5 MARKS

CHAPTER 17

CORRELATION AND REGRESSION

BY: SHIVANI SHARMA

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

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)						

- 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.

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

- **No . of cells = m x n**

  where ,

  m = no. of class interval of x

  n = no. of class interval of y

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

- No . of Marginal Distributions in Bivariate data = 2

CONDITIONAL DISTRIBUTION

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

- **No . of Conditional Distributions = m +n**

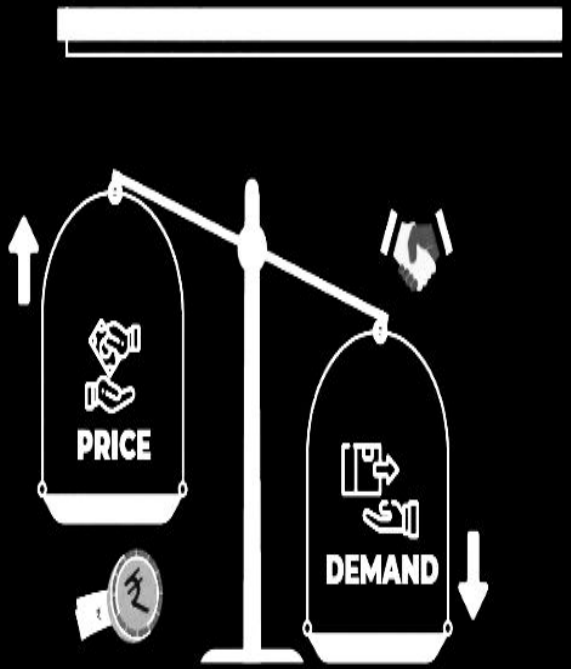
  where ,

  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.

- **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.

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**.

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).

Ans : d

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.

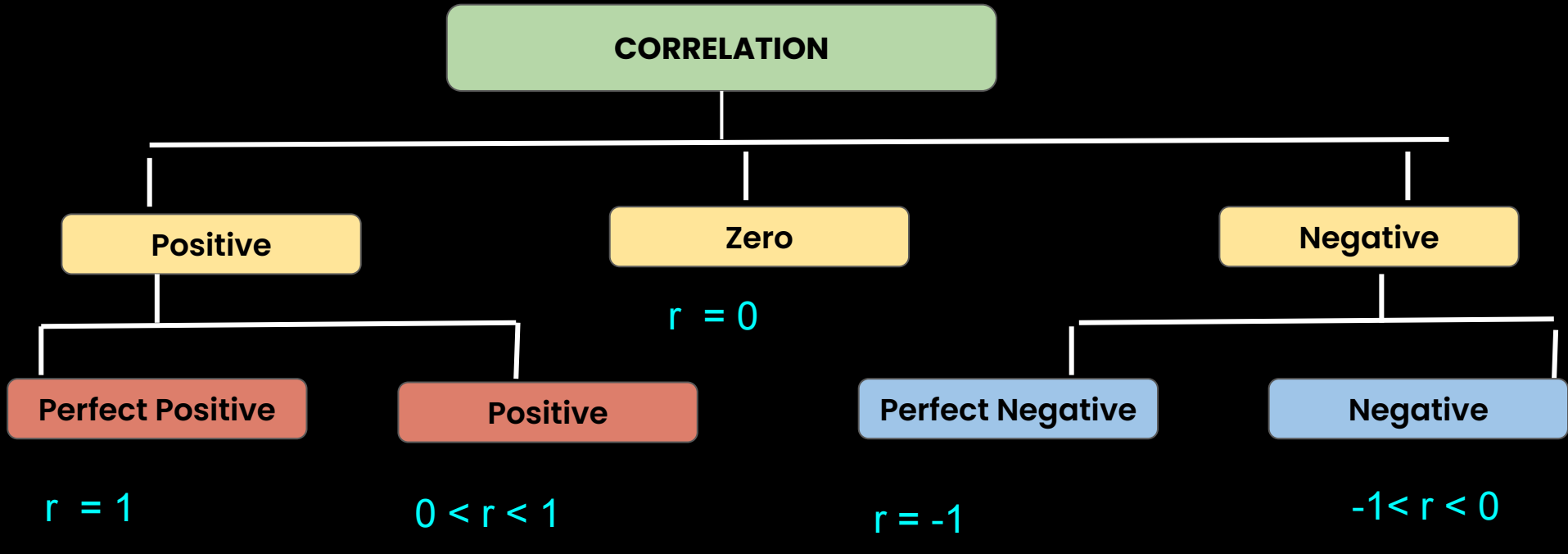
(c) It is the correlation between two variables having no causal relation.

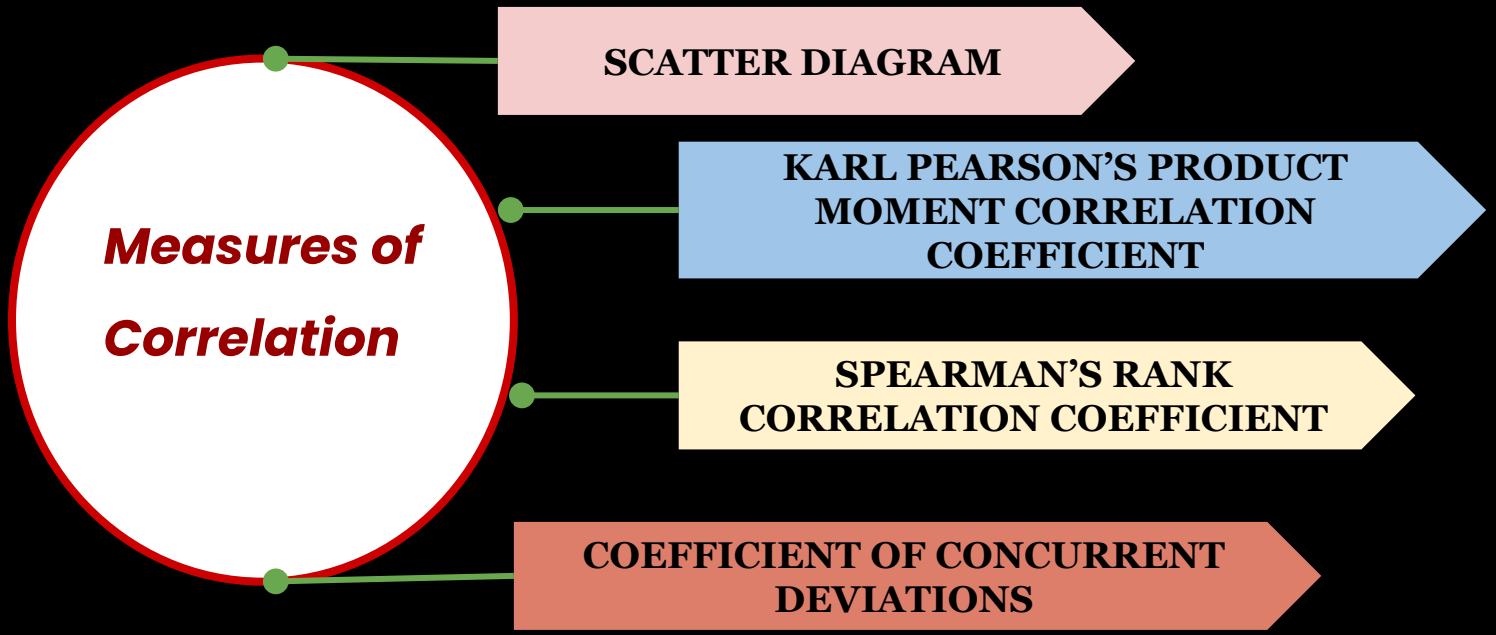
(d) It is a negative correlation.

Ans : c

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

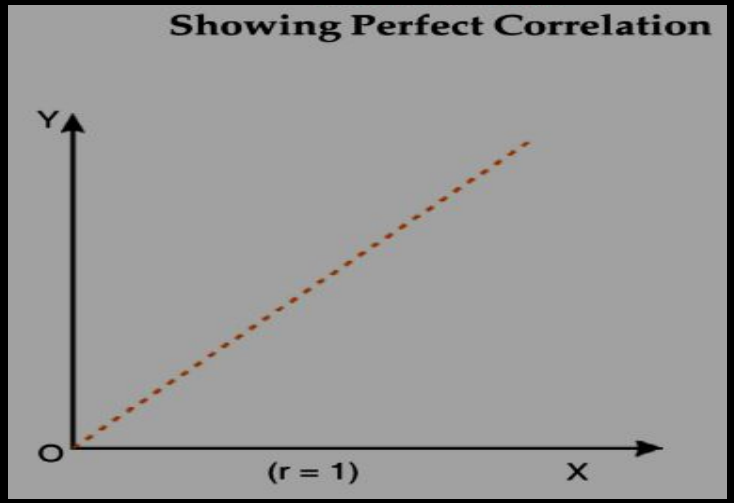
SCATTER DIAGRAM

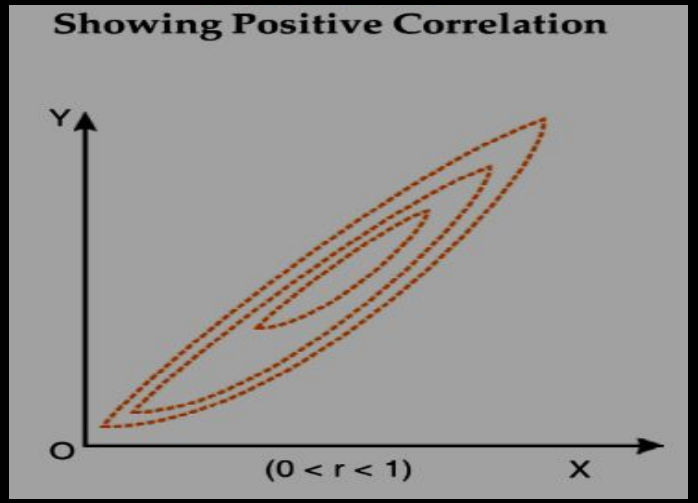
- This is a simple diagrammatic method to establish correlation between a pair of variables.

- 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.

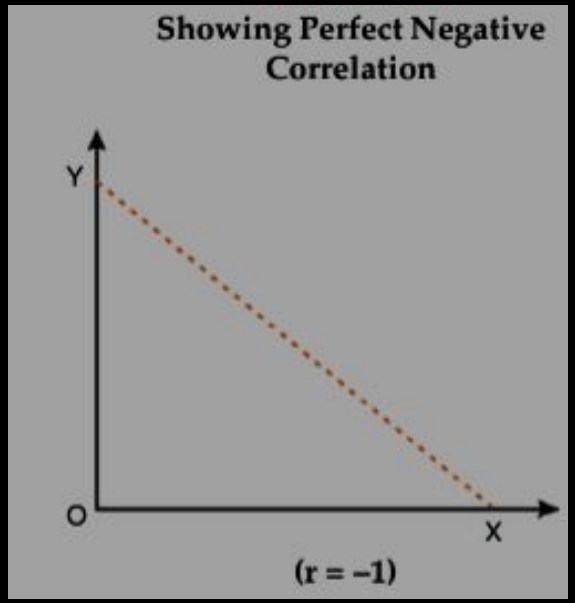
SCATTER DIAGRAM

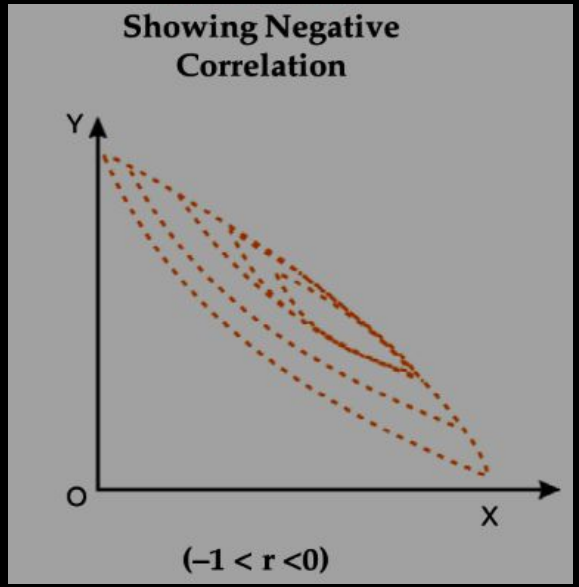




- The plotted points lie from lower left corner to upper right corner

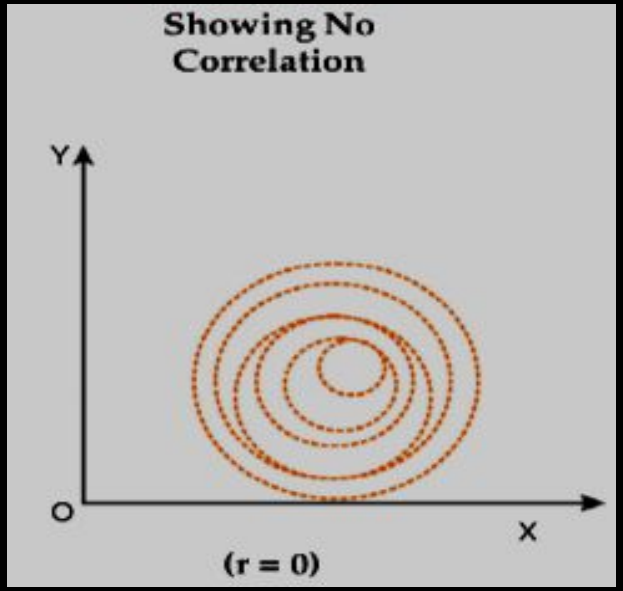
SCATTER DIAGRAM

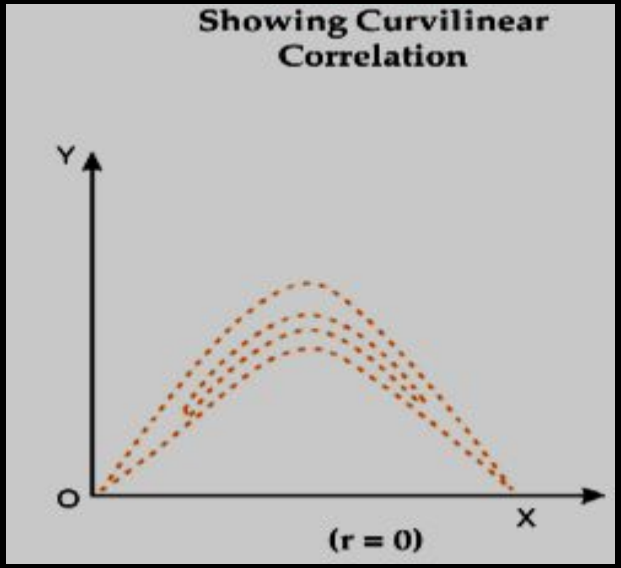




- The plotted points concentrate from upper left to lower right

SCATTER DIAGRAM





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

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).

Ans :d

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.

Ans :c

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).

Ans : a

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).

Ans :d

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).

Ans : a

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

KARL PEARSON'S PRODUCT MOMENT CORRELATION COEFFICIENT

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

where

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

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

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 $\Sigma y = 50$

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

June 2011

Que. If the covariance between variables X and Y is 8.4 and variance of X and Y are respectively 25 and 36, then the coefficient of correlation is

(a) 0.82

(b) 0.28

(c) 0.01

(d) 0.09

Ans : b

JAN 2021

Que. For the set of observations {(1, 2), (2, 5), (3, 7), (4, 8), (5, 10)} the value of karl-pearson's coefficient of correlation is approximately given by

(a) 0.755

(b) 0.655

(c) 0.525

(d) 0.985

ANS : d

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 $-1 \leq r \leq 1$.

Que. 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.

Ans : b

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 .

Que. 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.

Ans : c

Write the correct answers. Each question carries 1 mark.

Que. **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.

Ans : b

Write the correct answers. Each question carries 1 mark.

Que. 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.

Ans : c

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 $4v - 16y + 11 = 0$

DEC 2022

Que. Pearson's Correlation coefficient between x and y is:-

(a) $\frac{cov(x,y)}{S_x S_y}$ (b) $\frac{cov^2(x,y)}{S_x S_y}$

(c) $\frac{S_x S_y)^2}{cov(x,y)}$ (d) $\frac{S_x S_y}{cov(x,y)}$

Ans : a

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.

Ans : d

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.

Ans : c

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

Ans : d

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.

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

Example 17.12 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?

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 .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

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 .

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

where c is the number of concurrent deviations (same direction)

m is number of pairs compared , m = n-1

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

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.

Write the correct answers. Each question carries 1 mark.

Que. **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).

Ans : d

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

$x = a + by$

Write the correct answers. Each question carries 1 mark.

Que. **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).

Ans : a

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

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

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

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 17.15 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.

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.

PROPERTIES REGRESSION LINES / COEFFICIENTS

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

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

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

Original		change
x	→	u
y	→	v
b_{xy}	→	b_{uv}
b_{yx}	→	b_{vu}

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 $(\bar{x}, \bar{y})$ mean 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.

Que. 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)

Ans : a

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.

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

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

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

Ans : c

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.

NOTE

- Product of the regression coefficient must be numerically less than unity .
- 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

Que. 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.

Ans : b

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.

$$r^2$$

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.

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

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

(a) 0.4

(b) –0.6

(c) 0.36

(d) 0.64

Ans : d

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).

Ans : a

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.

Ans : a

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).

Ans : d

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.

Ans : d

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.

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