Linear Statistical Models

Week-8: Graded Assignment

Subjective Assignment: (Manual-grading) **Max. Marks: 45**

Note: *R is required for this assignment.*

1. Consider the Linear Model:

$$y_i = \beta_0 + \beta_1\, x_i + \epsilon_i \quad ; \quad 1 \leq i \leq 1000$$

Feel free to use the draft code available in the 'Supplementary Contents' → 'Week-8' on the portal.

(a) If $\epsilon_i \sim N(0, \sigma^2)$, then find the distribution of y_i. Elaborate on your answer. [2 Marks]

(b) Using R sample one value each for β_0, β_1 and one set of x_i as generated in code draft provided. [1 Mark]
Draft code:

```
beta_0 <- rdunif(1, 12, 5)
beta_1<- rdunif(1,10, -10)
N <-1000
x <- runif(N, -5,5)
```

(c) Using R, generate the dataset, i.e. y_i, and store it in a data frame. [2 Marks]
Draft code:

```
#Generating dataset
error=rnorm(N,0,1)
y<- beta_0 +beta_1*x+error
dataset = data.frame(x,y)
```

(d) Perform 200 iterations of the below and store the values of $\hat{\beta}_1$, R_1^2, R_0^2 and $R_1^2 - R_0^2$ computed in a data frame called estimates.

i. Sample $n = 100$ members for the data. Find the least square estimates of β_0 and β_1 for this sampled data. [2 Marks]
Draft code:

```
#Least Square Estimates
n=100
(x,y)= data[sample(1:N,n),]
```

```
x_bar = mean(x)
y_bar = mean(y)
beta_1_hat<-sum((y-y_bar)*(x-x_bar))/sum((x-x_bar)^2)
beta_0_hat<- y_bar - beta_1_hat*x_bar

#Can also use matrix approach:
a = c(replicate(n, 1))
X <- matrix(c(a, x), ncol = 2)
Y <- matrix(c(y), ncol = 1)
Betas<- inv((t(X) %*% X)) %*% t(X) %*% y
Betas
```

ii. Using R, compute the following: [3 Marks]

A. Residual Sum of squares, i.e. R_0^2

B. Total sum of squares, i.e. R_1^2

C. $R_1^2 - R_0^2$

Draft code:

```
R_1_square=sum((y-y_bar)^2)
R_0_square=sum((y-y_bar)^2)-
  (sum((y-y_bar)*(x-x_bar)))^2/sum((x-x_bar)^2)
Difference = R_1_square - R_0_square
```

iii. Find the estimate value of σ^2. And, write down the distribution of $\hat{\beta}_1$ [2 Marks]

Draft code:

```
#Estimating Variance
sigma_2_hat = R_0_square/(n-2)
sigma_2_hat
```

(e) Using R, plot the histogram of values obtained for each of $\hat{\beta}_1$, R_0^2, R_1^2 and $R_1^2 - R_0^2$. [4 Marks]

Draft code:

```
#Use 'hist()' to plot respective histograms
hist(data) #Change parameter as per need
```

(f) In each of the above plotted histogram, fit the density curve of the respective distribution to verify the following: [8 Marks]

i. $\hat{\beta}_1 \sim Normal$ distribution.

ii. $R_0^2 \sim \chi^2$ distribution with degree of freedom $(n-2)$.

iii. $R_1^2 \sim \chi^2$ distribution with degree of freedom $(n-1)$.

iv. $R_1^2 - R_0^2 \sim \chi^2$ distribution with degree of freedom 1.

Also, comment on each of the obtained histograms.

Draft code:

```
#Fitting Chi-square density curves
library(ggplot2)
ggplot(data.frame(x = c(0, 20)), aes(x = x)) +
  stat_function(fun = dchisq, args = list(df = #Enter df))
```

2. Consider the Linear Model:

$$y_i = \beta_0 + \beta_1\, x_{1i} + \beta_2\, x_{2i} + \epsilon_i \quad ; \quad 1 \le i \le 1000$$

Feel free to use the draft code available in the 'Supplementary Contents' → 'Week-8' on the portal.

(a) If $\epsilon_i \sim N(0, \sigma^2)$, then find the distribution of y_i. Elaborate on your answer. [2 Marks]

(b) Using R sample one value each for β_0, β_1, β_2 and one set of x_{1i} and x_{2i} as generated in code draft provided. [1 Mark]
Draft code:

```
beta_0 <- rdunif(1, 12, 5)
beta_1 <- rdunif(1,10, -10)
beta_2<- rdunif(1,7, 2)
N <-1000
x_1 <- runif(N, -5,5)
x_2 <- runif(N,0,10)
```

(c) Using R, generate the dataset, i.e. y_i, and store it in a data frame. [2 Marks]
Draft code:

```
#Generating dataset
error=rnorm(N,0,1)
y<- beta_0 +beta_1*x_1 + beta_2*x_2 +error
dataset = data.frame(x_1, x_2, y)
```

(d) Perform 200 iterations of the below and store the values of R_1^2, R_0^2 and $R_1^2 - R_0^2$ computed in a data frame called estimates_2.

i. Sample $n = 100$ members for the data. Find the least square estimates of β_0, β_1 and β_2 for this sampled data. [2 Marks]

ii. Using R, compute the following: [3 Marks]

A. Residual Sum of squares, i.e. R_0^2

B. Total sum of squares, i.e. R_1^2

C. $R_1^2 - R_0^2$

(e) Using R, plot the histogram of values obtained for each of R_0^2, R_1^2 and $R_1^2 - R_0^2$. [3 Marks]

(f) In each of the above plotted histogram, fit the density curve of the respective distribution to verify the following: [8 Marks]

i. $R_1^2 \sim \chi^2$ with degree of freedom $(n-1)$.

ii. $R_1^2 - R_0^2 \sim \chi^2$ with degree of freedom $= 2$.

iii. $R_0^2 \sim \chi^2$ with degree of freedom $= n - 3$.

Also, comment on each of the obtained histograms.