2.2 Correlation & Regression

2.2.1 Introduction

"If it is proved true that in a large number of instances two variables tend always to fluctuate in the same or in opposite directions, we consider that the fact is established and that a relationship exists. This relationship is called correlation."

(1) Univariate distribution : **These are the distributions in which there is only one variable such as the heights of the students of a class.**

(2) Bivariate distribution : **Distribution involving two discrete variable is called a bivariate distribution. For example, the heights and the weights of the students of a class in a school.**

(3) Bivariate frequency distribution : **Let x and y be two variables. Suppose x takes the values $x_1, x_2, ..., x_n$ and y takes the values $y_1, y_2, ..., y_n$, then we record our observations in the form of ordered pairs (x_1, y_1), where $1 \le i \le n, 1 \le j \le n$. If a certain pair occurs f_{ij} times, we say that its frequency is f_{ij}.**

The function which assigns the frequencies f_{ij}'s to the pairs (x_i, y_j) is known as a bivariate frequency distribution.

Example: 1 **The following table shows the frequency distribution of age (x) and weight (y) of a group of 60 individuals**

x (yrs) / y (yrs.)	40 – 45	45 – 50	50 – 55	55 – 60	60 – 65
45 – 50	2	5	8	3	0
50 – 55	1	3	6	10	2
55 – 60	0	2	5	12	1

Then find the marginal frequency distribution for x and y.

Solution: **Marginal frequency distribution for x**

x	40 – 45	45 – 50	50 – 55	55 – 60	60 – 65
f	3	10	19	25	3

Marginal frequency distribution for y

y	45 – 50	50 – 55	55 – 60
f	18	22	20

2.2.2 Covariance

Let $(x_1, x_i); i = 1, 2, ..., n$ be a bivariate distribution, where $x_1, x_2, ..., x_n$ are the values of variable x and $y_1, y_2, ..., y_n$ those of y. Then the covariance $Cov\ (x, y)$ between x and y is given by

$$Cov(x, y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \text{ or } Cov(x, y) = \frac{1}{n}\sum_{i=1}^{n}(x_i y_i - \bar{x}\,\bar{y}) \text{ where, } \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \text{ and } \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \text{ are}$$

means of variables x and y respectively.

Covariance is not affected by the change of origin, but it is affected by the change of scale.

Example: 2 Covariance (x, y) between x and y, if $\sum x = 15$, $\sum y = 40$, $\sum x.y = 110$, $n = 5$ is

(a) 22 (b) 2 (c) – 2 (d) None of these

Solution: (c) Given, $\sum x = 15, \sum y = 40$

$$\sum x.y = 110, \; n = 15$$

We know that, $$Cov(x, y) = \frac{1}{n}\sum_{i=1}^{n} x_i . y_i - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) = \frac{1}{n}\sum x.y - \left(\frac{1}{n}\sum x\right)\left(\frac{1}{n}\sum y\right)$$

$$= \frac{1}{5}(110) - \left(\frac{15}{5}\right)\left(\frac{40}{5}\right) = 22 - 3 \times 8 = -2.$$

2.2.3 Correlation

The relationship between two variables such that a change in one variable results in a positive or negative change in the other variable is known as correlation.

(1) **Types of correlation**

(i) Perfect correlation : If the two variables vary in such a manner that their ratio is always constant, then the correlation is said to be perfect.

(ii) Positive or direct correlation : If an increase or decrease in one variable corresponds to an increase or decrease in the other, the correlation is said to be positive.

(iii) Negative or indirect correlation : If an increase or decrease in one variable corresponds to a decrease or increase in the other, the correlation is said to be negative.

(2) **Karl Pearson's coefficient of correlation : The correlation coefficient $r(x, y)$, between two variable x and y is given by,** $r(x, y) = \dfrac{Cov(x, y)}{\sqrt{Var(x)}\sqrt{Var(y)}}$ or $\dfrac{Cov(x, y)}{\sigma_x \sigma_y}$,

$$r(x, y) = \frac{n\left(\sum_{i=1}^{n} x_i y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}}$$

$$r = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\sqrt{\Sigma(x - \bar{x})^2}\sqrt{\Sigma(y - \bar{y})^2}} = \frac{\Sigma\, dxdy}{\sqrt{\Sigma\, dx^2}\sqrt{\Sigma\, dy^2}}.$$

(3) **Modified formula :** $r = \dfrac{\sum dxdy - \dfrac{\sum dx \sum dy}{n}}{\sqrt{\left\{\sum dx^2 - \dfrac{\left(\sum dx\right)^2}{n}\right\}\left\{\sum dy^2 - \dfrac{\left(\sum dy\right)^2}{n}\right\}}}$, **where** $dx = x - \bar{x}; dy = y - \bar{y}$

Also $r_{xy} = \dfrac{Cov(x, y)}{\sigma_x \sigma_y} = \dfrac{Cov(x, y)}{\sqrt{\text{var}(x).\text{var}(y)}}$.

Example: 3 For the data

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

The Karl Pearson's coefficient is [Kerala (Engg.) 2002]

(a) $\frac{63}{\sqrt{94\times 66}}$ (b) 63 (c) $\frac{63}{\sqrt{94}}$ (d) $\frac{63}{\sqrt{66}}$

Solution: (a) Take $A = 5$, $B = 5$

x_i	y_i	$u_i = x_i - 5$	$v_i = y_i - 5$	u_i^2	v_i^2	$u_i v_i$
4	5	– 1	0	1	0	0
7	8	2	3	9	9	6
8	6	3	1	1	1	3
3	3	– 2	– 2	4	4	4
4	5	– 1	0	0	0	0
Total		$\sum u_i = 1$	$\sum v_i = 2$	$\sum u_i^2 = 19$	$\sum v_i^2 = 14$	$\sum u_i v_i = 13$

$$\Theta \quad r(x,y) = \frac{\sum u_i v_i - \frac{1}{n}\sum u_i \sum v_i}{\sqrt{\sum u_i^2 - \frac{1}{n}\left(\sum u_i\right)^2}\sqrt{\sum v_i - \frac{1}{n}\left(\sum v_i\right)^2}} = \frac{13 - \frac{1\times 2}{5}}{\sqrt{19 - \frac{1^2}{5}}\sqrt{14 - \frac{2^2}{5}}} = \frac{63}{\sqrt{94}\sqrt{66}}.$$

Example: 4 Coefficient of correlation between observations (1, 6),(2, 5),(3, 4), (4, 3), (5, 2), (6, 1) is [Pb. CET 1997; Him. CET 2001; DCE 2002]

(a) 1 (b) – 1 (c) 0 (d) None of these

Solution: (b) Since there is a linear relationship between x and y, *i.e.* $x + y = 7$

$\therefore$ Coefficient of correlation = – 1.

Example: 5 The value of co-variance of two variables x and y is $-\frac{148}{3}$ and the variance of x is $\frac{272}{3}$ and the variance of y is $\frac{131}{3}$. The coefficient of correlation is

(a) 0.48 (b) 0.78 (c) 0.87 (d) None of these

Solution : (d) We know that coefficient of correlation $= \frac{Cov(x,y)}{\sigma_x . \sigma_y}$

Since the covariance is – *ive*.

$\therefore$ Correlation coefficient must be – *ive*. Hence (d) is the correct answer.

Example: 6 The coefficient of correlation between two variables x and y is 0.5, their covariance is 16. If the S.D of x is 4, then the S.D. of y is equal to [AMU 1988, 89, 90]

(a) 4 (b) 8 (c) 16 (d) 64

Solution: (b) We have, $r_{xy} = 0.5$, $Cov(x,y) = 16$. S.D of x *i.e.*, $\sigma_x = 4$, $\sigma_y = ?$

We know that, $r(x,y) = \frac{Cov(x,y)}{\sigma_x . \sigma_y}$

$$0.5 = \frac{16}{4.\sigma_y}; \quad \therefore \ \sigma_y = 8.$$

Example: 7 For a bivariate distribution (x, y) if $\sum x = 50$, $\sum y = 60$, $\sum xy = 350$, $\bar{x} = 5, \bar{y} = 6$ variance of x is 4, variance of y is 9, then $r(x,y)$ is [AMU 1991; Pb. CET 1998; DCE 1998]

(a) 5/6 (b) 5/36 (c) 11/3 (d) 11/18

Solution: (a) $\bar{x} = \frac{\sum x}{n} \Rightarrow 5 = \frac{50}{n} \Rightarrow n = 10.$

$$\therefore Cov(x,y) = \frac{\sum xy}{n} - \bar{x}\bar{y} = \frac{350}{10} - (5)(6) = 5.$$

$\therefore \ r(x,y) = \frac{Cov(x,y)}{\sigma_x . \sigma_y} = \frac{5}{\sqrt{4}.\sqrt{9}} = \frac{5}{6}$.

Example: 8 A, B, C, D are non-zero constants, such that

(i) both A and C are negative. (ii) A and C are of opposite sign.

If coefficient of correlation between x and y is r, then that between $AX+B$ and $CY+D$ is

(a) r (b) $-r$ (c) $\frac{A}{C}r$ (d) $-\frac{A}{C}r$

Solution : (a,b) (i) Both A and C are negative.

Now $Cov(AX+B, CY+D) = AC\,Cov(X,Y)$

$\sigma_{AX+B} = |A|\sigma_x$ and $\sigma_{CY+D} = |C|\sigma_y$

Hence $\rho(AX+B, CY+D) = \frac{AC\,Cov(X,Y)}{(|A|\sigma_x)(|C|\sigma_y)} = \frac{AC}{|AC|}\rho(X,Y) = \rho(X,Y) = r, \quad (\because AC > 0)$

(ii) $\rho(AX+B, CY+D) = \frac{AC}{|AC|}\rho(X,Y), \quad (\because AC < 0)$

$$= \frac{AC}{-AC}\rho(X,Y) = -\rho(X,Y) = -r.$$

2.2.4 Rank Correlation

Let us suppose that a group of n individuals is arranged in order of merit or proficiency in possession of two characteristics A and B.

These rank in two characteristics will, in general, be different.

For example, if we consider the relation between intelligence and beauty, it is not necessary that a beautiful individual is intelligent also.

Rank Correlation : $\rho = 1 - \frac{6\sum d^2}{n(n^2-1)}$, which is the Spearman's formulae for rank correlation coefficient.

Where $\sum d^2$ = sum of the squares of the difference of two ranks and n is the number of pairs of observations.

Note : ❑ We always have, $\sum d_i = \sum(x_i - y_i) = \sum x_i - \sum y_i = n(\bar{x}) - n(\bar{y}) = 0$, $\quad (\because \bar{x} = \bar{y})$

If all ds are zero, then $r = 1$, which shows that there is perfect rank correlation between the variable and which is maximum value of r.

❑ If however some values of x_i are equal, then the coefficient of rank correlation is given by

$$r = 1 - \frac{6\left[\sum d^2 + \left(\frac{1}{12}\right)(m^3 - m)\right]}{n(n^2-1)}$$

where m is the number of times a particular x_i is repeated.

Positive and Negative rank correlation coefficients

Let r be the rank correlation coefficient then, if

- $r > 0$, it means that if the rank of one characteristic is high, then that of the other is also high or if the rank of one characteristic is low, then that of the other is also low. *e.g.,* if the two characteristics be height and weight of persons, then $r > 0$ means that the tall persons are also heavy in weight.

- $r = 1$, it means that there is perfect correlation in the two characteristics *i.e.,* every individual is getting the same ranks in the two characteristics. Here the ranks are of the type (1, 1), (2, 2),....., (*n, n*).
- $r < 1$, it means that if the rank of one characteristics is high, then that of the other is low or if the rank of one characteristics is low, then that of the other is high. *e.g.,* if the two characteristics be richness and slimness in person, then $r < 0$ means that the rich persons are not slim.
- $r = -1$, it means that there is perfect negative correlation in the two characteristics *i.e,* an individual getting highest rank in one characteristic is getting the lowest rank in the second characteristic. Here the rank, in the two characteristics in a group of *n* individuals are of the type $(1, n), (2, n-1), ..., (n, 1)$.
- $r = 0$, it means that no relation can be established between the two characteristics.

Important Tips

- ☞ *If $r = 0$, the variable x and y are said to be uncorrelated or independent.*
- ☞ *If $r = -1$, the correlation is said to be negative and perfect.*
- ☞ *If $r = +1$, the correlation is said to be positive and perfect.*
- ☞ *Correlation is a pure number and hence unitless.*
- ☞ *Correlation coefficient is not affected by change of origin and scale.*
- ☞ *If two variate are connected by the linear relation $x + y = K$, then x, y are in perfect indirect correlation. Here $r = -1$.*
- ☞ *If x, y are two independent variables, then $\rho(x+y, x-y) = \frac{\sigma_x^2 - \sigma_y^2}{\sigma_x^2 + \sigma_y^2}$.*
- ☞ $r(x,y) = \frac{\sum u_i v_i - \frac{1}{n}\sum u_i \cdot \sum v_i}{\sqrt{\sum u_i^2 - \frac{1}{n}\left(\sum u_i\right)^2}\sqrt{\sum v_i^2 - \frac{1}{n}\left(\sum v_i\right)^2}}$, where $u_i = x_i - A, v_i = y_i - B$.

Example: 9 Two numbers within the bracket denote the ranks of 10 students of a class in two subjects (1, 10), (2, 9), (3, 8), (4, 7), (5, 6), (6, 5), (7, 4), (8, 3), (9, 2), (10, 1). The rank of correlation coefficient is **[MP PET 1996]**

(a) 0 (b) – 1 (c) 1 (d) 0.5

Solution: (b) Rank correlation coefficient is $r = 1 - 6.\frac{\sum d^2}{n(n^2-1)}$, Where $d = y - x$ for pair (x, y)

$\therefore \sum d^2 = 9^2 + 7^2 + 5^2 + 3^2 + 1^2 + (-1)^2 + (-3)^2 + (-5)^2 + (-7)^2 + (-9)^2 = 330$

Also $n = 10$; $\therefore r = 1 - \frac{6 \times 330}{10(100-1)} = -1$.

Example : 10 Let $x_1, x_2, x_3, ..., x_n$ be the rank of *n* individuals according to character *A* and $y_1, y_2, ..., y_n$ the ranks of same individuals according to other character *B* such that $x_i + y_i = n + 1$ for $i = 1, 2, 3, ..., n$. Then the coefficient of rank correlation between the characters *A* and *B* is

(a) 1 (b) 0 (c) – 1 (d) None of these

Solution: (c) $x_i + y_i = n + 1$ for all $i = 1, 2, 3, ..., n$

Let $x_i - y_i = d_i$. Then, $2x_i = n + 1 + d_i \Rightarrow d_i = 2x_i - (n+1)$

$$\therefore \sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} [2x_i - (n+1)]^2 = \sum_{i=1}^{n} [4x_i^2 + (n+1)^2 - 4x_i(n+1)]$$

$$\sum_{i=1}^{n} d_i^2 = 4\sum_{i=1}^{n} x_i^2 + (n)(n+1)^2 - 4(n+1)\sum_{i=1}^{n} x_i = 4\frac{n(n+1)(2n+1)}{6} + (n)(n+1)^2 - 4(n+1)\frac{n(n+1)}{2}$$

$$\sum_{i=1}^{n} d_i^2 = \frac{n(n^2-1)}{3}.$$

$$\therefore\ r = 1 - \frac{6\sum d_i^2}{n(n^2-1)} = 1 - \frac{6(n)(n^2-1)}{3(n)(n^2-1)}\ \ i.e.,\ r = -1.$$

Regression

2.2.5 Linear Regression

If a relation between two variates x and y exists, then the dots of the scatter diagram will more or less be concentrated around a curve which is called the curve of regression. If this curve be a straight line, then it is known as line of regression and the regression is called linear regression.

Line of regression: The line of regression is the straight line which in the least square sense gives the best fit to the given frequency.

2.2.6 Equations of lines of Regression

(1) Regression line of y on x : If value of x is known, then value of y can be found as

$$y - \bar{y} = \frac{Cov(x,y)}{\sigma_x^2}(x - \bar{x}) \quad \text{or} \quad y - \bar{y} = r\frac{\sigma_y}{\sigma_x}(x - \bar{x})$$

(2) Regression line of x on y : It estimates x for the given value of y as

$$x - \bar{x} = \frac{Cov(x,y)}{\sigma_y^2}(y - \bar{y}) \quad \text{or} \quad x - \bar{x} = r\frac{\sigma_x}{\sigma_y}(y - \bar{y})$$

(3) Regression coefficient : (i) Regression coefficient of y on x is $b_{yx} = \frac{r\sigma_y}{\sigma_x} = \frac{Cov(x,y)}{\sigma_x^2}$

(ii) Regression coefficient of x on y is $b_{xy} = \frac{r\sigma_x}{\sigma_y} = \frac{Cov(x,y)}{\sigma_y^2}$.

2.2.7 Angle between Two lines of Regression

Equation of the two lines of regression are $y - \bar{y} = b_{yx}(x - \bar{x})$ **and** $x - \bar{x} = b_{xy}(y - \bar{y})$

We have, $m_1 =$ **slope of the line of regression of y on x** $= b_{yx} = r.\frac{\sigma_y}{\sigma_x}$

$m_2 =$ **Slope of line of regression of x on y** $= \frac{1}{b_{xy}} = \frac{\sigma_y}{r.\sigma_x}$

$$\therefore\ \tan\theta = \pm\frac{m_2 - m_1}{1 + m_1 m_2} = \pm\frac{\frac{\sigma_y}{r\sigma_x} - \frac{r\sigma_y}{\sigma_x}}{1 + \frac{r\sigma_y}{\sigma_x}.\frac{\sigma_y}{r\sigma_x}} = \pm\frac{(\sigma_y - r^2\sigma_y)\sigma_x}{r\sigma_x^2 + r\sigma_y^2} = \pm\frac{(1-r^2)\sigma_x\sigma_y}{r(\sigma_x^2 + \sigma_y^2)}.$$

Here the positive sign gives the acute angle θ, because $r^2 \le 1$ and σ_x, σ_y are positive.

$$\therefore\quad \tan\theta = \frac{1 - r^2}{r}.\frac{\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2} \qquad .....(i)$$

Note : ❑ **If $r = 0$, from (i) we conclude $\tan\theta = \infty$ or $\theta = \pi/2$ *i.e.,* two regression lines are at right angels.**

❑ **If $r = \pm 1$, $\tan\theta = 0$ *i.e.,* $\theta = 0$, since θ is acute *i.e.,* two regression lines coincide.**

2.2.8 Important points about Regression coefficients b_{xy} and b_{yx}

(1) $r = \sqrt{b_{yx}.b_{xy}}$ *i.e.* the coefficient of correlation is the geometric mean of the coefficient of regression.

(2) If $b_{yx} > 1$, then $b_{xy} < 1$ *i.e.* if one of the regression coefficient is greater than unity, the other will be less than unity.

(3) If the correlation between the variable is not perfect, then the regression lines intersect at $(\bar{x}, \bar{y})$.

(4) b_{yx} is called the slope of regression line y on x and $\frac{1}{b_{xy}}$ is called the slope of regression line x on y.

(5) $b_{yx} + b_{xy} > 2\sqrt{b_{yx}b_{xy}}$ or $b_{yx} + b_{xy} > 2r$, *i.e.* the arithmetic mean of the regression coefficient is greater than the correlation coefficient.

(6) Regression coefficients are independent of change of origin but not of scale.

(7) The product of lines of regression's gradients is given by $\frac{\sigma_y^2}{\sigma_x^2}$.

(8) If both the lines of regression coincide, then correlation will be perfect linear.

(9) If both b_{yx} and b_{xy} are positive, the r will be positive and if both b_{yx} and b_{xy} are negative, the r will be negative.

Important Tips

- ☞ ***If $r = 0$, then $\tan\theta$ is not defined i.e. $\theta = \frac{\pi}{2}$. Thus the regression lines are perpendicular.***
- ☞ ***If $r = +1$ or -1, then $\tan\theta = 0$ i.e. $\theta = 0$. Thus the regression lines are coincident.***
- ☞ ***If regression lines are $y = ax + b$ and $x = cy + d$, then $\bar{x} = \frac{bc + d}{1 - ac}$ and $\bar{y} = \frac{ad + b}{1 - ac}$.***
- ☞ ***If b_{yx}, b_{xy} and $r \ge 0$ then $\frac{1}{2}(b_{xy} + b_{yx}) \ge r$ and if b_{xy}, b_{yx} and $r \le 0$ then $\frac{1}{2}(b_{xy} + b_{yx}) \le r$.***
- ☞ ***Correlation measures the relationship between variables while regression measures only the cause and effect of relationship between the variables.***
- ☞ ***If line of regression of y on x makes an angle α, with the +ive direction of X-axis, then $\tan\alpha = b_{yx}$.***
- ☞ ***If line of regression of x on y makes an angle β, with the +ive direction of X-axis, then $\cot\beta = b_{xy}$.***

Example : 11 The two lines of regression are $2x - 7y + 6 = 0$ and $7x - 2y + 1 = 0$. The correlation coefficient between x and y is

[DCE 1999]

(a) – 2/7 (b) 2/7 (c) 4/49 (d) None of these

Solution: (b) The two lines of regression are $2x - 7y + 6 = 0$ .....(i) and $7x - 2y + 1 = 0$ ......(ii)

If (i) is regression equation of y on x, then (ii) is regression equation of x on y.

We write these as $y = \frac{2}{7}x + \frac{6}{7}$ and $x = \frac{2}{7}y - \frac{1}{7}$

$\therefore\ b_{yx} = \frac{2}{7},\ b_{xy} = \frac{2}{7};\quad \therefore\ b_{yx}.b_{xy} = \frac{4}{49} < 1$, So our choice is valid.

$\therefore\ r^2 = \frac{4}{49} \Rightarrow r = \frac{2}{7}$. $[\because\ b_{yx} > 0, b_{xy} > 0]$

Example: 12 Given that the regression coefficients are – 1.5 and 0.5, the value of the square of correlation coefficient is

[Kurukshetra CEE 2002]

(a) 0.75 (b) 0.7

(c) – 0.75 (d) – 0.5

Solution: (c) Correlation coefficient is given by $r^2 = b_{yx}.b_{xy} = (-1.5)(0.5) = -0.75$.

Example: 13 In a bivariate data $\sum x = 30, \sum y = 400, \sum x^2 = 196, \sum xy = 850$ and $n = 10$. The regression coefficient of y on x is

[Kerala (Engg.) 2002]

(a) – 3.1 (b) – 3.2 (c) – 3.3 (d) – 3.4

Solution: (c) $Cov(x, y) = \frac{1}{n}\sum xy - \frac{1}{n^2}\sum x.\sum y = \frac{1}{10}(850) - \frac{1}{100}(30)(400) = -35$

$Var(x) = \sigma_x^2 = \frac{1}{n}\sum x^2 - \left(\frac{\sum x}{n}\right)^2 = \frac{196}{10} - \left(\frac{30}{10}\right)^2 = 10.6$

$b_{yx} = \frac{Cov(x, y)}{Var(x)} = \frac{-35}{10.6} = -3.3.$

Example: 14 If two lines of regression are $8x - 10y + 66 = 0$ and $40x - 18y = 214$, then $(\bar{x}, \bar{y})$ is

(a) (17, 13) (b) (13, 17) (c) (– 17, 13) (d) (– 13, – 17)

Solution: (b) Since lines of regression pass through $(\bar{x}, \bar{y})$, hence the equation will be $8\bar{x} - 10\bar{y} + 66 = 0$ and $40\bar{x} - 18\bar{y} = 214$

On solving the above equations, we get the required answer $\bar{x} = 13, \bar{y} = 17$.

Example: 15 The regression coefficient of y on x is $\frac{2}{3}$ and of x on y is $\frac{4}{3}$. If the acute angle between the regression line is θ, then $\tan\theta =$ [DCE 1995]

(a) $\frac{1}{18}$ (b) $\frac{1}{9}$ (c) $\frac{2}{9}$ (d) None of these

Solution: (a) $b_{yx} = \frac{2}{3}, b_{xy} = \frac{4}{3}$. Therefore, $\tan\theta = \left|\frac{b_{xy} - \frac{1}{b_{yx}}}{1 + \frac{b_{xy}}{b_{yx}}}\right| = \left|\frac{\frac{4}{3} - \frac{3}{2}}{1 + \frac{4/3}{2/3}}\right| = \frac{1}{18}$.

Example: 16 If the lines of regression of y on x and x on y make angles 30^o and 60^o respectively with the positive direction of X-axis, then the correlation coefficient between x and y is

(a) $\frac{1}{\sqrt{2}}$ (b) $\frac{1}{2}$

(c) $\frac{1}{\sqrt{3}}$ (d) $\frac{1}{3}$

Solution: (c) Slope of regression line of y on x = $b_{yx} = \tan 30^o = \frac{1}{\sqrt{3}}$

Slope of regression line of x on y = $\frac{1}{b_{xy}} = \tan 60^o = \sqrt{3}$

$\Rightarrow b_{xy} = \frac{1}{\sqrt{3}}$. Hence, $r = \sqrt{b_{xy}.b_{yx}} = \sqrt{\left(\frac{1}{\sqrt{3}}\right)\left(\frac{1}{\sqrt{3}}\right)} = \frac{1}{\sqrt{3}}$.

Example: 17 If two random variables x and y, are connected by relationship $2x + y = 3$, then $r_{xy} =$

(a) 1 (b) – 1 (c) – 2 (d) 3

Solution: (b) Since $2x + y = 3$

$\therefore\ 2\bar{x} + \bar{y} = 3$; $\therefore\ y - \bar{y} = -2(x - \bar{x})$. So, $b_{yx} = -2$

Also $x-\bar{x}=-\frac{1}{2}(y-\bar{y})$, $\therefore$ $b_{xy}=-\frac{1}{2}$

$\therefore$ $r_{xy}^2=b_{yx}.b_{xy}=(-2)\left(-\frac{1}{2}\right)=1 \Rightarrow$ $r_{xy}=-1$. ($\because$ both b_{yx}, b_{xy} are $-ive$)

2.2.9 Standard error and Probable error

(1) Standard error of prediction : **The deviation of the predicted value from the observed value is known as the standard error prediction and is defined as** $S_y=\sqrt{\left\{\frac{\sum(y-y_p)^2}{n}\right\}}$

where y is actual value and y_p is predicted value.

In relation to coefficient of correlation, it is given by

(i) Standard error of estimate of x is $S_x=\sigma_x\sqrt{1-r^2}$ **(ii) Standard error of estimate of y is** $S_y=\sigma_y\sqrt{1-r^2}$.

(2) Relation between probable error and standard error : **If r is the correlation coefficient in a sample of n pairs of observations, then its standard error S.E. $(r)=\frac{1-r^2}{\sqrt{n}}$ and probable error P.E. (r) = 0.6745 (S.E.)= 0.6745 $\left(\frac{1-r^2}{\sqrt{n}}\right)$. The probable error or the standard error are used for interpreting the coefficient of correlation.**

(i) If $r<P.E.(r)$, there is no evidence of correlation.

(ii) If $r>6P.E.(r)$, the existence of correlation is certain.

The square of the coefficient of correlation for a bivariate distribution is known as the "Coefficient of determination".

Example: 18 If $Var(x)=\frac{21}{4}$ and $Var(y)=21$ and $r=1$, then standard error of y is

(a) 0 (b) $\frac{1}{2}$ (c) $\frac{1}{4}$ (d) 1

Solution: (a) $S_y=\sigma_y\sqrt{1-r^2}=\sigma_y\sqrt{1-1}=0$.

ASSIGNMENT

1. For the bivariate frequency table for x and y

y \ x	0 – 10	10 – 20	20 – 30	30 – 40	Sum
0 – 10	3	2	4	2	11
10 – 20	–	1	3	1	5
20 – 30	3	2	–	–	5
30 – 40	–	6	7	–	13
Sum	6	11	14	3	34

Then the marginal frequency distribution for y is given by

(a)

0 – 10	6
10 – 20	11
20 – 30	14
30 – 40	3

(b)

0 – 10	11
10 – 20	5
20 – 30	5
30 – 40	13

(c)

0 – 10	10
10 – 20	12
20 – 30	11
30 – 40	1

(d) None of these

2. The variables x and y represent height in *cm* and weight in *gm* respectively. The correlation between x and y has the unit

[MP PET 2003]

(a) *gm* (b) *cm* (c) *gm.cm* (d) None of these

3. The value of $\sum[(x-\bar{x})(y-\bar{y})]$ is

(a) $n.r_{xy}.\sigma_x\sigma_y$ (b) $r_{xy}.\sigma_x^2\sigma_y^2$ (c) $r_{xy}\sqrt{\sigma_x\sigma_y}$ (d) None of these

4. Karl Pearson's coefficient of correlation is dependent [MP PET 1993]

(a) Only on the change of origin and not on the change of scale (b) Only on the change of scale and not on the change of origin

(b) On both the change of origin and the change of scale (d) Neither on the change of scale nor on the change of origin

5. If X and Y are independent variable, then correlation coefficient is

(a) 1 (b) – 1 (c) $\frac{1}{2}$ (d) 0

6. The value of the correlation coefficient between two variable lies between [Kurukshetra CEE 1998]

(a) 0 and 1 (b) – 1 and 1 (c) 0 and ∞ (d) $-\infty$ and 0

7. The coefficient of correlation between two variables x and y is given by

(a) $r=\frac{\sigma_x^2+\sigma_y^2+\sigma_{x-y}^2}{2\sigma_x\sigma_y}$ (b) $r=\frac{\sigma_x^2+\sigma_y^2-\sigma_{x-y}^2}{2\sigma_x\sigma_y}$ (c) $r=\frac{\sigma_x^2+\sigma_y^2+\sigma_{x-y}^2}{\sigma_x\sigma_y}$ (d) $r=\frac{\sigma_x^2+\sigma_y^2-\sigma_{x-y}^2}{\sigma_x\sigma_y}$

8. If r is the correlation coefficient between two variables, then [MP PET 1995; Pb. CET 1995]

(a) $r \geq 1$ (b) $r \leq 1$ (c) $|r| \leq 1$ (d) $|r| \geq 1$

9. When the correlation between two variables is perfect, then the value of coefficient of correlation r is

(a) –1 (b) +1 (c) 0 (d) ± 1

10. If correlation between x and y is r, then between y and x correlation will be

(a) $-r$ (b) $\frac{1}{r}$ (c) r (d) 1- r

11. If r is the coefficient of correlation and $Y=a+bX$, then $|r|=$

(a) $\frac{a}{b}$ (b) $\frac{b}{a}$ (c) 1 (d) None of these

12. If coefficient of correlation between the variables x and y is zero, then

(a) Variables x and y have no relation (b) y decreases as x increases

(c) y increases as x increases (d) There may be a relation between x and y

13. When the origin is changed, then the coefficient of correlation

(a) Becomes zero (b) Varies (c) Remains fixed (d) None of these

14. If $r=-0.97$, then

(a) Correlation is negative and curved (b) Correlation is linear and negative

(c) Correlation is in third and fourth quadrant (d) None of these

15. In a scatter diagram, if plotted points form a straight line running from the lower left to the upper right corner, then there exists a

(a) High degree of positive correlation (b) Perfect positive correlation

(c) Perfect negative correlation (d) None of these

16. If the two variables x and y of a bivariate distribution have a perfect correlation, they may be connected by [Kurukshetra CEE

(a) $xy=1$ (b) $\frac{a}{x}+\frac{b}{y}=1$ (c) $\frac{x}{a}+\frac{y}{b}=1$ (d) None of these

17. If x and y are related as $y-4x=3$, then the nature of correlation between x and y is [AMU 1998]

(a) Perfect positive (b) Perfect negative (c) No correlation (d) None of these

18. If $\sum x=15, \sum y=36, \sum xy=110, n=5$ then $Cov(x,y)$ equals [AI CBSE 1991]

(a) $\frac{1}{5}$ (b) $\frac{-1}{5}$ (c) $\frac{2}{5}$ (d) $-\frac{2}{5}$

19. For a bivariable distribution (x, y), if $\sum xy = 350, \sum x = 50, \sum y = 60, \bar{x} = 5, \bar{y} = 6,$ then $Cov(x, y)$ equals [Pb. CET 1997, AMU 1992]

(a) 5 (b) 6 (c) 22 (d) 28

20. For covariance the number of variate values in the two given distribution should be [AMU 1989]

(a) Unequal (b) Any number in one and any number in the other

(c) Equal (d) None of these

21. If x and y are independent variables, then [AMU 1994]

(a) $Cov(x, y) = 1$ (b) $Cov(x, y) = -1$ (c) $Cov(x, y) = 0$ (d) $Cov(x, y) = \pm\frac{1}{2}$

22. If

x:	3	4	8	6	2	1
y:	5	3	9	6	9	2

then the coefficient of correlation will be approximately [AI CBSE 1990]

(a) 0.49 (b) 0. 40 (c) – 0. 49 (d) – 0. 40

23. The coefficient of correlation for the following data

x	20	25	30	35	40	45
y	16	10	8	20	5	10

will be [AI CBSE 1988]

(a) 0. 32 (b) – 0.32 (c) 0. 35 (d) None of these

24. Coefficient of correlation from the following data

x:	1	2	3	4	5
y:	2	5	7	8	10

will be [DSSE 1983, AI CBSE 1991]

(a) 0. 97 (b) – 0.97 (c) 0. 90 (d) None of these

25. Coefficient of correlation between x and y for the following data

x:	15	16	17	17	18	20	10
y:	12	17	15	16	12	15	11

will be approximately [DSSE 1979, 81; AI CBSE 1990]

(a) 0. 50 (b) 0. 53 (c) – 0. 50 (d) – 0. 53

26. Karl Pearson's coefficient of correlation between x and y for the following data [AISSE 1983, 85, 90]

x:	3	4	8	9	6	2	1
y:	5	3	7	7	6	9	2

(a) 0. 480 (b) – 0. 480 (c) 0. 408 (d) – 0. 408

27. The coefficient of correlation for the following data

x:	1	2	3	4	5	6	7	8	9	10
y:	3	10	5	1	2	9	4	8	7	6

will be [AISSE 1986, 1990]

(a) 0. 224 (b) 0. 240 (c) 0. 30 (d) None of these

28. Karl Pearson's coefficient of correlation between the marks in English and Mathematics by ten students

Marks in English	20	13	18	21	11	12	17	14	19	15
Marks in	17	12	23	25	14	8	19	21	22	19

Maths										

will be [AISSE 1979, 82]

(a) 0. 75 (b) – 0. 75 (c) 0. 57 (d) None of these

29. Coefficient of correlation between x and y for the following data

x	– 4	–3	–2	–1	0	1	2	3	4
y	16	9	4	1	0	1	4	9	16

will be [Mathematics Olympiad 1981; DSSE 1980]

(a) 1 (b) –1 (c) 0 (d) None of these

30. If the variances of two variables x and y are respectively 9 and 16 and their covariance is 8, then their coefficient of correlation is

[MP PET 1998]

(a) $\frac{2}{3}$ (b) $\frac{8}{3\sqrt{2}}$ (c) $\frac{9}{8\sqrt{2}}$ (d) $\frac{2}{9}$

31. If the co-efficient of correlation between x and y is 0. 28, covariance between x and y is 7.6 and the variance of x is 9, then the S.D. of y series is

(a) 9.8 (b) 10. 1 (c) 9.05 (d) 10. 05

32. If $Cov(x, y) = 0$, then $\rho(x, y)$ equals [AMU 1993]

(a) 0 (b) 1 (c) – 1 (d) $\pm\frac{1}{2}$

33. Karl Pearson's coefficient of correlation between the heights (in inches) of teachers and students corresponding to the given data

Height of teachers x :	66	67	68	69	70
Height of students y :	68	66	69	72	70

is [MP PET 1993]

(a) $\frac{1}{\sqrt{2}}$ (b) $\sqrt{2}$ (c) $-\frac{1}{\sqrt{2}}$ (d) 0

34. The coefficient of correlation between x and y is 0.6, then covariance is 16. Standard deviation of x is 4, then the standard deviation of y is

(a) 5 (b) 10 (c) 20/3 (d) None of these

35. If $Cov(u, v) = 3, \sigma_u^2 = 4.5, \sigma_v^2 = 5.5$, then $\rho(u, v)$ is [AMU 1988]

(a) 0.121 (b) 0.603 (c) 0.07 (d) 0.347

36. Given $n = 10, \sum x = 4, \sum y = 3, \sum x^2 = 8, \sum y^2 = 9$ and $\sum xy = 3$, then the coefficient of correlation is [Pb. CET 1999]

(a) $\frac{1}{4}$ (b) $\frac{7}{12}$ (c) $\frac{15}{4}$ (d) $\frac{14}{3}$

37. Let r_{xy} be the coefficient of correlation between two variables x and y. If the variable x is multiplied by 3 and the variable y is increased by 2, then the correlation coefficient of the new set of variables is

(a) r_{xy} (b) $3r_{xy}$ (c) $3r_{xy} + 2$ (d) None of these

38. Coefficient of correlation between the two variates X and Y is

X	1	2	3	4	5
Y	5	4	3	2	1

(a) 0 (b) –1 (c) 1 (d) None of these

39. The coefficient of correlation between two variables X and Y is 0.5, their covariance is 15 and $\sigma_x = 6$, then $\sigma_y =$ [AMU 1998

(a) 5 (b) 10 (c) 20 (d) 6

40. Karl Pearson's coefficient of rank correlation between the ranks obtained by ten students in Mathematics and Chemistry in a class test as given below

Rank in Mathematics :	1	2	3	4	5	6	7	8	9	10
Rank in Chemistry :	3	10	5	1	2	9	4	8	7	6

is [AISSE 1990]

(a) 0.224 (b) 0.204 (c) 0.240 (d) None of these

41. The sum of squares of differences in ranks of marks obtained in Physics and Chemistry by 10 students in a test is 150, then the co-efficient of rank-correlation is given by

(a) 0.909 (b) 0.091 (c) 0.849 (d) None of these

42. If a, b, h, k are constants, while U and V are $U = \frac{X-a}{h}$, $V = \frac{Y-b}{k}$, then [DCE 1999]

(a) $Cov(X, Y) = Cov(U, V)$ (b) $Cov(X, Y) = hk\, Cov(U, V)$

(c) $Cov(X, Y) = ab\, Cov(U, V)$ (d) $Cov(U, V) = hk\, Cov(X, Y)$

43. Let X, Y be two variables with correlation coefficient $\rho(X, Y)$ and variables U, V be related to X, Y by the relation $U = 2X$, $V = 3Y$, then $\rho(U, V)$ is equal to [AMU 1999]

(a) $\rho(X, Y)$ (b) $6\rho(X, Y)$ (c) $\sqrt{6}\rho(X, Y)$ (d) $\frac{3}{2}\rho(X, Y)$

44. If X and Y are two uncorrelated variables and if $u = X + Y$, $v = X - Y$, then $r(u, v)$ is equal to [DCE 1998]

(a) $\frac{\sigma_x^2 + \sigma_y^2}{\sigma_x^2 - \sigma_y^2}$ (b) $\frac{\sigma_x^2 - \sigma_y^2}{\sigma_x^2 + \sigma_y^2}$ (c) $\frac{\sigma_x^2 + \sigma_y^2}{\sigma_x \sigma_y}$ (d) None of these

45. If $\bar{x} = \bar{y} = 0, \sum x_i y_i = 12, \sigma_x = 2, \sigma_y = 3$ and $n = 10$, then the coefficient of correlation is [MP PET 1999]

(a) 0.4 (b) 0.3 (c) 0.2 (d) 0.1

46. Let X and Y be two variables with the same variance and U and V be two variables such that $U = X + Y$, $V = X - Y$. Then $Cov(U, V)$ is equal to

(a) $Cov(X, Y)$ (b) 0 (c) 1 (d) –

47. If there exists a linear statistical relationship between two variables x and y, then the regression coefficient of y on x is [MP PET 1998]

(a) $\frac{cov(x, y)}{\sigma_x \cdot \sigma_y}$

(b) $\frac{cov(x, y)}{\sigma_y^2}$

(c) $\frac{cov(x, y)}{\sigma_x^2}$

(d) $\frac{cov(x, y)}{\sigma_x}$, where σ_x, σ_y are standard deviations of x and y respectively.

48. If $ax + by + c = 0$ is a line of regression of y on x and $a_1 x + b_1 y + c_1 = 0$ that of x on y, then

(a) $a_1 b \le a b_1$ (b) $a a_1 = b b_1$ (c) $a b_1 \le a_1 b$ (d) None of these

49. Least square lines of regression give best possible estimates, when $\rho(X, Y)$ is [DCE 1996]

(a) < 1 (b) > -1 (c) -1 or 1 (d) None of these

50. Which of the following statement is correct [Kurukshetra CEE 1995]

(a) Correlation coefficient is the arithmetic mean of the regression coefficient

(b) Correlation coefficient is the geometric mean of the regression coefficient

(c) Correlation coefficient is the harmonic mean of the regression coefficient

(d) None of these

51. The relationship between the correlation coefficient r and the regression coefficients b_{xy} and b_{yx} is[MP PET 2003; Pb. CET 19

(a) $r=\frac{1}{2}(b_{xy}+b_{yx})$ (b) $r=\sqrt{b_{xy}.b_{yx}}$ (c) $r=(b_{xy}b_{yx})^2$ (d) $r=b_{xy}+b_{yx}$

52. If the coefficient of correlation is positive, then the regression coefficients [Pb. CET 1998; PU CET 2002]

(a) Both are positive

(b) Both are negative

(c) One is positive and another is negative

(d) None of these

53. If b_{yx} and b_{xy} are both positive (where b_{yx} and b_{xy} are regression coefficients), then [MP PET 2001]

(a) $\frac{1}{b_{yx}}+\frac{1}{b_{xy}}<\frac{2}{r}$ (b) $\frac{1}{b_{yx}}+\frac{1}{b_{xy}}>\frac{2}{r}$

(c) $\frac{1}{b_{yx}}+\frac{1}{b_{xy}}<\frac{r}{2}$ (d) None of these

54. If x_1 and x_2 are regression coefficients and r is the coefficient of correlation, then

(a) $x_1-x_2>r$ (b) $x_1+x_2<r$ (c) $x_1+x_2\geq 2r$ (d) None of these

55. If one regression coefficient be unity, then the other will be

(a) Greater than unity (b) Greater than or equal to unity (c) Less than or equal to unity (d)

56. If one regression coefficient be less than unity, then the other will be

(a) Less than unity (b) Equal to unity (c) Greater than unity (d) All of the above

57. If regression coefficient of y on x is 2, then the regression coefficient of x on y is [AMU 1990]

(a) 2 (b) $\frac{1}{2}$ (c) $\leq\frac{1}{2}$ (d) None of these

58. The lines of regression of x on y estimates [AMU 1993]

(a) x for a given value of y(b) y for a given value of x (c) x from y and y from x (d) None of these

59. The statistical method which helps us to estimate or predict the unknown value of one variable from the known value of the related variable is called [Pb. CET 1995]

(a) Correlation (b) Scatter diagram (c) Regression (d) Dispersion

60. The coefficient of correlation between two variables x and y is 0.8 while regression coefficient of y on x is 0.2. Then the regression coefficient of x on y is [MP PET 1993]

(a) –3.2 (b) 3.2 (c) 4 (d) 0.16

61. If the lines of regression coincide, then the value of correlation coefficient is

(a) 0 (b) 1 (c) 0.5 (d) 0.33

62. Two lines of regression are $3x+4y-7=0$ and $4x+y-5=0$. Then correlation coefficient between x and y is[AI CBSE 1991]

(a) $\frac{\sqrt{3}}{4}$ (b) $-\frac{\sqrt{3}}{4}$ (c) $\frac{3}{16}$ (d) $-\frac{3}{16}$

63. If the two lines of regression are $4x+3y+7=0$ and $3x+4y+8=0$, then the means of x and y are [AI CBSE 1990]

(a) $-\frac{4}{7},-\frac{11}{7}$ (b) $-\frac{4}{7},\frac{11}{7}$ (c) $\frac{4}{7},-\frac{11}{7}$ (d) 4, 7

64. The two regression lines for a bivariate data are $x+y+50=0$ and $2x+3y+K=0$. If $\bar{x}=0$, then $\bar{y}$ is

[BCA Delhi Entrance Exam. 1999]

(a) 50 (b) $K-100$ (c) – 50 (d) $50+K$

65. The two regression lines are $2x-9y+6=0$ and $x-2y+1=0$. What is the correlation coefficient between x and y [DCE 1999]

(a) $-\frac{2}{3}$ (b) $\frac{2}{3}$ (c) $\frac{4}{9}$ (d) None of these

66. If the two regression coefficient between x and y are 0.8 and 0.2, then the coefficient of correlation between them is [MP PET

(a) 0.4 (b) 0.6 (c) 0.3 (d) 0.5

67. The two lines of regression are given by $3x+2y=26$ and $6x+y=31$. The coefficient of correlation between x and y is [DCE 2000]

(a) $-\frac{1}{3}$ (b) $\frac{1}{3}$ (c) $-\frac{1}{2}$ (d) $\frac{1}{2}$

68. If the lines of regression be $x-y=0$ and $4x-y-3=0$ and $\sigma_x^2=1$, then the coefficient of correlation is

(a) – 0.5 (b) 0.5 (c) 1.0 (d) – 1.0

69. A student obtained two regression lines as $L_1 \equiv x-5y+7=0$ and $L_2 \equiv 3x+y-8=0$. Then the regression line of y on x is

(a) L_1 (b) L_2 (c) Neither of the two (d) $x-5y=0$

70. If b_{yx} and b_{xy} are regression coefficients of y on x and x on y respectively, then which of the following statement is true

[Pb. CET 1996]

(a) $b_{xy}=1.5, b_{yx}=1.4$ (b) $b_{xy}=1.5, b_{yx}=0.9$ (c) $b_{xy}=1.5, b_{yx}=0.8$ (d) $b_{xy}=1.5, b_{yx}=0.6$

71. Angle between two lines of regression is given by [Kurukshetra CEE 2000; DCE 1998]

(a) $\tan^{-1}\left(\frac{b_{yx}-\frac{1}{b_{xy}}}{1+\frac{b_{xy}}{b_{yx}}}\right)$ (b) $\tan^{-1}\left(\frac{b_{yx}-b_{xy}-1}{b_{yx}+b_{xy}}\right)$ (c) $\tan^{-1}\left(\frac{b_{xy}-\frac{1}{b_{yx}}}{1+\frac{b_{xy}}{b_{yx}}}\right)$ (d) $\tan^{-1}\left(\frac{b_{yx}-b_{xy}}{1+b_{yx}.b_{xy}}\right)$

72. If acute angle between the two regression lines is θ, then

(a) $\sin\theta \ge 1-r^2$ (b) $\tan\theta \ge 1-r^2$ (c) $\sin\theta \le 1-r^2$ (d) $\tan\theta \le 1-r^2$

73. If the angle between the two lines of regression is 90°, then it represents [DCE 1999]

(a) Perfect correlation (b) Perfect negative correlation (c) No linear correlation (d) None of these

74. If $2x+y=7$ and $x+2y=7$ are the two regression lines respectively, then the correlation co-efficient between x and y is

[DCE 1983; AMU 1993]

(a) + 1 (b) –1 (c) $+\frac{1}{2}$ (d) $-\frac{1}{2}$

75. For a perfect correlation between the variables x and y, the line of regression is $ax+by+c=0$ where $a, b, c>0$; then $\rho(x,y)=$

[AMU 1999]

(a) 0 (b) –1 (c) 1 (d) None of these

76. If two random variables X and Y of a bivariate distribution are connected by the relationship $3x+2y=4$, then correlation coefficient r_{xy} equals [AMU 1999]

(a) 1 (b) –1 (c) 2/3 (d) –2/3

77. Two variables x and y are related by the linear equation $ax+by+c=0$. The coefficient of correlation between the two is +1, if

[DCE 2002]

(a) a is positive (b) b is positive (c) a and b both are positive (d) a and b are of opposite sign

78. **If the two lines of regression are $5x+3y=55$ and $7x+y=45$, then the correlation coefficient between x and y is[AMU 1998]**

(a) $+1$ (b) -1 (c) $-\sqrt{\frac{5}{21}}$ (d) $-\sqrt{\frac{21}{5}}$

79. **The error of prediction of x from the required line of regression is given by,**

(where ρ is the co-efficient of correlation) **[AMU 1992]**

(a) $\sigma_x(1-\rho^2)$ (b) $n\sigma_x^2(1-\rho^2)$ (c) $\sigma_x^2(1-\rho^2)$ (d) $n\sigma_y^2(1-\rho^2)$

80. **Probable error of r is**

(a) $0.6745\left(\frac{1-r^2}{\sqrt{n}}\right)$ (b) $0.6754\left(\frac{1+r^2}{\sqrt{n}}\right)$ (c) $0.6547\left(\frac{1-r^2}{n}\right)$ (d) $0.6754\left(\frac{1-r^2}{n}\right)$

81. **For the following data**

	x	y
Mean	65	67
Standard deviation	5.0	2.5
Correlation coefficient	0.8	

Then the equation of line of regression of y on x is

(a) $y-67=\frac{2}{5}(x-65)$ (b) $y-67=\frac{1}{5}(x-65)$ (c) $x-65=\frac{2}{5}(y-67)$ (d) $x-65=\frac{1}{5}(y-67)$

82. **If the lines of regression of y on x and that of x on y are $y=kx+4$ and $x=4y+5$ respectively, then**

(a) $k\le 0$ (b) $k\ge 0$ (c) $0\le k\le\frac{1}{4}$ (d) $0\le k\le 1$

83. **From the following observations $\{(x,y)\}=\{(1,7),(4,5),(7,2),(10,6),(13,5)\}$. The line of regression of y on x is[AI CBSE 1991]**

(a) $7x+30y-187=0$ (b) $7x-30y-187=0$ (c) $7x-30y+187=0$ (d) None of these

84. **If the variance of $x=9$ and regression equations are $4x-5y+33=0$ and $20x-9y-10=0$, then the coefficient of correlation between x and y and the variance of y respectively are** **[AMU1997, 2002]**

(a) 0.6; 16 (b) 0.16; 16 (c) 0.3; 4 (d) 0.6; 4

85. **If the two lines of regression are $x+4y=3$ and $3x+y=15$, then value of x for $y=3$ is** **[DCE 1998]**

(a) 4 (b) –9 (c) – 4 (d) None of these

86. **Which of the following two sets of regression lines are the true representative of the information from the bivariate population**

I. $x+4y=15$ and $y+3x=12$, $\bar{x}=3$, $\bar{y}=3$ **II.** $3x+4y=9$ and $4x+y=1$, $\bar{x}=-\frac{5}{10}$, $\bar{y}=\frac{30}{13}$ **[AMU 2000]**

(a) Both I and II (b) II only (c) I only (d) None of these

87. **Out of the two lines of regression given by $x+2y=4$ and $2x+3y-5=0$, the regression line of x on y is[Kurukshetra CEE 1996**

(a) $x+2y=4$ (b) $2x+3y-5=0$

(c) The given lines cannot be the regression lines (d) $x+2y=0$

88. **Regression of savings (S) of a family on income Y may be expressed as $S=a+\frac{Y}{m}$, where a and m are constants. In a random sample of 100 families the variance of savings is one-quarter of the variance of incomes and the correlation coefficient is found to be 0.4. The value of m is**

(a) 2 (b) 5 (c) 8 (d) None of these

✻ ✻ ✻