Module - 4

Correlation And Regression

Definition:

Correlation is a statistical analysis which measures and analysis the degree or extent to which two variables fluctuate with reference to each other.

The correlation expresses the relationship or independence of two sets of variables upon each other.

One variable may be called the subject (independent) and other relative (dependent).

(or)

In a bivariate distribution if the change in one variable effects the change in other variable then the variables are called Correlated.

Types Of Correlation

1) Positive and Negative
2) Simple and Multiple.
3) Partial and Total
4) Linear and Non Linear

Positive and Negative Correlation:

It depends upon the direction of change of Variables

- If two variables tend to move together in same direction ~~th~~ i.e an increase in the value of one variable is accompanied by an increase in the value of other Variable or decrease in the value of one Variable accompanied by a decrease in the Value of other Variable then the correlation is called +ve or direct correlation
- If 2 Variables tend to move together in opp direction so that an increase or decrease in the values of one variable is accompanied by a decrease or increase in the value of the other variable. this correlation is called -ve or inverse Correlation.

Ex: Height & weight, rainfall & yeild of crops, prize & supply. are examples of +ve correlation.

Simple & Multiple:

When we study only 2 variables the relationship is described as simple correlation. But in a multiple correlation we study more than 2 Variables simultaneously.

Ex: Quantity of money & prize level, demand & Prize etc.. are examples of simple correlation. The relationship of prize, demand & supply of a commidity.

Partial & Total:

The study of 2 variables excluding some other variables is called partial correlation.

Ex: we study prize & demand eleminating the supply

- In total correlation all the facts are taken into the account.

Linear & Non Linear:

If the ratio of change b/w 2 Variables is uniform then there will be linear correlation b/w them.

Ex:

A	2	7	12	17
B	3	9	15	21

Here the ratio of change b/w the variables is uniform.

- If we plot the graph we get a straight line.

Non Linear (or) Curve Linear:

If the amount of change in one Variable doesn't bare a constant ratio of amount of change in other ratio is called non linear Correlation.

- The graph of a non linear relationship will be a curve.

Methods Of studing Correlation:

1. Graphic Method
2. Mathematical Method

Graphic Methods are scatter diagram and simple graph.

Mathematical Methods.

1. Karl Pearson's Coefficient Of Correlation:
2. Spearman's rank Correlation:
3. Coefficient Of Concurrent Deviation.

4. Method of List Squares.

Karl Pearsons Coefficient of Correlation:

Coefficient of Correlation:

Correlation is a statistical technique used for analysing the behaviour of ~~to~~ 2 or more variables. Its analysis deals with the association b/w 2 or more variables.

Statistical measures of correlation relates to co variation b/w series but not function or casual relationship.

It is a mathemathical method suggested by Karl Pearson. It is for measuring the magnitude of Linear relationship b/w 2 variables. It is Known as Pearsons. It is denoted by 'r'. This method is ~~to~~ also called product moment correlation coeff. There are several formulas to calculate 'r'.

① $r = \dfrac{\text{covarience of } xy}{\Sigma\, \sigma_x \times \sigma_y} = \dfrac{(\Sigma XY.N) - (\Sigma X \cdot \Sigma Y)}{\sqrt{\Sigma X^2.N - (\Sigma X)^2}}$

② $r = \dfrac{\Sigma xy}{N \sigma_x \sigma_y}$

③ $r = \dfrac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}}$

where $X = x - \bar{X}$ ($\bar{X}$ mean of series of x)
$Y = y - \bar{Y}$ ($\bar{Y}$ mean of series of y)

and σ_x = standard deviation of series x.
σ_y = SD of series y.

Note: We use formula ③ more.

Note: ① The coeff of correlation lies b/w -1 & 1
i.e $(-1 \leq r \leq 1)$

② If $r = 1$ correlation is perfect & +ve

③ If $r = -1$ then correlation is perfect & -ve.

④ If $r = 0$ then there is no relationship b/w the variables.

Problem

① Calculate the coeff of cr correlation for following data

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

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

$$\bar{x} = \frac{12 + 9 + 8 + 10 + 11 + 13 + 7}{7}$$

$= 10$

$$\bar{y} = \frac{14 + 8 + 6 + 9 + 11 + 12 + 13}{7}$$

$= 10.4$.

x	y	X $(x-\bar{x})$	Y $(y-\bar{y})$	X^2	Y^2	XY
12	14	2	3.6	4	12.96	7.2
9	8	-1	-2.4	1	5.76	2.4
8	6	-2	-4.4	4	19.36	8.8
10	9	0	-1.4	0	1.96	0
11	11	1	0.6	1	0.36	0.6
13	12	3	1.6	9	2.56	4.8
7	13	-3	2.6	9	6.76	-7.8

$\Sigma X^2 = 28$ $\Sigma Y^2 = 49.72$ $\Sigma XY = 16$

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}} = 0.43$$

There is a +ve correlation b/w X & Y

② Find if there is any significant correlation b/w hights & weights given below.

height	57	59	62	63	64	65	55	58	57
weight	113	117	126	126	130	129	111	116	112

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

$$\bar{x} = \frac{57+59+62+63+64+65+55+58+57}{9}$$

$= 60$

$$\bar{y} = \frac{113+117+126+126+130+129+111+116+112}{9}$$

$= \cancel{60}\ 120.$

x	y	X	Y	X^2	Y^2	XY
57	113	-3	-7	9	49	21
59	117	-1	-3	1	9	3
62	126	2	6	4	36	12
63	126	3	6	9	36	18
64	130	4	10	16	100	40
65	129	5	9	25	81	45
55	111	-5	-9	25	81	45
58	116	-2	-4	4	16	8
57	112	-3	-8	9	64	24
				102	472	216

$$r = \frac{\sum XY}{\sqrt{\sum X^2 \cdot \sum Y^2}}$$

$$= \frac{216}{\sqrt{102 \times 472}} = 0.98$$

3) Calculate quotient of correlation b/w age of car and annual maintanence cars & comment.

Age of cars (x) (years)	2	4	6	7	8	10	12
Maintanance cost (y)	1600	1500	1800	1900	1700	2100 ~~200~~	2000

Sol: $X = x - \bar{x}$ $Y = y - \bar{y}$

$\bar{X} = \frac{49}{7} = 7$ $\bar{y} = 1800$

x	y	X	Y	XY	X^2	Y^2
2	1600	-5	-200	1000	25	40000
4	1500	-3	-300	900	9	90000
6	1800	-1	0	0	1	0
7	1900	0	100	0	0	10000
8	1700	1	-100	-100	1	10000
10	2100	3	300	900	9	90000
12	2000	+5	200	+1000	25	40000
				3700 ~~100~~ ~~1900~~	70	280000

$$r = \frac{\sum XY}{\sqrt{\sum X^2 \sum Y^2}}$$

$$= \frac{3700}{\sqrt{70(280000)}} = 0.835$$

Comment:

We can observe that there is a high degree of +ve corellation b/w age of cost & annual maintanaince.

Q) Byscological tests of integlegence & engineering ability were applied to 10 students here is a record of ungrouped data showing intellengence ratio (IR) & engineering ability ratio (ER). Calculate coeff of correlation

Student	A	B	C	D	E	F	G	H	I	J
IR	105	104	102	101	100	99	98	96	93	92
ER	101	103	100	98	95	96	104	92	97	94

Soli: $X = x - \bar{x}$, $Y = y - \bar{y}$ $\bar{x} = 99$, $\bar{y} = 98$

x	y	~~XY~~	X	Y	XY	X^2	Y^2
105	101	~~6 106~~	6	3	18	36	9
104	103		5	5	25	25	25
102	100		3	2	6	9	4
101	98		2	0	0	4	0
100	95		1	-3	-3	1	9
99	96		0	-2	0	0	4
98	104		-1	6	-6	1	36
96	92		-3	-6	18	9	36
93	97		-6	-1	6	36	1
92	94		-7	-4	28	49	16
					92	170	140

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}} = \frac{92}{\sqrt{170(140)}}$$

$$= 0.596$$

When deviations are taken from an assumed mean:

When actual mean is not a whole number, but a fraction or when the series is large the calculation by direct method will involve a lot of time to avoid such situation we can use the assumed mean method.

$$r = \frac{\sum XY - \frac{\sum X . \sum Y}{N}}{\sqrt{\left[\sum X^2 - \frac{(\sum X)^2}{N}\right]\left[\sum Y^2 - \frac{(\sum Y)^2}{N}\right]}}$$

where X = deviation of items of x series from an assumed mean i.e

$$X = x - A$$

$$\text{Ily, } Y = y - A$$

$$N = \text{no. of items.}$$

Q)

X	28	41	40	38	35	33	40	32	36	33
Y	23	34	33	34	30	26	28	31	36	38

Sol:

$$\bar{x} = \frac{\sum x}{N} = 35.6 \simeq$$

$$\bar{y} = \frac{\sum y}{N} = 31.3 \simeq 31$$

Q) Find a suitable coeff of correlation for following data.

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

$\bar{x} = \frac{\Sigma x}{N} = 29$

$\bar{y} = \frac{\Sigma y}{N} = 117.25 \simeq 117$

x	y	X	Y	XY	X^2	Y^2
15	85	-14	-32	448	196	1024
18	93	-11	-24	264	121	576
20	95	-9	-22	198	81	484
24	105	-5	-12	60	25	144
30	120	1	3	3	1	9
35	130	6	13	78	36	169
40	150	11	33	363	121	1089
50	160	21	43	903	441	1849
		$\boxed{0}$	$\boxed{2}$	$\boxed{2317}$	$\boxed{1022}$	$\boxed{5344}$

$$r = \frac{\Sigma XY - \frac{\Sigma X . \Sigma Y}{N}}{\sqrt{\left(\Sigma X^2 - \frac{(\Sigma X)^2}{N}\right)\left(\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}\right)}} = \frac{2317 - \frac{0(2)}{8}}{\sqrt{\left(1022 - \frac{(0)^2}{8}\right)\left(5344 - \frac{(2)^2}{8}\right)}}$$

$$= \frac{2317}{\sqrt{(1022)(5343.5)}} = 0.99$$

Rank Correlation Coefficient.

A british psychologist spheremen found out the method of finding coeff of correlation by ranks. This method is based on rank and is useful in dealing with ~~polytative~~ qualitative characteristics such as molarity, character and intellegence. It cannot be measure quantitatively as in the case of pearson's coeff.

It is based on ranks given to the observation. Rank correlation is applicable only to the individual observations.

The formula for the spheremans rank correlation is given by

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

where ~~rows~~ ρ = rank correlation coeff.

D = difference of 2 ranks.

N = no. of ~~paid~~ paired observations

Properties Of Ranks Correlation Coeff

1) The value of ρ lies b/w 1 & −1
2) If $\rho = 1$ there is complete agreement in the order if the ranks and the direction of the rank is same.
3) If $\rho = -1$ then there is complete disagreement in the order of the rank and they are in opp direction.

Procedure:

When ranks are given we can calculate the difference of 2 ranks (D) and obtain the

rank correlation coeff by using formula
- When ranks are not given we must give the ranks to the given observations.
- We can give ranks by taking heighest as 1 or lowest as 1 and we can follow the order to the next observations.

Q) Following are the rank obtained by 10 students in the subjects statistics and mathematics to what extent the knowledge of the student in the subject is related.

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

Sol:

x	y	$D = x - y$	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
			40

$$P = 1 - \frac{6\Sigma D^2}{N(N^2-1)} \Rightarrow 1 - \frac{6(40)}{10((10)^2-1))}$$

$$= 0.76$$

Q) The ranks of 16 students in mathematics & 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 coeff for this grp in maths & statistics.

Sol:

x	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
			136

$$P = 1 - \frac{6(136)}{16(16^2 - 1)}$$

$$= 0.8$$

Q) A random sample of 5 college students is selected and their grades in maths & statistics are found to be.

	1	2	3	4	5
maths	85	60	73	40	90
statistics	93	75	65	50	80

8/ Calculate Spearman rank correlation coeff.

Sol:

marks in maths	Ranks (x)	marks in statistics	Ranks (y)	D = x − y	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
					4

$$P = 1 - \frac{6(4)}{5(5^2-1)}$$

$$= 0.8$$

Q) 10 competitors in a musical test were ranked by 3 judges A, B, C in the following order.

Ranks by (A)	1	6	5	10	3	2	4	9	7	8
Ranks by (B)	3	5	8	4	7	10	2	1	6	9
Ranks by (C)	6	4	9	8	1	2	3	10	5	7

using rank correlation method discuss which pair of judges has the nearest approach for common liking in music

Sol:

x (A)	(y) B	(z) C	$D_1 = x-y$	D_1^2	$D_2 = y-z$	D_2^2	$D_3 = x-z$	D_3^2
1	3	6	-2	4	-3	9	-5	25
6	5	4	1	1	1	1	2	4
5	8	9	-3	9	-1	1	-4	16
10	4	8	6	36	-4	16	2	4
3	7	1	-4	16	6	36	2	4
2	10	2	-8	64	8	64	0	0
4	2	3	2	4	-1	1	1	1
9	1	10	8	64	-9	81	1	1
7	6	5	1	1	1	1	2	4
8	9	7	-1	1	2	4	1	1
				200		214		60

$$P_1 = 1 - \frac{6(200)}{10(10^2-1)} = -0.21$$

$$P_2 = 1 - \frac{6(214)}{10(10^2-1)} = -0.29$$

$$P_3 = 1 - \frac{6(60)}{10(10^2-1)} = 0.63$$

P_3 getting nearest approach

Equal or Repeated Rank

If any two or more person getting more than one item with the same value then spheremans formula for calculating the rank correlation coeff brakes down.

In this case common ranks are given to repeated items. The common rank is the avg of the ranks which these items would have assumed, if they were diff from each other & the next item will get the rank next to ranks already assumed.

Example: If 2 individuals are placed in the 7th place each of them are given the rank 7.5. and the next rank will be 9. similarly if 3 persons are ranked equal at the 7th place then they are given the rank 8. which is common rank, assign to each and the next rank will be 10.

the formula for calculating (repeated rank) Rank correlation coeff.

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

where m = the no. of items whose ranks are common.

Q) From the following data Calculate the ranks correlation coeff.

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

Sol:

X	Ranks of (x)	Y	Ranks of (y)	D = x − y	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	0.5	0.25
57	9	9	8	1	1

$\Sigma D^2 = 41$

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

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

In x series 16 is repeated 3 times
$m = 3$.

y series 6 is repeated 2 times $m = 2$
and 13 is repeated 2 times $m = 2$

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

$= 0.733$

Q) Obtain the rank correlation coeff 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:

X	Ranks of (x)	Y	Rank of (y)	D = x - y	D^2
68	7	62	6	1	1
64	5	58	4	1	1
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
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

$\sum D^2 = 72$

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

$= 0.545$ //

Q) A sample of 12 fathers and their elder son gave the following data. Calculate the coeff of rank correlation.

x	65	63	67	64	68	62	70	66	68	67	69	71
y	68	66	68	65	69	66	68	65	71	67	68	70

Sol:

x	Ranks (x)	y	Ranks (y)	$D = x - y$	D^2
65	4	68	7.5	-3.5	12.25
63	2	66	3.5	-1.5	2.25
67	6.5	68	7.5	-1	1
64	3	65	1.5	1.5	2.25
68	8.5	69	10	-1.5	2.25
62	1	66	3.5	-2.5	6.25
70	11	68	7.5	3.5	12.25
66	5	65	1.5	3.5	12.25
68	8.5	71	12	-3.5	12.25
67	6.5	67	5	1.5	2.25
69	10	68	7.5	2.5	6.25
71	12	70	11	1	1
					72.5

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

$= 0.722$ //

Scanned with ACE Scanner

Q) The ranks of 15 students in 2 sub A & B are given below the 2 numbers within the brackets denoting the ranks of students 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 spheremans formula to calculate the ranks correlation coeff

Sol:

x	y	D	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
			272

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

$$= 1 - \frac{6(272)}{15(15^2-1)}$$

$$= 0.514$$

Q. S.T the max value of rank correlation coff is 1 or the coff of correlation lies b/w −1 & 1. Let x and y be deviations of X & Y series from their mean. Let Σx and Σy be their respective S.D.

Let $\Sigma\left[\frac{x}{\sigma_x} + \frac{y}{\sigma_y}\right]^2 = \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} + 2\frac{xy}{\sigma_{xy}}$

$= \frac{\Sigma x^2}{\sigma x^2} + \frac{\Sigma y^2}{\sigma y^2} + 2\frac{\Sigma xy}{\sigma_{xy}}$

but $\frac{\Sigma x^2}{\sigma x^2} = N$

$= 2N + 2rN + 2N(1+r)$

$\frac{\Sigma y^2}{\sigma y^2} = N$

$r = \frac{\Sigma xy}{N.\sigma x \sigma y}$

but $\left(\frac{x}{\sigma_x} + \frac{y}{\sigma_y}\right)^2$ is the sum of the squares of real quantities and as such it cannot be negative so atmost it can. be zero. $1 + r \geq 0$

$r \leq -1$

Hence r cannot be less than −1 at most it can be −1 similarly by expanding $\Sigma\left[\frac{x}{\sigma x} - \frac{y}{\sigma y}\right]$ it can be S.T $\Sigma\left(\frac{x}{\sigma x} - \frac{y}{\sigma y}\right)^2 = 2N(1-r)$

this again cannot be negative at most it can be zero $2N(1-r) \geq 0$

$1 - r \geq 0$

$r \leq 1$

Hence $-1 \leq r \leq 1$

Scanned with ACE Scanner

Regression:

In regression we can estimate the value of one variable with the value of other variable which is known.

The statistical method which helps us to estimate the unknown value of 1 variable from the known value of the related variable is called Regression.

The line described in the avg relationship b/w 2 variables is known as line of regression.

Now a days we are using the term estimating line instead of regression line.

The regression establishes a functional relation b/w dependent & independent variables.

In regression x is a random variable and y is fixed variable.

Regression analysis is a mathematical measure method of avg relationship b/w 2 or more variables in terms of the original units of the data.

The variable whose value is influenced or is to be predicted is called dependent variable and the variable which influences the value or is used for the prediction is called independent variable.

Q) A ~~parallel~~ Panel of 2 judges P and Q graded 7 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

~~marks to~~ the 8 performance which judge Q would not attend was awarded 37 marks by judge P. If judge Q has also be in present how many marks would be expected to have been awarded by him to the 8 performance.

Sol: given:

marks given by ~~ja~~ judges P and Q to 7 performances. we have to calculate marks given by Q (y) when marks given by P (x) = 37) per 8 performances.

$y = ?$ when $x = 37$

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

$\bar{x} = \frac{\Sigma x}{N} = 43$, $\bar{y} = \frac{\Sigma y}{N} = 38$.

x	y	$X = x - \bar{x}$	$Y = y - \bar{y}$	X^2	XY
46	40	3	2	9	6
42	38	-1	0	1	0
44	36	1	-2	1	-2
40	35	-3	-3	9	9
43	39	0	1	0	0
41	37	-2	-1	4	2
45	41	2	3	4	6
				28	21

$b_{yx} = r \cdot \frac{\sigma_y}{\sigma_x} = \frac{\Sigma xy}{\Sigma x^2} = 0.75$ $\quad y = Y - \bar{Y} = 38$

$y - \bar{y} = 0.75(37 - 43)$ $\quad Y = 38 - 4.5$

$= -4.5$ $\quad = 33.5$

Hence if Q would attend he would have awarded marks 33.5 for 8th performance

Q) If θ is the angle b/w 2 regression lines and standard deviation of y is twice the SD of x and ~~SD~~ $r = 0.25$ find $\tan\theta$.

given : $r = 0.25$

~~SD~~ of y is

$\sigma_y = 2\sigma_x$

if θ is the angle b/w 2 regression lines.

then $\tan\theta = \frac{1 - r^2}{r} \cdot \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}$

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

$= 3.75 \left(\frac{2\cancel{\sigma_x^2}}{5\cancel{\sigma_x^2}} \right) = 1.5$

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

given : $\sigma_x = \sigma_y = \sigma$

$\theta = \tan^{-1}\left(\frac{4}{3}\right)$

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

$\frac{4}{3} = \frac{1 - r^2}{r} \left(\frac{\sigma^2}{\sigma^2 + \sigma^2} \right)$

$\frac{4}{3} = \frac{1 - r^2}{r} \left(\frac{\cancel{\sigma^2}}{2\cancel{\sigma^2}} \right)$

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

$$8r = 3 - 3r^2.$$

$$3r^2 + 8r - 3 = 0$$

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