HW5

Keith MItchell

2/27/2020

QUESTION 1: Multiple Linear Regression

1. Find the least squares estimate beta_hat:

```r
Z <- as.matrix(cbind(1, c(2,2,2,0,-1,-2,-3), c(1,-2,1,0,1,-2,1)))

Z
```

```
##      [,1] [,2] [,3]
## [1,]    1    2    1
## [2,]    1    2   -2
## [3,]    1    2    1
## [4,]    1    0    0
## [5,]    1   -1    1
## [6,]    1   -2   -2
## [7,]    1   -3    1
```

```r
y <- as.matrix(cbind(c(1,0,1,0,-1,0,-1)))
y
```

```
##      [,1]
## [1,]    1
## [2,]    0
## [3,]    1
## [4,]    0
## [5,]   -1
## [6,]    0
## [7,]   -1
```

```r
beta_hat <- solve(t(Z)%*%Z)%*%t(Z)%*%y
beta_hat
```

```
##           [,1]
## [1,] 0.0000000
## [2,] 0.3076923
## [3,] 0.0000000
```

2. Find the R^2 statistic:

```r
mean(y)
```

```
## [1] 0
```

```r
y_hat <- Z%*%beta_hat
y_hat
```

```
##            [,1]
## [1,]  0.6153846
## [2,]  0.6153846
## [3,]  0.6153846
## [4,]  0.0000000
## [5,] -0.3076923
## [6,] -0.6153846
## [7,] -0.9230769
```

```r
ess <- (y_hat-mean(y))^2
sum(ess)
```

```
## [1] 2.461538
```

```r
tss <- (y-mean(y))^2
sum(tss)
```

```
## [1] 4
```

```r
r_sq <- sum(ess)/sum(tss)
r_sq
```

```
## [1] 0.6153846
```

3. Find $\hat{\sigma^2}$ and $\hat{Cov}(\vec{\beta})$

n is the number of observed instances (or data to learn from) for each r.

This is equal to the number of rows in Z = 7

r is the number of observed variables for which each n has a value.

So this is number of columns of data in Z = 2

```r
n <- 7
r <- 2
y_hat <- Z%*%beta_hat
eps_hat <- y-y_hat
sigma_sq <- (1/(n-r-1))*(sum(eps_hat^2))
(sum(eps_hat^2))
```

```
## [1] 1.538462
```

```r
sigma_sq
```

```
## [1] 0.3846154
```

```r
cov_hat_of_beta <- sigma_sq*solve(t(Z)%*%Z)
cov_hat_of_beta
```

```
##            [,1]      [,2]       [,3]
## [1,] 0.05494505 0.0000000 0.00000000
## [2,] 0.00000000 0.0147929 0.00000000
## [3,] 0.00000000 0.0000000 0.03205128
```

```r
# TODO make sure nothing else dependent on this
```

4. Find the 95% confidence interval for β_1

```
alpha <-0.05
half_dis <- sqrt(cov_hat_of_beta[2,2])*abs(qt(alpha/2,n-r-1))
B_j_exist_upper <- beta_hat[2] + half_dis
B_j_exist_lower <- beta_hat[2] - half_dis
B_j_exist_lower
```

```
## [1] -0.02999578
```

```
B_j_exist_upper
```

```
## [1] 0.6453804
```

5. Find the 95% simultaneous confidence intervals for β_0, β_1, β_2 based on the confidence region

```
for (val in c(1,2,3))
{
  c_range <- (cov_hat_of_beta[val,val])^(1/2)*(((r+1)*(qf(1-alpha,r+1,n-r-1)))^(1/2))
  B_j_exist_upper <- beta_hat[val] + c_range
  B_j_exist_lower <- beta_hat[val] - c_range
  print (sprintf("Beta-%s", val-1))
  beta_hat[val]
  print(B_j_exist_lower)
  print(B_j_exist_upper)
}
```

```
## [1] "Beta-0"
## [1] -1.042349
## [1] 1.042349
## [1] "Beta-1"
## [1] -0.2331561
## [1] 0.8485407
## [1] "Beta-2"
## [1] -0.7961072
## [1] 0.7961072
```

6. Find the 95% simultaneous confidence intervals for β_0, β_1, β_2 based on Bonferroni Correction

```
for (val in c(1,2,3))
{
  t_range <- (cov_hat_of_beta[val,val])^(1/2)*(abs((qt((alpha)/(2*(r+1)), n-r-1))))
  B_j_exist_upper <- beta_hat[val] + t_range
  B_j_exist_lower <- beta_hat[val] - t_range
  print (sprintf("Beta-%s", val-1))
  beta_hat[val]
  print(B_j_exist_lower)
  print(B_j_exist_upper)
}
```

```
## [1] "Beta-0"
## [1] -0.9284227
## [1] 0.9284227
```

```
## [1] "Beta-1"
## [1] -0.1740426
## [1] 0.7894272
## [1] "Beta-2"
## [1] -0.7090945
## [1] 0.7090945
```

7. Test H_0: $\beta_1 = \beta_2 = 0$ at the level of α=0.05

```r
C<- as.matrix(rbind(c(0,1,0), c(0,0,1)))
new_beta_hat <- C%*%beta_hat
new_beta_hat
```

```
##            [,1]
## [1,] 0.3076923
## [2,] 0.0000000
```

```r
Y <- C%*%solve(t(Z)%*%Z)%*%t(C)
Y
```

```
##             [,1]       [,2]
## [1,] 0.03846154 0.00000000
## [2,] 0.00000000 0.08333333
```

```r
q=0
r=2
lhs <- (1/sigma_sq)*t(new_beta_hat)%*%solve(Y)%*%new_beta_hat
cval <- (r-0)*qf(1-alpha, df1=2, df2=4)
cval
```

```
## [1] 13.88854
```

```r
lhs
```

```
##      [,1]
## [1,]  6.4
```

The since the statistic is less then the critical value we fail to reject the null hypothesis.

8. Find a 95% confidence interval for the mean response $E(Y_0) = \beta_0 + \beta_1\bar{z}_1 + \beta_2\bar{z}_2$

TODO: write out the latex equation for this part

```r
z_0 <- as.matrix(colMeans(Z))
z_0_beta_hat <- t(z_0)%*%beta_hat
z_0_beta_hat
```

```
##      [,1]
## [1,]    0
```

```r
statistic<-qt(1-(0.05/2),n-r-1)
ans<-t(z_0)%*%(solve(t(Z)%*%Z))%*%z_0
interval <- sqrt(sigma_sq)*statistic*sqrt(ans)
int_low <- z_0_beta_hat - interval
int_up <- z_0_beta_hat + interval
int_low
```

```
##            [,1]
## [1,] -0.6508088
```

```
int_up
```

```
##           [,1]
## [1,] 0.6508088
```

9. Find a 95% prediction interval for the mean response Y__0 corresponding to $\bar{z}_1$, $\bar{z}_2$

TODO: write out the latex equation for this part

```
interval <- sqrt(sigma_sq)*statistic*sqrt(1+ans)
int_low <- z_0_beta_hat - interval
int_up <- z_0_beta_hat + interval
int_low
```

```
##           [,1]
## [1,] -1.840765
```

```
int_up
```

```
##          [,1]
## [1,] 1.840765
```

QUESTION 2: Multiple Linear Regression

1. Find the least squares estimate beta__hat:

```
Z <- as.matrix(cbind(1, c(15.31,15.20,16.25,14.33,14.57,17.33,
                          14.48,14.91,15.25,13.89,15.18,14.44,
                          14.87,18.63,15.20,25.76,19.05,15.37,
                          18.06,16.35),
                        c(57.3,63.8,65.4,57.0,63.8,63.2,60.2,
                          57.7,56.4,55.6,62.6,63.4,60.2,67.2,
                          57.1,89.6,68.6,60.1,66.3,65.8)))

Z
```

```
##       [,1]  [,2] [,3]
##  [1,]    1 15.31 57.3
##  [2,]    1 15.20 63.8
##  [3,]    1 16.25 65.4
##  [4,]    1 14.33 57.0
##  [5,]    1 14.57 63.8
##  [6,]    1 17.33 63.2
##  [7,]    1 14.48 60.2
##  [8,]    1 14.91 57.7
##  [9,]    1 15.25 56.4
## [10,]    1 13.89 55.6
## [11,]    1 15.18 62.6
## [12,]    1 14.44 63.4
## [13,]    1 14.87 60.2
## [14,]    1 18.63 67.2
```

```
## [15,]    1 15.20 57.1
## [16,]    1 25.76 89.6
## [17,]    1 19.05 68.6
## [18,]    1 15.37 60.1
## [19,]    1 18.06 66.3
## [20,]    1 16.35 65.8
```

```
y <- as.matrix(cbind(c(74.8,74.0,72.9,70.0,74.9,76.0,72.0,73.5,
                       74.5,73.5,71.5,71.0,78.9,86.5,68.0,102,
                       84,69,88,76)))
y
```

```
##        [,1]
##  [1,]  74.8
##  [2,]  74.0
##  [3,]  72.9
##  [4,]  70.0
##  [5,]  74.9
##  [6,]  76.0
##  [7,]  72.0
##  [8,]  73.5
##  [9,]  74.5
## [10,]  73.5
## [11,]  71.5
## [12,]  71.0
## [13,]  78.9
## [14,]  86.5
## [15,]  68.0
## [16,] 102.0
## [17,]  84.0
## [18,]  69.0
## [19,]  88.0
## [20,]  76.0
```

```
beta_hat <- solve(t(Z)%*%Z)%*%t(Z)%*%y
beta_hat
```

```
##             [,1]
## [1,] 30.96656634
## [2,]  2.63439962
## [3,]  0.04518386
```

2. Find the R^2 statistic:

```
y_hat <- Z%*%beta_hat
y_hat
```

```
##           [,1]
##  [1,]  73.88826
##  [2,]  73.89217
##  [3,]  76.73058
##  [4,]  71.29299
##  [5,]  72.23250
##  [6,]  79.47633
##  [7,]  71.83274
```

```
##  [8,]  72.85257
##  [9,]  73.68953
## [10,]  70.07060
## [11,]  73.78526
## [12,]  71.87195
## [13,]  72.86016
## [14,]  83.08179
## [15,]  73.58944
## [16,] 102.87717
## [17,]  84.25149
## [18,]  74.17284
## [19,]  81.53951
## [20,]  77.01210
```

```r
ess <- (y_hat-mean(y))^2
sum(ess)
```

```
## [1] 1032.875
```

```r
tss <- (y-mean(y))^2
sum(tss)
```

```
## [1] 1237.87
```

```r
r_sq <- sum(ess)/sum(tss)
r_sq
```

```
## [1] 0.834397
```

3. Find $\hat{\sigma}^2$ and $\hat{Cov}(\vec{\beta})$

n is the number of observed instances (or data to learn from) for each r.

This is equal to the number of rows in Z = 20

r is the number of observed variables for which each n has a value.

So this is number of columns of data in Z = 2

```r
n <- 20
r <- 2
y_hat <- Z%*%beta_hat
eps_hat <- y-y_hat
sigma_sq <- (1/(n-r-1))*(sum(eps_hat^2))
(sum(eps_hat^2))
```

```
## [1] 204.9949
```

```r
sigma_sq
```

```
## [1] 12.05853
```

```r
cov_hat_of_beta <- sigma_sq*solve(t(Z)%*%Z)
cov_hat_of_beta
```

```
##           [,1]       [,2]        [,3]
## [1,] 62.129210  3.0680449 -1.76475975
## [2,]  3.068045  0.6171653 -0.20739543
## [3,] -1.764760 -0.2073954  0.08132918
```

4. Find the 95% confidence interval for β_1

```
alpha <-0.05
B_j_exist_upper <- beta_hat[2] + (sigma_sq*cov_hat_of_beta[2,2])^(1/2)*qt(1-alpha/2,n-r-1)
B_j_exist_lower <- beta_hat[2] - (sigma_sq*cov_hat_of_beta[2,2])^(1/2)*qt(1-alpha/2,n-r-1)
B_j_exist_lower
```

```
## [1] -3.121224
```

```
B_j_exist_upper
```

```
## [1] 8.390023
```

5. Find the 95% simultaneous confidence intervals for β_0, β_1, β_2 based on the confidence region

```
for (val in c(1,2,3))
{
  c_range <- (cov_hat_of_beta[val,val])^(1/2)*(((r+1)*(qf(1-alpha,r+1,n-r-1)))^(1/2))
  B_j_exist_upper <- beta_hat[val] + c_range
  B_j_exist_lower <- beta_hat[val] - c_range
  print (sprintf("Beta-%s", val-1))
  beta_hat[val]
  print(B_j_exist_lower)
  print(B_j_exist_upper)
}
```

```
## [1] "Beta-0"
## [1] 6.556739
## [1] 55.37639
## [1] "Beta-1"
## [1] 0.2015372
## [1] 5.067262
## [1] "Beta-2"
## [1] -0.8379773
## [1] 0.9283451
```

6. Find the 95% simultaneous confidence intervals for β_0, β_1, β_2 based on Bonferroni Correction

```
for (val in c(1,2,3))
{
  t_range <- (cov_hat_of_beta[val,val])^(1/2)*(abs((qt((alpha)/(2*(r+1)), n-r-1))))
  B_j_exist_upper <- beta_hat[val] + t_range
  B_j_exist_lower <- beta_hat[val] - t_range
  print (sprintf("Beta-%s", val-1))
  beta_hat[val]
  print(B_j_exist_lower)
  print(B_j_exist_upper)
}
```

```
## [1] "Beta-0"
## [1] 10.03934
## [1] 51.89379
## [1] "Beta-1"
```

```
## [1] 0.5486385
## [1] 4.720161
## [1] "Beta-2"
## [1] -0.711975
## [1] 0.8023427
```

7. Test H_0: $\beta_1 = \beta_2 = 0$ at the level of α=0.05

```
C<- as.matrix(rbind(c(0,1,0), c(0,0,1)))
new_beta_hat <- C%*%beta_hat
new_beta_hat
```

```
##             [,1]
## [1,] 2.63439962
## [2,] 0.04518386
```

```
Y <- C%*%solve(t(Z)%*%Z)%*%t(C)
Y
```

```
##             [,1]         [,2]
## [1,]  0.05118083 -0.017199069
## [2,] -0.01719907  0.006744537
```

```
q=0
r=2
lhs <- (1/sigma_sq)*t(new_beta_hat)%*%solve(Y)%*%new_beta_hat
cval <- (r-0)*qf(1-alpha, df1=2, df2=4)
cval
```

```
## [1] 13.88854
```

```
lhs
```

```
##          [,1]
## [1,] 85.65517
```

The since the statistic is greater then the critical value we can reject the null hypothesis.

8. Find a 95% confidence interval for the mean response $E(Y_0) = \beta_0 + \beta_1\bar{z}_1 + \beta_2\bar{z}_2$

write out the latex equation for this part

```
z_0 <- as.matrix(colMeans(Z))
z_0_beta_hat <- t(z_0)%*%beta_hat
z_0_beta_hat
```

```
##       [,1]
## [1,] 76.55
```

```
statistic<-qt(1-(0.05/2),n-r-1)
ans<-t(z_0)%*%(solve(t(Z)%*%Z))%*%z_0
interval <- sqrt(sigma_sq)*statistic*sqrt(ans)
int_low <- z_0_beta_hat - interval
int_up <- z_0_beta_hat + interval
int_low
```

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

```
## [1,] 74.91176
```

```
int_up
```

```
##           [,1]
## [1,] 78.18824
```

9. Find a 95% prediction interval for the mean response Y__0 corresponding to $\bar{z_1}$, $\bar{z_2}$

TODO: write out the latex equation for this part

```
interval <- sqrt(sigma_sq)*statistic*sqrt(1+ans)
int_low <- z_0_beta_hat - interval
int_up <- z_0_beta_hat + interval
int_low
```

```
##           [,1]
## [1,] 69.04266
```

```
int_up
```

```
##           [,1]
## [1,] 84.05734
```