VIT®
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

Activity 5:

Name: Preyash **Date: 08/02/2022**

Registration Number: 20BPS1022

Aim: To use R to study regression

Tools Used: R Studio

Syntax/ Commands Used:

- **lm()** - In R, the lm(), or "linear model," function can be used to create a simple regression model.
- **abline()** - The R function abline() can be used to add vertical, horizontal or regression lines to a graph.
- **pch :** numeric values (from 0 to 25) or character symbols ("+", ".", ";", etc) specifying the point symbols (or shapes).
- **cex :** numeric values indicating the point size.

Questions:

Task 1:

1.Write R code to calculate the regression coefficient and obtain the lines of regression for the following data

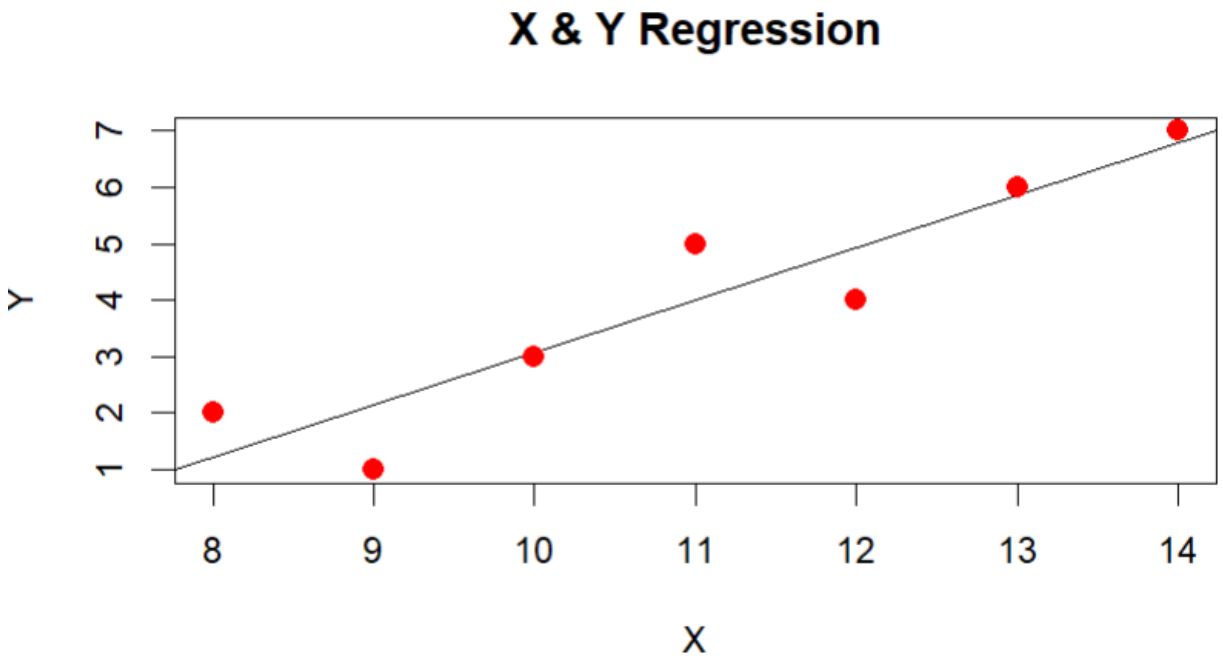
X	1	2	3	4	5	6	7
Y	9	8	10	12	11	13	14

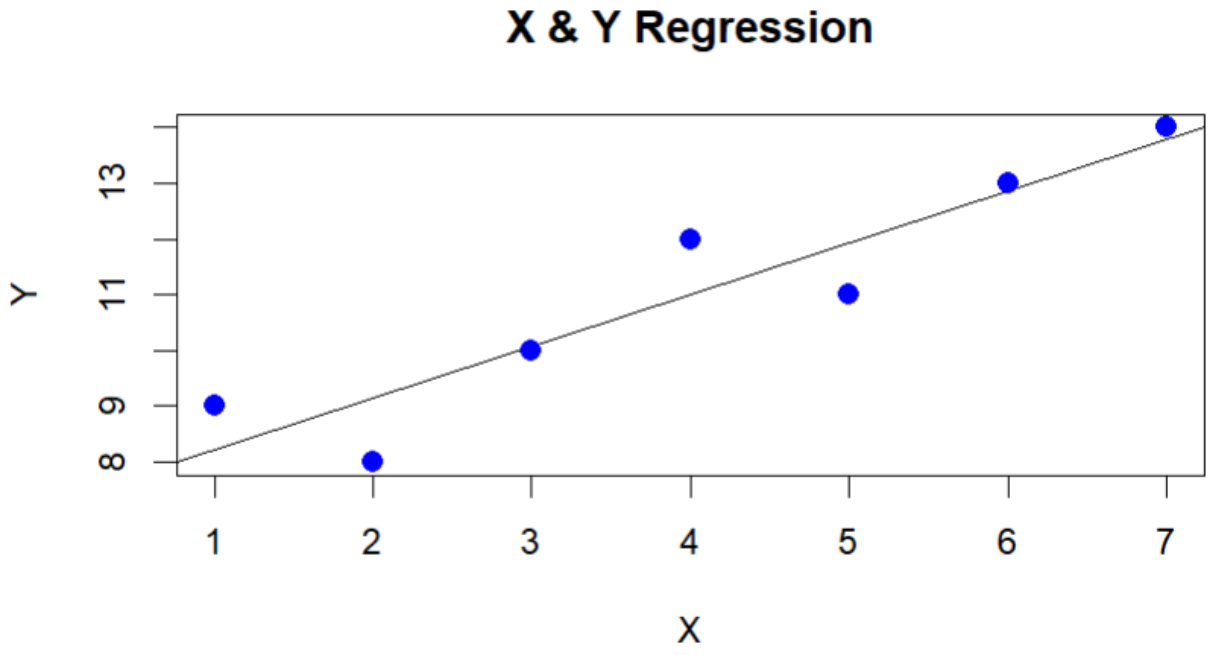
Code:

```
x<-c(1,2,3,4,5,6,7)

y<-c(9,8,10,12,11,13,14)

model<-lm(x~y)

summary(model)

plot(y, x, col = "red", main = "X & Y Regression", abline(lm(x~y)), cex = 1.3,
pch = 16, xlab =

    "X", ylab = "Y")

model=lm(y~x)

summary(model)

plot(x, y, col = "blue", main = "X & Y Regression", abline(lm(y~x)), cex = 1.3,
pch = 16, xlab =

    "X", ylab = "Y")
```

Output:

```
Console  Terminal  Jobs
R 4.1.2 · ~/
> x<-c(1,2,3,4,5,6,7)
> y<-c(9,8,10,12,11,13,14)
> model<-lm(x~y)
> summary(model)

Call:
lm(formula = x ~ y)

Residuals:
      1       2       3       4       5       6       7
-1.14286  0.78571 -0.07143 -0.92857  1.00000  0.14286  0.21429

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -6.2143     1.8558  -3.349  0.02036 *
y             0.9286     0.1660   5.594  0.00252 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8783 on 5 degrees of freedom
Multiple R-squared:  0.8622,    Adjusted R-squared:  0.8347
F-statistic:  31.3 on 1 and 5 DF,  p-value: 0.002519

> plot(y, x, col = "red", main = "X & Y Regression", abline(lm(x~y)), cex = 1.3, p
ch = 16, xlab =
+         "X", ylab = "Y")
> model=lm(y~x)
> summary(model)

Call:
lm(formula = y ~ x)

Residuals:
      1       2       3       4       5       6       7
 0.78571 -1.14286 -0.07143  1.00000 -0.92857  0.14286  0.21429

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.2857     0.7423   9.815 0.000187 ***
x             0.9286     0.1660   5.594 0.002519 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8783 on 5 degrees of freedom
Multiple R-squared:  0.8622,    Adjusted R-squared:  0.8347
F-statistic:  31.3 on 1 and 5 DF,  p-value: 0.002519

> plot(x, y, col = "blue", main = "X & Y Regression", abline(lm(y~x)), cex = 1.3,
 pch = 16, xlab =
+         "X", ylab = "Y")
>
```


X & Y Regression
Y
1
2
3
4
5
6
7
8
9
10
11
12
13
14
X


X & Y Regression
Y
8
9
11
13
1
2
3
4
5
6
7
X

Task 2:

2. Write R code to obtain regression equation of *Y* on *X* and estimate *Y* when *X*=55 from the following

X	40	50	38	60	65	50	35
Y	38	60	55	70	60	48	30

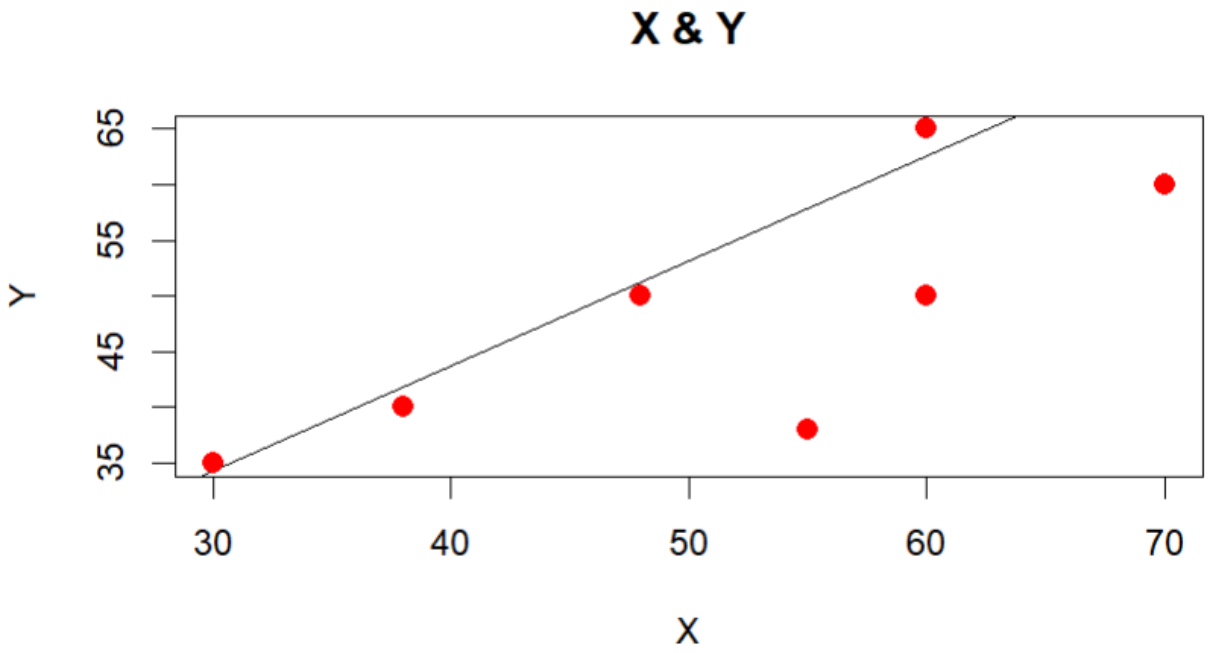
Also visualize it.

Code:

```
x<-c(40,50,38,60,65,50,35)

y<-c(38,60,55,70,60,48,30)

model<-lm(y~x)

plot(y,x,col = "red",main = "X & Y",abline(lm(y~x)),cex = 1.3,pch = 16,xlab = "X",ylab = "Y")

a<-data.frame(x=55)

result<-predict(model,a)

print(result)
```

Output:

```
Console  Terminal  Jobs
R 4.1.2 · ~/
> x<-c(40,50,38,60,65,50,35)
> y<-c(38,60,55,70,60,48,30)
> model<-lm(y~x)
> plot(y,x,col = "red",main = "X & Y",abline(lm(y~x)),cex = 1.3,pch = 16,xlab = "X",
ylab = "Y")
> a<-data.frame(x=55)
> result<-predict(model,a)
> print(result)
       1
57.89878
>
```



Task 3:

3. The following table shows the sales and advertisement expenditure of a form

	Sales	Advertisement expenditure (Rs. Crores)
Mean	40	6
SD	10	1.5

Coefficient of correlation *r*= 0.9. Estimate the likely sales for a proposed advertisement expenditure of Rs. 10 crores using R Code.

Code:

```
meanx=40

meany=6

SDx=10

SDy=1.5
```

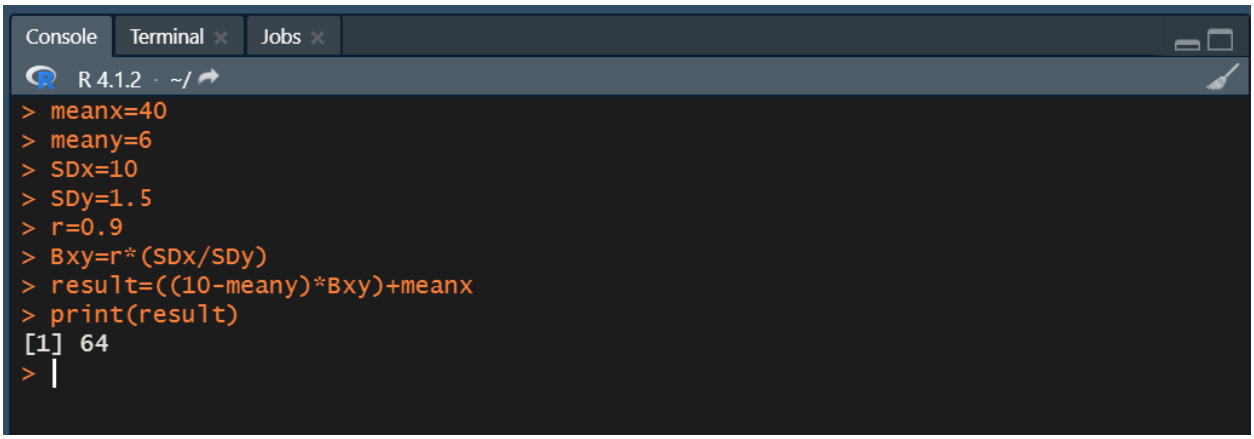
r=0.9

Bxy=r*(SDx/SDy)

result=((10-mean)*Bxy)+meanx

print(result)

Output:



Task 4 :

4. The below table displays the details of the Age of cars in years (X) and the annual maintenance cost (in Rupees) (Y).

Age of Cars in Years (X)	1	3	5	7	9
Annual Maintenance Cost (Y)	15	18	21	23	22

The factory manager wants to display the following calculation on their report summary.
Use R studio software to obtain the result.

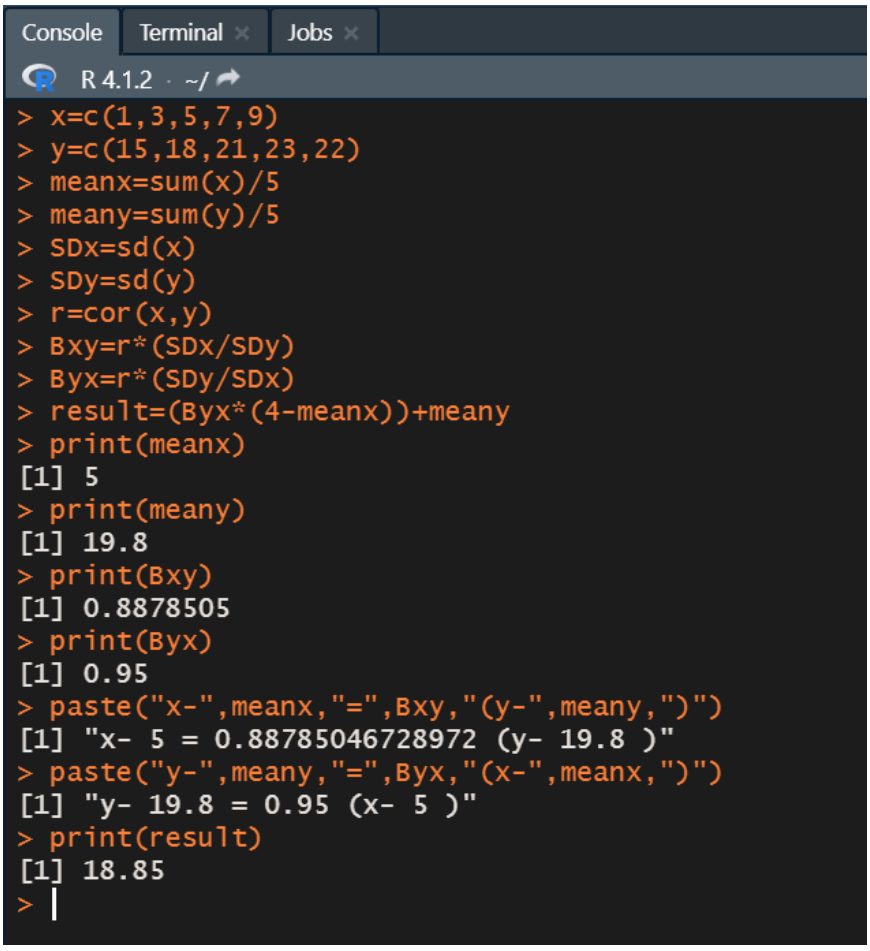
(i) Mean of X
(ii) Mean of Y
(iii) Regression coefficients using the formula
(iv) Regression lines of equation
(v) Estimate the maintenance cost for a 4-year-old after obtaining the regression equation.

Code:

```
x=c(1,3,5,7,9)

y=c(15,18,21,23,22)

meanx=sum(x)/5

meany=sum(y)/5

SDx=sd(x)

SDy=sd(y)

r=cor(x,y)

Bxy=r*(SDx/SDy)

Byx=r*(SDy/SDx)

result=(Byx*(4-meanx))+meany

print(meanx)
```

```
print(meany)
print(Bxy)
print(Byx)
paste("x-",meanx,"=",Bxy,"(y-",meany,")")
paste("y-",meany,"=",Byx,"(x-",meanx,")")
print(result)
```

Output:



Result: We successfully used R to study regression.