hm4stat632

Jessica Grover

3/18/2022

#a.

library(MASS)  
attach(Boston)

Y <- matrix(medv, ncol=1)

This is response vector

Let us now design a matrix

X <- cbind(Intercept = 1, dis, rm, tax, chas)  
X <- as.matrix(X)  
X[1:4,]

## Intercept dis rm tax chas  
## [1,] 1 4.0900 6.575 296 0  
## [2,] 1 4.9671 6.421 242 0  
## [3,] 1 4.9671 7.185 242 0  
## [4,] 1 6.0622 6.998 222 0

Manually calculating mean square estimates

(betaHat <- solve(t(X) %\*% X) %\*% t(X) %\*% Y)

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

Comparing with lm()

lmBoston <- lm(medv ~ dis + rm + tax + chas, data=Boston)  
coef(lmBoston)

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

We can conclude that results are same as the parameter estimates provided by the lm() function.

#b.

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

seBetaHat <- sqrt(diag(covBetaHat))  
seBetaHat

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

#Comparing with lm()

summary(lmBoston)$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

vcov(lmBoston)

## (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

summary(lmBoston)$coef[,2]

## (Intercept) dis rm tax chas   
## 2.91730838 0.15277528 0.40208969 0.00193904 1.07305924

After checking we got that the square root of the diagonal entries of this matrix are the same as the standard errors provided by the lm() function.