YData: Introduction to Data Science

Class 20: Confidence intervals

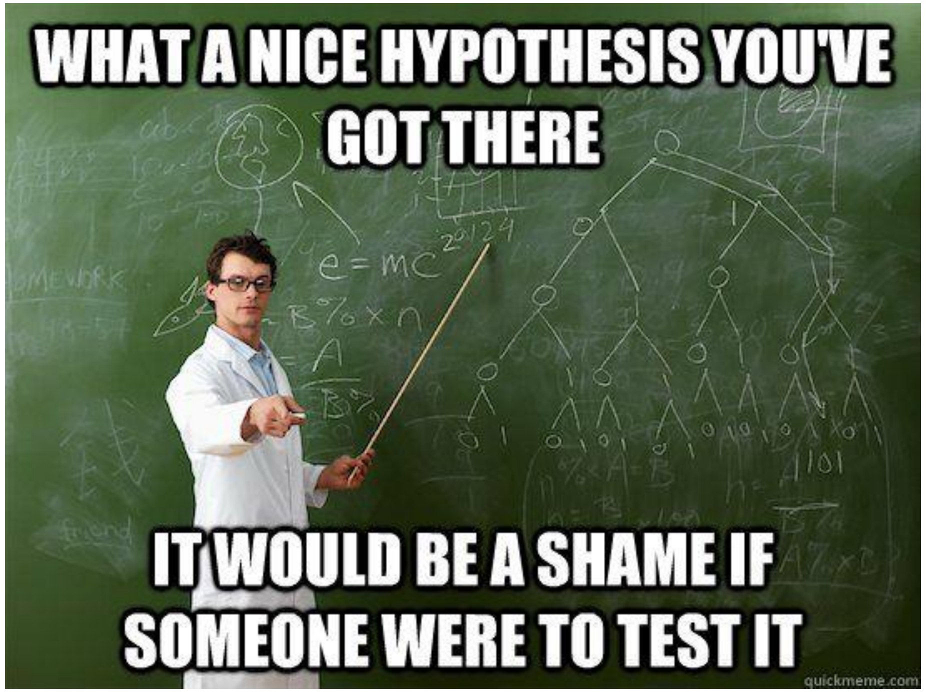
Overview

Review/continuation of hypothesis tests for correlation

Visual hypothesis tests

Two-sided hypothesis tests

Confidence intervals



Reminder: keep working on your class project

Homework 8 is due on Sunday November 10th

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

Review of Statistical Inference

Review: Statistical Inference

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

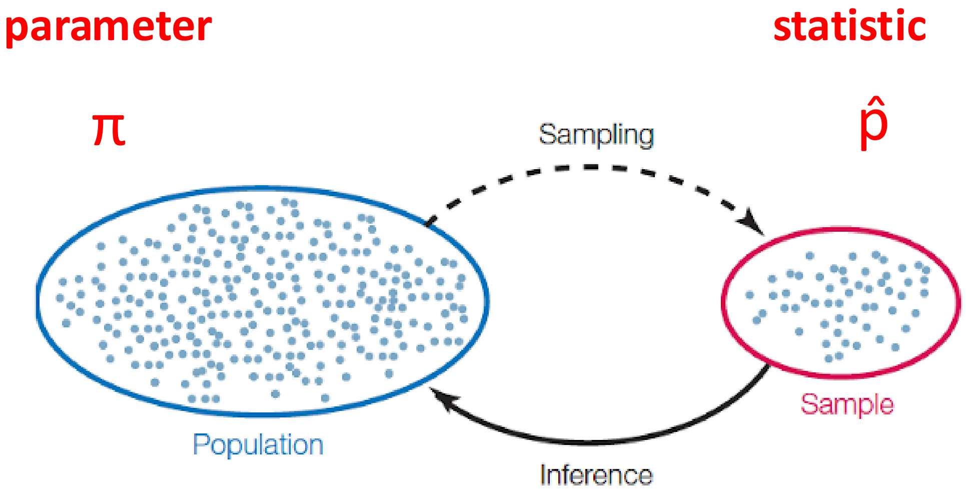
A parameter is number associated with the population

- 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



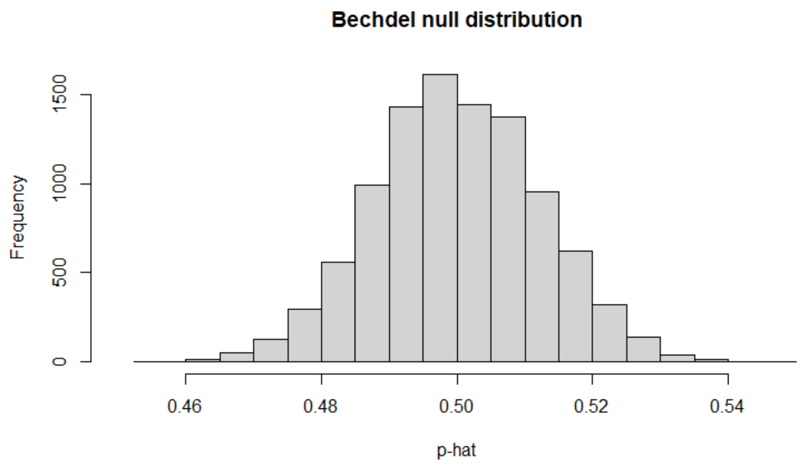
	Sample Statistic	Population Parameter
Mean	$\bar{x}$	μ
Proportion	$\hat{p}$	π
Correlation	r	ρ

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

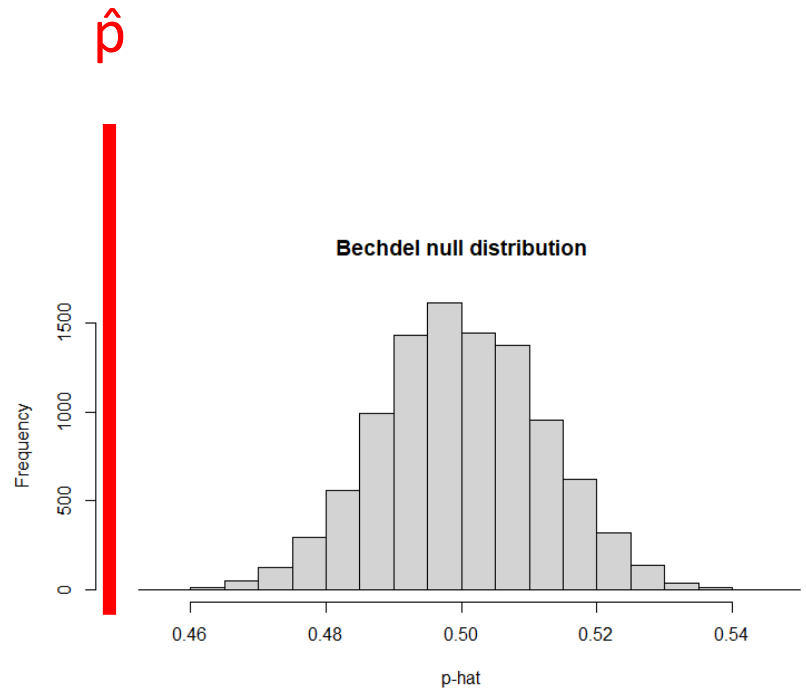
- The hypothesis we believe in (would like to see true)
- E.g., H_A: $\pi < 0.5$

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:

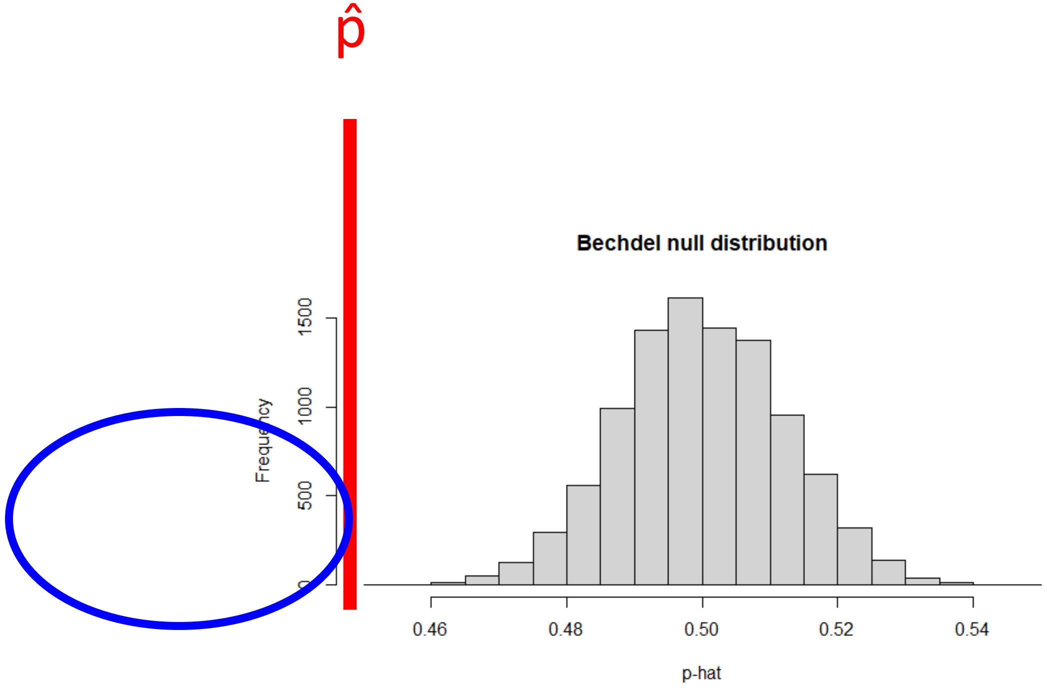
- We compare the observed test statistic to the statistic values in the null distribution
- If the observed statistic is not consistent with the null distribution, then we can reject the null hypothesis
  - E.g., H_0: $\pi = 0.5$
- And we accept the alternative hypothesis
  - E.g., H_A: $\pi < 0.5$



The p-value

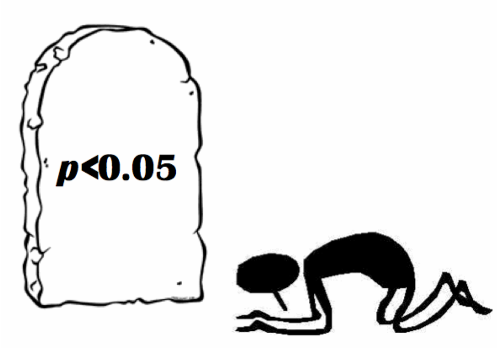
The p-value is the probability, that we get a statistic as or more extreme than the observed statistic from the null distribution

- P(Null_Stat ≤ obs_stat | H_0)



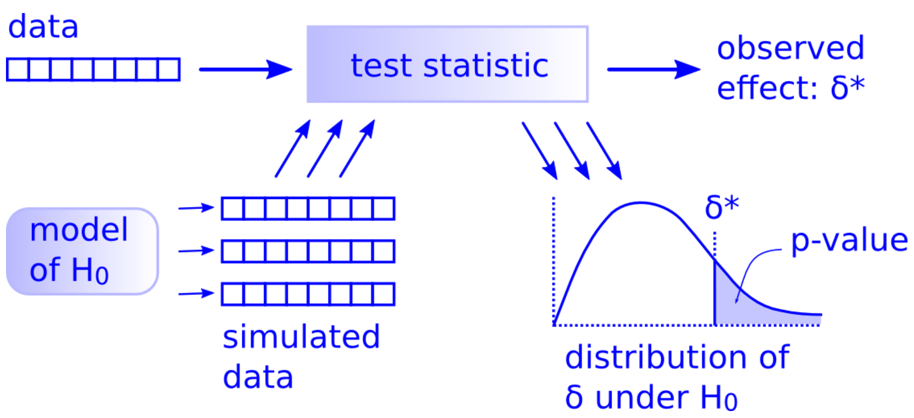
If the P-value is small, this is evidence against the null hypothesis and the results are often called "statistically significant"

- Convention, p-value < 0.05

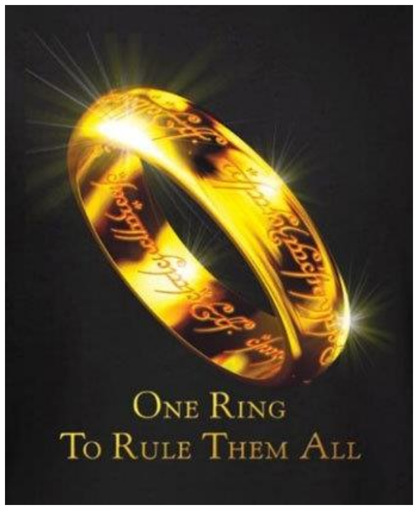


Steps needed to run a hypothesis test

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



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

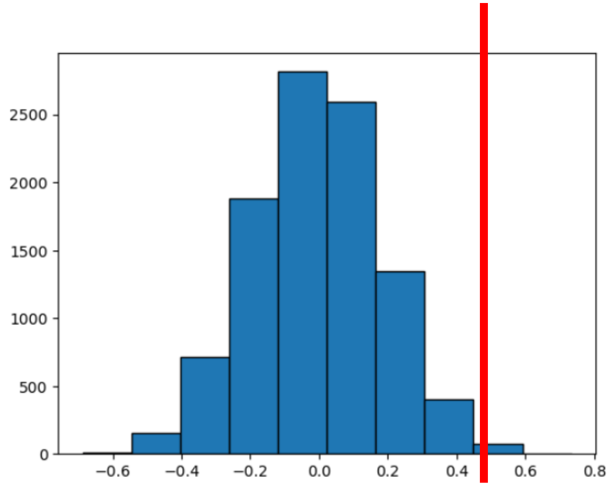
Leslie Foster, Larry Clapp, Marleigh Erickson, Bahman Jabbari

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

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

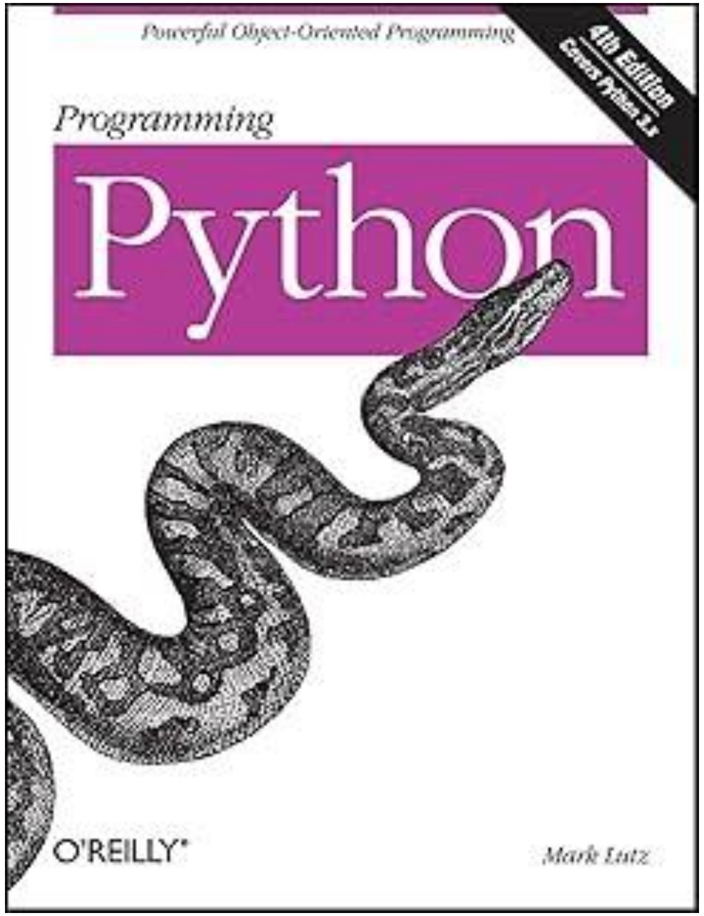
Alternative hypothesis:

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

In symbols:

H_0: $\rho = 0$

H_A: $\rho > 0$



Has 1626 pages

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?

Has 1012 pages

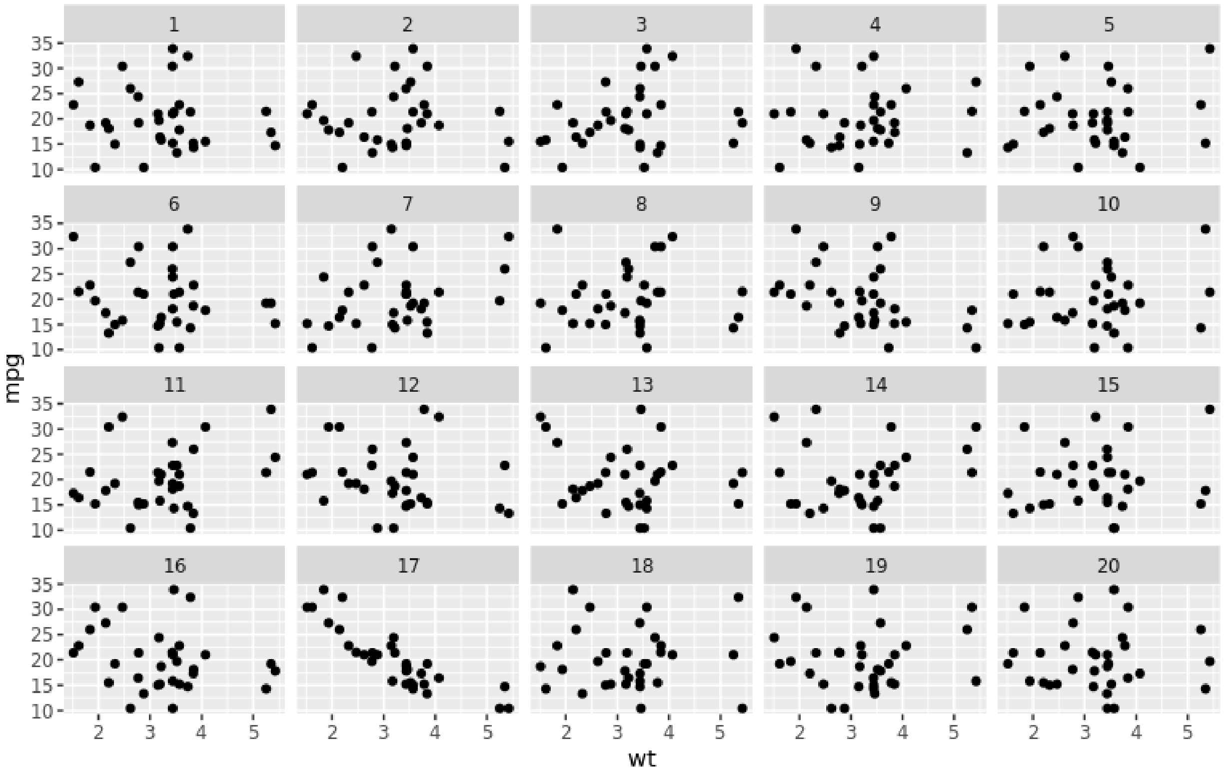
Visual hypothesis test

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

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:

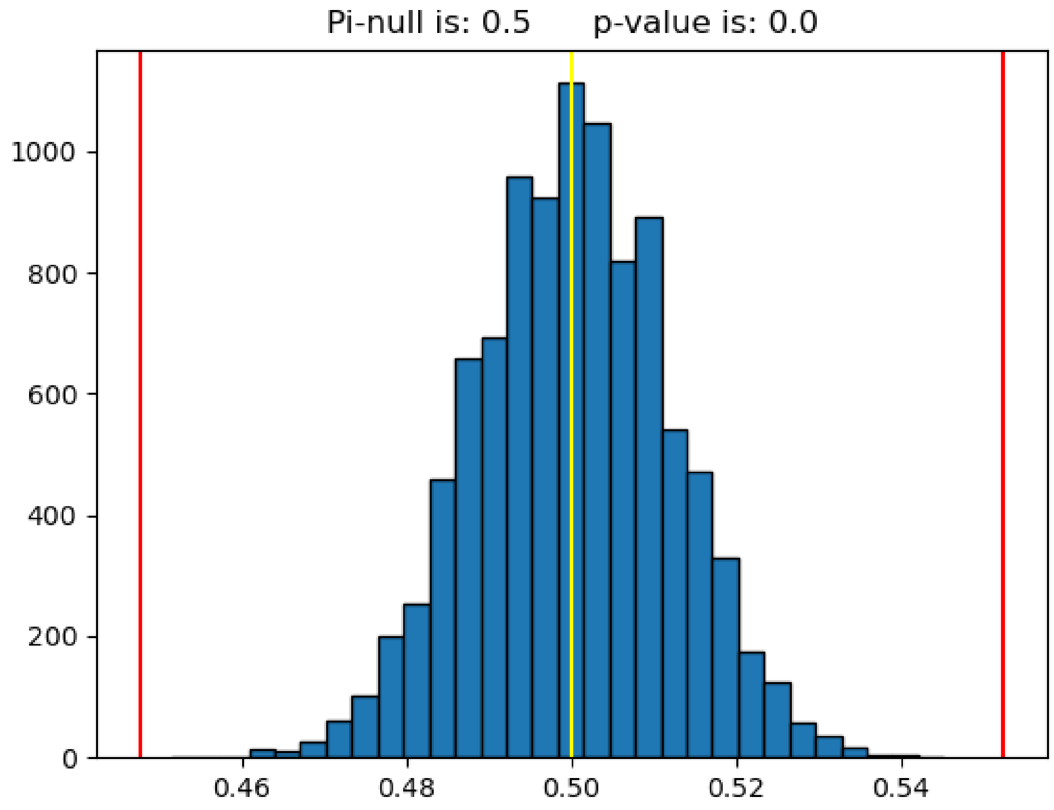
- H_A: $\pi_{Bechdel} \neq .5$
- H_A: $\mu_{non\text{-}smoke} \neq \mu_{smoke}$

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!



Confidence intervals

Interval estimate based on a margin of error

Null hypothesis tests tell us if a particular parameter value is implausible

- E.g., in the Bechdel data we rejected $\pi = .5$

An **interval estimate** give a range of plausible values for a population parameter

Example: 42% of American approve of Biden's job performance, plus or minus 3%

How do we interpret this?

Says that the population parameter π lies somewhere between 39% to 45%

- i.e., if they sampled all Americans the true population proportion would be likely be in this range

Confidence Intervals

A **confidence interval** is an interval computed by a method that will contain the *parameter* a specified percent of times

- i.e., if the estimation were repeated many times, the interval will have the parameter x% of the time

The **confidence level** is the percent of all intervals that contain the parameter

Think ring toss…

Parameter exists in the ideal world

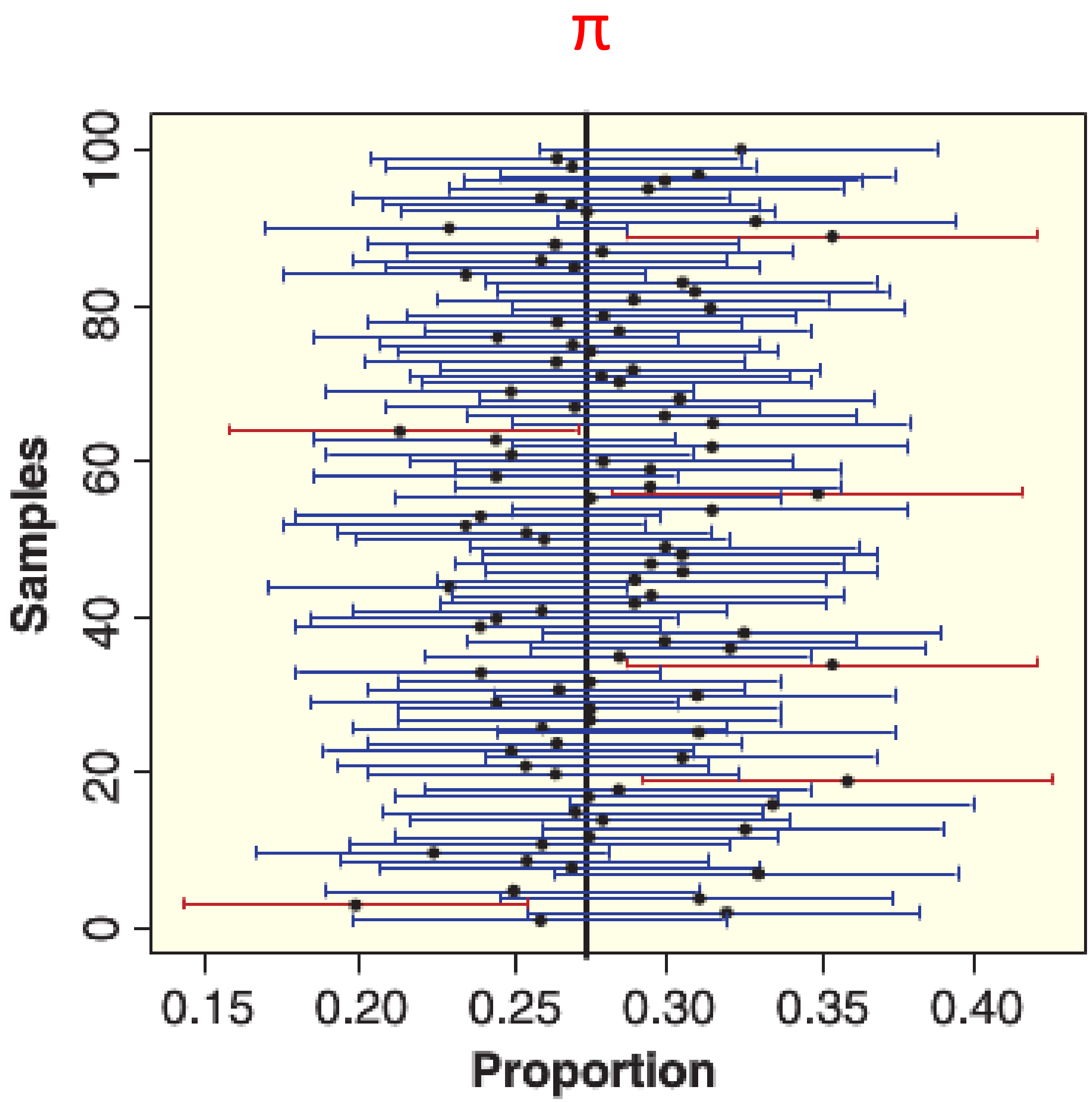
We toss intervals at it

95% of those intervals capture the parameter

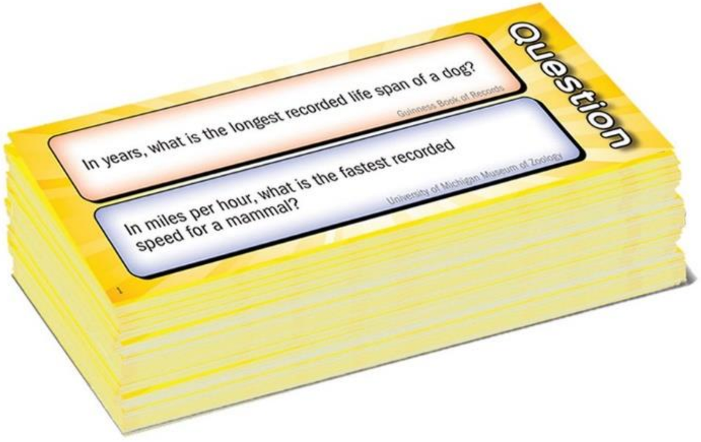
Confidence Intervals

For a **confidence level** of 95%...

95% of the **confidence intervals** will have the parameter in them



Wits and Wagers:
90% confidence interval estimator



I will ask 10 questions that have numeric answers

Please come up with a range of values that contains the true value in it for 9 out of the 10 questions

- i.e., be a 90% confidence interval estimator

Wits and Wagers…

Question 1: What is the diameter of the moon (in miles)?

Question 2: How many years passed between the first NBA game and the first WNBA game?

Question 3: What percent of U.S. land area does Alaska make up?

Wits and Wagers…

Question 4: On average, how many baseballs are used in a Major League Baseball season?

Question 5: How many rooms are there in the White House?

Question 6: How many votes were cast in the 2012 U.S. presidential election?

Question 7: Out of the 538 electoral votes, how many did Ronald Reagan receive in the 1984 presidential election ?

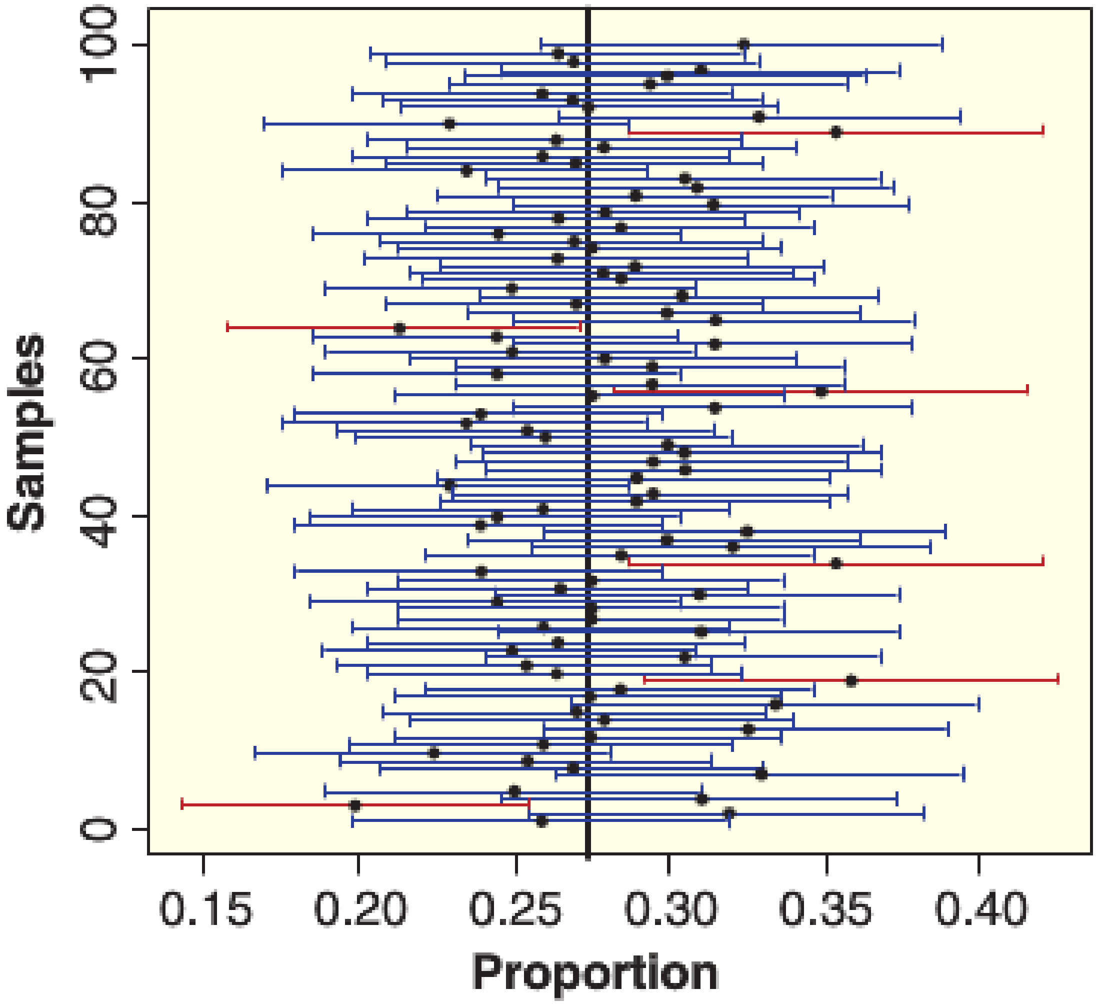
Wits and Wagers…

Question 8: How many cases of human spontaneous combustion appeared in medical journals between the years of 1600 and 1900?

Question 9: How many Academy Award nominations did *The Lord of the Rings* movie trilogy receive?

Question 10: In feet, how long was the largest whale ever recorded?

π



We all have 9 out of 10 correct?!

Note

For any given confidence interval we compute, we don't know whether it has really captured the parameter

But we do know that if we do this 100 times, 90 of these intervals will have the parameter in it

(for a 90% confidence interval)

Tradeoff between interval size and confidence level

There is a tradeoff between the **confidence level** (percent of times we capture the parameter) and the **confidence interval size**

Using hypothesis tests to construct confidence intervals

Constructing confidence intervals

There are several methods that can be used to construct confidence intervals including

- “Parametric methods” that use probability functions
  - E.g., confidence intervals based on the normal distribution
- A “bootstrap method” where data is resampled from our original sample to approximate a sampling distribution

To learn more about these methods, take Introductory Statistics!

Constructing confidence intervals

We are going to use a less conventional method to get confidence intervals based on the relationship between confidence intervals and hypothesis tests

- The method we will discuss is valid, but can be more computationally expensive than other methods

What we will do is to run a series of hypothesis test with different null hypothesis parameter values

Our confidence interval will be all parameters values where we **fail to reject** the null hypothesis

$H_0: \pi = \pi_0$

Failure to reject $\pi = \pi_0$ means π_0 is plausible

Motivation: Bechdel Confidence Interval

From running a hypothesis test on the Bechdel data, we saw that H_0: $\pi = .5$ is unlikely

- i.e., it was not plausible that 50% of movies pass the Bechdel test

But what is a reasonable range of values for the population proportion of movies that pass the Bechdel test?

We can create a confidence interval for $\pi_{Bechdel}$ to find out!

Very quick review of using hypothesis tests to construct confidence intervals

All parameter values where we fail to reject the null hypothesis make up the confidence interval

- Using a threshold of p-value < 0.05 yields a 95% CI
- Using a threshold of p-value < 0.01 yields a 99% CI

π	**p-values**
~~0.4~~	0
~~0.41~~	0.0013
~~0.42~~	0.0179
0.43	0.1361
0.44	0.5269
0.45	0.85
0.46	0.296
0.47	0.0614
~~0.48~~	0.0067
~~0.49~~	0.0004

Let's explore this in Jupyter!