Regression Analysis

Regression is a statistical method depends upon a Bivariate Data. Correlation determines the degree and nature of the relationship between two variables ie. Correlation of Bivariate Data is a numerical measure of the strength of association between two variables. But Regression Analysis is concerned with the estimation or prediction of most likely value of the variable (dependent variable) for some given value of the other variable (Independent variable). The statistical method by the help of which between two correlated values it is possible to find the unknown value of one variable (called dependent variable) from the known value of the another variable (called Independent variable) is known as Regression.

Regression deals with the determination of an equation for estimation of the variable to be estimated is called the Dependent variable, when the other variable known as Independent variable. This is simple Regression.

the mathematical measure of the average relationship between the variables is expressed in terms of suitable equations known as Regression equations.

Regression lines

The regression equations represents the Regression lines. there are two Regression equations, there are:

i) y on x :

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

where b_{yx} is the Regression co-efficient of y on x.

and,

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

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

ii) x on y :

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

where, b_{xy} is the Regression co-efficient of x on y.

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

r is the correlation co-efficient and σ_x, σ_y are the S.D of ; x and y resp.

Properties of Regression lines

1) The Product of two Regression co-efficients is equal to the square of the co-efficient of correlation.

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

2) b_{xy}, b_{yx} and r have the Same Sign. If b_{yx} and b_{xy} are both negative r is negative, if both Positive r is Positive.

3) The Regression lines always intersect at the Point $(\bar{x}, \bar{y})$. Where, $\bar{x}$ is the mean value of the variable x and $\bar{y}$ is the mean value of the variable y.

4) Two Regression lines will be Perpendicular to each other if r is zero. When $r = \pm 1$, they will coincide.

3) Problem - 1

Obtain the lines of Regression for the following

Data:

x	1	2	3	4	5	6	7	8	9
y	9	8	10	12	11	13	14	16	15

Table

x	y	$u = x-5$	$v = y-12$	u^2	v^2	uv
1	9	-4	-3	16	9	12
2	8	-3	-4	9	16	12
3	10	-2	-2	4	4	4
4	12	-1	0	1	0	0
5	11	0	-1	0	1	0
6	13	1	1	1	1	1
7	14	2	2	4	4	4
8	16	3	3	9	16	12
9	15	4	4	16	9	12

$n = 9$, $\sum x = 45$, $\sum y = 108$, $\sum u = 0$,

$\sum v = 0$, $\sum u^2 = 60$, $\sum v^2 = 60$,

$\sum uv = 57$.

$$b_{uv} = \frac{\frac{1}{n}\sum uv - \frac{\sum u}{n}\frac{\sum v}{n}}{\frac{\sum v^2}{n} - \left(\frac{\sum v}{n}\right)^2} = 0.95$$

$$b_{vu} = \frac{\frac{\sum uv}{n} - \frac{\sum u}{n}\cdot\frac{\sum v}{n}}{\frac{\sum u^2}{n} - \left(\frac{\sum u}{n}\right)^2} = 0.95$$

Problem-3

The lines of Regression of y on x and x on y are $y = x+5$ and $16x = 9y - 94$ resp. Find the variance of x if the variance of y is 16. Also find covariance of x and y.

Solution: The lines of Regression of y on x is,

$y = x + 5$.

$\therefore$ $b_{yx} = 1$.

The line of regression of x on y is,

$16x = 9y - 94$.

$\Rightarrow x = \left(\frac{9}{16}\right)y - \frac{94}{16}$.

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

$r^2 = b_{yx} \times b_{xy}$

$\Rightarrow r = \pm \frac{3}{4}$.

Since, b_{xy} and b_{yx} both are positive r is posit

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

$\Rightarrow 1 = \frac{3}{4} \cdot \frac{4}{\sigma_x}$

$\Rightarrow \sigma_x = 3$ $\Rightarrow$ $\boxed{\sigma_x^2 = 9}$

Variance of $x = \sigma_x^2 = 9$;

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

$\Rightarrow = \frac{3}{4} \cdot 3 \cdot 4 = 9$.

Problem-3

The lines of Regression of y on x and x on y are $y = x+5$ and $16x = 9y - 94$ resp. Find the variance of x if the variance of y is 16. Also find covariance of x and y.

Solution: The lines of Regression of y on x is,

$y = x + 5$.

$\therefore$ $b_{yx} = 1$.

The line of regression of x on y is,

$16x = 9y - 94$.

$\Rightarrow x = \left(\frac{9}{16}\right)y - \frac{94}{16}$.

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

$r^2 = b_{yx} \times b_{xy}$

$\Rightarrow r = \pm \frac{3}{4}$.

Since, b_{xy} and b_{yx} both are positive r is positive

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

$\Rightarrow 1 = \frac{3}{4} \cdot \frac{4}{\sigma_x}$

$\Rightarrow \sigma_x = 3 \Rightarrow \boxed{\sigma_x^2 = 9}$

Variance of $x = \sigma_x^2 = 9$;

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

$= \frac{3}{4} \cdot 3 \cdot 4 = 9$