Regression:

Y = a + bx (*)

Normal equations are

$$\sum y = na + b\sum x \quad (1)$$

And

$$\sum xy = a\sum x + b\sum x^2 \quad (2)$$

Now by solving above two equations for a and b

Multiply equation (1) by $\sum x$ and (2) by n

We can write:

$$\sum x \sum y = na\sum x + b\sum x\sum x$$

And

$$n\sum xy = na\sum x + nb\sum x^2$$

Subtracting these two equations

We get

$$\sum x\sum y - n\sum x\,y = b\left(\left(\sum x\right)^2 - n\sum x^2\right)$$

i.e.

$$n\sum x\,y - \sum x\sum y = b\left(n\sum x^2 - \left(\sum x\right)^2\right)$$

i.e.

$$b = \frac{n\sum x\,y - \sum x\sum y}{(n\sum x^2 - (\sum x)^2)}$$

Now,

Equation (1) can be written as

$$\frac{\sum y}{n} = a + \frac{b\sum x}{n}$$

i.e.

$$\bar{y} = a + b\bar{x} \qquad (3)$$

Now from (*) and (3), we can write

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

Regression line of y on x:

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

Where

$b_{yx} = \frac{n\sum xy - \sum x \sum y}{(n\sum x^2 - (\sum x)^2)}$ (using actual data: direct method)

$b_{yx} = \frac{\sum xy}{\sum x^2}$ $where\ x = x - \bar{x}\ and\ y = y - \bar{y}$ (using deviation taking actual mean)

$b_{yx} = \frac{n\sum d_x d_y - \sum d_x \sum d_y}{n\sum x^2 - (\sum x)^2}$ (using deviation form assumed mean)

Also, it is defined as,

$$\boldsymbol{y - \bar{y} = r_{xy}\frac{\sigma_y}{\sigma_x}(x - \bar{x})}$$

Where

$\boldsymbol{b_{yx} = r_{xy}\frac{\sigma_y}{\sigma_x}}$ **is known as regression coefficient of y on x**

Regression line of x on y:

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

Where

$b_{xy} = \frac{n\sum xy - \sum x \sum y}{(n\sum y^2 - (\sum y)^2)}$ (using actual data: direct method)

$b_{xy} = \frac{\sum xy}{\sum y^2}$ $where\ x = x - \bar{x}\ and\ y = y - \bar{y}$ (using deviation taking actual mean)

$b_{xy} = \frac{n\sum d_x d_y - \sum d_x \sum d_y}{n\sum y^2 - (\sum y)^2}$ (using deviation form assumed mean)

Also defined as

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

Where,

$\boldsymbol{b_{xy} = r_{yx}\frac{\sigma_x}{\sigma_y}}$ **is known as regression coefficient of x on y.**

Relation between coefficient of correlation r and regression coefficient, $\boldsymbol{b_{xy}\ and\ b_{yx}}$

$$\boldsymbol{r = \sqrt{b_{xy} * b_{yx}}}$$

Note: **<u>value of r is -1 to 1</u>**