QUESTION 1 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 tend to increase or decrease together

Two variables X and Y

Positive covariance: Variables move in the same direction

Negative covariance: One variable increases while the other decreases

Zero covariance: No linear relationship

Difference Between Covariance and Correlation		Scale & Interpretation
Aspect	**Covariance**	**Correlation**
Scale	Depends on the units of XXX and YYY	Unit-free (dimensionless)
Range	No fixed range	Always between –1 and +1
Interpretation	Indicates direction only	Indicates both direction and strength
Comparability	Cannot be compared across datasets	Can be compared across datasets
Effect of Unit Change	Changes if units change	Remains unchanged

Covariance is useful for understanding the direction of a relationship.

Correlation offers more practical insight than covariance in most applications.

QUESTION 2		What does a positive, negative, and zero covariance indicate about the relationship between two variables?	
ANSWER		Covariance describes the direction of the linear relationship between two variables. Its sign tells us how the variables move relative to each other	
		Positive Covariance	
		Both variables tend to move in the same direction	
		When one variable increases, the other also increases; when one decreases, the other decreases	
		Example:	Height and weight
		Negative Covariance	
		Variables tend to move in opposite directions	
		When one variable increases, the other decreases	
		Example	Price and quantity demanded
		Zero Covariance	
		There is no linear relationship between the variables.	
		Changes in one variable do not systematically relate to changes in the other.	
		Example:	Shoe size and intelligence

QUESTION 3		Discuss the limitations of covariance as a measure of relationship between two variables. Why is correlation preferred in many cases?
ANSWER		Covariance indicates whether two variables move together (positive covariance) or in opposite directions (negative covariance).
	No Standardized Scale	Covariance has no fixed range.
		Its value depends on the units of measurement of the variables
		A large covariance does not necessarily imply a strong relationship
Difficult to Interpret Magnitude		The scale varies, we cannot compare covariances across different datasets.
		Why Correlation Is Preferred
		Correlation overcomes most of the drawbacks of covariance
		Standardized Measure
		Correlation is unit-free.
		It is calculated by dividing covariance by the product of standard deviations.
		Fixed Range
		Correlation values lie between –1 and +1.

1	positive relationshio
-1	nagetive relationship
0	no liner 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

Both Pearson’s correlation coefficient and Spearman’s rank correlation coefficient measure the relationship between two variables

Pearson’s Correlation Coefficient (r)

Pearson’s correlation measures the strength and direction of a linear relationship between two quantitative (continuous) variable

Based on actual data values
Measures linear relationships only

Example: Relationship between height and weight.

Spearman’s Rank Correlation Coefficient (ρ or r□)

Spearman’s correlation measures the strength and direction of a monotonic relationship using the ranks of data rather than actual values

Based on ranks, not raw data
Measures monotonic (increasing or decreasing) relationships
Non-parametric
Less affected by outliers

When to Prefer Spearman’s Correlation

Data is ordinal (ranked data)
he relationship is non-linear but monotonic
Data contains outliers

Use Pearson’s correlation for linear relationships with normally distributed data.

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 not?
ANSWER		Interpretation of a Correlation Coefficient of 0.85
		(r) = 0.85 indicates a strong positive linear relationship between variables X and Y.
		X increases, Y tends to increase as well.
		The points on a scatter plot would lie close to an upward-sloping straight line
		The relationship is strong, though not perfect (which would be r = 1).
		Can You Infer Causation from This Value?
		No, causation cannot be inferred from correlation alone
		X may affect Y, Y may affect X, or both may influence each other
		Hidden variable may be causing changes in both X and Y.
		A correlation coefficient of 0.85 shows a strong positive association between X and Y.

QUESTION 6

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

X	2	4	6	8
Y	3	7	5	10

ANSWER

No of observation , N= 4

CALCULATE THE MEAN=		**X= 2+4+6+8/4**	**20/4 =5**
		y= 3+7+5+10/4	**25/4=6.25**

Calculate Deviations and Their products

X	**Y**	**(X-MEAN)**	**(Y-MEAN)**	**PRODUCT**
2	3	-3	-3.25	9.75
4	7	-1	0.75	-0.75
6	5	1	-1.25	-1.25
8	10	3	3.75	11.25

SUM OF PRODUCTS

PRODUCT
9.75
-0.75
-1.25
11.25
TOTAL = 19

CALCULATE THE CO VARIANCE

population variance	(X,Y)19/4= 4.75
sample variance	(X,Y) 19/3= 6.33

Population covariance = 4.75

Sample covariance = 6.33

QUESTION 7

Compute the Pearson correlation coefficient between variables A and B:

A	10	20	30	40	50
B	8	14	18	24	28

ANSWER

NUMBERS OF OBSERVATION =5

calculate the mean A= 10+20+30+40+50/5= 150/5= 30	
	B= 8+14+18+24+28/5= 92/5= 18.4

CALCULATE THE STD

A	**B**	**(A-MEAN)**	**(B MEAN)**	**PRODUCT**	**(A-A)*2**	**(B-B)*2**
10	8	-20	-10.4	208	400	108.16
20	14	-10	4.4	44	100	19.36
30	18	0	-0.4	0	0	0.16
40	24	10	5.6	56	100	31.36
50	28	20	9.6	192	400	92.16

∑(A−A¯)(B−B¯)=208+44+0+56+192=500

∑(A−A¯)2=400+100+0+100+400=1000

r=∑(A−A¯)∑(B−B¯)2∑ / (A−A¯)(B−B¯)

r= 500/√1000*25

r=500/√251200=

			r=500/501.2					
			r=0.998					

QUESTION 8

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

HEIGHT	150	160	165	170	180
WEIGHT	50	55	58	62	70

ANSWER

Numbers of observation =5

calculate the mean		150+160+165+170+180/5= 825/5=165		
		50+55+58+62+70/=295/5=59		

calculate the std

students	Heigts	Weight	(Height-mean)	(Weight-mean)	products	(x-x)*2	(Y-Y)*2
1	150	50	-15	-9	135	225	81
2	160	55	-5	-4	20	25	16
3	165	58	0	-1	0	0	1
4	170	62	5	3	15	25	9
5	180	70	15	11	165	225	121

∑(X−X¯)(Y−Y¯)=135+20+0+15+165=335

∑(X−X¯)2=225+25+0+25+225=500

∑(Y−Y¯)2=81+16+1+9+121=228

r=∑(x-x)-(Y-Y)/√∑(X-)X)*2∑(Y-Y)*2

			r=335/√500*228						
			r=335/√114000						
			r=335/√337.64						
			r=0.99						

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.)

X	1	2	3	4	5
Y	15	12	9	7	3

ANSWER

As X increases from 1 to 5,

Y consistently decreases from 15 to 3.

When one variable goes up, the other goes down.
There is a negative correlation between X and Y.

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.

YEAR	PORTFOLIO A	PORTFOLIO B
1	**8**	**6**
2	**10**	**9**
3	**12**	**11**
4	**9**	**8**
5	**11**	**10**

ANSWER

calculate the mean	
X=	8+10+12+9+11/5= 50/5= 10
Y=	6+9+11+8+10/5= 44/8= 8.8

YEAR	A	B	(X-MEAN)	(Y-MEAN)	PRODUCT	(X-X)*2	(Y-Y)*2
1	8	6	-2	-2.8	5.6	4	7.84
2	10	9	0	0.2	0	0	0.04
3	12	11	2	2.2	4.4	4	4.84
4	9	8	-1	0.8	0.8	1	0.64
5	11	10	1	1.2	1.2	1	1.44

∑(X−X⁻)(Y−Y⁻)=5.6+0+4.4+0.8+1.2=12

∑(X−X⁻)2=4+0+4+1+1=10

∑(Y−Y⁻)2=7.84+0.04+4.84+0.64+1.44=14.8

COV (X,Y) = 12/N-1=12/4= 3

r= ∑(X−X⁻)(Y−Y⁻)/√∑(x-x)*2∑(y-y)*2

r=12/√10*14.8

12/√148

12/12.17 **r=0.99**