1. Dependent vs. Independent Variable:

* Independent variable: Also known as the predictor variable, it's the variable you manipulate or measure in an experiment or study. It causes changes in the dependent variable.
* Dependent variable: Also known as the response variable, it's the variable you observe and that changes in response to the independent variable. It's what you're trying to predict or explain.

2. Simple Linear Regression:

It's a statistical technique to model the relationship between a single independent variable and a single dependent variable using a straight line. Example: Predicting house prices based on their square footage.

3. Slope in Linear Regression:

The slope in a linear regression equation (y = mx + b) represents the change in the dependent variable (y) for every unit change in the independent variable (x). Positive values indicate a direct relationship (as one increases, the other does too), while negative values indicate an inverse relationship (as one increases, the other decreases).

4. Graph Slope:

Calculating the slope (m) involves:

m = (y2 - y1) / (x2 - x1)

Substituting (2, 2) for (y1, x1) and (2, 2) for (y2, x2), we get:

m = (2 - 2) / (2 - 2)

Since both points share the same coordinates, the slope is undefined (division by zero).

5. Positive Slope:

In linear regression, a positive slope indicates a direct relationship between the variables. As the value of the independent variable increases, the value of the dependent variable also increases.

6. Negative Slope:

A negative slope signifies an inverse relationship. As the independent variable increases, the dependent variable decreases.

7. Multiple Linear Regression:

It extends simple linear regression to model the relationship between a dependent variable and multiple independent variables. It uses an equation with multiple terms, each representing the impact of an independent variable.

8. Number of Squares Due to Error (SSE):

It measures the sum of squared differences between the actual values of the dependent variable and the values predicted by the regression model. It represents the unexplained variation.

9. Number of Squares Due to Regression (SSR):

It measures the sum of squared differences between the predicted values and the mean of the dependent variable. It represents the explained variation.

10. Multicollinearity:

It occurs when two or more independent variables are highly correlated, making it difficult to assess their individual contribution to the model. It can lead to unreliable coefficient estimates and inaccurate predictions.

11. Heteroskedasticity:

It indicates unequal variance in the error terms across different values of the independent variable. It violates the assumption of constant variance, impacting the reliability of the model.

12. Ridge Regression:

It addresses multicollinearity by adding a penalty term to the cost function that shrinks the coefficients of correlated variables towards zero. This reduces the overall complexity of the model and improves its stability.

13. Lasso Regression:

Similar to ridge regression, it penalizes the model complexity, but it uses L1 regularization, which shrinks some coefficients to zero, resulting in feature selection (sparsification). It can be useful for interpreting models with many independent variables.

14. Polynomial Regression:

It models non-linear relationships by transforming the independent variable using polynomial terms (e.g., x^2, x^3) in the regression equation. This allows the model to capture more complex curves and patterns.

15. Basis Function:

In polynomial regression, each transformed term (e.g., x^2) acts as a basis function, creating a new feature space where the relationship becomes linear. Predictions are made by combining the results of linear regressions in this new space.

16. Logistic Regression:

It's a statistical technique for modeling the probability of a binary outcome (e.g., yes/no, spam/not spam) based on one or more independent variables. It uses a sigmoid function to transform the linear model's output into a probability between 0 and 1.