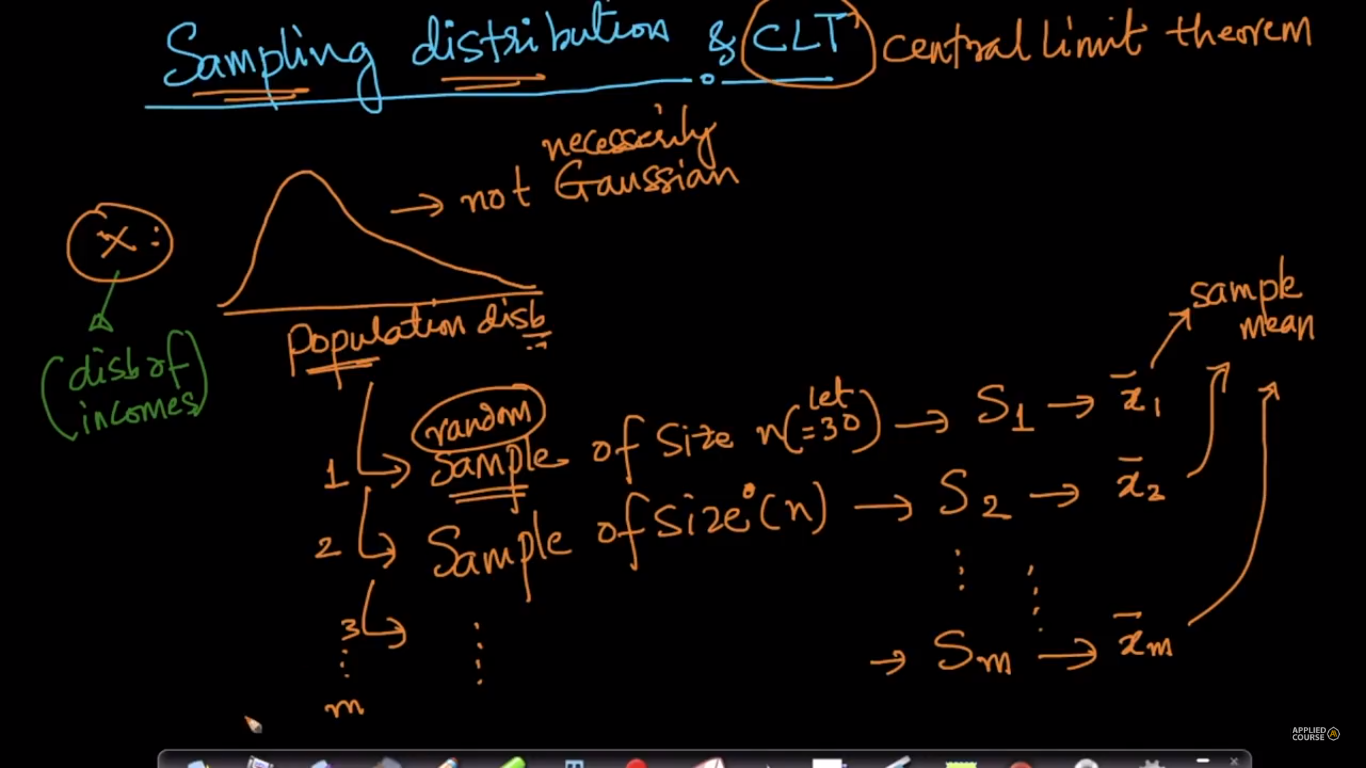
**Sampling distribution:**

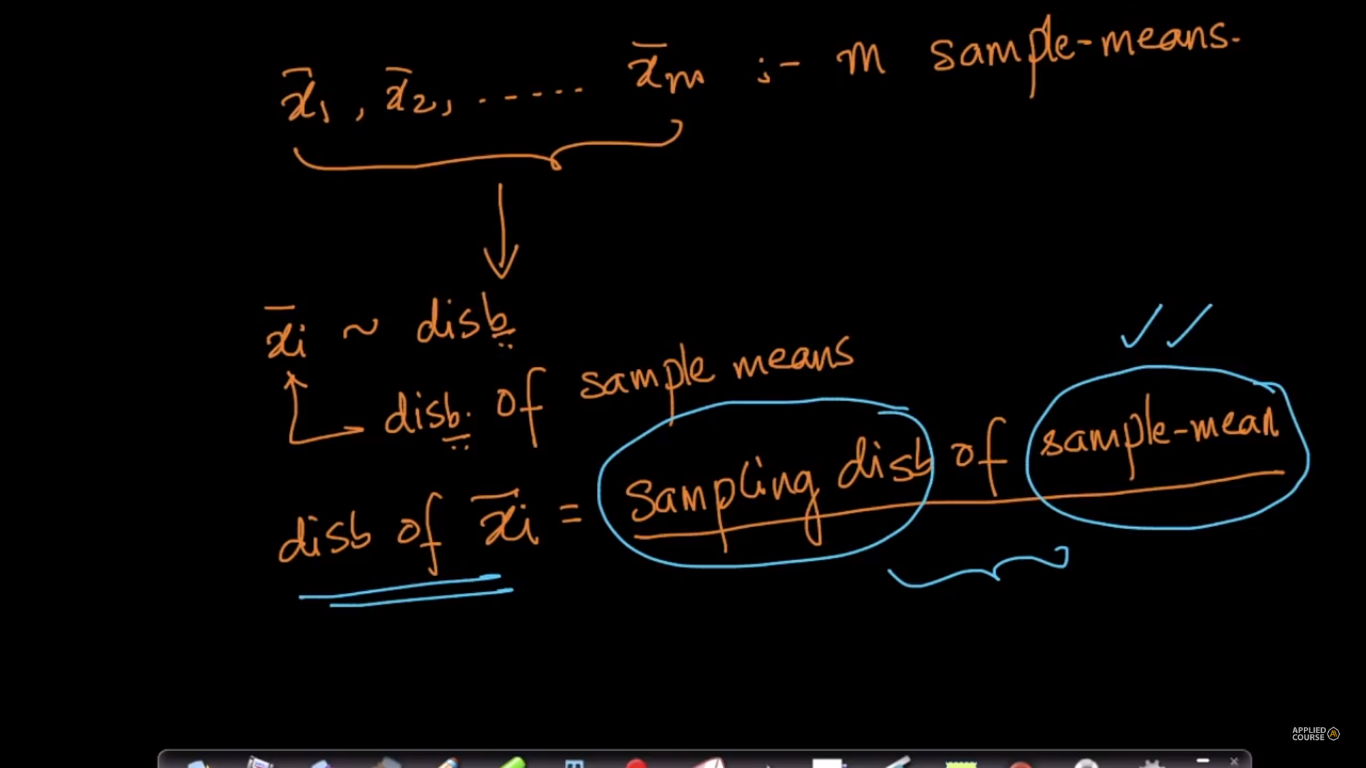
Suppose we have a random variable(X) which is not need to be gaussian/normal distributed, this X is a whole population.

Now we take sample of uniform size (let say 30 here) from this population.

Then we calculate the mean of each sample taken from population and this mean is called **Sample Mean.**

Suppose we take **m**  no of samples, so now we have **m**  no sample mean, now we will find or draw the distribution from obtained sample means and this distribution is known as **Sampling distribution of sample mean.**





**Central limit theorem(CLT):**

The central limit theorem states that the sampling distribution of the mean of any [independent](https://www.stattrek.com/help/glossary.aspx?target=independent), [random variable](https://www.stattrek.com/help/glossary.aspx?target=random_variable)(not necessarily be gaussian distributed) will be normal or nearly normal distribution, if the sample size is large enough.

How large is "large enough"? The answer depends on two factors.

* Requirements for accuracy. The more closely the sampling distribution needs to resemble a normal distribution, the more sample points will be required.
* The shape of the underlying population. The more closely the original population resembles a normal distribution, the fewer sample points will be required.

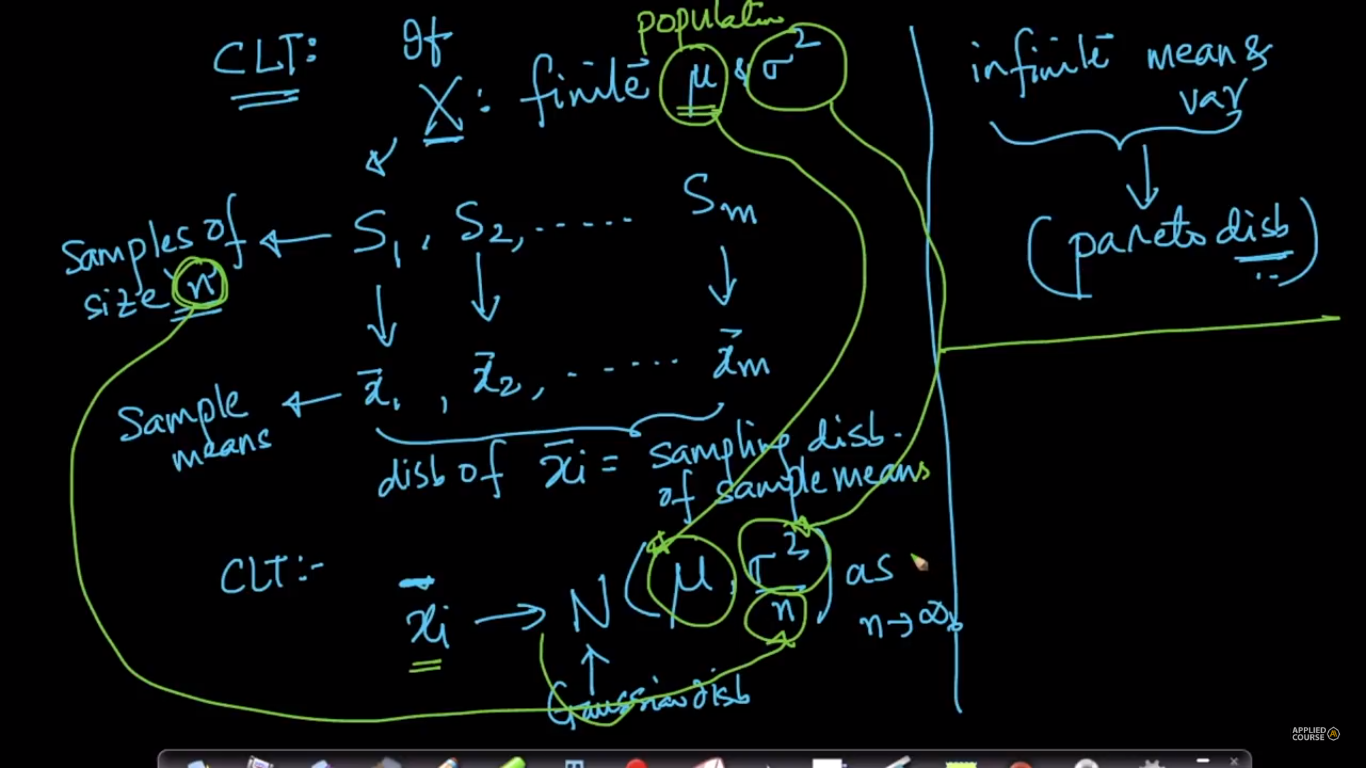
In practice, some statisticians say that a sample size of 30 is large enough when the population distribution is roughly bell-shaped. Others recommend a sample size of at least 40. But if the original population is distinctly not normal (e.g., is badly skewed, has multiple peaks, and/or has outliers), researchers like the sample size to be even larger.

In laymen term we can say that for given random variable X (which can follow any distribution not necessarily gaussian distribution), then we can easily find the mean and the variance of X by using CLT as:

1. Take **m** samples from X, called sample means
2. Draw distribution of these sample means
3. Then the distribution we obtained by drawing distribution of these sample means will be Normal distribution.
4. And the mean of this **Sampling distribution of sample mean** will be equal to mean of given population or X.
5. And the variance of this **Sampling distribution of sample mean**  will be equal to variance of population or X / n, here **n** is size of each sample.

That means,

variance of **Sampling distribution of sample mean =** Variance of X / size of each sample.



**How to calculate mean and variance of given random variable using CLT.**

Suppose as stated in the video if we take 1000 samples of size 30 each.We calculate 1000 different means and plot a single graph which will be Gaussian distributed.   
Now, It is stated in the video that mean (mu) of xi(bar)'s will be approx: equal to population mean(as 'n' tends to infinity).   
So this means:  
1)We computes 1000 different means==> 1000 distinct values  
2)Take the mean of these 1000 means==> 1 value  
  
The mean obtained in step 2 will be approx: equal to Population mean(as n==>infinity).  
3) We calculate the variance using the mean obtained in step 2 and using 1000 mean values from step 1.  
if we multiply the variance obtained in step 3 with 'n'(n=30 here) then we will obtain the population variance(as n==>infinity).  
  
So using the samples we will get to know the population mean and variance

**Conclusions:**

|  |
| --- |
| The Central Limit Theorem states that regardless of the shape of the population distribution, the distribution of sample means will be approximately normal. |

From the central limit theorem, the following is true:

1. Population distributions that have no skew will lead to distributions of sample means that have no skew.

2. Population distributions that are skewed right will lead to distributions of sample means that have no skew.

3. Population distributions that are skewed left will lead to distributions of sample means that have no skew.

The distribution of sample means will become more normal as its sample size increases.

Good rule of thumb: sample distributions will usually be approximately normal if their sample size is n = 30 or larger.