Correlation& Regression

Type I: Karl Pearson's coefficient of correlation (r)

1. Find the coefficient of correlation between X and Y

x	14	8	10	11	9	13	5
y	14	9	11	13	11	12	4

[N13/Biot/6M]

Solution:

x	y	$d_x = x - A$	$d_y = y - B$	d_x^2	d_y^2	$d_x d_y$
14	14	4	3	16	9	12
8	9	-2	-2	4	4	4
10	11	0	0	0	0	0
11	13	1	2	1	4	2
9	11	-1	0	1	0	0
13	12	3	1	9	1	3
5	4	-5	-7	25	49	35
70	74	0	-3	56	67	56

$$\overline{x} = \frac{\sum x}{N} = \frac{70}{7} = 10 \ (A = 10)$$

$$\overline{y} = \frac{\sum y}{N} = \frac{74}{7} = 10.57 \ (B = 11)$$

$$cov(x, y) = \frac{\sum d_x d_y}{N} - \left(\frac{\sum d_x}{N}\right)\left(\frac{\sum d_y}{N}\right) = \frac{56}{7} - \left(\frac{0}{7}\right)\left(-\frac{3}{7}\right) = 8$$

$$\sigma_x = \sqrt{\frac{\sum d_x^2}{N} - \left(\frac{\sum d_x}{N}\right)^2} = \sqrt{\frac{56}{7} - \left(\frac{0}{7}\right)^2} = 2.8284$$

$$\sigma_y = \sqrt{\frac{\sum d_y^2}{N} - \left(\frac{\sum d_y}{N}\right)^2} = \sqrt{\frac{67}{7} - \left(\frac{-3}{7}\right)^2} = 3.0639$$

$$r = \frac{cov\,(x,y)}{\sigma_x \sigma_y} = \frac{8}{2.8284 \times 3.0639} = 0.9232$$

2. Find the coefficient of correlation between X and Y

X	62	64	65	69	70	71	72	74
Y	126	125	139	145	165	152	180	208

[M14/ChemBiot/6M][N16/ChemBiot/6M]

Solution:

x	y	$d_x = x - A$	$d_y = y - B$	d_x^2	d_y^2	$d_x d_y$
62	126	-8	-29	64	841	232
64	125	-6	-30	36	900	180
65	139	-5	-16	25	256	80
69	145	-1	-10	1	100	10
70	165	0	10	0	100	0
71	152	1	-3	1	9	-3
72	180	2	25	4	625	50
74	208	4	53	16	2809	212
547	1240	-13	0	147	5640	761

$$\overline{x} = \frac{\sum x}{N} = \frac{547}{8} = 68.375\ (A = 70)$$

$$\overline{y} = \frac{\sum y}{N} = \frac{1240}{8} = 155\ (B = 155)$$

$$cov(x, y) = \frac{\sum d_x d_y}{N} - \left(\frac{\sum d_x}{N}\right)\left(\frac{\sum d_y}{N}\right) = \frac{761}{8} - \left(\frac{-13}{8}\right)\left(\frac{0}{8}\right) = 95.125$$

$$\sigma_x = \sqrt{\frac{\sum d_x^2}{N} - \left(\frac{\sum d_x}{N}\right)^2} = \sqrt{\frac{147}{8} - \left(-\frac{13}{8}\right)^2} = 3.9667$$

$$\sigma_y = \sqrt{\frac{\sum d_y^2}{N} - \left(\frac{\sum d_y}{N}\right)^2} = \sqrt{\frac{5640}{8} - \left(\frac{0}{8}\right)^2} = 26.5518$$

$$r = \frac{cov\,(x,y)}{\sigma_x \sigma_y} = \frac{95.125}{3.9667 \times 26.5518} = 0.9032$$

3. Calculate the correlation coefficient from the following data:
x: 23,27,28,29,30,31,33,35,36,39
y: 18,22,23,24,25,26,28,29,30,32
[N14/CompIT/6M][N14/ChemBiot/6M]
Solution:

x	y	$d_x = x - A$	$d_y = y - B$	d_x^2	d_y^2	$d_x d_y$
23	18	-7	-7	49	49	49
27	22	-3	-3	9	9	9
28	23	-2	-2	4	4	4
29	24	-1	-1	1	1	1
30	25	0	0	0	0	0
31	26	1	1	1	1	1
33	28	3	3	9	9	9
35	29	5	4	25	16	20
36	30	6	5	36	25	30
39	32	9	7	81	49	63
311	257	11	7	215	163	186

$$\overline{x} = \frac{\sum x}{N} = \frac{311}{10} = 31.1\ (A = 30)$$

$$\overline{y} = \frac{\sum y}{N} = \frac{257}{10} = 25.7\ (B = 25)$$

$$cov(x,y) = \frac{\sum d_x d_y}{N} - \left(\frac{\sum d_x}{N}\right)\left(\frac{\sum d_y}{N}\right) = \frac{186}{10} - \left(\frac{11}{10}\right)\left(\frac{7}{10}\right) = 17.83$$

$$\sigma_x = \sqrt{\frac{\sum d_x^2}{N} - \left(\frac{\sum d_x}{N}\right)^2} = \sqrt{\frac{215}{10} - \left(\frac{11}{10}\right)^2} = 4.5044$$

$$\sigma_y = \sqrt{\frac{\sum d_y^2}{N} - \left(\frac{\sum d_y}{N}\right)^2} = \sqrt{\frac{163}{10} - \left(\frac{7}{10}\right)^2} = 3.9762$$

$$r = \frac{cov\,(x,y)}{\sigma_x \sigma_y} = \frac{17.83}{4.5044 \times 3.9762} = 0.9955$$

4. Find coefficient of correlation for the following data

X	100	200	300	400	500
Y	30	40	50	60	70

[M15/CompIT/6M][N16/CompIT/5M]

Solution:

x	y	$x-\overline{x}$	$y-\overline{y}$	$(x-\overline{x})^2$	$(y-\overline{y})^2$	$(x-\overline{x})(y-\overline{y})$
100	30	-200	-20	40000	400	4000
200	40	-100	-10	10000	100	1000
300	50	0	0	0	0	0
400	60	100	10	10000	100	1000
500	70	200	20	40000	400	4000
1500	250	0	0	100000	1000	10000

$$\overline{x}=\frac{\sum x}{N}=\frac{1500}{5}=300$$

$$\overline{y}=\frac{\sum y}{N}=\frac{250}{5}=50$$

$$cov(x,y)=\frac{\sum(x-\overline{x})(y-\overline{y})}{N}=\frac{10000}{5}=2000$$

$$\sigma_x=\sqrt{\frac{\sum(x-\overline{x})^2}{N}}=\sqrt{\frac{100000}{5}}=100\sqrt{2}$$

$$\sigma_y=\sqrt{\frac{\sum(y-\overline{y})^2}{N}}=\sqrt{\frac{1000}{5}}=10\sqrt{2}$$

$$r=\frac{cov\,(x,y)}{\sigma_x\sigma_y}=\frac{2000}{100\sqrt{2}\times10\sqrt{2}}=1$$

5. Find Karl Pearson's coefficient of correlation for the following data

X	28	45	40	38	35	33	40	32	36	33
Y	23	34	33	34	30	26	28	31	36	35

[M15/AutoMechCivil/6M]

Solution:

x	y	$x-\overline{x}$	$y-\overline{y}$	$(x-\overline{x})^2$	$(y-\overline{y})^2$	$(x-\overline{x})(y-\overline{y})$
28	23	-8	-8	64	64	64
45	34	9	3	81	9	27
40	33	4	2	16	4	8
38	34	2	3	4	9	6
35	30	-1	-1	1	1	1
33	26	-3	-5	9	25	15
40	28	4	-3	16	9	-12
32	31	-4	0	16	0	0
36	36	0	5	0	25	0
33	35	-3	4	9	16	-12
360	310	0	0	216	162	97

$$\overline{x}=\frac{\sum x}{N}=\frac{360}{10}=36$$

$$\overline{y}=\frac{\sum y}{N}=\frac{310}{10}=31$$

$$cov(x,y)=\frac{\sum(x-\overline{x})(y-\overline{y})}{N}=\frac{97}{10}=9.7$$

$$\sigma_x=\sqrt{\frac{\sum(x-\overline{x})^2}{N}}=\sqrt{\frac{216}{10}}=4.6476$$

$$\sigma_y=\sqrt{\frac{\sum(y-\overline{y})^2}{N}}=\sqrt{\frac{162}{10}}=4.0249$$

$$r=\frac{cov\,(x,y)}{\sigma_x\sigma_y}=\frac{9.7}{4.6476\times 4.0249}=0.5185$$

6. The following marks have been obtained by a class of students in statistics

Paper I	80	45	55	56	58	60	65	68	70	75	85
Paper II	81	56	50	48	60	62	64	65	70	74	90

Compute the coefficient of correlation for the above data

[M15/ChemBiot/8M]

Solution:

x	y	$d_x = x - A$	$d_y = y - B$	d_x^2	d_y^2	$d_x d_y$
80	81	15	16	225	256	240
45	56	-20	-9	400	81	180
55	50	-10	-15	100	225	150
56	48	-9	-17	81	289	153
58	60	-7	-5	49	25	35
60	62	-5	-3	25	9	15
65	64	0	-1	0	1	0
68	65	3	0	9	0	0
70	70	5	5	25	25	25
75	74	10	9	100	81	90
85	90	20	25	400	625	500
717	720	2	5	1414	1617	1388

$$\overline{x} = \frac{\sum x}{N} = \frac{717}{11} = 65.1818\ (A = 65)$$

$$\overline{y} = \frac{\sum y}{N} = \frac{720}{11} = 65.4545\ (B = 65)$$

$$cov(x, y) = \frac{\sum d_x d_y}{N} - \left(\frac{\sum d_x}{N}\right)\left(\frac{\sum d_y}{N}\right) = \frac{1388}{11} - \left(\frac{2}{11}\right)\left(\frac{5}{11}\right) = 126.0992$$

$$\sigma_x = \sqrt{\frac{\sum d_x^2}{N} - \left(\frac{\sum d_x}{N}\right)^2} = \sqrt{\frac{1414}{11} - \left(\frac{2}{11}\right)^2} = 11.3363$$

$$\sigma_y = \sqrt{\frac{\sum d_y^2}{N} - \left(\frac{\sum d_y}{N}\right)^2} = \sqrt{\frac{1617}{11} - \left(\frac{5}{11}\right)^2} = 12.1158$$

$$r = \frac{cov\,(x,y)}{\sigma_x \sigma_y} = \frac{126.0992}{11.3363 \times 12.1158} = 0.9181$$

7. Calculate coefficient of correlation from the following data

x	30	33	25	10	33	75	40	85	90	95	65	55
y	68	65	80	85	70	30	55	18	15	10	35	45

[N15/CompIT/6M]

Solution:

x	y	$x-\overline{x}$	$y-\overline{y}$	$(x-\overline{x})^2$	$(y-\overline{y})^2$	$(x-\overline{x})(y-\overline{y})$
30	68	-23	20	529	400	-460
33	65	-20	17	400	289	-340
25	80	-28	32	784	1024	-896
10	85	-43	37	1849	1369	-1591
33	70	-20	22	400	484	-440
75	30	22	-18	484	324	-396
40	55	-13	7	169	49	-91
85	18	32	-30	1024	900	-960
90	15	37	-33	1369	1089	-1221
95	10	42	-38	1764	1444	-1596
65	35	12	-13	144	169	-156
55	45	2	-3	4	9	-6
636	576	0	0	8920	7550	-8152

$$\overline{x} = \frac{\sum x}{N} = \frac{636}{12} = 53$$

$$\overline{y} = \frac{\sum y}{N} = \frac{576}{12} = 48$$

$$cov(x,y) = \frac{\sum(x-\overline{x})(y-\overline{y})}{N} = \frac{-8152}{12} = -679.3333$$

$$\sigma_x = \sqrt{\frac{\sum(x-\overline{x})^2}{N}} = \sqrt{\frac{8920}{12}} = 27.2641$$

$$\sigma_y = \sqrt{\frac{\sum(y-\overline{y})^2}{N}} = \sqrt{\frac{7550}{12}} = 25.0832$$

$$r = \frac{cov\,(x,y)}{\sigma_x \sigma_y} = \frac{-679.3333}{27.2641 \times 25.0832} = -0.9934$$

8. Soil temperature (x) and germination interval (y) for winter wheat in 12 places are as follows: Find r.

$x\ (in\ °F)$	57	42	38	42	45	42	44	40	46	44	43	40
$y\ (days)$	10	26	41	29	27	27	19	18	19	31	29	33

[N15/AutoMechCivil/6M]

Solution:

x	y	$d_x = x - A$	$d_y = y - B$	d_x^2	d_y^2	$d_x d_y$
57	10	13	-16	169	256	-208
42	26	-2	0	4	0	0
38	41	-6	15	36	225	-90
42	29	-2	3	4	9	-6
45	27	1	1	1	1	1
42	27	-2	1	4	1	-2
44	19	0	-7	0	49	0
40	18	-4	-8	16	64	32
46	19	2	-7	4	49	-14
44	31	0	5	0	25	0
43	29	-1	3	1	9	-3
40	33	-4	7	16	49	-28
523	309	-5	-3	255	737	-318

$$\overline{x} = \frac{\sum x}{N} = \frac{523}{12} = 43.5833\ (A = 44)$$

$$\overline{y} = \frac{\sum y}{N} = \frac{309}{12} = 25.75\ (B = 26)$$

$$cov(x, y) = \frac{\sum d_x d_y}{N} - \left(\frac{\sum d_x}{N}\right)\left(\frac{\sum d_y}{N}\right) = \frac{-318}{12} - \left(\frac{-5}{12}\right)\left(\frac{-3}{12}\right) = -26.6042$$

$$\sigma_x = \sqrt{\frac{\sum d_x^2}{N} - \left(\frac{\sum d_x}{N}\right)^2} = \sqrt{\frac{255}{12} - \left(\frac{-5}{12}\right)^2} = 4.5909$$

$$\sigma_y = \sqrt{\frac{\sum d_y^2}{N} - \left(\frac{\sum d_y}{N}\right)^2} = \sqrt{\frac{737}{12} - \left(\frac{-3}{12}\right)^2} = 7.8329$$

$$r = \frac{cov\,(x,y)}{\sigma_x \sigma_y} = \frac{-26.6042}{4.5909 \times 7.8329} = -0.7398$$

9. Calculate coefficient of correlation between X and Y from the following

X	8	8	7	5	6	2
Y	3	4	10	13	22	8

[N17/Comp/6M]

Solution:

x	y	x^2	y^2	xy
8	3	64	9	24
8	4	64	16	32
7	10	49	100	70
5	13	25	169	65
6	22	36	484	132
2	8	4	64	16
36	60	242	842	339

$$\overline{x} = \frac{\sum x}{N} = \frac{36}{6} = 6$$

$$\overline{y} = \frac{\sum y}{N} = \frac{60}{6} = 10$$

$$cov(x, y) = \frac{\sum xy}{N} - \overline{x}.\overline{y} = \frac{339}{6} - (6)(10) = -3.5$$

$$\sigma_x = \sqrt{\frac{\sum x^2}{N} - (\overline{x})^2} = \sqrt{\frac{242}{6} - (6)^2} = 2.0817$$

$$\sigma_y = \sqrt{\frac{\sum y^2}{N} - (\overline{y})^2} = \sqrt{\frac{842}{6} - (10)^2} = 6.3509$$

$$r = \frac{cov\,(x,y)}{\sigma_x \sigma_y} = \frac{-3.5}{2.0817 \times 6.3509} = -0.2647$$

Type II: Spearman's Rank Correlation (R)

1. What is rank correlation? Obtain the rank correlation coefficient for the following data:

   X: 10 12 18 18 15 40

   Y: 12 18 25 25 50 25

   [N13/Chem/7M][M14/AutoMechCivil/6M]

   Solution:

   Rank correlation is the method of finding out relation between two variables by means of ranking.

X	Y	R_1	R_2	$d_i = R_1 - R_2$	d_i^2
10	12	6	6	0	0
12	18	5	5	0	0
18	25	2.5	3	-0.5	0.25
18	25	2.5	3	-0.5	0.25
15	50	4	1	3	9
40	25	1	3	-2	4
Total					13.50

Here, $m_1 = 2, m_2 = 3$

$$R = 1 - \frac{6\left[\sum d_i^2 + \frac{(m_1^3 - m_1)}{12} + \frac{(m_2^3 - m_2)}{12}\right]}{N^3 - N}$$

$$R = 1 - \frac{6\left[13.5 + \frac{(2^3 - 2)}{12} + \frac{(3^3 - 3)}{12}\right]}{6^3 - 6}$$

$$R = 0.5429$$

2. Calculate the value of rank correlation coefficient form the following data regarding the marks of 6 students in statistics and accountancy in a test

Marks in Statistics	40	42	45	35	36	39
Marks in Accountancy	46	43	44	39	40	43

[N14/AutoMechCivil/6M]

Solution:

X	Y	R_1	R_2	$d_i = R_1 - R_2$	d_i^2
40	46	3	1	2	4
42	43	2	3.5	-1.5	2.25
45	44	1	2	-1	1
35	39	6	6	0	0
36	40	5	5	0	0
39	43	4	3.5	0.5	0.25
Total					6.50

Here, $m_1 = 2$

$$R = 1 - \frac{6\left[\sum d_i^2 + \frac{\left(m_1^3 - m_1\right)}{12}\right]}{N^3 - N}$$

$$R = 1 - \frac{6\left[6.5 + \frac{(2^3 - 2)}{12}\right]}{6^3 - 6}$$

$$R = 0.8$$

3. Calculate Spearman's Rank Correlation from the following :

X:	36	56	20	42	33	44	50	15	60
Y:	50	35	70	58	75	60	45	80	38

[N15/ChemBiot/6M][M16/ChemBiot/6M]

Solution:

X	Y	R_1	R_2	$d_i = R_1 - R_2$	d_i^2
36	50	6	6	0	0
56	35	2	9	-7	49
20	70	8	3	5	25
42	58	5	5	0	0
33	75	7	2	5	25
44	60	4	4	0	0
50	45	3	7	-4	16
15	80	9	1	8	64
60	38	1	8	-7	49
Total					228

$$R = 1 - \frac{6\sum d_i^2}{N^3 - N}$$

$$R = 1 - \frac{6 \times 228}{9^3 - 9}$$

$$R = -0.9$$

4. From the following data calculate the coefficient of rank correlation between X and Y

X	32	55	49	60	43	37	43	49	10	20
Y	40	30	70	20	30	50	72	60	45	25

[M16/AutoMechCivil/6M]

Solution:

X	Y	R_1	R_2	$d_i = R_1 - R_2$	d_i^2
32	40	8	6	2	4
55	30	2	7.5	-5.5	30.25
49	70	3.5	2	-1.5	2.25
60	20	1	10	-9	81
43	30	5.5	7.5	-2	4
37	50	7	4	3	9
43	72	5.5	1	4.5	20.25
49	60	3.5	3	0.5	0.25
10	45	10	5	5	25
20	25	9	9	0	0
Total					176

Here, $m_1 = 2, m_2 = 2, m_3 = 2$

$$R = 1 - \frac{6\left[\sum d_i^2 + \frac{(m_1^3 - m_1)}{12} + \frac{(m_2^3 - m_2)}{12} + \frac{(m_3^3 - m_3)}{12}\right]}{N^3 - N}$$

$$R = 1 - \frac{6\left[176 + \frac{(2^3 - 2)}{12} + \frac{(2^3 - 2)}{12} + \frac{(2^3 - 2)}{12}\right]}{10^3 - 10}$$

$$R = -0.0758$$

5. Compute Spearman's Rank correlation coefficient from the following data:

X	18	20	34	52	12
Y	39	23	35	18	46

[M16/CompIT/5M]

Solution:

X	Y	R_1	R_2	$d_i = R_1 - R_2$	d_i^2
18	39	4	2	2	4
20	23	3	4	-1	1
34	35	2	3	-1	1
52	18	1	5	-4	16
12	46	5	1	4	16
Total					38

$R = 1 - \frac{6\sum d_i^2}{N^3 - N}$

$R = 1 - \frac{6\times 38}{5^3 - 5}$

$R = -0.9$

6. Calculate Spearman's coefficient of rank correlation from the data on height and weight of 8 students

X	60	62	64	66	68	70	72	74
Y	92	83	101	110	128	119	137	146

[N16/AutoMechCivil/6M] [M17/AutoMechCivil/6M]

Solution:

X	Y	R_1	R_2	$d_i = R_1 - R_2$	d_i^2
60	92	8	7	1	1
62	83	7	8	-1	1
64	101	6	6	0	0
66	110	5	5	0	0
68	128	4	3	1	1
70	119	3	4	-1	1
72	137	2	2	0	0
74	146	1	1	0	0
Total					4

$$R = 1 - \frac{6\sum d_i^2}{N^3 - N}$$

$$R = 1 - \frac{6 \times 4}{8^3 - 8}$$

$$R = 0.9524$$

7. Calculate R and r from the following data:

X	12	17	22	27	32
Y	113	119	117	115	121

[M17/CompIT/6M]

Solution:

X	Y	R_1	R_2	$d_i = R_1 - R_2$	d_i^2
12	113	5	5	0	0
17	119	4	2	2	4
22	117	3	3	0	0
27	115	2	4	-2	4
32	121	1	1	0	0
Total					8

$R = 1 - \frac{6\sum d_i^2}{N^3 - N}$

$R = 1 - \frac{6\times 8}{5^3 - 5}$

$R = 0.6$

x	y	$x-\overline{x}$	$y-\overline{y}$	$(x-\overline{x})^2$	$(y-\overline{y})^2$	$(x-\overline{x})(y-\overline{y})$
12	113	-10	-4	100	16	40
17	119	-5	2	25	4	-10
22	117	0	0	0	0	0
27	115	5	-2	25	4	-10
32	121	10	4	100	16	40
110	585	0	0	250	40	60

$\overline{x} = \frac{\sum x}{N} = \frac{110}{5} = 22$

$\overline{y} = \frac{\sum y}{N} = \frac{585}{5} = 117$

$cov(x, y) = \frac{\sum(x-\overline{x})(y-\overline{y})}{N} = \frac{60}{5} = 12$

$\sigma_x = \sqrt{\frac{\sum(x-\overline{x})^2}{N}} = \sqrt{\frac{250}{5}} = 5\sqrt{2}$

$\sigma_y = \sqrt{\frac{\sum(y-\overline{y})^2}{N}} = \sqrt{\frac{40}{5}} = 2\sqrt{2}$

$r = \frac{cov\,(x,y)}{\sigma_x \sigma_y} = \frac{12}{5\sqrt{2}\times 2\sqrt{2}} = 0.6$

Type III: Regression

1. Obtain the line of regression of cost on age from the following table giving the age of a car of certain make and the annual maintenance cost. Also find the maintenance cost if age of car is 9 years.

Age of car (in years) : x	2	4	6	8
Maintenance cost (in thousands) : y	5	7	8.5	11

[M14/CompIT/6M]

Solution:

x	y	x^2	y^2	xy
2	5	4	25	10
4	7	16	49	28
6	8.5	36	72.25	51
8	11	64	121	88
20	31.5	120	267.25	177

$$\overline{x} = \frac{\sum x}{N} = \frac{20}{4} = 5$$

$$\overline{y} = \frac{\sum y}{N} = \frac{31.5}{4} = 7.875$$

$$b_{yx} = \frac{\sum xy - \frac{\sum x . \sum y}{N}}{\sum x^2 - \frac{(\sum x)^2}{N}} = \frac{177 - \frac{20 \times 31.5}{4}}{120 - \frac{(20)^2}{4}} = 0.975$$

Equation of y on x,

$(y - \overline{y}) = b_{yx}(x - \overline{x})$

$(y - 7.875) = 0.975\ (x - 5)$

$y = 0.975x - 4.875 + 7.875$

$y = 0.975x + 3$

If $x = 9$,

Then, $y = 11.775$

2. Obtain two lines of regression and coefficient of correlation from the following data:

X	65	66	67	67	68	69	70	72
Y	67	68	65	66	72	72	69	71

[M14/AutoMechCivil/8M][N14/AutoMechCivil/8M]

Solution:

x	y	$d_x = x - A$	$d_y = y - B$	d_x^2	d_y^2	$d_x d_y$
65	67	-3	-1	9	1	3
66	68	-2	0	4	0	0
67	65	-1	-3	1	9	3
67	66	-1	-2	1	4	2
68	72	0	4	0	16	0
69	72	1	4	1	16	4
70	69	2	1	4	1	2
72	71	4	3	16	9	12
544	550	0	6	36	56	26

$$\overline{x} = \frac{\sum x}{N} = \frac{544}{8} = 68 \ (A = 68)$$

$$\overline{y} = \frac{\sum y}{N} = \frac{550}{8} = 68.75 \ (B = 68)$$

$$b_{yx} = \frac{\sum d_x d_y - \frac{\sum d_x \sum d_y}{N}}{\sum d_x^2 - \frac{(\sum d_x)^2}{N}} = \frac{26 - \frac{0\times 6}{8}}{36 - \frac{(0)^2}{8}} = 0.7222$$

$$b_{xy} = \frac{\sum d_x d_y - \frac{\sum d_x \sum d_y}{N}}{\sum d_y^2 - \frac{\left(\sum d_y\right)^2}{N}} = \frac{26 - \frac{0\times 6}{8}}{56 - \frac{(6)^2}{8}} = 0.5049$$

Equation of y on x,

$$(y - \overline{y}) = b_{yx}(x - \overline{x})$$
$$y - 68.75 = 0.7222(x - 68)$$
$$y = 0.7222x - 49.1096 + 68.75$$
$$y = 0.7222x + 19.6404$$

Equation of x on y,

$$(x - \overline{x}) = b_{xy}(y - \overline{y})$$
$$x - 68 = 0.5049(y - 68.75)$$
$$x = 0.5049y - 34.7119 + 68$$
$$x = 0.5049y + 33.2881$$

Coefficient of correlation,

$$r = \sqrt{b_{yx} \times b_{xy}} = \sqrt{0.7222 \times 0.5049}$$
$$r = 0.6039$$

3. Find the two equations of lines of regression from the following data

x:	1	2	3	4	5	6	7
y:	5	9	8	10	11	9	11

Also estimate the value of y for $x = 8$

[M14/ChemBiot/6M]

Solution:

x	y	x^2	y^2	xy
1	5	1	25	5
2	9	4	81	18
3	8	9	64	24
4	10	16	100	40
5	11	25	121	55
6	9	36	81	54
7	11	49	121	77
28	63	140	593	273

$$\overline{x} = \frac{\sum x}{N} = \frac{28}{7} = 4$$

$$\overline{y} = \frac{\sum y}{N} = \frac{63}{7} = 9$$

$$b_{yx} = \frac{\sum xy - \frac{\sum x.\sum y}{N}}{\sum x^2 - \frac{(\sum x)^2}{N}} = \frac{273 - \frac{28\times63}{7}}{140 - \frac{(28)^2}{7}} = 0.75$$

$$b_{xy} = \frac{\sum xy - \frac{\sum x.\sum y}{N}}{\sum y^2 - \frac{(\sum y)^2}{N}} = \frac{273 - \frac{28\times63}{7}}{593 - \frac{(63)^2}{7}} = 0.8077$$

Equation of y on x,

$(y - \overline{y}) = b_{yx}(x - \overline{x})$

$y - 9 = 0.75(x - 4)$

$y = 0.75x - 3 + 9$

$y = 0.75x + 6$

If $x = 8$,

Then, $y = 12$

Equation of x on y,

$(x - \overline{x}) = b_{xy}(y - \overline{y})$

$x - 4 = 0.8077(y - 9)$

$x = 0.8077y - 7.2693 + 4$

$x = 0.8077y - 3.2693$

4. State true or false with justification: if two lines of regression are $x + 3y - 5 = 0$ and $4x + 3y - 8 = 0$ then the correlation coefficient is $+0.5$.
[M14/CompIT/5M]
Solution:
Let the equation of y on x be $x + 3y - 5 = 0$
i.e. $3y = -x + 5$
$y = -\frac{1}{3}x + \frac{5}{3}$
$\therefore b_{yx} = -\frac{1}{3}$
Let the equation of x on y be $4x + 3y - 8 = 0$
i.e. $4x = -3y + 8$
$x = -\frac{3}{4}y + \frac{8}{4}$
$\therefore b_{xy} = -\frac{3}{4}$
Correlation coefficient,
$r = \sqrt{b_{yx} \times b_{xy}}$
$r = \sqrt{-\frac{1}{3} \times -\frac{3}{4}}$
$r = -0.5$
Thus, the given statement is false. If the values of b_{yx} and b_{xy} are negative then r is also negative.

5. The regression lines of a sample are $x + 6y = 6$ and $3x + 2y = 10$. Find (i) sample means (ii) coefficient of correlation between x and y. **[N15/CompIT/4M]**. Also estimate y when x = 12. **[N14/CompIT/5M][N16/CompIT/6M]**

Solution:

We have,

$x + 6y = 6$ ………….(1)

$3x + 2y = 10$ …………(2)

Solving the two equations, we get

$\overline{x} = 3, \overline{y} = \frac{1}{2}$

Let the equation of y on x be $x + 6y = 6$

i.e. $6y = -x + 6$

$y = -\frac{1}{6}x + 1$

$\therefore b_{yx} = -\frac{1}{6}$

Let the equation of x on y be $3x + 2y = 10$

i.e. $3x = -2y + 10$

$x = -\frac{2}{3}y + \frac{10}{3}$

$\therefore b_{xy} = -\frac{2}{3}$

Correlation coefficient,

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

$r = \sqrt{-\frac{1}{6} \times -\frac{2}{3}}$

$r = -\frac{1}{3}$

Putting $x = 12$ in eqn (1), we get

$12 + 6y = 6$

$y = -1$

6. If the tangent of the angle made by the line of regression of y on x is 0.6 and $\sigma_y = 2\sigma_x$. Find the coefficient of correlation between x & y.
[M15/AutoMechCivil/5M]
Solution:
$tan\theta$ of line y on x is 0.6 i.e slope of line y on x is 0.6
$\therefore b_{yx} = 0.6$
We have,
$b_{yx} = r\frac{\sigma_y}{\sigma_x}$
$0.6 = r\frac{(2\sigma_x)}{\sigma_x}$ since $\sigma_y = 2\sigma_x$ (given)
$0.6 = 2r$
$\therefore r = 0.3$

7. It is given that the means of x and y are 5 and 10. If the line of regression of y on x is parallel to the line $20y = 9x + 40$, estimate the value of y for $x = 30$.
[M15/CompIT/5M][N15/AutoMechCivil/5M]
Solution:
Given that,
$\overline{x} = 5, \overline{y} = 10$
It is also given that the line of regression of y on x is parallel to the line $20y = 9x + 40$. Thus, the slope of the two lines are same
Slope of line y on x is b_{yx}
$\therefore b_{yx} = \frac{9}{20}$
Equation of y on x,
$(y - \overline{y}) = b_{yx}(x - \overline{x})$
$(y - 10) = \frac{9}{20}(x - 5)$
$y = 0.45x - 2.25 + 10$
$y = 0.45x + 7.75$
When $x = 30$,
$y = 0.45(30) + 7.75$
$y = 21.25$

8. The equations of two regression lines are $3x + 2y = 26$ and $6x + y = 31$. Find (i) the means of x and y (ii) coefficient of correlation between x and y (iii) σ_y if $\sigma_x = 3$

[M16/AutoMechCivil/8M]

Solution:

We have,

$3x + 2y = 26$ ………….(1)

$6x + 2y = 31$ …………(2)

Solving the two equations, we get

$\overline{x} = \frac{5}{3}, \overline{y} = \frac{21}{2}$

Let the equation of y on x be $3x + 2y = 26$

i.e. $2y = -3x + 26$

$y = -\frac{3}{2}x + 13$

$\therefore b_{yx} = -\frac{3}{2}$

Let the equation of x on y be $6x + y = 31$

i.e. $6x = -y + 31$

$x = -\frac{1}{6}y + \frac{31}{6}$

$\therefore b_{xy} = -\frac{1}{6}$

Correlation coefficient,

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

$r = \sqrt{-\frac{3}{2} \times -\frac{1}{6}}$

$r = -\frac{1}{2}$

we have,

$b_{yx} = r\frac{\sigma_y}{\sigma_x}$

$-\frac{3}{2} = -\frac{1}{2} . \frac{\sigma_y}{3}$

$\therefore \sigma_y = 9$

9. Find the equations of lines of regression for the following data. Also find r.

x	5	6	7	8	9	10	11
y	11	14	14	15	12	17	16

[M16/CompIT/6M]

Solution:

x	y	x^2	y^2	xy
5	11	25	121	55
6	14	36	196	84
7	14	49	196	98
8	15	64	225	120
9	12	81	144	108
10	17	100	289	170
11	16	121	256	176
56	99	476	1427	811

$$\overline{x} = \frac{\sum x}{N} = \frac{56}{7} = 8$$

$$\overline{y} = \frac{\sum y}{N} = \frac{99}{7} = 14.1429$$

$$b_{yx} = \frac{\sum xy - \frac{\sum x . \sum y}{N}}{\sum x^2 - \frac{(\sum x)^2}{N}} = \frac{811 - \frac{56 \times 99}{7}}{476 - \frac{(56)^2}{7}} = 0.6786$$

$$b_{xy} = \frac{\sum xy - \frac{\sum x . \sum y}{N}}{\sum y^2 - \frac{(\sum y)^2}{N}} = \frac{811 - \frac{56 \times 99}{7}}{1427 - \frac{(99)^2}{7}} = 0.7074$$

Equation of y on x,

$$(y - \overline{y}) = b_{yx}(x - \overline{x})$$

$$y - 14.1429 = 0.6786(x - 8)$$

$$y = 0.6786x - 5.4288 + 14.1429$$

$$y = 0.6786x + 8.7141$$

Equation of x on y,

$$(x - \overline{x}) = b_{xy}(y - \overline{y})$$

$$x - 8 = 0.7074(y - 14.1429)$$

$$x = 0.7074y - 10.0047 + 8$$

$$x = 0.7074y - 2.0047$$

Correlation coefficient,

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

$$r = \sqrt{0.6786 \times 0.7074}$$

$$r = 0.6929$$

10. A chemical engineer is investigating the effect of process operating temperature X on product yield Y. The study results in the following data:

X:	100	110	120	130	140	150	160	170	180	190
Y:	45	51	54	61	66	70	74	78	85	89

Find the equation of the least square line which will enable to predict yield on the basis of temperature. Find also the degree of relation between the temperature and the yield.

[N16/AutoMechCivil/8M][M17/AutoMechCivil/8M]

Solution:

x	y	$d_x = x - A$	$d_y = y - B$	d_x^2	d_y^2	$d_x d_y$
100	45	-45	-22	2025	484	990
110	51	-35	-16	1225	256	560
120	54	-25	-13	625	169	325
130	61	-15	-6	225	36	90
140	66	-5	-1	25	1	5
150	70	5	3	25	9	15
160	74	15	7	225	49	105
170	78	25	11	625	121	275
180	85	35	18	1225	324	630
190	89	45	22	2025	484	990
1450	673	0	3	8250	1933	3985

$$\overline{x} = \frac{\sum x}{N} = \frac{1450}{10} = 145 \ (A = 145)$$

$$\overline{y} = \frac{\sum y}{N} = \frac{673}{10} = 67.3 \ (B = 67)$$

$$b_{yx} = \frac{\sum d_x d_y - \frac{\sum d_x \sum d_y}{N}}{\sum d_x^2 - \frac{(\sum d_x)^2}{N}} = \frac{3985 - \frac{0\times3}{10}}{8250 - \frac{(0)^2}{10}} = 0.483$$

$$b_{xy} = \frac{\sum d_x d_y - \frac{\sum d_x \sum d_y}{N}}{\sum d_y^2 - \frac{\left(\sum d_y\right)^2}{N}} = \frac{3985 - \frac{0\times3}{10}}{1933 - \frac{(3)^2}{10}} = 2.0625$$

Equation of y on x,

$$(y - \overline{y}) = b_{yx}(x - \overline{x})$$

$$y - 67.3 = 0.483(x - 145)$$

$$y = 0.483x - 70.035 + 67.3$$

$$y = 0.483x - 2.735$$

Coefficient of correlation,

$$r = \sqrt{b_{yx} \times b_{xy}} = \sqrt{0.483 \times 2.0625}$$

$$r = 0.9981$$

11. State whether the following statement is true or false with reasoning: "The regression coefficients between 2x and 2y are same as those between x and y"
[M17/CompIT/5M]
Solution:
We know that, regression coefficient between x and y is

$$b_{xy} = \frac{\sum xy - \frac{\sum x \sum y}{N}}{\sum y^2 - \frac{(\sum y)^2}{N}}$$

Now, let $u = 2x$ and $v = 2y$
Then regression coefficient between u and v becomes

$$b_{uv} = \frac{\sum uv - \frac{\sum u \sum v}{N}}{\sum v^2 - \frac{(\sum v)^2}{N}}$$

$$= \frac{\sum(2x)(2y) - \frac{\sum 2x \sum 2y}{N}}{\sum(2y)^2 - \frac{(\sum 2y)^2}{N}}$$

$$= \frac{4\left[\sum xy - \frac{\sum x \sum y}{N}\right]}{4\left[\sum y^2 - \frac{(\sum y)^2}{N}\right]}$$

$$= \frac{\sum xy - \frac{\sum x \sum y}{N}}{\sum y^2 - \frac{(\sum y)^2}{N}}$$

$$b_{uv} = b_{xy}$$

Thus, the given statement is true that the regression coefficients between 2x and 2y are same as those between x and y

12. Find the equations of line of regression of Y on X for the following data.

x	5	6	7	8	9	10	11
y	11	14	14	15	12	17	16

[N17/CompIT/6M]

Solution:

x	y	x^2	y^2	xy
5	11	25	121	55
6	14	36	196	84
7	14	49	196	98
8	15	64	225	120
9	12	81	144	108
10	17	100	289	170
11	16	121	256	176
56	99	476	1427	811

$$\overline{x} = \frac{\sum x}{N} = \frac{56}{7} = 8$$

$$\overline{y} = \frac{\sum y}{N} = \frac{99}{7} = 14.1429$$

$$b_{yx} = \frac{\sum xy - \frac{\sum x.\sum y}{N}}{\sum x^2 - \frac{(\sum x)^2}{N}} = \frac{811 - \frac{56 \times 99}{7}}{476 - \frac{(56)^2}{7}} = 0.6786$$

Equation of y on x,

$$(y - \overline{y}) = b_{yx}(x - \overline{x})$$

$$y - 14.1429 = 0.6786(x - 8)$$

$$y = 0.6786x - 5.4288 + 14.1429$$

$$y = 0.6786x + 8.7141$$

13. The equations of two regression lines are $3x + 2y = 26$ and $6x + y = 31$. Find (i) the means of x and y (ii) coefficient of correlation between x and y **[N17/AutoMechCivil/5M]**

Solution:

We have,

$3x + 2y = 26$ ………….(1)

$6x + 2y = 31$ …………(2)

Solving the two equations, we get

$\overline{x} = \frac{5}{3}, \overline{y} = \frac{21}{2}$

Let the equation of y on x be $3x + 2y = 26$

i.e. $2y = -3x + 26$

$y = -\frac{3}{2}x + 13$

$\therefore b_{yx} = -\frac{3}{2}$

Let the equation of x on y be $6x + y = 31$

i.e. $6x = -y + 31$

$x = -\frac{1}{6}y + \frac{31}{6}$

$\therefore b_{xy} = -\frac{1}{6}$

Correlation coefficient,

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

$r = \sqrt{-\frac{3}{2} \times -\frac{1}{6}}$

$r = -\frac{1}{2}$