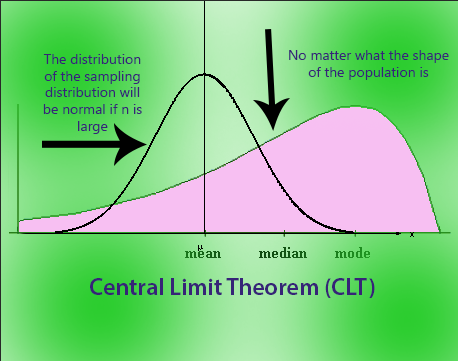
**Understanding Central Limit Theorem (CLT)**

The Central limit theorem states that given a sufficiently large sample size, the sampling distribution of the mean for a variable will approximate a normal distribution regardless of the variable’s distribution in the population.

In simple terms, the theorem states that the sampling distribution of the [mean](https://corporatefinanceinstitute.com/resources/knowledge/other/mean/) approaches a normal distribution as the size of the sample increases, regardless of the shape of the original population distribution.



**Sampling Distribution of the Mean**

The definition for central limit theorem also refers to “the sampling distribution of the mean.” What’s that?

Typically, you perform a study once, and you might calculate the mean of that one sample. Now, imagine that you repeat the study many times and collect the same sample size for each one. Then, you calculate the mean for each of these samples and graph them on a histogram. The histogram displays the distribution of sample means, which [statisticians](https://statisticsbyjim.com/glossary/statistics/) refer to as the sampling distribution of the mean.

These samples should be sufficient in size. The distribution of sample means, calculated from repeated sampling, will tend to normality as the size of your samples gets larger. This fact holds especially true for sample sizes over 30.

**Let us understand the significance of CLT**

The central limit theorem tells us that no matter what the distribution of the population is, the shape of the sampling distribution will approach [normality](https://www.simplypsychology.org/normal-distribution.html) as the sample size (N) increases.

This is useful, as the research never knows which mean in the sampling distribution is the same as the population mean, but by selecting many random samples from a population the sample means will cluster together, allowing the research to make a very good estimate of the population mean.

### **Example of Central Limit Theorem**

Consider that there are 15 sections in the science department of a university and each section hosts around 100 students. Our task is to calculate the average weight of students in the science department. Sounds simple, right?

* First, measure the weights of all the students in the science department
* Add all the weights
* Finally, divide the total sum of weights with a total number of students to get the average

But what if the size of the data is humongous? Does this approach make sense? Not really – measuring the weight of all the students will be a very tiresome and long process. So, what can we do instead? Let’s look at an alternate approach.

* First, draw groups of students at random from the class. We will call this a sample. We’ll draw multiple samples, each consisting of 30 students.
* Calculate the individual mean of these samples
* Calculate the mean of these sample means
* This value will give us the approximate mean weight of the students in the science department
* Additionally, the histogram of the sample mean weights of students will resemble a bell curve (or normal distribution)

**Application of CLT**

* Political/election polls are prime CLT applications. These polls estimate the percentage of people who support a particular candidate. You might have seen these results on news channels that come with confidence intervals. The central limit theorem helps calculate that
* Confidence interval, an application of CLT, is used to calculate the mean family income for a particular region
* The CLT is useful when examining the returns of an individual stock or broader indices, because analysis is simple, due to the relative ease of generating the necessary financial data. Consequently, investors of all types rely on the CLT to analyze stock returns, construct portfolios, and manage risk.

An investor is interested in estimating the return of stock market that is comprised of 10,000 stocks. Due to the large size of the [stock](https://corporatefinanceinstitute.com/resources/knowledge/trading-investing/dow-jones-industrial-average-djia/) market, the investor is unable to analyze each stock independently and instead chooses to use random sampling to get an estimate of the overall return of the index.

The investor picks random samples of the stocks, with each sample comprising at least 30 stocks. The samples must be random, and any previously selected samples must be replaced in subsequent samples to avoid bias.

If the first sample produces an average return of 7.5%, the next sample may produce an average return of 7.8%. With the nature of randomized sampling, each sample will produce a different result. As you increase the size of the sample size with each sample you pick, the sample means will start forming their own distributions.

The distribution of the sample means will move towards normal as the value of n increases. The average return of the stocks in the sample index estimates the return of the whole index of 100,000 stocks, and the average return is normally distributed.

**CLT in Life**

CLT says that as long as the sample size is large, any data would eventually shape up normally. So when it comes to our life just increase your sample size i.e. the time horizon you are looking at, and there’s a high chance that life would seem normal.

Say a person thinks that his life isn’t normal. Then there is a high chance that the person is not thinking on a broader horizon. All he needs to do is look back at a broader time frame (draw a larger sample number of days) and feel the difference in the perception of his current state of affairs.

**Formula of CLT**

The Central Limit Theorem is the sampling distribution of the sampling means approaches a normal distribution as the sample size gets larger, no matter what the shape of the data distribution. An essential component of the Central Limit Theorem is the average of sample means will be the population mean.

Similarly, if you find the average of all of the standard deviations in your sample, you will find the actual standard deviation for your population.

* Mean of sample is same as mean of the population.
* Standard deviation of the sample is equal to standard deviation of the population divided by square root of sample size.

Central limit theorem is applicable for a sufficiently large sample sizes (n≥30). The formula for central limit theorem can be stated as follows:

μx¯¯¯=μ

and

σx¯¯¯=σ/√n

Where,  
μ = Population mean  
σ = Population standard deviation  
μx¯¯¯ = Sample mean  
σx¯¯¯ = Sample standard deviation  
n = Sample size

### **Solved Examples**

Question 1: The record of weights of male population follows normal distribution. Its mean and standard deviation are 70 kg and 15 kg respectively. If a researcher considers the records of 50 males, then what would be the mean and standard deviation of the chosen sample?  
  
Solution:

Mean of the population μ = 70 kg  
Standard deviation of the population = 15 kg  
sample size n = 50  
Mean of the sample is given by:  
μx¯¯¯ = 70 kg  
Standard deviation of the sample is given by:  
σx¯¯¯ = σn√  
σx¯¯¯ = 1550√  
σx¯¯¯ = 2.121 = 2.1 kg (approx)