Correlation and Covariance Answer

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 (X and Y) change together.
If covariance is positive, it means both variables move in the same direction.
If covariance is negative, they move in opposite directions.
If it is close to zero, it means there is little or no linear relationship.

Difference from Correlation

Covariance depends on scale, so its value can be very large or very small depending on the units of X and Y.
Correlation does not depend on scale because it is a standardized value between –1 and +1, which makes it easier to interpret.

Question 2. What does a positive, negative, and zero covariance indicate?

- **Positive covariance (Cov > 0):** The two variables tend to increase or decrease together (move in the same direction).
- **Negative covariance(Cov) < 0):** The two variables tend to move in opposite directions (as one increases, the other tends to decrease).
- **Zero covariance(Cov = 0):** There is no linear relationship between the two variables.

Question3. Limitations of covariance and why correlation is preferred.

- Limitations of Covariance:
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.

Why Correlation is Preferred:
Correlation removes the effect of units by dividing covariance by standard deviations. This gives a value between –1 and +1, which makes it easy to understand the strength and direction of the relationship. So correlation is more clear and comparable

Question 4. Difference between Pearson correlation and Spearman correlation.

- Pearson correlation: Measures linear relationship and uses actual values.

- Spearman correlation: Measures rank-based relationship and checks monotonic trend.

Use Spearman when:

- Data is not normally distributed.
- Data has outliers.
- Relationship is not linear.

Question 5. Interpretation of correlation 0.85 between X and Y. Can we infer causation?

- A correlation of **0.85** shows a **strong positive relationship** between X and Y.
But **correlation does not mean causation**.
We cannot say X causes Y; they may be related due to other factors.

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

-

X	2	4	6	8
Y	3	7	5	1

Formula

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

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

Subtract the mean from each value and multiply

X	Y	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})(Y - \bar{Y})$
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	10-6.25=3.75	(3)*(3.75)=11.25

Sum the products

$\sum=(X-X)(Y-Y)=9.75-0.75-1.25+11.25=19.00$

Calculate the covariance

$$\text{Cov}(X,Y)=\frac{19.00}{4-1}=\frac{19.00}{3}\approx 6.33.$$

Covariance between X and Y = 6.33 (approximately).

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

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

Calculate the mean(A and B)

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

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

Calculate components for the numerator and denominator

Ai	Bi	Ai -A	Bi -B	(Ai -A)(Bi -B)	(Ai -A)2	B$i - B)2$
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
sum		0	0	500	1000	251.2

Calculate the Correlation Coefficient

Numerator:$\sum$(A$i - A)(Bi - B) = 500$

Denominator: $\sqrt{\sum(\text{A}i - A)2 \sum (Bi - B)2} = \sqrt{(1000) * (251.2)}$

r=$\frac{500}{\sqrt{251200}} \approx \frac{500}{501.1985} \approx 0.9976$.

The Pearson correlation coefficient between A and B is approximately 0.9976. This indicates 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(G)	50	55	58	62	70

Calculate the means(HandW)

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

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

Calculate components

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

Calculate the correlation coefficient

Numerator: $\sum$(H$i - H)(Wi - W) = 335$

Denominator: $\sqrt{\sum(Hi - H)2 \sum(Wi - W)2} = \sqrt{(500) * (228)}$

r=$\frac{335}{\sqrt{114000}} \approx \frac{335}{337.6389} \approx 0.9922$

The correlation coefficient between Height and Weight is approximately 0.9922.

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**

Reasoning:

1.Observe the trend of X: X values are increasing (1, 2, 3, 4, 5).
2.Observe the trend of Y: Y values are decreasing (15, 12, 9, 7, 3).
3.Since X and Y are moving in opposite directions (as X increases, Y decreases), the relationship is a negative correlation.

The correlation is Negative.

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	1	12	11	9
PortfolioB(B)	6	9	11	1	8

Calculate the means (A and B)

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

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

Ai	Bi	A$i - A$	B$i - B$	(A$i - A)(Bi - B)$	(A$i - A)2$	(B$i - B)2$
8	6	-2	-2.8	5.6	4	7.84
10	9	0	0.2	0	0	0.04
12	11	2	2.2	4.4	4	4.84
11	10	1	1.2	1.2	1	1.44
9	8	-1	-0.8	0.8	1	0.64
sum		0	0	12	10	14.8

Covariance:

cov(A,B)=$\frac{\sum(Ai-A)(Bi-B)}{n-1}$=$\frac{12}{5-1}$=$\frac{12}{4}$=3.0

Calculate the Correlation Coefficient

Numerator: 12

Denominator : $\sqrt{\sum(\text{A}i - A)2 \sum(Bi - B)2} = \sqrt{(10) * (14.8)}$

r=$\frac{12}{\sqrt{148}} \approx \frac{12}{12.1655} \approx 0.9864.$