Central Limit Theorem (CLT)

The central limit theorem (CLT) is a fundamental concept in statistics. It describes the behavior of the **sampling distribution of the mean** as the sample size increases.

Imagine you have a large population: Let's say it's all the students in a school district, and you're interested in their average height.

- **Sampling:** It's impractical to measure everyone, so you take a smaller sample (a random group of students).
- **Sampling Distribution:** Now, imagine repeating this process many times, each time taking a new random sample of the same size. The **sampling distribution** refers to the collection of all these sample means. It shows the probability of getting different average heights across these random samples.
- **The Magic of Large Samples:** Here's where the CLT comes in. The theorem states that **as the sample size gets larger**, the sampling distribution of the mean starts to resemble a **normal distribution (bell curve)**, regardless of the original population's distribution (assuming it's not extremely skewed).

The parameters of the sampling distribution of the mean are determined by the parameters of the population:

• The mean of the sampling distribution is the mean of the population.

$$\mu_{\bar{x}} = \mu$$

• The standard **deviation** of the sampling distribution is the standard deviation of the population divided by the square root of the sample size.

$$\sigma_{\bar{x}} \text{ (sigma bar)} = \sigma / \sqrt{n}$$

Conditions of the central limit theorem

The central limit theorem states that the sampling distribution of the mean will always follow a normal distribution under the following conditions:

1. The sample size is **sufficiently large**. This condition is usually met if the sample size is $n \geq 30$.
2. The samples are **independent and identically distributed random variables**. This condition is usually met if the sampling is random.
3. The population's distribution has **finite variance**. Central limit theorem doesn't apply to distributions with infinite variance, such as the Cauchy distribution. Most distributions have finite variance.