## Q1

Dependent vs. Independent Variables in a Linear Equation:

The dependent variable is the one you are trying to predict or explain. It depends on other variables.

The independent variable is the variable you believe has an influence on the dependent variable. It is not influenced by other variables in the context of the model.

## Q2

Simple Linear Regression:

Simple linear regression is a statistical method used to model the relationship between a single independent variable (predictor) and a dependent variable (response) by fitting a linear equation to observed data.

Example: Predicting a student's test score (dependent variable) based on the number of hours they studied (independent variable).

## Q3

Slope in Linear Regression:

In a linear regression equation (y = mx + b), the slope (m) represents the change in the dependent variable for a one-unit change in the independent variable. It measures the steepness or direction of the linear relationship between the variables.

## Q4

Calculating the Slope:

The slope (m) of a line passing through two points (x₁, y₁) and (x₂, y₂) is given by: m = (y₂ - y₁) / (x₂ - x₁).

Using the points (3, 2) and (2, 2), the slope is: m = (2 - 2) / (2 - 3) = 0 / (-1) = 0.

## Q5

Conditions for a Positive Slope in Linear Regression:

A positive slope indicates that as the independent variable increases, the dependent variable also increases.

The data points should generally exhibit an upward trend.

## Q6

Conditions for a Negative Slope in Linear Regression:

A negative slope indicates that as the independent variable increases, the dependent variable decreases.

The data points should generally exhibit a downward trend.

## Q7

Multiple Linear Regression:

Multiple linear regression extends simple linear regression to model the relationship between multiple independent variables and a dependent variable.

It uses a linear equation of the form: y = b₀ + b₁x₁ + b₂x₂ + ... + bₖxₖ, where there are multiple predictors (x₁, x₂, ..., xₖ).

## Q8

Sum of Squares Due to Error (SSE) in Multiple Linear Regression:

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

## Q9

Sum of Squares Due to Regression (SSR) in Multiple Linear Regression:

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

## Q10

Multicollinearity:

Multicollinearity refers to a situation where two or more independent variables in a regression model are highly correlated with each other. It can make it challenging to determine the individual effect of each variable on the dependent variable.

## Q11

Heteroskedasticity:

Heteroskedasticity occurs when the variability of the residuals (the differences between observed and predicted values) is not constant across all levels of the independent variable. It violates the assumption of constant variance in regression.

## Q12

Ridge Regression:

Ridge regression is a variant of linear regression that adds a penalty term to the linear regression objective function to address multicollinearity. It helps prevent overfitting by shrinking the coefficients of correlated predictors.

## Q13

Lasso Regression:

Lasso regression is another variant of linear regression that adds a penalty term to the objective function but uses L1 regularization. It not only addresses multicollinearity but also performs variable selection by setting some coefficients to exactly zero.

## Q14

Polynomial Regression:

Polynomial regression is a form of linear regression where the relationship between the independent and dependent variables is modeled as an nth-degree polynomial. It can capture nonlinear relationships in data.

## Q15

Basis Function:

In polynomial regression, a basis function is a mathematical function used to transform the original features into a higher-dimensional space to capture nonlinear relationships. Common basis functions include polynomials and radial basis functions.

## Q16

Logistic Regression:

Logistic regression is a classification algorithm used for binary and multiclass classification problems. It models the probability that an input belongs to a specific class. It uses the logistic function (sigmoid) to map the linear combination of input features to probabilities.