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

- ➢ Covariance is a statistical measure that tells us how two variables change together.
  The variable of a positive covariance has the tendency to move in the same direction.
  The variable of a negative covariance has the tendency to move in the opposite direction. And the zero covariance doesn't tend to move together in a predictable way.
  Covariance differs from correlation:
  1. In terms of scale : Covariance has an infinite range and the value depends on the units of measured Variables whereas correlation does not get affected by Scale of the Variable.
  2. In terms of interpretation : In Covariance the tendency of a positive value is to move in the Same direction while a negative, value move in opposite direction, whereas in Corrententia a value of +1 indicates a perfect positive linear relation, -1 indicates a perfect negative linear relation and 0 indicates no linear relation.

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

- ➢ The positive covariance moves in the same direction . The variables increase or decrease together .
- ➢ The negative covariance moves in the opposite direction. If a variable increases the other variable tends to decrease.
- ➢ The zero covariance does not move together in a linear pattern. The variables are not co-related.

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

- ➢ Covariance depends on the units of the variables, so its value can change if we change the measurement units. Because of this, we cannot compare covariance across different datasets, and a high or low value doesn’t clearly show the strength of the relationship.

  Correlation is preferred as it solves the problem of scale dependence. It is unitless and is always, .Bounded between -1 and +1, which make it a standardised measure. Because of the fixed scale it can easily interpret the strength of the 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?

Pearson's correlation coefficient	Spearman's correlation coefficient
• It measures the strength and direction of the linear relationship between two variables. • In this the data should be normally distributed and measured on ratio scale.	• It measures the strength and direction of the monotonic relationship. • It can be used for ordinal data or for the data that are not normally distributed.

- Use of spearman's rank correlation:

    - ➢ When the continuous data is not normally distributed.
    - ➢ It can be a good choice for initial analysis where the exact form of relationship is not certain.
    - ➢ When one or both the variables are ordidinal, i.e. they can be ranked but the distance between them is not fixed.

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?

➢ As 0.85 is very close to +1 and the sign is also positive, the correlation coefficient of 0.85 indicates a strong positive linear relationship between the two variables, X and Y.
No, we cannot infer causation from a correlation coefficient value of 0.85, no matter how strong the correlation is. Because r=0.85, only tells that two variables are strongly associated or related, but it does not provide any evidence that a change in one variable causes a change in the other too.

Question 6: Using the dataset below, calculate the covariance between X and Y.

X	2	4	6	8
Y	3	7	5	10

➢ Mean of X:

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

Mean of Y:

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

X	Y	X-$\overline{X}$	Y-$\overline{Y}$	(X-$\overline{X}$)(Y-$\overline{Y}$)
2	3	2-5=-3	3-6.25=-3.25	(-3)*(-3.25)=9.75
4	7	4-5=-1	7-6.25=0.75	(-1)*(0.75)=-0.75
6	5	6-5=1	5-6.25=-1.25	(1)*(-1.25)= -1.25
8	1	8-5=3	1-6.25=-5.25	(3)*(3.75)=11.25

Calculation of covariance:

Σ=(X-X)(Y-Y)=9.75−0.75−1.25+11.25=19.00

Cov(X,Y)=$\frac{19}{4-1}$ = 19/3 = 6.33

Questions 7.Compute the Pearson correlation coefficient between variables A and B:

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

Mean of A:

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

Mean of B:

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

A	B	A-X	B-Y	(A-X)^2	(B-Y)^2	(A-X)(B-Y)
10	8	-20	-10.4	400	108.06	208
20	14	-10	-4.4	100	19.36	44
30	18	0	-0.4	0	0.16	0
40	24	10	5.6	100	31.36	56
50	28	20	9.6	200	92.16	192
Total		0	0	1000	251.2	500

r = $\frac{\Sigma(A-X)(B-Y)}{\sqrt{\Sigma(A-X)2*\Sigma(B-Y)2}}$ = $\frac{500}{\sqrt{1000*251.2}}$ = $\frac{500}{501.1985}$ = 0.9976

Hence, the Pearson correlation coefficient between A and B is approximately 0.9976, which shows an extremely strong positive linear relationship.

Question 8. The Pearson correlation coefficient between A and B is approximately 0.9976. This indicates an extremely strong positive linear relationship.

Height(H)	**150**	**160**	**165**	**170**	**180**
Weight(W)	**50**	**55**	**58**	**62**	**70**

Mean of H:

h = $\frac{150+160+165+170+180}{5}$ = $\frac{825}{5}$ =165cm

Mean of W:

w = $\frac{50+55+58+62+70}{5}$ = $\frac{295}{5}$ = 59kg

H	W	H-h	W-w	(H-h)2	(W-w)2	(H-h)(W-w)
150	50	-15	-9	225	81	135
160	55	-5	-4	25	16	20
165	58	0	-1	0	1	0
170	62	5	3	25	9	15
180	70	15	11	225	121	165
Total		0	0	500	228	335

r = $\frac{\Sigma(H-h)(W-w)}{\sqrt{\Sigma(H-h)2*\Sigma(W-w)2}}$ = $\frac{335}{\sqrt{500*228}}$ = $\frac{335}{337.6389}$ = 0.9922(approx).

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**

➢ There is a negative correlation between the two variables (x and y). Because there is an increase in one variable (value of X is increasing from 1 to 5), whereas there is a decrease in the other variable (value of Y is decreasing from 15 down to 3).

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	1	2	3	4	5
PortfolioA(A)	**8**	**10**	**12**	**9**	**11**
PortfolioB(B)	**6**	**9**	**11**	**8**	**10**

Mean of A:

$$a = \frac{8+10+12+9+11}{5} = 50/5 = 10$$

Mean of B:

$$b = \frac{6+9+11+8+10}{5} = 44/5 = 8.8$$

A	B	A-a	B-b	(A-a)2	(B-b)2	(A-a)(B-b)
8	6	-2	-2.8	4	7.84	5.6
10	9	0	0.2	0	0.04	0
12	11	2	2.2	4	4.84	4.4
9	8	-1	-0.8	1	0.64	0.8
11	10	1	1.2	1	1.44	1.2
Total		0	0	10	14.8	12

Covariance:

cov(A,B) = 12/4 = 3

Correlation:

$$r = \frac{\Sigma(A-a)(B-b)}{\sqrt{\Sigma(A-a)2 * \Sigma(B-b)2}} = \frac{12}{\sqrt{10 * 14.8}} = 12/1655 = 0.9864$$