27/10/2025

UNIT-5

Correlation: It is a statistical measure. It can be used to measure the direction and degree of the relationship between the variables.

Definition: Two variables X and Y are said to be correlated if change in one variable results in a corresponding change in another variable.

(1) If the variables X and Y deviate in the same direction then we say that there is a positive correlation between X and Y.

↑ ↑
x y

↓ ↓
x y

(2) If the variables X and Y deviate in the opposite direction then we say that there is a -ve correlation between X and Y.

↑ ↓
x y

↓ ↑
x y

If the relationship between X and Y is of $Y = A + BX$, then we say there is a linear relationship between X and Y, otherwise we say that it is non-linear.

Measures of correlation between X and Y

① Karl Pearson's coefficient of correlation:

The coefficient of correlation between X and Y is given by r (or) r_{XY} (or) $r(X,Y) = \dfrac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y}$ → ①

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

(This is covariance between X and Y)

$$\sigma_X = \sqrt{\frac{1}{n} \Sigma (x_i - \bar{x})^2}, \quad \sigma_Y = \sqrt{\frac{1}{n} \Sigma (y_i - \bar{y})^2}$$

∴ ~~From ①, we get~~

$$r = \frac{\frac{1}{n} \Sigma (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \Sigma (x_i - \bar{x})^2} \sqrt{\frac{1}{n} \Sigma (y_i - \bar{y})^2}}$$

$$r = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma (x_i - \bar{x})^2 \Sigma (y_i - \bar{y})^2}} = \frac{\Sigma dx\, dy}{\sqrt{\Sigma dx^2 \cdot \Sigma dy^2}}$$

where $dx = x_i - \bar{x}$, $dy = y_i - \bar{y}$

It can be used to measure the direction and degree of the relationship between the variables.

Note: ① It lies between -1 & 1. That is:

$-1 \leq r \leq 1$

(2) If $r = 1$, then there is a perfect +ve correlation b/n X & Y.

If $r = -1$, then there is a perfect -ve correlation b/n X & Y.

If $r = 0$, then there is no correlation between X & Y. That is: X, Y are independent.

If r lies between 0.7 and 1, then there is a strong +ve. correlation b/n X and Y.

If r lies between 0.3 & 0.7, then there is a moderate +ve correlation between X & Y.

If r lies between 0 and 0.3, then there is a weak correlation between X & Y.

(applicable for -ve side too)

Q1) Find the coeff of correlation b/n X & Y for following data:

X:	1	2	3	4	5	6	7	8	9	(x_i)
Y:	10	11	12	~~13~~ 14	~~14~~ 13	15	16	17	18	(y_i)

2 forms

$$\bar{x} = \frac{\sum x_i}{n}, \quad \bar{y} = \frac{\sum y_i}{n}$$

Sol: $\bar{x} = \frac{\sum x_i}{n} = \frac{1+2+3+4+5+6+7+8+9}{9}$

$$= \frac{45}{9} = 5.$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{10+11+12+14+13+15+16+17+18}{9}$$

$$= \frac{126}{9} = 14$$

x_i	$d_i = x_i - \bar{x}$	dx^2	y_i	$dy = y_i - \bar{y}$	dy^2	$dxdy$
1	-4	16	10	-4	16	16
2	-3	9	11	-3	9	9
3	-2	4	12	-2	4	4
4	-1	1	~~13~~ 14	0	0	0
5	0	0	~~15~~ 13	-1	1	0
6	1	1	~~16~~ 15	1	1	1
7	2	4	16	2	4	4
8	3	9	17	3	9	~~3~~ 9
9	4	16	18	4	16	16
		60			60	59
		$\sum dx^2$			$\sum dy^2$	$\sum dxdy$

$$r = \frac{\sum dx\,dy}{\sqrt{\sum dx^2 \sum dy^2}} = \frac{59}{\sqrt{60 \cdot 60}}$$

$$= \frac{59}{60} = 0.9833$$

Strong +ve correlation

23/10/2025

Spearman's Rank Correlation Coefficient

If $(x_1, y_1), (x_2, y_2) \ldots (x_n, y_n)$ be the ranks of n individuals in two characteristics A and B, then the rank correlation coefficient between A and B is given by

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

Here n = number of pairs

$d = x_i - y_i$ = (difference between ranks).

Prob: The ranks of same 15 students in 2 subjects A and B are given below:

(1,10), (2,7), (3,2), (4,6), (5,4), (6,8), (7,3), (8,1), (9,11) (10,15) (11,9), (12,5), (13,4) (14,12), (15,13)

Use spearman's formula to find rank correlation coefficient.

Sol: Here, $n = 15$

Ranks in A (x_i)	Ranks in B (y_i)	$d = x_i - y_i$	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

$\Sigma d^2 = 272$

$\therefore$ Rank correlation coefficient is

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

$$= \frac{1 - 6(272)}{15(15^2-1)} = 0.514$$

Moderate, +ve correlation

Prob: Calculate Spearman's rank correlation coefficient between X and Y from the following data:

(advertisement cost)

X: 39, 65, 62, 90, 82, 75, 25, 98, 36, 78

Y: 47, 53, 58, 86, 62, 68, 60, 91, 51, 84

(sales in rupees)

Sol:

X	Ranks in X (x_i)	Y		$d = x_i - y_i$	d^2
39	8	47	10	-2	4
65	6	53	8	-2	4
62	7	58	7	0	0
90	2	86	2	0	0
82	3	62	5	-2	4
75	5	68	4	+1	1
25	10	60	6	4	16
98	1	91	1	0	0
36	9	51	9	0	0
78	4	84	3	1	1

$n = 10$ $\qquad \Sigma d^2 = 30$

Rank correlation coefficient is:

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

$$= 1 - \frac{6[30]}{10[10^2 - 1]} = 0.819$$

Spearman's Rank correlation coefficient for repeated ranks is

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

$\Sigma \frac{m(m^2-1)}{12} \longrightarrow$ Correction factor

m = number of times an item is repeated

Q) From the following data, calculate rank correlation coefficient after making adjustment for tied ranks.

common rank, this mean

X : 48, 33, 40, 9, 16, 16, 65, 24, 16, 57

Y : 13, 13, 24, 6, 15, 4, 20, 9, 6, 19

X	10 Rank of x_i	Y	Rank of y_i (x_i)	$d = x_i - y_i$	d^2
48	3	13	5.5	-2.5	6.25
33	5	13	5.5	-0.5	~~6.25~~ 0.25
40	4	24	1	3	9
9	10	6	8.5	1.5	2.25
16	8	15	4	4	16
16	8	4	4	-2	4
65	1	20	2	-1	1
24	6	9	7	-1	1
16	8	6	8.5	-0.5	~~0.~~ 0.25
57	2	19	3	1	1

$\Sigma d^2 = 41$

C.F

Repeated value	(m) Number of times m item is repeated	C.F $= \frac{m(m^2-1)}{12}$
X: 16	3	$\frac{3(3^2-1)}{12} = 2$
Y: 13	2	$\frac{2(2^2-1)}{12} = 0.5$
Y: 6	2	$\frac{2(2^2-1)}{12} = 0.5$
		$\frac{\Sigma m(m^2-1)}{12} = 3$

∴ Rank correlation coefficient is

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

$$= 1 - \frac{6[41+3]}{10[10^2-1]} = 0.7334$$

strong +ve correlation

24/10/2025

Q) From the following table, calculate the coefficient of correlation by Karl Pearson's method.

X :	6	2	10	4	8
Y :	9	11	?	8	7

Arithmetic Mean of X and Y series are 6 and 8 respectively.

$\bar{y} = 8$

$$r = \frac{\Sigma dx\, dy}{\sqrt{\Sigma dx^2\, \Sigma dy^2}}$$

$$\frac{9+11+x+8+7}{5} = 40$$

$\boxed{x=5}$

$\bar{x} = 6$

x_i	$d_i = x_i - \bar{x}$	dx^2	y_i	$dy = y_i - \bar{y}$	dy^2
6	0	0	9	1	1
2	-4	16	11	3	9
10	4	16	5	-3	9
4	-2	4	8	0	0
8	2	4	7	-1	1
		40			~~24~~ 20

dx dy

0

-12

-12

0

-2

-26

$$\frac{-26}{\sqrt{800}}$$

$$\frac{-26}{28.2842}$$

$$r = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\frac{1}{n}\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\frac{1}{n}\Sigma(x_i-\bar{x})^2}\sqrt{\frac{1}{n}\Sigma(y_i-\bar{y})^2}}$$

$$= \boxed{-0.919}$$

a) The coefficient of correlation between X and Y is 0.48, The covariance is 36. The variance of X is 16. Find the standard deviation of Y.

$$\text{cov}(X,Y) = \frac{1}{n}\Sigma(x_i-\bar{x})(y_i-\bar{y}) = 36$$

$$r = 0.48$$

$$r = \frac{36}{\sigma_X \sigma_Y}$$

$$0.48 = \frac{\cancel{36}\,9}{\cancel{4}\,\sigma_Y} \qquad \frac{\cancel{0.48}}{9}$$

$$\sigma_Y = \frac{9}{0.48} = \boxed{18.75} \quad 18.75$$

Given: $r = 0.48$, $\text{Cov}(X,Y) = 36$, $\sigma_X^2 = 16$

$\sigma_Y = ?$

Q3) From the following data, calculate the rank correlation coefficient :

X : 24, 29, 19, 14, 30, 19, 27, 30, 20, 28, 11

Y : 37, 35, 16, 26, 23, 27, 19, 20, 16, 11, 21

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

X	x_i	Y	y_i	d
24	6	37	1	5
29	3	35	2	1
19	8.5	16	9.5	−1
14	~~9~~ 10	26	4	6
30	1.5	23	5	−3.5
19	8.5	27	3	5.5
27	5	19	8	−3
30	1.5	20	7	−5.5
20	7	16	9.5	−2.5
28	4	11	11	−7
11	11	21	6	5

$$\frac{1+2}{2} = \frac{3}{2} = 1.5$$

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

$$\frac{9+10}{2} = \frac{19}{2} = 9.5$$

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

$$\rho = 1 - \frac{6\left[225 + 0.5 + 0.5 + 0.5\right]}{11(11^2 - 1)}$$

$$= -0.0295$$

$(X_1, Y_1), (X_2, Y_2) \ldots (X_n, Y_n)$
of X, Y

Scatter Diagram Method

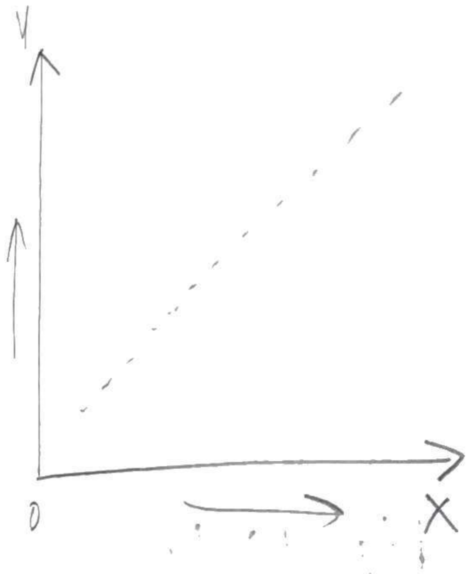


Fig 1: perfect +ve correlation

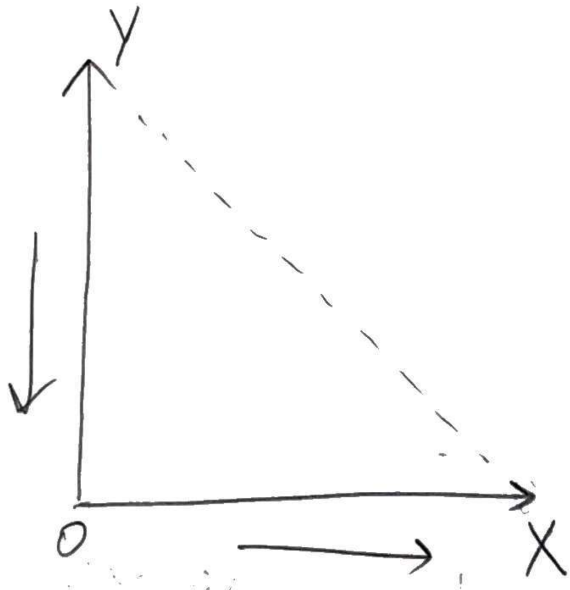


Fig 2: Perfect -ve correlation

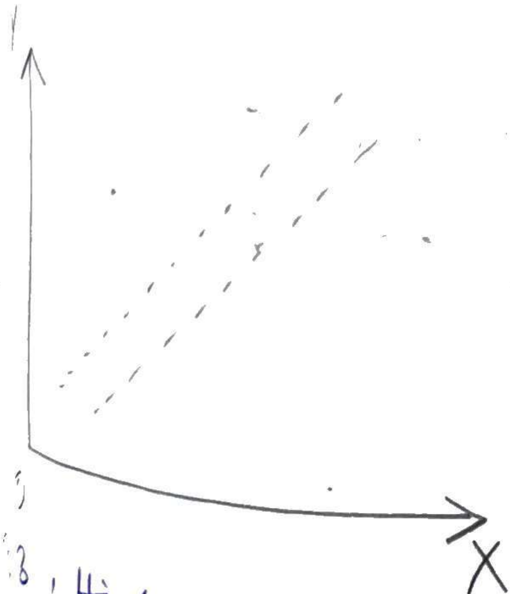


Fig 3: Higher degree, +ve correlation

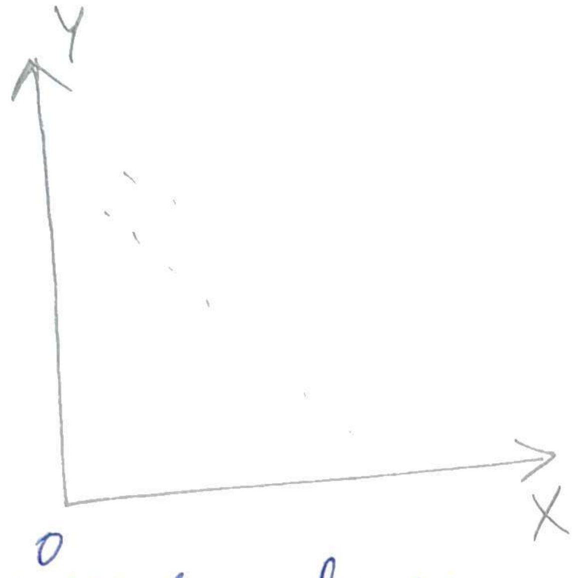


Fig 4: Higher degree -ve correlation

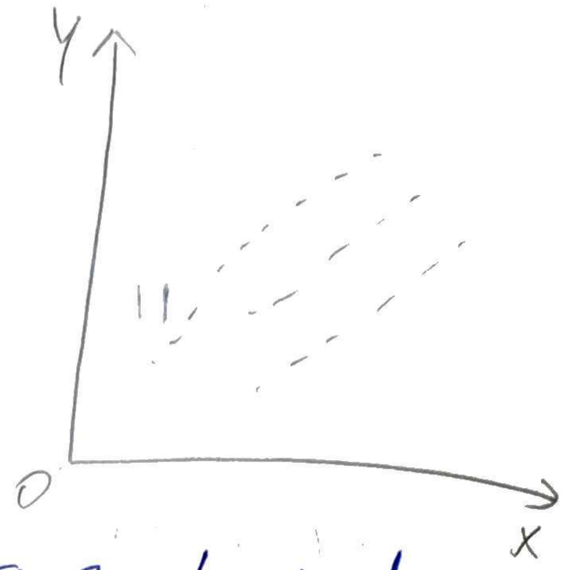


Fig 5: Lower degree, +ve correlation

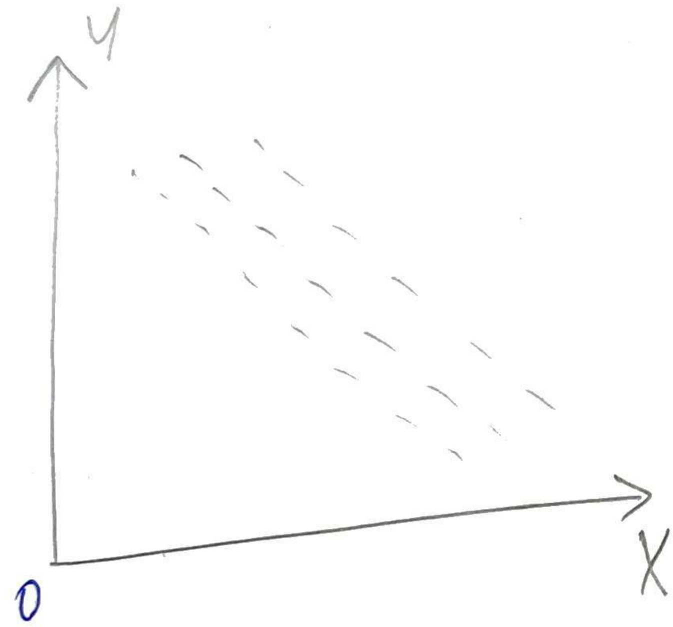


Fig 6: Lower degree, -ve correlation

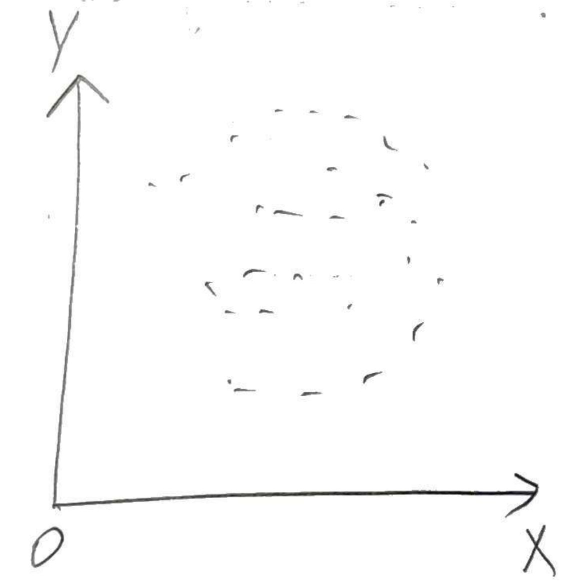


Fig 7: No correlation

Q) X: 65 66 67 67 68 69 70 72

Y: 67 68 65 68 72 72 69 71

Find $r = ?$

$d_i = x_i - \bar{x}$	dx^2	y_i	$dy = y_i - \bar{y}$	dy^2	$dx\,dy$
-3	9	67	-2	4	6
-2	4	68	-1	1	2
-1	1		-4	16	4
0	0	68	-1	1	0
1		72	3	9	3
2	4	72	3	9	6
4	16	69	0	0	0
-1	1	71	2	4	-2
					21
					19

$$\bar{y} = \frac{65+66+67+67+68+69+70+72}{8} = \frac{544}{8} = 68$$

0.6

$$\bar{x} = \frac{552}{8} = 69$$

$$r = \frac{19}{\sqrt{36 \times 44}} = \frac{19}{\sqrt{1584}} = \frac{19}{39.79}$$

$= 0.478$

X & Y shud be in order only.

28/10/2025

Curve Fitting by the method of Least Squares

1) Fitting a st. line $y = a+bx$

To fit a st. line $y=a+bx$ to the given data: $(x_1,y_1),(x_2,y_2)\ldots(x_n,y_n)$, we proceed as follows:

1. Form the normal equations

$$\Sigma y = na + b\Sigma x$$

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

These eqns are calculated by the method of least squares i.e by minimising the sum of squares of deviations (errors) between the observed values and the values predicted by the curve.

observed values

Predicted values by the curve

$x_1 \rightarrow y_1$ $\qquad \hat{y}_1 = a+bx_1$

$x_2 \rightarrow y_2$ $\qquad \hat{y}_2 = a+bx_2$

$\vdots$

$x_n \rightarrow y_n$ $\qquad \hat{y}_n = a+bx_n$

$$\left\{\begin{aligned} e_1 &= y_1 - \hat{y}_1 \\ e_2 &= y_2 - \hat{y}_2 \\ &\vdots \\ e_n &= y_n - \hat{y}_n \end{aligned}\right.$$

$$S = \underbrace{e_1^2 + e_2^2 +}\cdots + e_n^2$$

$\frac{dy}{dx}$

Step 2 : Solve the normal equation for a and b.

Step 3 : Substitute the values of a & b in ③ to get the required st. line. $y = bx + a$.

Q) Fit a straight line using the method of least squares for the following data:

x :	-2	-1	0	1	2
y :	1	2	3	3	4

Sol: Let $\underbrace{y = a + bx}$ be the st. line to be fitted for the given data.

Then the Normal eqns are

$$\Sigma y = na + b\Sigma x$$
$$\Sigma xy = a\Sigma x + b\Sigma x^2$$

x	y	xy	x^2
-2	1	-2	4
-1	2	-2	1
0	3	0	0
1	3	3	1
2	4	8	4

$\Sigma x = 0$

$\Sigma y = 13$

$\Sigma xy = 7$

$\Sigma x^2 = 10$

$n = 5$

Then the Normal eqns are

$$\Sigma y = na + b\Sigma x$$
$$\Sigma xy = a\Sigma x + b\Sigma x^2$$

$$13 = 5a + b(0)$$
$$7 = a(0) + b(10)$$

Solving these, we get

$a = 13/5$, $b = 7/10$.

Substituting these in (I), we get

$$\boxed{y = \frac{13}{5} + \frac{7x}{10}}$$

~~b) Curve Fitting~~

b) Fitting a parabola $y = a + bx + cx^2$

To fit a parabola $y = a + bx + cx^2$ — (I) to the given data:

$(x_1, y_1), (x_2, y_2) \ldots (x_n, y_n)$, we proceed as follows.

1. Form the Normal equations

$$\Sigma y = na + b\Sigma x + c\Sigma x^2$$
$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3$$
$$\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4$$

(2) Solve the normal eqns for a, b and c.

(3) Substitute the values of a, b and c in (I) to

get the required parabola $y = a + bx + cx^2$.

Q) Fit a parabola for the following data by the method of least square

x :	0	1	2
y :	1	6	17

Let $y = a + bx + cx^2$ be the parabola to be fitted for the given data.

Then the Normal eqns are

$$\Sigma y = na + b\Sigma x + c\Sigma x^2$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3$$

$$\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4$$

x	y	x^2	x^3	x^4	xy	x^2y
[illegible]	1	0	0	0	0	0
[illegible]	6	1	1	1	6	6
[illegible]	17	4	8	16	34	68
3	24	5	9	17	40	74
$=\Sigma x$	$=\Sigma y$	$=\Sigma x^2$				

$\boxed{n = 3}$

$$24 = 3a + 3b + 5c$$
$$40 = 3a + 5b + 9c$$
$$74 = 5a + 9b + 7c$$

$a = 1, b = 2, c = 3$

Now substituting these values in ①, we get

$$y' = 1 + 2x + 3x^3$$

Q) Fit a straight line for the given data by using the method of least squares

x :	10	12	13	16	17	20	25
y :	10	22	24	27	29	33	37

$\Sigma y = 182$, $\Sigma x = 113$, $n = 7$

$\Sigma xy = 3186$, $\Sigma x^2 = 1983$

$7a + 113b = 182$

$113a + 1983b = 3186$

$a = 0.82$, $b = 1.56$

0.79

$y = 0.82 + 1.56x$

29/10/2025

Curve Fitting

$\log_{10} \to \log$

$\log_e \to \ln$ ← natural logarithm

Fitting exponential curve $y = ab^x$

Consider $y = ab^x$ —(1)

Taking log's on both sides, we get

$$\log_{10} y = \log_{10}(ab^x) = \log_{10} a + \log_{10} b^x$$

$$\log(ab) = \log a + \log b, \quad \log(x^m) = m \log x$$

$$\log_{10} y = \log_{10} a + x . \log_{10} b$$

Putting $\log_{10} y = Y$, $\log_{10} a = A$, $\log_{10} b = B$

We get $Y = A + Bx$ which is a st. line.

The Normal eqns are

$$\Sigma Y = nA + B \Sigma x$$

$$\Sigma xY = A \Sigma x + B \Sigma x^2$$

Solving these, we get A and B.

Then we get

$$a = 10^A, \; b = 10^B.$$

$\therefore$ The required curve $\boxed{y = ab^x}$.

Q) Fit exponential curve of the form $y = ab^x$ for the following data by the method of least squares

x :	2	3	4	5	6
y :	8.3	15.4	33.1	65.2	127.4

Sol: The ~~curve~~ exponential curve to be fitted for the given data is:

$$y = ab^x \quad -① $$

Taking log's on both sides, we get

$$Y = A + Bx \quad -② \quad \text{where } Y = \log_{10} y$$

$$A = \log_{10} a$$

$$B = \log_{10} b$$

To fit this, the Normal eqns are:

$$\Sigma Y = nA + B\Sigma x \quad -③$$

$$\Sigma xY = A\Sigma x + B\Sigma x^2 \quad -④$$

x	y	x^2	$Y = \log_{10} y$	xY
2	8.3	4	0.9190	1.8380
3	15.4	9	1.1875	3.5625
4	33.1	16	1.5198	6.0792
5	65.2	25	1.8142	9.071
6	127.4	36	2.1051	12.6306
20		90	7.5456	33.1813
Σx		Σx^2	ΣY	ΣxY

$7.5456 = 5A + 20B$

$33.1813 = 20A + 90B$

Solving these, we get $A = 0.30956$

$B = 0.29989$

$a = 10^A = 2.0391$

$b = 10^B = 1.9947$

The required case is $\boxed{y = ab^x}$

$\Rightarrow y = (2.0396)(1.9947)^x$

Estimate the value of y for $x = 3.5$

Substitute $x = 3.5$.

$\Rightarrow \quad y = 22.8621$

Create graph by plotting points – x & y to decide which curve fitting to use..

Q) Fit an exponential curve of the form $\boxed{y = ae^{bx}}$ to the following data :

x :	1	5	7	9	12
y :	10	15	12	15	21

by using the method of least squares.

$$y = ae^{bx}$$

$$\ln y = \ln (ae^{bx})$$

$$\ln y = \ln a + \ln e^{bx} \qquad \ln_e e = 1$$

$$\ln (y) = \ln (a) + bx$$

$$Y = A + bx, \text{ where } Y = \ln (y)$$

$$A = \ln (a)$$

$$\Rightarrow a = e^{A}$$

$$\Sigma Y = nA + b\Sigma x$$

$$\Sigma xY = A\Sigma x + b\Sigma x^2$$

30/10/2025

Linear Regression

① Line of regression of y on x:

The line of regression of y on x is $y = a + bx$, where the unknowns a and b are calculated by solving the following normal equations obtained by the method of least squares

$$\sum y = na + b\sum x$$
$$\sum xy = a\sum x + b\sum x^2$$

Another form is

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

Here $\bar{x} = \frac{\sum x_i}{n}$, $\bar{y} = \frac{\sum y_i}{n}$

$$b_{yx} = \boxed{\frac{cov(x,y)}{\sigma_x^2}}$$ (This is known as coefficient of line of regression of y on x).

$$= \frac{r\,\sigma_x\,\sigma_y}{\sigma_x^2} \qquad \left(\because r = \frac{cov(x,y)}{\sigma_x \sigma_y}\right)$$

$$= \boxed{\frac{r\,\sigma_y}{\sigma_x}}$$

(OR)

$$b_{yx} = \frac{\frac{1}{n}\sum(x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n}\sum(x_i - \bar{x})^2} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

$$b_{yx} = \frac{\Sigma dx\, dy}{\Sigma dx^2}$$

where $dx = x_i - \bar{x}$

$dy = y_i - \bar{y}$

Line of regression of x on y

Form I: $x = A + By$

where A, B are calculated by the following normal equations.

$$\Sigma x = nA + B\Sigma y \quad (\cancel{Aa + b\Sigma y})$$

Multiplying with y & applying Σ

$$\Sigma xy = A\Sigma y + B\Sigma y^2$$

Form II:

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

where $\boxed{b_{xy} = \frac{\text{cov}(x,y)}{\cancel{\Sigma y}\, \sigma_y^2}}$ (This is the coefficient of line of regression of x on y).

$$\Rightarrow b_{xy} = \boxed{\frac{r\sigma_x}{\sigma_y}} = \boxed{\frac{\Sigma dx\, dy}{\Sigma dy^2}}$$

where $dx = x_i - \bar{x}$, $dy = y_i - \bar{y}$

Note:

(1) $$y - \bar{y} = \frac{r\sigma_y}{\sigma_x}(x - \bar{x}) \quad — (1)$$

$$x - \bar{x} = \frac{r\sigma_x (y - \bar{y})}{\sigma_y} \quad — (2)$$

From these, it is clear that $(\bar{x}, \bar{y})$ is a point of intersection of lines of regression.

(2) If $r = 0$ (No correlation), then the lines of regression are

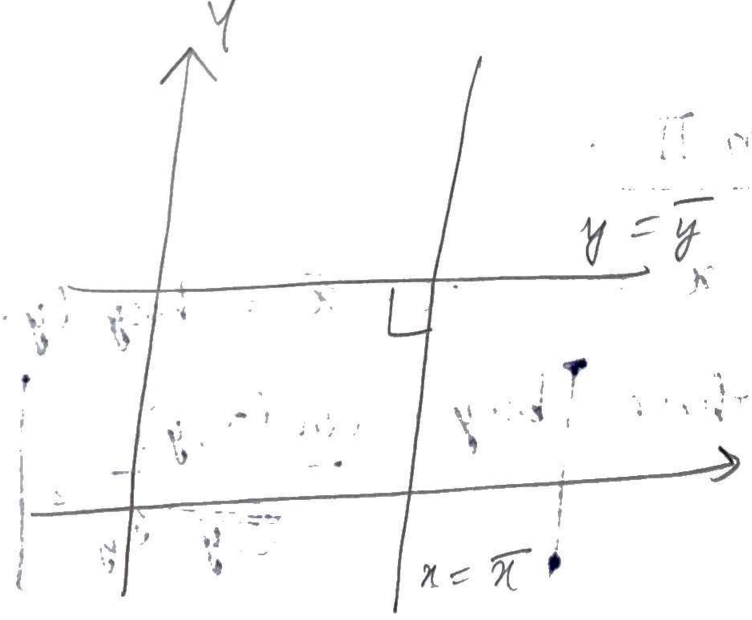
$y - \bar{y} = 0$, $x - \bar{x} = 0$

$y = \bar{y}$, $x = \bar{x}$

$\bar{x}, \bar{y} \to$ Means

↓

perpendicular lines



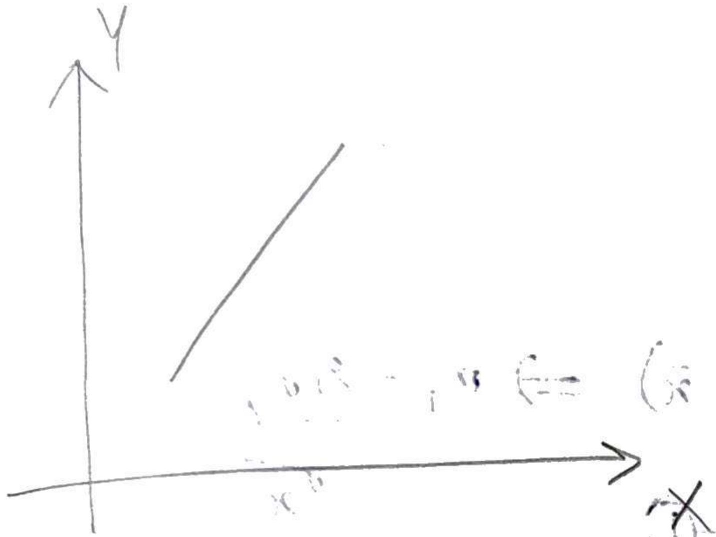
which are ⊥lar to each other.

(3) If $r = \pm 1$ (perfect correlation), then the regression lines are

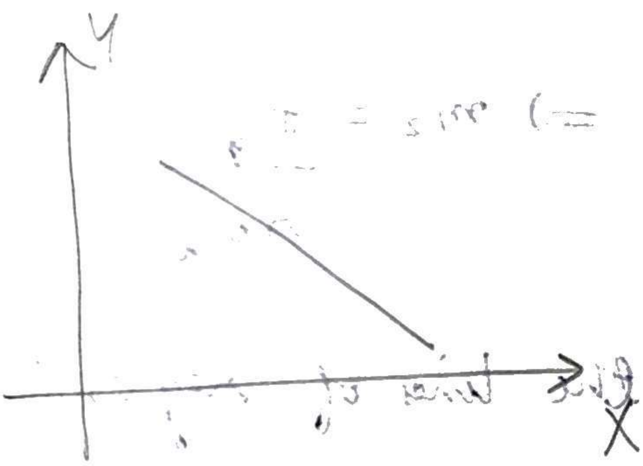
$$y - \bar{y} = \pm \frac{\sigma_y}{\sigma_x}(x - \bar{x})$$

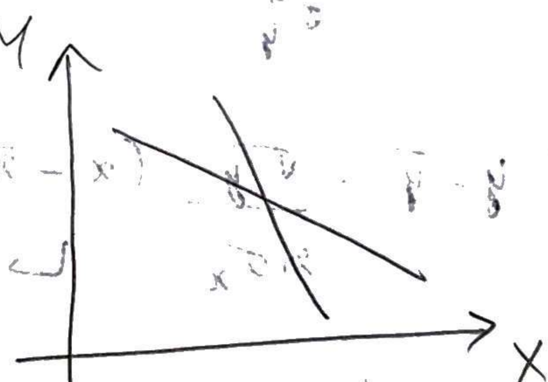
$$x-\bar{x} = \pm\frac{\sigma_x}{\sigma_y}(y-\bar{y}) \Rightarrow y-\bar{y} = \pm\frac{\sigma_y}{\sigma_x}(x-\bar{x})$$

That is:
$\therefore$ Lines of regression ~~are coincid~~ coincide.

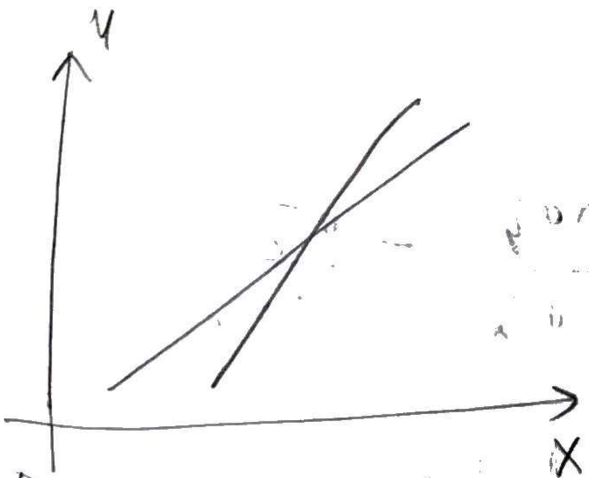


perfect +ve correlation

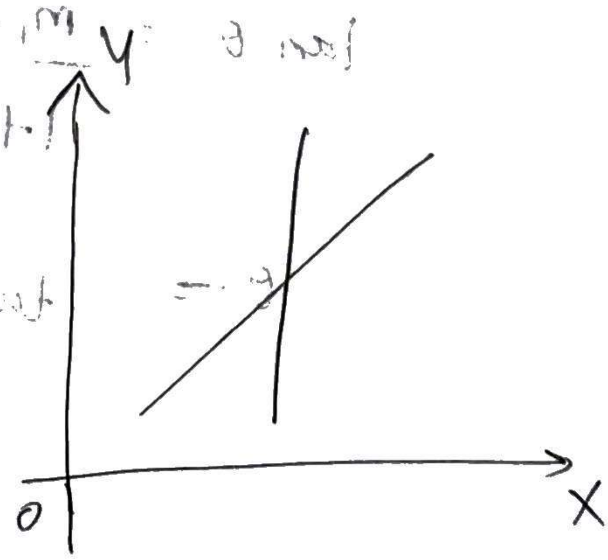


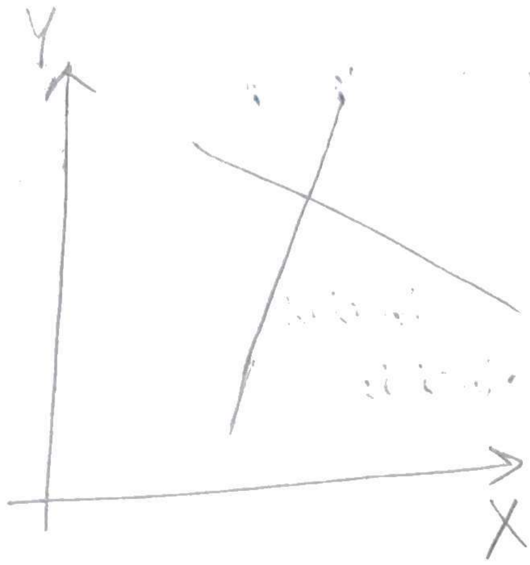


perfect -ve correlation



Strong +ve





Note: 4 ~~Angle~~

$$y - \bar{y} = \frac{r\sigma_y}{\sigma_x}(x - \bar{x}) \Rightarrow m_1 = \frac{r\sigma_y}{\sigma_x}$$

$\llcorner$ ①

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

$$\hookrightarrow y - \bar{y} = \frac{\sigma_y}{r\sigma_x}(x - \bar{x}) \Rightarrow m_2 = \frac{\sigma_y}{r\sigma_x}$$

$\llcorner$ ②

If 'θ' is angle between the lines of regression, we get:

$$\tan\theta = \frac{m_1 - m_2}{1 + m_1 m_2}$$

$$\cancel{\theta =} \quad \tan\theta = \frac{\frac{r\sigma_y}{\sigma_x} - \frac{\sigma_y}{r\sigma_x}}{1 + \frac{r\sigma_y}{\sigma_x}\frac{\sigma_y}{r\sigma_x}}$$

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

$$= \frac{\frac{\sigma_y}{\sigma_x}\left[\frac{r^2-1}{r}\right]}{\frac{\sigma_x^2+\sigma_y^2}{\sigma_x^2}}$$

$$= \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}\left[\frac{r^2-1}{r}\right]$$

Note ⑤ : From the above, we have

$$b_{yx} = \frac{cov(x,y)}{\sigma_x^2}, \quad b_{xy} = \frac{cov(x,y)}{\sigma_y^2},$$

$$r = \frac{cov(x,y)}{\sigma_x \sigma_y}$$

$\Rightarrow b_{yx} \cdot b_{xy} = r^2$

Since $\sigma_x > 0$, $\sigma_y > 0$, then the signs of b_{yx}, b_{xy} and r only depend on the sign of $cov(x,y)$.

That is:

$$\left.\begin{array}{l} cov(x,y) \text{ is +ve, then } b_{yx}, b_{xy}, r \text{ are +ve.} \\ cov(x,y) \text{ is -ve, then } b_{yx}, b_{xy}, r \text{ are -ve.} \end{array}\right\}$$

From these, it is clear that b_{yx}, b_{xy} and r have same sign.

That is:

If b_{yx}, b_{xy} are +ve, then r is also +ve.

If b_{yx}, b_{xy} are −ve, then r is also −ve.

31/10/2025

Prob ①: From the following ~~table~~ data, obtain the two lines of regression. Also find the coefficient of correlation between sales (x) and purchase (y).

Sales (x):	91	97	108	121	67	124	51	73	111	57
Purchase (y):	71	75	69	97	70	91	39	61	80	47

We Know that the lines of regression are:

$$y - \bar{y} = b_{yx}(x - \bar{x}) \quad -①$$

$$x - \bar{x} = b_{xy}(y - \bar{y}) \quad -②$$

Here $b_{yx} = \frac{\operatorname{cov}(x,y)}{\sigma_x^2} = \frac{r\sigma_y}{\sigma_x} = \frac{\Sigma dx\,dy}{\Sigma dx^2}$

~~Here~~ and $b_{xy} = \frac{\operatorname{cov}(x,y)}{\sigma_y^2} = \frac{r\sigma_x}{\sigma_y} = \frac{\Sigma dx\,dy}{\Sigma dy^2}$

$= \frac{\Sigma x_i}{n}$, $\bar{y} = \frac{\Sigma y_i}{n} = \frac{700}{10} = 70$

$= \frac{900}{10}$

$= 90$

	y	$dx = x - \bar{x}$	$dy = y - \bar{y}$	$dxdy$	dx^2	dy^2
1	71	1	1	1	1	1
7	75	7	5	35	49	25
08	69	18	-1	-18	324	1
21	97	31	27	837	961	729
7	70	-23	0	0	529	0
24	91	34	21	714	1156	441
1	39	-39	-31	1209	1521	961
	61	-17	-9	153	289	81
	80	21	10	210	441	100
	47	-33	-23	759	1089	529

$dxdy = 3900$, $\Sigma dx^2 = 6360$, $\Sigma dy^2 = 2868$

$b_{yx} = \frac{\Sigma dxdy}{\Sigma dx^2} = \frac{3900}{6360} = 0.6132$

$b_{xy} = \frac{\Sigma dxdy}{\Sigma dy^2} = \frac{3900}{2868} = 1.3598$

The lines of regression are

$y - 70 = 0.6132(x - 90)$

$$x - 90 = 1.3598(y - 70)$$

$$y = 70 + 0.6132(x - 90)$$
$$x = 90 + 1.3598(y - 70)$$

Coefficient of correlation is given by

$$r^2 = byx - bxy$$
$$r = \pm \sqrt{byx - bxy}$$
$$= \pm \sqrt{(0.6132)(1.3598)} = \pm 0.9131$$

$r = 0.9131$ ($\because byx > 0,\ bxy > 0$)

Q2) The data about the sales and advertisement expenditure of a company is given below:

~~Sales~~

	Sales	Adv exp
Means	: 40	6
Standard deviation	: 10	1.5

Coefficient of correlation : $r = 0.9$

a) Estimate the likely sales for a proposed advertisement expenditure of ₹10 cr.

b) What should be the advertisement expenditure if the company proposes a sales of 60 or of rupees)

Find $x = ?$ for $y = 10$

Find $y = ?$ for $x = 60$.

Sol: The line of regression of y on x:

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

Here $\bar{x} = 40, \bar{y} = 6, \sigma_x = 10, \sigma_y = 1.5, r = 0.9$

$$y - 6 = \frac{(0.9)(1.5)}{10}(x - 40)$$

$$y = 6 + 0.135(x - 40) \quad \text{--- } (1)$$

Here, $x = 60$, $y = 6 + 0.135(60 - 40) = 8.7$

The line of regression of x on y is

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

$$x - 40 = \frac{0.9 \times 10}{1.5}(y - 6)$$

$$x = 40 + 6(y - 6) \quad \text{--- } (2)$$

when $x = 10$, we get

$$x = 40 + 6(10 - 6) = 64$$

Q3) From the following data, find :

① The 2 regression coefficients

② " to equation

③ Coefficients of correlation b/n the mark in Economics & Statistics

④ The most likely marks in statistics when Economics marks r 30.

Given data :

x : (Economics)	25	28	35	32	31	36	29	38	34	32
y : (Stats)	43	46	49	41	36	32	31	30	33	39

$\bar{x} = 32$, $\bar{y} = 38$, $\sigma_x =$, $\sigma_y =$

$140 = \Sigma dx^2$

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

$\Sigma dy^2 = 398$

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

$\Sigma dx\, dy =$

$$b_{yx} = \frac{cov(x,y)}{\sigma_x^2} = \frac{r\sigma_y}{\sigma_x} = \frac{\Sigma dx\, dy}{\Sigma dx^2} = -0.6643$$

$$b_{xy} = \frac{cov(x,y)}{\sigma_y^2} = \frac{r\sigma_x}{\sigma_y}$$

$= -0.2337$

$r = -0.394$

$y = 39$, $x = 30$

$= 34.2$

x	y	$dx = x - \bar{x}$	$dy = y - \bar{y}$	$dxdy$	dx^2	dy^2
25	[illegible]3					
28	[illegible]6					
35	[illegible]9					
32	[illegible]1					
3[illegible]						
3[illegible]						
2[illegible]						
3[illegible]						
3[illegible]						
32						