Set 22 - Confidence Intervals

March 17, 2023 1:26 PM

Stat 260 Lecture Notes

Set 22 - Confidence Intervals

→ calculate 1 number

Recall: A point estimate is a single valued statistic used to estimate a population parameter. $\overline{x}$ is a point estimate for μ.

no sense of how accurate it is

sample mean — population mean

The downside of using a point estimate such as $\overline{x}$ is that we don't know how accurate our estimate is. Another method of estimation is to give a range of possible values - an interval estimate.

↳ anything in interval is good estimate

A confidence interval (CI) for μ is an interval $[L, U]$ which gives an estimate for the population mean μ with some degree of certainty.

Lower — Upper

A 95% confidence interval for μ has $P(L \leq \mu \leq U) = 0.95$

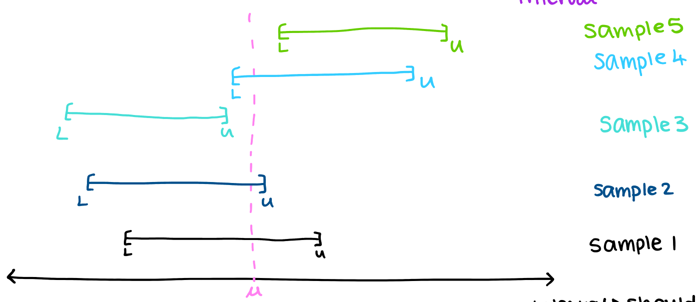
A 99% confidence interval for μ has $P(L \leq \mu \leq U) = 0.99$.

We find the numerical values for L and U by sampling. From each round of sampling we get an estimated range of values that μ might be contained in.

Interpretation: For a 95% CI, in the long run 95% of the CIs we create by sampling will actually contain μ.

↳ confidence interval

* Width of intervals not always the same

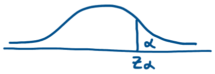


95% of these intervals should contain μ in the long run.

As long as μ in interval = good job

1

How do we find L and U?



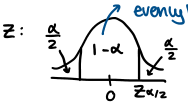
$z_{\alpha/2}$ → means same thing, will just be a number

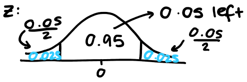
A $(1-\alpha)\%$ CI for μ is

$$[L, U] = \Big[\underbrace{\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}_{L},\ \underbrace{\bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}_{U}\Big].$$

subscript

α% leftover, distributed evenly between each side





Sometimes we write this as

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad \text{or} \quad \bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}.$$

If don't know σ can use s

estimate ± (critical value) · (standard error)

(or estimated)

↳ to get rid of ambiguity

σ = std. dev. for x

$\frac{\sigma}{\sqrt{n}}$ = std. dev. for $\bar{x}$ = standard error

$\frac{s}{\sqrt{n}}$ = estimated standard error

$\underbrace{3 \qquad\qquad 2}_{\text{e.s.e on formula sheet}}$

estimate ± (c.v.)(e.s.e)

↳ on formula sheet, this is a pattern.

When can we use this formula?

Whenever $\dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ or $\dfrac{\bar{X} - \mu}{s/\sqrt{n}}$ is normally distributed:

- (set 20) $X_1, X_2, \ldots, X_n$ from a normal distribution and we know σ
- (set 21) CLT $X_1, X_2, \ldots, X_n$ from any distribution and n is big ($n \geq 30$) and we know σ
- (set 21) $X_1, X_2, \ldots, X_n$ from any distribution and n is big ($n \geq 30$) and we don't know σ (so we use the estimate s instead)

$\bar{x}$ is normal

will be written in words → n=15 "have a sample of 15"

Example 1: $X_1, X_2, \ldots, X_{15}$ from a normal distribution with $\bar{x} = 147.33$ and $\sigma = 40$. Find a 95% confidence interval for μ and find a 99% confidence interval for μ.

Most problems will be finding confidence intervals

→ work is finding $z_{\alpha/2}$

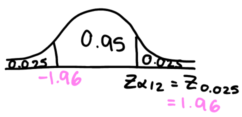
$z_{\alpha/2}$ always positive

95% CI: $\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

$= 147.33 \pm (1.96 \cdot \frac{40}{\sqrt{15}})$ order of operations matter!!!

$= [\underset{L}{127.09}, \underset{U}{167.57}]$

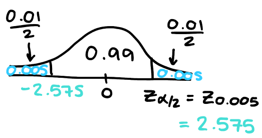
↳ need to calculate the lower and upper limits



99% CI: $\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

$= 147.33 \pm 2.575 \cdot \frac{40}{\sqrt{15}}$

$= [120.74, 173.92]$



Example 2: How does changing σ, n, and the confidence level $(1-\alpha)\%$ affect the CI?

$\bar{x}$	σ	n	confidence level	$\alpha/2$	$z_{\alpha/2}$	$[L, U]$
147.33	40	15	95%	.025	1.96	[127.09, 167.57]
★ 147.33	40	**50**	95%	.025	1.96	[136.25, 158.42]
★ 147.33	40	15	**99%**	.005	2.575	[120.74, 173.92] ← CI wider
147.33	**80**	15	95%	.025	1.96	[106.85, 187.82]

increasing the confidence level (%) ⇒ CI wider

↳ be more sure, pick wider interval, μ more likely to be in there

★ increasing n ⇒ narrower CI

increasing σ ⇒ wider CI

↳ σ is how much data is spread apart

*increasing n ⇒ narrower CI

increasing σ ⇒ wider CI

↳ σ is how much data is spread apart

3

Example 3: How did we get the formula

$$[L, U] = \left[\overline{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \ \ \overline{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right] \quad ?$$

Omit

Idea: $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ rearrange and solve for μ.

Look at the 95% confidence interval, where we have $z_{\alpha/2} = z_{0.025} = 1.96$. We can say

$P(L \leq \mu \leq U) = 0.95$ this is okay!

$$P(\overline{x} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}) = 0.95$$ this is also okay!

but once we fill in values for $\overline{x}$, σ, and $\sqrt{n}$ we can no longer use this probability.

this is not okay!!! not a variable

So for example we can't say $P(127.09 \leq \mu \leq 167.57) \neq 0.95$ with our confidence interval $[127.09, 167.58]$ because either μ is in this interval, or it isn't. (There are no variables here so there is no chance on where the value of μ sits. The probability is 1 if the value is in the interval, or 0 if it isn't.)

4