HW5

Keith MItchell

2/27/2020

Multiple Linear Regression

1. Find the least squares estimate beta_hat:

```
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
```

```
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
```

```
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:

```
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
ess
```

```
##            [,1]
## [1,] 0.37869822
## [2,] 0.37869822
## [3,] 0.37869822
## [4,] 0.00000000
## [5,] 0.09467456
## [6,] 0.37869822
## [7,] 0.85207101
```

```r
sum(ess)
```

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

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

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

```r
sum(tss)
```

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

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

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

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

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.

Beta_hat size = rank(Z) = r + 1 = 3

```r
n <- 7
r <- 3
sigma_sq <- (1/(n-r-1))*((norm(y - Z%*%beta_hat))^2)
sigma_sq
```

```
## [1] 2.556213
```

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

```
##           [,1]       [,2]      [,3]
## [1,] 0.3651733 0.00000000 0.0000000
## [2,] 0.0000000 0.09831589 0.0000000
## [3,] 0.0000000 0.00000000 0.2130178
```

4. Find the 95% confidence interval for β_1

```r
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)
beta_hat[2]
```

```
## [1] 0.3076923
```

```r
B_j_exist_lower
```

```
## [1] -1.287715
```

```r
B_j_exist_upper
```

```
## [1] 1.903099
```

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

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

```
## [1] "Beta-"
## [1] -5.834563
## [1] 5.834563
## [1] "Beta-"
## [1] -2.719714
## [1] 3.335098
## [1] "Beta-"
## [1] -4.456221
## [1] 4.456221
```

6. Find the 95% simultaneous confidence intervals for beta_0, beta_1, beta_2 based on Bonferroni Correction