Linear Regression

Regression Analysis

Regression analysis refers to assessing the relationship between the outcome variable and one or more variables. The outcome variable is known as the dependent or response variable.
The dependent variable is shown by "*y*" and independent variables are shown by "*x*" in regression analysis.

Linear Regression

Linear regression is a linear approach to modeling the relationship between the scalar components and one or more independent variables.
It means that the linear regression strives to show the relationship between two variables by applying a linear equation to observed data. One variable is supposed to be an independent variable, and the other is to be a dependent variable.

For example, the weight of the person is linearly related to his height. Hence this shows a linear relationship between the height and weight of the person. As the height is increased, the weight of the person also gets increased.

If the regression has one independent variable, then it is known as a **simple linear regression**.
If it has more than one independent variables, then it is known as **multiple linear regression**.

Simple linear regression

- One dependent variable i.e. Y (interval or ratio)
- One independent variable i.e. X (interval or ratio)

The equation of Simple Linear Regression

As we known, linear regression is used to model the relationship between two variables.
Linear regression is the most basic and commonly used predictive analysis.

One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable. For example, a modeler might want to relate the weights of individuals to their heights using a linear regression model.

Thus, the formula for the simple linear regression equation can be written as:

$$\mathbf{Y = a + b\,X}$$

Where *X* is the independent variable and plotted along the x-axis
Y is the dependent variable and plotted along the y-axis

The slope of the line is b, and a is the intercept (the value of *y* when x = 0).

a (intercept) $= \bar{Y} - b\bar{X}$ or $\frac{\sum Y}{n} - b\frac{\sum X}{n}$

$b(slop) = N\sum XY - \left(\sum X\right)\left(\sum Y\right)/[N\sum X^2 - \left(\sum X\right)^2]$

Where, *x* and *y* are two variables on the regression line.
b = Slope of the line.
a = *y*-intercept of the line.
x = Values of the first data set.
y = Values of the second data set.

Example: - Find linear regression equation for the following two sets of data

x	2	4	6	8
y	3	7	5	10

Solution:
Construct the following table:

x	**y**	**x^2**	**xy**
2	3	4	6
4	7	16	28
6	5	36	30
8	10	64	80
∑x = 20	**∑y = 25**	**∑x2 = 120**	**∑xy = 144**

$b = n\sum xy - (\sum x)(\sum y)/[n\sum x^2 - (\sum x)^2]$

b = [4×144–(20×25)]/[4 (120)–400]

b = 0.95

$a = \frac{\sum Y}{n} - b\frac{\sum X}{n}$

$a = \frac{25}{4} - 0.95\frac{20}{4}$

a = 1.5

Linear regression is given by: $y = a + b\,x$

$y = 1.5 + 0.95\,x$

Regression Coefficient

In the linear regression line, we have seen the equation is given by;

$$Y = \beta_0 + \beta_1 X$$

Where, β_0 is a constant, β_1 is the regression coefficient.

Now, let us see the formula to find the value of the regression coefficient.

$$\beta_1 = b_1 = \sum [\,(x_i - \bar{x})(y_i - \bar{y})\,] / \sum (x_i - \bar{x})^2$$

Where x_i and y_i are the observed data sets.

And $\bar{x}$, $\bar{y}$ are the mean values.

Multiple Regression Definition

Multiple regression analysis is a statistical technique that analyzes the relationship between two or more variables and uses the information to estimate the value of the dependent variables.

In multiple regression, the objective is to develop a model that describes a dependent variable *y* to more than one independent variable.

Multiple Regression Formula

In linear regression, there is only one independent and dependent variable involved. But, in the case of multiple regression, there will be a set of independent variables that helps us to explain better or predict the dependent variable y.

The multiple regression equation is given by

$$y = a + b_1 x_1 + b_2 x_2 + ... + b_k x_k$$

where x_1, x_2, ….x_k are the ***k*** independent variables and ***y*** is the dependent variable.

Multiple Regression Analysis Definition

Multiple regression analysis permits to control explicitly for many other circumstances that concurrently influence the dependent variable.

The objective of regression analysis is to model the relationship between a dependent variable and one or more independent variables.

Let *k* represent the number of variables and denoted by x_1, x_2, x_3,…, x_k. Such an equation is useful for the prediction of value for ***y*** when the values of x are known.

Finally, the purpose of a simple linear regression, to predict scores on a dependent variable from scores on a single independent variable.

And the purpose of a multiple regression, is to predict scores on a dependent variable from scores on multiple independent variables.