CHANAKYA 2.0

For CA Foundation

Regression Analysis

QUANTITATIVE APTITUDE

By Anurag Chauhan

TOPICS TO BE COVERED

01 Meaning

02 Regression Lines

03 Probable Error

Regression Analysis

Regression analysis is a **method (Statistical Tool)** to **predict the value** of a **variable based** on the value of **another variable.**

Statistical tool

⇓

Prediction of Dependent variable

⇓

with the help of independent variable.

mrx

variable

Independent

Dependent

Adv	Sales
100 →	1000
120 →	1500
140 →	2000
150 →	3000
200 →	?

Sales (y)
Adv. (x)

$$y = 1000x + 200$$ → Regression y on x

(Relation)

find estimated value of Sales when Adv. budget is ₹30?

Sol:

$x = 30$

Now

$$y = 1000x + 200$$
$$= 1000(30) + 200$$
$$= 30,000 + 200 = 30,200$$

Linear Regression Equation

y on x

$\Downarrow$

$$\boxed{y' = a + bx'}$$

$\Downarrow$

x is independent

y is Dependent

Linear Equation

$\Downarrow$

Power of x & y is one

Linear Regression Equation

x on y

$\Downarrow$

$$\boxed{x' = a + by'}$$

$\Downarrow$

y is independent

x is Dependent

Lines are Derived using Least Square Method

Eg Regression Equation

y on x : $y = 2x + 3$

x on y : $x = 0.6y + 10$

i> find value of x if $y = 20$

ii> find value of y if $x = 50$

$\boxed{\text{Sol:}}$ i) $y = 20$ (Independent)

Dependent $x = 0.6y + 10$

$= 0.6(20) + 10$

$= 12 + 10$

$x = 22$

ii> If $x = 50$ → Independent

$y = 2x + 3$

Depend

$= 2(50) + 3$

$= 100 + 3$

$= 103$

How to find two Regression Equations?

(Line of Best fit)
Regression Equation (Line)
y on x

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

$\Downarrow$ Simplify $\Downarrow$

$$\boxed{y = a + bx}$$

Standard form

(Line of Best fit)
Regression Equation (Line)
x on y

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

$\Downarrow$ Simplify $\Downarrow$

$$\boxed{x = a + by}$$

Standard form

$\bar{x}$ = mean of x
$\bar{y}$ = mean of y
b_{yx} = Regression coefficient
b_{xy} = Regression coefficient

If $\bar{x} = 10$ & $\bar{y} = 15$

$b_{yx} = 0.6$ & $b_{xy} = 0.8$

Find two regression Lines

Also find the value of x at which y=35

Sol: $\bar{x} = 10$, $\bar{y} = 15$, $b_{yx} = 0.6$ & $b_{xy} = 0.8$

Regression Line y on x

$$y - \bar{y} = b_{yx}(x - \bar{x})$$
$$y - 15 = 0.6(x - 10)$$
$$y - 15 = 0.6x - 6$$
$$y = 0.6x - 6 + 15$$
y on x: $$\boxed{y = 0.6x + 9}$$

Regression Line x on y

$$x - \bar{x} = b_{xy}(y - \bar{y})$$
$$x - 10 = 0.8(y - 15)$$
$$x - 10 = 0.8y - 12$$
$$x = 0.8y - 12 + 10$$
x on y: $$\boxed{x = 0.8y - 2}$$

at $y = 35$

$$x = 0.8y - 2$$
$$x = 0.8(35) - 2$$
$$= 28 - 2$$
$$x = 26$$

Regression Line

y on x

$\Downarrow$

$y = a + bx$

$\Downarrow$

$b = b_{yx} \Rightarrow$ Regression cofficients $\Rightarrow b_{xy} = b$

$\Downarrow$

Approx change in y due to one unit change in x

Regression Line

x on y

$\Downarrow$

$x = a + by$

$\Downarrow$

$\Downarrow$

Approx change in x due to one unit change in y

Calculations of Regression coefficents

$$\# \quad b_{yx} = \frac{\text{cov}\cdot(x, y)}{\sigma_x^{2}}$$

$$\# \quad b_{yx} = \frac{\Sigma (x_i - \bar{x})\cdot(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2}$$

$$\# \quad b_{yx} = \frac{\Sigma xy - \frac{\Sigma x \times \Sigma y}{N}}{\Sigma x^2 - \frac{(\Sigma x)^2}{N}}$$

$$\# \quad b_{yx} = r\,\frac{\sigma_y}{\sigma_x}$$

$$\# \quad b_{xy} = \frac{\text{cov}\cdot(x, y)}{\sigma_y^{2}}$$

$$\# \quad b_{xy} = \frac{\Sigma (x_i - \bar{x})\cdot(y_i - \bar{y})}{\Sigma (y_i - \bar{y})^2}$$

$$\# \quad b_{xy} = \frac{\Sigma xy - \frac{\Sigma x \times \Sigma y}{N}}{\Sigma y^2 - \frac{(\Sigma y)^2}{N}}$$

$$\# \quad b_{xy} = r\,\frac{\sigma_x}{\sigma_y}$$

Find two regression Lines if

X: 2 4 6 8

Y: 3 6 9 12

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
2	3	4	9	6
4	6	16	36	24
6	9	36	81	54
8	12	64	144	96
20	30	120	270	180

$\bar{X} = \frac{\Sigma x_i}{N} = \frac{20}{4} = 5$

$\bar{y} = \frac{\Sigma y_i}{N} = \frac{30}{4} = 7.5$

$$b_{yx} = \frac{\Sigma xy - \frac{\Sigma x \times \Sigma y}{N}}{\Sigma x^2 - \frac{(\Sigma x)^2}{N}}$$

$$= \frac{180 - \frac{20 \times 30}{4}}{120 - \frac{(20)^2}{4}}$$

$$= \frac{30}{20}$$

$$b_{yx} = 1.5$$

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

$$= \frac{180 - \frac{20 \times 30}{4}}{270 - \frac{(30)^2}{4}}$$

$$= \frac{30}{270 - 225}$$

$$= \frac{30}{45}$$

$$b_{xy} = \frac{2}{3} = 0.6666$$

$$\boxed{1.5 \times \frac{2}{3} = \frac{3}{3} = 1}$$

Line y on x

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

$$y - 7.5 = 1.5(x - 5)$$

$$y - \cancel{7.5} = 1.5x - \cancel{7.5}$$

y on x $\boxed{y = 1.5x}$

Line x on y

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

$$x - 5 = \frac{2}{3}(y - 7.5)$$

$$x - \cancel{5} = \frac{2}{3}y - \cancel{5}$$

x on y $\boxed{x = \frac{2}{3}y}$

Find the regression coefficient x on y (b_{xy})

If $\sum x = 40, \sum y = 20, \sum xy = 800, \sum x^2 = 2500$

$\sum y^2 = 1500$ & N=10

A. -0.493

B. $\boxed{0.493}$

C. -0.749

D. None

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

$$= \frac{800 - \frac{40 \times 20}{10}}{1500 - \frac{(20)^2}{10}} = \frac{720}{1460} = 0.493$$

Find the regression coefficient of $\boxed{y \text{ on } x}$

	Price (x_i)	Demand (y_i)
Arithmetic Mean	$\bar{x} = 20$	$55 = \bar{y}$
Standard deviation	$2 = \sigma_x$	$5 = \sigma_y$
Correlation Coefficient	$\boxed{r=0.6}$	

A. 1.5

B. -3

C. 3

D. None

$$b_{yx} = r\frac{\sigma_y}{\sigma_x}$$

$$= (0.6) \times \frac{5}{2}$$

$$b_{yx} = 1.5$$

$$b_{xy} = r\frac{\sigma_x}{\sigma_y}$$

$$= 0.6 \times \frac{2}{5}$$

$$= 0.24$$

1.5×0.24

$= 0.36 < 1$

Find the regression coefficient of y on x

Covariance of x and y = 121

SD of x =15 & SD of y=14

A. 0.54

B. 0.65

C. 0.6875

D. None

$$b_{yx} = \frac{\text{cov.}(x, y)}{\sigma_x^2}$$

$$= \frac{121}{(15)^2}$$

$$= 0.53777$$

or 0.54

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

$$= \frac{121}{(14)^2}$$

$$= 0.6173$$

Q. Find the regression equation of y on x

	Price (x)	Demand (y)
Arithmetic Mean	$\bar{x} = 44$	$\bar{y} = 58$
Standard deviation	5.60	6.30
Correlation Coefficient		r=0.48

(A) Y=1.0745x-4.71

(B) Y=1.0745x-5.71

(C) Y=1.0725x-3.71

(D) None

$$b_{yx} = r\frac{\sigma_y}{\sigma_x}$$

$$= 0.48 \times \frac{6.30}{5.60}$$

$$= 0.54$$

Rg. Eq. y on x

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

$$y - 58 = 0.54(x - 44)$$

$$y - 58 = 0.54x - 23.76$$

$$\boxed{y = 0.54x + 34.24}$$

$$b_{yx} = r\frac{\sigma_y}{\sigma_x} \quad \& \quad b_{xy} = r\frac{\sigma_x}{\sigma_y}$$

$$b_{yx} \times b_{xy} = r\frac{\cancel{\sigma_y}}{\cancel{\sigma_x}} \times r\frac{\cancel{\sigma_x}}{\cancel{\sigma_y}}$$

$$b_{yx} \times b_{xy} = r^2$$

$$\boxed{r = \pm\sqrt{b_{yx} \times b_{xy}}}$$

$$\boxed{r^2 \leq 1} \quad \boxed{b_{yx} \times b_{xy} \leq 1}$$

b_{yx} & b_{xy} both have same sign

If b_{yx} & b_{xy} are positive then 'r' will be positive

If b_{yx} & b_{xy} are negative then 'r' will be negative.

CA WALLAH

$$\boxed{r = \pm\sqrt{b_{yx} \times b_{xy}}}$$

'r' is the Geometric mean of two Regression coefficients

If $b_{xy} = 0.8$ & $b_{yx} = 0.46$

Then the value of r ?

$$r = +\sqrt{b_{yx} \times b_{xy}}$$

$$= \sqrt{0.8 \times 0.46}$$

$$= 0.6066$$

A. 0.26

B. 0.607

C. 0.75

D. None

If $b_{xy} = -0.25$ & $b_{yx} = -0.37$

Then the value of r ?

A. -0.304

B. 0.304

C. 0

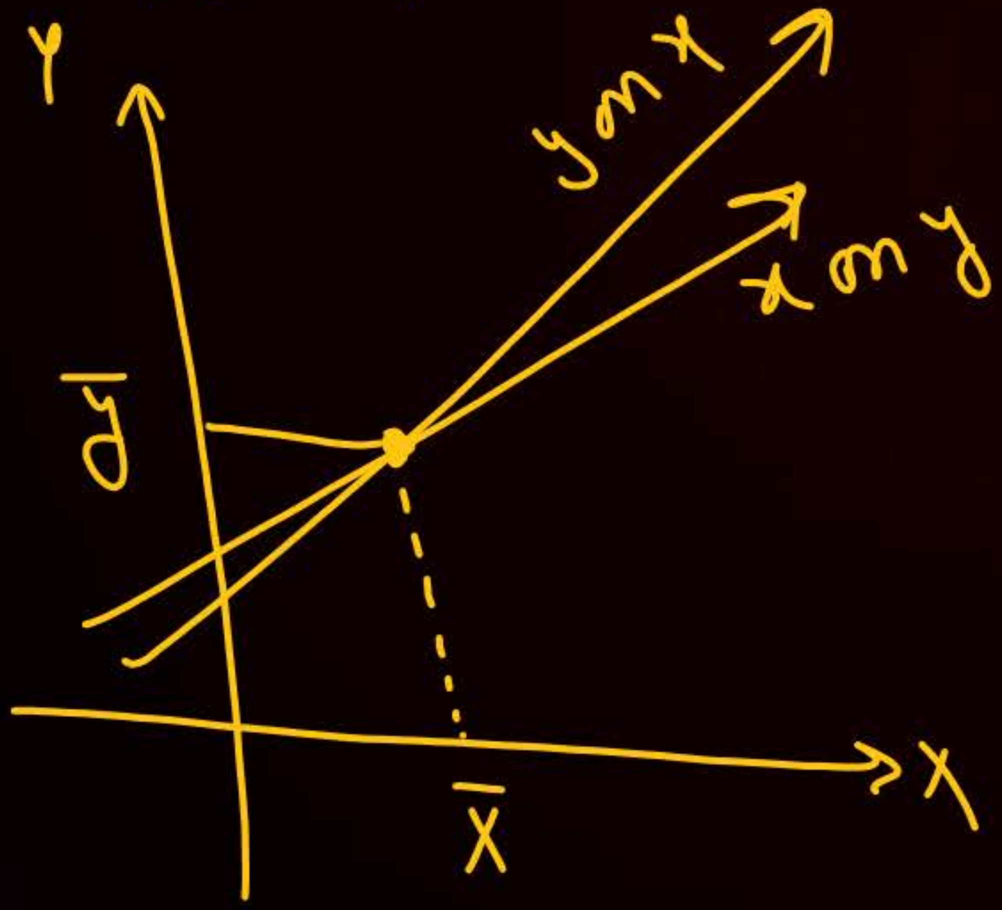
D. None

$$r = -\sqrt{(-0.25)(-0.37)}$$

$$= -\sqrt{0.0925}$$

$$= -0.304$$

Two Regression Lines intersect each other at the point $(\bar{x}, \bar{y})$ i.e. their Arithmetic Means



Find the mean values of X and Y when the equations of the two lines of regression are:

$$X + 2Y - 5 = 0 \ \ and \ \ 2X + 3Y - 8 = 0$$

sol:

$x + 2y = 5$ —— ① $\times 3$

$2x + 3y = 8$ —— ② $\times 2$

$3x + 6y = 15$

$4x + 6y = 16$

$-\quad -\quad -$

$-x = -1$

$x = 1$

Now

$x + 2y = 5$

$1 + 2y = 5$

$2y = 4$

$y = 2$

$\bar{X} = 1$

$\bar{y} = 2$

If two Regression Equations are given, how to identify which Equation is 'y on x' & which one is "x on y"

⇓

?

$\underbrace{y \text{ on } x}$

$\Downarrow$

$A_1 x + B_1 y + C_1 = 0$

$\Downarrow$

convert instandard form

$(y = a + bx)$

$B_1 y = -C_1 - A_1 x$

$y = -\frac{C_1}{B_1} - \frac{A_1}{B_1} x$

$\# \boxed{b = b_{yx} = -\frac{A_1}{B_1}}$

$\underbrace{x \text{ on } y}$

$\Downarrow$

$A_2 x + B_2 y + C_2 = 0$

$\Downarrow$

convert in standard form

$(x = a + by)$

$A_2 x = -C_2 - B_2 y$

$x = -\frac{C_2}{A_2} - \frac{B_2}{A_2} y$

$\# \boxed{b = b_{xy} = -\frac{B_2}{A_2}}$

Eg y on x : $\boxed{y = 0.5x + 10}$

x on y : $\boxed{x = 0.6y + 20}$

find b_{yx} & b_{xy}

Sol:

y on x	x on y
$\Downarrow$	$\Downarrow$
$y = a + bx$	$x = a + by$
$b_{yx} = 0.5$	$b = b_{xy} = 0.6$

If 7x-3y=18 be a regression equation of y on x then regression coefficient is

y on x : $\boxed{7x - 3y = 18}$

$$\Downarrow$$

$$b_{yx} = -\frac{A}{B} = -\frac{(7)}{(-3)} = \frac{7}{3}$$

or

$$7x - 3y = 18$$

$$-3y = -7x + 18$$

$$y = \frac{7}{3}x + \frac{18}{-3}$$

A. 7/3

B. -7/3

C. 3/7

D. None

Eg Regression Line

y on x: $2x+3y+7=0$ & x on y: $5x+2y+18=0$

find two Regression coefficients.

Sol:

y on x

⇓

$\underset{A}{2}x+\underset{B}{3}y+\underset{C}{7}=0$

$b_{yx} = -\frac{A}{B} = -\frac{2}{3}$

x on y

⇓

$\underset{A}{5}x+\underset{B}{2}y+\underset{C}{18}=0$

$b_{xy} = -\frac{B}{A} = -\frac{2}{5}$

Now

$r = \pm\sqrt{b_{yx} \times b_{xy}}$

$= -\sqrt{\left(-\frac{2}{3}\right)\left(-\frac{2}{5}\right)}$

$= -\sqrt{\frac{4}{15}}$

$= -0.516$

Find the means of X and Y variables and the coefficient of correlation between them from the following two regression equations:

$$2Y - X - 50 = 0 \quad and \quad 3Y - 2X - 10 = 0$$

Sol:

$-x + 2y = 50$ — ① $\times 3$

$-2x + 3y = 10$ — ② $\times 2$

$-3x + 6y = 150$

$-4x + 6y = 20$

$+ \quad - \quad -$

$x = 130$

Now

$-x + 2y = 50$

$-130 + 2y = 50$

$2y = 180$

$y = 90$

$\boxed{\bar{x} = 130 \\ \bar{y} = 90}$

Let

y on x: $-x + 2y = 50$

x on y: $-2x + 3y = 10$

$b_{yx} = -\frac{A}{B} = -\frac{(-1)}{2} = \frac{1}{2}$

$b_{xy} = -\frac{B}{A} = -\frac{3}{(-2)} = \frac{3}{2}$

$b_{yx} \times b_{xy} = \frac{1}{2} \times \frac{3}{2} = \frac{3}{4}$

$= 0.75 < 1$

$$r = \sqrt{b_{yx} \times b_{xy}}$$

$$= +\sqrt{\frac{1}{2} \times \frac{3}{4}}$$

$$= +\sqrt{\frac{3}{4}}$$

$$= +\sqrt{0.75}$$

$$= 0.866$$

The equations of two lines of regression are:

$$8X - 10Y + 66 = 0 \quad and \quad 40X - 18Y = 214$$

The variance of X is 9. Find

(1) The mean value of X and Y.

(2) Correlation coefficient between X and Y.

Standard deviation of Y.

i) Sol:

$8x - 10y = -66$ — ① $\times 5$

$40x - 18y = 214$ — ② $\times 1$

$\cancel{40x} - 50y = -330$

$\pm \cancel{40x} \mp 18y = \pm 214$

$-32y = -544$

$y = \frac{544}{32}$

$y = 17$

now

$8x - 10y = -66$

$8x - 10(17) = -66$

$8x = -66 + 170$

$8x = 104$

$x = 13$

$$\boxed{\begin{aligned} \bar{X} &= 13 \\ \bar{y} &= 17 \end{aligned}}$$

ii) Wt

y on x

$8x - 10y + 66 = 0$

$b_{yx} = -\left(\frac{A}{B}\right)$

$= -\left(\frac{8}{-10}\right)$

$b_{yx} = \frac{4}{5}$

x on y

$40x - 18y = 214$

$b_{xy} = -\left(\frac{B}{A}\right)$

$= -\frac{(-18)}{40}$

$= +\frac{9}{20}$

Now $b_{yx} \times b_{xy} = \frac{4}{5} \times \frac{9}{20} = \frac{36}{100} < 1$

$r = \sqrt{b_{yx} \times b_{xy}}$

$r = \sqrt{\frac{4}{5} \times \frac{9}{20}}$

$= \sqrt{0.36}$

$\boxed{r = 0.6}$

iii) $\sigma_x^2 = 9$

$\sigma_x = 3$

$r = 0.6$

$b_{yx} = \frac{4}{5}$

Now

$b_{yx} = r\frac{\sigma_y}{\sigma_x}$

$\frac{4}{5} = 0.6 \times \frac{\sigma_y}{3}$

$\sigma_y = \frac{0.8}{0.2} = 4$

Change Of Origin

Eg

x_i	y_i	x_i^2	y_i^2	x_iy_i
1	1	1	1	1
2	3	4	9	6
3	5	9	25	15
6	9	14	35	22

$$b_{yx} = \frac{\Sigma xy - \frac{\Sigma x \times \Sigma y}{N}}{\Sigma x^2 - \frac{(\Sigma x)^2}{N}}$$

$$= \frac{22 - \frac{6 \times 9}{3}}{14 - \frac{(6)^2}{3}} = \frac{4}{2} = 2$$

$U_i = x_i + 1$	$V_i = y_i + 2$	U^2	V^2	UV
2	3	4	9	6
3	5	9	25	15
4	7	16	49	28
9	15	29	83	49

$$b_{VU} = \frac{\Sigma UV - \frac{\Sigma U \times \Sigma V}{N}}{\Sigma U^2 - \frac{(\Sigma U)^2}{N}}$$

$$= \frac{49 - \frac{9 \times 15}{3}}{29 - \frac{(9)^2}{3}} = \frac{4}{2} = 2$$

Regression cofficients are independent of change of origin

If $u_i = x_i - A$ & $v_i = y_i - B$

$$b_{yx} = b_{vu} \quad \& \quad b_{xy} = b_{uv}$$

Find The regression Coefficient Of x on y = bxy = ?

If $\sum(x-58)=56$ $\sum(x-58)^2=3186$

$\sum(y-58)=12$ $\sum(y-58)^2=383$

$\sum(x-58)(y-58)=1195$ & N=7

A. -2.197

B. $\boxed{3.03}$

C. 0.372

D. None

$U = x - 58$ & $v_i = y_i - 58$

$\Sigma u = 56$

$\Sigma u^2 = 3186$

$\Sigma v = 12$

$\Sigma v^2 = 383$

$\Sigma uv = 1195$

$N = 7$

b_{xy}

$= b_{uv}$

$$= \frac{\Sigma uv - \frac{\Sigma u \times \Sigma v}{N}}{\Sigma v^2 - \frac{(\Sigma v)^2}{N}}$$

$$= \frac{1195 - \frac{56 \times 12}{7}}{383 - \frac{(12)^2}{7}}$$

$$= \frac{1099}{362.42} = 3.0323$$

Change Of Scale

Eg

x_i	y_i	x_i^2	y_i^2	x_iy_i
1	1	1	1	1
2	3	4	9	6
3	5	9	25	15
6	9	14	35	22

$$b_{yx} = \frac{\Sigma xy - \frac{\Sigma x \times \Sigma y}{N}}{\Sigma x^2 - \frac{(\Sigma x)^2}{N}}$$

$$b_{yx} = \frac{22 - \frac{6\times 9}{3}}{14 - \frac{(6)^2}{3}} = \frac{4}{2} = 2$$

$U_i = 3x_i$ & $V_i = 5y_i$

U_i	V_i	U_i^2	V_i^2	U_iV_i
3	5	9	25	15
6	15	36	225	90
9	25	81	625	225
18	45	126	875	330

$$b_{VU} = \frac{\Sigma UV - \frac{\Sigma U \times \Sigma V}{N}}{\Sigma U^2 - \frac{(\Sigma U)^2}{N}}$$

$$= \frac{330 - \frac{18\times 45}{3}}{126 - \frac{(18)^2}{3}} = \frac{60}{18} = \frac{10}{3}$$

$b_{yx} = 2$

$u_i = 3x_i$ & $v_i = 5y_i$

$b_{vu} = \frac{10}{3}$

$b_{vu} \neq b_{yx}$

$\boxed{b_{vu} = \frac{\text{scale of } y}{\text{scale of } x} \times b_{yx}}$

$\frac{10}{3} = \frac{5}{3} \times 2$

$\boxed{b_{uv} = \frac{\text{scale of } x}{\text{scale of } y} \times b_{xy}}$

Eg $b_{yx} = 3$

$v_i = 2y_i + 3$

$u_i = 10x_i + 20$

$b_{vu} = ?$

Sol:

$$b_{vu} = \frac{\text{scale of } y}{\text{scale of } x} \times b_{yx}$$

$$= \frac{2}{10} \times 3$$

$$b_{vu} = 0.6$$

Q. If the relation between two variable a and u is $u+3x=-15$ & other two variables v and y is $2y+5v=23$. If the regression coefficient of y on x is known as 0.80 What would be the regression coefficient of v on u ?

Sol: $u+3x=-15$ & $2y+5v=23$, $b_{yx}=0.80$

$b_{vu}=?$

$u=-\frac{3}{1}x-15$

$5v=-2y+23$

$v=-\frac{2}{5}y+\frac{23}{5}$

$b_{vu}=\frac{\text{Scale of } y}{\text{Scale of } x}\times b_{yx}$

$b_{vu}=\frac{\left(-\frac{2}{5}\right)}{\left(-\frac{3}{1}\right)}\cdot 0.80$

$=-\frac{2}{5}\times\left(-\frac{1}{3}\right)\times\frac{80}{100}$

$=\frac{16}{150}=\frac{8}{75}$

A. 8/75

B. 75/8

C. -8/75

D. None

y on x

$y = a + bx$

$\Downarrow$

Regression coefficient $= b_{yx}$

$\Downarrow$

Slope $(m_1) = b_{yx}$

x on y

$x = a + by$

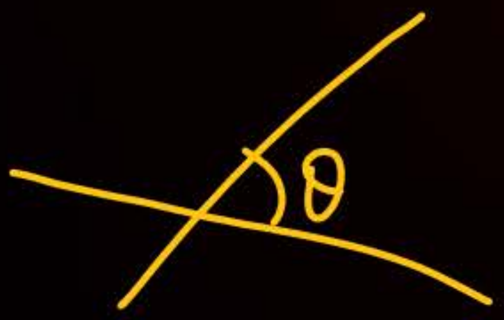
$\Downarrow$

Regression coefficient $= b_{xy}$

$\Downarrow$

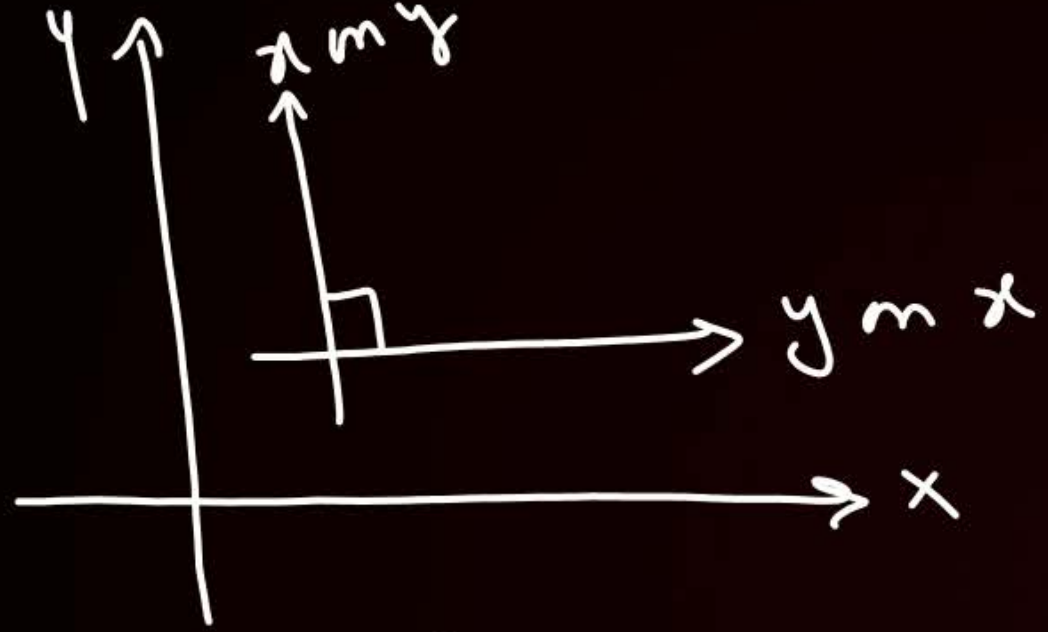
Slope $(m_2) = \dfrac{1}{b_{xy}}$

Angle b/w two Lines



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

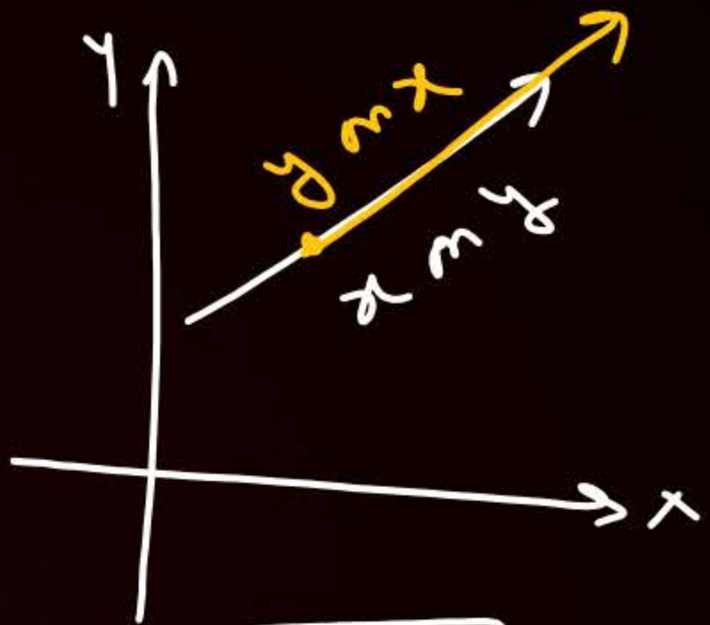
When two Regression lines are Perpendicular

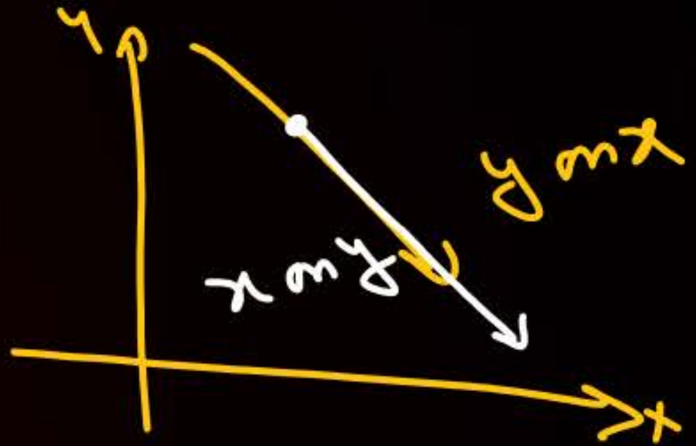


⇓

correlation $(r) = 0$

When two Lines are coincident to each other





$\boxed{r = 1}$ $\boxed{r = -1}$

$\boxed{r = \pm 1}$

$\boxed{r = \pm\sqrt{b_{yx} \times b_{xy}}}$

$\Downarrow$

GM of b_{yx} & b_{xy} is r

$r < \dfrac{b_{yx} + b_{xy}}{2}$

$r <$ AM of two Regression coefficient

Eg $b_{yx} = 0.8$ & $b_{xy} = 0.6$

$r = \sqrt{0.8 \times 0.6} = 0.6920$

$AM = \dfrac{0.8 + 0.6}{2} = 0.70$

$r < AM$

Eg $b_{yx} = 4$ & $b_{xy} = \frac{1}{4} = 0.25$

$r = \sqrt{4 + \frac{1}{4}} = \sqrt{1} = 1$

$AM = \dfrac{4 + 0.25}{2} = 2.125$

Concept Of probable Error

Population

$\Downarrow$

$(r - P.E.,\ r + P.E.)$ $\longrightarrow$ Sample

$$S.E. = \frac{1 - r^2}{\sqrt{n}}$$

$\Downarrow$

Standard Error

$$\text{Probable Error } (P.E.) = 0.6745\left(\frac{1 - r^2}{\sqrt{N}}\right)$$

$$P.E. = 0.6745 \times (S.E)$$

sample r $\longrightarrow$ Population

$$P.E. = 0.6745\,(S.E.)$$

where

$$S.E. = \frac{1-r^2}{\sqrt{N}}$$

$\Downarrow$

Correlation of population lies in the internal

$r - P.E., \; r + P.E.$

Range for the Correlation of Population

$$= \left[r - PE, \; r + P.E\right]$$

Compute the probable error assuming correlation is 0.8 from a sample of 25 pairs of items

$r = 0.8$

$N = 25$

$P.E = 0.6745 \times \frac{1 - r^2}{\sqrt{N}}$

$= 0.6745 \times \frac{[1 - (0.8)^2]}{\sqrt{25}}$

$P.E. = 0.048564$

A. 0.8406

B. 0.0486

C. 0.0086

D. 0.8456

If r=0.7 and n=64 , Find out the probable error of correlation and determine the limit for the correlation of the population

Sol:

$r = 0.7$

$n = 64$

$$P.E = 0.6745 \times \frac{1-r^2}{\sqrt{N}}$$

$$= 0.6745 \times \left(\frac{1-0.7^2}{\sqrt{64}}\right)$$

$$= 0.6745 \times \left[\frac{1-0.49}{8}\right]$$

$$P.E. = 0.043$$

Correlation of population

$$= (r - P.E,\ r + P.E.)$$

$$= [0.7 - 0.043,\ 0.7 + 0.043]$$

$$= [0.657,\ 0.743]$$

$$\boxed{r < 6\,(P.E)}$$

$\Downarrow$

No Evidence of correlation

$$\boxed{r > 6\,(P.E)}$$

Evidence of correlation exists.

Eg $r = 0.8$

$P.E = 0.2$

$6(P.E) = 6 \times 0.2 = 1.2$

$r < 6(PE)$ No Evidence

Eg $r = 0.9$

$P.E = 0.1$

$6(PE) = 0.6$

$r > 6(PE)$

Evidence of correlation exists

Bivariate Data

Univariate Data

⇓

one variable

	marks in maths
A	6
B	4
C	5
D	2

Bivariate Data

	(x_i) Maths	(y_i) Stats
A	8	9
B	7	6
C	3	4
D	10	8

Marks of 15 students of class 12^{th} in Maths and physics are awarded out of 15 and 40 respectively

maths = 15 physics = 40

Let x and y represents marks in maths and physics
Prepare a bivariate frequency table

(14,38) (12,37) , (3 ,8) ,(8,18) (9,35)
(1,6) (7,28) , (8,35) , (11, 31) ,(1, 12)
(13,2) ,(14, 29) ,(14,19) ,(6,39) ,(8,7)

<table>
<tr><th>Maths \ Physics</th><th>0-10</th><th>10-20</th><th>20-30</th><th>30-40</th><th>Total</th></tr>
<tr><td>0-5</td><td>|| = 2</td><td>| = 1</td><td>0</td><td>0</td><td>3</td></tr>
<tr><td>5-10</td><td>| = 1</td><td>| = 1</td><td>| = 1</td><td>||| = 3</td><td>6</td></tr>
<tr><td>10-15</td><td>| = 1</td><td>| = 1</td><td>| = 1</td><td>||| = 3</td><td>6</td></tr>
<tr><td>Total</td><td>4</td><td>3</td><td>2</td><td>6</td><td>15</td></tr>
</table>

Marginal Distribution of marks in maths

Marks	No of students
0-5	3
5-10	6
10-15	5

Marginal Distribution of marks in Physics

Marks	No of students
0-10	4
10-20	3
20-30	2
30-40	6

Conditional Distribution of marks in maths when student score '0-10' in Physics

Marks in maths	f_i
0-5	2
5-10	1
10-15	1
	4

Condition Dist of marks in Physics when student score $\boxed{10-15}$ in maths

Marks in Physics	f_i
0-10	1
10-20	1
20-30	1
30-40	3
	6


Savings
FINANCE
THANK
YOU
KEEP REVISING
&
STAY MOTIVATED !!