Homework 4 STAT 632

Winnie Lu

2022-03-12

Exercise 1

MLR: $Y = X\beta + e$ $\quad Var(e) = \sigma^2 I \quad$ I = n×n identity matrix

$\hat{Y} = X\hat{\beta} = X(X'X)^{-1}X'Y = HY$; $H = X(X'X)^{-1}X'$

(a) $HH' = HH = H$

∴ H is idempotent & symmetric so $HH = H$ & $H = H'$

∴ $HH' = HH$ $\quad$ ∴ some prove $HH = H$

$$X \underbrace{(X'X)^{-1}} \underbrace{X' \cdot X} (X'X)^{-1}X'$$

∴ identity matrix I

$$= X(X'X)^{-1}X'$$

∴ which is just equal to H.

(b) $E(\hat{Y}) = X\beta$

$Y = X\beta + e$

$E(HY) = X(X'X)^{-1}X' \cdot E(Y)$ $\quad$ ∴ $E(Y) = E(X\beta + \epsilon)$

$= $ ∴ $XX'(X'X)^{-1} = I$ $\quad = E(X\beta) + E(\epsilon)$

$= I \cdot X\beta$ $\quad = X\beta + 0$

$= X\beta$ $\quad = X\beta$

(c) $Var(\hat{Y}) = \sigma^2 H$ $\quad$ ∴ $\hat{Y} = HY$

$Var(HY) = H\,Var(Y)\,H'$ $\quad$ ∴ by linearity of Expectation: $Var(AX) = A\,Var(X)\,A'$

$Var(\hat{Y}) = \sigma^2 H$ $\quad$ ∴ H is idempotent & symmetric so $H \cdot H' = H$

∴ $Var(Y) = Var(X\beta + \epsilon) = Var(\epsilon) + \cancel{Var(X\beta)}$ $\quad$ ∴ $X\beta$ = fixed

$= Var(\epsilon)$

$= \sigma^2 I_n$

Exercise 2

Exercise 2.

$Var(\hat{B}) = \sigma^2 (X'X)^{-1}$

$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \qquad X' = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{bmatrix}$

So, $X'X = \underbrace{\begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{bmatrix}}_{2 \times n} \underbrace{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}}_{n \times 2} = \underbrace{\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}}_{2 \times 2}$

$Var(\hat{B}) = \sigma^2 (X'X)^{-1}$

$\therefore$ if $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ then, $A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

$\hookrightarrow \frac{\sigma^2}{n \sum x_i^2 - (\sum x_i)^2} \begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{bmatrix}$

$\therefore$ compare it w/ $Var(\hat{B})$: $\quad Var(\hat{B}) = \begin{bmatrix} Var(\hat{B}_0) & Cov(\hat{B}_0, \hat{B}_1) \\ Cov(\hat{B}_1, \hat{B}_0) & Var(\hat{B}_1) \end{bmatrix}$

- $$\mathrm{Var}(\hat{B}_0) = \frac{\sigma^2 \sum x_i^2}{n \sum x_i^2 - (\sum x_i)^2} = \frac{\sigma^2 \sum x_i^2}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}$$

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

$$= \frac{\sigma^2 \sum x_i^2}{n \cdot Sxx}$$

$$= \sigma^2 \left[\frac{\sum x_i^2 - n\bar{x}^2 + n\bar{x}}{n \cdot Sxx} \right]$$

$$= \sigma^2 \left[\frac{Sxx + n\bar{x}^2}{n\,Sxx} \right]$$

$$\therefore \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{Sxx} \right] \qquad \therefore \mathrm{Var}(\hat{B}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{Sxx} \right)$$

$$\therefore Sxx = \sum_{i=1}^{y} (x_i - \bar{x})$$

$$\hookrightarrow \sum x_i^2 - \sum \bar{x}^2$$

- $$\mathrm{Var}(\hat{B}_1) = \frac{\sigma^2 n}{n \sum x_i^2 - (\sum x_i)^2} \qquad \therefore \bar{x} = \frac{\sum x_i}{n}$$

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

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

$$= \frac{\sigma^2}{Sxx} \qquad \therefore \mathrm{Var}(\hat{B}_1) = \frac{\sigma^2}{Sxx}$$

Exercise 3

```
library(MASS)
head(Boston)
```

```
##      crim zn indus chas   nox    rm  age    dis rad tax ptratio  black lstat
## 1 0.00632 18  2.31    0 0.538 6.575 65.2 4.0900   1 296    15.3 396.90  4.98
## 2 0.02731  0  7.07    0 0.469 6.421 78.9 4.9671   2 242    17.8 396.90  9.14
## 3 0.02729  0  7.07    0 0.469 7.185 61.1 4.9671   2 242    17.8 392.83  4.03
## 4 0.03237  0  2.18    0 0.458 6.998 45.8 6.0622   3 222    18.7 394.63  2.94
## 5 0.06905  0  2.18    0 0.458 7.147 54.2 6.0622   3 222    18.7 396.90  5.33
## 6 0.02985  0  2.18    0 0.458 6.430 58.7 6.0622   3 222    18.7 394.12  5.21
##   medv
## 1 24.0
## 2 21.6
## 3 34.7
## 4 33.4
## 5 36.2
## 6 28.7
```

(a)

- The results are the same as the parameter estimates provided by the lm() function.

```
#response vector
Y <-matrix(Boston$medv, ncol=1)

#design matrix
X <-cbind(Intercept =1, Boston[ , c('dis', 'rm', 'tax', 'chas')])
X <-as.matrix(X)


#manually calculate least squares estimate
betahat <-solve(t(X) %*% X) %*% t(X) %*% Y
betahat
```

```
##                  [,1]
## Intercept -20.16720221
## dis        -0.10656777
## rm          7.88589232
## tax        -0.01647039
## chas        3.87901205
```

```
#compare with lm()
lm1 <-lm(medv ~ dis + rm + tax + chas, data=Boston)
coef(lm1)
```

```
##  (Intercept)          dis           rm          tax         chas
## -20.16720221  -0.10656777   7.88589232  -0.01647039   3.87901205
```

(b)

- The square root of the diagonal entries of the computed var-cov matrix is the same as the standard errors provided by the lm() function.

```
#manually calculate standard errors for least squares estimate
n <-nrow(Boston)
p <- 4
resid <- as.numeric(Y - X %*% betahat)
sigmahat2 <- sum(resid^2) / (n-p-1)
covbetahat <- sigmahat2 * solve(t(X) %*% X)
covbetahat
```

```
##              Intercept          dis           rm           tax         chas
## Intercept  8.510688175 -0.1236974635 -1.0632321031 -3.155777e-03  0.014868410
## dis       -0.123697464  0.0233402861 -0.0045240274  1.516224e-04  0.023912819
## rm        -1.063232103 -0.0045240274  0.1616761190  1.644565e-04 -0.040647554
## tax       -0.003155777  0.0001516224  0.0001644565  3.759877e-06  0.000171935
## chas       0.014868410  0.0239128186 -0.0406475538  1.719350e-04  1.151456124
```

```
#compare with lm()
vcov(lm1)
```

```
##              (Intercept)          dis           rm           tax         chas
## (Intercept)  8.510688175 -0.1236974635 -1.0632321031 -3.155777e-03  0.014868410
## dis         -0.123697464  0.0233402861 -0.0045240274  1.516224e-04  0.023912819
## rm          -1.063232103 -0.0045240274  0.1616761190  1.644565e-04 -0.040647554
## tax         -0.003155777  0.0001516224  0.0001644565  3.759877e-06  0.000171935
## chas         0.014868410  0.0239128186 -0.0406475538  1.719350e-04  1.151456124
```

```
sebetahat <- sqrt(diag(covbetahat))
sebetahat
```

```
##  Intercept        dis         rm        tax       chas
## 2.91730838 0.15277528 0.40208969 0.00193904 1.07305924
```

```
#compare with lm()

summary(lm1)$coef
```

```
##                Estimate Std. Error    t value     Pr(>|t|)
## (Intercept) -20.16720221 2.91730838 -6.9129484 1.450596e-11
## dis          -0.10656777 0.15277528 -0.6975459 4.857848e-01
## rm            7.88589232 0.40208969 19.6122719 5.663118e-64
## tax          -0.01647039 0.00193904 -8.4940923 2.286843e-16
## chas          3.87901205 1.07305924  3.6149095 3.308341e-04
```