REGRESSION

A study of measuring the relationship between associated variables, wherein one variable is dependent on another independent variable, called as Regression. It is developed by **Sir Francis Galton in 1877** to measure the relationship of height between parents and their children.

Regression analysis is a statistical tool to study the nature and extent of functional relationship between two or more variables and to estimate (or predict) the unknown values of dependent variable from the known values of independent variable. The variable that forms the basis for predicting another variable is known as the Independent Variable and the variable that is predicted is known as dependent variable.

For example, if we know that two variables price (X) and demand (Y) are closely related we can find out the most probable value of X for a given value of Y or the most probable value of Y for a given value of X. Similarly, if we know that the amount of tax and the rise in the price of a commodity are closely related, we can find out the expected price for a certain amount of tax levy.

Uses of Regression Analysis:

1. It provides estimates of values of the dependent variables from values of independent variables.

2. It is used to obtain a measure of the error involved in using the regression line as a basis for estimation.

3. With the help of regression analysis, we can obtain a measure of degree of association or correlation that exists between the two variables.

4. It is highly valuable tool in economies and business research, since most of the problems of the economic analysis are based on cause and effect relationship.

Distinction between Correlation and Regression

Sl No	Correlation	Regression
1	It measures the degree and direction of relationship between the variables.	It measures the nature and extent of average relationship between two or more variables in terms of the original units of the data
2	It is a relative measure showing association between the variables.	It is an absolute measure of relationship.
3	Correlation Coefficient is independent of change of both origin and scale.	Regression Coefficient is independent of change of origin but not scale.
4	Correlation Coefficient is independent of units of measurement.	Regression Coefficient is not independent of units of measurement.
5	Expression of the relationship between the variables ranges from –1	Expression of the relationship between the variables may be in any
	to +1.	of the forms like: $Y = a + bX$ $Y = a + bX + cX^2$
6	It is not a forecasting device.	It is a forecasting device which can be used to predict the value of dependent variable from the given value of independent variable.
7	There may be zero correlation such as weight of wife and income of husband.	There is nothing like zero regression.

Regression Lines and Regression Equation:

Regression lines and regression equations are used synonymously. Regression equations are algebraic expression of the regression lines. Let us consider two variables: X & Y. If y depends

on x, then the result comes in the form of simple regression. If we take the case of two variable X and Y, we shall have two regression lines as the regression line of X on Y and regression line of Y on X. The regression line of Y on X gives the most probable value of Y for given value of X and the regression line of X on Y given the most probable value of X for given value of Y. Thus, we have two regression lines. However, when there is either perfect positive or perfect negative correlation between the two variables, the two regression line will coincide, i.e. we will have one line. If the variables are independent, r is zero and the lines of regression are at right angles i.e. parallel to X axis and Y axis. Therefore, with the help of simple linear regression model we have the following two regression lines

1. Regression line of Y on X: This line gives the probable value of Y (Dependent variable) for any given value of X (Independent variable).

Regression line of Y on X : $Y - \dot{Y} = byx\ (X - \dot{X})$ OR : $Y = a + bX$

2. Regression line of X on Y: This line gives the probable value of X (Dependent variable) for any given value of Y (Independent variable).

Regression line of X on Y : $X - \dot{X} = bxy\ (Y - \dot{Y})$ OR : $X = a + bY$

In the above two regression lines or regression equations, there are two regression parameters, which are “a” and “b”. Here “a” is unknown constant and “b” which is also denoted as “byx” or “bxy”, is also another unknown constant popularly called as regression coefficient. Hence, these “a” and “b” are two unknown constants (fixed numerical values) which determine the position of the line completely. If the value of either or both of them is changed, another line is determined.