Regression Lines – (\bar{x}, \bar{y}) and Slope

1. Equation of straight line y on x:

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

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

2. Equation of straight line x on y:

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

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

Note: $b_{xy} \neq b_{yx}$

Regression Lines – Least square method

Consider the equation of a straight line y on x:

$$y = bx + a$$

Normal equations are

$$\sum y = \mathbf{b} \sum x + \mathbf{a}n$$
$$\sum yx = \mathbf{b} \sum x^2 + \mathbf{a} \sum x$$

Solving the above two equations we get the values of a and b

$$y = bx + a$$

Regression Lines – Least square method

Consider the equation of a straight line x on y:

$$x = dy + c$$

Normal equations are

$$\sum x = d \sum y + cn$$

$$\sum xy = d \sum y^2 + c \sum y$$

Solving the above two equations we get the values of *c* and *d*

$$x = dy + c$$

Problems

1. From the following data obtain the two regression equations:

X: 6 2 10 4 8 Y: 9 11 5 8 7

Solution:

Computation of	Regression	Equations
----------------	------------	-----------

Comparation		201	V2		
v	v	XY	X^2	I.	
Α	1	54	36	81	
6	9	DARGOLES PROJECTION	4	121	
2	11	22		THE RESERVE OF THE PARTY OF THE	
10	5	50	100	25	
10	0	32	16	64	
4	0		64	49	
8	7	56		EV2 240	
$\Sigma X = 30$	$\Sigma Y = 40$	$\Sigma XY = 214$	$\Sigma X^2 = 220$	$\Sigma Y^2 = 340$	

Regression equation of Y on X is given by

$$Y = a + bX$$

... (i)

Following two normal equations are required to be solved to determine the values of a and b.

$$\Sigma Y = Na + b\Sigma X$$

$$\Sigma XY = a \Sigma X + b\Sigma X^{2}$$

Substituting the values, we get
$$40 = 5$$

$$40 = 5a + 30b$$
 ... (ii)
 $214 = 30a + 220b$... (iii)

$$240 = 30a + 180b$$

$$214 = 30a + 220b$$
 ... (v)

Subtracting equation (v) from (iv), we get

$$-40b = 26 \implies b = -0.65$$

Substituting the value of b in equation (ii), we have

$$40 = 5a + 30 (-0.65) \implies 5a = 40 + 19.5 = 59.5 \implies a = 11.9$$

Putting the values of a and b in equation (i), we get the regression equation of Y on X as

$$Y = 11.9 - 0.65X$$

... (iv)

Regression equation of X on Y is given by

$$X = a + bY$$

... (vi)

and the two normal equations are:

$$\Sigma X = Na + b\Sigma Y$$

$$\Sigma XY = a\Sigma Y + b\Sigma Y^2$$

Substituting the values, we get 30 = 5a + 40b

$$30 = 5a + 40b$$

$$214 = 40a + 340b$$

Multiplying equation (vii) by 8,
$$240 = 40a + 320b$$

$$240 = 40a + 320b$$

$$214 = 40a + 340b$$

Subtracting equation (x) from (ix), we get

$$-20b = 26 \implies b = -1.3$$

Substituting the value of b in equation (vii), we have

$$30 = 5a + 40 (-1.3)$$
 \Rightarrow $5a = 30 + 52 = 82$ \Rightarrow $a = 16.4$

Putting the values of a and b in equation (vi), we get the regression equation of X on Y as

$$X = 16.4 - 1.3Y$$
.

2. A study of prices at Chennai and Vellore gave the following data:

	Chennai	Vellore
Mean	19.5	17.75
S.D.	1.75	2.5

Also the **coefficient of correlation** between the two **is 0.8**. Estimate the most likely price of rice (i) at Chennai corresponding to the price of 18 at Vellore, and (ii) at Vellore corresponding to the price of 17 at Chennai.

Let the prices of rice at Chennai and Vellore be denoted by X and Y respectively. Then from the data,

$$\bar{x} = 19.5$$
, $\bar{y} = 17.75$, $\sigma_X = 1.75$, $\sigma_Y = 2.5$ and $r_{XY} = 0.8$.

Regression line of X on Y is

$$x - \overline{x} = \frac{r \, \sigma_X}{\sigma_Y} (y - \overline{y})$$

i.e.,
$$x - 19.5 = \frac{0.8 \times 1.75}{2.5} (y - 17.75)$$

 \therefore When y = 18,

$$x = 19.5 + \frac{0.8 \times 1.75}{2.5} \times 0.25$$
$$= 19.64$$

Regression line of Y on X is

$$y - \overline{y} = \frac{r \, \sigma_{\gamma}}{\sigma_{\chi}} (x - \overline{x})$$

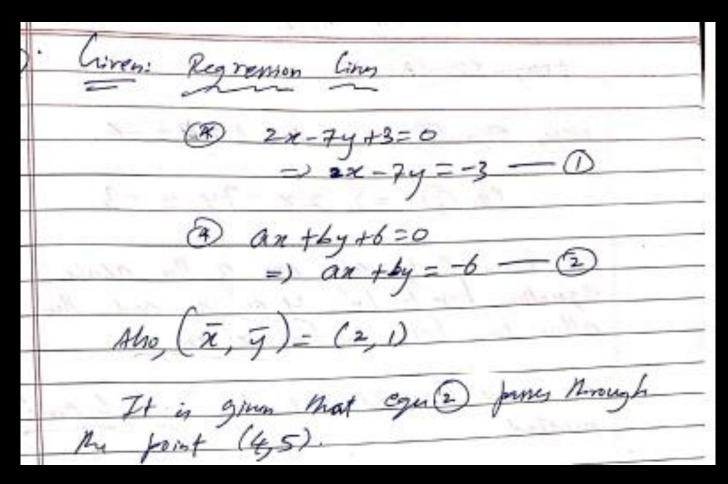
i.e., $y - 17.75 = \frac{0.8 \times 2.5}{0.1.75} (x - 19.5)$

... When
$$x = 17$$
,

$$y = 17.75 + \frac{0.8 \times 2.5}{1.75} \times (-2.5)$$

at a contract of the contract

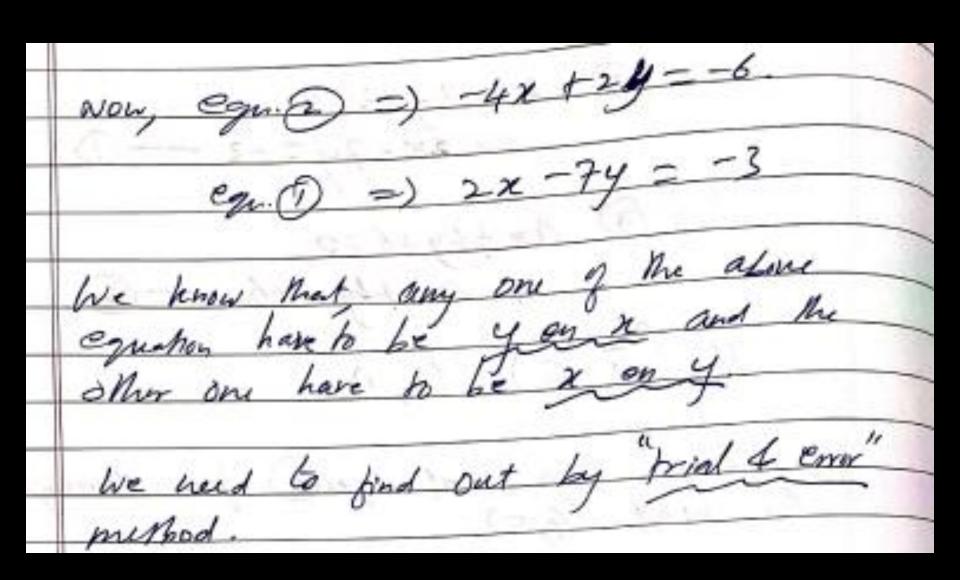
3. For two variables X and Y, we have two regression lines 2X-7Y+3=0 and aX+bY+6=0 for some constants a and b. Suppose we know that mean of X is 2 and mean of Y is 1. Also the second regression line passes through (4, 5). Find the correlation coefficient of these two lines.



We know that the regression lines forms

Morough the forat (x, y).

=) Equite fames money (x, y) also. Thuy, (2) => an +by =-6. Point (2,1) paring Many 2 => 2a+b=-6. Print (4,5) paning Money => 4 a + 56=-6. a = -4 Now, 2a+b=-6. b = 2



ge our anumy Dios

Now let 1 be you x. -7y = -2x-3 4= (2/2)x+3 and let (2) he x on y. -42 = -24 -6 n= (/2) 4 +3/2 Ti = byn- byg = (/2) (/2) = /2. 72 = 1/2 < 1 (Valid) =) 71 = V/7 =) 7 = 0.37796