UNIT-III

Correlation and Regression

PART-B

1. A random sample of 5 college students are selected and their grades in mathematics and statistics are found to be

	1	2	3	4	5
Mathematics	85	60	73	40	90
Statistics	93	75	65	50	80

Calculate Spearman's rank correlation coefficient.

Sol:

Mathematics X	Ranks R_1	Statistics	R_2	$D = R_1 - R_2$	D^2
85	2	93	1	1	1
60	4	75	3	1	1
73	3	65	4	-1	1
40	5	50	5	0	0
90	1	80	2	-1	1
					$\Sigma D^2 = 4$

Here $N = 5$, $\Sigma D^2 = 4$

$\therefore$ Pearman's Rank correlation $= 1 - \dfrac{6\Sigma D^2}{N(N^2-1)}$

$= 1 - \dfrac{6 \times 4}{5(5^2-1)}$

$= 0.8$

2. Calculate the coefficient of correlation from the following data.

x	12	9	8	10	11	13	7
y	14	8	6	9	11	12	13

Sol:

x	y	$X = x - \bar{x}$	$Y = y - \bar{y}$	X^2	Y^2	XY
12	14	2	3.6	4	12.9	7.2
9	8	-1	-2.4	1	5.7	2.4
8	6	-2	-4.4	4	19.3	8.8
10	9	0	-1.4	0	1.9	0
11	11	1	0.6	1	0.3	0.6
13	12	3	1.6	9	2.5	4.8
7	13	-3	2.6	9	6.7	-7.8
$\Sigma x = 70$	$\Sigma y = 63$			$\Sigma X^2 = 28$	$\Sigma Y^2 = 49.3$	$\Sigma XY = 16$

$\bar{x} = \dfrac{\Sigma x}{n} = \dfrac{70}{7} = 10$, $\quad \bar{y} = \dfrac{\Sigma y}{n} = \dfrac{63}{7} =$

$\therefore$ Correlation coefficient $= \dfrac{\Sigma xy}{\sqrt{\Sigma x^2 \, \Sigma y^2}} = \dfrac{16}{\sqrt{28 \times 49.3}} = 0.43$

3. The following data gives the marks in obtained by 10 Students in accountancy and statistics.

R.No.	1	2	3	4	5	6	7	8	9	10
Acc	45	70	65	30	90	40	50	75	85	60
Stats	35	90	70	40	95	40	80	80	80	50

Find the coefficient of correlation

Sol:

x	y	$X = x-\bar{x}$	$Y = y-\bar{y}$	X^2	Y^2	XY
45	35	−16	−31	256	961	496
70	90	9	24	81	576	216
65	70	4	4	16	16	16
30	40	−31	−26	961	676	806
90	95	29	29	841	841	841
40	40	−21	−26	441	676	546
50	80	−11	14	121	196	154
75	80	14	14	196	196	196
85	80	24	14	576	196	336
60	50	−1	−16	1	256	16
$\Sigma x = 610$	$\Sigma y = 660$			$\Sigma X^2 = 3490$	$\Sigma Y^2 = 4590$	3623

$\bar{x} = \dfrac{\Sigma x}{n} = 61$, $\bar{y} = \dfrac{\Sigma y}{n} = 66$

$$\text{correlation coefficient } (r) = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \times \Sigma Y^2}}$$

$$= \frac{3623}{\sqrt{3490 \times 4590}}$$

$$= 0.9$$

4. Calculate the Karl pearson's coefficient of correlation from the following data.

wages	100	101	102	102	100	99	97	98	96	95
cost of living	98	99	99	97	95	92	95	94	90	91

Sol:

x	y	$X = x - \bar{x}$	X^2	$Y = y - \bar{y}$	Y^2	XY
100	98	1	1	3	9	3
101	99	2	4	4	16	8
102	99	3	9	4	16	12
102	97	3	9	2	4	6
100	95	1	1	0	0	0
99	92	0	0	-3	9	0
97	95	-2	4	0	0	0
98	94	-1	1	-1	1	1
96	90	-3	9	-5	25	15
95	91	-4	16	-4	16	16
$\Sigma x = 990$	$\Sigma y = 950$		$\Sigma X^2 = 54$		$\Sigma Y^2 = 96$	$\Sigma XY = 61$

$\bar{x} = \frac{\Sigma x}{n} = 99$, $\bar{y} = \frac{\Sigma y}{n} = 95$

$\therefore$ coefficient of correlation $r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \cdot \Sigma Y^2}}$

$= \frac{61}{\sqrt{54 \times 96}}$

$=$

5. Find a suitable coefficient of correlation for the following data:

Fertiliser used (tonnes)	15	18	20	24	30	35	40	50
Productivity (tonnes)	85	93	95	105	120	130	150	160

Sol:

x	y	$X = x - \bar{x}$	X^2	$Y = y - \bar{y}$	Y^2	XY
15	85	-14	196	-34	1156	476
18	93	-11	121	-26	676	286
20	95	-9	81	-24	576	216
24	105	-5	25	-14	196	70
30	120	1	1	1	1	1
35	130	6	36	11	121	66
40	150	11	121	31	961	341
50	160	21	441	41	1681	861
$\Sigma x = 232$	$\Sigma y = 938$		$\Sigma X^2 = 1022$		$\Sigma Y^2 = 5368$	$\Sigma XY = 2317$

$\bar{x} = \frac{\Sigma x}{n} = 29$, $\bar{y} = 119$

$\therefore$ coefficient of correlation $(r) = \dfrac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$

$= \dfrac{2317}{\sqrt{1022 \times 5368}}$

$= 0.99$

6. The following table give the distribution of the total Population and those who are totally partially blind among them. Find out if there is any relation b/w age and blindness.

Age	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of Persons (000)	100	60	40	36	24	11	6	3
Blind	55	40	40	40	36	22	18	15

Sol

Age(x)	Mid value	$X = \frac{m-45}{10}$	X^2	Blind per lakh (y)
0-10	5	-4	16	$\frac{55}{100000} \times 100000 = 55$
10-20	15	-3	9	$\frac{40}{60000} \times 100000 = 67$
20-30	25	-2	4	$\frac{40}{40000} \times 100000 = 100$
30-40	35	-1	1	$\frac{40}{36000} \times 100000 = 111$
40-50	45	0	0	$\frac{36}{24000} \times 100000 = 150$
50-60	55	1	1	$\frac{22}{11000} \times 100000 = 200$
60-70	65	2	4	$\frac{18}{6000} \times 100000 = 300$
70-80	75	3	9	$\frac{15}{3000} \times 100000 = 500$
			$\Sigma X^2 = 44$	$\Sigma y =$

$\bar{y} = \frac{\Sigma y}{n} = 185$

$Y = y - \bar{y}$	Y^2	XY
-130	16900	520
-118	13924	354
-85	7225	170
-74	5476	74
-35	1225	0
15	225	15
115	13225	230
315	99225	945
	$\Sigma Y^2 = 157425$	$\Sigma XY = 2308$

$\therefore$ coefficient of correlation $(r) = \dfrac{\Sigma XY}{\sqrt{\Sigma X^2 \cdot \Sigma Y^2}} = \dfrac{2308}{\sqrt{44 \times 157425}}$

7. Following are the ranks obtained by 10 students in two subjects, statistics and mathematics. To what extent the knowledge of the students in two subjects is related?

Statistics	1	2	3	4	5	6	7	8	9	10
Mathematics	2	4	1	5	3	9	7	10	6	8

Sol:

R_1	R_2	$D = R_1 - R_2$	D^2
1	2	-1	1
2	4	-2	4
3	1	2	4
4	5	-1	1
5	3	2	4
6	9	-3	9
7	7	0	0
8	10	-2	4
9	6	3	9
10	8	2	4
			$\Sigma D^2 = 40$

Given $N = 10$, $\Sigma D^2 = 40$

$\therefore$ Rank correlation coefficient $(\rho) = 1 - \dfrac{6 \Sigma D^2}{N(N^2 - 1)}$

$= 1 - \dfrac{6 \times 40}{10(10^2 - 1)}$

$= 0.75$

8. The ranks of 16 students in mathematics and statistics are as follows (1,1), (2,10), (3,3), (4,4), (5,5), (6,7), (7,2), (8,6), (9,8), (10,11), (11,15), (12,9), (13,14), (14,12), (15,16), (16,13). Calculate the rank correlation coefficient for proficiencies of this group in mathematics and statistics.

Soln

R_1	R_2	$D = R_1 - R_2$	D^2
1	1	0	0
2	10	-8	64
3	3	0	0
4	4	0	0
5	5	0	0
6	7	-1	1
7	2	5	25
8	6	2	4
9	8	1	1
10	11	-1	1
11	15	-4	16
12	9	3	9
13	14	-1	1
14	12	2	4
15	16	-1	1
16	13	3	9
			$\Sigma D^2 = 136$

$N = 16$.

$\therefore$ Rank correlation coefficient $(\rho) = 1 - \dfrac{6\Sigma D^2}{N(N^2-1)}$

$= 1 - \dfrac{6 \times 136}{16(16^2-1)}$

$= 0.8$

9. A sample of 12 fathers and their elder sons gave the following data about their elder sons. Calculate the coefficient of rank correlation

Fathers	65	63	67	64	68	62	70	66	68	67	69	71
Sons	68	66	68	65	69	66	68	65	71	67	68	70

(7)

X	R_1	Y	R_2	$D = R_1 - R_2$	D^2
65	9	68	5.5	3.5	12.25
63	11	66	9.5	1.5	2.25
67	6.5	68	5.5	1.0	1
64	10	65	11.5	-1.5	2.25
68	4.5	69	3	1.5	2.25
62	12	66	9.5	2.5	6.25
70	2	68	5.5	-3.5	12.25
66	8	65	11.5	3.5	12.25
68	4.5	71	1	-3.5	12.25
67	6.5	67	8	-1.5	2.25
69	3	68	5.5	-2.5	6.25
71	1	70	2	-1	1

$\Sigma D^2 = 72.5$

Given $N = 12$, $m = 2$, $m = 2$ (X)

$m = 4$, $m = 2$, $m = 2$ (Y)

Rank correlation coefficient (ρ)

$$= 1 - \frac{6\left[\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m)\right]}{N(N^2 - 1)}$$

$$= 1 - \frac{6\left[72.5 + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(4^3 - 4) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2)\right]}{12(12^2 - 1)}$$

$= 0.72$

10. Following are the rank obtained by 10 students in two subjects, Statistics and Mathematics. To what extent the knowledge of the students in two subjects are related?

	48	33	40	9	16	16	65	24	16	57
Mathematics							20	9	6	19
Statistics	13	13	24	6	15	4				

Sol:

X	R_1	Y	R_2	$D=R_1-R_2$	D^2
48	8	13	5.5	2.5	6.25
33	6	13	5.5	0.5	0.25
40	7	24	10	-3	9
9	1	6	2.5	-1.5	2.25
16	3	15	7	4	16
16	3	4	1	2	4
65	10	20	9	1	1
24	5	9	4	1	1
16	3	6	2.5	5	25
57	9	19	8	1	1

$\Sigma D^2 = 41$

Given $N=10$, $m=3$, $m=2$, $m=2$

$\therefore$ Rank correlation coefficient $(\rho) = 1 - \dfrac{6\left[\Sigma D^2 + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m)\right]}{N(N^2-1)}$

$= 1 - \dfrac{6\left[41 + \frac{1}{12}(3^3-3) + \frac{1}{12}(2^3-2) + \frac{1}{12}(2^3-2)\right]}{10(10^2-1)}$

$= \underline{\underline{0.73}}$

11. Determine the regression equation which best fit to the following data

x	10	12	13	16	17	20	25
y	10	22	24	27	29	33	37

Sol:

x	y	x^2	xy
10	10	100	100
12	22	144	264
13	24	169	312
16	27	256	432
17	29	289	493
20	33	400	660
25	37	625	925
$\Sigma x = 113$	$\Sigma y = 182$	$\Sigma x^2 = 1983$	$\Sigma xy = 3186$

& $n = 7$

$\therefore$ The regression equation of y on x is

$y = a + bx$ —— ①

$\therefore$ The normal equations are

$\Sigma y = na + b\Sigma x \Rightarrow 182 = 7a + 113b$ —— ②

$\Sigma xy = a\Sigma x + b\Sigma x^2 \Rightarrow 3182 = 113a + 1983b$ —— ③

Solve (2) & (3)

we get $a = 0.7985$, $b = 1.5611$

$\therefore$ Substitute these values in (1), we get

$\therefore y = 0.7985 + 1.5611x$

12. In the following table S is weight of Potassium bromide which will dissolve in 100 gms, of water at $v^\circ C$. Fit an equation of the form $S = mT + b$ by the method of least squares. Use this relation to estimate S when $T = 50^\circ$.

T	0	20	40	60	80
S	54	65	75	85	96

Sol:

S	T	ST	S^2
54	0	0	2916
65	20	1300	4225
75	40	3000	5625
85	60	5100	7225
96	80	7680	9216
		17080	29207

Given $n = 5$, $\Sigma T = 200$, $\Sigma ST = 10660$

$\Sigma S = 375$, $\Sigma S^2 =$

13. From a sample of 200 pairs of observation the following quantities were calculated.

$\Sigma X = 11.34$, $\Sigma Y = 20.78$, $\Sigma X^2 = 12.16$, $\Sigma Y^2 = 84.96$, $\Sigma XY = 22.13$

From the above data show how to compute the coefficients of the equation $Y = a + bX$

Sol: Given equation is $Y = a + bX$ —— (1)

Normal equations are $\Sigma Y = na + b\Sigma X \Rightarrow 20.78 = 200a + 11.34b$ —— (2)

$\Sigma XY = a\Sigma X + b\Sigma X^2 \Rightarrow 22.13 = 11.34a + 12.16b$ —— (3)

Solve (2) & (3), we get $a = 0.0005$, $b = 1.82$

$\therefore Y = 0.0005 + 1.82X$

14. If $\sigma_x = \sigma_y = \sigma$ and the angle b/w the regression lines is $\tan^{-1}(4/3)$. Find r.

Sol: Given $\sigma_x = \sigma_y = \sigma$, $\theta = \tan^{-1}(4/3)$

W.K.T $\tan\theta = \left(\frac{1-r^2}{r}\right)\left(\frac{\sigma_x \sigma_y}{\sigma_x^2+\sigma_y^2}\right)$

$$\theta = \tan^{-1}\left[\left(\frac{1-r^2}{r}\right)\cdot\left(\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\right)\right]$$

$$\tan^{-1}\left(\frac{4}{3}\right) = \tan^{-1}\left[\left(\frac{1-r^2}{r}\right)\cdot\frac{1}{2}\right] \quad [\because \sigma_x = \sigma_y = \sigma]$$

$$\Rightarrow \frac{4}{3} = \frac{1-r^2}{r}$$

$$\Rightarrow r = 1/3 \text{ (or) } -3$$

15. Give the following data multiple coefficient of correlation of X_3 on X_1 and X_2.

X_1	3	5	6	8	12	14
X_2	16	10	7	4	3	2
X_3	90	72	54	42	30	12

Sol:

X_1	$x_1=X_1-\bar{X}_1$	x_1^2	X_2	$x_2=X_2-\bar{X}_2$	x_2^2	X_3	$x_3=X_3-\bar{X}_3$	x_3^2	x_1x_2	x_2x_3	x_3x_1
3	-5	25	16	9	81	90	40	1600	-45	360	-200
5	-3	9	10	3	9	72	22	484	-9	66	-66
6	-2	4	7	0	0	54	4	16	0	0	-8
8	0	0	4	-3	9	42	-8	64	0	24	0
12	4	16	3	-4	16	30	-20	400	-16	80	-80
14	6	36	2	-5	25	12	-38	1444	-30	90	-228
		90			140			4008	-100	-582	720

$$r_{12} = \frac{\Sigma x_1 x_2}{\sqrt{\Sigma x_1^2 \cdot \Sigma x_2^2}} = \frac{-100}{\sqrt{90\times 140}} = -0.89$$

$$r_{13} = \frac{\Sigma x_1 x_3}{\sqrt{\Sigma x_1^2 \cdot \Sigma x_3^2}} = \frac{-582}{\sqrt{90\times 4008}} = -0.97$$

$$r_{23} = \frac{\Sigma x_2 x_3}{\sqrt{\Sigma x_2^2 \cdot \Sigma x_3^2}} = \frac{720}{\sqrt{140 \times 4008}} = 0.96$$

$$R_{3.12} = \sqrt{\frac{r_{13}^2 + r_{23}^2 - 2 r_{13} r_{23} r_{12}}{1 - r_{12}^2}} = 0.987$$

16. For 20 army personal the regression of weight of kidneys(y) on weight of heart (x) is $y = 0.399x = 6.394$ and the regression of weight of heart on weight of kidneys is $x = 1.212y + 2.461$. Find correlation coefficient.

Sol Given $y = 0.399x + 6.394$

$x = 1.212x + 2.461$

Let $\bar{x}$ & $\bar{y}$ be the means

we have $\bar{y} = 0.399\bar{x} + 6.394$ —— (1)

$\bar{x} = 1.212\bar{y} + 2.461$ —— (2)

Solve (1) & (2), we get

$\bar{x} = 197720.03$, $\bar{y} = 78890.28$

Regression coefficient of y on x is 0.399

& that of x on y is 1.212

$\therefore$ Correlation coefficient $(r) = \sqrt{0.399 \times 1.212}$

$= 0.6953$

17. Find the most likely production corresponding to a rainfall 40 from the following data:

	Rainfall (x)	Production (y)
Average	30	500 kgs
S.D	5	100 kgs
coefficient of correlation	0.8	-

Sol we have to calculate the value of y when $x = 40$

So, find regression equation of y on x

~~Random variable is any ... associated with a real number is called~~

Given Mean of x is $\bar{x} = 30$

Mean of y is $\bar{y} = 500$

$\sigma_x = 5, \quad \sigma_y = 100$

$\therefore$ Regression of y on x

$$y - \bar{y} = r \cdot \frac{\sigma_x}{\sigma_y}(x - \bar{x})$$

$$\Rightarrow y - 500 = (0.8) \cdot \frac{5}{100}(x - 30)$$

$$\Rightarrow y = 0.04x + 498.8 \quad \text{——} \textcircled{1}$$

when $x = 40$, in ①, we get

$$y = 0.04(40) + 498.8$$

$$\therefore y = 500.4$$

18. The heights of mothers and daughters are given in the following table. From the two tables of regression estimate the expected average height of daughter when the height of the mother is 64.5 inches.

Height of mother	62	63	64	64	65	66	68	70
Height of daughter	64	65	61	69	67	68	71	65

Sol

x	y	$X = x - \bar{x}$	$Y = y - \bar{y}$	XY	X^2
62	64	-3.25	-2.25	7.3125	10.5625
63	65	-2.25	-1.25	2.8125	5.0625
64	61	-1.25	-5.25	6.5625	1.5625
64	69	-1.25	2.75	-3.4375	1.5625
65	67	-0.25	0.75	-0.1875	0.0625
66	68	0.75	1.75	1.3125	0.5625
68	71	2.75	4.75	13.0625	7.5625
70	65	4.75	-1.25	-5.9375	22.5625
$\Sigma x = 522$	$\Sigma y = 530$			21.5	49.5

$$\bar{x} = \frac{\Sigma x}{n} = \frac{522}{8} = 65.25, \quad \bar{y} = \frac{\Sigma y}{n} = 66.25$$

$\Sigma XY = 21.5$, $\Sigma X^2 = 49.5$

$$b_{yx} = \frac{\Sigma XY}{\Sigma X^2} = \frac{21.5}{49.5} = 0.43$$

Regression line of y on x is

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

$$(y-66.25) = 0.43(x-65.25)$$

$$y = 0.43x + 38.1925$$

when $x = 64.5$

$$\therefore y = 0.43(64.5) + 38.1925$$

$$= 65.9275$$

19. A panel of two judges P and Q graded seven dramatic performances by independently awarding marks as follows:

Performance	1	2	3	4	5	6	7
Marks by P	46	42	44	40	43	41	45
Marks by Q	40	38	36	35	39	37	41

would not attend, was a The eight performance, which judge Q has also been awarded 37 marks by judge P. If judge Q had been expected to have been present, how many marks would be awarded by him to the eight performance.

Soln

x	y	$X = x-\bar{x}$	$Y = y-\bar{y}$	XY	X^2	Y^2
46	40	3	2	6	9	4
42	38	-1	0	0	1	0
44	36	1	-2	-2	1	4
40	35	-3	-3	9	9	9
43	39	0	1	0	0	1
41	37	-2	-1	2	4	1
45	41	2	3	6	4	9

$$\bar{x} = \frac{\Sigma x}{n} = \frac{301}{7} = 43$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{266}{7} = 38$$

$$b_{yx} = \frac{\Sigma xy}{\Sigma x^2} = \frac{21}{28} = 0.75$$

Regression line of y on x is $y - \bar{y} = b_{yx}(x - \bar{x})$ — (1)

$$y - 38 = 0.75(x - 43)$$

$$\Rightarrow y = 0.75x + 5.75$$

If $x = 37$, $y = 0.75(37) + 5.75$

$$\underline{y = 33.5}$$

20. Find the multiple linear regression of X_1 on X_2 and X_3 from the data given below:

X_1	11	17	26	28	31	35	41	49	63	69
X_2	2	4	6	5	8	7	10	11	13	14
X_3	2	3	4	5	6	7	9	10	11	13

Sol: The regression equation of X_1 on X_2 and X_3 is

$$X_1 = a_{1.23} + b_{12.3} X_2 + b_{13.2} X_3$$

Normal equations are

$$\Sigma X_1 = N a_{1.23} + b_{12.3} \Sigma X_2 + b_{13.2} \Sigma X_3$$

$$\Sigma X_1 X_2 = a_{1.23} \Sigma X_2 + b_{12.3} \Sigma X_2^2 + b_{13.2} \Sigma X_2 X_3$$

$$\Sigma X_1 X_3 = a_{1.23} \Sigma X_3 + b_{12.3} \Sigma X_2 X_3 + b_{13.2} \Sigma X_3^2$$

X_1	X_2	X_3	X_1X_2	X_1X_3	X_2X_3	X_1^2	X_2^2	X_3^2
11	2	2	22	22	4	121	4	4
17	4	3	68	51	12	289	16	9
26	6	4	156	104	24	676	36	16
28	5	5	140	140	25	784	25	25
31	8	6	248	186	48	961	64	36
35	7	7	245	245	49	1225	49	49
41	10	9	410	369	90	1681	1000	81
49	11	10	539	490	110	2401	121	1000
63	13	11	819	693	143	3969	169	121
69	14	13	966	897	182	4761	196	169
370	80	70	3613	3197	687	780	610	16868

Sub in
Normal equations are

$$370 = 10a_{1.23} + 8b_{12.3} + 70b_{13.2}$$

$$3613 = 80a_{1.23} + 780b_{12.3} + 687b_{13.2}$$

$$3197 = 70a_{1.23} + 687b_{12.3} + 610b_{13.2}$$

Solving $a_{1.23} = 0.561$, $b_{12.3} = 1.735$, $b_{13.2} = 3.223$

Regression eqn is $X_1 = 0.561 + 1.735X_2 + 3.223X_3$

PART C

1. Find coefficient of correlation between X & Y for the following data.

X	10	12	18	24	23	27
Y	13	18	12	25	30	10

Sol:-

X	Y	X^2	Y^2	XY
10	13	100	269	130
12	18	144	324	216
18	12	324	144	216
24	25	476	624	600
23	30	529	900	690
27	10	729	100	270
		2402	2362	2122

$\therefore$ the correlation coefficient $r = \dfrac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}}$

$$= \frac{2122}{\sqrt{2402 \times 2362}}$$

$$= 0.89$$

2. Ten competitors in a musical test were ranked by the three judges A, B and C in the following order.

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

Using rank correlation method, discuss which pair of judges has the nearest approach to common likings in music.

Sol:

R_1	R_2	R_3	$D_1=R_1-R_2$	$D_2=R_1-R_3$	$D_3=R_2-R_3$	D_1^2	D_2^2	D_3^2
1	3	6	-2	-5	-3	4	25	9
6	5	4	1	2	1	1	4	1
5	8	9	-3	-4	-1	9	16	1
10	4	8	6	2	-4	36	4	16
3	7	1	-4	2	6	16	4	36
2	10	2	-8	0	8	64	0	64
4	2	3	2	1	-1	4	1	1
9	1	10	8	-1	-9	64	1	81
7	6	5	1	2	1	1	4	1
8	9	7	-1	1	2	1	1	4
						200	60	214

$$\rho_1(R_1,R_2) = 1-\frac{6\Sigma D_1^2}{N(N^2-1)} = 1-\frac{6\times 200}{10\times 99} = -7/33$$

$$\rho_2(R_1,R_3) = 1-\frac{6\Sigma D_2^2}{N(N^2-1)} = 1-\frac{6\times 60}{10\times 99} = 7/11$$

$$\rho_3(R_2,R_3) = 1-\frac{6\Sigma D_3^2}{N(N^2-1)} = 1-\frac{6\times 214}{10\times 99} = \frac{-49}{165}$$

Since $\rho_2(R_1,R_3)$ is maximum.

we conclude that the pair of judges A&C has the nearest approach to common likings in music

3. Obtain the rank correlation coefficient for the following data.

X	68	64	75	50	64	80	75	40	55	64
Y	62	58	68	45	81	60	68	48	50	70

Sol:-

X	Y	R_1	R_2	$D=R_1-R_2$	D^2
68	62	4	5	-1	1
64	58	6	7	-1	1
75	68	2.5	3.5	-1	1
50	45	9	10	-1	1
64	81	6	1	5	25
80	60	1	6	-5	25
75	68	2.5	35	-1	1
40	48	10	9	1	1
55	50	8	8	0	0
64	70	6	2	4	16

$\Sigma D^2 = 72$

$\therefore$ Rank correlation coefficient $= 1 - \dfrac{6\left[\Sigma D^2 + \frac{1}{12}(m^2-m) + \frac{1}{12}(m^2-m)\right]}{N(N^2-1)}$

$= 1 - \dfrac{6\left[72 + \frac{1}{12}(2^2-2) + \frac{1}{12}(2^2-2)\right]}{10(10^2-1)}$

$= 0.545$

4. Prove that the coefficient of correlation lies b/w -1 & 1.

Sol: coefficient of correlation is $\gamma = \dfrac{cov(x,y)}{\sigma_x \cdot \sigma_y}$ —— (1)

where $cov(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y-\bar{y})$

$$\sigma_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$\sigma_y = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2}$$

Sub these in (1), we get

$$\gamma = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right]\left[\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2\right]}}$$

Squaring on both sides

$$\gamma^2 = \frac{\left[\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\right]^2}{\left[\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2\right]}$$

Let $x_i - \bar{x} = A_i$, $y_i - \bar{y} = B_i$

$$\Rightarrow r^2 = \frac{\left[\sum_{i=1}^{n} A_i B_i\right]^2}{\sum_{i=1}^{n} A_i^2 \sum_{i=1}^{n} B_i^2}$$

$$\Rightarrow r^2 = \frac{\cancel{\sum_{i=1}} \left[A_1 B_1 + A_2 B_2 + \dots + A_n B_n\right]^2}{\left[A_1^2 + A_2^2 + \dots + A_n^2\right]\left[B_1^2 + B_2^2 + \dots + B_n^2\right]}$$

By Schwartz Inequality

$$\because \left[A_1^2 + A_2^2 + \dots + A_n^2\right]\left[B_1^2 + B_2^2 + \dots + B_n^2\right] \geq \left[A_1 B_1 + A_2 B_2 + \dots + A_n B_n\right]^2$$

$\therefore r^2 \leq 1$

$\therefore -1 \leq r \leq 1$

5. The ranks of the 15 students in two subjects A & B are given below, the two numbers within the brackets denoting the ranks of the same student in A & B respectively. (1,10), (2,7), (3,2), (4,6), (5,4), (6,8), (7,3), (8,1), (9,11) (10,15), (11,9), (12,5), (13,14), (14,12), (15,13). Use Spearman's formula to find the rank correlation coefficient.

Sol:

R_1	R_2	$D = R_1 - R_2$	D^2
1	10	-9	81
2	7	-5	25
3	2	1	1
4	6	-2	4
5	4	1	1
6	8	-2	4
7	3	4	16
8	1	7	49
9	11	-2	4
10	15	-5	25
11	9	2	4
12	5	7	49
13	14	-1	1
14	12	2	4
15	13	2	4
			$\Sigma D^2 = 308$

$\therefore$ Rank correlation coefficient

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

$$= 1 - \frac{6 \times 308}{15(15^2-1)}$$

$$= 0.45$$

6. Prove that the angle between the two regression lines.

Sol: Regression line of y on x

$$(y-\bar{y}) = r \frac{\sigma_y}{\sigma_x}(x-\bar{x}) \quad \text{--- (1)}$$

Regression line of x on y

$$(x-\bar{x}) = r \cdot \frac{\sigma_x}{\sigma_y}(y-\bar{y})$$

$$\Rightarrow (y-\bar{y}) = \frac{1}{r}\frac{\sigma_y}{\sigma_x}(x-\bar{x})$$

m_1 from (1) $\Rightarrow m_1 = r \cdot \frac{\sigma_y}{\sigma_x}$

m_2 from (2) $\Rightarrow m_2 = \frac{1}{r} \cdot \frac{\sigma_y}{\sigma_x}$

Then the angle between them is

$$\tan\theta = \left|\frac{m_2 - m_1}{1+m_1 m_2}\right|$$

$$\tan\theta = \frac{\frac{1}{r}\frac{\sigma_y}{\sigma_x} - r\frac{\sigma_y}{\sigma_x}}{1 + r\frac{\sigma_y}{\sigma_x} * \frac{1}{r}\cdot\frac{\sigma_y}{\sigma_x}}$$

$$\tan\theta = \left(\frac{1-r^2}{r}\right)\frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}$$

7) If $\sigma_x = \sigma_y = \sigma$ and the angle b/w the regression lines are $\theta = \tan^{-1}(3)$. obtain r.

Sol: Given $\sigma_x = \sigma_y = \sigma$

$$\theta = \tan^{-1}(3)$$

W.K.T $\tan\theta = \left(\frac{1-r^2}{r}\right)\left(\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\right)$

$$\theta = \tan^{-1}\left\{\left(\frac{1-r^2}{r}\right)\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\right\}$$

$$\tan^{-1}3 = \tan^{-1}\left(\frac{1-r^2}{2r}\right)$$

$$\Rightarrow 3 = \frac{1-r^2}{2r}$$

$$\Rightarrow 6r = 1 - r^2$$

$$\Rightarrow r^2 + 6r - 1 = 0$$

$$r = 0.162$$

8. If θ is the angle b/w two regression lines and S.D of Y is twice the S.D of X and $r = 0.25$, find $\tan\theta$.

Sol:- Given $r = 0.25$ & $\sigma_y = 2\sigma_x$

$\therefore$ the angle b/w two regression lies are

$$\tan\theta = \left(\frac{1-r^2}{r}\right)\frac{\sigma_x \cdot \sigma_y}{\sigma_x^2 + \sigma_y^2}$$

$$= \left[\frac{1-(0.25)^2}{0.25}\right] \cdot \frac{\sigma_x \cdot 2\sigma_x}{\sigma_x^2 + 4\sigma_x^2}$$

$$= \left[\frac{1-0.0625}{0.25}\right] \cdot \left(\frac{2}{5}\right)$$

$$= 1.5$$

9. Find the multiple linear regression equation of X_1 on X_2 & X_3 from the data given below.

X_1	2	4	6	8
X_2	3	5	7	9
X_3	4	6	8	10

Sol:- The regression equation of X_1 on X_2 & X_3 is

$$X_1 = a_{1.23} + b_{12.3}X_2 + b_{13.2}X_3 \quad \text{——} (1)$$

Normal equations are

$$\Sigma X_1 = N a_{1.23} + b_{12.3}\Sigma X_2 + b_{13.2}\Sigma X_3 \quad \text{——} (2)$$

$$\Sigma X_1 X_2 = a_{1.23}\Sigma X_2 + b_{12.3}\Sigma X_2^2 + b_{13.2}\Sigma X_2 X_3 \quad - (3)$$

$$\Sigma X_1 X_3 = a_{1.23}\Sigma X_3 + b_{12.3}\Sigma X_2 X_3 + b_{13.2}\Sigma X_3^2 \quad - (4)$$

Sol:

X_1	X_2	X_3	X_1X_2	X_1X_3	X_2X_3	X_1^2	X_2^2	X_3^2
2	3	4	6	8	12	4	9	16
4	5	6	20	24	30	16	25	36
6	7	8	42	48	56	36	49	64
8	9	10	72	80	90	64	81	100
20	24	28	140	160	188	120	164	216

Substituting the values in (2), (3) & (4)

$$6a_{1.23} + 24b_{12.3} + 28b_{13.2} = 20 \quad \text{— (5)}$$

$$24a_{1.23} + 164b_{12.3} + 188b_{13.2} = 140 \quad \text{— (6)}$$

$$28a_{1.23} + 188b_{12.3} + 216b_{13.2} = 160 \quad \text{— (7)}$$

Solving (5), (6) & (7), we get

$a_{1.23} = 0$, $b_{12.3} = 2$, $b_{13.2} = -1$

∴ The regression equation of X_1 on X_2 and X_3

$$X_1 = 2X_2 - X_3$$

10. Calculate the regression equation of Y on X from the data given below, taking deviations from actual means of X & Y.

Price (Rs.)	10	12	13	12	16	15
Amount Demanded	40	38	43	45	37	43

Estimate the likely demand when the price is Rs. 20.

Sol:

x	y	$X = x-\bar{x}$	X^2	$Y = y-\bar{y}$	Y^2	XY
10	40	-3	9	-1	1	3
12	38	-1	1	-3	9	3
13	43	0	0	2	4	0
12	45	-1	1	4	16	-4
16	37	3	9	-4	16	-12
15	43	2	4	2	4	4

$\bar{x} = \frac{\Sigma x}{n} = 13$, $\bar{y} = \frac{\Sigma y}{n} = 41$

Regression equation of y on x is

$$y - \bar{y} = b_{yx}(x - \bar{x}) \quad \text{— (1)}$$

where $b_{yx} = \frac{\Sigma xy}{\Sigma x^2} = \frac{-6}{24} = -0.25$

Sub the values in ①, we get

$$y - 41 = (-0.25)(x - 13)$$

$$y = -0.25x + 44.25 \quad \text{—— ②}$$

Regression equation when $x = 20$, in ②

$$\Rightarrow y = (-0.25)(20) + 44.25$$

$$\therefore y = 39.25$$

when the Price is 20, the likely demand is 39.25