CORRELATION ANALYSIS.

The relationship between two series when measured quantitatively is known as Correlation.

Types of Correlation:

(a) positive and Negative Correlation

(b) Linear and Non-Linear Correlation.

(c) Simple and Multiple Correlation.

Methods to Calculate Correlation:

(i) Graphical Methods:

Scatter Diagram or Scattergram.

(ii) Mathematical Methods:

a) Karl pearson's coefficient of Correlation Method

(or)

Karl pearson's method.

b) Spearman's Coefficient of Correlation Method

(or)

Spearman's Method and

(c) Concurrent Deviation Method.

Karl pearson's Method:

a) Direct Method:

This method is used

when given variables are small in magnitude.

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

where N is number of pairs.

Σx is the sum of terms of I series

Σy is the sum of terms of II series

Σx^2 is the sum of squares of terms of I series.

Σy^2 is the sum of squares of terms of II series.

Σxy is the sum of products of corresponding terms.

b) Shortcut Method:

In case the mean is a whole number above method is simple. But when the mean is in fractions, short-cut method is used. In this method, the deviations are calculated from assumed mean.

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

where $\Sigma dx = \Sigma(X - A_x)$

$\Sigma dy = \Sigma(Y - A_y)$

N = Number of pairs.

EX: 1 Calculate karl pearson's coefficient of correlation between the age and weight of the children.

Age (yrs):	1	2	3	4	5
weight (kg):	3	4	6	7	12

Solution:

The given variables are small. so applied direct method.

$N = 5.$

Age (x)	weight (y)	x^2	y^2	xy
1	3	1	9	3
2	4	4	16	8
3	6	9	36	18
4	7	16	49	28
5	12	25	144	60
$\sum_1^5 x = 15$	$\sum_1^5 y = 32$	$\sum_1^5 x^2 = 55$	$\sum_1^5 y^2 = 254$	$\sum_1^5 xy = 117$

Karl pearson's method:

$$r = \frac{N\sum_1^5 xy - \sum_1^5 x \sum_1^5 y}{\sqrt{N\sum_1^5 x^2 - (\sum_1^5 x)^2}\sqrt{N\sum_1^5 y^2 - (\sum_1^5 y)^2}}$$

$$r = \frac{5(117) - 15 \times 32}{\sqrt{5(55) - (15)^2}\sqrt{5(254) - (32)^2}}$$

$$r = \frac{585 - 480}{\sqrt{275 - 225}\sqrt{1270 - 1024}}$$

$$= \frac{105}{\sqrt{50 \times 246}}$$

$$= \frac{105}{\sqrt{12300}}$$

$$= \frac{105}{110.90}$$

$$\boxed{r = 0.9467}$$

EX : 2 calculate coefficient of correlation between profits of two firms x and y in a particular year using karl pearson's Method.

profit of Firm X :	14	12	14	16	16	17	16	15
profit of Firm Y :	13	11	10	15	15	9	14	17

solution

$N = 8$.

Let $A_x = 15$ and $A_y = 14$.

X	Y	$dx = X - A_x$	dx^2	$dy = X - A_y$	dy^2	$dx\,dy$
14	13	-1	1	-1	1	1
12	11	-3	9	-3	9	9
14	10	-1	1	-4	16	4
16	15	1	1	1	1	1
16	15	1	1	1	1	1
17	9	2	4	-5	25	-10
16	14	1	1	0	0	0
15	17	0	0	3	9	0
		$\Sigma dx = 0$	$\Sigma dx^2 = 18$	-8	62	6

Karl pearson's Method.

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

$$= \frac{8(6) - (0)(-8)}{\sqrt{8 \times 18 - (0)^2} \sqrt{8 \times 62 - (8)^2}}$$

$$= \frac{48 - 0}{\sqrt{144} \sqrt{496 - 64}} = \frac{48}{\sqrt{144} \sqrt{432}}$$

$$= \frac{48}{\sqrt{62208}} = \frac{48}{249.41} = .192$$

$\therefore r = 0.19.$

The important conclusion from the value of $r = 0.19$ is that there is hardly any significant relationship between the two variables i.e., the profits earned by x and y.

Properties of karl pearson's coefficient of correlation:

1. Karl pearson's coefficient of correlation lies between 1 and -1. i.e., $1 \geq r \geq -1$.

2. Karl pearson's coefficient is independent of change of scale.

3. Karl pearson's coefficient of correlation is independent of origin.

4. If b_{xy} and b_{yx} are two regression coefficients ; karl pearson's coefficient of correlation is $\sqrt{b_{xy} \times b_{yx}}$

5. r is independent of unit of measurement.

Spearman's coefficient of rank correlation :

professor charles spearman gave the formula by judging their rank as follow. It is known as Rank correlation formula.

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

where r is coefficient of Rank correlation
N is number of pairs
D is sum of squares of difference in respective ranks.

when ranks are given.

Ex: 3 Following are given the ranks of 8 pairs Find r.

Rank X:	4	2	7	5	3	1	8	6
Rank Y:	8	3	6	5	1	2	7	4

Solution.

Now, $N = 8$

X	Y	D	D^2
4	8	-4	16
2	3	-1	1
7	6	1	1
5	5	0	0
3	1	2	4
1	2	-1	1
8	7	1	1
6	4	2	4
			$\sum D^2 = 28$

Spearman's coefficient of rank correlation

$$r = 1 - \frac{6 \sum D^2}{N(N^2-1)}$$

$$= 1 - \frac{6(28)}{8(8^2-1)}$$

$$= 1 - \frac{168}{8(64-1)}$$

$$= 1 - \frac{168}{504}$$

$$= 1 - 0.33$$

$$\boxed{r = 0.67}$$

when ranks are not given.

EX: 4 Find out Rank correlation from the following

X	56	66	49	55	64	68	46	50
Y	40	70	50	60	80	75	49	62

Solution:

$N = 8.$

X	R_1	Y	R_2	$D = R_1 - R_2$	D^2
56	4	40	8	-4	16
66	2	70	3	-1	1
49	7	50	6	1	1
55	5	60	5	0	0
64	3	80	1	2	4
68	1	75	2	-1	1
46	8	49	7	1	1
50	6	62	4	2	4
					$\Sigma D^2 = 28$

Spearman's Method

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

$$= 1 - \frac{6(28)}{8(64-1)}$$

$$= \boxed{1 - \frac{168}{504}}$$

$$r = 1 - 0.333 \quad = \quad +0.667.$$

Repeated Ranks :

when some terms in the series are equal we use another formula.

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

where m is the no. of terms whose ranks are Equal.

Ex: 5:

Eight students have obtained the following marks in Accountancy and Economics. calculate the Rank co-efficient of correlation.

Accountancy x	25	30	38	22	50	70	30	90
Economics y	50	40	60	40	30	20	40	70

Solution

Accountancy Economics

X	R_1	Y	R_2	$D = R_1 - R_2$	D^2
25	2	50	6	-4	16
30	3.5	40	4	-0.5	0.25
38	5	60	7	-2	4
22	1	40	4	-3	9
50	6	30	2	4	16
70	7	20	1	6	36
30	3.5	40	4	-0.5	0.25
90	8	70	8	0	0
				$\sum D^2 =$	81.5

$\Sigma D^2 = 81.5$ and $N = 8$.

$$\therefore \quad r = 1 - \frac{6\left[\Sigma D^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2)\right]}{N(N^2-1)}$$

As item 30 is repeated 2 times in X. So
$m_1 = 2$

In series Y the item 40 is repeated 3 times so $m_2 = 3$

$$\therefore r = 1 - \frac{6\left\{81.5 + \frac{1}{12}(2^3-2) + \frac{1}{12}(3^3-3)\right\}}{8(64-1)}$$

$$= 1 - \frac{6(81.5 + 0.5 + 2)}{504}$$

$$= 1 - \frac{6(84)}{504}$$

$$= 1 - \frac{504}{504}$$

$$= 1 - 1$$

$$\boxed{r = 0}$$ //.

When series are more than two

Ex: 6 Following were the ranks given by three judges in a beauty contest. Find degree of correlation between 1^{st} and 2nd, 2nd and 3^{rd},

3^{rd} and 1^{st}. And also mention which pair of judges agree or disagree the most.

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

Solution.

R_1	R_2	R_3	D_1 (R_1-R_2)	D_1^2	D_2 (R_2-R_3)	D_2^2	D_3 (R_3-R_1)	D_3^2
1	7	4	-6	36	-3	9	-3	9
3	5	10	-2	4	-5	25	-7	49
7	4	3	3	9	1	1	4	16
9	6	9	3	9	-3	9	0	0
2	1	2	1	1	-1	1	0	0
4	2	8	2	4	-6	36	-4	16
10	3	1	7	49	2	4	9	81
8	8	5	0	0	3	9	3	9
6	10	7	-4	16	3	9	-1	1
5	9	6	-4	16	3	9	-1	1
				144		112		182

$$r = 1 - \frac{6\sum D^2}{N(N^2-1)} \quad ; N = 10$$

Correlation between 1^{st} and 2^{nd}

$$r_{12} = 1 - \frac{6\sum D_1^2}{N(N^2-1)}$$

$$= 1 - \frac{6(144)}{10(10^2-1)}$$

$$= 1 - \frac{864}{990}$$

$$= 1 - 0.873$$

$$r_{12} = 0.127$$ (Low degree +ive correlation)

Correlation between 2^{nd} and 3^{rd}

$$r_{23} = 1 - 6\frac{\sum_1 D_2^2}{N(N^2-1)}$$

$$= 1 - \frac{6(112)}{990}$$

$$= 1 - \frac{672}{990} = 1 - 0.679$$

$$r_{23} = 0.321$$ (Moderate degree +ive correlation)

Correlation between 3^{rd} and 1^{st}

$$r_{31} = 1 - 6\frac{\sum_1 D_3^2}{N(N^2-1)}$$

$$= 1 - \frac{6(182)}{990}$$

$$= 1 - \frac{1092}{990}$$

$$= 1 - 1.103$$

$$r_{31} = -0.103$$ (-ive correlation)

r_{23} is highest. hence 2^{nd} and 3^{rd} judges agree the most, whereas 3^{rd} and 1^{st} judges disagree the most. r_{31} being lowest.

Merits of Rank difference coefficient of correlation:

1. It is easy to calculate
2. It is simple to understand
3. It can be applied to any type of data.