A2

Jefferson Li, Arib Shaikh

15/02/2021

Set Seed

```
set.seed(1005057368)
```

Q1

a)

Note that

$$\hat{\beta}_1 \sim N(\beta_1, \sigma^2/Sxx)$$

We know $\beta_1 = 4, \sigma^2 = 25$.

We can calculate Sxx from the following code

```
Xs = c(4, 8, 12, 16, 20)
Sxx = 0
for (x in Xs) {
  Sxx = Sxx + (x - mean(Xs))^2
}
Sxx
```

```
## [1] 160
```

Therefore

$$\hat{\beta}_1 \sim N(4, 25/160)$$

We want to calculate,

$$
\begin{align}
P(|\hat{\beta}_1 - \beta_1| > 1) &= P(|\hat{\beta}_1 - 4| > 1) \tag{1}\\
&= P(\hat{\beta}_1 - 4 > 1) \quad or \quad P(\hat{\beta}_1 - 4 < -1) \tag{2}\\
&= P(\hat{\beta}_1 > 5) + P(\hat{\beta}_1 < 3) \qquad \text{Since they are disjoint} \tag{3}
\end{align}
$$

This probability can be calculated in r with the following code

```
prob = pnorm(3, 4, sqrt(25/Sxx)) + (1-pnorm(5, 4, sqrt(25/Sxx)))
```

Therefore,

$$P(|\hat{\beta}_1 - \beta_1| > 1) = 0.011412$$

1

b)

```r
Xs = c(4, 8, 12, 16, 20)
```

```r
errors = rnorm(5,0,5)
errors
```

```
## [1]  5.910658  3.719423 -4.960113 -1.684409  5.930103
```

```r
Ys = rep(0, 5)
for (i in 1:length(Ys)) {
    Ys[i] = 20 + 4*Xs[i] + errors[i]
}
Ys
```

```
## [1]   41.91066   55.71942   63.03989   82.31559 105.93010
```

```r
Sxy = 0
for (i in 1:length(Ys)) {
    Sxy = Sxy + (Xs[i] - mean(Xs))*(Ys[i] - mean(Ys))
}
Sxy
```

```
## [1] 618.5402
```

```r
Bhat1 = Sxy / Sxx
Bhat0 = mean(Ys) - Bhat1*mean(Xs)
Bhat1
```

```
## [1] 3.865876
```

```r
Bhat0
```

```
## [1] 23.39261
```

```r
X0 = 10
Yhat0 =  Bhat0 + Bhat1 * X0
Yhat0
```

```
## [1] 62.05138
```

```r
fit = lm(formula = Ys ~ Xs)
CI = predict(fit, data.frame(Xs=10), interval="confidence")
```

The 95% confidence interval for $E(Y0)$ when $X0 = 10$ is given by

$$[53.6645619, 70.4381973]$$

c)

```
Bhat0s = rep(0,1000)
Bhat1s = rep(0,1000)
lowers = rep(0,1000)
uppers = rep(0,1000)

Xs = c(4, 8, 12, 16, 20)

for (i in 1:length(Bhat0s)) {
  errors = rnorm(5,0,5)

  Ys = rep(0, 5)
  for (j in 1:length(Ys)) {
      Ys[j] = 20 + 4*Xs[j] + errors[j]
  }
  fit = lm(formula = Ys ~ Xs)

  Bhat0s[i] = fit$coefficients[1]
  Bhat1s[i] = fit$coefficients[2]
  CI = predict(fit, data.frame(Xs=10), interval="confidence")
  lowers[i] = CI[2]
  uppers[i] = CI[3]
 }
```
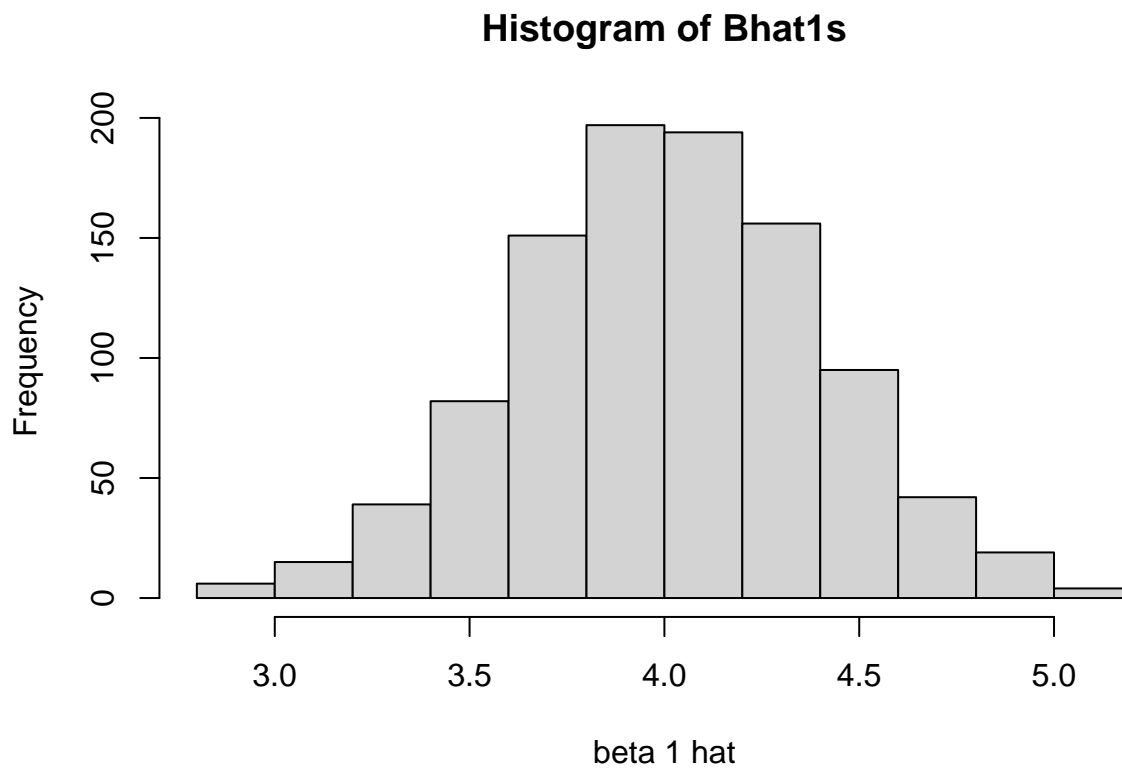
```
hist(Bhat1s, xlab = "beta 1 hat")
```


Histogram of Bhat1s
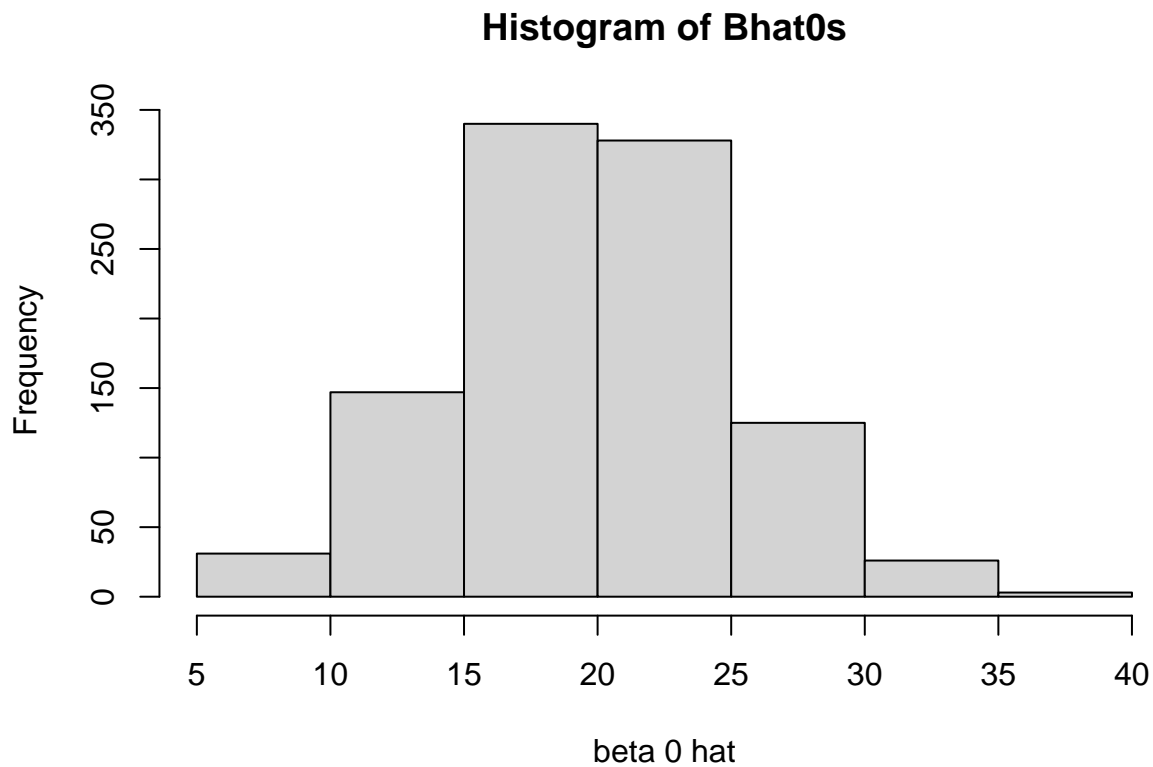
```r
mean(Bhat1s)
```

```
## [1] 4.01303
```

```r
sd(Bhat1s)
```

```
## [1] 0.389916
```

The sample mean of 4.0130298 is consistent with the theoretical mean of 4.

The sample standard deviation of 0.389916 is consistent with the theoretical standard deviation of $\sqrt{\sigma^2/Sxx} = \sqrt{25/160} = 0.3952$.

```r
hist(Bhat0s, xlab = "beta 0 hat")
```

Histogram of Bhat0s


```r
mean(Bhat0s)
```

```
## [1] 19.83268
```

```r
sd(Bhat0s)
```

```
## [1] 5.184647
```

The sample mean of 19.8326773 is consistent with the theoretical mean of 20.

The sample standard deviation of 5.1846469 is consistent with the theoretical standard deviation of $\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{Sxx}}\sigma = \sqrt{\frac{1}{5} + \frac{12^2}{160}}5 = 5.244$.

d)

```
numsWithinCI = 0

for (i in 1:length(lowers)) {
  Ey = 20 + 4*X0
  if (Ey > lowers[i] && Ey < uppers[i]){
    numsWithinCI = numsWithinCI + 1
  }
}
numsWithinCI
```

```
## [1] 951
```

This result of 951/1000 is consistent with the theoretical proportion of 950/1000

Q4

a)

1)

```
Xs = c(3,5,4,6,7)
Ys = c(4,6.5,5,7,7.5)
fit = lm(formula = Ys ~ Xs,x=TRUE)
designMatrix = matrix(model.matrix(fit), nrow = 5, ncol = 2, byrow = FALSE)
designMatrix
```

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

2)

```
coeffs = solve(t(designMatrix)%*%designMatrix)%*%t(designMatrix)%*%Ys
coeffs
```

```
##      [,1]
## [1,]  1.5
## [2,]  0.9
```

3)

```
varCovarMatrix = solve(t(designMatrix)%*%designMatrix)
varCovarMatrix
```

```
##      [,1] [,2]
## [1,]  2.7 -0.5
## [2,] -0.5  0.1
```

b)

1)

```
varCovarMatrix[1,2]
```

```
## [1] -0.5
```

2)

```
varCovarMatrix[1,1]
```

```
## [1] 2.7
```

3)

```
varCovarMatrix[2,2]
```

```
## [1] 0.1
```

c)

```
H = designMatrix%*%solve(t(designMatrix)%*%designMatrix)%*%t(designMatrix)
H
```

```
##                [,1] [,2]          [,3]          [,4]          [,5]
## [1,]  6.000000e-01  0.2 4.000000e-01 4.440892e-16 -2.000000e-01
## [2,]  2.000000e-01  0.2 2.000000e-01 2.000000e-01  2.000000e-01
## [3,]  4.000000e-01  0.2 3.000000e-01 1.000000e-01  3.330669e-16
## [4,]  4.440892e-16  0.2 1.000000e-01 3.000000e-01  4.000000e-01
## [5,] -2.000000e-01  0.2 4.440892e-16 4.000000e-01  6.000000e-01
```

d)

```
sigma2 = 4
VarE = (1 - H)*sigma2
VarE
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]  1.6  3.2  2.4  4.0  4.8
## [2,]  3.2  3.2  3.2  3.2  3.2
## [3,]  2.4  3.2  2.8  3.6  4.0
## [4,]  4.0  3.2  3.6  2.8  2.4
## [5,]  4.8  3.2  4.0  2.4  1.6
```