Q1. Pearson correlation coefficient is a measure of the linear relationship between two variables. Suppose

you have collected data on the amount of time students spend studying for an exam and their final exam

scores. Calculate the Pearson correlation coefficient between these two variables and interpret the result.

Ans:

1. Organize your data into pairs (x, y), where x represents the amount of time spent studying, and y represents the final exam scores.
2. Calculate the mean (average) of x (denoted as $\bar{x}$) and the mean of y (denoted as $\bar{y}$).
3. For each pair (xi, yi), calculate the deviations from the mean for both x ($x_i - \bar{x}$) and y ($y_i - \bar{y}$).
4. Multiply the deviations for each pair: $(x_i - \bar{x})(y_i - \bar{y})$ for all pairs.
5. Sum up all the products from step 4.
6. Calculate the standard deviation of x (denoted as s_x) and the standard deviation of y (denoted as s_y).
7. Divide the sum from step 5 by the product of s_x and s_y.
8. The result is the Pearson correlation coefficient (r).

The formula for the Pearson correlation coefficient (r) is:

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

Once you have the value of the correlation coefficient (r), you can interpret it as follows:

- $r = 1$: Perfect positive correlation (as one variable increases, the other variable increases proportionally).
- $r = -1$: Perfect negative correlation (as one variable increases, the other variable decreases proportionally).
- $r = 0$: No linear correlation.

Q2. Spearman's rank correlation is a measure of the monotonic relationship between two variables. Suppose you have collected data on the amount of sleep individuals get each night and their overall job satisfaction level on a scale of 1 to 10. Calculate the Spearman's rank correlation between these two variables and interpret the result.

Ans:- Rank the data for both variables separately. Assign ranks starting from 1 to the smallest value, and continue in ascending order. In case of ties, assign the average rank.

Calculate the differences between the ranks of corresponding pairs (d = rank_sleep - rank_satisfaction).

Square each difference (d^2).

Sum up all the squared differences.

Use the formula for Spearman's rank correlation coefficient (ρ):

Q3. Suppose you are conducting a study to examine the relationship between the number of hours of exercise per week and body mass index (BMI) in a sample of adults. You collected data on both variables for 50 participants. Calculate the Pearson correlation coefficient and the Spearman's rank correlation between these two variables and compare the results.

Ans:-

1. **Pearson Correlation Coefficient (r):**

   Use the Pearson correlation coefficient formula:

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

   Follow the steps mentioned earlier for calculating Pearson correlation.

2. **Spearman's Rank Correlation Coefficient (ρ):**

   Follow these steps:

   a. Rank the data for both variables separately.

   b. Calculate the differences between the ranks of corresponding pairs (d = rank_hours - rank_BMI).

   c. Square each difference (d^2).

   d. Sum up all the squared differences.

   e. Use the formula for Spearman's rank correlation coefficient:

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

   where n is the number of pairs.

Now, compare the results:

- If the Pearson correlation coefficient (r) is close to 1 or -1, it suggests a strong linear relationship.
- If the Spearman's rank correlation coefficient (ρ) is close to 1 or -1, it suggests a strong monotonic relationship.

Q4. A researcher is interested in examining the relationship between the number of hours individuals spend watching television per day and their level of physical activity. The researcher collected data on both variables from a sample of 50 participants. Calculate the Pearson correlation coefficient between these two variables.

Ans :-

1. **Calculate the means ($\bar{x}$ and $\bar{y}$):**

   $\bar{x} = \frac{\sum x}{n}$

   $\bar{y} = \frac{\sum y}{n}$

2. **Calculate the deviations from the means ($x_i - \bar{x}$ and $y_i - \bar{y}$):**

   $\text{Deviation from mean of } x = x_i - \bar{x}$

   $\text{Deviation from mean of } y = y_i - \bar{y}$

3. **Multiply the deviations for each pair: $(x_i - \bar{x})(y_i - \bar{y})$ for all pairs.**

4. **Sum up all the products from step 3.**

5. **Calculate the standard deviation of x (s_x) and the standard deviation of y (s_y):**

   $s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$

   $s_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n}}$

6. **Divide the sum from step 4 by the product of s_x and s_y:**

   $r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{s_x \cdot s_y}$

The result r is the Pearson correlation coefficient.

5. A survey was conducted to examine the relationship between age and preference for a particular brand of soft drink. The survey results are shown below:

Age(Years)

25 42 37 19 31 28

Soft drink

Preference Coke Pepsi Mountain dew Coke Pepsi Coke

Ans:-

have the age data paired with the corresponding soft drink preferences. To analyze the relationship between age and soft drink preference, we can create a frequency distribution for the soft drink preferences within different age groups. Here's the organized data:

Age (Years)Soft Drink Preference25Coke42Pepsi37Mountain Dew19Coke31Pepsi28CokeAge (Years)25 4237193128Soft Drink PreferenceCokePepsiMountain DewCokePepsiCoke

Now, let's calculate the frequency distribution for each soft drink preference:

Coke: 3 occurrences (25, 19, 28 years old)

Pepsi: 2 occurrences (42, 31 years old)

Mountain Dew: 1 occurrence (37 years old)

Q6. A company is interested in examining the relationship between the number of sales calls made per day and the number of sales made per week. The company collected data on both variables from a sample of 30 sales representatives. Calculate the Pearson correlation coefficient between these two variables.

Ans:-

1. **Calculate the means ($\bar{x}$ and $\bar{y}$):**
   $\bar{x} = \frac{\sum x}{n}$
   $\bar{y} = \frac{\sum y}{n}$
2. **Calculate the deviations from the means ($x_i - \bar{x}$ and $y_i - \bar{y}$):**
   $\text{Deviation from mean of } x = x_i - \bar{x}$
   $\text{Deviation from mean of } y = y_i - \bar{y}$
3. **Multiply the deviations for each pair: $(x_i - \bar{x})(y_i - \bar{y})$ for all pairs.**
4. **Sum up all the products from step 3.**
5. **Calculate the standard deviation of x (s_x) and the standard deviation of y (s_y):**
   $s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$
   $s_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n}}$
6. **Divide the sum from step 4 by the product of s_x and s_y:**
   $r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{s_x \cdot s_y}$

The result r is the Pearson correlation coefficient.

Interpretation:

- $r > 0$: Positive correlation (as the number of sales calls made per day increases, the number of sales made per week tends to increase).