Name: Abhishek Dixit

Batch: Data Analytics Dec Live Batch

Assignment: Covariance & Correlation

Assignment

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

Sol)

- Covariance only tells direction not strength.
- It tells us the direction of relationship (Positive & Negative).

Difference Between Covariance and Correlation:

Basis	**Covariance**	**Correlation**
Scale	Depends on units	Unit-free
Range	No fixed range	-1 to +1
Interpretation	Hard to interpret magnitude	Easy to interpret strength

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

Sol)

Positive Covariance → Variables move in same direction
(If X increases, Y increases)

Negative Covariance → Variables move in opposite direction
(If X increases, Y decreases)

Zero Covariance → No linear relationship.

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

Sol)

Limitations:

1. No fixed range
2. Depends on measurement units
3. Difficult to interpret strength
4. Cannot compare different datasets

That's why correlation is preferred because it is standardized and ranges from 1 to +1.

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

Sol)

Pearson	**Spearman**
• Measures linear relationship	• Measures monotonic relationship
• Uses actual values	• Uses ranks
• Sensitive to outliers	• Less sensitive to outliers
• Requires normal distribution	• No strict distribution assumption

When to Use Spearman:

- When data is not normally distributed
- When dealing with ranks
- When relationship is non-linear but monotonic

Q5. 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 not?

Sol)

0.85 indicates ***strong positive correlation.***

As X increases, Y increases strongly.

We cannot infer causation from this value

Because:

- Correlation does not imply causation
- There may be third variable involved