POLS0008: Introduction to Quantitative Research Methods
Week 6 Seminar: Confidence Intervals

James Rice

UCL Department of Political Science

- **Lecture Recap (15 min):** what a CI is, how to compute it, how to interpret it
- **Group work:** Seminar tasks Q1-Q4
- **Review:** Step-by-step answers

Why we need confidence intervals

Goal: learn about an unknown population parameter (e.g., the true mean change in blood pressure after an intervention, or the true share of people who improved).

From population to sample: we usually cannot measure everyone, so we take a *random sample*. A sample gives us a **point estimate**:

$$\bar{x} \text{ (sample mean)}, \qquad \hat{p} \text{ (sample proportion)}$$

Key fact: even with the same population, different random samples produce different $\bar{x}$ and $\hat{p}$. This is **sampling variability**.

Confidence intervals (CIs) convert a point estimate into a *range* that reflects sampling variability:

- narrow CI $\Rightarrow$ precise estimate (less sampling noise)
- wide CI $\Rightarrow$ imprecise estimate (more sampling noise)

Bottom line: CIs quantify uncertainty from sampling, under explicit statistical assumptions.

The general CI recipe

Most CIs follow the same structure:

$$\text{estimate} \;\pm\; (\text{critical value}) \times (\text{standard error})$$

Standard error (SE): the estimated SD of the estimator across repeated samples.

Critical value: how many SEs we go out from the estimate to achieve a target long-run coverage rate.

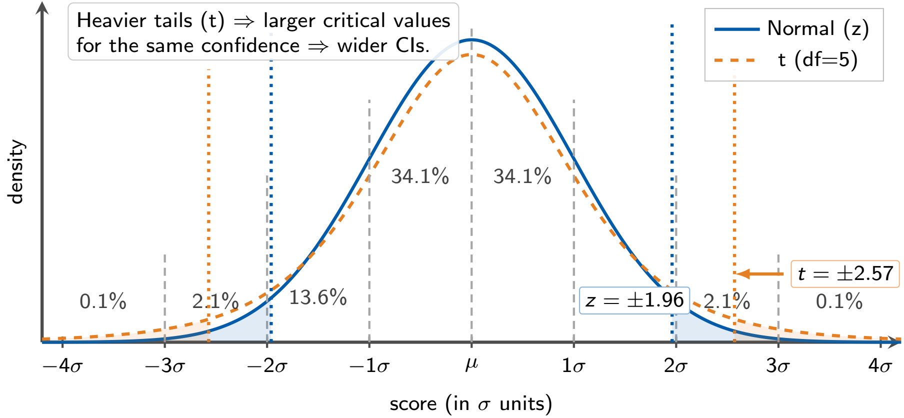
$$95\%: \; z_{0.025} \approx 1.96 \qquad 99\%: \; z_{0.005} \approx 2.576$$

z critical values

A z-score is "how many standard deviations (σ) from the mean (μ)". On the normal distribution, areas under the curve correspond to probabilities.

- Rough landmarks (the **68–95–99.7 rule**): $\mu \pm 1\sigma$ contains about **68**% of observations (about 34.1% on each side of μ), $\mu \pm 2\sigma$ contains about **95**% of observations, while $\mu \pm 3\sigma$ contains about **99.7**% of observations.
    - A **95**% CI uses $z_{0.025} \approx$ **1.96**: we go about $\pm 2\sigma$ from the estimate (leaving $\approx 5\%$ in the two tails).
    - A **99**% CI uses $z_{0.005} \approx$ **2.576**: wider interval (leaving $\approx 1\%$ in the two tails).

t and z visualization



Each curve is on a **standard-deviation scale** (so 1σ is one SD from the mean μ); for a **two-tailed** $\alpha = 0.05$ CI the normal uses $z = \pm 1.96$, while with small df the heavier-tailed t distribution needs a **larger critical value** (e.g., $t = \pm 2.57$, which corresponds to **two-tailed** $\alpha = 0.01$), so t-based confidence intervals are wider.

Interpreting a CI

If we repeatedly draw samples *the same way* and compute a CI each time, then *95% of those intervals will contain the true parameter*.

Why this matters: the parameter is fixed; the interval is random (it changes from sample to sample).

Avoid this common phrasing (not the standard interpretation):
"There is a 95% probability the true value lies in this specific interval."

Decision-style heuristic (not a magic rule):

- Mean difference: CI excluding 0 suggests evidence for a non-zero mean effect.
- Proportion: CI excluding a benchmark (e.g., 0.5) suggests evidence it differs from that benchmark.

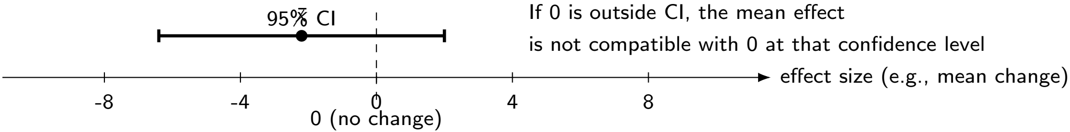
CI for a mean

Mean CI (large sample / normal approximation):

$$\text{CI for } \mu: \quad \bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$$

What each piece means:

- $\bar{x}$: best estimate of the population mean μ from this sample.
- s: spread of individual outcomes in the sample (more spread $\Rightarrow$ more uncertainty about μ).
- $\frac{s}{\sqrt{n}}$: uncertainty of the *mean* (averaging reduces noise by $\sqrt{n}$).
- $z_{\alpha/2}$: chosen to hit 95% or 99% long-run coverage under the normal approximation.



CI for a proportion

Proportion CI (normal approximation):

$$\text{CI for } p: \quad \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Interpretation target: the unknown population proportion p (not $\hat{p}$). $\hat{p}$ is our estimate; the CI says how much $\hat{p}$ would vary across repeated samples.

Why the term $\hat{p}(1-\hat{p})$ appears:

- It is the estimated variance of a Bernoulli outcome (success/failure).
- It peaks at 0.25 when $\hat{p} = 0.5 \Rightarrow$ maximum uncertainty around 50/50.

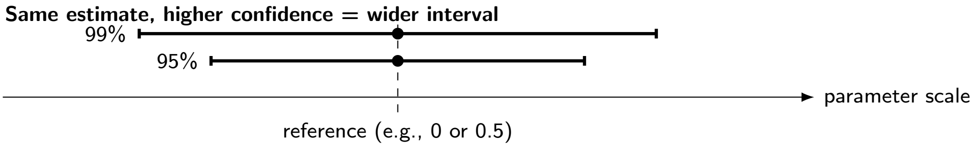
z vs t (and why 99% CIs get wider)

For means, why do we sometimes replace z with t?

When the population SD σ is unknown (typical), we estimate it with s. That adds uncertainty, especially for small n.

$$\text{CI for } \mu: \quad \bar{x} \pm t_{\alpha/2,\, n-1} \frac{s}{\sqrt{n}}$$

- df $= n - 1$: with low df, t has heavier tails $\Rightarrow$ larger critical values.
- As n increases, $t_{\alpha/2,n-1} \to z_{\alpha/2}$.
- Higher confidence (99% vs 95%) $\Rightarrow$ larger critical value $\Rightarrow$ wider CI (more coverage requires more width).



Seminar tasks

Q1. You sample 120 people and measure their blood pressure before and after an intervention and find that the mean change is -5.09 with a standard deviation of 16.71. Find the 95% and 99% confidence intervals of the mean change. Does the intervention reduce blood pressure in the population?

Q2. In the same sample of 120 people you find that 61% showed a decrease in blood pressure. Find the 95% and 99% confidence intervals of the proportion. Does this confirm the effect of the intervention on blood pressure?

Q3. A random sample of 25 countries finds that the mean life expectancy is 64 years for women with a standard deviation of 21 years and 62 years for men with a standard deviation of 14 years. Using a t-distribution, find the 95% and 99% confidence intervals of the mean for men and for women. Interpret your results.

Q4. In your own words, describe the difference between a t-score and a z-score. When do we use a normal distribution and when do we use a t-distribution? At about what number of observations does the difference in the distributions become meaningless?

Answers — Q1 and Q2

Q1 (mean change; z): $n = 120,\ \bar{x} = -5.09,\ s = 16.71.$

$$SE = \frac{16.71}{\sqrt{120}} \approx 1.525$$

$$95\%:\ -5.09 \pm 1.96(1.525) \Rightarrow [-8.08,\ -2.10] \qquad 99\%:\ -5.09 \pm 2.576(1.525) \Rightarrow [-9.02,\ -1.16]$$

Interpretation: both exclude 0 ⇒ consistent with a population reduction.

Q2 (proportion; z): $n = 120,\ \hat{p} = 0.61.$

$$SE = \sqrt{\frac{0.61(0.39)}{120}} \approx 0.0445$$

$$95\%:\ 0.61 \pm 1.96(0.0445) \Rightarrow [0.52,\ 0.70] \qquad 99\%:\ 0.61 \pm 2.576(0.0445) \Rightarrow [0.50,\ 0.72]$$

Interpretation: plausible true majority; at 99% it's borderline vs 0.5.

t-table: df $= 24$, two-tailed $\alpha = 0.05$ and 0.01

For confidence intervals we typically use **two-tailed** critical values:

$$t_{\alpha/2,\ \text{df}} \qquad (95\% :\ \alpha = 0.05 \Rightarrow \alpha/2 = 0.025;\quad 99\% :\ \alpha = 0.01 \Rightarrow \alpha/2 = 0.005)$$

df	$\alpha = 0.10$ ($t_{0.05}$)	$\alpha = 0.05$ ($t_{0.025}$)	$\alpha = 0.02$ ($t_{0.01}$)	$\alpha = 0.01$ ($t_{0.005}$)
20	1.725	2.086	2.528	2.845
21	1.721	2.080	2.518	2.831
22	1.717	2.074	2.508	2.819
23	1.714	2.069	2.500	2.807
24	**1.711**	**2.064**	**2.492**	**2.797**
25	1.708	2.060	2.485	2.787

Answers — Q3 and Q4

Q3 (means; t, df=24): $\sqrt{25} = 5$.
Women: $\bar{x} = 64,\ s = 21,\ SE = 21/5 = 4.2$.

$$95\% :\ 64 \pm 2.064(4.2) = [55.33, 72.66], \quad 99\% :\ 64 \pm 2.797(4.2) = [52.25, 75.75]$$

Men: $\bar{x} = 62,\ s = 14,\ SE = 14/5 = 2.8$.

$$95\% :\ 62 \pm 2.064(2.8) = [56.22, 67.78], \quad 99\% :\ 62 \pm 2.797(2.8) = [54.17, 69.83]$$

Q4 (concept):

- **z** uses the normal distribution; typical for large n (or known σ).
- **t** uses df-based critical values; typical when σ unknown and n smaller.
- t has heavier tails $\Rightarrow$ slightly wider CIs; by $\sim n = 30$ difference often small.