

# Confidence Intervals Pt. 1

## Using Standard Error to Estimate Things

Grinnell College

## Normal distribution

- ▶ unimodal + symmetric bell-curve
- ▶ probabilities

## Central Limit Theorem:

1. If variable  $X$  has mean  $\mu$  and std.dev.  $\sigma$ , and
2. If the number of observations in the sample ( $n$ ) is large
3. then the sampling distribution for  $\bar{X}$  (sample mean) is Normal with mean  $\mu$  and standard error  $\sigma/\sqrt{n}$ .

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

## Review – Sources of Variation

**Pop. Standard Deviation:** Description of the variability in our *population*. It is often denoted  $\sigma$

**Sample Standard Deviation:** Description of the variability in our *observations* (sample). It is often denoted  $s$

**Standard Error:** Description of variability in our *estimates* of a parameter (such as the mean). We will denote standard error as  $SE$ , with  $SE = \sigma/\sqrt{n}$ , where  $n$  is the number of observations in our sample

# Outline

We saw that the statistic is not going to be exactly equal to the parameter

- ▶ sampling bias
- ▶ sampling variability

So... we can't just provide a single value for our estimate of the parameter

- ▶ We also need to quantify how far away our guess is

This is why we came up with the *standard error (SE)*, now we need to figure out how to use it.

**Goal:** We are going to spend today learning how to estimate population means

## Example – COVID Vaccines

According to the U.S. Census Bureau, as of October 11, 2021:

“83.3% (+/- 0.5%) of U.S. adults 18 years and older have received at least one dose of a COVID-19 vaccine.” This is based on a representative sample of civilians aged 18 and over. “Margins of error shown at 90% confidence.”

What does margin of error mean?

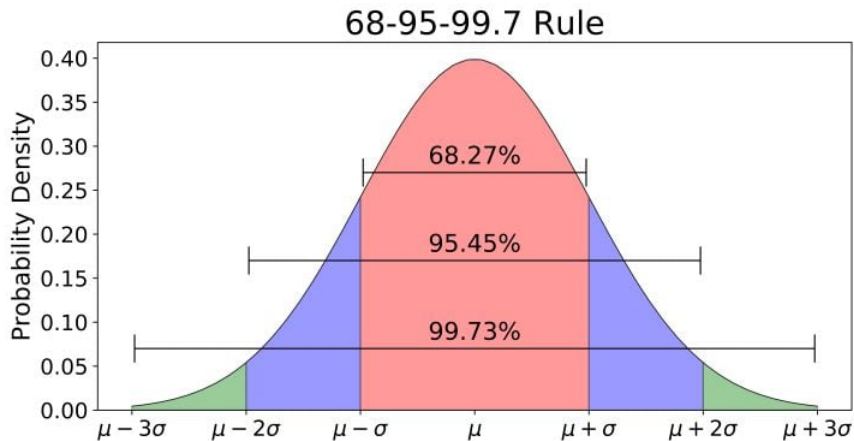
# Intervals

For the rest of these slides, our goal is to determine the mean of a *population*.

We cannot rely on only our **point estimate**  $\bar{X}$ , but perhaps we can find a range of reasonable values that looks like:

Point Estimate  $\pm$  Margin of Error

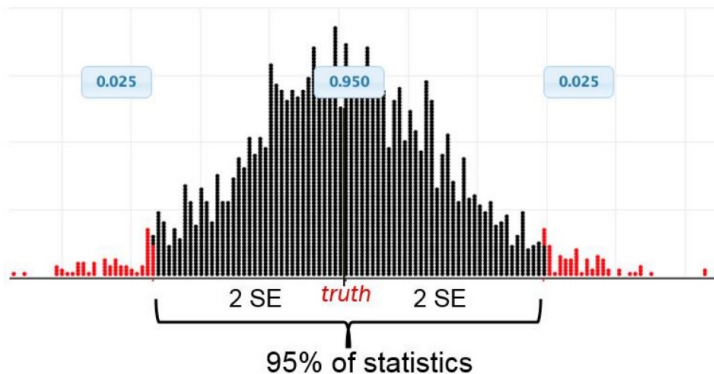
## A good place to start:



## Good place to start:

The sampling distribution for the sample mean looks  $N(\mu, \sigma^2/n)$

- ▶ Central Limit Theorem



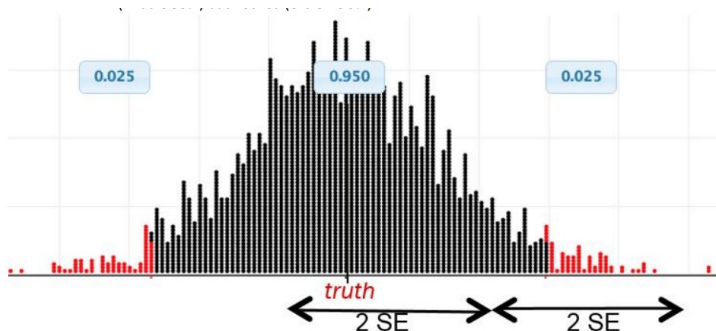
11

In the sampling distribution, 95% of statistics will be within  $2 \times SE$  of the pop. mean  $\mu$

## Good place to start:

The sampling distribution for the sample mean looks  $N(\mu, \sigma^2/n)$

- ▶ Central Limit Theorem



Equivalently: The interval 'statistic  $\pm 2 \times \text{SE}$ ' will contain  $\mu$  for 95% of the statistics

# Confidence Interval

We are going to call this interval a "95% Confidence Interval":

- ▶ 95% comes from the fact that 95% of statistics are within 2SE's of the mean
- ▶ Confidence refers to the fact that this is a range of plausible values for the parameter

Formula for a 95% confidence interval for estimating a pop. mean ( $\mu$ ) is:

$$\bar{x} \pm 2 \times SE$$

# Confidence Interval

Formula for a 95% confidence interval for estimating a pop. mean ( $\mu$ ) is:

$$\bar{x} \pm 2 \times SE$$

**Margin of Error** tells us how wide our interval is.

- ▶ ME = half the width (or length) of the interval
- ▶ for 95% CI  $\rightarrow$  ME =  $2 \times SE = 2 \frac{\sigma}{\sqrt{n}}$

This makes our (final) formula for a 95% confidence interval for estimating a pop. mean ( $\mu$ ):

$$\bar{x} \pm 2 \times \frac{\sigma}{\sqrt{n}}$$

# CI Interpretation

Remember our goal: we are trying to estimate the *population mean*

## Confidence Interval Interpretation

- ▶ mention confidence level
- ▶ specify the values we got
- ▶ use context for the population mean when able

"We are 95% confident that (the population mean) is between (lower value) and (upper value)."

## Example – Movie Budgets

Hollywood movie budget data:  $\mu = 51.38$ ,  $\sigma = 57.93$

From a sample of 50 movies we find  $\bar{x} = 45.65$ .

Construct a 95% Confidence Interval for the pop. mean movie budget.

## Example – Movie Budgets

Hollywood movie budget data:  $\mu = 51.38$ ,  $\sigma = 57.93$

From a sample of 50 movies we find  $\bar{x} = 45.65$ .

Construct a 95% Confidence Interval for the pop. mean movie budget.

- ▶ When we know  $\sigma$  we can use the formula directly

$$\bar{x} \pm 2 \frac{\sigma}{\sqrt{n}} \rightarrow 45.65 \pm 2 \times \frac{57.93}{\sqrt{50}} \rightarrow 45.65 \pm 16.39$$
$$(29.26, 62.04)$$

## Example – Movie Budgets

The 95% CI for the population mean is (29.26, 62.04).

**Interpretation:**

We are 95% confident that (the population mean) is between (lower value) and (upper value).

## Example – Movie Budgets

The 95% CI for the population mean is (29.26, 62.04).

**Interpretation:**

We are 95% confident that the population mean movie budget is between 29.26 and 62.04 million dollars.

## Example – Fish Mercury Levels

Data was collected from 53 lakes in Florida. For each lake, the mercury level (parts per million) was computed for a large mouth bass.

Based on this sample of fish, the mean mercury level was 0.527 ppm with a standard deviation of .118

**Construct a 95% CI for pop. mean:**

- ▶ Issue: we don't know  $\sigma$

## Example – Fish Mercury Levels

Data was collected from 53 lakes in Florida. For each lake, the mercury level (parts per million) was computed for a large mouth bass.

Based on this sample of fish, the mean mercury level was 0.527 ppm with a standard deviation of 0.118

**Construct a 95% CI for pop. mean:**

$$\bar{x} \pm 2 \frac{\sigma}{\sqrt{n}} \rightarrow \bar{x} \pm 2 \times \frac{s}{\sqrt{n}} \rightarrow 0.527 \pm 2 \times \frac{0.118}{\sqrt{53}}$$
$$(0.495, 0.560)$$

We are 95% confident that the true pop. mean mercury level of fish in Florida lakes is between 0.495ppm and 0.560ppm

## More on "Confidence"

We are 95% confident that...

Let's really dig into what this 'confidence' part means

## More on "Confidence"

A **confidence interval** is an interval that has the following properties:

- ▶ Point estimate  $\pm$  Margin of Error
- ▶ It is made with the intention of giving a plausible range of values for a *parameter* based on a *statistic*
- ▶ There is no probability associated with a confidence interval; *it is either correct or it is incorrect*

### Movie Budgets

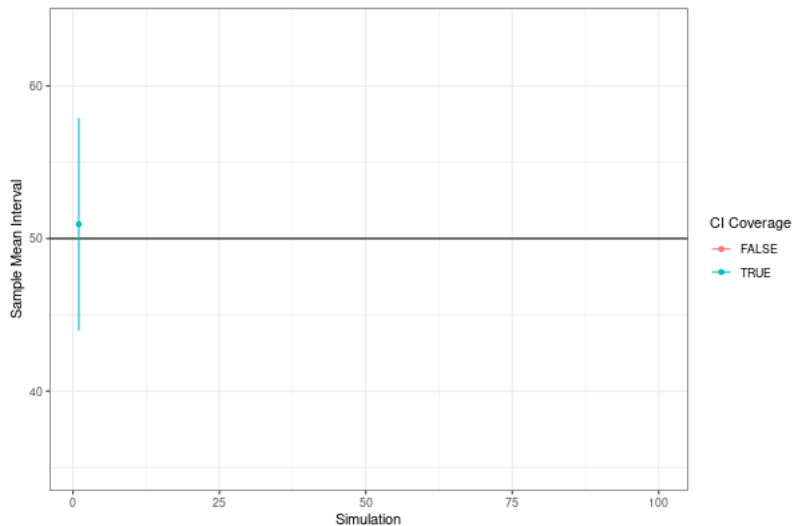
- ▶ It was made to present a reasonable range of values for the *parameter*  $\mu$  as estimated by the *statistic*  $\bar{X}$
- ▶ The interval was (29.26, 62.04). As our true mean is  $\mu = 51.38$ , this interval *is* correct in the sense that it *contains* our true parameter

## More on "Confidence"

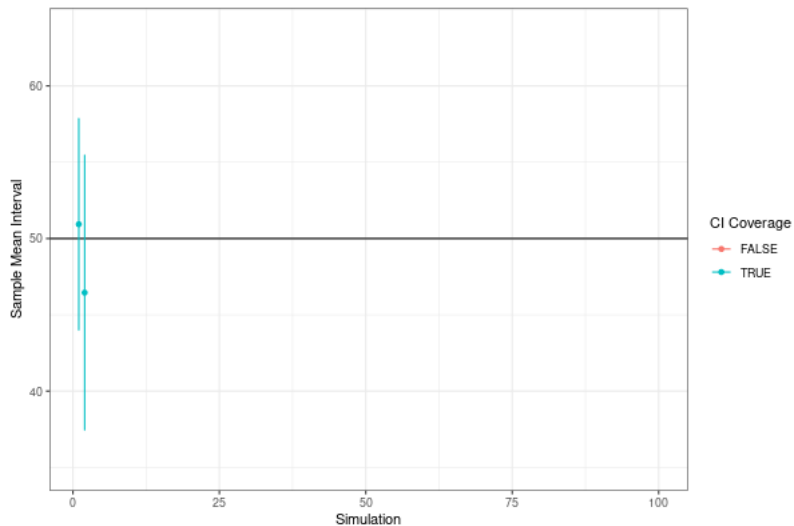
When we say something has a 95% confidence interval, what we mean is:

*The process that constructed this interval has the property that, on average, it contains the true value of the parameter 95 times out of 100*

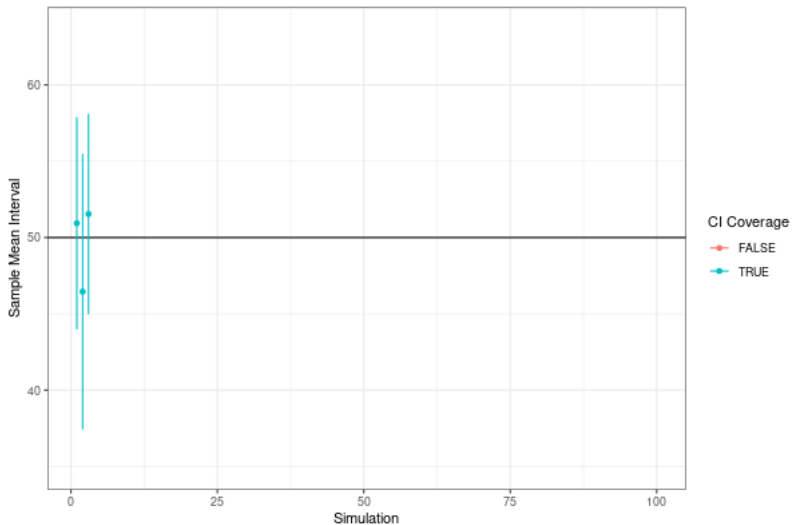
# Coverage



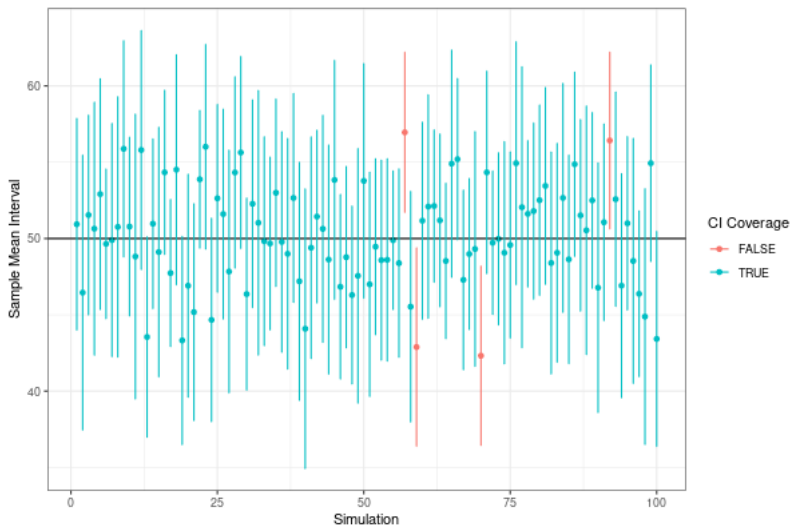
# Coverage



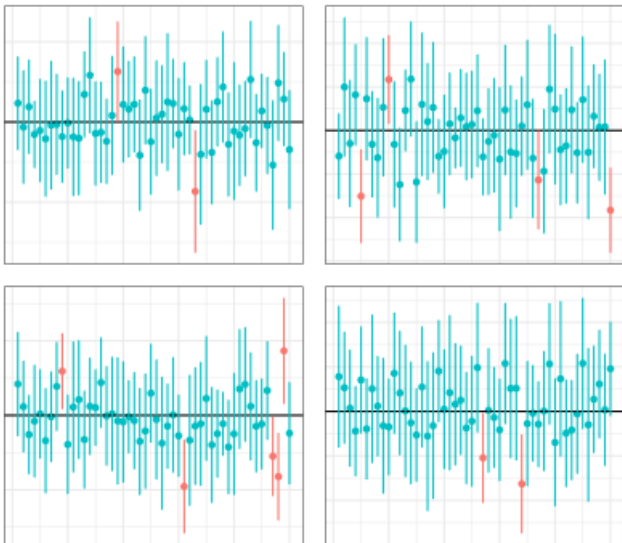
# Coverage



# Coverage



# Coverage



# Wrapping Up

Confidence Intervals quantify sampling variability

- ▶ Range of plausible values for the statistic
- ▶ Range is determined by how much the statistics vary from sample to sample

Confidence Intervals **DO NOT** account for bias in the samples

- ▶ We will never be able to quantify the bias in our samples
- ▶ It is important to mention possible sources of bias in our final conclusions