Correlation and Regression

Page: Subhash
Date:

Correlation → We know that the quantity of measure of relationship between X and Y is given by covariance between X and Y

Here Covariance between X and Y = $Cov(X,Y) = E(XY) - E(X)E(Y)$

and $-\infty < Cov(X,Y) < \infty$,

Correlation blw X and Y = $r(X,Y) = \frac{Cov(X,Y)}{\sigma_x \sigma_y}$

where σ_x = SD of X and σ_y = SD of Y

Limits for correlation coefficient

Let $E(X) = \mu_x$ and $E(Y) = \mu_y$

Now consider $E\left[\left\{\frac{X-\mu_x}{\sigma_x} \pm \frac{Y-\mu_y}{\sigma_y}\right\}^2\right] \geq 0$

$$E\left[\left(\frac{X-\mu_x}{\sigma_x}\right)^2 + \left(\frac{Y-\mu_y}{\sigma_y}\right)^2 \pm 2\left(\frac{X-\mu_x}{\sigma_x}\right)\left(\frac{Y-\mu_y}{\sigma_y}\right)\right] \geq 0$$

$$E\left[\left(\frac{X-\mu_x}{\sigma_x}\right)^2\right] + E\left[\left(\frac{Y-\mu_y}{\sigma_y}\right)^2\right] \pm 2E\left[\frac{(X-\mu_x)(Y-\mu_y)}{\sigma_x \sigma_y}\right] \geq 0$$

$$1 + 1 \pm 2\rho(X,Y) \geq 0$$

$$1 \pm r(X,Y) \geq 0$$

$$\boxed{-1 \leq r(X,Y) \leq 1}$$

$E\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right]$

$\frac{1}{\sigma_x^2} E\left[(x-\mu_x)^2\right]$

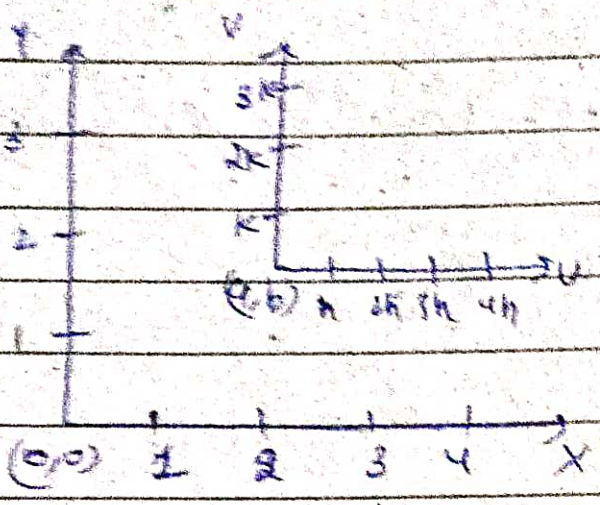
$\frac{\sigma_x^2}{\sigma_x^2} = 1$

Page No. Date

Note Correlation coefficient is independent of origin and scale

Let $U = \frac{X-a}{h}$ $V = \frac{Y-b}{k}$ where a, b, h, k are constants

then $r(X,Y) = r(U,V)$



Formulas to calculate correlation coefficient

(i) For ungrouped data → If X takes the values x_i, $i=1,2,\ldots n$ and Y takes the values y_i, $i=1,2,\ldots n$ then (x_i, y_i), $i=1,2,\ldots n$ are the n pairs of observations.

$$E(X) = \frac{\sum_{i=1}^{n} x_i}{n} \qquad E(Y) = \frac{\sum_{i=1}^{n} y_i}{n} \qquad E(XY) = \frac{\sum_{i=1}^{n} x_i y_i}{n}$$

$$\sigma_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - (E(X))^2} \qquad \sigma_y = \sqrt{\frac{1}{n}\sum_{i=1}^{n} y_i^2 - (E(Y))^2}$$

Correlation coefficient b/w X and Y

$$\rho(X,Y) = \frac{E(XY) - E(X)E(Y)}{\sigma_x \sigma_y}$$

② For grouped data (Bivariate frequency distribution table)

Let x_i, $i = 1, 2, \ldots n$ be the mid points of the n class intervals defined for X with respective frequencies f_{x_i}, $i = 1, 2, \ldots n$

Let y_j, $j = 1, 2, \ldots m$ be the mid point of m class interval defined for Y with respective frequencies f_{y_j} $j = 1, 2, \ldots m$

Let f_{ij} be the frequency of i^{th} class interval of X correlated with j^{th} class interval of Y

Let $N = \sum_{i=1}^{n} f_{x_i} = \sum_{j=1}^{m} f_{y_j} = \sum_{i=1}^{n}\sum_{j=1}^{m} f_{ij}$

Now we can calculate

$$E(X) = \frac{\sum_{i=1}^{n} f_{x_i} x_i}{N} \qquad E(Y) = \frac{\sum_{j=1}^{m} f_{y_j} y_j}{N} \qquad E(XY) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} f_{ij} x_i y_j}{N}$$

$$\sigma_x = \sqrt{\frac{1}{N}\sum_{i=1}^{n} f_{x_i} x_i^2 - [E(X)]^2} \qquad \sigma_y = \sqrt{\frac{1}{N}\sum_{j=1}^{n} f_{y_j} y_j^2 - [E(X)]^2}$$

correlation coeff b/w X and Y = $\rho(X,Y) = \frac{E(XY) - E(X)E(Y)}{\sigma_x \sigma_y}$

Problems

Q1. Calculate the correlation coefficient for the following heights in inches of fathers (X) and their children (Y)

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

Sol^n

$$r(X,Y) = \frac{E(XY) - E(X)E(Y)}{\sigma_x \sigma_y}$$

$n=8$ 8 pairs of information

$$E(X) = \frac{\sum_{i=1}^{8} x_i}{n} = \frac{544}{8} \qquad E(Y) = \frac{\sum_{i=1}^{8} y_i}{8} = \frac{552}{8}$$

$$E(XY) = \frac{\sum_{i=1}^{8} x_i y_i}{n} = \frac{3756}{8}$$

$\Sigma x_i^2 = 37028$

$\Sigma y_i^2 = 38132$

$E(X) = 68 \quad E(Y) = 69 \quad E(XY) = 4695$

$\sigma_x = 2.1213 \quad \sigma_y = 2.3452$

$\therefore r(X,Y) = 0.603$

Q2 To calculate the correlation coefficient btw two variables X and Y from 25 pairs of observations, the following data is obtained.

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

$\Sigma XY = 508$

Later it was discovered that the two pairs of data values (6,14), (8,6) are considered instead of

Page: Subhash
Date:

correct pair of values (8, 18), (6, 8) are wrong. Hence obtain correlation coff for the data.

Sol: $\Sigma X = 125$ and $\Sigma Y = 100$ will remain same

as $6+8=14$ or $8+6=14$

$14+6=20$ or $12+8=20$

$\Sigma X^2 = 650$ remain same

$8^2 = 64$, $6^2 = 36$

$\Sigma Y^2 = 460 - 196 - 36 + 144 + 64$

$\Sigma Y^2 = 436$

$\Sigma XY = 508 - 84 - 48 + 96 + 48 = 520$

$E(X) = \frac{\Sigma X}{n} = \frac{125}{25} = 5$

$E(Y) = \frac{\Sigma Y}{n} = \frac{100}{25} = 4$

$E(XY) = \frac{520}{n} = \frac{104}{5} = 20.8$

$\sigma_x = \sqrt{\frac{1}{n}\Sigma x_i^2 - (E(X))^2}$

$= \sqrt{\frac{650}{25} - 25}$

$\sqrt{26-25} = 1$

$\sigma_y = \sqrt{\frac{1}{n}\Sigma y_i^2 - (E(Y))^2}$

$\sigma_y = \sqrt{\frac{436}{25} - 16}$

$\sigma_y = \sqrt{19.81 - 16} = \sqrt{3.81}$

$= 1.95$

$r(X,Y) = \frac{20.8 - 20}{1 \times 1.95} = \frac{0.8}{1.95} = 0.41$

Page Subhash

Date

Q3 Calculate the correlation coefficient for the following bivariate frequency distribution table.

mid y	20 (y_1)	30 (y_2)	40 (y_3)	50 (y_4)	60 (y_5)	70 (y_6)	Total freq of x
mid x \ y	15-25	25-35	35-45	45-55	55-65	65-75	f_{x_i}
20	15-25	1 (f_{11})	1 (f_{12})				2 (f_{x_1})
30	25-35	2 (f_{21})	21 (f_{22})	1 (f_{23})			24 (f_{x_2})
40	35-45		4	10	1		15 (f_{x_3})
50	45-55			3	6	1	10 (f_{x_4})
60	55-65		-		2	4	2	8 (f_{x_5})
70	65-75					1	2	3 (f_{x_6})
Total freq of y		3 (f_{y_1})	26 (f_{y_2})	14 (f_{y_3})	9 (f_{y_4})	6 (f_{y_5})	4 (f_{y_6})	

$N = 62$

- $E(X) = \frac{\sum_{i=1}^{n} f_{x_i} x_i}{N}$ $\quad E(Y) = \frac{\sum_{j=1}^{m} f_{y_j} y_j}{N}$

$E(XY) = \sum_{i=1}^{n}\sum_{j=1}^{m} f_{ij} x_i y_j$

$n = 6$ $\quad m = 6$

$\sigma_x = \sqrt{\frac{1}{N}\sum_{i=1}^{n} f_{x_i} x_i^2 - [E(X)]^2}$ $\quad \sigma_y = \sqrt{\frac{1}{N}\sum_{j=1}^{m} f_{y_j} y_j^2 - [E(Y)]^2}$

$\sum_{i=1}^{n} f_{x_i} x_i = 2280$ $\quad \sum_{j=1}^{m} f_{y_j} y_j = 2220$

$\sum_{i=1}^{n}\sum_{j=1}^{m} f_{ij} x_i y_j = 103800$

$\sum_{i=1}^{n} f_{x_i} x_i^2 = 106800$ $\quad \sum_{j=1}^{m} f_{y_j} y_j^2 = 102600$

Q. Calculate the correlation coefficient for the following bivariate frequency distribution table.

Y → / X ↓	59-62 (y_1, 60.5)	63-66 (y_2, 64.5)	67-70 (y_3, 68.5)	71-74 (y_4, 72.5)	75-78 (y_5, 76.5)	Total freq. of X
100 (x_1)	2	1				3 f_{x_1}
120 (x_2)	7	8	4	2		21 f_{x_2}
140 (x_3)	5	15	22	7	1	50 f_{x_3}
160 (x_4)	2	12	63	19	5	101 f_{x_4}
180 (x_5)		7	28	32	12	79 f_{x_5}
200 (x_6)		2	10	20	7	39 f_{x_6}
220 (x_7)			1	4	2	7 f_{x_7}
Total freq. of Y	16	45				

$n=7$	$m=5$

Lines of Regression

Let (x_i, y_i), $i = 1, 2, \ldots n$ be the given set of n points

Aim → We want to get a straight line passing through the given set of n points.

If not, we want to get a best fit of straight line for the given set of n point.

Let $Y = a + bX$ be the best fit of straight line for the given set of n points if X is dependent on X and this best fit of straight line is known as regression lines of Y on X.

That means, we want to find values of a and b such that

$$S = \sum_{i=1}^{n} (y_i - (a + bx_i))^2 \text{ is minimum}$$

Here x_i and y_i are known and (a, b) are unknown

① $f(x)$ → $f'(x) = 0$ → roots

$$f''(x)\Big|_{\text{roots}} = \begin{cases} > 0 & \text{at that root } f(x) \text{ is min} \\ < 0 & \text{at that root } f(x) \text{ is max} \end{cases}$$

② $f(x, y)$ → $\boxed{\frac{\partial f}{\partial x} = 0, \frac{\partial f}{\partial y} = 0}$ → solving this we get stationary points

$\boxed{\frac{\partial^2 f}{\partial x^2}, \frac{\partial^2 f}{\partial y^2}, \frac{\partial^2 f}{\partial x \partial y}}$ → evaluate them at stationary point.

(i) if $\left.\frac{\partial^2 f}{\partial x^2}\right|_{\text{st.point}} > 0$, $\left[\frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2\right]_{\text{st. point}} > 0$

↓

at that st. point $f(x,y)$ is minimum

(ii) if $\left.\frac{\partial^2 f}{\partial x^2}\right|_{\text{st.point}} < 0$, $\left[\frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2\right]_{\text{st.point}} > 0$

↓

at that st. point $f(x,y)$ is maximum

Here we have

$$S = \sum_{i=1}^{n} (y_i - (a+bx_i))^2 = S(a,b)$$

Here S is a function of a and b

for S to be minimum, the necessary condition are $\frac{\partial S}{\partial a} = 0$, $\frac{\partial S}{\partial b} = 0$.

$$\frac{\partial S}{\partial a} = 0 \Rightarrow na + b\Sigma x_i = \Sigma y_i \quad -(1)$$

$$\frac{\partial S}{\partial b} = 0 \Rightarrow a\Sigma x_i + b\Sigma x_i^2 = \Sigma x_i y_i \quad -(2)$$

Solving (1) and (2) we get a and b

$$b = r\frac{\sigma_y}{\sigma_x} \qquad r = \text{coeff of correlation} = \frac{Cov(X,Y)}{\sigma_x \sigma_y}$$

$$a = \frac{1}{n}(\Sigma y_i - b\Sigma x_i)$$

$$a = \frac{\Sigma y_i}{n} - b\frac{\Sigma x_i}{n} = E(Y) - bE(X)$$

$$\Rightarrow E(Y) = a + bE(X)$$

$$y = a + bx$$

Here $(E(X), E(Y))$ is a point of st. line $y = a+bx$

Now we'll check condition for minimum by using values of a and b

⇒ For these values of (a,b), S will be minimum if

$$\left.\frac{\partial^2 S}{\partial a^2}\right|_{(a,b)} > 0 \quad \text{and} \quad \left[\frac{\partial^2 S}{\partial a^2}\frac{\partial^2 S}{\partial b^2} - \left(\frac{\partial^2 S}{\partial a \partial b}\right)^2\right]_{(a,b)} > 0$$

Here

$$\frac{\partial^2 S}{\partial a^2} = n > 0 \text{ and}$$

$$\left[\frac{\partial^2 S}{\partial a^2}\frac{\partial^2 S}{\partial b^2} - \left(\frac{\partial^2 S}{\partial a \partial b}\right)^2\right]_{(a,b)} = n\Sigma x_i^2 - (\Sigma x_i)^2 = n^2\left[\frac{1}{n}\Sigma x_i^2 - \left(\frac{\Sigma x_i}{n}\right)^2\right]$$

$$= n^2\sigma_x^2 > 0$$

Therefore at the values of (a,b), S is minimum.

Note

① If $\bar{x} = \frac{\Sigma x_i}{n}$ and $\bar{y} = \frac{\Sigma y_i}{n}$ then regression line of Y on X is given by $y - \bar{y} = r\frac{\sigma_y}{\sigma_x}(x - \bar{x})$

② If X is dependent on Y and the best fit of straight line $X = c + dY$ is known as regression lines of X on Y

The values of c and d can be found by using relations $d = r\frac{\sigma_x}{\sigma_y}$ and $c = \frac{1}{n}(\Sigma x_i - d\Sigma y_i)$

Also we can get the regression line of X and Y by using the equation

Page 1 Subhash
Date / / /

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

(3) The point $(\bar{x}, \bar{y})$ is the point of intersection of both the regression lines

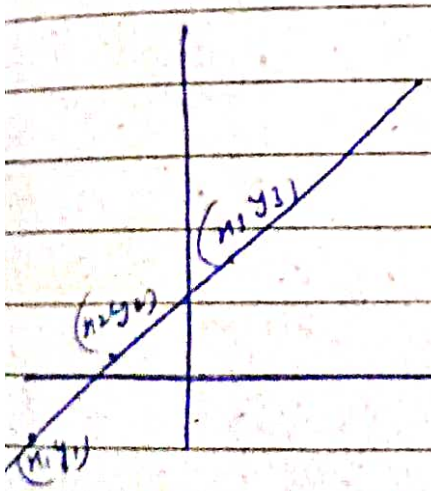
(4) The correlation coefficient is the G.M (Geometric mean) of two regression coefficient (Slope of regression lines)

$r = \pm\sqrt{r\left(\frac{\sigma_x}{\sigma_x}\right)\left(r\frac{\sigma_x}{\sigma_y}\right)} = \pm\sqrt{r^2} = \pm r$

If Both the regression coefficients are +ve, then we'll take $r > 0$

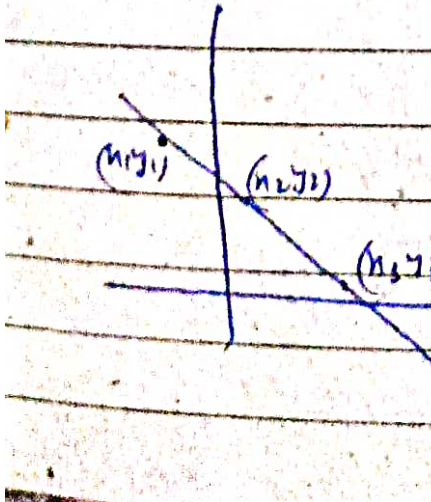
If both the regression coefficients are -ve, then we'll take $r < 0$

$r > 0$ if x and y are directly proportional



if x_1, x_2, x_3 --- ↑
then y_1, y_2, y_3 --- ↑

if x_1, x_2, x_3 --- ↓
then y_1, y_2, y_3 --- ↓



if x_1, x_2, x_3 --- ↓
then y_1, y_2, y_3 --- ↑

if x_1, x_2, x_3 --- ↑
then y_1, y_2, y_3 --- ↓

$r < 0$ if x and y are inversely proportional

Problems on Regression lines

① Find the correlation coefficient b/w X and Y for the following data. Also find regression lines

X	1	2	3	4	5	6	7	8	9	10
Y	10	12	16	28	25	36	41	49	40	50

To find

(1) $r = \frac{E(XY) - E(X)E(Y)}{\sigma_x \sigma_y}$

(2) Regression line of Y on X $\Rightarrow$ $y - \bar{y} = r \frac{\sigma_y}{\sigma_x}(x - \bar{x})$

(3) Regression line of X on Y $\Rightarrow$ $x - \bar{x} = r \frac{\sigma_x}{\sigma_y}(y - \bar{y})$

$\Sigma x = 55 \quad \Sigma y = 307 \quad \Sigma x^2 = 385 \quad \Sigma y^2 = 11387$

$\Sigma xy = 2074$

$n = 10$

$\bar{x} = E(X) = \frac{\Sigma x}{n} = 5.5 \qquad \bar{y} = E(Y) = \frac{\Sigma y}{n} = 30.7$

$E(XY) = \frac{\Sigma xy}{n} = 207.4$

$\sigma_x = \sqrt{\frac{1}{n}\Sigma x_i^2 - (E(X))^2} = 2.872$

$\sigma_y = \sqrt{\frac{1}{n}\Sigma y_i^2 - (E(Y))^2} = 14.007$

$r = 0.9583$

Regression lines of y on x $\Rightarrow$ $y - \bar{y} = r \frac{\sigma_y}{\sigma_x}(x - \bar{x})$

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

Q. Develop the estimating equation that best describes the following data and predict Y for X = 10, 15, 20

Page | Subhash | Date

X	13	16	14	11	17	9	13	17	18	12
Y	6.2	8.6	7.2	4.5	9.0	3.5	6.5	9.3	9.5	5.7

Soln We have to find regression line of Y on X

$$Y = a + bX \quad \text{or} \quad y - \bar{y} = r\frac{\sigma_y}{\sigma_x}(x - \bar{x})$$

$\Sigma X = 140 \quad \Sigma Y = 70 \quad \Sigma x^2 = 2038 \quad \Sigma XY = 1035$

reg line $\Rightarrow \quad Y = -2.871 + 0.7051X$

$$b = \frac{\frac{1}{n}\Sigma x_i y_i - \left(\frac{\Sigma x_i}{n}\right)\left(\frac{\Sigma y_i}{n}\right)}{\frac{1}{n}\Sigma x_i^2 - \left(\frac{\Sigma x_i}{n}\right)^2} = \frac{Cov(X,Y)}{\sigma_x^2} = \frac{Cov(X,Y)}{\sigma_x \sigma_y}\cdot\frac{\sigma_y}{\sigma_x} = r\frac{\sigma_y}{\sigma_x}$$

$$a = \frac{1}{n}\left(\Sigma y_i - b\Sigma x_i\right)$$

$$b = \frac{\frac{1}{10}\times 1035 - \left(\frac{140}{10}\right)\left(\frac{70}{10}\right)}{\frac{1}{10}\times 2038 - \left(\frac{140}{10}\right)^2} = \frac{103.5 - 98}{203.8 - 196} = \frac{5.5}{7.8} \approx 0.705$$

$$a = \frac{1}{10}(70 - 0.705 \times 140) = \frac{1}{b}(70 - 98.7) = -2.87$$

$$\boxed{Y = -2.87 + 0.705X}$$

Q3 Cost accountants often estimate overhead based on the level of Production. At the standard Knitting Co., they have collected information on overhead expenses and units produced at different plants and want to estimate a regression equation to predict future overhead.

Overhead expenses (Y)	191	171	272	155	280	173	234	116	153	178
Units (X)	40	42	53	35	56	34	48	30	37	40

Develop the regression equation for the cost accountants

Soln regression lines of Y on X — x is independent, y is dependent

If we are not able to figure out which is dependent or independent then we find both regression lines of Y on X as well as X on Y

$n=10$ $\Sigma X = 420$ $\Sigma Y = 1922$ $\Sigma XY = 84541$

$\Sigma X^2 = 18228$

Calculate a and b: $a = -80.443$, $b = 6.491$

$$Y = a + bX$$

Q4 If it is given that variance of $X = 9$ and the two regression lines are $8X - 10Y + 66 = 0$, $40X - 18Y - 214 = 0$

Calculate

(i) the mean values of X and Y

(ii) the correlation coeff b/w X and Y

(iii) the standard deviation of Y

Variance $X = 9$

$\sigma_x^2 = 9$

$\sigma_x = 3$

Page 1 Subhash

Date 1 1 1

(i) the mean values of X and Y

and we know mean values of X and Y are the intersection of these two lines

$8X - 10Y + 66 = 0 \quad \times 5$

$40X - 18Y - 214 = 0$

$40X - 50Y + 330 = 0$

$40X - 18Y - 214 = 0$

$-32Y + 544 = 0$

$Y = \frac{544}{32} = 17$

$Y = 17$

$8X - 10 \times 17 + 66 = 0$

$8X - 170 + 66 = 0$

$8X = 104$

$X = \frac{104}{8} = 13$

$X = 13$

$E(X) = 13$ $\quad$ $E(Y) = 17$

(ii)

Let take $8X - 10Y + 66 = 0$ as reg line X on Y

$8X = 10Y - 66$

$X = \frac{10}{8}Y - 66$

$\frac{10}{8} = b = r\frac{\sigma_x}{\sigma_y}$

Based on this

$40X - 18Y - 214 = 0$ is reg line Y on X

$18Y = 40X - 214$

$Y = \frac{40}{18}X - \frac{214}{18}$

$\frac{40}{18} = b = r\frac{\sigma_y}{\sigma_x}$

Since both slope are +ve so $r > 0$

$r = \sqrt{\left(r\frac{\sigma_x}{\sigma_y}\right)\left(r\frac{\sigma_y}{\sigma_x}\right)} = \sqrt{\frac{10}{8} \times \frac{40}{18}} = \sqrt{\frac{25}{9}} = \sqrt{2.7} > 1$

But this is wrong as r can not be > 1

So our prediction is wrong

Another case $8X - 10Y + 66 = 0$ as reg line Y on X

$$8X + 66 = 10Y$$

$$Y = \frac{8}{10}X + \frac{66}{10}$$

On this Basis

$40X - 18Y - 214 = 0$ as reg line X on Y

$$40X = 18Y + 214$$

$$X = \frac{18Y}{40} + \frac{214}{40}$$

Since both slopes are +ve So r will also be +ve

$$r = \sqrt{\frac{8}{10} \times \frac{18}{40}} = \sqrt{\frac{9}{25}} < 1$$

$$\frac{3}{5} = 0.6 < 1$$

(iii) SD of Y (σ_Y)

from reg line Y on X, slope is $\frac{8}{10}$

$$\frac{8}{10} = r\frac{\sigma_Y}{\sigma_X}$$

$$\frac{8}{10} = 0.6\frac{\sigma_Y}{3} \Rightarrow \frac{8}{10} \times \frac{3}{0.6} \times 5 = \sigma_Y$$

$\sigma_Y = 4$ Ans

Q The variable X and Y are connected by the equation $aX+bY+c=0$. Show that the correlation coefficient b/w them is -1 if signs of a and b are alike and +1 if they are different

Soln Connected by the equation means every point lies on this straight line $aX+bY+c=0$

$$aX+bY+c=0 \quad -①$$

$$E[aX+bY+c] = E(0)$$

$$aE(X)+bE(Y)+c=0 \quad -②$$

① − ②

$$a(X-E(X))+b(Y-E(Y))=0$$

$$a(X-E(X)) = -b(Y-E(Y))$$

$$X-E(X) = -\frac{b}{a}[Y-E(Y)]$$

ρ = correlation coff $= \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-E(X))(Y-E(Y))]}{\sigma_X \sigma_Y}$

Now $E[(X-E(X))(Y-E(Y))]$

$$\Rightarrow E\left[-\frac{b}{a}(Y-E(Y))(Y-E(Y))\right] = -\frac{b}{a}E[(Y-E(Y))^2]$$

$$= -\frac{b}{a}\sigma_Y^2$$

$$\sigma_X^2 = E[(X-E(X))^2] = E\left[\frac{b^2}{a^2}(Y-E(Y))^2\right]$$

$$\sigma_X^2 = \frac{b^2}{a^2}E[(Y-E(Y))^2] = \frac{b^2}{a^2}\sigma_Y^2$$

$$\sigma_X = \left|\frac{b}{a}\right|\sigma_Y$$

Page 1 Subhash

Date 1

$$\rho = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{-\frac{b}{a}\sigma_Y^2}{\sigma_X \sigma_Y} = \frac{-\frac{b}{a}\sigma_Y^2}{\left|\frac{b}{a}\right| \sigma_Y^2} =$$

if b and a of same sign

$$\frac{b}{a} > 0 \Rightarrow \boxed{\left|\frac{b}{a}\right| = \frac{b}{a}}$$

$$\rho = \frac{-\frac{b}{a}}{\frac{b}{a}} = (-1)$$ if a and b of same sign

if a and b of diff sign

$$\frac{b}{a} < 0 \Rightarrow \left|\frac{b}{a}\right| = -\frac{b}{a}$$

$$\rho = \frac{-\frac{b}{a}}{-\frac{b}{a}} = (+1)$$ if a and b are of diff sign.

HP