Moh Jaiswal
500916860
moh.jaiswal@ryerson

ECN 627 Econometrics I
Section 01/02/03/04
Fall 2020
Programming Assignment
Due Date: Nov 30, 2020

Instructions:
Please answer all questions using R-studio and paste the corresponding outputs to this word file for each question. Outputs can be found at the Console window located at the left bottom of R-studio interface. You may either directly copy and paste the outputs or take screenshots of them and paste the pictures. When pasting the outputs, please make sure you have included the commands and **ONLY** the commands that produce these outputs.

Upload the filled word file to D2L – Assessments – Programming Assignment.

Here are two sample questions and their solutions:

Sample Question 1: How much is the logarithm of 10?

Solution 1 (screenshots):
Output:

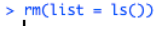
```
> log(10)
[1] 2.302585
```

Solution 2 (copy and paste):
Output:

```
> log(10)
[1] 2.302585
```

Sample Question 2: Clear the memory.

Solution 1 (screenshots):
Output:



Solution 2 (copy and paste):
Output:

```
> rm(list = ls())
```

Note that in both outputs, we have lines staring with ">" (i.e. ">log(10)" and ">rm(list=ls())"). Such lines correspond to your R codes that generate the results. Please make sure such commands are included in your answer for each question. Some questions may need multiple lines of commands. Do not include unnecessary lines.

Grading Policy: Each question from 1 to 12 is worth 2 marks. You get 1 mark for correct commands and 1 mark for correct results. Question 13 is worth 6 marks. You get 3 marks for correct codes and 3 marks for correct graph. The total is 30 marks. The teaching assistant, Hari, will grade the programming assignment and determine how to award partial credits.

Here is the assignment question:

Background: We want to predict the effect of temperature on the burned area of forest fires in the northeast region of Portugal. We obtain a dataset, "forestfires.csv", with the following variables.

1. X - x-axis spatial coordinate within the Montesinho park map: 1 to 9
2. Y - y-axis spatial coordinate within the Montesinho park map: 2 to 9
3. month - month of the year: 'jan' to 'dec'
4. day - day of the week: 'mon' to 'sun'
5. FFMC - FFMC index from the FWI system: 18.7 to 96.20
6. DMC - DMC index from the FWI system: 1.1 to 291.3
7. DC - DC index from the FWI system: 7.9 to 860.6
8. ISI - ISI index from the FWI system: 0.0 to 56.10
9. temp - temperature in Celsius degrees: 2.2 to 33.30
10. RH - relative humidity in %: 15.0 to 100
11. wind - wind speed in km/h: 0.40 to 9.40
12. rain - outside rain in mm/m2 : 0.0 to 6.4
13. area - the burned area of the forest (in ha): 0.00 to 1090.84

Please download the dataset and address the following questions.

Question 1: Clear memory, set the directory as the folder where you save the dataset "forestfires.csv", and check if the directory is correctly specified.
Output:

```
> rm(list = ls())
> setwd("~/Desktop/R")
> dir()
[1] "forestfires.csv"                "Programming Assignment.docx"
```

Question 2: Load dataset using the "read.csv" command.
Output:

```
> data1 = read.csv("~/Desktop/R/forestfires.csv")
> data1
    X Y month day FFMC   DMC    DC  ISI temp RH wind rain area
1   7 5   mar fri 86.2  26.2  94.3  5.1  8.2 51  6.7  0.0    0
2   7 4   oct tue 90.6  35.4 669.1  6.7 18.0 33  0.9  0.0    0
3   7 4   oct sat 90.6  43.7 686.9  6.7 14.6 33  1.3  0.0    0
4   8 6   mar fri 91.7  33.3  77.5  9.0  8.3 97  4.0  0.2    0
5   8 6   mar sun 89.3  51.3 102.2  9.6 11.4 99  1.8  0.0    0
6   8 6   aug sun 92.3  85.3 488.0 14.7 22.2 29  5.4  0.0    0
7   8 6   aug mon 92.3  88.9 495.6  8.5 24.1 27  3.1  0.0    0
[illegible]
12  [illegible]  13.0 22.6 [illegible]  0.0
13  6 5   aug fri 63.5  70.8 665.3  0.8 17.0 72  6.7  0.0    0
14  6 5   sep mon 90.9 126.5 686.5  7.0 21.3 42  2.2  0.0    0
```

Question 3: Calculate the mean of variable "area".

Output:

```
> mean(data1$area)
[1] 12.84729
```

Question 4: Calculate the variance of variable "area".

Output:

```
> var(data1$area)
[1] 4052.063
```

Question 5: Generate a new variable "log_area", as the logarithm of ("area" + 0.000001).

Output:

```
> log_area1 = log10(data1$area + 0.000001)
> log_area1
  [1] -6.000000000 -6.000000000 -6.000000000
  [9] -6.000000000 -6.000000000 -6.000000000
 [17] -6.000000000 -6.000000000 -6.000000000
 [25] -6.000000000 -6.000000000 -6.000000000
```

Question 6: Regress "log_area" on "temp" and summarize the regression results.

Output:

```
> ols1 = lm(log10(data1$area + 0.000001) ~ data1$temp)
> ols1

Call:
lm(formula = log10(data1$area + 1e-06) ~ data1$temp)

Coefficients:
(Intercept)   data1$temp
   -3.24807      0.04235
```

```
> summary(ols1)

Call:
lm(formula = log10(data1$area + 1e-06) ~ data1$temp)

Residuals:
    Min      1Q  Median      3Q     Max
-4.124 -3.523  2.041  3.222  5.223

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.24807    0.51363  -6.324 5.54e-10 ***
data1$temp   0.04235    0.02599   1.629    0.104
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.429 on 515 degrees of freedom
Multiple R-squared:  0.005128,  Adjusted R-squared:  0.003196
F-statistic: 2.654 on 1 and 515 DF,  p-value: 0.1039
```

Question 7: Regress "log_area" on "temp" , "rain", "X" and "Y".

Output:

```
> ols2 = lm(log10(data1$area + 0.000001) ~ data1$temp + data1$rain + data1$X + data1$Y)
> ols2

Call:
lm(formula = log10(data1$area + 1e-06) ~ data1$temp + data1$rain +
    data1$X + data1$Y)

Coefficients:
(Intercept)   data1$temp   data1$rain      data1$X      data1$Y
   -3.97615      0.04359      0.20291      0.07566      0.08071
```

Question 8: List all the coefficients of the second regression.

Output:

28 :

```
> regstats$coefficients
                Estimate Std. Error    t value     Pr(>|t|)
(Intercept) -3.97614606 0.74794222 -5.3161140 1.585690e-07
data1$temp   0.04358840 0.02610193  1.6699299 9.554443e-02
data1$rain   0.20291468 0.51253232  0.3959061 6.923392e-01
data1$X      0.07565782 0.07770952  0.9735978 3.307160e-01
data1$Y      0.08070613 0.14578220  0.5536076 5.800891e-01
```

Question 9:

```
> summary(ols2)

Call:
lm(formula = log10(data1$area + 1e-06) ~ data1$temp + data1$rain +
    data1$X + data1$Y)

Residuals:
   Min     1Q Median     3Q    Max
-4.521 -3.472  1.865  3.273  5.248

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.97615    0.74794  -5.316 1.59e-07 ***
data1$temp   0.04359    0.02610   1.670   0.0955 .
```

Question 10: Display the standard error of the OLS estimate of the coefficient of "temp" in the second regression.

Output:

```
> summary(ols2)

Call:
lm(formula = log10(data1$area + 1e-06) ~ data1$temp + data1$rain +
    data1$X + data1$Y)

Residuals:
   Min     1Q Median     3Q    Max
-4.521 -3.472  1.865  3.273  5.248

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.97615    0.74794  -5.316 1.59e-07 ***
data1$temp   0.04359    0.02610   1.670   0.0955 .
```

Question 11: Compute and display the value of the t-statistic of

H0: the coefficient "temp" in the second regression is 0.1

H1: the coefficient "temp" in the second regression is not 0.1

Output:

```
> betah = regstats$coefficients[2]
> betah
[1] 0.0435884
> betah0 = 0.1
> sigbetah = regstats$coefficients[7]
> sigbetah
[1] 0.02610193
> tstat = (betah - betah0)/sigbetah
> tstat
[1] -2.161204
```

Question 12: Compute and display the p-value of the hypothesis testing in question 12.

Output:

```
> pval = pnorm(-abs(tstat))*2
> pval
[1] 0.03067958
```

Question 13: Plot the "log_area" against "temp" and add a regression line of "log_area" on "temp" (or, equivalently, the regression line in question 6). Put "log_area" at the vertical axis and the "temp" at the horizontal axis. In the output area below, you need to display both the results and the graph.

Output:

```
> plot(y = log_area1, x = data1$temp)
>
> ols1 = lm(log10(data1$area + 0.000001) ~ data1$temp)
> ols1

Call:
lm(formula = log10(data1$area + 1e-06) ~ data1$temp)

Coefficients:
(Intercept)   data1$temp
   -3.24807      0.04235

> abline(lm(log10(data1$area + 0.000001) ~ data1$temp))
```