1.In a linear equation, what is the difference between a dependent variable and an independent variable?

Ans.

The independent variable is the variable that is manipulated or controlled in an experiment, while the dependent variable is the variable being measured or observed and affected by the independent variable.

2.What is the concept of simple linear regression? Give a specific example.

Ans.

Simple linear regression is a statistical method that helps to model the relationship between two variables by fitting a linear equation to the observed data. For example, a simple linear regression can be used to model the relationship between the weight of a person and their height.

3.In a linear regression, define the slope.

Ans.

The slope in linear regression is the measure of how steep the line is. It represents the rate of change of the dependent variable for each unit increase in the independent variable.

4.Determine the graph's slope, where the lower point on the line is represented as (3, 2) and the higher point is represented as (2, 2).

Ans.

The slope of the graph is undefined because the two points have the same y-coordinate and therefore lie on a horizontal line.

5.In linear regression, what are the conditions for a positive slope?

Ans.

In linear regression, a positive slope indicates that the dependent variable increases as the independent variable increases. The conditions for a positive slope are that the correlation between the two variables is positive and that the residual errors are normally distributed around the regression line.

6.In linear regression, what are the conditions for a negative slope?

Ans.

In linear regression, a negative slope indicates that the dependent variable decreases as the independent variable increases. The conditions for a negative slope are that the correlation between the two variables is negative and that the residual errors are normally distributed around the regression line.

7.What is multiple linear regression and how does it work?

Ans.

Multiple linear regression is a statistical method used to model the relationship between a dependent variable and two or more independent variables. It works by fitting a linear equation to the observed data, where the coefficients of the equation represent the effect of each independent variable on the dependent variable.

8.In multiple linear regression, define the number of squares due to error.

Ans.

The number of squares due to error in multiple linear regression is a measure of the unexplained variability in the dependent variable that is not accounted for by the independent variables. It represents the sum of the squared differences between the actual values of the dependent variable and the predicted values based on the regression equation.

9.In multiple linear regression, define the number of squares due to regression.

Ans.

The number of squares due to regression in multiple linear regression is a measure of the explained variability in the dependent variable that is accounted for by the independent variables. It represents the sum of the squared differences between the predicted values based on the regression equation and the mean value of the dependent variable.

10.In a regression equation, what is multicollinearity?

Ans.

Multicollinearity in a regression equation is a phenomenon where two or more independent variables are highly correlated with each other, leading to inaccurate and unstable estimates of the regression coefficients.

11.What is heteroskedasticity, and what does it mean?

Ans.

Heteroskedasticity is a phenomenon in which the variance of the error terms in a regression equation is not constant, but instead varies with the values of the independent variables. It means that the errors are not evenly distributed and can result in biased estimates of the regression coefficients.

12.Describe the concept of ridge regression.

Ans.

Ridge regression is a regularization technique used to prevent overfitting in a regression model by adding a penalty term to the sum of squared residuals. The penalty term shrinks the regression coefficients towards zero and helps to reduce the variance of the estimates, improving the model's predictive accuracy.

13.Describe the concept of lasso regression.

Ans.

Lasso regression is a regularization technique used to select the most important independent variables in a regression model by adding a penalty term to the sum of the absolute values of the regression coefficients. The penalty term encourages sparsity in the model, leading to a simpler and more interpretable model.

14.What is polynomial regression and how does it work?

Ans.

Polynomial regression is a type of regression analysis where the relationship between the dependent variable and the independent variable is modeled as an nth-degree polynomial. It works by fitting a curve to the observed data, where the degree of the polynomial determines the complexity of the curve.

15.Describe the basis function.

Ans.

A basis function is a mathematical function used to transform the input data in a regression model into a higher-dimensional space. The basis function can be a polynomial, exponential, or any other function that can map the input data into a space where the relationship between the dependent variable and the independent variables is more easily modeled.

16.Describe how logistic regression works.

Ans.

Logistic regression is a statistical method used to model the relationship between a binary dependent variable and one or more independent variables. It works by fitting a logistic function to the observed data, where the logistic function transforms the linear combination of the independent variables into a probability of the dependent variable being one of the two possible outcomes.