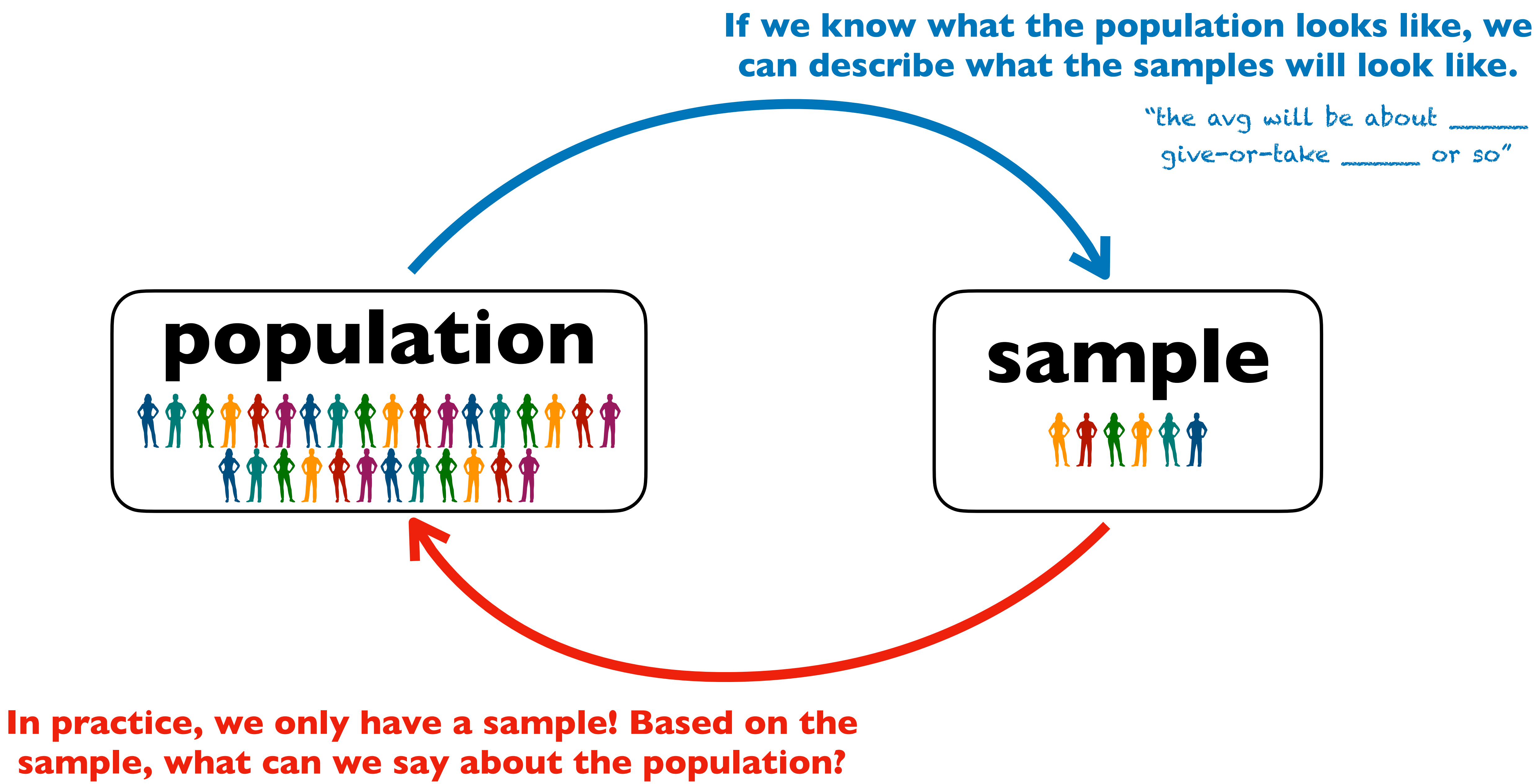
Confidence Intervals for Percentages

Reasoning from Sample to Population


If we know what the population looks like, we can describe what the samples will look like.
"the avg will be about ______ give-or-take ______ or so"
population
sample
In practice, we only have a sample! Based on the sample, what can we say about the population?

Scenario

There are 13 million registered voters in FL. We survey a randomly-selected 400 of them and ask: "Do you approve or disapprove of the job that Ron DeSantis is doing as governor?"

Population Percent

61.5%

8 million people approve
5 million do not approve
(either disapprove or refused to answer)

Sample Percent

hypothetical; something we might do in the future; we can characterize the samples very well.

Population Percent

unknown; it's the reason we're doing the study

Sample Percent

60.25%

observe one sample;
241 of 400 respondents approve

Claim

Suppose we know the population:

The sample percentage will be within two standard errors of the population percentage about 95% of the time.

Suppose we only know the sample:

The sample percentage will be within two standard errors of the population percentage about 95% of the time.

Claim

unknown

Suppose we know the population:

The sample percentage will be within two standard errors of the population percentage about 95% of the time.

$$\frac{8 \text{ million}}{13 \text{ million}} \times 100\% = 62\%$$

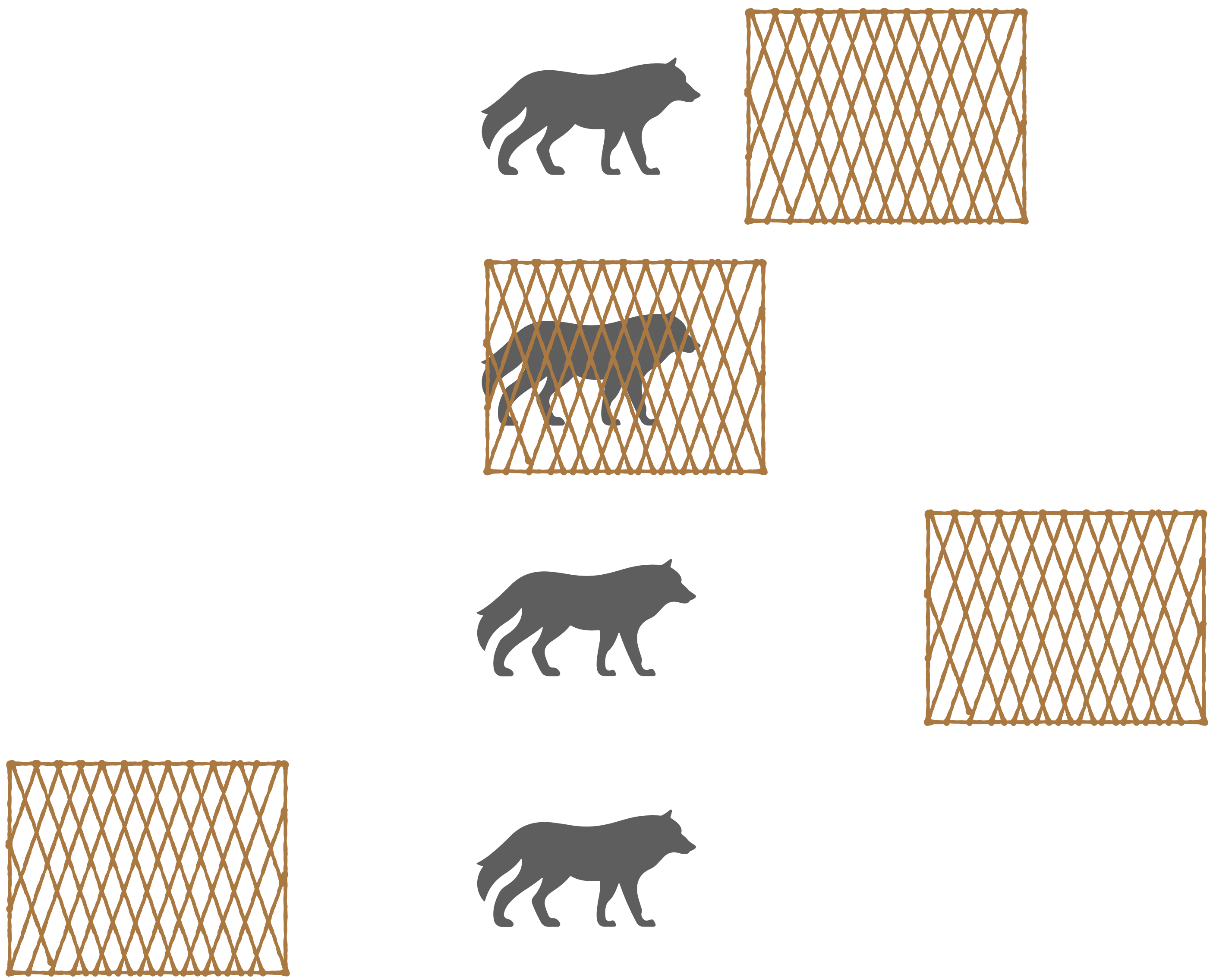
$$\frac{241}{400} \times 100\% = 60\%$$

Suppose we only know the sample:

The sample percentage will be within two standard errors of the population percentage about 95% of the time.

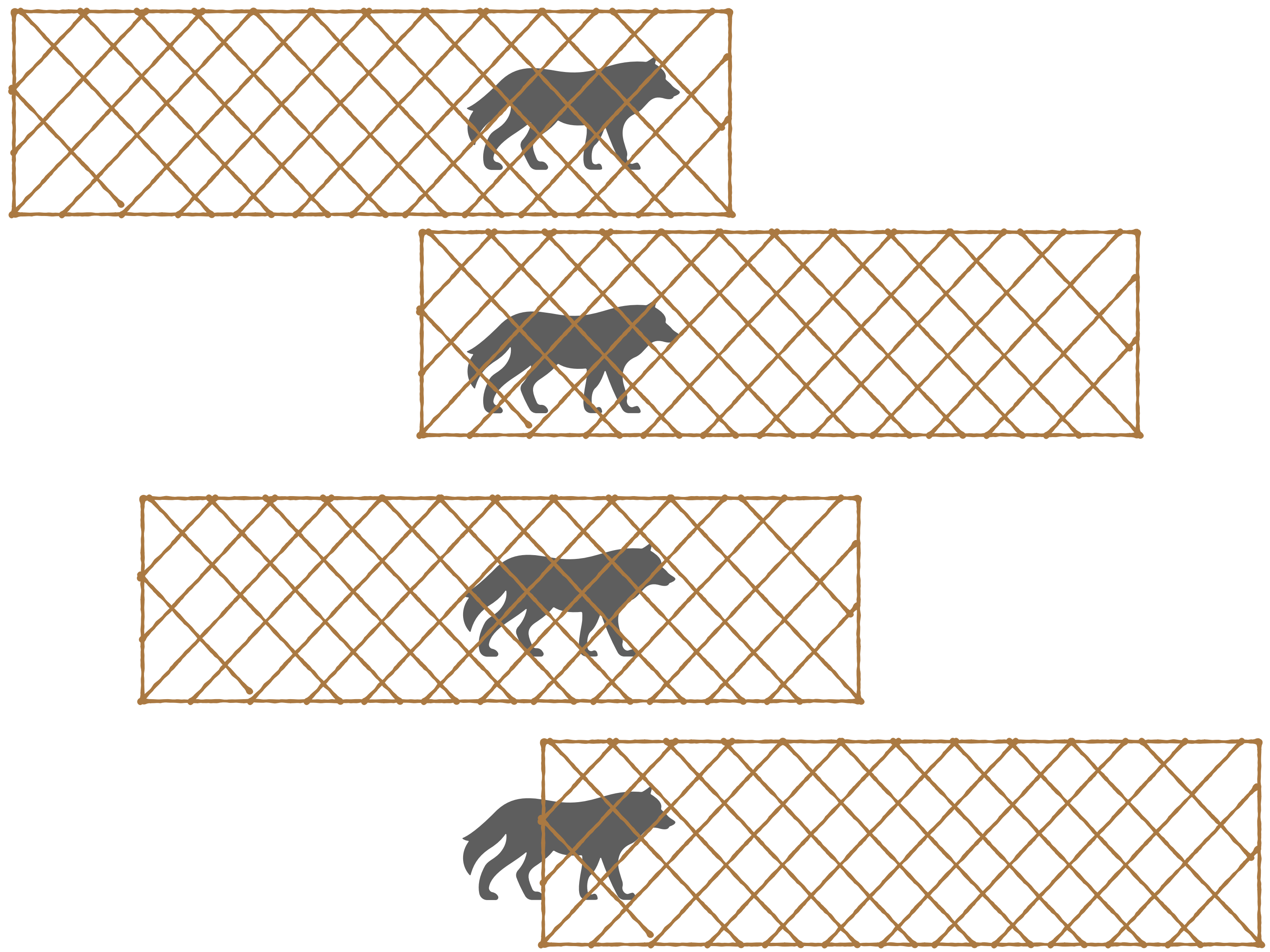
unknown

Too Narrow



Just Right

miss left about 2.5% of the time, miss right about 2.5% of the time
"almost always" catch the wolf



So how wide does our net need to be?

Like nets capturing wolves, we want to add "arms" to our sample percent.

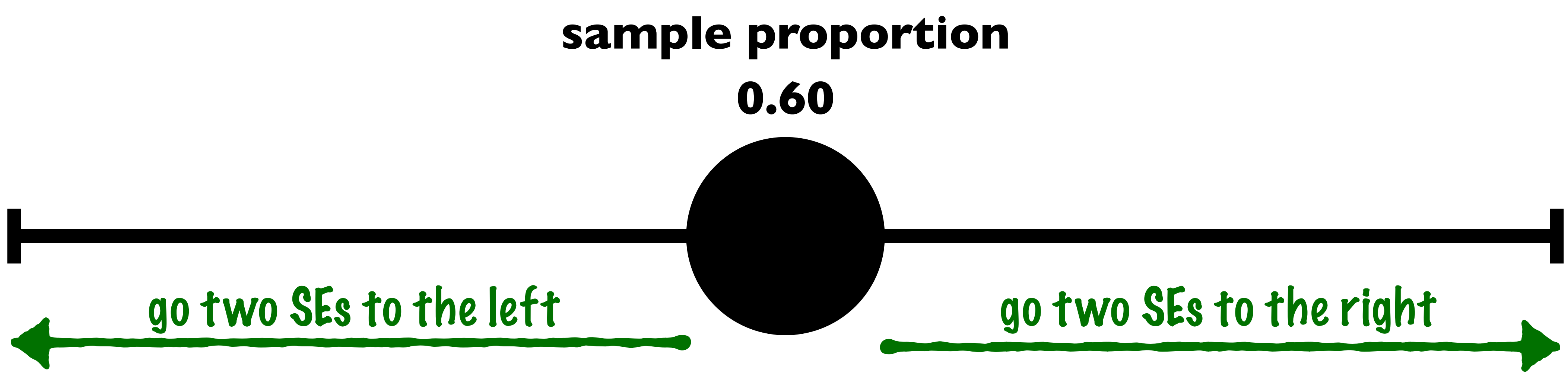
One arm goes left; the other goes right.

These arms should be long enough to "almost always" (i.e., 95% of the time) capture the population percent.

Big Idea

If the sample is within 2 SEs of the population 95% of the time,

then the population will be within 2 SEs of the sample 95% of the time.



95% CI: sample proportion ± 2 standard errors

This 95% confidence interval will capture the population percent in about 95% of repeated samples.

95% CI: sample proportion ± 2 standard errors

The sample proportion is 0.60.

What is the standard error?

$$\text{standard error for proportion} = \frac{\overbrace{\sqrt{\text{frac. 1s in box} \times \text{frac. 0s in box}}}^{\text{SD of box}}}{\sqrt{\text{sample size}}}$$

(Remember to do math on proportions; convert to percents at the end.)

Chance Process

Sampling 400 respondents from a population 13 million people (but an unknown number of approvers and non-approvers) and computing the proportion of approvers in the sample.

Box Model

...is like drawing _____ times with replacement from the box ____________ and averaging the draws.

To get the SE, we need the population proportion.

If we have the population proportion, we don't need the SE.

95% CI: sample percent ± 2 standard errors

The sample proportion is 0.60.

What is the standard error?

$$\text{standard error for proportion} = \frac{\overbrace{\sqrt{\text{frac. 1s in box} \times \text{frac. 0s in box}}}^{\text{SD of box}}}{\sqrt{\text{sample size}}}$$

ideas?

rather than use the fraction of 0s and 1s in the box (or population), let's use the fraction of 0s and 1s in the sample—that should be a pretty good guess.

KEY IDEA : plug-in principle

95% CI: sample percent ± 2 standard errors

The sample proportion is 0.60.

What is the standard error?

$$\text{standard error for proportion} = \frac{\overbrace{\sqrt{\text{frac. 1s in } \cancel{\text{box}}^{\text{sample}} \times \text{frac. 0s in } \cancel{\text{box}}^{\text{sample}}}}^{\text{SD of } \cancel{\text{box}}^{\text{sample}}}}{\sqrt{\text{sample size}}}$$

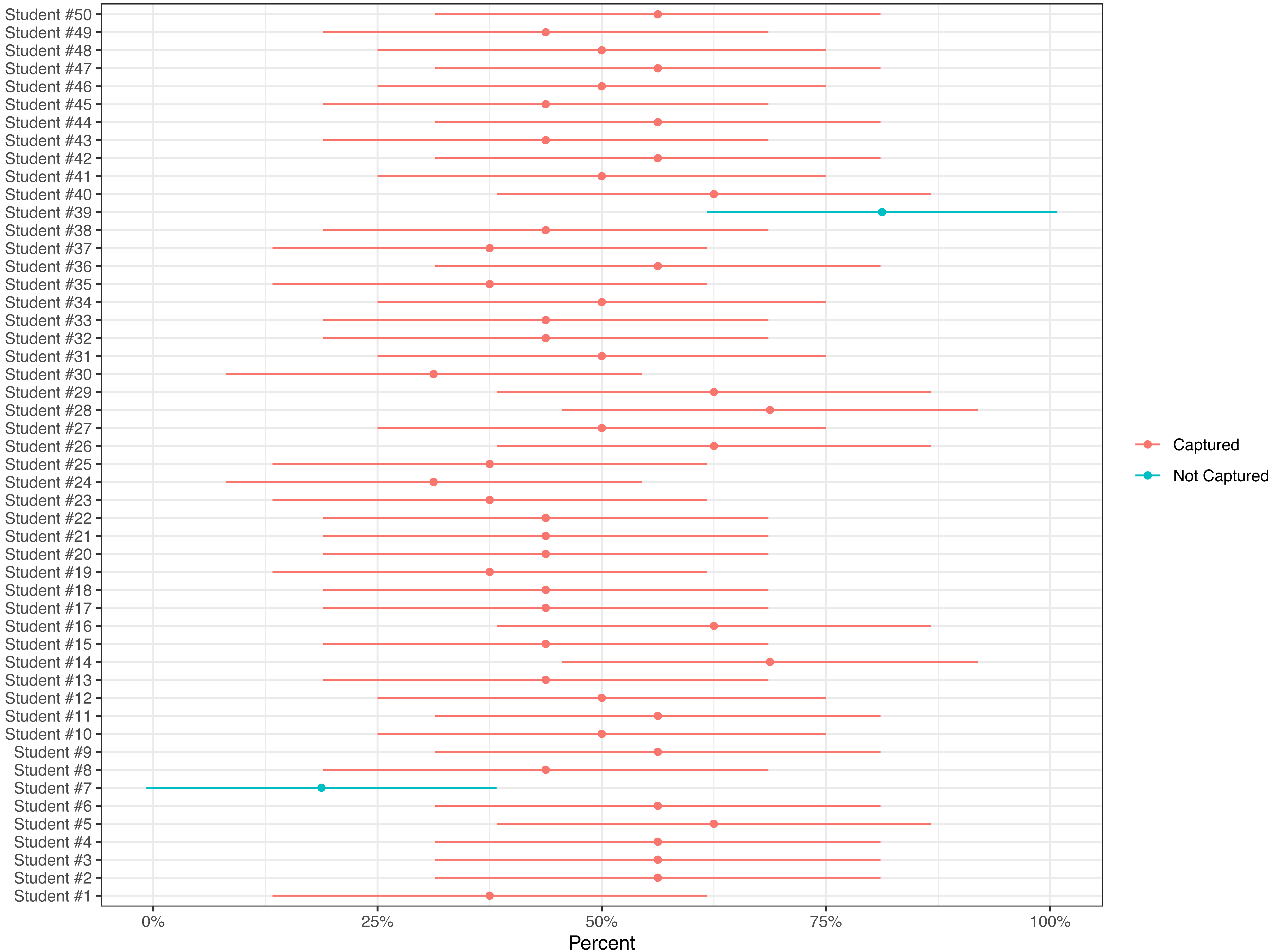
$\boxed{\begin{array}{c} SE = sqrt(.6 \times .4)/sqrt(400) = 0.025 \\ 95\%\ CI = 0.60 \pm (2 \times 0.025) = \\ [0.55, 0.65] \text{ or } [55\%, 65\%] \end{array}}$

Poll a random sample of 16 Americans and ask: "With respect to the abortion issue, would you consider yourself pro-life or pro-choice?" Compute a 95% confidence interval.

Exercise

Do it! We can't sample Americans, but we can toss a coin. Let a Head be like sampling a pro-life respondent. Let a Tail be like sampling a pro-choice respondent.

1. **Toss the coin 16 times and find the percent of "pro-life respondents." Compute the 95% confidence interval for the percent of the population that's pro-life. (It's actually 50%.) Did you capture the truth?**
   a. compute the standard error (using the plug-in principle!)
   b. go two SEs to the left and two SEs to the right
2. **Add your confidence interval to the white board.**
3. **If you have time, do it again.**


Student #50
Student #49
Student #48
Student #47
Student #46
Student #45
Student #44
Student #43
Student #42
Student #41
Student #40
Student #39
Student #38
Student #37
Student #36
Student #35
Student #34
Student #33
Student #32
Student #31
Student #30
Student #29
Student #28
Student #27
Student #26
Student #25
Student #24
Student #23
Student #22
Student #21
Student #20
Student #19
Student #18
Student #17
Student #16
Student #15
Student #14
Student #13
Student #12
Student #11
Student #10
Student #9
Student #8
Student #7
Student #6
Student #5
Student #4
Student #3
Student #2
Student #1
0%
25%
50%
75%
100%
Percent
Captured
Not Captured

Key Points

- **net gun analogy.** We want the 95% CI to be exactly wide enough so that the "net" or "arms" capture the truth 95% of the time.

- **plug-in principle.** Find the SE by using the usual formula, but plug-in the sample SD as an estimate of the population SD—this works well!

- **two to the left; two to the right.** Build the confidence intervals by going two SEs to the left of the sample percent and two to the right.