Large-Sample CI's for any Pop. Mean/Proportion

Engineering Statistics
Section 7.2

Josh Engwer

TTU

01 April 2016

z Critical Values

A key component to some CI's is the z **critical value**:

Definition

$z^*_{\alpha/2}$ is called a z **critical value** of the std normal distribution such that its upper-tail probability is exactly its subscript value $\alpha/2$:

$$\mathbb{P}(Z > z^*_{\alpha/2}) = \alpha/2$$

IMPORTANT: Do not confuse z critical value $z^*_{\alpha/2}$ with the z percentile $z_{\alpha/2}$:

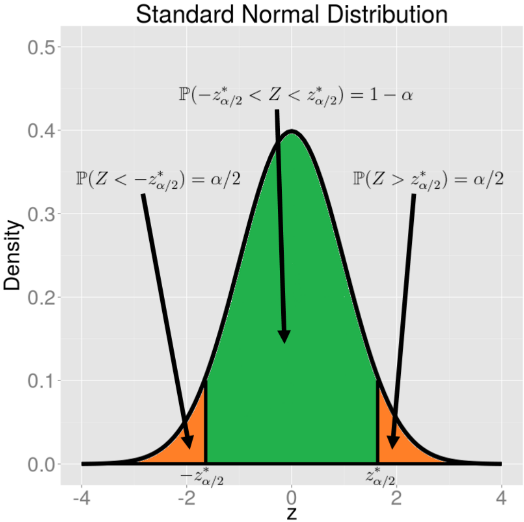
$$\mathbb{P}(Z \le z_{\alpha/2}) = \alpha/2$$

Finally, notice that $z^*_{\alpha/2}$ is always **positive**.

Common z critical values:

	90% CI	95% CI	99% CI
$z^*_{\alpha/2}$	1.64	1.96	2.58

z Critical Values



Common z critical values:

	90% CI	95% CI	99% CI
$z^*_{\alpha/2}$	1.64	1.96	2.58

Large-Sample CI for Population Mean μ (Motivation)

Given any population with mean μ.
Let $\mathbf{X} := (X_1, \ldots, X_n)$ be a random sample from the population.
Then, construct the $100(1-\alpha)\%$ CI for parameter μ:

1. Produce a suitable **pivot**: Let $Q(\mathbf{X};\mu) = \dfrac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$
2. Determine the pdf of the pivot: Provided $n > 30$, then by the CLT,
$$\overline{X} \overset{approx}{\sim} \text{Normal}(\mu, \sigma/\sqrt{n}) \implies Q(\mathbf{X};\mu) \overset{approx}{\sim} \text{Normal}(0,1)$$
3. Find constants $a < b$ such that $\mathbb{P}(a < Q(\mathbf{X};\mu) < b) = 1 - \alpha$
   Since the std normal pdf is symmetric, $a = -z^*_{\alpha/2}$ and $b = z^*_{\alpha/2}$
4. Manipulate the inequalities to isolate parameter μ:
$$-z^*_{\alpha/2} < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < z^*_{\alpha/2} \implies \overline{X} - z^*_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z^*_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$
5. Let $\mathbf{x} := (x_1, \ldots, x_n)$ be a sample taken from the population.
6. Replace estimator $\overline{X}$ & parameter σ with $\bar{x}$ & s computed from sample:
$$\bar{x} - z^*_{\alpha/2} \cdot \frac{s}{\sqrt{n}} < \mu < \bar{x} + z^*_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$
   Since s estimates σ, further restrict condition on sample size n: $n > 40$

Large-Sample CI for Population Mean μ

Proposition

Given any population with mean μ.
Let $x_1, \ldots, x_n$ be a large sample $(n > 40)$ taken from the population.
*Then the $100(1-\alpha)\%$ **large-sample CI for** μ is approximately*

$$\left(\overline{x} - z^*_{\alpha/2} \cdot \frac{s}{\sqrt{n}},\ \ \overline{x} + z^*_{\alpha/2} \cdot \frac{s}{\sqrt{n}}\right)$$

— OR WRITTEN MORE COMPACTLY —

$$\overline{x} \pm z^*_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

*If a **half-width** of w is desired for the $100(1-\alpha)\%$ CI yielding $(\bar{x} - w, \bar{x} + w)$, then the minimum sample size required to achieve this is:*

$$n = \left\lceil \left(\frac{z^*_{\alpha/2} \cdot s}{w}\right)^2 \right\rceil,$$ *where s is the "best guess" for the sample std deviation.*

Wald CI for Population Proportion p (Motivation)

Given any population with proportion p of some "success."
Let $\mathbf{X} := (X_1, \ldots, X_n)$ be a random sample from the population.
Let $\mathbf{x} := (x_1, \ldots, x_n)$ be a "small" sample less than 10% of population.

Then, construct the $100(1-\alpha)\%$ **Wald CI** for parameter p:

1. Produce a **pivot**: $Q(\mathbf{X};p) := \dfrac{\widehat{p} - \mu_{\widehat{p}}}{\widehat{\sigma}_{\widehat{p}}} = \dfrac{\widehat{p} - p}{\sqrt{\widehat{p}\widehat{q}/n}}$ $\quad (X \sim$ Binomial$(n, \widehat{p}))$
2. Determine pdf of pivot: Provided $\min\{n\widehat{p}, n\widehat{q}\} \geq 10$, then by the CLT,
$$X \overset{approx}{\sim} \text{Normal}(n\widehat{p}, n\widehat{p}\widehat{q}) \implies Q(\mathbf{X};p) \overset{approx}{\sim} \text{Normal}(0,1)$$
3. Find constants $a < b$ such that $\mathbb{P}(a < Q(\mathbf{X};p) < b) = 1 - \alpha$

   Since the std normal pdf is symmetric, $a = -z^*_{\alpha/2}$ and $b = z^*_{\alpha/2}$
4. Manipulate the inequalities to isolate parameter p:
$$-z^*_{\alpha/2} < \frac{\widehat{p} - p}{\sqrt{\widehat{p}\widehat{q}/n}} < z^*_{\alpha/2} \implies \widehat{p} - z^*_{\alpha/2} \cdot \sqrt{\widehat{p}\widehat{q}/n} < p < \widehat{p} + z^*_{\alpha/2} \cdot \sqrt{\widehat{p}\widehat{q}/n}$$

Wald CI for Population Proportion p

Proposition

Given any population with proportion p of some "success."
Let $x_1, \ldots, x_n$ be a sample taken from the population s.t. $\min\{n\widehat{p}, n\widehat{q}\} \geq 10$.

*Then the $100(1-\alpha)\%$ **Wald CI for** p is approximately*

$$\left(\widehat{p} - z^*_{\alpha/2} \cdot \sqrt{\widehat{p}\widehat{q}/n},\ \ \widehat{p} + z^*_{\alpha/2} \cdot \sqrt{\widehat{p}\widehat{q}/n}\right)$$

— OR WRITTEN MORE COMPACTLY —

$$\widehat{p} \pm z^*_{\alpha/2} \cdot \sqrt{\widehat{p}\widehat{q}/n}$$

where $\widehat{p} := \dfrac{X}{n} \equiv \dfrac{\text{\# Successes in Sample}}{\text{Sample Size}}$ *and* $\widehat{q} := 1 - \widehat{p}$

Unfortunately, Wald CI's turn out to be very unreliable unless p is close to 0.5.
See page 290 in the textbook for the details why.

Fortunately, a more reliable CI for p is available: Wilson Score CI

Wilson Score CI for Population Proportion p

To construct the Wilson Score CI, form the pivot using the std error of $\widehat{p}$, $\sigma_{\widehat{p}}$, rather than the estimated standard error p, $\widehat{\sigma}_{\widehat{p}}$:

Proposition

Given any population with proportion p of some "success."
Let $x_1, \ldots, x_n$ be a sample (of any size) taken from the population.

Then the $100(1-\alpha)\%$ ***Wilson score CI for*** *p is approximately*

$$\frac{n\widehat{p} + 0.5(z^*_{\alpha/2})^2}{n + (z^*_{\alpha/2})^2} \pm z^*_{\alpha/2} \cdot \frac{\sqrt{n\widehat{p}\widehat{q} + 0.25(z^*_{\alpha/2})^2}}{n + (z^*_{\alpha/2})^2}$$

where $\widehat{p} := \frac{X}{n} \equiv \frac{\text{\# Successes in Sample}}{\text{Sample Size}}$ *and* $\widehat{q} := 1 - \widehat{p}$

Textbook Logistics for Section 7.2

- Difference(s) in Notation:

CONCEPT	TEXTBOOK NOTATION	SLIDES/OUTLINE NOTATION
Probability of Event	$P(A)$	$\mathbb{P}(A)$
Estimated Std Error of $\widehat{\theta}$	$s_{\widehat{\theta}}$	$\widehat{\sigma}_{\widehat{\theta}}$
z Critical Value	$z_{\alpha/2}$	$z^*_{\alpha/2}$

- Ignore ”One-Sided Confidence Intervals” (pg 292)
  - One-sided CI’s are only occasionally encountered in applications.
  - One-Sided Confidence Intervals will not be considered in this course.
  - Going forward, mentioning a CI implicitly means a **two-sided** CI.

Fin.