DIGITAL ASSIGNMENT – 2

ARIJIT DE 23BEC0215

SLOT : L3+L4

1. A researcher collects data on student's hours spent on a new training program (X) and their test performance scores (Y) to analyse the relationship between them.

Person	1	2	3	4	5	6	7	8	9	10
Training Hours (X)	2.5	3	4.5	5	6.5	7	8	9.5	10	11.5
Test Score (Y)	55	60	63	68	72	75	78	82	85	90

Find the Pearson correlation coefficient and between training hours and test scores.

```
x=c(2.5,3,4.5,5,6.5,7,8,9.5,10,11.5)
> y=c(55,60,63,68,72,75,78,82,85,90)
> cor.test(x,y,method="pearson")

        Pearson's product-moment correlation

data:  x and y
t = 26.11, df = 8, p-value = 4.972e-09
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9746595 0.9986751
sample estimates:
      cor
0.9941838
```

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Console Terminal Background Jobs

R 4.4.2 · ~/

```
R version 4.4.2 (2024-10-31 ucrt) -- "Pile of Leaves"
Copyright (C) 2024 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64

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.

[Workspace loaded from ~/.RData]

> x=c(2.5,3,4.5,5,6.5,7,8,9.5,10,11.5)
> y=c(55,60,63,68,72,75,78,82,85,90)
> cor.test(x,y,method="pearson")

        Pearson's product-moment correlation

data:  x and y
t = 26.11, df = 8, p-value = 4.972e-09
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9746595 0.9986751
sample estimates:
      cor
0.9941838
```

2. An educational researcher, is conducting a study to understand the relationship between students' performance in Economics and Statistics. The researcher believes that the students who perform well in Statistics may also excel in Economics. The researcher collects data from a group of 12 students, recording their scores in both subject (out of 50).

Student	1	2	3	4	5	6	7	8	9	10	11	12
Economics (X)	48	42	39	45	37	50	41	44	36	49	38	43
Statistics (Y)	45	40	38	42	35	47	39	41	33	46	36	40

Find the least squares regression equation of Economics scores (X) on Statistics scores (Y).

```
x <- c(48, 42, 39, 45, 37, 50, 41, 44, 36, 49, 38, 43)

> y <- c(45, 40, 38, 42, 35, 47, 39, 41, 33, 46, 36, 40)

> input_data=data.frame(x,y)

> input_data

   x  y

1  48 45

2  42 40

3  39 38

4  45 42

5  37 35

6  50 47

7  41 39

8  44 41

9  36 33

10 49 46

11 38 36

12 43 40

> RegModel <- lm(y~x,data=input_data)

> RegModel


Call:
```

```
lm(formula = y ~ x, data = input_data)


Coefficients:
(Intercept)            x
     0.9877       0.9183
summary(RegModel)


Call:
lm(formula = y ~ x, data = input_data)


Residuals:
     Min       1Q   Median       3Q      Max
-1.04496 -0.32970  0.02725  0.17984  1.20027


Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.98774    1.59118   0.621    0.549
x            0.91826    0.03709  24.760 2.64e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 0.5801 on 10 degrees of freedom
Multiple R-squared:  0.984,	Adjusted R-squared:  0.9823
F-statistic: 613.1 on 1 and 10 DF,  p-value: 2.639e-10


> coefficients <- coef(RegModel)
>
> cat("Regression equation: Y =", coefficients[1], "+", coefficients[2], "* X\n")
Regression equation: Y = 0.9877384 + 0.9182561 * X
```

```
2  42 40
3  39 38
4  45 42
5  37 35
6  50 47
7  41 39
8  44 41
9  36 33
10 49 46
11 38 36
12 43 40
> RegModel <- lm(y~x,data=input_data)
> RegModel

Call:
lm(formula = y ~ x, data = input_data)

Coefficients:
(Intercept)            x
     0.9877       0.9183

> summary(RegModel)

Call:
lm(formula = y ~ x, data = input_data)

Residuals:
     Min       1Q   Median       3Q      Max
-1.04496 -0.32970  0.02725  0.17984  1.20027

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.98774    1.59118   0.621    0.549
x            0.91826    0.03709  24.760 2.64e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5801 on 10 degrees of freedom
Multiple R-squared:  0.984,     Adjusted R-squared:  0.9823
F-statistic: 613.1 on 1 and 10 DF,  p-value: 2.639e-10

> coefficients <- coef(RegModel)
> cat("Regression equation: Y =", coefficients[1], "+", coefficients[2], "* X\n")
Regression equation: Y = 0.9877384 + 0.9182561 * X
>
```

3. An environmental scientist wants to study how traffic volume (X_1) and temperature (X_2) affects air pollution levels (Y) in different areas of a city. Increased vehicle movement and temperature fluctuations are believed to influence pollution levels, but the exact relationship is unclear. By analysing data from multiple locations, the scientist aims to develop a predicative model to estimate air pollution based on these factors. The collected data form 8 locations is given below

Location	1	2	3	4	5	6	7	8
Air Pollution index (Y)	150	120	180	100	140	160	110	175
Traffic volume (X_1) (vehicles / hour)	500	350	600	300	450	550	320	580
Temperature (X_2) (°C)	30	28	32	25	29	31	27	33

Develop a multiple regression equation to predict Air pollution index based on traffic volume and temperature.

```
> Y=c(150, 120, 180, 100, 140, 160, 110, 175)

> X1=c(500, 350, 600, 300, 450, 550, 320, 580)

>

> X2=c(30, 28, 32, 25, 29, 31, 27, 33)

>

> data <- data.frame(Y,X1,X2)

> RegModel <- lm(Y~X1+X2,data=data)

> RegModel


Call:

lm(formula = Y ~ X1 + X2, data = data)


Coefficients:

(Intercept)          X1          X2

  -37.6863      0.1727      3.4299


> summary(RegModel)
```

```
Call:
lm(formula = Y ~ X1 + X2, data = data)

Residuals:
      1       2       3       4       5       6       7       8 
-1.5755  1.1934  4.2919  0.1196  0.4908 -3.6418 -0.1948 -0.6835 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept) -37.68626   28.06904  -1.343  0.23712   
X1            0.17273    0.03149   5.486  0.00275 **
X2            3.42994    1.41680   2.421  0.06005 . 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.696 on 5 degrees of freedom
Multiple R-squared:  0.9941,	Adjusted R-squared:  0.9918 
F-statistic: 423.7 on 2 and 5 DF,  p-value: 2.635e-06

> coefficients <- coef(RegModel)
> cat("Regression equation: Y =", coefficients[1], "+", coefficients[2], "* X1 +", coefficients[3], "* X2\n")
Regression equation: Y = -37.68626 + 0.1727273 * X1 + 3.429938 * X2
```

```
> Y=c(150, 120, 180, 100, 140, 160, 110, 175)
> X1=c(500, 350, 600, 300, 450, 550, 320, 580)
>
> X2=c(30, 28, 32, 25, 29, 31, 27, 33)
>
> data <- data.frame(Y,X1,X2)
> RegModel <- lm(Y~X1+X2,data=data)
> RegModel

Call:
lm(formula = Y ~ X1 + X2, data = data)

Coefficients:
(Intercept)           X1           X2
   -37.6863       0.1727       3.4299

> summary(RegModel)

Call:
lm(formula = Y ~ X1 + X2, data = data)

Residuals:
      1       2       3       4       5       6       7       8
-1.5755  1.1934  4.2919  0.1196  0.4908 -3.6418 -0.1948 -0.6835

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -37.68626   28.06904  -1.343  0.23712
X1            0.17273    0.03149   5.486  0.00275 **
X2            3.42994    1.41680   2.421  0.06005 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.696 on 5 degrees of freedom
Multiple R-squared:  0.9941,    Adjusted R-squared:  0.9918
F-statistic: 423.7 on 2 and 5 DF,  p-value: 2.635e-06

> coefficients <- coef(RegModel)
> cat("Regression equation: Y =", coefficients[1], "+", coefficients[2], "* X1 +", coefficients[3],
"* X2\n")
Regression equation: Y = -37.68626 + 0.1727273 * X1 + 3.429938 * X2
>
>
```