What are Assumptions in Regression?

Regression is a parametric approach. **'Parametric' means it makes assumptions about data for the purpose of analysis.** Due to its parametric side, regression is restrictive in nature. It fails to deliver good results with data sets which doesn't fulfil its assumptions. Therefore, for a successful regression analysis, it's essential to validate these assumptions.

What are Assumptions of Linear Regression?

Violations of assumptions of linear regression can lead to biased or inefficient estimates, and it is important to assess and address these violations for accurate and reliable regression results. Assumptions of linear regression include:

1. **Linearity:** The relationship between the dependent and independent variables is linear.
2. **Independence:** The observations are independent of each other.
3. **Homoscedasticity:** The variance of the errors is constant across all levels of the independent variables.
4. **Normality:** The errors follow a normal distribution.
5. **No multicollinearity:** The independent variables are not highly correlated with each other.

6. **No endogeneity:** There is no relationship between the errors and the independent variables.

Multiple Linear Regression | A Quick Guide (Examples)

Regression models are used to describe relationships between variables by fitting a line to the observed data. Regression allows you to estimate how a dependent variable changes as the independent variable(s) change.

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

How to perform a multiple linear regression

Multiple linear regression formula

The formula for a multiple linear regression is:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} + \epsilon$$

where, for $i = n$ observations:

y_i = dependent variable

x_i = expanatory variables

β_0 = y-intercept (constant term)

β_p = slope coefficients for each explanatory variable

ϵ = the model's error term (also known as the residuals)

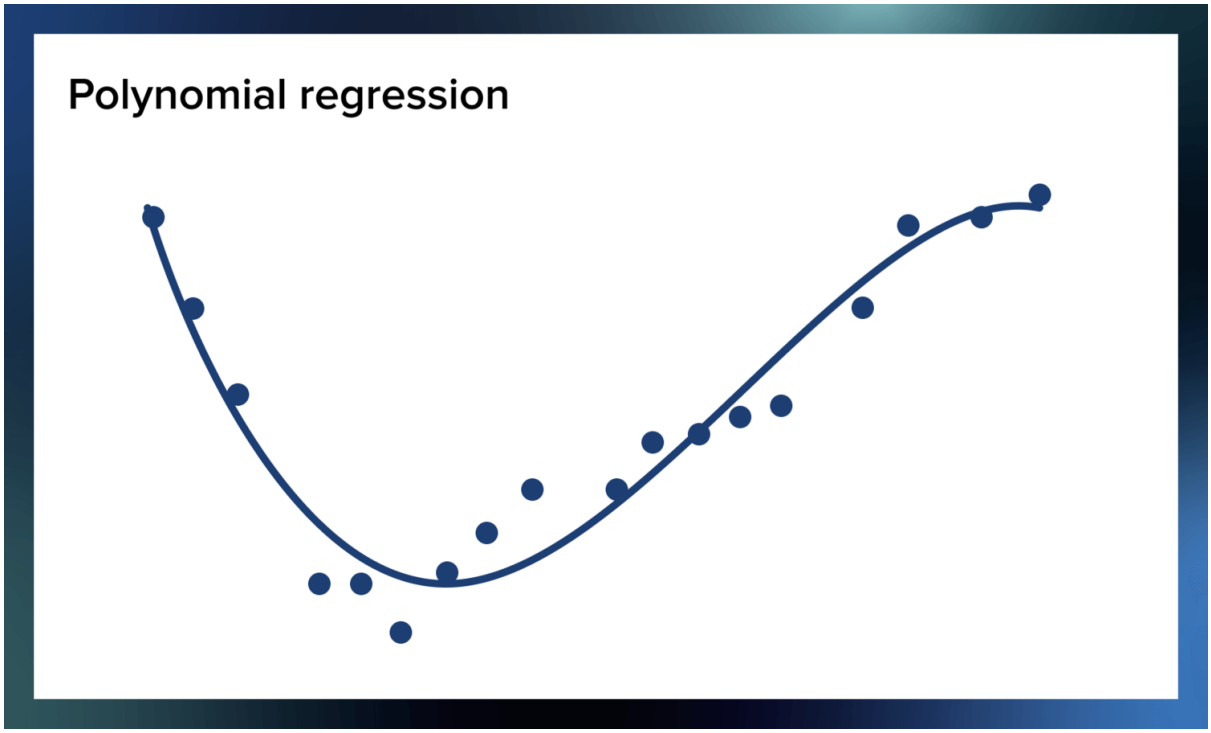
Why Polynomial Regression?

A simple linear regression algorithm only works when the relationship between the data is linear. But suppose we have non-linear data, then linear regression will not be able to draw a best-fit line. Simple regression analysis fails in such conditions.

How Does Polynomial Regression Handle Non-Linear Data?

Polynomial regression is a form of Linear regression **where only due to the Non-linear relationship between dependent and independent variables,** we add some polynomial terms to linear regression to convert it into Polynomial regression.

In polynomial regression, the relationship between the dependent variable and the independent variable is modeled as an nth-degree polynomial function. **When the polynomial is of degree 2, it is called a quadratic model;** when the degree of a polynomial is 3, it is called a **cubic model,** and so on.


Polynomial regression