Linear regression is a technique to predict the relation between dependent variable and independent variables. It is commonly used type of predictive analysis. There are many names for a regression’s dependent variable.  It may be called an outcome variable.  The independent variables can be called exogenous variables, predictor variables. Three major uses for regression analysis are (1) determining the strength of predictors,

(2) Forecasting an effect,

(3) Trend forecasting.

***Y = a + bX –* Linear Regression Formula**

First, the regression might be used to identify the strength of the effect that the independent variable(s) have on a dependent variable.

Second, it can be used to forecast effects or impact of changes.  That is, the regression analysis helps us to understand how much the dependent variable changes with a change in one or more independent variables.

Third, regression analysis predicts trends and future values.  The regression analysis can be used to get point estimates.

There are several types of linear regression analyses –

1. Simple linear regression
2. Multiple Linear Regression
3. Logistic Regression

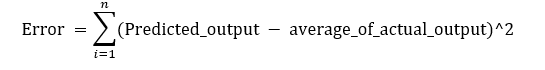
**Metrics for model evaluation -**

**R-Squared value**

This value ranges from 0 to 1. Value ‘1’ indicates predictor perfectly accounts for all the variation in Y. Value ‘0’ indicates that predictor ‘x’ accounts for no variation in ‘y’.

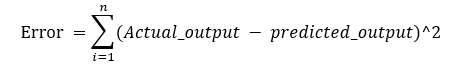
1. Regression sum of squares (SSR)

This gives information about how far estimated regression line is from the horizontal ‘no relationship’ line (average of actual output).



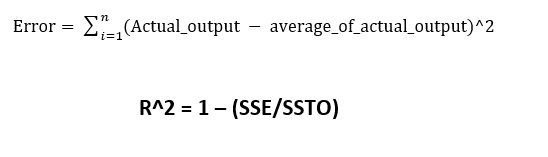
2. Sum of Squared error (SSE)

How much the target value varies around the regression line (predicted value).



3. Total sum of squares (SST)

This tells how much the data point move around the mean.



Let’s explain the concept of residue through an example. Consider, we have a dataset which predicts sales of juice when given a temperature of place. Value predicted from regression equation will always have some difference with the actual value. Sales will not match exactly with the true output value. This difference is called as residue.

Residual plot helps in analysing the model using the values of residues. It is plotted between predicted values and residue. Their values are standardized. The distance of the point from 0 specifies how bad the prediction was for that value. If the value is positive, then the prediction is low. If the value is negative, then the prediction is high. 0 value indicates prefect prediction. Detecting residual pattern can improve the model.

**Multiple Linear Regression:** This analysis will have any number of independent variables

 y = b1\*x1 + b2\*x2 +b3\*x3+…+ c