Linear Regression Assumptions

Machine Learning

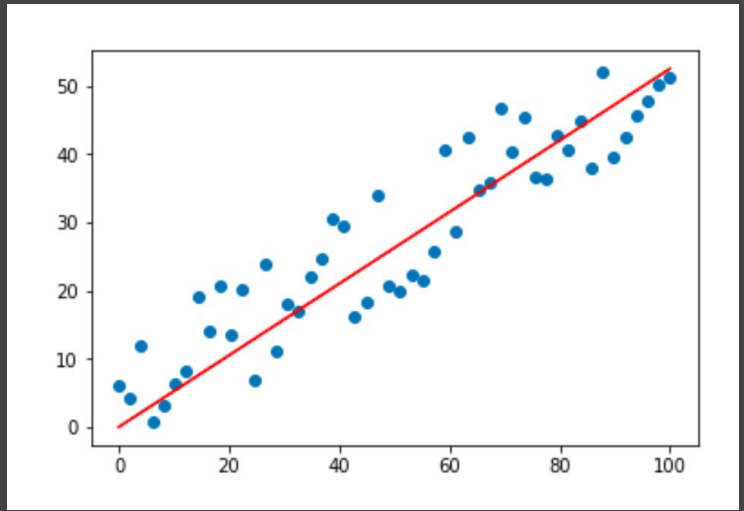
By Megha Singhal | 2020

Linear Relationship

Linear regression is a linear approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables).

Linear relationships can be expressed either in a graphical format or as a mathematical equation of the form

$y = mx + c.$



No Autocorrelation

The autocorrelation function is one of the tools used to find patterns in the data. Specifically, the autocorrelation function tells you the correlation between points separated by various time lags.

Test needed : Durbin - Watson Test.

Normality of Residual

The second assumption is Normality of Residuals.

For this we do Jarque Bera test. For a good model, the residuals should be normally distributed.The higher the value of Jarque Bera test , the lesser the residuals are normally distributed.We generally prefer a lower value of Jarque bera test.

The Jarque–Bera test is a goodness-of-fit test of whether sample data have the skewness and kurtosis matching a normal distribution.

Linearity of Residuals

Linearity means that the predictor variables in the regression have a straight-line relationship with the outcome variable.

To detect nonlinearity one can inspect plots of observed vs. predicted values or residuals vs. predicted values.

Homoscedasticity Test

Homoscedasticity - If the residuals are symmetrically distributed across the trend, then it is called as homoscedacious.

Heteroscedasticity - If the residuals are not symmetric across the trend, then it is called as heteroscedacious.

H0 = constant variance among residuals. (Homoscedasticity)

Ha = Heteroscedasticity.

No Multicollinearity

Multicollinearity is a condition in which the independent variables are highly correlated (r = 0.8 or greater) such that the effects of the independents on the outcome variable cannot be separated.

There should be no multicollinearity.

One way to handle multicollinearity is to remove highly correlated predictors from the model. If you have two or more factors with a high VIF, remove one from the model.

Next tutorials on Implementation of Linear Regression

Like and Subscribe

Thank you and Keep Learning

Linear Regression Assumptions