Confidence Interval

A confidence interval (CI) is a statistical method used to estimate a range of values that likely contains the true population parameter (usually the mean) with a certain level of confidence. In simpler terms, it tells you how precise your estimate of the population mean is based on your sample data.

When do you use confidence intervals?

You can calculate confidence intervals for many kinds of statistical estimates, including:

- Proportions
- Population means
- Differences between population means or proportions
- Estimates of variation among groups

There are two main scenarios for calculating confidence intervals:

1. Standard Normal Distribution (Population Standard Deviation Known):

- CI = Confidence Interval
- $\bar{X}$ (mu bar) = Population Mean (which is usually unknown in real-world applications)
- Z* = Critical value from the standard normal z-distribution (based on the chosen confidence level)
- σ (sigma) = Population Standard Deviation (assumed to be known)
- √n = Square root of the Sample Size

The formula:

$$CI = \bar{X} \pm Z^* (\sigma / \sqrt{n})$$

2. Unknown Population Standard Deviation (Estimated Using Sample Data):

- CI = Confidence Interval
- ẍ (x bar with a hat) = Sample Mean (used as an estimate for the population mean)
- t* = Critical value from the student t-distribution (based on the chosen confidence level and degrees of freedom, which is n-1)
- s (sigma with a hat) = Sample Standard Deviation (used as an estimate for the population standard deviation)
- √n = Square root of the Sample Size

The formula:

$$CI = \ddot{x} \pm t^* (s / \sqrt{n})$$

When to use a confidence interval for proportions:

- You're studying a categorical variable with only two outcomes (often success/failure or yes/no). For example, the proportion of voters who support a particular candidate.
- You have a random sample from the population of interest.

Formula for confidence interval for proportions:

- CI = Confidence Interval
- p̂ (p hat) = Sample proportion (successes divided by total sample size)
- z* = Critical value from the standard normal z-distribution (based on the chosen confidence level)
- √(p̂ * (1 - p̂) / n) = Margin of error (square root of the product of estimated success probability (p̂) and estimated failure probability (1-p̂) divided by the sample size (n))

The formula:

$$CI = \hat{p} \pm z^* \sqrt{\hat{p} * (1 - \hat{p}) / n}$$