Confidence Interval Based on a Single Sample

An Example

$$X_1, X_2, \ldots, X_n \sim N(\mu, 2)$$

Point estimator: $\hat{\mu} = \bar{X}$
$P(\bar{X} = \mu) =??$

$P(\mu \in (\bar{X} - l, \bar{X} + l)) =??$

How to determine l?

Confidence level

Random confidence interval

A sample: 2, 3, 1, 6, 5, 7, 10, 4, 9, 8

Confidence interval

Interpreting a confidence interval

$100(1-\alpha)\%$ **confidence interval for** μ

$$(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$$

Precision and Choice of sample size

$$w = 2 \cdot z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Example: Suppose $\sigma = 25$, what sample size is necessary to ensure that the resulting 95% CI has a width of (at most) 10?

General $100(1-\alpha)\%$ CIs

$$X_1, X_2, \ldots, X_n \sim f(x;\theta)$$

$$(l(X_1, X_2, \ldots \ldots, X_n), u(X_1, X_2, \ldots, X_n))$$

is a $100(1-\alpha)\%$ CI for θ if

$$P(l(X_1, X_2, \ldots \ldots, X_n), u(X_1, X_2, \ldots, X_n)) = 1-\alpha$$

CIs for μ with σ unknown

$$X_1, X_2, \ldots, X_n \sim N(\mu, \sigma)$$

Similarly, $P(\bar{X} - l \le \mu \le \bar{X} + l) = 1 - \alpha$

How to determine l??

A new distribution: t distribution

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

Basic properties

Let t_ν denote the density function curve for ν df.

1. Each t_ν curve is bell-shaped and centered at 0.
2. Each t_ν curve is more spread out than the standard normal curve.
3. As ν increases, the spread of the corresponding t_ν curve decreases.
4. As $\nu \to \infty$, the sequence of t_ν curves approaches the standard normal curve (so the z curve is often called the t curve with df=∞)

t critical values

Let $t_{\alpha,\nu}$ = the number for which the area under the t curve with ν df to the right of $t_{\alpha,\nu}$ is called a t critical value.

Examples

One sample t confidence interval

The $100(1-\alpha)\%$ CI for μ is

$$(\bar{x} - t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}}, \bar{x} - t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}})$$

Example: 2, 3, 1, 6, 5, 7, 10, 4, 9, 8 $\sim N(\mu, \sigma)$

Prediction interval

suppose

$$X_1, X_2, \ldots, X_n \sim N(\mu, \sigma)$$

Use an interval to predict the next observation X_{n+1}

A similar question

$$P(\bar{X} - l \le X_{n+1} \le \bar{X} + l) = 1 - \alpha$$

Equivalently,

$$P(-l \le \bar{X} - X_{n+1} \le l) = 1 - \alpha$$

$E(\bar{X} - X_{n+1})$
$V(\bar{X} - X_{n+1})$
Result: A prediction interval (PI) for a single observation t be selected from a normal distribution is

$$(\bar{x} - t_{\alpha/2,n-1} \cdot \sqrt{1 + 1/n}\ \bar{x} + t_{\alpha/2,n-1} \cdot \sqrt{1 + 1/n})$$

The prediction level is $100(1 - \alpha)\%$.

Example: n=10, $\bar{x} = 21.90$, $s = 4.134$, construct the 95% PI for the next observation.

Confidence intervals without normality assumption
A large-sample interval for μ
If n is sufficiently large ($n > 40$), the standardized variable

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

has approximately a standard normal distribution. This implies that

$$(\bar{x} - z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{s}{\sqrt{n}})$$

is a large-sample confidence interval for μ with confidence level approximately $100(1-\alpha)\%$.

A large-sample CI for a population proportion p

Point estimator: $\hat{p} = X/n$

$E(\hat{p}) = p$

$V(\hat{p}) =$

When n is large, $\hat{p}$ is approximately normally distributed

$$P(-z_{\alpha/2} \leq \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \leq z_{\alpha/2}) \approx 1-\alpha$$

Hence, a CI with confidence level approximately $100(1-\alpha)\%$ has

$$\text{left endpoint} = \frac{\hat{p} + z_{\alpha/2}^2/2n - z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n + z_{\alpha/2}^2/4n^2}}{1+(z_{\alpha/2}^2)/n}$$

and

$$\text{right endpoint} = \frac{\hat{p} + z_{\alpha/2}^2/2n + z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n + z_{\alpha/2}^2/4n^2}}{1+(z_{\alpha/2}^2)/n}$$