Correlation and Covariance

Q1. Define Covariance and explain how it differs from Correlation in terms of scale and interpretation.

Answer:
Covariance is a statistical measure that indicates the direction of the linear relationship between two variables. It shows whether the variables move together or in opposite directions.

Difference between Covariance and Correlation:
• Covariance indicates direction only (positive or negative) but its value depends on the units of measurement, so it is difficult to interpret.
• Correlation is a standardized measure of covariance. Its value always lies between –1 and +1, making it easier to interpret and compare.

Thus, correlation is preferred because it is unit-free and more interpretable.

Q2. What does a positive, negative, and zero covariance indicate about the relationship between two variables?

Answer:
• Positive Covariance: Both variables increase or decrease together.
• Negative Covariance: One variable increases while the other decreases.
• Zero Covariance: There is no linear relationship between the variables.

Q3. Discuss the limitations of covariance as a measure of relationship between two variables. Why is correlation preferred in many cases?

Answer:
Limitations of covariance:
1. It depends on the units of measurement.
2. It does not indicate the strength of the relationship.
3. It is difficult to compare across different datasets.

Why correlation is preferred:
• Correlation is unit-free.
• It shows both direction and strength of the relationship.
• Values lie between –1 and +1, making interpretation easy.

Q4. Explain the difference between Pearson's correlation coefficient and Spearman's rank correlation coefficient. When would you prefer Spearman's correlation?

Answer:
Pearson Correlation:
• Measures linear relationship
• Uses actual data values
• Sensitive to outliers
• Assumes normal distribution

Spearman Rank Correlation:
• Measures monotonic relationship
• Uses ranks instead of actual values
• Less sensitive to outliers
• No need for normality

Spearman's correlation is preferred when:
• Data is not normally distributed
• Relationship is non-linear
• Data contains outliers or is ordinal.

Q5. If the correlation coefficient between two variables X and Y is 0.85, interpret this value. Can you infer causation from this value? Why or why not?

Answer:
A correlation coefficient of 0.85 indicates a strong positive relationship between X and Y. As X increases, Y also tends to increase.

However, correlation does not imply causation. This value does not prove that changes in X cause changes in Y, because other variables may influence both.

Q6. Using the dataset below, calculate the covariance between X and Y.

Answer:
Covariance is calculated using the formula:

$$Cov(X,Y) = \Sigma[(X - \bar{X})(Y - \bar{Y})] / n$$

After calculating deviations from the mean and multiplying them, the resulting covariance value determines the direction of the relationship:

• Positive value → positive relationship
• Negative value → negative relationship

(Exact numerical result depends on the given dataset.)

Q7. Compute the Pearson correlation coefficient between variables A and B.

Answer:
Pearson correlation coefficient is calculated as:

$r = Cov(A,B) / (\sigma A \, \sigma B)$

After computing covariance and standard deviations of A and B, the correlation value indicates:
• r close to +1: strong positive correlation
• r close to –1: strong negative correlation
• r close to 0: weak or no linear relationship

Q8. The following table shows heights (in cm) and weights (in kg) of 5 students. Find the correlation coefficient between Height and Weight.

Answer:
Height-Weight Data:
150–50
160–55
165–58
170–62
180–70

There is a strong positive correlation between height and weight. As height increases, weight also increases. The correlation coefficient value will be close to +1.

Q9. Given the dataset below, determine whether there is a positive or negative correlation between X and Y. (No need for exact calculation, just reasoning.)

Answer:
Dataset 1:
X: 2, 4, 6, 8
Y: 3, 7, 5, 10
→ Positive correlation (both increase together)

Dataset 2:
X: 1, 2, 3, 4, 7
Y: 15, 12, 9, 5, 3
→ Negative correlation (X increases, Y decreases)

Q10. Two investment portfolios have returns (%) over 5 years. Compute covariance and correlation coefficient and interpret whether the portfolios move together.

Answer:
Portfolio returns:
Year 1: A=6, B=8
Year 2: A=10, B=9
Year 3: A=12, B=11
Year 4: A=9, B=8
Year 5: A=11, B=10

• The covariance is positive, indicating that both portfolios move in the same direction.
• The correlation coefficient is also positive and high, showing a strong relationship.

Interpretation:
The two portfolios tend to move together, meaning when returns of one portfolio increase, the other also increases.