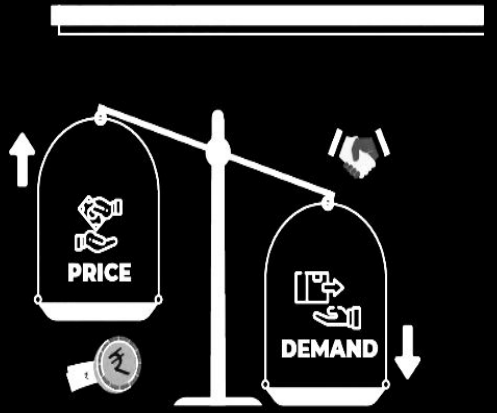
Demand (X)	Price (Y)

CORRELATION



TYPES OF DATA

UNIVARIATE DATA

One variable at a time

BIVARIATE DATA

- **When data are collected and analysed on two variables simultaneously, they are known as bivariate**

BIVARIATE DATA

- When data are collected on two variables simultaneously, they are known as bivariate data and the corresponding frequency distribution, derived from it, is known as Bivariate Frequency Distribution.

Example Prepare a Bivariate Frequency table for the following data relating to the marks in Statistics (x) and Mathematics (y):

(15, 13),	(1, 3),	(2, 6),	(8, 3),	(15, 10),	(3, 9),	(13, 19),
(10, 11),	(6, 4),	(18, 14),	(10, 19),	(12, 8),	(11, 14),	(13, 16),
(17, 15),	(18, 18),	(11, 7),	(10, 14),	(14, 16),	(16, 15),	(7, 11),
(5, 1),	(11, 15),	(9, 4),	(10, 15),	(13, 12)	(14, 17),	(10, 11),
(6, 9),	(13, 17),	(16, 15),	(6, 4),	(4, 8),	(8, 11),	(9, 12),
(14, 11),	(16, 15),	(9, 10),	(4, 6),	(5, 7),	(3, 11),	(4, 16),
(5, 8),	(6, 9),	(7, 12),	(15, 6),	(18, 11),	(18, 19),	(17, 16)
(10, 14)						

Take mutually exclusive classification for both the variables, the first class interval being 0-4 for both.

Bivariate Frequency Distribution of Marks in Statistics and Mathematics.

		MARKS IN MATHS					
Y / X		0-4	4-8	8-12	12-16	16-20	Total
MARKS IN STATS	0–4	I (1)	I (1)	II (2)			4
	4–8	I (1)	IIII (4)	~~IIII~~ (5)	I (1)	I (1)	12
	8–12	I (1)	II (2)	IIII (4)	~~IIII~~ I (6)	I (1)	14
	12–16		I (1)	III (3)	II (2)	~~IIII~~ (5)	11
	16–20			I (1)	~~IIII~~ (5)	III (3)	9
	Total	3	8	15	14	10	50

NOTE

- From the above Bivariate Frequency Distribution, we can obtain two types of univariate distributions which are known as:

MARGINAL DISTRIBUTION

CONDITIONAL DISTRIBUTION

MARGINAL DISTRIBUTION

Marginal Distribution of Marks in Statistics

Marks	No. of Students
0-4	4
4-8	12
8-12	14
12-16	11
16-20	9
Total	50

Marginal Distribution of Marks in Mathematics

CONDITIONAL DISTRIBUTION

Conditional Distribution of Marks in Mathematics for those who scored 12 -16 in stats

Conditional Distribution of Marks in Statistics for Students having Mathematics Marks between 8 to 12

Marks	No. of Students
0-4	2
4-8	5
8-12	4
12-16	3
16-20	1
Total	15

BIVARIATE DATA

Marginal Distribution

- No . of Marginal Distributions in Bivariate data = 2
- No . of cells = m x n

BIVARIATE DATA

Conditional Distribution

- No . of Conditional Distributions = m +n

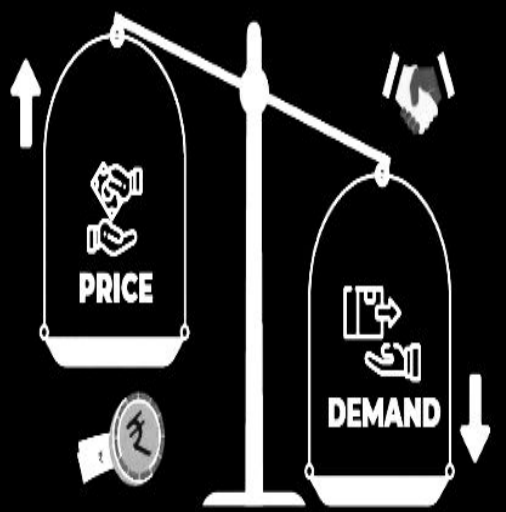
  where ,

  m = no. of class interval of x

  n = no. of class interval of y

Correlation

- In a bivariate data , if change in one variable causes change in another variable either directly or inversely, then the two variables are known to be associated or correlated.



TYPES OF CORRELATION

POSITIVE CORRELATION

- If two variables move in the same direction i.e. an increase (or decrease) on the part of one variable introduces an increase (or decrease) on the part of the other variable, then the two variables are known to be positively correlated.

- As for example, yield and rainfall, are positively correlated.

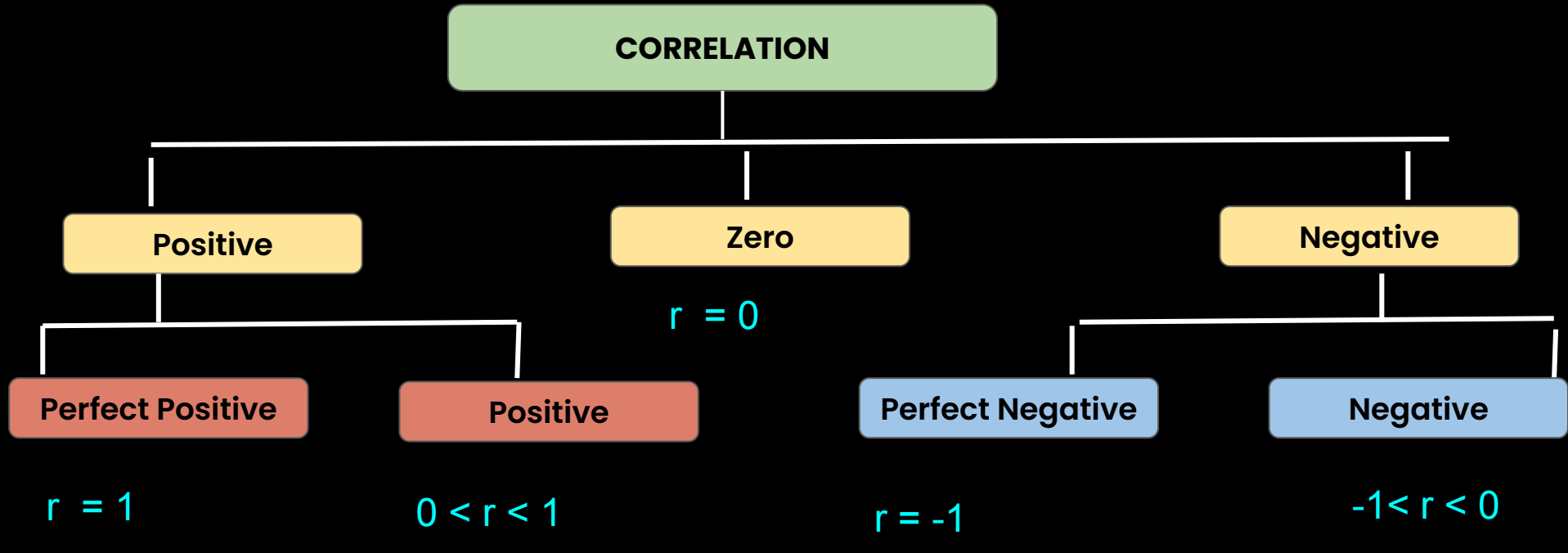
NEGATIVE CORRELATION

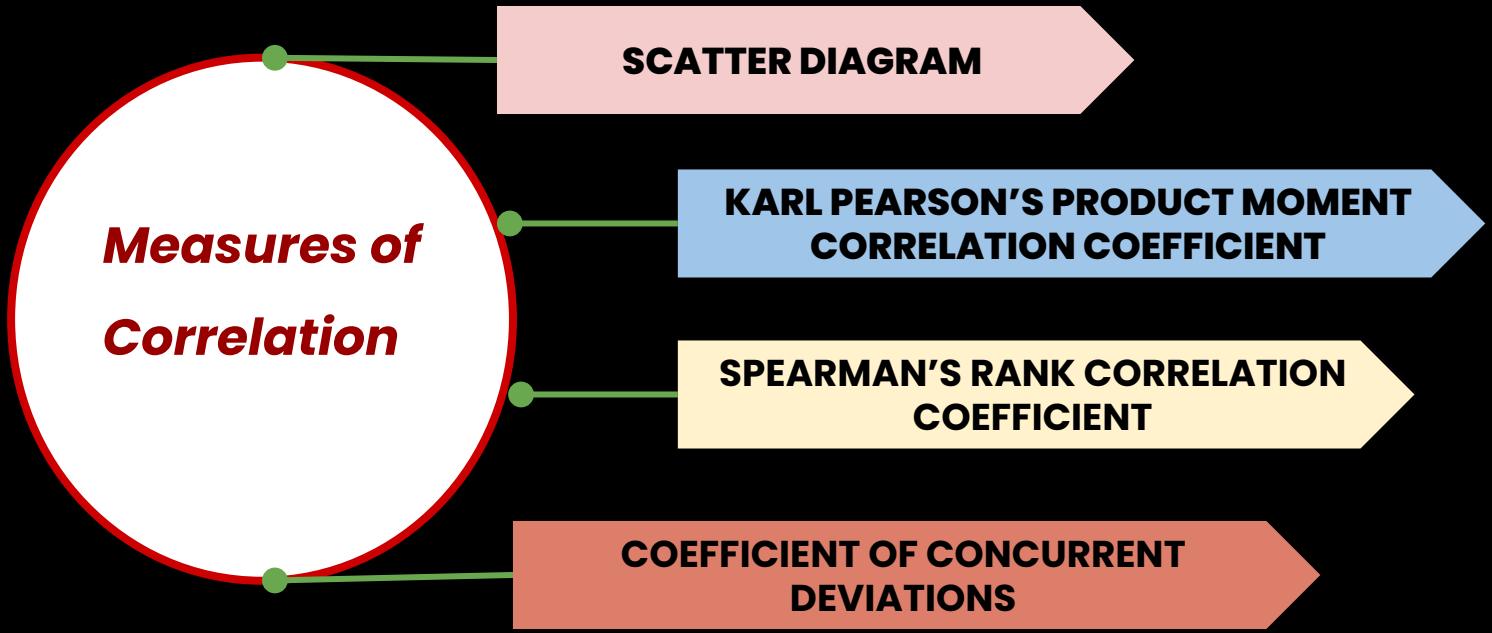
- if the two variables move in the opposite directions i.e. an increase (or a decrease) on the part of one variable results a decrease (or an increase) on the part of the other variable, then the two variables are known to have a negative correlation.

- As for example , the price and demand of an item, is negative correlation.

Correlation

- Correlation is expressed using r
- The value of correlation ranges from - 1 to 1 , both inclusive $-1 \leq r \leq 1$.




Measures of Correlation
SCATTER DIAGRAM
KARL PEARSON'S PRODUCT MOMENT CORRELATION COEFFICIENT
SPEARMAN'S RANK CORRELATION COEFFICIENT
COEFFICIENT OF CONCURRENT DEVIATIONS

SCATTER DIAGRAM

- This is a simple diagrammatic method to establish correlation between a pair of variables.
- scatter diagram can be applied for any type of correlation –linear as well as non-linear i.e. curvilinear.
- Scatter diagram can distinguish between different types of correlation although it fails to measure the extent of relationship between the variables.

SCATTER DIAGRAM

- Each data point, which in this case a pair of values (x_i, y_i) is represented by a point in the rectangular axes of coordinates.

- The totality of all the plotted points forms the scatter diagram. The pattern of the plotted points reveals the nature of correlation.

SCATTER DIAGRAM

POSITIVE CORRELATION

- The plotted points lie from lower left corner to upper right corner

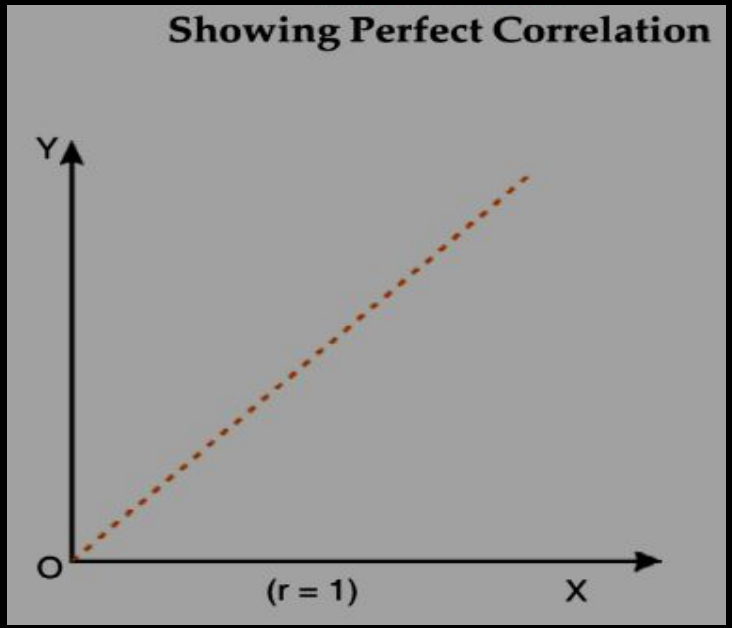
NEGATIVE CORRELATION

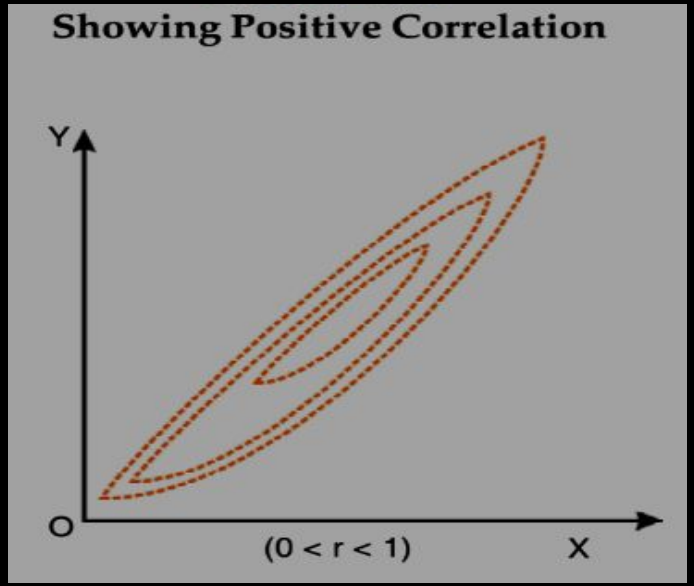
- The plotted points concentrate from upper left to lower right

ZERO CORRELATION

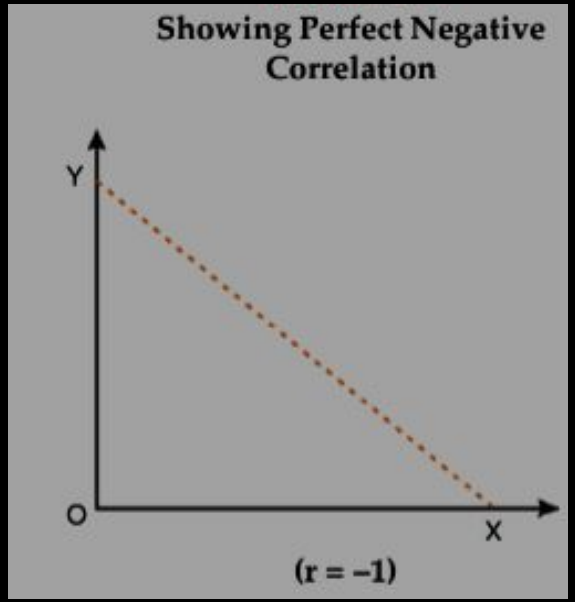
- The plotted points would be equally distributed without depicting any particular pattern.

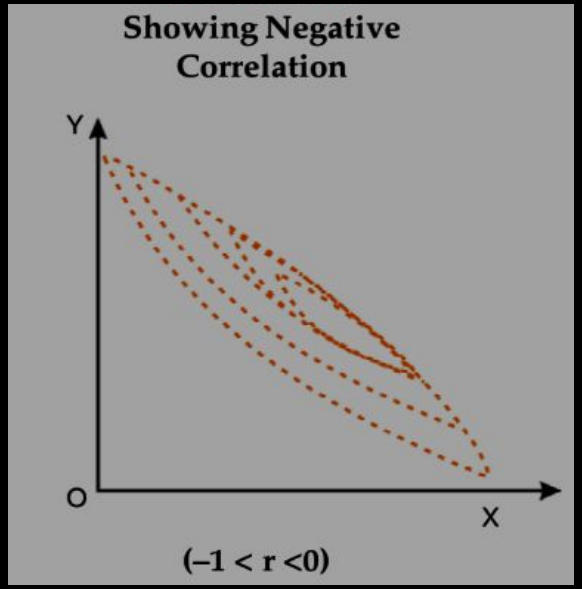
SCATTER DIAGRAM



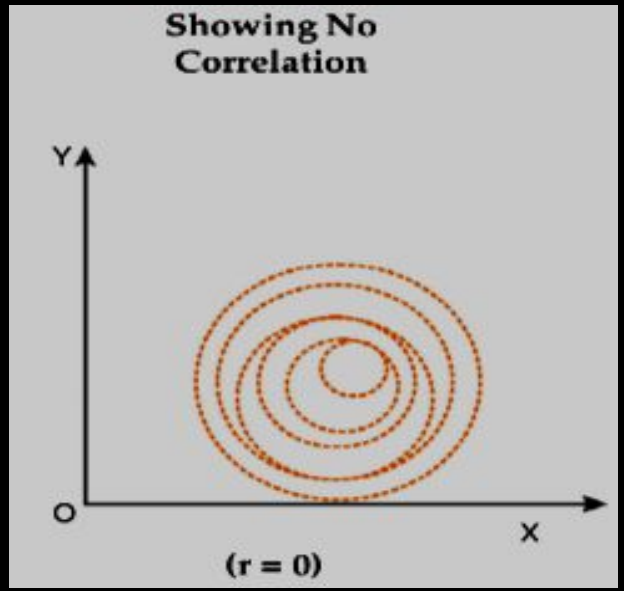


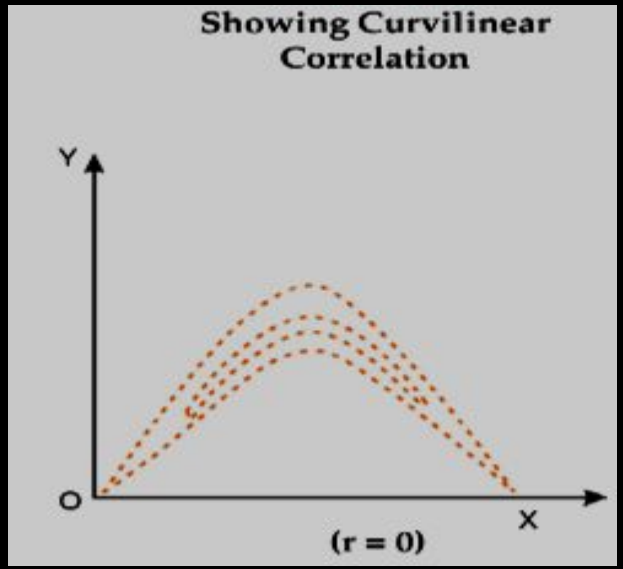
SCATTER DIAGRAM





SCATTER DIAGRAM





KARL PEARSON'S PRODUCT MOMENT CORRELATION COEFFICIENT

- This is by for the best method for finding correlation between two variables **provided the relationship between the two variables is linear**

PRODUCT MOMENT : CORRELATION COEFFICIENT

KARL PEARSON'S PRODUCT MOMENT CORRELATION COEFFICIENT

- $$r = r_{xy} = \frac{\mathrm{Cov}(x, y)}{S_x \times S_y} \ldots$$

where

$$\mathrm{cov}\,(x, y) = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n} = \frac{\Sigma x_i y_i}{n} - \bar{x}\bar{y} \ldots$$

$$S_x = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}} = \sqrt{\frac{\Sigma x_i^2}{n} - \bar{x}^2} \ldots$$

KARL PEARSON'S PRODUCT MOMENT CORRELATION COEFFICIENT

- and $S_y = \sqrt{\frac{\Sigma\left(y_i - \bar{y}\right)^2}{n}} = \sqrt{\frac{\Sigma y_i^2}{n} - \bar{y}^2}$ ..

A single formula for computing correlation coefficient is given by

$$r = \frac{n\Sigma x_i y_i - \Sigma x_i \times \Sigma y_i}{\sqrt{n\Sigma x_i^2 - \left(\Sigma x_i\right)^2}\sqrt{n\Sigma y_i^2 - (\Sigma y_i)^2}}$$

Example 17.2 Compute the correlation coefficient between x and y from the following data $n = 10$, $\Sigma xy = 220$, $\Sigma x^2 = 200$, $\Sigma y^2 = 262$

$\Sigma x = 40$ and $\Sigma y = 50$

Solution

Thus there is a good amount of positive correlation between the two variables x and y.

$$r = \frac{n\sum xy - \sum x \times \sum y}{\sqrt{n\sum x^2 - (\sum x)^2} \times \sqrt{n\sum y^2 - (\sum y)^2}}$$

$$= \frac{10 \times 220 - 40 \times 50}{\sqrt{10 \times 200 - (40)^2} \times \sqrt{10 \times 262 - (50)^2}}$$

$$= \frac{2200 - 2000}{\sqrt{2000 - 1600} \times \sqrt{2620 - 2500}}$$

$$= \frac{200}{20 \times 10.9545}$$

$$= 0.91$$

Example 17.3 Find product moment correlation coefficient from the following information:

x :	2	3	5	5	6	8
y :	9	8	8	6	5	3

Solution In order to find the covariance and the two standard deviation, we prepare the following table:

Table 17.3

Computation of Correlation Coefficient

x_i (1)	y_i (2)	$x_i y_i$ (3)= (1) x (2)	x_i^2 (4)= $(1)^2$	y_i^2 (5)= $(2)^2$
2	9	18	4	81
3	8	24	9	64
5	8	40	25	64
5	6	30	25	36
6	5	30	36	25
8	3	24	64	9
29	39	166	163	279

We have

$$\bar{x} = \frac{29}{6} = 4.8333 \quad \bar{y} = \frac{39}{6} = 6.50$$

$$\text{cov}(x, y) = \frac{\sum x_i y_i}{n} - \bar{x}\,\bar{y}$$

$$= 166/6 - 4.8333 \times 6.50 = -3.7498$$

$$= \sqrt{\frac{\Sigma x_i^2}{n} - (\bar{x})^2}$$

$$= \sqrt{\frac{163}{6} - (4.8333)^2}$$

$$= \sqrt{27.1667 - 23.3608} = 1.95$$

$$S_y = \sqrt{\frac{\sum y_i^2}{n} - (\bar{y})^2}$$

$$= \sqrt{\frac{279}{6} - (6.50)^2}$$

$$= \sqrt{46.50 - 42.25} = 2.0616$$

Thus the corre … x and y in given by

$$r = \frac{\text{cov}(x, y)}{S_x \times S_y}$$

$$= \frac{-3.7498}{1.9509 \times 2.0616}$$

$$= -0.93$$

- **We find a high degree of negative correlation between x and y**

PROPERTIES OF CORRELATION COEFFICIENT

(i) The Coefficient of Correlation is a unit-free measure.

(ii) The coefficient of correlation always lies between –1 and 1, including both the limiting values $-1 \leq r \leq 1$.

PROPERTIES OF CORRELATION COEFFICIENT

(iii) If two variables are related by a linear equation , then correlation coefficient will always be perfect +1 or -1 depends on the sign of slope of equation .

(iv) The coefficient of correlation remains invariant under a change of origin and/or scale of the variables under consideration depending on the sign of scale factors.

PROPERTIES OF CORRELATION COEFFICIENT

- Change of Origin : NO Impact
- Change of Scale : No Impact of value but affected by sign

Original		Change
x	⟶	u
y	⟶	v
r_{xy}	⟶	r_{uv}

- If sign of both change of scale are same

$$r_{uv} = r_{xy}$$

- If sign of both change of scale are different

$$r_{uv} = -r_{xy}$$

Example 17.8 Given that the correlation coefficient between x and y is 0.8, write down the correlation coefficient between u and v where

(i) $2u + 3x + 4 = 0$ and $4v + 16y + 11 = 0$

(ii) $2u - 3x + 4 = 0$ and $4v + 16y + 11 = 0$

(iii) $2u - 3x + 4 = 0$ and $4v - 16y + 11 = 0$

(iv) $2u + 3x + 4 = 0$ and $4v - 16y + 11 = 0$

SPEARMAN'S RANK CORRELATION COEFFICIENT

- When we need finding correlation between two qualitative characteristics, say, beauty and intelligence, we take recourse to using rank correlation coefficient.

- Rank correlation can also be applied to find the level of agreement (or disagreement) between two judges so far as assessing a qualitative characteristic is concerned..

SPEARMAN'S RANK CORRELATION COEFFICIENT

- Spearman's rank correlation coefficient is given by

$$r_R = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

where r_R denotes rank correlation coefficient and it lies between –1 and 1 inclusive of these two values.

$d_i = x_i - y_i$ represents the difference in ranks for the i-th individual and n denotes the number of individuals.

SPEARMAN'S RANK CORRELATION COEFFICIENT

IN CASE OF TIED VALUE

$$r_R = 1 - \frac{6\left[\sum_i d_i^2 + \sum_j \frac{\left(t_j^3 - t_j\right)}{12}\right]}{n\left(n^2 - 1\right)}$$

Example 17.10 Compute rank correlation from the following data relating to ranks given by two judges in a contest:

Serial No. of Candidate :	1	2	3	4	5	6	7	8	9	10
Rank by Judge A :	10	5	6	1	2	3	4	7	9	8
Rank by Judge B :	5	6	9	2	8	7	3	4	10	1

Example 17.9 compute the coefficient of rank correlation between sales and advertisement expressed in thousands of rupees from the following data:

Sales :	90	85	68	75	82	80	95	70
Advertisement :	7	6	2	3	4	5	8	1

Solution Let the rank given to sales be denoted by x and rank of advertisement be denoted by y. We note that since the highest sales as given in the data, is 95, it is to be given rank 1, the second highest sales 90 is to be given rank 2 and finally rank 8 goes to the lowest sales, namely 68. We have given rank to the other variable advertisement in a similar manner. Since there are no ties, we apply formula .

Table 17.7

Computation of Rank correlation between Sales and Advertisement.

Sales (x_i)	Advertisement (y_i)	Rank for Sales (x_i)	Rank for Advertisement (y_i)	$d_i = x_i - y_i$	d_i^2
90	7	2	2	0	0
85	6	3	3	0	0
68	2	8	7	1	1
75	3	6	6	0	0
82	4	4	5	–1	1
80	5	5	4	1	1
95	8	1	1	0	0
70	1	7	8	–1	1
Total	—	—	—	0	4

Since n = 8 and $\sum d_i^2 = 4$, applying formula (17.11), we get.

$$r_R = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 4}{8(8^2-1)}$$

$$= 1 - 0.0476$$

$$= 0.95$$

The high positive value of the rank correlation coefficient indicates that there is a very good amount of agreement between sales and advertisement.

Example 17 .11 Compute the coefficient of rank correlation between Eco. marks and stats. Marks as given below:

Eco Marks :	80	56	50	48	50	62	60
Stats Marks :	90	75	75	65	65	50	65

Example For a group of 8 students, the sum of squares of differences in ranks for Mathematics and Statistics marks was found to be 50 what is the value of rank correlation coefficient?

Example While computing rank correlation coefficient between profits and investment for 10 years of a firm, the difference in rank for a year was taken as 7 instead of 5 by mistake and the value of rank correlation coefficient was computed as 0.80. What would be the correct value of rank correlation coefficient after rectifying the mistake?

COEFFICIENT OF CONCURRENT DEVIATIONS

- A very simple and casual method of finding correlation when we are not serious about the magnitude of the two variables is the application of concurrent deviations .

COEFFICIENT OF CONCURRENT DEVIATIONS

- Denoting the number of concurrent deviation by c and total number of deviations as m (which must be one less than the number of pairs of x and y values), the coefficient of concurrent deviation is given by

$$r_C = \pm\sqrt{\pm\frac{(2c-m)}{m}}$$

COEFFICIENT OF CONCURRENT DEVIATIONS

- If $(2c-m) > 0$, then we take the positive sign both inside and outside the radical sign and if $(2c-m) < 0$, we are to consider the negative sign both inside and outside the radical sign.

- Like Pearson's correlation coefficient and Spearman's rank correlation coefficient, the coefficient of concurrent deviations also lies between -1 and 1, both inclusive.

Example 17.15 Find the coefficient of concurrent deviations from the following data.

Year :	1990	1991	1992	1993	1994	1995	1996	1997
Price :	25	28	30	23	35	38	39	42
Demand :	35	34	35	30	29	28	26	23

Set A

Write the correct answers. Each question carries 1 mark.

Que. 1 Bivariate Data are the data collected for

(a) Two variables irrespective of time

(b) More than two variables

(c) Two variables at the same point of time

(d) Two variables at different points of time.

Write the correct answers. Each question carries 1 mark.

Que. 2 For a bivariate frequency table having (p + q) classification the total number of cells is

(a) p

(b) p + q

(c) q

(d) pq

Write the correct answers. Each question carries 1 mark.

Que. 3 Some of the cell frequencies in a bivariate frequency table may be

(a) Negative

(b) Zero

(c) a or b

(d) None of these

Write the correct answers. Each question carries 1 mark.

Que. 4 For a p × q bivariate frequency table, the maximum number of marginal distributions is

(a) p

(b) p + q

(c) 1

(d) 2

Write the correct answers. Each question carries 1 mark.

Que. 5 For a p × q classification of bivariate data, the maximum number of conditional distributions is

(a) p

(b) p + q

(c) pq

(d) p or q

Write the correct answers. Each question carries 1 mark.

Que. 6 Correlation analysis aims at

(a) Predicting one variable for a given value of the other variable

(b) Establishing relation between two variables

(c) Measuring the extent of relation between two variables

(d) Both (b) and (c).

Write the correct answers. Each question carries 1 mark.

Que. 7 Regression analysis is concerned with

(a) Establishing a mathematical relationship between two variables

(b) Measuring the extent of association between two variables

(c) Predicting the value of the dependent variable for a given value of the independent variable

(d) Both (a) and (c).

Write the correct answers. Each question carries 1 mark.

Que. 8 What is spurious correlation?

(a) It is a bad relation between two variables.

(b) It is very low correlation between two variables.

(c) It is the correlation between two variables having no causal relation.

(d) It is a negative correlation.

SPURIOUS CORRELATION

- There are some cases when we may find a correlation between two variables although the two variables are not causally related. This is due to the existence of a third variable which is related to both the variables under consideration. Such a correlation is known as **spurious correlation or nonsense correlation**.
- As an example, there could be a positive correlation between production of rice and that of iron in India for the last twenty years due to the effect of a third variable time on both these variables. It is necessary to eliminate the influence of the third variable before computing correlation between the two original variables.

Write the correct answers. Each question carries 1 mark.

Que. 9 Scatter diagram is considered for measuring

(a) Linear relationship between two variables

(b) Curvilinear relationship between two variables

(c) Neither (a) nor (b)

(d) Both (a) and (b).

Write the correct answers. Each question carries 1 mark.

Que. 10 If the plotted points in a scatter diagram lie from upper left to lower right, then the correlation is

(a) Positive

(b) Zero

(c) Negative

(d) None of these.

Write the correct answers. Each question carries 1 mark.

Que. 11 If the plotted points in a scatter diagram are evenly distributed, then the correlation is

(a) Zero

(b) Negative

(c) Positive

(d) (a) or (b).

Write the correct answers. Each question carries 1 mark.

Que. 12 If all the plotted points in a scatter diagram lie on a single line, then the correlation is

(a) Perfect positive

(b) Perfect negative

(c) Both (a) and (b)

(d) Either (a) or (b).

Write the correct answers. Each question carries 1 mark.

Que. 13 The correlation between shoe-size and intelligence is

(a) Zero

(b) Positive

(c) Negative

(d) None of these.

Write the correct answers. Each question carries 1 mark.

Que. 14 The correlation between the speed of an automobile and the distance travelled by it after applying the brakes is

(a) Negative

(b) Zero

(c) Positive

(d) None of these.

Write the correct answers. Each question carries 1 mark.

Que. 15 Scatter diagram helps us to

(a) Find the nature of correlation between two variables

(b) Compute the extent of correlation between two variables

(c) Obtain the mathematical relationship between two variables

(d) Both (a) and (c).

Write the correct answers. Each question carries 1 mark.

Que. 16 Pearson's correlation coefficient is used for finding

(a) Correlation for any type of relation

(b) Correlation for linear relation only

(c) Correlation for curvilinear relation only

(d) Both (b) and (c).

Write the correct answers. Each question carries 1 mark.

Que. 17 Product moment correlation coefficient is considered for

(a) Finding the nature of correlation

(b) Finding the amount of correlation

(c) Both (a) and (b)

(d) Either (a) and (b).

Write the correct answers. Each question carries 1 mark.

Que. 18 If the value of correlation coefficient is positive, then the points in a scatter diagram tend to cluster

(a) From lower left corner to upper right corner

(b) From lower left corner to lower right corner

(c) From lower right corner to upper left corner

(d) From lower right corner to upper right corner.

Write the correct answers. Each question carries 1 mark.

Que. 19 When r = 1, all the points in a scatter diagram would lie

(a) On a straight line directed from lower left to upper right

(b) On a straight line directed from upper left to lower right

(c) On a straight line

(d) Both (a) and (b).

Write the correct answers. Each question carries 1 mark.

Que. 20 **Product moment correlation coefficient may be defined as the ratio of**

(a) The product of standard deviations of the two variables to the covariance between them

(b) The covariance between the variables to the product of the variances of them

(c) The covariance between the variables to the product of their standard deviations

(d) Either (b) or (c).

Write the correct answers. Each question carries 1 mark.

Que. 21 The covariance between two variables is

(a) Strictly positive

(b) Strictly negative

(c) Always 0

(d) Either positive or negative or zero.

Write the correct answers. Each question carries 1 mark.

Que. 22 The coefficient of correlation between two variables

(a) Can have any unit.

(b) Is expressed as the product of units of the two variables

(c) Is a unit free measure

(d) None of these.

Write the correct answers. Each question carries 1 mark.

Que. 23 What are the limits of the correlation coefficient?

(a) No limit

(b) –1 and 1, excluding the limits

(c) 0 and 1, including the limits

(d) –1 and 1, including the limits

Write the correct answers. Each question carries 1 mark.

Que. 24 In case the correlation coefficient between two variables is 1, the relationship between the two variables would be

(a) $y = a + bx$

(b) $y = a + bx, b > 0$

(c) $y = a + bx, b < 0$

(d) $y = a + bx$, both a and b being positive.

Write the correct answers. Each question carries 1 mark.

Que. 25 If the relationship between two variables x and y is given by $2x + 3y + 4 = 0$, then the value of the correlation coefficient between x and y is

(a) 0

(b) 1

(c) –1

(d) negative.

Write the correct answers. Each question carries 1 mark.

Que. 26 For finding correlation between two attributes, we consider

(a) Pearson's correlation coefficient

(b) Scatter diagram

(c) Spearman's rank correlation coefficient

(d) Coefficient of concurrent deviations.

Write the correct answers. Each question carries 1 mark.

Que. 27 For finding the degree of agreement about beauty between two Judges in a Beauty Contest, we use

(a) Scatter diagram

(b) Coefficient of rank correlation

(c) Coefficient of correlation

(d) Coefficient of concurrent deviation.

Write the correct answers. Each question carries 1 mark.

Que. 28 If there is a perfect disagreement between the marks in Geography and Statistics, then what would be the value of rank correlation coefficient?

(a) Any value

(b) Only 1

(c) Only –1

(d) (b) or (c)

Write the correct answers. Each question carries 1 mark.

Que. 29 When we are not concerned with the magnitude of the two variables under discussion, we consider

(a) Rank correlation coefficient

(b) Product moment correlation coefficient

(c) Coefficient of concurrent deviation

(d) (a) or (b) but not (c).

Write the correct answers. Each question carries 1 mark.

Que. 30 What is the quickest method to find correlation between two variables?

(a) Scatter diagram

(b) Method of concurrent deviation

(c) Method of rank correlation

(d) Method of product moment correlation

Write the correct answers. Each question carries 1 mark.

Que. 33 Since Blood Pressure of a person depends on age, we need to consider

(a) The regression equation of Blood Pressure on age

(b) The regression equation of age on Blood Pressure

(c) Both (a) and (b)

(d) Either (a) or (b).

Write the correct answers. Each question carries 1 mark.

Que. 34 The method applied for deriving the regression equations is known as

(a) Least squares

(b) Concurrent deviation

(c) Product moment

(d) Normal equation.

Write the correct answers. Each question carries 1 mark.

Que. 35 The difference between the observed value and the estimated value in regression analysis is known as

(a) Error

(b) Residue

(c) Deviation

(d) (a) or (b).

Write the correct answers. Each question carries 1 mark.

Que. 36 The errors in case of regression equations are

(a) Positive

(b) Negative

(c) Zero

(d) All these.

Write the correct answers. Each question carries 1 mark.

Que. 37 The regression line of y on x is derived by

(a) The minimisation of vertical distances in the scatter diagram

(b) The minimisation of horizontal distances in the scatter diagram

(c) Both (a) and (b)

(d) (a) or (b).

Write the correct answers. Each question carries 1 mark.

Que. 38 The two lines of regression become identical when

(a) $r = 1$

(b) $r = -1$

(c) $r = 0$

(d) (a) or (b).

Write the correct answers. Each question carries 1 mark.

Que. 39 What are the limits of the two regression coefficients?

(a) No limit

(b) Must be positive

(c) One positive and the other negative

(d) Product of the regression coefficient must be numerically less than unity.

Write the correct answers. Each question carries 1 mark.

Que. 40 The regression coefficients remain unchanged due to a

(a) Shift of origin

(b) Shift of scale

(c) Both (a) and (b)

(d) (a) or (b).

Write the correct answers. Each question carries 1 mark.

Que. 41 If the coefficient of correlation between two variables is –0.9, then the coefficient of determination is

(a) 0.9

(b) 0.81

(c) 0.1

(d) 0.19.

Write the correct answers. Each question carries 1 mark.

Que. 42 If the coefficient of correlation between two variables is 0.7 then the percentage of variation unaccounted for is

(a) 70%

(b) 30%

(c) 51%

(d) 49%

Set B

Que. 1 If for two variable x and y, the covariance, variance of x and variance of y are 40, 16 and 256 respectively, what is the value of the correlation coefficient?

(a) 0.01

(b) 0.625

(c) 0.4

(d) 0.5

Que. 2 If cov(x, y) = 15, what restrictions should be put for the standard deviations of x and y?

(a) No restriction.

(b) The product of the standard deviations should be more than 15.

(c) The product of the standard deviations should be less than 15.

(d) The sum of the standard deviations should be less than 15.

Que. 3 If the covariance between two variables is 20 and the variance of one of the variables is 16, what would be the variance of the other variable?

(a) $S^2_y \geq 25$

(b) More than 10

(c) Less than 10

(d) More than 1.25

Que. 4 If $y = a + bx$, then what is the coefficient of correlation between x and y?

(a) 1

(b) –1

(c) 1 or –1 according as $b > 0$ or $b < 0$

(d) none of these.

Que. 5 If r = 0.6 then the coefficient of non-determination is

(a) 0.4

(b) –0.6

(c) 0.36

(d) 0.64

Que. 6 If u + 5x = 6 and 3y – 7v = 20 and the correlation coefficient between x and y is 0.58 then what would be the correlation coefficient between u and v?

(a) 0.58

(b) –0.58

(c) –0.84

(d) 0.84

Que. 7 If the relation between x and u is $3x + 4u + 7 = 0$ and the correlation coefficient between x and y is -0.6, then what is the correlation coefficient between u and y?

(a) -0.6

(b) 0.8

(c) 0.6

(d) -0.8

Que. 8 From the following data

x:	2	3	5	4	7
y:	4	6	7	8	10

The coefficient of correlation was found to be 0.93. What is the correlation between u and v as given below?

u:	–3	–2	0	–1	2
v:	–4	–2	–1	0	2

(a) –0.93

(b) 0.93

(c) 0.57

(d) –0.57

Que. 9 Referring to the data presented in Q. No. 8, what would be the correlation between u and v?

u:	10	15	25	20	35
v:	–24	–36	–42	–48	–60

(a) –0.6

(b) 0.6

(c) –0.93

(d) 0.93

Que. 10 If the sum of squares of difference of ranks, given by two judges A and B, of 8 students is 21, what is the value of rank correlation coefficient?

(a) 0.7

(b) 0.65

(c) 0.75

(d) 0.8

Que. 11 If the rank correlation coefficient between marks in management and mathematics for a group of student is 0.6 and the sum of squares of the differences in ranks is 66, what is the number of students in the group?

(a) 10

(b) 9

(c) 8

(d) 11

Que. 12 While computing rank correlation coefficient between profit and investment for the last 6 years of a company the difference in rank for a year was taken 3 instead of 4. What is the rectified rank correlation coefficient if it is known that the original value of rank correlation coefficient was 0.4?

(a) 0.3

(b) 0.2

(c) 0.25

(d) 0.28

Que. 13 For 10 pairs of observations, No. of concurrent deviations was found to be 4. What is the value of the coefficient of concurrent deviation?

(a) $\surd 0.2$

(b) $\surd - 0.2$

(c) 1/3

(d) –1/3

Que. 14 The coefficient of concurrent deviation for p pairs of observations was found to be $1/\sqrt{3}$.

If the number of concurrent deviations was found to be 6, then the value of p is.

(a) 10

(b) 9

(c) 8

(d) none of these

Que. 15 What is the value of correlation coefficient due to Pearson on the basis of the following data:

x:	–5	–4	–3	–2	–1	0	1	2	3	4	5
y:	27	18	11	6	3	2	3	6	11	18	27

(a) 1

(b) –1

(c) 0

(d) –0.5

Que. 16 Following are the two normal equations obtained for deriving the regression line of y and x:

$5a + 10b = 40$

$10a + 25b = 95$

The regression line of y on x is given by

(a) $2x + 3y = 5$

(b) $2y + 3x = 5$

(c) $y = 2 + 3x$

(d) $y = 3 + 5x$

Que. 17 If the regression line of y on x and of x on y are given by $2x + 3y = -1$ and $5x + 6y = -1$ then the arithmetic means of x and y are given by

(a) (1, –1)

(b) (–1, 1)

(c) (–1, –1)

d) (2, 3)

Que. 18 Given the regression equations as $3x + y = 13$ and $2x + 5y = 20$, which one is the regression equation of y on x?

(a) 1st equation

(b) 2nd equation

(c) both (a) and (b)

(d) none of these.

Que. 19 Given the following equations: $2x - 3y = 10$ and $3x + 4y = 15$, which one is the regression equation of x on y ?

(a) 1st equation

(b) 2nd equation

(c) both the equations

(d) none of these

Que. 20 If $u = 2x + 5$ and $v = -3y - 6$ and regression coefficient of y on x is 2.4, what is the regression coefficient of v on u?

(a) 3.6

(b) −3.6

(c) 2.4

(d) −2.4

Que. 21 If $4y - 5x = 15$ is the regression line of y on x and the coefficient of correlation between x and y is 0.75, what is the value of the regression coefficient of x on y?

(a) 0.45

(b) 0.9375

(c) 0.6

(d) none of these

Que. 22 If the regression line of y on x and that of x on y are given by $y = -2x + 3$ and $8x = -y + 3$ respectively, what is the coefficient of correlation between x and y?

(a) 0.5

(b) $-1/\sqrt{2}$

(c) −0.5

(d) none of these

Que. 23 If the regression coefficient of y on x, the coefficient of correlation between x and y and variance of y are $-3/4$, $\sqrt{3}/2$ and 4 respectively, what is the variance of x?

(a) $2/\sqrt{3}/2$

(b) $16/3$

(c) $4/3$

(d) 4

Que. 24 If $y = 3x + 4$ is the regression line of y on x and the arithmetic mean of x is -1, what is the arithmetic mean of y?

(a) 1

(b) -1

(c) 7

(d) none of these

REGRESSION

REGRESSION ANALYSIS

- In regression analysis, we are concerned with the estimation of one variable for a given value of another variable on the basis of an average mathematical relationship between the two variables.

Estimation of Y when X is given → **Y on X**

Y : Dependent

X : Independent

$y = a + bx$

Estimation of X when Y is given → **X on Y**

X : Dependent

Y : Independent

$x = a + by$

REGRESSION

Estimation of Y when X is given

Regression line of Y on X

$$Y - \overline{Y} = b_{yx}(X - \overline{X})$$

Estimation of X when Y is given

Regression line of X on Y

$$X - \overline{X} = b_{xy}(Y - \overline{Y})$$

METHOD OF LEAST SQUARES

REGRESSION COEFFICIENT

Regression Coefficient of Y on X

$$b_{yx} = \frac{Cov(x,y)}{Var\ of\ x}$$

$$b_{yx} = r \cdot \frac{SD_y}{SD_x}$$

REGRESSION COEFFICIENT

Regression Coefficient of X on Y

$$b_{xy} = \frac{Cov(x,y)}{Var\ of\ y}$$

$$b_{xy} = r \cdot \frac{SD_x}{SD_y}$$

Example 17.17 The following data relate to the mean and SD of the prices of two shares in a stock Exchange:

Share	Mean (in ₹)	SD (in ₹)
Company A	44	5.60
Company B	58	6.30

Coefficient of correlation between the share prices = 0.48

Find the most likely price of share A corresponding to a price of ` 60 of share B and also the most likely price of share B for a price of ` 50 of share A.

Example 17.15 Find the two regression equations from the following data:

x:	2	4	5	5	8	10
y:	6	7	9	10	12	12

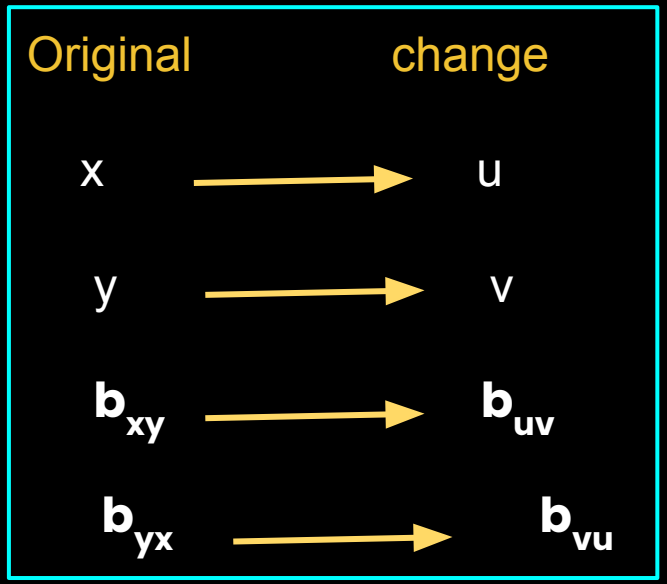
Hence estimate y when x is 13 and estimate also x when y is 15.

PROPERTIES REGRESSION LINES / COEFFICIENTS

(i) The regression coefficients remain unchanged due to a shift of origin but change due to a shift of scale.

- $b_{uv} = b_{xy} \cdot \dfrac{\text{change of scale of x}}{\text{Change of scale of y}}$

- $b_{vu} = b_{yx} \cdot \dfrac{\text{change of scale of y}}{\text{Change of scale of x}}$



Example 17.19 If the relationship between two variables x and u is $u + 3x = 10$ and between two other variables y and v is $2y + 5v = 25$, and the regression coefficient of y on x is known as 0.80, what would be the regression coefficient of v on u?

PROPERTIES REGRESSION LINES / COEFFICIENTS

(ii) The two lines of regression intersect at the point $(\bar{x}, \bar{y})$, where x and y are the variables under consideration.

According to this property, the point of intersection of the regression line of y on x and the regression line of x on y is $(\bar{x}, \bar{y})$ i.e. the solution of the simultaneous equations in x and y.

PROPERTIES REGRESSION LINES / COEFFICIENTS

(iii) The coefficient of correlation between two variables x and y is the simple geometric mean of the two regression coefficients. The sign of the correlation coefficient would be the common sign of the two regression coefficients.

This property says that if the two regression coefficients are denoted by b_{yx} and b_{xy} then the coefficient of correlation is given by

$$r = \pm \sqrt{b_{yx} \times b_{xy}}$$

If both the regression coefficients are negative, r would be negative and if both are positive, r would assume a positive value.

Example 17.20 For the variables x and y, the regression equations are given as $7x - 3y - 18 = 0$ and $4x - y - 11 = 0$

(i) Find the arithmetic means of x and y.

(ii) Identify the regression equation of y on x.

(iii) Compute the correlation coefficient between x and y.

(iv) Given the variance of x is 9, find the SD of y.

Coefficient of Determination / Explained Variance / Accounted Variance

- Correlation coefficient measuring a linear relationship between the two variables indicates the amount of variation of one variable accounted for by the other variable.
- A better measure for this purpose is provided by the square of the correlation coefficient, Known as ‘coefficient of determination’.
- This can be interpreted as the ratio between the explained variance to total variance i.e.

$$r^2 = \frac{\text{Explained variance}}{\text{Total variance}}$$

Coefficient of Non- Determination / Unexplained Variance / Unaccounted Variance

- The 'coefficient of non-determination' is given by $(1-r^2)$ and can be interpreted as the ratio of unexplained variance to the total variance.

$$\text{Coefficient of non-determination} = (1-r^2)$$

NOTE

- Regression analysis is concerned with establishing a functional relationship between two variables and using this relationship for making future projection.
- This can be applied, unlike correlation for any type of relationship linear as well as curvilinear.
- The two lines of regression coincide i.e. become identical when r = –1 or 1 or in other words, there is a perfect negative or positive correlation between the two variables under discussion.
- If r = 0 Regression lines are perpendicular to each other

Set C

Write down the correct answers. Each question carries 5 marks.

Que. 1 What is the coefficient of correlation from the following data?

x:	1	2	3	4	5
y:	8	6	7	5	5

(a) 0.75

(b) –0.75

(c) –0.85

(d) 0.82

Que. 2 The coefficient of correlation between x and y where

x:	64	60	67	59	69
y:	57	60	73	62	68

is

(a) 0.655

(b) 0.68

(c) 0.73

(d) 0.758

Que. 3 What is the coefficient of correlation between the ages of husbands and wives from the following data?

Age of husband (year):	46	45	42	40	38	35	32	30	27	25
Age of wife (year):	37	35	31	28	30	25	23	19	19	18

(a) 0.58

(b) 0.98

(c) 0.89

(d) 0.92

Que. 4 The following results relate to bivariate data on (x, y):

$\Sigma xy = 414$, $\Sigma x = 120$, $\Sigma y = 90$, $\Sigma x^2 = 600$, $\Sigma y^2 = 300$, $n = 30$. Later or, it was known that two pairs of observations (12, 11) and (6, 8) were wrongly taken, the correct pairs of observations being (10, 9) and (8, 10). The corrected value of the correlation coefficient is

(a) 0.752

(b) 0.768

(c) 0.846

(d) 0.953

Que. 5 The following table provides the distribution of items according to size groups and also the number of defectives:

Size group:	9-11	11-13	13-15	15-17	17-19
No. of items:	250	350	400	300	150
No. of defective items:	25	70	60	45	20

The correlation coefficient between size and defectives is

(a) 0.25

(b) 0.12

(c) 0.14

(d) 0.07

Que. 6 For two variables x and y, it is known that cov (x, y) = 8, r = 0.4, variance of x is 16 and sum of squares of deviation of y from its mean is 250. The number of observations for this bivariate data is

(a) 7

(b) 8

(c) 9

(d) 10

Que. 7 Eight contestants in a musical contest were ranked by two judges A and B in the following manner:

Serial Number of the contestants:	1	2	3	4	5	6	7	8
Rank by Judge A:	7	6	2	4	5	3	1	8
Rank by Judge B:	5	4	6	3	8	2	1	7

The rank correlation coefficient is

(a) 0.65

(b) 0.63

(c) 0.60

(d) 0.57

Que. 8 Following are the marks of 10 students in Botany and Zoology:

Serial No.:	1	2	3	4	5	6	7	8	9	10
Marks in Botany:	58	43	50	19	28	24	77	34	29	75
Marks in Zoology:	62	63	79	56	65	54	70	59	55	69

The coefficient of rank correlation between marks in Botany and Zoology is

(a) 0.65

(b) 0.70

(c) 0.72

(d) 0.75

Que. 9 What is the value of Rank correlation coefficient between the following marks in Physics and Chemistry:

Roll No.:	1	2	3	4	5	6
Marks in Physics:	25	30	46	30	55	80
Marks in Chemistry:	30	25	50	40	50	78

(a) 0.782

(b) 0.696

(c) 0.932

(d) 0.857

Que. 10 What is the coefficient of concurrent deviations for the following data:

Supply:	68	43	38	78	66	83	38	23	83	63	53
Demand:	65	60	55	61	35	75	45	40	85	80	85

(a) 0.82

(b) 0.85

(c) 0.89

(d) –0.81

Que. 11 What is the coefficient of concurrent deviations for the following data:

Year:	1996	1997	1998	1999	2000	2001	2002	2003
Price:	35	38	40	33	45	48	49	52
Demand:	36	35	31	36	30	29	27	24

(a) –1

(b) 0.43

(c) 0.5

(d) √2

Que. 12 The regression equation of y on x for the following data:

x	41	82	62	37	58	96	127	74	123	100
y	28	56	35	17	42	85	105	61	98	73

Is given by

(a) $y = 1.2x - 15$

(b) $y = 1.2x + 15$

(c) $y = 0.93x - 14.68$

(d) $y = 1.5x - 10.89$

Que. 13 The following data relate to the heights of 10 pairs of fathers and sons:

(175, 173), (172, 172), (167, 171), (168, 171), (172, 173), (171, 170), (174, 173), (176, 175) (169, 170), (170, 173)

The regression equation of height of son on that of father is given by

(a) $y = 100 + 5x$

(b) $y = 102.60 + 0.405x$

(c) $y = 89.653 + 0.582x$

(d) $y = 88.758 + 0.562x$

Que. 14 The two regression coefficients for the following data:

x:	38	23	43	33	28
y:	28	23	43	38	8

are

(a) 1.2 and 0.4

(b) 1.6 and 0.8

(c) 1.7 and 0.8

(d) 1.8 and 0.3

Que. 15 For y = 25, what is the estimated value of x, from the following data:

X:	11	12	15	16	18	19	21
Y:	21	15	13	12	11	10	9

(a) 15

(b) 13.926

(c) 6.07

(d) 14.986

Que. 16 Given the following data:

Variable:	x	y
Mean:	80	98
Variance:	4	9

Coefficient of correlation = 0.6

What is the most likely value of y when x = 90 ?

(a) 90

(b) 103

(c) 104

(d) 107

Que. 17 The two lines of regression are given by

$8x + 10y = 25$ and $16x + 5y = 12$ respectively.

If the variance of x is 25, what is the standard deviation of y?

(a) 16

(b) 8

(c) 64

(d) 4

Que. 18 Given below the information about the capital employed and profit earned by a company over the last twenty five years:

	Mean	SD
Capital employed (0000 ₹)	62	5
Profit earned (000 ₹)	25	6

Correlation coefficient between capital employed and profit = 0.92. The sum of the Regression coefficients for the above data would be:

(a) 1.871

(b) 2.358

(c) 1.968

(d) 2.346

Que. 19 The coefficient of correlation between cost of advertisement and sales of a product on the basis of the following data:

Ad cost (000 ₹):	75	81	85	105	93	113	121	125
Sales (000 000 ₹):	35	45	59	75	43	79	87	95

is

(a) 0.85

(b) 0.89

(c) 0.95

(d) 0.98