YData: Introduction to Data Science

Class 20: Hypothesis tests for two proportions and means

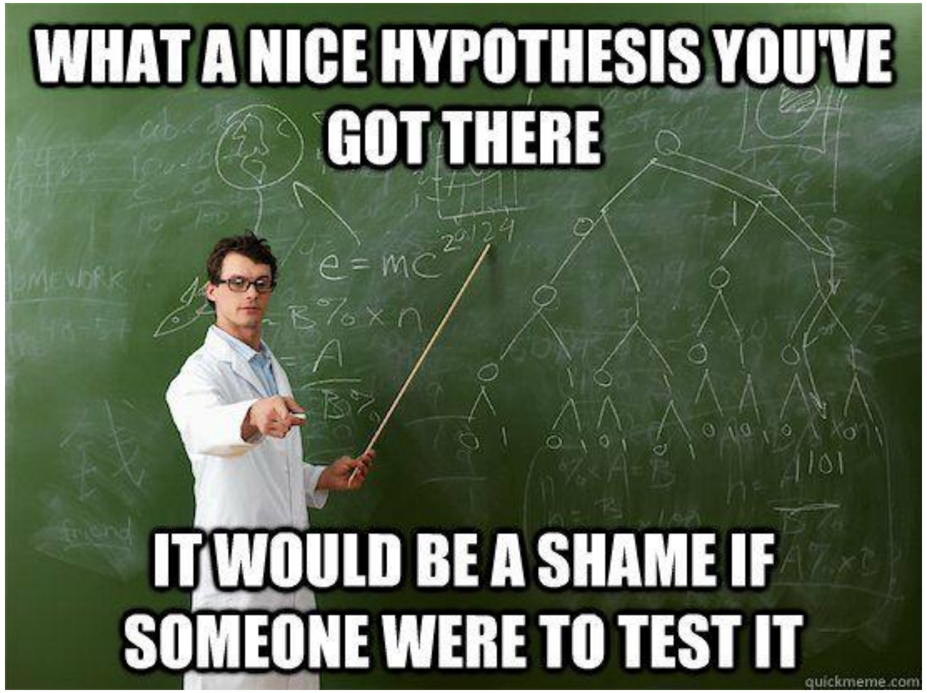
Overview

Causality and hypothesis tests for two proportions

Hypothesis tests for two means

If there is time:

- Hypothesis tests for correlation
- Visual hypothesis tests



Reminder: keep working on your class project

Homework 8 is due on Sunday November 9th

A polished draft of the project is due on November 16th

Review of Statistical Inference

Review: Statistical Inference

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

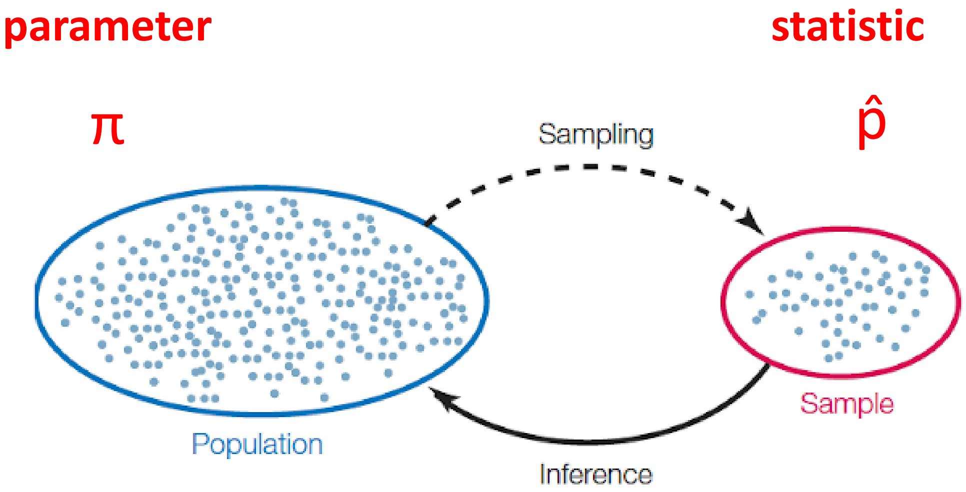
A parameter is number associated with the population

- We use Greek symbols to denote parameters

A **statistic** is number calculated from the sample

- We use Latin symbols to denote statistics

A statistic can be used as an estimate of a parameter



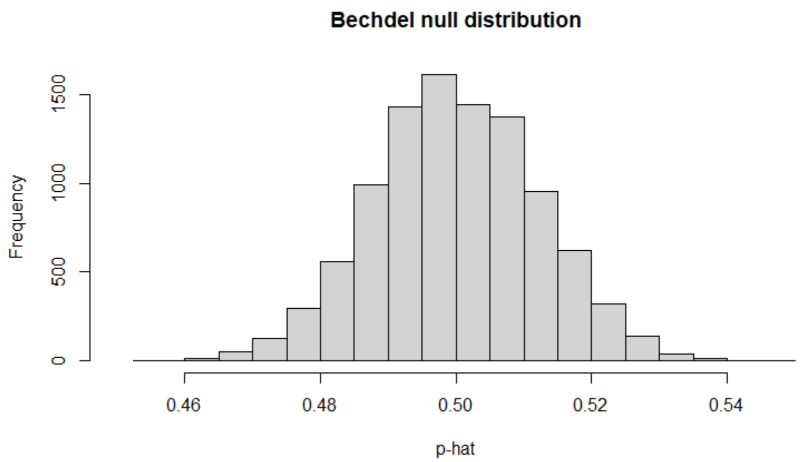
	Sample Statistic	Population Parameter
Mean	$\bar{x}$	μ
Proportion	$\hat{p}$	π
Correlation	r	ρ

Hypothesis tests

Null and Alternative hypotheses

Null hypothesis

- A hypothesis where "nothing interesting" happened
  - E.g., our experiment failed
  - E.g., H_0: $\pi = 0.5$
- We can simulate data under the assumptions of this model to get a "null distribution" of statistics



Alternative hypothesis

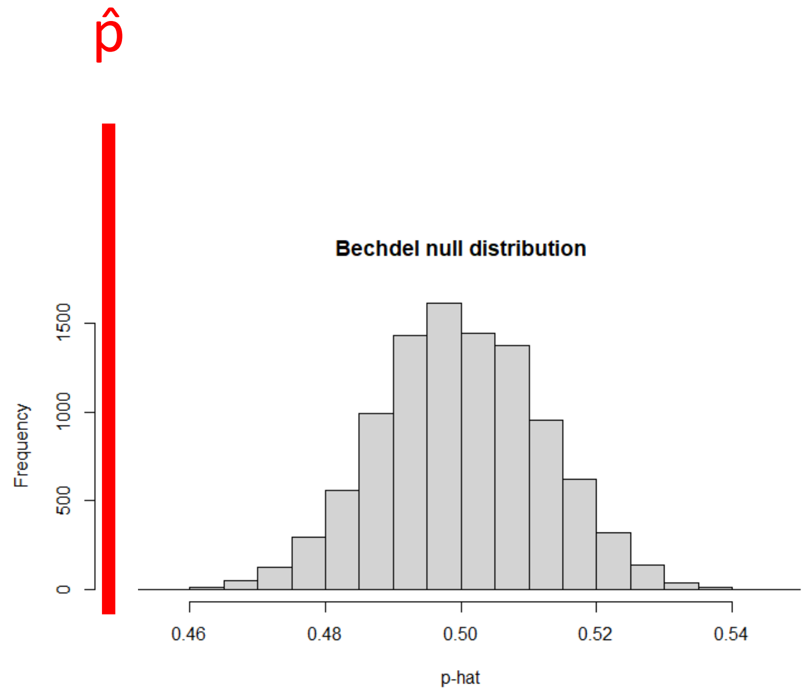
- The hypothesis we believe in (would like to see true)
- E.g., H_A: $\pi < 0.5$

A test statistic is the statistic we choose to simulate in order decide between the two hypotheses

Testing the null hypothesis

To resolve choice between null and alternative hypotheses:

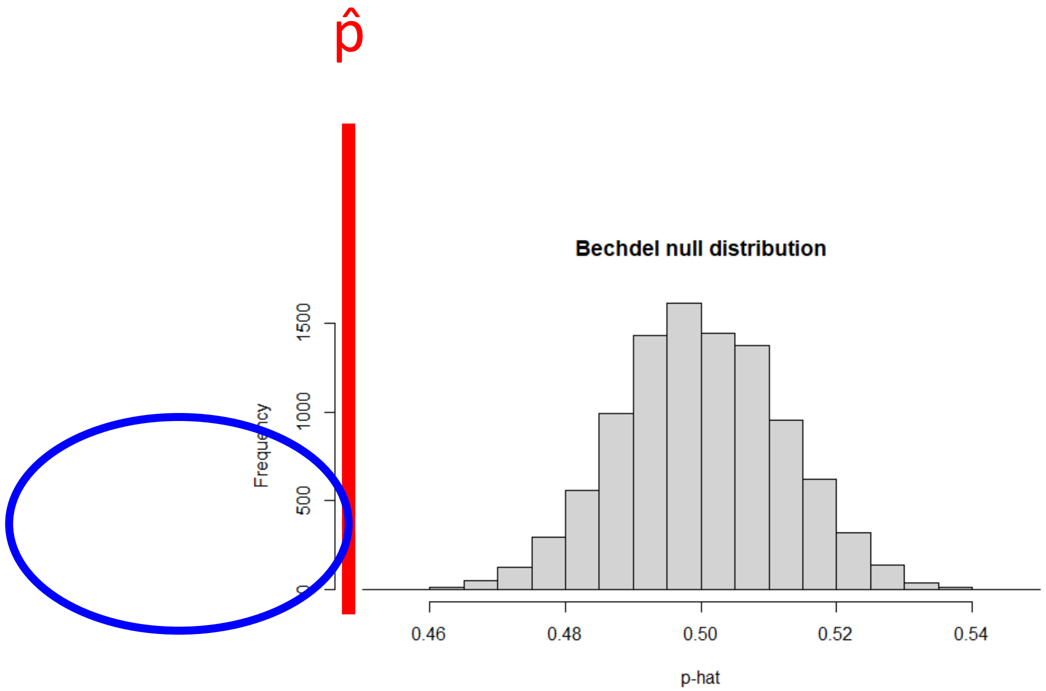
- We compare the observed test statistic to the statistic values in the null distribution
- If the observed statistic is not consistent with the null distribution, then we can reject the null hypothesis
  - E.g., $H_0: \pi = 0.5$
- And we accept the alternative hypothesis
  - E.g., $H_A: \pi < 0.5$



The p-value

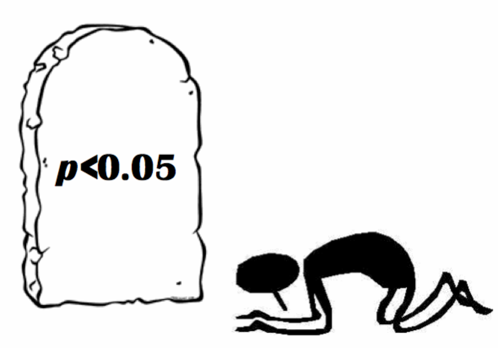
The p-value is the probability, that we get a statistic as or more extreme than the observed statistic from the null distribution

- $P(Null\_Stat \leq obs\_stat \mid H_0)$



If the P-value is small, this is evidence against the null hypothesis and the results are often called "statistically significant"

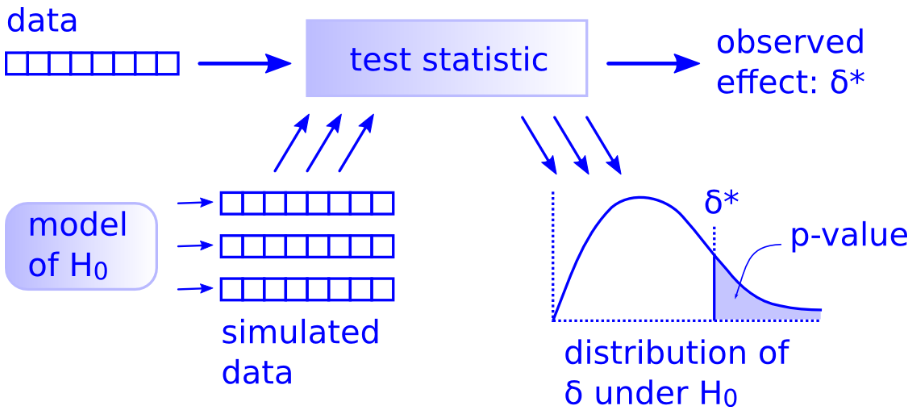
- Convention, p-value < 0.05

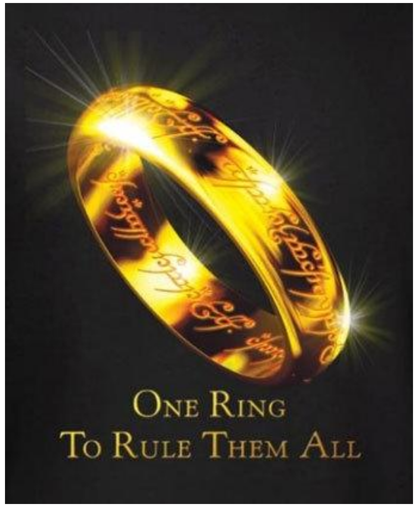


Steps needed to run a hypothesis test

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





Assessing causal relationships

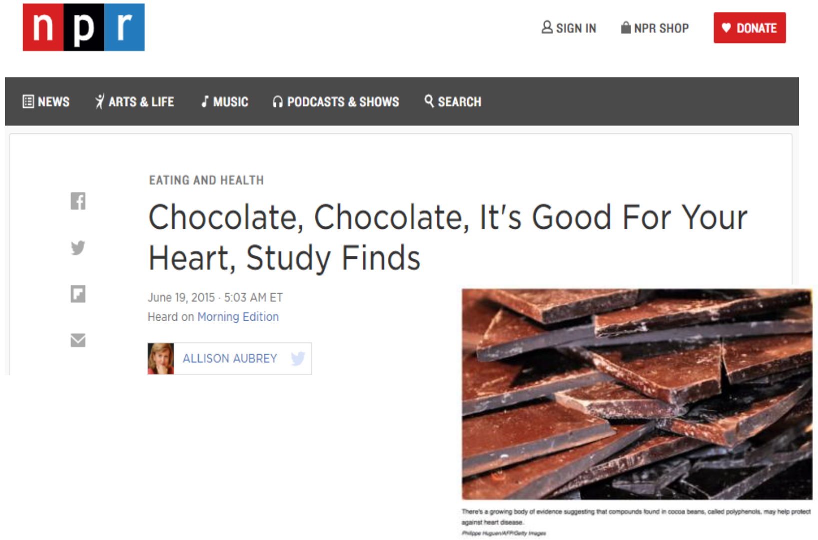
Causality

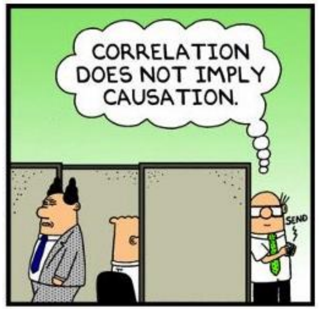
An association is the presence of a reliable relationship between the treatments an outcome

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


npr
SIGN IN NPR SHOP DONATE
NEWS ARTS & LIFE MUSIC PODCASTS & SHOWS SEARCH
EATING AND HEALTH
Chocolate, Chocolate, It's Good For Your Heart, Study Finds
June 19, 2015 · 5:03 AM ET
Heard on Morning Edition
ALLISON AUBREY



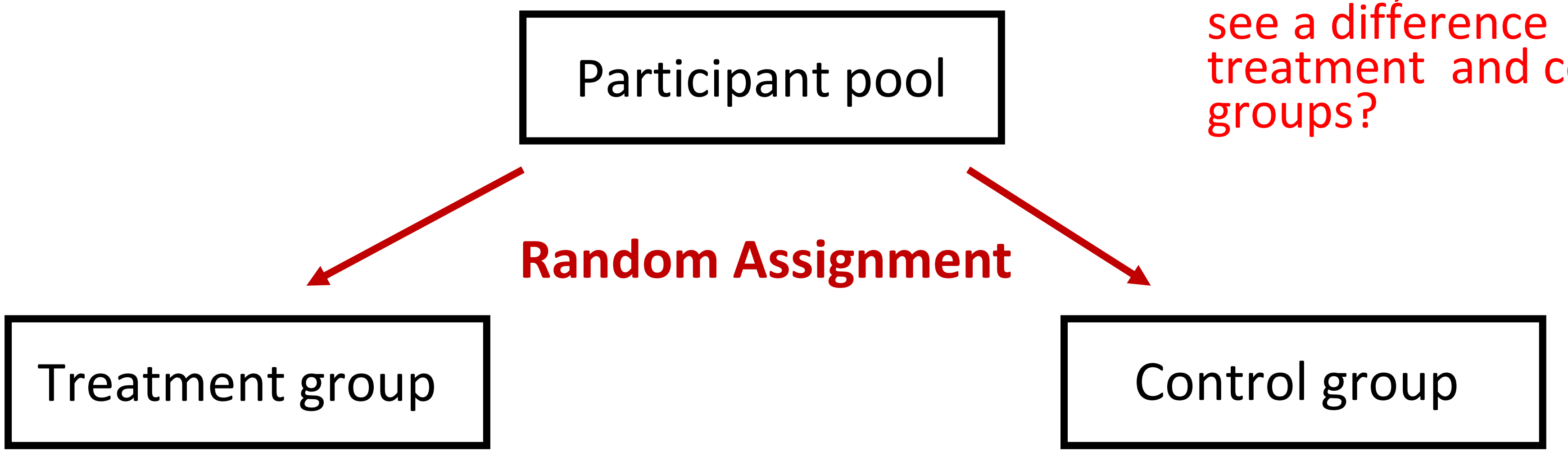
Lurking variable

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

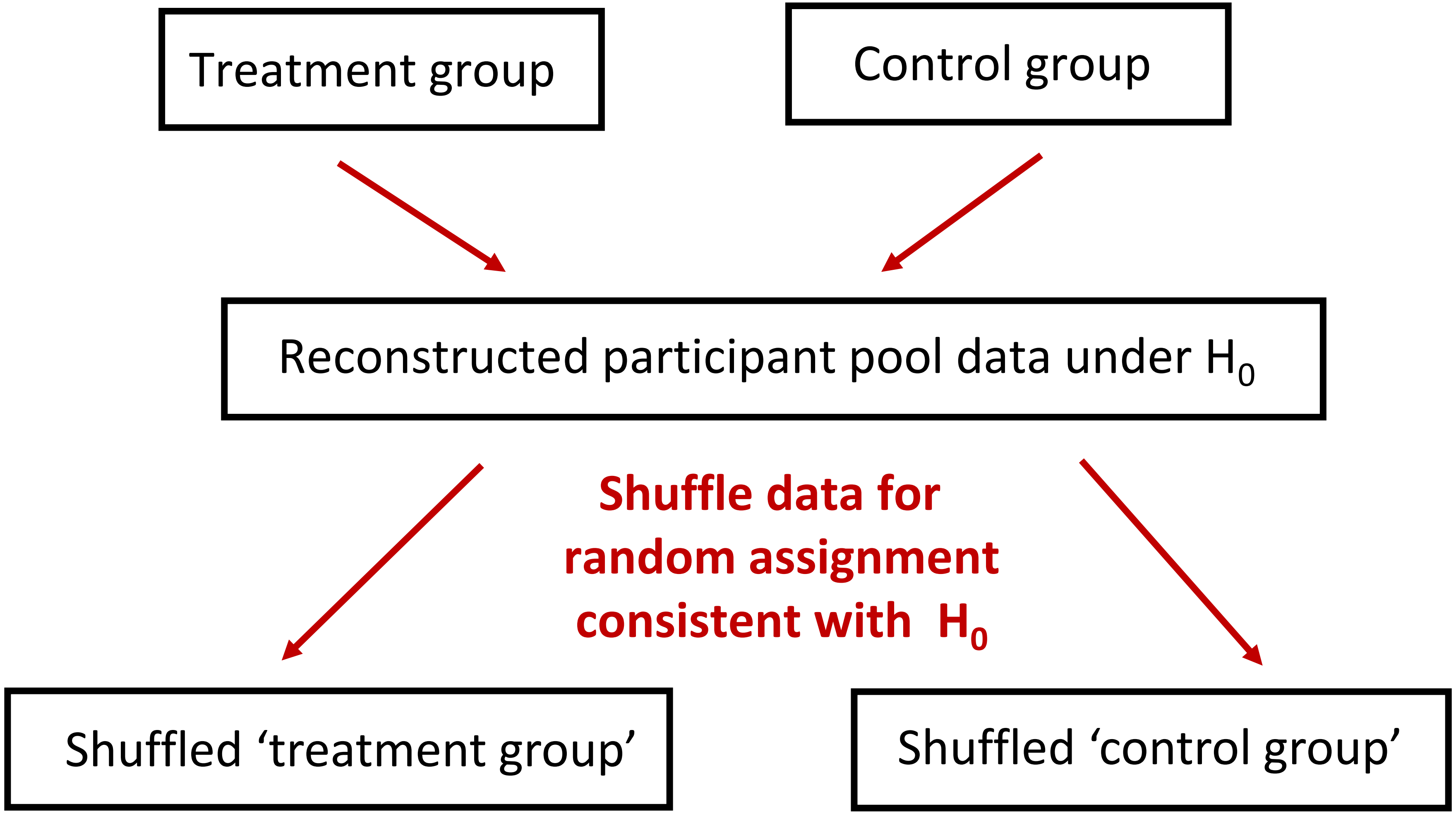
Observed statistic is: $\hat{p}_{treat} - \hat{p}_{control}$

$= 9/15 - 2/16$

$= 0.475$

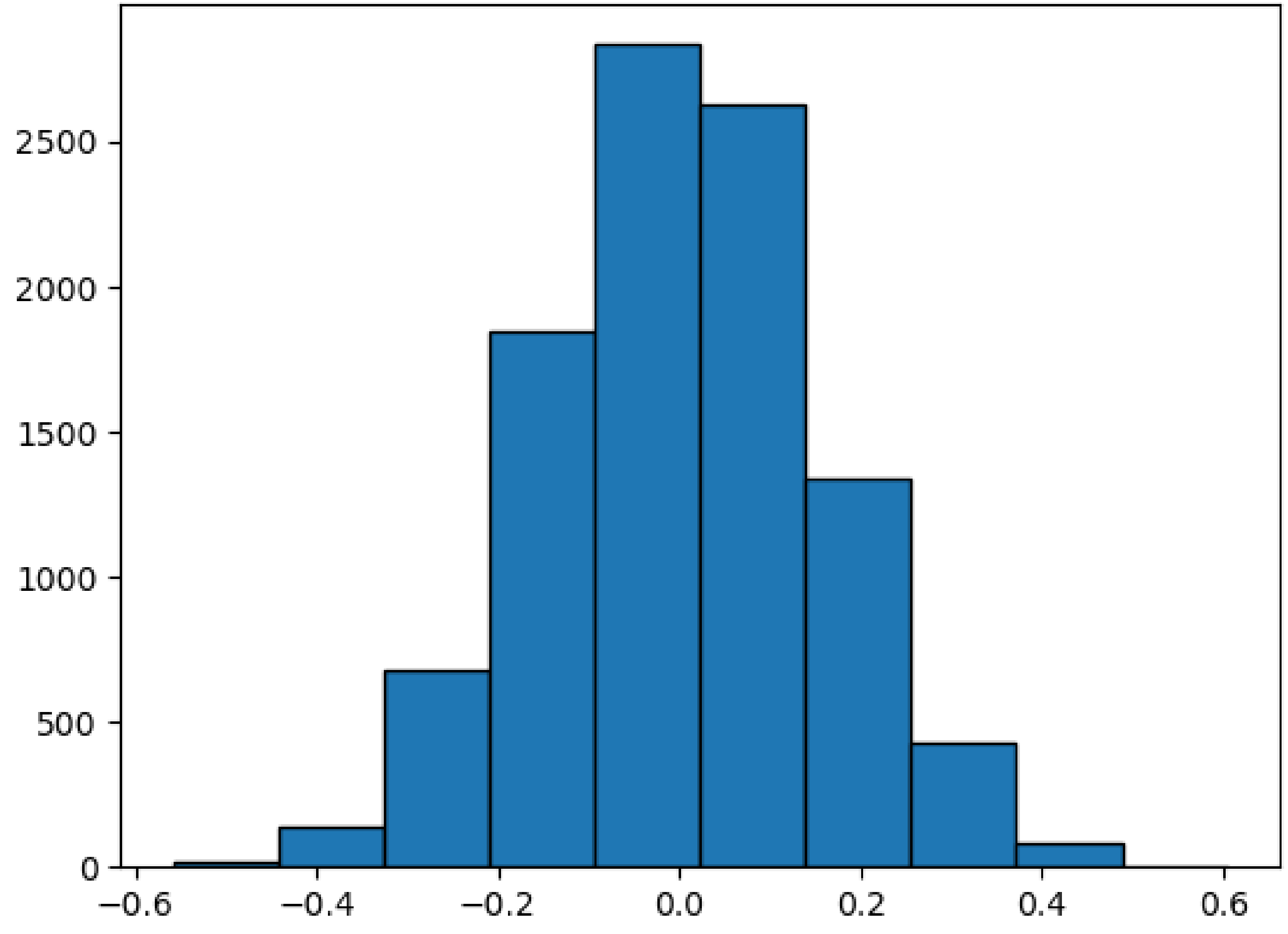
	Group	Result
19	Treatment	1.0
7	Control	0.0
6	Control	0.0
26	Treatment	0.0
17	Treatment	1.0
9	Control	0.0
13	Control	0.0
3	Control	0.0
1	Control	1.0
30	Treatment	0.0
28	Treatment	0.0

3. Create the null distribution!

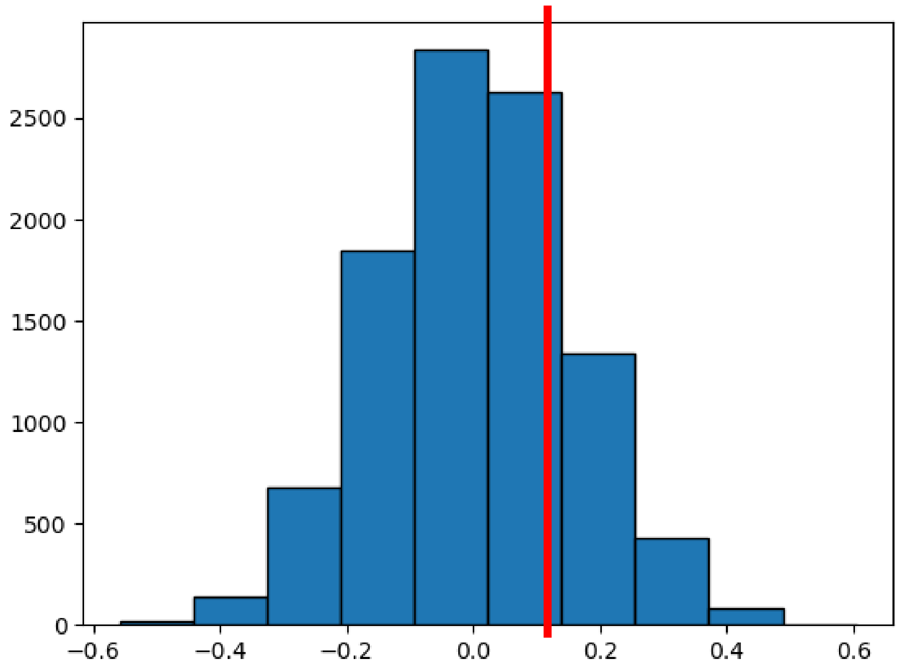


One null distribution statistic: $\hat{p}_{Shuff\_Treatment} - \hat{p}_{Shuff\_control}$

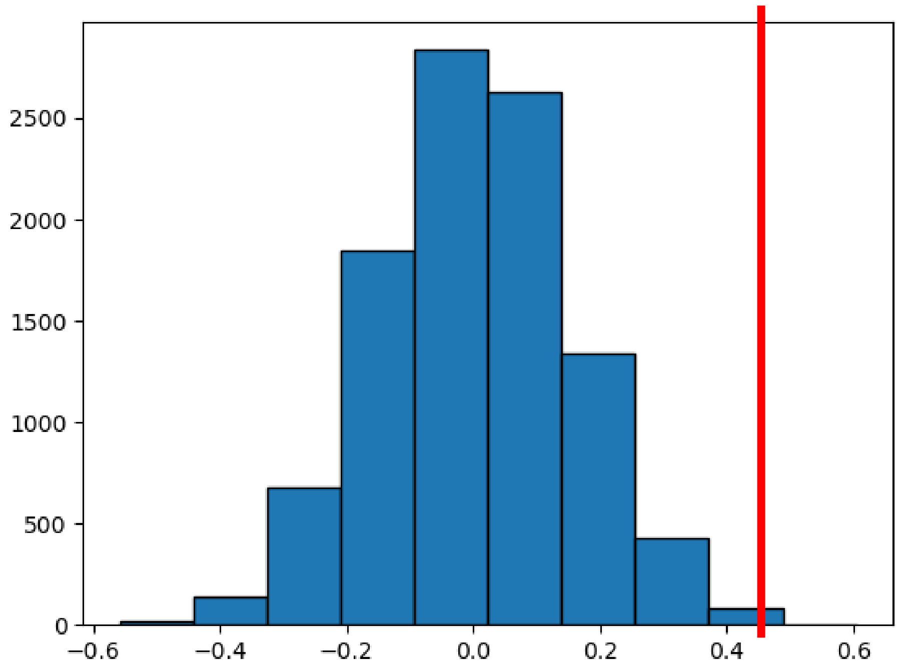
Step 3: Create a null distribution



Step 4: Calculate the p-value

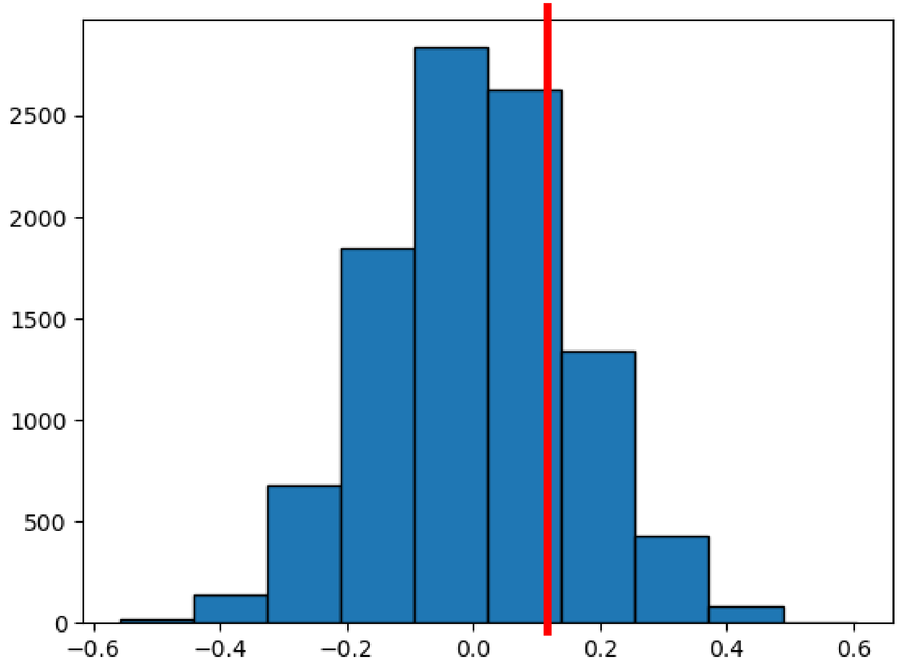


If $\hat{p}_{treat}$ - $\hat{p}_{control}$ = 0.1 what would the p-value be?

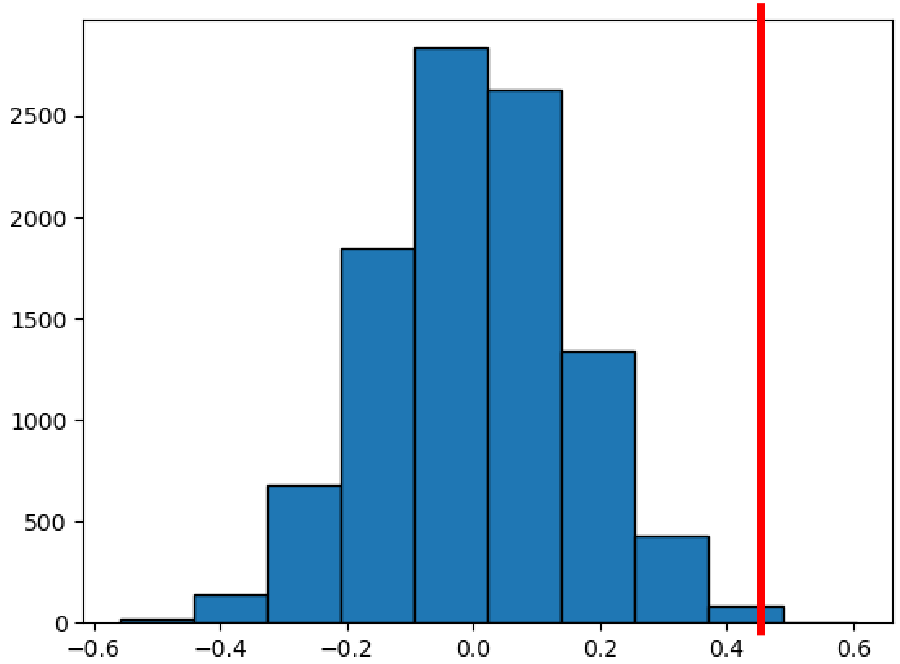


If $\hat{p}_{treat}$ - $\hat{p}_{control}$ = 0.5 what would the p-value be?

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

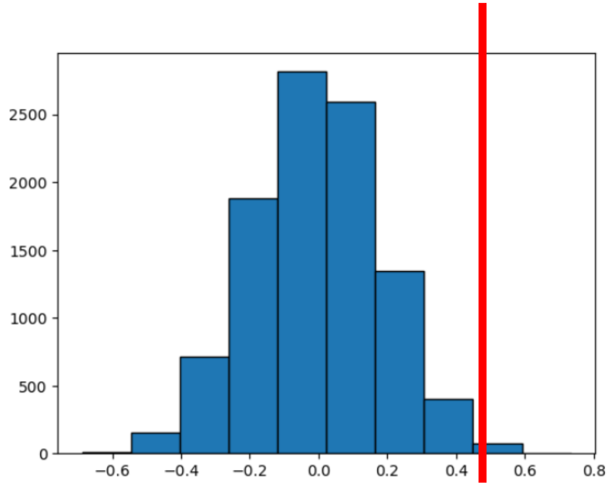
Leslie Foster, Larry Clapp, Marleigh Erickson, Bahman Jabbari

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 explore this in Jupyter!

Baby birth weights

Question: Is the average weight of babies at birth affected by whether a mother smokes?

To gain insight into this question let’s compare:

A. Birth weights of babies of mothers who smoked during pregnancy

B. Birth weights of babies of mothers who didn't smoke

Step 1: State the null and alternative hypotheses

Null hypothesis:

- In the population, the distributions of the birth weights of the babies in the two groups are the same

Alternative hypothesis:

- In the population, the babies of the mothers who didn't smoke were heavier, on average, than the babies of the smokers

How can we write these hypotheses using symbols we have discussed?

H_0: $\mu_{\text{non-smoker}} = \mu_{\text{smoker}}$ or $\mu_{\text{non-smoker}} - \mu_{\text{smoker}} = 0$

H_A: $\mu_{\text{non-smoker}} > \mu_{\text{smoker}}$ or $\mu_{\text{non-smoker}} - \mu_{\text{smoker}} > 0$

Step 2: Compute the observed statistic

Let's look at a data set from 1236 mother-baby pairs that was collected between 1960 and 1967 among women in the Kaiser Foundation Health Plan in the San Francisco East Bay area

- 742 mothers who did not smoke
- 484 mothers who smoked

Statistic: Difference between average baby weights

- $\bar{x}_{\text{non-smokers}} - \bar{x}_{\text{smoker}}$

Large values of this statistic favor the alternative

Step 3: Create the null distribution

If the null is true, all rearrangements of the birth weights among the two groups are equally likely

Plan:

- Shuffle all the birth weights
- Assign some to "Group A" and the rest to "Group B", maintaining the two sample sizes
- Find the difference between the averages of the two shuffled groups
- Repeat

Create the null distribution!

Group A

Group B

Reconstructed data pool

Shuffle data for random assignment consistent with the null hypothesis

Shuffled 'Non-smokers'

Shuffled 'Smokers'

One null distribution statistic: $\bar{x}_{\text{shffle-non-smokers}} - \bar{x}_{\text{shuffle -smoker}}$

Let's explore this in Jupyter!

Hypothesis tests for correlation

Hypothesis tests for correlation

Is there a positive correlation between the number of pages in a book and the price of the book?

What is the population parameter and the statistic of interest?

Hypothesis testing for correlation

1. Write down the null and alternative in symbols and words

Null hypothesis:

- There is no correlation between book price and the number of pages

Alternative hypothesis:

- There is a positive correlation between book price and the number of pages

In symbols:

H_0: $\rho = 0$

H_A: $\rho > 0$

Significance tests for correlation

Let's look at the books from Amazon.com

Title	List.Price	NumPages
1,001 Facts that Will Scare the S#*t Out of You	12.95	304
21: Bringing Down the House	15.00	273
100 Best-Loved Poems	1.50	96
1421: The Year China Discovered America	15.99	672

```
amazon = pd.read_csv("amazon.csv")
```

Try this in Python!

Step 2: What is the observed statistic?

- Also say whether you think you will be able to reject the null hypothesis based on a plot of your data

Step 3: Create the null distribution

- To start with: how we can create one point in the null distribution?
  - Hint: think about shuffling the data

Step 4: What is the p-value that you get?

Step 5: What decision would you make?

Visual hypothesis test

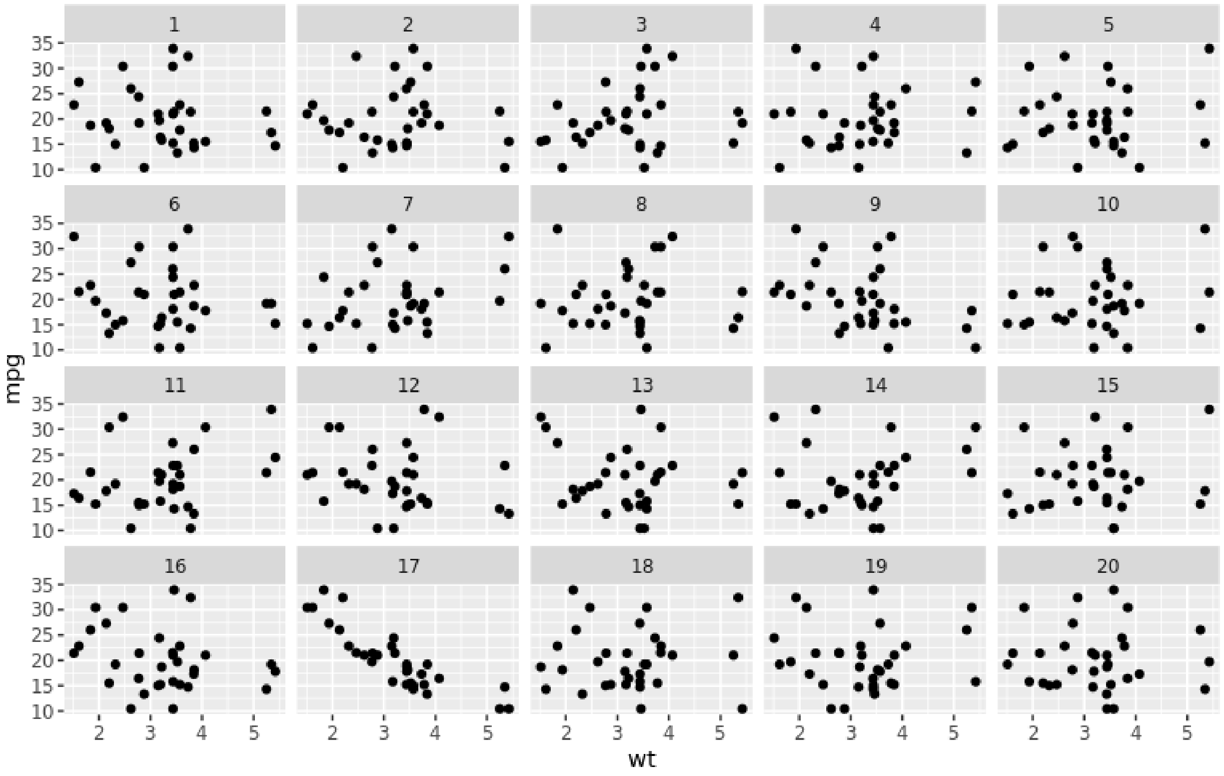
Visual hypothesis test

In visual hypothesis tests, we create data visualizations to try to assess whether particular relationships exist in our data.

- One way this is done through a visual lineup

Visual hypothesis test

Which plot shows the true relationship between a car's weight and the number of miles per gallon a car gets?



Let’s try it in Jupyter

Brief mention: two-sided hypothesis tests

Brief mention: two-sided hypothesis tests

So far we have always had a specific prediction for the effect we observed

For example:

- We believed that *less than* 50% of movies passed the Bechdel test
- We believed that babies or mothers who did not smoke would weigh ***more*** (on average) than babies of mothers who smoked

This directionality was reflected in our alternative hypotheses

- H_A: $\pi_{\text{Bechdel}} < .5$
- H_A: $\mu_{\text{non-smoke}} > \mu_{\text{smoke}}$

Brief mention: two-sided hypothesis tests

Sometimes we do not know the direction of an effect, we only know that the value specified in the null hypothesis is not correct

For example:

- We just know that 50% of movies do not pass the Bechdel test
  - But it could be than more 50% or less than 50%

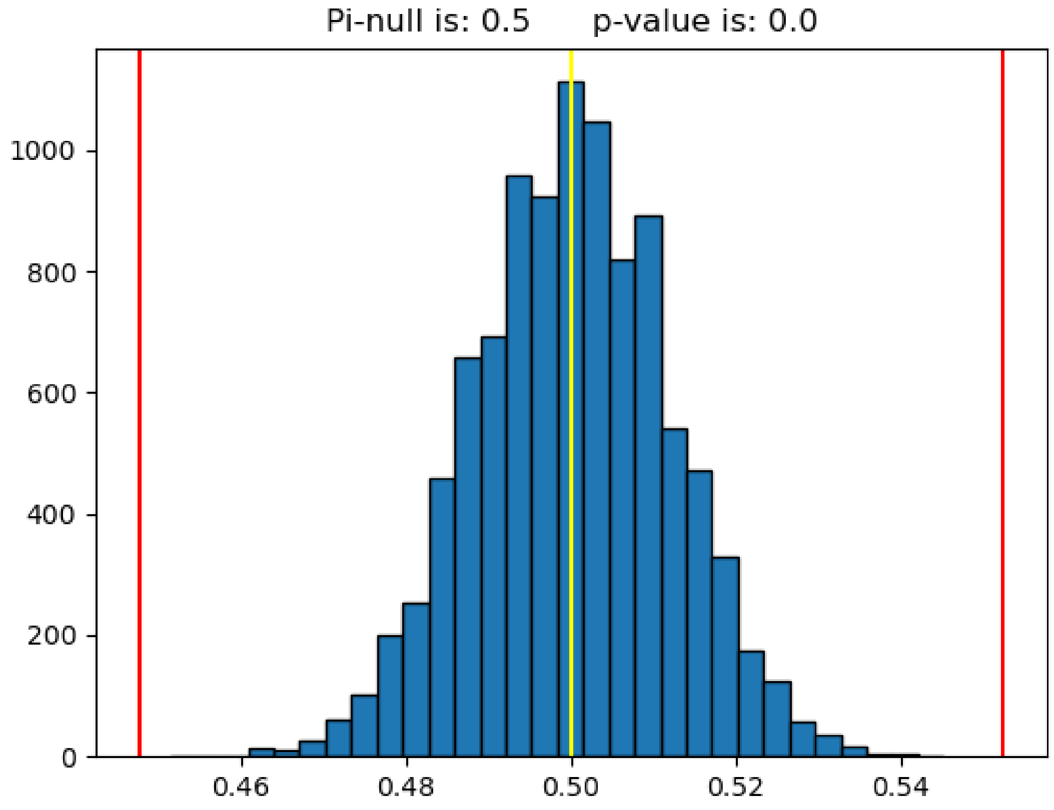
We would then write our alternative hypotheses as:

- H_A: $\pi_{Bechdel} \neq .5$
- H_A: $\mu_{non\text{-}smoke} \neq \mu_{smoke}$

Brief mention: two-sided hypothesis tests

When we have a “two-sided” alternative hypothesis, we need to calculate the the statistics that are “more extreme” than the observed statistic to get the p-value

- i.e., we need to look at both tails of our null distribution to get the p-value



Let's explore this in Jupyter!