LINEAR REGRESSION

INTRODUCTION

As we know that with the help of Correlation studies we can find out the linear relationship between two variables. Therefore Correlation coefficient measures the strength of linear relationship and direction of the correlation whether it is positive or negative.

But in case of two or more than two variable in which one variable is considered as an independent variable and another as dependent variable, and if we are interested in estimating the value of dependent variable for a particular value of independent variable, **we study regression analysis.**

For example we might be interested in estimation of production of a crop for particular amount of rainfall or in prediction of demand on the price or prediction of marks on the basis of study hours of students. In these types of cases, regression would be the choice of statisticians or researchers. In general sense, regression analysis means estimation or prediction of the unknown value of one variable from the other variable.

CONCEPT OF LINEAR REGRESSION

Regression analysis describes how the independent variable(s) is (are) related to the dependent variable i.e. regression analysis measures the average relationship between independent variables and dependent variable. The literal meaning of regression is "**stepping back towards the average**" which was used by British Biometrician Sir Francis Galton (1822-1911) regarding the height of parents and their offspring's.

Regression analysis is a mathematical measure of the average relationship between two or more variables.

There are two types of variables in regression analysis:

1. Independent variable

The variable which is used for prediction is called independent variable. It is also known as regressor or predictor or explanatory variable.

2. Dependent variable

The variable whose value is predicted by the independent variable is called dependent variable. It is also known as regressed or explained variable.

Regression can also be classified according to number of variables being used. If only two variables are being used this is considered as **simple regression** whereas the involvement of more than two variables in regression is categorized as **multiple regression**.

LINES OF REGRESSION

Regression lines are the lines of best fit which express the average relationship between variables. Here, the concept of lines of best fit is based on principle of least squares.

When two variables are considered in regression analysis, there are two regression lines

(i) Regression line of y on x and

(ii) Regression line of x on y.

Regression line of y on x is used to estimate or predict the value of dependent variable y for the given value of independent variable x. Estimate of y obtained by this line will be best because this line minimizes the sum of squares of the errors of the estimates in y. If x is considered as dependent variable and y as independent variable then regression line of x on y is used to estimate or predict the value of variable x for the given value of y. Estimate of x obtained by regression line of x on y will be best because it minimizes the sum of squares of the errors of the estimates in x.

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

This is known as regression line of y on x.

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

This is known as regression line of x on y.

When two regression lines cut each other at right angle (at the angle of 90 degree), it shows no correlation between y and x.

REGRESSION COEFFICIENTS

If regression line of y on x is

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

Then $\frac{r\sigma_y}{\sigma_x}$ is called the regression coefficient of y on x and it is denoted by b_{yx}. Thus, Regression coefficient of y on x ,

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

PROPERTIES OF REGRESSION COEFFICIENTS

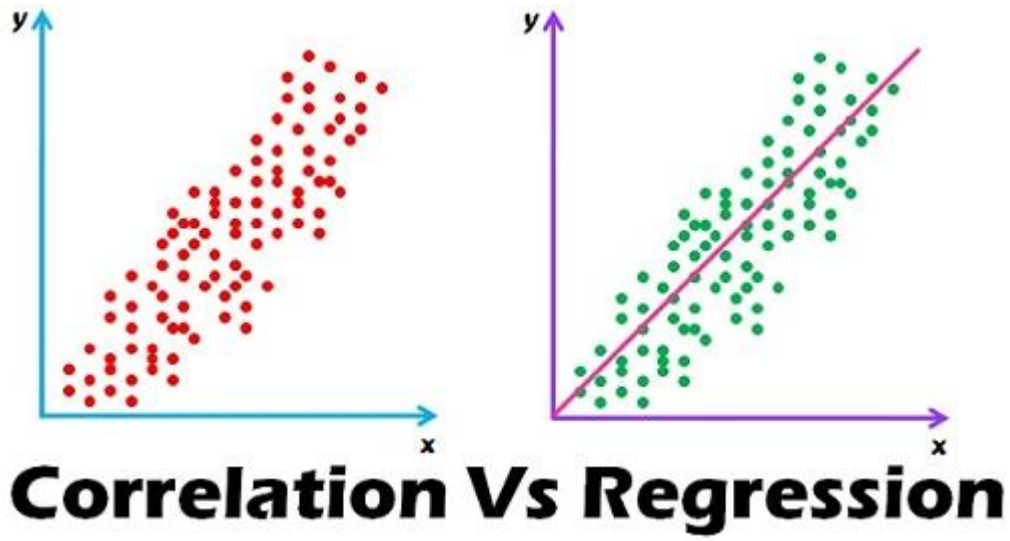
Property 1: Geometric mean of the regression coefficients is correlation coefficient.

Property 2: If one of the regression coefficients is greater than one, then other must be less than one.

Property 3: Arithmetic mean of the regression coefficients is greater than the correlation coefficient.

DISTINCTION BETWEEN CORRELATION AND REGRESSION

S.No.	CORRELATION	REGRESSION
1	Correlation studies the linear relationship between two variables.	Regression analysis is a mathematical measure of the average relationship between two or more variables
2	Correlation has limited application because it gives the strength of linear relationship.	The purpose of regression is to "predict" the value of the dependent variable for the given values of one or more independent variables.
3	Correlation makes no distinction between independent and dependent variables.	Linear regression does it, i.e. correlation does not consider the concept of dependent and independent variables while in regression analysis one variable is considered as dependent variable and other(s) is/are as independent variable(s).



Please Follow my LinkedIn Profile

https://www.linkedin.com/in/sumit-upadhyay-818836b2/