Statistics-Central Limit Theorem

Central Limit Theorem (CLT):

Given a large enough sample size (n), the sampling distribution of a statistic (e.g., mean, proportion) will be approximately normally distributed, regardless of the shape of the population distribution from which the samples are drawn.

Key points:

1. **Large sample size**: The CLT requires that the sample size (n) is sufficiently large, typically $n \geq 30$.
2. **Sampling distribution**: The sampling distribution refers to the distribution of a statistic across many randomly drawn samples from a population.
3. **Normality**: Despite the population distribution being skewed or non-normal, the sampling distribution will be approximately normal.

Example:

Suppose we want to estimate the average height (in cm) of a population using a sample of 36 adults. The heights in this population are known to follow a lognormal distribution, which is skewed and not normally distributed.

Let's consider two scenarios:

Scenario 1: Sample size = 6 In this case, the sampling distribution of the mean height would likely be very skewed, mirroring the shape of the population distribution.

Scenario 2: Sample size = 36 (larger sample) Using the CLT, we can expect that the sampling distribution of the mean height will be approximately normally distributed, even though the underlying population distribution is skewed. This means that our estimate of the average height based on this larger sample would likely be more accurate and follow a normal distribution.

In summary, the Central Limit Theorem tells us that with a sufficiently large sample size ($n \geq 30$), the sampling distribution of a statistic will be approximately normally distributed, allowing us to make inferences about population parameters.

Curated by Brajesh Kumar