Probability and Statistics

Alvin Lin

Probability and Statistics: January 2017 - May 2017

Basic Properties of Confidence Intervals

$$B - r \leq A \leq B + r$$

Add $-B$ to each side:

$$-r \leq -B + A \leq r$$

Add $-A$ to each side:

$$-A - r \leq -B \leq -A + r$$

Multiply both sides by -1:

$$A + r \geq B \geq A - r$$

The original statement is logically equivalent to the following:

$$A - r \leq B \leq A + r$$

That is:

$$B - r \leq A \leq B + r \quad \leftrightarrow \quad A - r \leq B \leq A + r$$

Theorem

Consider a random sample $X_1, X_2, \ldots, X_n$ from a normal distribution with mean value μ (population mean) and standard deviation σ (population standard deviation). Then the sample mean

$$\overline{X} = \frac{X1 + X_2 + \cdots + X_n}{n}$$

is normally distributed with with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$. The random variable $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ is a standard normal random variable.

Example

Consider the following probability:

$$\begin{aligned}
P(-1.96 \leq Z \leq 1.96) &= \Phi(1.96) - \Phi(-1.96) \\
&= 0.9750 - 0.0250 \\
&= 0.95 \\
&= P(-1.96 \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}) \\
&= P(-1.96\frac{\sigma}{\sqrt{n}} \leq \overline{X} - \mu \leq 1.96\frac{\sigma}{\sqrt{n}}) \\
&= P(\mu - 1.96\frac{\sigma}{\sqrt{n}} \leq \overline{X} \leq \mu + 1.96\frac{\sigma}{\sqrt{n}}) \\
&= P(\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + 1.96\frac{\sigma}{\sqrt{n}})
\end{aligned}$$

This represents the probability that the interval $(\overline{X} - 1.96\frac{\sigma}{\sqrt{n}}, \overline{X} + 1.96\frac{\sigma}{\sqrt{n}})$ contains the population mean μ. After observing $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, we compute the observed sample mean $\overline{x}$.

$$\left(\overline{x} - 1.96\frac{\sigma}{\sqrt{n}}, \overline{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

This interval is a 95% confidence interval for μ.

Example

Suppose we repeat taking a sample of size n, finding $\overline{x}$, and calculating the interval 100 times. How many intervals do not contain the population mean μ? It could be any number. If N (the # of repetitions of sampling) $\rightarrow \infty$, then 5% of the 95% confidence intervals do not contain μ.

Definition

$100(1-\alpha)\%$ confidence interval for the population mean μ of a normal population when the population standard deviation σ is known is:

$$\overline{x} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$
\begin{aligned}
1-\alpha &= P(Z_{\alpha/z} \le Z \le Z_{\alpha/2}) \\
&= P(\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) \\
&= P(\textit{the interval } \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \textit{ contains } \mu)
\end{aligned}
$$

The sample size necessary for the confidence interval to have a width w is

$$n = (2Z_{\alpha/2}\frac{\sigma}{w})^2$$

Example

What is the confidence level for the interval $\overline{x} \pm 2.81\frac{\sigma}{\sqrt{n}}$?

$$
\begin{aligned}
Z_{\alpha/2} &= 2.81 \\
\alpha &= 0.0050 \\
100(1-\alpha)\% &= 99.5\%
\end{aligned}
$$

You can find all my notes at `http://omgimanerd.tech/notes`. If you have any questions, comments, or concerns, please contact me at alvin@omgimanerd.tech