Chapter – 9

Correlation and Regression Analysis

Ex 9.1

Question 1.
Calculate the correlation co-efficient for the following data:

X	5	10	5	11	12	4	3	2	7	1
Y	1	6	2	8	5	1	4	6	5	2

Solution:

X	Y	$x = X - \bar{X}$ $= X - 6$	$y = Y - \bar{Y}$ $= Y - 4$	x^2	y^2	xy
5	1	−1	−3	1	9	3
10	6	4	2	16	4	8
5	2	−1	−2	1	4	2
11	8	5	4	25	16	20
12	5	6	1	36	1	6
4	1	−2	−3	4	9	6
3	4	−3	0	9	0	0
2	6	−4	2	16	4	−8
7	5	1	1	1	1	1
1	2	−5	−2	25	4	10
60	40	0	0	134	52	48

$N = 10, \Sigma X = 60, \Sigma Y = 40, \Sigma x^2 = 134, \Sigma y^2 = 52, \Sigma xy = 48, \bar{X} = \frac{\Sigma X}{N} = \frac{60}{10} = 6, \bar{Y} = \frac{\Sigma Y}{N} = \frac{40}{10} = 4,$

Coefficient of correlation

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}, \text{ Where } x = X - \bar{X}, y = Y - \bar{Y}$$

$$r = \frac{48}{\sqrt{134 \times 52}} = 0.575$$

Question 2.
Find the coefficient of correlation for the following:

Cost (₹)	14	19	24	21	26	22	15	20	19
Sales (₹)	31	36	48	37	50	45	33	41	39

Solution:

X	Y	$x = X - \bar{X}$ $= X - 20$	$y = Y - \bar{Y}$ $= Y - 40$	x^2	y^2	xy
14	31	−6	−9	36	81	54
19	36	−1	−4	1	16	4
24	48	4	8	16	64	32
21	37	1	−3	1	9	−3
26	50	6	10	36	100	60
22	45	2	5	4	25	10
15	33	−5	−7	25	49	35
20	41	0	1	0	1	0
19	39	−1	−1	1	1	1
180	360	0	0	120	346	193

$N = 9$, $\Sigma X = 180$, $\bar{X} = \frac{\Sigma X}{N} = \frac{180}{9} = 20$, $\Sigma Y = 360$, $\bar{Y} = \frac{\Sigma Y}{N} = \frac{360}{9} = 40$, $\Sigma x^2 = 25$, $\Sigma y^2 = 49$,
$\Sigma xy = 193$

$$\text{Correlation coefficient} = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$$

$$r = \frac{193}{\sqrt{120 \times 346}}$$

$$= 0.947$$

Question 3.
Calculate the coefficient of correlation for the ages of husbands and their respective wives:

Age of husbands	23	27	28	29	30	31	33	35	36	39
Age of wives	18	22	23	24	25	26	28	29	30	32

Solution:

Without deviation:

Age of husbands (X)	Age of wives (Y)	X^2	Y^2	XY
23	18	529	324	414
27	22	729	484	594
28	23	784	529	644
29	24	841	576	696
30	25	900	625	750
31	26	961	676	806
33	28	1089	784	924
35	29	1225	841	1015
36	30	1296	900	1080
39	32	1521	1024	1248
311	257	9875	6763	8171

$\Sigma X = 311, \Sigma Y = 257, \Sigma x^2 = 9875, \Sigma y^2 = 6763, \Sigma xy = 8171$

N = 10

Coefficient of correlation

$$r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{N\Sigma X^2 - (\Sigma X)^2}\sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$$

$$= \frac{10 \times 8171 - 311 \times 257}{\sqrt{10 \times 9875 - (311)^2}\sqrt{10 \times 6763 - (257)^2}}$$

$$= \frac{81710 - 79927}{\sqrt{98750 - 96721}\sqrt{67630 - 66049}}$$

$$= \frac{1783}{45 \times 39.76}$$

$$= 0.9965$$

Note: We can do the above problem using deviations taken from arithmetic means of X and Y. i.e., using

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}}$$

Question 4.
Calculate the coefficient of correlation between X and Y series from the following data:

Description	X	Y
Number of pairs of observation	15	15
Arithmetic mean	25	18
Standard deviation	3.01	3.03
Sum of squares of deviation from the arithmetic mean	136	138

The summation of product deviations of X and Y series from their respective arithmetic means is 122.

Solution:

$N = 15$, $\overline{X} = 25$, $\overline{Y} = 18$, $x = X - \overline{X}$, $y = Y - \overline{Y}$, $\Sigma x^2 = 136$, $\Sigma y^2 = 138$, $\Sigma xy = 122$

Correlation coefficient

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$$

$$= \frac{122}{\sqrt{136 \times 138}} = \frac{122}{136.996} = 0.891$$

Question 5.
Calculate the correlation coefficient for the following data:

X	25	18	21	24	27	30	36	39	42	48
Y	26	35	48	28	20	36	25	40	43	39

Solution:

X	Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	x^2	y^2	xy
25	26	6	–8	36	64	48
18	35	–13	1	169	1	– 13
21	48	–10	14	100	196	– 140
24	28	–7	– 6	49	36	42
27	20	– 4	–14	16	196	56
30	36	–1	2	1	4	– 2
36	25	5	–9	25	81	– 45
39	40	8	6	64	36	48
42	43	11	9	121	81	99
48	39	17	5	289	25	85
310	340	0	0	870	720	178

$N = 10, \Sigma X = 310, \Sigma Y = 340, \Sigma x^2 = 870, \Sigma y^2 = 720, \Sigma xy = 178$

$$\bar{X} = \frac{\Sigma X}{N} = \frac{310}{10} = 31, \ \bar{Y} = \frac{\Sigma Y}{N} = \frac{340}{10} = 34$$

Correlation coefficient

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$$

$$= \frac{178}{\sqrt{870 \times 720}} = \frac{178}{791.5} = 0.225$$

Question 6.
Find the coefficient of correlation for the following:

X	78	89	96	69	59	79	68	62
Y	121	72	88	60	81	87	123	92

Solution:

X	Y	$dx = X - 75$	$dy = Y - 90$	dx^2	dy^2	$dxdy$
78	121	3	31	9	961	93
89	72	14	–18	196	324	–252
96	88	21	–2	441	4	–42
69	60	–6	–30	36	900	180
59	81	–16	–9	256	81	144
79	87	4	–3	16	9	–12
68	123	–7	33	49	1089	–231
62	92	–13	2	169	4	–26
600	724	0	4	1172	3372	–146

$N = 8$, $\Sigma X = 600$, $\Sigma Y = 724$, $\Sigma dx^2 = 1172$, $\Sigma dy^2 = 3372$, $\Sigma dxdy = -146$

Correlation coefficient

$$r = \frac{N\Sigma dxdy - \Sigma dx \Sigma dy}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2}\sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$$

$$= \frac{8(-146) - 0(4)}{\sqrt{8 \times 1172 - 0}\sqrt{8 \times 3372 - 16}}$$

$$= \frac{-1168}{96.83 \times 164.2} = -0.0735$$

Question 7.

An examination of 11 applicants for an accountant post was taken by a finance company. The marks obtained by the applicants in the reasoning and aptitude tests are given below.

Applicant	A	B	C	D	E	F
Reasoning test	20	50	28	25	70	90
Aptitude test	30	60	50	40	85	90

Applicant	G	H	I	J	K
Reasoning test	76	45	30	19	26
Aptitude test	56	82	42	31	49

Calculate Spearman's rank correlation coefficient from the data given above.

Solution:

X	Y	R_X	R_Y	$d = R_X - R_Y$	d^2
20	30	10	11	–1	1
50	60	4	4	0	0
28	50	7	6	1	1
25	40	9	9	0	0
70	85	3	2	1	1
90	90	1	1	0	0
76	56	2	5	–3	9
45	82	5	3	2	4
30	42	6	8	–2	4
19	31	11	10	1	1
26	49	8	7	1	1
					$\Sigma d^2 = 22$

N = 11, $\Sigma d^2 = 22$

Rank correlation

$$\rho = 1 - \frac{6\Sigma d^2}{N(N^2-1)}$$

$$= 1 - \frac{6 \times 22}{11(121-1)}$$

$$= 1 - \frac{12}{120} = 1 - 0.1 = 0.9$$

Question 8.
The following are the ranks obtained by 10 students in commerce and accountancy are given below:

Commerce	6	4	3	1	2	7	9	8	10	5
Accountancy	4	1	6	7	5	8	10	9	3	2

To what extent is the knowledge of students in the two subjects related?

Solution:

R_X	R_Y	$d = R_X - R_Y$	d^2
6	4	2	4
4	1	3	9
3	6	–3	9
1	7	–6	36
2	5	–3	9
7	8	–1	1
9	10	–1	1
8	9	–1	1
10	3	7	49
5	2	3	9
			$\Sigma d^2 = 128$

N = 11, $\Sigma d^2 = 128$
Rank correlation

$$\rho = 1 - \frac{6\Sigma d^2}{N(N^2 - 1)}$$

$$= 1 - \frac{6 \times 128}{10(100 - 1)}$$

$$= 1 - \frac{6 \times 128}{990} = 0.224$$

Question 9.
A random sample of recent repair jobs was selected and the estimated cost and actual cost were recorded.

Estimated cost	300	450	800	250	500	975	475	400
Actual cost	273	486	734	297	631	872	396	457

Calculate the value of spearman's correlation coefficient.

Solution:

X	Y	R_X	R_Y	$d = R_X - R_Y$	d^2
300	273	7	8	–1	1
450	486	5	4	1	1
800	734	2	2	0	0
250	297	8	7	1	1
500	631	3	3	0	0
975	872	1	1	0	0
475	396	4	6	–2	4
400	457	6	5	1	1
					$\Sigma d^2 = 8$

$N = 8$, $\Sigma d^2 = 8$
Rank correlation

$$= 1 - \frac{6\Sigma d^2}{N(N^2 - 1)}$$

$$= 1 - \frac{6 \times 8}{8(64 - 1)} = 1 - \frac{6}{63} = 0.905$$

Question 10.
The rank of 10 students of the same batch in two subjects A and B are given below. Calculate the rank correlation coefficient.

Rank of A	1	2	3	4	5	6	7	8	9	10
Rank of B	6	7	5	10	3	9	4	1	8	2

Solution:

R_X	R_Y	$d = R_X - R_Y$	d^2
1	6	–5	25
2	7	–5	25
3	5	–2	4
4	10	–6	36
5	3	2	4
6	9	–3	9
7	4	3	9
8	1	7	49
9	8	1	1
10	2	8	64
			$\Sigma d^2 = 226$

$N = 10$, $\Sigma d^2 = 226$
Rank correlation

$$\rho = 1 - \frac{6\Sigma d^2}{N(N^2-1)}$$

$$= 1 - \frac{6 \times 226}{10(100-1)}$$

$$= 1 - \frac{6 \times 226}{10 \times 99}$$

$$= -0.37$$

Ex 9.2

Question 1.
From the data given below:

Marks in Economics:	25	28	35	32	31
Marks in Statistics:	43	46	49	41	36
Marks in Economics:	36	29	38	34	32
Marks in Statistics:	32	31	30	33	39

Find
(a) The two regression equations
(b) The coefficient of correlation between marks in Economics and statistics
(c) The most likely marks in Statistics when the marks in Economics is 30.

Solution:

Marks in Economics X	Marks in Statistics Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	x^2	y^2	xy
25	43	–7	5	49	25	–35
28	46	–4	8	16	64	–32
35	49	3	11	9	121	33
32	41	0	3	0	9	0
31	36	–1	–2	1	4	2
36	32	4	–6	16	36	–24
29	31	–3	–7	9	49	21
38	30	6	–8	36	64	–48

34	33	2	–5	4	25	–10
32	39	0	1	0	1	0
320	380	0	0	140	398	–93

$N = 10, \Sigma X = 320, \Sigma Y = 380, \Sigma x^2 = 140, \Sigma y^2 = 398, \Sigma XY = -93, \bar{X} = \frac{320}{10} = 32, \bar{Y} = \frac{380}{10} = 38$

(a) Regression equation of X on Y.

$$b_{xy} = r\frac{\sigma_x}{\sigma_y} = \frac{\Sigma xy}{\Sigma y^2} = \frac{-93}{398} = -0.234$$

$$X - \bar{X} = b_{xy}(Y - \bar{Y})$$
$$X - 32 = -0.234(Y - 38)$$
$$X = -0.234\ Y + 8.892 + 32$$
$$X = -0.234\ Y + 40.892$$

Regression equation of Y on X.

$$Y - \bar{Y} = b_{yx}(X - \bar{X})$$
$$b_{yx} = r\frac{\sigma_y}{\sigma_x} = \frac{\Sigma xy}{\Sigma x^2} = \frac{-93}{140} = -0.664$$
$$Y - 38 = -0.664\ (X - 32)$$
$$Y = -0.664X + 21.248 + 38$$
$$Y = -0.664X + 59.248$$

(b) Coefficient of correlation $r = \pm\sqrt{b_{xy} \times b_{yx}}$

$= \sqrt{(-0.234)(-0.664)}$

= -0.394

(c) When X = 30, Y = ?

Y = -0.664(30) + 59.248

= -19.92 + 59.248

= 39.328.

Question 2.

The heights (in cm.) of a group of fathers and sons are given below.

Heights of fathers:	158	166	163	165	167	170	167	172	177	181
Heights of Sons :	163	158	167	170	160	180	170	175	172	175

Find the lines of regression and estimate the height of son when the height of the father is 164 cm.

Solution:

Heights of fathers (X)	Heights of sons (Y)	$dx = X - 168$	$dy = Y - 169$	dx^2	dy^2	$dxdy$
158	163	–10	– 6	100	36	– 60
166	158	–2	–11	4	121	22
163	167	–5	–2	25	4	10
165	170	–3	1	9	1	–3
167	160	–1	–9	1	81	9
170	180	2	11	4	121	22
167	170	–1	1	1	1	–1
172	175	4	6	16	36	24
177	172	9	3	81	9	27
181	175	13	6	169	36	78
1686	1690	6	0	410	446	248

$N = 10$, $\Sigma X = 1686$, $\Sigma Y = 1690$, $\Sigma dx^2 = 410$, $\Sigma dy^2 = 446$, $\Sigma dxdy = 248$, $\bar{X} = \frac{1686}{10} = 168.6$,

$$\bar{Y} = \frac{1690}{10} = 169$$

$$b_{xy} = \frac{N\Sigma dxdy - (\Sigma dx)(\Sigma dy)}{N\Sigma dy^2 - (\Sigma dy)^2}$$

$$= \frac{10(248) - 6(0)}{(10)(446) - 0^2}$$

$$= \frac{2480}{4460} = 0.556$$

Regression equation of X on Y

$$= \frac{(10)(248)-(6)(0)}{(10)(410)-(6)^2} = \frac{2480}{4100-36}$$

$$= \frac{2480}{4064} = 0.610$$

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

$$X - 168.6 = 0.556\,(Y - 169)$$

$$= 0.556\,Y - 93.964 + 168.6$$

$$= 0.556\,Y + 74.636$$

$$b_{yx} = \frac{N\Sigma xy - (\Sigma dx)(\Sigma dy)}{N\Sigma dx^2 - (\Sigma dx)^2}$$

Regression equation of Y on X

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

Y – 169 = 0.610 (X – 168.6)

Y – 169 = 0.610X – 102.846

Y = 0.610X – 102.846 + 169

Y = 0.160X + 66.154 ....(1)

To get son's height (Y) when the father height is X = 164 cm.

Put X = 164 cm in equation (1) we get

Son's height = 0.610 × 164 + 66.154

= 100.04 + 66.154 cm

= 169.19 cm.

Question 3.
The following data give the height in inches (X) and the weight in lb. (Y) of a random sample of 10 students from a large group of students of age 17 years:

X	61	68	68	64	65	70	63	62	64	67
Y	112	123	130	115	110	125	100	113	116	125

Estimate weight of the student of a height 69 inches.

Solution:

Height (X)	Weight (Y)	$dx = X - 65$	$dy = Y - 117$	dx^2	dy^2	$dxdy$
61	112	−4	−5	16	25	20
68	123	3	6	9	36	18
68	130	3	13	9	169	39
64	115	−1	−2	1	4	2
65	110	0	−7	0	49	0
70	125	5	8	25	64	40
63	100	−2	−17	4	289	34
62	113	−3	−4	9	16	12
64	116	−1	−1	1	1	1
67	125	2	8	4	64	16
652	1169	2	−1	78	717	182

$N = 10, \Sigma X = 652, \Sigma Y = 1169, \Sigma dx = 2, \Sigma dy = -1, \Sigma dx^2 = 78, \Sigma dy^2 = 717, \Sigma dxdy = 182, \bar{X} = \frac{652}{10} = 65.2,$

$$\bar{Y} = \frac{1169}{10} = 116.9$$

$$b_{yx} = \frac{N\Sigma dxdy - (\Sigma dx)(\Sigma dy)}{N\Sigma dx^2 - (\Sigma dx)^2}$$

$$= \frac{10(182) - (2)(-1)}{10(78) - (2)^2} = \frac{1822}{776}$$

$$= 2.3479$$

Regression equation of Y on X

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

Y – 117 = 2.3479 (X – 65.2)
Y – 117 = 2.3479X – (2.3479)(65.2)
Y = 2.3479X – 153.08308 + 117
Y = 2.3479 – 36.08308
When the height X = 69 inches
Weight, Y = 2.3479(69) – 36.08308
= 162.0051 – 36.08308
= 125.92202
= 125.92 lb

Question 4.
Obtain the two regression lines from the following data N = 20, ΣX = 80, ΣY = 40, $ΣX^2$ = 1680, $ΣY^2$ = 320 and ΣXY = 480.

Solution:
ΣX = 80, ΣY = 40, $ΣX^2$ = 1680, $ΣY^2$ = 320, ΣXY = 480, N = 20

$$\bar{X} = \frac{\Sigma X}{N} = \frac{80}{20} = 4,$$

$$\bar{Y} = \frac{\Sigma Y}{N} = \frac{40}{20} = 2$$

$$b_{yx} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{N\Sigma X^2 - (\Sigma X)^2}$$

$$= \frac{20(480) - (80)(40)}{20(1680) - (80)^2} = \frac{9600 - 3200}{33600 - 6400}$$

$$= \frac{6400}{27200} = 0.235 = 0.24$$

Regression line of Y on X

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

$$Y - 2 = 0.24(X - 4)$$

$$Y = 0.24\,X - 0.96 + 2$$

$$Y = 0.24\,X + 1.04$$

$$b_{xy} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{N\Sigma Y^2 - (\Sigma Y)^2}$$

$$= \frac{20(480)-(80)(40)}{20(320)-(40)^2} = \frac{9600-3200}{6400-1600}$$

$$= \frac{6400}{4800} = 1.33$$

Regression line of X on Y

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

X – 4 = 1.33(Y – 2)

X = 1.33Y – 2.66 + 4

X = 1.33Y + 1.34

Question 5.
Given the following data, what will be the possible yield when the rainfall is 29″

Details	Rainfall	Production
Mean	25``	40 units per acre
Standard Deviation	3``	6 units per acre

The coefficient of correlation between rainfall and production is 0.8.

Solution:

$\overline{X} = 25$, $\sigma_x = 3$, $\overline{Y} = 40$, $\sigma_y = 6$, $r = 0.8$

$$b_{yx} = r\frac{\sigma_y}{\sigma_x} = 0.8 \times \frac{6}{3} = 1.6$$

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

$$Y - 40 = 1.6\,(X - 25)$$

$$Y - 40 = 1.6\,X - (1.6)(25)$$

$$Y - 40 = 1.6\,X - 40$$

$$\therefore\ Y = 1.6\,X$$

To find the yield when the rainfall is 29″ is,

Put X = 29 in the above equation we get yield,

Y = 1.6 × 29 = 46.4 units/acre

Question 6.
The following data relate to advertisement expenditure (in lakh of rupees) and their corresponding sales (in crores of rupees)

Advertisement expenditure	40	50	38	60	65	50	35
Sales	38	60	55	70	60	48	30

Estimate the sales corresponding to advertising expenditure of ₹ 30 lakh.

Solution:

X	Y	X^2	Y^2	XY
40	38	1600	1444	1520
50	60	2500	3600	3000
38	55	1444	3025	2090
60	70	3600	4900	4200
65	60	4225	3600	3900
50	48	2500	2304	2400
35	30	1225	900	1050
338	361	17094	19773	18160

$N = 7$, $\Sigma X = 338$, $\Sigma Y = 361$, $\Sigma X^2 = 17094$, $\Sigma Y^2 = 19773$, $\Sigma XY = 18160$.

$$\bar{X} = \frac{\Sigma X}{N} = \frac{338}{7} = 48.29$$

$$\bar{Y} = \frac{\Sigma Y}{N} = \frac{361}{7} = 51.57$$

$$b_{yx} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{N\Sigma X^2 - (\Sigma X)^2}$$

$$= \frac{7(18160) - (338)(361)}{7(17094) - (338)^2}$$

$$= \frac{127120 - 122018}{119658 - 114244}$$

$$= \frac{5102}{5414} = 0.942$$

Regression equation of Y on X

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

Y – 51.57 = 0.942(X – 48.29)

Y – 51.57 = 0.942X – 0.942 × 48.29

Y – 51.57 = 0.942X – 45.48918

Y = 0.942X + 51.57 – 48.29

Y = 0.942X + 6.081

To find the sales, when the advertising is X = ₹ 30 lakh in the above equation we get,
Y = 0.942(30) + 6.081
= 28.26 + 6.081
= 34.341
= ₹ 34.34 crores

Question 7.
You are given the following data:

Details	X	Y
Arithmetic Mean	36	85
Standard Deviation	11	8

If the Correlation coefficient between X and Y is 0.66, then find
(i) the two regression coefficients,
(ii) the most likely value of Y when X = 10.

Solution:

$\bar{X} = 36$, $\bar{Y} = 85$, $\sigma_x = 11$, $\sigma_y = 8$, $r = 0.66$

(i) The two regression coefficients are,

$$b_{yx} = r\frac{\sigma_y}{\sigma_x} = 0.66 \times \frac{8}{11} = 0.48$$

$$b_{xy} = r\frac{\sigma_x}{\sigma_y} = 0.66 \times \frac{11}{8} = 0.9075$$

(ii) Regression equation of X on Y:

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

X – 36 = 0.91(Y – 85)

X – 36 = 0.91Y – 77.35

X = 0.91Y – 77.35 + 36

X = 0.91Y – 41.35

Regression line of Y on X:

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

Y – 85 = 0.48(X – 36)

Y = 0.48X – 17.28 + 85

Y = 0.48X + 67.72

The most likely value of Y when X = 10 is

$Y = 0.48(10) + 67.72 = 72.52$.

Question 8.

Find the equation of the regression line of Y on X, if the observations (X_i, Y_i) are the following (1, 4) (2, 8) (3, 2) (4, 12) (5, 10) (6, 14) (7, 16) (8, 6) (9, 18).

Solution:

X	Y	X^2	Y^2	XY
1	4	1	16	4
2	8	4	64	16

3	2	9	4	6
4	12	16	144	48
5	10	25	100	50
6	14	36	196	84
7	16	49	256	112
8	6	64	36	48
9	18	81	324	162
45	90	285	1140	530

$N = 9, \Sigma X = 45, \Sigma Y = 90, \Sigma X^2 = 285, \Sigma Y^2 = 1140, \Sigma XY = 530, \bar{X} = \frac{\Sigma X}{N} = \frac{45}{9} = 5, \bar{Y} = \frac{\Sigma Y}{N} = \frac{90}{9} = 10$

$$b_{yx} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{N(\Sigma X^2) - (\Sigma X)^2}$$

$$= \frac{9(530) - (45)(90)}{9(285) - (45)^2} = \frac{4770 - 4050}{2565 - 2025}$$

$$= \frac{720}{540} = 1.33$$

Regression line of Y on X:

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

Y – 10 = 1.33(X – 5)

Y = 1.33X – 6.65 + 10

Y = 1.33X + 3.35

Question 9.

A survey was conducted to study the relationship between expenditure on accommodation (X) and expenditure on Food and Entertainment (Y) and the following results were obtained:

Details	Mean	SD
Expenditure on Accommodation (₹)	178	63.15
Expenditure on Food and Entertainment (₹)	47.8	22.98

Coefficient of Correlation	0.43

Write down the regression equation and estimate the expenditure on Food and Entertainment, if the expenditure on accommodation is ₹ 200.

Solution:

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

$$Y - 47.8 = 0.1565(X - 178)$$

$$Y = 0.1565X - 27.857 + 47.8$$

$$Y = 0.1565X + 19.94$$

When the expenditure on accommodation is ₹ 200 the expenditure on food and entertainments is,
$Y = 0.1565X + 19.94$
$Y = 0.1565(200) + 19.94$
$= 31.3 + 19.94$
= ₹ 51.24.

Question 10.
For 5 observations of pairs of (X, Y) of variables X and Y the following results are obtained.
$\Sigma X = 15$, $\Sigma Y = 25$, $\Sigma X^2 = 55$, $\Sigma Y^2 = 135$, $\Sigma XY = 83$. Find the equation of the lines of regression and estimate the values of X and Y if Y = 8; X = 12.

Solution:
$N = 5$, $\Sigma X = 15$, $\Sigma Y = 25$, $\Sigma X^2 = 55$, $\Sigma Y^2 = 135$, $\Sigma XY = 83$,

$$\overline{X} = \frac{15}{5} = 3, \overrightarrow{Y} = \frac{25}{5} = 5.$$

$$b_{yx} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{N(\Sigma X^2) - (\Sigma X)^2} = \frac{5(83) - (15)(25)}{5(55) - (15)^2}$$

$$= \frac{415 - 375}{275 - 225} = \frac{40}{50} = 0.8$$

Regression line of Y on X:

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

Y – 5 = 0.8(X – 3)

Y = 0.8X – 2.4 + 5

Y = 0.8X + 2.6.

WhenX = 12, Y = 0.8X + 2.6

Y = (0.8)12 + 2.6

= 9.6 + 2.6

= 12.2

$$b_{xy} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{N\Sigma Y^2 - (\Sigma Y)^2}$$

$$= \frac{5(83) - (15)(25)}{5(135) - (25)^2} = \frac{415 - 375}{675 - 625} = \frac{40}{50} = 0.8$$

Regression line of X on Y:

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

X – 3 = 0.8(Y – 5)

X = 0.8Y – 4 + 3

X = 0.8Y – 1

When Y = 8, X = 0.8Y – 1

X = (0.8)8 – 1
= 6.4 – 1
= 5.4

Question 11.
The two regression lines were found to be 4X – 5Y + 33 = 0 and 20X – 9Y – 107 = 0. Find the mean values and coefficient of correlation between X and Y.

Solution:
To get mean values we must solve the given lines.
4X – 5Y = -33 ……(1)

$20X - 9Y = 107$ ....(2)
(1) $\times 5 \Rightarrow 20X - 25Y = -165$
$20X - 9Y = 107$
Subtracting (1) and (2), $-16Y = -272$

$Y = \frac{272}{16} = 17$

i.e., $\overline{Y} = 17$

Using $Y = 17$ in (1) we get, $4X - 85 = -33$

$4X = 85 - 33$

$4X = 52$

$X = 13$

i.e., $\overline{X} = 13$

Mean values are $\overline{X} = 13$, $\overline{Y} = 17$,

Let regression line of Y on X be

$4X - 5Y + 33 = 0$

$5Y = 4X + 33$

$Y = \frac{1}{5}(4X + 33)$

$Y = \frac{4}{5}X + \frac{33}{5}$

$Y = 0.8X + 6.6$

$\therefore b_{yx} = 0.8$

Let regression line of X on Y be

$20X - 9Y - 107 = 0$

$20X = 9Y + 107$

$X = \frac{1}{20}(9Y + 107)$

$X = \frac{9}{20}Y + \frac{107}{20}$

$X = 0.45Y + 5.35$

$\therefore b_{xy} = 0.45$

Coefficient of correlation between X and Y is

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

$r = \pm\sqrt{0.8 \times 0.45}$

= ±0.6

= 0.6

Both byx and bxy is positive take positive sign.

Question 12.
The equations of two lines of regression obtained in a correlation analysis are the following 2X = 8 – 3Y and 2Y = 5 – X . Obtain the value of the regression coefficients and correlation coefficient.

Solution:
Let regression line of Y on X be,
2Y = 5 – X
Y = -0.5X + 2.5
b_{yx} = -0.5

i.e., $b_{yx} = -\frac{1}{2}$

Let regression line of X on Y be

2X = 8 – 3Y

X = -1.5Y + 4

b_{xy} = -1.5

i.e., $b_{xy} = -\frac{3}{2}$

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

$= \pm\sqrt{1.5 \times 0.5}$

= -0.866

Both b_{xy} and b_{yx} is negative so take negative sign.

Ex 9.3

Choose the correct answer.

Question 1.
An example of a positive correlation is:
(a) Income and expenditure
(b) Price and demand
(c) Repayment period and EMI
(d) Weight and Income

Answer:
(a) Income and expenditure

Question 2.
If the values of two variables move in the same direction then the correlation is said to be:
(a) Negative
(b) Positive
(c) Perfect positive
(d) No correlation

Answer:
(b) Positive

Question 3.
If the values of two variables move in the opposite direction then the correlation is said to be:
(a) Negative
(b) Positive
(c) Perfect positive
(d) No correlation

Answer:
(a) Negative

Question 4.
Correlation co-efficient lies between:
(a) 0 to ∞

(b) -1 to +1
(c) -1 to 0
(d) -1 to ∞

Answer:
(b) -1 to +1

Question 5.
If r(X, Y) = 0 the variables X and Y are said to be:
(a) Positive correlation
(b) Negative correlation
(c) No correlation
(d) Perfect positive correlation

Answer:
(c) No correlation

Question 6.
The correlation coefficient from the following data N = 25, ΣX = 125, ΣY = 100, $\Sigma X^2 = 650$, $\Sigma Y^2 = 436$, ΣXY = 520:
(a) 0.667
(b) -0.006
(c) -0.667
(d) 0.70

Answer:
(a) 0.667
Hint:

$$r = \frac{N\Sigma XY - \Sigma X \Sigma Y}{\sqrt{N\Sigma X^2 - (\Sigma X)^2}\sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$$

$$= \frac{25(520) - 125 \times 100}{\sqrt{25 \times 650 - (125)^2}\sqrt{25 \times 436 - (100)^2}}$$

$$= \frac{13000 - 12500}{\sqrt{16250 - 15625}\sqrt{10900 - 10000}}$$

$$= \frac{500}{\sqrt{625}\sqrt{900}} = \frac{500}{25 \times 30} = \frac{2}{3}$$
$$= 0.6666$$
$$= 0.667$$

Question 7.
From the following data, N = 11, ΣX = 117, ΣY = 260, $\Sigma X^2 = 1313$, $\Sigma Y^2 = 6580$, ΣXY = 2827. the correlation coefficient is:
(a) 0.3566
(b) -0.3566
(c) 0
(d) 0.4566

Answer:
(a) 0.3566
Hint:

$$r = \frac{N\Sigma XY - \Sigma X\Sigma Y}{\sqrt{N\Sigma X^2 - (\Sigma X)^2}\sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$$
$$= \frac{11 \times 2827 - 117 \times 260}{\sqrt{11 \times 1313 - (117)^2}\sqrt{11 \times 6580 - (260)^2}}$$
$$= \frac{31097 \times 30420}{\sqrt{14443 - 13689}\sqrt{72380 - 67600}}$$
$$= \frac{677}{\sqrt{754}\sqrt{4780}}$$
$$= \frac{677}{\sqrt{3604120}} = \frac{677}{1898.45} = 0.3566$$

Question 8.
The correlation coefficient is:

(a) $r(X,Y) = \frac{\sigma_x \sigma_y}{\operatorname{cov}(x,y)}$

(b) $r(X,Y)= \frac{\text{cov}(x,y)}{\sigma_x \sigma_y}$

(c) $r(X,Y)= \frac{\text{cov}(x,y)}{\sigma_y}$

(d) $r(X,Y)= \frac{\text{cov}(x,y)}{\sigma_x}$

Answer:

(b) $r(X, Y) = \frac{cov(x,y)}{\sigma_x \sigma_y}$

Question 9.
The variable whose value is influenced or is to be predicted is called:
(a) dependent variable
(b) independent variable
(c) regressor
(d) explanatory variable

Answer:
(a) dependent variable

Question 10.
The variable which influences the values or is used for prediction is called:
(a) Dependent variable
(b) Independent variable
(c) Explained variable
(d) Regressed

Answer:
(b) Independent variable

Question 11.
The correlation coefficient:

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

(b) $r=\frac{1}{b_{xy} \times b_{yx}}$

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

(b) $r = \frac{1}{b_{xy} \times b_{yx}}$

Answer:

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

Question 12.
The regression coefficient of X on Y:

(a) $b_{xy} = \frac{N\Sigma dx\,dy - (\Sigma dx)(\Sigma dy)}{N\,\Sigma dy^2 - (\Sigma dy)^2}$

(b) $b_{yx} = \frac{N\Sigma\,dx\,dy - (\Sigma dx)(\Sigma dy)}{N\,\Sigma dy^2 - (\Sigma dy)^2}$

(c) $b_{xy} = \frac{N\Sigma dxdy - (\Sigma dx)(\Sigma dy)}{N\,\Sigma dx^2 - (\Sigma dx)^2}$

(d) $b_{y} = \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{N\Sigma x^2 - (\Sigma x)^2} \times \sqrt{N\Sigma y^2 - (\Sigma y)^2}}$

Answer:

(a) $b_{xy} = \frac{\mathrm{N}\Sigma dxdy - (\Sigma dx)(\Sigma dy)}{\mathrm{N}\Sigma dy^2 - (\Sigma dy)^2}$

Question 13.
The regression coefficient of Y on X:

(a) $b_{xy} = \frac{N\Sigma dx\,dy - (\Sigma dx)(\Sigma dy)}{N\,\Sigma dy^2 - (\Sigma dy)^2}$

(b) $b_{yx} = \frac{N\Sigma dxdy - (\Sigma dx)(\Sigma dy)}{N\Sigma dy^2 - (\Sigma dy)^2}$

(c) $b_{yx} = \dfrac{N\Sigma dx\,dy - (\Sigma dx)(\Sigma dy)}{N\Sigma dx^2 - (\Sigma dx)^2}$

(d) $b_{xy} = \dfrac{N\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{N\Sigma x^2 - (\Sigma x)^2} \times \sqrt{N\Sigma y^2 - (\Sigma y)^2}}$

Answer:

(c) $b_{yx} = \frac{N\Sigma dxdy - (\Sigma dx)(\Sigma dy)}{N\Sigma dx^2 - (\Sigma dx)^2}$

Question 14.
When one regression coefficient is negative, the other would be:
(a) Negative
(b) Positive
(c) Zero
(d) None of these

Answer:
(a) Negative

Question 15.
If X and Y are two variates, there can be at most:
(a) one regression line
(b) two regression lines
(c) three regression lines
(d) more regression lines

Answer:
(b) two regression lines

Question 16.
The lines of regression of X on Y estimates:
(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

Answer:
(a) X for a given value of Y

Question 17.
Scatter diagram of the variate values (X, Y) give the idea about:
(a) functional relationship
(b) regression model
(c) distribution of errors
(d) no relation

Answer:
(a) functional relationship

Question 18.
If regression co-efficient of Y on X is 2, then the regression co-efficient of X on Y is:
(a) $\leq \frac{1}{2}$
(b) 2
(c) $> \frac{1}{2}$
(d) 1

Answer:
(a) $\leq \frac{1}{2}$

Question 19.
If two variables move in a decreasing direction then the correlation is:
(a) positive
(b) negative
(c) perfect negative
(d) no correlation

Answer:
(a) positive

Question 20.
The person suggested a mathematical method for measuring the magnitude of the linear relationship between two variables say X and Y is:
(a) Karl Pearson
(b) Spearman
(c) Croxton and Cowden

(d) Ya Lun Chou

Answer:
(a) Karl Pearson

Question 21.
The lines of regression intersect at the point:
(a) (X, Y)
(b) $(\bar{X}, \bar{Y})$
(c) (0, 0)
(d) (σ_x, σ_y)

Answer:
(b) $(\bar{X}, \bar{Y})$

Question 22.
The term regression was introduced by:
(a) R.A Fisher
(b) Sir Francis Galton
(c) Karl Pearson
(d) Croxton and Cowden

Answer:
(b) Sir Francis Galton

Question 23.
If r = -1, then correlation between the variables:
(a) perfect positive
(b) perfect negative
(c) negative
(d) no correlation

Answer:
(b) perfect negative

Question 24.
The coefficient of correlation describes:

(a) the magnitude and direction
(b) only magnitude
(c) only direction
(d) no magnitude and no direction

Answer:
(a) the magnitude and direction

Question 25.
If Cov(x, y) = -16.5, σ_x^2 = 2.89, σ_y^2 = 100. Find correlation coeffient.
(a) -0.12
(b) 0.001
(c) -1
(d) -0.97

Answer:
(d) -0.97