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

Answer:

Covariance

Definition of Covariance

Covariance is a statistical measure that indicates the direction of the relationship between two variables.

- A positive covariance means both variables increase or decrease together
- A negative covariance means one variable increases while the other decreases
- A covariance of zero indicates no linear relationship

Difference Between Covariance and Correlation

1. Scale

- Covariance:
  The value of covariance depends on the units of measurement of the variables, so it does not have a fixed range.
- Correlation:
  Correlation is unit-free and always lies between –1 and +1.

2. Interpretation

- Covariance:
  Shows only the direction of the relationship, not the strength.
- Correlation:
  Shows both the direction and strength of the relationship.

Summary Table

Covariance	**Correlation**
Depends on units	Unit-free
No fixed range	Range from –1 to +1
Indicates direction only	Indicates direction and strength
Harder to interpret	Easier to interpret

Answer

Covariance measures the direction of the relationship between two variables but depends on the scale of measurement. Correlation standardizes covariance, making it easier to interpret the strength and direction of the relationship.

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

Answer:

Interpretation of Covariance

Positive Covariance

- A positive covariance indicates that both variables move in the same direction.
- When one variable increases, the other also tends to increase.

Example:
Income and expenditure.

Negative Covariance

- A negative covariance indicates that the variables move in opposite directions.
- When one variable increases, the other tends to decrease.

Example:
Price of a product and its demand.

Zero Covariance

- A zero covariance indicates no linear relationship between the variables.
- Changes in one variable do not show a consistent pattern with the other.

Example:
Shoe size and exam marks.

Answer

A positive covariance shows that variables move together, a negative covariance shows they move in opposite directions, and a zero covariance indicates 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. Depends on Scale and Units

Covariance is affected by the units of measurement of the variables.
Changing the scale (for example, from meters to centimeters) changes the value of covariance, making it difficult to compare results.

2. No Fixed Range

Covariance does not have a fixed range, so its magnitude cannot be easily interpreted as weak or strong relationship.

3. Difficult Interpretation

Covariance only shows the direction of the relationship (positive or negative) but does not show the strength of the relationship clearly.

Why Correlation Is Preferred

Correlation is preferred because:

- It is unit-free
- It has a fixed range between –1 and +1
- It clearly shows both direction and strength of the relationship
- It allows easy comparison between different datasets

Answer

Covariance has limitations because it depends on the scale of measurement and does not indicate the strength of a relationship clearly. Correlation overcomes these limitations by standardizing covariance, making it easier to interpret and compare, which is why correlation is preferred in many cases.

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

Answer:

Pearson's vs Spearman's Correlation

Pearson's Correlation Coefficient

- Measures the linear relationship between two continuous variables
- Assumes data is normally distributed
- Sensitive to outliers
- Uses actual data values

Example:
Relationship between height and weight.

Spearman's Rank Correlation Coefficient

- Measures the monotonic relationship between variables
- Does not require normal distribution
- Less affected by outliers
- Uses ranks of data instead of actual values

Example:
Relationship between students' ranks and their marks.

Key Differences

Pearson's Correlation	Spearman's Correlation
Uses actual values	Uses ranked values
Measures linear relationship	Measures monotonic relationship
Requires normality	No normality assumption
Sensitive to outliers	Less sensitive to outliers

When Spearman's Correlation Is Preferred

Spearman's correlation is preferred when:

- Data is not normally distributed
- Relationship is non-linear but monotonic
- Data is ordinal or ranked
- Outliers are present

Answer

Pearson's correlation measures linear relationships using actual data values, while Spearman's correlation measures monotonic relationships using ranks. Spearman's correlation is preferred when data is non-normal, ranked, or contains outliers.

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?

Answer:

Interpretation of Correlation Coefficient (r = 0.85)

Interpretation

A correlation coefficient of 0.85 indicates a strong positive relationship between variables X and Y.

This means:

- As X increases, Y also tends to increase
- The relationship between the variables is strong, but not perfect

Causation vs Correlation

Even though the correlation is high, causation cannot be inferred.

This is because:

- Correlation only measures association, not cause and effect
- A third variable may influence both X and Y
- The relationship could be coincidental

Answer

A correlation coefficient of 0.85 shows a strong positive relationship between X and Y. However, this does not imply causation because correlation does not establish a cause-and-effect relationship, and other factors may be involved.

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

X	2	4	6	8
Y	3	7	5	10

Answer:

Given Data

X 2 4 6 8

Y 3 7 5 10

Number of observations n=4

Step 1: Calculate the Mean of X and Y

$$\bar{x}=\frac{2+4+6+8}{4}=\frac{20}{4}=5$$
$$\bar{y}=\frac{3+7+5+10}{4}=\frac{25}{4}= 6.25$$

Step 2: Calculate Deviations and Their Products

X	Y	$X-\bar{x}$	$Y-\bar{y}$	$(X-\bar{X})(Y-\bar{Y})$
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:

9.75−0.75−1.25+11.25=19

Step 3: Calculate Covariance

$$\text{cov}(x,y) = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n}$$
$$=\frac{19}{4}= 4.75$$

Final Answer

The covariance between X and Y is 4.75.

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

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

Answer:

Given Data

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

Number of observations n=5

Step 1: Calculate Means

$$\overline{A} = \frac{10+20+30+40+50}{5} = \frac{150}{5} = 30$$

$$\overline{B} = \frac{8+14+18+24+28}{5} = \frac{92}{5} = 18.4$$

Step 2: Calculate Deviations and Products

A	B	A−$\overline{A}$	B−$\overline{B}$	Product
10	8	-20	-10.4	208
20	14	-10	-4.4	44
30	18	0	-0.4	0

40	24	10	5.6	56
50	28	20	9.6	192

$\sum(A-\overline{A})(B-\overline{B})=500$

Step 3: Calculate Sum of Squares

$\sum(A-\overline{A})^2=1000$

$\sum(B-\overline{B})^2=251.2$

Step 4: Pearson Correlation Coefficient

$$r=\frac{\Sigma(A-\overline{A})(B-\overline{B})}{\sqrt{\Sigma(A-\overline{A})^2\,\Sigma(B-\overline{B})^2}}$$

$$r=\frac{500}{\sqrt{1000\times 251.2}}\approx 0.998$$

Final Answer

The Pearson correlation coefficient between A and B is approximately 0.998.

Q8. 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:

Given Data

Height (cm)	150	160	165	170	180
Weight (kg)	50	55	58	62	70

Number of observations n=5

Step 1: Calculate Means

$$\overline{H} = \frac{150+160+165+170+180}{5} = \frac{825}{5} = 165$$
$$\overline{W} = \frac{50+55+58+62+70}{5} = \frac{295}{5} = 59$$

Step 2: Calculate Deviations and Products

H	W	H−$\overline{H}$	W−$\overline{W}$	Product
150	50	-15	-9	135
160	55	-5	-4	20
165	58	0	-1	0
170	62	5	3	15
180	70	15	11	165

∑(H−$\overline{H}$)(W−$\overline{W}$)=335

Step 3: Calculate Sum of Squares

$\sum(H - \overline{H})^2$=500

$\sum(W - \overline{W})^2$=228

Step 4: Pearson Correlation Coefficient

r=$\frac{\Sigma(H-\overline{H})(W-\overline{W})}{\sqrt{\Sigma(H-\overline{H})^2\Sigma(W-\overline{W})^2}}$

r =$\frac{335}{\sqrt{500\times228}}\approx\frac{335}{337.7}\approx0.99$

Final Answer

The correlation coefficient between Height and Weight is approximately 0.99.

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.)

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

Answer:

Given Data

X 1 2 3 4 5

Y 15 12 9 7 3

Reasoning

As the values of X increase from 1 to 5, the values of Y decrease from 15 to 3.
This shows that the two variables move in opposite directions.

Final Answer

There is a negative correlation between X and Y.

Q10. 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:

Given Data

Year	Portfolio A	Portfolio B
1	8	6
2	10	9
3	12	11
4	9	8
5	11	10

Number of observations n=5

Step 1: Calculate Means

$$\overline{A} = \frac{8+10+12+9+11}{5} = \frac{50}{5} = 10$$

$$\overline{B} = \frac{6+9+11+8+10}{5} = \frac{44}{5} = 8.8$$

Step 2: Deviations and Products

A	B	A−$\overline{A}$	B−$\overline{B}$	Product
8	6	-2	-2.8	5.6
10	9	0	0.2	0
12	11	2	2.2	4.4
9	8	-1	-0.8	0.8
11	10	1	1.2	1.2

$$\sum(A-\overline{A})(B-\overline{B})=12$$

Step 3: Covariance

$$\mathrm{Cov}(A,B)=\frac{12}{5}=2.4$$

Step 4: Correlation Coefficient

$$\sum(A-\overline{A})^2 = 10$$

$$\sum(B-\overline{B})^2 = 14.8$$

$$r=\frac{12}{\sqrt{10\times14.8}} \approx 0.99$$

Final Answer

- Covariance = 2.4
- Correlation coefficient ≈ 0.99