Correlation:

Correlation is the degree of the relationship between two or more variables It does not explain the cause behind the relationship.

eg: price and supply, demand and supply income and expenditure are correlated.

Types of correlation

Positive correlation: Correlation is said to be positive when two variables moves in same direction.

Ex.1 If income and expenditure of a Household may be increasing or decreasing simultaneously If so, there is positive correlation.

Ex.2 $y = a + bx$.

Negative correlation: Corrdation is said to be negative when two variables move in opposite direction.

eg Determinants of Quantity demanded.

$Qd = f(P$

Ex-1 Price and demand for a commodity move in the opposite direction.

Ex 2 $y = a - bx$.

Type II Based on number of variable.

i) Simple correlation.

ii) Multiple correlation.

iii) Partial correlation.

Simple correlation :

When two variables are studied it as simple correlation.

Multiple correlation :

When atleast these variables are studied and their relationships are simultaneously worked out it is a case of multiple correlation.

eg :- Determinants of Quantity demanded.

$Qd = f(P, Pc, Ps, t, y)$

where Qd stands for Quantity demanded f stands for function P is the price of the goods, Pc is the price of competitive goods Ps is the price of substituting goods, t is the taste and preference, y is the income

Partial correlation :

When more than two variables are studied keeping the other variables constant, it is called Partial correlation.

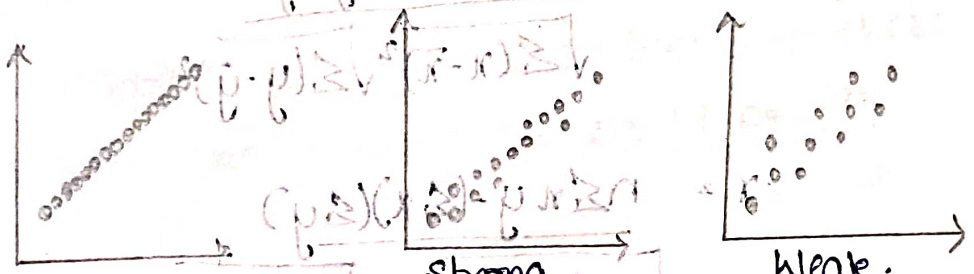
Type III Based on change in proportion.

linear correlation :- correlation is said to be linear when one variable move with the other variable in fixed proportion.

Non linear correlation :- correlation is said to be non linear when one variable move with the other variable in changing proportion.

eg : $y = a + bx^2$.

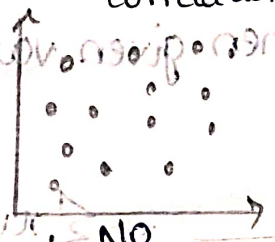
Scale of correlation coefficient	value.
$0 < r \leq 0.19$	Very low correlation.
$0.2 \leq r \leq 0.39$	low correlation.
$0.4 \leq r \leq 0.59$	moderate correlation.
$0.6 \leq r \leq 0.79$	High correlation
$0.8 \leq r \leq 1.0$	Very high correlation.



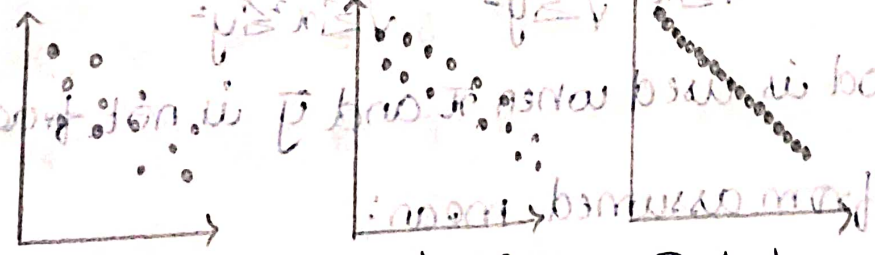
Perfect Positive correlation

Strong Positive correlation.

Weak. Positive. correlation.



No correlation.



Weak negative correlation

Strong Negative correlation

Perfect Negative correlation.

Karl person's coefficient correlation is a ... Popular method of calculating correlation. Arthemetic mean and standard deviation are the basis for it calculation. The correlation coeff (r) also called as the linear correlation coefficient measures the strength and direction of a linear relationship between two variables The value of r lies between -1 and 1.

Deviation from Actual mean

Indirect method: $$r = \frac{\Sigma(x-\bar{x})\Sigma(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\sqrt{\Sigma(y-\bar{y})^2}}$$

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

(This method is used when given variable are small magnitude)

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

(This method is used when $\bar{x}$ and $\bar{y}$ is not fraction)

Deviation from assumed mean:

Direct method

$$r = \frac{n\Sigma d_x d_y - (\Sigma d_x)(\Sigma d_y)}{\sqrt{n\Sigma d_x^2 - (\Sigma d_x)^2}\sqrt{n\Sigma d_x^2 - \Sigma d_y)^2}}$$

(This method is used when $\bar{x}$ and $\bar{y}$ is fraction or

Karl person's coefficient of correlation:

coefficient correlation is a mathematical and most popular method of calculating correlation. Arthemetic mean and standard deviation are the basis for it calculation. The correlation coefficient (r) also called as the linear correlation coefficient measures the strength and direction of a linear relationship between two variables. The value of r lies between -1 and 1.

Deviation from Actual mean

Indirect method: $$r = \frac{\Sigma(x-\bar{x})\Sigma(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\sqrt{\Sigma(y-\bar{y})^2}}$$

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

(This method is used when given variable are small magnitude)

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

(This method is used when $\bar{x}$ and $\bar{y}$ is not fraction)

Deviation from assumed mean:

Direct method

$$r = \frac{n\Sigma d_x d_y - (\Sigma d_x)(\Sigma d_y)}{\sqrt{n\Sigma d_x^2 - (\Sigma d_x)^2}\sqrt{n\Sigma d_y^2 - (\Sigma d_y)^2}}$$

(This method is used when $\bar{x}$ and $\bar{y}$ is fraction or the

series is large)

$d_x = x - A$, $d_y = y - B$ where A, B are constants.

n = number of pair of observations

$\sum d_x$ = Sum of deviation of x

$\sum dy$ = sum of deviation of y

$\sum d_x^2$ = sum of deviation. Square of x.

$\sum d_y^2$ = sum of deviation square of y.

$\sum d_x dy$ = sum of product of deviation x and y.

The pearson correlation is the most used measuremen for a linear relationship between two variables. The stronger the correlation between these two data sets, the closerit "lbeto +1 or -1".

Calculate the KPCC. from the following data.

x	2	4	5	6	8	11
y	18	12	10	8	7	5

x	y	x^2	y^2	xy.
2	18	4	324	36
4	12	16	144	48.
5	10	25	100	50.
6	8	36	64	48
8	7	64	49	56
11	5	121	25	55
$\sum x = 36$	$\sum y = 60$	$\sum x^2 = 266$	$\sum y^2 = 706$	$\sum xy = 293$

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

$$\frac{6(298) - 6(36)(60)}{\sqrt{6(266) - (36)^2}\cdot\sqrt{6(706) - (60)^2}}$$

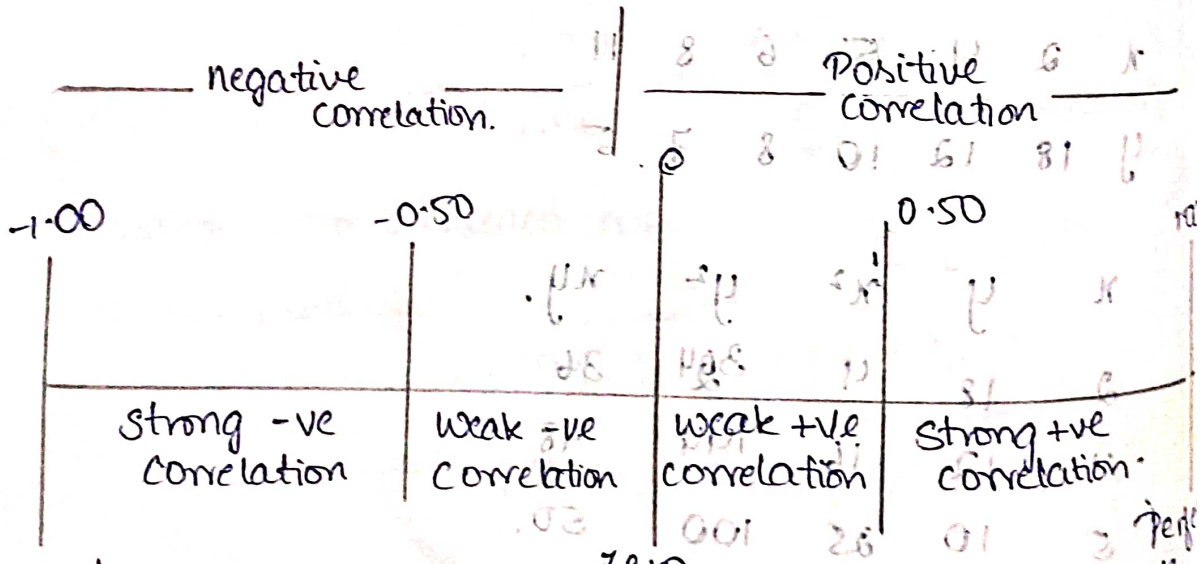
$$\frac{1758 - 2160}{\sqrt{1596 - 1296}\cdot\sqrt{4236 - 3600}}$$

$$\frac{1758 - 2160}{\sqrt{300}\cdot\sqrt{636}}$$

$$\frac{-402}{(17.38)(25.2)} \approx -0.92$$

$$\frac{-402}{680} = -0.59$$

Note:-



Karl Pearson correlation coefficient b/w Age and weights.

Age	1	2	3	4	5
weight	3	4	6	7	12.

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
$\Sigma x = 15$	$\Sigma y = 32$	$\Sigma x^2 = 55$	$\Sigma y^2 = 254$	$\Sigma xy = 117$.

n = 5

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

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

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

$$\frac{105}{\sqrt{50}\cdot\sqrt{246}}$$

$$\frac{95}{(7.07)(15.68)}$$

$$\frac{105}{110.85} = 0.95 \Rightarrow \text{shows +ve correlation}$$

Calculate correlation coefficient b/w death and birth rate for the data.

Birth rate	Death	$x = x_i - \bar{x}$ $x_i - 30$	$y = y_i - \bar{y}$ $\bar{y} = 22$	x^2	y^2	xy
24	15	-6	-7	36	49	42
26	20	-4	-2	16	4	8
32	22	2	0	4	0	0
33	24	3	2	9	4	6
35	27	5	5	25	25	25
30	24	0	2	0	4	0
				90	86	81

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2} \cdot \sqrt{\Sigma y^2}} = \frac{r = 81}{\sqrt{90} \cdot \sqrt{86}}$$

$$\frac{81}{9.4 \times 9.2}$$

$$\frac{81}{86.48} = 0.93.$$

Problem: Psychological tests of intelligence and business analytics ability were applied to 10 students. There is a record of ungrouped data showing intelligence Ratio (IR) and business analytics Ability Ratio (BAAR). Calculate the coefficient of correlation.

Students	IR (x_i)	BAAR (y_i)	$\bar{x} = x_i - \bar{x} = x_i - 99$	$y = y - \bar{y}_1$ $y_i - 98$	x^2	y^2	xy
A	105	101	6	3	36	9	18
B	104	103	5	5	25	25	25
C	102	100	3	2	9	4	6
D	101	98	2	0	4	0	0
E	100	95	1	-3	1	9	-3
F	99	96	0	-2	0	4	0
G	98	104	-1	+6	1	36	-6
H	96	92	-3	-6	9	36	18
I	93	97	-6	-1	36	1	6
J	92	94	-7	-4	49	16	28
					170	140	92

$$\bar{x} = \frac{\sum x_i}{n} = \frac{990}{10} \Rightarrow \bar{x} = 99.$$

$$\bar{y} = \frac{980}{10} \Rightarrow \bar{y} = 98.$$

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \cdot \sum y^2}} = \frac{92}{\sqrt{170 \cdot 140}} = \frac{92}{154} = 0.59$$

Problem
between
Height x_i
57
59
62
63
64
65
55
58
57
$\bar{x} =$
$\bar{y} =$
Σ

Problem: Psychological tests of intelligence and business analytics ability were applied to 10 students. There is a record of ungrouped data showing Intelligence Ratio (IR) and business analytics Ability Ratio (BAAR). Calculate the coefficient of correlation.

Students	IR (x_i)	BAAR (y_i)	$\bar{x} = x_i - \bar{x}$ $= x_i - 99$	$y = y_i - \bar{y}$ $y_i - 98$	x^2	y^2	xy
A	105	101	6	3	36	9	18
B	104	103	5	5	25	25	25
C	102	100	3	2	9	4	6
D	101	98	2	0	4	0	0
E	100	95	1	-3	1	9	-3
F	99	96	0	-2	0	4	0
G	98	104	-1	+6	1	36	-6
H	96	92	-3	-6	9	36	18
I	93	97	-6	-1	36	1	6
J	92	94	-7	-4	49	16	28
					170	140	93

$$\bar{x} = \frac{\sum x_i}{n} = \frac{990}{10} \Rightarrow \bar{x} = 99.$$

$$\bar{y} = \frac{980}{10} \Rightarrow \bar{y} = 98.$$

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \cdot \sum y^2}} = \frac{92}{\sqrt{170 \cdot 140}} = \frac{92}{154} = 0.59$$

Problem Find if there is any significant correlation between heights and weights for following data.

Height x_i	weight y_i	$x_i - \bar{x}$ $x_i - 60$	$y_i - \bar{y}$ $y_i - 120$	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

$$\bar{x} = \frac{\sum x_i}{n} = \frac{540}{90} = 60$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{1080}{90} = 120$$

$$\sum xy = \frac{216}{\sqrt{102 \cdot 472}} = \frac{216}{\sqrt{48144}} = \frac{216}{219.41} = 0.98$$

Spearman's rank correlation coeffi...

Ranks are not equal :

$$R = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

where $d = R_1 - R_2$

d is difference in a pair of ranks.

R is rank correlation coefficient

R_1 is rank of observations with respect to first variable.

R_2 is rank of observations with respect to second variable.

n is the number of paired observations.

Ranks are equal :

$$R = 1 - \frac{6\left[\sum d^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \frac{1}{12}(m_{1,2}^3 - m_{1,2})\right]}{n(n^2-1)}$$

Calculate SRCC for the following data.

Ranking economics	Ranking statistics	$d = R_1 - R_2$	d^2
1	4	-3	9
2	8	-6	36
3	2	1	1
4	3	1	1
5	5	0	0
6	7	-1	1
7	6	1	1
8	9	-1	1
9	10	-1	1
10	1	9	81
			132

$n = 10$

$\sum d = 132$

$$R = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

$$1 - \frac{6 \times 132}{10(100-1)}$$

$$1 - \frac{792}{990}$$

$$1 - 0.8 = 0.2$$

$R = 0.2$

Calculate.

x	17	13	15	16	6	11	14	9	7	12
y	36	46	35	24	12	18	27	22	2	8.

x	y	R_1	R_2	$d = R_1 - R_2$	d^2
17	36	1	2.	-1	1
13	46	5	1	4	16
15	35	3	3	0	0
16	24	2	5	-3	9
6	12	10	8.	2	4
11	18	7	7	0	0
14	27	4	4	0	0
9	22	8	6	2	4
7	2	9	10	-1	1
12.	8.	6	9.	-3	9

$n = 10$

$\sum d^2 = 44.$

$$R = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

$$1 - \frac{6 \times 44}{10(100-1)}$$

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

$$1 - 0.26$$

$$R = 0.74$$

(9)

x	40	42	35	45	30	50
y	110	120	125	130	128	160.

x	y	R_1	R_2	$d = R_1 - R_2$	d^2
40	110	4	6	-2	4
42	120	3	5	-2	4
35	125	5	4	1	1
45	130	2	2	0	0
30	128	6	3	3	9
50	160	1	1	0	0

$n = 6$

$\sum d^2 = 18$

$$R = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

$$1 - \frac{6 \times 18}{6(36-1)}$$

$$1 - \frac{108}{210}$$

$$= 1 - 0.51$$

$$= 0.49.$$

Calculate between 2 variables. x and y.

x	y	R_1	R_2	$d = R_1 - R_2$	d^2
68	62	4	5	1	1
	58	6	7	1	1
64	68	2.5	3.5	-1	1
75	45	9	10	1	1
50	81	6	1	5	25
64	60	1	6	5	25
80	68	2.5	3.5	-1	1
75	48	10	9	1	1
40	50	8	8	0	0
55	70	6	2	4	16
64					72

$m_1 = 2$

$m_2 = 3$

$m_3 = 2$

$$1 - \frac{6\left[72 + \frac{1}{12}(8-2) + \frac{1}{12}(27-3) + \frac{1}{12}(8-2)\right]}{100(100-1)}$$

$$1 - \frac{6\left[72 + \frac{6}{12} + \frac{24}{12} + \frac{6}{12}\right]}{990}$$

$$1 - \frac{6[75]}{990} \quad 1 - \frac{6(72)}{990}$$

$$1 - 0.43$$

$$= 0.57$$

Problem : The ranks of 15 students in two subjects A and B are given below. The two numbers within brackets denote the ranks of a student in A and B.

(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)

Find Spearman's rank correlation coefficient.

Rank A (R_1)	Rank (B) 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
			$\Sigma d^2 = 272$

$$R = \frac{1 - 6\Sigma d^2}{n(n^2-1)}$$

$$1 - \frac{6 \times 272}{15(225-1)} = 1 - \frac{6 \times 272}{15(224)} = 1 - \frac{1632}{3360} = 1 - 0.48$$

$$R = 0.52$$

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

Use the rank correlation coefficient to determine which pair of judges has the nearest approach for judgement of beauties.

	Judge (R_1)										Judge (R_2)										Judge (R_3)									
d_3^2	25	4	16	4	4	0	1	1	4	1.	$\sum d_3^2$ = 60.																			
d_3, R_3	5	-2	4	-2	-2	0	1	-1	-2	1																				
d_2^2	9	1	1	16	36	64	1	81	1	49	$\sum d_{1,2}^2$ = 274																			
$d_2^2 - R_2 - R_3$	-3	1	-1	-4	6	8	-1	9	-1	7																				
d_1^2	4	1	9	36	16	64	4	64	1	1	$\sum d_1^2$ = 200																			
$d_1 - R_1 - R_2$	-2	1	-3	6	-4	-8	2	8	-1	-1																				
Judge 3.	6	4	9	8	1	2	3	10	5	7.																				
Judge 2	3	5	8	4	7	10	2	1	6	9.																				
Judge 1	1	6	5	10	3	2	4	9	7	8.																				

n = 10

$$R_{12} = 1 - \frac{6\sum d_1^2}{n(n^2-1)}$$

$$R_{12} = 1 - \frac{6(200)}{10(99)}$$

$$R_{12} = 0.21.$$

$$R_{23} = 1 - \frac{6\sum d_2^2}{n(n^2-1)}$$

$$1 - \frac{6 \times 214}{990}$$

$$R_{23} = -0.29.$$

$$R_{31} = 1 - \frac{6\sum d_3^2}{n(n^2-1)}$$

$$1 - \frac{6(60)}{990} \Rightarrow R_{31} = 0.64.$$

∴ The correlation coefficient $R_{31} = 0.64$ is highest, the judges 3 and 1 have nearest approach for judgement of Beauties.

Find the RCC from the following data

marks in calculus	15	20	28	12	40	60	20	80.
mark in statistics	40	30	50	30	20	10	30	60.

marks in calculus.	Mark in Statistics	R_1	R_2.	$d = R_1 - R_2$	d^2
15	40	7	3	4	16
20	30	5.5	.5	0.5	0.25
28	50	4	2	2.	4
12	30	8	5	3.	9
40	20	3	7	-4	16
60	10	2.	8.	-6	36
20	30	5.5	.5	0.5	0.25
80.	60.	1	1	0.	0.

$\Sigma d^2 = 81.5$

$m_1 = 2.$

$m_2 = 3$

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

$$1 - 6\left[\frac{81.5 + \frac{1}{12}(8-2) + \frac{1}{12}(27-3)}{8(64-1)}\right]$$

$$1 - \frac{6[81.5 + 0.5 + 2]}{8(64-1)} = \frac{1 - 6[84]}{8(63)} = 1 - \frac{504}{504.}$$

$1 - 1 = 0$

16) Calculate RCC. after making adjustment at tied rank.

x	y	R_1	R_2	d	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	14	2	3	-1	1
					41

$m_1 = 3$

$\frac{7+8+9}{3} = \frac{24}{3} = 8$

$m_2 = 2$

$\frac{5+6}{2} = \frac{11}{2} = 5.5$

$m_3 = 2$

$\frac{8+9}{2} = \frac{17}{2} = 8.5$

$$R = 1 - 6\left[\sum d^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \frac{1}{12}(m_3^3 - m_3)\right] / n(n^2-1)$$

$$1 - 6\left[41 + \frac{24}{12} + \frac{6}{12} + \frac{6}{12}\right] / 990$$

$$= 1 - 6\left[41 + 2 + 0.5 + 0.5\right] / 990 = 1 - 6[44]/990$$

$$1 - \frac{264}{990} = 1 - 0.26$$

$$= 0.74.$$

Regression Analysis

Regression is the measure of the average relationship between two or more variable in terms of the original units of the data. It is a statistical tools with the help of which the unknown values of one variable can be estimated from known values of another variable.

Kinds of regression may be studied on the basis of.

Changes in Proportions:-

linear regression:- Regression is said to be linear when one variable move with other variable in fixed proportion.

Non linear regression:- Regression is said to be non linear when one variable move with other variable in changing proportions.

Number of variation.

Simple regression: when only two variables are studied it is a single regression.

eg:- $y = a + bx + \varepsilon$.

where y = Dependent variable.

x = Independent (explanatory) variable.

a = Intercept.

b = slope.

ε = Error term.

Multiple regression: When atleast three variables are studied and their relationship are simultaneously worked out, it is a case of multiple regression.

$$y = a + bx_1 + cx_2 + dx_3 + \varepsilon$$

where y = Dependent variable.

x_1, x_2, x_3 = Independent (explanatory) variable.

a = Intercept.

b, c, d = slope.

ε = Error term.

Partial regression:

When more than two variables are studied. keeping the other variables constant, it is called Partial regression.

The methods of least squares:

$$y = a + bx.$$

where y = Dependent variable.

x = Independent (explanatory) variable.

a = Intercept.

b = slope.

regression equation of on x $\quad \Sigma y = na + b\Sigma x.$

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

Regression equation x on y : $\Sigma x = na + b\Sigma y$.

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

Short cut method :

$$S_{xy} = \Sigma xy - \frac{\Sigma x \Sigma y}{n}.$$

$$S_{xx} = \Sigma x^2 - \frac{(\Sigma x)^2}{n}.$$

$$b = \frac{S_{xy}}{S_{xx}}$$

$$a = \bar{y} - b\bar{x}.$$

The regression equation is $y = a + bx$.

Deviations taken from Assumed mean values of x and y :

Regression equation x and $\bar{y}$: $x - \bar{x} = b_{xy}(y - \bar{y})$

where $b_{xy} = \frac{n\Sigma dxdy - (\Sigma dx)(\Sigma dy)}{n\Sigma dy^2 - (\Sigma dy)^2}$.

n = number of observations.

$dx = x - A$, A is assumed mean of x.

$dy = y - B$, B is assumed mean of y.

Regression equation y on x : $y - \bar{y} = b_{yx}(x - \bar{x})$.

where $b_{yx} = \frac{n\Sigma dxdy - (\Sigma dx)(\Sigma dy)}{n\Sigma dx^2 - (\Sigma dx)^2}$.

n = number of observation of x.

$dx = x - A$, A is assumed mean of x

$dy = y - B$, B is assumed mean of y.

Deviation taken from actual mean values of x. and y.

Regression equation x and on y : $x - \bar{x} = b_{xy}$

where $b_{xy} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(y-\bar{y})^2}$.

Regression equation y on x : $y - \bar{y} = b_{yx}(x - \bar{x})$

where $b_{yx} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$

Regression coefficient in term of correlation coefficient :

Regression equation x on y : $x - \bar{x} = r\frac{\sigma_x}{\sigma_y}(y-\bar{y})$

where $\bar{x}$ = mean of x , $\bar{y}$ = mean of y.

r = correlation between the x and y variables.

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

The correlation coefficient is the geometric mean of two regression coefficients $r = \sqrt{b_{xy} \times b_{yx}}$.

From the following data find two regression equations.

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

Sol

x	y	x^2	y^2	xy
1	2	1	4	2
2	4	4	16	8
3	7	9	49	21
4	6	16	36	24
5	5	25	25	25
6	6	36	36	36
7	5	49	25	35
$\Sigma x = 28$	$\Sigma y = 35$	$\Sigma x^2 = 140$	$\Sigma y^2 = 191$	$\Sigma xy = 151$

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

$$\frac{7(151)-(28)(35)}{7(191)-(35)^2}$$

$$= \frac{1057-980}{1337-1225} = \frac{77}{112} = 0.69 = b_{xy}$$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{28}{7} = 4$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{35}{7} = 5$$

Regression equation x on y

$$x - \bar{x} = b_{xy}(y - \bar{y})$$
$$x - 4 = 0.69(y - 5)$$
$$x - 4 = 0.69y - 3.45$$
$$x = 0.69y - 3.45 + 4$$
$$x = 0.69y + 0.55$$

$b_{yx} = 0.39$.

Regression equation y on x.

$$y - \bar{y} = b_{yx}(x - \bar{x})$$
$$y - 5 = 0.39(x - 4)$$
$$y - 5 = 0.39x - 1.56$$
$$y = 0.39x - 1.56 + 5$$
$$y = 0.39x + 3.44.$$

2) From the following data find two regression equation

x	3	4	8	7	2
y	11	12	9	3	5

x	y	x^2	y^2	xy
3	11	9	121	33
4	12	16	144	48
8	9	64	81	72
7	3	49	9	21
2	5	4	25	10
$\Sigma x = 24$	$\Sigma y = 40$	$\Sigma x^2 = 142$	$\Sigma y^2 = 380$	$\Sigma xy = 184$

$$b_{xy} = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum y^2 - (\sum y)^2}$$

$$\frac{5(184) - (24)(40)}{5(380) - (40)^2} = \frac{920 - 960}{1900 - 1600}$$

$$\frac{-40}{300} = -0.13$$

$$b_{xy} = -0.13$$

$$\bar{x} = \frac{\sum x}{n} = \frac{24}{5} = 4.8$$

$$\bar{y} = \frac{\sum y}{n} = \frac{40}{5} = 8$$

Regression eq x and y

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

$$x - 4.8 = -0.13(y - 8)$$

$$x - 4.8 = -0.13y + 1.04$$

$$x = -0.13y + 1.04 + 4.8$$

$$x = -0.13y + 5.84$$

$$b_{yx} = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2}$$

$$\frac{5(184) - (24)(40)}{5(142) - (24)^2} = \frac{920 - 960}{710 - 576}$$

$$\Rightarrow \frac{-40}{134}$$

$$\Rightarrow -0.29$$

$$b_{yx} = -0.29$$

Regression equation y on x

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

$$y - 8 = -0.29(x - 4.8)$$

$$y - 8 = -0.29x + 1.392$$

$$y = -0.29x + 9.39.$$

Problem: Price indices of cotton and wool are given below for twelve months of a year. Obtain the equation of the lines of regression between the indices

Price index of cotton: 78 77 85 88 87 82 81 77 76 83

Price index of wool: 84 82 82 85 89 90 88 92 83 89 98.

x	y	$d_x = x_i - A$ $= x_i - 83$	d_x^2	$d_y = y_i - B$ $y_i - 88$.	d_y^2	$d_x d_y$
78	84	-5	25	-4	16	20
77	82	-6	36	-6	36	36
85	82	2.	4	-6	36	-12
88	85	5.	25	-3	9	-18
87	89	4.	16	1	1	4
82	90	-1	1	2.	4	-2
81	88 - B	-2	4	0	0	0
77.	92	-6	36	4	16	-24
76	83	-7	49	-5.	25	35
83 - A	89	0	0	1	1	0
97 - A	98	14	196	10	100	140
93	99 - B.	10.	100	11	121	110
1004	1061.	8.	492.	5.	365.	292.

$$byx = \frac{n\Sigma dxdy - \Sigma dx \cdot (\Sigma dy)}{n\Sigma dx^2 - (\Sigma dx)^2}$$

$$\frac{12(292) - (8)(5)}{12(492) - (8)^2}$$

$$\frac{3504 - 40}{5904 - 64}$$

$$\frac{3464}{5840} = 0.59.$$

$byx = 0.59.$

Regression equation y on x:

$$y - \bar{y} = byx(x - \bar{x})$$

$$y - 88.4 = 0.59(x - 83.6)$$

$$y = 0.59x - 49.3 + 88.4$$

$$y = 0.59x + 39.1$$

$$bxy = \frac{n\Sigma dxdy - (\Sigma dx)(\Sigma dy)}{n\Sigma dy^2 - (\Sigma dy)^2}$$

$$\frac{12(292) - (8)(5)}{12(365) - (5)^2}$$

$$\frac{3504 - 40}{4380 - 25}$$

$$\frac{3464}{4355} = 0.79.$$

$bxy = 0.79.$

Regression equation x on y.

$$x - \bar{x} = bxy(y - \bar{y})$$

$$x - 83.6 = 0.79(y - 88.4) \Rightarrow x = 0.79y - 69.8 + 83.6$$

$$x = 0.79y + 13.8$$

Problem: The following data relate to the scores obtain by a salesman of a company in an intelligence test and their weekly sales (Rs in 1000)

Sales man	A	B	C	D	E	F	G	H	I	J
Test scores	50	60	50	60	80	50	80	40	70	
Weekly sales	30	60	40	50	60	30	70	50	60	

x	y	$dx = x_i - A$ $= -60$	dx^2	$dy = y_i - A$ -50	dy^2	$dxdy$
50	30	-10	100	-20	400	200
60	60	0	0	10	100	0
50	40	-10	100	-10	100	100
60	50	0	0	0	0	0
80	60	20	400	10	100	200
50	30	-10	100	-20	400	200
80	70	20	400	20	400	400
40	50	-20	400	0	0	0
70	60	10	100	10	100	100
540	450	0	1600	0	1600	1200

$$b_{yx} = \frac{n\Sigma dxdy - \Sigma dx \cdot \Sigma dy}{n\Sigma dx^2 - (\Sigma dx)^2}$$

$$\frac{9(1200) - 0}{9(1600)} = \frac{9(1200)}{9(1600)} = \frac{3}{4} = 0.75$$

① Regression equation y on x

$$y - \bar{y} = b_{yx}(x - \bar{x})$$
$$y - 50 = 0.75(x - 60)$$
$$y - 50 = 0.75x - 45$$
$$y = 0.75x - 45 + 50$$
$$y = 0.75x + 5$$

when $x = 65$

$$y = 0.75(65) + 5$$
$$48.75 + 5$$
$$y = 53.75$$

Prob: The following data based on 450 students are given for marks in statistics and Economics at a certain exam

Mean marks in statistics: 40

Mean marks in Economics: 48

S.D of marks in Economics: 25

Sum of the product of deviation of marks from their respective mean: 42075

Obtain equation of the two lines of regression and estimate the average marks in Economics of candidates who obtained 50 marks in statistics.

8d) Given $n = 450$

Let the marks in statistics and economics be x, y

$\bar{x} = 40$, $\bar{y} = 48$.

$\sigma_x = 12$, $\sigma_y^2 = 256$

$\sigma_y = \sqrt{256} = 16$.

$\Sigma dx dy = 42075$.

$r = \frac{\Sigma dx dy}{n \times \sigma_x \times \sigma_y}$.

$r = \frac{42075}{450 \times 16 \times 12} = 0.45$

Regression equation y on x

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

$y - 48 = 0.48 \cdot \left(\frac{16}{12/3}\right)(x - 40)$

$y - 48 = 0.048\left(\frac{4}{3}\right)(x - 40)$

$y - 48 = 0.48(1.3)(x - 40)$

$y - 48 = 0.624(x - 40)$

$y - 48 = 0.624x - 24.96$

$y = 0.624(50) - 24.96 + 48$

$y = 31.2 + 23.04$

$y = 58.06$.