Correlation

It means the quantitative relationship between two variables. Two variables are said to be ~~correlated~~ correlated when the value of one variable changes with the change in the value of other variables.

Types of Correlation

1) Positive & Negative Correlation :- If an increase or decrease in the values of one variable is associated with increase or decrease in the values of other variable, the correlation b/w them is said to be positive or direct.

If the values of two variables move in the opposite direction, the correlation is said to be negative or inverse.

2) Simple & Multiple Correlation :- In simple Correlation, the study relates to two variables only. Eg the study of correlation b/w income & saving, price & demand etc.

In multiple Correlation, If there are more than two variables & one variable is related to a number of variables, the study of relationship b/w one variable & all other variables taken together is called multiple correlation. Eg, the study of relationship b/w production of a crop (X) & rainfall (Y), use of fertilizer (Z) taken together falls under multiple correlation.

3) <u>Linear</u> & <u>Non-linear</u> <u>Correlation</u>:-

Two variables are said to have linear correlation if corresponding to a unit change in one variable, there is a constant change in the other variable over the whole distribution.

Eg

X:	1	2	3	4	5
Y:	3	5	7	9	11

Here for a unit change in the value of X there is a constant change of 2 in the corresponding value of Y.

Two variables are said to have non-linear correlation if corresponding to unit change in one variable, the other variable doesnot change at a constant rate.

Eg

X:	1	2	3	4	5
Y:	5	7	10	17	28

Methods of Correlation Analysis

1) Karl Pearson's coefficient of correlation or Co-variance Method.

2) Spearman's coefficient of correlation or Rank Correlation coefficient.

1) Karl Pearson's coefficient of correlation:

Let $(x_1, y_1), (x_2, y_2) - - , (x_n, y_n)$ are n pairs of observations of variables X & Y then the Karl-Pearson's coefficient of correlation is given by:

$$r \text{ or } r_{xy} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

where $Cov(X, Y)$ is covariance of X & Y & σ_X & σ_Y are S.D of X & Y resp.

$$Cov(X, Y) = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{n} = \frac{\Sigma xy}{n}$$

$$\sigma_X = \sqrt{\frac{\Sigma(X-\bar{X})^2}{n}} = \sqrt{\frac{\Sigma x^2}{n}}, \quad \sigma_Y = \sqrt{\frac{\Sigma(Y-\bar{Y})^2}{n}} = \sqrt{\frac{\Sigma y^2}{n}}$$

where $x = X-\bar{X}$, $y = Y-\bar{Y}$

$$\Rightarrow r = r_{xy} = \frac{\frac{1}{n}\Sigma(X-\bar{X})(Y-\bar{Y})}{\sqrt{\frac{\Sigma x^2}{n}}\sqrt{\frac{\Sigma y^2}{n}}}$$

$$\Rightarrow \boxed{r_{xy} = \frac{\Sigma xy}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}}}$$

Q1] The coefficient of correlation b/w two variables X & Y is 0.48. The covariance is 36. The variance of X is 16. Find the S.D. of Y

Sol^n Given $r_{xy} = 0.48$, $Cov(X,Y) = 36$

$\sigma_x^2 = 16 \Rightarrow \sigma_x = 4$, $\sigma_y = ?$

Using $\boxed{r_{xy} = \frac{Cov(X,Y)}{\sigma_x \sigma_y}}$ $\quad 0.48 = \frac{36}{4\sigma_y}$

$\Rightarrow \sigma_y = \frac{36 \times 100}{4 \times 48} = 18.75$

Q2] Calculate coeff. of correlation b/w X & Y for data:

X:	1	3	4	5	7	8	10
Y:	2	6	8	10	14	16	20

Sol^n

X	Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	xy	x^2	y^2
1	2	-4.43	-8.86	39.25	19.62	78.49
3	6	-2.43	-4.86	11.81	5.90	23.62
4	8	-1.43	-2.86	4.09	2.04	8.18
5	10	-0.43	-.86	0.37	0.18	0.74
7	14	1.57	3.14	4.93	2.46	9.86
8	16	2.57	5.14	13.21	6.60	26.42
10	20	4.57	9.14	41.77	20.88	83.54
				$\Sigma xy = 115.43$	$\Sigma x^2 = 57.68$	$\Sigma y^2 = 230.82$

$\bar{X} = \frac{\Sigma X}{n} = 5.43$

$\bar{Y} = \frac{\Sigma Y}{n} = 10.86$

$r_{xy} = \frac{\Sigma xy}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}} = \frac{115.43}{\sqrt{57.68}\sqrt{230.82}} = \frac{115.43}{115.39} = 1.$

$\Rightarrow \boxed{r_{xy} = 1}$

Scanned by CamScanner

Q37 Find the coeff. of correlation b/w X & Y

X :	9	8	7	6	5	4	3	2	1
Y :	15	16	14	13	11	12	10	8	9

Solⁿ $\bar{X} = \frac{\Sigma X}{n} = \frac{45}{9} = 5$ $\bar{Y} = \frac{\Sigma Y}{n} = \frac{108}{9} = 12$

X	Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	xy	x^2	y^2
9	15	4	3	12	16	9
8	16	3	4	12	9	16
7	14	2	2	4	4	4
6	13	1	1	1	1	1
5	11	0	-1	0	0	1
4	12	-1	0	0	1	0
~~30~~ 3	10	-2	-2	4	4	4
2	8	-3	-4	12	9	16
1	9	-4	-3	12	16	9
				$\Sigma xy = 57$	$\Sigma x^2 = 60$	$\Sigma y^2 = 60$

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \, \Sigma y^2}} = \frac{57}{\sqrt{60}\sqrt{60}} = 0.95$$

$$\Rightarrow \boxed{r = 0.95}$$

Q4] Find the coeff. of correlation for the following

X:	10	14	18	22	26	30
Y:	18	12	24	6	30	36

Soln Do yourself.

$\boxed{\bar{x}=20, \bar{y}=21, \Sigma xy=252, \Sigma x^2=280, \Sigma y^2=630, r=0.6}$

Q5]: Calculate Karl Pearson's coeff. of correlation:

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

Soln Do yourself.

$\boxed{\bar{x}=68, \bar{y}=69, r=0.603}$

Q6] Calculate Karl Pearson's coeff. of correlation!

X:	1	2	3	4	5	6	7	8	9	10
Y:	3	10	5	1	2	9	4	8	7	6.

Soln Do yourself.

$\boxed{\bar{x}=5.5, \bar{y}=5.5, r=0.22}$

Spearman's Rank Correlation

Rank correlation coefficient is used for measuring the relationship b/w two qualitative variables such as honesty, beauty, tastes etc.

Rank correlation coeff. is denoted by r_s or ρ (rho).

$$\boxed{\rho = 1 - \frac{6\sum D^2}{n(n^2-1)}} \quad \text{—— (1)}$$

where D is the difference in ranks of the variables & n is the no. of pair of observations.

The following situations may arise for the calculation of correlation coefficient:

1) when ranks are given.
2) when ranks are not given.
3) when items or ranks repeat in data.

1) When ranks are given:-

a) Compute difference of ranks of two variables & denote it by D.

b) Calculate $\sum D^2$.

c) Put values in eqn (1) to obtain rank correlation coefficient.

2) when ranks are not given:- we may rank the values either by assigning 1 to the smallest item then 2, 3, 4 -- in ascending order. when ranks are assigned, we apply ① formula.

3) when items or Ranks repeat:- when two or more items in a series have equal value they are assigned equal ranks or average ranks. in this case the formula becomes:

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

where $m_1, m_2, \ldots$ refers to the no. of times an items is repeated.

When ranks are given:-

Q1) 12 entries in painting competition were ranked by two judges as shown below:

Entry	A	B	C	D	E	F	G	H	I	J	K	L
Judge I	5	2	3	4	1	6	8	7	10	9	12	11
Judge II	4	5	2	1	6	7	10	9	11	12	3	8

Find the coefficient of rank correlation.

Sol^n

R_I	R_{II}	$D = R_I - R_{II}$	D^2
5	4	1	1
2	5	-3	9
3	2	1	1
4	1	3	9
1	6	-5	25
6	7	-1	1
8	10	-2	4
7	9	-2	4
10	11	-1	1
9	12	-3	9
12	3	9	81
11	8	3	9
			$\Sigma D^2 = 154$

$$\Rightarrow \rho = 1 - \frac{6\Sigma D^2}{n(n^2-1)} = 1 - \frac{6(154)}{12(12^2-1)} = 0.46$$

Q2] In fancy dress competition, two judges accord the following rank to ten participants :-

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

Calculate rank correlation coefficient.

Sol^n Do yourself $\Sigma D^2 = 264$, r_k or $\rho = -0.6$

when ranks are not given! -

Ques 1):- In a poetry competition, the participant were accorded following marks by two different judges X & Y :

Judge X :	10	15	12	17	13	16	24	14	22
Judge Y :	30	42	45	46	33	34	40	35	39

Calculate the rank correlation co-efficient

Sol^n:

X	R_I	Y	R_{II}	$D = R_I - R_{II}$	D^2
10	1	30	1	0	0
15	5	42	7	-2	4
12	2	45	8	-6	36
17	7	46	9	-2	4
13	3	33	2	1	1
16	6	34	3	3	9
24	9	40	6	-3	9
14	4	35	4	0	0
22	8	39	5	3	9

$\Sigma D^2 = 72$

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

$= 1 - \frac{6 \times 72}{9(81-1)}$

$\Rightarrow \boxed{r = 0.4}$

Q2] Ten students got the following percentage of marks in chemistry & Physics:

Marks in chemistry: 78 36 98 25 75 82 90 62 65 39
Marks in Physics: 84 51 91 60 68 62 86 58 63 47

Calculate the rank correlation co-efficient

Ans $\rho = 0.84$.

Sol^n Do yourself

When items or ranks repeat in data.

Q17: Obtain the rank correlation co-efficient 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^n

X	R_I	Y	R_{II}	$D = R_I - R_{II}$	D^2
68	7	62	6	1	1
64	5	58	4	1	1
$\boxed{75}$	8.5	68	7.5	1	1
50	2	45	1	1	1
64	5	81	10	-5	25
80	10	60	5	5	25
$\boxed{75}$	8.5	68	7.5	1	1
40	1	48	2	-1	1
55	3	50	3	0	0
64	5	70	9	-4	16
					$\Sigma D^2 = 72$

$$\boxed{\rho = 1 - \frac{6\left[\Sigma D^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2)\right]}{n(n^2-1)}} \Rightarrow \rho = 1 - 6\left[72 + \frac{1}{12}\right.$$

P.T.O

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

$$\Rightarrow \rho = 1 - \frac{6\left[(72) + \frac{1}{12}\times 3(3^2-1) + \frac{1}{12}(2)(2^2-1) + \frac{1}{12}(2)(2^2-1)\right]}{10(10^2-1)}$$

$$\Rightarrow \rho = 1 - \frac{6[75]}{990} \Rightarrow \boxed{\rho = 0.545}$$

Q2] A sample of 12 fathers & their eldest sons gave the following data about their heights in 'inches.'

Father:	65	63	67	64	68	62	70	66	68	67	69	71
Son:	68	66	68	65	69	66	68	65	71	67	68	70

Calculate the co-efficient of rank correlation.

Solⁿ Do yourself.

$\boxed{\text{Ans } \rho = 0.722}$

Regression:-

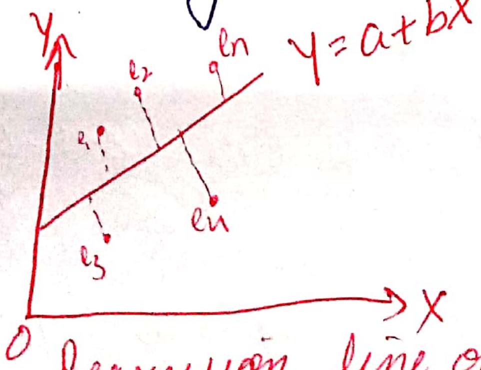
Regression is the estimation or prediction of unknown values of one variable from known values of another variable.

After establishing the fact of correlation b/w two variables, it is natural curiosity to know the extent to which one variable varies in response to a given variation in the other variable. i.e. one is interested to know the nature of relationship b/w the two variables.

Regression lines:- A line of regression is the straight line which gives the best fit in the least square sense to the given frequency.

Regression line of Y on X:-

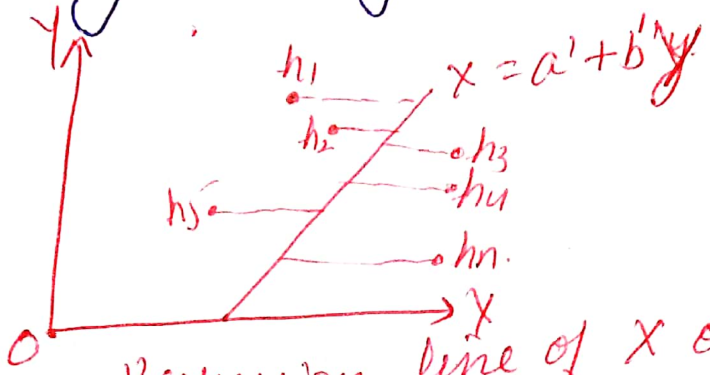
The regression line of Y on X is that line from which we get the best estimated value of Y corresponding to a given value of X.



Regression line of Y on X

Regression line of X on Y :-

The regression line of X on Y is that line from which we get the best estimated value of X corresponding to a given value of Y.



Regression line of X on Y.

Types of Regression Analysis

(1) Simple & Multiple Regression
(2) Total & Partial Regression
(3) Linear & Non-linear Regression.

Linear Regression

When the functional relationship between X & Y is expressed as the first degree equation, it is known as linear regression.

Regression Equation of X on Y is.

$$X = a + bY$$

$$\Sigma X = Na + b\Sigma Y \quad \text{—(1)}$$
$$\Sigma XY = a\Sigma Y + b\Sigma Y^2 \quad \text{—(2)}$$

By method of least square.

Scanned by CamScanner

From 1, we get:

$$a = \frac{\Sigma X}{N} - \frac{b\Sigma Y}{N}$$

$$\Rightarrow \boxed{a = \bar{X} - b\bar{Y}} \quad \text{— } (*)$$

From 2, we get:-

$$b = \frac{\Sigma XY}{\Sigma Y^2} - a\frac{\Sigma Y}{\Sigma Y^2}$$

$$\Rightarrow b = \frac{\Sigma XY}{\Sigma Y^2} - \left(\frac{\Sigma X}{N} - b\frac{\Sigma Y}{N}\right)\frac{\Sigma Y}{\Sigma Y^2}$$

After solving, we get:-

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

where b_{XY} is regression coeff. of X on Y.

Illy regression coeff. of Y on X is given by:

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

Relation b/w coeff. of correlation & coeff. of regression:

$$\boxed{r = \pm\sqrt{b_{XY}\, b_{YX}}}$$

Also regression co-efficient of $\boxed{Y \text{ on } X}$ is given by:

$$\boxed{b_{YX} = r\frac{\sigma_Y}{\sigma_X}}$$

& the line of regression of Y on X is given by:

$$Y - \bar{Y} = r\frac{\sigma_Y}{\sigma_X}(X - \bar{X})$$

Illy regression co-efficient of X on Y is given by:

$$\boxed{b_{XY} = r\frac{\sigma_X}{\sigma_Y}}$$

& the line of regression of X on Y is given by:

$$X - \bar{X} = r\frac{\sigma_X}{\sigma_Y}(Y - \bar{Y})$$

*Remarks:

1. If $r = 0$, lines of regression are mutually perpendicular.
2. If $r = \pm 1$, lines of regression will coincide.

Q1) Calculate the co-efficient of correlation & obtain the least square regression line of Y on X for data:

X:	1	2	3	4	5	6	7	8	9
Y:	9	8	10	12	11	13	14	16	15

Also obtain an estimate of Y which should correspond on the average to X = 6.2

Sol^n

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

$\bar{X} = \frac{\Sigma X}{n} = 5$, $\bar{Y} = \frac{\Sigma Y}{n} = 12$

$$r_{xy} = \frac{\Sigma xy}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}} = \frac{57}{\sqrt{60}\sqrt{60}} = 0.95$$

Equation of regression line of Y on X is:

$$Y - \bar{Y} = \frac{r\,\sigma_Y}{\sigma_X}(X - \bar{X}) \Rightarrow Y - 12 = \frac{(0.95)\sqrt{\frac{60}{9}}}{\sqrt{\frac{60}{9}}}(X - 5)$$

$$\Rightarrow Y - 12 = (0.95)(X - 5)$$

$$\Rightarrow \boxed{Y = 0.95X + 7.25}$$

When X = 6.2, estimated value of Y

$Y = 0.95(6.2) + 7.25$

$\Rightarrow \boxed{Y = 13.14}$

Q2] Obtain two regression coeff, regression equation, & the coeff. of correlation from the following data:

Sol^n

X:	8	1	3	4	7	9	-2
Y:	7	0	4	3	6	8	1

Sol^n $\bar{X} = \frac{\Sigma X}{N} = \frac{30}{7} = 4.3$

$\bar{Y} = \frac{\Sigma Y}{N} = \frac{29}{7} = 4.14$

X	Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	xy	x^2	y^2
8	7	3.7	2.86	10.58	13.69	8.18
1	0	3.3	-4.14	-13.66	10.89	17.14
3	4	$\cancel{0}$.3	0.14	-0.18	1.69	0.02
4	3	-0.3	-1.14	0.34	0.09	1.30
7	6	2.7	1.86	5.02	7.29	3.46
9	8	4.7	3.87	18.19	22.09	14.98
-2	1	-6.3	-3.14	19.78	39.69	9.86
				$\Sigma xy = 40.07$	$\Sigma x^2 = 95.43$	$\Sigma y^2 = 54.94$

Coeff. of correlation is given by:-

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}} = \frac{40.07}{\sqrt{95.43}\sqrt{54.94}} = \frac{40.07}{(9.77)(7.41)} = 0.55$$

$$\boxed{\sigma_Y = \sqrt{\frac{\Sigma x^2}{n}} = 3.69 \qquad \sigma_X = \sqrt{\frac{\Sigma y^2}{n}} = 2.80}$$

regression coeff. of Y on X is given by:-

$$b_{YX} = r\frac{\sigma_Y}{\sigma_X} = \frac{0.55(3.69)}{2.80} = 0.72$$

Regression coeff. of X on Y is given by:-

$$b_{xy} = r\frac{\sigma_x}{\sigma_y} = \frac{(0.55)(2.80)}{3.69} = 0.42$$

Regression eqn of Y on X is given by:-

$$Y - \bar{Y} = b_{yx}(X - \bar{X}) \Rightarrow Y - 4.14 = 0.72(X - 4.3)$$

$$\Rightarrow \boxed{Y = 0.72X + 1.044}$$

Regression eqn of X on Y is given by:-

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

$$\Rightarrow X - 4.3 = 0.42(Y - 4.14)$$

$$\Rightarrow \boxed{X = 0.42Y + 2.5612}$$

Q37:- Obtain two regression coeff, regression eqn & Coeff. of correlation to the following data.

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

Soln Do yourself.

$$\boxed{\begin{array}{l} b_{xy} = 0.8, \; b_{yx} = 0.8, \; r = 0.8 \\ X = 3.6 + 0.8Y, \; Y = 1.8 + 0.8X \end{array}}$$

Scanned by CamScanner

Q4] Calculate coeff. of correlation & obtain regression lines for the following data:

X:	18	19	20	21	22	23	24	25	26	27
Y:	17	17	18	18	18	19	19	20	21	22

Sol Do yourself.

Q5] Find the regression lines:

X:	5	10	15	20	25	30	35	40
Y:	4	7	21	47	53	24	12	6