**2019 solved**

## 2.What is the Multiplication Rule of Probability?

According to the multiplication rule of probability, the probability of occurrence of both the events A and B is equal to the product of the probability of B occurring and the conditional probability that event A occurring given that event B occurs.

If A and B are dependent events, then the probability of both events occurring simultaneously is given by:

|  |
| --- |
| **P(A ∩ B) = P(B) . P(A|B)** |

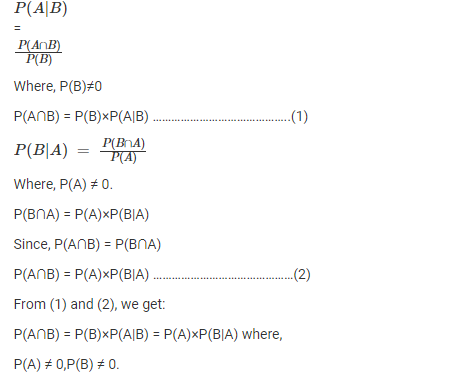
If A and B are two independent events in an experiment, then the probability of both events occurring simultaneously is given by:

|  |
| --- |
| **P(A ∩ B) = P(A) . P(B)** |

## Proof

We know that the conditional probability of event A given that B has occurred is denoted by P(A|B) and is given by

P(A|B)=



The above result is known as the multiplication rule of probability.

For independent events A and B, P(B|A) = P(B). The equation (2) can be modified into,

P(A∩B) = P(B) × P(A)

## Multiplication Theorem of Probability

We have already learned the multiplication rules we follow in probability, such as;

P(A∩B) = P(A)×P(B|A) ; if P(A) ≠ 0

P(A∩B) = P(B)×P(A|B) ; if P(B) ≠ 0

Let us learn here the multiplication theorems for independent events A and B.

If A and B are two independent events for a random experiment, then the probability of simultaneous occurrence of two independent events will be equal to the product of their probabilities. Hence,

P(A∩B) = P(A).P(B)

Now, from multiplication rule we know;

P(A∩B) = P(A)×P(B|A)

Since A and B are independent, therefore;

P(B|A) = P(B)

Therefore, again we get;

P(A∩B) = P(A).P(B)

Hence, proved.

1. **Discuss statistical and mathematical definition of probability**

Probability denotes the possibility of the outcome of any random event. The meaning of this term is to check the extent to which any event is likely to happen. For example, when we flip a coin in the air, what is the possibility of getting a head? The answer to this question is based on the number of possible outcomes. Here the possibility is either head or tail will be the outcome. So, the probability of a head to come as a result is 1/2.

The probability is the measure of the likelihood of an event to happen. It measures the certainty of the event. The formula for probability is given by;

****P(E) = Number of Favourable Outcomes/Number of total outcomes****

**P(E) = n(E)/n(S)**

Here,

n(E) = Number of event favourable to event E

n(S) = Total number of outcomes

1. **What do you understand by measure of central tendency ?**

In statistics, the central tendency is the descriptive summary of a data set. Through the single value from the dataset, it reflects the centre of the data distribution. Moreover, it does not provide information regarding individual data from the dataset, where it gives a summary of the dataset. Generally, the central tendency of a dataset can be defined using some of the measures in statistics.

## Definition

The central tendency is stated as the statistical measure that represents the single value of the entire distribution or a dataset. It aims to provide an accurate description of the entire data in the distribution.

## Measures of Central Tendency

The central tendency of the dataset can be found out using the three important measures namely [mean, median and mode](https://byjus.com/maths/mean-median-mode/).

## IMG_256

## Mean

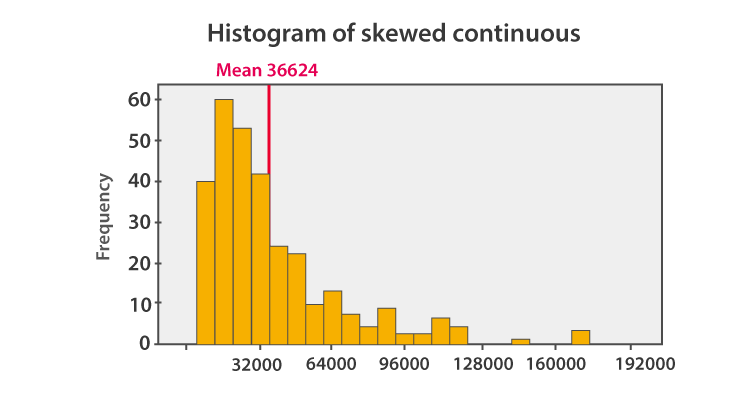
The mean represents the average value of the dataset. It can be calculated as the sum of all the values in the dataset divided by the number of values. In general, it is considered as the arithmetic mean. Some other measures of mean used to find the central tendency are as follows:

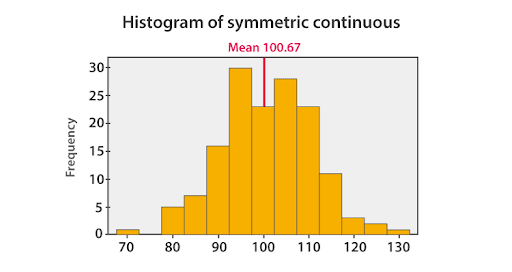
* Geometric Mean
* Harmonic Mean
* Weighted Mean

It is observed that if all the values in the dataset are the same, then all geometric, arithmetic and harmonic mean values are the same. If there is variability in the data, then the mean value differs. Calculating the mean value is completely easy. The formula to calculate the mean value is given by:

Mean=x1+x2+..+xnn

The histogram given below shows that the mean value of symmetric continuous data and the skewed continuous data.



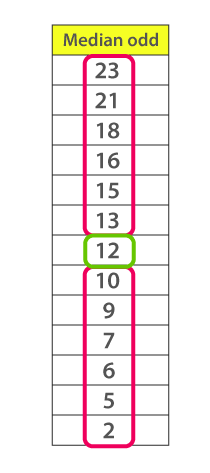


In symmetric data distribution, the mean value is located accurately at the centre. But in the skewed continuous data distribution, the extreme values in the extended tail pull the mean value away from the centre. So it is recommended that the mean can be used for the symmetric distributions.

### Median

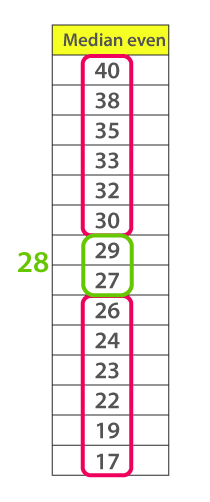
Median is the middle value of the dataset in which the dataset is arranged in the ascending order or in descending order. When the dataset contains an even number of values, then the median value of the dataset can be found by taking the mean of the middle two values.

Consider the given dataset with the odd number of observations arranged in descending order – 23, 21, 18, 16, 15, 13, 12, 10, 9, 7, 6, 5, and 2



Here 12 is the middle or median number that has 6 values above it and 6 values below it.

Now, consider another example with an even number of observations that are arranged in descending order – 40, 38, 35, 33, 32, 30, 29, 27, 26, 24, 23, 22, 19, and 17



When you look at the given dataset, the two middle values obtained are 27 and 29.

Now, find out the mean value for these two numbers.

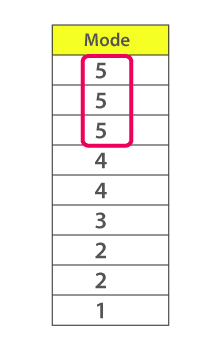
i.e.,(27+29)/2 =28

Therefore, the median for the given data distribution is 28.

### Mode

The mode represents the frequently occurring value in the dataset. Sometimes the dataset may contain multiple modes and in some cases, it does not contain any mode at all.

Consider the given dataset 5, 4, 2, 3, 2, 1, 5, 4, 5



Since the mode represents the most common value. Hence, the most frequently repeated value in the given dataset is 5.

Based on the properties of the data, the measures of central tendency are selected.

* If you have a symmetrical distribution of continuous data, all the three measures of central tendency hold good. But most of the times, the analyst uses the mean because it involves all the values in the distribution or dataset.
* If you have skewed distribution, the best measure of finding the central tendency is the median.
* If you have the original data, then both the median and mode are the best choice of measuring the central tendency.
* If you have categorical data, the mode is the best choice to find the central tendency.

1. **Define geometric mean with properties.**

In mathematics and statistics, the summary that describes the whole data set values can be easily described with the help of measures of central tendencies. The most important measures of central tendencies are mean, median, mode and the range. Among these, the mean of the data set will provide the overall idea of the data. The mean defines the average of numbers. The different types of mean are Arithmetic Mean (AM), Geometric Mean (GM) and [Harmonic Mean](https://byjus.com/maths/harmonic-mean/) (HM). In this article, let us discuss the definition, formula, properties, applications, the relation between AM, GM, and HM with solved examples in detail.

****Table of Contents:****

* [Definition](https://byjus.com/" \l "Definition)
* [Formula](https://byjus.com/" \l "Formula)
* [Difference Between AM and GM](https://byjus.com/" \l "Difference Between AM and GM)
* [Relation Between AM, GM and HM](https://byjus.com/" \l "Relation Between AM, GM and HM)
* [Properties](https://byjus.com/" \l "Properties)
* [Applications](https://byjus.com/" \l "Applications)
* [Examples](https://byjus.com/" \l "Examples)
* [FAQs](https://byjus.com/" \l "FAQs)

## Geometric Mean Definition

In Mathematics, the ****Geometric Mean (GM)**** is the average value or mean which signifies the central tendency of the set of numbers by finding the product of their values. Basically, we multiply the numbers altogether and take the nth root of the multiplied numbers, where n is the total number of data values. For example: for a given set of two numbers such as 3 and 1, the geometric mean is equal to √(3×1) = √3 = 1.732.

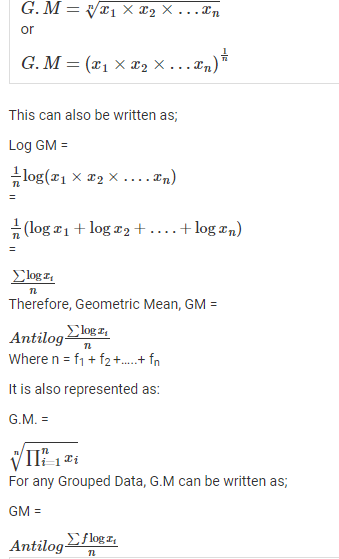
In other words, the geometric mean is defined as the nth root of the product of n numbers. It is noted that the geometric mean is different from the arithmetic mean. Because, in arithmetic mean, we add the data values and then divide it by the total number of values. But in geometric mean, we multiply the given data values and then take the root with the radical index for the total number of data values. For example, if we have two data, take the square root, or if we have three data, then take the cube root, or else if we have four data values, then take the 4th root, and so on. 

## Geometric Mean Formula

The formula to calculate the geometric mean is given below:

The Geometric Mean (G.M) of a series containing n observations is the nth root of the product of the values.

Consider, if x1, x2 …. Xn are the observation, then the G.M is defined as:



1. **How will you calculate median in case of ungrouped data ?**

## How to calculate Median step by step?

You can use the following steps to calculate the median.

**For ungrouped data:**

Step 1. Arrange the given values in the ascending order.

Step 2. Find the number of observations in the given set of data. It is denoted by n.

Step 3. If n is odd, the median equals the [(n+1)/2]th observation.

Step 4. If n is even, then the median is given by the mean of (n/2)th observation and [(n/2)+1]th observation.

## Formula

**For ungrouped data:**

Median = [(n+1)/2]th observation, if n is odd.

Median = mean of (n/2)th observation and [(n/2)+1]th observation, if n is even.

1. **Define correlation coefficient.**

**Correlation Coefficient is a statistical concept, which helps in establishing a relation between predicted and actual values obtained in a statistical experiment. The calculated value of the correlation coefficient explains the exactness between the predicted and actual values.**

**Correlation Coefficient value always lies between -1 to +1. If correlation coefficient value is positive, then there is a similar and identical relation between the two variables. Else it indicates the dissimilarity between the two variables.**

**The covariance of two variables divided by the product of their standard deviations gives Pearson’s correlation coefficient. It is usually represented by ρ (rho).**

**ρ (X,Y) = cov (X,Y) / σX.σY.**

**Here cov is the covariance. σX is the standard deviation of X and σY is the standard deviation of Y. The given equation for correlation coefficient can be expressed in terms of means and expectations.**

**ρ(X,Y)=E(X−μx)(Y−μy)σx.σy**

**μx and μy are mean of x and mean of y respectively. E is the expectation.**

1. **What do you understand by correlation between two variables**

The correlation coefficient is measured on a scale that varies from + 1 through 0 to – 1. Complete correlation between two variables is expressed by either + 1 or -1. When one variable increases as the other increases the correlation is positive; when one decreases as the other increases it is negative.

1. **Define regression coefficient**

## **What are Regression Coefficients?**

Regression coefficients can be defined as estimates of some unknown parameters to describe the relationship between a predictor variable and the corresponding response. In other words, regression coefficients are used to predict the value of an unknown [variable](https://www.cuemath.com/algebra/variables-constants-and-expressions/) using a known variable. Linear regression is used to quantify how a unit change in an independent variable causes an effect in the dependent variable by determining the equation of the best-fitted [straight line](https://www.cuemath.com/geometry/straight-line/). This process is known as regression analysis.

## **Formula for Regression Coefficients**

The goal of linear regression is to find the equation of the straight line that best describes the relationship between two or more variables. For example, suppose a simple regression equation is given by y = 7x - 3, then 7 is the [coefficient](https://www.cuemath.com/algebra/coefficient/), x is the predictor and -3 is the constant term. Suppose the equation of the best-fitted line is given by Y = aX + b then, the regression coefficients formula is given as follows:

a = n(∑xy)−(∑x)(∑y)n(∑x2)−(∑x)2n(∑xy)−(∑x)(∑y)n(∑x2)−(∑x)2

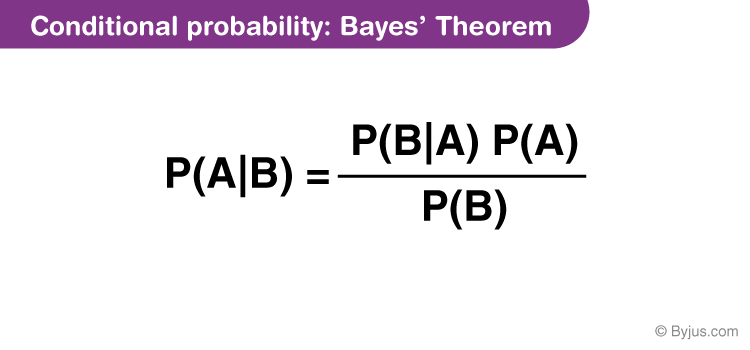
b = (∑y)(∑x2)−(∑x)(∑xy)n(∑x2)−(∑x)2(∑y)(∑x2)−(∑x)(∑xy)n(∑x2)−(∑x)2

here, n refers to the number of data points in the given data sets.

**SECTION C**

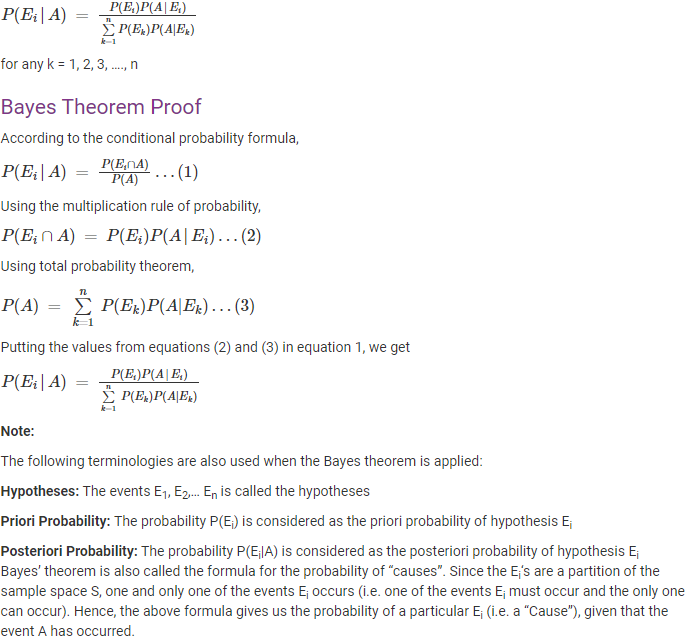
1. **Explain bayes theorem with an example**

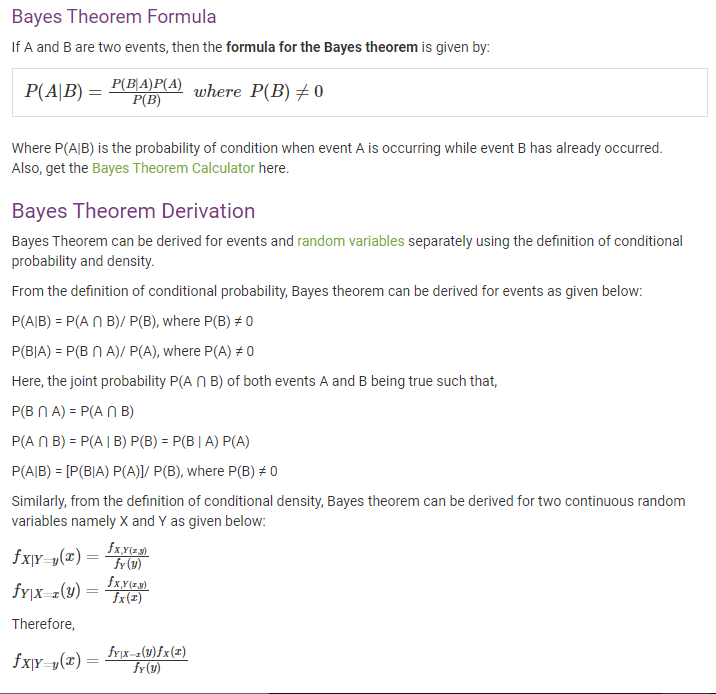
 Bayes theorem is also known as the formula for the probability of “causes”. For example: if we have to calculate the probability of taking a blue ball from the second bag out of three different bags of balls, where each bag contains three different colour balls viz. red, blue, black.

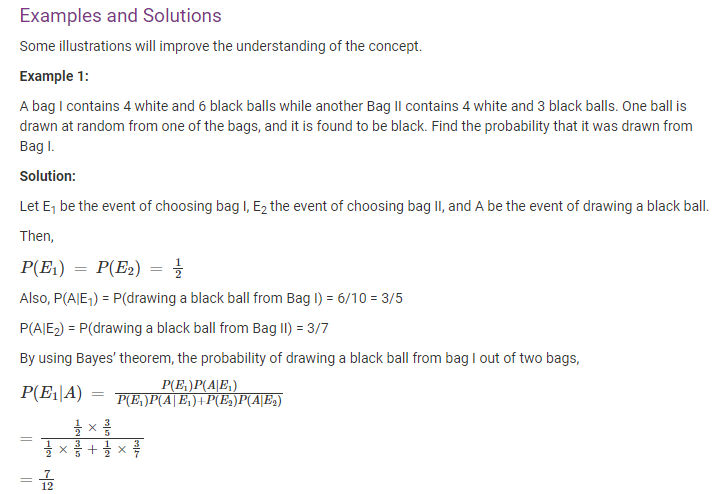


## Bayes Theorem Statement

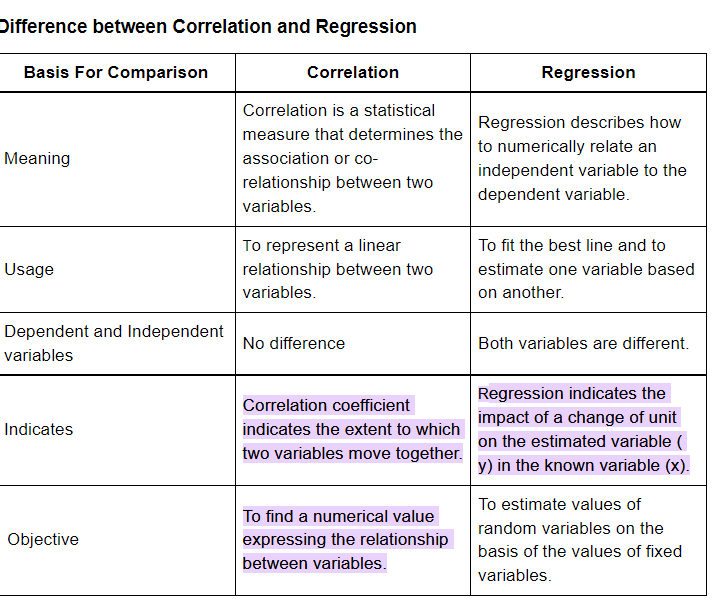
Let E1, E2,…, En be a set of events associated with a sample space S, where all the events E1, E2,…, En have nonzero probability of occurrence and they form a partition of S. Let A be any event associated with S, then according to Bayes theorem,







1. Define regression coefficient and how it differ from correlation.



Defination -

