

YData: Introduction to Data Science

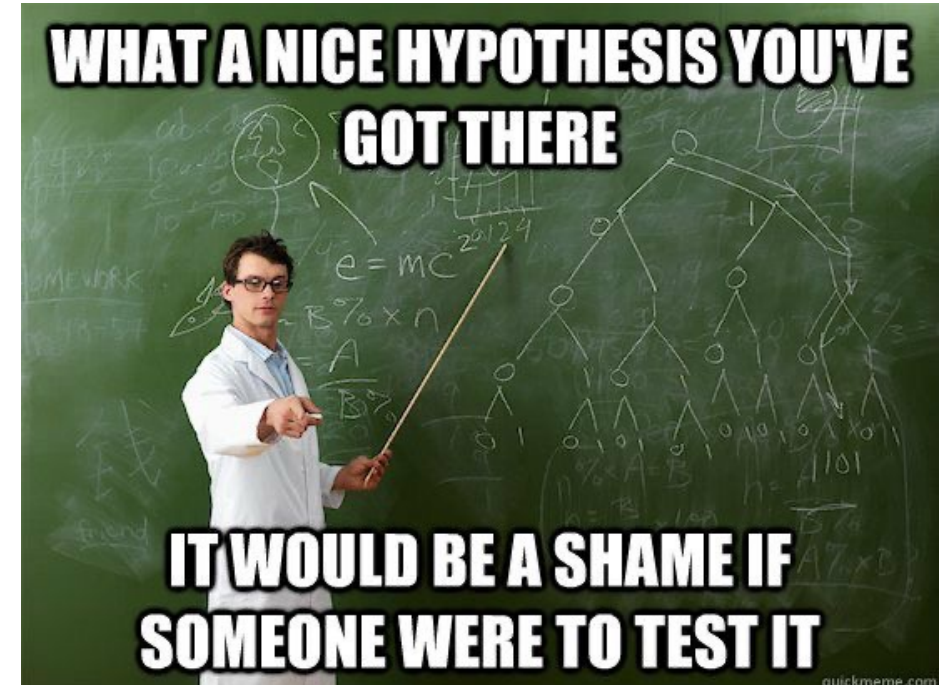


Class 20: Hypothesis tests for assessing causation

Overview

Quick review of parameters, statistics, sampling, and hypothesis tests for a single proportion

Hypothesis tests for two means and assessing causality



Project timeline

Sunday, April 7th

- Projects are due on Gradescope at 11pm
- Also, email a pdf of your project to your peer reviewers
 - A list of whose paper you will review is on Canvas

Wednesday, April 17th

- Jupyter notebook files with your reviews need to be sent to the authors
- A template for doing your review will be available

Sunday, April 28th

- Project is due on Gradescope
 - Add peer reviews to the Appendix of your project



Review of Statistical Inference

Review: Statistical Inference

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

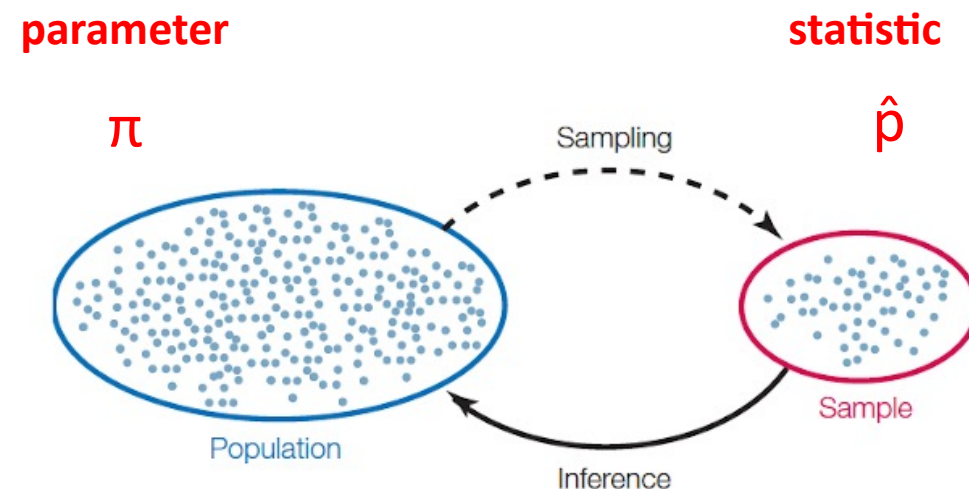
A parameter is number associated with the population

- e.g., population proportion π
- e.g., the proportion of all voters who voted for Biden

A statistic is number calculated from the sample

- e.g., sample proportion \hat{p}
- e.g., the proportion of Biden's vote out of 1,000 people in a sample

A statistic can be used as an estimate of a parameter



	Sample Statistic	Population Parameter
Mean	\bar{x}	μ
Proportion	\hat{p}	π

Probability distribution of a statistic

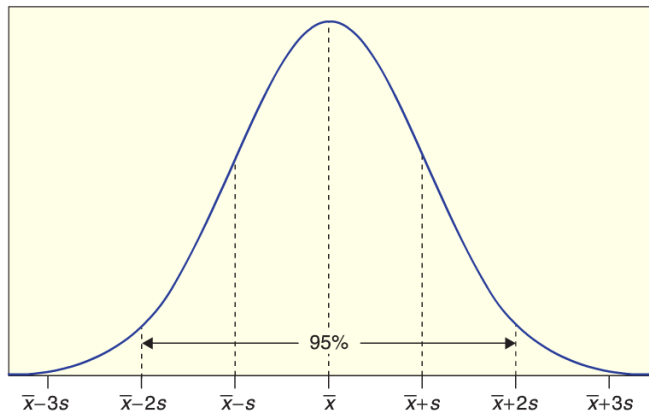
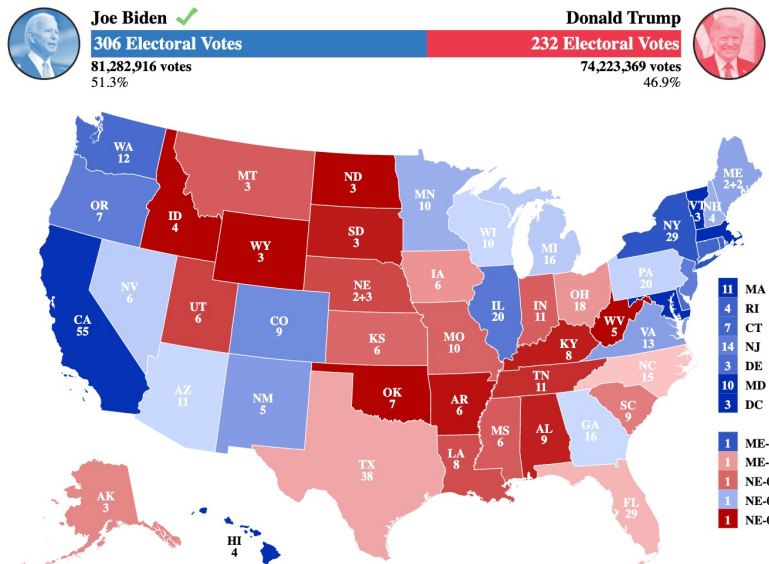
Values of a statistic vary because random samples vary

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 simulating statistics

π Biden

n = 1,000



Sampling distribution!



\hat{p}_{Biden}



\hat{p}_{Biden}



\hat{p} Biden

Simulating random proportions (\hat{p} 's)

We can ***simulate random proportions*** \hat{p} consistent with a population proportion π by:

1. Generated n random numbers uniformly distributed between 0 and 1
 - `rand_nums = np.random.rand(1000)`
2. Marking points less than π as being **True**, and greater π than as being **False**
 - `rand_binary = rand_nums <= pi_value`
3. Calculating the proportion of points to get a \hat{p}
 - `rand_phat = np.mean(rand_binary)`



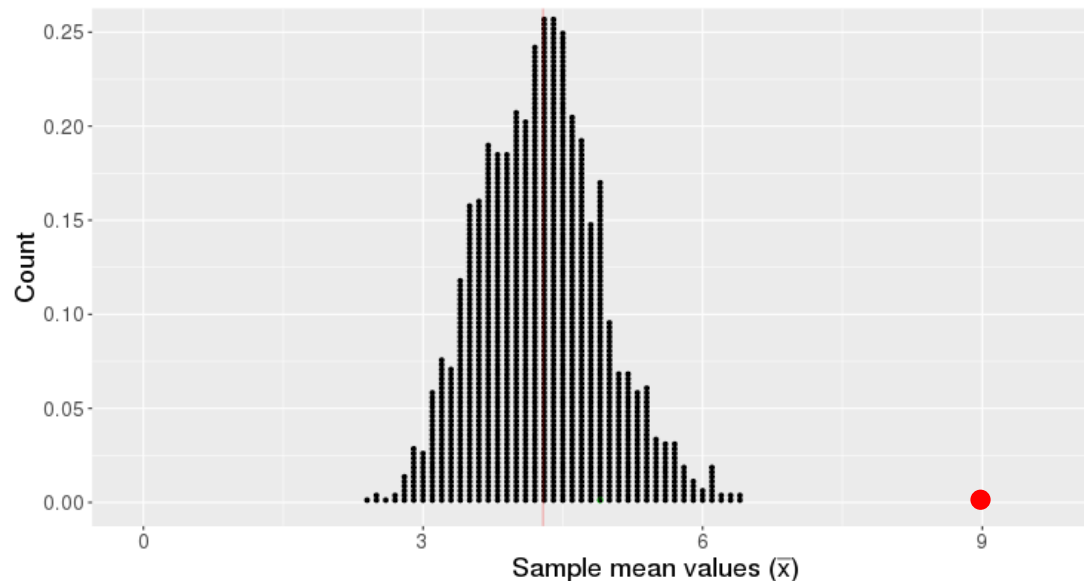
Hypothesis tests

Basic hypothesis test logic

We start with a claim about a population parameter

- E.g., $\mu = 4$

This claim implies we should get a certain distribution of statistics



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

Null and Alternative hypotheses

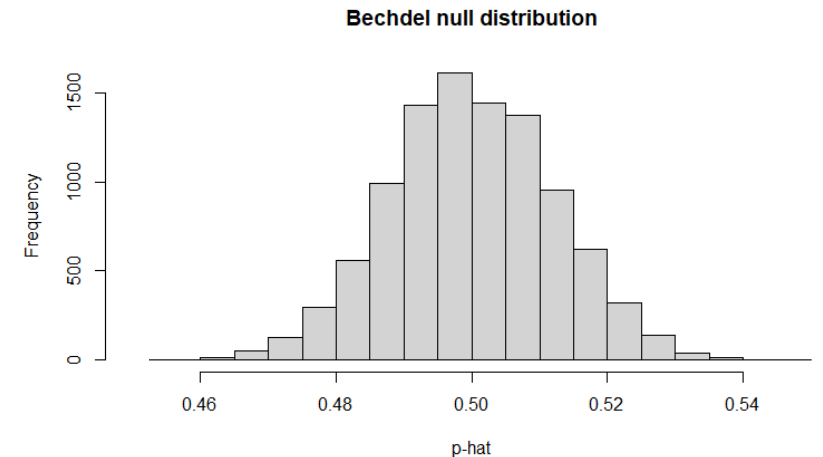
Null hypothesis

- 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., $H_A: \pi < 0.5$

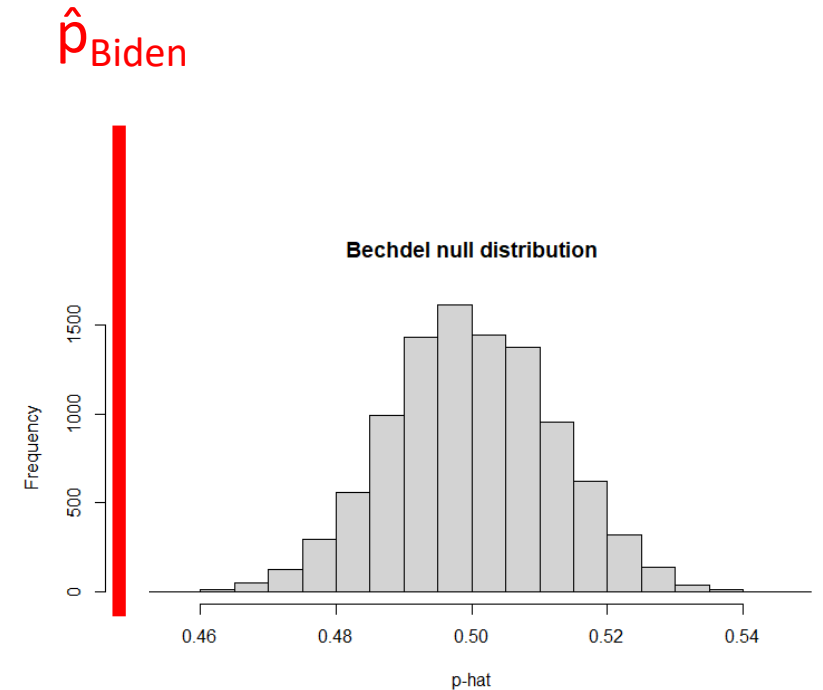
A **test statistic** is the statistic we choose to simulate in order to decide between the two hypotheses



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., $H_A: \pi < 0.5$



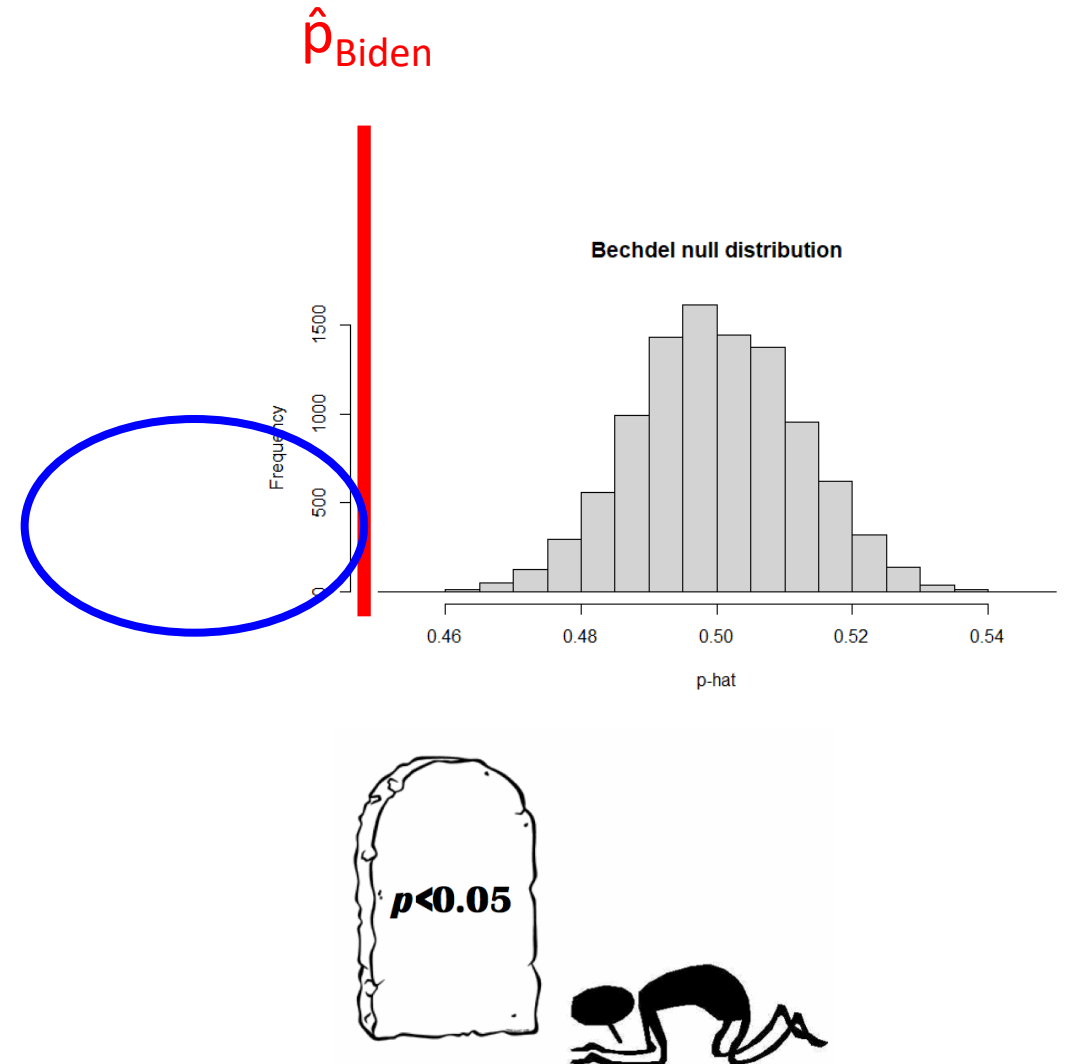
The p-value

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

- $P(\text{Null_Stat} \leq \text{obs_stat} \mid H_0)$

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

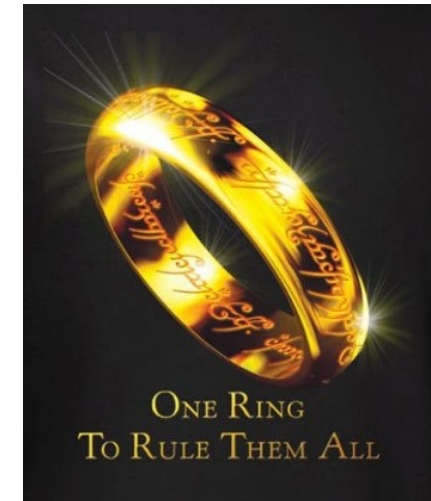
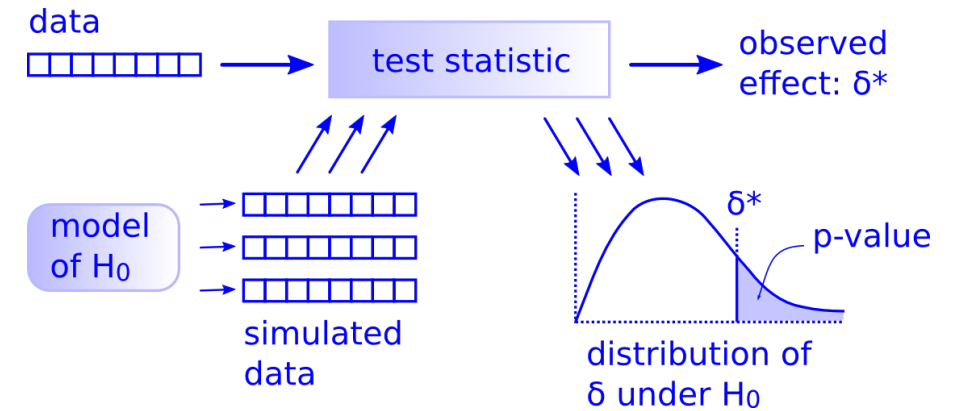
- Convention, $p\text{-value} < 0.05$



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



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

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

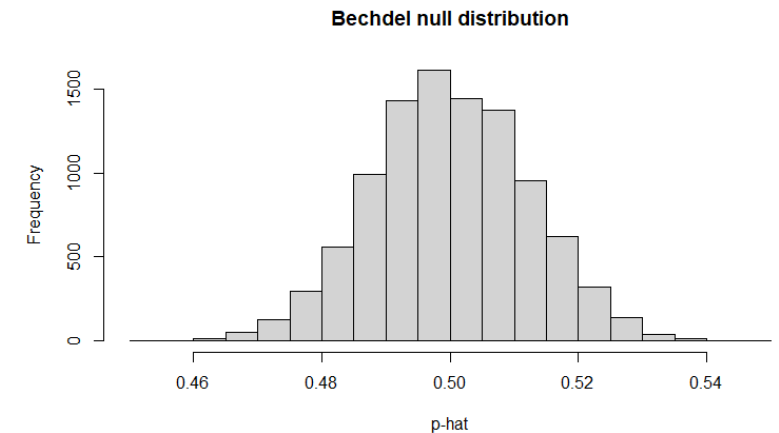
- What is the probability that only 803 of 1794 movies would pass the Bechdel test ($\hat{p} = .448$) if the null hypothesis was true?
- i.e., what is the p-value?

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'



$$\hat{p} = .448$$



One more 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



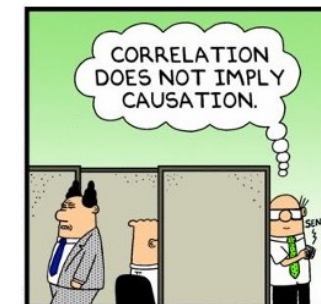
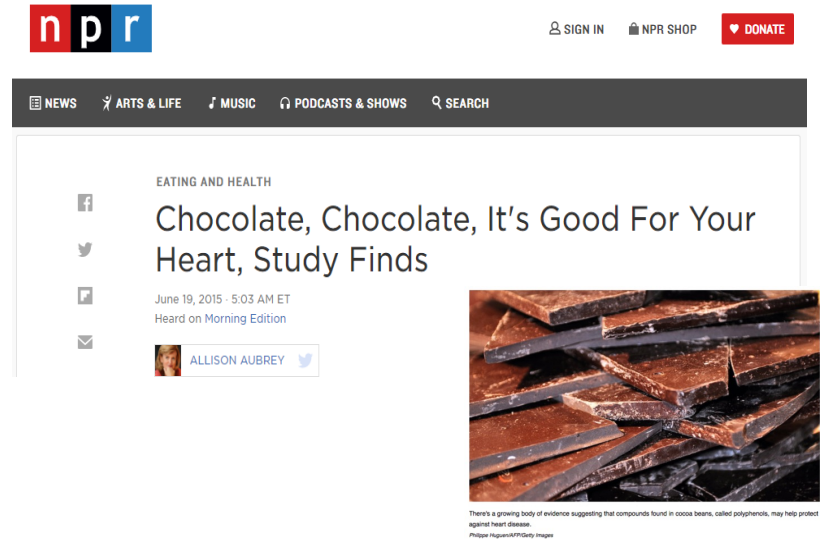
Let's explore this in Jupyter!

Assessing causal relationships

Review: Causality

Recall from class 2:

- **An association** is the presence of a reliable relationship between the treatments and 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

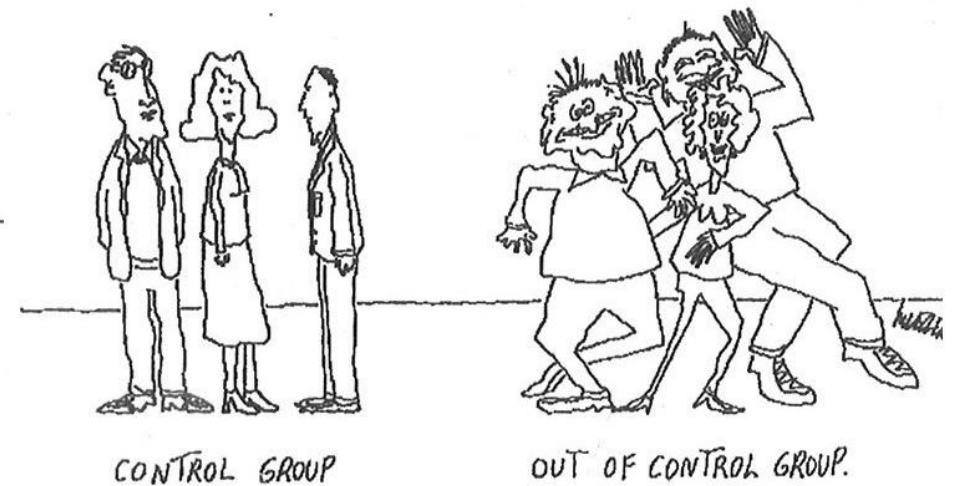
Sample B: treatment group

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

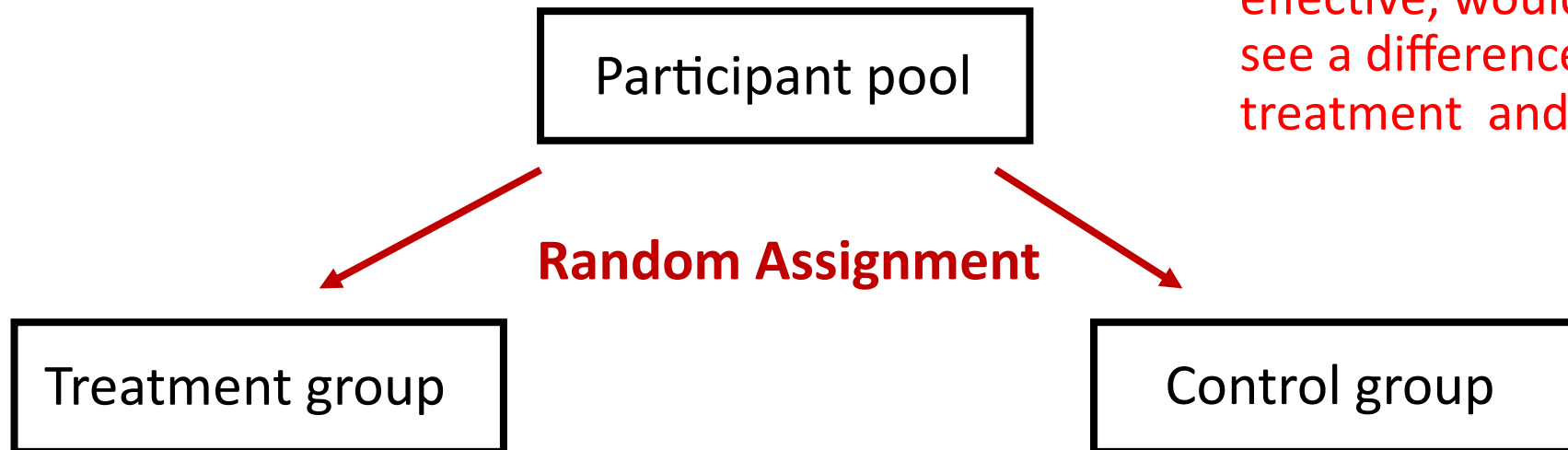
Randomly assigning participants to treatment and control groups allows us to separate what expected by chance and consequently what is due to the treatment



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.
 - $H_0: \pi_{\text{treat}} = \pi_{\text{control}}$

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
 - $H_A: \pi_{\text{treat}} > \pi_{\text{control}}$

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:

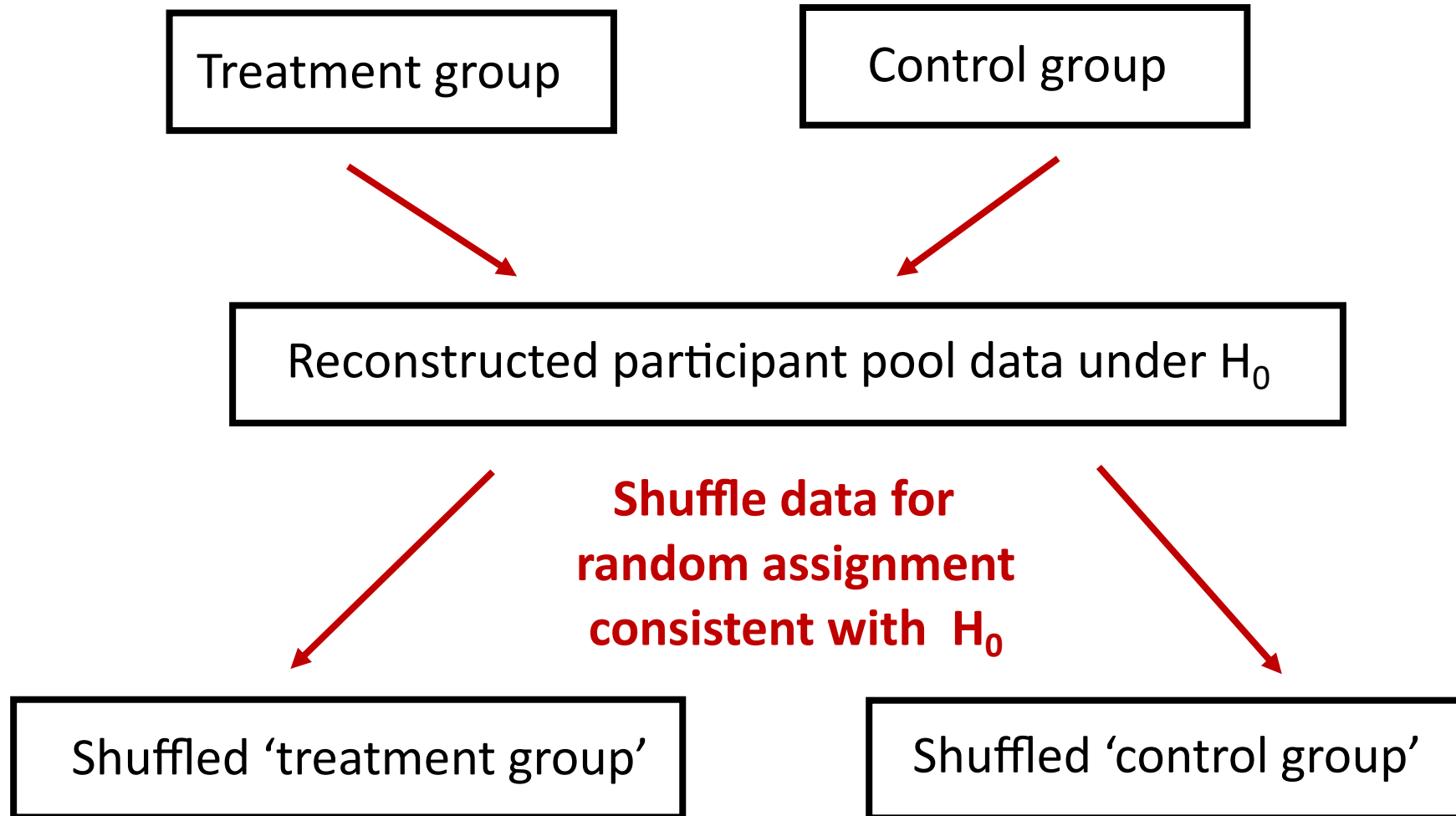
Let's have our observed statistic mirror our hypotheses

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

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

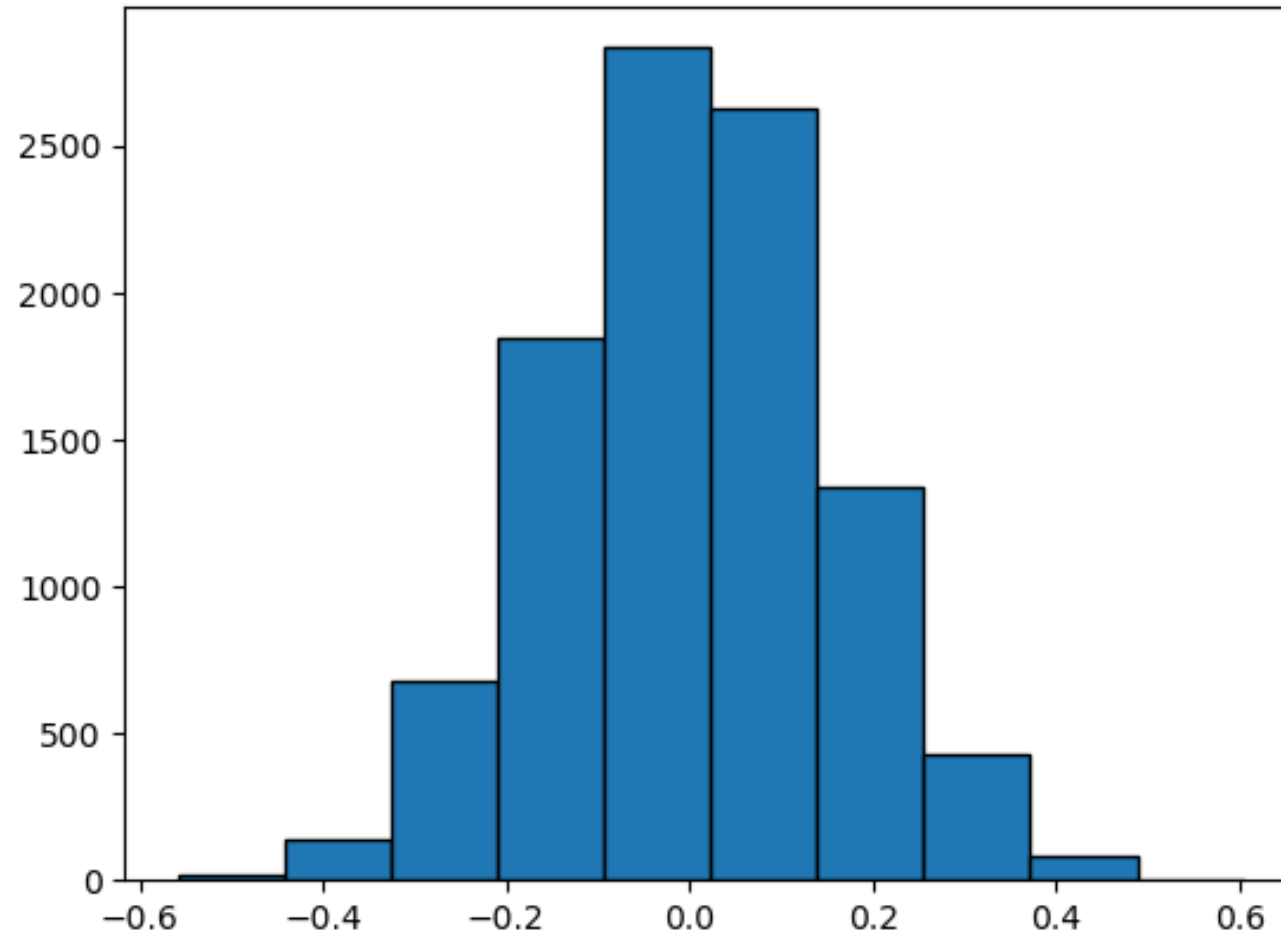
	Group	Result
19	Treatment	1.0
7	Control	0.0
6	Control	0.0
26	Treatment	0.0
17	Treatment	1.0
9	Control	0.0
13	Control	0.0
3	Control	0.0
1	Control	1.0
30	Treatment	0.0
28	Treatment	0.0

3. Create the null distribution!

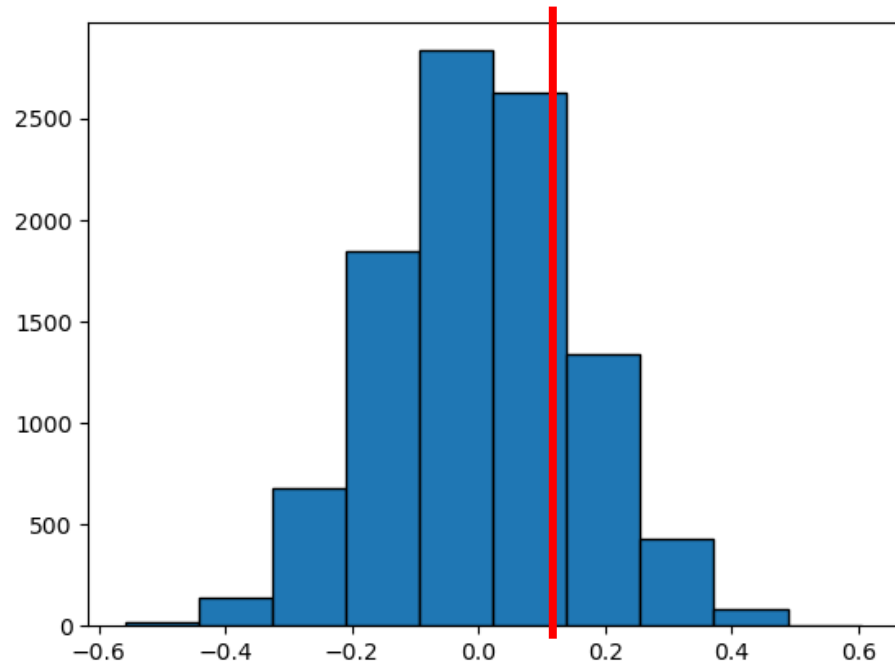


One null distribution statistic: $\hat{p}_{\text{Shuff_Treatment}} - \hat{p}_{\text{Shuff_control}}$

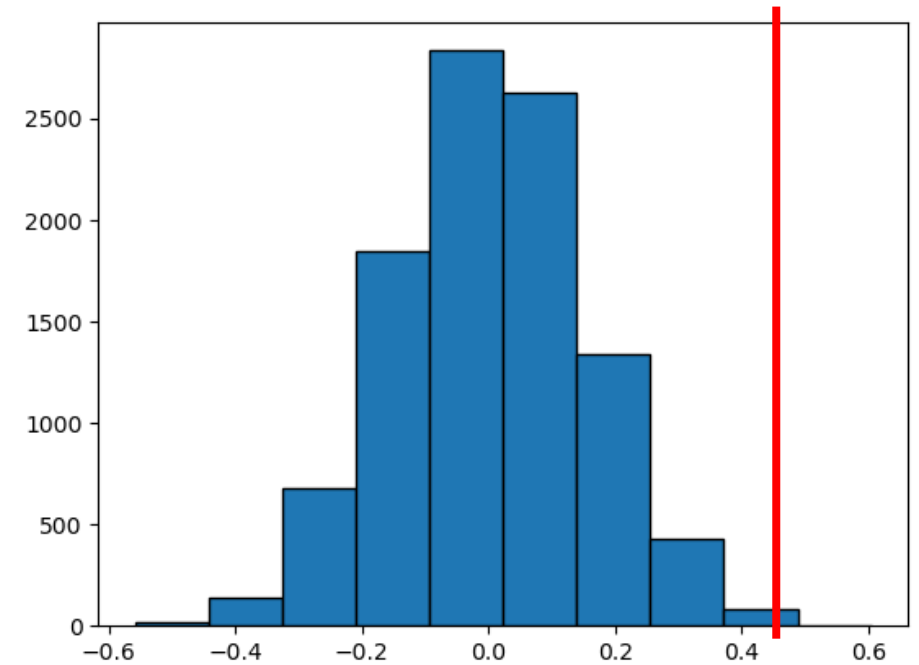
Step 3: Create a null distribution



Step 4: Calculate the p-value

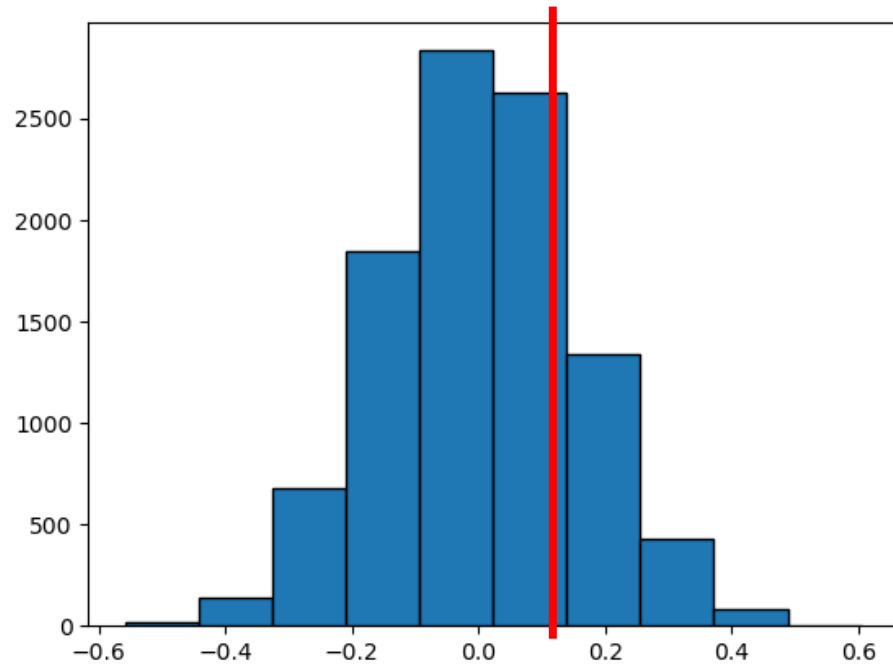


If $\hat{p}_{\text{treat}} - \hat{p}_{\text{control}} = 0.1$ what would the p-value be?

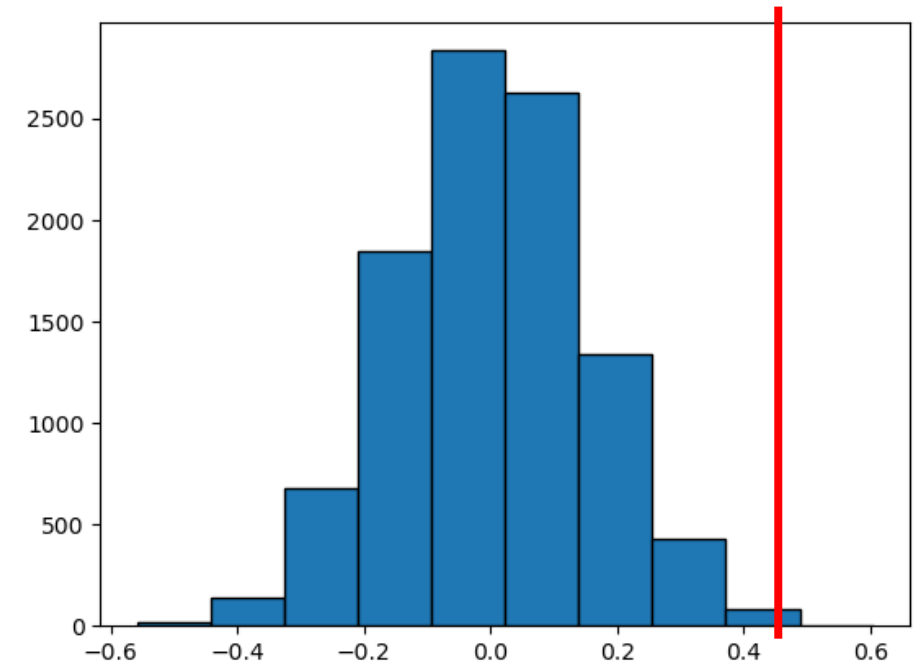


If $\hat{p}_{\text{treat}} - \hat{p}_{\text{control}} = 0.5$ what would the p-value be?

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?



Let's explore this in Jupyter!

Summary: BTA for back pain relief

1. State the null hypothesis and the alternative hypothesis

- BTA does not lead to an increase in pain relief: $H_0: \pi_{\text{treat}} = \pi_{\text{control}}$
- BTA leads to an increase in pain relief: $H_A: \pi_{\text{treat}} > \pi_{\text{control}}$

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

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

- The $\hat{p}_{\text{treat}} - \hat{p}_{\text{control}}$ 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 $\hat{p}_{\text{treat}} - \hat{p}_{\text{control}}$ 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

Neurology®

May 22, 2001; 56 (10) ARTICLES

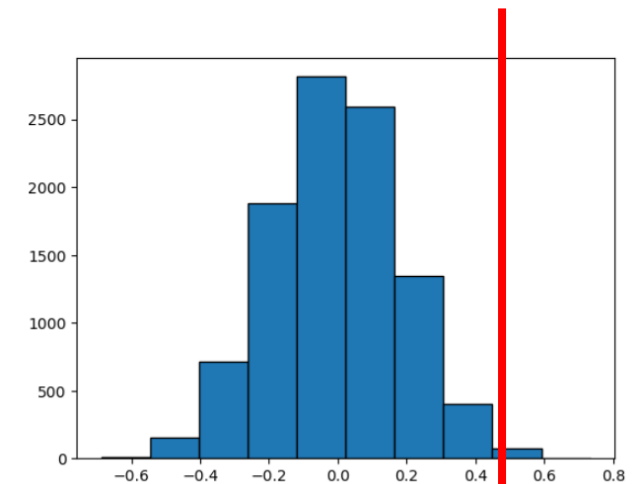
Botulinum toxin A and chronic low back pain

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>

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



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?

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

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

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 weights

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

Large values of this statistic favor the alternative



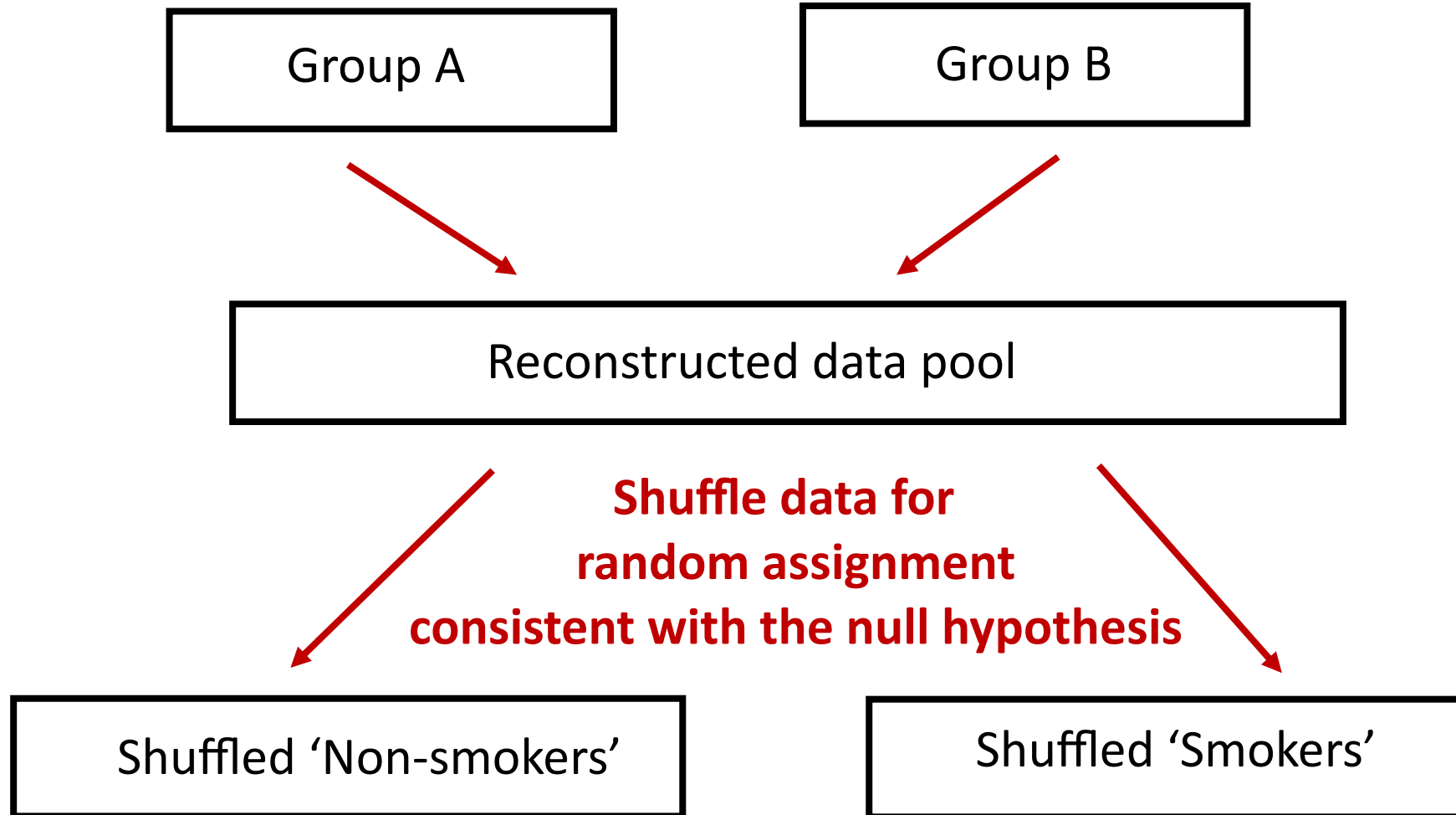
Step 3: Create the null distribution

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!



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



Let's explore this in Jupyter!