Confidence Intervals for Percentages

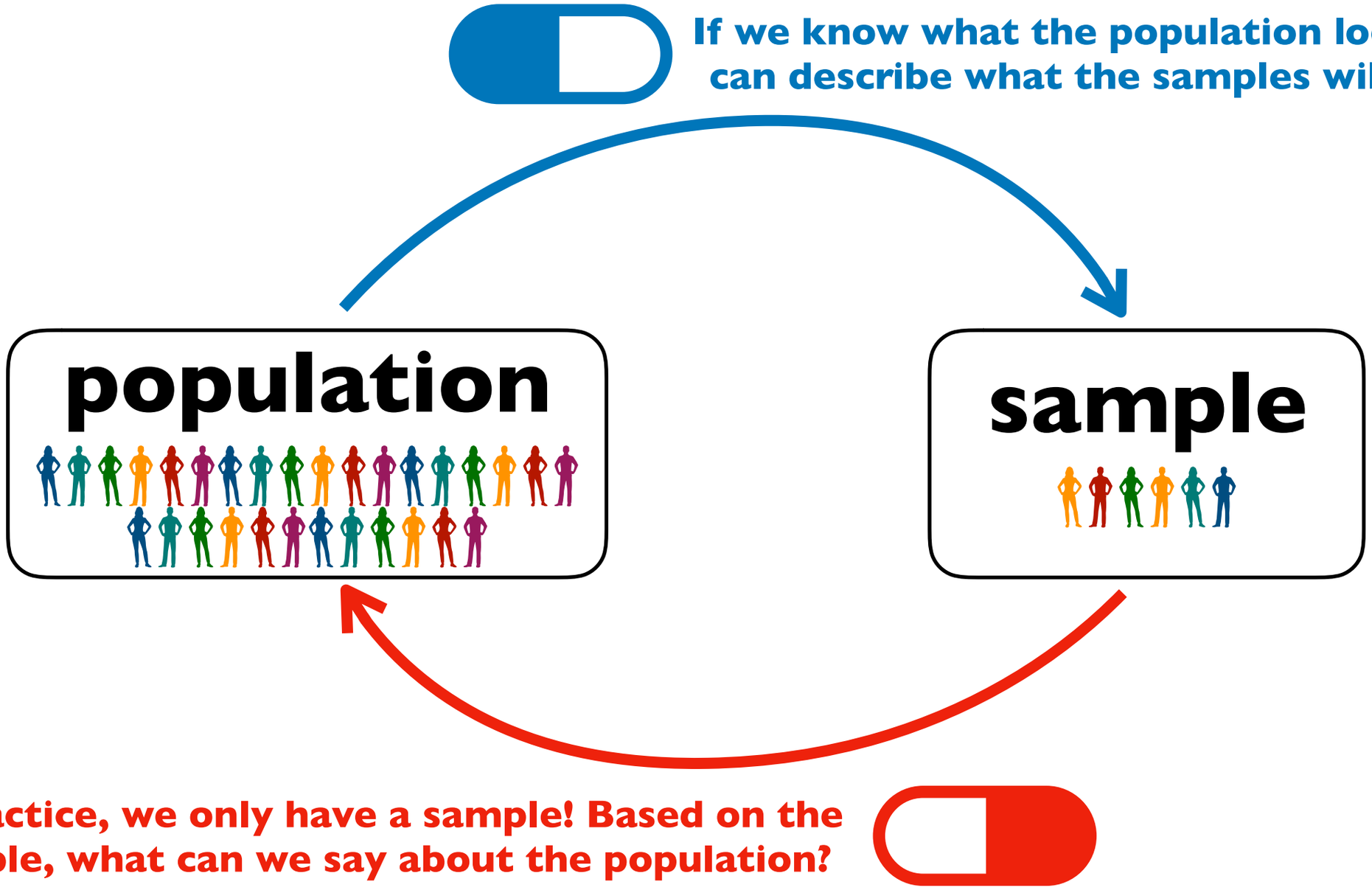
Reasoning from Sample to Population

“This is your last chance. After this there is no turning back. You take the blue pill: the story ends, you wake up in your bed and believe whatever you want to believe. You take the red pill: you stay in Wonderland and I show you how deep the rabbit hole goes.”

—Morpheus, The Matrix

harsh reality

imagined utopia


If we know what the population looks like, we can describe what the samples will look like.
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.


Corporate needs you to find the differences
between this picture and this picture.
They're the same picture.
imgflip.com

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.

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

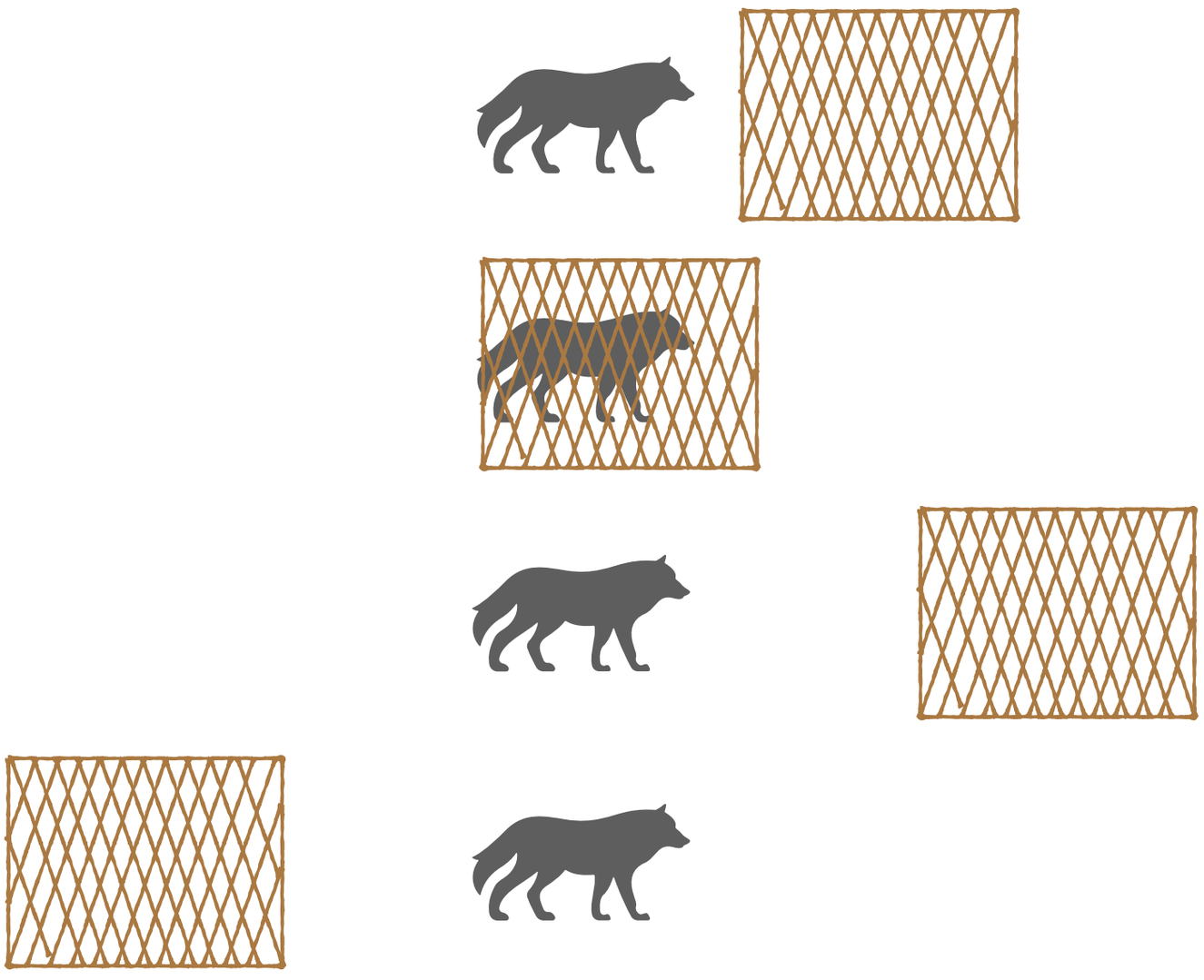
$$\frac{241}{400} \times 100\% = 60\%$$

Suppose we only know the sample:

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

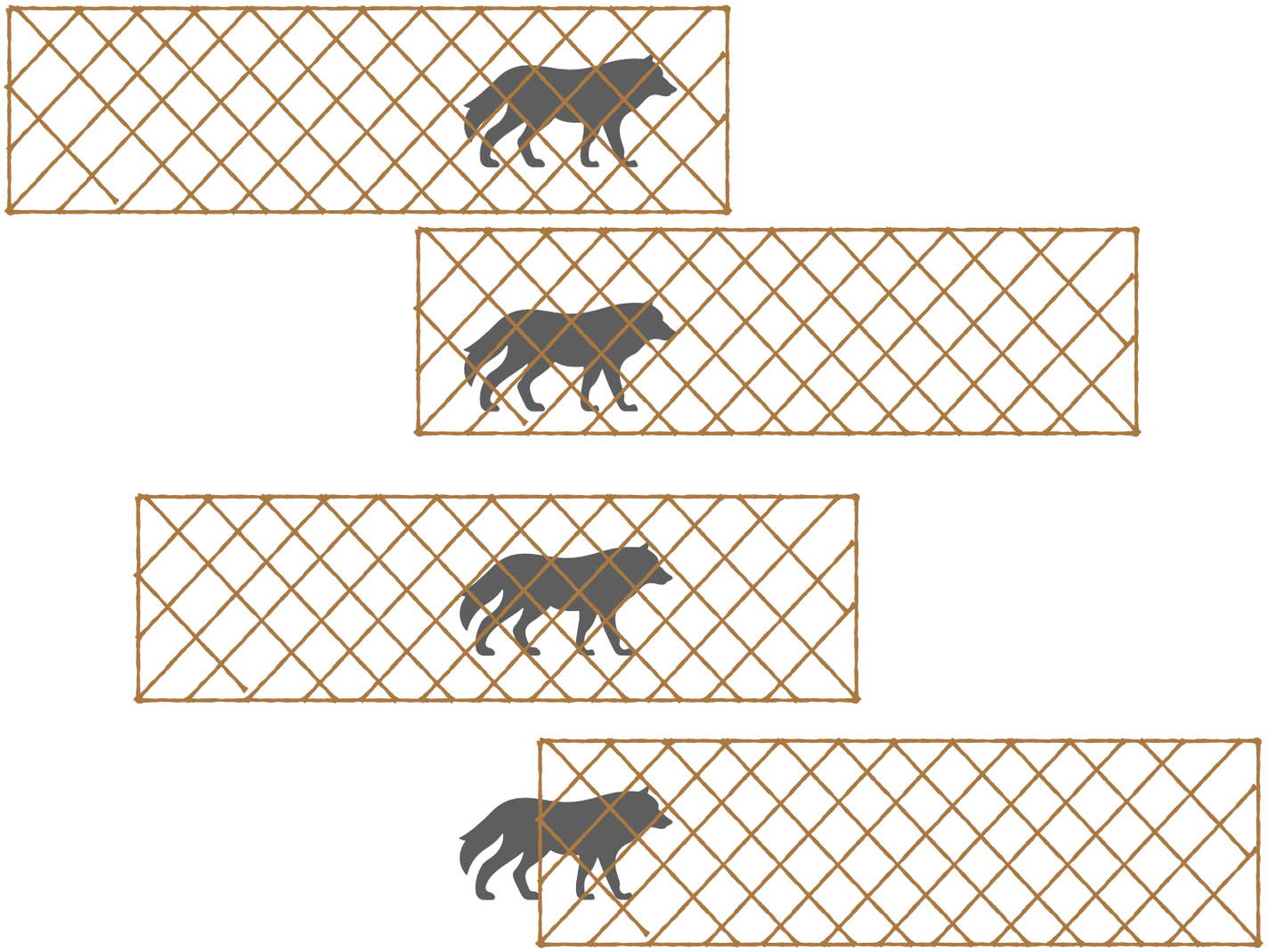
unknown

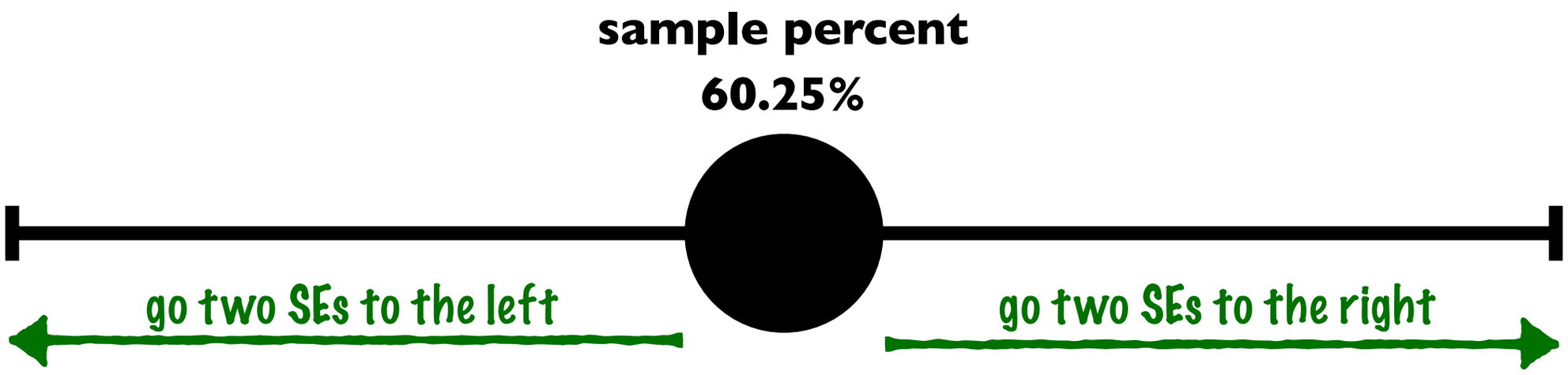
Too Narrow



Just Right

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





95% CI: sample percent ± 2 standard errors

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

95% CI: sample percent ± 2 standard errors

The sample percent is 60.25%.

What is the standard error (for the percent)?

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

$$\text{standard error for percent} = \frac{\text{standard error for sum}}{\text{number of draws}} \times 100\,\%$$

Chance Process

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

Box Model

…is like drawing _____ times with replacement from the box _____ and summing the draws.

To get the SE, we need the population percent.

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

95% CI: sample percent ± 2 standard errors

The sample percent is 60.25%.

What is the standard error (for the percent)?

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

$$\text{standard error for percent} = \frac{\text{standard error for sum}}{\text{number of draws}} \times 100\,\%$$

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.

95% CI: sample percent ± 2 standard errors

The sample percent is 60.25%.

What is the standard error (for the percent)?

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

sample

sample

sample

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?**

2. **Add your confidence interval to the white board.**

3. **If you have time, do it again.**