

Date: \_\_\_\_\_

Question: (Correlation coefficient)

A software development team is interested in investigating the relationship between lines of code written in a project and the execution time in milliseconds of the resulting program. For this purpose, a sample of 10 programs is selected. The dataset is as follows:

Lines of code (X)	Execution time (ms) Y
50	120
70	160
60	135
100	210
90	185
80	170
110	220
130	255
120	245
140	270

- Calculate correlation b/w the two variables & interpret the result.
- Test at 5% level of significance that the execution time and lines of code are positively related.
- Calculate the % of variability explained in Y due to the independent variable.

Date: \_\_\_\_\_

Sol:

a-  $r = \frac{n\sum XY - \sum X \sum Y}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$

$$= \frac{10(201100) - (950)(1970)}{\sqrt{[10(98500) - (950)^2][10(411800) - (1970)^2]}}$$
$$= 0.9974$$

Strong Positive Correlation.

b-  $H_0: \rho \leq 0$

$H_1: \rho > 0$

2-  $d = 5Y$ .

3- Test-Statistics:  $t = \frac{r \sqrt{n-2}}{\sqrt{1-r^2}}$  with  $n-2$  df.

$$= \frac{(0.9974) \sqrt{10-2}}{\sqrt{1 - (0.9974)^2}} = \frac{2.8211}{\sqrt{0.0052}}$$

$\neq$

4- Critical Region:  $t > t_{\alpha(v)}$  one tail test  
 $39.12 > t_{0.05(8)}$  as  $H_1$  shows  
 $39.12 > 1.86$   $\rho > 0$

~~AR 1.86 CR. 39.12~~

5- Conclusion: Since calculated value falls in the critical Region so we are rejecting  $H_0$  and conclude that lines of code and execution time are positively related.

c-  $R^2 = \text{Corr} = (0.9974)^2 = 0.9948$

99.48% variability in dependent variable is explained due to independent Variable

KING'S  
NOTES

## Question (correlation coefficient)

A digital Market Analyst is studying the impact of Product Price on the number of units sold for a new line of headphones. The data collected over 10 weeks is presented below:

week	Price (USD)	units sold
1	120	400
2	110	450
3	100	510
4	90	570
5	80	630
6	130	380
7	95	560
8	85	600
9	125	390
10	105	490

Task:-

⇒ Test whether there is a negative relationship between Price & units sold. (use sample correlation)

Sol:

$$\text{Sample correlation} = r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}$$

$$n=10, \sum xy = 504000, \sum x = 1040, \sum y = 4980$$

$$\sum x^2 = 111800, \sum y^2 = 2550600$$

Put these sums in the above formula

$$r = \frac{(10)(504000) - (1040)(4980)}{\sqrt{[(10)(111800) - (1040)^2][(10)(2550600) - (4980)^2]}}$$

$$r = \boxed{-0.9295}$$

<sup>strong</sup> negative relationship.

## Hypothesis Testing

①  $H_0: \rho \geq 0$  (No correlation or Positive correlation)

$H_1: \rho < 0$  (negatively related or Inverse correlation)

②  $\alpha = 5\%$

③ Test-Statistics

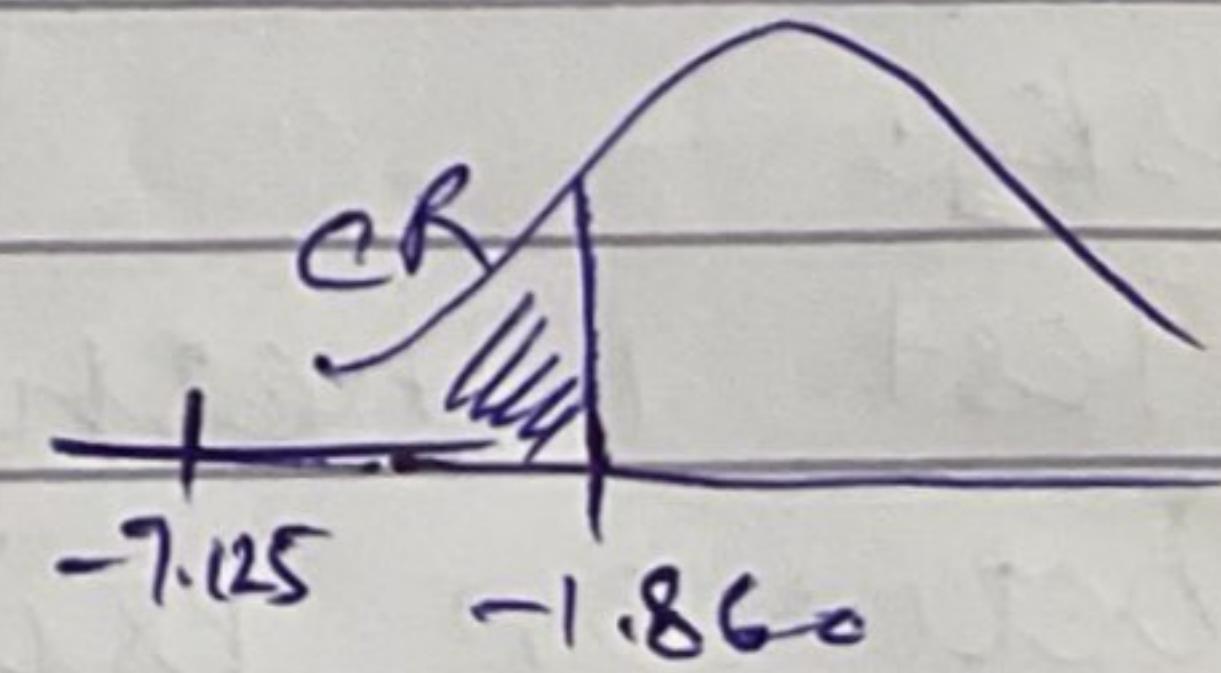
$$t = \frac{r \sqrt{n-2}}{\sqrt{1-r^2}} = \frac{-0.9295 \sqrt{8}}{\sqrt{1-(-0.9295)^2}}$$

$$t = -7.125$$

④ Critical Region:

$$t \leq -t_{0.05, 8}$$

$$t \leq -t_{0.05, 8} \Rightarrow -7.125 \leq -1.860$$



⑤ Conclusion:

It is observed from the above

Procedure that  $t_{\text{cal}}$  falls in

the Critical Region which leads to

the rejection of  $H_0$ . So it is concluded that Price (USD) & the units sold are inversely related. (Indirect/negative correlation).