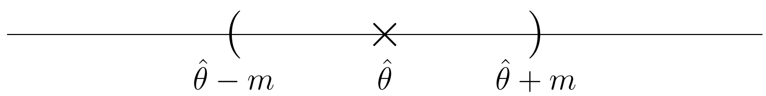
Confidence Intervals

Since a point estimate of a population parameter θ is a value of a continuous random variable (the point estimator) based on a particular sample, the probability that $\hat{\theta} = \theta$ is 0. The goal is to find a short interval with a high confidence level that the true θ value is contained within that interval. For a given confidence level α and some value m:

$$P\left(\theta \in (\hat{\theta} - m, \hat{\theta} + m)\right) = 1 - \alpha$$



Definition: Confidence Interval

Let $\hat{\theta}$ be a point estimate/estimator for a population parameter θ and let $1 - \alpha$ be a desired confidence level. The random interval:

$$\hat{\theta} \pm m$$

such that:

$$P\left(\theta \in (\hat{\theta} - m, \hat{\theta} + m)\right) = 1 - \alpha$$

is called $1 - \alpha$ *confidence interval* for θ. The value $1 - \alpha$ is called the *confidence level* and the value m is called the *margin of error*.

Theorem

Let $X_i \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ such that μ is unknown but σ is known. The margin for error m for the $1 - a$ confidence interval for $\bar{X}$ from a sample of size n is given by:

$$m = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

Hence:

$$P\left(\mu \in (\bar{X} - m, \bar{X} + m)\right) = 1 - \alpha$$

Proof.

$$
\begin{aligned}
&P\left(\mu \in (\bar{X}-m, \bar{X}+m)\right) = 1-\alpha \\
&P\left(-m \le \bar{X}-\mu \le m)\right) = 1-\alpha \\
&P\left(-\frac{m}{\frac{\sigma}{\sqrt{n}}} \le \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \le \frac{m}{\frac{\sigma}{\sqrt{n}}}\right) = 1-\alpha \\
&P\left(-\frac{m}{\frac{\sigma}{\sqrt{n}}} \le Z \le \frac{m}{\frac{\sigma}{\sqrt{n}}}\right) = 1-\alpha \\
&z_{\frac{\alpha}{2}} = \frac{m}{\frac{\sigma}{\sqrt{n}}}
\end{aligned}
$$

$$
\therefore m = z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}
$$

■

Example

A sample carton of a dozen brown eggs from a farm has $\bar{x} = 65.5$ and $\sigma = 2$. Find the 95% confidence for μ.

$$
\begin{aligned}
&\alpha = 1 - 0.95 = 0.05 \\
&z_{0.025} = \Phi(0.975) = 1.96
\end{aligned}
$$

$$
\bar{x} \pm z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} = 65.5 \pm 1.96\frac{2}{\sqrt{12}} = 65.5 \pm 1.1 = (64.4, 66.6)
$$

Correct interpretation:

- $(64.4, 66.6)$ is a 95% confidence interval for μ.
- We are 95% confident that the true value of μ is contained by this interval.

Incorrent interpretation:

- The probability that μ is contained by this interval is 0.95.

Note that *probability* predicts something that hasn't happened yet, whereas *confidence* describes the faith in a particular outcome.

As sample size n increases, the width m of the confidence interval decreases.

Example

From the previous example, increase the sample size to $n = 48$. Find the new 95% confidence interval.

$$\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 65.5 \pm 1.96 \frac{2}{\sqrt{48}} = 65.5 \pm 0.6 = (64.9, 66.1)$$

How large must the sample size be in order for the margin of error to be 0.2?

$$n = \left(z_{\frac{\alpha}{2}} \frac{\sigma}{m}\right)^2 = \left(1.96 \frac{2}{0.2}\right)^2 = 384.2$$

Thus, the minimum necessary sample size is $n = 385$.

As confidence level $1 - a$ increases, α decreases and $z_{\frac{\alpha}{2}}$ increases, thus increasing the width m of the confidence interval.

Example

From the previous example with $n = 12$, find the 90% and 99% confidence intervals.

For 90%:

$$\alpha = 1 - 0.90 = 0.10$$
$$z_{0.05} = \Phi(0.95) = 1.645$$

$$\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 65.5 \pm 1.645 \frac{2}{\sqrt{12}} = 65.5 \pm 0.9 = (64.6, 66.4)$$

For 99%:

$$\alpha = 1 - 0.99 = 0.01$$
$$z_{0.005} = \Phi(0.995) = 2.575$$

$$\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 65.5 \pm 2.575 \frac{2}{\sqrt{12}} = 65.5 \pm 1.5 = (64.0, 67.0)$$

As expected, 99% CI is wider that 95% CI is wider than 90% CI.