Introduction to Regression Analysis

Regression analysis, in general sense, is concerned with the estimation or prediction of the unknown value of one variable when the value of other variable is known.

The term regression was first used by British Biometrician Sir Francis Galton (1877) while studying the relationship between the height of fathers and sons.

While correlation shows strength and direction of linear relationship between variables, regression provides equations, models and predictions. Regression analysis is one of the most important statistical tools which is extensively used in almost all sciences – Natural, Social and Physical.

Correlation vs. Regression

Correlation	Regression
• We consider two variables symmetrically. This means we can interchange between variables.	• There is an asymmetry in the way dependent and independent variables are treated. Here the dependent variable is random and independent variable is fixed and can't be interchanged.
• It can't tell us the amount of change in one variable to a unit change in other variable and hence can't predict future values.	• It tells us the amount of change in the dependent variable to a unit change in the independent variable, hence can predict future values.
• There may be nonsense relation in correlation.	• There is no such relation in regression.
• Correlation is not very useful for further mathematical treatment.	• Regression is widely used for further mathematical treatment.
• It indicates only linear relationship between variables.	• It indicates any type of relationship.

Regression Analysis

Regression analysis is a mathematical measure of the relationship between one dependent variable with one/more independent variables in terms of the original units of the data which can predict about the future values.

According to M.M. Blair, "Regression is the measure of the average relationship between two or more variables in terms of the original units of the data."

Regression is used to denote estimation or prediction of the average value of one variable for a specified value of the other variable. Regression expresses the relationship in the form of an equation.

In regression analysis, there are two types of variables.

- Dependent variable
- Independent variable

Dependent Variable (Y): The variable whose value is influenced or is to be predicted is called dependent variable.

Independent Variable (X): The variable which influences the values or is used for prediction is called independent variable.

Dependent Variable

Dependent variable is also known as-

- Regressed Variable
- Explained Variable
- Response Variable
- Endogenous Variable

Independent Variable

Independent variable is also known as-

- Regressor
- Predictor
- Explanatory Variable
- Exogenous Variable

Types of Regression

Simple Regression: It involves only two variables.

Multiple regression: It involves more than two variables.

Linear regression: If the regression curve is a straight line then there is a linear regression between the variables. Regression curve is a smooth curve which fitted to the set of paired data in regression analysis.

Non-linear regression: If the regression curve is not a straight line then there is a non-linear or curvilinear regression between the variables.

Advantages of Regression Analysis

- Regression analysis estimates the relationship that exists, on the average between the dependent variable and the independent variables.

- It shows the nature of relationship between two or more variables.

- It determines the effect of each of the explanatory variables on the dependent variable, controlling the effects of all other explanatory variables.

- Regression predicts the value of dependent variable for a given value of the independent variable.

- Regression analysis is used for forecasting, time series modelling and finding the causal effect between the variables.