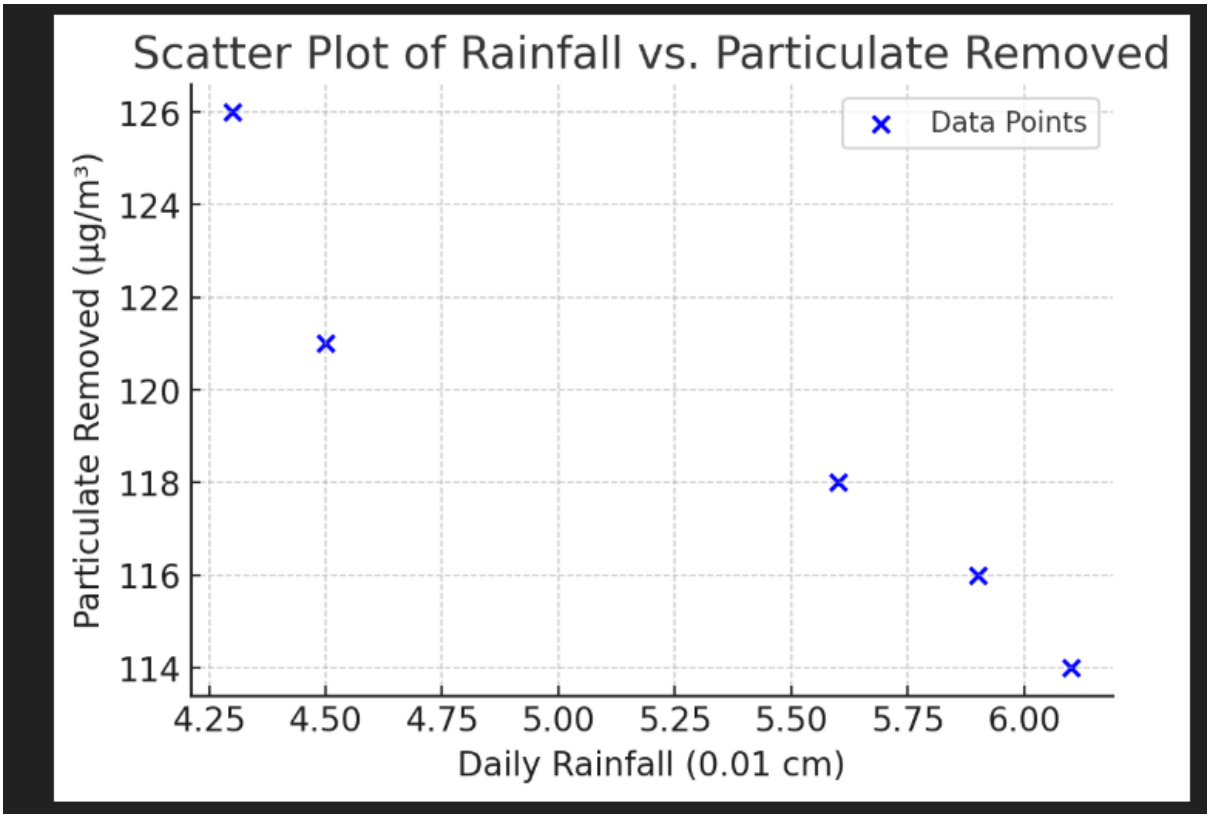
TUTORIAL-6

1.

Given Data:

Daily Rainfall, x (0.01 cm)	4.3	4.5	5.9	5.6	6.1
Particulate Removed, y (μg/m³)	126	121	116	118	114

Now, I'll generate the scatter plot and compute the correlation coefficient.



Step-by-Step Calculation of Correlation Coefficient (r)

The Pearson correlation coefficient is calculated using the formula:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}}$$

where:

- x_i and y_i are the given data points,
- $\bar{x}$ and $\bar{y}$ are the means of x and y,
- $\sum(x_i - \bar{x})(y_i - \bar{y})$ is the sum of the product of deviations,
- $\sum(x_i - \bar{x})^2$ is the sum of squared deviations for x,
- $\sum(y_i - \bar{y})^2$ is the sum of squared deviations for y.

Step 1: Given Data

$$x = [4.3, 4.5, 5.9, 5.6, 6.1]$$

$$y = [126, 121, 116, 118, 114]$$

Step 2: Compute Means

$$\bar{x} = \frac{4.3 + 4.5 + 5.9 + 5.6 + 6.1}{5} = 5.28$$

$$\bar{y} = \frac{126 + 121 + 116 + 118 + 114}{5} = 119$$

Step 3: Compute Deviations

$$x_i - \bar{x} = [-0.98, -0.78, 0.62, 0.32, 0.82]$$

$$y_i - \bar{y} = [7, 2, -3, -1, -5]$$

Step 4: Compute Summations

$$\sum(x_i - \bar{x})(y_i - \bar{y}) = (-0.98 \times 7) + (-0.78 \times 2) + (0.62 \times -3) + (0.32 \times -1) + (0.82 \times -5) = -14.7$$

$$\sum(x_i - \bar{x})^2 = 2.728$$

$$\sum(y_i - \bar{y})^2 = 88$$

Step 5: Compute r

$$r = \frac{-14.7}{\sqrt{2.728} \times \sqrt{88}}$$

$$= \frac{-14.7}{1.651 \times 9.380}$$

$$= \frac{-14.7}{15.49} = -0.949$$

Final Answer:

$$r = -0.949$$

(Strong negative correlation)

2. While calculating correlation coefficient between two variables x and y from 30 pairs of observations, the following results were obtained: n=30, $\sum x$=120, $\sum x^2$=600, $\sum y$=90, $\sum y^2$=250, $\sum xy$=356. Later it was discovered at the time of checking that the pair of values

x	y
18	12
10	17

were copied down as

x	y
8	10
12	7

Obtain the correct value of correlation coefficient.

2.

Step 1: Given Data

- $n = 30$
- $\sum x = 120$
- $\sum x^2 = 600$
- $\sum y = 90$
- $\sum y^2 = 250$
- $\sum xy = 356$

Step 2: Identify Incorrectly Recorded Pairs

- Actual values: $(18, 12)$ and $(10, 17)$
- Incorrectly recorded as: $(8, 10)$ and $(12, 7)$

Step 3: Adjust Summations

Adjust $\sum x$

$$\sum x_{\text{corrected}} = \sum x - (8 + 12) + (18 + 10)$$

$$= 120 - 20 + 28 = 128$$

Adjust $\sum y$

$$\sum y_{\text{corrected}} = \sum y - (10 + 7) + (12 + 17)$$

$$= 90 - 17 + 29 = 102$$

Adjust $\sum x^2$

$$\sum x^2_{\text{corrected}} = \sum x^2 - (8^2 + 12^2) + (18^2 + 10^2)$$

$$= 600 - (64 + 144) + (324 + 100)$$

$$= 600 - 208 + 424 = 816$$

Adjust $\sum y^2$

$$\sum y^2_{\text{corrected}} = \sum y^2 - (10^2 + 7^2) + (12^2 + 17^2)$$

$$= 250 - (100 + 49) + (144 + 289)$$

$$= 250 - 149 + 433 = 534$$

Adjust $\sum xy$

$$\sum xy_{\text{corrected}} = \sum xy - (8 \times 10 + 12 \times 7) + (18 \times 12 + 10 \times 17)$$

$$= 356 - (80 + 84) + (216 + 170)$$

$$= 356 - 164 + 386 = 578$$

Step 4: Compute Corrected Correlation Coefficient r

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2)}}$$

Substituting values:

$$r = \frac{30(578) - (128)(102)}{\sqrt{(30 \times 816 - 128^2)(30 \times 534 - 102^2)}}$$

$$= \frac{17340 - 13056}{\sqrt{(24480 - 16384)(16020 - 10404)}}$$

$$= \frac{4284}{\sqrt{8096 \times 5616}}$$

$$= \frac{4284}{\sqrt{45483216}}$$

$$= \frac{4284}{6745.36}$$

$$= 0.635$$

Date: STUDENT NAME:

3. The value of Karlpearson's correlation (r) for the following data is **0.636**.

x:	0.05	0.14	0.24	0.30	0.47	0.52	0.57	0.61	0.67	0.72
y:	1.08	1.15	1.27	1.33	1.41	1.46	1.54	2.72	4.01	9.63

(i) Calculate the Spearman's rank correlation for this data.

(ii) What advantage of ρ brought out in this problem?

3. **Solution:**

Spearman's rank correlation coefficient is given by:

$$\rho = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

where:

- d = difference between ranks of x and y,
- n = number of data points.

Step 1: Assign Ranks to x and y

Ranking x-values

$$x = [0.05, 0.14, 0.24, 0.30, 0.47, 0.52, 0.57, 0.61, 0.67, 0.72]$$

Assigning ranks (smallest to largest):

$$R_x = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]$$

Ranking y-values

$$y = [1.08, 1.15, 1.27, 1.33, 1.41, 1.46, 1.54, 2.72, 4.01, 9.63]$$

Assigning ranks:

$$R_y = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]$$

Step 2: Compute d^2

Since the rankings of x and y are identical, the differences $d = R_x - R_y$ are all zero.

$$\sum d^2 = 0$$

Step 3: Compute Spearman's Rank Correlation

$$\rho = 1 - \frac{6(0)}{10(10^2 - 1)}$$

$$= 1 - 0 = 1$$

Thus, **Spearman's Rank Correlation** $\rho = 1.00$.

Step 4: Advantage of Spearman's Rank Correlation

- Spearman's ρ is perfect (1.00), indicating a **monotonic** relationship.
- Pearson's $r = 0.636$ shows a moderate linear relationship, but Spearman's measure suggests that the relationship is **strongly increasing**, even if not strictly linear.
- This highlights **Spearman's advantage:** it captures the strength of monotonic relationships, even when Pearson's r is not very high.

4. Twelve recruits were subjected to a selection test to ascertain their suitability for a certain course of training. At the end of training there was a proficiency test given.
   The marks secured by recruits in the selection test (x) and the proficiency test (Y) are given below
   x: 65 63 67 64 68 62 70 66 68 67 69 71
   y: 68 66 68 65 69 66 68 65 71 67 68 70
   Calculate coefficient of rank correlation.

4.

Step 1: Rank the Scores

First, we need to rank the scores for both the selection test (x) and the proficiency test (y). If there are ties, we assign the average rank.

Selection Test (x) Scores:

65, 63, 67, 64, 68, 62, 70, 66, 68, 67, 69, 71

Proficiency Test (y) Scores:
68, 66, 68, 65, 69, 66, 68, 65, 71, 67, 68, 70

Let's rank these scores.

5. First, we need to rank the scores for both the selection test (x) and the proficiency test (y). If there are ties, we assign the average rank.
6. **Selection Test (x) Scores:**
   65, 63, 67, 64, 68, 62, 70, 66, 68, 67, 69, 71
7. **Proficiency Test (y) Scores:**
   68, 66, 68, 65, 69, 66, 68, 65, 71, 67, 68, 70
8. Let's rank these scores.

Step 2: Assign Ranks

For x:

- Sort x in ascending order: 62, 63, 64, 65, 66, 67, 67, 68, 68, 69, 70, 71
- Assign ranks:
  - 62: 1
  - 63: 2
  - 64: 3
  - 65: 4
  - 66: 5
  - 67: 6.5 (tie)
  - 67: 6.5 (tie)
  - 68: 8.5 (tie)
  - 68: 8.5 (tie)
  - 69: 10
  - 70: 11
  - 71: 12

For y:

- Sort y in ascending order: 65, 65, 66, 66, 67, 68, 68, 68, 68, 69, 70, 71
- Assign ranks:
  - 65: 1.5 (tie)
  - 65: 1.5 (tie)
  - 66: 3.5 (tie)
  - 66: 3.5 (tie)
  - 67: 5
  - 68: 7.5 (tie)
  - 68: 7.5 (tie)
  - 68: 7.5 (tie)
  - 68: 7.5 (tie)
  - 69: 10
  - 70: 11
  - 71: 12

Step 3: Calculate Differences and Square Them

Now, for each recruit, find the difference between their ranks in x and y, then square the differences.

Let's create a table for clarity:

Recruit	x Rank	y Rank	Difference (d)	d^2
1	4	7.5	-3.5	12.25
2	2	3.5	-1.5	2.25
3	6.5	7.5	-1	1
4	3	1.5	1.5	2.25
5	8.5	10	-1.5	2.25
6	1	3.5	-2.5	6.25
7	11	7.5	3.5	12.25
8	5	1.5	3.5	12.25
9	8.5	12	-3.5	12.25
10	6.5	5	1.5	2.25
11	10	11	-1	1
12	12	12	0	0

$\sum di2=12.25+2.25+1+2.25+2.25+6.25+12.25+12.25+12.25+2.25+1+0=64.5$

Now, apply the Spearman's rank correlation formula:

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} = 1 - \frac{6 \times 64.5}{12(144 - 1)} = 1 - \frac{387}{1716} \approx 1 - 0.2255 = 0.7745$$

5. In the accompanying table, x is the tensile force applied to a steel specimen in thousands of pounds, and y is the resulting elongation in thousandths of an inch:

   X: 1 2 3 4 5 6

   Y: 14 33 40 63 76 85

   a) Graph the data to verify that it is reasonable to assume that the regression of Y on x is linear.

   b) Find the equation of the least squares line, and use it to predict the elongation when the tensile force is 3.5 thousand pounds.

a.PLEASE U CAN DRAW GRAPH BY OWN

Plot the given x*x* and y*y* values on a scatter plot. The points should roughly form a straight line, indicating a linear relationship.

b.

b) Find the Least Squares Line and Predict Elongation

1. **Calculate the necessary sums:**
   - $n = 6$
   - $\sum x = 1 + 2 + 3 + 4 + 5 + 6 = 21$
   - $\sum y = 14 + 33 + 40 + 63 + 76 + 85 = 311$
   - $\sum xy = (1 \times 14) + (2 \times 33) + (3 \times 40) + (4 \times 63) + (5 \times 76) + (6 \times 85) = 14 + 66 + 120 + 252 + 380 + 510 = 1342$
   - $\sum x^2 = 1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2 = 1 + 4 + 9 + 16 + 25 + 36 = 91$
2. **Calculate the slope (b) and intercept (a) of the least squares line:**
   - $b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = \frac{6 \times 1342 - 21 \times 311}{6 \times 91 - 21^2} = \frac{8052 - 6531}{546 - 441} = \frac{1521}{105} \approx 14.4857$
   - $a = \frac{\sum y - b\sum x}{n} = \frac{311 - 14.4857 \times 21}{6} = \frac{311 - 304.2}{6} \approx \frac{6.8}{6} \approx 1.1333$
3. **Equation of the least squares line:**
   - $y = 1.1333 + 14.4857x$
4. **Predict elongation for $x = 3.5$:**
   - $y = 1.1333 + 14.4857 \times 3.5 \approx 1.1333 + 50.7 \approx 51.8333$

So, the predicted elongation is approximately **51.83** thousandths of an inch.

6. . A professor in the school of business in a university polled a dozen colleagues about the number of professional meetings professors attended in the past five years (x) and the number of papers submitted by those to referred journals (y) during the same period. The summary data are given as follows: n=12, $\overline{x}$ = 4,$\overline{y}$ = 12,∑ xi 2 = 232,∑ xiyi = 318. n i=1 Fit a straight line to the given data.

To fit a straight line $y = a + bx$ to the given data, we use the least squares method. The formulas for the slope b and intercept a are:

$$b = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

$$a = \bar{y} - b\bar{x}$$

Given Data:

- $n = 12$
- $\bar{x} = 4$
- $\bar{y} = 12$
- $\sum x_i^2 = 232$
- $\sum x_i y_i = 318$

First, we compute $\sum x_i$ and $\sum y_i$:

$$\sum x_i = n\bar{x} = 12 \times 4 = 48$$

$$\sum y_i = n\bar{y} = 12 \times 12 = 144$$

First, we compute $\sum x_i$ and $\sum y_i$:

$$\sum x_i = n\bar{x} = 12 \times 4 = 48$$

$$\sum y_i = n\bar{y} = 12 \times 12 = 144$$

Now, we compute the slope b:

$$b = \frac{12(318) - (48)(144)}{12(232) - (48)^2}$$

$$b = \frac{3816 - 6912}{2784 - 2304}$$

$$b = \frac{-3096}{480} = -6.45$$

Now, we compute the intercept a:

$$a = 12 - (-6.45 \times 4)$$

$$a = 12 + 25.8 = 37.8$$

Final Equation:

$$y = 37.8 - 6.45x$$

VIVA:

1. **What is linear regression, and how is it used in statistics?**
   - Linear regression is a statistical method used to model the relationship between a dependent variable (y) and one or more independent variables (x). It is used to predict y based on x by fitting a straight-line equation:

     $$y = a + bx$$

     where a is the intercept and b is the slope.
2. **How is the correlation coefficient interpreted?**
   - The correlation coefficient (r) measures the strength and direction of a linear relationship between two variables:
     - $r = 1$ → Perfect positive correlation
     - $r = -1$ → Perfect negative correlation
     - $r = 0$ → No correlation
       Values closer to ±1 indicate a stronger relationship.
3. **State the difference between Linear and Non-linear regression.**
   - **Linear Regression:** Assumes a straight-line relationship between variables ($y = a + bx$).
   - **Non-Linear Regression:** Models a curved relationship using polynomial, exponential, or other complex functions.