Linear Models and Regression Analysis Homework 4

Yifei Dong

9/23/2020

0. Data Preparation

setwd("/Users/yifei/Documents/Teachers College/Linear Models and Regression/Week 4/hw4") getwd()

1. Flavor Deterioration

The results shown below were obtained in a small-scale experiment to study the relation between $^{\circ}F$ of storage temperature (X) and number of weeks before flavor deterioration of a food product begins to occur (Y).

Use the data from Exercises 5.4 on p.210. Using ***matrix methods***, find the following:

```
# Get data from Exercise 5.4 on p.210
mydata<-read.table(
  "http://users.stat.ufl.edu/~rrandles/sta4210/Rclassnotes/data/textdatasets/KutnerData/Chapter%20%205%
# Renames the variable
names(mydata)<-c("Y","X")
mydata
```

```
##      Y  X
## 1  7.8  8
## 2  9.0  4
## 3 10.2  0
## 4 11.0 -4
## 5 11.7 -8
```

(1). $Y'Y$

```
# Create the design matrix Y
Y<-matrix(mydata$Y,ncol = 1,nrow(mydata))
Y
```

```
##      [,1]
## [1,]  7.8
## [2,]  9.0
## [3,] 10.2
## [4,] 11.0
## [5,] 11.7
```

```
# Create the design matrix X
X<-cbind(rep(1,nrow(mydata)),mydata$X)
X
```

```
##      [,1] [,2]
```

```
## [1,]    1    8
## [2,]    1    4
## [3,]    1    0
## [4,]    1   -4
## [5,]    1   -8
```

```
# Y'Y
t(Y) %*% Y
```

```
##        [,1]
## [1,] 503.77
```

(2). $X'X$

```
# X'X
t(X) %*% X
```

```
##      [,1] [,2]
## [1,]    5    0
## [2,]    0  160
```

(3). $(X'X)^{-1}$

```
#(X'X)^(-1)
solve(t(X) %*% X)
```

```
##      [,1]    [,2]
## [1,]  0.2 0.00000
## [2,]  0.0 0.00625
```

(4). Vector $\mathbf{b}$ of estimated regression coefficients

```
# b=((X'X)^(-1))X'Y
b<-solve(t(X) %*% X) %*% t(X) %*% Y
b
```

```
##        [,1]
## [1,]  9.940
## [2,] -0.245
```

```
# Check and compare to built-in lm function
reg<-lm(mydata$Y~mydata$X)
summary(reg)
```

```
##
## Call:
## lm(formula = mydata$Y ~ mydata$X)
##
## Residuals:
##     1     2     3     4     5
## -0.18  0.04  0.26  0.08 -0.20
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.94000    0.09933  100.07  2.2e-06 ***
```

```
## mydata$X    -0.24500    0.01756  -13.95 0.000797 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2221 on 3 degrees of freedom
## Multiple R-squared:  0.9848, Adjusted R-squared:  0.9798
## F-statistic: 194.7 on 1 and 3 DF,  p-value: 0.0007971
```

(5). Vector of residuals

```
# Predicted Values: Xb
y.hat<-X %*% b
y.hat
```

```
##       [,1]
## [1,]  7.98
## [2,]  8.96
## [3,]  9.94
## [4,] 10.92
## [5,] 11.90
```

```
# Cross-check with build-in function
cbind(predict(reg),y.hat)
```

```
##    [,1]  [,2]
## 1  7.98  7.98
## 2  8.96  8.96
## 3  9.94  9.94
## 4 10.92 10.92
## 5 11.90 11.90
```

The predicted values calculated using the matrix methods are equal to the the results with build-in function

```
# Vector of Residuals
e<-Y-y.hat
e
```

```
##       [,1]
## [1,] -0.18
## [2,]  0.04
## [3,]  0.26
## [4,]  0.08
## [5,] -0.20
```

```
# Cross-check with build-in function
residuals(reg)
```

```
##     1     2     3     4     5
## -0.18  0.04  0.26  0.08 -0.20
```

```
# Sample size
n<-nrow(mydata)
# Identity matrix
I<-diag(rep(1,n))
# H matrix: X(X'X)^(-1)X'
H<-X %*% solve(t(X) %*% X) %*% t(X)
```

```r
# Matrix methods for Residuals
(I-H) %*% Y
```

```
##       [,1]
## [1,] -0.18
## [2,]  0.04
## [3,]  0.26
## [4,]  0.08
## [5,] -0.20
```

(6). SSR

```r
# J matrix
J<-matrix(rep(1,n^2),n,n)
# SSR=Y'(H-(1/n)J)Y
SSR<-t(Y) %*% (H-(J/n)) %*% Y
SSR
```

```
##       [,1]
## [1,] 9.604
```

(7). SSE

```r
# SSE=e'e=Y'Y-b'X'Y
SSE<-t(Y) %*% Y-t(b) %*% t(X) %*% Y
SSE
```

```
##       [,1]
## [1,] 0.148
```

```r
# Check with ANOVA Table
anova(reg)
```

```
## Analysis of Variance Table
##
## Response: mydata$Y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## mydata$X   1  9.604  9.6040  194.68 0.0007971 ***
## Residuals  3  0.148  0.0493
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# The results are the same.
```

(8). Cov(b)

```r
# Covariance matrix of b
# MSE
MSE<-SSE/(n-2)
MSE
```

```
##            [,1]
## [1,] 0.04933333
```

```r
# cov(b)=MSE(X'X)^(-1)
s.sq.b<-drop(MSE)*solve(t(X) %*% X)
s.sq.b
```

```
##              [,1]         [,2]
## [1,] 0.009866667 0.0000000000
## [2,] 0.000000000 0.0003083333
```

(9). Point estimate of $E\{Y_h\}$ when $X_h = -6$

```r
x.new<-c(1,-6)
y.hat.h<-t(x.new) %*% b
y.hat.h
```

```
##       [,1]
## [1,] 11.41
```

(10). Hat matrix H

```r
# H matrix: X(X'X)^(-1)X'
H<-X %*% solve(t(X) %*% X) %*% t(X)
H
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]  0.6  0.4  0.2  0.0 -0.2
## [2,]  0.4  0.3  0.2  0.1  0.0
## [3,]  0.2  0.2  0.2  0.2  0.2
## [4,]  0.0  0.1  0.2  0.3  0.4
## [5,] -0.2  0.0  0.2  0.4  0.6
```