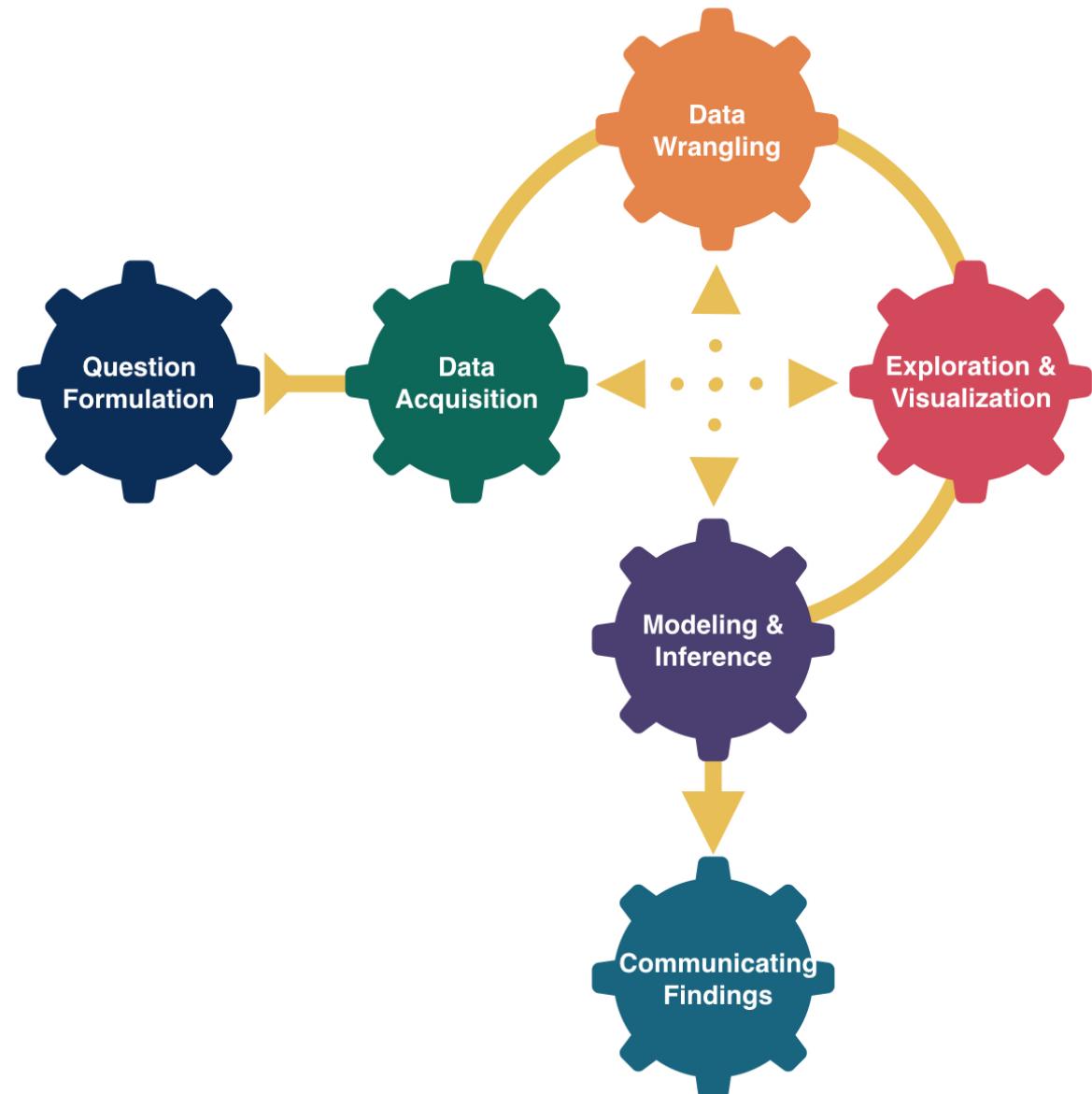


Hypothesis Testing



Kelly McConville
Stat 100
Week 9 | Fall 2023

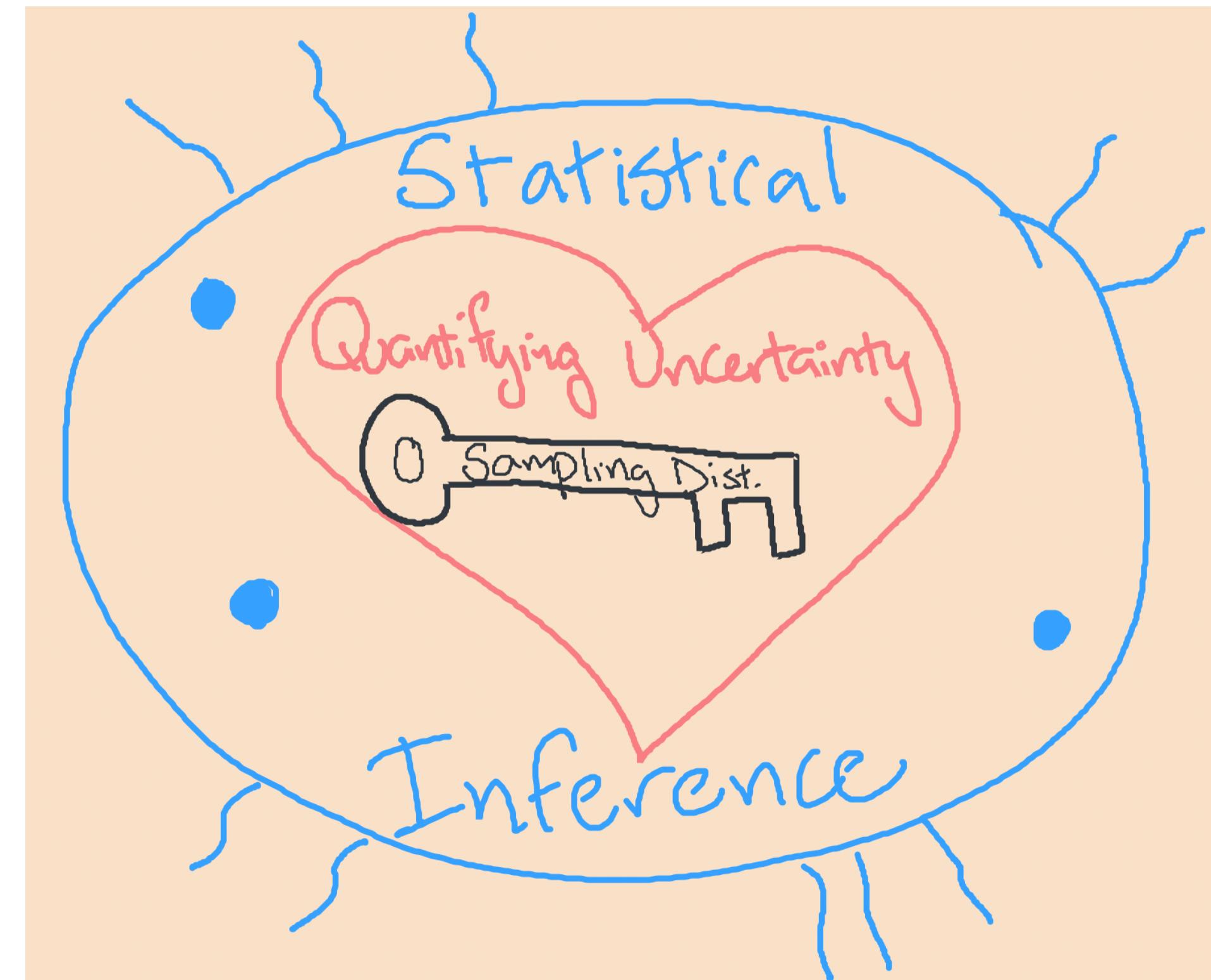
Announcements

- Don't forget that the midterm exam rewrites are due on Thursday at 5pm on Gradescope.
 - Make sure to use the Quarto doc in the Midterm Exam (Rewrites) project on Posit Cloud.
- 🎉 We are now accepting Course Assistant/Teaching Fellow applications for Stat 100 for next semester. To apply, fill out [this application](#) by **Nov 15th**.
 - About 10-12 hours of work per week.
 - Primary responsibilities: Attend weekly team meetings, lead a discussion section, hold office hours, grade assessments.

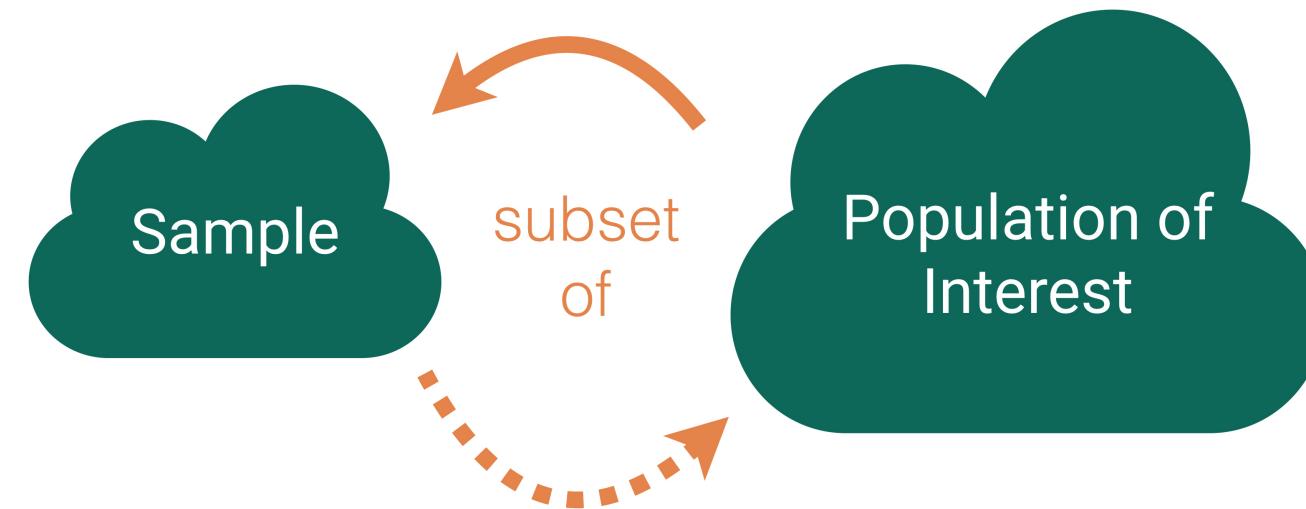
Goals for Today

- Set-up the structure of **hypothesis testing**
- Determine if Harvard students have ESP!

Question: How do sampling distributions help us quantify uncertainty?



Statistical Inference



Goal: Draw conclusions about the population based on a sample.

Main Flavors:

- Estimating numerical quantities.
- Testing conjectures.

Example: Does Extrasensory Perception (ESP) exist?



Daryl Bem and Ben Honorton

Bem and Honorton conducted extrasensory perception studies:

- A “sender” randomly chooses an object out of 4 possible objects and sends that information to a “receiver”.
- The “receiver” is then given a set of 4 possible objects and they must decide which one most resembles the object sent to them.

Out of 329 trials, the “receivers” correctly identified the object 106 times.

ESP Example

Let's consider the following questions:

- a. If ESP does not exist and the “receivers” are guessing, how often would we expect them to be correct?
- b. For each sample (set of 329 trials), do we expect the proportion of correct guesses to be equal? Why or why not?
- c. Is it possible to randomly guess correctly 106 out of 329 times (i.e., 32% of the time)?
- d. How unusual is it to guess correctly 106 out of 329 times if ESP doesn’t exist?

To help us answer d., we need a **sampling distribution** for the sample proportion where **we assume the “receivers” were purely guessing!**

Sampling Distribution of a Statistic

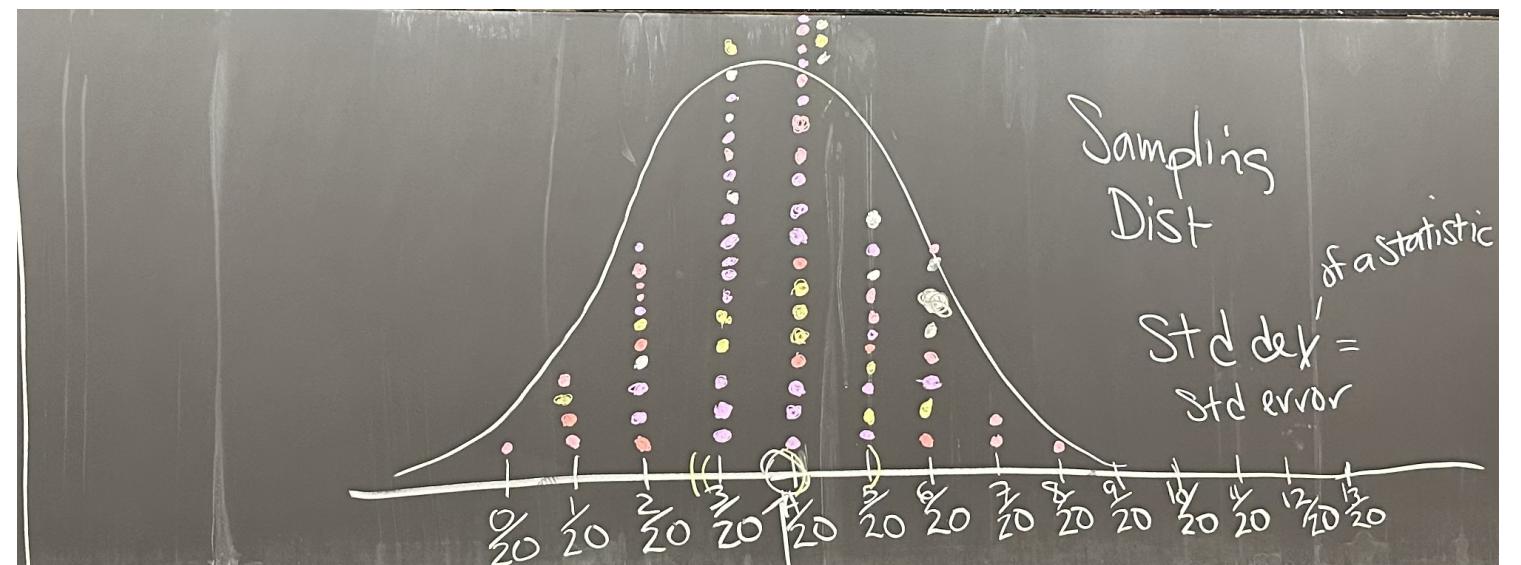
Steps for (Approximate) Distribution:

1. Decide on a sample size, n .
2. Randomly select a sample of size n from the population.
3. Compute the sample statistic.
4. Put the sample back in.
5. Repeat Steps (2) - (4) many (1000+) times.

Sampling Distribution of a Statistic

Steps for (Approximate) Distribution:

1. Decide on a sample size, n .
2. Randomly select a sample of size n from the population.
3. Compute the sample statistic.
4. Put the sample back in.
5. Repeat Steps (2) - (4) many (1000+) times.



Sampling Distribution Under No ESP

Steps for (Approximate) Distribution:

1. Decide on a sample size, n .
2. Randomly select a sample of size n from the population.

```
1 library(mosaic)
2 rflip(n = 329, prob = 0.25)
```

Flipping 329 coins [Prob(Heads) = 0.25] ...

```
T T T T T T T T T T T T T T T T H T H T T T T H T T T T T T T T
T H T T H T H T T T T T T T T H T T T T T H H H H T H H H T H T H H T
H H T T T T T T H T T T T H T T T T H H T T H T T T T H H T T T H
T T T T T H H H T T T T H T T T H T T H T T T T T T T T T T T T
T T H T T H T H T T T T H T T T H T T T T T H T T T T T T T T T
T H H T H T T T T T T H T T T H H T T T H T T T H T T T H T T
T T T H H T T T T T H T T T T H T H H T T T H H T T T H H T T
H T T T H H T T T T T H T T T T T H T T T H H T T T H H T T T H
T H T H T T T T T H H T T T H H H H T T T H H H H T T T H H H H
T T T H T
```

Number of Heads: 81 [Proportion Heads: 0.246200607902736]

Sampling Distribution Under No ESP

3. Compute the sample statistic.

```
1 rflip(n = 329, prob = 0.25, summarize = TRUE)
```

	n	heads	tails	prob
1	329	78	251	0.25

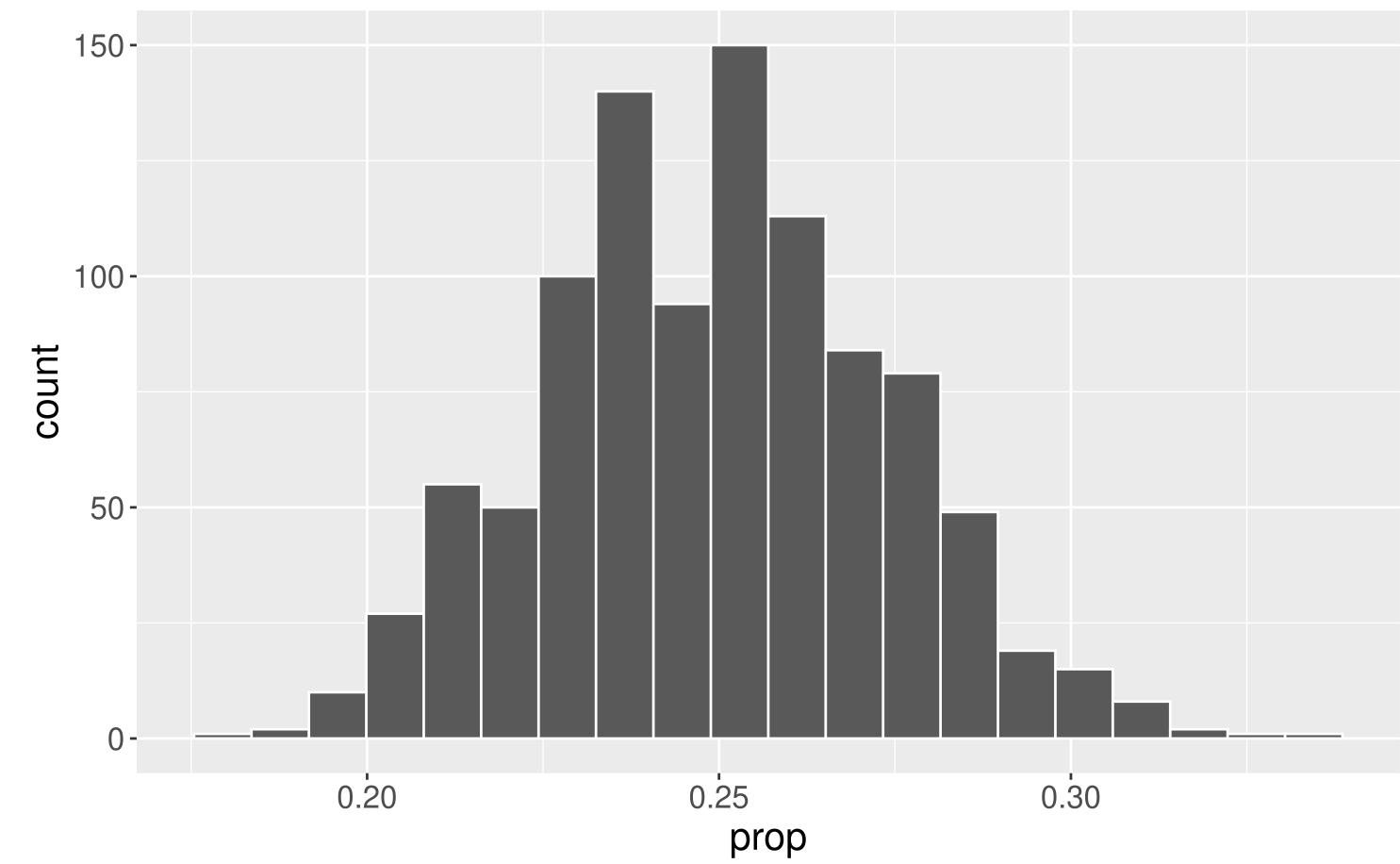
4. Put the sample back in.

5. Repeat Steps (2) - (4) many (1000+) times.

	n	heads	tails	prop
1	329	79	250	0.2401216
2	329	81	248	0.2462006
3	329	76	253	0.2310030
4	329	76	253	0.2310030
5	329	85	244	0.2583587
6	329	88	241	0.2674772
7	329	74	255	0.2249240
8	329	78	251	0.2370821
9	329	76	253	0.2310030
10	329	84	245	0.2553191
11	329	84	245	0.2553191
12	329	92	237	0.2796353
13	329	85	244	0.2583587
14	329	77	252	0.2340426
15	329	86	243	0.2610000

Sampling Distribution Under No ESP

```
1 ggplot(data = guess_sampling_dist,  
2         mapping = aes(x = prop)) +  
3         geom_histogram(color = "white",  
4                             bins = 20)
```

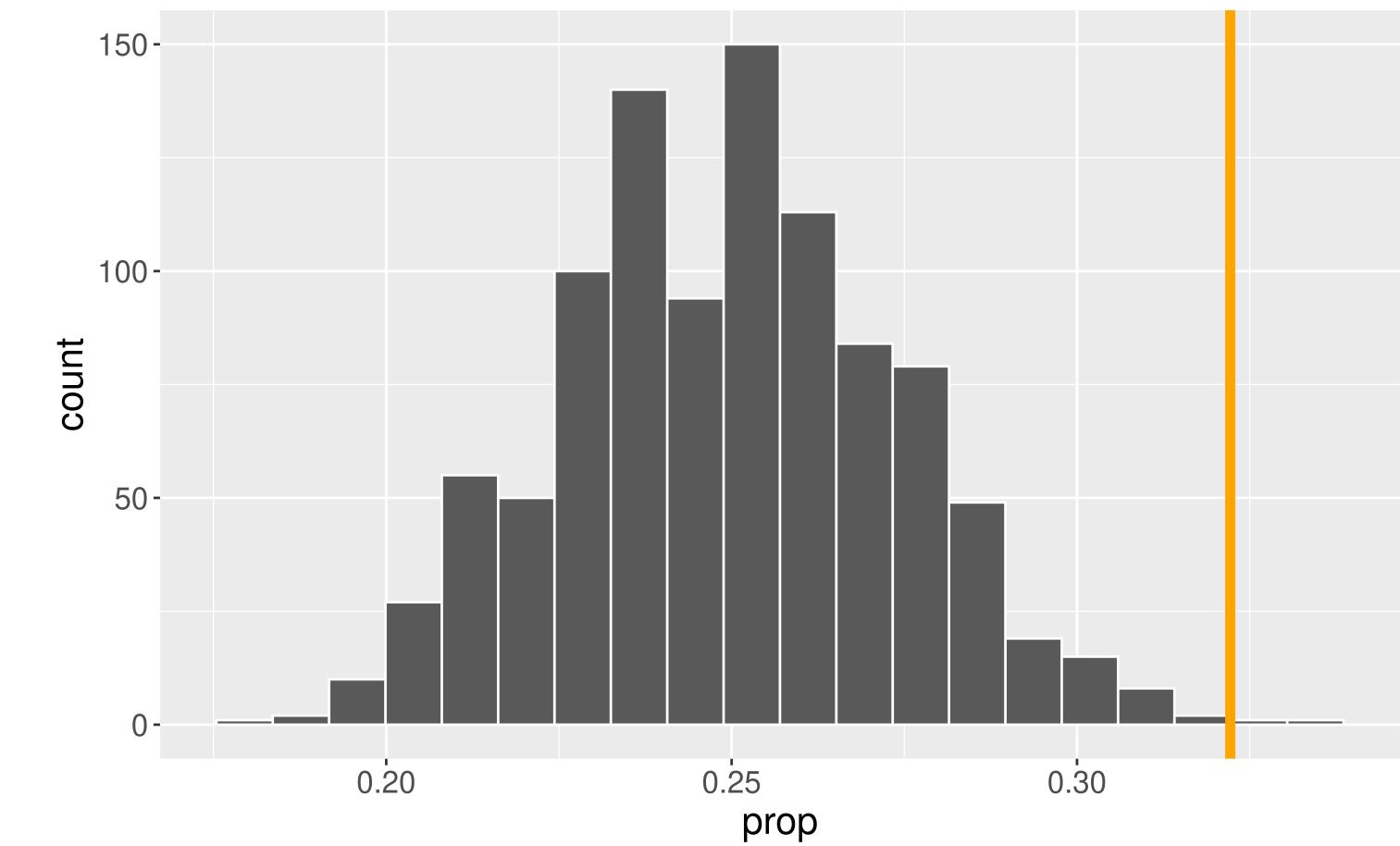


What value should our sampling distribution be centered around if the receivers are just guessing?

Sampling Distribution Under No ESP

- How do the study results compare to the sampling distribution under no ESP?
 - How unusual is it to guess correctly 106 out of 329 times if ESP doesn't exist?

```
1 p_hat <- 106/329
2 ggplot(data = guess_sampling_dist,
3         mapping = aes(x = prop)) +
4         geom_histogram(color = "white",
5                         bins = 20) +
6         geom_vline(xintercept = p_hat,
7                     color = "orange",
8                     size = 2)
```

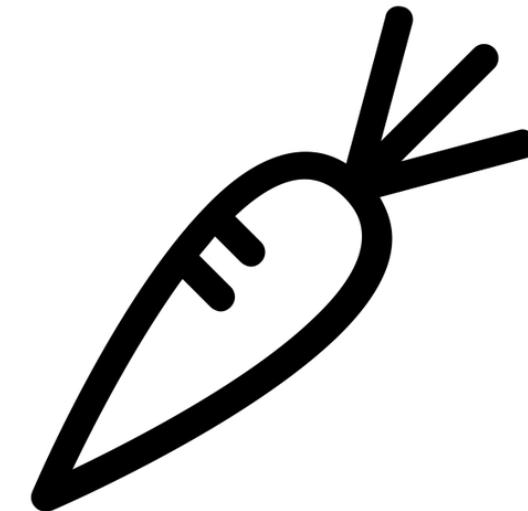


- Do Bem and Honorton have evidence that ESP exists?

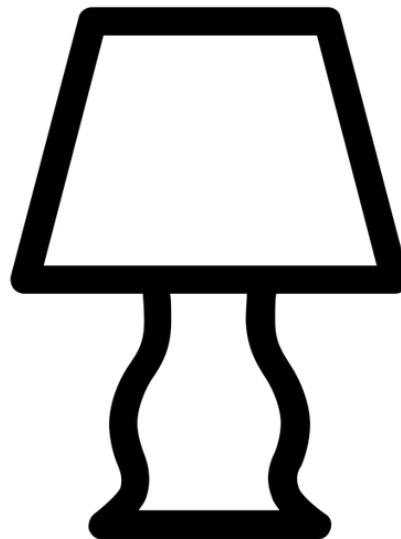
Do Harvardians Have ESP?

In pairs:

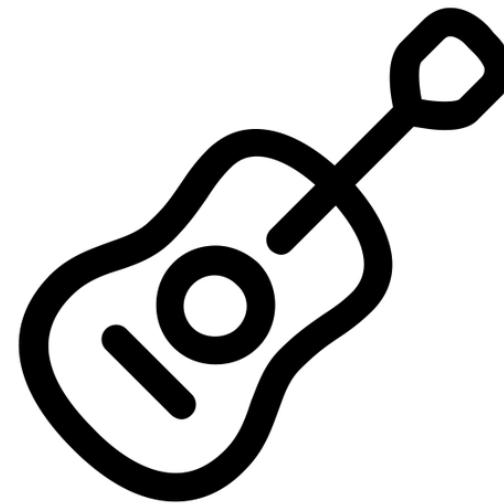
- Decide who is going to be the sender and who is going to be the receiver.
- **Sender:** Think of one of these images.
- **Receiver:** Guess which image the sender was thinking of.
- Now switch roles and do it again!
- Once you have both played each role, each person should add a tally mark on the chalkboard.



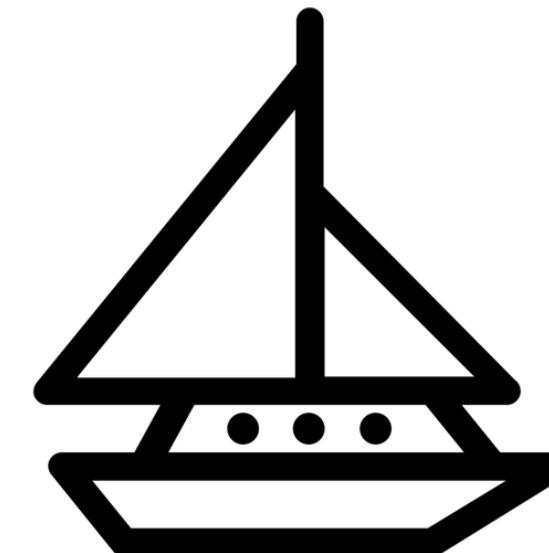
Created by Focus Lab
from Noun Project



Created by Focus Lab
from Noun Project



Created by Focus Lab
from Noun Project



Created by Focus Lab
from Noun Project

Do Harvardians Have ESP?

What do we need to modify in the code to answer the question?

```
1 guess_sampling_dist <- do(1000)*rflip(n = 329, prob = 0.25)
2 p_hat <- 106/329
3 ggplot(data = guess_sampling_dist, mapping = aes(x = prop)) +
4   geom_histogram(color = "white", bins = 20) +
5   geom_vline(xintercept = p_hat, color = "orange", size = 2)
```

Hypothesis Testing

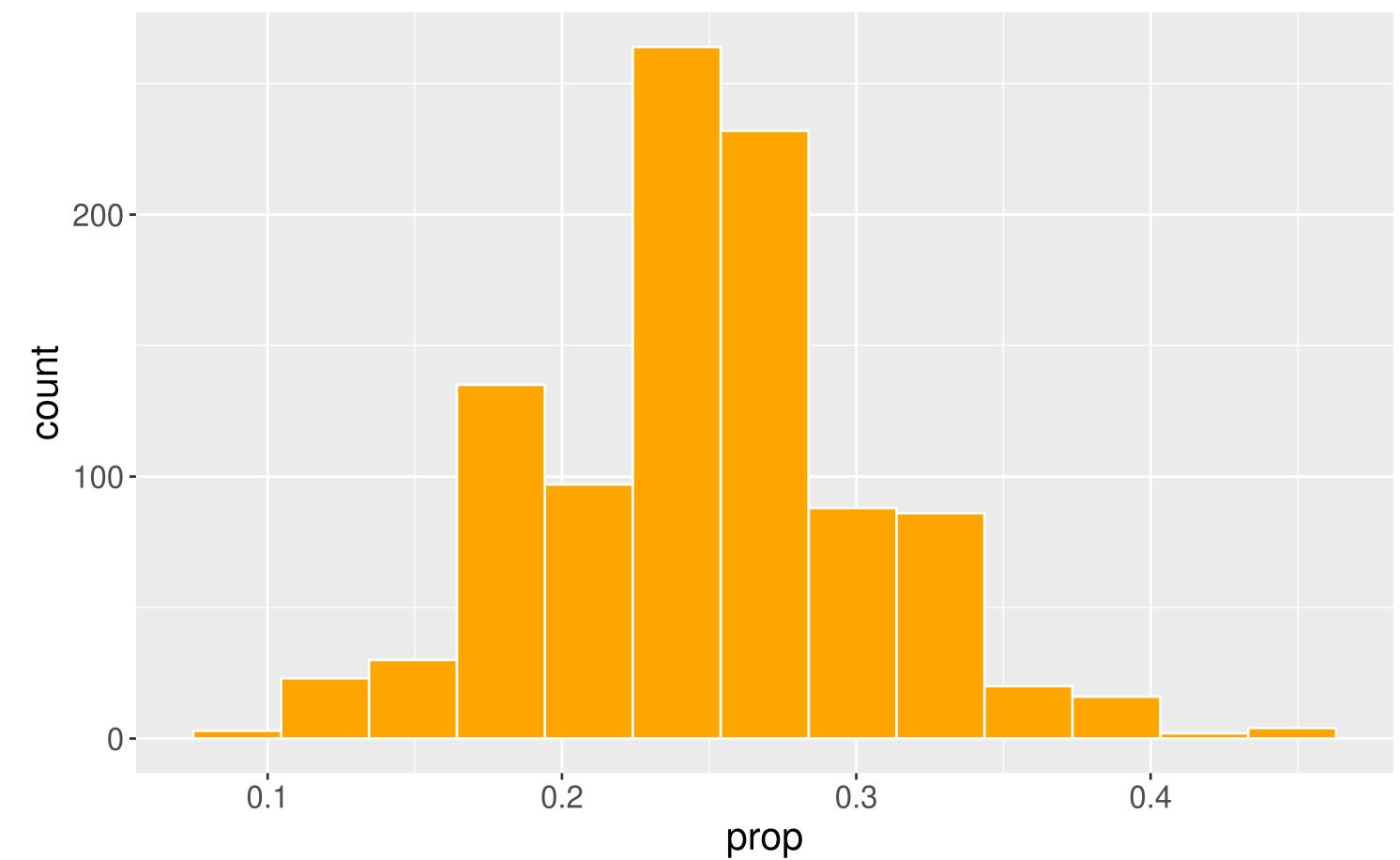
Big Idea:

- Make an assumption about the population **parameter**.
- Generate a sampling distribution for a *test statistic* based on that assumption.
 - Called a **null distribution**
- See if the test statistic based on the observed sample aligns with the generated sampling distribution or not.
- If it does, then we didn't learn much.
 - (Didn't prove the parameter equals the assumed value but it is still plausible)
- If it doesn't, then we have evidence that our assumption about the parameter was wrong.

ESP Example

Big Idea:

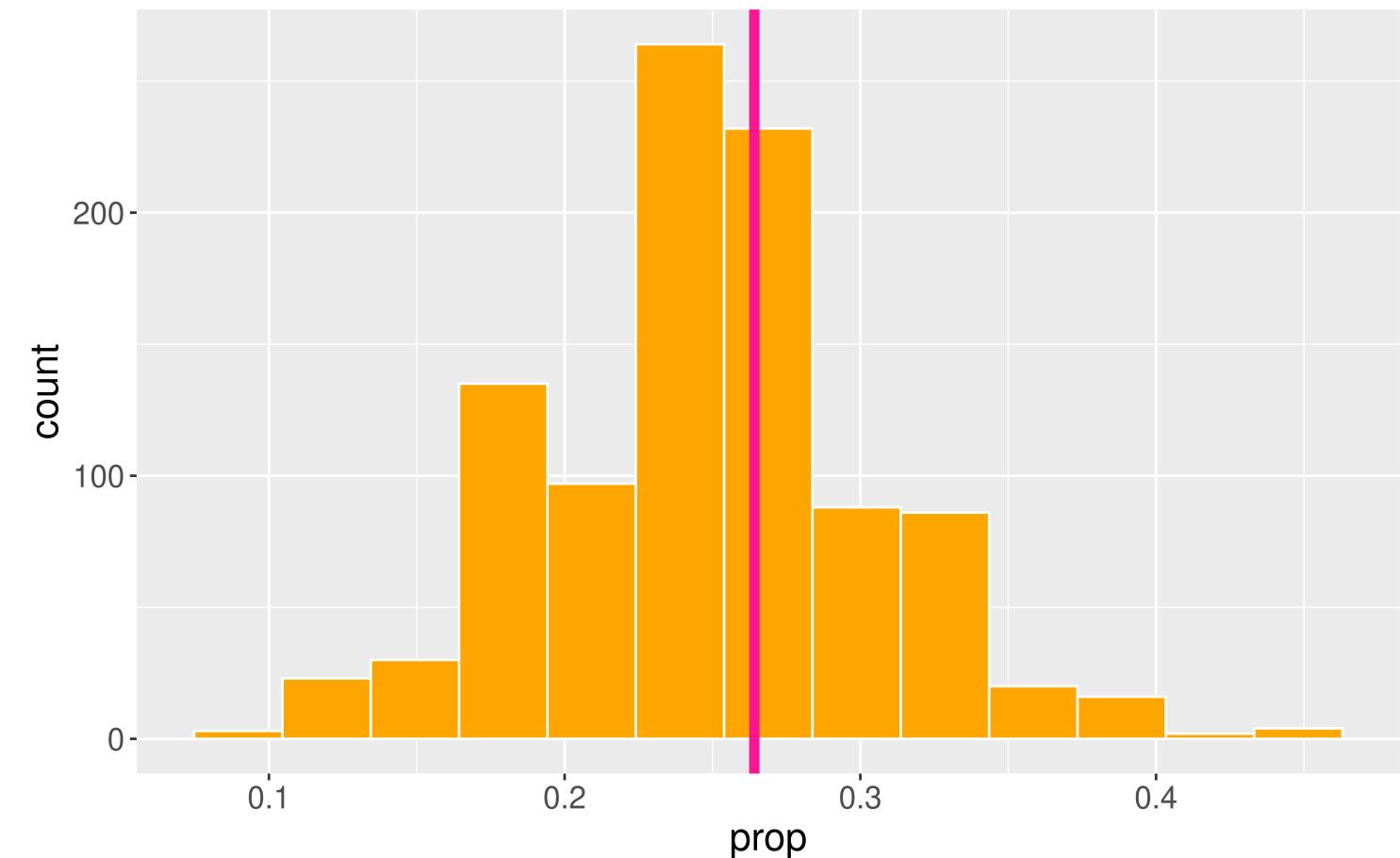
- Make an assumption about the population parameter.
 - Ex: ESP doesn't exist. p , probability of guessing correctly, equals 0.25.
- Generate a sampling distribution for a *test statistic* based on that assumption.
 - Called a **null distribution**



ESP Example

Big Idea:

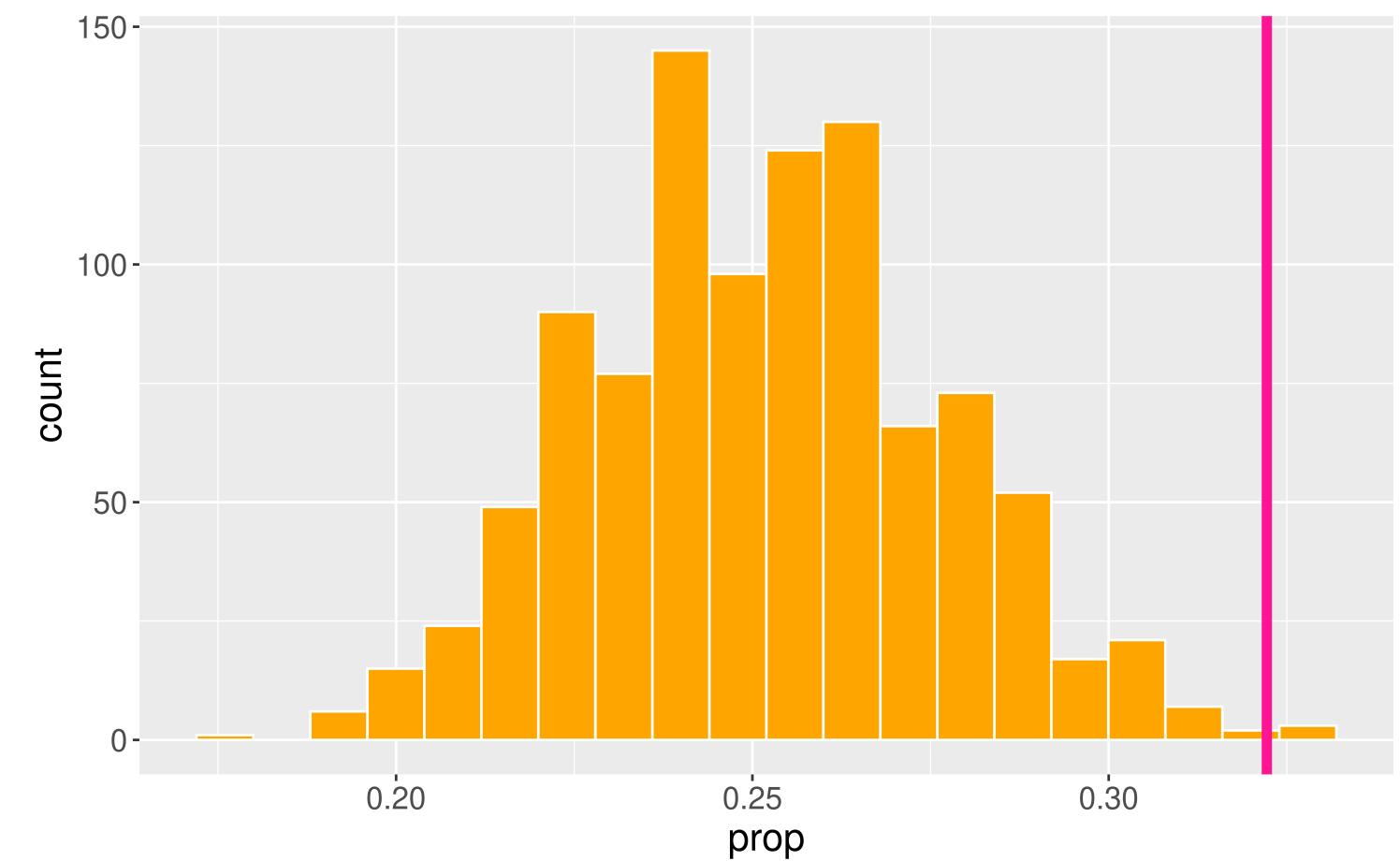
- See if the test statistic based on the observed sample aligns with the generated sampling distribution or not.
 - Ex: It is in the center-ish of the distribution. It isn't an unusual value.
- If it does, then we didn't learn much. (Didn't prove the parameter equals the assumed value but it is still plausible)
 - It is still possible that ESP doesn't exist.



ESP Example

Big Idea:

- See if the test statistic based on the observed sample aligns with the generated sampling distribution or not.
 - It is far in the tails of the distribution. It is an unusual value.
- If it doesn't, then we have evidence that our assumption about the parameter was wrong.
 - We have evidence ESP exists.



Let's Take a Step Back from Our Last Statement...

- Two important words in data analysis:
 - Reproducibility
 - Replicability
- **Reproducibility**: If I give you the raw data and my write-up, you will get to the exact same final numbers that I did.
- By using **Quarto** Documents, we are learning a **reproducible** workflow.
- **Replicability**: If you follow my study design but collect new data (i.e. repeat my study on new subjects), you will come to the same conclusions that I did.

Replication Crisis

- Science is going through a **replication crisis** right now.
 - In cancer science, many “discoveries” don’t hold up
 - Estimating the reproducibility of psychological science
 - Psychology Is Starting To Deal With Its Replication Problem
- And, sadly, **replication** studies of Bem and Honorton’s ESP trials typically failed to find evidence of ESP.

Reminders:

- Don't forget that the midterm exam rewrites are due on Thursday at 5pm on Gradescope.
 - Make sure to use the Quarto doc in the Midterm Exam (Rewrites) project on Posit Cloud.
- 🎉 We are now accepting Course Assistant/Teaching Fellow applications for Stat 100 for next semester. To apply, fill out [this application](#) by **Nov 15th**.
 - About 10-12 hours of work per week.
 - Primary responsibilities: Attend weekly team meetings, lead a discussion section, hold office hours, grade assessments.

