Future Connect Media

Machine Learning Part-B1

Part of Future Connect Media's IT Course

By Abhishek Sharma

Future Connect Media

Assumption of Linear Regression

Why?

Future Connect Media

Linear regression assumptions are important because they form the foundation of the statistical inference of the regression results. Violations of these assumptions can lead to biased and unreliable estimates/results, and can compromise the validity of the regression analysis. By considering and verifying these assumptions, we can ensure the accuracy and reliability of the linear regression models.

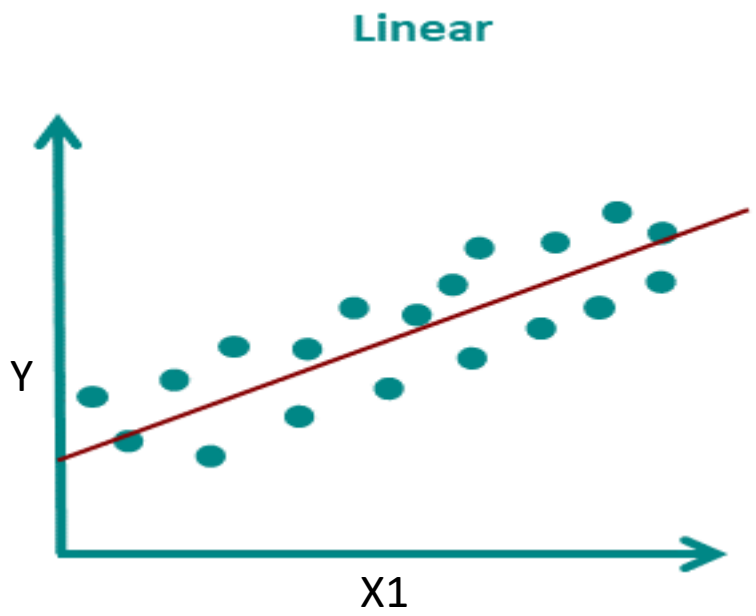
5 Assumptions of Linear Regression

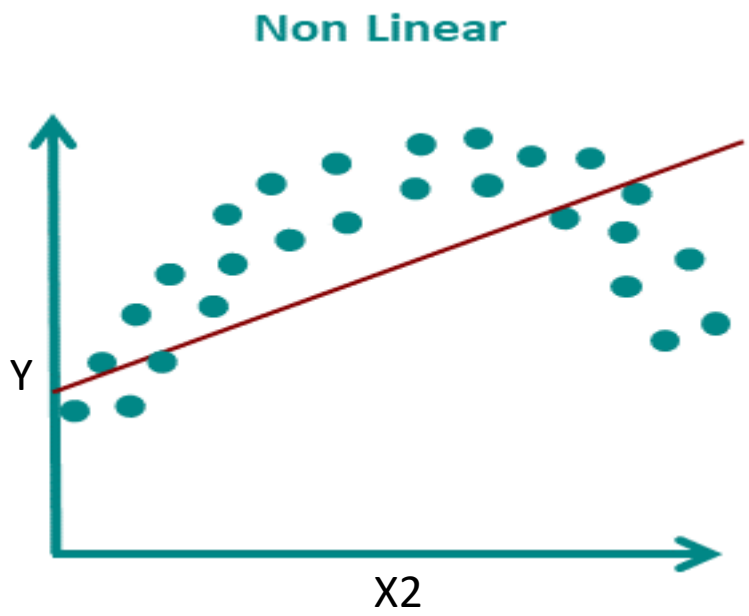
Future Connect Media

- Linear Relationship
- No or Little Multicollinearity
- Normality of Residuals
- Homoscedasticity
- Independence of Errors

Linear Relationship

The relationship between the dependent variable and the independent variables is assumed to be linear. This means that the change in the dependent variable is proportional to the change in the independent variables.





No or Little MultiCollinearity

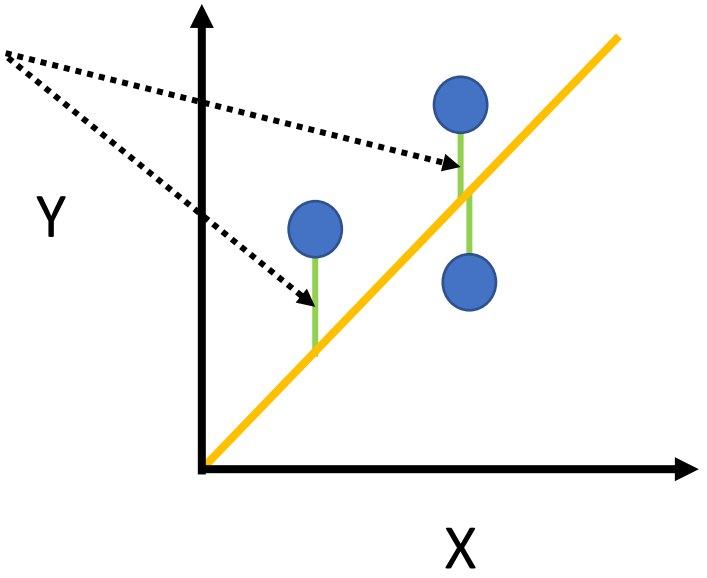
There should be no perfect multicollinearity among the independent variables. Multicollinearity occurs when there is a high correlation between two or more independent variables, making it difficult to separate their individual effects on the output(Dependent) variable.

MultiCollinearity between the Independent Variable can be found by calculating the Variation Inflation Factor (VIF). Variance inflation factor can estimate how much the variance of a regression coefficient is inflated due to multicollinearity.

Residuals

Future Connect Media

These **Green** lines
are called **Residuals**

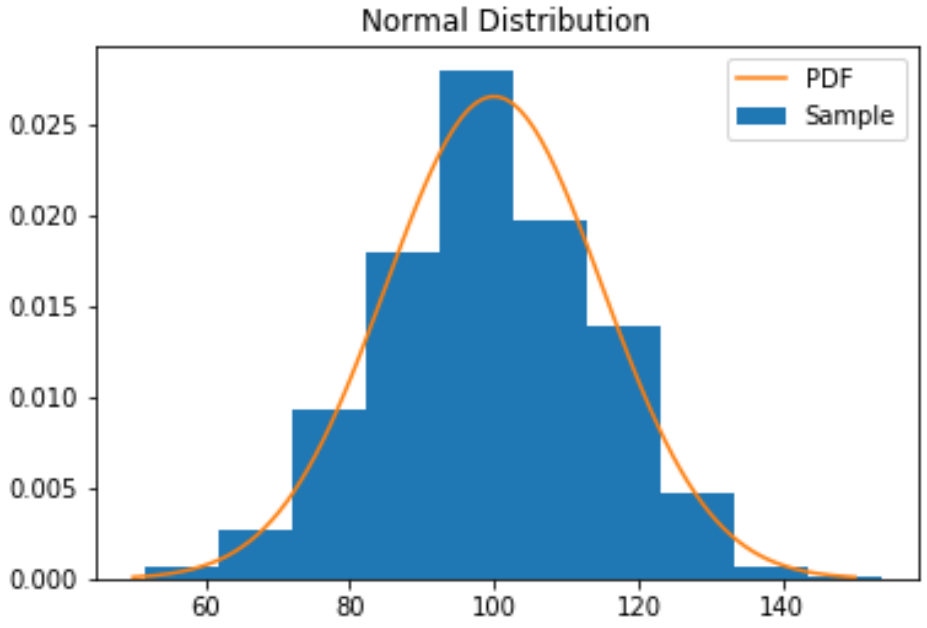


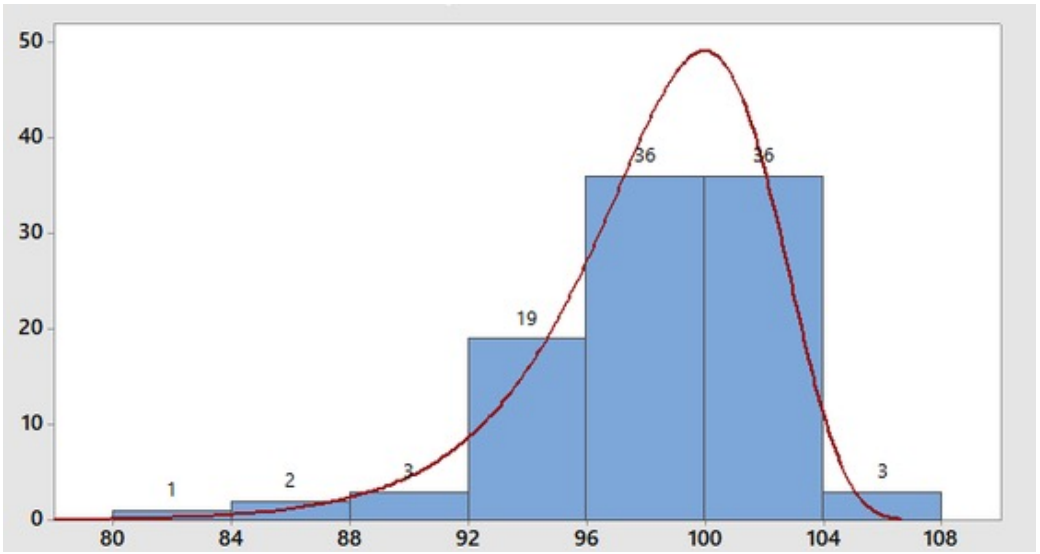
Residuals = Observed value – Predicted value

Normality of Residuals

Future Connect Media

Normality of residuals means that residuals are distributed symmetrically around zero with no skewness. This assumption implies that the model captures the main patterns and sources of variation in the data.

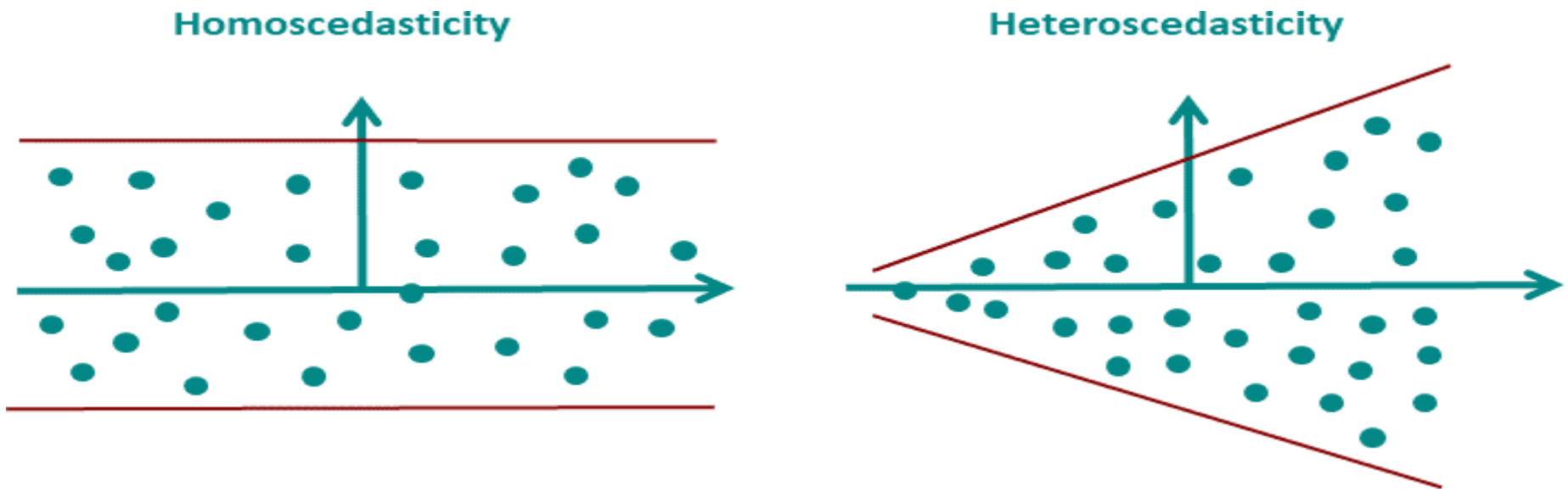




Homoscedasticity

Future Connect Media

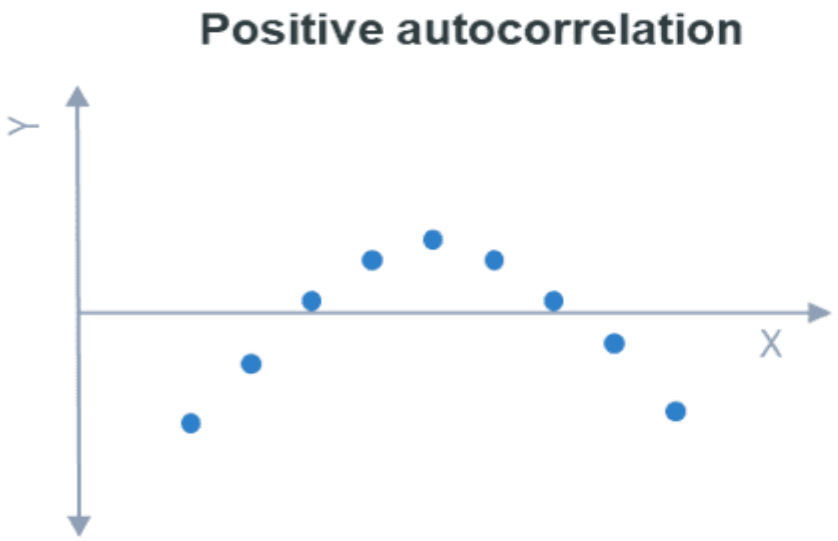
Homoscedasticity assumes that the variance of the residuals is constant across all levels of the independent variables. In other words, the error term does not vary much as the value of the predictor variable changes.

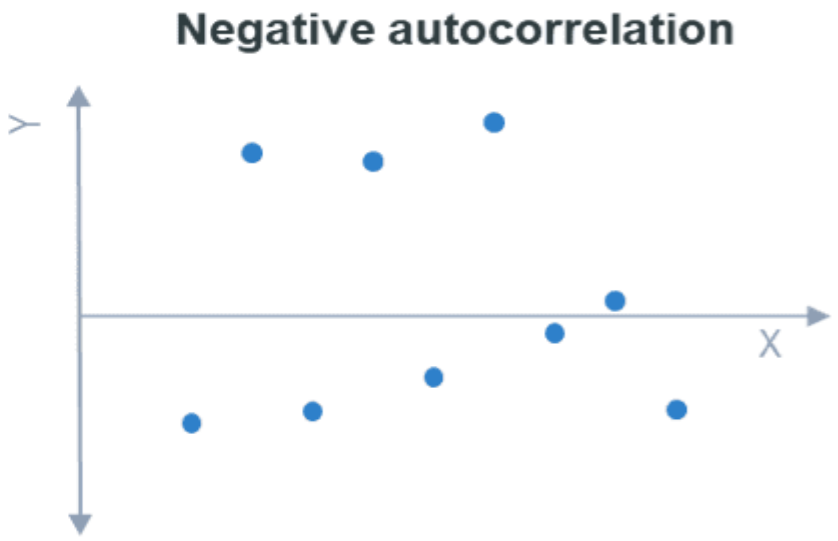


Independence of Errors

Future Connect Media

The observations should be independent of each other. There should be no systematic relationship or correlation between the residuals (the differences between the observed and predicted values) of the model.





Consequences of Violation

- **Inaccurate predictions**: If the relationship between the independent and dependent variables is not linear, the model will not be able to accurately predict the values of the dependent variable.
- **Underestimated standard errors:** If the residuals are not normally distributed, the standard errors of the regression coefficients will be underestimated, which could lead to incorrect inferences about the significance of the coefficients.
- **Inconclusive results:** If the residuals are not independent, the results of the analysis may be inconclusive.

Future
Connect
Media

Thank You