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

Answer: The independent variable is the variable that is manipulated or controlled in a study, and its values are chosen or observed by the experimenter. The dependent variable, on the other hand, is the outcome or response that is being measured. In a linear equation, the independent variable is often plotted on the x-axis, while the dependent variable is plotted on the y-axis.

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

Answer: Simple linear regression is a statistical method used to model the relationship between two variables, where one variable (the independent variable) is used to predict the values of another variable (the dependent variable). An example could be predicting a person's salary (dependent variable) based on their years of experience (independent variable).

**3. Question: In a linear regression, define the slope.**

Answer: The slope in a linear regression represents the change in the dependent variable's value for a unit change in the independent variable. Mathematically, it's the coefficient that indicates the rate of change between the two variables.

**4. Question: 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).**

Answer: In this case, the slope would be undefined because the points have the same y-coordinate (2), meaning that the line is horizontal. A horizontal line has a slope of 0.

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

Answer: A positive slope in linear regression indicates that as the independent variable increases, the dependent variable also increases. This relationship holds when the correlation between the two variables is positive, meaning that higher values of one variable are associated with higher values of the other.

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

Answer: A negative slope in linear regression indicates that as the independent variable increases, the dependent variable decreases. This relationship holds when the correlation between the two variables is negative, meaning that higher values of one variable are associated with lower values of the other.

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

Answer: Multiple linear regression is an extension of simple linear regression that involves multiple independent variables to predict a single dependent variable. It models a linear relationship between the dependent variable and multiple predictors by finding the best-fitting linear equation that minimizes the sum of squared differences between the observed and predicted values.

**8. Question: In multiple linear regression, define the sum of squares due to error.**

Answer: The sum of squares due to error (SSE) is a measure of the total variation that cannot be explained by the regression model. It quantifies the difference between the actual observed values and the predicted values from the regression equation.

**9. Question: In multiple linear regression, define the sum of squares due to regression.**

Answer: The sum of squares due to regression (SSR) is a measure of the variation in the dependent variable that is explained by the regression model. It represents the sum of squared differences between the predicted values and the mean of the dependent variable.

**10. Question: In a regression equation, what is multicollinearity?**

Answer: Multicollinearity occurs when two or more independent variables in a regression model are highly correlated with each other. This can lead to issues in interpreting the individual effects of each variable because their impacts become difficult to distinguish. It can also make the estimates of the regression coefficients unstable.

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

Answer: Heteroskedasticity refers to a situation in regression analysis where the variability of the residuals (the differences between the observed and predicted values) is not constant across all levels of the independent variable(s). This violates the assumption of homoskedasticity, which assumes that the variability of the residuals is constant.

**12. Question: Describe the concept of ridge regression.**

Answer: Ridge regression is a regularization technique used in linear regression to mitigate the effects of multicollinearity. It adds a penalty term to the regression equation that discourages large coefficients, thus reducing the impact of highly correlated variables. This helps stabilize the regression estimates and can improve model performance.

**13. Question: Describe the concept of lasso regression.**

Answer: Lasso (Least Absolute Shrinkage and Selection Operator) regression is another regularization technique used in linear regression. Similar to ridge regression, it adds a penalty term to the regression equation. However, lasso has the additional property of setting some regression coefficients exactly to zero, effectively performing variable selection and leading to a sparse model.

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

Answer: Polynomial regression is a type of regression analysis where the relationship between the independent and dependent variables is modeled as a polynomial function of a higher degree (e.g., quadratic, cubic). This allows the model to capture nonlinear relationships between the variables. The regression equation includes powers of the independent variable in addition to the linear term.

**15. Question: Describe the basis function.**

Answer: Basis functions are mathematical functions used in polynomial regression and other types of regression models to transform the input data into a higher-dimensional space. They allow the model to capture complex relationships between variables by creating additional features that are nonlinear transformations of the original features.

**16. Question: Describe how logistic regression works.**

Answer: Logistic regression is used for binary classification tasks, where the goal is to predict the probability that an instance belongs to a certain class. It models the relationship between the independent variables and the log-odds of the dependent variable using a logistic function (S-shaped curve). The logistic function maps the linear combination of input variables to a value between 0 and 1, which represents the predicted probability of belonging to the positive class.