**LINEAR REGRESSION**

Definition:- In the most simple words, **Linear Regression** is the supervised Machine Learning model in which the **model finds the best fit linear line between the independent and dependent variable**

(or)

It finds the linear relationship between the dependent and independent variable

Advantages of linear regression:-

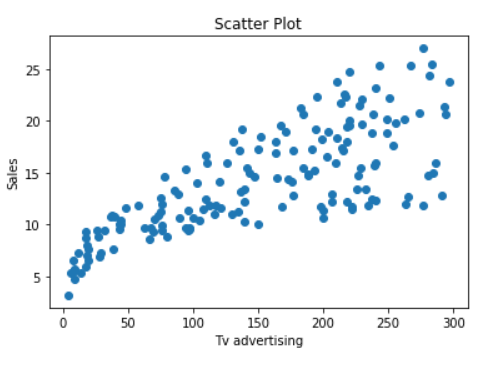
* Linear Regression is simple to implement and easier to interpret the output coefficients.
* If we know the relationship dependent and independent variable if the relationship is linear then this algorithm is best to use because of its less complexity when compared to other algorithm
* It  is the extrapolation beyond a specific data set

Disadvantages of linear regression:-

* Linear regression is quite sensitive to outliers
* Data must be independent
* It only looks at mean of dependent variable

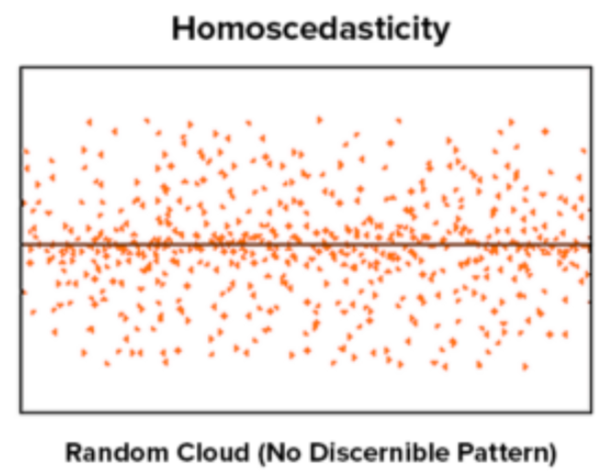
Assumptions of linear regression:-

1. **Linear relationship:-**  It captures the linear relationship between the feature and target. This can be validate by plotting a scatter plot between feature and the target

****

The above scatter plot of the feature TV vs Sales tells us that as the money invested on Tv advertisement increases the sales also increases linearly

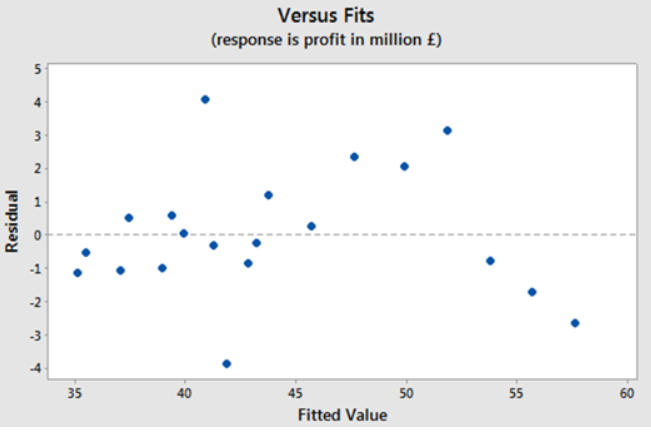
1. **Homoscedasticity:-**  It describes a situation in which the noise or random disturbance in the relationship between the feature and target is the same across the all values of independent variables



From the above graph we can observe that there is no specific pattern that is no constant among the

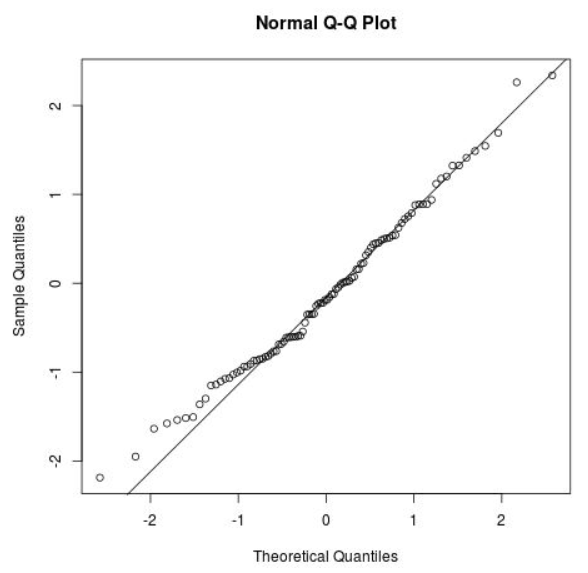
residuals

1. **Independence:-**  Here the residuals are independent, this is mostly relevant when we are working with the time series data. We don’t want any pattern among the consecutive residuals.



From the above we cannot have any positive auto correlation because we cannot see any cyclic pattern. Also it does not show any negative auto correlation.

1. **Normality:-** In this the residuals are normally distributed . We can check is assumption visually by using Q-Q plot means quantile-quantile plot it is a type of plot where we determine whether the residual model follows a normal distribution or not.



From the above Q-Q plot we can say that the residuals roughly follows normal distribution

**Types of linear regression:-** There are two types of linear regressions they are 1) Simple linear regression

2) Multiple linear regression

**Simple linear regression:-** With simple linear regression when we have a single input, we can use statistics to estimate

the coefficient . This requires that you calculate statistical properties from the data such as means, standard deviations,

correlations and covariance. All of the data must be available to traverse and calculate statistics.

Multiple linear regression:- **Multiple linear regression** is used to estimate the relationship between **two or more independent variables**and**one dependent variable**.