Confidence Intervals

A **Confidence Interval** (CI) is a range of values, derived from a sample, that is likely to contain the true population parameter with a certain level of confidence, typically, 95%. The confidence level reflects how often the interval would contain the true parameter if the experiment were repeated many times.

For a population proportion (mean) π, the confidence interval is:

$$CI_{0.95} = \left[\hat{\pi} - z_{\alpha/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}},\ \hat{\pi} + z_{\alpha/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}\right]$$

where$\hat{\pi}$ is the sample proportion (mean), $z_{\alpha/2}$ is the critical value from the standard normal distribution corresponding to the confidence level, and n is the sample size.

Example

In a study of 200 patients, 140 showed improvement after treatment. The sample proportion is $\hat{\pi} = 140/200 = 0.70$. For a 95% confidence level, the critical value 1.96, and the margin of error is $1.96\sqrt{\frac{0.3*0.7}{200}} = 0.06351$. Therefore, the 95% confidence interval is

$$CI_{0.95} = 0.6365, 0.7635$$

meaning we are 95% confident that the true proportion of patients who show improvement lies between 63.66% and 76.35%.

Highest Posterior Density Intervals

A **Highest Posterior Density Interval** (HPDI) is an interval that contains the most probable values of a parameter, given the observed data and prior distribution, with a specified posterior probability, typically 95%. Unlike confidence intervals, the HPDI is derived from the posterior distribution in Bayesian inference and always contains the densest region of the posterior probability.

There is no single formula for HPDI, as it depends on the posterior distribution $p(\theta \mid data)$. The HPDI for a parameter θ with posterior distribution is the shortest interval $[a, b]$ such that:

$$HPDI_{0.95} = \int_a^b p(\theta \mid data) d\theta = \gamma$$

where γ is the desired credible level (e.g., 95%).

Example

In a study of 200 patients, 140 showed improvement. Assume a $Beta(2,2)$ prior (commonly used to represent weak prior knowledge). The posterior distribution is $Beta(\alpha + x, \beta + n - x)$.
We want to compute the 95% HPDI. The HPDI for the $Beta(142, 62)$ distribution can be found by solving for the interval $[a, b]$ such that 95% of the posterior probability lies within this interval, and the posterior density at any point inside the interval is higher than outside the interval. Using numerical methods, the 95% HPDI is approximately

$$HPDI_{0.95} = 0.6365, 0.7635$$

This means we are 95% confident that the true proportion of patients who improve after treatment lies between 63.66% and 76.35%., based on the data and prior beliefs.

Confidence Intervals: simulations