YData: Introduction to Data Science

Class 19: Hypothesis tests for proportions

Overview

Review of parameters, statistics, sampling distributions

Review and continuation of hypothesis tests for a single proportion

If there is time:

- Hypothesis for two proportions
- Hypothesis tests for two means

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

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

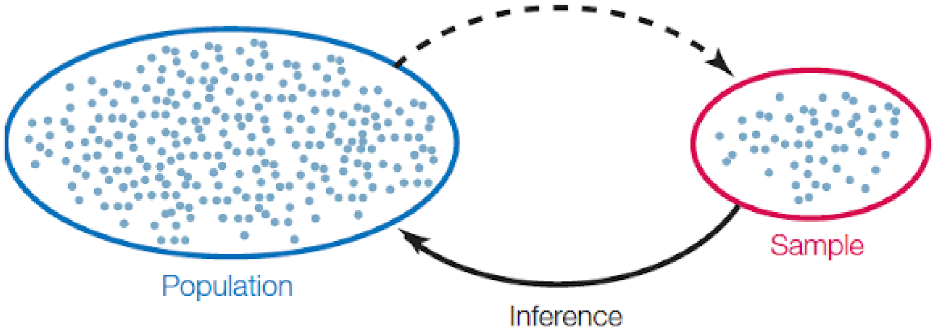
A **statistic** is number calculated from the sample

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

A statistic can be used as an estimate of a parameter

parameter **statistic**

π $\hat{p}$



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

Probability distribution of a statistic

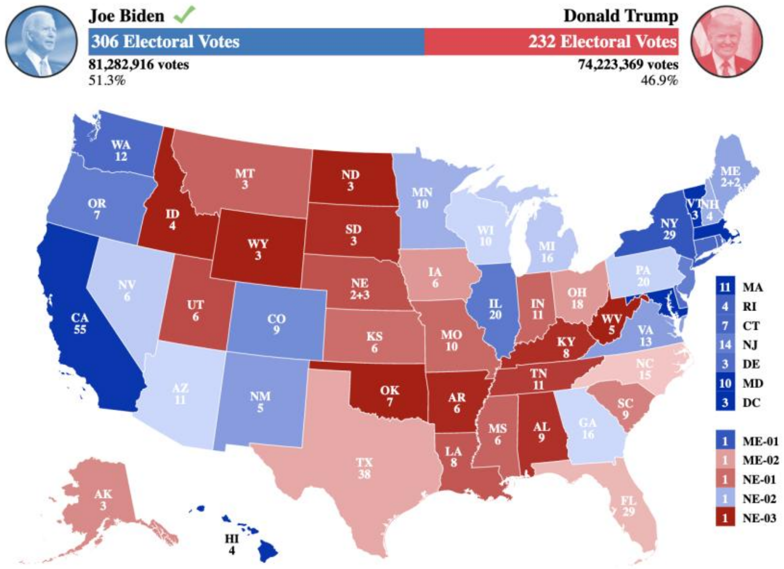
Values of a statistic vary because random samples vary

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_{Trump}

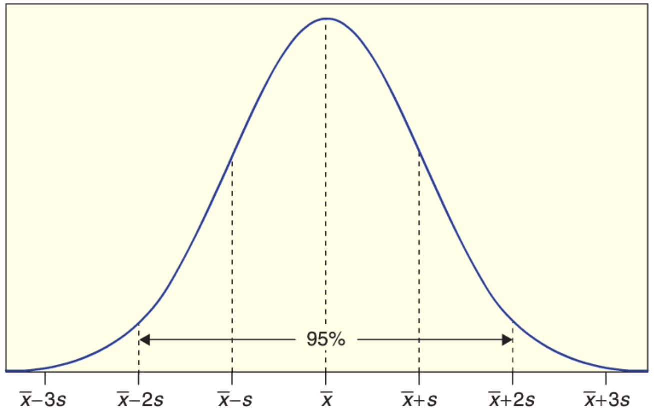
n = 1,000



$\hat{p}_{Trump}$

$\hat{p}_{Trump}$

$\hat{p}_{Trump}$



Sampling distribution!

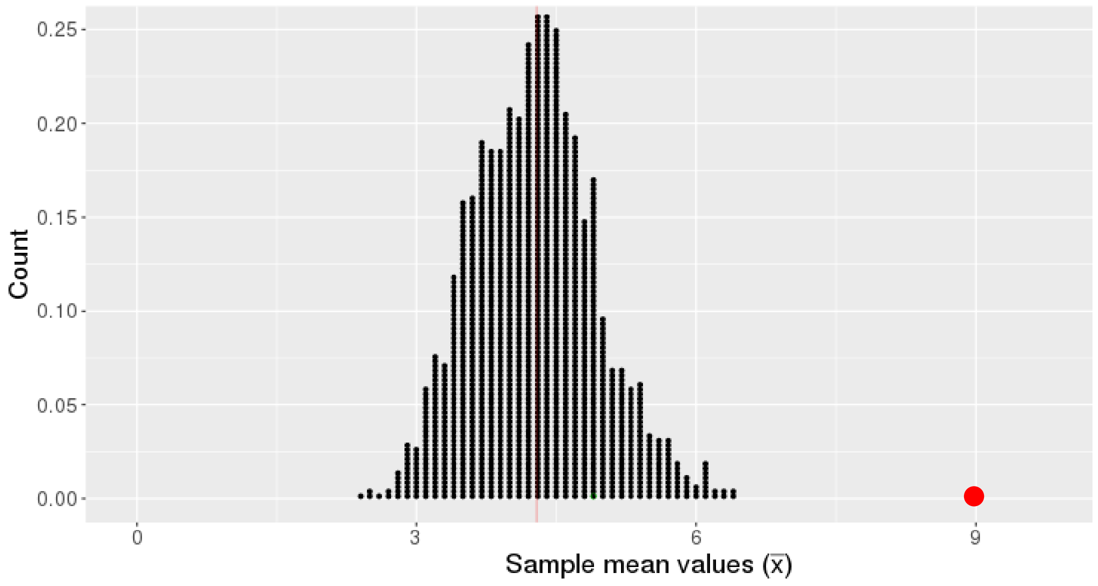
Hypothesis tests

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

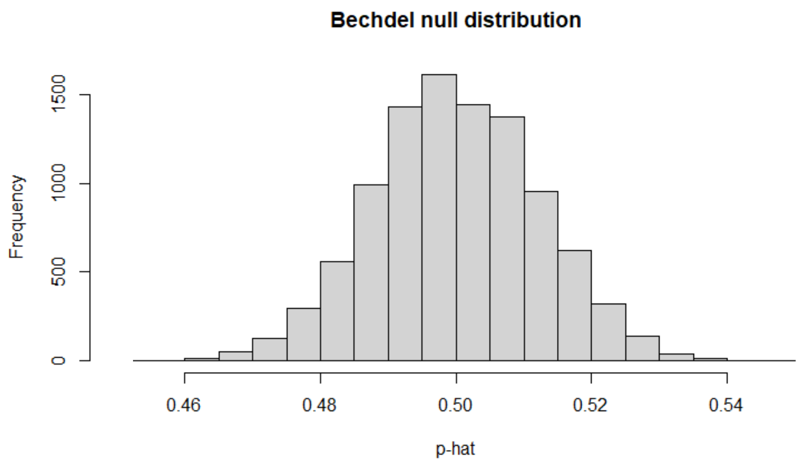


If our observed statistic is highly unlikely, we reject the claim

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

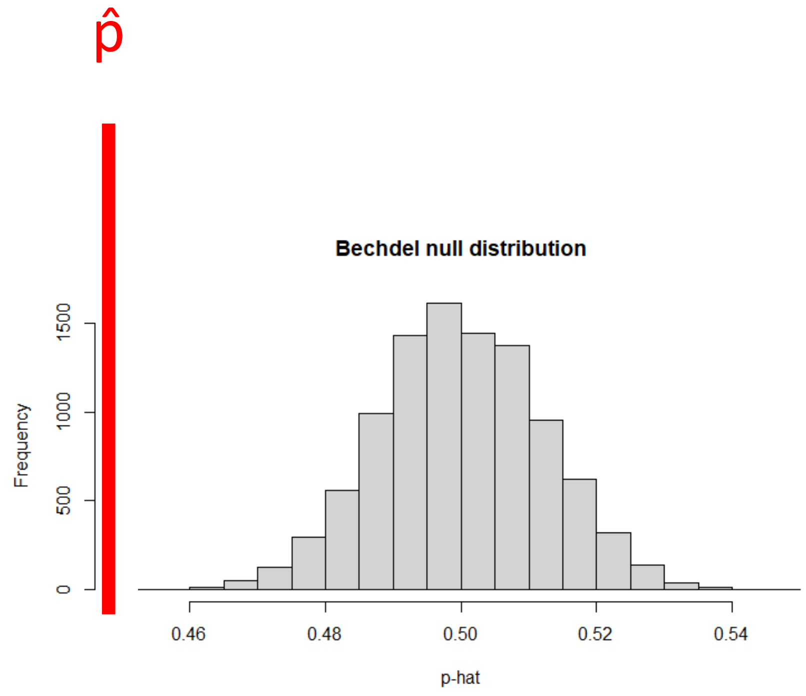
- The hypothesis we believe in (would like to see true)
- E.g., H_A: $\pi < 0.5$

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

Testing the null hypothesis

To resolve choice between null and alternative hypotheses:

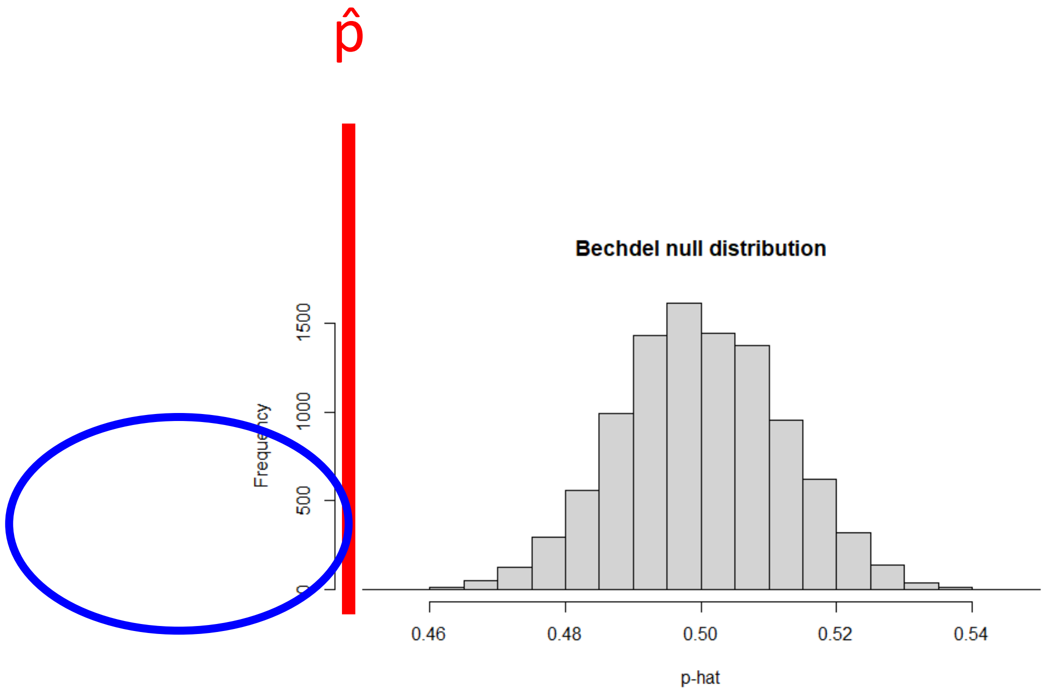
- We compare the observed test statistic to the statistic values in the null distribution
- If the observed statistic is not consistent with the null distribution, then we can reject the null hypothesis
    - E.g., H_0: $\pi = 0.5$
- And we accept the alternative hypothesis
    - E.g., H_A: $\pi < 0.5$



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 ≤ obs_stat | H_0)



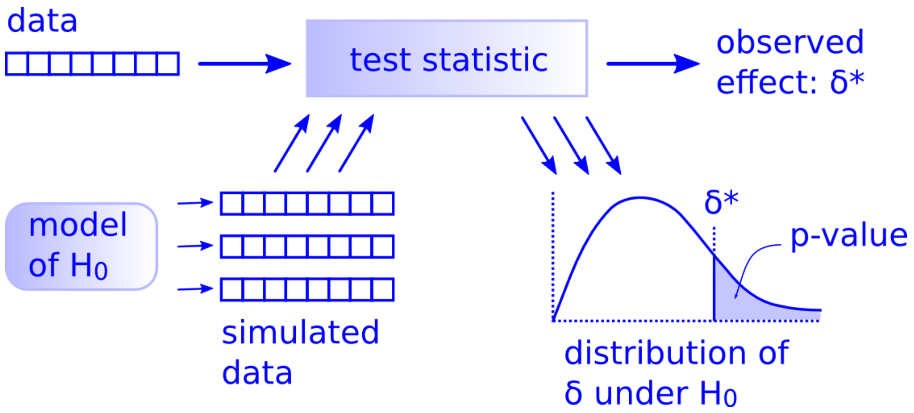
If the P-value is small, this is evidence against the null hypothesis and the results are often called "statistically significant"

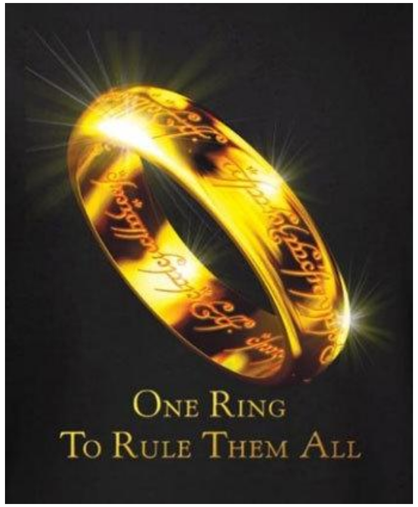
- Convention, p-value < 0.05

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





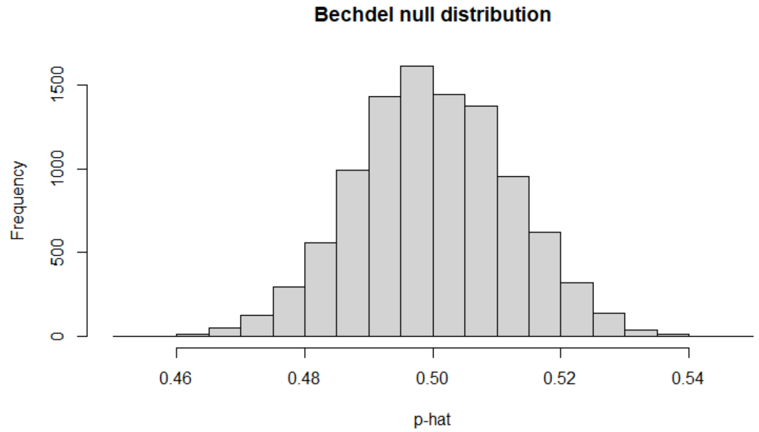
Bechdel (hypothesis) test

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 explore this in Jupyter!

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 explore this in Jupyter!

Assessing causal relationships

Causality

An association is the presence of <u>a reliable relationship</u> between the treatments an outcome

A causal relationship is when changing the value of a treatment variable <u>influences</u> 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

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?

Participant pool

Random Assignment

Treatment group

Control group

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!

Treatment group

Control group

Reconstructed participant pool data under H_0

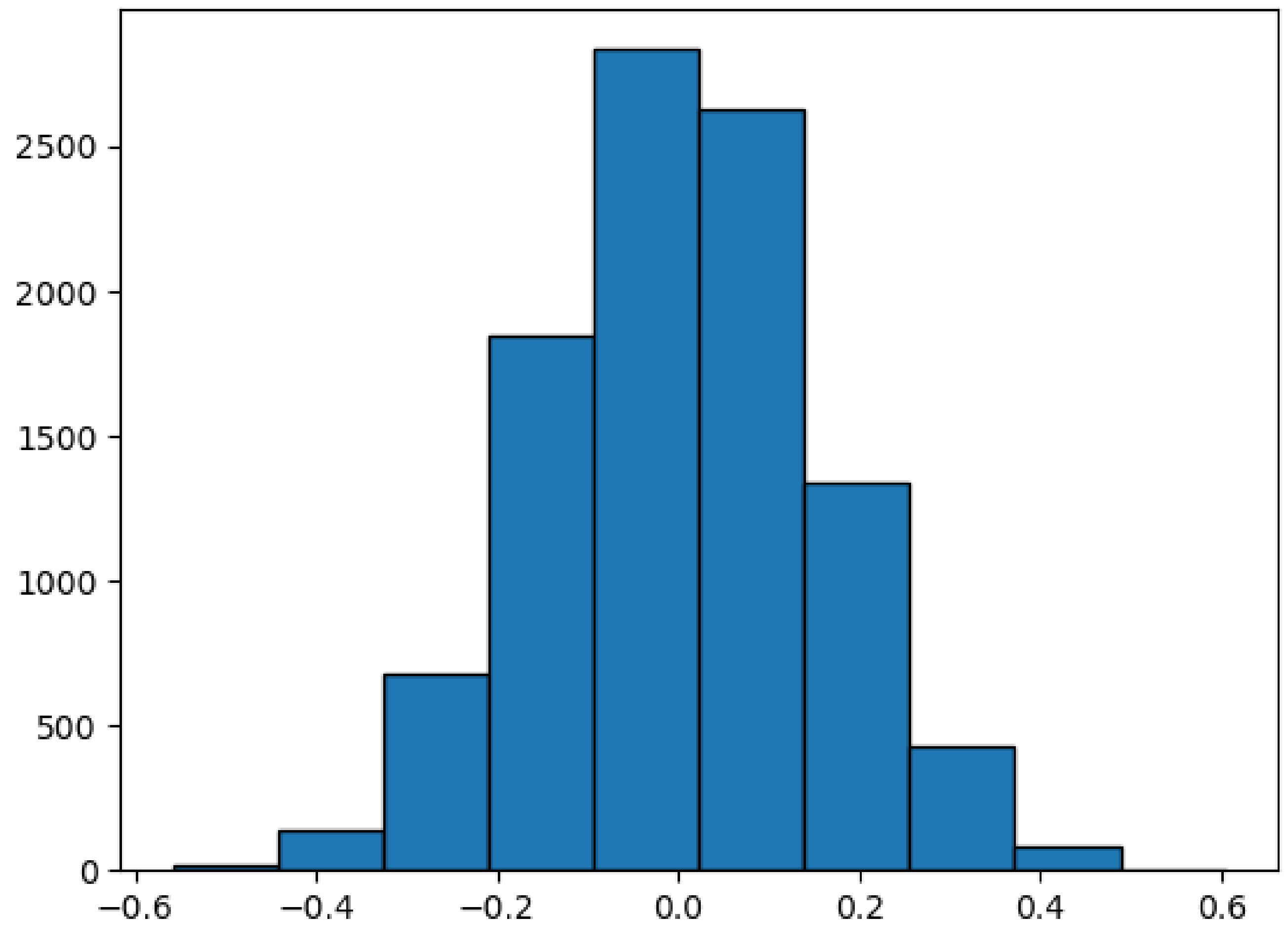
Shuffle data for random assignment consistent with H_0

Shuffled ‘treatment group’

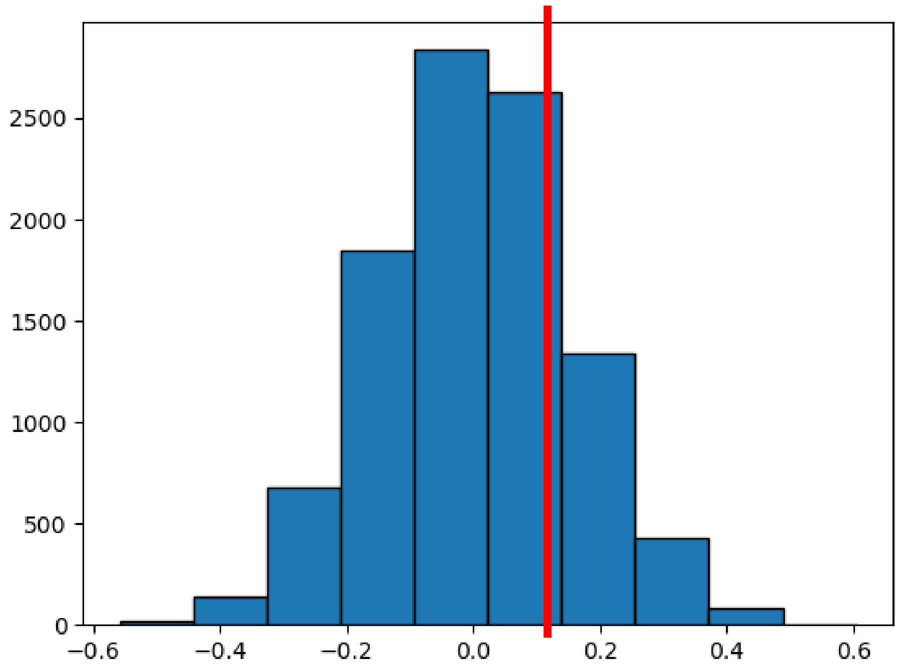
Shuffled ‘control group’

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

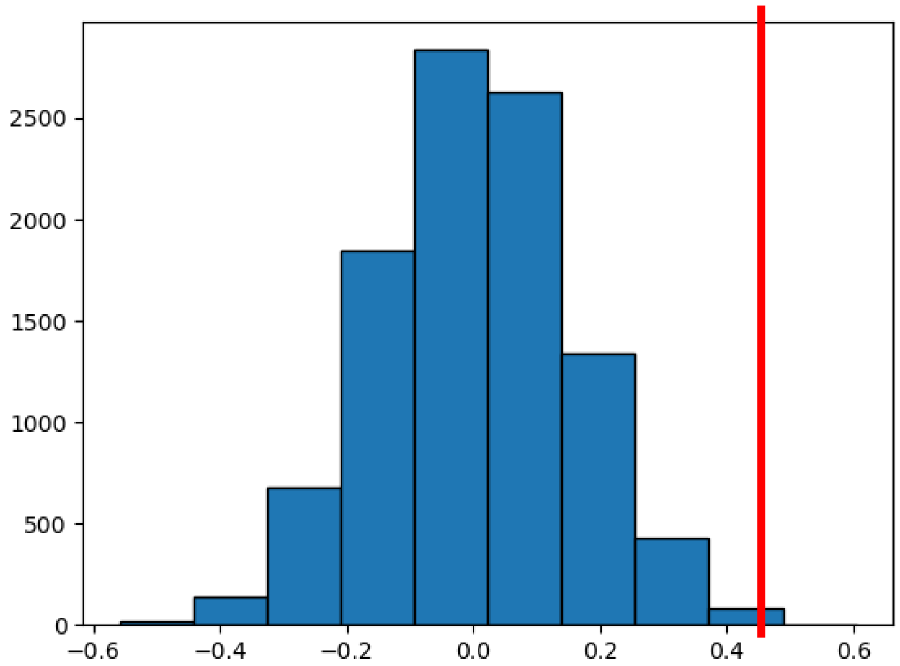
Step 3: Create a null distribution



Step 4: Calculate the p-value

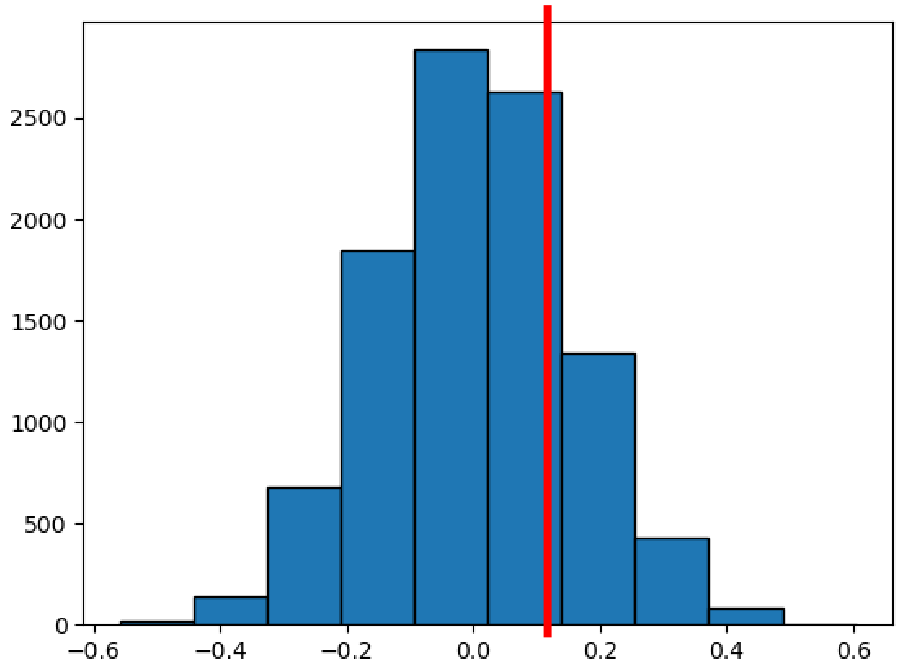


If $\hat{p}_{treat} - \hat{p}_{control} = 0.1$ what would the p-value be?

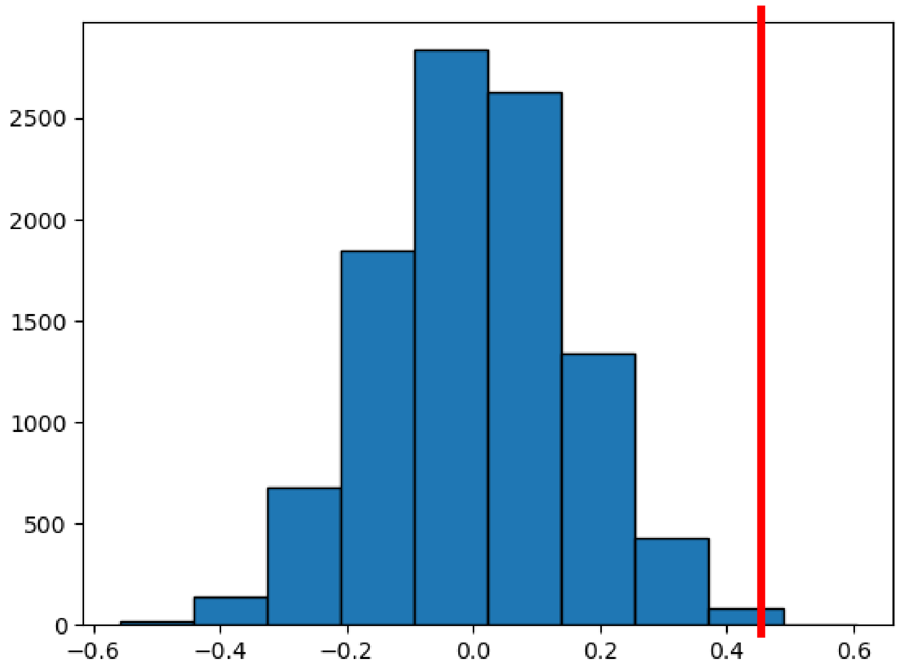


If $\hat{p}_{treat} - \hat{p}_{control} = 0.5$ what would the p-value be?

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

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

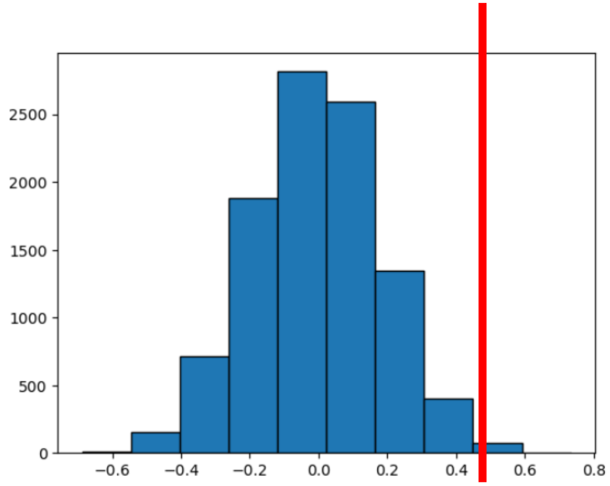
First published May 22, 2001, DOI: https://doi.org/10.1212/WNL.56.10.1290

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

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

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 baby weights

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

Large values of this statistic favor the alternative

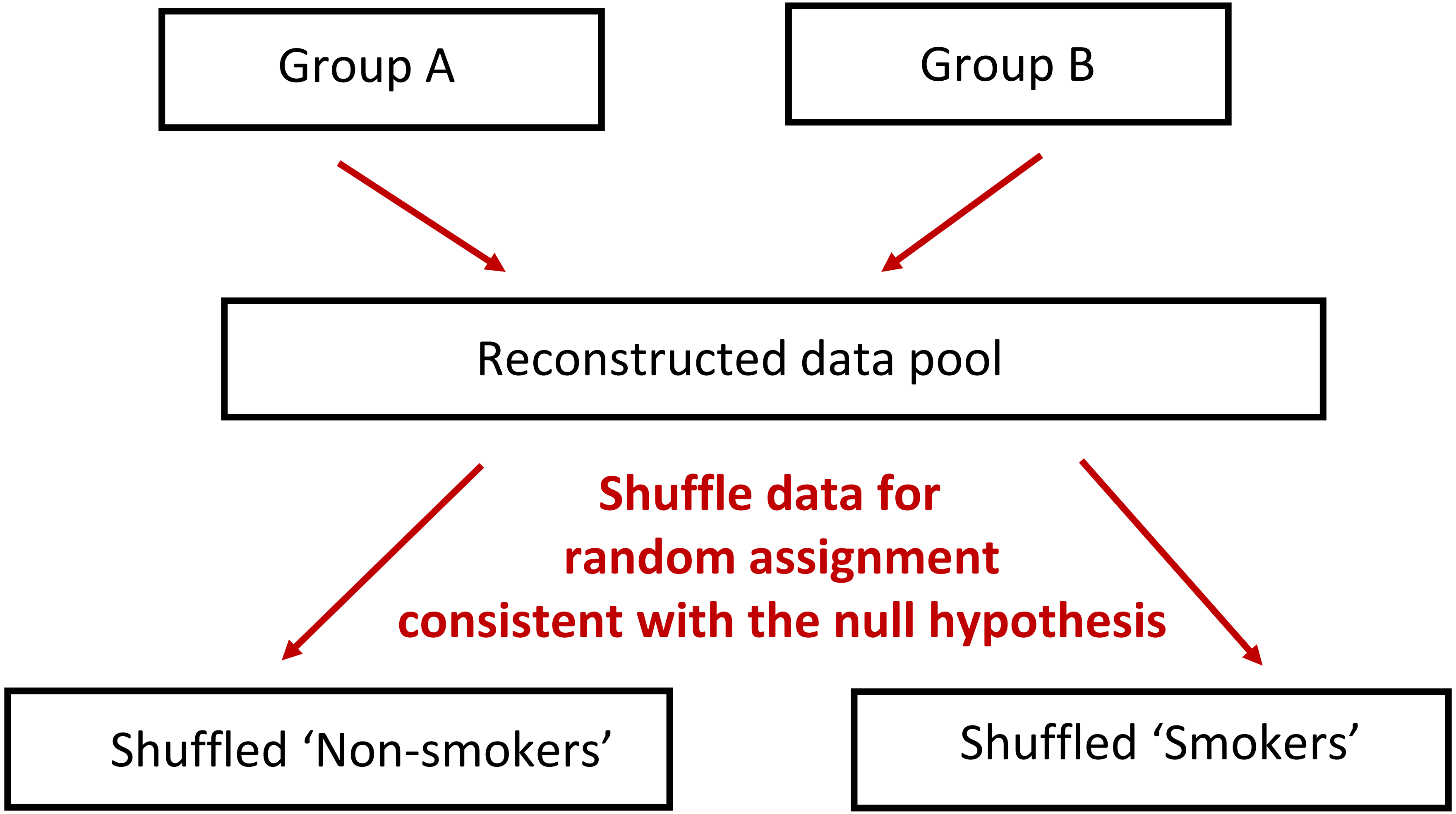
Step 3: Create the null distribution

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!



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

Let's explore this in Jupyter!