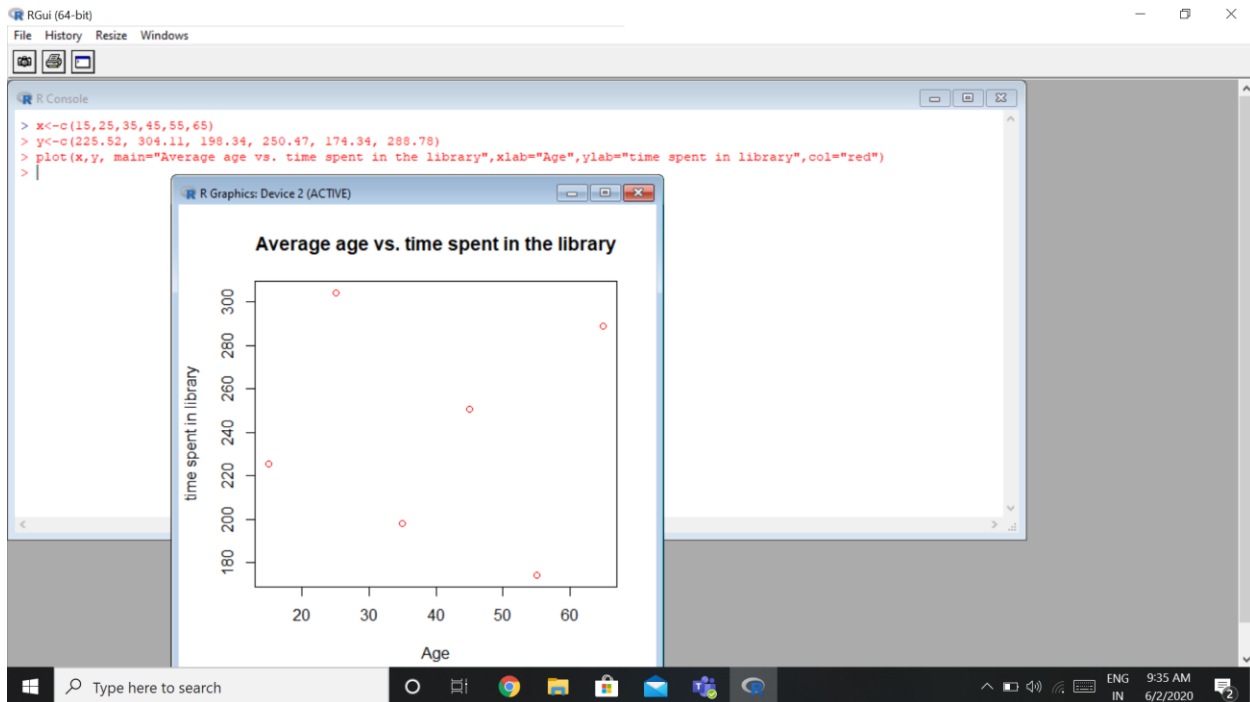


Statistics DA 2

Name: Kulvir Singh

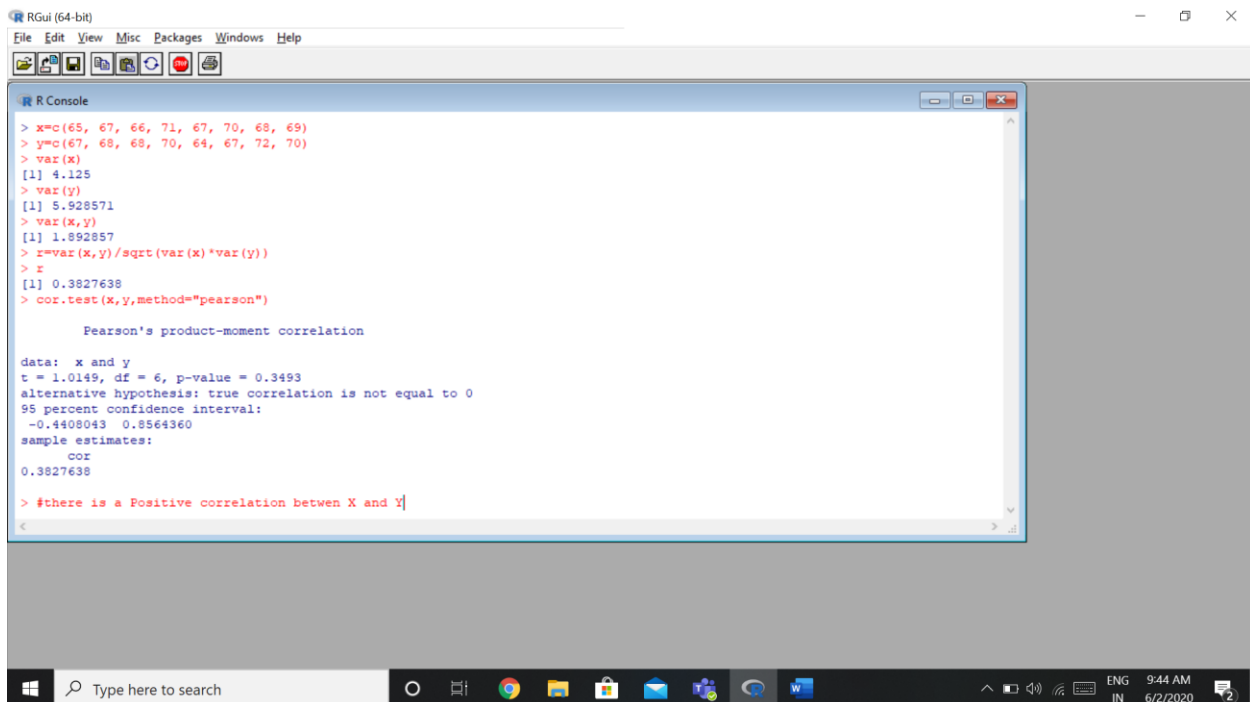
Reg No. : 19BCE2074

Q1. Illustrate the relationship between the average age versus the time spent in the library, by using scatterplot: Age Group 10 – 19 20 – 29 30 – 39 40 – 49 50 – 59 60 – 69 Representative age 15 25 35 45 55 65 Hours spend in the library 225.52 304.11 198.34 250.47 174.34 288.78



Q2. Compute the coefficients of correlation between X and Y using the following data: X 65 67 66 71 67 70 68 69

Y 67 68 68 70 64 67 72 70



The screenshot shows the RGui (64-bit) interface. The R Console window displays the following code and output:

```
> x=c(65, 67, 66, 71, 67, 70, 68, 69)
> y=c(67, 68, 68, 70, 64, 67, 72, 70)
> var(x)
[1] 4.125
> var(y)
[1] 5.928571
> var(x,y)
[1] 1.892857
> r=var(x,y)/sqrt(var(x)*var(y))
> r
[1] 0.3827638
> cor.test(x,y,method="pearson")

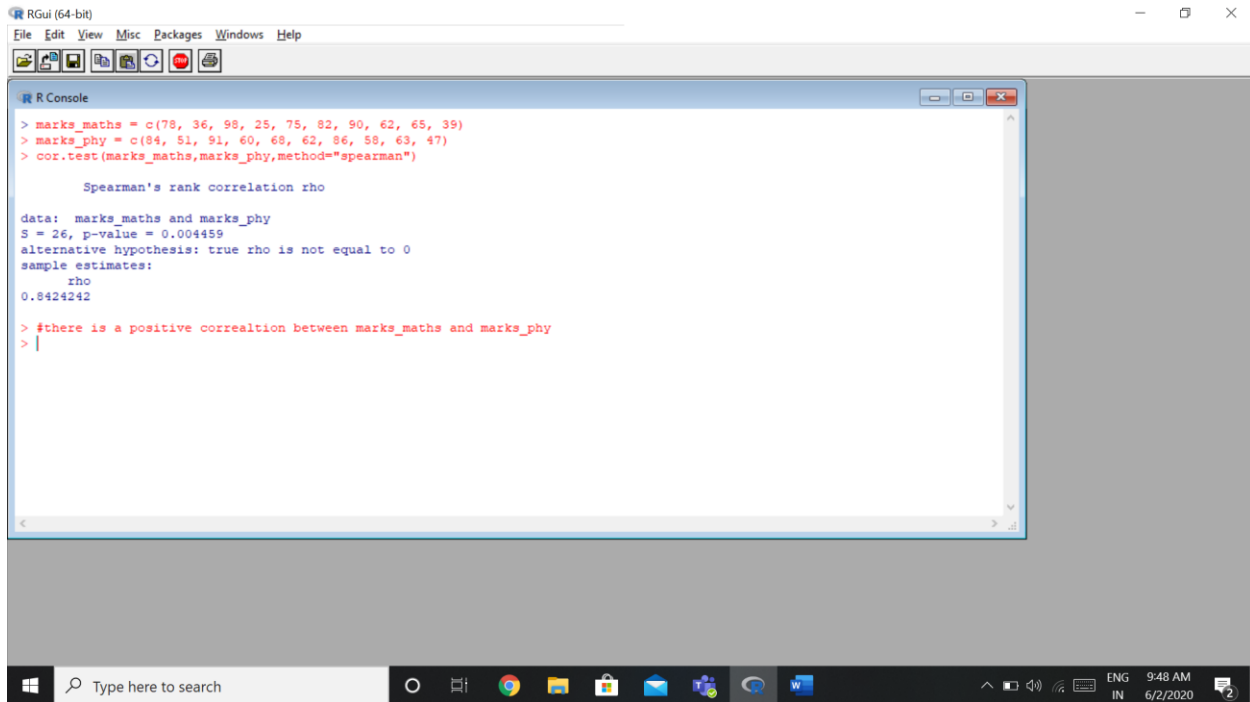
Pearson's product-moment correlation

data: x and y
t = 1.0149, df = 6, p-value = 0.3493
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
-0.4408043  0.8564360
sample estimates:
cor
0.3827638

> #there is a Positive correlation between X and Y
```

The Windows taskbar at the bottom shows the search bar, taskbar icons, and system tray with the date 6/2/2020 and time 9:44 AM.

Q3. Ten students got the following percentage of marks in Mathematics and Physical Sciences:
Students 1 2 3 4 5 6 7 8 9 10 Marks in Mathematics 78 36 98 25 75 82 90 62 65 39 Marks in Physical
Sciences 84 51 91 60 68 62 86 58 63 47 Calculate the rank correlation coefficient.



```
RGui (64-bit)
File Edit View Misc Packages Windows Help

R Console
> marks_maths = c(78, 36, 98, 25, 75, 82, 90, 62, 65, 39)
> marks_phy = c(84, 51, 91, 60, 68, 62, 86, 58, 63, 47)
> cor.test(marks_maths, marks_phy, method="spearman")

Spearman's rank correlation rho

data: marks_maths and marks_phy
S = 26, p-value = 0.004459
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.8424242

> #there is a positive correaltion between marks_maths and marks_phy
> |
```

Q4. The following data refers to the daily sales of tomatoes (in kg) at different prices (in Rs) observed on different days in a market. Let us carry out linear regression analysis for this data: Price 4.5 5.5 4.5 4.5 4 5.5 5.5 6.5 5 5.5 6 4.5 Quantity Sold 125 115 140 140 150 150 130 120 130 100 105 150

```

RGui (64-bit) - [R Console]
File Edit View Misc Packages Windows Help

> x=c(4.5, 5.5, 4.5, 4.5, 4, 5.5, 5.5, 6.5, 5, 5.5, 6, 4.5)
> y=c(125, 115, 140, 140, 150, 150, 130, 120, 130, 100, 105, 150)
> fit = lm(y~x)
> fit

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)          x
      205.14       -14.74

> summary(fit)

Call:
lm(formula = y ~ x)

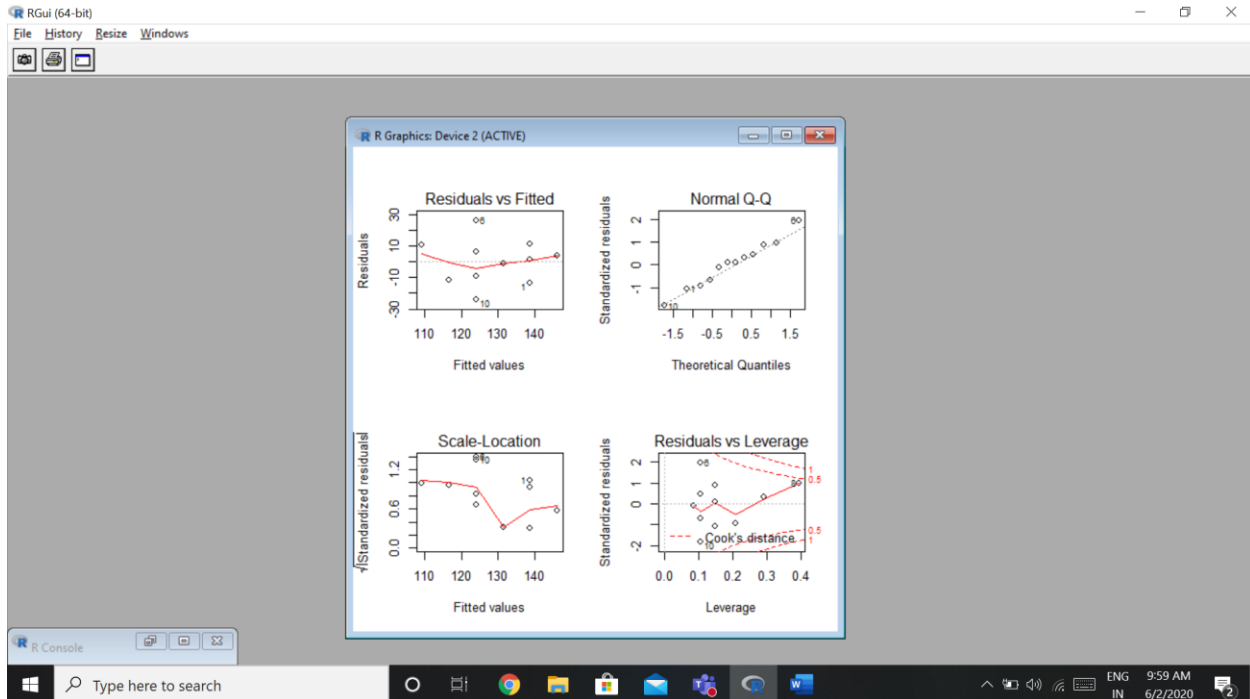
Residuals:
    Min       1Q   Median       3Q      Max
-24.055  -9.712   1.203   7.131  25.945

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  205.137    29.384    6.981 3.8e-05 ***
x           -14.742     5.679   -2.596  0.0267 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.98 on 10 degrees of freedom
Multiple R-squared:  0.4026,    Adjusted R-squared:  0.3428
F-statistic: 6.738 on 1 and 10 DF, p-value: 0.02668

> par(mfrow=c(2,2));
> plot(fit)
> par(mfrow=c(1,1))

```



Q5. The following data represent the chemistry grades for a random sample of 12 freshmen at a certain college along with their scores on an intelligence test administered while they were still seniors in high school. The number of class periods is also given: Students 1 2 3 4 5 6 7 8 9 10 11 12 Chemistry Grade (Y) 85 74 76 90 85 87 94 98 81 91 76 74 Test Score (X1) 65 50 55 65 55 70 65 70 55 70 50 55 50 55 Classes Missed (X2) 1 7 5 2 6 3 2 5 4 3 1 4 Fit a multiple linear regression equation of the form $\hat{Y} = b_0 + b_1X_1 + b_2X_2$

```

RGui (64-bit) - [R Console]
File Edit View Misc Packages Windows Help

> y=c(85, 74, 76, 90, 85, 87, 94, 98, 81, 91, 76, 74)
> x1=c(65, 50, 55, 65, 55, 70, 65, 70, 55, 70, 50, 55)
> x2=c(1, 7, 5, 2, 6, 3, 2, 5, 4, 3, 1, 4)
> input_data=data.frame(y,x1,x2)
> input_data
  y x1 x2
1 85 65 1
2 74 50 7
3 76 55 5
4 90 65 2
5 85 55 6
6 87 70 3
7 94 65 2
8 98 70 5
9 81 55 4
10 91 70 3
11 76 50 1
12 74 55 4
> RegModel<-lm(y~x1+x2,data=input_data)
> RegModel

Call:
lm(formula = y ~ x1 + x2, data = input_data)

Coefficients:
(Intercept)          x1          x2
    27.5467      0.9217      0.2842

> summary(RegModel)

Call:
lm(formula = y ~ x1 + x2, data = input_data)

Residuals:
    Min       1Q   Median       3Q      Max
-5.9169 -2.9701  0.0018  2.6925  5.9757

Coefficients:
(Intercept)          x1          x2
    27.5467      0.9217      0.2842

t-values Pr(>|t|)
(Intercept)  2.204 0.054975
x1          4.960 0.000781 ***
x2          0.377 0.714772

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.517 on 9 degrees of freedom
Multiple R-squared:  0.7478,    Adjusted R-squared:  0.6918
F-statistic: 13.34 on 2 and 9 DF,  p-value: 0.002031

> #Now the regression model is Y = 27.5467 + 0.9217*X1 + 0.2842*X2

```

```

Residuals:
    Min       1Q   Median       3Q      Max
-5.9169 -2.9701  0.0018  2.6925  5.9757

Coefficients:
(Intercept)          x1          x2
    27.5467      0.9217      0.2842

t-values Pr(>|t|)
(Intercept)  2.204 0.054975
x1          4.960 0.000781 ***
x2          0.377 0.714772

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.517 on 9 degrees of freedom
Multiple R-squared:  0.7478,    Adjusted R-squared:  0.6918
F-statistic: 13.34 on 2 and 9 DF,  p-value: 0.002031

> #Now the regression model is Y = 27.5467 + 0.9217*X1 + 0.2842*X2

```