151A HW 2

Name: Jiyoon Jeong (Clover)
SID: 3032235220

1. (9.6) $V(b) = \sigma_\varepsilon^2 (X^T X)^{-1}$ simple regression: $y = b_0 + b_1 x$ or $= A + Bx$

$p = 2$

Let $X = \begin{bmatrix} 1 & x_{1,1} \\ \vdots & \vdots \\ 1 & x_{n,1} \end{bmatrix}$ then

$$X^T X = \underset{p\times n}{\begin{bmatrix} 1 & \cdots & 1 \\ x_{1,1} & \cdots & x_{n,1} \end{bmatrix}} \underset{n\times p}{\begin{bmatrix} 1 & x_{1,1} \\ \vdots & \vdots \\ 1 & x_{n,1} \end{bmatrix}} = \begin{bmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^n x_{i,1}^2 \end{bmatrix}$$

$$(X^T X)^{-1} = \frac{1}{n\sum_{i=1}^n x_{i,1}^2 - (n\bar{x})^2} \cdot \begin{bmatrix} \sum_{i=1}^n x_{i,1}^2 & -n\bar{x} \\ -n\bar{x} & n \end{bmatrix}$$

$$* \; n\sum_{i=1}^n x_{i,1}^2 - (n\bar{x})^2 = n\sum_{i=1}^n x_{i,1}^2 - n\cdot\bar{x}\cdot\sum_{i=1}^n x_{i,1} = n\left(\sum_{i=1}^n x_{i,1}^2 - \bar{x}\sum_{i=1}^n x_{i,1}\right)$$

$$= n\left(\sum_{i=1}^n x_{i,1}(x_{i,1}-\bar{x})\right) = n\left(\sum (x_{i,1}-\bar{x})(x_{i,1}-\bar{x})\right)$$

$$= n\sum (x_{i,1}-\bar{x})^2$$

$\therefore$ covariance matrix $V(b) = \sigma_\varepsilon^2 (X^T X)^{-1}$

$$= \frac{\sigma_\varepsilon^2}{n\sum(x_i-\bar{x})^2}\begin{bmatrix} \sum^n x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{bmatrix}$$

Since $V(b)$ is covariance matrix between b_i, b_j,

$$V(b) = \begin{bmatrix} \underset{var(b_0)}{cov(b_0,b_0)} & cov(b_0,b_1) \cdots & cov(b_0,b_p) \\ \vdots & \ddots & \\ cov(b_p,b_0) & \cdots & \underset{var(b_p)}{cov(b_p,b_p)} \end{bmatrix} = \Sigma$$

the diagonal entries of $V(b)$ are $var(b_0)$ and $var(b_1)$ in simple regression case,

$$\therefore V(A) = V(b_0) = \frac{\sigma_\varepsilon^2}{n\sum(x_i-\bar{x})^2}\cdot\sum^n x_i^2$$

$$V(B) = V(b_1) = \frac{\sigma_\varepsilon^2}{n\sum(x_i-\bar{x})^2}\cdot n = \frac{\sigma_\varepsilon^2}{\sum(x_i-\bar{x})^2}$$

2-(11.2) Let $H = X(X^TX)^{-1}X^T$

① $H = H^T$

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

② $H^2 = H$

$$H^2 = (X(X^TX)^{-1}X^T)(X(X^TX)^{-1}X^T)$$
$$= X(X^TX)^{-1}X^TX(X^TX)^{-1}X^T = X(X^TX)^{-1}X^T = H$$

$\therefore$ ①② proved

3-(a) consider normal equation : $(X^TX)\hat{\beta} = X^Ty$ ... ①

Let $X = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ \vdots & x_{2,1} & \cdots & x_{2,p} \\ 1 & x_{n,1} & \cdots & x_{n,p} \end{bmatrix}_{n\times(p+1)}$ so,

From ①, $\overbrace{\begin{bmatrix} 1 & \cdots & 1 \\ x_{1,1} & \cdots & x_{n,1} \\ \vdots & & \\ x_{1,p} & \cdots & x_{n,p} \end{bmatrix}_{(p+1)\times n} y}^{X^Ty} = \overbrace{\begin{bmatrix} 1 & \cdots & 1 \\ x_{1,1} & \cdots & x_{n,1} \\ x_{1,p} & \cdots & x_{n,p} \end{bmatrix}_{(p+1)\times n} X\hat{\beta}}^{X^T(X\beta)}$

$\Rightarrow [1 \cdots 1]y = [1 \cdots 1] X\hat{\beta}$

since $\hat{y} = X\hat{\beta}$, $[1 \cdots 1]y = [1 \cdots 1]\hat{y}$

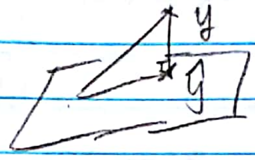
$\Rightarrow [1 \cdots 1]\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = [1 \cdots 1]\begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_n \end{bmatrix}$

$\Rightarrow y_1 + \cdots + y_n = \hat{y}_1 + \cdots + \hat{y}_n \Rightarrow \bar{y} = \overline{\hat{y}_1}$

<15'

3-(b) Show $y^T\hat{y} = \hat{y}^T\hat{y} \Rightarrow (y^T - \hat{y}^T)\hat{y} = 0$

$\Rightarrow (y-\hat{y})^T\hat{y} = 0$



prove $(y-\hat{y})^T\hat{y} = 0$. Let $\hat{y} = X\hat{\beta}$ then

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

$= (y^T X - (X\hat{\beta})^T X)\hat{\beta} = (y^T X - \hat{\beta}^T X^T X)\hat{\beta}$ ... ①

Since normal equation is $X^TX\hat{\beta} = X^Ty \Rightarrow X^Ty - X^TX\hat{\beta} = 0$

Transpose both sides lead $y^TX - \hat{\beta}^TX^TX = 0^T = 0$

Therefore, ① $= 0 \cdot \hat{\beta} = 0$

$\therefore (y-\hat{y})^T\hat{y} = 0$. Proved

3-(c) $cov(y,\hat{y}) = \frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})(\hat{y}_i-\bar{\hat{y}}) = \frac{1}{n}(y-\bar{y})^T(\hat{y}-\bar{\hat{y}})$

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

$= \frac{1}{n}(\underbrace{\hat{y}^T\hat{y}}_{\text{From (b)}} - \underbrace{\boxed{y^T\bar{y}}}_{\text{from (a)}} - \underbrace{\hat{y}^T\hat{y}}_{\text{from (a)}} + \underbrace{\bar{\hat{y}}^T\bar{y}}_{\text{from (a)}})$

$var(\hat{y}) = \frac{1}{n}\sum(\hat{y}_i-\bar{\hat{y}})^2 = \frac{1}{n}(\hat{y}-\bar{\hat{y}})^T(\hat{y}-\bar{\hat{y}})$

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

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

check $y^T\bar{y} = \hat{y}^T\bar{y} \Rightarrow [y_1 \cdots y_n]\begin{bmatrix}\bar{y}\\ \vdots\\ \bar{y}\end{bmatrix} = y^T\bar{y}$

$= (y_1+\cdots+y_n)\bar{y}$ ... ①

$\hat{y}^T\bar{y} = [\hat{y}_1 \cdots \hat{y}_n]\begin{bmatrix}\bar{y}\\ \vdots\\ \bar{y}\end{bmatrix} = (\hat{y}_1+\cdots+\hat{y}_n)\bar{y}$ ... ②

divide ① and ② by n leads

$\frac{1}{n}(y_1 + \cdots + y_n)\bar{y} = y^T\bar{y} \Rightarrow \bar{y}\cdot\bar{y} = \bar{y}\cdot\bar{y}$

$\frac{1}{n}(\hat{y}_1 + \cdots + \hat{y}_n)\bar{y} = \hat{y}^T\bar{y} \Rightarrow \bar{\hat{y}}\cdot\bar{y} = \bar{y}\cdot\bar{y}$ from (a)

$\therefore\ y^T\bar{y} = \hat{y}^T\bar{y} \Rightarrow cov(y, \hat{y}) = var(\hat{y})$

4. Im set null Hypothesis as $H_0: \beta_1 = \cdots = \beta_5 = 0$,

Then our small model m is $y_i = \beta_0 + e_i$

In this case, $RSS(m) = \sum^n (y_i - \bar{y})^2 = SS_{total}$

Then F-statistics is $\frac{(SS_{total} - SS_{res})/P}{SS_{res}/(n-p-1)}$ ... ①

(since $q=0$)

$n-p-1 = 246$ and $p=5 \Rightarrow n = 246+6 = 252$

$$① = \frac{\frac{SS_{total} - SS_{res}}{SS_{total}}}{\frac{SS_{res}}{SS_{total}}} \cdot \frac{(n-p-1)}{p} = \frac{R^2}{1-R^2} \cdot \frac{(n-p-1)}{p}$$

$$= \frac{0.7228}{1-0.7228} \cdot \frac{246}{5} \approx 128.289$$

Since F-statistics $\sim F_{5,246}$ $(F_{p-0,\, n-p-1})$

P-val $(F_{5,246} > 128.289) \approx 0 \Rightarrow$ so we reject null hypothesis.

Blank ①: 128.289 Blank ③: ≈ 0

" ②: 5

1 숙제

H

5-(a) Let $x_0^T = [1 \; x_{01} \; \cdots \; x_{0k}]$

If $\hat{Y}_0 = x_0^T b$ is estimate of $E(Y_0)$,
error in estimation is $\delta = \hat{Y}_0 - E(Y_0)$

(If model is correct, $Y_0 = x_0^T\beta + e_0$) Q
and e is random variable and independent of x_0,
Also, $E(Y_0) = E(x_0^T\beta + e_0) = E(x_0^T\beta) + E(e_0) = x_0^T\beta$

$$E(\delta) = E(\hat{Y}_0 - E(Y_0)) = E(\hat{Y}_0) - E(Y_0) = E(x_0^T b) - x_0^T\beta$$
$$= E(x_0^T(X^TX)^{-1}X^TY) - x_0^T\beta$$
$$= x_0^T(X^TX)^{-1}X^T E(Y) - x_0^T\beta$$
$$= x_0^T(X^TX)^{-1}X^T X\beta - x_0^T\beta = x_0^T\beta - x_0^T\beta = 0$$

$$var(\delta) = var(\hat{Y}_0 - E(Y_0)) = var(\hat{Y}_0) = var(x_0^T b)$$
$$= var(x_0^T(X^TX)^{-1}X^TY)$$
$$= x_0^T(X^TX)^{-1}X^T \, var(Y) \, X(X^TX)^{-1}x_0$$
$$= x_0^T(X^TX)^{-1}X^T \sigma_\varepsilon^2 I_n X(X^TX)^{-1}x_0$$
$$= \sigma_\varepsilon^2 \cdot x_0^T(X^TX)^{-1}x_0$$

5-(b) Forecasted Actual value : $Y_0 = x_0^T\beta + \varepsilon_0$
Error in forecast : $D = \hat{Y}_0 - Y_0 = x_0^T b - (x_0^T\beta + \varepsilon_0)$
$= x_0^T(b-\beta) - \varepsilon_0$

From (a)?

$$E(D) = E(\hat{Y}_0 - Y_0) = E(x_0^T(b-\beta) - \varepsilon_0) =$$
$$= E(x_0^T(b-\beta)) - E(\varepsilon_0) = x_0^T E(b-\beta) = x_0^T \cdot 0 = 0$$

$$var(D) = var(\hat{Y}_0 - Y_0) = var(x_0^T(b-\beta) - \varepsilon_0)$$
$$= var(x_0^T(b-\beta)) - 2cov(x_0^T(b-\beta), \varepsilon_0) + var(\varepsilon_0)$$
$$= x_0^T var(b-\beta) x_0 - 2x_0^T cov(b-\beta, \varepsilon_0) + var(\varepsilon_0)$$

$= x_0^T var(b) x_0 - 2x_0^T cov(b, \varepsilon_0) + 6_{\varepsilon_0}^2$

$= x_0^T var(b) x_0 - 2x_0^T cov((X^TX)^{-1}X^Ty, \varepsilon_0) + 6_{\varepsilon_0}^2$

$= x_0^T var(b) x_0 \underbrace{- 2x_0^T (X^TX)^{-1}X^T cov(y, \varepsilon_0)}_{\cdots ①} + 6_{\varepsilon_0}^2$

$① = -2x_0^T(X^TX)^{-1}X^T \; cov(X\beta + e, e_0)$

$= -2x_0^T(X^TX)^{-1}X^T \; cov(e, e_0)$

Since e_i are iid, $cov(e, e_0) = 0$

$\therefore ① = 0$

$= x_0^T var((X^TX)^{-1}X^Ty) x_0 + 6_{\varepsilon_0}^2$

$= x_0^T (X^TX)^{-1}X^T var(y)) X (X^TX)^{-1} x_0 + 6_{\varepsilon_0}^2$

$= 6_\varepsilon^2 x_0^T (X^TX)^{-1} x_0 + \underline{6_{\varepsilon_0}^2} \rightarrow 6_\varepsilon^2$

$= 6_\varepsilon (x_0^T (X^TX)^{-1} x_0 + 1)$

Q Q: why $var(D) > var(\delta)$?

: Since δ is the error when we estimate y given specific number x_0. D is error when we estimate random variable. Since mean is always more efficiently estimated because it is a summary of distribution, $var(\delta)$ is smaller.

Reading

5-(c) From 5.2, $P = -6.7943 + 4.1866\,Edu + 0.0013136\,Inc - 0.0089052\,W$

$S_E = 7.846$

If $\varepsilon_i \sim iid\ N(0, \sigma_\varepsilon^2) \Rightarrow \varepsilon = \begin{pmatrix}\varepsilon_1 \\ \vdots \\ \varepsilon_n\end{pmatrix} \sim N(0, \sigma_\varepsilon^2 I_n)$

$b = (X^TX)^{-1}X^Ty = (X^TX)^{-1}X^T(X\beta+\varepsilon) = \beta + (X^TX)^{-1}X^T\varepsilon$

$\therefore b \sim N(\beta, \sigma_\varepsilon^2 (X^TX)^{-1})$

$\left(\begin{aligned} &var(\beta + (X^TX)^{-1}X^T\varepsilon) = var((X^TX)^{-1}X^T\varepsilon) \\ &= (X^TX)^{-1}X^T\, var(\varepsilon)\, X(X^TX)^{-1} = \sigma_\varepsilon^2 (X^TX)^{-1} \end{aligned} \right)$

$\therefore \hat{Y}_0 = X_0^Tb \sim N(X_0^T\beta, \sigma_\varepsilon^2 X_0^T(X^TX)^{-1}X_0)$

$\left(\begin{aligned} &E(X_0^Tb) = X_0^T \cdot E(b) = X_0^T\beta \\ &var(X_0^Tb) = X_0^T var(b) X_0 = \sigma_\varepsilon^2 X_0^T(X^TX)^{-1}X_0 \end{aligned} \right)$

$\therefore \dfrac{\hat{Y}_0 - X_0^T\beta}{\sigma_\varepsilon\sqrt{X_0^T(X^TX)^{-1}X_0}} \sim N(0,1)$ ... ①

don't know

$\because \dfrac{RSS}{\sigma_\varepsilon^2} \sim \chi^2_{n-p-1}$ and $\sqrt{\dfrac{RSS}{n-p-1}} = RSE$

$\Rightarrow \dfrac{RSS}{n-p-1} = RSE^2 \Rightarrow RSS = RSE^2(n-p-1)$

$\therefore \dfrac{RSE^2(n-p-1)}{\sigma_\varepsilon^2} \sim \chi^2_{n-p-1} \Rightarrow RSE^2 \sim \dfrac{\chi^2_{n-p-1}}{n-p-1}\sigma_\varepsilon^2$

$Q(\hat{\sigma}_\varepsilon^2)$

$\therefore ① = \dfrac{\hat{Y}_0 - X_0^T\beta}{\hat{\sigma}_\varepsilon\sqrt{X_0^T(X^TX)^{-1}X_0}} = \dfrac{\dfrac{\hat{Y}_0 - X_0^T\beta}{\sqrt{X_0^T(X^TX)^{-1}X_0}}}{\hat{\sigma}_\varepsilon}$ $\Rightarrow$

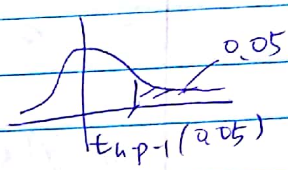
$$= \frac{N(0, \sigma_\varepsilon^2)}{\sigma_\varepsilon \sqrt{\frac{\chi^2_{n-p-1}}{n-p-1}}} \sim \frac{N(0,1)}{\sqrt{\frac{\chi^2_{n-p-1}}{n-p-1}}} \sim t_{n-p-1}$$

Q

let $t_{n-p-1}(0.05)$ is the value that

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

$P(T_{n-p-1} \geq t_{n-p-1}(0.05)) = 0.05$



$$0.9 = 90\% = P\left(\left| \frac{\hat{Y}_0 - x_0^T\beta}{\hat{\sigma}_\varepsilon \sqrt{x_0^T (X^TX)^{-1} x_0}} \right| \leq t_{n-p-1}(0.05) \right)$$

$$= P\left(-\hat{\sigma}_\varepsilon t_{n-p-1}(0.05) \cdot \sqrt{x_0^T (X^TX)^{-1} x_0} + \hat{Y}_0 \leq x_0^T\beta \leq \hat{\sigma}_\varepsilon t_{n-p-1}(0.05) \sqrt{x_0^T (X^TX) x_0} + \hat{Y}_0\right)$$

$\Rightarrow x_0^T\beta = E(Y_0)$ 90% confidence Interval is

$\hat{Y}_0 \pm \hat{\sigma}_\varepsilon t_{n-p-1}(0.05) \sqrt{x_0^T (X^TX)^{-1} x_0}$

$$\hat{Y}_0 - Y = \hat{Y}_0 - (x_0^T\beta + \varepsilon_0) = D \sim N(0, \sigma_\varepsilon^2(1 + x_0^T (X^TX)^{-1} x_0))$$

$$\frac{D}{\sigma_\varepsilon \sqrt{1 + x_0^T (X^TX)^{-1} x_0}} \sim N(0,1)$$

$$\frac{\frac{D}{\sqrt{1 + x_0^T (X^TX)^{-1} x_0}}}{\hat{\sigma}_\varepsilon} = \frac{N(0, \sigma_\varepsilon^2)}{\sigma_\varepsilon \sqrt{\frac{\chi^2_{n-p-1}}{n-p-1}}} = \frac{N(0,1)}{\sqrt{\frac{\chi^2_{n-p-1}}{n-p-1}}} = t_{n-p-1}$$

$\Rightarrow \hat{Y}_0 \pm \hat{\sigma}_\varepsilon t_{n-p-1}(0.05) \sqrt{1 + x_0^T (X^TX)^{-1} x_0}$

is CI for $Y_0 = (x_0^T\beta + \varepsilon_0)$

151AHW2

Jiyoon Clover Jeong

9/21/2017

5-(c)

From 5.2.2, we know that

Prestige = -6.7943 + 4.1866 Education + 0.0013136 Income - 0.0089052 Woman

$\sigma_{\hat{e}} = 7.846$

degrees of freedom for t-distribution : 102 - 3 - 1 = 98 (n-p-1)

$x_0^{\intercal} = (1, 13, 12000, 50)$.

```
data <- matrix(0, nrow=4, ncol=4)
colnames(data)<- c("Pres", "Educ", "Inc", "%W")
rownames(data)<- c("Pres", "Educ", "Inc", "%W")

  uppertri <- c(253618, 55326, 12513, 37748108, 8121410,
  6534383460, 131909, 32281, 14093097, 187312)

  data[upper.tri(data, diag=TRUE)] <- uppertri
  data[lower.tri(data)] <- t(data)[lower.tri(data)]

  finalsum <- matrix(c(102, 1095, 693386, 2956), nrow=4, ncol = 1)
  # we don't need prestige, replace sum of prestige as n = 102 (sum of 1*102)

  XtX <- cbind(finalsum, rbind(finalsum[-1],data[-1,-1]))
  XtXinv <- solve(XtX)



  t_0.05 <- qt(p=0.95, df=102-3-1)
  beta <- c(-6.7943, 4.1866, 0.0013136, -0.0089052)
  x_0 <- c(1, 13, 12000, 50)
  std_err <- 7.846



  var_delta <- std_err^2 * x_0 %*% XtXinv %*% x_0      # var(Y_hat0 - E(Y0))
  var_D <- std_err^2 * (1 + x_0 %*% XtXinv %*% x_0 )  # var(Y_hat0 - Y0)

  cat("delta = Y_hat0 - E(Y0)","D = Y_hat0 - Y0\n",sep = "\n")
```

```
## delta = Y_hat0 - E(Y0)
## D = Y_hat0 - Y0
```

```
  cat("var(delta) : ", var_delta, "\nvar(D): ", var_D,"\n\n")
```

```
## var(delta) :  2.754159
## var(D):  64.31387
```

```r
Y_hat0 <- x_0 %*% beta;
cat("Y_hat0 (Point estimate) : \n")
```

```
## Y_hat0 (Point estimate) :
```

```r
as.numeric(Y_hat0)
```

```
## [1] 62.94944
```

```r
cat("Confidence interval for E(Y_0)\n")
```

```
## Confidence interval for E(Y_0)
```

```r
Y_hat0 + c(-1,1) * t_0.05 * sqrt(var_delta)
```

```
## [1] 60.19365 65.70523
```

```r
cat("Confidence interval for Y_0\n")
```

```
## Confidence interval for Y_0
```

```r
Y_hat0 + c(-1,1) * t_0.05 * sqrt(var_D)
```

```
## [1] 49.63249 76.26639
```

5 - (d)

```r
x_0 <- c(1, 0, 50000, 100)
Y_hat <- as.numeric(x_0 %*% beta)
cat("Y_hat (Estimated point) :", Y_hat,"\n")
```

```
## Y_hat (Estimated point) : 57.99518
```

```r
var_delta <- std_err^2 * x_0 %*% XtXinv %*% x_0
var_D <- std_err^2 * (1 + x_0 %*% XtXinv %*% x_0 )
cat("var(delta) : ", var_delta, "\nvar(D): ", var_D)
```

```
## var(delta) :  271.7183
## var(D):  333.278
```

Estimated variance of the forecast error (Var(D) = Var(Y_hat - Y)) is 64.3 in part (c) while it is 333.278 in part(d). Since Var(D) in part(d) is almost 5 times greater than Var(D) in part(c), it suggests that error is very large and so does uncertainty. It is understandable since there is no occupation which has similar characteristics as the given data.

(given data : an occupation with an average income of $50,000, an average education of 0 years, and 100% women.)

6-(a)

```r
dat <- read.csv("/Users/cloverjiyoon/2017Fall/Stat 151A/Lab/Lab3/bodyfat.csv")
n = dim(dat)[1]
p <- 4
q <- 1

fit <- lm(bodyfat ~ Knee + Thigh + Hip + Ankle, data = dat)
RSS_M <- sum(resid(fit)^2)
```

$$H_0 : \beta_{knee} + \beta_{thigh} = \beta_{hip} + \beta_{ankle}$$

.

Since $\beta_{knee} = -\beta_{thigh} + \beta_{hip} + \beta_{ankle}$, the model can then be rewritten as

$$\text{bodyfat} = \beta_0 + \beta_{thigh}(\text{thigh} - \text{knee}) + \beta_{hip}(\text{hip} + \text{knee}) + \beta_{ankle}(\text{ankle} + \text{knee})$$

Use formula

$$\frac{(\text{RSS}(m) - \text{RSS}(M))/q}{\text{RSS}(M)/(n-p-1)}$$

where q is the number of dropped variable.

```
fit_m <- lm(bodyfat ~ I(Thigh - Knee) + I(Hip + Knee) + I(Ankle + Knee), data=dat)
RSS_m <- sum(resid(fit_m)^2)
Fstat1 <- ((RSS_m - RSS_M) /q) / (RSS_M / (n - p - 1))

cat("F statistics is : ", Fstat1,"\n")
```

```
## F statistics is :  0.002813842
```

```
cat("Degree of freedom is  (1, 247)\n")
```

```
## Degree of freedom is  (1, 247)
```

```
cat("P value is : ", 1- pf(Fstat1,1,n-p-1),"\n")
```

```
## P value is :  0.9577384
```

General way to solve problem ==

Can write hypothesis as $H_0 : L\beta = 0$, where

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

Use formula

$$\frac{(L\hat{\beta} - c)^\top [L(X^\top X)^{-1} L^\top]^{-1} (L\hat{\beta} - c)/q}{\text{RSS}(M)/(n-p-1)}.$$

```
X <- as.matrix(cbind(1, dat[,c("Knee", "Thigh", "Hip", "Ankle")]))
y <- as.numeric(dat$bodyfat)
beta_hat <- solve(t(X) %*% X) %*% t(X) %*% y
y_hat <- X %*% beta_hat
L <- matrix(c(0,1,1,-1,-1), nrow = 1)
Fstat2 <- (t(L %*% beta_hat) %*% solve(L %*% solve(t(X) %*% X) %*% t(L)) %*% (L %*% beta_hat) / q) / (s


cat("F statistics is : ", Fstat2,"\n")
```

```
## F statistics is :  0.002813842
```

```
cat("Degree of freedom is  (1, 247)\n")
```

```
## Degree of freedom is  (1, 247)
```

```
cat("P value is : ", 1- pf(Fstat2,1,n-p-1) ,"\n")
```

```
## P value is :  0.9577384
```

Check that they match.

```
Fstat1
```

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

```
Fstat2
```

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

Since P value is fairly large, we do not reject the null hypothesis.

6-(b)

```
XtXinv <- solve(t(X) %*% X)
S <- sqrt(sum(resid(fit)^2)/(n-p-1))    # RSE = sqrt(RSS/n-p-1)
var_mat <- S^2 * XtXinv
numerator <- beta_hat[2] + beta_hat[3] - beta_hat[4] - beta_hat[5]
denominator <- sqrt(sum(diag(var_mat)[2:5]) + 2*var_mat[2,3] - 2*var_mat[2,4] - 2*var_mat[2,5] - 2*var_
                      2*var_mat[4,5])



t <- as.numeric(numerator / denominator)
cat("T statistics is: ", t, "\n")
```

```
## T statistics is:  0.05304566
```

```
cat("Degree of Freedom : ", n-p-1, "\n")
```

```
## Degree of Freedom :  247
```

```
cat("P value : ", 2*(1-pt(t,247)), "\n")
```

```
## P value :  0.9577384
```

6-(c)

```
cat("Square of T statistics : ", t^2, "\n")
```

```
## Square of T statistics :  0.002813842
```

```
cat("F statistics : ", Fstat1, "\n")
```

```
## F statistics :  0.002813842
```

As we can see, square value of T statistics is equal to the F statistics value.