Name – MD Anas

Course Name - Data analytics with GenAi

Topic - correlation and covariance

Email id - anasanwar1120@gmail.com

Question 1 Define Covariance and explain how it differs from Correlation in terms of scale and interpretation.

Answer Covariance

- Definition: A statistical measure of the joint variability of two random variables, showing if they tend to increase or decrease together.
- Scale: Varies widely; its value depends on the units of the variables (e.g., dollars, hours) and their scales, so a large covariance doesn't necessarily mean a strong relationship.
- Interpretation: A positive value means variables move in the same direction; negative means opposite directions; zero means no *linear* relationship. Magnitude is hard to interpret.

Correlation

- Definition: A standardized measure derived from covariance, expressing the strength and direction of the linear relationship.
- Scale: Unitless and bounded between -1 and +1.
- Interpretation:
    - +1: Perfect positive linear relationship.
    - -1: Perfect negative linear relationship.
    - 0: No linear relationship.
    - Values closer to 1 or -1 indicate stronger relationships; values near 0 indicate weaker ones

Question 2What does a positive, negative, and zero covariance indicate about the relationship between two variables?

Answer Positive Covariance:

- Indicates: A tendency for variables to increase or decrease together.
- Example: Height and weight (taller people tend to be heavier).

- Negative Covariance:
  - Indicates: A tendency for variables to move in opposite directions (when one goes up, the other goes down).
  - Example: Outdoor temperature and heating bills (as temperature rises, bills fall).
- Zero Covariance:
  - Indicates: No *linear* relationship between the variables; changes in one are not consistently associated with changes in the other.
  - Note: This doesn't rule out *non-linear* relationships, just linear ones.

Question 3. Discuss the limitations of covariance as a measure of relationship between two variables. Why is correlation preferred in many cases

AnswerCovariance has several drawbacks that limit its usefulness as a measure of relationship:

- Scale Dependence:
  The magnitude of covariance depends on the units of measurement. For example, income in dollars vs. rupees will produce very different covariance values, even if the relationship is the same.
- Unbounded Values:
  Covariance can take any value from negative infinity to positive infinity. This makes it hard to judge whether a relationship is "strong" or "weak".
- Interpretation Difficulty:
  A positive covariance only tells us that variables move in the same direction, and a

negative one that they move in opposite directions. It does not quantify the strength of the relationship in a standardized way.

- Not Comparable Across Datasets:
  Because covariance depends on scale and units, comparing relationships across different datasets or variables is unreliable.

✅ Why Correlation is Preferred

Correlation addresses these limitations by normalizing covariance:

- Unit-Free Measure:
  Correlation divides covariance by the product of the standard deviations of the two variables, removing dependence on units.
- Bounded Range (-1 to +1):
  This makes interpretation straightforward:
  - +1 → perfect positive relationship
  - -1 → perfect negative relationship
  - 0 → no linear relationship
- Comparability:
  Correlation values are standardized, so relationships across different datasets or variables can be meaningfully compared.
- Strength & Direction:
  Unlike covariance, correlation clearly indicates both the direction and strength of the linear relationship.

Question 4. Explain the difference between Pearson's correlation coefficient and Spearman's rank correlation coefficient. When would you prefer to use Spearman's correlation
Answer

Why Correlation is Preferred

Correlation addresses these limitations by normalizing covariance:

- Unit-Free Measure:
  Correlation divides covariance by the product of the standard deviations of the two variables, removing dependence on units.

- Bounded Range (-1 to +1):
  This makes interpretation straightforward:
  - +1 → perfect positive relationship
  - -1 → perfect negative relationship
  - 0 → no linear relationship
- Comparabi lity:
  Correlation values are standardized, so relationships across different datasets or variables can be meaningfully compared.
- Strength & Direction:
  Unlike covariance, correlation clearly indicates both the direction and strength of the linear relationship.

Question 5 If the correlation coefficient between two variables X and Y is 0.85, interpret this value in context. Can you infer causation from this value? Why or why no

Answer interpretation of r = 0.85

- Strength: The value 0.85 is close to 1 (the maximum positive correlation), signifying a strong relationship.
- Direction: The positive sign means it's a positive correlation, so the variables move in the same direction.
- Context: If X were study hours and Y were test scores, a 0.85 correlation would mean more study time strongly predicts higher scores.

Why causation cannot be inferred

- Correlation ≠ Causation: This is a fundamental rule in statistics; a strong association doesn't prove one variable *causes* the other.
- Confounding Variables (Z): A third, unmeasured variable (Z) might be causing both X and Y to change. For example, good nutrition (Z) could lead to better health (X) and higher grades (Y).
- Reverse Causality: The assumed cause (X) might actually be the effect (Y). More sleep (Y) might lead to better performance (X), not the other way around.
- Coincidence: Sometimes, strong correlations occur purely by chance, especially with large datasets.

Question 6 Using the dataset below, calculate the covariance between X and Y
Answer correlation^0covariance.xlsx

Question 7 Compute the Pearson correlation coefficient between variables A and B
Answer correlation^0covariance.xlsx

Question 8 . The following table shows heights (in cm) and weights (in kg) of 5 students. Find the correlation coefficient between Height and Weight
Answer correlation^0covariance.xlsx

Question 9 Given the dataset below, determine whether there is a positive or negative correlation between X and Y. (No need for exact calculation, just reasoning.
Answercorrelation^0covariance.xlsx

Question 10 Two investment portfolios have the following returns (%) over 5 years. Compute the covariance and correlation coefficient, and interpret whether the portfolios move together.
Answercorrelation^0covariance.xlsx