Confidence intervals

If we have enought samples (enough to be condifent that the sampling distribution of the mean is normal), then the confidence interval is:

$$\text{CI} = \bar{X} \pm (q \times \frac{\sigma}{\sqrt{N}})$$

where q corresponds to the (1 - CI/2)th quantile from the normal distrubtion (for a 95% CI, this number is 1.96)

This formula requires us to know the population variance (σ). Since we don't know that most of the time, we can:

- instead of a normal distribution, find q for a **t-distribution** with degrees of freedom $N - 1$
- replace σ with $\hat{\sigma}$

Interpretation

- It is **NOT** correct to say that a 95% confidence interval means that there is a 95% chance that the true mean lies in that interval
  - that interpretation is bayesian, confidence intervals are frequentist
- The more correct interpretation is to say that if we replicated the experiment over and over again and computed a 95% confidence interval for each trial, then 95% of those intervals will contain the true mean.
  - Alternatively, 95% of all confidence intervals constructed using this procedure will contain the true mean.

References

- Learning statistics with R section 10.5