1

* Correlation ⇒ If the change in one variable affects a change in other variable are said to be correlated.

Ex ⇒ Height Vs. weight, Temperature Vs. Ice cream sales.

* If the increase (or decrease) in one results in a corresponding increase (or decrease) in other, correlation is said to be direct or positive.

* If increase (or decrease) in one results in corresponding decrease (or increase) in other correlation is said to be diverse or negative.

Ex ⇒ the income and expenditure is positive

Ex 1 ⇒ The correlation b/w (i) the heights and weight of a group of a persons and (ii) the income and expenditure is positive.

Ex 2 ⇒ The correlation b/w (i) the price and demand of a commodity and (ii) volume and pressure of a perfect gas is negative.

* Scatter diagram ⇒ It is the simplest way of the diagrammatic representation of bivariate data. Thus for the bivariate distribution (x_i, y_i); $i = 1, 2, \ldots, n$, if the values of the variable x and y are plotted along the x-axis and y-axis respectively in the x-y plane, the diagram of dots so obtained is known as scatter diagram.

* Karl pearson's coefficient of correlation ⇒ As a measure of intensity or degree of linear relationship b/w two variables, karl pearson, a British Biometrician, developed a formula called correlation coefficient.

* Correlation Coefficient b/w two random variables X and Y, usually denoted by $r(X,Y)$ or r_{XY} is a numerical measure of linear relationship b/w them and is defined as

$$r(X,Y) = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\frac{1}{n}\Sigma XY - \bar{X}\bar{Y}}{\sqrt{\left(\frac{1}{n}\Sigma X^2 - \bar{X}^2\right)\left(\frac{1}{n}\Sigma Y^2 - \bar{Y}^2\right)}}$$

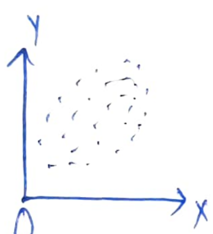
If (x_i, y_i); $i = 1, 2, \ldots, n$ is the bivariate distribution then

$$\text{Cov.}(X,Y) = \frac{1}{n}\Sigma(x_i - \bar{x})(y_i - \bar{y})$$

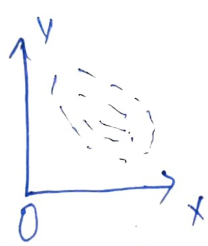
$$\sigma_X^2 = \frac{1}{n}\Sigma(x_i - \bar{x})^2$$

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

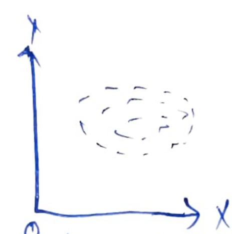
* Remark Following are the figures of the standard data for $r > 0$, < 0, $= 0$ and $r = \pm 1$



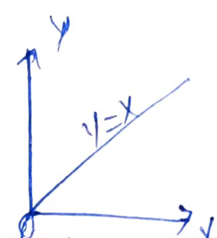
$(r > 0)$



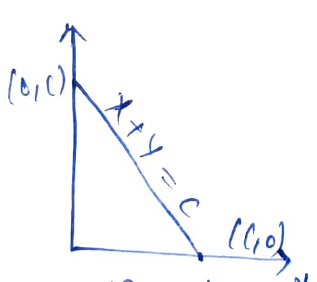
$(r < 0)$



$(r = 0)$



$(r = \pm 1)$



$(r = -1)$

The Marks obtained by 10 students in Mathematics (X) and Statistics (Y) are given below. Find the coefficient Correlation b/w X and Y

S. No.	1	2	3	4	5	6	7	8	9	10
X	75	30	60	80	53	35	15	40	38	48
Y	85	45	54	91	58	63	35	43	45	44

* Rank Correlation → Let (x_i, y_i); $i = 1, 2, \dots, n$ be the ranks of the ith individual in two characteristics A and B respectively. Pearsonian coefficient of Correlation b/w the rank x_i's & y_i's is called the rank Correlation coefficient b/w A and B for that group of individuals.

F) Spearman's Rank correlation coefficient ⇒ Assuming that no two individuals are bracketed equal in either classification, each of the variables X and Y takes the values 1, 2, — n

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

where $\sum d_i = \sum(x_i - y_i)$

Ques The Ranks of some 16 students in Mathematics and Physics are as follows. Two no. within brackets denote the ranks of the students in Mathematics and Physics.

(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 coefficient for proficiencies of this group in Mathematics and Physics.

Ranks in Maths. (x) 1 2 3 4 5 6 7 8 9 10

Ranks in Maths. (x)	Ranks 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
		0	136

Rank correlation coefficient is given by

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

$$= 1 - \frac{6 \times 136}{16\left((16)^2 - 1\right)}$$

$$\Rightarrow 1 - \frac{816}{16(255)}$$

$$\Rightarrow \frac{3264}{4080}$$

$$\Rightarrow 0.81$$

Ques Calculate the coefficients of correlation for ranks from the following data:

(x, y): (5, 8) (10, 3) (6, 2) (3, 9) (19, 12) (5, 3), (6, 17) (12, 18) (8, 22) (2, 12) (10, 17), (19, 20)

Ten competitors in a Musical test were ranked by the three judges A, B & C in the following order:

nk by A :	1	6	5	10	3	2	4	9	7	8
ank 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

* Repeated Ranks

Que Obtain the rank Correlation coefficient for the following data:

X :	68	64	75	50	64	80	75	40	55	64
Y :	62	58	68	45	81	60	~~48~~ 68	~~50~~ 48	50	70

X	Y	Rank X (x)	Rank Y (y)
68	62	4	5
64	58	6	7
75	68	2.5	3.5
50	45	9	
64	81	6	
80	60	1	
75	68	2.5	
40	48	10	
55	50	8	
64	70	6	

* The Case when epetation occur in variable th

have formula

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

Que..

A computer while calculating correlation coefficients b/w variables X and Y from 25 pairs of observations obtained the following result.

$n = 25$, $\Sigma X = 125$, $\Sigma X^2 = 650$, $\Sigma Y = 100$, $\Sigma Y^2 = 460$

$\Sigma XY = 508$

Sol^n

$\bar{X} = \frac{1}{n} \Sigma X = \frac{125}{25} = 5$

$\bar{Y} = \frac{1}{n} \Sigma Y = \frac{100}{25} = 4$

$r(X,Y) = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\frac{1}{n}\Sigma XY - \bar{X}\bar{Y}}{\sqrt{\left(\frac{1}{n}\Sigma X^2 - \bar{X}^2\right)\left(\frac{1}{n}\Sigma Y^2 - \bar{Y}^2\right)}}$

$\Rightarrow \frac{\frac{1}{25} \times 508 - 5 \times 4}{\sqrt{\left(\frac{1}{25} \times 650 - 25\right)\left(\frac{1}{25} \times 460 - 16\right)}} \Rightarrow \frac{20.32 - 20}{\sqrt{1(2.4)}}$

$\Rightarrow \frac{20.32 - 20}{\sqrt{(81.25 - 25)(57.5 - 16)}} \Rightarrow \frac{0.32}{1.5491}$

$\Rightarrow \frac{0.32}{\sqrt{(56.25)(41.5)}} = \frac{0.32}{48.36} \Rightarrow 0.2065$

Practice questions $\Rightarrow$

Ques 1 Calculate the coefficient of correlation b/w X and Y for the following:

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

(*) Calculate the Correlation coefficients for the following (in Inches) of fathers (X) and their sons (Y):

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

Sol^n

	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
Total	544	552	37028	38132	37560

$$\bar{X} = \frac{1}{n}\Sigma X = \frac{544}{8} = 68$$

$$\bar{Y} = \frac{1}{n}\Sigma Y = \frac{1}{8} \times 552 = 69$$

$$r(X,Y) = \frac{Cov(X,Y)}{\sigma_x \sigma_y} = \frac{\frac{1}{n}\Sigma XY - \bar{X}\bar{Y}}{\sqrt{\left(\frac{1}{n}\Sigma X^2 - \bar{X}^2\right)\left(\frac{1}{n}\Sigma Y^2 - \bar{Y}^2\right)}}$$

$$\Rightarrow \frac{\frac{1}{8} \times \{37560 - 68 \times 69\}}{\sqrt{\left\{\frac{37028}{8} - (68)^2\right\}\left\{\frac{38132}{8} - (69)^2\right\}}}$$

$$\Rightarrow \frac{4695 - 4692}{\sqrt{(4628.5 - 4624)(4766.5 - 4761)}} = \frac{3}{\sqrt{4.5 \times 5.5}} = 0.603$$

Regression ⇒ Regression analysis is a Mathematical measure of the average relationship b/w two or more variables in terms of the original units of the data.

* **Linear Regression** ⇒ If the variables in a bivariate distribution are related, we will find that the points in the scatter diagram will cluster round some curve called curve of regression.

If the curve is a straight line, it is called the line of regression.

* The line of regression is the line which gives the best estimate to the value of one variable for any specific value of the other variable. Thus the line of regression is the line of "best fit" and is obtained by the principle of Least Square.

* **Curve fitting by Least Square Methods.**

i) **fitting a straight line** ⇒ To fit a curve in straight line

$$y = a + bx \quad —(*)$$

$$\Sigma y = \Sigma a + \Sigma bx$$

$$\Sigma y = na + b\Sigma x \quad —(1)$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 \quad —(2)$$

Solve Eq^n (1) & (2) for a & b and then put value of a & b in (*) we get the straight line.

Ex-) for 10 randomly selected observations, the following were recorded:

x	y	xy	x^2
1	2	2	1
1	7	7	1
2	7	14	4
2	10	20	4
3	8	24	9
3	12	36	9
4	10	40	16
5	14	70	25
6	11	66	36
7	14	98	49
34	95	377	154

$$95 = 10a + b \cdot 34$$

$$377 = 34a + b \cdot 154$$

$\Rightarrow$ solve for a & b above Eqn we get

$$3230 = 340a + 1156b$$

$$3770 = 340a + 1540b$$

$$- \quad\quad - \quad\quad -$$

$$+540 = +384b$$

$$b = \frac{540}{384} \Rightarrow 1.4$$

lowing ✓

$= 100a + 34(1.4)$

$5 - 47.4 = 10a$

$47.6 = 10a \Rightarrow 4.76$

$a = 4.76$

put value of a & b in (*) we get

$$\boxed{y = 4.76 + 1.4x}$$

Ques

x:	1	3	4	6	8	9	11	14
y:	1	2	4	4	5	7	8	9

fit a straight line. $y = 0.64x + 0.55$

Ques fit a straight line to the following data

x:	0	1	2	3	4
y:	1	1.8	3.3	4.5	6.3

$$\boxed{y = 0.72 + 1.33x}$$

Ques fit a straight line to the following data

x:	3	7	9	10
y:	168	120	72	63

$y = -15.46x + 217.9$

Que. fit a second degree parabola to the following :-

x :	0	1	2	3	4
x :	1	1.8	1.3	2.5	6.3

$y = 1.42 + (-1.07)x + 0.55x^2$

Que. fit a second degree parabola to the following :

x	0	1	2	3	4
y	1	3	4	5	6

$a + bx + cx^2$

$y = 1.113 + (1.7716)x + (-0.1429)x^2$

Que. fit a second degree parabola to the following data

X	1	2	3	4	5
Y	10	12	13	16	19

$ax^2 + bx + c$

$y = 0.29x^2 + 0.49x + 9.4$

$10a + 34b + 154c = 95$

$34a + 154b + 820c = 377$

$\frac{a}{1} = \frac{-b}{4} = \frac{c}{E} = \frac{-1}{}$

following:- fitting a parabola.

$$y = a + bx + cx^2 \quad \text{—(1)}$$

$$\Sigma y = na + b\Sigma x + c\Sigma x^2 \quad \text{—(2)}$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \quad \text{—(3)}$$

$$\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 \quad \text{—(4)}$$

Solve Eq^n (2), (3) & (4) for a, b, c and put value of a, b, c in (1) we get a second degree parabola.

Ques. for 10 randomly selected observations, the following data were recorded:-

X :	1	1	2	2	3	3	4	5	6	7
Y :	2	7	7	10	8	12	10	14	11	14

X	Y	x^2	x^3	x^4	xy	x^2y
1	2	1	1	1	2	2
1	7	1	1	1	7	7
2	7	4	8	16	14	28
2	10	4	8	16	20	40
3	8	9	27	81	24	72
3	12	9	27	81	36	108
4	10	16	64	256	40	160
5	14	25	125	625	70	350
6	11	36	216	1296	66	396
7	14	49	343	2401	98	686
4	95	154	820	4774	377	1845

$$10a + 34b + 154c = 95$$

$$34a + 154b + 820c = 377$$

$$154a + 820b + 4774c = 1849$$

Solve for a, b & c we get

$$a = 1.8 \quad b = 3.48, \quad c = -0.27$$

$$y = 1.80 + 3.48x - 0.27x^2$$

~~Ques~~ (x) fitting a power curve.

$$y = ax^b \quad \text{---} \quad (*)$$

$$\log y = \log(ax^b)$$

$$\log y = \log a + b \log x \Rightarrow U = A + bV \quad \text{---} \quad (\circledast)$$

$$\left[\because U = \log y, \quad A = \log a \atop V = \log x \right]$$

$$\Sigma U = nA + b\Sigma V \quad \text{---} \quad (1)$$

$$\Sigma UV = A\Sigma V + b\Sigma V^2 \quad \text{---} \quad (2)$$

find value A, b from Eqn (1) & (2) and put in Eqn (*) we get power curve.

Ques fit a power curve of the form $y = ax^b$ to the following data!

x	y	$U = \log y$	$V = \log x$ ~~$A = \log a$~~	UV	V^2
1	1.0	0	0	0	0
2	1.2	0.0792	0.3010	0.0238	0.09
3	1.3	0.1139	0.4771	0.0543	0.2275
4	1.8	0.2552	0.6020	0.1536	0.362404 ~~0.0924~~
5	2.1	0.3222	0.6989	0.2251	0.48846
6	7.1	0.851	0.7781	0.6621631	0.6054

1.6215, $\Sigma V = 2.857$, $\Sigma UV = 1.1189$, $\Sigma V^2 = 1.7737$

$1.62 = 6A + b \cdot 2.85$ — (1)

$1.118 = 2.85A + b \cdot 1.77$ — (2)

Solve these two Eqn for A & b we get.

$b = 0.83$

$A = -0.124$

$A = \log a$

$-0.124 = \log a$

$a = e^{-0.124}$

$a = 0.883$

Put value of a & b in Eqn (A)

$y = 0.8x^{0.8}$

(4) Exponential curve

$y = ab^x$ — (A)

$\log y = \log a + x \log b$

$U = A + xB$

$\left\{ \begin{array}{l} U = \log y \\ A = \log a \\ B = \log b \end{array} \right\}$ — (**)

$\Sigma U = nA + B\Sigma x$ — (1)

$\Sigma Ux = A\Sigma x + B\Sigma x^2$ — (2)

Solve Eqn (1) & (2) for A & B and then put value of A & B in (**) and get value of a & b

Que Fit an exponential curve of the form $y = ab^x$ to the following data:

X :	1	2	3	4	5	6	7	8
Y :	1.0	1.2	1.8	2.5	3.6	4.7	6.6	9.1

Sol^n

X	Y	U = log Y	XU	X^2
1	1.0	0.0	0.0	1
2	1.2	0.079	0.158	4
3	1.8	0.2553	0.765	9
4	2.5	0.3979	1.591	16
5	3.6	0.556	2.781	25
6	4.7	0.672	4.032	36
7	6.6	0.819	5.736	49
8	9.1	0.959	7.672	64
36	30.5	3.739	22.7385	204

from Eq^n (1) & (2) we get

$$3.739 = 8A + 36B \quad \text{---(3)}$$
$$22.73 = 36A + 204B \quad \text{---(4)}$$

Solve these two Eq^n for A & B we get

$B = 0.1408 \qquad A = -0.166$

$A = \log a \qquad B = \log b$

$-0.166 = \log a \qquad 0.1408 = \log b$

$a = e^{-0.166} \qquad b = e^{0.1408}$

$a = 0.847 \qquad b = 1.1571$

put value of a & b in (*) we get

$$Y = (0.847)(1.157)^{x}$$