**[What the hell is "Probability Distributions?" {Part-3} Layman Terms!](https://medium.com/@saikirandasari41/what-the-hell-is-probability-distributions-part-3-layman-terms-816bb29a4182)**

After a successful explanation of my previous 2 parts of statistics blogs, Here I bring Part 3 Which I feel is highly recommended at this stage after Part 2 at this time!

[**Part1**](https://medium.com/@saikirandasari41/statistics-by-saikiran-dasari-for-data-science-f995cebd2fd1), [**Part2**](https://medium.com/@saikirandasari41/how-descriptive-stats-can-be-this-structured-6697b4f22202)

 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Tip for this Blog, how much Maths & Stats do we need to know?

The myth: You need to have very deep knowledge of mathematics and statistics in order to become a Data Scientist or Machine Learning engineer

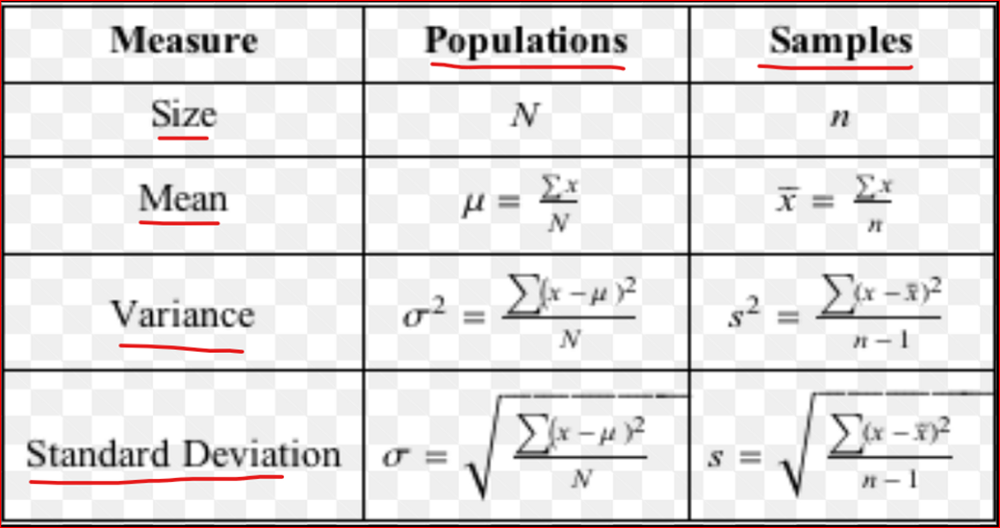
The truth: {I always believe in!}

You need to have a good understanding of statistics and be confident in using the various statistical concepts!! You are not a Statistician or a Mathematician!! Your job is to have an understanding of the concepts and know how to use the tools based on the concepts to achieve results and that's where I come in and BLOG!

 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

**Recap: Note the Formula for Population and Samples below:**

N, µ, σ, σ² | n, x bar, s², s - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -



**Table of Contents:**

What is a Random Variable?

a. Types of Random Variables (Discrete and Continuous)

2. Probability and Probability Distribution Functions

-> What is it? Various Formulas! and a couple of Examples

3. Probability Distributions

-> What is a distribution function with one example?

4. What is Probability Density Function (PDF), Probability Mass Function (PMF), and Cumulative Density Function (CDF)?

5. Discrete Probability Distribution

a. Bernoulli Distribution or Bernoulli Random Variable?

b. Binomial Distribution?

c. Poisson Distribution?

5. Continuous Probability Distribution

a. Uniform Distribution?

b. What is "Not a Uniform Distribution" then?

c. NORMAL Distribution? End to End!

d. Standard Normal Distribution or SND?

e. Student t-Test Distribution?

Without further any adieu let's jump into the Detailed Contents:

 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

**1. What is a Random variable?**

Anything which cannot be predicted certainly is called RV

Eg: Somebody tossing a coin / Rolling a Die. So, before tossing a die / rolling a coin we cannot tell anything. So, there are so many things that cannot be predicted before conducting an experiment!

**A) Explain Discrete Random Variable and Continuous RV (Continuous Distribution)**

Types of Random Variables

Random variables are classified into discrete and continuous variables. The main difference between the two categories is the type of possible values that each variable can take.

**Discrete Random Variable**: A discrete random variable is a (random) variable whose values take only a finite number of values.

Eg: No. of Defective light bulbs in a box, No. of Children in a family

**2. Continuous Random Variable**

: Unlike discrete variables, continuous random variables can take on an infinite number of possible values.

Eg: Time takes to complete a race, the length of time between arrival at the hospital

 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

**2. Probability and Probability Distribution Functions**

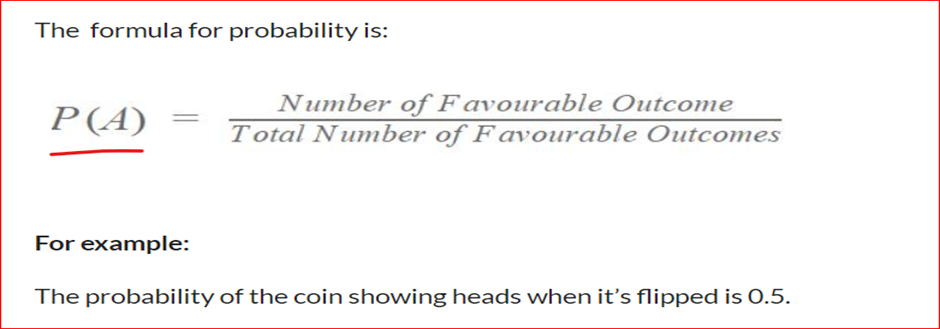
**a) What is Probability?**

Most events are difficult to predict precisely and therefore what we can do is to find the likelihood that the event will or not occur-this is called Probability.

The concept of probability is: It means how likely an event is about to occur or the chance of the occurrence of an event.

Eg: You cannot say that based on my analysis, the company will certainly increase its sales by 25% by the end of the next quarter, anything can happen. But you can safely say there is a higher probability that the company will increase its sales by 25% by the end of the next quarter.

**b) The formulas for probability are:**



**-Standard Formula after I've analysed every material!!-> 1. Classical Model:**

* **1. Classical Model:**

Eg: So, when I flip a coin and I am interested in finding out the probability of getting a head, the number of outcomes in which the event occurred will be 1 and the total number of possible outcomes is it could be either head or it could be a tail.

So, the total number of possible outcomes is two. So, in this case, the probability of getting head will be

= 1 (No. of outcomes in which event occurs) / 2 (Total No. of possible outcomes of an experiment)

* **2. Relative Frequency of Occurrence:**

Eg: Here, in this case, let me take an example of going from my home to my office and my plan is to reach at 6:30 AM in the morning.

Sometimes I reach at 6:25, sometimes I react at 6:30, sometimes I reach at 6:35, and so on.

So, let's say now if I want to find out what is the probability that I reach the office at 6:30 or before that.

So, what I will do is I will list down all the times I reached the office. So, let's say I note down my time of reaching the office for 100 days, and out of those hundred days, I see that 35 days I've reached on or before 6:30.

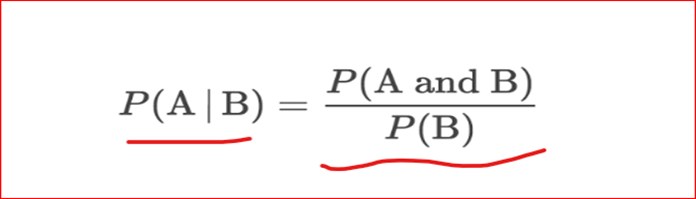
So here, the probability of reaching the office on or before 6:30 will be

= 35(No. of times an event occurred) / 100(Total opportunities for an event to occur)

* **3. Conditional Probability**

Conditional probability is the probability of an event occurring provided another event has already occurred.

The formula of conditional probability



P(A|B): Above denotes means (Probability of A given B is true)

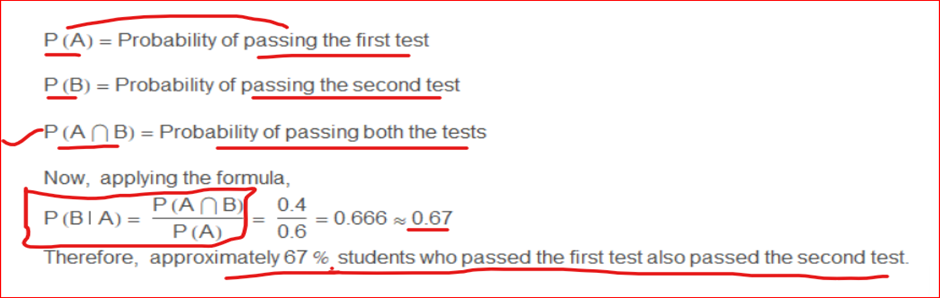
**For example:**

**Given**} The students of a class have been given two tests in the subject of mathematics. In the first test, 60% of the students pass

while only 40% of the students passed both tests.

**Find**} What percentage of students who passed the first test, cleared the second test?

Solving the Example!

 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

**3. Probability Distributions!**

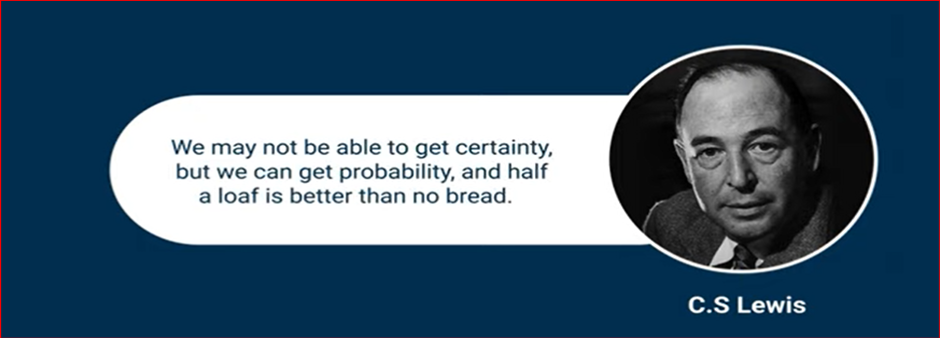
**First of all, let's discuss What is a distribution function?**

A distribution function is a mathematical expression that describes the probability of different possible outcomes for an experiment.

It is denoted as **Variable ~ Type (Characteristics)**

**Eg**: Experiment of tossing a fair coin: If we use X to denote the events, the probability distribution of X would take the value 0.5 for X = heads, and 0.5 for X= tails.

**C.S Lewis (British Writer)**



PROBABILITY is always P <= 100% not more than 100%

PROBABILITY always lies between 0 to 1 {0<=P<=1}

and Total Probability is always =1

 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

**4. What is Probability Density Function (PDF), Probability Mass Function (PMF), and Cumulative Density Function (CDF)?**

Before deep-diving into the types of distributions, it is important to revise the fundamental concepts like Probability Density Function (PDF), Probability Mass Function (PMF), and Cumulative Density Function (CDF).

**a) Probability Density Function (PDF):**

It describes the probability distribution of a continuous random variable. The probability associated with a single value is always Zero.

**b) Probability Mass Function (PMF)**

It is a statistical term that describes the probability distribution of a discrete random variable.

**c) Cumulative Distribution Function (CDF)**

It is another method to describe the distribution of a random variable (either continuous or discrete).

 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

**5. Discrete Probability Distribution we have:**

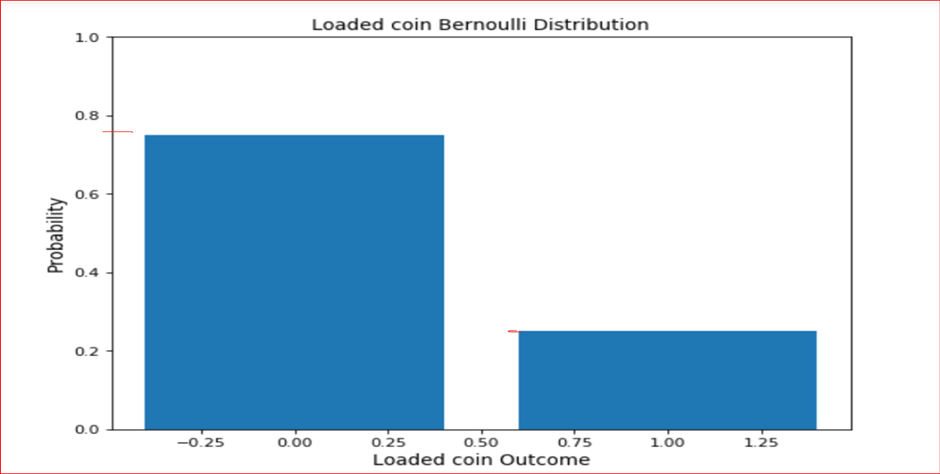
**a) Bernoulli Distribution or Bernoulli Random Variable**

Bernoulli Distribution: Single-Trial with Two Possible Outcomes

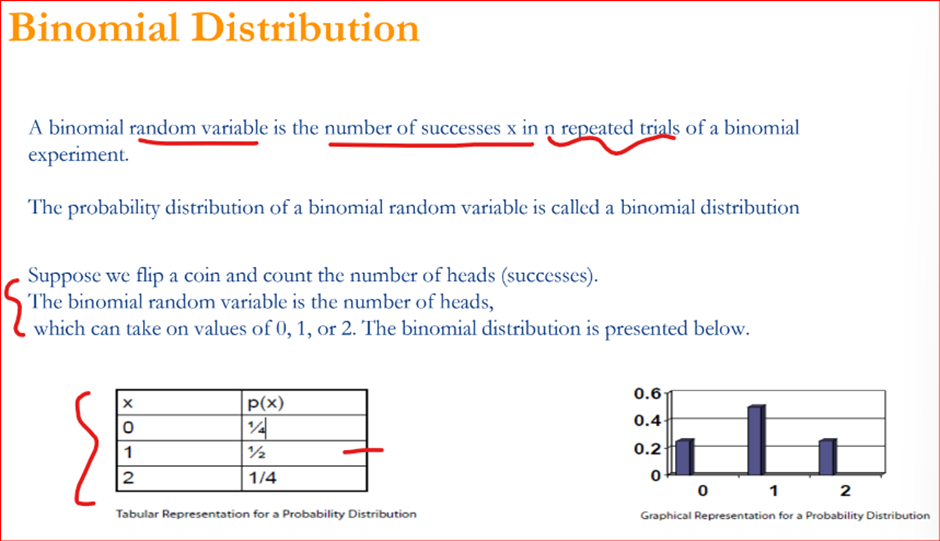
It must have exactly 2 outcomes:

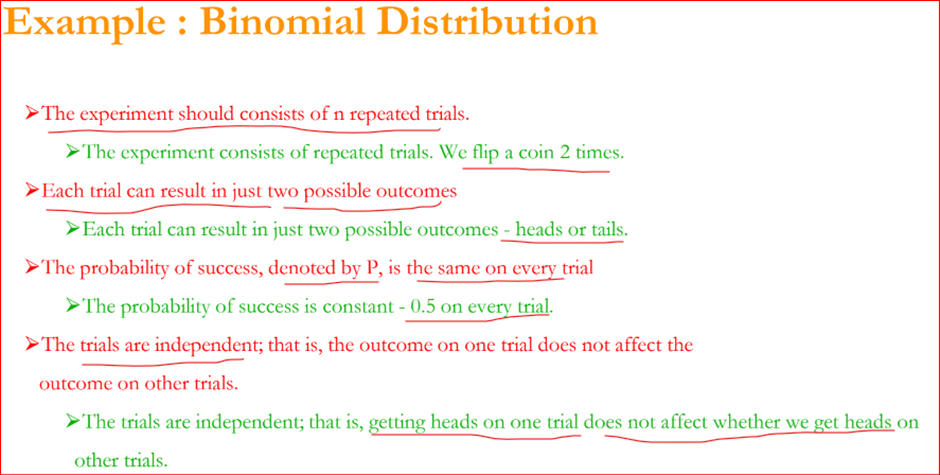
Eg: There are Biased coins where P(H) = 2/3 and P(T) = 1/3

Image representing Coin Outcomes:



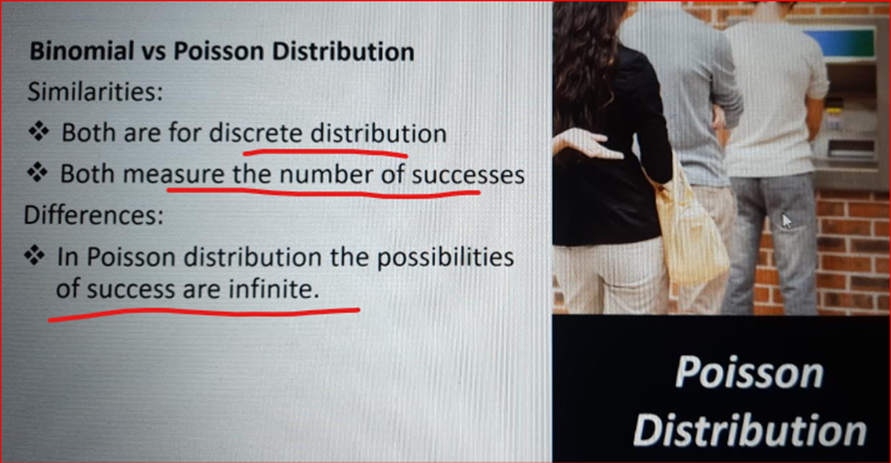
**b) Binomial Distribution:**

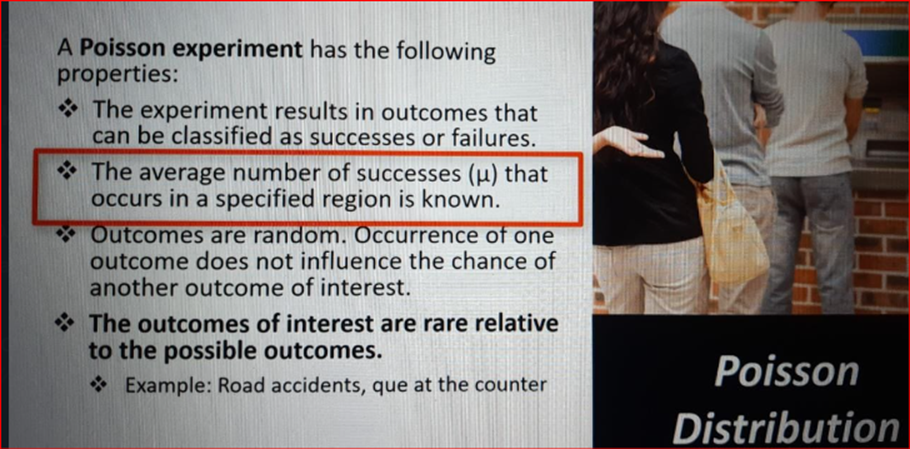




**c) Poisson Distribution:**

Image representing Poisson Experiment:





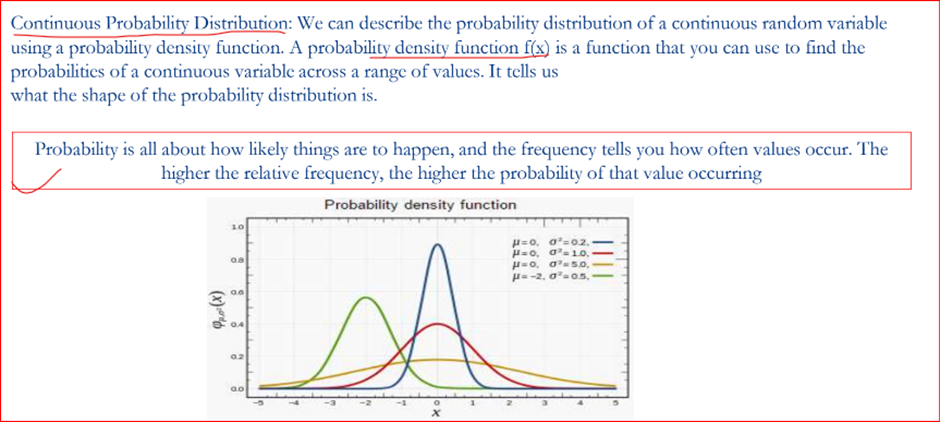
**6) Continuous Distributions**

Whenever the data is Continuous that data is called continuously distributed data

**Eg**: Suppose someone says tomorrow exactly at 12:00 PM what is the P(T = 24 degrees Celcius)

There will be infinitely many temp values considering past/ historic data, so the answer is 1/infinity = 0 (un-countable)

NOTE: In CD probability of each point is Zero



**Type of Continuous distribution:**

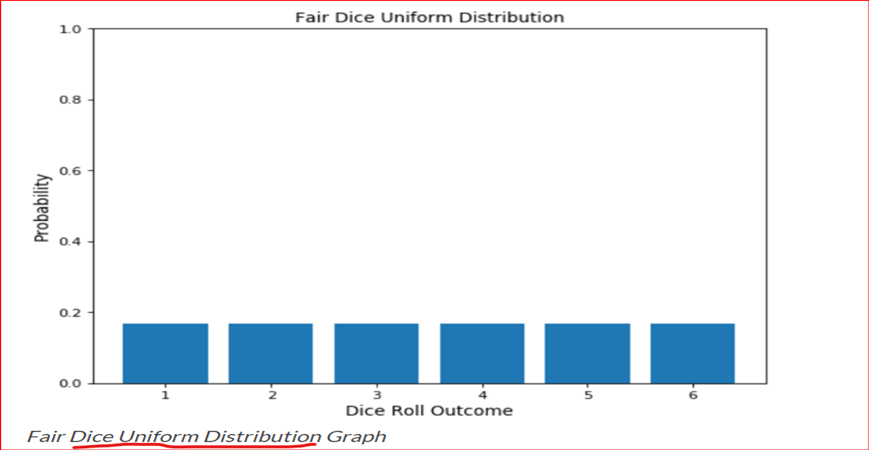
**a) Uniform Distribution:**

Uniform distribution refers to a statistical distribution in which all outcomes are equally likely.

The uniform distribution graph contains bars of equal height representing each outcome

eg: Fair Dice: 1/6 in all the 6 events = 1/6 = 0.166

Here, DENSITY is same for each region and we uniformly going to choose data from 0 to 1



**b) What is "Not a Uniform Distribution"?**

Here, Density will help us in telling if we have more chances of data coming from certain Intervals (In this distribution the question can be formed as eg: What are the temperatures that lie in the interval 25 to 30 in a week)

If the density is high we have more data

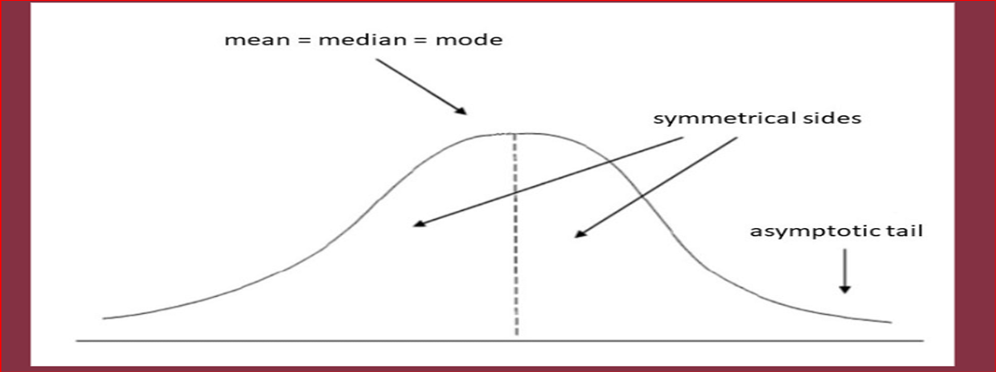
**The total area under the curve is always = 1**

**c) NORMAL Distribution:**

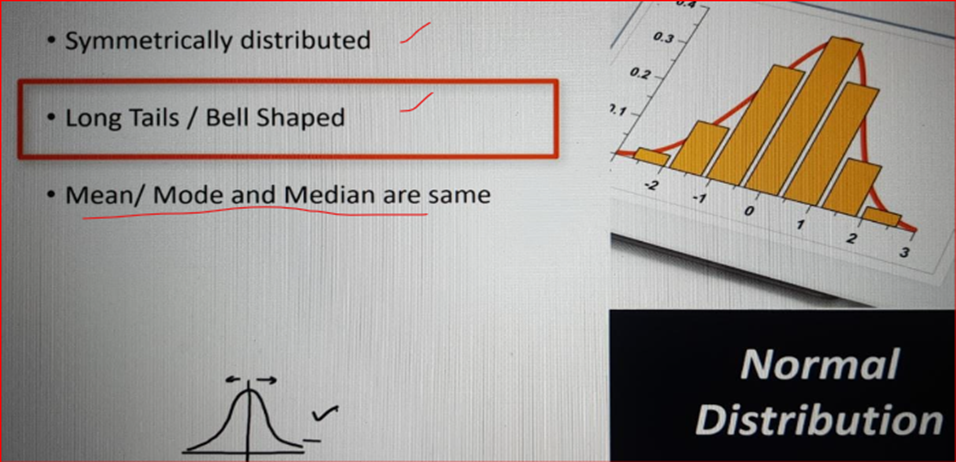
The normal distribution is an important Continuous distribution because a good number of random variables that occur in practice can be approximated to it

ND, also known as Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than the data far from the mean.

ND is used only for Continuous data, whereas previous Binomial and Poison are for discrete data



Eg: Salaries of people/ Height of people where the majority of people lie in middle and some in lower & higher



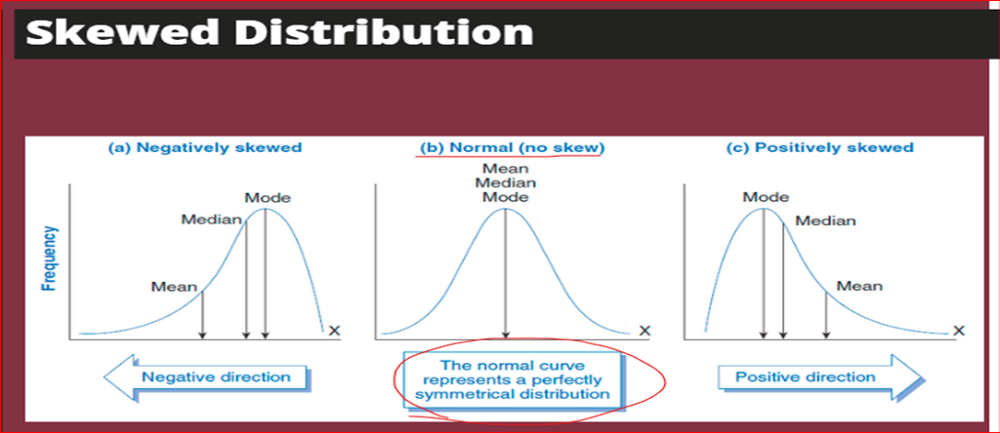
**There are 2 Types of parameters that drives the Normal Distribution shape of the curve:**

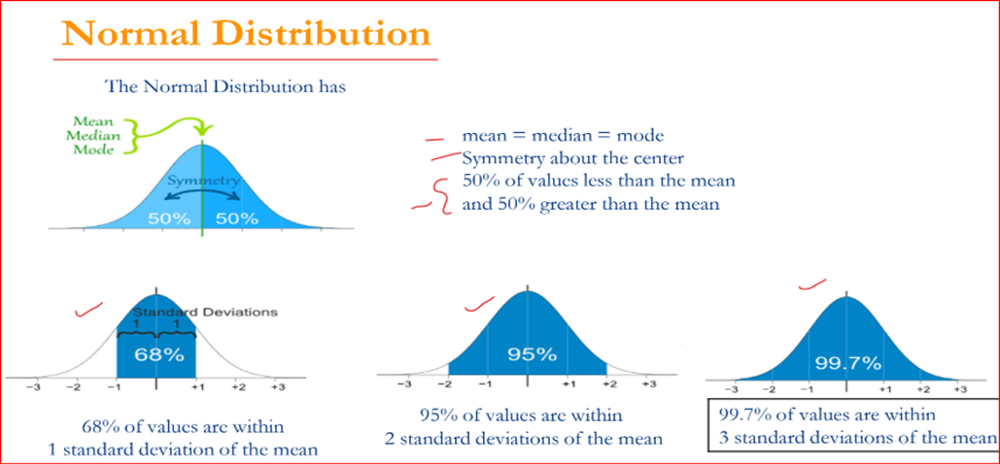
1st Parameter is Variance (σ²)

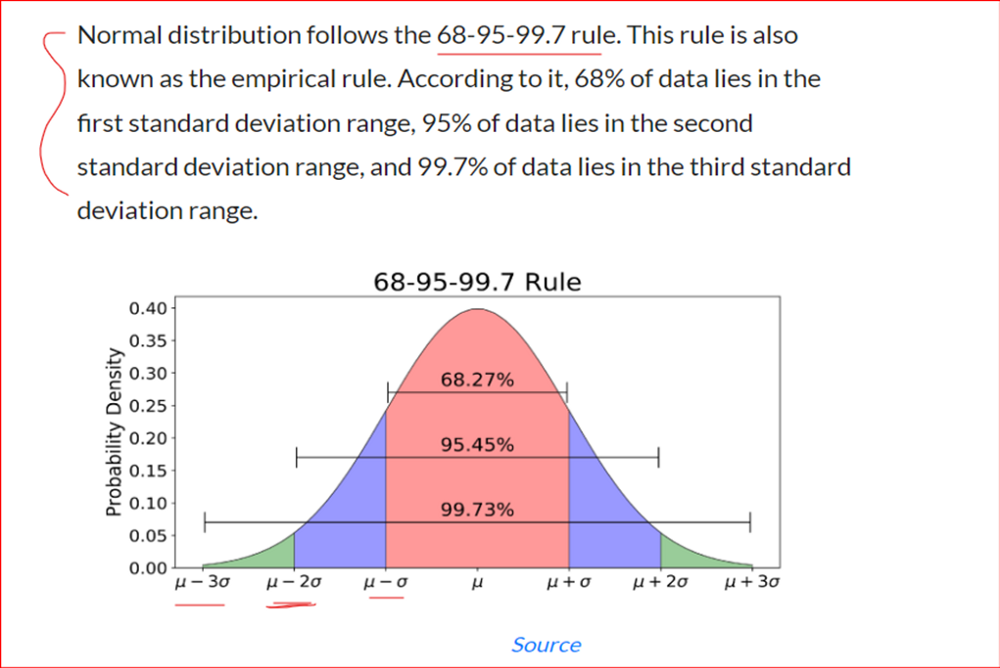
If the Spread is less -> Variance is low i.e less density

If the Spread is more -> Variance is High i.e More density

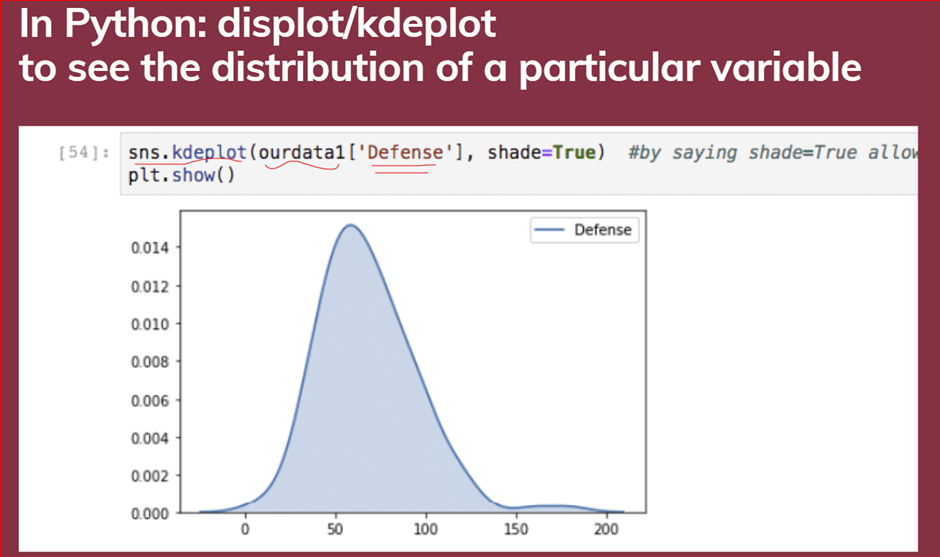
2nd parameter is Mean (µ)







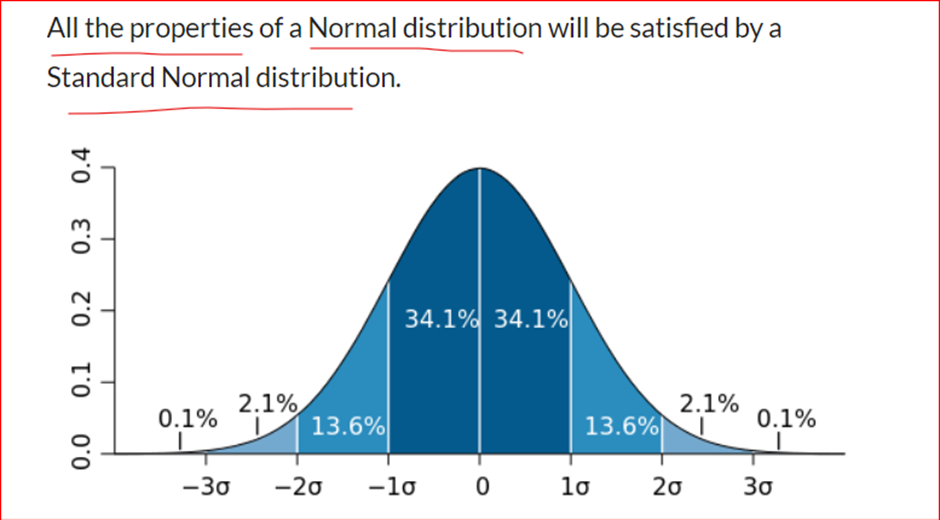
**65–95–99.7 Rule**



**d) Standard Normal Distribution or SND:**

It is denoted as Z ~ N(0, 1). And is read as X is a continuous random variable that follows Normal distribution with mean 0 and variance 1.

It is a transformation of Normal distribution in such a way that Mean = 0, and standard deviation 1.



And in addition, there exists a table that summarizes the most commonly used values of a CDF of s Standard Normal Distribution. This table is known as a Z-score table.

The formula for standardization is Z = (X-μ)/σ

**e) Student t-Test Distribution:**

Small sample size approximation of a normal distribution

It is denoted as X ~ t(k). And is read as X is a continuous random variable that follows Student's T distribution with parameter k.

where k is the degrees of freedom. If the sample size is n, then k = n-1.

Eg: For example, suppose we deal with the total number of apples sold by a shopkeeper in a month. In that case, we will use the normal distribution. Whereas, if we are dealing with the total amount of apples sold in a day, i.e., a smaller sample, we can use the t distribution.

 - - - - - - - - - - - -END - - - - - - - - - - - - - - - - - - - - -

I have tried to write a detailed article and I hope that I am successful in doing so. I'll try to keep on adding More Content that will link to each other!!

Below are the ways where you could contact me or take a look at my work.