Sample Surveys
and the Box Model
Connecting the Two

Review

basic rules of chance

addition rule

multiplication rule

chance processes

especially those we can model with a box model

expected value and standard error

The sum will be about _____ give or take _____ or so.

normal approximation

Scenario

Scenario

There are 13 million registered voters in FL.

Scenario

There are 13 million registered voters in FL.

We have enough research money to survey 400 of them.

Scenario

There are 13 million registered voters in FL.

We have enough research money to survey 400 of them.

We decide to ask

"Do you approve or disapprove of the job that Ron DeSantis is doing as governor?"

Scenario

There are 13 million registered voters in FL.

We have enough research money to survey 400 of them.

We decide to ask

"Do you approve or disapprove of the job that Ron DeSantis is doing as governor?"

Assumptions

Scenario

There are 13 million registered voters in FL.

We have enough research money to survey 400 of them.

We decide to ask

"Do you approve or disapprove of the job that Ron DeSantis is doing as governor?"

Assumptions

8 million people approve

Scenario

There are 13 million registered voters in FL.

We have enough research money to survey 400 of them.

We decide to ask

"Do you approve or disapprove of the job that Ron DeSantis is doing as governor?"

Assumptions

8 million people approve

5 million do not approve

(either disapprove or refused to answer)

Scenario

There are 13 million registered voters in FL.

We have enough research money to survey 400 of them.

We decide to ask

"Do you approve or disapprove of the job that Ron DeSantis is doing as governor?"

Assumptions

8 million people approve

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

Question

Scenario

There are 13 million registered voters in FL.

We have enough research money to survey 400 of them.

We decide to ask

"Do you approve or disapprove of the job that Ron DeSantis is doing as governor?"

Assumptions

8 million people approve

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

Question

The number of approvers in our sample will be about _____ give or take _____ or so.

Sampling 400 respondents from a population of 8 million approvers and 5 million non-approvers and counting the approvers in the sample…

Sampling 400 respondents from a population of 8 million approvers and 5 million non-approvers and counting the approvers in the sample…

…is like…

Sampling 400 respondents from a population of 8 million approvers and 5 million non-approvers and counting the approvers in the sample…

…is like…

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

Sampling 400 respondents from a population of 8 million approvers and 5 million non-approvers and counting the approvers in the sample…

…is like…

…drawing 400 times with replacement from the box _____ and summing the draws.

Sampling 400 respondents from a population of 8 million approvers and 5 million non-approvers and counting the approvers in the sample…

…is like…

…drawing 400 times with replacement from the box 8m 1s, 5m 0s and summing the draws.

Sampling 400 respondents from a population of 8 million approvers and 5 million non-approvers and counting the approvers in the sample…

…is like…

…drawing 400 times with replacement from the box 8m 1s, 5m 0s and summing the draws.

The number of approvers in our sample will be about _____ give or take _____ or so.

Sampling 400 respondents from a population of 8 million approvers and 5 million non-approvers and counting the approvers in the sample…

…is like…

…drawing 400 times with replacement from the box 8m 1s, 5m 0s and summing the draws.

The number of approvers in our sample will be about _____ give or take _____ or so.

$$\textbf{expected value for sum} = 400 \times \frac{8m}{13m} \approx 246$$

$$\textbf{standard error for sum} = \sqrt{400} \times \sqrt{\frac{8}{13} \times \frac{5}{13}} \approx 10$$

Sampling 400 respondents from a population of 8 million approvers and 5 million non-approvers and counting the approvers in the sample…

…is like…

…drawing 400 **times with replacement from the box** 8m 1s, 5m 0s **and summing the draws.**

The number of approvers in our sample will be about 246 give or take _____ or so.

$$\textbf{expected value for sum} = 400 \times \frac{8m}{13m} \approx 246$$

$$\textbf{standard error for sum} = \sqrt{400} \times \sqrt{\frac{8}{13} \times \frac{5}{13}} \approx 10$$

Sampling 400 respondents from a population of 8 million approvers and 5 million non-approvers and counting the approvers in the sample…

…is like…

…drawing 400 **times with replacement from the box** 8m 1s, 5m 0s **and summing the draws.**

The number of approvers in our sample will be about 246 give or take 10 or so.

$$\textbf{expected value for sum} = 400 \times \frac{8m}{13m} \approx 246$$

$$\textbf{standard error for sum} = \sqrt{400} \times \sqrt{\frac{8}{13} \times \frac{5}{13}} \approx 10$$

The number of approvers in our sample will be about 246 give or take 10 or so.

The number of approvers in our sample will be about 246 give or take 10 or so.

Question

The **percent** of approvers in our sample will be about _____ give or take _____ or so.

The number of approvers in our sample will be about 246 give or take 10 or so.

Question

The **percent** of approvers in our sample will be about _____ give or take _____ or so.

$$\begin{aligned}
\textbf{expected value for percent} &= \frac{\textbf{expected value for sum}}{\textbf{number of draws}} \times 100\,\% \\
&= \frac{246}{400} \times 100\,\% \\
&= 0.615 \times 100\,\% \\
&\approx 62\,\%
\end{aligned}$$

$$\begin{aligned}
\textbf{standard error for percent} &= \frac{\textbf{standard error for sum}}{\textbf{number of draws}} \times 100\,\% \\
&= \frac{10}{400} \times 100\,\% \\
&= 0.025 \times 100\,\% \\
&\approx 3\,\%
\end{aligned}$$

The number of approvers in our sample will be about 246 give or take 10 or so.

Question

The **percent** of approvers in our sample will be about 62% give or take _____ or so.

$$\begin{aligned}
\textbf{expected value for percent} &= \frac{\textbf{expected value for sum}}{\textbf{number of draws}} \times 100\,\% \\
&= \frac{246}{400} \times 100\,\% \\
&= 0.615 \times 100\,\% \\
&\approx 62\,\%
\end{aligned}$$

$$\begin{aligned}
\textbf{standard error for percent} &= \frac{\textbf{standard error for sum}}{\textbf{number of draws}} \times 100\,\% \\
&= \frac{10}{400} \times 100\,\% \\
&= 0.025 \times 100\,\% \\
&\approx 3\,\%
\end{aligned}$$

The number of approvers in our sample will be about 246 give or take 10 or so.

Question

The **percent** of approvers in our sample will be about 62% give or take 3% or so.

$$\begin{aligned}
\textbf{expected value for percent} &= \frac{\textbf{expected value for sum}}{\textbf{number of draws}} \times 100\,\% \\
&= \frac{246}{400} \times 100\,\% \\
&= 0.615 \times 100\,\% \\
&\approx 62\,\%
\end{aligned}$$

$$\begin{aligned}
\textbf{standard error for percent} &= \frac{\textbf{standard error for sum}}{\textbf{number of draws}} \times 100\,\% \\
&= \frac{10}{400} \times 100\,\% \\
&= 0.025 \times 100\,\% \\
&\approx 3\,\%
\end{aligned}$$

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

Assume 300m Americans and that 50% are pro-life and 50% are pro-choice.

Questions

1. What's the expected value and standard error for the number in your sample that are pro-life?

2. What's the expected value and standard error for the percent in the sample that are pro-life?

3. 95% of the time, the sample percentage will fall between ____ and _____.

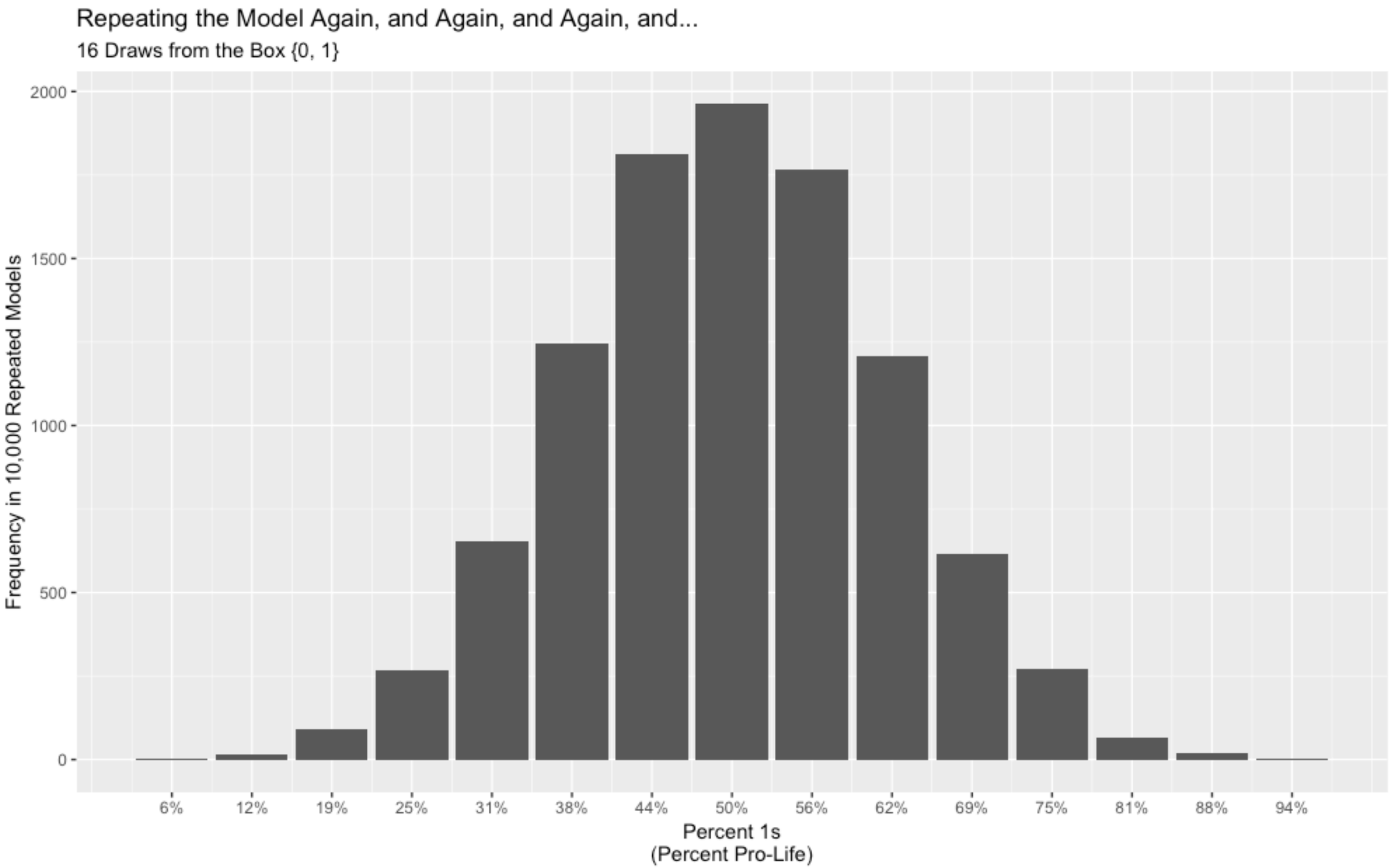
Exercise

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." Round to the nearest whole percent.**

2. **Add a yellow sticky note to the top of the column for your percent.**

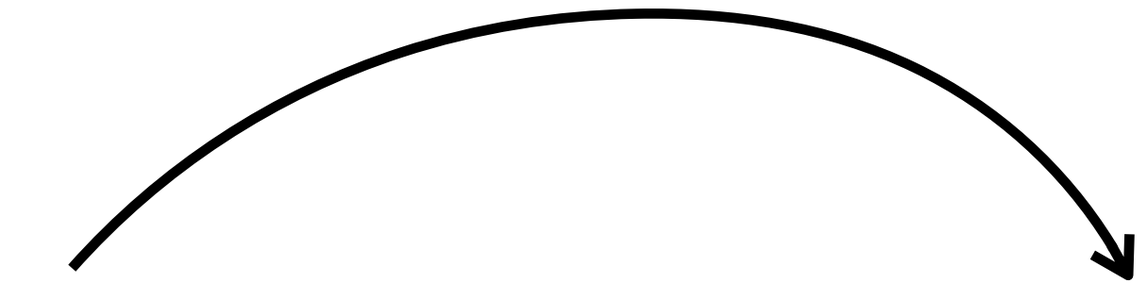
3. **If you have time, do it again!**


Repeating the Model Again, and Again, and Again, and...
16 Draws from the Box {0, 1}
Frequency in 10,000 Repeated Models
2000
1500
1000
500
0
6%
12%
19%
25%
31%
38%
44%
50%
56%
62%
69%
75%
81%
88%
94%
Percent 1s
(Percent Pro-Life)

Concluding Observations

population **sample**

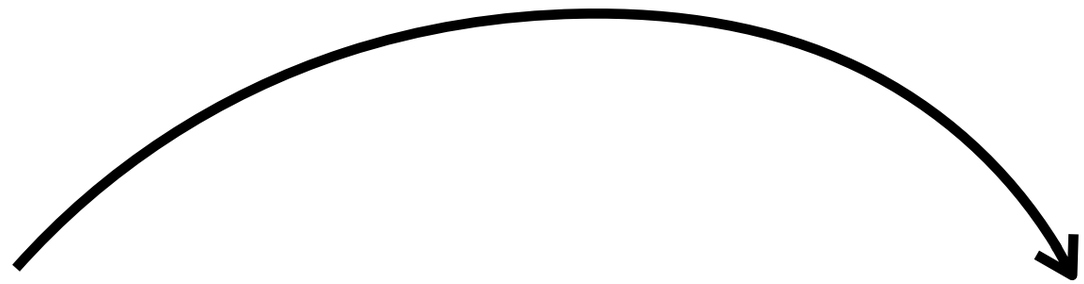
Concluding Observations



population **sample**

Concluding Observations

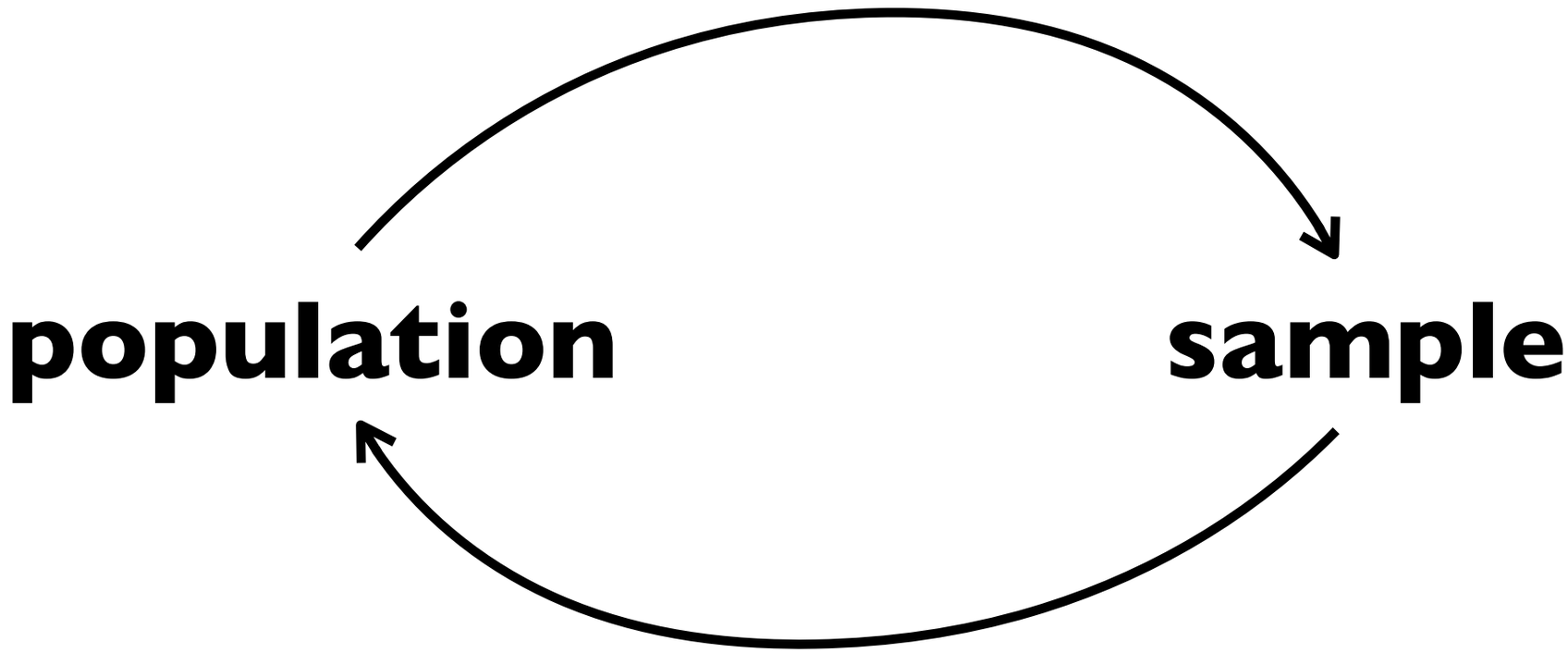
If we know what the population looks like, we can describe what the samples will look like.



population **sample**

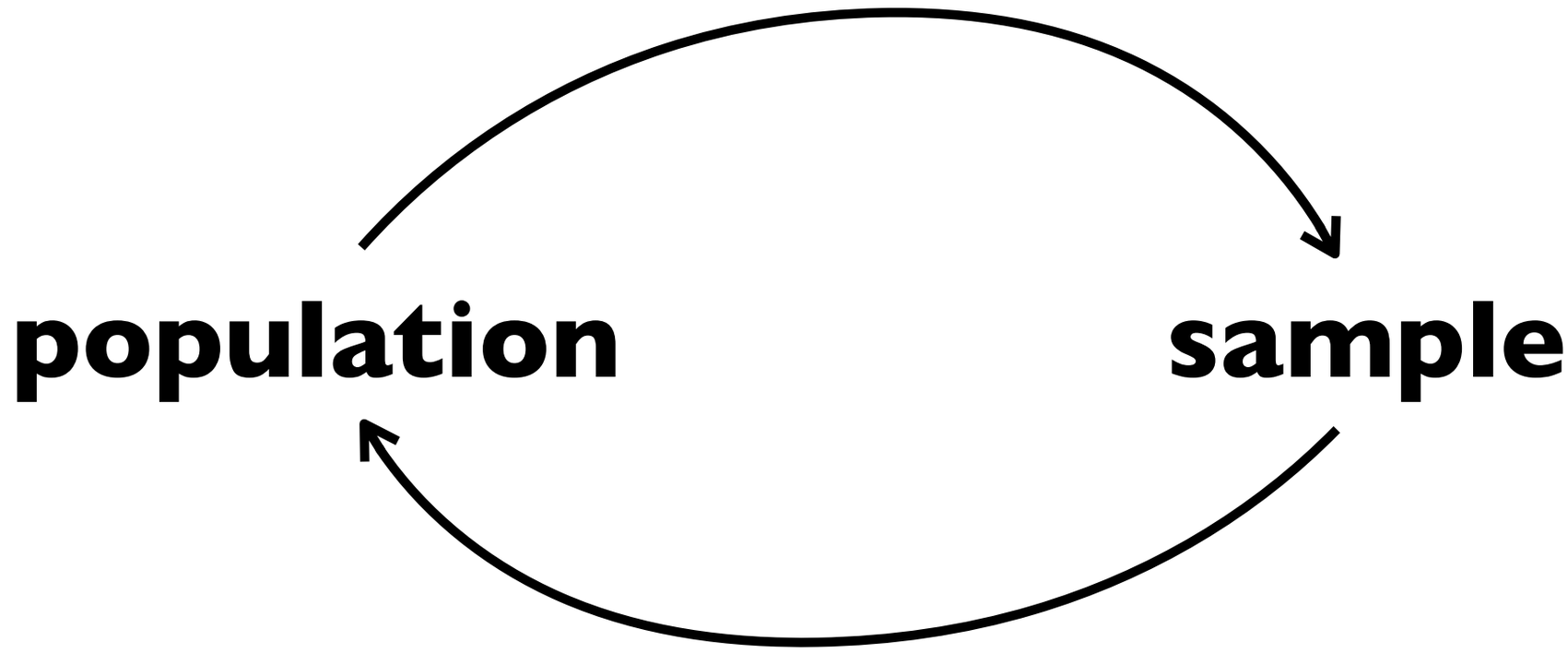
Concluding Observations

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



Concluding Observations

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



In practice, we only have a sample!
Based on the sample, what can we say about the population?

Concluding Observations

The expected sample percentage equals the population percentage.

Concluding Observations

The expected sample percentage equals the population percentage.

$$
\begin{aligned}
\textbf{expected value for percent} &= \frac{\textbf{expected value for sum}}{\textbf{number of draws}} \times 100\,\% \\
&= \frac{\textbf{number of draws} \times \textbf{avg. of box}}{\textbf{number of draws}} \times 100\,\% \\
&= \textbf{avg. of box} \times 100\,\% \\
&= \frac{\textbf{number of 1s}}{\textbf{total number}} \times 100\,\% \\
&= \textbf{fraction of 1s in population} \times 100\,\% \\
&= \textbf{percent of 1s in population}
\end{aligned}
$$

Concluding Observations

The sample percentage is about the population percentage give or take a SE or so.

Concluding Observations

population

The ~~sample~~ percentage is about the ~~population~~ percentage give or take a SE or so.

sample