Regression.

Least Square Method: Regression of y on x is $y=a+bx$

The normal equations are $\Sigma y = na + b\Sigma x$

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

Pbm. Find regression of y on x for the following data.

x	1	2	3	4	5
y	10	12	15	14	15

x	y	xy	x^2
1	10	10	1
2	12	24	4
3	15	45	9
4	14	56	16
5	15	75	25
15	66	210	55

$\Sigma y = na + b\Sigma x \Rightarrow 66 = 5a + 15b \rightarrow (1)$

$\Sigma xy = a\Sigma x + b\Sigma x^2 \Rightarrow 210 = 15a + 55b \rightarrow (2)$

$(1)\times 3 \Rightarrow \mp 198 = \mp 15a \mp 45b$

$12 = 10b$

$b = \frac{12}{10} = 1.2$

$(1) \Rightarrow 5a + 15b = 66$

$5a + 15(1.2) = 66$

$5a + 18 = 66$

$5a = 66 - 18 = 48$

$a = 48/5 = 9.6$

Regression of y on x is $y = a + bx$

$\boxed{y = 9.6 + 1.2x}$

The regression equation of x on y is $x = a' + b'y$, where the normal equations are

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

$$\Sigma xy = a'\Sigma y + b'\Sigma y^2$$

Regression of y on x is

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

$y-\bar{y}=b_{yx}(x-\bar{x})$, where $b_{yx}\to$ regression coefficient of y on x

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

$$=\frac{\Sigma xy-\frac{\Sigma x\Sigma y}{n}}{\Sigma x^2-\frac{(\Sigma x)^2}{n}}$$

Regression of x on y is

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

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

$$=\frac{\Sigma xy-\frac{\Sigma x\Sigma y}{n}}{\Sigma y^2-\frac{(\Sigma y)^2}{n}}$$

Pbm Find the two regression equations when $\bar{x} = 15$, $\bar{y} = 20$, $b_{yx} = 0.3$, $b_{xy} = 0.7$

Regression equation of y on x is

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

$$y - 20 = 0.3(x - 15)$$

$$y - 20 = 0.3x - 4.5$$

$$y = 0.3x - 4.5 + 20$$

$$\boxed{y = 0.3x + 15.5}$$

Regression equation of x on y is

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

$$x - 15 = 0.7(y - 20)$$

$$= 0.7y - 14$$

$$x = 0.7y - 14 + 15$$

$$\boxed{x = 0.7y + 1}$$

Pbm Find the regression of y on x for the following data. $n=10$, $\bar{x}=30$, $\bar{y}=50$, $\sigma_x=4$, $\sigma_y=5$, $r_{xy}=0.7$. Estimate y when $x=37$.

Regression of y on x is

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

$$y-50=0.7\left(\frac{5}{4}\right)(x-30)$$

$$=0.875(x-30)$$

$$y-50=0.875x-26.25$$

$$y=0.875x-26.25+50$$

$$\boxed{y=0.875x+23.75}$$

When $x=37$, $y=0.875(37)+23.75=32.375+23.75$
$=56.125$

Pbm. Following data is available for 15 students regarding the time spent in studies everyday and percentage marks in an examinations. Find the expected percentage when a student studies for 3.5 hours every day.

	Time (hr) x	Percentage y
Average	$\bar{x} = 4$	$\bar{y} = 90$
SD	$\sigma_x = 0.5$	$\sigma_y = 7$
coefficient of correlation.	$r = 0.55$	

TF y when $x = 3.5$

Regression of y on x is

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

$$y - 90 = 0.55\left(\frac{7}{0.5}\right)(x - 4)$$

$$= 7.7(x-4)$$

$$y - 90 = 7.7x - 30.8$$

$$y = 7.7x - 30.8 + 90$$

$$= 7.7x + 59.2$$

When $x = 3.5$

$$y = 7.7(3.5) + 59.2$$

$$= 26.95 + 59.2$$

$$= 86.15$$

Find the two regression equations given the following data.

$\bar{x} = 70$, $\bar{y} = 80$

Regression coefficient of y on $x = 1.5$ $= b_{yx}$

Regression coefficient of x on $y = 0.6$ $= b_{xy}$

Soln:

Given, $\bar{x} = 70$, $\bar{y} = 80$, $b_{yx} = 1.5$, $b_{xu} = 0.6$

Regression of y on x is

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

$$y - 80 = 1.5(x - 70)$$
$$= 1.5x - 105$$
$$y = 1.5x - 105 + 80$$
$$\boxed{y = 1.5x - 25}$$

Regression of x on y is

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

$$x - 70 = 0.6(y - 80)$$
$$= 0.6y - 48$$
$$x = 0.6y - 48 + 70$$
$$\boxed{x = 0.6y + 22}$$

Pbm Find the regression equation of y on x when $\Sigma x = 420$, $\Sigma y = 1922$, $\Sigma xy = 84541$, $\Sigma x^2 = 18228$, $n=10$

regression equation of y on x is $y - \bar{y} = b_{yx}(x - \bar{x})$

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

$$= \frac{84541 - \frac{(420)(1922)}{10}}{18228 - \frac{(420)^2}{10}} = \frac{84541 - 80724}{18228 - 17640}$$

$$= \frac{3817}{588} = 6.49$$

regression equation of y on x is $y - \bar{y} = b_{yx}(x - \bar{x})$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{1922}{10} = 192.2$$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{420}{10} = 42$$

$$y - 192.2 = 6.49(x - 42)$$

$$y - 192.2 = 6.49x - 272.64$$

$$y = 6.49x - 272.64 + 192.2$$

$$= 6.49x - 80.44$$

The following data give the number of years in the pit and the daily earnings of 10 miners

Years in pit x	5	6	9	5	8	3	7	5	1	1
Daily earnings (₹) y	32	25	30	34	30	39	26	23	15	11

Find the regression of daily earnings on the no of years in the pit.

TF: Regression of y on x.

$$y - \bar{y} = b_{yx}(x - \bar{x}), \quad \bar{y} = \frac{\Sigma y}{n}$$

$$\bar{x} = \frac{\Sigma x}{n}$$

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

x	y	xy	x^2
5	32	160	25
6	25	150	36
9	30	270	81
5	34	170	25
8	30	240	64
3	39	117	9
7	26	182	49
5	23	115	25
1	15	15	1
1	11	11	1
50	265	1430	316

$$\bar{x} = \frac{\Sigma x}{n} = \frac{50}{10} = 5$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{265}{10} = 26.5$$

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

$$= \frac{1430 - \frac{(50)(265)}{10}}{316 - \frac{(50)^2}{10}}$$

$$= \frac{1430 - 1325}{316 - 250}$$

$$= \frac{105}{66} = 1.59$$

Regression of y on x is

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

$$y - 26.5 = 1.59(x - 5)$$

$$= 1.59x - 7.95$$

$$y = 1.59x - 7.95 + 26.5$$

$$\boxed{y = 1.59x + 18.54}$$

Properties of Regression equations:

1) Regression equations represent two straight lines. These two lines coincide when there is a perfect correlation ($r = \pm 1$)
They are perpendicular to each other when there is no correlation. ($r=0$)

2) The intersection of both regression equations is $(\bar{x}, \bar{y})$.
$(\bar{x}, \bar{y})$ is the only point which satisfy both the equations simultaneously.
If both the regression equations are given, we can solve both equations and find the arithmetic mean of x & y.

3) We have $b_{yx} = r\dfrac{\sigma_y}{\sigma_x}$, $b_{xy} = r\dfrac{\sigma_x}{\sigma_y}$

$$b_{yx} \cdot b_{xy} = r\frac{\cancel{\sigma_y}}{\cancel{\sigma_x}} \cdot r\frac{\cancel{\sigma_x}}{\cancel{\sigma_y}}$$
$$= r^2$$
$$r = \sqrt{b_{yx} \cdot b_{xy}}$$

Both regression coefficients have same sign.

r positive $\iff$ both b_{xy} & b_{yx} are positive

r negative $\iff$ both b_{xy} & b_{yx} are negative

Pbm. The regression of y on x for certain bivariate data was found to be $10y = 3x + 155$ and that of x on y was $10x = 7y + 10$. Find $\bar{x}$ and $\bar{y}$.

$(\bar{x}, \bar{y})$ is the point of intersection of both regression equations.

$10y = 3x + 155 \Rightarrow -3x + 10y = 155 \rightarrow (1)$

$10x = 7y + 10 \Rightarrow 10x - 7y = 10 \rightarrow (2)$

$(1) \times 10 \Rightarrow -30x + 100y = 1550$

$(2) \times 3 \Rightarrow 30x - 21y = 30$

$79y = 1580$

$y = \frac{1580}{79} = 20$

$(2) \Rightarrow 10x - 7y = 10$

$10x - 7(20) = 10$

$10x - 140 = 10$

$10x = 10 + 140 = 150$

$x = \frac{150}{10} = 15$

Soln: $\bar{x} = 15$, $\bar{y} = 20$

1) Regression analysis is used to determine cause and effect relationship.

2) If the regression equation of y on x is $10y = 5 - 3x$, the relationship of y on x is linear and direct.

3) If the slope of a regression line is positive, then the value of r is the positive square root of r^2.

4) Suppose there is a direct relationship between price and demand, price tends to be high, when demand is high.

5) For a data, $r = 0.3$, $b_{yx} = 0.24$ and $b_{xy} = -0.375$ -
False. both the regression equations have the same sign.

1) If two variables x and y are highly correlated then y can be estimated for a given value of x using regression equation of y on x.

2) If the regression equation of x on y is $2x + 7y = 135$, the estimated value of x when $y=17$ is _____

$$2x + 7(17) = 135$$

$$2x + 119 = 135$$

$$2x = 135 - 119 = 16$$

$$x = \frac{16}{2} = 8$$

Ans: $x = 8$

3) If the two regression coefficients are $b_{yx}=3$ and $b_{xy}=1/3$, the correlation coefficent is _______

$$r = \sqrt{b_{yx} \cdot b_{xy}} = \sqrt{3 \cdot \frac{1}{3}} = \sqrt{1} = 1.$$

Ans

$r = 1.$

4) for 10 pairs x and y values, the regression coefficient can be _______

a) 1.2 and 0.8 b) 1.2 and 1.1 c) 1.2 and -0.2

Ans: a) 1.2 and 0.8

5) If the two regression coefficients are positive, then the value of correlation coefficient is positive and vice versa.

6) If the values of regression coefficients are 0.4 and 0.9, then the value of correlation coefficient is

$$r = \sqrt{b_{yx} \cdot b_{xy}} = \sqrt{0.4 \times 0.9} = \sqrt{0.36} = 0.6$$

7) If the values of regression coefficients are both 0.7 each, then the value of correlation coefficient is

$$r = \sqrt{b_{yx} \cdot b_{xy}} = \sqrt{0.7 \times 0.7} = \sqrt{(0.7)^2} = 0.7$$

8) The point $(\bar{x}, \bar{y})$ lies on both regression lines x on y and y on x.

9) If the values obtained from 10 pairs of x and y are $b_{xy} = 2$, $r = \frac{1}{3}$, SD of $x = 3$, then SD of y is ___

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

$$2 = \frac{1}{\cancel{3}} \cdot \frac{\cancel{3}}{\sigma_y}$$

$$2\sigma_y = 1$$

$$\therefore \boxed{\sigma_y = \frac{1}{2}}$$

10) If the values obtained from 10 pairs of x and y are $b_{yx} = \frac{2}{3}$, $r = \frac{1}{2}$, SD of $x = 3$, then SD of y is ___

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

$$\frac{2}{3} = \frac{1}{2} \cdot \frac{\sigma_y}{3} \Rightarrow \frac{2}{\cancel{3}} \times 2 \times \cancel{3} = \sigma_y$$

$$\Rightarrow \boxed{\sigma_y = 4}$$

11) The two regression lines intersect each other at the point $(\bar{x}, \bar{y})$

12) If the correlation coefficient is zero, the two regression lines are perpendicular.

13) If the correlation coefficient is +1 or −1, the two regression lines coincide with each other.

14) The value of correlation coefficient is Geometric Mean of the two regression equations.