Sample Surveys

(and the Numbered Ticket Model)

Connecting the Two

Review

basic rules of chance

chance processes
especially those we can model with a NTM

expected value and standard error
The average will be about _____ give or take _____ or so.

normal approximation

These are our theoretical tools!

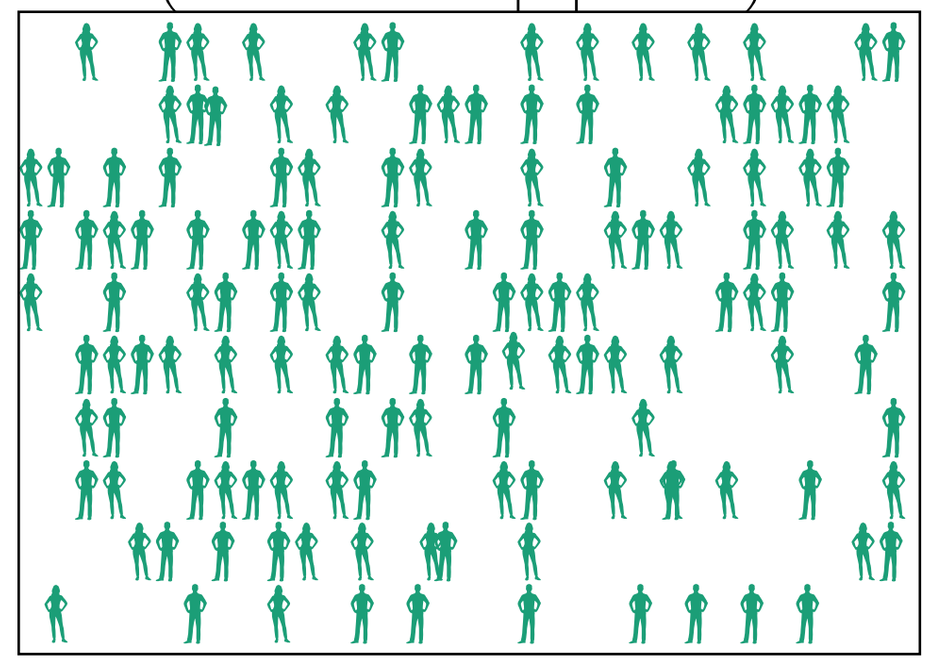
Sample Surveys

Six Key Concepts

population
(e.g., adults living in the US)

sampling method
(e.g., simple random sample, convenience sample)

sample
(a subset of the population)



parameter
(e.g., percent that approve)

unknown

inference
(i.e., how close can we expect the sample statistic to be to the population parameter?)

statistic
(e.g., percent that approve)

40%

population **sampling method** **sample**

parameter **inference** **statistic**

For the details, refer to section 12.1 of the notes.

Simple Random Sample

simple random sample: draw your sample from the population at random without replacement.

With a simple random sample, we randomly select individuals from the population without replacement, so each individual in the population is **equally-likely** to be included in the sample.

Extensions

- Alternatives to the simple random sample: notes, sections12.4 and 12.5
- A particularly bad (infamous) sample survey: notes, exercise 12.2
- A summary of YouGov's method: exercise 12.3
- A recent academic summary of survey methods: exercise 12.4

Now let's apply our theoretical tools to

sample surveys

(with a simple random sample*).

*the exact equations will vary across the sampling method, but the answers will not be too different, and the intuition applies.

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 proportion 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 computing proportion of approvers in the sample…

…is like…

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

$$\textbf{expected value} = \frac{8m}{13m} = \frac{8}{13} \approx 0.62$$

$$\textbf{standard error} = \frac{\sqrt{\frac{8}{13} \times \frac{5}{13}}}{\sqrt{400}} \approx \frac{\sqrt{0.62 \times 0.28}}{20} \approx \frac{0.49}{20} = 0.024$$

The proportion of approvers in our sample will be about 0.62 give or take 0.024 or so.

The percent of approvers in our sample will be about 62% give or take 2.4% or so.

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 <u>proportion</u> in the sample that are pro-life?
2. 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. (Each of these 16 tosses represents a "study.")
2. Add a yellow sticky note to the top of the column for your percent.
3. If you have time, do it again!
4. Let's see what the distribution of responses looks like across the several studies. Even though the studies are use the exact same methods, the answers will differ. The key is that we be able to guess the size of this variation (see the "Questions" in the middle of this slide).

We can do this with a computer, too.

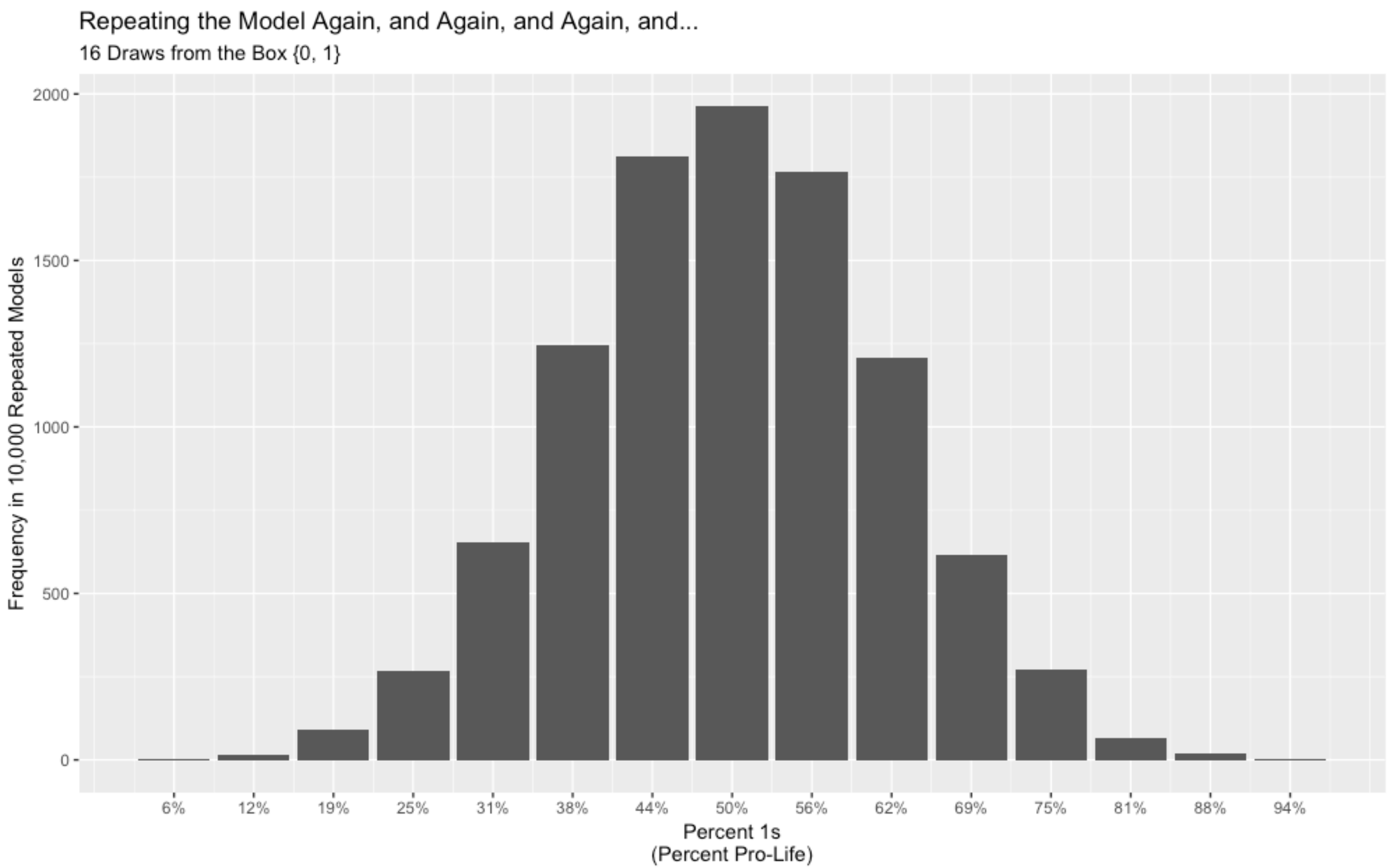
```r
library(foreach)

n_draws <- 16
box <- c(0, 1)

times(10) %do%
  mean(sample(box, size = n_draws, replace = TRUE))
```

```
> library(foreach)
>
> n_draws <- 16
> box <- c(0, 1)
>
> times(10) %do%
+   mean(sample(box, size = n_draws, replace = TRUE))
 [1] 0.2500 0.3750 0.4375 0.3125 0.5000 0.5625 0.3125 0.4375 0.3750 0.6875
```

```
> library(foreach)
>
> n_draws <- 16
> box <- c(0, 1)
>
> times(10000) %do%
+   mean(sample(box, size = n_draws, replace = TRUE))
   [1] 0.5625 0.5000 0.4375 0.5000 0.6875 0.6250 0.5625 0.5625 0.3125 0.6250 0.6875 0.6250 0.5625 0.3125 0.6250 0.3750 0.5000 0.5000 0.4375 0.5000
  [21] 0.3750 0.6875 0.3750 0.5625 0.6250 0.3750 0.5000 0.3125 0.4375 0.5625 0.8125 0.3125 0.6250 0.4375 0.5625 0.6875 0.8125 0.3750 0.5000 0.6875
  [41] 0.3750 0.2500 0.5000 0.3750 0.5000 0.6250 0.6250 0.1875 0.3750 0.6875 0.3750 0.5000 0.5625 0.6875 0.6250 0.5000 0.3750 0.5625 0.6875 0.5625
  [61] 0.5000 0.5000 0.5625 0.7500 0.6250 0.6250 0.4375 0.4375 0.5625 0.3750 0.5000 0.5625 0.7500 0.5000 0.4375 0.5625 0.6875 0.5625 0.5000 0.3750
  [81] 0.5625 0.3750 0.6250 0.5625 0.4375 0.5000 0.6875 0.4375 0.5000 0.8125 0.4375 0.5625 0.5000 0.4375 0.6875 0.6250 0.7500 0.7500 0.4375 0.5000
 [101] 0.5000 0.3125 0.5625 0.3125 0.3750 0.6250 0.4375 0.3125 0.5000 0.4375 0.6250 0.2500 0.6250 0.5625 0.6250 0.5000 0.5625 0.6250 0.5000 0.6250
 [121] 0.7500 0.4375 0.6250 0.6875 0.3125 0.4375 0.5625 0.1875 0.5000 0.5625 0.8125 0.4375 0.5625 0.5000 0.3750 0.5625 0.5625 0.5625 0.6250 0.5000
 [141] 0.5625 0.3125 0.5625 0.4375 0.7500 0.6250 0.5000 0.5000 0.5625 0.6250 0.4375 0.3750 0.5625 0.6250 0.5625 0.3750 0.5000 0.5625 0.4375 0.5000
 [161] 0.4375 0.4375 0.5000 0.5000 0.5625 0.5000 0.3750 0.6250 0.3750 0.4375 0.3750 0.5625 0.3750 0.3125 0.4375 0.7500 0.4375 0.5625 0.5000 0.6250
 [181] 0.6250 0.5625 0.5625 0.5625 0.5625 0.4375 0.5000 0.5000 0.3125 0.5625 0.3750 0.5000 0.3750 0.5000 0.6250 0.5625 0.5625 0.2500 0.5000 0.4375
 [201] 0.6250 0.4375 0.5000 0.5000 0.5000 0.2500 0.6875 0.6250 0.2500 0.6250 0.6250 0.4375 0.5625 0.3750 0.6250 0.4375 0.5000 0.5000 0.4375 0.5000
 [221] 0.5625 0.6250 0.6250 0.5625 0.4375 0.5625 0.5000 0.5625 0.5625 0.5625 0.3125 0.5625 0.3125 0.5625 0.5000 0.6250 0.6875 0.6875 0.6250 0.3125
 [241] 0.5625 0.6250 0.5000 0.4375 0.6875 0.6875 0.3750 0.5625 0.6250 0.4375 0.5000 0.2500 0.5625 0.6250 0.5625 0.5625 0.5000 0.5625 0.6250 0.7500
 [261] 0.5000 0.3125 0.3125 0.5000 0.5625 0.6250 0.5625 0.1875 0.4375 0.5000 0.3750 0.5625 0.4375 0.6250 0.7500 0.3125 0.4375 0.6250 0.3750 0.2500
 [281] 0.6875 0.3125 0.3750 0.4375 0.6875 0.4375 0.3125 0.3750 0.3750 0.6875 0.5625 0.4375 0.5625 0.6250 0.4375 0.6250 0.5000 0.4375 0.5000 0.4375
 [301] 0.5000 0.3750 0.5625 0.5625 0.5000 0.4375 0.4375 0.4375 0.6250 0.2500 0.5000 0.4375 0.5000 0.3750 0.5000 0.3750 0.4375 0.3750 0.5625 0.5000
 [321] 0.5625 0.7500 0.4375 0.3125 0.3750 0.4375 0.5000 0.1875 0.5625 0.5000 0.5625 0.8125 0.5625 0.5625 0.5625 0.6875 0.5000 0.3750 0.6250 0.3750
 [341] 0.5625 0.5625 0.3125 0.4375 0.5625 0.3750 0.7500 0.3125 0.3125 0.6250 0.5625 0.4375 0.3750 0.6250 0.4375 0.4375 0.3750 0.3125 0.5000 0.4375
 [361] 0.6250 0.4375 0.5000 0.6875 0.3125 0.5000 0.3750 0.5625 0.4375 0.4375 0.6250 0.4375 0.2500 0.4375 0.5000 0.2500 0.6875 0.5000 0.6250 0.3125
 [381] 0.4375 0.4375 0.3750 0.3750 0.5000 0.3750 0.4375 0.3125 0.2500 0.3750 0.5625 0.4375 0.3750 0.1875 0.5000 0.6875 0.5625 0.5625 0.3125 0.3125
 [401] 0.5625 0.4375 0.5625 0.6250 0.6250 0.3750 0.3750 0.5625 0.6250 0.5625 0.5000 0.3750 0.6250 0.4375 0.5625 0.6250 0.3125 0.3750 0.6875 0.5625
 [421] 0.5000 0.3750 0.4375 0.3750 0.4375 0.6875 0.5625 0.3125 0.5000 0.5625 0.3125 0.6250 0.7500 0.5625 0.5000 0.6875 0.6250 0.8750 0.6250 0.6250
 [441] 0.5000 0.4375 0.3750 0.3750 0.2500 0.2500 0.5625 0.3750 0.6875 0.4375 0.2500 0.3125 0.3750 0.6250 0.5000 0.4375 0.4375 0.4375 0.3750 0.1875
 [461] 0.3750 0.6250 0.6875 0.6250 0.6250 0.5000 0.5625 0.6250 0.5625 0.6250 0.3750 0.3125 0.3750 0.3750 0.5000 0.5625 0.5000 0.4375 0.4375 0.2500
 [481] 0.5000 0.6875 0.5625 0.5000 0.5000 0.5000 0.6250 0.5000 0.5000 0.5000 0.4375 0.5000 0.6875 0.6250 0.5000 0.5000 0.2500 0.6250 0.5000 0.4375
 [501] 0.4375 0.5000 0.6250 0.4375 0.5625 0.4375 0.4375 0.5625 0.6250 0.7500 0.5000 0.4375 0.7500 0.6875 0.5000 0.3125 0.4375 0.5625 0.6250 0.5625
 [521] 0.5625 0.5625 0.7500 0.6875 0.6250 0.3750 0.6250 0.3750 0.5625 0.6875 0.6250 0.5000 0.3125 0.4375 0.5625 0.6250 0.4375 0.3125 0.1875 0.3750
 [541] 0.4375 0.6875 0.3750 0.6875 0.5000 0.3750 0.5000 0.3125 0.4375 0.6250 0.4375 0.3750 0.3750 0.3125 0.5000 0.4375 0.3750 0.7500 0.5000 0.3750
 [561] 0.5000 0.3125 0.3750 0.6250 0.5625 0.4375 0.4375 0.4375 0.5000 0.3750 0.6875 0.5000 0.6250 0.5000 0.4375 0.5625 0.7500 0.3125 0.3125 0.5625
 [581] 0.8125 0.4375 0.5000 0.6875 0.2500 0.5625 0.2500 0.3750 0.3750 0.5625 0.5000 0.5625 0.3750 0.8125 0.4375 0.1875 0.4375 0.5000 0.5000 0.5000
 [601] 0.2500 0.5000 0.5000 0.5000 0.5625 0.6250 0.6250 0.3125 0.6250 0.4375 0.5625 0.5000 0.8750 0.3125 0.5625 0.4375 0.4375 0.5000 0.6250 0.3750
 [621] 0.3750 0.6875 0.5000 0.8125 0.6250 0.4375 0.5000 0.5625 0.6250 0.4375 0.5000 0.6250 0.3750 0.6875 0.5625 0.5625 0.4375 0.5625 0.3750 0.6250
 [641] 0.5625 0.5000 0.5000 0.3750 0.5000 0.5000 0.3750 0.3750 0.6875 0.5000 0.5000 0.4375 0.5000 0.5000 0.3125 0.6250 0.5000 0.7500 0.5000 0.5000
 [661] 0.6875 0.3750 0.4375 0.4375 0.2500 0.5000 0.5000 0.5000 0.7500 0.4375 0.6250 0.5625 0.5000 0.5000 0.5000 0.3750 0.6250 0.5000 0.6250 0.3750
 [681] 0.5625 0.5000 0.3750 0.5625 0.4375 0.3125 0.3750 0.5625 0.5000 0.4375 0.4375 0.3125 0.3750 0.6875 0.5000 0.6875 0.4375 0.5625 0.6875 0.4375
 [701] 0.6875 0.5000 0.4375 0.5625 0.8125 0.6250 0.5000 0.5625 0.3750 0.5000 0.5625 0.6875 0.3750 0.6250 0.3125 0.4375 0.4375 0.6250 0.6250 0.6250
 [721] 0.5625 0.4375 0.3125 0.4375 0.5625 0.5000 0.5625 0.4375 0.5625 0.5625 0.3750 0.4375 0.6250 0.4375 0.3750 0.3125 0.6875 0.6875 0.3750 0.5000
 [741] 0.6875 0.3750 0.5000 0.3125 0.4375 0.4375 0.4375 0.5000 0.3750 0.5625 0.4375 0.3750 0.5000 0.4375 0.4375 0.3125 0.5625 0.6250 0.5000 0.3750
 [761] 0.4375 0.5625 0.2500 0.3750 0.3750 0.3750 0.5625 0.2500 0.5625 0.6250 0.6250 0.3125 0.5625 0.4375 0.6250 0.2500 0.5625 0.3125 0.3125 0.5000
 [781] 0.6250 0.3750 0.4375 0.5000 0.3750 0.3750 0.6875 0.3750 0.6875 0.5625 0.3750 0.4375 0.5625 0.5000 0.6875 0.5625 0.4375 0.4375 0.6250 0.5000
 [801] 0.6250 0.5625 0.5625 0.5625 0.8125 0.3750 0.5625 0.5625 0.6250 0.7500 0.5000 0.3750 0.3125 0.5625 0.3750 0.3750 0.3750 0.4375 0.4375 0.4375
 [821] 0.5000 0.5000 0.5625 0.4375 0.8125 0.5000 0.4375 0.6250 0.6250 0.5625 0.5625 0.6875 0.4375 0.4375 0.4375 0.5000 0.3750 0.6250 0.6875 0.6250
 [841] 0.4375 0.4375 0.5625 0.4375 0.5000 0.5000 0.6250 0.5000 0.8125 0.5000 0.6250 0.3125 0.3750 0.5625 0.6875 0.4375 0.6250 0.6250 0.5625 0.5000
 [861] 0.5625 0.4375 0.5625 0.5625 0.5625 0.3750 0.4375 0.5625 0.5000 0.5000 0.8125 0.5000 0.5000 0.3750 0.5625 0.5625 0.6875 0.3750 0.5000 0.3750
 [881] 0.5000 0.5625 0.6250 0.3125 0.6250 0.6875 0.4375 0.6250 0.4375 0.3125 0.4375 0.4375 0.3750 0.2500 0.6875 0.6250 0.6250 0.3750 0.5000 0.5000
 [901] 0.5625 0.6875 0.3750 0.3125 0.6250 0.8125 0.4375 0.5000 0.4375 0.3125 0.5000 0.4375 0.6250 0.3750 0.5625 0.6250 0.5000 0.3125 0.5000 0.3125
 [921] 0.5000 0.4375 0.3750 0.4375 0.3750 0.3750 0.3125 0.6875 0.5000 0.3125 0.4375 0.5625 0.5000 0.6250 0.4375 0.7500 0.6875 0.3125 0.5000 0.5625
 [941] 0.4375 0.3125 0.3750 0.5625 0.5000 0.3125 0.3125 0.5625 0.5625 0.5000 0.5000 0.3125 0.5000 0.4375 0.3750 0.3750 0.6875 0.2500 0.4375 0.5625
 [961] 0.5000 0.2500 0.6250 0.3125 0.5625 0.4375 0.5000 0.3750 0.5000 0.6250 0.5000 0.7500 0.6250 0.3750 0.3125 0.6875 0.4375 0.6250 0.5000 0.6875
 [981] 0.6250 0.5000 0.3750 0.5000 0.5000 0.4375 0.3750 0.5000 0.5625 0.6875 0.3750 0.3125 0.3750 0.3750 0.4375 0.5625 0.5625 0.6250 0.6250 0.5000
[ reached getOption("max.print") -- omitted 9000 entries ]
```


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)

An App to Repeatedly Execute the Box Model

Box Model: Draw _____ times from the box _____ with replacement and average the draws.

Number of Draws

16

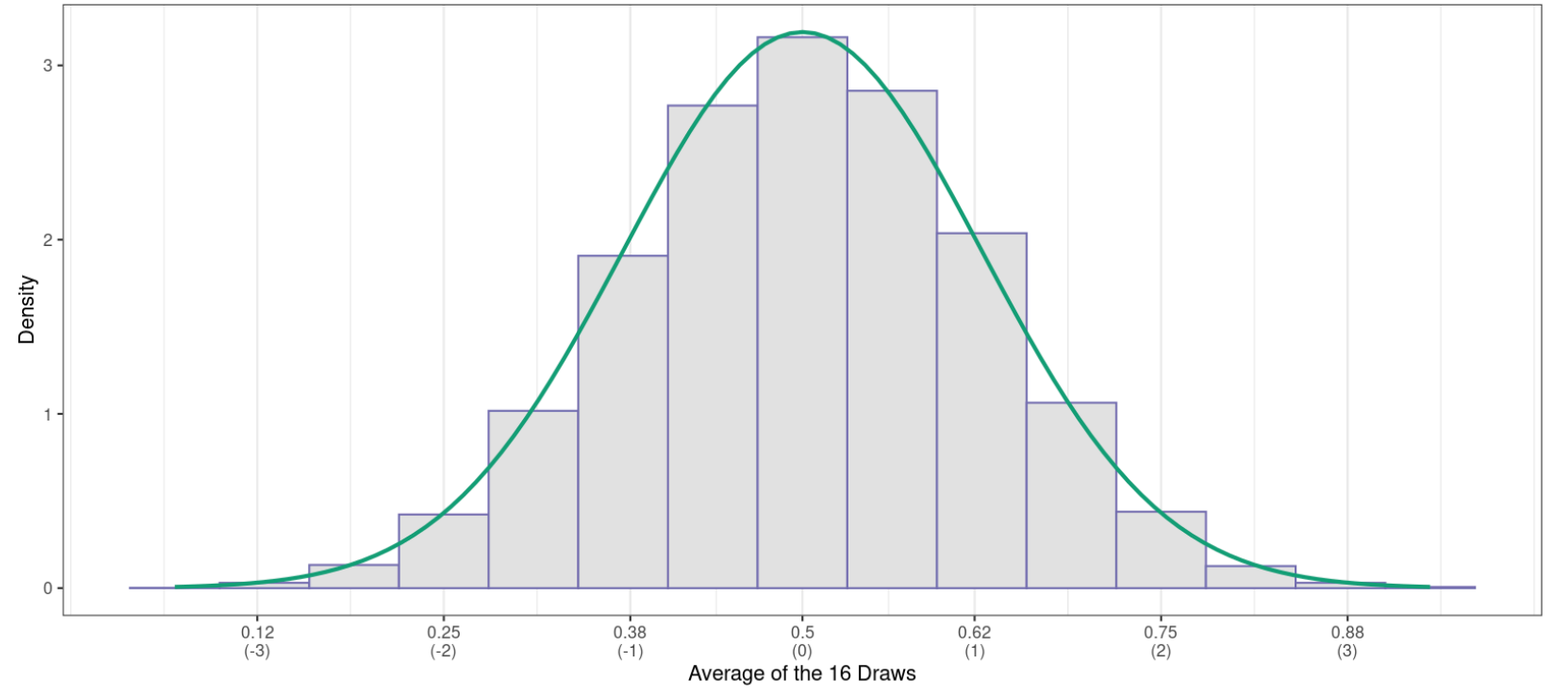
Box (separate with commas; e.g., 1, 2, 2, -14):

0, 1

Quantity	Value
Number of Draws	16
Box	0, 1
Five Example Executions (Average-of-the-Draws)	0.4375, 0.6875, 0.75, 0.625, 0.3125
Average of Box	0.5
SD of Box	0.5
Expected Value	0.5
Standard Error	0.12

A Histogram of Average-of-the-Draws

...if we execute the box model 10,000 times.



Concluding Observations

Concluding Observations

The sample proportion will be about the population percentage give or take a SE or so.

Concluding Observations

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?