} - y^T\bar{y} - \bar{y}^T\hat{y} + \bar{y}^T\bar{y})$

$= \frac{1}{n}(\hat{y}^T\hat{y} - y^T\bar{y} - \bar{y}^T\hat{y} + \bar{y}^T\bar{y})$

So, I need to prove $\hat{y}^T\bar{y} = y^T\bar{y}$.

$\Longrightarrow$ c.t.

$\Rightarrow$ c.t.

prove $y^T\bar{y} = \hat{y}^T\bar{y}$.

So, $y^T \begin{bmatrix} \bar{y} \\ \vdots \\ \bar{y} \end{bmatrix} = \hat{y}^T \begin{bmatrix} \bar{y} \\ \vdots \\ \bar{y} \end{bmatrix}$.

So, $\bar{y}\, y^T \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \bar{y}\, \hat{y}^T \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$.

So, $\bar{y} \sum_i y_i = \bar{y} \sum_i \hat{y}_i$.

So, $\bar{y} \cdot \frac{1}{n} \sum_i y_i = \bar{y} \cdot \frac{1}{n} \sum_i \hat{y}_i$.

So, $\bar{y} \cdot \bar{y} = \bar{y} \cdot \bar{\hat{y}}$.

And, as proved in part a),

$\bar{\hat{y}} = \bar{y}$.

So, <u>proof</u> is finished.

#4. (Body fat data $\Rightarrow 252 \times 15$)

Our $H_0: \beta_1 = \cdots = \beta_5 = 0. \rightarrow$ lm()

For denominator, we divide RSS by $n-p-1 = 252-(5+1) = 246$.

For numerator, we divide RegSS by p (# of variables dropped) $= 5$.

So, F test has degree of freedom of 5 and 246.

$$F_0 = \frac{\frac{RegSS}{p}}{\frac{RSS}{n-p-1}} = \frac{\frac{TSS-RSS}{p}}{\frac{RSS}{n-p-1}}.$$

But how do I make it easier?

Since $RSE = \sqrt{\frac{RSS}{n-p-1}}$,

$(RSE)^2 \cdot (n-p-1) = (4.122)^2 \cdot 246 = 4179.757 = RSS$.

Also, $R^2 = \frac{TSS-RSS}{TSS}$.

So, $TSS = TSS \cdot R^2 + RSS$.

It implies $TSS \cdot R^2 = RegSS$.

$$\text{So, } F_0 = \frac{\frac{TSS \cdot R^2}{p}}{\frac{RSS}{n-p-1}} = \frac{TSS \cdot R^2 (n-p-1)}{p \cdot RSS}$$

$$= \boxed{\frac{R^2}{1-R^2} \cdot \frac{n-p-1}{p}}, \text{ as } R^2 = 1 - \frac{RSS}{TSS}.$$

$$\sim F_{p, n-p-1} = F_{5, 246}.$$

$\Longrightarrow$ c.t.

$\Longrightarrow$ c.t. #4.

So, $F = \dfrac{0.7228}{1-0.7228} \cdot \dfrac{246}{5}$

$\approx \boxed{128.29}$

p-value $= P(F_{5,246} > 128.29)$

$\approx 0.$

So, we reject the null.

So, there is linear relationship in that model.

So, the blank spaces are

① 128.29
② 5
③ 0

Jin Kwon

#5. a)

I need to understand the meaning of 'correct model.'

So, original model applies to x_0 such that $Y_0 = x_0^T\beta + e_0$, where $e_0 \sim N(0, \sigma^2)$ and $e_i \perp\!\!\!\perp e_j$ for $i \neq j$.

'Model is correct' does not mean all these assumptions, but 'model' is referring to these assumptions.

In ch9, we assume $y_i = x_i^T\beta + e$ for some unknown β and e is indepent Gaussian with variance σ^2.

This set-up is assumed in Ch.9.

So, in this problem, we try to find if the model is correct, then we assume x_0 have the same model, not just $x_1, \dots, x_n$

So, I need to prove that if the model is correct, then $E(\delta) = E(\hat{Y}_0 - E(Y_0)) = 0$ and $var(\hat{Y}_0 - E(Y_0)) = \sigma^2 x_0^T (X^TX)^{-1} x_0$.

As I said above, for the model to be correct, error is a random variable with $N(0, \sigma^2)$ and independent with x_0 and $e_i \perp\!\!\!\perp e_j$ for $i \neq j$, where $Y_0 = x_0^T\beta + e_0$

Then, $E(Y_0) = E(x_0^T\beta + e_0) = E(x_0^T\beta) + E(e_0) = x_0^T\beta$

So, $E(\hat{Y}_0 - E(Y_0)) = E(\hat{Y}_0) - E(E(Y_0))$

$= E(x_0^T\hat{\beta}) - E(E(x_0^T\beta + e_0)) = E(x_0^T\hat{\beta}) - E(x_0^T\beta) =$

$= E(x_0^T(X^TX)^{-1}X^Ty) - x_0^T\beta = x_0^T(X^TX)^{-1}X^TE(y) - x_0^T\beta$

$= x_0^T(X^TX)^{-1}X^TE(X\beta + e) - x_0^T\beta = x_0^T(X^TX)^{-1}X^TX\beta - x_0^T\beta$

$= x_0^T\beta - x_0^T\beta = 0.$

$\Longrightarrow$

$\Rightarrow$

$V(\delta) = var(\hat{Y}_0 - E(Y_0)) = var(X_0^T\hat{\beta} - X_0^T\beta)$

$= var(X_0^T\hat{\beta})$, as $X_0^T\beta$ is fixed.

$= X_0^T var(\hat{\beta}) X_0 = X_0^T var((X^TX)^{-1}X^Ty)X_0$

$= X_0^T (X^TX)^{-1}X^T . var(y) [(X^TX)^{-1}X^T]^T X_0$

$= X_0^T (X^TX)^{-1}X^T . \sigma^2 . X (X^TX)^{-1} X_0$

$= \sigma^2 X_0^T (X^TX)^{-1} (X^TX) (X^TX)^{-1} X_0$

$= \sigma^2 X_0^T (X^TX)^{-1} X_0$

$\therefore$ So, it is proved when the model is correct, x_0 have the same model.

---

b) Now, we are interested in working on the error in forecasting the actual value, Y_0.

The question proves that

$$D = \hat{Y}_0 - Y_0 = X_0^T(\hat{\beta} - \beta) - e_0.$$

$E(D) = E[X_0^T(\hat{\beta} - \beta) - e_0] =$

$= E[X_0^T(\hat{\beta} - \beta)] - E(e_0)$

$= X_0^T E(\hat{\beta} - \beta) - E(e_0) = X_0^T E((X^TX)^{-1}X^Ty - \beta)$

$= X_0^T (X^TX)^{-1} X^T E(y) - X_0^T E(\beta)$

$= X_0^T (X^TX)^{-1} X^T E(X\beta + e) - X_0^T\beta$

$= X_0^T (X^TX)^{-1} X^T X\beta - X_0^T\beta$

$= X_0^T\beta - X_0^T\beta = 0$

$var(D) = var(X_0^T(\hat{\beta} - \beta) - e_0)$

$= var(X_0^T(\hat{\beta} - \beta)) + var(e_0) - 2cov(X_0^T(\hat{\beta} - \beta), e_0)$

$= X_0^T var(\hat{\beta} - \beta) X_0 + var(e_0) - 2X_0^T cov(\hat{\beta} - \beta, e_0)$

$\Longrightarrow$

$\Longrightarrow$

$= X_0^T var(\hat{\beta} - \beta) X_0 + \sigma^2$
$\quad - 2X_0^T [cov(\hat{\beta}, e_0) - cov(\beta, e_0)]$

$= X_0^T var(\hat{\beta}) X_0 + \sigma^2 - 2X_0^T cov(\hat{\beta}, e_0)$, as β is fixed.

$= X_0^T var((X^TX)^{-1}X^Ty) X_0 + \sigma^2$
$\quad - 2X_0^T cov(\hat{\beta}, e_0)$

$= X_0^T (X^TX)^{-1} X^T var(y) X (X^TX)^{-1} X_0$
$\quad + \sigma^2 - 2X_0^T cov((X^TX)^{-1}X^Ty, e_0)$

$= \sigma^2 X_0^T (X^TX)^{-1} X_0 + \sigma^2$
$\quad - 2X_0^T (X^TX)^{-1} X^T cov(y, y_0 - X_0^T\beta)$

$= \sigma^2 X_0^T (X^TX)^{-1} X_0 + \sigma^2$
$\quad - 2X_0^T (X^TX)^{-1} X^T cov(X\beta + e, e_0)$

$= \sigma^2 X_0^T (X^TX)^{-1} X_0 + \sigma^2$
$\quad - 2X_0 (X^TX)^{-1} X^T cov(e, e_0)$

and as $e_i \perp\!\!\!\perp e_j$ for $i \neq j$,
$cov(e, e_0) = 0$.

$= \sigma^2 X_0^T (X^TX)^{-1} X_0 + \sigma^2$

$= \sigma^2 (X_0^T (X^TX)^{-1} X_0 + 1) \geq var(\delta)$

The intuition is that the actual value Y_0 always has some noise, so I need to include variance for the noise as well, but $E(Y_0)$ does not have noise. Also, $var(\delta)$ can be small since mean is more efficiently estimated as it is summary of distribution and as we have some information we can get less variance.

$\Longrightarrow$

Jin Kweon

$\Rightarrow$ C.t. #5

☆ $E(Y_0)$ is fixed and Y_0 is r.v., and C.I. is random interval that contains $E(Y_0)$ and Y_0 with certain probability ☆

c) $X_0^T = [1, 13, 12000, 50]$

part i)

$E(Y_0) = E(X_0^T\beta + e_0) = E(X_0^T\beta) = X_0^T\beta$.

And, my goal is to construct 90% confidence interval of $X_0^T\beta$.

And, $E(X_0^T\hat{\beta}) = X_0^T E((X^TX)^{-1}X^TY)$

$= X_0^T(X^TX)^{-1}X^T E(X\beta + e)$

$= X_0^T(X^TX)^{-1}X^TX\beta = X_0^T\beta$

And,

$var(X_0^T\hat{\beta}) = X_0^T var(\hat{\beta}) X_0$

$= X_0^T var((X^TX)^{-1}X^TY) X_0$

$= X_0^T(X^TX)^{-1}X^T var(y) X (X^TX)^{-1} X_0$

$= \sigma^2 X_0^T(X^TX)^{-1}X_0$.

As y is normally distributed, $\hat{\beta}$ is normally distributed, so $X_0^T\hat{\beta}$ is normally distributed as well.

So, $\hat{Y}_0 = X_0^T\hat{\beta} \sim N(X_0^T\beta, \sigma^2 X_0^T(X^TX)^{-1}X_0)$

So, $\dfrac{X_0^T\hat{\beta} - X_0^T\beta}{\sigma\sqrt{X_0^T(X^TX)^{-1}X_0}} \sim N(0,1)$.

Since I don't know σ, I need to estimate it.

$RSE = \dfrac{RSS}{n-p-1} = \dfrac{RSS}{\sigma^2} \cdot \dfrac{\sigma^2}{n-p-1} \sim \chi^2_{n-p-1} \cdot \dfrac{\sigma^2}{n-p-1}$
$(=S)$

So, $\dfrac{X_0^T\hat{\beta} - X_0^T\beta}{S\sqrt{X_0^T(X^TX)^{-1}X_0}} \sim \dfrac{X_0^T\hat{\beta} - X_0^T\beta}{\sqrt{\frac{\chi^2_{n-p-1}}{n-p-1}}\,\sigma\sqrt{X_0^T(X^TX)^{-1}X_0}}$

$\sim t_{n-p-1}$, as $\dfrac{Z}{\sqrt{\frac{U}{d}}} \sim t_d$ where $Z \sim N(0,1)$ and $U \sim \chi^2_d$ and Z and U are independent.

I know numerator is a function of $\hat{\beta}$ and denominator is a function of $\hat{e}$, and they are independent.

Since we want to build up the 90% C.I., we should have 5% on each side.

$90\% = 1 - 2P\left(\dfrac{X_0^T\hat{\beta} - X_0^T\beta}{S\sqrt{X_0^T(X^TX)^{-1}X_0}} \geq t_{n-p-1}(0.05)\right) =$

$= P\left(-S \cdot t_{n-p-1}(0.05)\sqrt{X_0^T(X^TX)^{-1}X_0} \leq X_0^T\beta - X_0^T\hat{\beta} \leq S \cdot t_{n-p-1}(0.05)\sqrt{X_0^T(X^TX)^{-1}X_0}\right)$

$= P\left(X_0^T\hat{\beta} - S \cdot t_{n-p-1}(0.05)\sqrt{X_0^T(X^TX)^{-1}X_0} \leq X_0^T\beta \leq X_0^T\hat{\beta} + S \cdot t_{n-p-1}(0.05)\sqrt{X_0^T(X^TX)^{-1}X_0}\right)$

So, $X_0^T\hat{\beta} \pm S \cdot t_{n-p-1}(0.05)\sqrt{X_0^T(X^TX)^{-1}X_0}$ is 90% C.I. for $X_0^T\beta$.

☆ Output is in Rmd file.

$\Longrightarrow$

part ii)

$\hat{Y}_0 - Y_0 = D \sim N(0, \sigma^2(1 + X_0^T(X^TX)^{-1}X_0))$

So, I can do similar work as I did in part i).

$$\frac{(\hat{Y}_0 - Y_0) - 0}{\sigma\sqrt{1 + X_0^T(X^TX)^{-1}X_0}} \sim N(0,1).$$

If I replace σ with S again,

$$\frac{\hat{Y}_0 - Y_0}{S\sqrt{1 + X_0^T(X^TX)^{-1}X_0}} = \frac{\hat{Y}_0 - Y_0}{\sqrt{\frac{RSS/\sigma^2}{n-p-1}}\ \sigma\sqrt{1 + X_0^T(X^TX)^{-1}X_0}}$$

$\sim t_{n-p-1}$

So, $90\% = 1 - 2P\left(\frac{\hat{Y}_0 - Y_0}{S\sqrt{1 + X_0^T(X^TX)^{-1}X_0}} \geq t_{n-p-1}(0.05)\right)$

And, $Y_0 \sim \hat{Y}_0 \pm t_{n-p-1}(0.05) \cdot S \cdot \sqrt{1 + X_0^T(X^TX)^{-1}X_0}$

So, I have $\hat{Y}_0 = X_0^T\hat{\beta} = [1 \ 13 \ 12000 \ 50]\begin{bmatrix} -6.794 \\ 4.187 \\ 0.001314 \\ -0.008905 \end{bmatrix}$

and $S = 7.846$ (from the textbook)

and $X^TX = \begin{bmatrix} 102 & 1095 & 693386 & 2956 \\ 1095 & 12513 & 8121410 & 32281 \\ 693386 & 8121410 & 6534383460 & 14093097 \\ 2956 & 32281 & 14093097 & 187312 \end{bmatrix}$

d)

For this question, we have $X_0^T = [1 \ 0 \ 50000 \ 100]$.

Find estimated variance of forecast error, using $S^2[1 + X_0^T(X^TX)^{-1}X_0]$.

More details in Rmd file.

Jin Kwean

#6.

a) I will use two different ways to do this problem.

① I use the formula given on pg. 219 from the textbook.

There is only one constraint

$\Rightarrow H_0: \beta_1 + \beta_2 = \beta_3 + \beta_4$

I can think of it as $\beta_1 + \beta_2 - \beta_3 - \beta_4 = 0$.

Now, use formula below:

$$\frac{(L\hat{\beta} - c)^T [L(X^TX)L^T]^{-1} (L\hat{\beta} - c)}{q\, S_E^2} \sim F_{q, n-p-1}.$$

$L = [0 \;\; 1 \;\; 1 \;\; -1 \;\; -1]$

and $q = 1$ (= # of constraint)

and $p = 4$.

② I re-organize $\beta_1 + \beta_2 = \beta_3 + \beta_4$ to $\beta_1 = \beta_3 + \beta_4 - \beta_2$.

I will use F-statistic:

$$\frac{[RSS(m) - RSS(M)]/q}{RSS(M)/(n-p-1)} \sim F_{q, n-p-1}$$

(or, I can say

$$\frac{[RegSS(M) - RegSS(m)]/q}{RSS(M)/n-p-1}$$).

So, get RSS(M) from the original model:

$$Bodyfat = \beta_0 + \beta_1(knee) + \beta_2(thigh) + \beta_3(hip) + \beta_4(ankle) + e.$$

Replace β_1 to $\beta_3 + \beta_4 - \beta_2$, and get RSS(m):

$$Bodyfat = \beta_0 + (\beta_3 + \beta_4 - \beta_2)(knee) + \beta_2(thigh) + \beta_3(hip) + \beta_4(ankle) + e$$

$$= \beta_0 + \beta_2(thigh - knee) + \beta_3(hip + knee) + \beta_4(knee + ankle) + e.$$

After I got RSS(M) and RSS(m), plug values into the statistic above, and calculate p-value.

$\Longrightarrow$

$\Rightarrow$

b)

Again, null hypothesis is

$H_0: \beta_1+\beta_2-\beta_3-\beta_4=0$,

by re-arranging.

Use t-test by doing

$$\frac{(\hat{\beta}_1+\hat{\beta}_2-\hat{\beta}_3-\hat{\beta}_4)-(\beta_1+\beta_2-\beta_3-\beta_4)}{S.E.(\hat{\beta}_1+\hat{\beta}_2-\hat{\beta}_3-\hat{\beta}_4)}$$

$$= \frac{\hat{\beta}_1+\hat{\beta}_2-\hat{\beta}_3-\hat{\beta}_4}{\sqrt{\widehat{Var}(\hat{\beta}_1+\hat{\beta}_2-\hat{\beta}_3-\hat{\beta}_4)}}$$

$$= \frac{\hat{\beta}_1+\hat{\beta}_2-\hat{\beta}_3-\hat{\beta}_4}{\sqrt{\hat{V}(\hat{\beta}_1)+\hat{V}(\hat{\beta}_2)+\hat{V}(\hat{\beta}_3)+\hat{V}(\hat{\beta}_4)+2\widehat{cov}(\hat{\beta}_1,\hat{\beta}_2)-2\widehat{cov}(\hat{\beta}_1,\hat{\beta}_3)-2\widehat{cov}(\hat{\beta}_1,\hat{\beta}_4)-2\widehat{cov}(\hat{\beta}_2,\hat{\beta}_3)-2\widehat{cov}(\hat{\beta}_2,\hat{\beta}_4)+2\widehat{cov}(\hat{\beta}_3,\hat{\beta}_4)}}$$

Since we know variance-covariance matrix of $\hat{\beta}$ is $\sigma^2(X^TX)^{-1}$.

Since we don't know σ, we get $\hat{\sigma}=RSE=\sqrt{\frac{RSS}{n-p-1}}$.

Be aware that $\sigma^2(X^TX)^{-1}$ includes intercept term, so

$\hat{V}(\hat{\beta}_1)=\hat{\sigma}^2(X^TX)^{-1}_{2,2}$.

c) As I can see on my Rmd file,

$t^2=F$.

Also the p-value are the same.

Jin Kweon (3032235207)_HW2

Jin Kweon

9/13/2017

I include the most of my calculation and theoretical works on the paper I attached. This Rmd file is basically used for calculation purpose only.

Problem 5 part c and d.

$x_0^T \ = \ [1 \ \ 13 \ \ 12000 \ \ 50]$ is used for both part i) and ii).

$S_E \ = \ 7.846$ is given in the text book on pg. 98.

The number of rows/individuals is 102. (So, n - p - 1 = 102 - 3 - 1 = 98)

And, the $\hat{\beta}^T \ = \ [-6.7943 \ \ 4.1866 \ \ 0.0013136 \ \ -0.0089052]$, including the intercept term.

Lasty,

$$X^T X = \begin{bmatrix} 102 & 1095 & 693386 & 2956 \\ 1095 & 12513 & 8121410 & 32281 \\ 693386 & 8121410 & 6534383460 & 14093097 \\ 2956 & 32281 & 14093097 & 187312 \end{bmatrix}$$

, which makes sense, since this matrix is symmetric.

part c part i

$$E(Y_o) \ = \ E(x_o^T \beta \ + \ e_o) \ = \ E(x_0^T \beta) \ = \ x_o^T \beta$$

So, for part i), I want to estimate 90% confidence interval for $E(Y_o) \ = \ x_o^T \beta$.

$$E(Y_o) \ = \ x_0^T \beta \ \sim \ (\hat{Y}_0 \ - \ S \ t_{n-p-1}(0.90) \sqrt{x_0^T \ (X^T X)^{-1} \ x_0} \ , \quad \hat{Y}_0 \ + \ S \ t_{n-p-1}(0.90) \sqrt{x_0^T \ (X^T X)^{-1} \ x_0} \ )$$

```
x0 <- c(1, 13, 12000, 50)
beta_hat <- c(-6.7943, 4.1866, 0.0013136, -0.0089052)
yhat <- as.numeric(x0 %*% beta_hat)
S <- 7.846
xtx <- matrix(c(102, 1095, 693386, 2956, 1095, 12513, 8121410, 32281, 693386, 8121410, 6534383460,
                14093097, 2956, 32281, 14093097, 187312), 4, 4)

low_bound <- as.numeric(yhat - (S * qt(0.95, 98) * sqrt(x0 %*% solve(xtx) %*% x0)))
high_bound <- as.numeric(yhat + (S * qt(0.95, 98) * sqrt(x0 %*% solve(xtx) %*% x0)))

paste("Confidence interval is from", low_bound, "to", high_bound)
```

```
## [1] "Confidence interval is from 60.1936459616152 to 65.7052340383848"
```

Be Careful! This is two-sided test!!!

part c part ii

For part ii), I want to estimate 90% confidence interval for actual Y_0. There is only small difference with what I used for part i).

$$Y_o \;=\; x_0^T\beta + \epsilon_0 \;\sim\; (\hat{Y}_0 - S\, t_{n-p-1}(0.90)\sqrt{1 \;+\; x_0^T\;(X^TX)^{-1}\;x_0}\;,\quad \hat{Y}_0 + S\, t_{n-p-1}(0.90)\sqrt{1 \;+\; x_0^T\;(X^TX)^{-1}\;x_0}\;)$$

```
low_bound2 <- as.numeric(yhat - (S * qt(0.95, 98) * sqrt(1 + x0 %*% solve(xtx) %*% x0)))
high_bound2 <- as.numeric(yhat + (S * qt(0.95, 98) * sqrt(1 + x0 %*% solve(xtx) %*% x0)))

paste("Confidence interval is from", low_bound2, "to", high_bound2)
```

```
## [1] "Confidence interval is from 49.632494791766 to 76.266385208234"
```

part d

Estimated variance of the forecast error is $S^2(1 \;+\; x_0^T\;(X^TX)^{-1}\;x_0)$

```
new_x0 <- c(1, 0, 50000, 100)

estvar <- as.numeric(S^2 * (1 + (t(new_x0) %*% solve(xtx) %*% new_x0)))
estvar
```

```
## [1] 333.278
```

Definitely, I can see that estimated variance of forecast erorr is super large! I think the variance of the forecast error adequately captures the uncertaintly in using the regression equation to predict Y with given x_0. Although the income is quite high, the fact of average education is 0 years and 100% women implies something extraordinary. So, this is telling variance of prestige gap might be super high.

Problem 6

The body density dataset includes the following 15 variables listed from left to right:

Density : Density determined from underwater weighing

Fat : Percent body fat from Siri's (1956) equation

Age : Age (years)

Weight : Weight (kg)

Height : Height (cm)

Neck : Neck circumference (cm)

Chest: Chest circumference (cm)

Abdomen : Abdomen circumference (cm)

Hip : Hip circumference (cm)

Thigh : Thigh circumference (cm)

Knee : Knee circumference (cm)

Ankle : Ankle circumference (cm)

Biceps : Biceps (extended) circumference (cm)

Forearm : Forearm circumference (cm)

Wrist : Wrist circumference (cm)

The above table shows the Body Density Dataset having 15 variables and 252 observations.

(Source for the data)Source

Or, Download through google docs using "=IMPORTHTML("http://wiki.stat.ucla.edu/socr/index.php/SOCR_Data_BMI_Regression","table",2)"

Or, download using download.file function: download.file("https://bcourses.berkeley.edu/courses/1463885/files/folder/labs/lab01/bodyfat.csv", destfile = "aaaa")

Import

```
setwd("/Users/yjkweon24/Desktop/Cal/2017 Fall/Stat 151a/HW/HW2")
data <- read.csv("bodyfat.csv", header = T, stringsAsFactors = F)

dim(data)
```

```
## [1] 252  15
```

```
data <- na.omit(data)
```

Part A

I will use two different ways to solve part a).

First way

```r
intercept <- rep(1, nrow(data))
x <- as.matrix(cbind(intercept, data[,c(11, 10, 9, 12)]))

a <- lm(bodyfat ~ Knee + Thigh + Hip + Ankle, data = data)
a
```

```
##
## Call:
## lm(formula = bodyfat ~ Knee + Thigh + Hip + Ankle, data = data)
##
## Coefficients:
## (Intercept)         Knee        Thigh          Hip        Ankle
##   -48.88229      0.14055      0.01472      0.76667     -0.64317
```

```r
#Just to learn more about how this model works.
summary(a)
```

```
##
## Call:
## lm(formula = bodyfat ~ Knee + Thigh + Hip + Ankle, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.3024  -3.8824  -0.2141   4.5077  18.9750
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -48.88229    7.19700  -6.792 8.20e-11 ***
## Knee          0.14055    0.32614   0.431   0.6669
## Thigh         0.01472    0.18258   0.081   0.9358
## Hip           0.76667    0.14200   5.399 1.57e-07 ***
## Ankle        -0.64317    0.30968  -2.077   0.0388 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.527 on 247 degrees of freedom
## Multiple R-squared:  0.4014, Adjusted R-squared:  0.3918
## F-statistic: 41.42 on 4 and 247 DF,  p-value: < 2.2e-16
```

```r
beta_hat2 <- a$coefficients
beta_hat2
```

```
##  (Intercept)         Knee        Thigh          Hip        Ankle
## -48.88228758   0.14055207   0.01471569   0.76666715  -0.64317133
```

```r
L <- c(0, 1, 1, -1, -1)
F1 <- ((L %*% beta_hat2) * solve(L %*% solve(t(x) %*% x) %*% L) * (L %*% beta_hat2)) / (1 * (6.527)^2)
# 6.527 = S_E, from summary(a)

#Some other ways to get RSE
#1.
#beta_hat <- solve(t(x) %*% x, t(x) %*% data$bodyfat)
#y_hat <- x %*% beta_hat
#sum((data$bodyfat - y_hat)^2)/247


#2.
#sum((data$bodyfat - a$fitted.values)^2)/247

F1 <- as.numeric(round(F1, 6))
F1
```

```
## [1] 0.002814
```

```r
pf(F1, 1, nrow(x)-4-1, lower.tail = F)
```

```
## [1] 0.9577372
```

Second way

$TSS = \frac{RSS}{(1 - R^2)}$, so, $\frac{(1 - R^2) * R^2}{(RSE)^2 * (n - p - 1)} = Regss$, as $Regss = TSS * R^2$ and $RSS = (RSE)^2 * (n - p - 1)$

```r
b <- lm(bodyfat ~ I(Thigh - Knee) + I(Hip + Knee) + I(Knee + Ankle), data = data)
summary(b)
```

```
##
## Call:
## lm(formula = bodyfat ~ I(Thigh - Knee) + I(Hip + Knee) + I(Knee +
##     Ankle), data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.3650  -3.8849  -0.2344   4.5095  18.8999
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -48.90788    7.16635  -6.825 6.72e-11 ***
## I(Thigh - Knee)   0.01118    0.16963   0.066   0.9475
## I(Hip + Knee)     0.77127    0.11213   6.878 4.91e-11 ***
## I(Knee + Ankle)  -0.63203    0.22710  -2.783   0.0058 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.514 on 248 degrees of freedom
## Multiple R-squared:  0.4014, Adjusted R-squared:  0.3942
## F-statistic: 55.44 on 3 and 248 DF,  p-value: < 2.2e-16
```

```r
regss1 <- ((1-0.4014)^2 * (0.4014)^2)/(6.527^2 * 247)
regss0 <- ((1-0.4014)^2 * (0.4014)^2)/(6.514^2 * 247)

rss1 <- sum((data$bodyfat - a$fitted.values)^2)
rss0 <- sum((data$bodyfat - b$fitted.values)^2)

F2 <- round((rss0 - rss1) / (rss1 / (nrow(data) - 5)),6)
F2
```

```
## [1] 0.002814
```

```r
pf(F2, 1, nrow(x)-4-1, lower.tail = F)
```

```
## [1] 0.9577372
```

F statistic value:

The F statistic value is approimately 0.002814.

Degree of freedom:

Degree of freedom will be 1 and 247.

p value:

The P-value is around 0.9577, and we do not reject the null.

Since we do not reject the null, we can accept $\beta_1 + \beta_2 = \beta_3 + \beta_4$. And, this means that we can say $\beta_1 = \beta_3 + \beta_4 - \beta_2$ and plug this into β_1. By doing that, we can have only three explanatory variables to get the linear model with response variable of Bodyfat.

Part B

```r
beta_hat2
```

```
##  (Intercept)         Knee        Thigh          Hip        Ankle
## -48.88228758   0.14055207   0.01471569   0.76666715  -0.64317133
```

```r
xtx <- solve(crossprod(x))

s <- sqrt(sum((a$model[,1] - a$fitted.values)^2)/247)

var_mat <- s^2 * xtx

numerator <- beta_hat2[2] + beta_hat2[3] - beta_hat2[4] - beta_hat2[5]

denominator <- sqrt(var_mat[2,2] + var_mat[3,3] + var_mat[4,4] + var_mat[5,5] + 2*var_mat[2,3] -
                      2*var_mat[2,4] - 2*var_mat[2,5] - 2*var_mat[3,4] -2*var_mat[3,5] +
                      2*var_mat[4,5])
```

```r
t <- as.numeric(numerator / denominator)
t
```

```
## [1] 0.05304566
```

```r
2 * pt(t, 247, lower.tail = F) #I'm multiplying by two since it is two-tailed.
```

```
## [1] 0.9577384
```

T statistic value:

T value is around 0.05304566.

Degree of freedom:

n - p - 1 = 252 - 4 - 1 = 247 will be our d.f.

p value:

The p-value is around 0.9577, which is the one I got from p-value for F test.

Part C

Conclusion:

```r
t^2
```

```
## [1] 0.002813842
```

```r
F2
```

```
## [1] 0.002814
```

```r
identical(round(t^2, 7), round(F2, 7))
```

```
## [1] FALSE
```

We have not proved in the class, but as we assumed that y is normally distributed and noise/error is iid, t^2 is equal to F value.