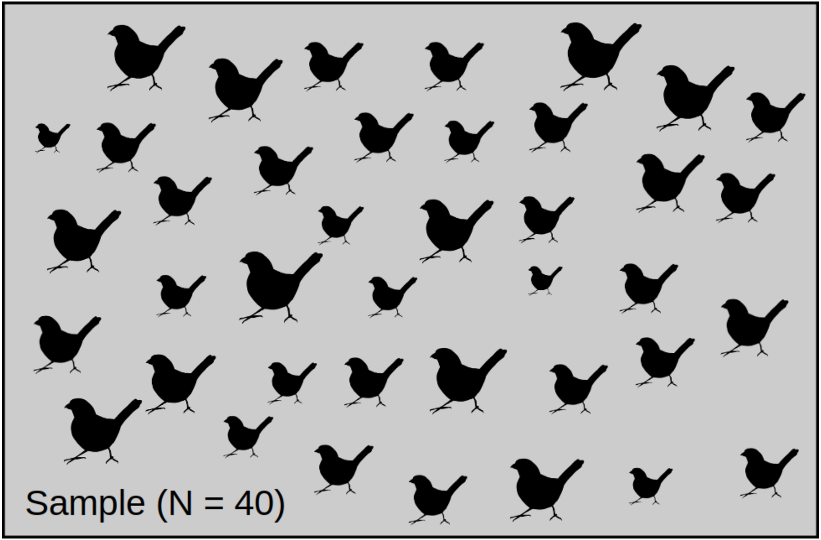
z- and t-intervals

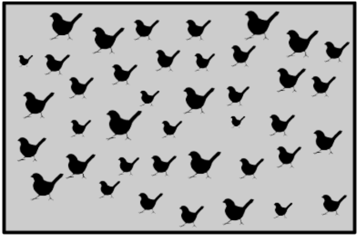
The goal of CIs is to contain uncertainty:

- How frequently will the CIs that we draw *contain* the population mean (μ_X)?
- Logic is again based on repeated resampling

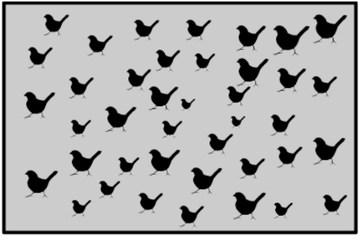
Consider the distribution of a sample of N = 40



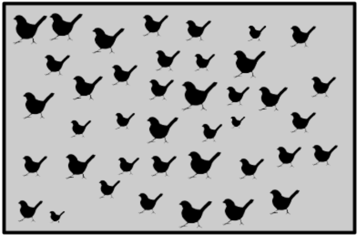
Sample means given true mean of 7.2 when N = 40



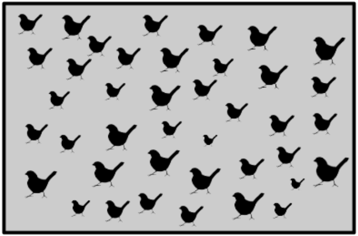
Sample mean = 6.59 cm



Sample mean = 7.45 cm

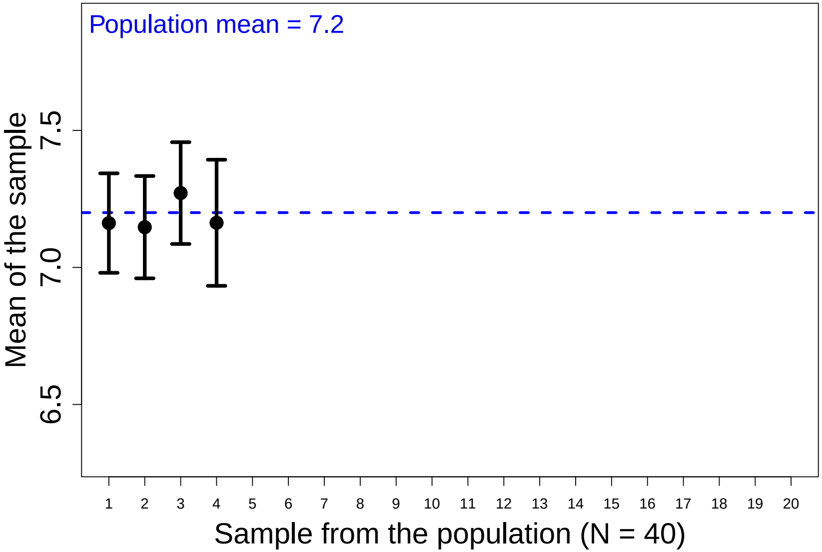


Sample mean = 7.03 cm

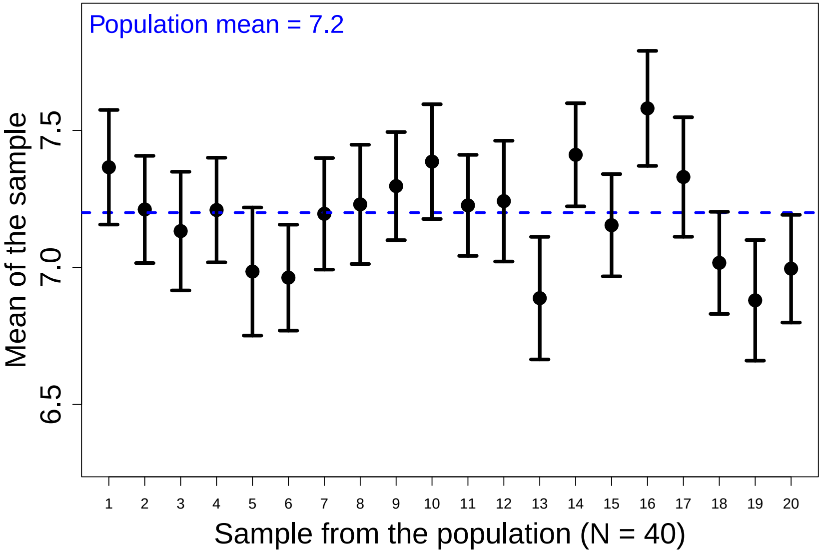


Sample mean = 6.50 cm

Confidence in a sampled mean value (80%)



Confidence in a sampled mean value (80%)



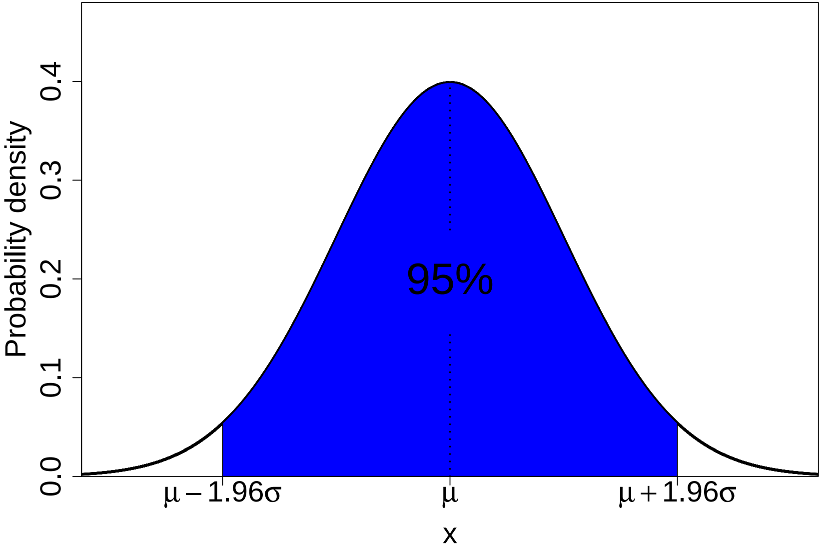
The goal of CIs is to contain uncertainty:

- How frequently will the CIs that we draw *contain* the population mean (μ_x)?
- Distribution probabilities (e.g., z-scores)
- **The standard error ($s/\sqrt{N}$)**

Calculating CIs

- $\text{LCI} = \bar{x} - z\left(\frac{s}{\sqrt{N}}\right)$
- $\text{UCI} = \bar{x} + z\left(\frac{s}{\sqrt{N}}\right)$

Probabilities calculated from the normal distribution



[1] https://bradduthie.github.io/stats/app/zandp/

Really assuming: $\mathbf{s} = \sigma$

- $\mathrm{LCI} = \bar{x} - z\left(\frac{\sigma}{\sqrt{N}}\right)$
- $\mathrm{UCI} = \bar{x} + z\left(\frac{\sigma}{\sqrt{N}}\right)$

Problem assuming: $\mathbf{s} = \sigma$

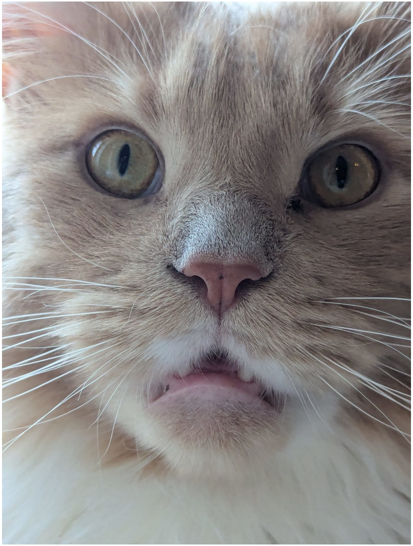
- Assuming no uncertainty in estimation of standard deviation (σ)
- If uncertainty in σ, our CIs will be wrong

Problem assuming: $s = \sigma$

- Problem less relevant[1] for large ($N > 30$) sample sizes
- Low sample size increases uncertainty of standard deviation

[1] Sokal, RR & FJ Rohlf. 1995. Biometry. 3rd ed. WH Freeman & Company, New York, USA.

Confidence in a sampled mean value



Uncertainty in the standard deviation?

Need to adjust the normal distribution

The normal distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

[1] You do not need to know this equation.

The t-distribution

$$f(t) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi v}\,\Gamma\left(\frac{v}{2}\right)}\left(1 + \frac{t^2}{v}\right)^{-\frac{v+1}{2}}$$

[1] Forget the equation, see: https://bradduthie.github.io/stats/app/t_score/

The t-distribution

$$f(t) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi v}\,\Gamma\left(\frac{v}{2}\right)} \left(1 + \frac{t^2}{v}\right)^{-\frac{v+1}{2}}$$

[1] Forget the equation, see: https://bradduthie.github.io/stats/app/t_score/

Confidence intervals with a normal distribution

Verbally

- LCI $=$ Mean – (z-score $\times$ standard error)
- UCI $=$ Mean $+$ (z-score $\times$ standard error)

Mathematically

- $\text{LCI} = \bar{x} - z\left(\frac{s}{\sqrt{N}}\right)$
- $\text{UCI} = \bar{x} + z\left(\frac{s}{\sqrt{N}}\right)$

[1] https://bradduthie.github.io/stats/app/zandp/

Confidence intervals with a t-distribution

Verbally

- LCI = Mean – (t-score × standard error)
- UCI = Mean + (t-score × standard error)

Mathematically

- $\text{LCI} = \bar{x} - t\left(\frac{s}{\sqrt{N}}\right)$
- $\text{UCI} = \bar{x} + t\left(\frac{s}{\sqrt{N}}\right)$

[1] https://bradduthie.github.io/stats/app/t_score/

Confidence intervals with normal distribution

x_1	x_2	x_3	x_4	x_5	x_6	x_7
17.1	15.2	14.9	12.6	15.2	10.3	12.7

$N = 7 \quad \bar{x} = 14 \quad s = 2.26 \quad z = 1.96 \quad df = \infty$

$$LCI = \bar{x} - z\left(\frac{s}{\sqrt{N}}\right)$$

$$= 14 - 1.96\left(\frac{2.26}{\sqrt{7}}\right)$$

$$= 12.34$$

Confidence intervals with t-distribution

x_1	x_2	x_3	x_4	x_5	x_6	x_7
17.1	15.2	14.9	12.6	15.2	10.3	12.7

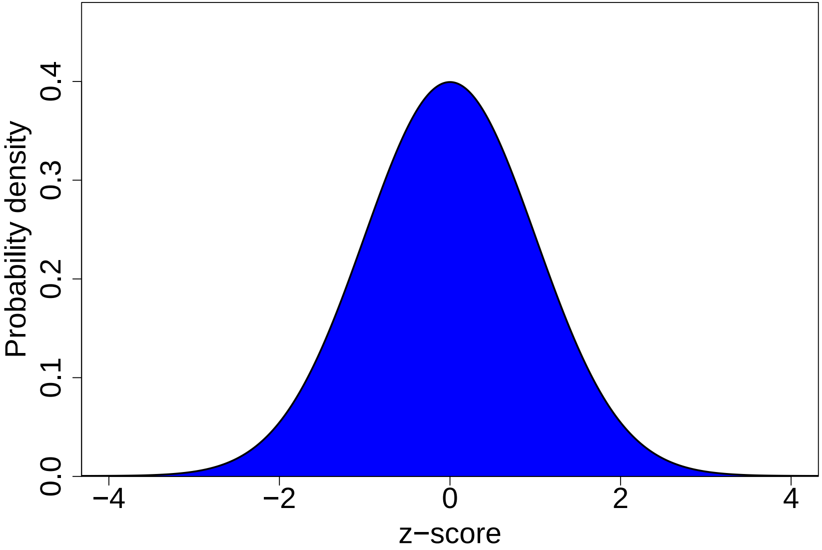
$N = 7 \quad \bar{x} = 14 \quad s = 2.26 \quad t = 2.45 \quad df = 6$

$$LCI = \bar{x} - t\left(\frac{s}{\sqrt{N}}\right)$$

$$= 14 - 2.45\left(\frac{2.26}{\sqrt{7}}\right)$$

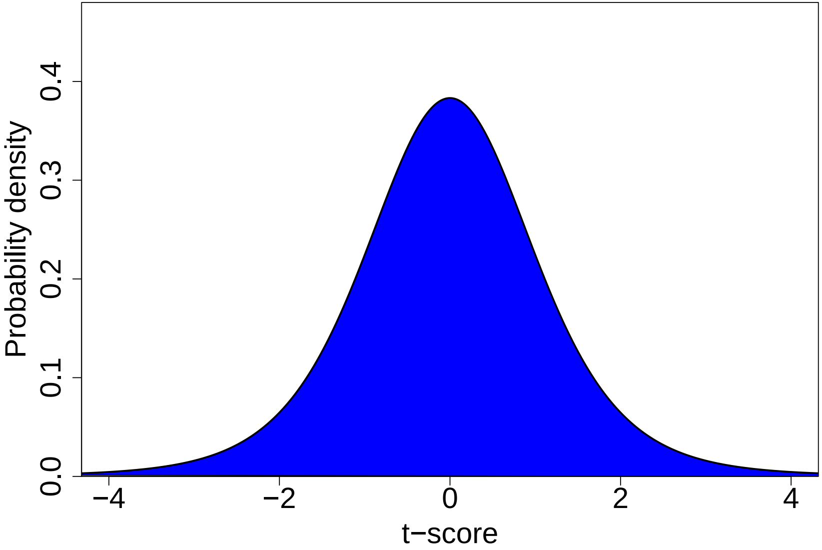
$$= 11.91$$

Standard normal distribution



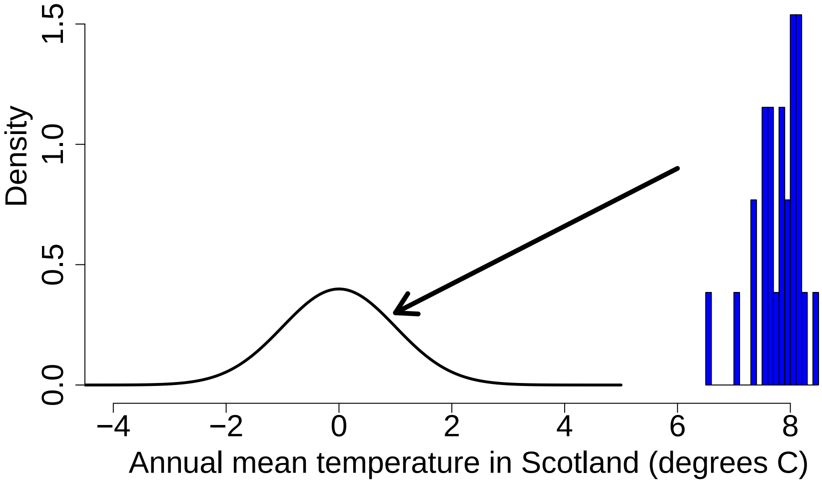
[1] https://bradduthie.github.io/stats/app/zandp/

The t-distribution (6 df)



[1] https://bradduthie.github.io/stats/app/t_score/

Turning a value into z-score



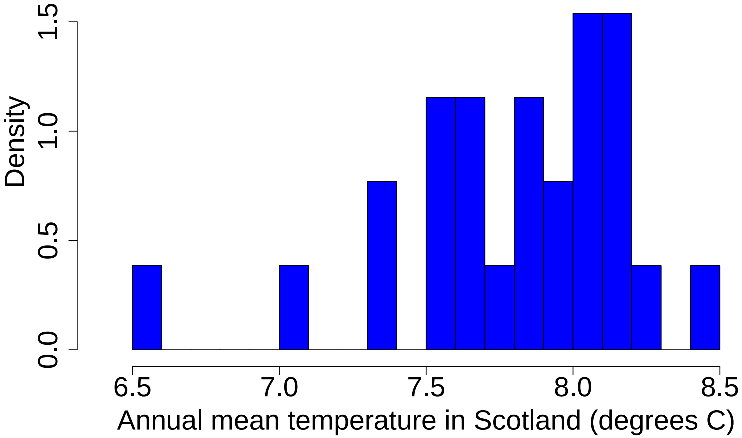
Translating a variable into a standard normal distribution

- Subtract mean
- Divide by standard deviation

$$z = \frac{x - \bar{x}}{s}$$

Can turn x-values into z-scores

What if the population mean is really 8.0?



$N = 26 \qquad \bar{x} = 7.789 \qquad s = 0.4124844$

[1] Met Office. (July 1, 2021). Annual mean temperature in Scotland from 1995 to 2020. In Statista. Retrieved February 10, 2026, from https://www.statista.com/statistics/367853/scotland-average-temperature/

Testing a hypothesis with the t-distribution

- Subtract hypothesised mean
- Divide by standard deviation

$$t = \frac{\bar{x} - \mu}{SE(x)}$$

If μ is the population mean, what is the probability of getting $\bar{x}$?

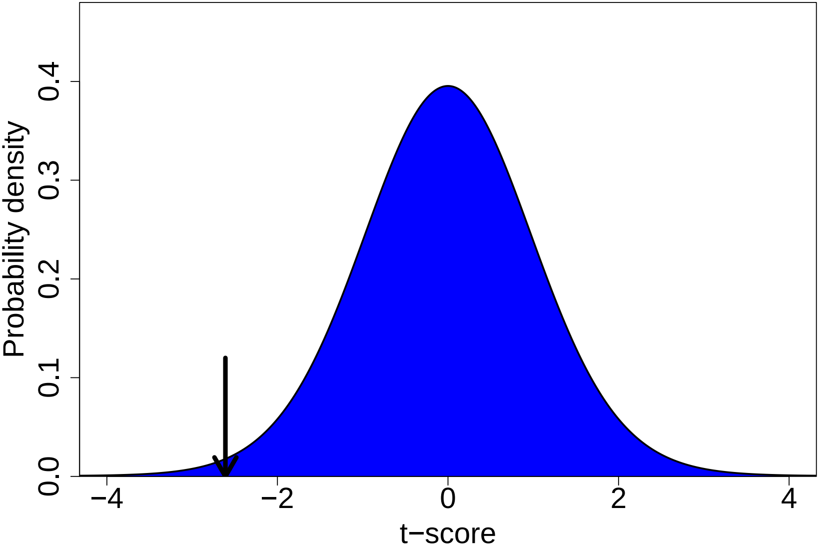
Testing a hypothesis with the t-distribution

- Subtract hypothesised mean
- Divide by standard deviation

$$t = \frac{7.789 - 8}{(0.4124844/\sqrt{26})} = -2.61$$

If 8 is the population mean, what is the probability of getting 7.789?

The t-distribution (25 df)



[1] https://bradduthie.github.io/stats/app/t_score/