NAME: SPARSH ARYA

REG NUMBER: 17BEC0656

SLOT:E2

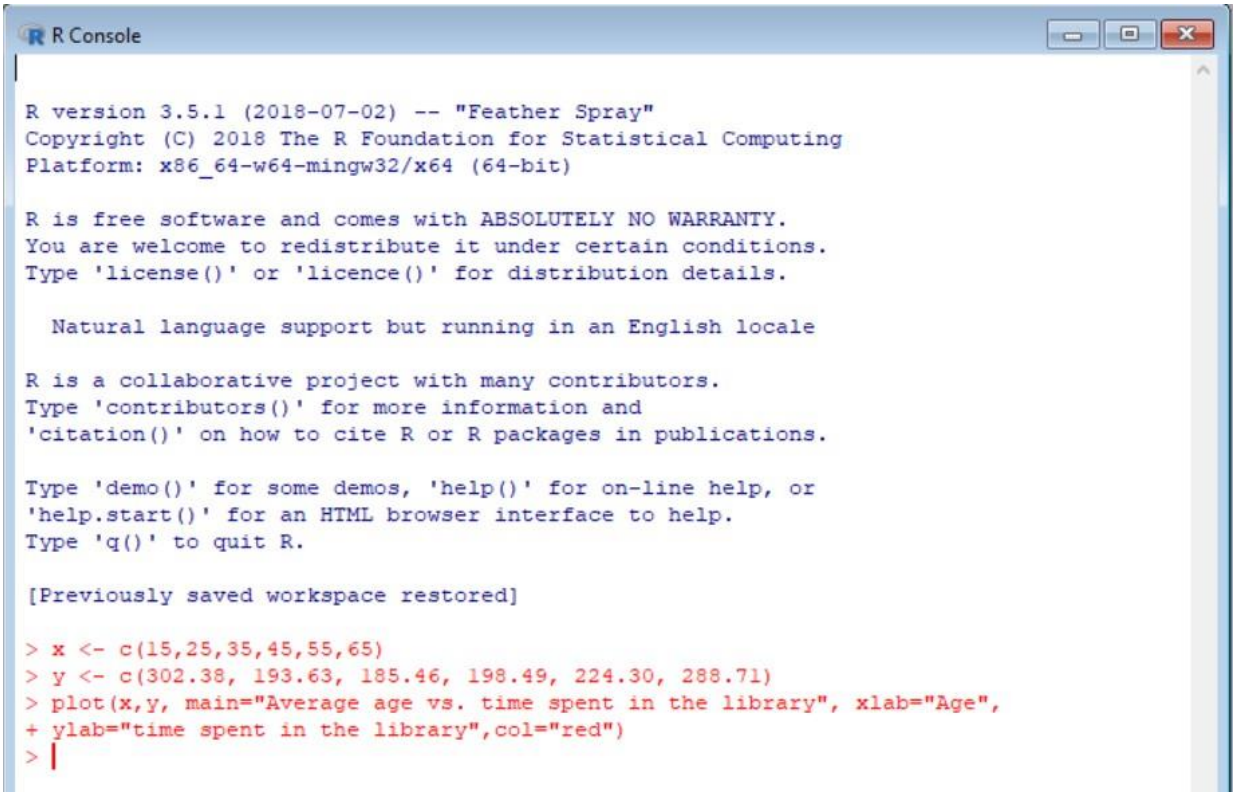
Measures of Correlation:

Scatter Diagram:

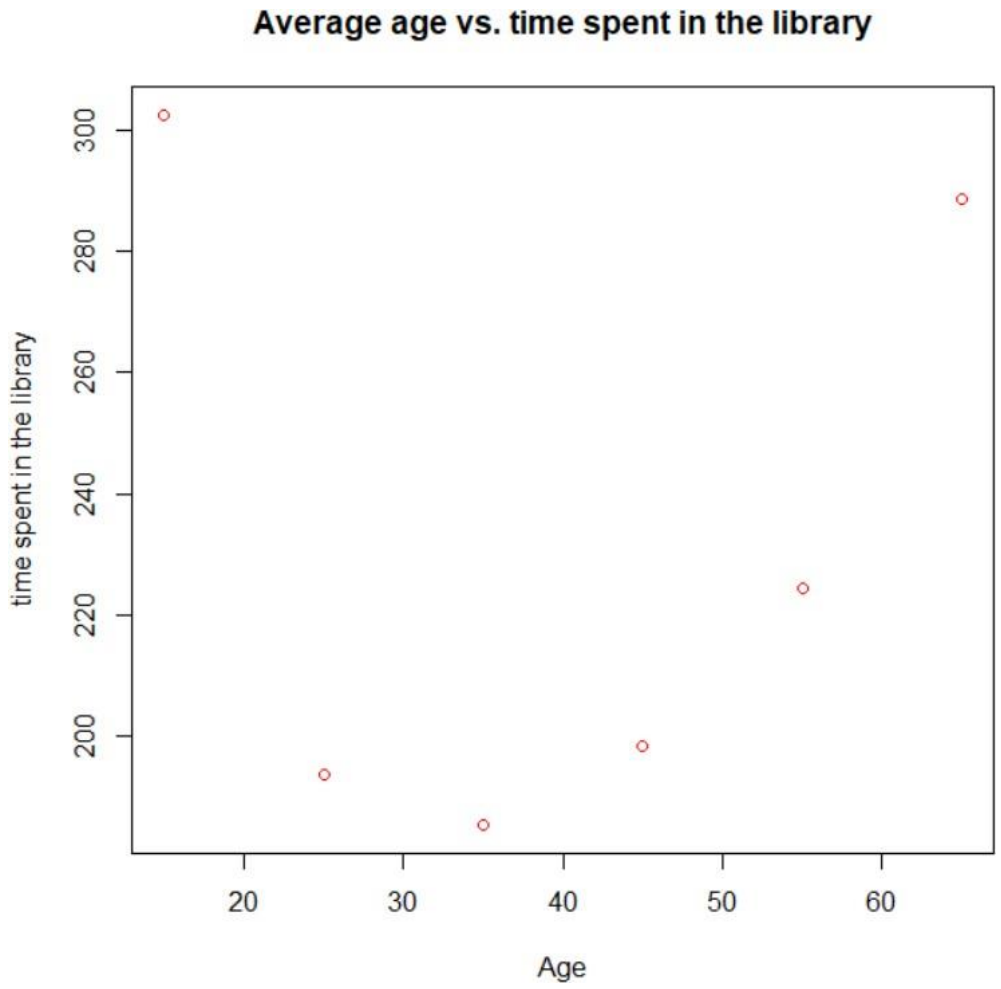
Problem 1-:

AGE GROUP	REPRESENTATIVE AGE	HOURS SPEND IN THE LOCAL LIBRARY
10-19	15	302.38
20-29	25	193.63
30-39	35	185.46
40-49	45	198.49
50-59	55	224.30
60-69	65	288.71

R code-


R Console
R version 3.5.1 (2018-07-02) -- "Feather Spray"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
[Previously saved workspace restored]
> x <- c(15,25,35,45,55,65)
> y <- c(302.38, 193.63, 185.46, 198.49, 224.30, 288.71)
> plot(x,y, main="Average age vs. time spent in the library", xlab="Age",
+ ylab="time spent in the library",col="red")
>

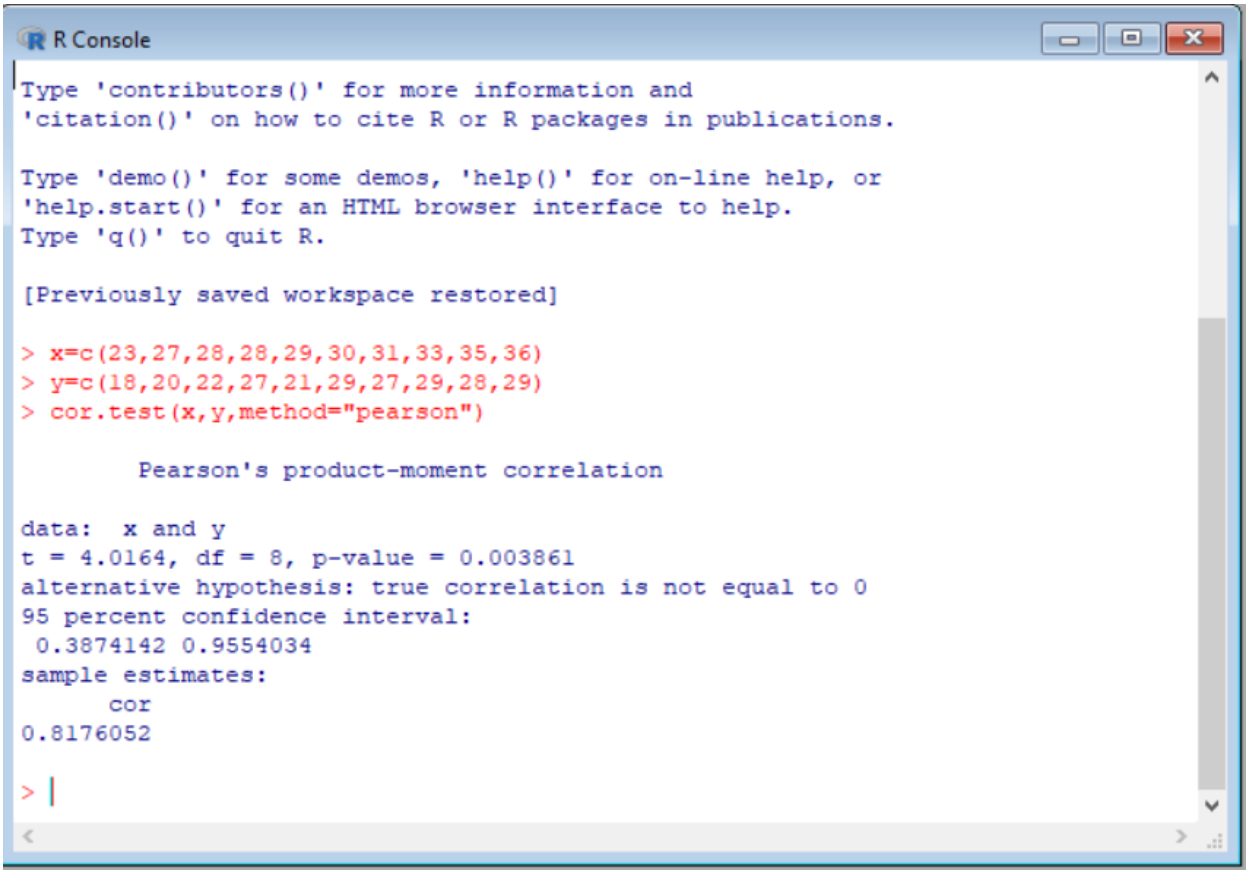
Output-:



KARL PEARSON'S COEFFICIENT OF CORRELATION

Problem2: Consider a set of 'n' pairs of observations (X1, Y1), (X2, Y2), … (Xn, Yn) on two variables X and Y. Then we have, Covariance between X and Y.

R code-



There is a Positive correlation between X and Y

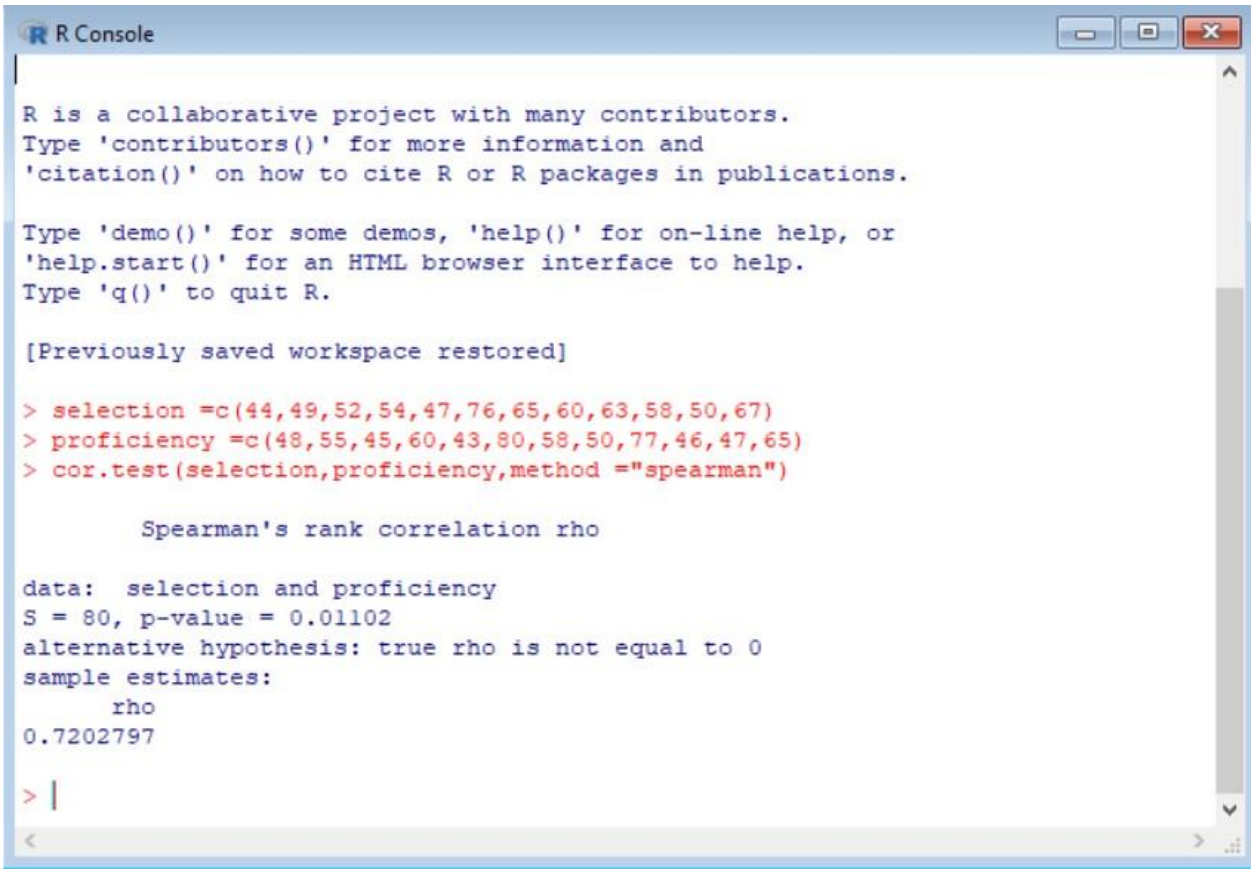
SPEARMAN'S RANK CORRELATION COEFFICIENT

Problem3 : Twelve recruits were subjected to selection test to ascertain their suitability for a certain course of training. At the end of training they were given a proficiency test. The marks scored by the recruits are recorded below :

Recruit	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
Selection Test Score	*44*	*49*	*52*	*54*	*47*	*76*	*65*	*60*	*63*	*58*	*50*	*67*
Proficiency Test Scrore	*48*	*55*	*45*	*60*	*43*	*80*	*58*	*50*	*77*	*46*	*47*	*65*

Calculate rank correlation coefficient and comment on your result.

R code-



There is a positive correlation between selection and Proficiency

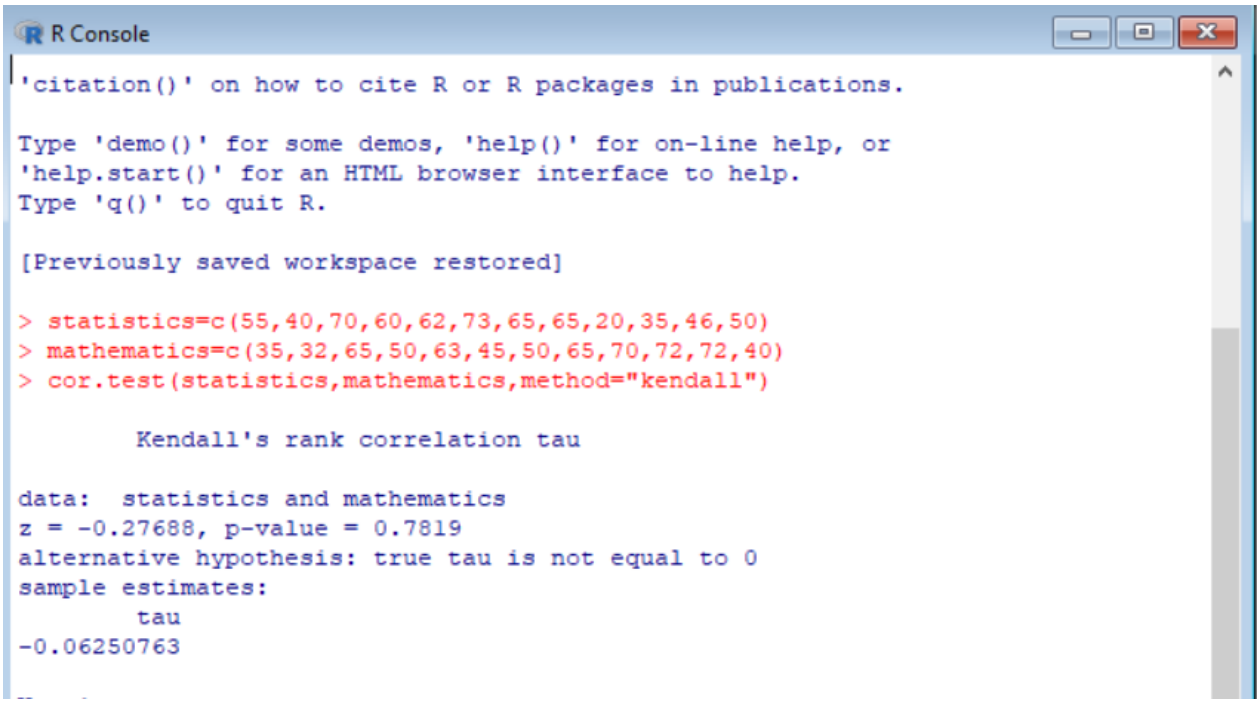
KENDALL'S COEFFICIENT OF CONCURRENT DEVIATIONS

Problem4: The following data gives the marks obtained by 12 students in statistics and computer science :

Students		*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
Marks	*Statistics*	*55*	*40*	*70*	*60*	*62*	*73*	*65*	*65*	*20*	*35*	*46*	*50*
	Computer Science	*35*	*32*	*65*	*50*	*63*	*45*	*50*	*65*	*70*	*72*	*72*	*40*

Compute the coefficient of correlation by the method of concurrent deviations.

R code:



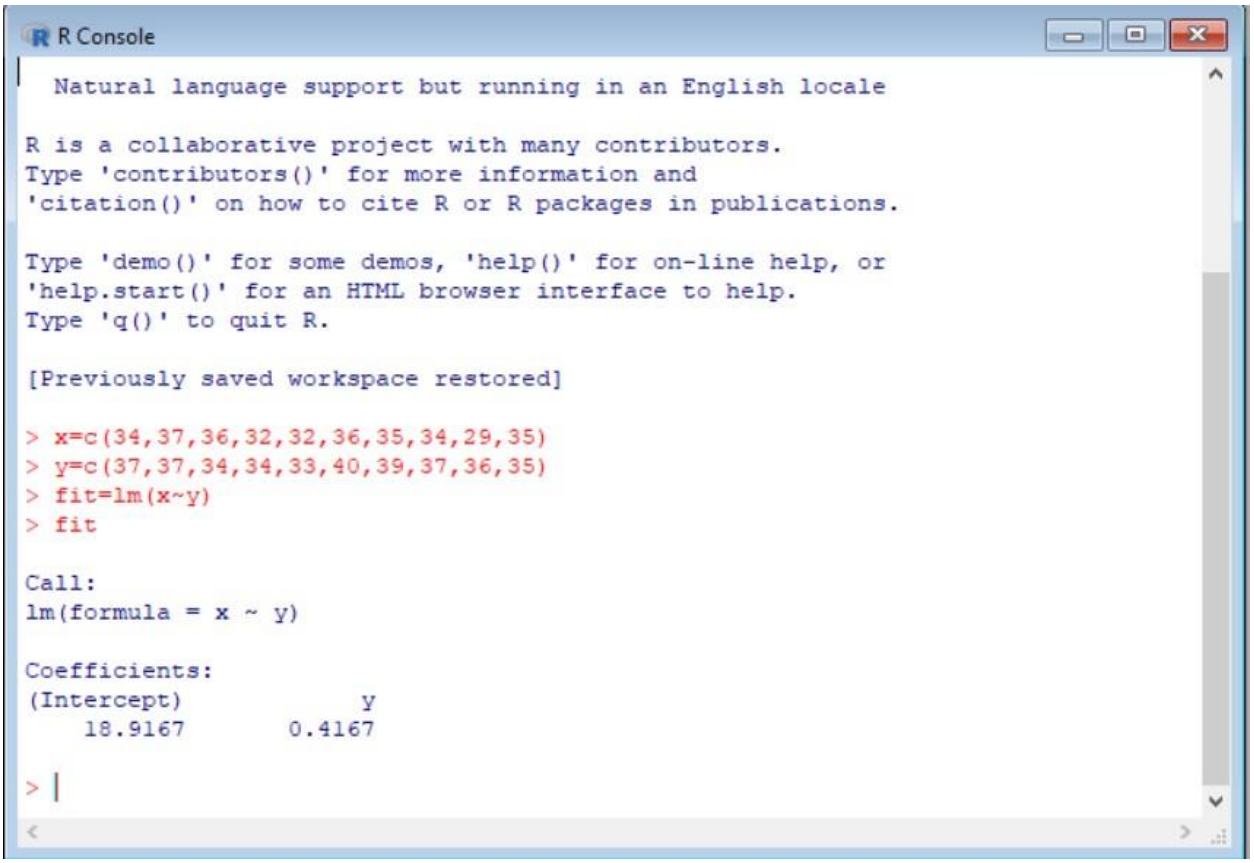
There is a negative correlation between mathematics and statistics.

Linear regression and Multiple Linear Regression

Problem 1: The following table shows the scores (X) of 10 students on Zoology test and scores (Y) on Botony test .The maximum score in each test was 50.Obtain least square equation of line of regression of X on Y. If it is known that the score of a student in Botony is 28,Estimate his/her score in Zoology.

X	34	37	36	32	32	36	35	34	29	35
Y	37	37	34	34	33	40	39	37	36	35

R code-

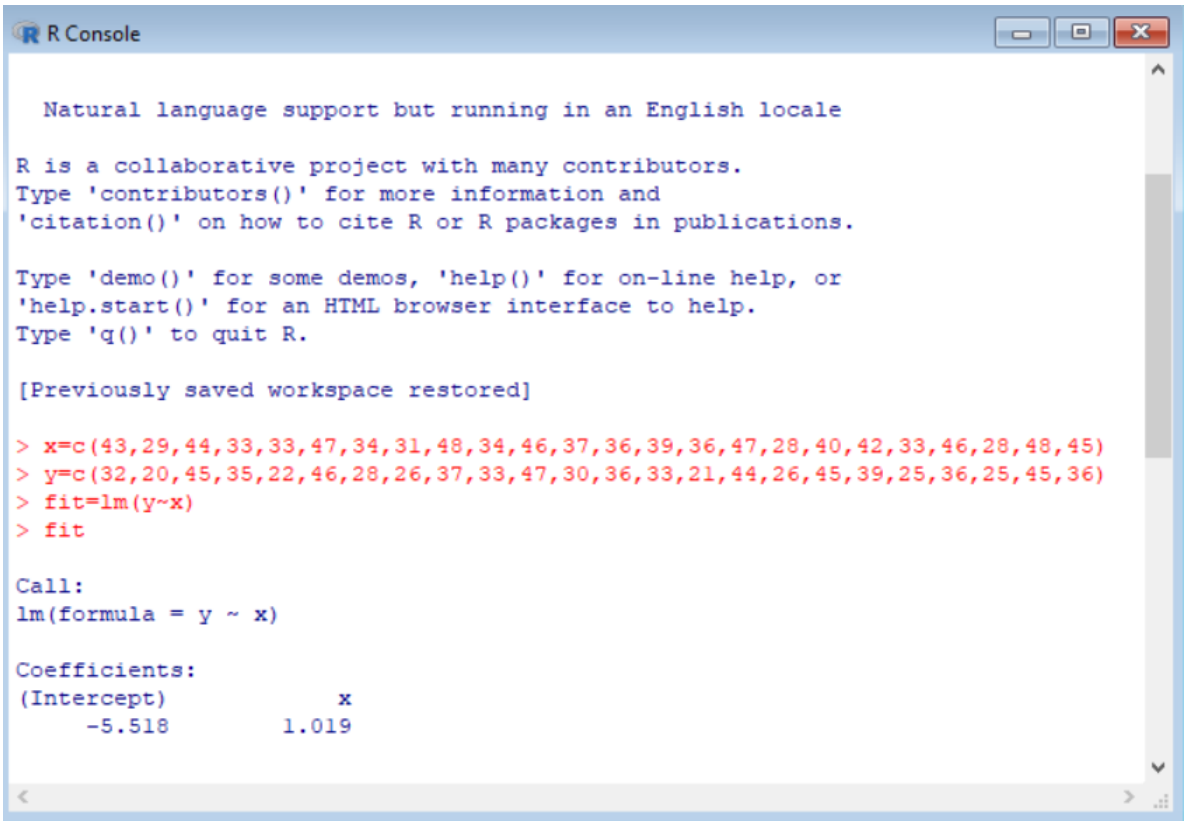


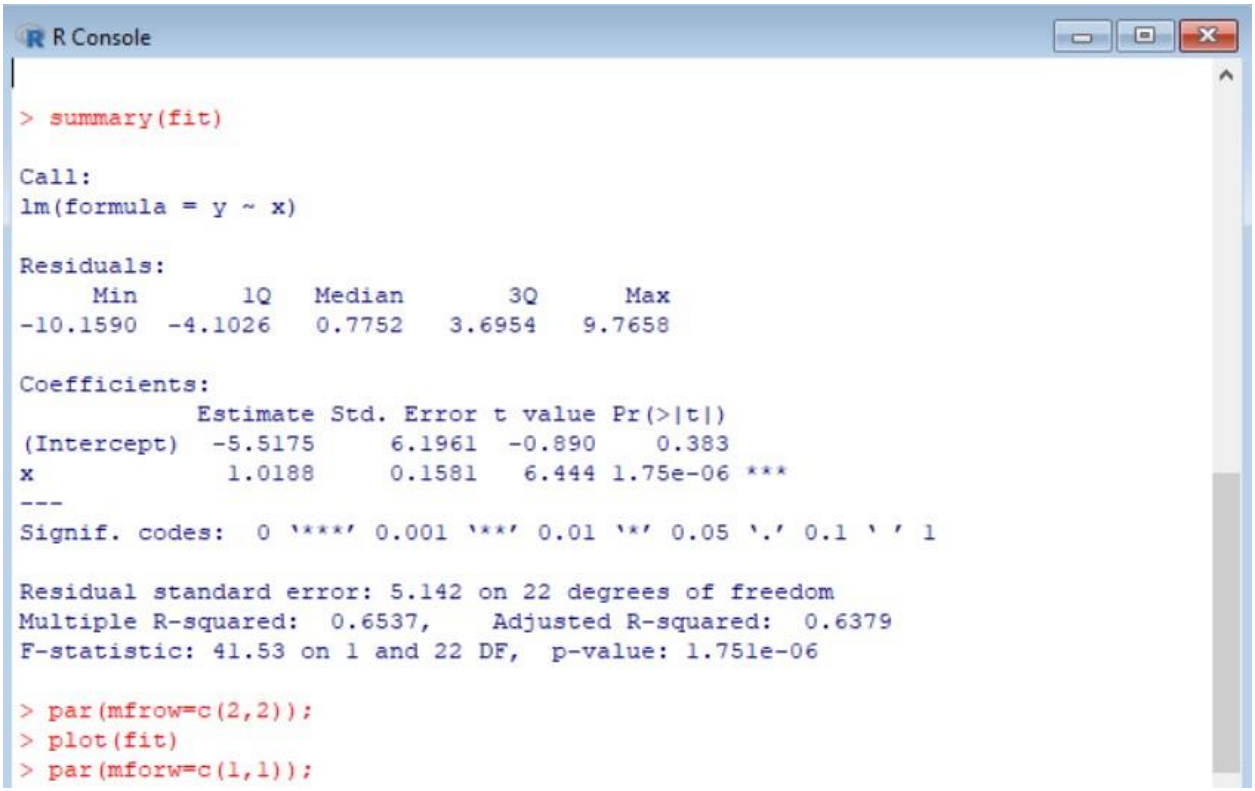
The equation of the line of regression of X and Y is X=18.9167+0.4167Y.
The required score of the student in Zoology is 30.58333

Problem 2 :- The following data pertain to the resistance in (ohms) and the failure times (minutes) of 24 overloaded resistors.

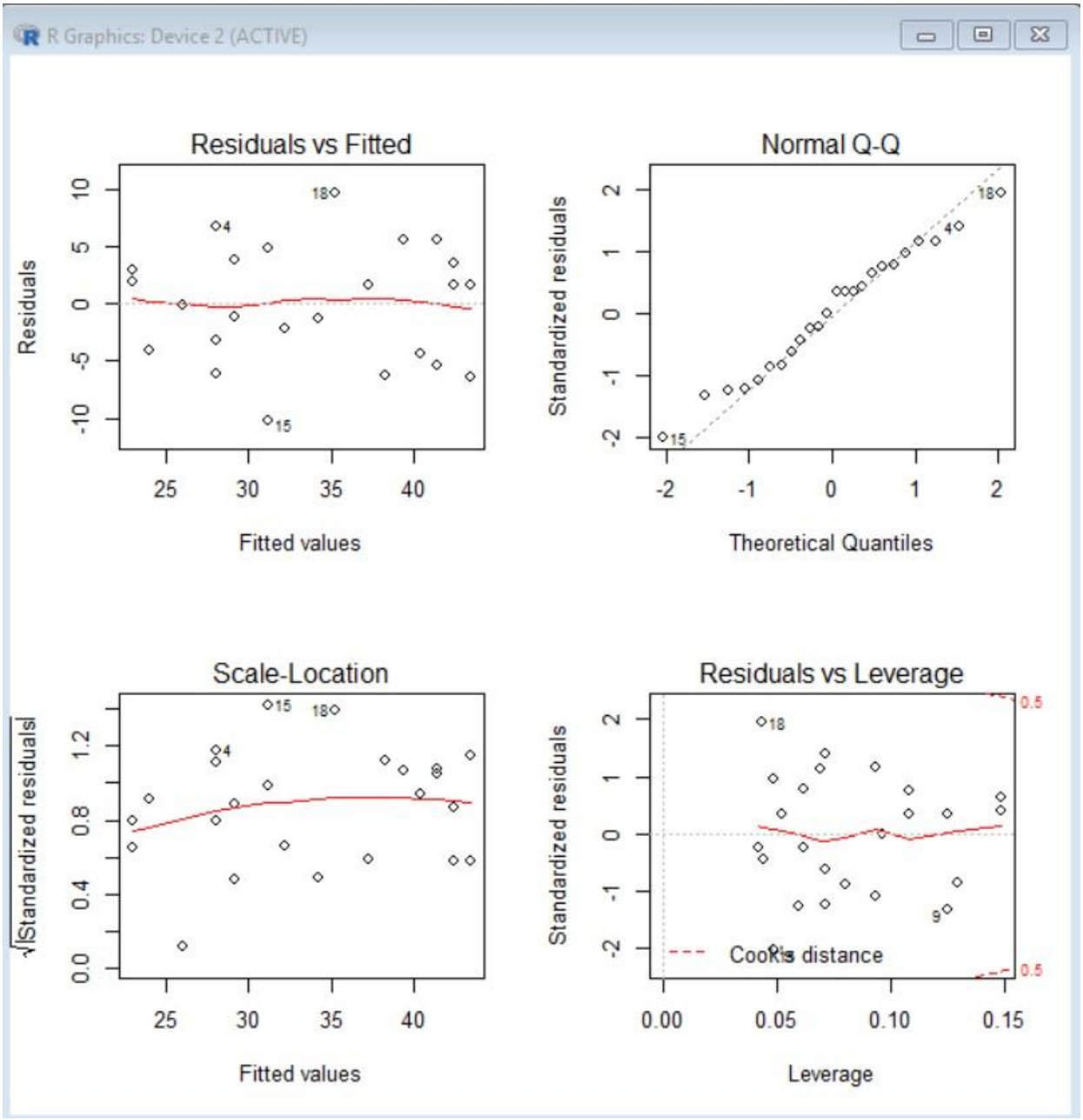
Resistance(x)	43	29	44	33	33	47	34	31	48
	34	46	37	36	39	36	47	28	40
	42	33	46	28	48	45			
Failure time(y)	32	20	45	35	22	46	28	26	37
	33	47	30	36	33	21	44	26	45
	39	25	36	25	45	36			

R code:-


R Console
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
[Previously saved workspace restored]
> x=c(43,29,44,33,33,47,34,31,48,34,46,37,36,39,36,47,28,40,42,33,46,28,48,45)
> y=c(32,20,45,35,22,46,28,26,37,33,47,30,36,33,21,44,26,45,39,25,36,25,45,36)
> fit=lm(y~x)
> fit
Call:
lm(formula = y ~ x)
Coefficients:
(Intercept)
x
-5.518
1.019


R Console
> summary(fit)
Call:
lm(formula = y ~ x)
Residuals:
Min
1Q
Median
3Q
Max
-10.1590
-4.1026
0.7752
3.6954
9.7658
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)
-5.5175
6.1961
-0.890
0.383
x
1.0188
0.1581
6.444 1.75e-06 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 5.142 on 22 degrees of freedom
Multiple R-squared: 0.6537, Adjusted R-squared: 0.6379
F-statistic: 41.53 on 1 and 22 DF, p-value: 1.751e-06
> par(mfrow=c(2,2));
> plot(fit)
> par(mforw=c(1,1));

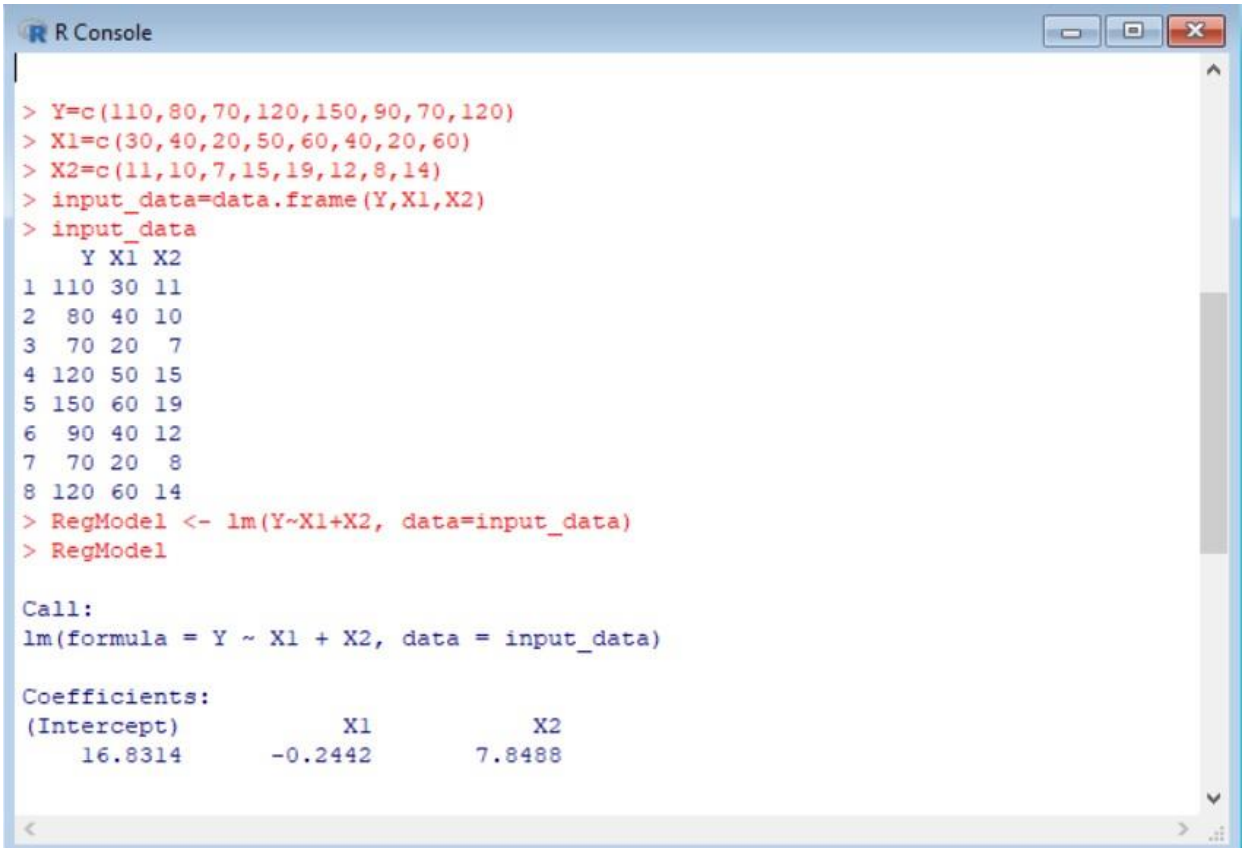

R Graphics: Device 2 (ACTIVE)
Residuals vs Fitted
Residuals
Fitted values
Normal Q-Q
Standardized residuals
Theoretical Quantiles
Scale-Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage

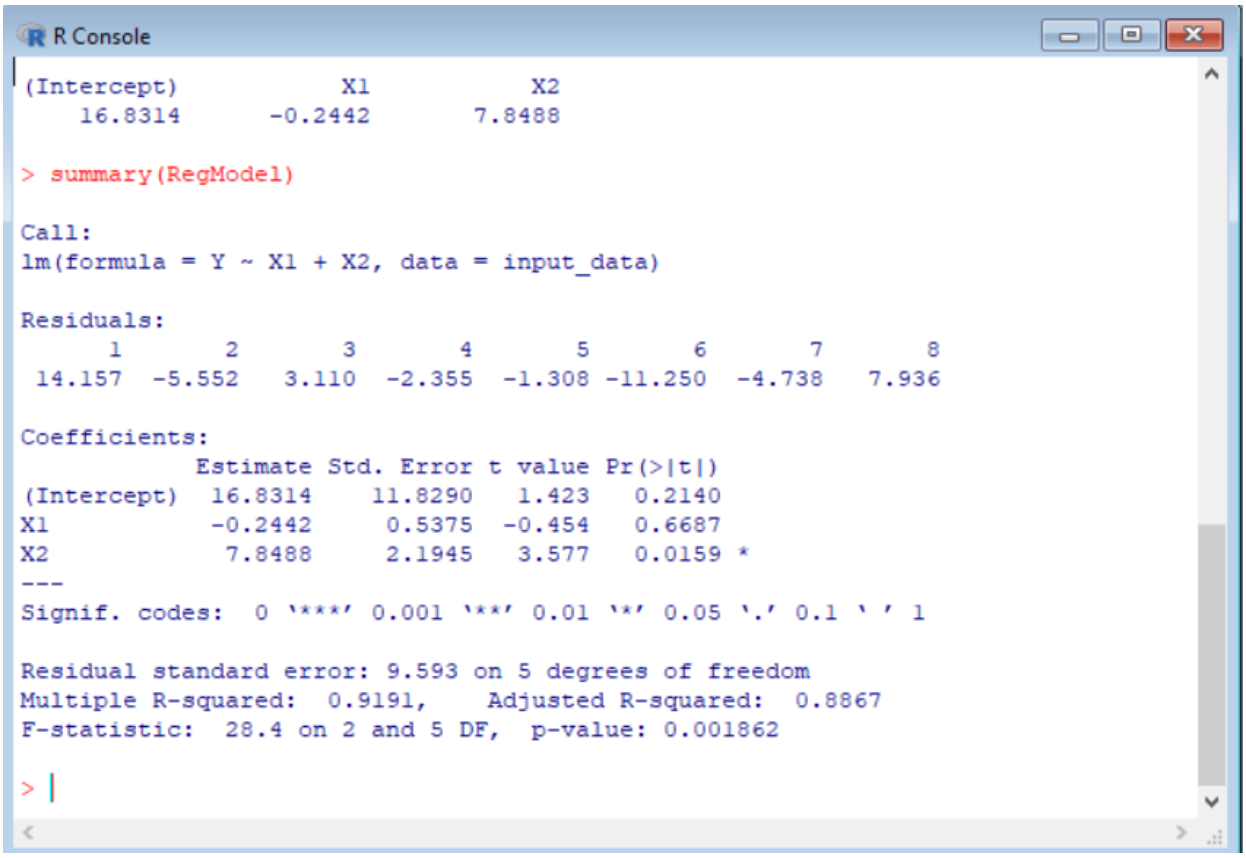
Diagnostic plots

Problem 3: The sale of a Product in lakhs of rupees(Y) is expected to be influenced by two variables namely the advertising expenditure X1 (in'OOO Rs) and the number of sales persons(X2) in a region. Sample data on 8 Regions of a state has given the following results

Area	Y	X1	X2
1	110	30	11
2	80	40	10
3	70	20	7
4	120	50	15
5	150	60	19
6	90	40	12
7	70	20	8
8	120	60	14

R-code-:




R Console
(Intercept) X1 X2
16.8314 -0.2442 7.8488
> summary(RegModel)
Call:
lm(formula = Y ~ X1 + X2, data = input_data)
Residuals:
1 2 3 4 5 6 7 8
14.157 -5.552 3.110 -2.355 -1.308 -11.250 -4.738 7.936
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 16.8314 11.8290 1.423 0.2140
X1 -0.2442 0.5375 -0.454 0.6687
X2 7.8488 2.1945 3.577 0.0159 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 9.593 on 5 degrees of freedom
Multiple R-squared: 0.9191, Adjusted R-squared: 0.8867
F-statistic: 28.4 on 2 and 5 DF, p-value: 0.001862
>