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

* If x and y are two variables in an algebraic equation and every value of x is linked with any other value of y, then 'y' value is said to be a function of x value known as an independent variable, and 'y' value is known as a dependent variable.

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

* Linear regression is an attempt to model the relationship between two variables by fitting a linear equation to observed data, where one variable is considered to be an explanatory variable and the other as a dependent variable.
* We could use the equation to predict weight if we knew an individual's height. In this example, if an individual was 70 inches tall, we would predict his weight to be: Weight = 80 + 2 x (70) = 220 lbs. In this simple linear regression, we are examining the impact of one independent variable on the outcome

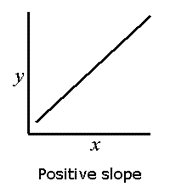
3. **In a linear regression, define the slope.**

* A linear regression line has an equation of the form Y = a + bX, where X is the explanatory variable and Y is the dependent variable. **The slope of the line is b**, and a is the intercept (the value of y when x = 0).

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).**

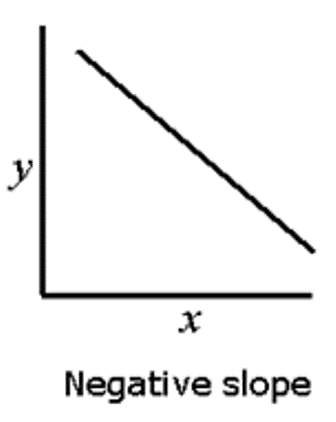
5. **In linear regression, what are the conditions for a positive slope?**

* In simpler words, a positive slope is one in which the variable x increases with the increase in variable y and/or variable y increases with the increase in variable x. Similarly, the variable x decreases with the decrease in variable y, and/or variable y decreases with the decrease in variable x.



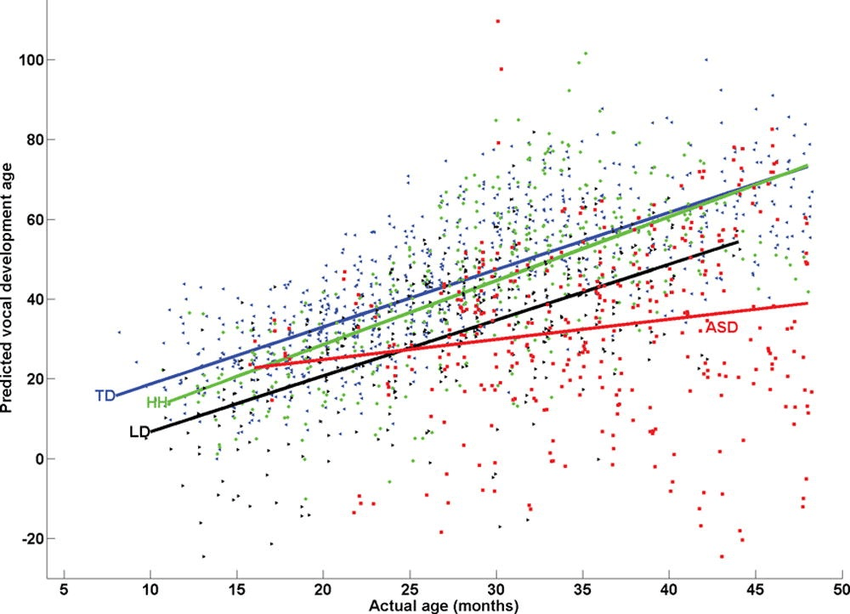
6. **In linear regression, what are the conditions for a negative slope?**

* If the slope is negative, y decreases as x increases and the function runs downhill.



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

* Multiple linear regression refers to a statistical technique that uses two or more independent variables to predict the outcome of a dependent variable. The technique enables analysts to determine the variation of the model and the relative contribution of each independent variable in the total variance.



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

* Residual sum of squares (also known as the sum of squared errors of prediction) The residual sum of squares essentially measures the variation of modeling errors. In other words, it depicts how the variation in the dependent variable in a regression model cannot be explained by the model.

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

* The sum of squares due to regression, or SSR is the sum of the differences between the predicted value and the mean of the dependent variable. Think of it as a measure that describes how well our line fits the data.

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

* Multicollinearity occurs when two or more independent variables are highly correlated with one another in a regression model. This means that an independent variable can be predicted from another independent variable in a regression model.

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

* Heteroscedasticity often occurs when there is a large difference among the sizes of the observations. A classic example of heteroscedasticity is that of income versus expenditure on meals. As one's income increases, the variability of food consumption will increase.

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

* Ridge regression is the method used for the analysis of multicollinearity in multiple regression data. It is most suitable when a data set contains a higher number of predictor variables than the number of observations.

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

* Lasso regression is **a regularization technique**. It is used over regression methods for a more accurate prediction. This model uses shrinkage. Shrinkage is where data values are shrunk towards a central point as the mean. The lasso procedure encourages simple, sparse models (i.e. models with fewer parameters).

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

* Polynomial Regression is a form of Linear regression known as a special case of Multiple linear regression which estimates the relationship as an nth degree polynomial. Polynomial Regression is sensitive to outliers so the presence of one or two outliers can also badly affect the performance.

15. **Describe the basis function.**

* In mathematics, a basis function is **an element of a particular basis for a function space**. Every function in the function space can be represented as a linear combination of basis functions, just as every vector in a vector space can be represented as a linear combination of basis vectors

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

* Logistic regression is a statistical analysis method to predict a binary outcome, such as yes or no, based on prior observations of a data set. A logistic regression model predicts a dependent data variable by analyzing the relationship between one or more existing independent variables.