MARUDHAR KESARI JAIN COLLEGE FOR WOMEN, VANIYAMBADI

PG & RESEARCH DEPARTMENT OF MATHEMATICS

CLASS : I BCA

SUBJECT CODE : STATISTICAL METHODS AND IT APPLICATIONS I

SUBJECT NAME : 23UECA12A

SYLLABUS

UNIT- V

Correlation - Karl Pearson – Spearman's Rank correlation - concurrent deviation methods. Regression Analysis: Simple Regression Equations.

2 CORRELATION CO-EFFICIENT

Introduction:

Bivariant distribution used to find out if there is any correlation or co-variation between two variables.

~~Bri~~

Bivariate distribution:

In certain series each term of the series may assume the values of two or more variables is called bivariate distribution.

eg: If we measure the height and weight of certain group of persons is known as bivariate distribution. One variate relative to height and other variable relative to Weight. It is used to study whether there is any correlation or co-variation between two variables.

Definitions:-

* **Correlations:**

Correlation defines if changing one variable affects change in the other variable. Correlation determines the relationship between two or more variable.

Positive correlation:

This correlation is said to be direct or positive if two variables deviate in the same direction. i.e, if one variable increase (decreases) and the other variable also Increases (decreases)

Eg: Income and Expenditure.

Negative Correlation:

This correlation is said to be diversed or negative if they constantly deviate in the opposite direction. i.e; if Increasing one result corresponding decrese in the other.

eg: Prize and demand of a commodity

Perfect Correlation:

The correlation is said to be perfect if the deviation in one variable is followed by corresponding professional deviation in the other.

* Scatter Diagram:

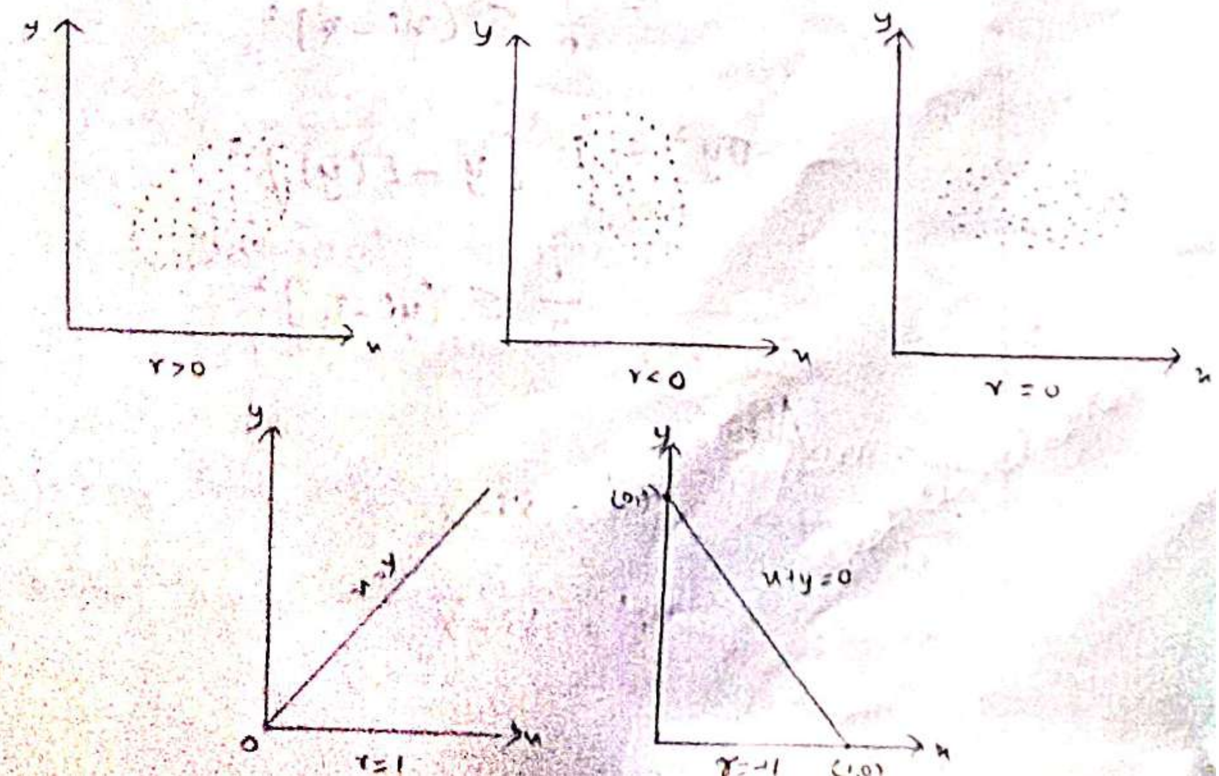
For a bivariate distribution (x_i, y_i)

$i = 1, 2, \ldots, n$. If the values of the variable x and y are plotted along x-axis and y axis respectively in the xy plane.

The diagram of dots so obtained is known as scattered diagram.

From scatterd diagram we can find whether the variables correlated or not correlated. i.e:
if the points are very dense (very close) to each other. We say that fairly good amount of correlation between the variables and if the points are widely scattered the poor correlation is expected.

The figures of scattered data for $r > 0$, $r < 0$, $r = 0$, $r = \pm 1$ are given below.



KARL PEARSON'S CO-EFFICIENT OF CORRELATION (or) CO-EFFICIENT OF CORRELATION:

Correlation between two random variables x and y usually denoted as $r(x,y)$ or r_{xy} is a numerical measure of linear relation between them and is defined by

$$r = r(x,y) = \frac{Cov(x,y)}{\sigma_x \sigma_y}$$

where,

$$Cov(x,y) = E[(x - E(x))(y - E(y))]$$

$$= \frac{1}{n} \Sigma (x_i - \bar{x})(y_i - \bar{y})$$

$$\sigma_x^2 = E[x - E(x)]^2$$

$$= \frac{1}{n} \Sigma (x_i - \bar{x})^2$$

$$\sigma_y^2 = E[y - E(y)]^2$$

$$= \frac{1}{n} \Sigma (y_i - \bar{y})^2$$

Also,

$$x_i' = x_i - \bar{x}$$

$$y_i' = y_i - \bar{y}$$

Karl Pearson's Coefficient Of Correlation:

By this method of measuring the magnitude of linear relationship between two variables.

Karl pearson's method is the most ~~widet~~ widely used method in practise and known as pearsonian coefficient of correlation is it denoted by small "r".

i) $r = \dfrac{\text{co-variance of } xy}{\sigma x \text{ or } \sigma y}$

ii) $r = \dfrac{\Sigma xy}{N \sigma x \times \sigma y}$

iii) $r = \dfrac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$

$\therefore x = (x - \bar{x})\ ;\ y = (y - \bar{y})$

σx = standard deviation of series x

σy = standard deviation of series y.

When the deviation of items take from the actual means.

iv) $r = \dfrac{N\Sigma xy - \Sigma x \Sigma y}{\sqrt{N\Sigma x^2 - (\Sigma x)^2} \cdot \sqrt{N\Sigma y^2 - (\Sigma y)^2}}$

v) $r = \dfrac{N\Sigma dx\, dy - \Sigma dx \Sigma dy}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2} \cdot \sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$

where,

$dx = x - A$ (Assumed mean)

$dx = x - \bar{x}$ (Assumed mean)

---

1. Find karl pearson's co-efficient of correlation from the following data:

Wages	100	101	102	102	100	99	97	98	96	9
Cost of living	98	99	99	97	95	92	95	94	90	9

Sol:

x	$dx = x - \bar{x}$ 99	Σx^2 (or) $d^2(x)$	y	$dy = y - \bar{y}$	Σy^2 (or) $d^2(y)$	$d(x)\, d(y)$
100	1	1	98	3	9	3
101	2	4	99	4	16	8
102	3	9	99	4	16	12
102	3	9	97	2	4	6
100	1	1	95	0	0	0
99	0	0	92	-3	9	0
97	-2	4	95	0	0	0

98	-1	1	94	-1	1	1
96	-3	9	90	-5	25	15
95	-4	16	91	-4	16	16
$\Sigma x = 990$		$\Sigma x^2 = 54$	$\Sigma y = 950$		$\Sigma y^2 = 96$	$\Sigma dx \Sigma dy = 61$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{990}{10}$$

$$\bar{x} = 99$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{950}{10}$$

$$\bar{y} = 95$$

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

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

$$= \frac{61}{\sqrt{5184}} = \frac{61}{72}$$

$$\boxed{r = 0.847}$$

2 Calculate karl pearson's co-efficient of correlation from the following table:

x	12	9	8	10	11	13	7
y	14	8	9	9	11	12	3

Sol:

The computation of karl pearson's co-efficient of correlation ;

x	y	x^2	y^2	xy
12	14	144	196	168
09	8	81	64	72
8	9	~~81~~ 64	81	72
10	9	100	81	90
11	11	121	121	121
13	12	169	144	156
7	3	49	9	21
$\Sigma x = 70$	$\Sigma y = 66$	$\Sigma x^2 = 728$	$\Sigma y^2 = 696$	$\Sigma xy = 700$

$$r = \frac{N \Sigma xy - \Sigma x \Sigma y}{\sqrt{N \Sigma x^2 - (\Sigma x)^2} \cdot \sqrt{N \Sigma y^2 - (\Sigma y)^2}}$$

$$= \frac{7(700) - 70 \times 66}{\sqrt{7 \times 728 - (70\cancel{0})^2} \cdot \sqrt{7 \times 696 - (66)^2}}$$

$$= \frac{4900 - 4620}{\sqrt{5096 - 4900} \cdot \sqrt{4872 - 4356}}$$

$$= \frac{280}{\sqrt{196} \cdot \sqrt{516}}$$

$$= \frac{280}{\sqrt{101136}}$$

$$= \frac{280}{318.01}$$

$$\boxed{r = 0.88}$$

(or) Using Assumed Mean:-

x	y	$dx = x - A$	dx^2	$dy = y - B$	dy^2	$dxdy$
12	14	2	4	5	25	10
9	8	-1	1	-1	1	1
8	9	-2	4	0	0	0
A 10	B 9	0	0	0	0	0
11	11	1	1	2	~~1~~ 4	2
13	12	3	9	3	9	9
7 ~~12~~	3	-3	9	-6	36	18
		$\Sigma dx = 0$	$\Sigma dx^2 = 28$	$\Sigma dy = 3$	$\Sigma dy^2 = 75$	$\Sigma dxdy = 40$

$$r = \frac{N \cdot \Sigma xy - \Sigma x \Sigma y}{\sqrt{N \Sigma x^2 - (\Sigma x)^2} \cdot \sqrt{N \Sigma y^2 - (\Sigma y)^2}}$$

$$= \frac{7 \times 40 - (0)(3)}{\sqrt{7 \times 28 - (0)} \cdot \sqrt{7 \times 75 - (3)^2}}$$

$$= \frac{280}{\sqrt{196} \cdot \sqrt{516}} = \frac{280}{\sqrt{101136}}$$

$$\boxed{r = 0.88}$$

3. ~~Find~~ calculate the co-efficient of correlation between x and y from the following data:

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

Sol:

x	y	$dx = x-A$	dx^2	$dy = y-B$	dy^2	$dxdy$
1	2	-3	9	-1	1	3
2	4	-2	4	1	1	-2
3	5	-1	1	2	4	-2
4 A	B 3	0	0	0	0	0
5	8	1	1	5	25	5
6	6	2	4	3	9	6
7	7	3	9	4	16	12
		$\Sigma dx = 0$	$\Sigma dx^2 = 28$	$\Sigma dy = 14$		$\Sigma dxdy = 22$

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

$$= \frac{7 \times 22 - 0(14)}{\sqrt{7(28) - 0^2} \cdot \sqrt{7(56) - (14)^2}}$$

$$= \frac{154}{\sqrt{196} \cdot \sqrt{196}}$$

$$= \frac{154}{196}$$

$$\boxed{r = 0.7857}$$

4. Find the co-efficient of correlation in the following case,

Height Of Father	65	66	67	67	68	69	71	73
Weight Of Son	67	68	64	68	72	70	69	70

Sol:-

x	y	dx = x - A	dx^2	dy = y - B	dy^2	dxdy
65	67	-3	9	-5	25	+15
66	68	-2	4	-4	16	8
67	64	-1	1	-8	64	8
67	68	-1	1	-4	16	4
A 68	B 72	0	0	0	0	0
69	70	1	1	-2	4	-2
71	69	3	9	-3	9	-9
73	70	5	25	-2	4	-10
		$\Sigma dx = 2$	$\Sigma dx^2 = 50$	$\Sigma dy = -28$	$\Sigma dy^2 = 138$	$\Sigma dxdy = 14$

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

$$= \frac{8 \times 14 - 2(-28)}{\sqrt{8(50) - (2)^2} \cdot \sqrt{8(138) - (-28)^2}}$$

$$= \frac{112 + 56}{\sqrt{396} \cdot \sqrt{1104 - 784}}$$

$$= \frac{168}{\sqrt{396} \cdot \sqrt{320}}$$

$$= \frac{168}{\sqrt{126720}}$$

$$= \frac{168}{355.97}$$

$$\boxed{r = 0.471}$$

5 Find if there is any significant correlation between the height and weight given below.

Height of (in Inches)	57	59	62	63	64	65	55	58	57
Weight (in Inches)	113	117	126	126	130	129	111	116	112

Sol:-

x	y	$dx = x - A$	dx^2	$dy = y - B$	dy^2	$dxdy$
57	113	-7	49	-17	289	119
59	117	-5	25	-13	169	65
62	126	-2	4	-4	16	8
63	126	-1	1	-4	16	4
A 64	130 B	0	0	0	0	0
65	129	1	1	-1	1	-1
55	111	-9	81	-19	361	171
58	116	-6	36	-14	196	84
57	112	-7	49	-18	324	126
		$\Sigma dx = -36$	$\Sigma dx^2 = 246$	$\Sigma dy = -90$	$\Sigma dy^2 = 1372$	$\Sigma dxdy = 576$

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

$$= \frac{9 \times 576 - (-36)(-90)}{\sqrt{9(246) - (-36)^2} \cdot \sqrt{9(1372) - (-90)^2}}$$

$$= \frac{5184 - 3240}{\sqrt{2214 - 1296} \cdot \sqrt{12348 - 8100}}$$

$$= \frac{1944}{\sqrt{918} \cdot \sqrt{4248}}$$

man

$\boxed{r = 0.9877}$

$$= \frac{1944}{1974.75}$$

$r = 0.9844$

Rank Correlation:

Let (x_i, y_i) $i = 1, 2, 3, \dots n$ be the rank of the i^{th} Individual in two characterics A and B respectively.

Pearson co-efficient of correlation between rank x_i's and y_i's is called rank correlation co-efficient between A and B for that group of Individuals.

$$\rho = 1 - \frac{6 \sum_{i=1}^{n} D_i^2}{n(n^2-1)}$$

Pearsons formula for Rank correlation Co-efficient

Tied Rank:

Let us suppose that 'm' of the Individual say $(k+1)^{th}, (k+2)^{th}, \dots (k+m)$ are tied then each of this Individual assigned a common rank which is the arithmetic mean of the ranks $[(k+1), (k+2), \dots (k+m)^{th}]$

$$p(x, y) = \frac{\frac{n(n^2-1)}{6} - \Sigma D^2 + Tx + Ty}{\left[\frac{n(n^2-1)}{6} - 2Tx\right]^{\frac{1}{2}} \left[\frac{n(n^2-1)}{6} - 2Ty\right]}$$

Where,

$$T_x = \frac{1}{12} \sum_{i=1}^{3} (m_i^3 - m_i)$$

$$T_y = \frac{1}{12} \sum_{i=1}^{3} (m_i^3 - m_i)$$

Problems:

1. The ranks of same 16 students in Mathematics and Physics 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 for proficiency of this group Maths and Physics.

Sol:-

$$\rho = 1 - \frac{6 \Sigma D^2}{n^3 - n}$$

Rank in Maths (x)	Rank in Physics (y)	D=x-y	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

$\Sigma D^2 = 136$

$\Sigma D^2 = 136$; $n = 16$; $n^3 = 4096$

$$\rho = 1 - \frac{6(136)}{4096 - 16}$$

$$= 1 - \frac{816}{4080}$$

$$= 1 - 0.2$$

$$\boxed{\rho = 0.8}$$

Rank correlation for proficiency of groups Maths & physics is 0.8.

---

2 Calculate the correlation co-efficient for the following height (in Inches) of fathers (x) and their sons (y)

x	66	66	67	67	68	69	70	72
y	67	68	65	68	72	72	69	71

Sol:-

x	y	x^2	y^2	xy
65	67	4225	4489	4355
66	68	4356	4624	4488
67	65	4489	4225	4355
67	68	4489	4624	4556
68	72	4624	5184	4896
69	72	4761	5184	4968
70	69	4900	4761	4830
72	71	5184	5041	5112
$\Sigma x =$ 544	$\Sigma y =$ 552	$\Sigma x^2 =$ 37028	$\Sigma y^2 =$ 38132	$\Sigma xy =$ 37560

$$\bar{x} = \frac{\Sigma x}{N} = \frac{544}{8}$$

$$\bar{x} = 68$$

$$\bar{y} = \frac{\Sigma x}{N} = \frac{552}{8}$$

$$\bar{y} = 69$$

Correlation Co-efficient:

$$r = \frac{Cov(x,y)}{\sigma_x \sigma_y}$$

$$Cov(x,y) = \frac{1}{n}\Sigma xy - \bar{x}\bar{y}$$

$$= \frac{1}{8}(37560) - (68)(69)$$

$$= 4695 - 4692$$

$$Cov(x,y) = 3$$

$$\sigma_x^2 = \frac{1}{n}\Sigma x_i^2 - \bar{x}^2$$

$$= \frac{1}{8}(37028) - (68)^2$$

$$= 4628.5 - 4624$$

$$= 4.5$$

$\sigma_x^2 = 4.5$,	$\sigma_x = 2.12$

$$\sigma_y^2 = \frac{1}{n}\Sigma y_i^2 - \bar{y}^2$$

$$= \frac{1}{8}(38132) - (69)^2$$

$$= 4766.5 - 4761$$

$\sigma_y^2 = 5.5$
$\sigma_y = 2.34$

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

$$= \frac{3}{(2.12)(2.34)}$$

$$= \frac{3}{4.9608}$$

$$\boxed{r(x,y) = 0.604.}$$

$\therefore$ Correlation co-efficient is 0.604

3 Find the co-efficient of correlation between two subjects the marks obtained by 10 students in Mathematics and statistics are gn below.

Maths	75	30	60	80	53	35	15	40	38	48
Stat.	85	45	54	91	58	63	35	43	45	44

Sol:

x	y	x^2	y^2	xy
75	85	5625	7225	6375
30	45	900	2025	1350
60	54	3600	2916	3240
80	91	6400	8281	7280
53	58	2809	3364	3074
35	63	1225	3969	2205
15	35	225	1225	525
40	43	1600	1849	1720
38	45	1444	2025	1710
48	44	2304	1936	2112
$\Sigma x = 474$	$\Sigma y = 563$	$\Sigma x^2 = 26132$	$\Sigma y^2 = 34815$	$\Sigma xy = 29591$

$$\bar{x} = \frac{\Sigma x}{N} = \frac{474}{10}$$

$$\bar{x} = 47.4$$

$$\bar{y} = \frac{\Sigma y}{N} = \frac{563}{10}$$

$$\bar{y} = 56.3$$

Correlation co-efficient:

$$r = \frac{cov(x,y)}{\ldots}$$

$$\mathrm{cov}(x,y) = \frac{1}{n} \Sigma x\dot{y} - \bar{x}\bar{y}$$

$$= \frac{1}{10}(29591) - (47.4)(56.3)$$

$$= 290.48$$

$$\mathrm{Cov}(x,y) = 290.48$$

$$\sigma x^2 = \frac{1}{n} \Sigma x_i^2 - \bar{x}^2$$

$$= \frac{1}{10}(26132) - (47.4)^2$$

$$= 366.44$$

$\sigma x^2 = 366.44$
$\sigma x = 19.142$

$$\sigma y^2 = \frac{1}{n} \Sigma y_i^2 - \bar{y}^2$$

$$= \frac{1}{10}(34815) - 3169.69$$

$\sigma y^2 = 311.81$
$\sigma y = 17.658$

$$r(x,y) = \frac{\mathrm{Cov}(x,y)}{\sigma x \, \sigma y}$$

$$= \frac{290.48}{(19.142)(17.658)}$$

$$= \frac{290.48}{338.009}$$

$\boxed{r = 0.859}$

REGRESSION:

Regression is the measure of the average of relationship between two or more variable in terms of the original units of the data.

Uses of Regression Analysis:

Regression analysis is used in statistics in all those field where two or more relative variables are having the tendency to go back to the average.

The regression analysis is highly useful and the regression line equation helps to estimate the value of dependent variable, when the values of Independent variables are used in the equation.

Regression analysis predicts the value of dependent variable from the values of Independent Variables.

We can calculate co-efficient of correlation [r] and the co-efficient of determination

[r^2] with the help of regression co-efficient.

Regression analysis in statistical estimation of demand curve, supply curves, production function, cost function, consumption function etc.,.

Difference b/w correlation and Regression:

Correlation	Regression
1. Both variables x and y are random variables.	x is a random variable and y is a fixed variable. Sometimes both the variables may be random variable
2. There may be nonsense correlation between two variables.	In regression there is no such nonsense regression.
3 It has limited application, because it is confined only to linear relationship between the variable	It has wider applications as it studies linear and non linear relationship between the variable
4. It is not very useful for further mathematical treatment.	It is widely used for further mathematical treatment

5. If the co-efficient of correlation is positive, then the two variables are positively correlated and vice versa.	The regression co-efficient explains that the decrease in one variable is associated with the increase in the Other variable.
6. It is Immaterial whether x depends upon y or y depends upon x.	There is a functional relationship between the two variables so that we may identify b/w the Independent and dependent variables

Method Of Studying Regression:

There are two methods of studying regression.

i) Graphical Method.

ii) Algebraic Method.

i) Graphic Method:-

* The points are plotted on a graph paper representating pairs Of values of the concerned Variables.

* In this diagram the Independent variables taken on horizontal axis and dependent variable on the Vertical axis

ii) Algebraic Method:

Regression line:-

★ A regression line is a straight line fitted to the data by the method of least squares.

★ It Indicate the best probable mean value of one variable corresponding to the mean value of the Other variable.

★ Therefore, one regression line shows regression of x upon y and the Other shows the regression of y upon x.

Regression Equation:-

★ Regression Equation of x and y.

Regression equation of x ~~and~~ on y

$$x(e) = a + by$$

By the least square method we can find out the value of small a and b and determine the regression line.

$$\Sigma x = Na + b\Sigma y$$

$$\Sigma xy = a\Sigma y + b\Sigma y^2$$

Where N is the number of Observed pairs Of Values

Regression Equation of Y on X :

$$Y_c = a + bx$$

$$\Sigma y = Na + b\Sigma x$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2$$

PROBLEM :

Determine the equation Of a straight line which best fits the data:

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

Sol:

Straight line $y = a + bx$

The two normal equations are

$$\Sigma y = b\Sigma x + Na$$

$$\Sigma xy = b\Sigma x^2 + a\Sigma x$$

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

Sub the Values:

$$\Sigma y = b\Sigma x + Na$$

$$\Sigma y = 182, \quad \Sigma x = 113, \quad N = 7$$

$$113b + 7a = 182 \longrightarrow ①$$

$$\Sigma xy = b\Sigma x^2 + a\Sigma x$$

$$\Sigma x = 113, \quad \Sigma xy = 3186, \quad \Sigma x^2 = 1983$$

$$1983b + 113a = 3186 \rightarrow ②$$

from ① & ②

① $\times 113 \Rightarrow 12769b + 791a = 20566$

② $\times 7 \Rightarrow \underset{(-)}{13881b} + \underset{(-)}{791a} = \underset{(-)}{22302}$

$$-1112b = -1736$$

$$b = \frac{1736}{1112}$$

$$\boxed{b = 1.56}$$

Sub b in equ (1)

$$113b + 7a = 182$$

$$176.28 + 7a = 182$$

$$7a = 182 - 176.28$$

$$a = \frac{5.72}{7}$$

$$\boxed{a = 0.817}$$

The equation of straight line is $y = a + bx$

here, $a = \cancel{0.82}$ 0.817, $b = 1.56$

$$\boxed{y = 0.82 + 1.56x} \rightarrow y = 0.817 + 1.56x$$

∴ The equation of the required straight line is

$y = 0.82 + 1.56x$

This is called regression equation of y on x.

Deviation taken from Arithmetic Mean of x on y:

Regression equation of x ~~an~~ on y

$$x - \bar{x} = r \frac{\sigma x}{\sigma y} (y - \bar{y})$$

$\bar{x}$ = Mean of x series

$\bar{y}$ = Mean of y series.

$r \frac{\sigma x}{\sigma y}$: The regression co-efficient of x on y (or)

b_{xy}.

$$b_{xy} = r \frac{\sigma x}{\sigma y}$$

(or)

$$b_{xy} = \frac{\Sigma xy}{\Sigma y^2}$$

It gives values by which one variable changes that is for a unit change in the other variable.

$$b_{xy} = r \frac{\sigma x}{\sigma y}$$

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

$$b_{xy} = \frac{\Sigma xy}{\Sigma y^2}$$

$r \frac{\sigma y}{\sigma x}$: The regression co-efficient of y on x (or)

b_{yx}

$$b_{yx} = r \frac{\sigma y}{\sigma x}$$

(or)

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

$$b_{yx} = r \frac{\sigma y}{\sigma x}$$

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

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

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

$$r^2 = b_{xy} \times b_{yx}$$

PROBLEM:

1. Calculate the two regression equation of x on y and y on x from the data given below, taking deviations from actual means of x and y.

Price(x)	10	12	13	12	16	15
Amount demanded	40	38	43	45	37	43

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

x	$x = x - \bar{x}$ $\bar{x} = 13$	x^2	y	$\bar{y} = 41$ $y = y - \bar{y}$	y^2	xy
10	-3	9	40	-1	1	3
12	-1	1	38	-3	9	3
13	0	0	43	2	4	0
12	-1	1	45	4	16	-4
16	3	9	37	-4	16	-12
15	2	4	43	2	4	4
$\Sigma x = 78$	$\Sigma x = 0$		$\Sigma y = 246$	$\Sigma y = 0$	$\Sigma y^2 = 50$	$\Sigma xy = -6$

Regression equation of x on y

$$\bar{x} = \frac{\Sigma x}{N} = \frac{78}{6}$$

$$\bar{x} = 13$$

$$\bar{y} = \frac{\Sigma x}{n} = \frac{246}{6}$$

$$\bar{y} = 41$$

$$r\frac{\sigma x}{\sigma y} = \frac{\Sigma xy}{\Sigma y^2} = \frac{-6}{50}$$

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

$$x - 13 = \frac{-6}{50}(y - 41)$$

$$x - 13 = -0.12(y - 41)$$

$$x - 13 = -0.12y + 4.92$$

$$x = -0.12y + 4.92$$

$$x = -0.12y + 4.92 + 13$$

$$x = -0.12y + 14.92$$

Regression equation of y on x

$$y - \bar{y} = r \frac{\sigma y}{\sigma x}(x - \bar{x})$$

$$r \frac{\sigma y}{\sigma x} = \frac{\Sigma xy}{\Sigma x^2} = \frac{-6}{24}$$

$$y - \bar{y} = r \frac{\sigma y}{\sigma x}(x - \bar{x})$$

$$y - 41 = \frac{-6}{24}(x - 13)$$

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

$$y = -0.25x + 3.25 + 41$$

$$y = -0.25x + 44.25$$

$\therefore$ When x is 20, y will be

$$y = -0.25x + 44.25$$

$$= -0.25(20) + 44.25$$

$$\boxed{y = 39.25}$$

When the price is Rs. 20 the likely demand is 39.25.

2 Given the following data. calculate the expected value of y, when $x = 12$.

regression method

	x	y
Average	7.6	14.8
Standard deviation	3.6 $\sigma(x)$	2.5 $\sigma(y)$
correlation $r = 0.99$		

Sol:- Gn:

Mean of x Series $\bar{x} = 7.6$

Mean of y series $\bar{y} = 14.8$

σ of x series $= 3.6$

σ of y series $= 2.5$

Coefficient of correlation $= 0.99$

Regression equation of y on x;

$$y - \bar{y} = r \frac{\sigma y}{\sigma x} (x - \bar{x})$$

$$y - 14.8 = 0.99 \times \frac{2.5}{3.6} (x - 7.6)$$

$$y - 14.8 = 0.99 \times 0.694 (x - 7.6)$$

$$y - 14.8 = 0.687 \times X - 5.22$$

$$y = 0.687x \; X - 5.22 + 14.8$$

$$y = 0.687x + 9.58$$

When x is 12

$$y = 0.687(12) + 9.58$$

$$= 8.244 + 9.58$$

$$y = 17.824$$

---

2. From the following data obtain the two regression equation.

takes	91	97	108	121	67	124	51	73	111	57
Purchase	71	75	69	97	70	91	39	61	80	47

<u>Sol:</u>

x	$X = x - \bar{x}$, $\bar{x} = 90$	x^2	y	$Y = y - \bar{y}$, $\bar{y} = 70$	$y2$	xy
91	1	1	71	1	1	1
97	7	49	75	5	25	35
108	18	324	69	-1	1	-18
121	31	961	97	27	729	837
67	-23	529	70	0	0	0
124	34	1156	91	21	441	714
51	-39	1521	39	-31	961	1209
73	-17	289	61	-9	81	153
111	21	441	80	10	100	210
57	-33	1089	47	-23	529	759
$\Sigma x = 900$	$\Sigma x = 0$	$\Sigma x^2 = 6360$	$\Sigma y = 700$	$\Sigma y^2 = 0$	$\Sigma y^2 = 2868$	$\Sigma xy = 3900$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{900}{10} = 90$$

$$\bar{y} = \frac{\Sigma x}{n} = \frac{700}{10} = 70$$

Regression equation of x on y

$$r\frac{\sigma x}{\sigma y} = \frac{\Sigma xy}{\Sigma y^2} = \frac{3900}{2868}$$

$$x-\bar{x} = r\frac{\sigma x}{\sigma y}(y-\bar{y})$$

$$x-90 = \frac{3900}{2868}(y-70)$$

$$x-90 = 1.36(y-70)$$

$$x-90 = 1.36y-95.2$$

$$x = 1.36y-95.2+90$$

$$x = 1.36y-5.2$$

Regression equation of y on x

$$y-\bar{y} = r\frac{\sigma y}{\sigma x}(x-\bar{x})$$

$$r\frac{\sigma y}{\sigma x} = \frac{\Sigma xy}{\Sigma x^2}(x-\bar{x})$$

$$y-70 = \frac{3900}{6360}[x-90]$$

$$y-70 = 0.61(x-90)$$

$$y = 0.61x-54.9+70$$

$$y = 0.61x+15.1$$

4. Find the most likely production corresponding to a rainfall $\overset{x}{40^\circ}$ from the following data :

	Rainfall	Production
Average	30°	500 kgs
Standard deviation	5°	100 kgs
$r = 0.8$		

Sol :

Mean of x series $\bar{x} = 30$

Mean of y series $\bar{y} = 500$ kgs

σ of x series $= 5$

σ of y series $= 100$ kgs

Coefficient of correlation $= 0.8$

Regression equation of y on x ;

$$y - \bar{y} = r \frac{\sigma y}{\sigma x} [x - \bar{x}]$$

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

$$y - 500 = 0.8 (20) (x - 30)$$

$$y - 500 = 16x - 480$$

$$y = 16x - 480 + 500 \text{ kg}$$

$$y = 16x - 20 \text{ kg}$$

When x is 40

$$y = 16(40) + 20kg$$

$$= 640 + 20kg$$

$$y = 660 kgs$$

5 You are given the following data:

	x	y
Arithmetic mean	36	85
Standard deviation	11	8
Correlation coefficient = 0.66		

Sol:

i) Find the 2 Regression equation

ii) Estimate the values of x when $y = 75$.

Mean of x series $\bar{x} = 36$

Mean of y series $\bar{y} = 85$

σ of x series $= 11$

σ of y series $= 8$

Coefficient of correlation $= 0.66$

Regression equation of x on y

$$x - \bar{x} = r \frac{\sigma x}{\sigma y}(y - \bar{y})$$

$$x - 36 = 0.66 \times \frac{11}{8}(y - 85)$$

$$x - 36 = 0.66 \times 1.375 (y - 85)$$

$$x - 36 = 0.90y - 76.5$$

$$x = 0.90y - 76.5 + 36$$

$$x = 0.90y - 40.5$$

Regression equation of y on x

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

$$y - 85 = 0.66 \times \frac{8}{11} (x - 36)$$

$$y - 85 = 0.66 (0.72) (x - 36)$$

$$y - 85 = 0.47x - 16.92$$

$$y = 0.47x - 16.92 + 85$$

$$y = 0.47x + 68.08$$

∴ When x on y = 75

$$x = 0.90y - 40.5$$

$$= 0.90 (75) - 40.5$$

$$= 67.5 - 40.5$$

$$\boxed{x = 27}$$

6. The following results where worked out from the scores in statistics and mathematics in a certain examination.

	Scores in Statistics (x)	Scores in Mathematics (y)
Mean	39.5	47.5
Standard deviation	10.8	17.8

Karl pearson correlation coefficient between x and $y = -0.42$.

Find both regression lines.

Use these regression equation and estimate the value of y for $x = 50$ and also estimate the value of x for $y = 30$.

Sol:

Mean of series $\bar{x} = 39.5$

Mean of series $\bar{y} = 47.5$

σ of x series $= 10.8$

σ of y series $= 17.8$

Coefficient of correlation $= -0.42$

Regression equation of x on y:

$$x - \bar{x} = r \frac{\sigma x}{\sigma y} (y - \bar{y})$$

$$x - 39.5 = -0.42 \times \frac{10.8}{17.8} (y - 47.5)$$

$$x - 39.5 = -0.42 \times 0.60 \, (y - 47.5)$$

$$x - 39.5 = -0.252y + 11.87$$

$$x = -0.252y + 11.87 + 39.5$$

$$x = -0.252y + 51.37$$

When $y = 30$

$$x = -0.25(30) + 51.37$$

$$= -7.5 + 51.37$$

$$\cancel{x = 58.87}$$

$$x = 43.87$$

Regression equation of y on x

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

$$y - 47.5 = -0.42 \times \frac{17.8}{10.8} (x - 39.5)$$

$$y - 47.5 = -0.42 \times 1.64 \, (x - 39.5)$$

$$y - 47.5 = -0.68 \, (x - 39.5)$$

$$y = -0.68x + 26.86 + 47.5$$

$$y = -0.68x + 74.56$$

When $x = 50$

$$y = -0.68(50) + 74.36$$

$$= -34 + 74.36$$

$$y = 40.36$$

7. From the following data obtain the two regression equations:

x	10	12	13	17	18
y	5	6	7	9	13

i) Calculate regression of the two equation.

ii) Estimate the value of x when $y = 20$

Sol:-

x	$x = x - \bar{x}$, $\bar{x} = 14$	x^2	y	$y = y - \bar{y}$, $\bar{y} = 8$	y^2	xy
10	-4	16	5	-3	9	12
12	-2	4	6	-2	4	4
13	-1	1	7	-1	1	1
17	3	9	9	1	1	3
18	4	16	13	5	25	20
$\Sigma x = 70$	$\Sigma x = 0$	$\Sigma x^2 = 46$	$\Sigma y = 40$	$\Sigma y = 0$	$\Sigma y^2 = 40$	$\Sigma xy = 40$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{70}{5}$$

$$\bar{x} = 14$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{40}{5}$$

$$\bar{y} = 8$$

Regression equation of x on y

$$r\frac{\sigma x}{\sigma y} = \frac{\Sigma xy}{\Sigma y^2} = \frac{40}{40}$$

$$x - \bar{x} = r\frac{\sigma x}{\sigma y}(y - \bar{y})$$

$$x - 14 = \frac{40}{40}(y - 8)$$

$$x = y - 8 + 14$$

$$x = y + 6$$

when $y = 20$

$$x = 20 + 6$$

$$x = 26$$

Regression equation of y on x

$$r\frac{\sigma y}{\sigma x} = \frac{\Sigma xy}{\Sigma x^2} = \frac{40}{46}$$

$$y - \bar{y} = r\frac{\sigma y}{\sigma x}(x - \bar{x})$$

$$y - 8 = \frac{40}{46}(x - 14)$$

$$y - 8 = 0.86(x - 14)$$

$$y = 0.86x - 12.04 + 8$$

$$y = 0.86x - 4.04$$

8. From the following data of the rainfall production of rice, find the most likely production corresponding to the rainfall of 40.

	Rainfall (inches)	Production (Quantity)
Mean	35	50
Standard deviation	5	8
Coefficient of Correlation = 0.8		

Sol:-

Mean of series $\bar{x} = 35$

Mean of series $\bar{y} = 50$

σ of series $x = 5$

σ of series $y = 8$

Coefficient of correlation $= 0.8$

Regression equation of y on x:

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

$$y - 50 = 0.8 \times 1.6 \,(x - 35)$$

$$y - 50 = 1.28\,(x - 35)$$

$$y = 1.28x - 44.8 + 50$$

$$y = 1.28x + 5.2$$

When $x = 40$

$$y = 1.28(40) + 5.2$$

$$y = 51.2 + 5.2$$

$$y = 56.4$$

---

9. From the following data calculate:

i) Correlation coefficient.

ii) Standard deviation of y; $b_{xy} = 0.85y$

$b_{yx} = 0.89x$

$\sigma_x = 3$

Sol:

i) Correlation coefficient:

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

$$= \sqrt{0.85 \times 0.89}$$

$$= \sqrt{0.7565}$$

$$\boxed{r = 0.86}$$

ii) Standard deviation of y:

$$b_{xy} = r \frac{\sigma x}{\sigma y}$$

$$0.85 = 0.87 \times \frac{3}{\sigma y}$$

$$0.85 \sigma y = 0.87 \times 3$$

$$0.85 \sigma y = 2.61$$

$$\sigma y = \frac{2.61}{0.85}$$

$$\boxed{\sigma y = 3.07}$$

10. Calculate the co-efficient of correlation and obtain the lines of regression for the following:

x	1	2	3	4	5	6	7	8	9
y	9	8	10	12	11	13	14	16	15

Sol :-

Obtain an estimate of y which should correspond to the average x = 2.6

x	$X = x - \bar{x}$, $\bar{x} = 5$	x^2	y	$Y = y - \bar{y}$, $\bar{y} = 12$	y^2	xy
1	-4	16	9	-3	9	12
2	-3	9	8	-4	16	12
3	-2	4	10	-2	4	4
4	-1	1	12	0	0	0
5	0	0	11	-1	1	0
6	1	1	13	1	1	1
7	2	4	14	2	4	4
8	3	9	16	4	16	12
9	4	16	15	3	9	12
$\Sigma x = 45$	$\Sigma X = 0$	$\Sigma x^2 = 60$	$\Sigma y = 108$	$\Sigma Y = 0$	$\Sigma y^2 = 60$	$\Sigma xy = 57$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{45}{9}$$

$$\bar{x} = 5$$

$$\bar{y} = \frac{\Sigma x}{n} = \frac{108}{9}$$

$$\bar{y} = 12$$

Regression equation of x on y

$$r\frac{\sigma x}{\sigma y} = \frac{\Sigma xy}{\Sigma y^2} = \frac{57}{60}$$

$$x - \bar{x} = r\frac{\sigma x}{\sigma y}(y - \bar{y})$$

$$x - 5 = \frac{57}{60}(y - 12)$$

$$x - 5 = 0.95(y - 12)$$

$$x = 0.95y - 11.4 + 5$$

$x = 0.95y - 6.4$

Regression equation y on x

$$r\frac{\sigma y}{\sigma x} = \frac{\Sigma xy}{\Sigma x^2} = \frac{57}{60}$$

$$y - \bar{y} = r\frac{\sigma y}{\sigma x}(x - \bar{x})$$

$$y - 12 = \frac{57}{60}(x - 5)$$

$$y - 12 = 0.95(x - 5)$$

$$y = 0.95x - 4.75 + 12$$

$$y = 0.95x + 7.25$$

When $x = 6.2$

$$y = 0.95(6.2) + 7.25$$

$$= 5.89 + 7.25$$

$$\boxed{y = 13.14}$$

Concurrent deviation:

$$r(c) = \pm \sqrt{\pm\left(\frac{2C-N}{N}\right)}$$

where,

r(c) - Co-efficient of correlation by the concurrent deviation method.

C - No. of. concurrent deviation

N - No. of pairs of deviation compared

1. Calculate the co-efficient of concurrent deviations from the gn data:

Month	Jan	Feb	Mar	April	May	June	July	Aug	Sep
Supply	160	164	172	182	166	170	178	192	186
Price	292	280	260	234	266	254	230	190	200

Sol:

$$r(c) = \pm \sqrt{\pm\left(\frac{2C-N}{N}\right)}$$

~~here, C=0~~

Month	Supply (x)	dx	Price (y)	dy	dxdy
Jan	160		292		
Feb	164	+	280	−	−
March	172	+	260	−	−
April	182	+	234	−	−
May	166	−	266	+	−
June	170	+	254	−	−
July	178	+	230	−	−
Aug	192	+	190	−	−
Sep	186	−	200	+	−
					dxdy = 0

$$r(c) = \pm\sqrt{\pm\left(\frac{2C-N}{N}\right)}$$

$$= \pm\sqrt{\pm\left(\frac{2(0)-8}{8}\right)}$$

$$= \pm\sqrt{-(-1)}$$

$$= \pm\sqrt{1}$$

$$\boxed{r(c) = 1}$$

Rank Correlation:

1. Ten competitors in a musical test were ranked by the 3 judges. A, B and C in following order:-

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

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

Sol:

x	y	z	$d_1 = x-y$	$d_2 = y-z$	$d_3 = z-x$	d_1^2	d_2^2	d_3^2
1	3	6	-2	-3	5	4	9	25
6	5	4	1	1	-2	1	1	4
5	8	9	-3	-1	4	9	1	16
10	4	8	6	-4	-2	36	16	4
3	7	1	-4	6	-2	16	36	4
2	10	2	-8	8	0	64	64	0
4	2	3	2	-1	-1	4	1	1
9	1	10	8	-9	1	64	81	1
7	6	5	1	1	-2	1	1	4
8	9	7	-1	2	-1	1	4	1
						$\Sigma d_1^2 =$ 200	$\Sigma d_2^2 =$ 214	$\Sigma d_3^2 =$ 60

$$\rho_1 = 1 - \frac{6 \Sigma D_1^2}{n(n^2-1) \rightarrow \text{(or)} (n^3-n)}$$

here, $\Sigma D_1^2 = 200$, $n = 10$, $n^3 = 1000$

$$\rho_1 = 1 - \frac{6(200)}{1000-10}$$

$$= 1 - \frac{1200}{990}$$

$$= 1 - 1.212$$

$$\boxed{\rho_1 = -0.21}$$

$$\rho_2 = 1 - \frac{6 \Sigma D_2^2}{(n^3-n)}$$

$$= 1 - \frac{6(214)}{990}$$

$$= 1 - \frac{1284}{990}$$

$$= 1 - 1.29$$

$$\boxed{\rho_2 = -0.29}$$

$$\rho_3 = 1 - \frac{6 \Sigma D_3^2}{(n^3-n)}$$

$$= 1 - \frac{6(60)}{990}$$

$$= 1 - \frac{360}{990}$$

$$= 1 - 0.36$$

$$\boxed{\ell_3 = 0.64}$$

∴ Judge B and C has the nearest approach of common liking in music.

---

2. Ten competitors in a beauty contest were ranked by the 3 judges A, B & C in following Order:

Judge 1 :	1	5	4	8	9	6	10	7	3	2
Judge 2	4	8	7	6	5	9	10	3	2	1
Judge 3	6	7	8	1	5	10	9	2	3	4

Using rank correlation discuss which part of judges has the nearest approach to common

Sol.:

x	y	z	$d_1 = x-y$	$d_2 = y-z$	$d_3 = z-x$	d_1^2	d_2^2	d_3^2
1	4	6	-3	-2	5	9	4	25
5	8	7	-3	1	2	9	1	4
4	7	8	-3	-1	4	9	1	16
8	6	1	2	5	-7	4	25	49
9	5	5	4	0	-4	16	0	16
6	9	10	-3	-1	4	9	1	16
10	10	9	0	1	-1	0	1	1
7	3	2	4	1	-5	16	1	25
3	2	3	1	-1	0	1	1	0
2	1	4	1	-3	2	1	9	4
						74	44	156

$$\ell_1 = 1 - \frac{6 \Sigma D_1^2}{(n^3 - n)}$$

here, $n = 10$ ∴ $n^3 = 1000$

$$\ell_1 = 1 - \frac{6(74)}{(1000 - 10)}$$

$$= 1 - \frac{444}{990}$$

$$= 1 - 0.44$$

$$\ell_1 = 0.56$$

$$\ell_2 = 1 - \frac{6(44)}{990}$$

$$= 1 - \frac{264}{990}$$

$$= 1 - 0.26$$

$$\ell_2 = 0.74$$

$$\ell_3 = 1 - \frac{6(156)}{990}$$

$$= 1 - \frac{936}{990}$$

$$= 1 - 0.94$$

$$\ell_3 = 0.06$$

Repeated Rank:

$$\rho = 1 - \frac{6\left\{\Sigma D^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \frac{1}{12}(m_3^3 - m_3)\right\}}{n^3 - n}$$

1. Obtain the rank correlation co-efficient for 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:-

R_1 - Rank 1

R_2 - Rank 2

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	3.5	-1	1
40	48	10	9	1	1
55	50	8	8	0	0
64	70	6	2	4	16
					$\Sigma D^2 = 72$

* x Series :

→ 75 is repeated 2 times

$$\frac{2+3}{2} = \frac{5}{2} = 2.5 \qquad (m_1 = 2)$$

→ 64 is repeated 3 times

$$\frac{5+6+7}{3} = \frac{18}{3} = 6 \qquad (m_2 = 3)$$

y Series :-

→ 68 repeated 2 times

$\frac{3+4}{2} = \frac{7}{2} = 3.5 \quad (m_3 = 2)$

Here : $m_1 = 2$, $m_2 = 3$, $m_3 = 2$; $n = 10$

$$\rho = 1 - \frac{6\left\{\sum D^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \frac{1}{2}(m_3^3 - m_3)\right\}}{n^3 - n}$$

$$= 1 - \frac{6\left\{72 + \frac{1}{12}(6) + \frac{1}{12}(24) + \frac{1}{12}(6)\right\}}{990}$$

$$= 1 - \frac{6\left\{72 + \frac{1}{2} + 2 + \frac{1}{2}\right\}}{990}$$

$$= 1 - \frac{6\{75\}}{990}$$

$$= 1 - \frac{450}{990}$$

$$= 1 - 0.4545$$

$$\boxed{\rho = 0.5455}$$

2 Obtain the Rank correlation co-efficient for the given data:

x	43	44	46	40	44	42	45	42	38	40	42	57
y	29	31	19	18	19	27	27	29	41	30	26	10

<u>Sol</u>:

x	y	R_1	R_2	$D = R_1 - R_2$	D^2
43	29	6	5.5	0.5	0.25
44	31	4.5	3	1.5	2.25
46	19	2	10.5	-8.5	72.25
40	18	10.5	12	-1.5	2.25
44	19	4.5	10.5	-6	36
42	27	8	7.5	0.5	0.25
45	27	3	7.5	-4.5	20.25
42	29	8	5.5	2.5	6.25
38	41	12	2	10	100
40	30	10.5	4	6.5	42.25
42	26	8	9	-1	1
57	10 ~~70~~	1	1	0	0
					$\Sigma D^2 =$ 415 ~~283~~

* x <u>Series:</u>

→ 44 repeated 2 times.

$$\frac{4+5}{2} = \frac{9}{2} = 4.5 \quad (m_1 = 2)$$

→ 42 repeated 3 times.

$$\frac{7+8+9}{3} = \frac{24}{3} = 8 \quad (m_2 = 3)$$

→ 40 repeated 2 times:

$$\frac{10+11}{2} = \frac{21}{2} = 10.5 \quad (m_3 = 2)$$

* **y series**

→ 29 repeated 2 times

$$\frac{5+6}{2} = \frac{11}{2} = 5.5 \quad (m_4 = 2)$$

→ 27 repeated 2 times

$$\frac{7+8}{2} = \frac{15}{2} = 7.5 \quad (m_5 = 2)$$

→ 19 repeated 2 times

$$\frac{10+11}{2} = \frac{21}{2} = 10.5 \quad (m_6 = 2)$$

$$\rho = 1 - \frac{6\left\{\Sigma D^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \frac{1}{12}(m_3^3 - m_3) + \frac{1}{12}(m_4^3 - m_4) + \frac{1}{12}(m_5^3 - m_5) + \frac{1}{12}(m_6^3 - m_6)\right.}{n^3 - n}$$

$$= 1 - \frac{6\left\{\overset{415}{283} + \frac{1}{12}(6) + \frac{1}{12}(24) + \frac{1}{12}(6) + \frac{1}{12}(6) + \frac{1}{12}(6) + \frac{1}{12}(6)\right.}{1728 - 12}$$

$$= 1 - \frac{6\left\{\overset{415}{283} + \frac{1}{2} + 2 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2}\right\}}{1716}$$

$$= 1 - \frac{6\left\{\overset{415}{\cancel{257}} + 2 + 2.5\right\}}{1716}$$

$$= 1 - \frac{6\left\{\overset{419.5}{\cancel{237.5}}\right\}}{1716}$$

$$= 1 - \frac{1725}{1716} \qquad = 1 - \frac{6\{419.5\}}{1716}$$

$$= 1 - 1.005 \qquad = 1 - \frac{2517}{1716}$$

$$\boxed{\rho = -0.005} \qquad = 1 - 1.466$$

$$\rho = -0.466$$

---

2 Obtain the Rank correlation co-efficient of the gn data:

x	48	33	40	9	16	16	65	24	16	57
y	13	13	24	6	15	4	20	9	6	19

<u>Sol</u>:-

x	y	R_1	R_2	$D = R_1 - R_2$	D^2
48	13	3	5.5	-2.5	6.25
33	13	5	5.5	-0.5	0.25
40	24	4	1	3	9
9	6	10	8.5	1.5	2.25
16	15	8	4	4	16
16	4	8	10	-2	4
65	20	1	2	-1	1
24	9	6	7	-1	1
16	6	8	8.5	-0.5	0.25
57	19	2	3	-1	1
					$\Sigma D^2 = 41$

* x Series:

→ 16 Repeated 3 times:

$$\frac{7+8+9}{3} = \frac{24}{3} = 8 \quad (m_1 = 3)$$

* y Series:

→ 13 Repeated 2 times:

$$\frac{5+6}{2} = \frac{11}{2} = 5.5 \quad (m_2 = 2)$$

→ 6 Repeated 2 times:

$$\frac{8+9}{2} = \frac{17}{2} = 8.5 \quad (m_3 = 2)$$

$$\rho = 1 - \frac{6\left\{\Sigma D^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \frac{1}{12}(m_3^3 - m_3)\right)}{n^3 - n}$$

$$= 1 - \frac{6\left\{41 + \frac{1}{12}(24) + \frac{1}{12}(6) + \frac{1}{12}(6)\right\}}{10^3 - 10}$$

$$= 1 - \frac{6\left\{41 + 2 + \frac{1}{2} + \frac{1}{2}\right\}}{990}$$

$$= 1 - \frac{6\{44\}}{990}$$

$$= 1 - \frac{264}{990}$$

$$= 1 - 0.26$$

$$\boxed{\rho = 0.74}$$