Regression Dependence of one variable upon other.

⇒ Dependence of one variable called dependent variable on one or more independent variable.

⇒ It provides an equation to be used for estimating or predicting the avg value of the dependent variable from known values of independent variable.

⇒ The dependent variable is assumed to be random variable, where as the independent variable is assumed to have fixed values i.e they are choosen non-randomly.

⇒ Relation b/w expected value of the dependent variable and independent variable is called a Regression Relation.

⇒ When the dependente is represented by a straight line equation, the Regressio is called linear.

⇒ Dependent variable is also called regressand, predicand, The response variable or the explained variable.

⇒ Independent variable is also called non-random variable, regressor, the predictor, the regression variable or the explanatory variable.

Examples :-

1) The blood pressure of a person depends upon one's weight, age e.t.c.

2) The yield of crop depends upon the fertility of land, fertilizer applied, rainfall and Quality of seed.

3) Heights of children depends on their parents

4) Houses prices depends on area e.t.c.

⇒ If the relationship is exact then it is a deterministic relationship e.g

Area of circle $= \pi r^2$

Or Relationship b/w celcius and farenheit

$$F = 32 + \frac{9}{5}C$$

Such relations cannot be studied by regression.

⇒ If relationship is not exact i.e non-deterministic or probabilistic modal.

$$Y = \alpha + \beta X + \epsilon$$

ϵ are unknown random errors.

Example 2: The table below shows the number of absences, x, in a Calculus course and the final exam grade, y, for 7 students. Find the correlation coefficient and interpret your result.

x	1	0	2	6	4	3	3
y	95	90	90	55	70	80	85

You may use the facts that (double check this for practice)

$$\sum x = 19, \quad \sum y = 565, \quad \sum x^2 = 75, \quad \sum y^2 = 46,775, \quad \sum xy = 1,380.$$

Calculate the numerator:

$$n\sum(xy) - \left(\sum x\right)\left(\sum y\right) = 7 \cdot 1380 - 19 \cdot 565 = -1075$$

Then calculate the denominator:

$$\sqrt{n\sum x^2 - \left(\sum x\right)^2}\sqrt{n\sum y^2 - \left(\sum y\right)^2} = \sqrt{7 \cdot 75 - (19)^2}\sqrt{7 \cdot 46775 - (565)^2}$$
$$= \sqrt{164}\sqrt{8200} \approx 1159.66$$

Now, divide to get $r \approx \dfrac{-1075}{1159.66} \approx -0.93$.

Interpret this result: There is a strong negative correlation between the number of absences and the final exam grade, since r is very close to -1. Thus, as the number of absences increases, the final exam grade tends to decrease.