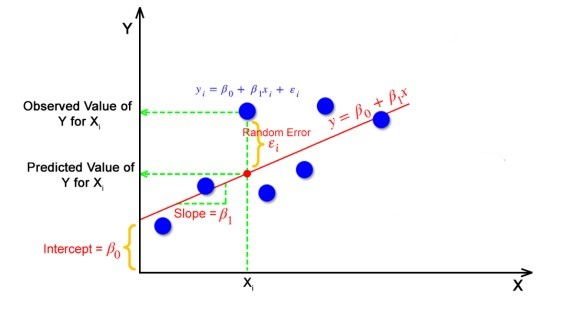
1. Using a graph to illustrate slope and intercept, define basic linear regression.

This algorithm explains the linear relationship between the dependent(output) variable y and the independent(predictor) variable X using a straight line  Y= B0 + B1 X.

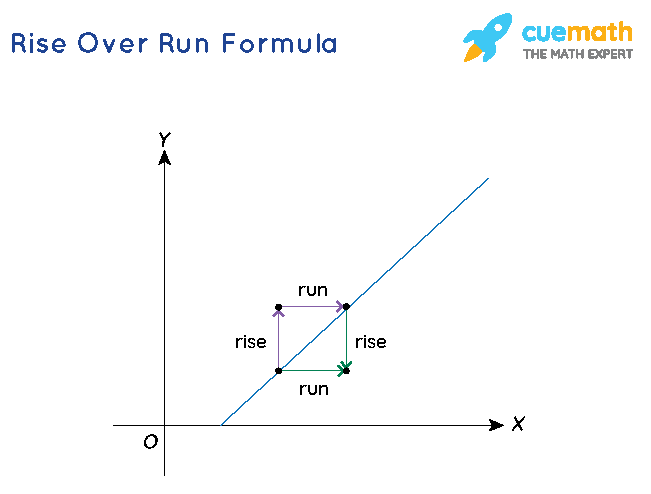


But how the linear regression finds out which is the best fit line?

The goal of the linear regression algorithm is to get the **best values for B0 and B1** to find the best fit line. The best fit line is a line that has the least error which means the error between predicted values and actual values should be minimum.

2. In a graph, explain the terms rise, run, and slope.

The difference between the y-coordinates of the two points is called the rise. The difference between the x-coordinates of the same two points is called the run. The **slope can be calculated by dividing the rise by run**.



3. Use a graph to demonstrate slope, linear positive slope, and linear negative slope, as well as the different conditions that contribute to the slope.

A linear equation in two variables is an equation of the form ax + by + c = 0 where a, b, c ∈ R, a, and b ≠ 0. A system of linear equations that has no solution is called an inconsistent pair of linear equations.

4. Use a graph to demonstrate curve linear negative slope and curve linear positive slope.

5. Use a graph to show the maximum and low points of curves.

6. Use the formulas for a and b to explain ordinary least squares.

7. Provide a step-by-step explanation of the OLS algorithm.

**OLS: Ordinary Least Square Method**

1. Set a difference between dependent variable and its estimation:
2. Square the difference:
3. Take summation for all data.
4. To get the parameters that make the sum of square difference become minimum, take partial derivative for each parameter and equate it with zero,

8. What is the regression's standard error? To represent the same, make a graph.

The standard error of the regression (S), also known as the standard error of the estimate, represents **the average distance that the observed values fall from the regression line**. Conveniently, it tells you how wrong the regression model is on average using the units of the response variable.

9. Provide an example of multiple linear regression.

For example, if you're doing a multiple regression to try to **predict blood pressure** (the dependent variable) from independent variables such as height, weight, age, and hours of exercise per week, you'd also want to include sex as one of your independent variables.

As an example, an analyst may want to know how the movement of the market affects the **price of ExxonMobil (XOM)**. In this case, their linear equation will have the value of the S&P 500 index as the independent variable, or predictor, and the price of XOM as the dependent variable.

10. Describe the regression analysis assumptions and the BLUE principle.

Assumption 1: Linear Relationship.

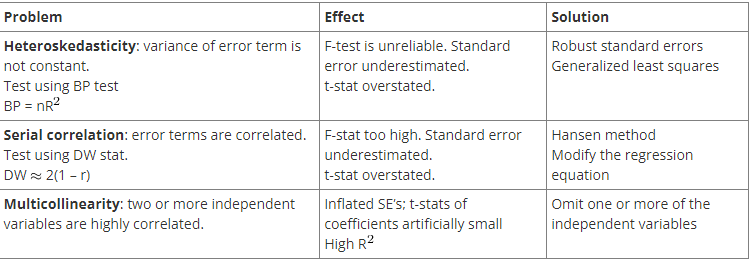
Assumption 2: Independence.

Assumption 3: Homoscedasticity.

Assumption 4: Normality.

BLUE is an acronym for the following: Best Linear Unbiased Estimator. In this context, the definition of “best” refers to the minimum variance or the narrowest sampling distribution.

11. Describe two major issues with regression analysis.

12. How can the linear regression model's accuracy be improved?

1. Add more data. Having more data is always a good idea. ...
2. Treat missing and Outlier values. ...
3. Feature Engineering. ...
4. Feature Selection. ...
5. Multiple algorithms. ...
6. Algorithm Tuning. ...
7. Ensemble methods.

13. Using an example, describe the polynomial regression model in detail.

In statistics, polynomial regression is a form of regression analysis in which the relationship between the independent variable x and the dependent variable y is modelled as an nth degree polynomial in x. For this reason, polynomial regression is considered to be a special case of multiple linear regression.

Polynomial regression is a special case of linear regression where we fit a polynomial equation on the data with a curvilinear relationship between the target variable and the independent variables.

14. Provide a detailed explanation of logistic regression.

Logistic regression is a statistical analysis method used to predict a data value based on prior observations of a data set. For example, a logistic regression could be **used to predict whether a political candidate will win or lose an election or whether a high school student will be admitted to a particular college**.

15. What are the logistic regression assumptions?

Basic assumptions that must be met for logistic regression include **independence of errors, linearity in the logit for continuous variables, absence of multicollinearity, and lack of strongly influential outliers**.

16. Go through the details of maximum likelihood estimation.

Maximum likelihood estimation is a method that will find the **values of μ and σ** that result in the curve that best fits the data. The goal of maximum likelihood is to find the parameter values that give the distribution that maximise the probability of observing the data.