Correlation and Covariance

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

Ans - Covariance tells us how two things change together.

- If both increase or decrease together → positive covariance
- If one increases while the other decreases → negative covariance
- If they don't move together → covariance near zero.

Covariance tells whether two things move together.

Correlation tells how strongly they move together and is easier to understand because it is always between -1 and +1.

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

Ans-

- Positive covariance:
  Both things go up or down together.
- Negative covariance:
  One thing goes up while the other goes down.
- Zero covariance:
  The two things have no connection at all.

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

Ans – 1. Covariance is hard to interpret because it has no fixed range.
2. It depends on the units of measurement.
3. It only shows the direction, not the strength, of the relationship.

Why Correlation is Preferred

1. Correlation has a fixed range from –1 to +1.
2. It is independent of units and easy to understand.
3. It shows both direction and strength of the relationship.

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

Ans - Difference between Pearson's and Spearman's correlation

Pearson's correlation measures the relationship between two variables using their actual values.

It works best when the relationship is linear and the data is normally distributed.

Spearman’s rank correlation measures the relationship using ranks instead of actual values. It checks whether the variables move in the same order.

When Spearman’s correlation is preferred

Spearman’s correlation is used when the data is not normally distributed, contains outliers, or when the relationship is not linear. It is also used for ranked or ordinal data.

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?

Ans - If the correlation coefficient between X and Y is 0.85, it means there is a strong positive relationship between them. When X increases, Y also tends to increase. causation cannot be inferred from this value. Correlation only shows that the variables move together, not that one causes the other. The relationship could be due to another factor or coincidence.

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

Ans-

$$\text{Cov}(X, Y) = \Sigma [(X_i - \bar{X}) * (Y_i - \bar{Y})] / (n - 1)$$

X – 2, 4, 6, 8

Y – 3, 7, 5, 10

$\bar{X}$ = 2+4+6+8=20/4 = 5

$\bar{Y}$ = 3+7+5+10+25/4 = 6.25

$X_i - \bar{X}$ $Y_i - \bar{Y}$

2-5=-3 3-6.5=-3.5

4-5=-1 7 -6.25= 0.75

6-5=1 5-6.25=-1.25

8-5=3 10-6.5=3.5

-3*-3.25=9.75

-1*0.75=-0.75

1*(-1.2.5)=-1.25

3*3.75=11.25

9.75+-0.75+-1.25+11.25 = 19

19/(4-1) = 19/3 = 6.33

So, the covariance is 6.33.

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

A -10,20,30,40,50

B-8,14,18,24,28

Let's find mean of A and B

A = 10+20+30+40+50/5 = 30

B = 8+14+18+24+28=18.4

Now, we will subtract each value with their mean.

For A,

10-30= -20

20-30= -10

30-30= 0

40-30= 10

50-30= 20

For B,

8-18.4= -10.4

14-18.4= -4.4

18-18.4= -0.4

24-18.4= 5.6

28-18.4= 9.6

Now we will multiple,

-20*-10.4= 208

-10*-4.4= 44

0*-0.4= 0

10*5.6= 56

20*9.6= 192

_____

Total = 500

Squares of A

-20=400

-10=100

0=0

10=100

20=400

_____

Total = 1000

Squares of B

-10.4=108

-4.4=19.36

-0.4=0.16

5.6=31.36

9.6=92.16

_____

Total = 251.2

= 500/ √1000*251.2

= 500/501.2

= 0.997 – Very strong positive correlation.

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

Mean of

Height - 165

Weight – 59

Now, we will subtract values with mean

Height,

150-165=-15

160-165=-5

165-165=0

170-165=5

180-165-15

Weight,

50-59=-9

55-59=-4

58-59=-1

62-59=3

70-59=11

H*W

-15*-9=135

-5*-4=20

0*-1=0

5*3=15

15*11=165

_____

= 335

Now, lets find square root of values,

Height,

-15=225

-5=25

0=0

5=25

15=225

Weight,

-9=81

-4=16

-1=1

3=9

11=121

_____

= 228

= 335/√500*228

= 335/√114000

=335/337.64

=0.992

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+5/5, Mean of x = 3

Y =15+12+9+7+3/5, Mean of Y = 9.2

X

1-3 = -2

2-3 = -1

3-3 = 0

4-3 = 1

5-3 = 2

Y

15-9.2 = 5.8

12-9.2 = 2.8

9-9.2 = -0.2

7-9.2 = -2.2

3-9.2 = -6.2

-2*5.8 = -11.6

-1*2.8 = -2.8

0*-0.2 = 0

1*-2.2 = -2.2

2*-6.2 = -12.4

_____

Total = -29

Sum of (X-mean of X)*(Y-mean of Y) is negative hence 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 = 1, 2, 3, 4, 5

Portfolio A = 8, 10, 12, 9, 11

Portfolio B = 6, 9, 11, 8, 10

Mean of Portfolio A = 10

Mean of portfolio B = 8.8

A

8-10=-2

10-10=0

12-10=2

9-10=1

11-10=1

B

6-8.8=-2.8

9-8.8=0.2

11-8.8=2.2

8-8.8=-0.8

10-8.8=1.2

-2*-2.8=5.6

0*0.2=0

2*2.2=4.4

-1*0.8=-0.8

1*1.2=1.2

_____

Total = 12/5

Covariance = 2.4

Now, let’s find square of value A and B

X

-2 = 4

0 = 0

2 = 4

-1 = 1

1 = 1

Total = 10/5

SD $\sqrt{2}$ = 1.414

Y

-2.8 = 7.84

0.2 = 0.04

2.2 = 4.84

-0.8 = 0.64

1.2 = 1.44

_____

Total = 14.8/5

SD$\sqrt{2.96}$ = 1.72

R = Covariance/SDA*SDB

R= 2.4/1.414*1.72

=2.4/2.43

=0.99

Covariance is positive and Correlation coefficient is close to +1.