6.2 Developing Confidence Intervals

Dr. Lauren Perry

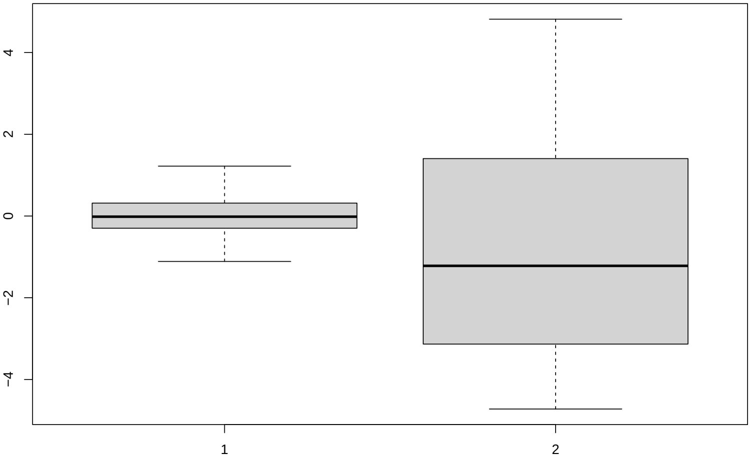
Goals

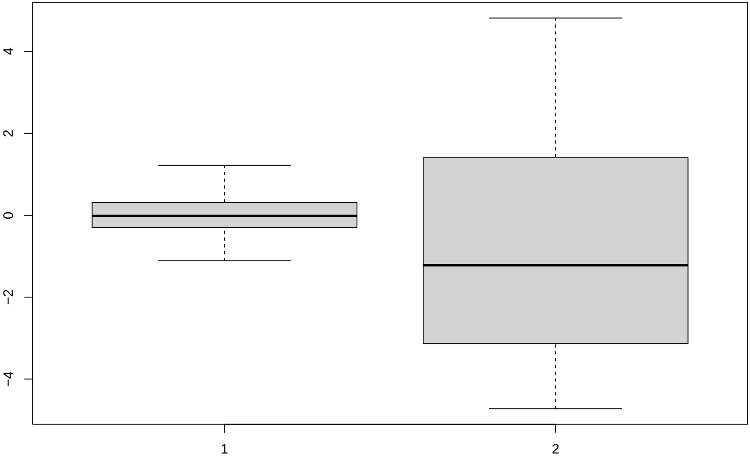
1. Find and interpret 95% confidence intervals for a mean when (A) the population is normal and (B) σ is known.

- A **point estimate** is a single-value estimate of a population parameter.
- We say that a statistic is an **unbiased estimator** if the mean of its distribution is equal to the population parameter.
  - Otherwise, it is a **biased estimator**.
- Ideally, we want estimates that are unbiased with small standard error.
  - For example, a sample mean (unbiased) with a large sample size (results in smaller standard error).

Point estimates are useful, but they only give us so much information. The variability of an estimate is also important!

Take a look at these two boxplots:





- Both samples are size $n = 100$ and have $\bar{x} = 0$
- Variable 1 has a standard deviation of $\sigma = 0.5$
- Variable 2 has standard deviation $\sigma = 5$

Confidence Intervals

A **confidence interval** is an interval of numbers based on the point estimate of the parameter (along with some other stuff).

- Say we want to be 95% confident about a statement.
- In Statistics, this means that we have arrived at our statement using a method that will give us a correct statement 95% of the time.

- Our best point estimate for μ (based on a random sample) is $\bar{x}$, so that value will make up the center of the interval.
- To create an interval around $\bar{x}$, we will construct what is called the **margin of error**.
  - We will use the variability of the data along with some normal distribution properties.
  - This will look like
  $$z \times \frac{\sigma}{\sqrt{n}}$$
  - The value z will come from the normal distribution and will be based on how confident we want to be, e.g., 95% confident.

Putting everything together, the 95% confidence interval is

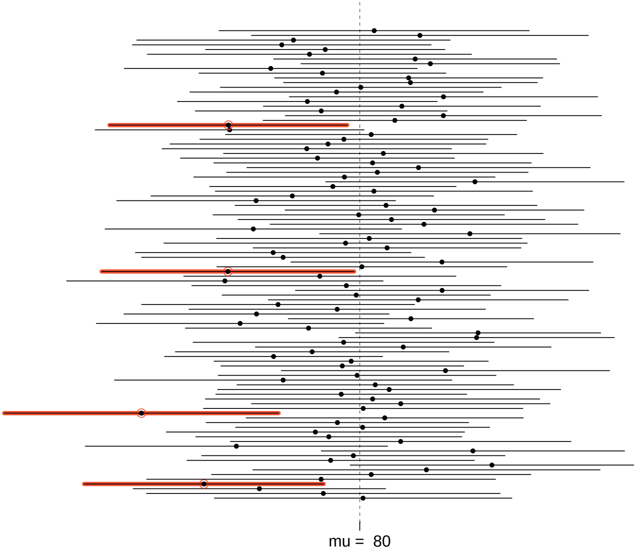
$$\left(\bar{x} - z_* \frac{\sigma}{\sqrt{n}}, \bar{x} + z_* \frac{\sigma}{\sqrt{n}}\right)$$

where $z_* = 1.96$.

The value 1.96 is chosen because $P(-1.96 < Z < 1.96) = 0.95$ (this is what makes it a 95% confidence interval!).

Interpreting a Confidence Interval

If an experiment is run infinitely many times, the true value of μ will be contained in 95% of the intervals.



Common mistakes

- It is NOT accurate to say that "the probability that μ is in the confidence interval is 0.95".
  - The parameter μ is some fixed quantity and it's either in the interval or it isn't.
- We are NOT "95% confident that $\bar{x}$ is in the interval".
  - The value $\bar{x}$ is some known quantity and it's always in the interval.

Example

Suppose I took a random sample of 50 Sac State students and asked about their SAT scores and found a mean score of 1112. Prior experience with SAT scores in the CSU system suggests that SAT scores are well-approximated by a normal distribution with standard deviation known to be 50.

1. Find a 95% confidence interval for Sac State SAT scores.
2. Interpret your interval in the context of the problem.

Checkpoint

The preferred keyboard height for typists is approximately normally distributed with $\sigma = 2.0$. A sample of size $n = 31$, resulted in a mean preferred keyboard height of $80cm$.

1. Find and interpret a 95% confidence interval for keyboard height.
2. What is the width of the interval? If we wanted a narrower interval, what could we do?