Sampling Distribution

Sampling distribution is a probability distribution that describes the statistical properties of a sample statistic (such as the sample mean or sample proportion) computed from multiple independent samples of the same size from a population.

- Population : The entire data.
- Sample : Some part of the data.

Why Sampling Distribution is important?

Sampling distribution is important in statistics because it allows us to estimate the variability of a sample statistic, which is useful for making inferences about the population. By analysing the properties of the sampling distribution, we can compute confidence intervals, perform hypothesis tests, and make predictions about the population based on the sample data.

Central Limit Theorem

The Central Limit Theorem (CLT) states that the distribution of the sample means of a large number of independent and identically distributed random variables will approach a normal distribution.

The conditions required for the CLT:

1. The sample size is large enough, typically greater than or equal to 30.
2. The random variables in the sample are independent and identically distributed.

Point Estimate

A point estimate is a single value, calculated from a sample, that serves as the best guess or approximation for an unknown population parameter, such as the mean or standard deviation. Point estimates are often used in statistics when we want to make inferences about a population based on a sample.

Confidence Interval

Confidence interval, in simple words, is a range of values within which we expect a particular population parameter, like a mean, to fall. It's a way to express the uncertainty around an estimate obtained from a sample of data.

Confidence level, usually expressed as a percentage like 95%, indicates how sure we are that the true value lies within the interval.

Confidence Interval = Point Estimate ± Margin of Error

Two Types of CI :

- Confidence Interval (Z Procedure)
- Confidence Interval (T Procedure)

1. Confidence Interval With Z Procedure:

Assumptions :

1. Random Sampling
2. Known population Standard Daviation
3. Normal Distribution , if not normal than apply central limit theorem to make distribution normal.

Formula : CI = $\overline{X} \pm Z\alpha/2 \ \frac{\delta}{\sqrt{n}}$

Confidence Interval With T Procedure:

Assumptions :

1. Random Sampling
2. Unknown population Standard Daviation
3. Normal Distribution , if not normal than apply central limit theorem to make distribution normal.

Formula : CI = $\overline{X} \pm t\alpha/2 \ \frac{S}{\sqrt{n}}$

Scientific Reason Why CI Work

Even with one sample, the CI works because:

1. **Sample is random and unbiased:**
   Each random sample is equally likely to be above or below the population mean.

2. **CLT guarantees:**
   The distribution of sample means is predictable (approximately normal).

3. **Probability interpretation:**
   If you repeated the experiment many times, 95% of the confidence intervals calculated from different samples would contain the true population mean.