

Confidence Intervals II

Grinnell College

October 12, 2026

Warm-up

A random sample of cats found around campus gives an approximate 95% confidence interval of (8, 12) pounds for the population mean weight

1. Does this mean that about 95% of campus cats weigh between 8 and 12 pounds? If not, what does the 95% describe?
2. Using the same sample, we construct an 80% confidence interval. Would its center change? Would it be wider or narrower?
3. Another group takes a different random sample of the same size and gets a 95% confidence interval of (10, 14) pounds. Our intervals overlap from 10 to 12. Can we report (10, 12) as a 95% confidence interval? Why or why not?

Standardization

Recall that we can standardize our sample mean creating a Z score, where

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

To construct a confidence interval with $M\%$ confidence, we simply need to find C such that

$$P(-C < Z < C) = M\%$$

Point $\pm$ Margin of Error

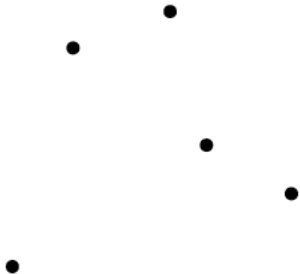
The key idea is this:

In finding a plausible range for our parameter μ , we want to create an interval, centered at our observed statistic, that takes the form

$$\bar{X} \pm C \times \left(\frac{\sigma}{\sqrt{n}} \right)$$

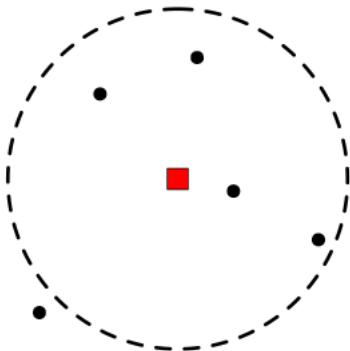
Our **confidence**, mediated by C , is the frequency with which an interval constructed in this way will contain the true value of the mean.

Start with a collection of different sample means, $\bar{X}_1, \bar{X}_2, \dots$

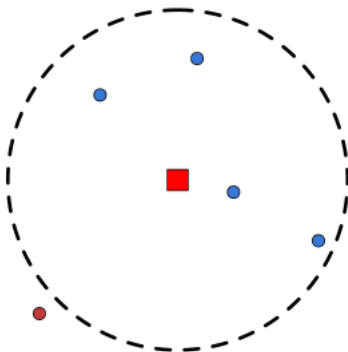


The CLT tells us how the sample means will be distributed:

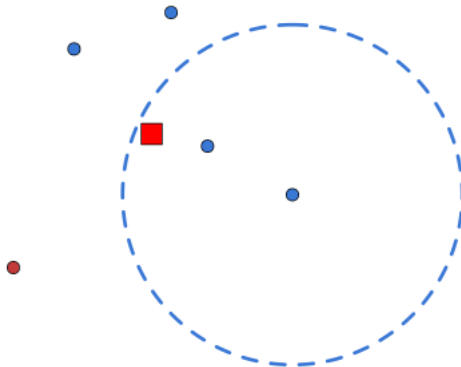
$$\bar{X} \sim N(\mu, \sigma/\sqrt{n})$$



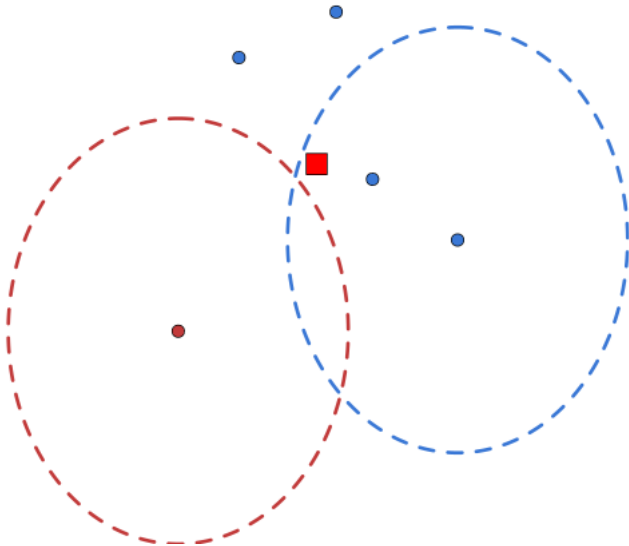
Depending on our choice of critical value, C , some percentage of sample means will fall within some interval of μ

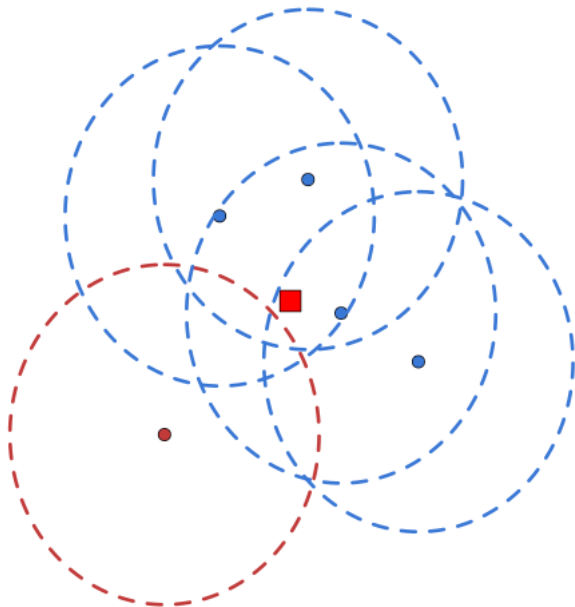


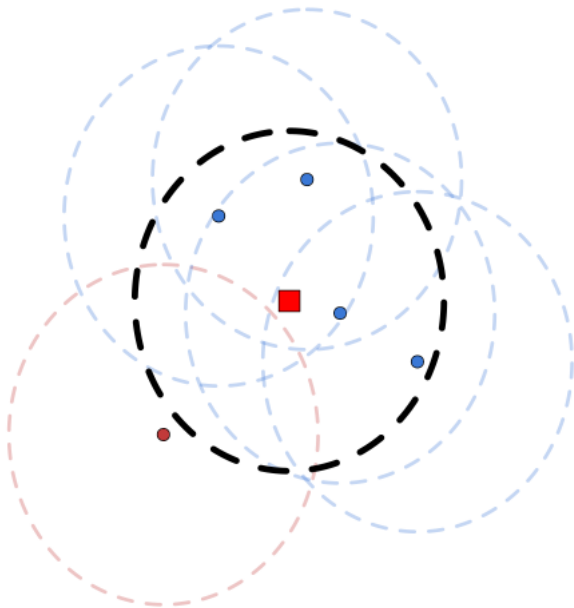
By *symmetry*, if some $\bar{X}$ falls in an interval around μ , μ will fall within an interval around that $\bar{X}$

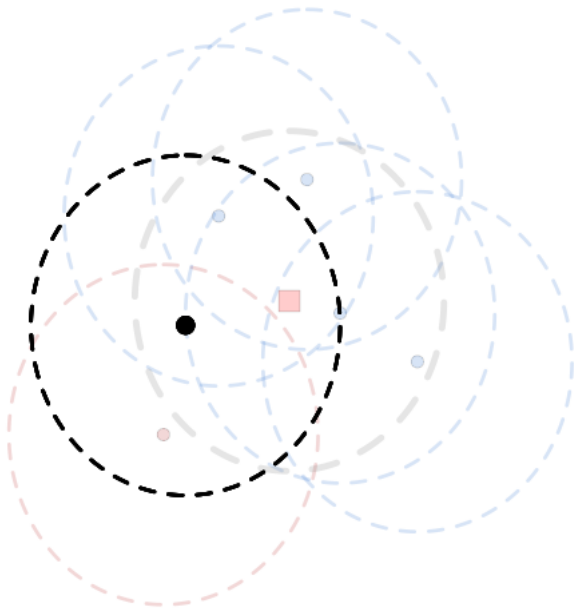


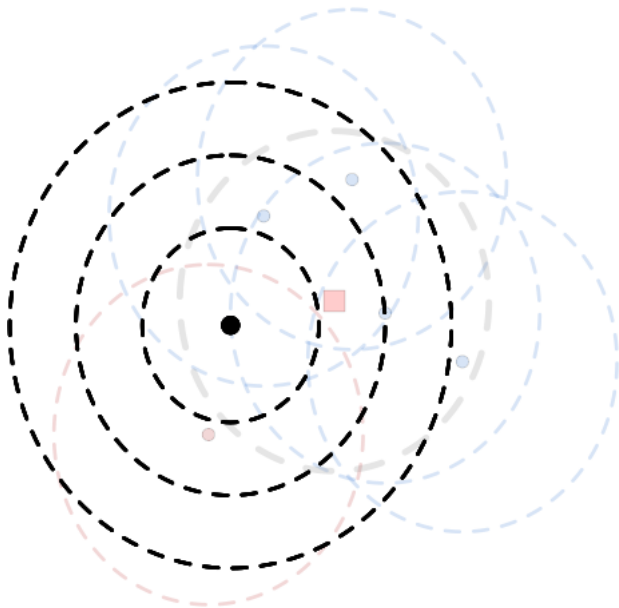
This is also true if a particular $\bar{X}$ does *not* fall within an interval around μ











Key Ideas Today

We need to have clear two separate but not independent ideas in our minds: the size of a an interval and the amount of confidence

CRITICAL VALUES ARE THE ONLY THING THAT DETERMINE CONFIDENCE ← This is a fact, never let anyone tell you otherwise

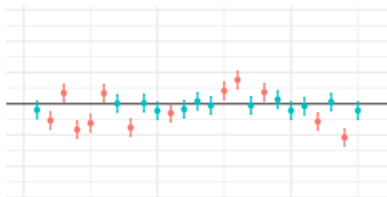
While the critical values are the only thing that determine confidence, there are in fact three pieces that determine the final size:

- ▶ Critical value, C (more confidence means larger intervals)
- ▶ The standard error ($\sigma/\sqrt{n}$), communicating the amount of uncertainty in $\bar{X}$ (more uncertainty also means larger intervals)

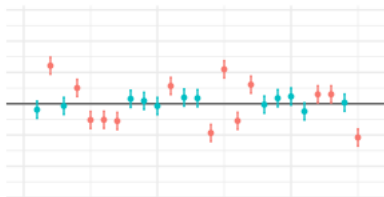
Confidence Intervals

It is also worth observing that we can *alter* our process to achieve different results. There is a trade-off between how frequently we are correct and how much uncertainty we allow in our prediction

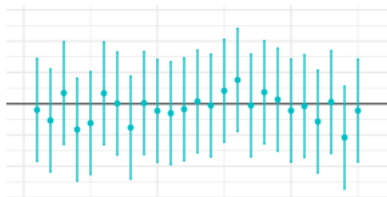
Mean ± 0.5 SE



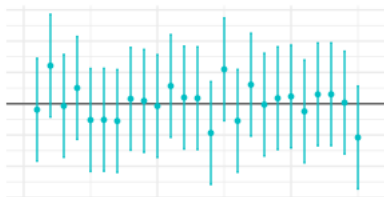
Mean ± 0.5 SE



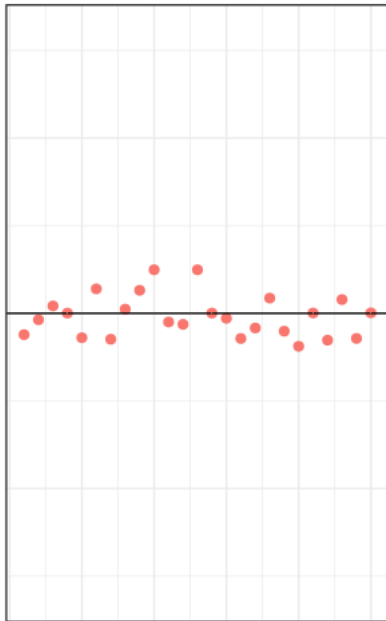
Mean ± 3 SE



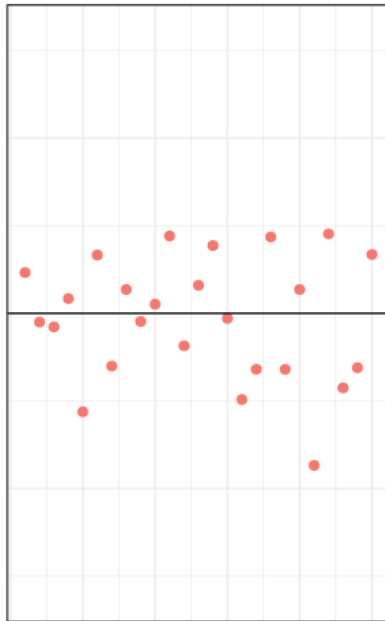
Mean ± 3 SE



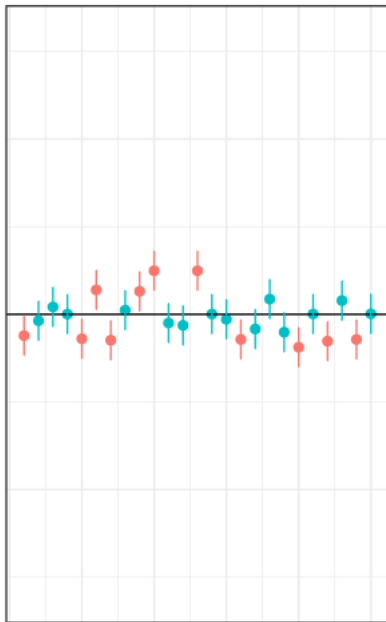
Low Variance



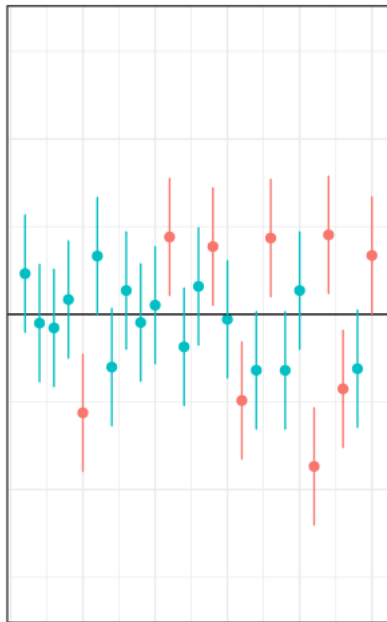
High Variance



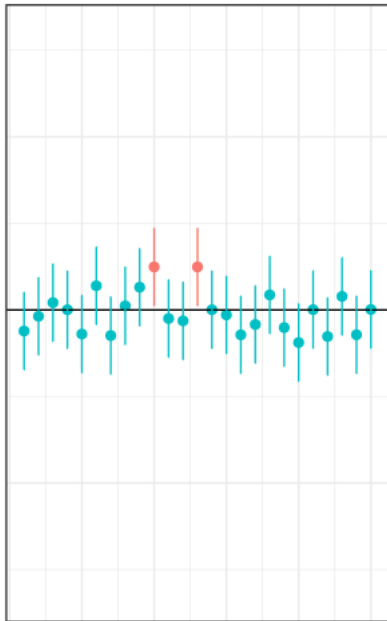
Low Variance, $C = 1$



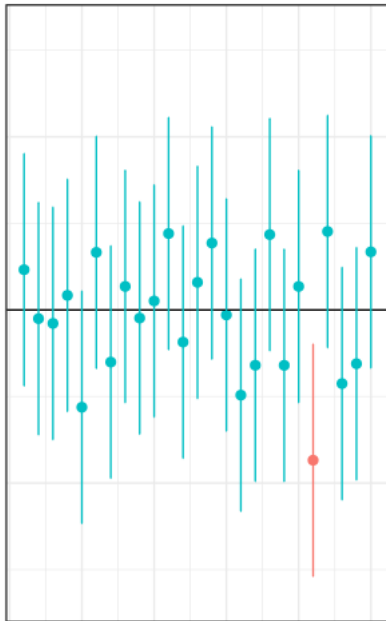
High Variance, $C = 1$



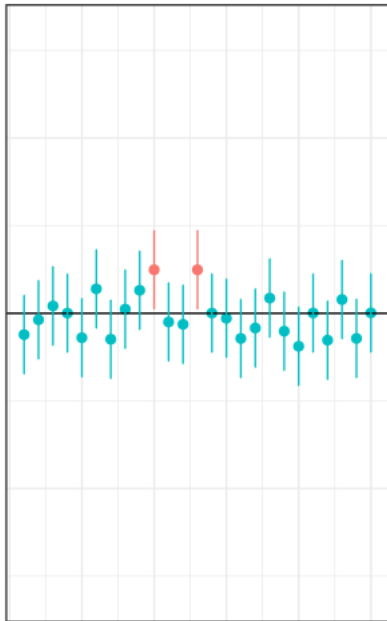
Low Variance, $C = 2$



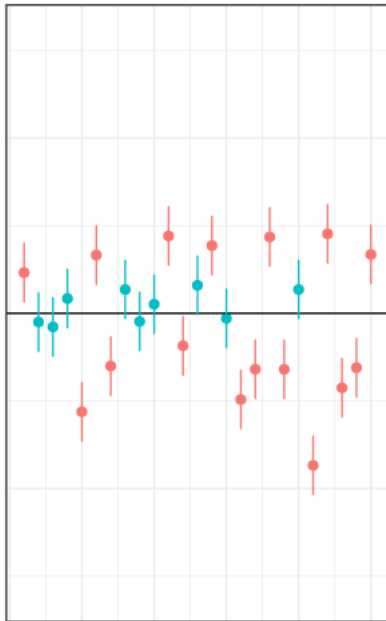
High Variance, $C = 2$



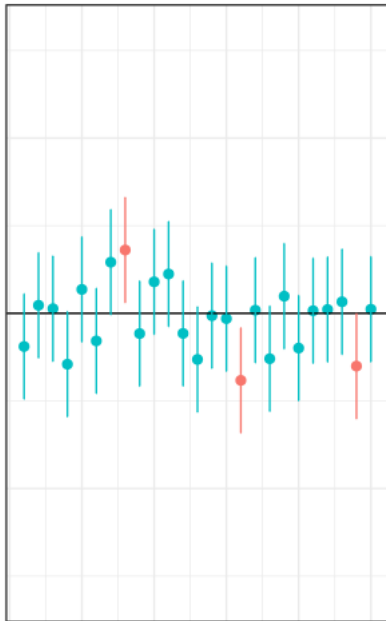
Low Variance, $C = 2$



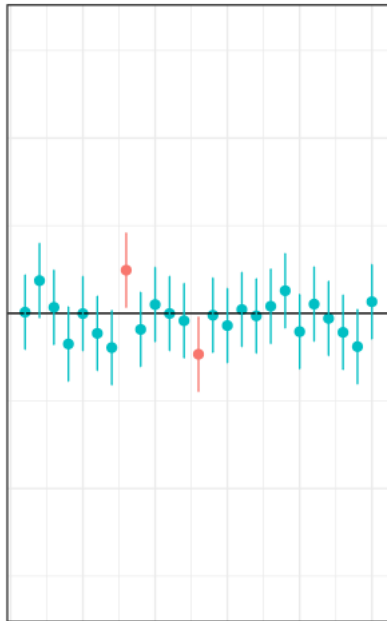
High Variance, $C = 0.5$



$n = 25$



$n = 50$



Scenarios

What kind of behavior should you expect in these scenarios:

Scenario 1: Small n , large σ , small C

Scenario 2: Large n , large σ , large C

Scenario 3: Small n , small σ , large C

Identify: what are sources of variability in $\bar{X}$, what determines my confidence

What are we doing:

1. Collecting a sample to estimate plausible interval for population parameter
2. Based on mean from sample and standard deviation
3. Interval constructed according to CLT

What impacts my intervals:

1. Critical value C *only thing that mediates confidence*. Larger C means larger interval (why)
2. More variability in $\bar{X}$ also means larger interval. The variability in $\bar{X}$ is measured by the *standard error*, $\sigma/\sqrt{n}$
 - ▶ σ represents variability in the population. If this is larger, uncertainty in $\bar{X}$ will be large (and vice versa)
 - ▶ n represents the sample size. Larger n means more data means more precise estimate of $\bar{X}$