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

Ans: The dependent variable is the output value and the independent variable is the input value

**Key Differences between Dependent and Independent Variables**

* A dependent variable is a variable whose value depends on another variable, whereas An Independent variable is a variable whose value never depends on another variable.
* The dependent variable is the presumed effect, On the other hand, the Independent variable is the presumed cause.
* Dependent variables are often referred to as the predicted variable, but independent variables are the predictors or regressors.
* Dependent variables are obtained from longitudinal research or by solving complex mathematical equations, On the contrary, Independent variables do not need any complex mathematical procedures and observations.
* Dependent variables are positioned vertically on the graph, while Independent variables are positioned horizontally on the graph.
* Any change in the dependent variable does not affect the independent variable, while Any change in the independent variable also affects the dependent variable.

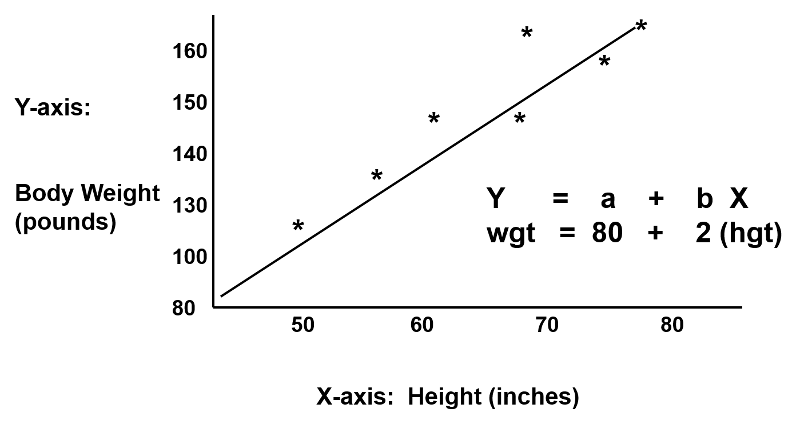
2. What is the concept of simple linear regression? Give a specific example.

Ans:

Simple linear regression is **a regression model that estimates the relationship between one independent variable and one dependent variable using a straight line**. Both variables should be quantitative.

In [statistics](https://en.wikipedia.org/wiki/Statistics), simple linear regression is a [linear regression](https://en.wikipedia.org/wiki/Linear_regression) model with a single [explanatory variable](https://en.wikipedia.org/wiki/Covariate). That is, it concerns two-dimensional sample points with [one independent variable and one dependent variable](https://en.wikipedia.org/wiki/Dependent_and_independent_variables) (conventionally, the x and y coordinates in a [Cartesian coordinate system](https://en.wikipedia.org/wiki/Cartesian_coordinate_system)) and finds a linear function (a non-vertical [straight line](https://en.wikipedia.org/wiki/Straight_line)) that, as accurately as possible, predicts the dependent variable values as a function of the independent variable. The adjective simple refers to the fact that the outcome variable is related to a single predictor.

For example, suppose that height was the only determinant of body weight. If we were to plot height (the independent or 'predictor' variable) as a function of body weight (the dependent or 'outcome' variable), we might see a very linear relationship, as illustrated below.



We could also describe this relationship with the equation for a line, Y = a + b(x), where 'a' is the Y-intercept and 'b' is the slope of the line. 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:

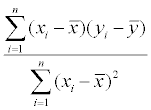
Weight = 80 + 2 x (70) = 220 lbs.

In this simple linear regression, we are examining the impact of one independent variable on the outcome. If height were the only determinant of body weight, we would expect that the points for individual subjects would lie close to the line. However, if there were other factors (independent variables) that influenced body weight besides height (e.g., age, calorie intake, and exercise level), we might expect that the points for individual subjects would be more loosely scattered around the line, since we are only taking height into account.

3. In a linear regression, define the slope.

Ans: In a regression line passing through a set of data points in data sets say X and Y, the slope is **the vertical distance divided by the horizontal distance between any two points on the line**. This ratio is also known as the rate of change along the line.

The slope of a [regression](https://www.chegg.com/learn/statistics/introduction-to-statistics/regression) [line](https://www.chegg.com/learn/calculus/calculus/lines-in-precalculus) (b) represents the rate of change in y as x changes. Because y is dependent on x, the slope describes the predicted values of y given x. When using the ordinary least squares method, one of the most common linear regressions, slope, is found by calculating b as the



covariance of x and y, divided by the sum of squares ([variance](https://www.chegg.com/learn/business/business-law/variance)) of x, . The slope must be calculated before the y-intercept when using a linear regression, as the intercept is calculated using the slope. The slope of a regression line is used with a [t-statistic](https://www.chegg.com/learn/statistics/introduction-to-statistics/t-statistic) to test the significance of a linear relationship between x and y.

First, the slope of a line is a measure of its steepness. In a line, slope is a ratio of the change in one variable to the change in the other. Usually, this refers to the change in y for each unit change in x, but sometimes other variables may be used.

| A red line in the 1st quadrant of a graph. The rise and run are shown in blue and labeled delta y and delta x respectively. |
| --- |

Slope is usually represented by the variable

m =Δy/Δx = y2−y1 / x2-−x1 = change in y/change in x

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: Slope is 2-2 /3-2 = 0

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

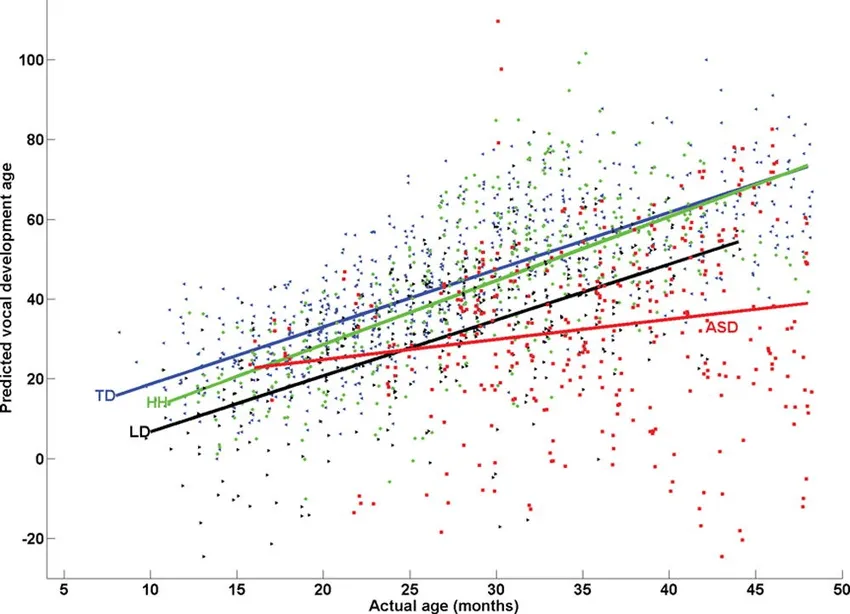
Ans: if the slope is positive, **y increases as x increases**, and the function runs "uphill" (going left to right)

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

Ans: 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?

Ans: Multiple linear regression (MLR), also known simply as multiple regression, is a statistical technique that **uses several explanatory variables to predict the outcome of a response variable.**

****

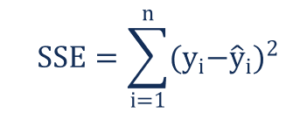
****

* yi​ is the dependent or predicted variable
* β0 is the y-intercept, i.e., the value of y when both xi and x2 are 0.
* β1 and β2 are the regression coefficients representing the change in y relative to a one-unit change in xi1 and xi2, respectively.
* βp is the slope coefficient for each independent variable
* ϵ is the model’s random error (residual) term.

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

Ans: 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. Generally, a lower residual sum of squares indicates that the regression model can better explain the data, while a higher residual sum of squares indicates that the model poorly explains the data.

The residual sum of squares can be found using the formula below:



Where:

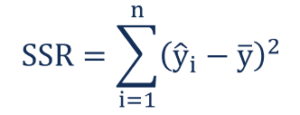
* yi – the observed value
* ŷi – the value estimated by the regression line

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

Ans:

The regression sum of squares describes how well a regression model represents the modeled data. A higher regression sum of squares indicates that the model does not fit the data well.

The formula for calculating the regression sum of squares is:



Where:

ŷi – the value estimated by the regression line

ȳ – the mean value of a sample

10. In a regression equation, what is multicollinearity?

Ans:

Multicollinearity exists **whenever an independent variable is highly correlated with one or more of the other independent variables in a multiple regression equation**. Multicollinearity is a problem because it will make the statistical inferences less reliable.

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

Ans:

Heteroskedasticity refers to **situations where the variance of the residuals is unequal over a range of measured values**. When running a regression analysis, heteroskedasticity results in an unequal scatter of the residuals (also known as the error term

12. Describe the concept of ridge regression.

Ans:

Ridge regression is **a method of estimating the coefficients of multiple-regression models in scenarios where the independent variables are highly correlated**.

It is particularly useful to mitigate the problem of [multicollinearity](https://en.wikipedia.org/wiki/Multicollinearity) in [linear regression](https://en.wikipedia.org/wiki/Linear_regression), which commonly occurs in models with large numbers of parameters. In general, the method provides improved [efficiency](https://en.wikipedia.org/wiki/Efficient_estimator) in parameter estimation problems in exchange for a tolerable amount of [bias](https://en.wikipedia.org/wiki/Bias_of_an_estimator). It is a method of [regularization](https://en.wikipedia.org/wiki/Regularization_(mathematics)) of [ill-posed problems](https://en.wikipedia.org/wiki/Ill-posed_problem).

Ridge regression was developed as a possible solution to the imprecision of least square estimators when linear regression models have some multicollinear (highly correlated) independent variables—by creating a ridge regression estimator (RR). This provides a more precise ridge parameters estimate, as its variance and mean square estimator are often smaller than the least square estimators.

This method performs L2 regularization. When the issue of multicollinearity occurs, least-squares are unbiased, and variances are large, this results in predicted values being far away from the actual values.

***Min(||Y – X(theta)||^2 + λ||theta||^2)***

Lambda is the penalty term. λ given here is denoted by an alpha parameter in the ridge function. So, by changing the values of alpha, we are controlling the penalty term. The higher the values of alpha, the bigger is the penalty and therefore the magnitude of coefficients is reduced.

## Bias and variance trade-off

Bias and variance trade-off is generally complicated when it comes to building ridge regression models on an actual dataset. However, following the general trend which one needs to remember is:

1. The bias increases as λ increases.
2. The variance decreases as λ increases.

13. Describe the concept of lasso regression.

Ans:

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..

**lasso** (**least absolute shrinkage and selection operator**; also **Lasso** or **LASSO**) is a [regression analysis](https://en.wikipedia.org/wiki/Regression_analysis) method that performs both [variable selection](https://en.wikipedia.org/wiki/Variable_selection) and [regularization](https://en.wikipedia.org/wiki/Regularization_(mathematics)).

Lasso can set coefficients to zero, while the superficially similar ridge regression cannot. This is due to the difference in the shape of their constraint boundaries. Both lasso and ridge regression can be interpreted as minimizing the same objective function

## Choice of regularization parameter

Choosing the regularization parameter () is a fundamental part of lasso. A good value is essential to the performance of lasso since it controls the strength of shrinkage and variable selection, which, in moderation can improve both prediction accuracy and interpretability. However, if the regularization becomes too strong, important variables may be omitted and coefficients may be shrunk excessively, which can harm both predictive capacity and inferencing. [Cross-validation](https://en.wikipedia.org/wiki/Cross-validation_(statistics)) is often used to find the regularization parameter.

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

Ans:

Polynomial regression, abbreviated E(y |x), **describes the fitting of a nonlinear relationship between the value of x and the conditional mean of y**. It usually corresponded to the least-squares method.According to the Gauss Markov Theorem, the least square approach minimizes the variance of the coefficients

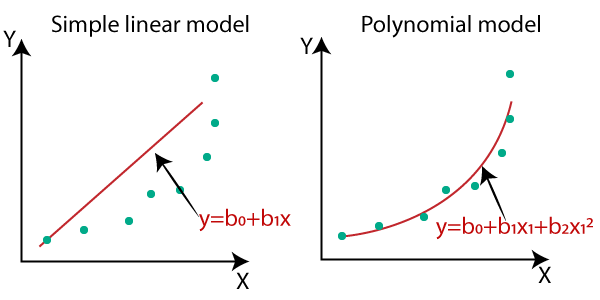
* Polynomial Regression is a regression algorithm that models the relationship between a dependent(y) and independent variable(x) as nth degree polynomial. The Polynomial Regression equation is given below:

y= b0+b1x1+ b2x12+ b2x13+...... bnx1n

* It is also called the special case of Multiple Linear Regression in ML. Because we add some polynomial terms to the Multiple Linear regression equation to convert it into Polynomial Regression.
* It is a linear model with some modification in order to increase the accuracy.
* The dataset used in Polynomial regression for training is of non-linear nature.
* It makes use of a linear regression model to fit the complicated and non-linear functions and datasets.
* **Hence, *"In Polynomial regression, the original features are converted into Polynomial features of required degree (2,3,..,n) and then modeled using a linear model."***

Need for Polynomial Regression:

* The need of Polynomial Regression in ML can be understood in the below points:
* If we apply a linear model on a linear dataset, then it provides us a good result as we have seen in Simple Linear Regression, but if we apply the same model without any modification on a non-linear dataset, then it will produce a drastic output. Due to which loss function will increase, the error rate will be high, and accuracy will be decreased.
* So for such cases, where data points are arranged in a non-linear fashion, we need the Polynomial Regression model. We can understand it in a better way using the below comparison diagram of the linear dataset and non-linear dataset.



* In the above image, we have taken a dataset which is arranged non-linearly. So if we try to cover it with a linear model, then we can clearly see that it hardly covers any data point. On the other hand, a curve is suitable to cover most of the data points, which is of the Polynomial model.
* Hence, if the datasets are arranged in a non-linear fashion, then we should use the Polynomial Regression model instead of Simple Linear Regression.

15. Describe the basis function.

Ans:

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.

**Basis functions** (called *derived features* in machine learning) are building blocks for creating more complex functions. In other words, they are a set of k standard functions, combined to estimate another function—one which is difficult or impossible to model exactly.

For example, individuals powers of x— the basis functions 1, x, x2, x3…— can be strung together to form a [polynomial function](https://www.statisticshowto.com/types-of-functions/polynomial-function/). The set of basis functions used to create the more complex function is called a basis set.

It’s possible to create many complex functions by hand; IDeally, you’ll want to work with a set of as few functions as possible. However, many real-life scenarios involve thousands of basis functions, necessitating the need for a computer.

## Common Basis Functions

The most common types of basis functions in calculus are:

1. **Polynomial basis**: 1, x, x2, x3…—
2. **B-Spline basis:** a set of k polynomial functions, each of a specified order d. An order is the number of constants required to define the function (Ramsay and Silverman, 2005; Ramsay et al., 2009). Popular for [non-periodic](https://www.statisticshowto.com/periodic-function/#aperiodic) data.
3. **Fourier basis:** a set of [sine functions](https://www.statisticshowto.com/sine-function/) and [cosine functions](https://www.statisticshowto.com/types-of-functions/trigonometric-function/#cosine): 1, sin(ωx), cos(ωx), sin(2ωx), cos(2ωx), sin(3ωx), cos(3ωx)&hellip;. These are often used to form [periodic functions](https://www.statisticshowto.com/periodic-function/). Derivatives for these functions are easy to calculate but aren’t suitable for modeling [discontinuous functions](https://www.statisticshowto.com/calculus-definitions/types-of-discontinuity/#definition) (Svishcheva et al., 2015).

In machine learning, popular methods include:

1. Artificial Neural Networks,
2. Basis function regression (including Radial Basis Functions),
3. k-Nearest Neighbors.

16. Describe how logistic regression works.

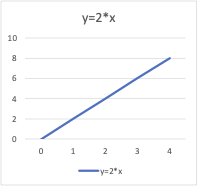
Ans:

Logistic regression is a data analysis technique that **uses mathematics to find the relationships between two data factors**. It then uses this relationship to predict the value of one of those factors based on the other. The prediction usually has a finite number of outcomes, like yes or no.

For example, let’s say you want to guess if your website visitor will click the checkout button in their shopping cart or not. Logistic regression analysis looks at past visitor behavior, such as time spent on the website and the number of items in the cart. It determines that, in the past, if visitors spent more than five minutes on the site and added more than three items to the cart, they clicked the checkout button. Using this information, the logistic regression function can then predict the behavior of a new website visitor.

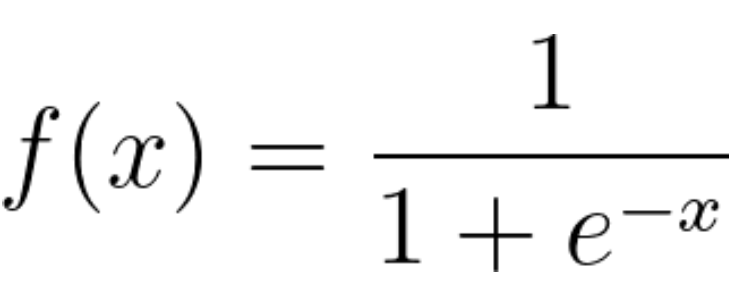
### Equations

In mathematics, equations give the relationship between two variables: *x* and *y*. You can use these equations, or functions, to plot a graph along the x-axis and y-axis by putting in different values of *x* and *y*. For instance, if you plot the graph for the function *y* = 2\**x*, you will get a straight line as shown below. Hence this function is also called a linear function.

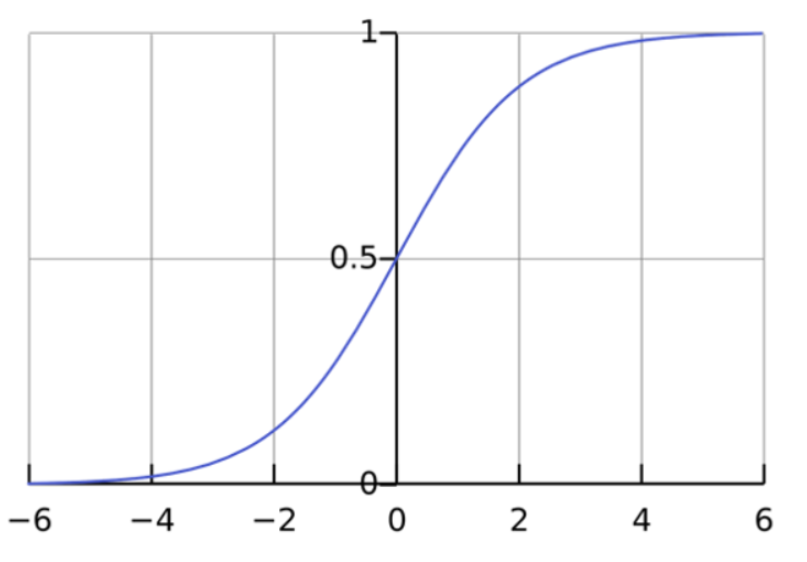


### Logistic regression function

### Logistic regression is a statistical model that uses the logistic function, or logit function, in mathematics as the equation between *x* and *y*. The logit function maps *y* as a sigmoid function of *x*.



If you plot this logistic regression equation, you will get an S-curve as shown below.



As you can see, the logit function returns only values between 0 and 1 for the dependent variable, irrespective of the values of the independent variable. This is how logistic regression estimates the value of the dependent variable. Logistic regression methods also model equations between multiple independent variables and one dependent variable.

### Logistic regression analysis with multiple independent variables

Logistic regression formulas assume a linear relationship between the different independent variables. You can modify the sigmoid function and compute the final output variable as

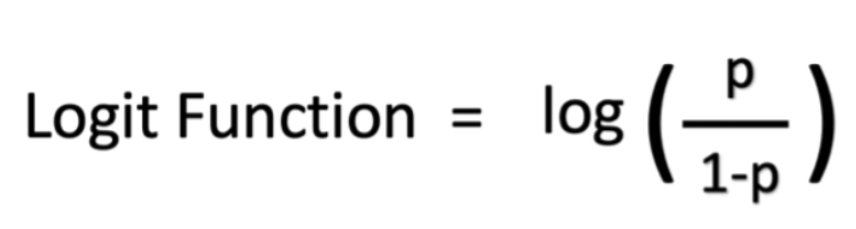
*y* = *f*(β0 + β1*x*1 + β2*x*2+… βn*x*n)

The symbol β represents the regression coefficient. The logit model can reverse calculate these coefficient values when you give it a sufficiently large experimental dataset with known values of both dependent and independent variables.

### Log odds

The logit model can also determine the ratio of success to failure or log odds. For example, if you were playing poker with your friends and you won four matches out of 10, your odds of winning are four sixths, or four out of six, which is the ratio of your success to failure. The probability of winning, on the other hand, is four out of 10.

Mathematically, your odds in terms of probability are *p*/(1 - *p*), and your log odds are log (*p*/(1 - *p*)). You can represent the logistic function as log odds as shown below:



## What are the types of logistic regression analysis?

There are three approaches to logistic regression analysis based on the outcomes of the dependent variable.

### Binary logistic regression

Binary logistic regression works well for binary classification problems that have only two possible outcomes. The dependent variable can have only two values, such as yes and no or 0 and 1.

Even though the logistic function calculates a range of values between 0 and 1, the binary regression model rounds the answer to the closest values. Generally, answers below 0.5 are rounded to 0, and answers above 0.5 are rounded to 1, so that the logistic function returns a binary outcome.

### Multinomial logistic regression

Multinomial regression can analyze problems that have several possible outcomes as long as the number of outcomes is finite. For example, it can predict if house prices will increase by 25%, 50%, 75%, or 100% based on population data, but it cannot predict the exact value of a house.

Multinomial logistic regression works by mapping outcome values to different values between 0 and 1. Since the logistic function can return a range of continuous data, like 0.1, 0.11, 0.12, and so on, multinomial regression also groups the output to the closest possible values.

### Ordinal logistic regression

Ordinal logistic regression, or the ordered logit model, is a special type of multinomial regression for problems in which numbers represent ranks rather than actual values. For example, you would use ordinal regression to predict the answer to a survey question that asks customers to rank your service as poor, fair, good, or excellent based on a numerical value, such as the number of items they purchase from you over the year.

## How does logistic regression compare to other ML techniques?

The two common data analysis techniques are linear regression analysis and deep learning.

### Linear regression analysis

As explained above, linear regression models the relationship between dependent and independent variables by using a linear combination. The linear regression equation is

*y*= β0*X*0 + β1*X*1 + β2*X*2+… βn*X*n+ ε, where β1 to βn and ε are regression coefficients.

#### Logistic regression vs. linear regression

Linear regression predicts a continuous dependent variable by using a given set of independent variables. A continuous variable can have a range of values, such as price or age. So linear regression can predict actual values of the dependent variable. It can answer questions like "What will the price of rice be after 10 years?"

Unlike linear regression, logistic regression is a classification algorithm. It cannot predict actual values for continuous data. It can answer questions like "Will the price of rice increase by 50% in 10 years?"

### Deep learning

Deep learning uses neural networks or software components that simulate the human brain to analyze information. Deep learning calculations are based on the mathematical concept of vectors.

#### Logistic regression vs. deep learning

Logistic regression is less complex and less compute intensive than deep learning. More importantly, deep learning calculations cannot be investigated or modified by developers, due to their complex, machine-driven nature. On the other hand, logistic regression calculations are transparent and easier to troubleshoot.