**Statistics**

# **Statistics for Datascience**

## **Descriptive Statistics**

Descriptive statistics are brief informational coefficients that summarize a given data set, which can be either a representation of the entire population or a sample of a population. Descriptive statistics are broken down into measures of central tendency and measures of variability (spread). Measures of central tendency include the [mean](https://www.investopedia.com/terms/m/mean.asp), [median](https://www.investopedia.com/terms/m/median.asp), and [mode](https://www.investopedia.com/terms/m/mode.asp), while measures of variability include [standard deviation](https://www.investopedia.com/terms/s/standarddeviation.asp), [variance](https://www.investopedia.com/terms/v/variance.asp), minimum and maximum variables, [kurtosis](https://www.investopedia.com/terms/k/kurtosis.asp), and [skewness](https://www.investopedia.com/terms/s/skewness.asp).

### **Key Takeaways**

* Descriptive statistics summarizes or describes the characteristics of a data set.
* Descriptive statistics consists of three basic categories of measures: measures of central tendency, measures of variability (or spread), and frequency distribution.
* Measures of central tendency describe the center of the data set (mean, median, mode).
* Measures of variability describe the dispersion of the data set (variance, standard deviation).
* Measures of frequency distribution describe the occurrence of data within the data set (count).

[Descriptive statistics resource](https://www.investopedia.com/terms/d/descriptive_statistics.asp)

## **Variability**

**Variability** describes how far apart data points lie from each other and from the center of a distribution. Along with measures of [central tendency](https://www.scribbr.com/statistics/central-tendency/), measures of variability give you [descriptive statistics](https://www.scribbr.com/statistics/descriptive-statistics/) that summarize your data.

Variability is also referred to as spread, scatter or dispersion. It is most commonly measured with the following:

* [**Range**](https://www.scribbr.com/statistics/range/)**:** the difference between the highest and lowest values
* [**Interquartile range**](https://www.scribbr.com/statistics/interquartile-range/)**:** the range of the middle half of a distribution
* [**Standard deviation**](https://www.scribbr.com/statistics/standard-deviation/)**:** average distance from the mean
* [**Variance**](https://www.scribbr.com/statistics/variance/)**:** average of squared distances from the mean

## Why does variability matter?

While the [central tendency](https://www.scribbr.com/statistics/central-tendency/), or average, tells you where most of your points lie, variability summarizes how far apart they are. This is important because the amount of variability determines how well you can [generalize](https://www.scribbr.com/research-bias/generalizability/) results from the sample to your population.

Low variability is ideal because it means that you can better predict information about the[population](https://www.scribbr.com/methodology/population-vs-sample/) based on sample data. High variability means that the values are less consistent, so it’s harder to make predictions.

Data sets can have the same central tendency but different levels of variability or [vice versa](https://www.scribbr.com/definitions/vice-versa/). If you know only the central tendency or the variability, you can’t say anything about the other aspect. Both of them together give you a complete picture of your data.

[Variability resource](https://www.scribbr.com/statistics/variability/)

## **Different types of correlation**

Correlation is a statistical calculation that indicates that two variables are parallelly related (which means that the variables change together at a constant rate). It is a simple and popularly used tool for defining relationships without delivering a statement concerning the cause and effect.

In simple words, correlation is a statistical calculation that estimates the point at which the two variables shift in relation to each other.

A positive and perfect correlation indicates that the coefficient correlation is exactly one. It indicates that when one variable moves upward or downward, the another variable moves in the same direction.

However, a negative and perfect correlation indicates that both the variables move in the opposite directions. When there is a zero correlation, it means that there is no relationship at all.

## Meaning of Linear Correlation

Linear correlation is referred to as the measure of relationship between two random variables with values ranging from -1 and 1. It is proportional to covariance and can be interpreted in the same way as covariance.

Linear correlation is also said to be based on a straight-line relationship between two random variables.

## –Meaning of Curvilinear Correlation

Non-linear or curvilinear correlation is said to occur when the ratio of change between two variables is not constant. It can happen that as the value of one variable increases, the value of another variable also increases. This will happen till a certain point, after which the increase in value of one variable will result in the decrease in value of the other variable.

The graphical representation of a curvilinear correlation is like an inverted U.

## Linear and Curvilinear Correlation

|  |  |
| --- | --- |
| **Positive Correlation** | **Negative Correlation** |
| Two variables are said to have a positive correlation when they move in the same direction i.e. change occurs in them in the same direction. In positive correlation, the increase in value of one variable results in an increase in the value of another variable.  Example,   1. The area under cultivation and agricultural production 2. The use of manure and the increase in output 3. The expenditure on advertisement and the increase in sales | Two variables are said to have a negative correlation when they move in the opposite direction i.e. change occurs in them in the opposite direction. In negative correlation, the increase in value of one variable results in decrease in the value of another variable.  Example,   1. The price of onions and the  demand of onions 2. The production of vegetables and the prices of vegetables 3. The time spent on video games and the marks in exams |

## Meaning of Linear Correlation

Linear correlation is referred to as the measure of relationship between two random variables with values ranging from -1 and 1. It is proportional to covariance and can be interpreted in the same way as covariance.

Linear correlation is also said to be based on a straight-line relationship between two random variables.

## Meaning of Curvilinear Correlation

Non-linear or curvilinear correlation is said to occur when the ratio of change between two variables is not constant. It can happen that as the value of one variable increases, the value of another variable also increases. This will happen till a certain point, after which the increase in value of one variable will result in the decrease in value of the other variable.

The graphical representation of a curvilinear correlation is like an inverted U.

## Linear and Curvilinear Correlation

|  |  |
| --- | --- |
| **Positive Correlation** | **Negative Correlation** |
| Two variables are said to have a positive correlation when they move in the same direction i.e. change occurs in them in the same direction. In positive correlation, the increase in value of one variable results in an increase in the value of another variable.  Example,   1. The area under cultivation and agricultural production 2. The use of manure and the increase in output 3. The expenditure on advertisement and the increase in sales | Two variables are said to have a negative correlation when they move in the opposite direction i.e. change occurs in them in the opposite direction. In negative correlation, the increase in value of one variable results in decrease in the value of another variable.  Example,   1. The price of onions and the  demand of onions 2. The production of vegetables and the prices of vegetables 3. The time spent on video games and the marks in exams |

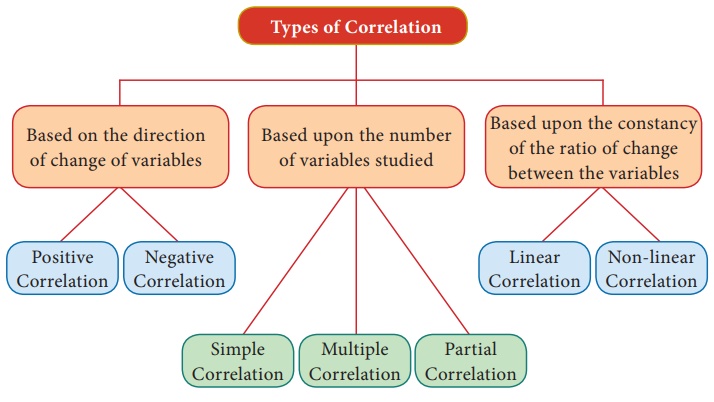
|  |  |
| --- | --- |
| **Linear Correlation** | **Curvilinear Correlation** |
| There exists a linear correlation if the ratio of change in the two variables is constant.  ●     If we plot these coordinates on a graph, we will get a straight line.  Linear Correlation | There exists a curvilinear correlation if the change in the variables is not constant.  ●     If we plot these coordinates on a graph, we will get a curve.  Curvilinear Correlation |

|  |
| --- |
| **Simple, Partial, and Multiple Correlations** |
| **Simple correlation:**When we consider only two variables and check the correlation between them, it is termed as simple correlation. For example, the radius and circumference of a circle.  **Multiple correlation:**When we consider three or more variables for correlation, it is termed as multiple correlation. For example, the price of cola drink, temperature, income, and the demand for cola.  **Partial correlation:**When one or more variables are kept constant and the relationship is studied between the others, it is termed as partial correlation. For example, if we keep the price of cola constant and check the correlation between temperature and the demand for cola, it is a partial correlation. |

|  |
| --- |
| **Simple, Partial, and Multiple Correlations** |
| **Simple correlation:**When we consider only two variables and check the correlation between them, it is termed as simple correlation. For example, the radius and circumference of a circle.  **Multiple correlation:**When we consider three or more variables for correlation, it is termed as multiple correlation. For example, the price of cola drink, temperature, income, and the demand for cola.  **Partial correlation:**When one or more variables are kept constant and the relationship is studied between the others, it is termed as partial correlation. For example, if we keep the price of cola constant and check the correlation between temperature and the demand for cola, it is a partial correlation. |

## Types of Correlation

Correlation is classified in   several different ways. Three of the most important ways of classifying correlation are



## Type I: Based on the direction of change of variables

Correlation is classified into two types as Positive correlation and Negative Correlation based on the direction of change of the variables.

**i. Positive Correlation:**

The correlation is said to be positive if the values of two variables move in the same direction.

**Ex 1:**If income and Expenditure ofa Household may be increasing or decreasing simultaneously. If so, there is positive correlation. Ex. Y= a + bx

**ii. Negative Correlation:**

The Correlation is said to be negative when the values of variables move in the opposite directions. Ex. Y= a – bx

**Ex 1:**Price and demand for a commoditymove in the opposite direction.

## Type II: Basedupon the number of variables studied.

There are three types based upon the number of variables studied as

i) Simple Correlation

ii) Multiple Correlation

iii) Partial Correlation

**i) Simple Correlation:**

If only two variables are taken for study then it is said to be simple correlation. Ex. Y= a + bx

**ii) Multiple Correlations:**

If three or more than three variables are studied simultaneously, then it is termed as multiple correlation.

**Ex:**Determinants of Quantity demanded

Qd= f (P, Pc, Ps, t, y)

Where Qd stands for Quantity demanded, f stands for function.

P is the price of the goods,

Pc is the price of competitive goods

Ps is the price of substituting goods

t is the taste and preference

y is the income.

**iii) Partial Correlation:**

If there are more than two variables but only two variables are considered keeping the other variables constant, then the correlation is said to be Partial Correlation

## Type III: Based upon the constancy of the ratio of change between the variables

Correlation is divided into two types as linear correlation and Non-Linear correlation based upon the Constancy of the ratio of change between the variables.

**i) Linear Correlation:**Correlation is said tobe linear when the amount of change in one variable tends to bear a constant ratio to the amount of change in the other.

Ex. Y= a + bx

**ii) Non Linear:**The correlation would benon -linear if the amount of change in one variable does not bear a constant ratio to the amount of change in the other variables.

Ex. Y= a + bx2

[Types of correlation](https://www.brainkart.com/article/Correlation_37173/)

## **Different type if covariance**

# Covariance

Covariance is a statistical relationship between two random variables, showing how they change relative to each other with time. Accordingly, covariance can be of two types: positive and negative.

They are used to analyze stock market trends, detect patterns in machines, and quantify income and expenditure.

## Types

### Positive Covariance

When two variables have positive covariance, they move in the same direction. This means if one variable increases, the other variable also increases, and vice versa.

Let us consider the relationship between weather temperature and ice cream sales. As the temperature rises, ice cream sales also increase, showing a positive covariance.

### Negative Covariance

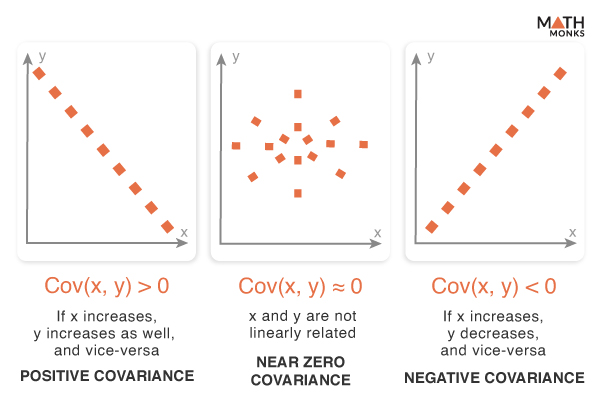
When two variables have negative covariance, they move in opposite directions. This means if one variable increases, the other tends to decrease, and vice versa.

For example, we can imagine the relationship between the speed of a car and the time taken to reach a destination. As the speed increases, the travel time decreases, showing a negative covariance.

### Zero Covariance

A covariance close to zero means the variables are not linearly related.

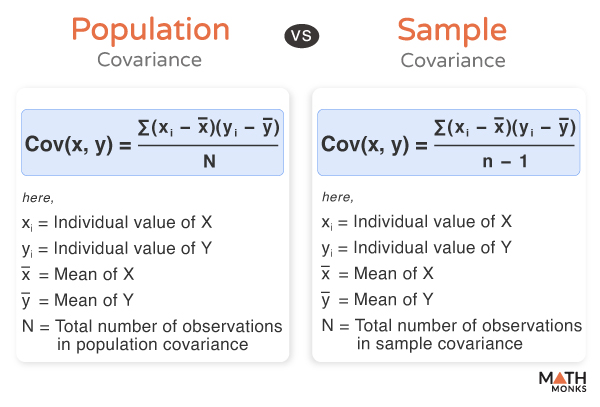
Now, let us observe the number of books read by a student and their shoe size. There is likely no linear relationship between these variables, resulting in a covariance close to zero.

[](https://mathmonks.com/wp-content/uploads/2025/01/Covariance-Types.jpg)

## Formulas

Covariance is differentiated into population covariance and sample covariance based on whether we are considering the entire dataset or a subset of the entire population.

If X and Y are two random variables, their covariance can be calculated using the following formulas:

[](https://mathmonks.com/wp-content/uploads/2025/01/Covariance.jpg)

### Population Covariance

Cov(x, y) = ∑(xi−x―)(yi−y―)N

Here,

* xi = Individual value of X
* yi = Individual value of Y
* x― = Mean of X
* y― = Mean of Y
* N = Total number of observations in population covariance

### Sample Covariance

Cov(x, y) = ∑(xi−x―)(yi−y―)n−1

Here,

* xi = Individual value of X
* yi = Individual value of Y
* x― = Mean of X
* y― = Mean of Y
* n = Total number of observations in sample covariance

[Types of covariance](https://mathmonks.com/covariance#Population_Covariance)

## **Probability Distribution**

In Statistics, the **probability distribution** gives the possibility of each outcome of a random experiment or event. It provides the probabilities of different possible occurrences. Also read, [events in probability](https://byjus.com/maths/types-of-events-in-probability/), here.

To recall, the **probability is a measure of uncertainty of various phenomena**. Like, if you throw a dice, the possible outcomes of it, is defined by the probability. This distribution could be defined with any random experiments, whose outcome is not sure or could not be predicted. Let us discuss now its definition, function, formula and its types here, along with how to create a table of probability based on random variables.

|  |
| --- |
| **Table of Contents:**   * [Definition](https://byjus.com/#definition) * [Probability distribution of random variables](https://byjus.com/#Probability%20distribution%20of%20random%20variables) * [Formulas](https://byjus.com/#Formulas) * [Types](https://byjus.com/#types)   + [Continuous Probability Distribution](https://byjus.com/#continuous-probability-distribution)   + [Discrete Probability Distribution](https://byjus.com/#discrete-probability-distribution) * [Negative Binomial Distribution](https://byjus.com/#negative-binomial-distribution) * [Poisson Probability Distribution](https://byjus.com/#poisson-probability-distribution) * [Probability Distribution Function](https://byjus.com/#probability-distribution-function) * [Probability Distribution Table](https://byjus.com/#probability-distribution-table) * [Prior Probability](https://byjus.com/#prior-probability) * [Posterior Probability](https://byjus.com/#posterior-probability) * [Probability Distribution Examples](https://byjus.com/#probability-distribution-examples) |

## What is Probability Distribution?

Probability distribution yields the possible outcomes for any random event. It is also defined based on the underlying sample space as a set of possible outcomes of any random experiment. These settings could be a set of real numbers or a set of vectors or a set of any entities. It is a part of probability and statistics.

Random experiments are defined as the result of an experiment, whose outcome cannot be predicted. Suppose, if we toss a coin, we cannot predict, what outcome it will appear either it will come as Head or as Tail. The possible result of a random experiment is called an outcome. And the set of outcomes is called a sample point. With the help of these experiments or events, we can always create a probability pattern table in terms of variables and probabilities.

## Probability Distribution of Random Variables

A random variable has a probability distribution, which defines the probability of its unknown values. Random variables can be discrete (not constant) or continuous or both. That means it takes any of a designated finite or countable list of values, provided with a probability mass function feature of the random variable’s probability distribution or can take any numerical value in an interval or set of intervals. Through a probability density function that is representative of the random variable’s probability distribution or it can be a combination of both discrete and continuous.

Two random variables with equal probability distribution can yet vary with respect to their relationships with other random variables or whether they are independent of these. The recognition of a random variable, which means, the outcomes of randomly choosing values as per the variable’s probability distribution function, are called **random variates.**

## Probability Distribution Formulas

|  |  |
| --- | --- |
| Binomial Distribution | P(X) = nCxaxbn-x  Where a = probability of success  b=probability of failure  n= number of trials  x=random variable denoting success |
| Cumulative Distribution Function | FX(x)=∫−∞xfX(t)dt |
| Discrete Probability Distribution | P(x)=n!r!(n−r)!⋅pr(1−p)n−rP(x)=C(n,r)⋅pr(1−p)n−r |

## Types of Probability Distribution

There are two types of probability distribution which are used for different purposes and various types of the data generation process.

* + 1. Normal or Cumulative Probability Distribution
    2. Binomial or Discrete Probability Distribution

[probability distribution](https://byjus.com/maths/probability-distribution/)

## **Regression**

## Correlation Analysis

Correlation analysis is applied in quantifying the association between two continuous variables, for example, an dependent and independent variable or among two independent variables.

## Regression Analysis

Regression analysis refers to assessing the relationship between the outcome variable and one or more variables. The outcome variable is known as the dependent or response variable and the risk elements, and co-founders are known as predictors or independent variables. The dependent variable is shown by “y” and independent variables are shown by “x” in regression analysis.

The sample of a correlation coefficient is estimated in the correlation analysis. It ranges between -1 and +1, denoted by r and quantifies the strength and direction of the linear association among two variables. The correlation among two variables can either be positive, i.e. a higher level of one variable is related to a higher level of another or negative, i.e. a higher level of one variable is related to a lower level of the other.

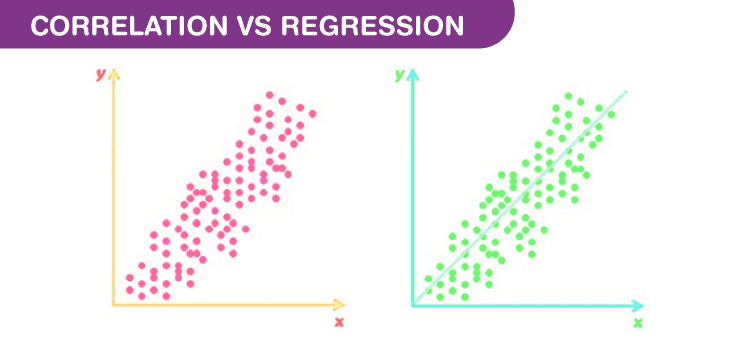
The sign of the coefficient of correlation shows the direction of the association. The magnitude of the coefficient shows the strength of the association.

For example, a correlation of r = 0.8 indicates a positive and strong association among two variables, while a correlation of r = -0.3 shows a negative and weak association. A correlation near to zero shows the non-existence of linear association among two continuous variables.

## Linear Regression

**Linear regression** is a linear approach to modelling the relationship between the scalar components and one or more independent variables. If the regression has one independent variable, then it is known as a simple linear regression. If it has more than one independent variable, then it is known as multiple linear regression. Linear regression only focuses on the[conditional probability](https://byjus.com/maths/conditional-probability-and-conditional-probability-examples/) distribution of the given values rather than the joint probability distribution. In general, all the real world regressions models involve multiple predictors. So, the term linear regression often describes multivariate linear regression.

### **Correlation and Regression Differences**



There are some differences between Correlation and regression.

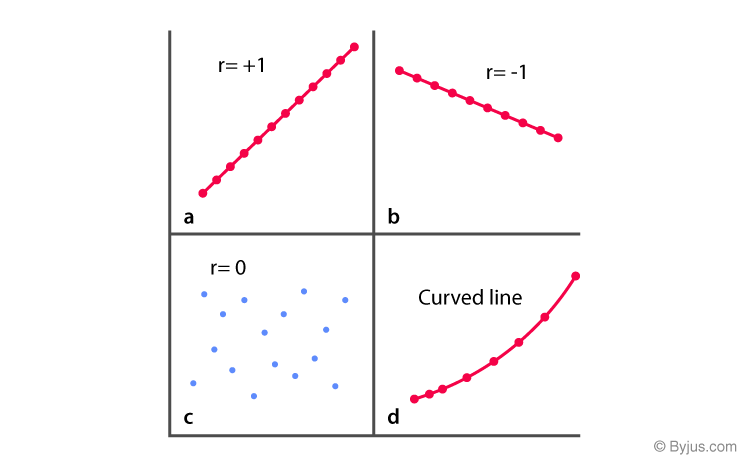
* Correlation shows the quantity of the degree to which two variables are associated. It does not fix a line through the data points. You compute a correlation that shows how much one variable changes when the other remains constant. When r is 0.0, the relationship does not exist. When r is positive, one variable goes high as the other goes up. When r is negative, one variable goes high as the other goes down.
* Linear regression finds the best line that predicts y from x, but Correlation does not fit a line.
* Correlation is used when you measure both variables, while linear regression is mostly applied when x is a variable that is manipulated.

#### **Comparison Between Correlation and Regression**

|  |  |  |
| --- | --- | --- |
| **Basis** | **Correlation** | **Regression** |
| Meaning | A statistical measure that defines co-relationship or association of two variables. | Describes how an independent variable is associated with the dependent variable. |
| Dependent and Independent variables | No difference | Both variables are different. |
| Usage | To describe a linear relationship between two variables. | To fit the best line and estimate one variable based on another variable. |
| Objective | To find a value expressing the relationship between variables. | To estimate values of a random variable based on the values of a fixed variable. |

### **Correlation and Regression Statistics**

The degree of association is measured by “r” after its originator and a measure of linear association. Other complicated measures are used if a curved line is needed to represent the relationship.



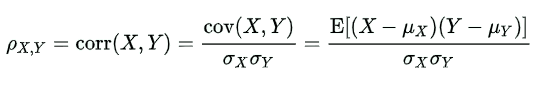
The above graph represents the correlation.

The coefficient of correlation is measured on a scale that varies from +1 to -1 through 0. The complete correlation among two variables is represented by either +1 or -1. The correlation is positive when one variable increases and so does the other; while it is negative when one decreases as the other increases. The absence of correlation is described by 0.

### **Correlation Coefficient Formula**

Let X and Y be the two random variables.

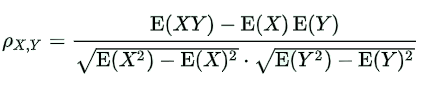
The population correlation coefficient for X and Y is given by the formula:



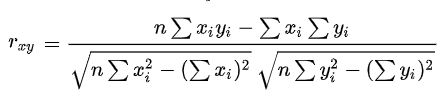
Where,

ρXY = Population correlation coefficient between X and Y  
μX = Mean of the variable X  
μY = Mean of the variable Y  
σX = Standard deviation of X  
σY = Standard deviation of Y  
E = Expected value operator  
Cov = Covariance

The above formulas can also be written as:



The sample correlation coefficient formula is:



The above formulas are used to find the correlation coefficient for the given data. Based on the value obtained through these formulas, we can determine how strong is the association between two variables.

**Simple Linear Regression Equation**

As we know, linear regression is used to model the relationship between two variables. Thus, a simple linear regression equation can be written as:  
Y = a + bX  
Where,

Y = Dependent variable

X = Independent variable

a = [(∑y)(∑x2) – (∑x)(∑xy)]/ [n(∑x2) – (∑x)2]

b = [n(∑xy) – (∑x)(∑y)]/ [n(∑x2) – (∑x)2]

### **Regression Coefficient**

In the linear regression line, the equation is given by:

Y = b0 + b1X

Here b0 is a constant and b1 is the regression coefficient.

The formula for the regression coefficient is given below.

b1 = ∑[(xi – x)(yi – y)]/ ∑[(xi – x)2]

The observed data sets are given by xi and yi. x and y are the mean value of the respective variables.

We know that there are two regression equations and two coefficients of regression.

The regression coefficient of y and x formula is:

byx = r(σy/σx)

The regression coefficient of x on y formula is:

bxy = r(σx/σy)

Where,

σx = Standard deviation of x

σy = Standard deviation of y

Some of the properties of a regression coefficient are listed below:

* The regression coefficient is denoted by b.
* The regression coefficient of y on x can be represented as byx. The regression coefficient of x on y can be represented as bxy. If one of these regression coefficients is greater than 1, then the other will be less than 1.
* They are not independent of the change of scale. They will change in the regression coefficient if x and y are multiplied by any constant.
* The arithmetic mean of both regression coefficients is greater than or equal to the coefficient of correlation.
* The geometric mean between the two regression coefficients is equal to the correlation coefficient.

If bxy is positive, then byx is also positive and vice versa.

[regression](https://byjus.com/maths/correlation-and-regression/)

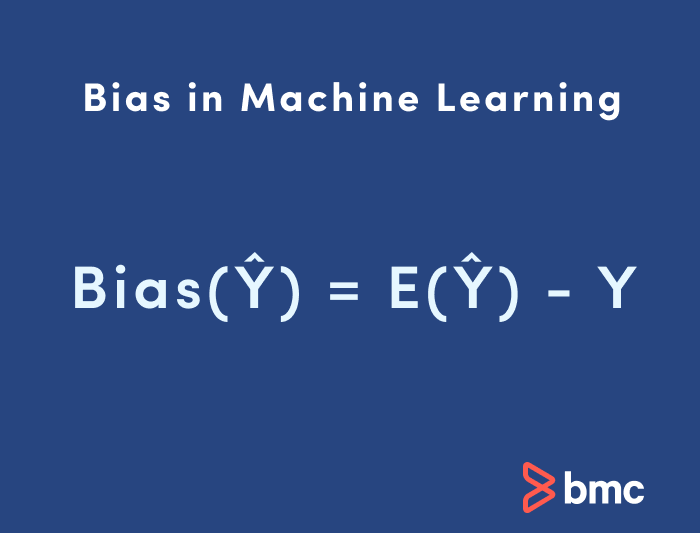
## **Bias / variance tradeoff**

[Bias-variance](https://medium.com/@sarita_68521/understanding-the-bias-variance-tradeoff-in-machine-learning-examples-and-solutions-5de459ddeabd)

Bias and variance are two sources of error in predictive models. Getting the right balance between the bias and variance tradeoff is fundamental to effective machine learning algorithms. Here is a quick explanation of these concepts:

* [**Bias.**](https://www.bmc.com/blogs/bias-variance-machine-learning/#bias-in-machine-learning) Bias refers to error caused by a model for solving complex problems that is over simplified, makes significant assumptions, and misses important relationships in your data.
* [**Variance.**](https://www.bmc.com/blogs/bias-variance-machine-learning/#variance-in-machine-learning) Variance is an error caused by an algorithm that is too sensitive to fluctuations in data, creating an overly complex model that sees patterns in data that are actually just randomness.
* [**Bias–variance tradeoff.**](https://www.bmc.com/blogs/bias-variance-machine-learning/#bias-variance-tradeoff) Minimizing errors caused by oversimplification and excessive complication requires finding the right balance or tradeoff between the two.

## **What is bias in machine learning?**



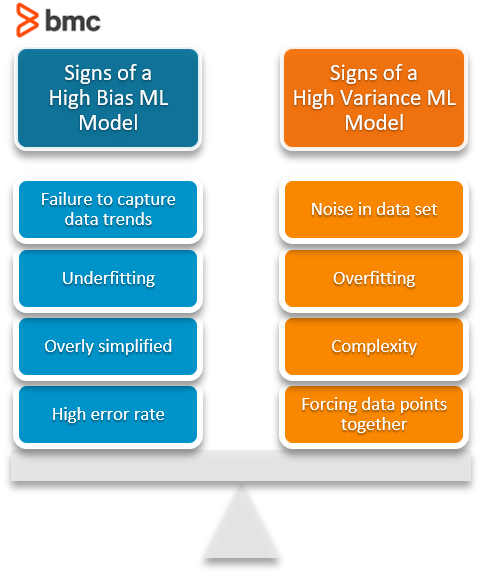
Bias in ML is sometimes called the “too simple” problem. Bias is considered a systematic error that occurs in the machine learning model itself due to incorrect assumptions in the ML process.

Technically, we can define bias as the error between average model prediction and the ground truth. Moreover, it describes how well the model matches the training data set:

* **High bias.**A model with a higher bias would not match the data set closely.
* **Low bias.** A low bias model will closely match the training data set.

Characteristics of a high bias model include:

* Failure to capture proper data trends
* Potential towards underfitting
* More generalized/overly simplified
* High error rate



[Bias\_variance tradeoff](https://www.bmc.com/blogs/bias-variance-machine-learning/)

## **Hypothesis testing**

The process of hypothesis testing is to draw inferences or some conclusion about the overall population or data by conducting some statistical tests on a sample. The same inferences are drawn for different machine learning models through T-test which I will discuss in this tutorial.

For drawing some inferences, we have to make some assumptions that lead to two terms that are used in the hypothesis testing.

* Null hypothesis: It is regarding the assumption that there is no anomaly pattern or believing according to the assumption made.
* Alternate hypothesis: Contrary to the null hypothesis, it shows that observation is the result of real effect.

## P value

It can also be said as evidence or level of significance for the null hypothesis or in machine learning algorithms. It’s the significance of the predictors towards the target.

Generally, we select the level of significance by 5 %, but it is also a topic of discussion for some cases. If you have a strong prior knowledge about your data functionality, you can decide the level of significance.

On the contrary of that if the p-value is less than 0.05 in a machine learning model against an independent variable, then the variable is considered which means there is heterogeneous behavior with the target which is useful and can be learned by the machine learning algorithms.

The steps involved in the hypothesis testing are as follow:

* Assume a null hypothesis, usually in machine learning algorithms we consider that there is no anomaly between the target and independent variable.
* Collect a sample
* Calculate test statistics
* Decide either to accept or reject the null hypothesis

[Hypothesis testing](https://www.datacamp.com/tutorial/hypothesis-testing-machine-learning)

# **Central Tendency Measures**

**Vector calculas**

[Vector calculas](https://medium.com/intuition/vector-calculus-for-machine-learning-5d8842d8cd23)

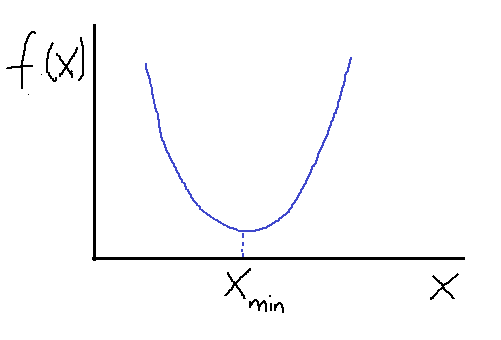
# Calculus in Data Science and Machine Learning

A machine learning algorithm (such as classification, clustering or regression) uses a training dataset to determine weight factors that can be applied to unseen data for predictive purposes. *Behind every machine learning model is an optimization algorithm that relies heavily on calculus*. In this article, we discuss one such optimization algorithm, namely, the Gradient Descent Approximation (GDA) and we’ll show how it can be used to build a simple linear regression estimator.

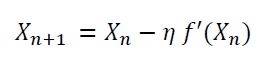
# Optimization Using the Gradient Descent Algorithm

## Derivatives and Gradients

In one-dimension, we can find the maximum and minimum of a function using derivatives. Let us consider a simple quadratic function*f(x)* as shown below.

  
**Figure 1**. Minimum of a simple function using gradient descent algorithm. Image by Author.

Suppose we want to find the minimum of the function *f(x)*. Using the gradient descent method with some initial guess, *X* gets updated according to this equation:



where the constant *eta*is a small positive constant called the learning rate. Note the following:

* when X\_n > X\_min, f’(X\_n) > 0: this ensures that X\_n+1 is less than X\_n. Hence, we are taking steps in the left direction to get to the minimum.
* when X\_n < X\_min, f’(X\_n) < 0: this ensures that X\_n+1 is greater than X\_n. Hence, we are taking steps in the right direction to get to X\_min.

The above observation shows that it doesn’t matter what the initial guess is, the gradient descent algorithm will always find the minimum. How many optimization steps it’s going to take to get to *X\_min* depends on how good the initial guess is. Sometimes if the initial guess or the learning rate is not carefully chosen, the algorithm can completely miss the minimum. This is often referred to as an “**overshoot**”. Generally, one could ensure convergence by adding a convergence criterion such as:

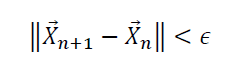
Calculus for Data Science

where *epsilon* is a small positive number.

In higher dimensions, a function of several variables can be optimized (minimized) using the gradient descent algorithm as well. In this case, we use the gradient to update the ***vector X***:

Calculus for Data Science

As in one-dimension, one could ensure convergence by adding a convergence criterion such as:



[Vectorcalculas](https://dev.to/anurag629/calculus-for-data-science-an-introduction-33lm)

**Univariate,Bivariate,Multivariate**

* **Univariate** statistics summarize only **one**[**variable**](https://www.scribbr.com/methodology/types-of-variables/) at a time.
* **Bivariate** statistics compare**two variables**.
* **Multivariate** statistics compare **more than two variables**.

univariate analysis examines individual variables, bivariate analysis explores relationships between two variables, and multivariate analysis investigates the interplay of multiple variables.

Here's a more detailed explanation:

Univariate Analysis:

Focuses on understanding a single variable at a time.

Uses descriptive statistics (e.g., mean, median, mode, standard deviation) to summarize the characteristics of the variable.

Examples: Analyzing the distribution of customer ages, the frequency of different product categories sold, or the average sales per month.

Bivariate Analysis:

Examines the relationship between two variables.

Helps determine if there's a correlation or association between the variables.

Examples: Investigating the relationship between advertising spend and sales, or the correlation between education level and income.

Multivariate Analysis:

Analyzes the relationships and interactions among three or more variables.

Used to understand complex patterns and dependencies within a dataset.

Examples: Predicting customer churn based on multiple factors (e.g., demographics, purchase history, website activity) or identifying clusters of similar customers based on several characteristics.

[uni,bi,multi](https://medium.com/analytics-vidhya/univariate-bivariate-and-multivariate-analysis-8b4fc3d8202c)

**Gradients**

In machine learning, a gradient is a vector that indicates the direction and magnitude of the steepest increase of a function, often used in optimization algorithms like gradient descent to minimize a loss function.

Here's a more detailed explanation:

What is a Gradient?

* **Direction of Steepest Ascent:** A gradient points in the direction where the function's value increases most rapidly.
* **Vector:** It's a vector, meaning it has both direction and magnitude (or slope).
* **Partial Derivatives:** In the context of machine learning, the gradient is calculated using partial derivatives of the loss function with respect to the model's parameters (like weights and biases).
* **Slope:** Think of it as the slope of a function at a particular point.

How Gradients are Used in Machine Learning

* **Gradient Descent:**

The most common application of gradients is in the gradient descent algorithm, a fundamental optimization technique used to train machine learning models.

* **Minimizing Loss:**

Gradient descent iteratively adjusts the model's parameters in the direction opposite to the gradient (the direction of steepest increase) to minimize the loss function.

* **Learning:**

By iteratively adjusting the parameters based on the gradient, the model learns to make better predictions.

* **Learning Rate:**

The "learning rate" determines the size of the steps taken in the direction opposite to the gradient during each iteration.

* **Cost Function:**

The cost function measures the accuracy of the model's predictions.

* **Iterative Process:**

The gradient descent algorithm works iteratively, calculating the gradient, updating the parameters, and repeating until the loss function is minimized.

In Summary

Gradients are crucial in machine learning because they provide the direction and magnitude of the steepest change in a function, enabling algorithms like gradient descent to efficiently find the optimal parameters for a model by minimizing the loss function

## What is Probability?

Probability, a branch of Math that deals with the likelihood of the occurrences of the given event. The probability values for the given experiment is usually defined between the range of numbers. The values lie between the numbers 0 and 1. The probability value cannot be a negative value. The basic rules such as addition, multiplication and complement rules are associated with the probability.

## Experimental Probability Vs Theoretical Probability

There are two approaches to study probability:

* Experimental Probability
* Theoretical Probability

### What is Experimental Probability?

Experimental probability, also known as Empirical probability, is based on actual experiments and adequate recordings of the happening of events. To determine the occurrence of any event, a series of actual experiments are conducted. Experiments which do not have a fixed result are known as random experiments. The outcome of such experiments is uncertain. Random experiments are repeated multiple times to determine their likelihood. An experiment is repeated a fixed number of times and each repetition is known as a trial. Mathematically, the formula for the experimental probability is defined by;

**Probability of an Event P(E) = Number of times an event occurs / Total number of trials.**

### What is Theoretical Probability?

In probability, the [theoretical probability](https://byjus.com/maths/theoretical-probability/) is used to find the probability of an event. Theoretical probability does not require any experiments to conduct. Instead of that, we should know about the situation to find the probability of an event occurring. Mathematically, the theoretical probability is described as the number of favourable outcomes divided by the number of possible outcomes.

**Probability of Event P(E) = No. of. Favourable outcomes/ No. of. Possible outcomes.**

### Experimental Probability Example

**Example:** You asked your 3 friends Shakshi, Shreya and Ravi to toss a fair coin 15 times each in a row and the outcome of this experiment is given as below:

|  |  |  |
| --- | --- | --- |
| **Coin Tossed By:** | **No. of. Heads** | **No. of. Tails** |
| Shakshi | 6 | 9 |
| Shreya | 7 | 8 |
| Ravi | 8 | 7 |

Calculate the probability of occurrence of heads and tails.

Solution: The experimental probability for the occurrence of heads and tails in this experiment can be calculated as:

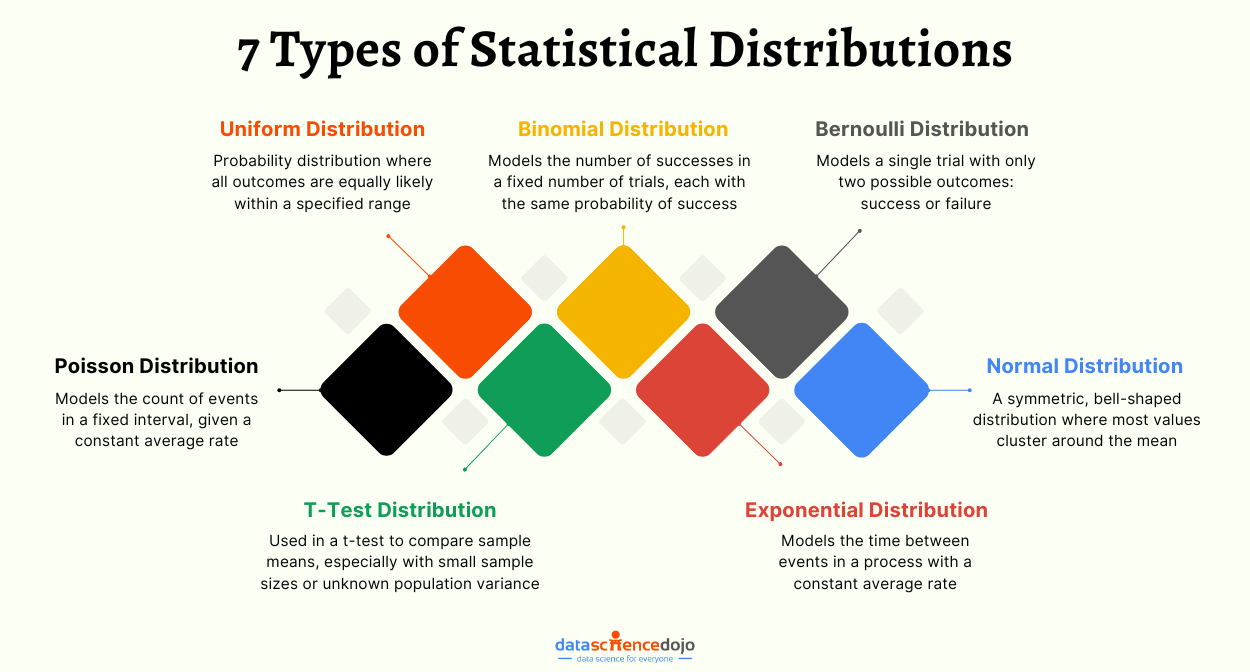
Experimental Probability of Occurrence of heads = Number of times head occurs/Number of times coin is tossed.

Experimental Probability of Occurrence of tails = Number of times tails occurs/Number of times coin is tossed.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Coin Tossed By:** | **No. of. Heads** | **No. of. Tails** | **Experimental Probability for the occurrence of Head** | **Experimental Probability for the occurrence of Tail** |
| Shakshi | 6 | 9 | 6/15 = 0.4 | 9/15 = 0.6 |
| Shreya | 7 | 8 | 7/15 = 0.47 | 8/15 = 0.53 |
| Ravi | 8 | 7 | 8/15 = 0.53 | 7/15 = 0.47 |

[probability](https://byjus.com/maths/experimental-probability/)

**Types of distribution**

Statistical distributions help us understand a problem better by assigning a range of possible values to the variables, making them very useful in [data science](https://datasciencedojo.com/data-science-bootcamp/) and [machine learning](https://datasciencedojo.com/blog/machine-learning-101/). 

[Types of distribution](https://datasciencedojo.com/blog/types-of-statistical-distributions-in-ml/)

**Random variable**

A random variable’s likely values may express the possible outcomes of an experiment, which is about to be performed or the possible outcomes of a preceding experiment whose existing value is unknown. They may also conceptually describe either the results of an “objectively” random process (like rolling a die) or the “subjective” randomness that appears from inadequate knowledge of a quantity.

The domain of a random variable is a sample space, which is represented as the collection of possible outcomes of a random event. For instance, when a coin is tossed, only two possible outcomes are acknowledged such as heads or tails.

**Also, read:**

|  |
| --- |
| Related Links |
| [Probability](https://byjus.com/maths/probability/) | [Probability Mass Function](https://byjus.com/maths/probability-mass-function/) |
| [Probability Density Function](https://byjus.com/maths/probability-density-function/) | [Mean and Variance of Random Variable](https://byjus.com/maths/mean-variance-random-variable/) |

## Random Variable Definition

A random variable is a rule that assigns a numerical value to each outcome in a [sample space](https://byjus.com/maths/sample-space/). Random variables may be either discrete or continuous. A random variable is said to be discrete if it assumes only specified values in an interval. Otherwise, it is continuous. We generally denote the random variables with capital letters such as X and Y. When X takes values 1, 2, 3, …, it is said to have a discrete random variable.

As a function, a random variable is needed to be measured, which allows probabilities to be assigned to a set of potential values. It is obvious that the results depend on some physical variables which are not predictable. Say, when we toss a fair coin, the final result of happening to be heads or tails will depend on the possible physical conditions. We cannot predict which outcome will be noted. Though there are other probabilities like the coin could break or be lost, such consideration is avoided.

## Variate

A variate can be defined as a generalization of the random variable. It has the same properties as that of the random variables without stressing to any particular type of probabilistic experiment. It always obeys a particular probabilistic law.

* A variate is called discrete variate when that variate is not capable of assuming all the values in the provided range.
* If the variate is able to assume all the numerical values provided in the whole range, then it is called continuous variate.

## **Types of Random Variable**

As discussed in the introduction, there are two random variables, such as:

* Discrete Random Variable
* Continuous Random Variable

Let’s understand these types of variables in detail along with suitable examples below.

## Discrete Random Variable

A discrete random variable can take only a finite number of distinct values such as 0, 1, 2, 3, 4, … and so on. The [probability distribution](https://byjus.com/maths/probability-distribution/) of a random variable has a list of probabilities compared with each of its possible values known as probability mass function.

In an analysis, let a person be chosen at random, and the person’s height is demonstrated by a random variable. Logically the random variable is described as a function which relates the person to the person’s height. Now in relation with the random variable, it is a probability distribution that enables the calculation of the probability that the height is in any subset of likely values, such as the likelihood that the height is between 175 and 185 cm, or the possibility that the height is either less than 145 or more than 180 cm. Now another random variable could be the person’s age which could be either between 45 years to 50 years or less than 40 or more than 50.

## Continuous Random Variable

A numerically valued variable is said to be continuous if, in any unit of measurement, whenever it can take on the values a and b. If the random variable X can assume an infinite and uncountable set of values, it is said to be a continuous random variable. When X takes any value in a given interval (a, b), it is said to be a continuous random variable in that interval.

Formally, a continuous random variable is such whose cumulative distribution function is constant throughout. There are no “gaps” in between which would compare to numbers which have a limited probability of occurring. Alternately, these variables almost never take an accurately prescribed value c but there is a positive probability that its value will rest in particular intervals which can be very small.

## Random Variable Formula

For a given set of data the [mean and variance random variable](https://byjus.com/maths/mean-variance-random-variable/) is calculated by the formula. So, here we will define two major formulas:

* Mean of random variable
* Variance of random variable

**Mean of random variable:**If X is the random variable and P is the respective probabilities, the mean of a random variable is defined by:

**Mean (μ) = ∑ XP**

where variable X consists of all possible values and P consist of respective probabilities.

**Variance of Random Variable:**The variance tells how much is the spread of random variable X around the mean value. The formula for the variance of a random variable is given by;

Var(X) = σ2 = E(X2) – [E(X)]2

where E(X2) = ∑X2P and E(X) = ∑ XP

[random variable](https://byjus.com/maths/random-variable/)

**Central Limit Theorem**

## What Is the Central Limit Theorem (CLT)?

In probability theory, the central limit theorem (CLT) states that the [distribution of a sample](https://www.investopedia.com/terms/s/sampling-distribution.asp) will approximate a normal distribution (i.e., a [bell curve](https://www.investopedia.com/terms/b/bell-curve.asp)) as the sample size becomes larger, regardless of the population's actual distribution shape.

Put another way, CLT is a [statistical](https://www.investopedia.com/terms/s/statistics.asp) premise that, given a sufficiently large sample size from a population with a finite level of variance, the mean of all sampled variables from the same population will be approximately equal to the mean of the whole population. Furthermore, these samples will approximate a [normal distribution](https://www.investopedia.com/terms/n/normaldistribution.asp), with their variances being approximately equal to the [variance](https://www.investopedia.com/terms/v/variance.asp) of the population as the sample size gets larger, according to the [law of large numbers](https://www.investopedia.com/terms/l/lawoflargenumbers.asp).

### **Key Takeaways**

* The central limit theorem (CLT) states that the distribution of sample means approximates a normal distribution as the sample size gets larger, regardless of the population's distribution.
* A sufficiently large sample size can predict the characteristics of a population more accurately.
* Sample sizes equal to or greater than 30 are often considered sufficient for the CLT to hold.
* A key aspect of CLT is that the average of the sample means and standard deviations will equal the population mean and standard deviation.
* CLT is useful in finance and investing when analyzing a large collection of securities to estimate portfolio distributions and traits for returns, risk, and correlation.

[CLT](https://www.investopedia.com/terms/c/central_limit_theorem.asp#:~:text=Key%20Takeaways-,The%20central%20limit%20theorem%20(CLT)%20states%20that%20the%20distribution%20of,regardless%20of%20the%20population's%20distribution.)

[CLT2](https://www.scribbr.com/statistics/central-limit-theorem/)

**Sampling**

When you conduct research about a group of people, it’s rarely possible to collect data from every person in that group. Instead, you select a **sample**. The sample is the group of individuals who will actually participate in the research.

To draw valid conclusions from your results, you have to carefully decide how you will select a sample that is representative of the group as a whole. This is called a **sampling method**. There are two primary types of sampling methods that you can use in your research:

* [**Probability sampling**](https://www.scribbr.com/methodology/probability-sampling/) involves random selection, allowing you to make strong statistical inferences about the whole group.
* [**Non-probability sampling**](https://www.scribbr.com/methodology/non-probability-sampling/) involves non-random selection based on convenience or other criteria, allowing you to easily collect data.

[sampling](https://www.scribbr.com/methodology/sampling-methods/#:~:text=Sampling%20means%20selecting%20the%20group,the%20characteristics%20of%20a%20population.)

**Optimization and optimization techniques**

[optimization](https://medium.com/@kumarvrsec/optimization-techniques-in-data-science-a-comprehensive-overview-part-1-dae20acaa3da)