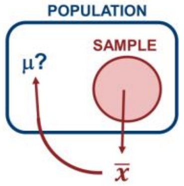
Introduction to confidence intervals

Estimating the population mean

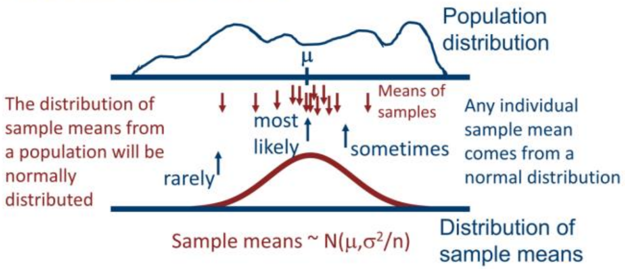


We want to know the population mean.
We can't measure the whole population.
We take a random sample.
We calculate a sample mean.

The sample mean is an estimate of the population mean, but sampling error makes it inexact.

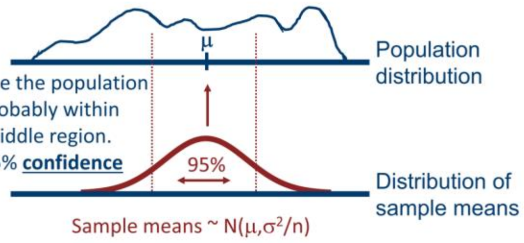
What can we say about the likely population mean, based on the sample mean?

The Central Limit Theorem



Confidence intervals for the population mean

95% chance the population mean is probably within the 95% middle region.
This is a 95% **confidence interval**.

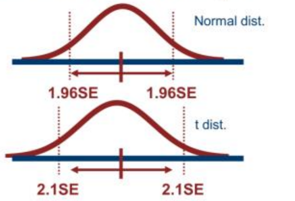


Confidence interval reality

Unfortunately in the real-world we would never have the population variance. We must estimate it from the sample variance.

To adjust for sampling error, we use the t-distribution instead of the normal.

The t-distribution is wider to account for the sampling error.



Confidence interval example (part 1)

e.g., if we take a sample of 16 values, $\bar{x} = 15, \sigma = 5$

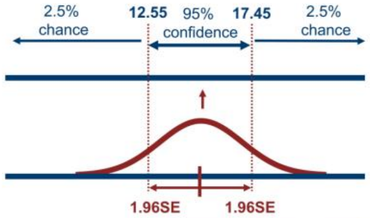
The 95% confidence interval is $\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}$

Which is

$15 \pm (1.96)\frac{5}{\sqrt{16}}$

15 ± 2.45

$\{12.55, 17.45\}$



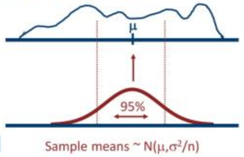
Calculating the confidence interval for the population mean

Take a sample of n values.

Calculate the mean, variance, and standard deviation.

Get middle region of the standard normal distribution corresponding to the degree of confidence desired.

This tells the width of the confidence interval in terms of number of standard errors above and below the sample mean



Standard error vs standard deviation

Population values ~ N(μ,σ²)
Sample values ~ N($\bar{x}$,s²)

Standard deviation describes the spread of the data values

Sample means ~ N(μ,σ²/n) $\quad s_{\bar{x}}^2 = \frac{\sigma^2}{n}$ therefore $s_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

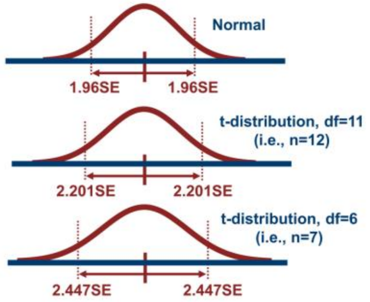
Standard error, $s_{\bar{x}}$, describes the spread of the sample means and, likewise, the spread of possible population means.

The t-distribution

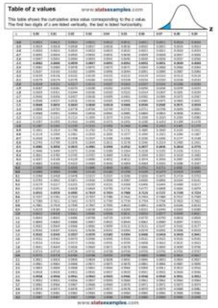
The t-distribution is wider to account for the sampling error.

As sample size increases, it narrows to become the normal.

There are t-tables like Z tables

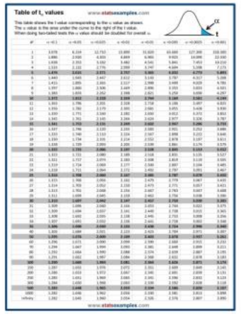


Z and t tables



Z tables *usually* describe area to the left (versatility).

t tables *usually* describe α area on outside (for confidence intervals) and have rows for each df (n-1)



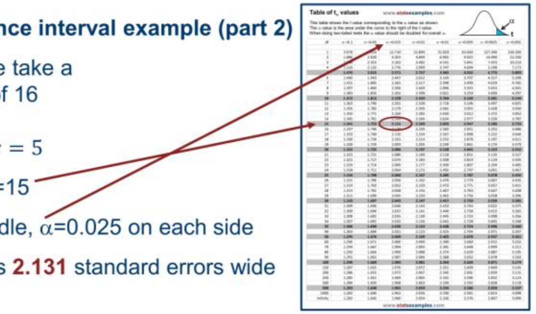
Confidence interval example (part 2)

e.g., if we take a sample of 16 values,
$\bar{x} = 15, s = 5$

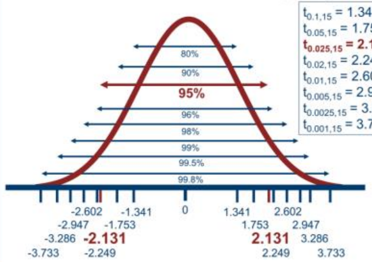
n=16, df=15

95% middle, α=0.025 on each side

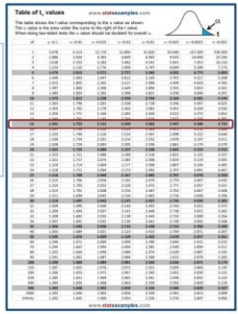
Interval is **2.131** standard errors wide



Confidence interval example (part 2)

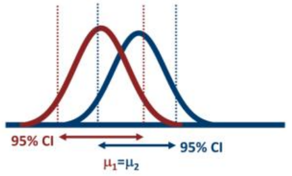


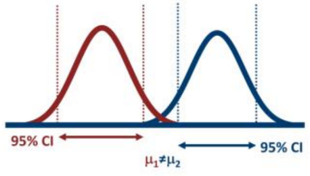
$t_{0.1,15} = 1.341$
$t_{0.05,15} = 1.753$
$t_{0.025,15} = 2.131$
$t_{0.02,15} = 2.249$
$t_{0.01,15} = 2.602$
$t_{0.005,15} = 2.947$
$t_{0.0025,15} = 3.286$
$t_{0.001,15} = 3.733$



What are confidence intervals used for?

- As a descriptive statistic for our estimate of the population mean.
- As the basis for the t-test, a test of whether two populations appear to have different means.



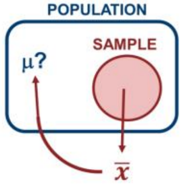


Which method to use to calculate CI

If we know μ and σ
No stats needed

If we know $\bar{x}$ and σ
Use normal distribution (unrealistic)

If we know $\bar{x}$ and *s*
Use t-distribution



Confidence intervals are the real-world version of significant figures

Confidence interval (CI) comparison

If we take a sample of 16 values, $\bar{x} = 15, s = 5$

Using the normal distribution:
The 95% CI: $\bar{x} \pm \mathbf{1.96}\frac{s}{\sqrt{n}}$, $\quad 15 \pm 2.45$, $\quad \{12.55, 17.45\}$

Using t-distribution:
The 95% CI: $\bar{x} \pm \mathbf{2.131}\frac{s}{\sqrt{n}}$, $\quad 15 \pm 2.66$, $\quad \{12.34, 17.66\}$

We should always use the t-distribution unless the sample size is large (e.g., df=25 gives ~5% error, 2.06 vs 1.96).

Confidence interval example (part 2)

e.g., if we take a sample of 16 values, $\bar{x} = 15, s = 5$

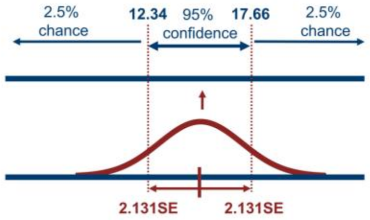
The 95% confidence interval is $\bar{x} \pm \mathbf{2.131}\frac{s}{\sqrt{n}}$

Which is

$15 \pm (2.131)\frac{5}{\sqrt{16}}$,

15 ± 2.66,

$\{12.34, 17.66\}$



StatsExamples.com