Regression.

The line of Regression is the line which gives the "best estimate" to the value of one variables for any specific value of the other variables Thus the line of regression is the line of "best fit" and is obtained by the "principle of least squares".

In a bivariate distribution (x_i, y_i) $i=1,2..n$ let y is dependent variable and x is independent variable.

Let $y=a+bx$ be the line of regression of y on x

According to the principle of least squares the constant a and b are to be determined such that.

$$E = \sum_{i=1}^{n} (y_i - a - bx_i)^2$$ is minimum equating to zero the partial derivates of E with respect to a and b.

we get,
$$\frac{dE}{da} = 0 \Rightarrow -2\sum_{i=1}^{n} (y_i - a - bx_i) = 0$$

$$\frac{dE}{db} = 0 \Rightarrow -2 \cdot \sum_{i=1}^{n} (y_i - a - bx_i)(x_i) = 0$$

Thus the normal equation for estimating

a and b are,

$$\sum_{i=1}^{n}(y_i - a - bx_i) = 0$$

$$\sum_{i=1}^{n} y_i - na - b\sum_{i=1}^{n} x_i = 0$$

$$\sum_{i=1}^{n} y_i = na + b\sum_{i=1}^{n} x_i \quad \text{--- (1)}$$

and $\sum_{i=1}^{n} x_i(y_i - a - bx_i) = 0$

$$\sum_{i=1}^{n} x_i y_i - a\sum_{i=1}^{n} x_i - b\sum_{i=1}^{n} x_i^2 = 0$$

$$\sum_{i=1}^{n} x_i y_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 \quad \text{--- (2)}$$

Dividing (1) by "n" we get

$$\bar{y} = a + b\bar{x} \quad \text{--- (3)}$$

Thus the line of regression of Y on X passes through the point $(\bar{x}, \bar{y})$

we know $\operatorname{cov}(x,y) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}$

$$\frac{1}{n}\sum_{i=1}^{n} x_i y_i = \operatorname{cov}(x,y) + \bar{x}\bar{y} \quad \text{--- (4)}$$

Also $\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\bar{x})^2$

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2 = \sigma_x^2 + (\bar{x})^2 \longrightarrow (4.a)$$

Dividing (2) by "n" we get

$$\frac{1}{n}\sum_{i=1}^{n} x_i y_i = \frac{a\sum_{i=1}^{n} x_i}{n} + \frac{b\sum x_i^2}{n}$$

using (4) and (4.a) we get

$$cov(x, y) + \bar{x}\bar{y} = a\bar{x} + b\{\sigma_x^2 + (\bar{x})^2\} \quad — (5)$$

$3 \times \bar{x}$

$$\cancel{\bar{x}\bar{y}} = \cancel{a\bar{x}} + \cancel{b\bar{x}^2}$$

$$(-) \quad (-) \quad (-)$$

$$cov(x, y) = b\sigma_x^2$$

$$b = \frac{cov(x, y)}{\sigma_x^2} \quad — (6)$$

"b" is the slope of the line of regression of y on x. and the line of Regression. Passes through $(\bar{x}, \bar{y})$

∴ The equation is

$$y - \bar{y} = b(x - \bar{x})$$

$$y - \bar{y} = \frac{cov(x, y)}{\sigma_x^2}(x - \bar{x}) \quad — (7)$$

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

$$\therefore cov(x, y) = r \cdot \sigma_x \sigma_y$$

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

$$y - \bar{y} = r \cdot \frac{\sigma_y}{\sigma_x}(x - \bar{x}) \quad — (8)$$

Starting with the equation.

$x = a + by$, and proceeding similarly (or) by simply interchanging variables x and y on, we get the line of regression of x on y as.

$$x - \bar{x} = r \cdot \frac{\sigma_x}{\sigma_y} (y - \bar{y}) \quad \text{—— (9)}$$

Always we have to line of regression one is y on x another one is x on y.

for the given value of x to find y are use the line of Regression of y on x

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

for a given value of y to find x we use the line of regression of x on y.

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

where $r = \pm 1$ Both lines reduces to.

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

two lines of regression coincide and thus we have only one line.

Regression coefficients

The slope of the line 'b' is also called the coefficient of regression. The coefficient of Regression "b" on the regression line of y on x represent the increment in the value of y variables corresponding to a unit change in the value of x variables.

b_{yx} = Regression coefficient of y on x.

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

$|||^{ly}$ the regression coefficient of x on y indicates the change in the value of x variables corresponding to a unit change in the value of y variable.

b_{xy} = Regression coefficient of x on y.

2m $r = \frac{cov(x,y)}{\sigma_y^2} = \frac{r \cdot \sigma_y \sigma_x}{\sigma_y^2} = r \frac{\sigma_x}{\sigma_y}$.

1m **Properties of Regression coefficients.**

(1) correlation coefficient is the geometric mean of between the regression coefficients.

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

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

$$(b_{yx})(b_{xy}) = \left(r \cdot \frac{\sigma_y}{\sigma_x}\right)\left(r \cdot \frac{\sigma_x}{\sigma_y}\right) = r^2$$

$$\therefore r = \pm\sqrt{(b_{xy})(b_{yx})}$$

(ie) correlation coefficient is the G.M. between Regression coefficient.

Remark:

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

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

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

The sign of correlation coefficient is the same as that of regression coefficient since the sign of each depends upon the covariance term. Thus if the regression coefficient are positive "r" is positive and if the regression coefficient and negative "r" is negative.

$$r = \pm\sqrt{(b_{xy})(b_{yx})}$$

The sign to be taken before the square root is that of the regression coefficient.

2) If one of the regression coefficient is greater than unity.

Proof:

Let one of the regression coefficients b_{yx} be greater than Unity, then we have to show that $b_{xy} < 1$.

$$b_{yx} > 1 \text{ (or) } \frac{1}{b_{yx}} < 1$$

2m

Also $r^2 \leq 1$

$$\Rightarrow (b_{yx})(b_{xy}) \leq 1$$

$$b_{xy} \leq \frac{1}{b_{yx}} < 1$$

(ie) If $b_{yx} >$ then $b_{xy} < 1$

3. Arithmetic mean of the Regression coefficient is greater than the correlation coefficient

Proof:

we have to prove that

$$\frac{1}{2}(b_{yx} + b_{xy}) \geq r.$$

$$\Rightarrow b_{yx} + b_{xy} \geq 2r$$

$$\Rightarrow (b_{yx} + b_{xy}) \geq 2 \pm \sqrt{b_{(xy)} b_{(yx)}}$$

$$b_{yx} + b_{xy} \pm 2\sqrt{(b_{xy})(b_{yx})} \geq 0$$

$$\text{ie } (\sqrt{b_{yx}} \pm \sqrt{b_{xy}})^2 \geq 0$$

which is always true since the square of a real quantity is always non-negative.

hence the arithmetic mean of the Regression coeff. is greater than the correlation coefficient

4) Regression coefficient are independent of the origin but not scale:

<u>Proof</u>:

$$U = \frac{X-a}{h} \quad \text{and} \quad V = \frac{y-b}{k}$$

$$Uh = X-a \quad , \quad Vk = y-b$$

$$X = a+uh \quad , \quad y = b+kv$$

where a, b, h and k are constants

$$E(X) = a+h\,E(u)$$

$$E(y) = b+k\,E(v)$$

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

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

$$X - E(X) = a+hu - \{a+h\,E(u)\}$$

$$= a+hu-a-h\,E(u)$$

$$= h\{u-E(u)\}$$

$$y - E(y) = b+k.v - \{b+k\,E(v)\}$$

$$= b+k.v-b-k\,E(v)$$

$$= k\{v-E(v)\}$$

$$b_{yx} = \frac{E[h\{u - E(u)\}\, k\,\{v - E(v)\}]}{E[h\{u - E(u)\}]^2}$$

$$= \frac{hk\, E[\{u - E(u)\}\{v - E(v)\}]}{h^2\, E\{u - E(u)\}^2}$$

$$= \frac{k}{h} \frac{\mathrm{cov}(u,v)}{\sigma_u^2} = \frac{k}{h} b_{vu}$$

Angle between two line of Regression:

Line of Regression of y on x is

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

Slope of this line is $r \cdot \frac{\sigma_y}{\sigma_x}$ (ie) $m_1 = r \cdot \frac{\sigma_y}{\sigma_x}$

Line of Regression of x on y is

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

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

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

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

Slope of this line is $\frac{\sigma_y}{r\,\sigma_x}$

If θ is the acute angle between the two lines of regression.

then $\tan\theta = \dfrac{m_1 \sim m_2}{1+m_1 m_2}$

$$\tan\theta = \frac{\dfrac{r\,\sigma_y}{\sigma_x} \sim \dfrac{\sigma_y}{r\,\sigma_x}}{1+\dfrac{r\sigma_y}{\sigma_x}\cdot\dfrac{\sigma_y}{r\sigma_x}}$$

$$= \frac{\dfrac{r^2\sigma_y\sigma_x \sim \sigma_x\sigma_y}{r\sigma_x^2}}{\dfrac{\sigma_x^2+\sigma_y^2}{\sigma_x^2}}$$

$$= \frac{\sigma_x\sigma_y\,(r^2 \sim 1)}{r\,\sigma_x^2} \times \frac{\sigma_x^2}{\sigma_x^2+\sigma_y^2}$$

Since $\boxed{r^2 \leq 1}$

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

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

If $r=0$, $\tan\theta = \infty \Rightarrow \theta = \pi/2$

Thus if the two variables are uncorrelated the lines of regression become perpendicular to each other.

If $r = \pm 1$, $\tan\theta = 0 \Rightarrow \theta = 0$ (or) π

In this case the two lines of regression either coincide or they are parallel to each.

other. But since both the line of regression pass through the points $(\bar{x}, \bar{y})$ they cannot be parallel. Hence in the case of perfect correlation positive or negative, the two lines of regression coincide.

28/6/19 Find the two regression equations for the following data

x:	10	14	15	28	35	48
y:	74	61	50	54	43	26

solution:

Regression equation for of x on y is

$$(x-\bar{x}) = r \cdot \frac{\sigma x}{\sigma y}(y-\bar{y})$$

Regression equation of y on x is

$$(y-\bar{y}) = r \cdot \frac{\sigma y}{\sigma x}(x-\bar{x})$$

b_{xy} b_{x}

$$r = \frac{\operatorname{cov}(x,y)}{\sigma x \cdot \sigma y}$$

$$\operatorname{cov}(x,y) = \frac{\Sigma xy}{N} - \bar{x}\,\bar{y}$$

$$\sigma x = \sqrt{\frac{\Sigma x^2}{N} - (\bar{x})^2}$$

$$\sigma y = \sqrt{\frac{\Sigma y^2}{N} - (\bar{y})^2}$$

x	Y	x^2	y^2	xy
10	74	100	5476	740
14	61	196	3721	854
15	50	225	2500	750
28	54	784	2916	1512
35	43	1225	1849	1505
48	26	2304	676	1248
150	308	4834	17138	6609

$\Sigma x = 150$ $\quad \Sigma x^2 = 4834$ $\quad \Sigma xy = 6609$

$\Sigma y = 308$ $\quad \Sigma y^2 = 17138$ $\quad N = 6$

$$\bar{X} = \frac{\Sigma x}{N} = \frac{150}{6} = 25$$

$$\bar{Y} = \frac{\Sigma y}{N} = \frac{308}{6} = 51.3334$$

$$\text{cov}(x,y) = \frac{\Sigma xy}{N} - (\bar{x})(\bar{y})$$

$$= \frac{6609}{6} - (25)(51.3333)$$

$$= 1101.5 - 1283.3325$$

$$\text{cov}(x,y) = -181.8325$$

$$\sigma x = \sqrt{\frac{\Sigma x^2}{N} - (\bar{x})^2}$$

$$= \sqrt{\frac{4834}{6} - (25)^2}$$

$$= \sqrt{805.6666 - 625}$$

$$= \sqrt{180.6666}$$

$$\sigma x = 13.4412$$

$$\sigma y = \sqrt{\frac{\Sigma y^2}{N} - (\bar{y})^2}$$

$$= \sqrt{\frac{17138}{6} - (51.3333)^2}$$

$$= \sqrt{2856.3333 - (2635.1077)}$$

$$= \sqrt{221.2256}$$

$$\sigma y = 14.8736$$

$$r = \frac{cov(x,y)}{\sigma x . \sigma y}$$

$$= \frac{-181.8325}{(13.4412)(14.8736)}$$

$$= \frac{-181.8325}{199.9190}$$

$$\boxed{r = -0.9095}$$

regression equation of y on x is

$$(x - \bar{x}) = r . \frac{\sigma x}{\sigma y} (y - \bar{y})$$

$$(x - 25) = (-0.9095)\left(\frac{13.4412}{14.8736}\right)(y - 51.3333)$$

$$(x - 25) = (-0.9095)(0.9036)(y - 51.3333)$$

$$(x - 25) = (-0.8218)(y - 51.3333)$$

$$(x - 25) = -0.8218y + 42.1857$$

$$x = -0.8218y + 42.1857 + 25$$

$$\boxed{x = -0.8218y + 67.1857}$$

regression equation of y on x is

$$(y-\bar{y}) = r \cdot \frac{\sigma y}{\sigma x}(x-\bar{x})$$

$$(y-51.3333) = (-0.9095)\left(\frac{14.8736}{13.4412}\right)(x-25)$$

$$(y-51.3333) = (-0.9095)(1.1065)(x-25)$$

$$(y-51.3333) = (-1.0063)(x-25)$$

$$(y-51.3333) = -1.0063x + 25.1575$$

$$y = -1.0063x + 25.1575 + 51.3333$$

$$\boxed{y = -1.0063x + 76.4908}$$

2\. For two variable x and y the equations of thee regression lines are

L $9y-x-288$ and $x-4y+38$. find

(i) The mean values of x and y

(ii) The correlation between x and y

(iii) The most probable value of y when $x=145$

(iv) The most probable value of x when $y=35$

solution :

(i)

$$9y - \cancel{x} = 288 \rightarrow (1)$$

$$-4y + \cancel{x} = -38 \rightarrow (2)$$

$$5y = 250$$

$$y = \frac{250}{5}$$

$$y = 50$$

$$\boxed{\therefore \bar{y} = 50}$$

Substitute $y = 50$ in (1)

$$9(50) - x = 288$$

$$450 - x = 288$$

$$-x = 288 - 450$$

$$-x = -162$$

$$x = 162$$

$$\boxed{\therefore \bar{x} = 162}$$

(ii) assumed that the regression equation of y on x has

$$9y - x - 288 = 0$$

$$-x = -9y + 288.$$

$$x = 9y - 288.$$

$$b_{xy} = 9.$$

$$9y = x + 288$$

$$y = x/9 + 288/9$$

$$y = 1/9\, x + 32$$

$$\boxed{\therefore b_{yx} = 1/9}$$

assumed that the equation of x on y has

$$x - 4y + 38 = 0$$

$$-4y = -x - 38.$$

$$y = \frac{x}{4} + \frac{38}{4}$$

$$x = 4y - 38$$

$$\boxed{\therefore b_{xy} = 4}$$

$$b_{yx} = 1/4$$

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

$$= \sqrt{1/9 \times 1/4}$$

$$= 0.08$$

$$= \pm\sqrt{4 \cdot \left(\frac{1}{9}\right)}$$

$$= \pm\sqrt{4 \cdot 0.1111}$$

$$= \pm\sqrt{0.4444}$$

$$\boxed{r = 0.6666}$$

(iii) The equation of y on x is

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

$$(y-50) = (0.1111)(x-162)$$

$$(y-50) = 0.1111x - 17.9982$$

$$y = 0.1111x - 17.9982 + 50$$

$$y = 0.1111x + \cancel{17.9982x}\ 32.00$$

$$y = 0.1111x + 32$$

when $x = 145$, $y = 0.1111(145) + 32$

$$y = 16.1095 + 32$$

$$\boxed{y = 48.1095}$$

(iv) The equation of x on y is when $y = 35$

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

$$(x-162) = 4(y-50)$$

$$(x-162) = 4y - 200$$

$$x = 4y - 200 - 162$$

$$x = 4y - 362$$

$$x = 4(35) - 362$$

$$x = 140 - 362$$

$$\boxed{x = \cancel{-222}}\ 102$$

Given that

	X	Y
Arithemetic mean :	36	85
Standard deviation :	11	8
correlation coefficient between x and y :	0.66	

arithmetic mean, Standard deviation

i) Find the two regression equation

ii) Estimate the values of x when $Y = 75$

Solution :

$\overline{X} = 36 \quad \overline{Y} = 85 \quad \sigma x = 11 \quad \sigma y = 8 \quad r = 0.66$

i) Find the two regression equation :-

The regression equation of x and y

$$(X - \overline{X}) = r \cdot \frac{\sigma x}{\sigma y} (Y - \overline{Y})$$

$$(x - 36) = 0.66 \left(\frac{11}{8}\right)(Y - 85)$$

$$(x - 36) = 0.66 (1.375)(Y - 85)$$

$$(x - 36) = 0.9075 (Y - 85)$$

$$(x - 36) = 0.9075Y - 8.2875$$

$$x = 0.9075Y - 8.2875 + 36$$

$$\boxed{x = 0.9075Y - 27.7125}$$

ii) x when $Y = 75$

$$x = 0.9075 (75) - 27.7125$$

$$x = 68.0625 - 27.7125$$

$$\boxed{x = 40.35}$$

The regression equation y on x

$$(y-\bar{y}) = r \cdot \frac{\sigma y}{\sigma x}(x-\bar{x})$$

$$(y-85) = 0.66\left(\frac{8}{11}\right)(x-36)$$

$$(y-85) = 0.66(0.7272)(x-36)$$

$$(y-85) = 0.4799(x-36)$$

$$(y-85) = 0.4799x - 17.2764$$

$$y = 0.4799x - 17.2764 + 85$$

$$\boxed{y = 0.4799x + 67.7236}$$

2. Let $9x-3y=165$ and $3x-4y=40$ are be two regression equations. find

i) $\bar{x}$ and $\bar{y}$

ii) Find standard deviation of y when $V(x)=25$

Solution:

$$9x-3y=165$$
$$3x-4y=40$$

$$\cancel{9x}-3y=165$$
$$\cancel{9x}-12y=120$$
$$(-)\quad(+)\quad(-)$$

$$9y=45$$
$$y=\frac{45}{9}$$
$$\boxed{\bar{y}=5}$$

$$9x-3(5)=165$$
$$9x-15=165$$
$$9x=\frac{165+15}{9}$$
$$x=\frac{180}{9}$$
$$\boxed{\bar{x}=20}$$

(ii)

$$9x-3y=165$$
$$9x=3y+65$$
$$x=\frac{3y}{9}+\frac{165}{9}$$
$$\boxed{b_{xy}=1/3}$$

$$3x-4y=40$$
$$-4y=-3x+40$$
$$y=\frac{3x}{4}-\frac{40}{4}$$
$$\boxed{b_{yx}=3/4}$$

$$r = \pm\sqrt{b_{xy} \cdot b_{yx}}$$
$$= \sqrt{\tfrac{1}{3} \times \tfrac{3}{4}}$$
$$= \sqrt{\tfrac{1}{4}}$$
$$\boxed{r = 0.5}$$

$$b_{xy} = r \cdot \frac{\sigma x}{\sigma y}$$

Given that $V(x) = 25$

$$\sigma x = \sqrt{V(x)}$$
$$= \sqrt{25}$$
$$= 5$$

$$b_{xy} = r \cdot \frac{\sigma x}{\sigma y}$$

$$\frac{1}{3} = 0.5 \times \frac{5}{\sigma y}$$

$$\sigma y = 0.5 \times 5 \times 3$$

$$\boxed{\sigma y = 7.5}$$

$$b_{yx} = r\,\frac{\sigma y}{\sigma x}$$

$$\frac{3}{4} = 0.5\,\frac{\sigma y}{5}$$

$$\sigma y = \frac{(0.75)(5)}{0.5}$$

$$\boxed{\sigma y = 7.5}$$

---

3. Two regression equations are $2x - 3y + 6 = 0$
$4y - 5x - 8 = 0$

i) calulate the mean of x and Y

ii) If the standard deviation of x is 3 determine the standard deviation of y

(iii) also find the correlation coefficiant

(i) $2x - 3y + 6 = 0$
$4y - 5x - 8 = 0$

$2x - 3y = -6 \times 4$
$4y - 5x = -8 \times 3$

$$8x - 12y = 24$$
$$-15x + 12y = \pm 24$$

$$7x = 48$$
$$\boxed{x = 6.8} \quad \bar{x} = 0$$

$2x-3y+6=0$

$4y-5x-8=0$

$2x-3y=-6 \times 5$

$-5x+4y=8 \times 2$

$$\begin{aligned} \cancel{10x}-30y &= -60 \\ -\cancel{10x}+8y &= 16 \\ \hline -22y &= -44 \end{aligned}$$

$y=\frac{-44}{-22}$

$\boxed{y=2}$

$4(2)-5x-8=0$

$8-5x=8$

$5x=0$

$\boxed{x=0}$

(iii) $2x-3y+6=0$

$2x=3y-6$

$x=\frac{3}{2}y-\frac{6}{2}$

$\therefore \; b_{xy}=3/2 \quad =2/3$

$\cancel{b_{yx}=2/3}$

$4y-5x-8=0$

$4y=5x+8$

$y=\frac{5}{4}x+\frac{8}{4}$

$\therefore \; b_{yx}=5/4 \quad =4/5$

$\cancel{b_{xy}\ 4/5}$

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

$\sqrt{2/3\cdot 4/5} \quad = \sqrt{3/2\cdot 5/4} \quad =\sqrt{1.5\times 1.25} \quad =\sqrt{}$

$=\sqrt{0.5333} \quad = \sqrt{1.875}$

$\boxed{0.7302 \quad = r}$

$\boxed{\sigma_y=2.1386}$

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

$\frac{2}{3}=0.7302\left(\frac{3}{\sigma_y}\right)$

$\sigma_y=\frac{(0.7302)(3)(3)}{2}=\frac{6.5718}{2}$

$\boxed{\sigma_y=3.2859}$

Can $y = 5 + 2.8x$ and $x = 3 - 0.5y$ be the estimated regression equations of y on x and x on y respectively

solution:

The regression equation of y on x is

$$y = 5 + 2.8x$$

$$b_{yx} = 2.8$$

The regression equation of x on y is

$$x = 3 - 0.5y$$

$$b_{xy} = -0.5$$

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

$$= \pm\sqrt{2.8 \times -0.5}$$

$$= \pm\sqrt{-1.4}$$

This is not possible, since each of the regression coefficiants b_{xy} and b_{yx} must have the same sine

The two regression coefficiants are -0.5 and -1.3. What would be the value of the correlation coefficiant.

solution:

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

since σ_x and σ_y are positive, r should b negative.

Regression coefficient $b_{xy} = -0.5$

$b_{yx} = -1.3$

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

$r = \pm\sqrt{-0.5 \times -1.3}$

$r = \pm\sqrt{0.65}$

$\boxed{r = -0.8062}$

The regression line of Y on X and X on Y are $y = x+5$ $16x - 9y = 94$. find

(i) find $V(x)$ if $V(Y)$ is 16

(ii) find $\mathrm{cov}(x, y)$

solution:

$y = x+5$

$\therefore b_{yx} = 1$

$16x = 9y + 94$

$x = \frac{9y}{16} + \frac{94}{16}$

$\therefore b_{xy} = \frac{9}{16}$

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

$= \pm\sqrt{\frac{9}{16} \times 1} = \sqrt{0.5625}$

$r = 0.75$

$V(y) = 16$

$\sigma_y = \sqrt{16}$

$\sigma_y = 4$

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

$$\frac{9}{16} = 0.75 \left(\frac{\sigma_x}{4}\right)$$

$$\frac{19 \times 4}{16 \times 0.75} = \frac{19 \times 4}{16 \times 0.75}$$

$$= \frac{19}{3}$$

$$\sigma_x = 6.333$$

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

$$1 = 0.75 \left(\frac{4}{\sigma_x}\right)$$

$$\boxed{\sigma_x = 3}$$

$$\boxed{V(X) = 9}$$

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

$$0.75 = \frac{cov(x,y)}{3 \times 4}$$

$$0.75 \times 12 = cov(x,y)$$

$$\boxed{9 = cov(x,y)}$$

Fin V (X,Y) b_{xy} From the following data

X :	45	55	56	58	60	65	68	70	75	80	85
Y :	56	50	48	60	62	64	65	70	74	82	90

solution:

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

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

$$cov(x,y) = \frac{\Sigma xy}{N} - \bar{x}\bar{y}$$

$$\sigma_x = \sqrt{\frac{\Sigma x^2}{N} - (\bar{x})^2} \qquad \bar{x} = \frac{\Sigma x}{N}$$

$$\sigma_y = \sqrt{\frac{\Sigma y^2}{N} - (\bar{y})^2} \qquad \bar{y} = \frac{\Sigma y}{N}$$

X	Y	x^2	Y^2	XY
45	56	2025	3136	2520
55	50	3025	2500	2750
56	48	3136	2304	2688
58	60	3364	3600	3480
60	62	3600	3844	3720
65	64	4225	4096	4160
68	65	4624	4225	4420
70	70	4900	4900	4900
75	74	5625	5476	5550
80	82	6400	6724	6560
85	90	7225	8100	7200
717	721	48149	48905	48398

$$\bar{x} = \frac{\Sigma x}{N} = \frac{717}{11} = 65.1818$$

$$\bar{Y} = \frac{\Sigma y}{N} = \frac{721}{11} = 65.5454$$

$$\text{cov}(x,y) = \frac{\Sigma xy}{N} - (\bar{x})(\bar{y})$$

$$= \frac{48398}{11} - (65.1818)(65.5454)$$

$$= 4399.8181 - 4272.3671$$

$$\text{cov}(x,y) = \cancel{76}\ 127.451$$

$$\sigma_x = \sqrt{\frac{\Sigma x^2}{N} - (\bar{x})^2}$$

$$= \sqrt{\frac{48149}{11} - (65.1818)^2}$$

$$= \sqrt{4377.1818 - 4248.6670}$$

$$= \sqrt{128.5148}$$

$$\sigma_x = 11.3364$$

$$\sigma y = \sqrt{\frac{\Sigma y^2}{N} - (\bar{y})^2}$$

$$= \sqrt{\frac{48905}{11} - (65.5454)^2}$$

$$= \sqrt{4445.9090 - 4296.1994}$$

$$= \sqrt{149.7096}$$

$$\sigma y = 12.2355$$

$$r = \frac{cov(x, y)}{\sigma x . \sigma y}$$

$$= \frac{127.451}{(11.3364)(12.2355)}$$

$$= \frac{127.451}{138.7065}$$

$$\boxed{r = 0.9188}$$

$$b_{xy} = r . \frac{\sigma x}{\sigma y}$$

$$= 0.9188 \times \frac{11.3364}{12.2355}$$

$$= 0.9188 \times 0.9265$$

$$\boxed{b_{xy} = 0.8512}$$

Given that $N = 25$, $\Sigma x = 125$, $\Sigma x^2 = 650$, $\Sigma xy = 508$, $\Sigma y = 100$ $\Sigma y^2 = 460$

Find the two regression lines

regression line of x on y is

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

regression line of y on x is

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

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

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

$$b_{yx} = \frac{25(508) - (125)(100)}{25(650) - (125)^2}$$

$$= \frac{12700 - 12500}{16250 - 15625}$$

$$= \frac{200}{625}$$

$$b_{yx} = 0.32$$

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

$$= \frac{25(508) - (125)(100)}{25(460) - (100)^2}$$

$$= \frac{12700 - 12500}{11500 - 10000}$$

$$= \frac{200}{1500}$$

$$b_{xy} = 0.1333$$

$$\bar{x} = \frac{125}{25} = 5 \qquad \bar{y} = \frac{100}{25} = 4$$

X on Y is $(X-\bar{X}) = b_{xy}(Y-\bar{Y})$

$$(X-5) = 0.1333(Y-4)$$

$$X-5 = 0.1333Y - 0.5332$$

$$\boxed{X = 0.1333Y + 4.4668}$$

Y on X is $(Y-\bar{Y}) = b_{yx}(X-\bar{X})$

$$(Y-4) = 0.32\ (X-5)$$

$$Y-4 = 0.32X - 1.6$$

$$\boxed{Y = 0.32X + 2.4}$$

The following data gives the mark in subjects A & B in a certain examination. The mean marks in A is 39.5 & mean mark in B is 47.5. The S.D in marks A is 10.8 & S.D in marks B is 16.8. $r = 0.42$. Estimate the marks in B in a certain examination, who secured 51 marks in A

$\bar{A} = 39.5$ $\bar{B} = 47.5$

$\sigma_A = 10.8$ $\sigma_B = 16.8$ $r = 0.42$

regression equation of $X - \bar{X} = r \cdot \frac{\sigma_x}{\sigma_y}(Y-\bar{Y})$

$$X - 39.5 = 0.42 \times \left(\frac{10.8}{16.8}\right)(Y-47.5)$$

$$X = 0.27(Y-47.5) + 39.5$$

$$X = 0.27Y - 12.825 + 39.5$$

$$X = 0.27(51) + 26.675$$

$$X = 13.77 + 26.675$$

$$X = 40.445$$