



Confidence Intervals

Confidence Interval Dynamics

1. Exploring the role of sample size With 95% level ($\alpha = 0.05$) and $s = 0.9$, the half-width is

$$\text{half-width} = t^* \frac{s}{\sqrt{n}}, \quad t^* = t_{0.025, df}, \quad df = n - 1.$$

Using standard t -quantiles:

- $n = 3$ ($df = 2$): $t^* \approx 4.303 \Rightarrow \text{half-width} \approx 1.295$
- $n = 5$ ($df = 4$): $t^* \approx 2.776 \Rightarrow \text{half-width} \approx 1.118$
- $n = 10$ ($df = 9$): $t^* \approx 2.262 \Rightarrow \text{half-width} \approx 0.643$
- $n = 15$ ($df = 14$): $t^* \approx 2.145 \Rightarrow \text{half-width} \approx 0.498$
- $n = 20$ ($df = 19$): $t^* \approx 2.093 \Rightarrow \text{half-width} \approx 0.421$
- $n = 30$ ($df = 29$): $t^* \approx 2.045 \Rightarrow \text{half-width} \approx 0.336$

2. Confidence level and t^*

- a) Fix $n = 10$ ($df = 9$), $s = 0.9$; two-tailed t^* :

- 80%: $t_{0.10,9} \approx 1.383$
- 90%: $t_{0.05,9} \approx 1.833$
- 95%: $t_{0.025,9} \approx 2.262$
- 99%: $t_{0.005,9} \approx 3.250$

- b) Higher confidence \Rightarrow smaller $\alpha \Rightarrow$ fatter central region \Rightarrow larger quantile t^* ; hence $2 t^* \frac{s}{\sqrt{n}}$ increases with the level.

- c) Width ratio (same n, s) is the ratio of the two values of t^* :

$$\frac{\text{width}_{99\%}}{\text{width}_{90\%}} = \frac{t_{0.005,9}}{t_{0.05,9}} \approx \frac{3.250}{1.833} \approx 1.77.$$

3. Trade-off between variability and sample size

a) At 95%, half-width = $t^* \frac{s}{\sqrt{n}}$.

- $(n, s) = (8, 0.6)$: $df = 7$, $t^* \approx 2.365$, so half-width $\approx 2.365 \cdot \frac{0.6}{\sqrt{8}} \approx 0.502$

- $(n, s) = (20, 1.0)$: $df = 19$, $t^* \approx 2.093$, so half-width $\approx 2.093 \cdot \frac{1.0}{\sqrt{20}} \approx 0.468$

Thus $(8, 0.6)$ is (slightly) wider—smaller n and larger t^* outweigh the smaller s .

b) Keep $s = 0.9$, level 95%. Need $t^* \frac{0.9}{\sqrt{n}} \leq 0.30$. Using $t_{0.025, n-1}$, the smallest n that satisfies

this is about $n \approx 37$ (since $t_{0.025, 36} \approx 2.03$ gives $2.03 \cdot 0.9/\sqrt{37} \approx 0.30$).

c) Hold $n = 12$ ($df = 11$), level 90% so $t^* = t_{0.05, 11} \approx 1.796$. Full width = 0.9 means

$$2 t^* \frac{s}{\sqrt{12}} = 0.9 \Rightarrow s = \frac{0.9 \sqrt{12}}{2 t^*} \approx \frac{0.9 \cdot 3.464}{2 \cdot 1.796} \approx 0.868.$$

Calculations

1. A researcher is analysing individuals' relative fear of being a victim of burglary on a 1-100 scale. A random sample of 9 individuals found a mean score of 47 on the scale with a sample variance of 158.76 for fear of being burgled.

a. What distribution would be used to calculate an 80% confidence interval around this mean?

A t-distribution as we don't know the population standard deviation and n is small.

b. Construct that interval.

$$\bar{x} = 47$$

$$n = 9$$

$$t \text{ from tables} = 1.397$$

$$s = \sqrt{158.76}$$

$$s = 12.6$$

Confidence interval formula

$$\bar{x} \pm t \times \frac{s}{\sqrt{n}}$$

Lower bound

$$47 - 1.397 \times \frac{12.6}{\sqrt{9}} = 47 - 5.867 = 41.13$$

Upper bound

$$47 + 1.397 \times \frac{12.6}{\sqrt{9}} = 47 + 5.867 = 52.87$$

2. We are investigating the height of men in the UK. For this we have obtained a random sample of 100 UK men and found they had a mean height of 180cm with a standard deviation of 10cm.
- a. Construct a 95% confidence interval for the mean height of UK males.

$$\bar{x} = 180$$

$$s = 10$$

$$n = 100$$

As the population standard deviation is not known, the t distribution and t need to be used.

Find the t-score for a 95% confidence interval in the t-table with 99 df.

$$t = 1.984$$

Confidence interval:

$$\bar{x} \pm t \times \frac{s}{\sqrt{n}}$$

Lower bound:

$$180 - 1.984 \times \frac{10}{\sqrt{100}} = 180 - 1.984 = 178.02$$

Upper bound:

$$180 + 1.984 \times \frac{10}{\sqrt{100}} = 181.98$$

- b. Select all true statements concerning the constructed confidence interval and justify your choice for each statement.
- The probability of the population mean being within the upper and lower bounds is 95%.
FALSE - The population mean is fixed but unknown and therefore can either be inside the bounds or outside. The Probability is therefore 50%.
 - 95% of men's heights fall between the upper and lower bound.
FALSE - The distribution calculated is not the distribution of men's height, but the sampling distribution of the mean male height.
 - 95% of the cases in the sample fall between the upper and lower bound.
FALSE - The distribution calculated is not of men's height in this sample, but the sampling distribution of the mean male height.
 - On average 95% of confidence intervals constructed would contain the population mean.
TRUE

- v. On average 95% of the means of samples with 100 respondents will fall within the upper and lower bands.
FALSE - This confidence interval is not making statements about various sample means but rather about the population mean.
- vi. On average 95% of the sample means equal the population mean.
FALSE - The confidence interval is a range and does not make claims about where the population mean is exactly.

Significance Testing

Exercises – The t-distribution

1. How df changes the shape and critical values

Using two-tailed, $\alpha = 0.05$ (so $t_{\text{crit}} = t_{0.025, df}$ with $df = n - 1$):

- $n = 3$ ($df = 2$): $t_{\text{crit}} \approx 4.303$
- $n = 5$ ($df = 4$): $t_{\text{crit}} \approx 2.776$
- $n = 10$ ($df = 9$): $t_{\text{crit}} \approx 2.262$
- $n = 20$ ($df = 19$): $t_{\text{crit}} \approx 2.093$
- $n = 50$ ($df = 49$): $t_{\text{crit}} \approx 2.009$

The normal reference is $z_{0.025} \approx 1.960$.

Within 0.05 of 1.96 occurs by $n \approx 50$ ($df = 49$, 2.009 is 0.049 above 1.96).

Explanation: as $df \rightarrow \infty$, the t distribution $\rightarrow N(0, 1)$, so tails thin and t_{crit} decreases toward 1.96.

2. One-tailed vs two-tailed critical regions

With $df = 14$ and $\alpha = 0.05$:

- Two-tailed: $t_{0.025, 14} \approx 2.145$, cut points at ± 2.145
- One-tailed (upper): $t_{0.05, 14} \approx 1.761$, single cut point at $+1.761$

For $H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0$, the one-tailed threshold $t_{0.05, 14}$ is relevant because only the upper tail provides evidence against H_0 .

In both cases, the shaded area equals α :

- two-tailed shades two symmetric regions each of area $\alpha/2$;
- one-tailed shades a single upper tail of area α .

Effect size, sample size, and power

1. Same effect size, different n (mapping t , n , and \hat{d})
 - a. With $t = 1.6$ and $n = 16$, $\hat{d} = t/\sqrt{n} = 1.6/4 = 0.40$. The power at $n = 16$ for $\hat{d} \approx 0.40$ and $\alpha = 0.05$ will be modest (typically well below 0.80).
 - b. Keeping \hat{d} fixed at 0.40 and raising n to 64 requires $t = \hat{d}\sqrt{n} = 0.40 \times 8 = 3.2$. The app will show the same \hat{d} but a higher power at $n = 64$.
 - c. As n increases while the underlying effect stays fixed, the standard error shrinks: $SE = s/\sqrt{n}$. The test statistic is $t = \frac{\bar{X} - \mu_0}{SE}$, so a smaller SE makes $|t|$ larger on average. For a fixed α , the critical cutoff (e.g., $t_{\alpha/2, df}$ for a two-sided test) is essentially fixed, so larger typical $|t|$ increases the chance that $|t| > t_{\alpha/2, df}$. Therefore, bigger $n \Rightarrow$ less sampling noise \Rightarrow tighter estimates \Rightarrow higher power.
2. Planning with a SESOI and α sensitivity
 - a. For $t = 2.0$ and $n = 30$, $\hat{d} = t/\sqrt{n} \approx 2.0/\sqrt{30} \approx 0.37$. The app will report the power at $n = 30$ using this \hat{d} (typically moderate at $\alpha = 0.05$).
 - b. Turning on the SESOI with $d = 0.5$ switches the curve to a fixed target effect. The orange marker shows the n giving 80% power at $\alpha = 0.05$; for $d = 0.5$ this is typically in the few-dozen range for a one-sample t -test (on the order of the 30s).
 - c. Increasing α to 0.10 lowers the critical threshold and reduces the required n for a given power; decreasing α to 0.01 raises the threshold and increases the required n . Formally, stricter α increases the critical t value, so a larger n is needed for the same probability of exceeding it under the alternative.
3. Was the study well powered? Post-hoc check and replication planning
 - a. With $t = 2.1$ and $n = 25$, $\hat{d} = 2.1/\sqrt{25} = 2.1/5 = 0.42$. The app will show the power at $n = 25$ for $\hat{d} \approx 0.42$ and $\alpha = 0.05$; this is typically only moderate, not comfortably high.
 - b. For replication planning, turn on SESOI and choose the closest option to \hat{d} (e.g., $d = 0.5$). The orange marker will give the n needed for 80% power at the chosen α ; expect a sample size in the few-dozen range for $d = 0.5$ at $\alpha = 0.05$.
 - c. A significant result with low power can be fragile because small shifts in sampling may miss the effect and estimates are noisier. Picking a SESOI (a target d tied to practical importance) and planning for 80% power helps ensure a replication has adequate sensitivity to detect a meaningfully sized effect.

Summarising Conclusions

- Power is the probability your test will detect a real effect (reject H_0 when the effect truly exists).
- The curve fixes an effect size d and shows how power increases with n .
- If you base d on your observed test ($\hat{d} = t/\sqrt{n}$), the curve answers: "If the effect really was the size we observed, how likely would the test be to detect that effect at other sample sizes n ?"
- If you base d on a meaningful minimum (SESOI), the curve answers: "How large should n be to reliably detect an effect that actually matters?"
- The 0.80 (80%) line is a convention (not a law). Many fields use it as a planning minimum; higher power (e.g., 90%) is desirable when stakes are high.
- Rule of thumb: because $t = d\sqrt{n}$, halving d requires about 4× the sample size to keep power roughly the same.

Why do many plots show an 80% power line?

- 80% power means choosing $\beta = 0.20$ (so power = $1 - \beta = 0.80$). This is a planning convention, not a mathematical law.
- The convention became common through methodological guidance (e.g., Cohen's power analysis texts) and is echoed in many applied fields.
- In higher-stakes settings (e.g., confirmatory clinical trials), targets of 80–90% power are typical; 90

Applied Exercises

See RScript in the [Online Companion](#)