Class 18: Hypothesis test for proportions

Halloween edition...

Overview

Hypothesis tests

- Hypothesis tests for a single proportion

If there is time

- Assessing causal relationships
- Hypothesis tests for two proportions

Reminder: keep working on your class project

Change in plan!

A polished draft of the project is due on November 17th

Homework 7 is due on Sunday November 3rd

Homework 8 will be due on Sunday November 10th

Statistical Inference

Inference

Population: all individuals/objects of interest

- E.g., all voters

A parameter is number associated with the <u>population</u>

- E.g., The proportion of all voters who voted for Biden: π_{Trump}

Sample: A subset of the population

- E.g., 1000 randomly sampled voters

A **statistic** is number calculated from the <u>sample</u>

- e.g., The proportion in the sample who voted for Biden: $\hat{p}_{Trump}$

Statistical Inference: Making conclusions about a population based on data in a sample

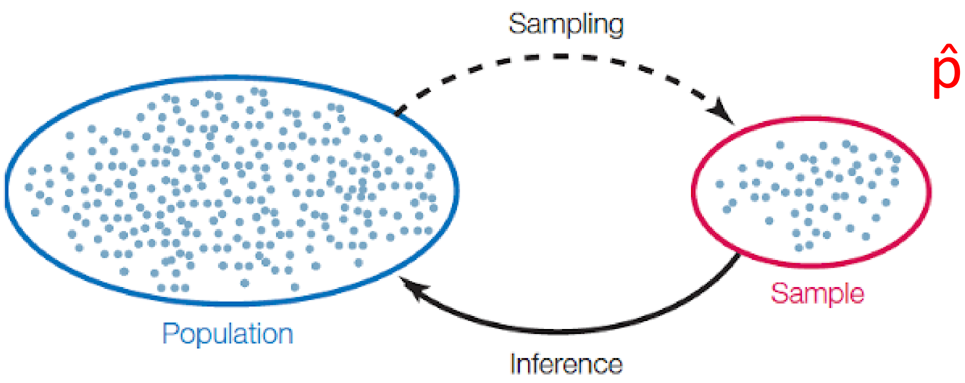
- E.g., using $\hat{p}_{Trump}$ as an estimate of π_{Trump}

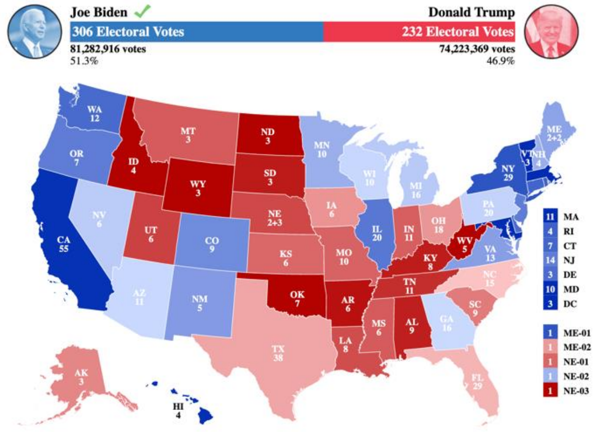
parameter

π

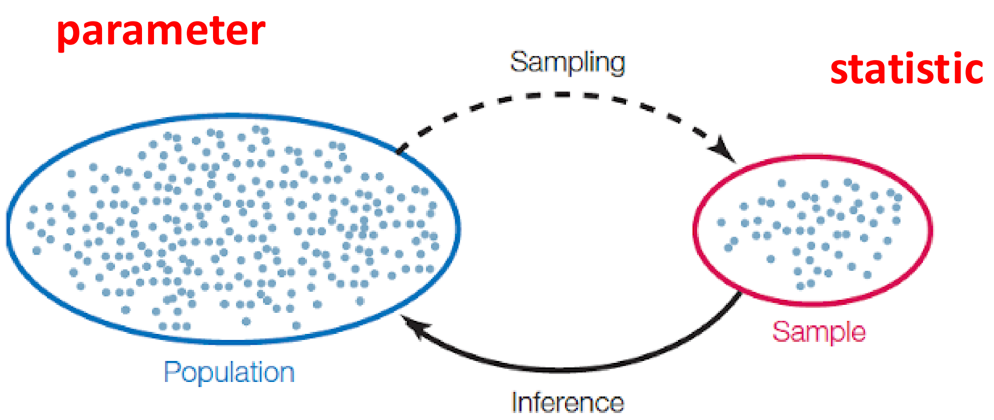
statistic

$\hat{p}$





Examples of parameters and statistics



	Population Parameter	Sample Statistic
Mean	μ	$\bar{x}$
Proportion	π	$\hat{p}$
Standard deviation	σ	s
Correlation	ρ	r

Probability distribution of a statistic

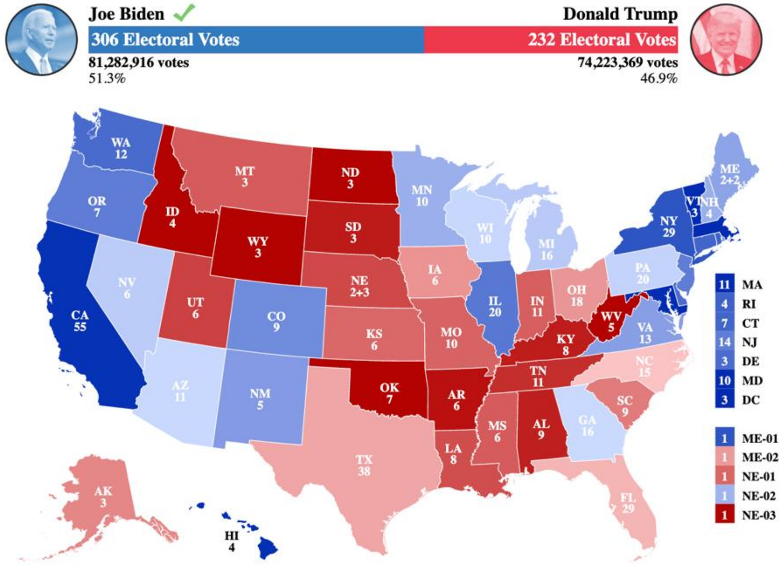
Values of a statistic vary because random samples vary

A **sampling distribution** is a probability distribution of *statistics*

- All possible values of the statistic and all the corresponding probabilities
- We can approximate a sampling distribution by a simulated statistics

π_{Trump}

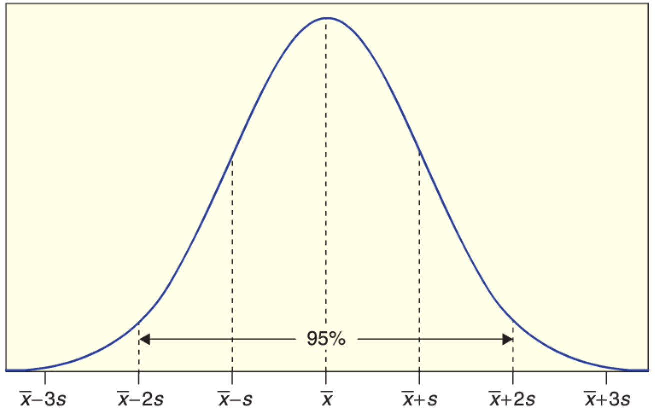
n = 1,000



$\hat{p}_{Trump}$

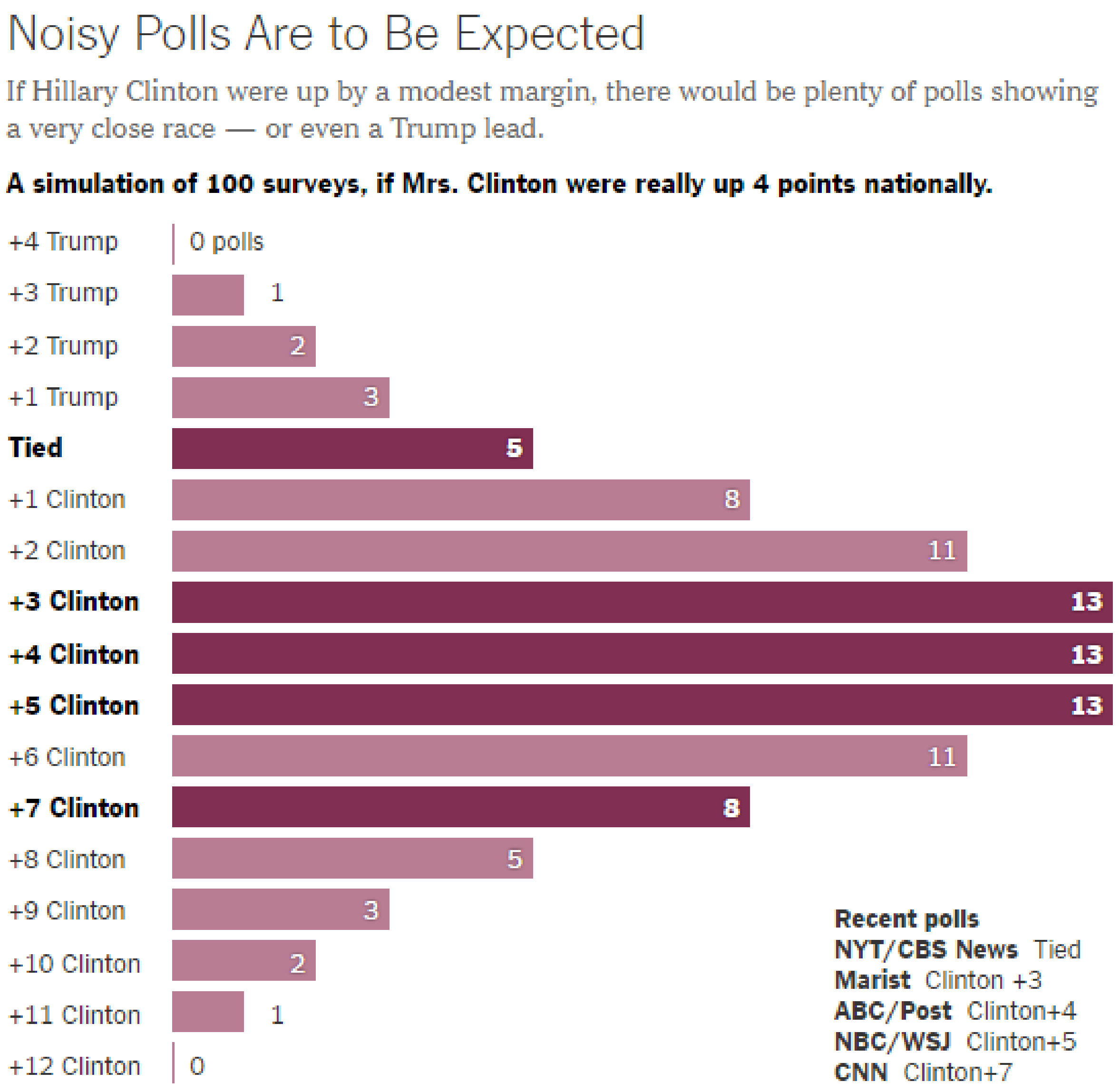
$\hat{p}_{Trump}$

$\hat{p}_{Trump}$



Sampling distribution!

Confused by Contradictory Polls? Take a Step Back



Sampling distribution of $\hat{p}_{Clinton}$

What parameter are they trying to estimate?

NY Times 7/18/16

Questions?

Hypothesis tests

Statistical tests (hypothesis test)

A **statistical test** uses data from a sample to assess a claim about a population (parameter)

Example 1: The average body temperature of humans is 98.6˚

How can we write this using symbols?

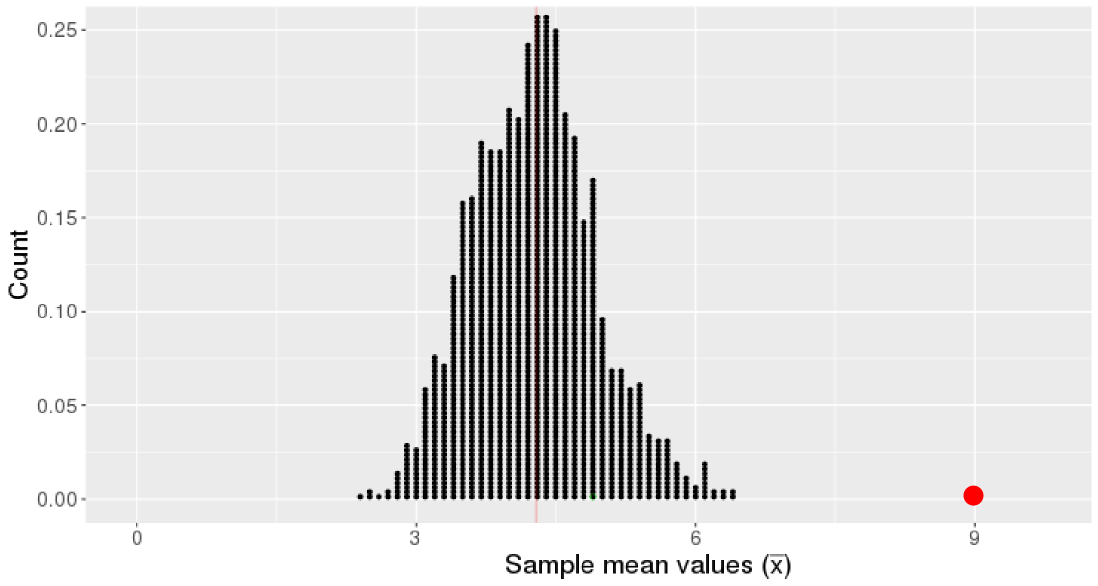
- $\mu = 98.6$

Basic hypothesis test logic

We start with a claim about a population parameter

- E.g., μ = ~~4~~

This claim implies we should get a certain distribution of statistics



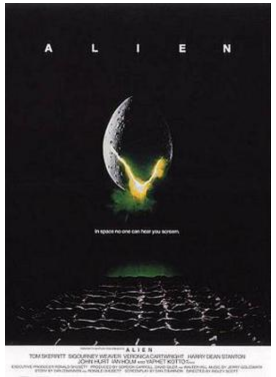
If our observed statistic is highly unlikely, we reject the claim

Motivating example: The Bechdel Test

Question: Do less than 50% of movies pass the Bechdel test?

Questions:

- What is the population/process?
- What is our parameter of interest?
  - What symbol should we use to denote it?
- What is out statistic of interest?
  - What symbol should we use to denote it?

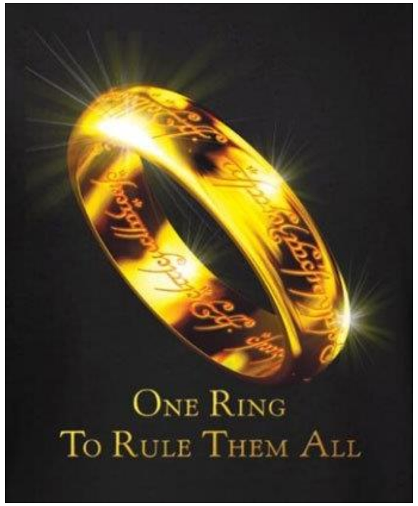


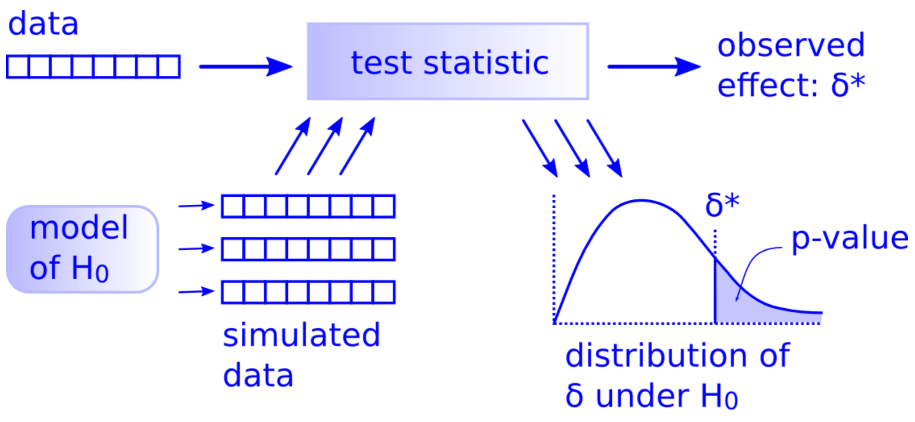
Steps needed to run a hypothesis test

To run a hypothesis test, we can use 5 steps:

1. State the null and alternative hypothesis
2. Calculate the observed statistic of interest
3. Create the null distribution
4. Calculate the p-value
5. Make a decision

Let's go through these steps now…





Do less than 50% of movies pass the Bechdel test?

Step 1: state the null and alternative hypotheses

If only 50% of the movies passed the Bechdel test, what would we expect the value of the parameter to be?

H_0: $\pi = 0.5$

If fewer than 50% of movies passed the Bechdel test, what would we expect the value of the parameter to be?

H_A: $\pi < 0.5$

Observed statistic value

Step 2: calculate the observed statistic

There are 1794 movies in our data set

Of these, 803 passed the Bechdel test

What is our observed statistic value and what symbol should we use to denote this value?

A: $\hat{p}$ = 803/1794 = 0.448

	title	binary
1	Dredd 3D	PASS
2	12 Years a Slave	FAIL
3	2 Guns	FAIL
4	42	FAIL
5	47 Ronin	FAIL
6	A Good Day to Die Hard	FAIL
7	About Time	PASS
8	Admission	PASS
9	After Earth	FAIL
10	American Hustle	PASS
11	August: Osage County	PASS
12	Beautiful Creatures	PASS
13	Blue Jasmine	PASS
14	Captain Phillips	FAIL

Step 3: Create a null distribution

How can we assess whether 803 out of 1794 movies passing the Bechdel test ($\hat{p}$ = 0.448) is consistent with what we would expect if 50% (or more) movies passed the Bechdel test?

- i.e., is $\hat{p}$ = 0.448 a likely value if π = 0.5?

If 50% of movies passed the Bechdel test, we can model movies passing the as a fair coin flip:

Heads (True) = passed the Bechdel test

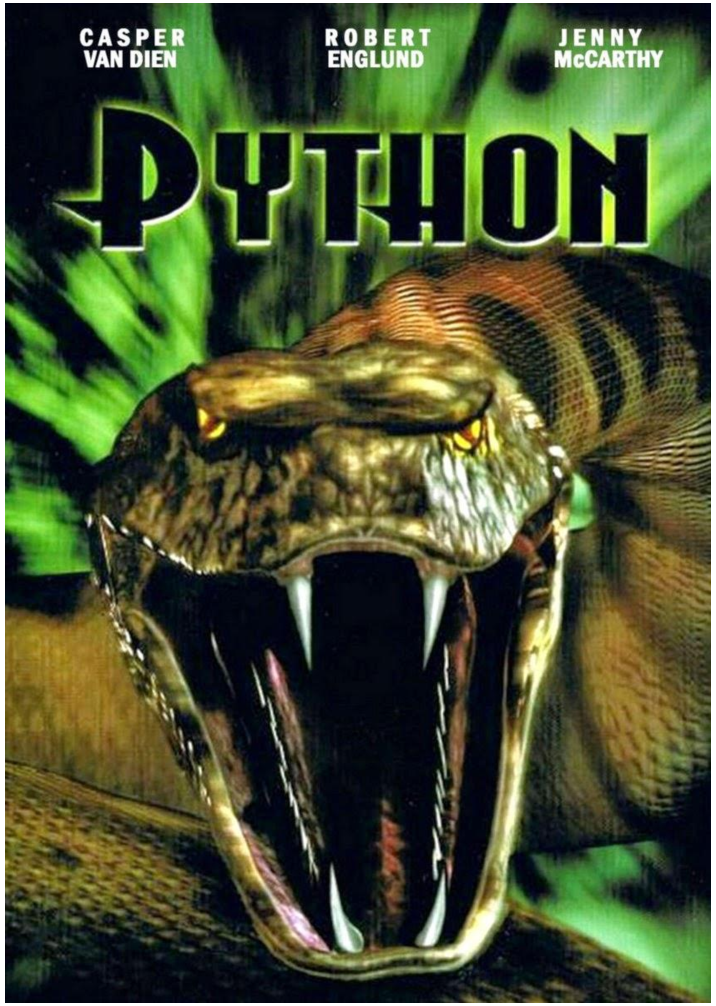
Tails (False) = failed to pass the Bechdel test

Let's simulate fliping a coin 1794 times and see how many times we get 803 **or fewer** heads

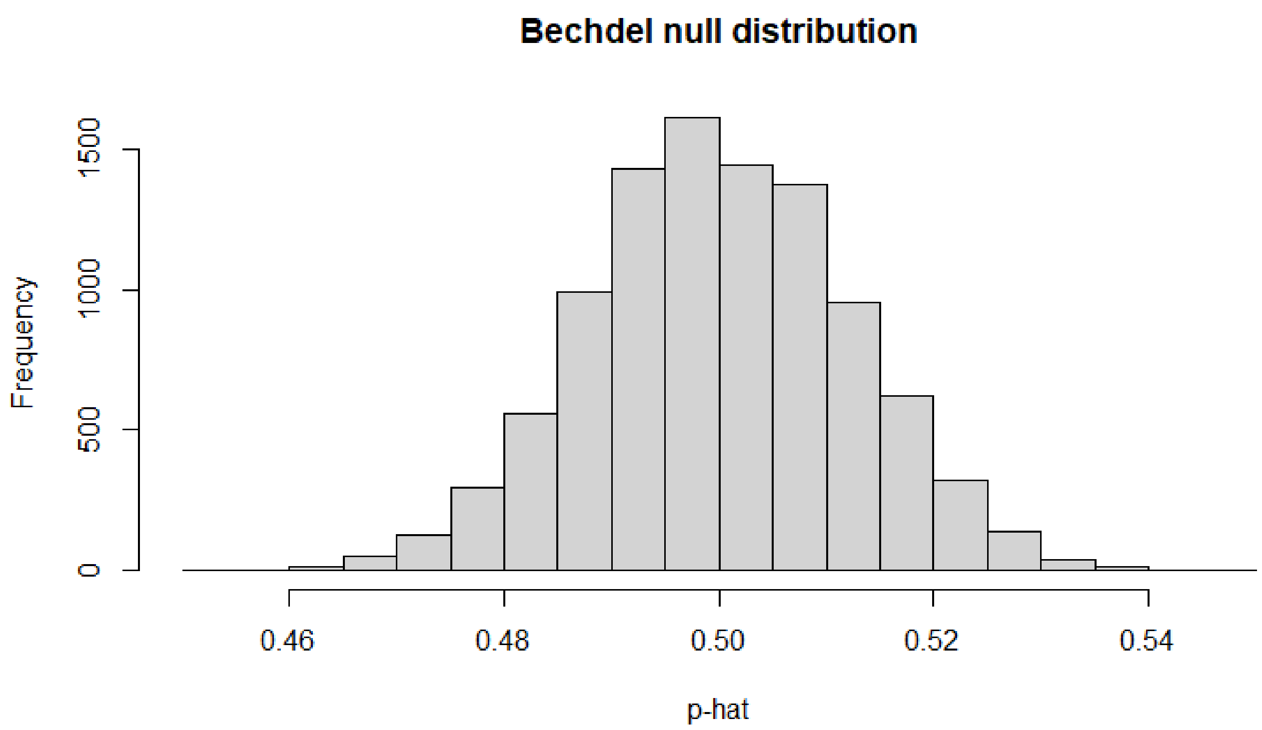
Chance models

To really be sure, how many repetitions of flipping a coin 1794 times should we do?

Any ideas how to do this?



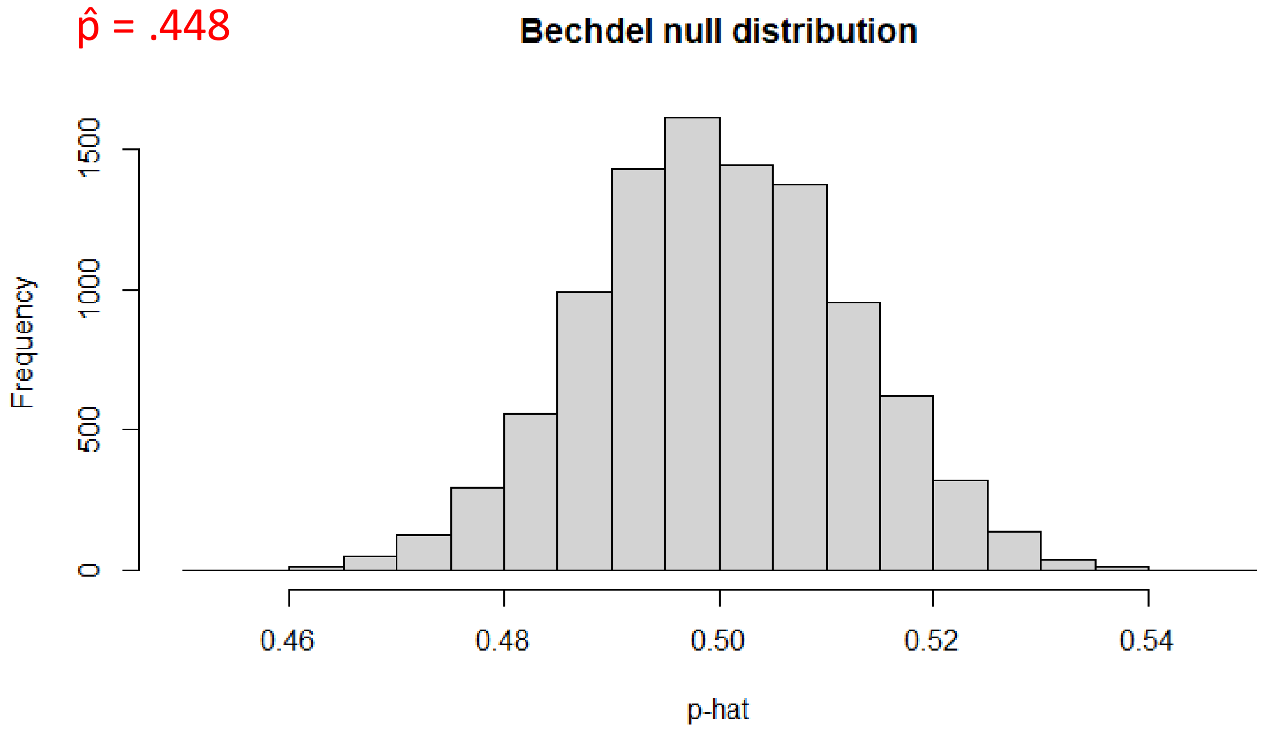
Simulating Flipping 1794 coins 10,000 times



Assuming the null hypothesis is true, the distribution of statistics we get is called the **null distribution**

Step 4: calculate the p-value

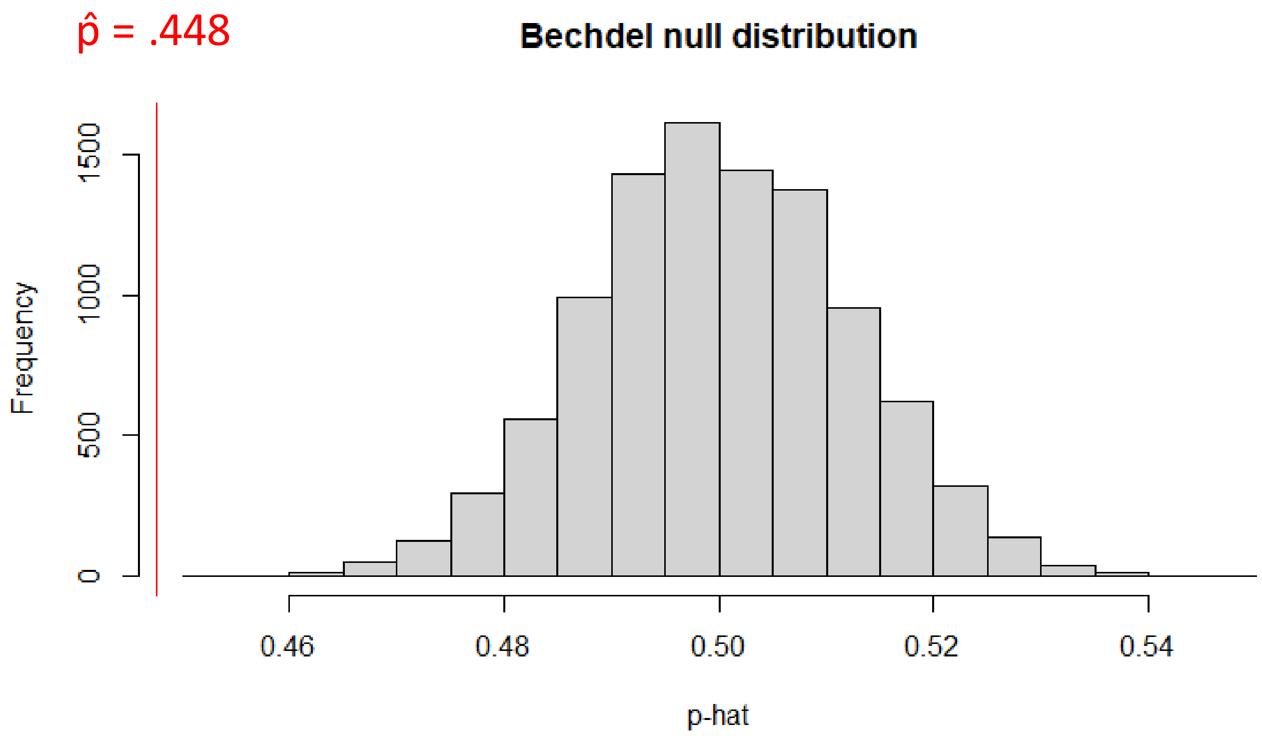
$\hat{p}$ = .448



Q: Is it likely that 50% of movies pass the Bechdel test?

- i.e., is it likely that $\pi = .5$?

Step 4: calculate the p-value

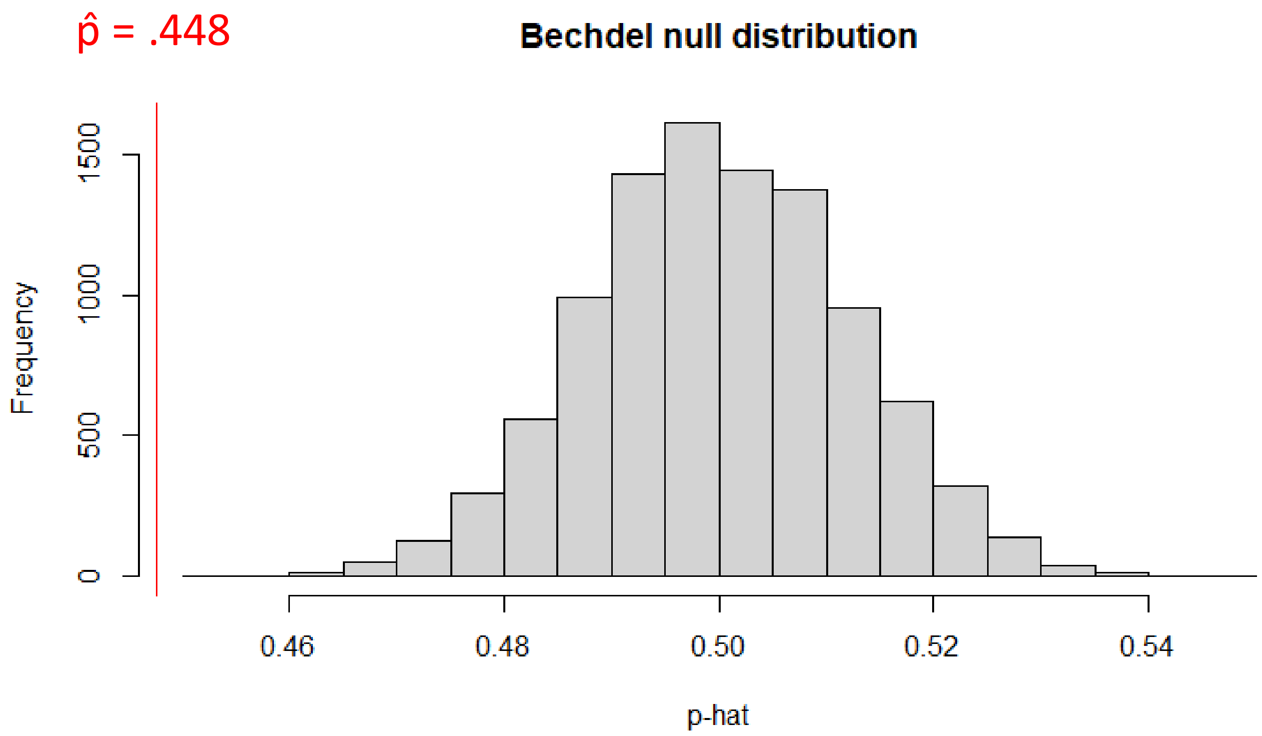


The **p-value** is the probability we will get a statistic as or more extreme than the observed statistic, if the null hypothesis was true

Q: What is the p-value here?

A: the p-value is 0

Step 5: Make a decision



If the observed data is very unlikely if the null hypothesis is true, we can reject the null hypothesis

- i.e., if p-value is very small we can reject the null hypothesis

Bechdel (hypothesis) test

1. State the null hypothesis and the alternative hypothesis

- 50% of the movies pass the Bechdel test: H_0: $\pi = 0.5$
- Less than 50% of movies pass the: H_A: $\pi < 0.5$

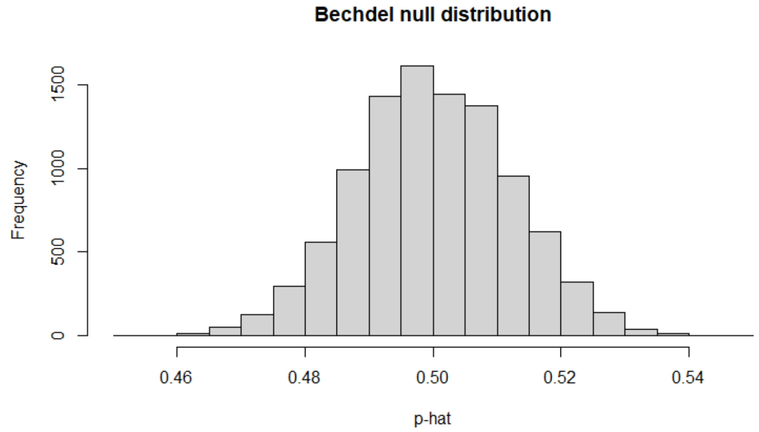
2. Calculate the observed statistic

- 803 out of 1794 movies passed the Bechdel test

$\hat{p}$ = .448

3. Create a null distribution that is consistent with the null hypothesis

- i.e., the statistics we expect if 50% of the movies passed the Bechdel test



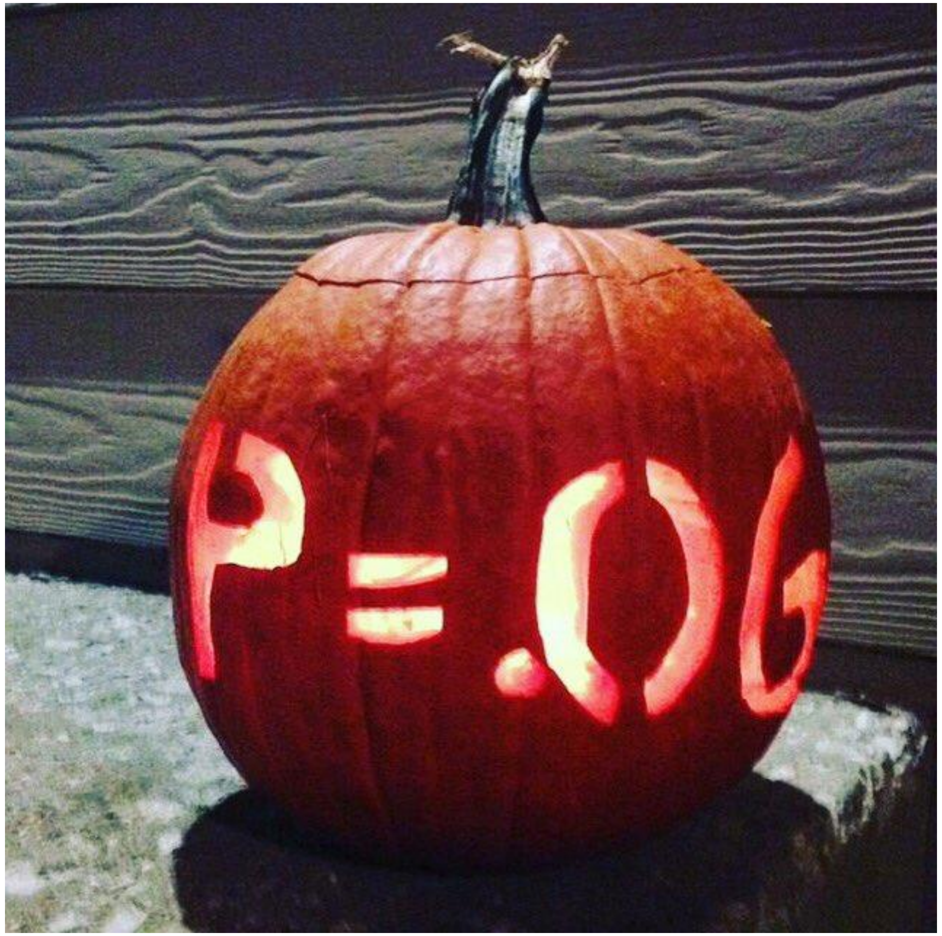
4. Examine how likely the observed statistic is to come from the null distribution

- What is the probability that only 803 of 1794 movies would pass the Bechdel test ($\hat{p}$ = .448) if the null hypothesis was true?
- i.e., what is the p-value?

5. Make a judgement

- A small p-value this means that $\pi = .5$ is unlikely, and so it is likely $\pi < .5$
- i.e., we say our results are 'statistically significant'

Let’s try it in Python!

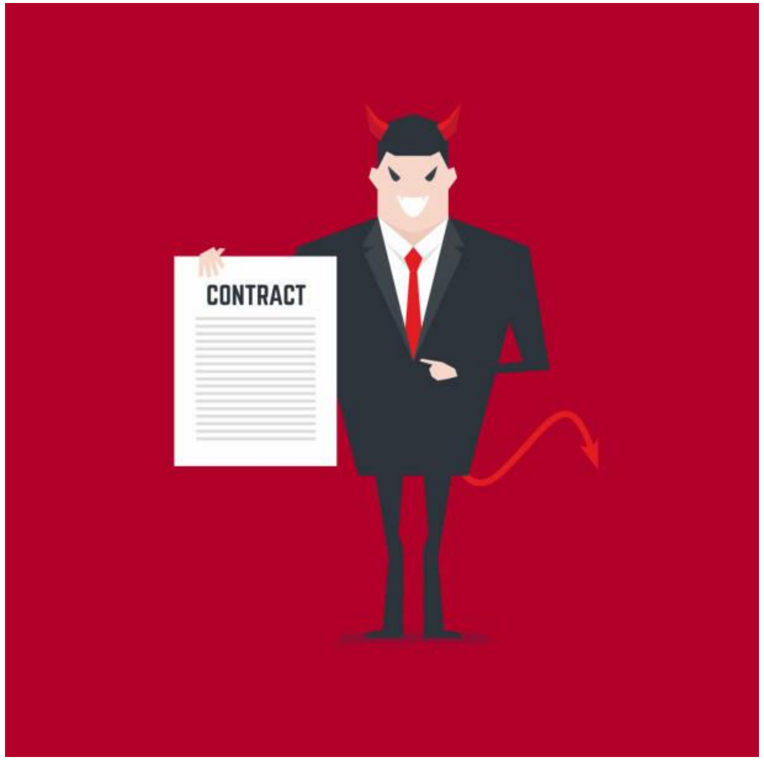


Another example: sinister lawyers

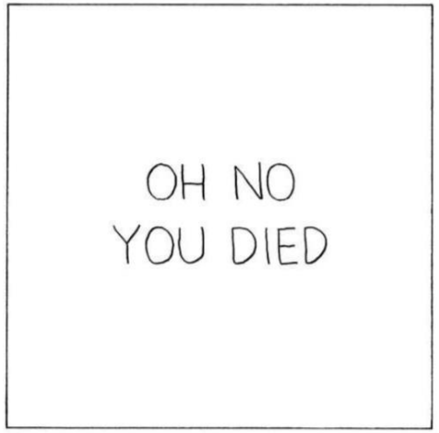
10% of American population on is left-handed

A study found that out of a random sample of 105 lawyers, 16 were left-handed

Use our 5 steps of hypothesis testing to assess whether the proportion of left-handed lawyers is greater than the proportion on found in the American population



Let’s try it in Python!



Assessing causal relationships

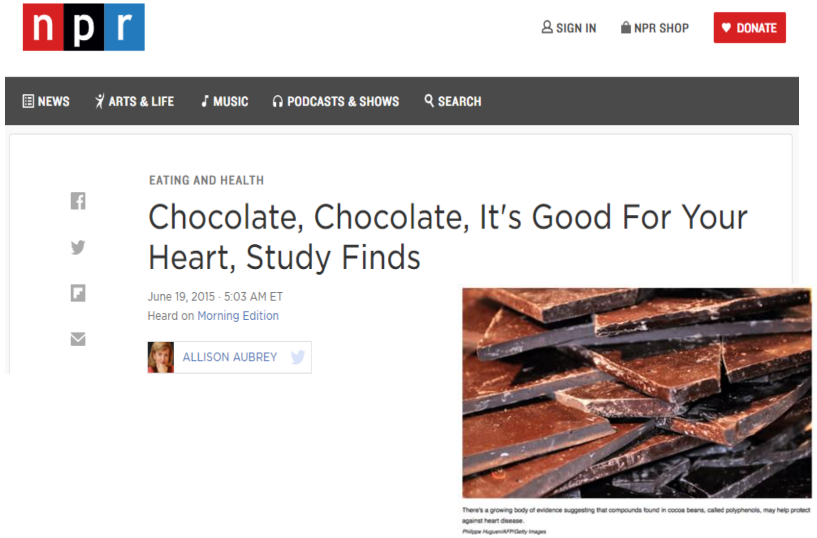
Causality

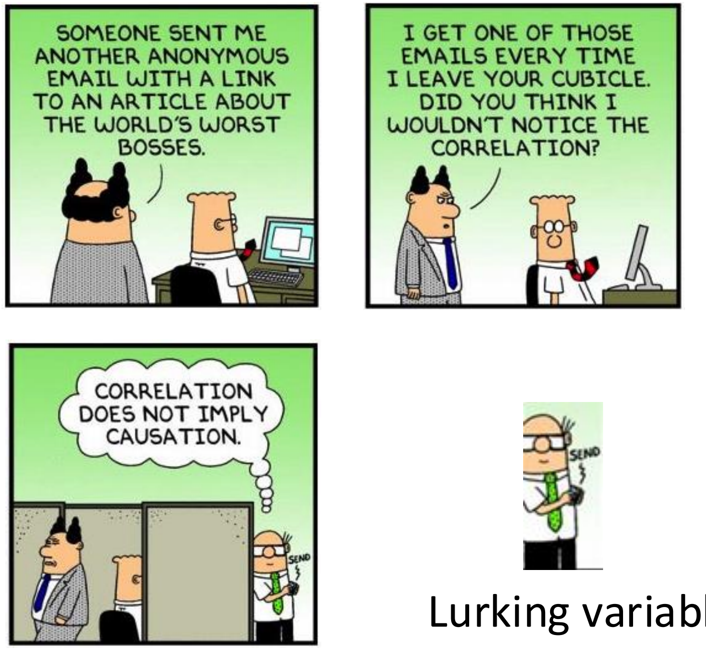
An association is the presence of a reliable relationship between the treatments an outcome

A causal relationship is when changing the value of a treatment variable influences the value outcome variable

A **confounding variable** (also known as a **lurking variable**) is a third variable that is associated with both the treatment (explanatory) variable and the outcome (response) variable

- A confounding variable can offer a plausible explanation for an association between the other two variables of interest





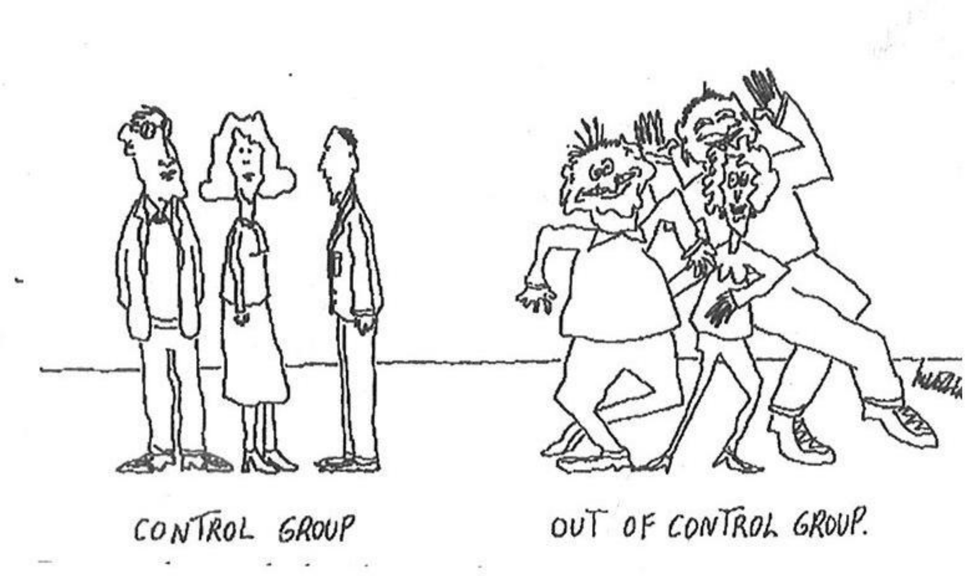
Lurking variable

Randomized Controlled Experiment

Sample A: control group

Sample B: treatment group

If members of the treatment and control groups are selected at random; this allows causal conclusions!



In particular, any difference in outcomes between the two groups could be due to:

- Chance
- The treatment

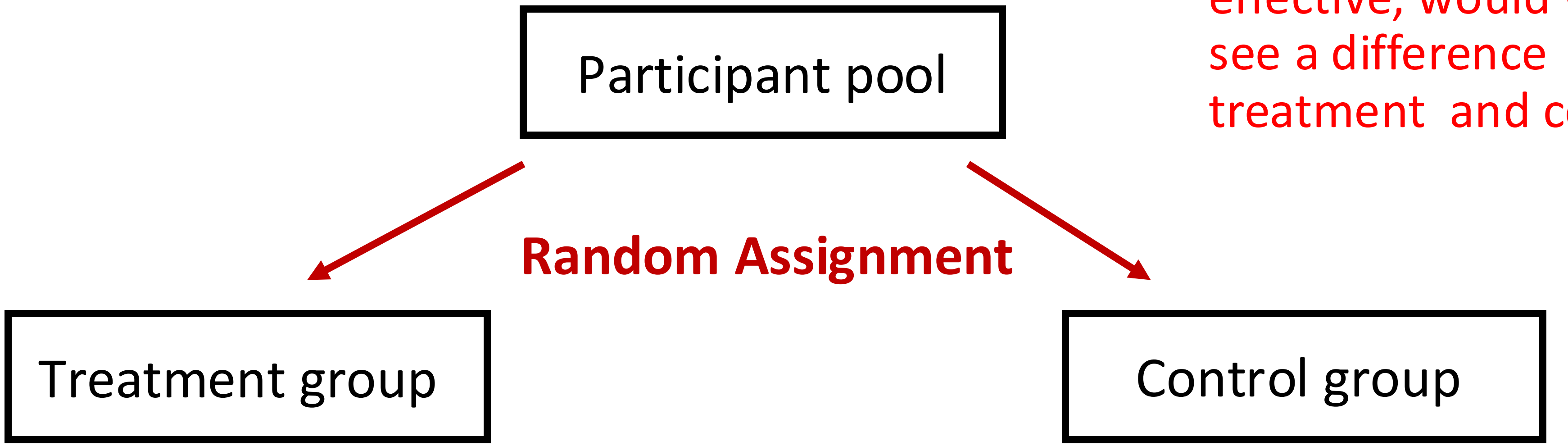
Randomly assigning participants to treatment and control groups allows us to separate what expected by chance and consequently what is due to the treatment

Randomized Controlled Experiment

Take a group of participant and ***randomly assign***:

- Half to a *treatment group* where they get chocolate
- Half in a *control group* where they get a fake chocolate (placebo)
- See if there is more improvement in the treatment group compared to the control group

Q: If the treatment was not effective, would we expect to see a difference between the treatment and control groups?



Case study

RCT to study Botulinum Toxin A (BTA) as a treatment to relieve chronic back pain

- 15 patients in the treatment group (received BTA)
- 16 in the control group (normal saline)

Trials were run double-blind: neither doctors nor patients knew which group they were in.

Results

- 2 patients in the control group had relief from pain (outcome=1)
- 9 patients in the treatment group had relief.

Can this difference be just due to chance?

Neurology®

May 22, 2001; 56 (10) ARTICLES

Botulinum toxin A and chronic low back pain

A randomized, double-blind study

Leslie Foster, Larry Clapp, Marleigh Erickson, Bahman Jabbari

First published May 22, 2001, DOI: https://doi.org/10.1212/WNL.56.10.1290

Step 1: The hypotheses

Null:

- BTA does not lead to an increase in pain relief
  - i.e., if many people were to get BTA and saline, the proportion of people who experienced pain relief would be the same in both groups.
  - H_0: $\pi_{treat} = \pi_{control}$

Alternative:

- BTA leads to an increase in pain relief
  - i.e., if many people were to get BTA and saline, the proportion of people who experienced pain relief would be higher for those who received BTA
  - H_A: $\pi_{treat} > \pi_{control}$

Neurology®

May 22, 2001; 56 (10) ARTICLES

Botulinum toxin A and chronic low back pain

A randomized, double-blind study

Leslie Foster, Larry Clapp, Marleigh Erickson, Bahman Jabbari

First published May 22, 2001, DOI: https://doi.org/10.1212/WNL.56.10.1290

Step 2: The observed statistic

To calculate an observed statistic we need data:

Let's have our observed statistic mirror our hypotheses

- H_0: $\pi_{treat} - \pi_{control} = 0$

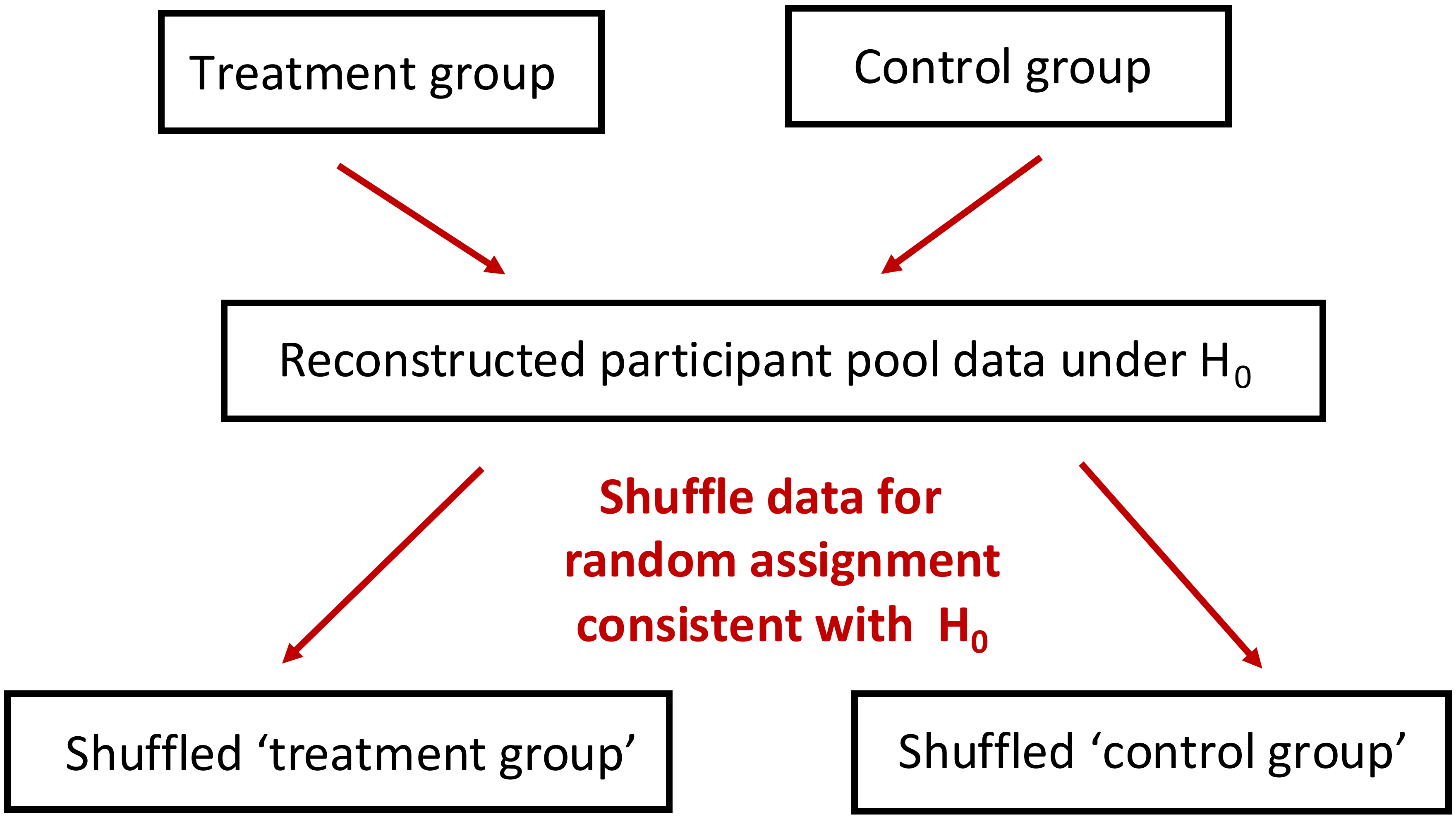
Observed statistic is: $\hat{p}_{treat} - \hat{p}_{control}$

$$= 9/15 - 2/16$$

$$= 0.475$$

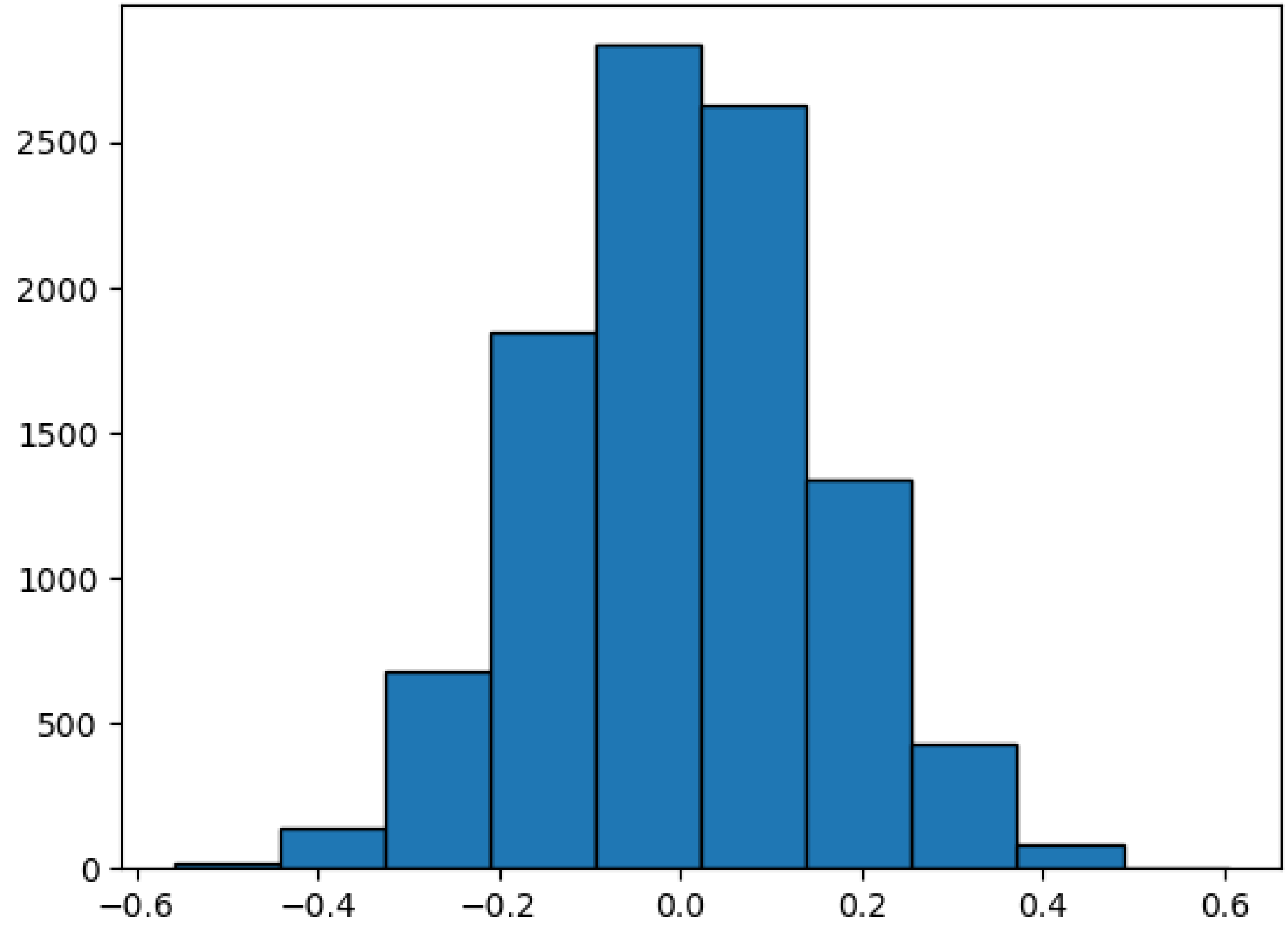
	Group	Result
19	Treatment	1.0
7	Control	0.0
6	Control	0.0
26	Treatment	0.0
17	Treatment	1.0
9	Control	0.0
13	Control	0.0
3	Control	0.0
1	Control	1.0
30	Treatment	0.0
28	Treatment	0.0

3. Create the null distribution!

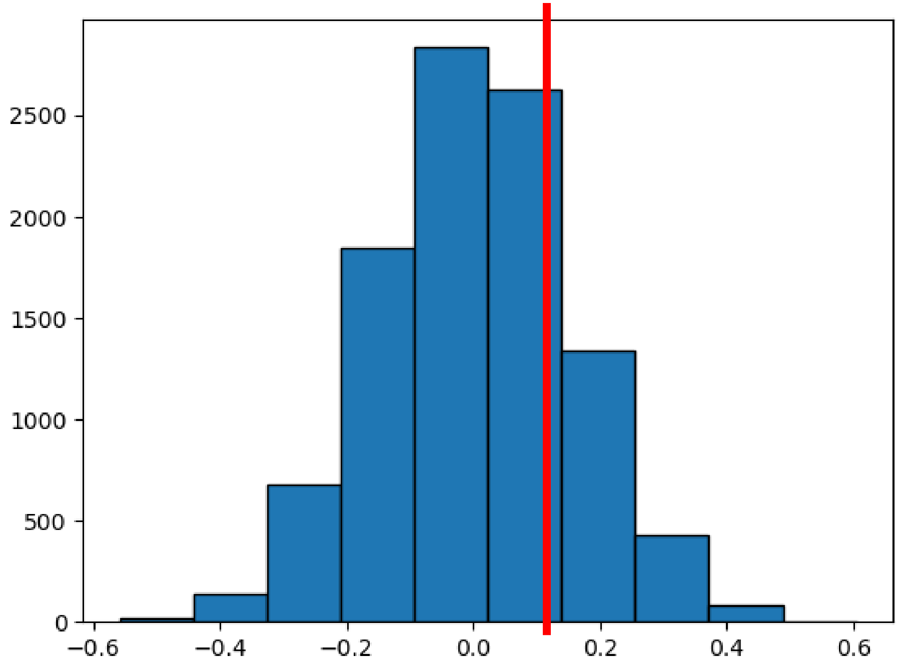


One null distribution statistic: $\hat{p}_{Shuff\_Treatment} - \hat{p}_{Shuff\_control}$

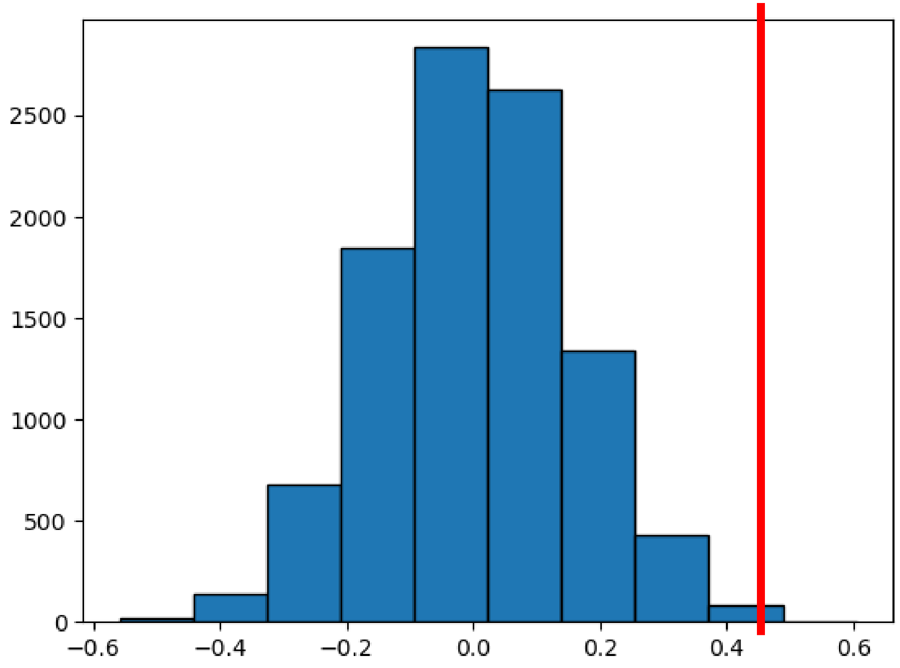
Step 3: Create a null distribution



Step 4: Calculate the p-value

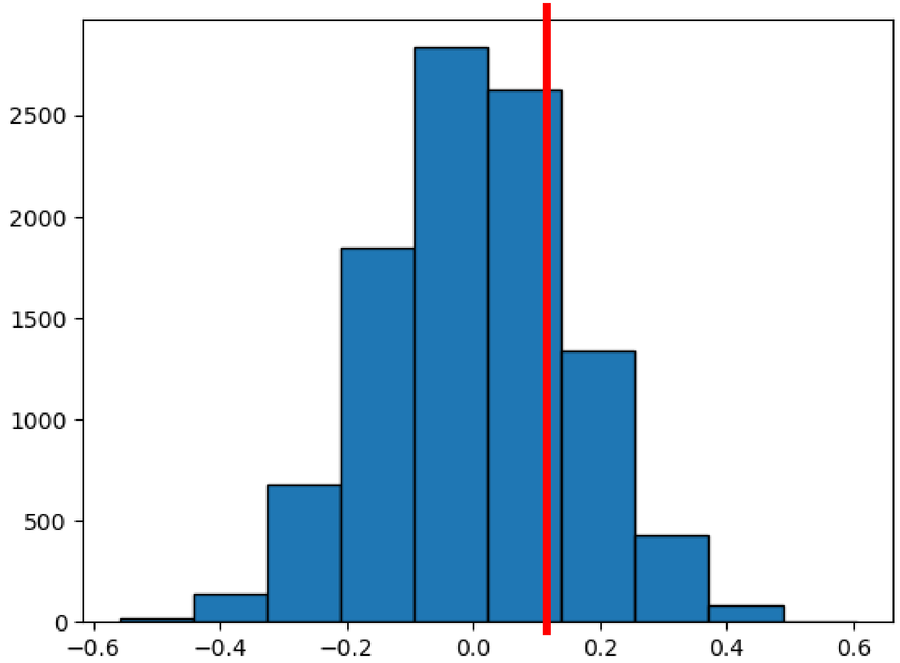


If $\hat{p}_{treat} - \hat{p}_{control} = 0.1$ what would the p-value be?

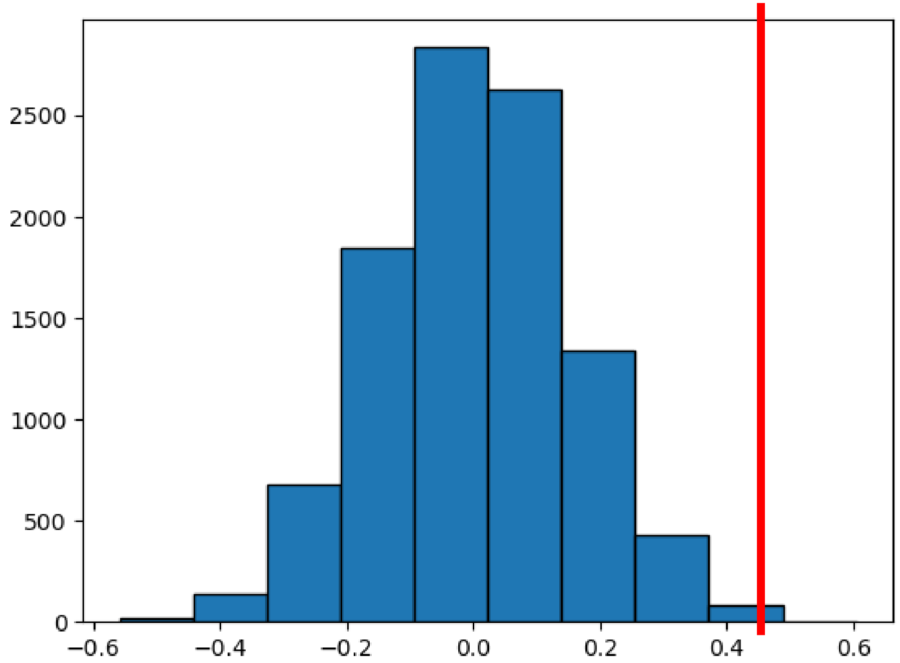


If $\hat{p}_{treat} - \hat{p}_{control} = 0.5$ what would the p-value be?

Step 5: Draw a conclusion



If the p-value was 0.19 what would we conclude?



If the p-value was 0.0007 what would we conclude?

Summary: BTA for back pain relief


Neurology

May 22, 2001; 56 (10) ARTICLES

Botulinum toxin A and chronic low back pain

A randomized, double-blind study

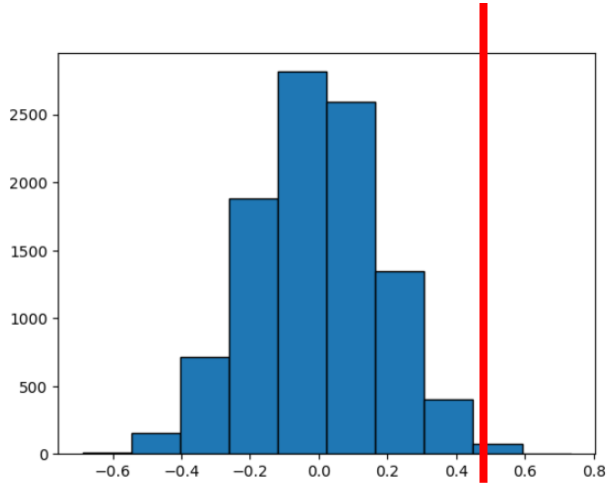
Leslie Foster, Larry Clapp, Marleigh Erickson, Bahman Jabbari

First published May 22, 2001, DOI: https://doi.org/10.1212/WNL.56.10.1290

1. State the null hypothesis and the alternative hypothesis
   - BTA does not lead to an increase in pain relief: H_0: $\pi_{treat} = \pi_{control}$
   - BTA leads to an increase in pain relief: H_A: $\pi_{treat} > \pi_{control}$

2. Calculate the observed statistic: $\hat{p}_{treat}$ - $\hat{p}_{control}$

$\hat{p}_{treat}$ - $\hat{p}_{control}$ = .475



3. Create a null distribution that is consistent with the null hypothesis
   - The $\hat{p}_{treat}$ - $\hat{p}_{control}$ statistics we expect if the null hypothesis was true
   - i.e., statistics we would expect if there was no difference in pain relief between the two groups

4. Examine how likely the observed statistic is to come from the null distribution
   - What is the probability that we would get a $\hat{p}_{treat}$ - $\hat{p}_{control}$ statistic larger than 0.475 if the null hypothesis was true?
   - i.e., what is the p-value?

5. Make a judgement
   - A small p-value this means that at the proportion of pain relief differed between the two groups
     - i.e., we say our results are 'statistically significant'
   - Because our analysis is based on a randomized controlled trial (using random assignment) we can say that BTA ***causes*** an increase in pain relief

Let’s try it in Python!