A2

Jefferson Li, Arib Shaikh

15/02/2021

Set Seed

```r
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

```r
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,

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

This probability can be calculated in r with the following code

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

Therefore,

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

b)

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

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
```

Since we know

$$E(Y_0) = \hat{Y_0} \sim N(\mu, \sigma^2), \text{where } \mu = \hat{\beta}_0 + \hat{\beta}_1 \cdot X_0 = 23.3926148 + 3.8658765 \cdot 10 = 62.0513796, \quad \sigma^2 = 25$$

```r
lower = qnorm(0.025, Bhat0+Bhat1 * X0, 5)
upper = qnorm(0.975, Bhat0+Bhat1 * X0, 5)
```

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

$$[52.2515597, 71.8511995]$$

c)

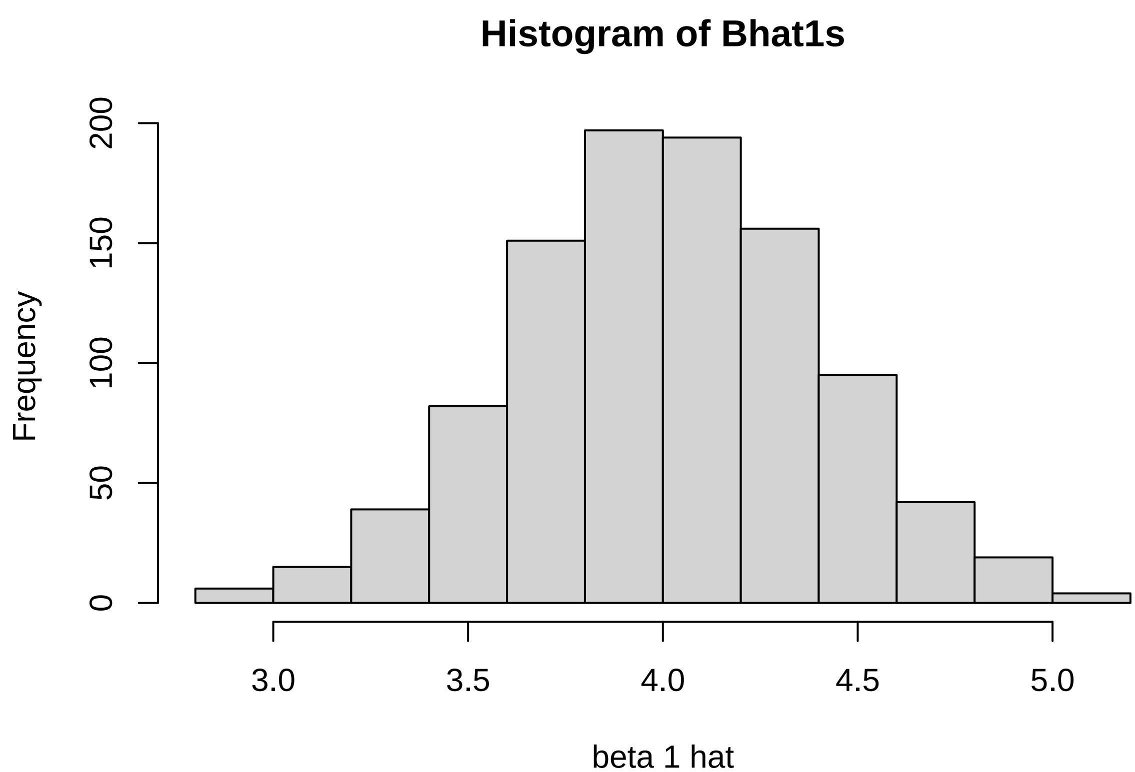
```r
Bhat0s = rep(0,1000)
Bhat1s = rep(0,1000)
Xs = c(4, 8, 12, 16, 20)

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]
}
```

```r
hist(Bhat1s, xlab = "beta 1 hat")
```




```r
mean(Bhat1s)
```

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

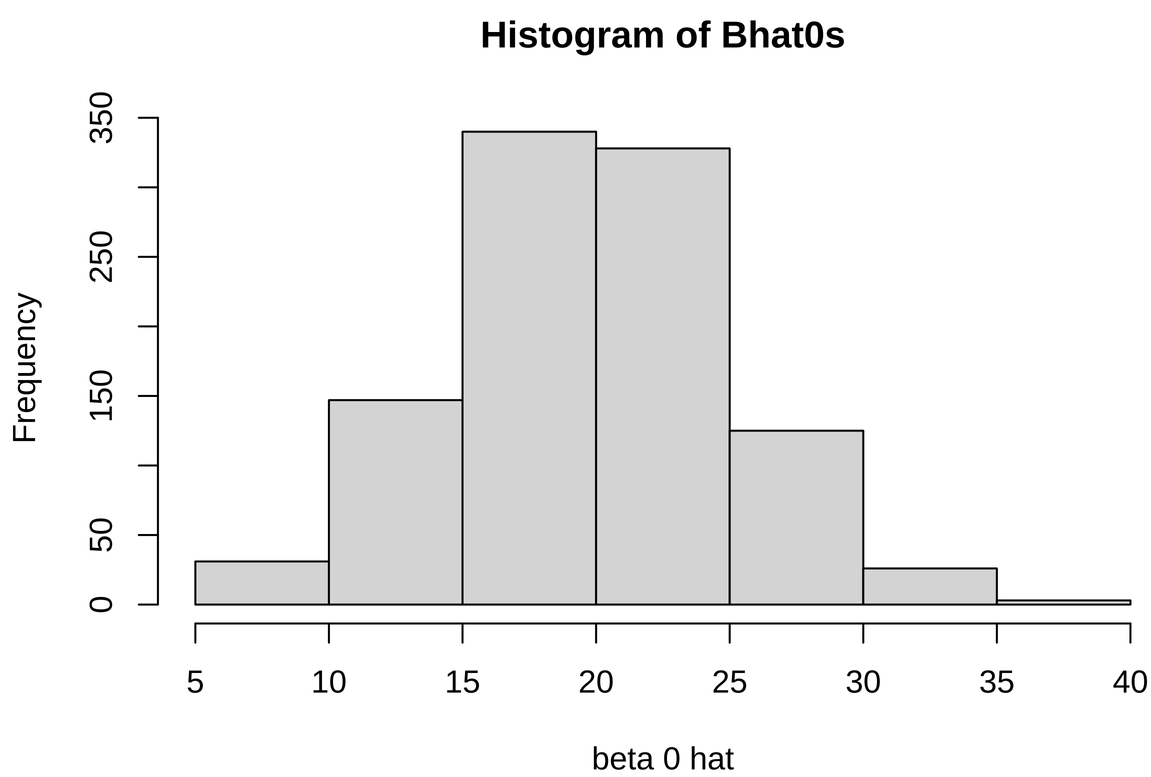
```r
sd(Bhat1s)
```

```
## [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")
```




```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)

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

$$[52.2515597, 71.8511995]$$

```r
numsWithinCI = 0

for (i in 1:length(Bhat0s)) {
  Y0 = Bhat0s[i] + Bhat1s[i] * X0

  if (Y0 > lower && Y0 < upper){
    numsWithinCI = numsWithinCI + 1
  }
  #print(Y0)
}
numsWithinCI
```

```
## [1] 1000
```

```r
Ey = 20 + 4 * X0

numsWithinCI = 0

for (i in 1:length(Bhat0s)) {
  lower = qnorm(0.025, Bhat0s[i]+ Bhat1s[i] * X0, 5)
  upper = qnorm(0.975, Bhat0s[i]+ Bhat1s[i] * X0, 5)


  if (Ey > lower && Ey < upper){
    numsWithinCI = numsWithinCI + 1
  }
}
numsWithinCI
```

```
## [1] 1000
```