

Organ donors: Parameter vs. statistic

Parameter, p : true rate of complication

Statistic, \hat{p} : rate of complication in the sample = $\frac{3}{62} = 0.048$



Correlation vs. causation

Is it possible to assess the consultant's claim using the data?



Correlation vs. causation

Is it possible to assess the consultant's claim using the data?

No. The claim is that there is a causal connection, but the data are observational. For example, maybe patients who can afford a medical consultant can afford better medical care, which can also lead to a lower complication rate.

While it is not possible to assess the causal claim, it is still possible to test for an association using these data.

For this question we ask, **how likely is it that the low complication rate observed of $\hat{p} = 0.048$ be due solely to chance?**

Two claims

- **Null hypothesis:** "There is nothing going on"

Complication rate for this consultant is no different than the US average of 10%



Two claims

- **Null hypothesis:** "There is nothing going on"

Complication rate for this consultant is no different than the US average of 10%

- **Alternative hypothesis:** "There is something going on"

Complication rate for this consultant is **lower** than the US average of 10%



Two claims

- **Null hypothesis:** "There is nothing going on"

Complication rate for this consultant is no different than the US average of 10%

- **Alternative hypothesis:** "There is something going on"

Complication rate for this consultant is **lower** than the US average of 10%

In statistical hypothesis testing we always first assume that the null hypothesis is true and then see whether we reject or fail to reject this claim.



Hypothesis testing as a court trial

- Null hypothesis, H_0 : Defendant is innocent
- Alternative hypothesis, H_a : Defendant is guilty



Hypothesis testing as a court trial

- Null hypothesis, H_0 : Defendant is innocent
- Alternative hypothesis, H_a : Defendant is guilty
- Present the evidence: Collect data



Hypothesis testing as a court trial

- Null hypothesis, H_0 : Defendant is innocent
- Alternative hypothesis, H_a : Defendant is guilty
- Present the evidence: Collect data
- Judge the evidence: "Could these data plausibly have happened by chance if the null hypothesis were true?"
 - Yes: Fail to reject H_0
 - No: Reject H_0



The hypothesis testing framework



The hypothesis testing framework

- 1 Start with two hypotheses about the population: the null hypothesis and the alternative hypothesis.



The hypothesis testing framework

- 1 Start with two hypotheses about the population: the null hypothesis and the alternative hypothesis.
- 2 Choose a (representative) sample, collect data, and analyze the data.



The hypothesis testing framework

- 1 Start with two hypotheses about the population: the null hypothesis and the alternative hypothesis.
- 2 Choose a (representative) sample, collect data, and analyze the data.
- 3 Figure out how likely it is to see data like what we observed, IF the null hypothesis were in fact true (called a **p-value**)



The hypothesis testing framework

- 1 Start with two hypotheses about the population: the null hypothesis and the alternative hypothesis.
- 2 Choose a (representative) sample, collect data, and analyze the data.
- 3 Figure out how likely it is to see data like what we observed, IF the null hypothesis were in fact true (called a **p-value**)
- 4 If our data would have been extremely unlikely if the null hypothesis were true, then we reject it in favor of the alternative hypothesis.

Otherwise, we cannot reject the null hypothesis



1

Defining the hypotheses

Remember, the null and alternative hypotheses are defined for **parameters**, not statistics

What will our null and alternative hypotheses be for this example?



1

Defining the hypotheses

Remember, the null and alternative hypotheses are defined for **parameters**, not statistics

What will our null and alternative hypotheses be for this example?

- H_0 : the true proportion of complications among her patients is **equal** to the US population rate
- H_a : the true proportion of complications among her patients is **lower** than the US population rate

1

Defining the hypotheses

Remember, the null and alternative hypotheses are defined for **parameters**, not statistics

What will our null and alternative hypotheses be for this example?

- H_0 : the true proportion of complications among her patients is **equal** to the US population rate
- H_a : the true proportion of complications among her patients is **lower** than the US population rate

Expressed in symbols:

- $H_0 : p = 0.10$
- $H_a : p < 0.10$

2

Collecting and summarizing data

With these two hypotheses, we now take our sample and summarize the data.

The choice of summary statistic calculated depends on the type of data. In our example, we use the sample proportion

$$\hat{p} = 3/62 \approx 0.048$$



3

Assessing the evidence observed

Next, we calculate the probability of getting data like ours, *or more extreme*, if H_0 were in fact actually true.

This is a conditional probability: "given that H_0 is true, $p = 0.1$, what would the probability of observing $\hat{p} = 3/62$ or less?"

This probability is known as the **p-value**.



Calculate the p-value using simulation



Simulating the null distribution

Let's return to the organ transplant scenario.

Since $H_0 : p = 0.10$, we need to simulate a distribution for \hat{p} under the null hypothesis such that the probability of complication for each patient is 0.10 for 62 patients.

This null distribution for \hat{p} represents the distribution of the observed proportions we might expect, if the null hypothesis were true.

When sampling from the null distribution, what is the expected proportion of complications?



Data

```
glimpse(organ_donor)
```

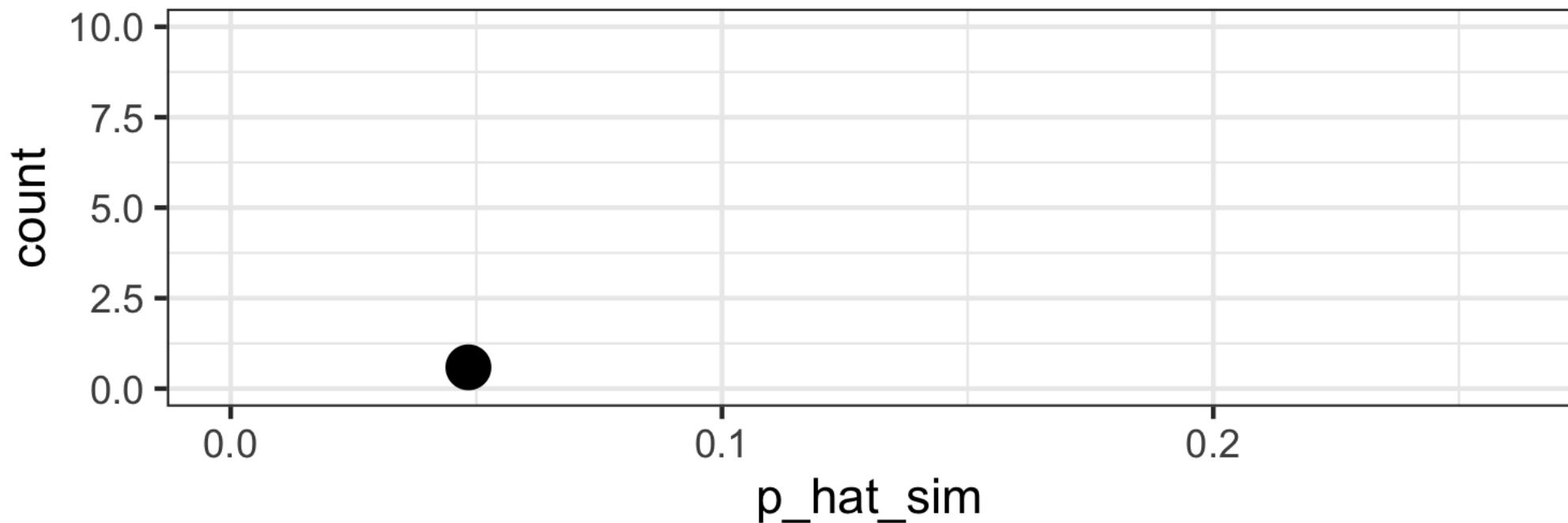
```
organ_donor %>%  
  count(outcome)
```

```
## # A tibble: 2 x 2
##   outcome      n
##   <chr>     <int>
## 1 complication     3
## 2 no complication 59
```



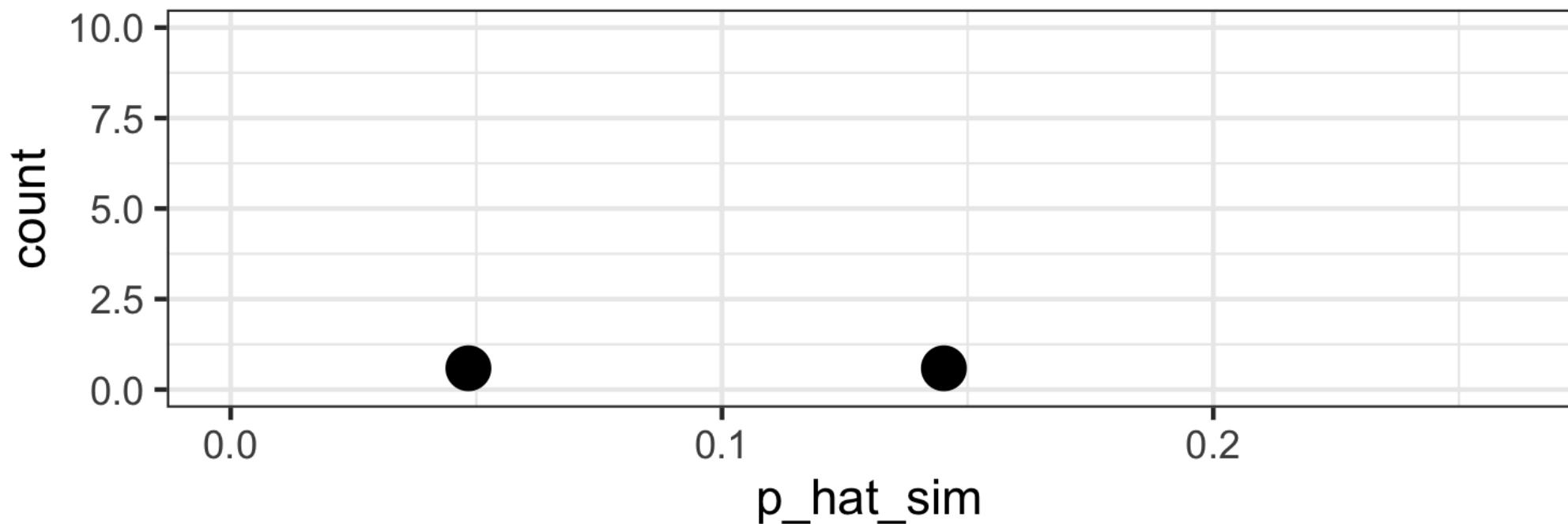
Simulation #1

```
## sim1  
##      complication no complication  
##            3                  59  
  
## [1] 0.0483871
```



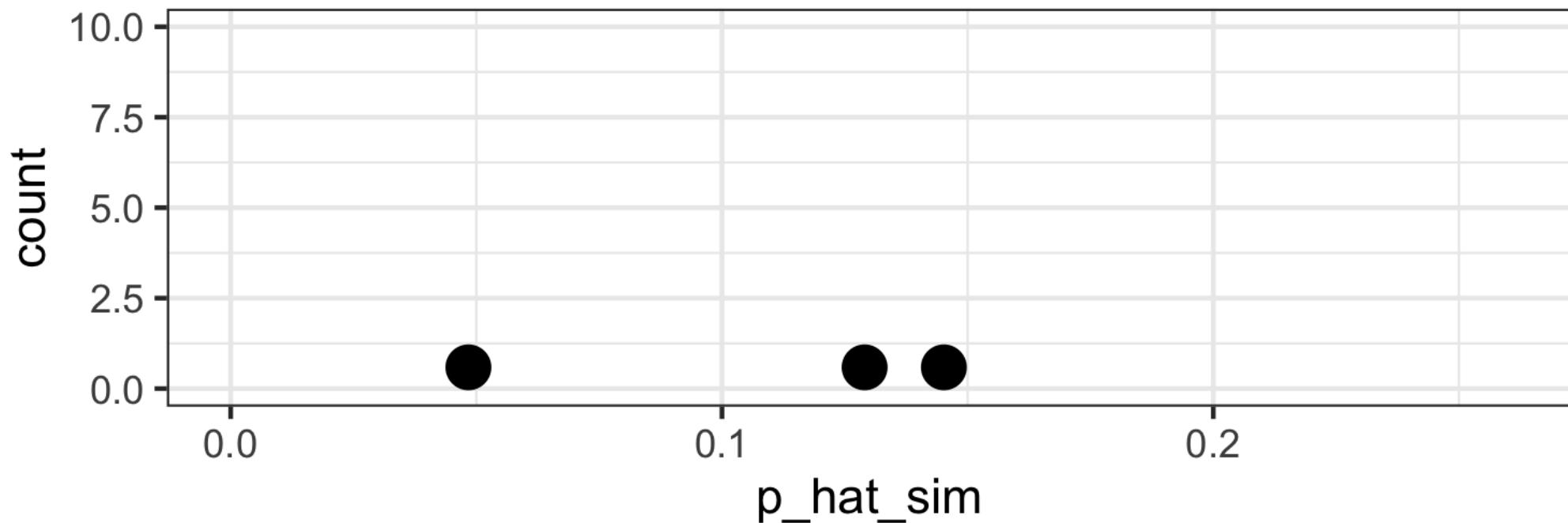
Simulation #2

```
## sim2  
##      complication no complication  
##          9                 53  
  
## [1] 0.1451613
```



Simulation #3

```
## sim3  
##      complication no complication  
##                      8                      54  
  
## [1] 0.1290323
```



This is getting boring...

We need a way to automate this process!



Using infer to generate the null distribution

```
null_dist <- organ_donor %>%
  specify(response = outcome, success = "complication") %>%
  hypothesize(null = "point",
              p = c("complication" = 0.10, "no complication" = 0.90))
  ) %>%
  generate(reps = 100, type = "simulate") %>%
  calculate(stat = "prop")
```



Specify

```
null_dist <- organ_donor %>%
  specify(response = outcome, success = "complication") %>%
  hypothesize(null = "point",
               p = c("complication" = 0.10, "no complication" = 0.90))
  ) %>%
  generate(reps = 100, type = "simulate") %>%
  calculate(stat = "prop")
```

- **response**: **outcome** in the **organ_donor** data frame
- **success**: "complication", the level of outcome we're interested in studying



Hypothesize

```
null_dist <- organ_donor %>%
  specify(response = outcome, success = "complication") %>%
  hypothesize(null = "point",
               p = c("complication" = 0.10, "no complication" = 0.90))
  ) %>%
  generate(reps = 100, type = "simulate") %>%
  calculate(stat = "prop")
```

- **null**: Since we're testing the point null hypothesis that $H_0 : p = 0.10$, we choose "**point**"
- Next, we provide the probability of "success" and "failure"
 - **"complication" = 0.10, "no complication" = 0.90**



Generate

```
null_dist <- organ_donor %>%
  specify(response = outcome, success = "complication") %>%
  hypothesize(null = "point",
               p = c("complication" = 0.10, "no complication" = 0.90))
  ) %>%
  generate(reps = 1000, type = "simulate") %>%
  calculate(stat = "prop")
```

- **reps**: We will generate 100 repetitions here
- **type**: Choose "**simulate**" for testing a point null
 - Choose **bootstrap** for estimation
 - Choose **permute** for testing independence

Calculate

```
null_dist <- organ_donor %>%
  specify(response = outcome, success = "complication") %>%
  hypothesize(null = "point",
               p = c("complication" = 0.10, "no complication" = 0.90))
  ) %>%
  generate(reps = 100, type = "simulate") %>%
  calculate(stat = "prop")
```

- Calculate a sample statistic. Here, the sample proportion.
 - **stat = "prop"**



Store simulated null distribution

```
null_dist <- organ_donor %>%
  specify(response = outcome, success = "complication") %>%
  hypothesize(null = "point",
              p = c("complication" = 0.10, "no complication" = 0.90))
  ) %>%
  generate(reps = 100, type = "simulate") %>%
  calculate(stat = "prop")
```



Store simulated null distribution

```
null_dist <- organ_donor %>%
  specify(response = outcome, success = "complication") %>%
  hypothesize(null = "point",
               p = c("complication" = 0.10, "no complication" = 0.90))
  ) %>%
  generate(reps = 100, type = "simulate") %>%
  calculate(stat = "prop")
```

```
## # A tibble: 100 x 2
##       replicate   stat
##           <dbl> <dbl>
## 1             1  0.161
## 2             2  0.081
## 3             3  0.048
## 4             4  0.081
## 5             5  0.081
## 6             6  0.065
## 7             7  0.065
```



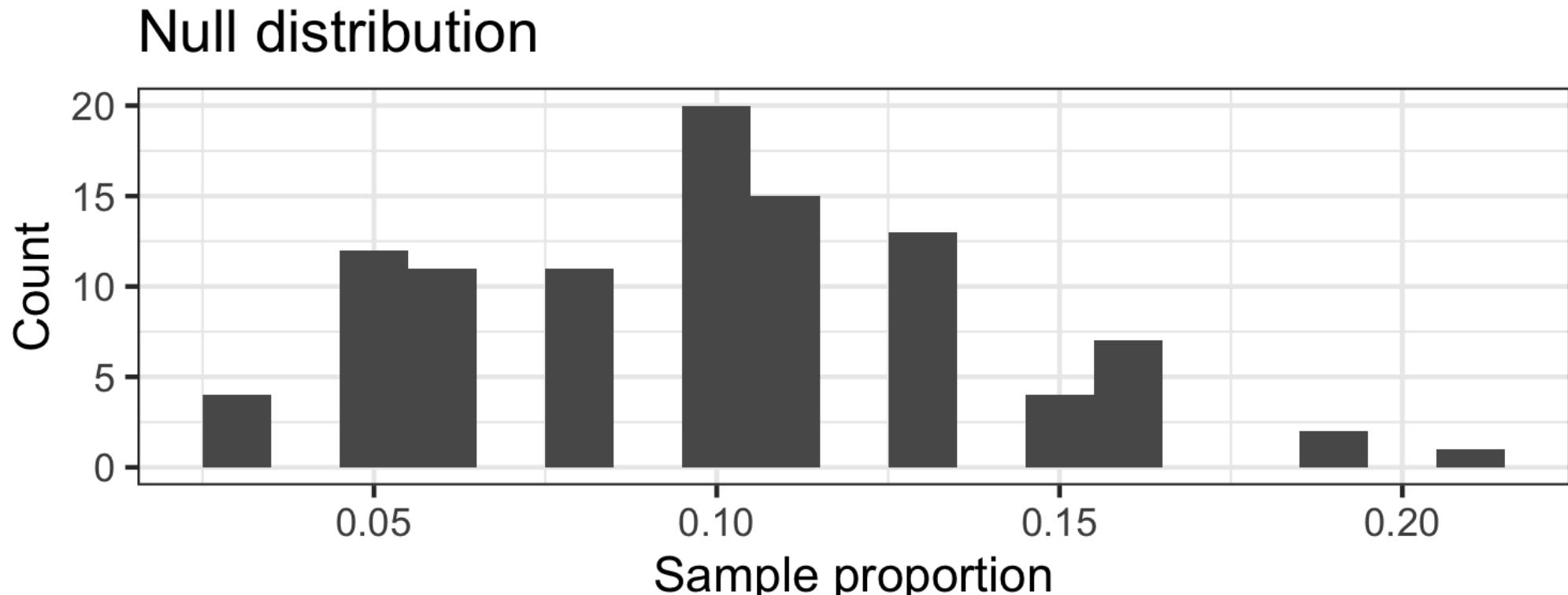
Visualizing the null distribution

What would you expect the center of the null distribution to be?



Visualizing the null distribution

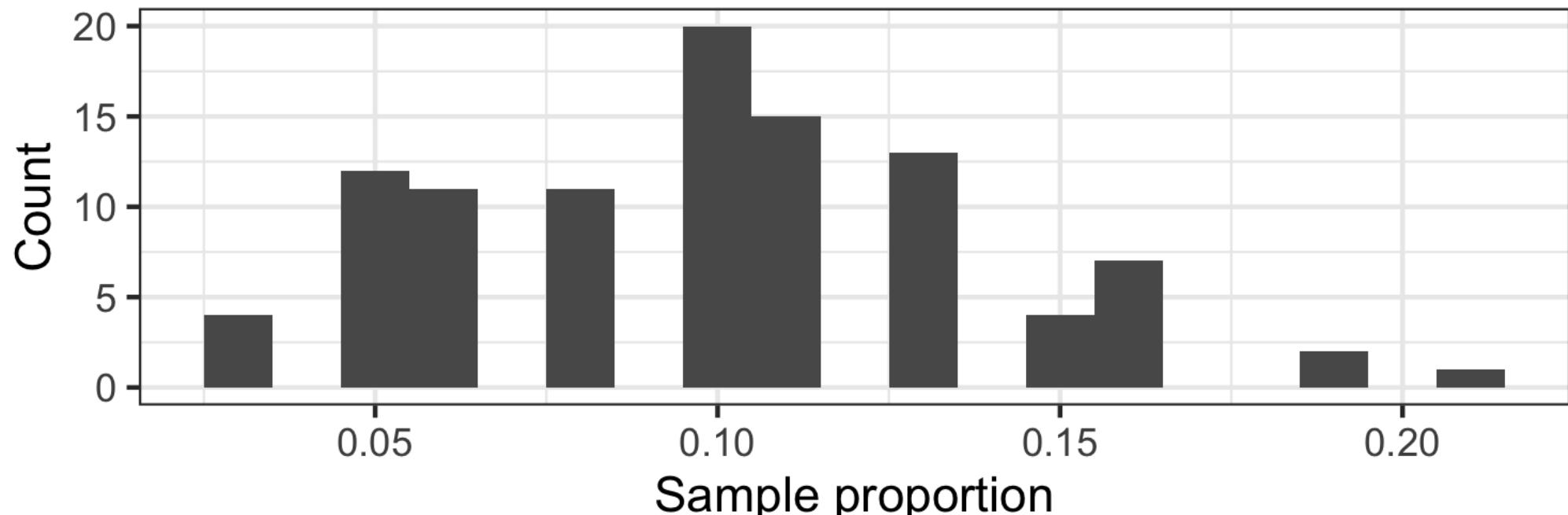
What would you expect the center of the null distribution to be?



Calculating the p-value, visually

What is the p-value (just eyeball it)?

Null distribution



Calculating the p-value, directly

```
null_dist %>%
  filter(stat <= (3/62)) %>%
  summarise(p_value = n()/nrow(null_dist))
```

```
## # A tibble: 1 x 1
##   p_value
##   <dbl>
## 1 0.16
```



4 Making a conclusion

We reject the null hypothesis if the p-value is probability is small enough, i.e. it is very unlikely to observe our data or more extreme if H_0 were actually true.



4

Making a conclusion

We reject the null hypothesis if the p-value is probability is small enough, i.e. it is very unlikely to observe our data or more extreme if H_0 were actually true.

What is "small enough"? We often consider a threshold (the **significance level** or α -level) defined *prior* to conducting the analysis.



Significance level

We often use 5% as the cutoff for whether the p-value is low enough that the data are unlikely to have come from the null model.

- If $p\text{-value} < \alpha$, reject H_0 in favor of H_a : - The data provide convincing evidence for the alternative hypothesis.
- If $p\text{-value} \geq \alpha$, fail to reject H_0 in favor of H_a
 - The data do not provide convincing evidence for the alternative hypothesis.



What if p -value $\geq \alpha$?

If p-value $\geq \alpha$ we fail to reject H_0 .



What if p -value $\geq \alpha$?

If p -value $\geq \alpha$ we fail to reject H_0 .

Importantly, we never "accept" the null hypothesis.



What if p -value $\geq \alpha$?

If p -value $\geq \alpha$ we fail to reject H_0 .

Importantly, we never "accept" the null hypothesis.

When we fail to reject the null hypothesis, we are stating that there is **insufficient evidence** to conclude that it is false. This could be due to any number of reasons:

- There truly is no effect
- There truly is an effect (and we happened to get unlucky with our sample or didn't have enough data to tell that there was one)

What can go wrong?

Suppose we test a certain null hypothesis, which can be either true or false (we never know for sure!). We make one of two decisions given our data: either reject or fail to reject H_0 .



What can go wrong?

Suppose we test a certain null hypothesis, which can be either true or false (we never know for sure!). We make one of two decisions given our data: either reject or fail to reject H_0 .

We have the following four scenarios:

Decision	H_0 is true	H_0 is false
Fail to reject H_0	Correct decision	<i>Type II Error</i>
Reject H_0	<i>Type I Error</i>	Correct decision

It is important to weigh the consequences of making each type of error.