STAT 5531 - LINEAR MODELS

CHAPTER 2 – ADDITIONAL EXCERCISE

Thanh Doan – Student ID 0159701

EXERCISE 2.19

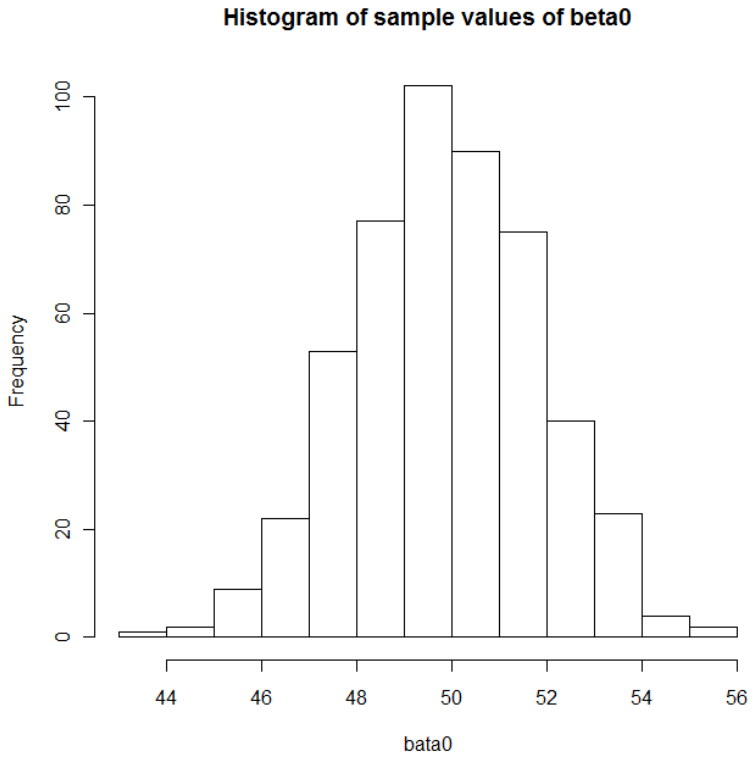
R code to generate 500 samples and compute sample estimates for question a, b, c, d

```
Generare 500 samples. Each sample has 20 data points
x = 1, 1.5,2,...,10.5 and y = 50 + 10x + e. where e ~ N(0,16)
For each sample, estimate slope, intercept, E(y|x), CIs
    then store these estimates in the estimates matrix

    n = 20;
    x = seq(1,10.5,0.5);
    estimates = matrix(0, nrow = 500, ncol=7, dimnames = list(c(1:500),
          c("beta0", "beta1", "meanY_x5", "slope.lwr", "slope.upr", "meanY.lwr", "meanY.upr")));

    for (i in c(1:500))  {
          e = rnorm(n, mean = 0, sd = 4);
          y = 50 + 10*x + e;
          sample.lm = lm(y ~ x);
          beta.hat = sample.lm$coef;
          estimates[i,1:2] = beta.hat;
          estimates[i,3] = sum(beta.hat * c(1,5));
          estimates[i,4:5] = confint(sample.lm, 'x', level=0.95);
          meanY.est = predict(sample.lm, list(x=5.00), interval='confidence', level=0.95);
          estimates[i,6:7] =  meanY.est[2:3];
    }
```

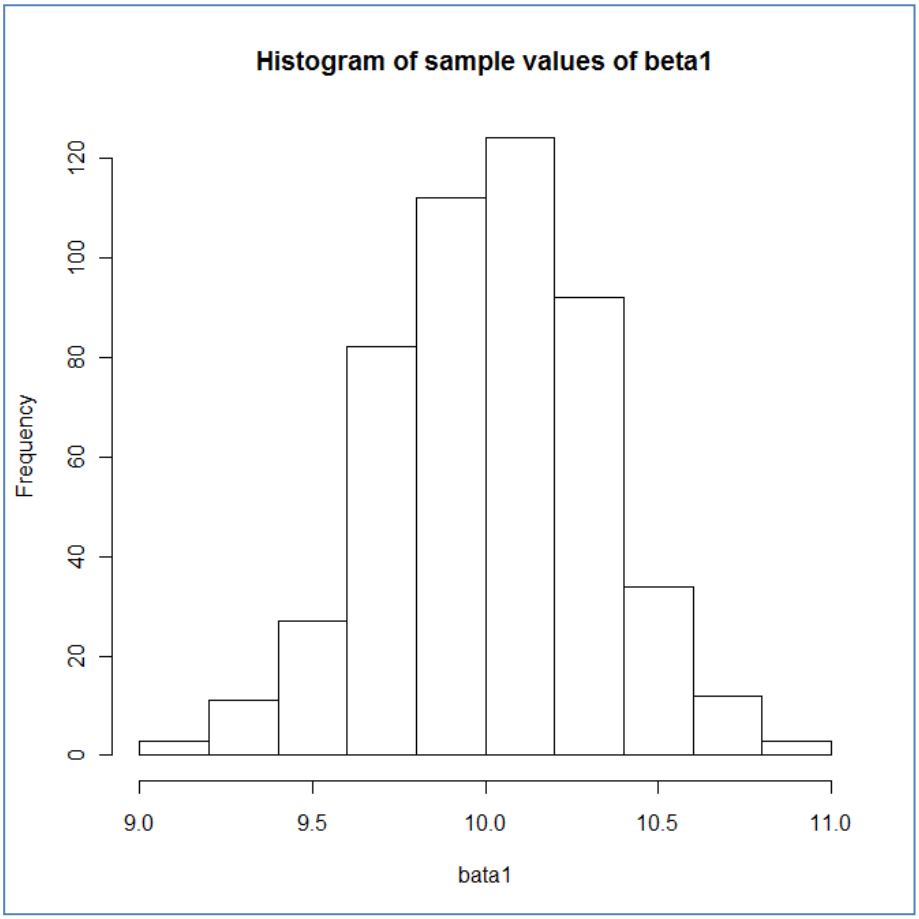
a – Histograms of sample values of beta0 and beta1



R code to draw histograms using estimate values generated by the R code in previous page

```
a) Now construct histogram of sample values of beta0.hat and beta1.hat
    beta0.hat = estimates[,1];
    hist(beta0.hat,main='Histogram of sample values of beta0', xlab='bata0');

    beta1.hat = estimates[,2];
    hist(beta1.hat,main='Histogram of sample values of beta1', xlab='bata1');
```

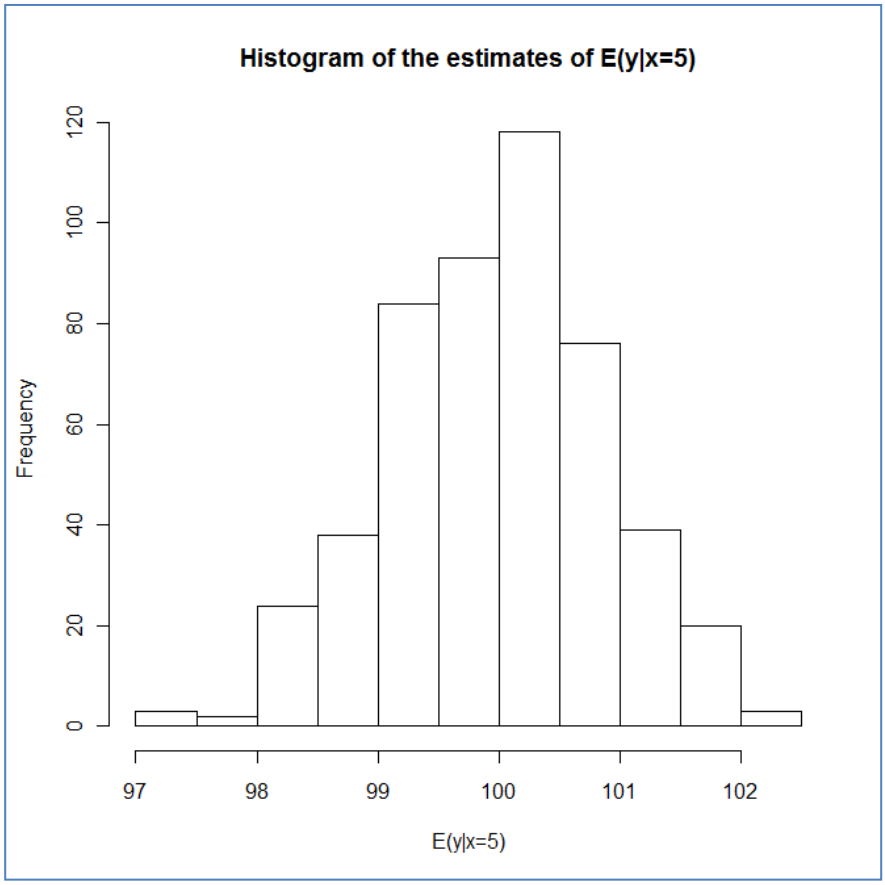


Discuss the shape of the histograms:

- The histogram for β_0 is centered around 50.
- The one for β_1 is centered around 10.
- These histograms are ***consistent*** with the model $y = 50 + 10x + \in$

b – Histogram of the 500 estimates of E(y | x = 5) from 500 samples

```
b)  Construct histogram of the estimates of E(y | x = 5)
    Ey_x5 =  estimates[,3];
    hist(Ey_x5, main='Histogram of the estimates of E(y|x=5)', xlab='E(y|x=5)');
```



Discuss the shape of the histogram:

- The histogram of the estimate values of $E\ (y \mid x = 5)$ is centered around 100.
- This histogram is ***consistent*** with the model $y = 50 + 10x + \in$

c – Compute 95% CI on the slope. How many intervals contain the true value of $\beta_1 = 10$

```
# c)  95% CIs for beta1 was computed and stored in estimates[,4:5] previously
#     Here is code to check how many contains the true value beta1 = 10
#       coverage_count count a number of CIs ontaining the true value beta1 = 10

total_samples = 500;
coverage_count = sum(estimates[,4] <= 10 & 10<= estimates[,5]);
coverage_percent = coverage_count/total_samples;
data.frame(coverage_count , total_samples, coverage_percent);
coverage_count total_samples coverage_percent
           475           500             0.95
```

Answer:

- 475 confidence intervals contain the true value of $\beta_1 = 10$
- 475 intervals out of 500 samples are 95%. It sounds like the number is too good to be true. But it is true number computed from the R code a above

d – Compute 95% CI on E(y | x=5).

How many intervals contain the true value of E(y | x=5) = 100

```
# d)  95% CIs for E(y | x=5 ) was computed and stored in estimates[,6:7] previously
#     Here is code to check how many contains the true value of  E(y | x=5 ) = 100
#       coverage_count count a number of CIs ontaining the true value

total_samples = 500;
coverage_count = sum(estimates[,6] <= 100 & 100<= estimates[,7]);
coverage_percent = coverage_count/total_samples;
data.frame(coverage_count , total_samples, coverage_percent);
coverage_count total_samples coverage_percent
           479           500            0.958
```

Answer:

- 479 intervals (95.8%) contain the true value of E(y | x=5) = 100
- This is ***consistent*** with the model $y = 50 + 10x + \in$