Introduction to hypothesis tests

Overview

Last quick review of confidence intervals

A brief note on probability

Introduction to hypothesis tests

If there is time: additional examples

Homework 5

Homework 5 has been posted

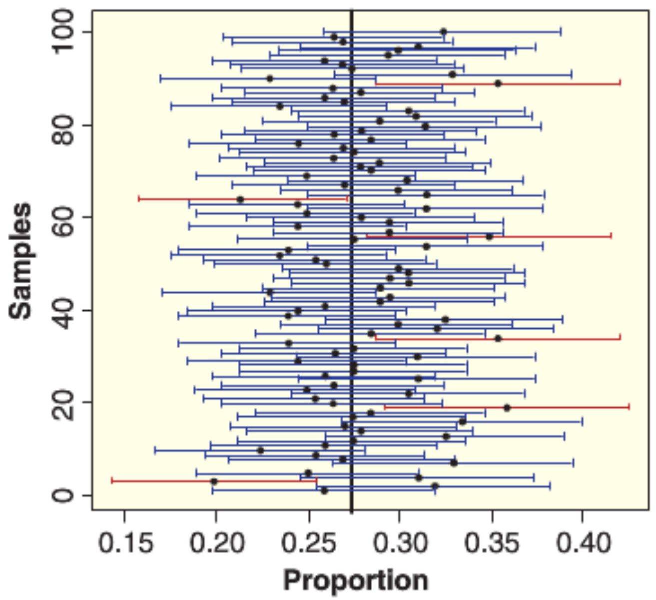
- Due on Gradescope at 11pm on Sunday February 25th

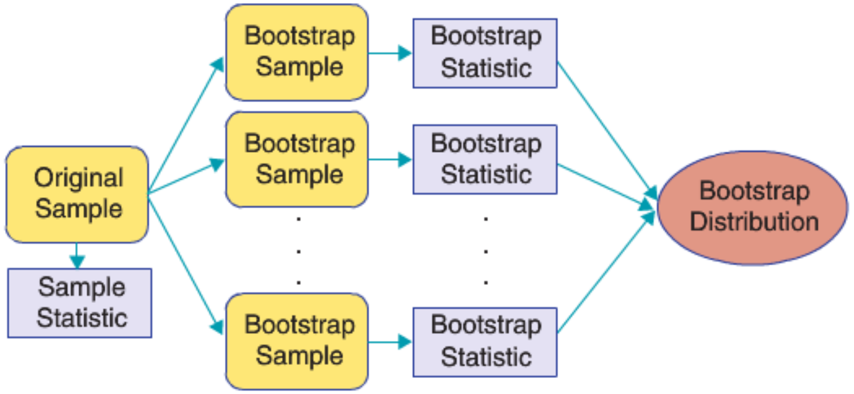
Any questions about homework 4?

Last review of bootstrap confidence intervals

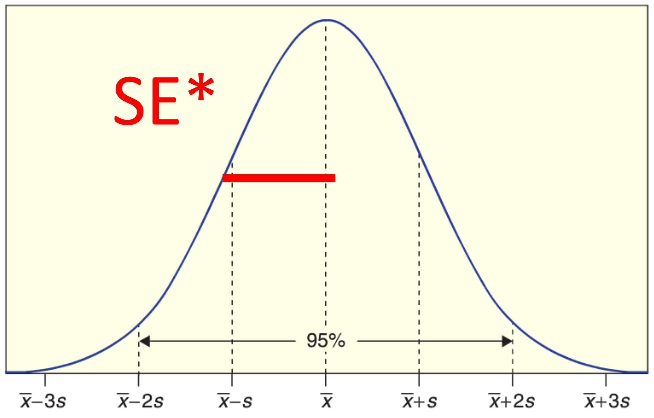
Very quick review of bootstrap confidence intervals

Confidence intervals give range of plausible values for a parameter





Bootstrap distribution



95% confidence interval: stat $\pm\ 2 \cdot \hat{SE}$

Are we feeling confident about confidence intervals?

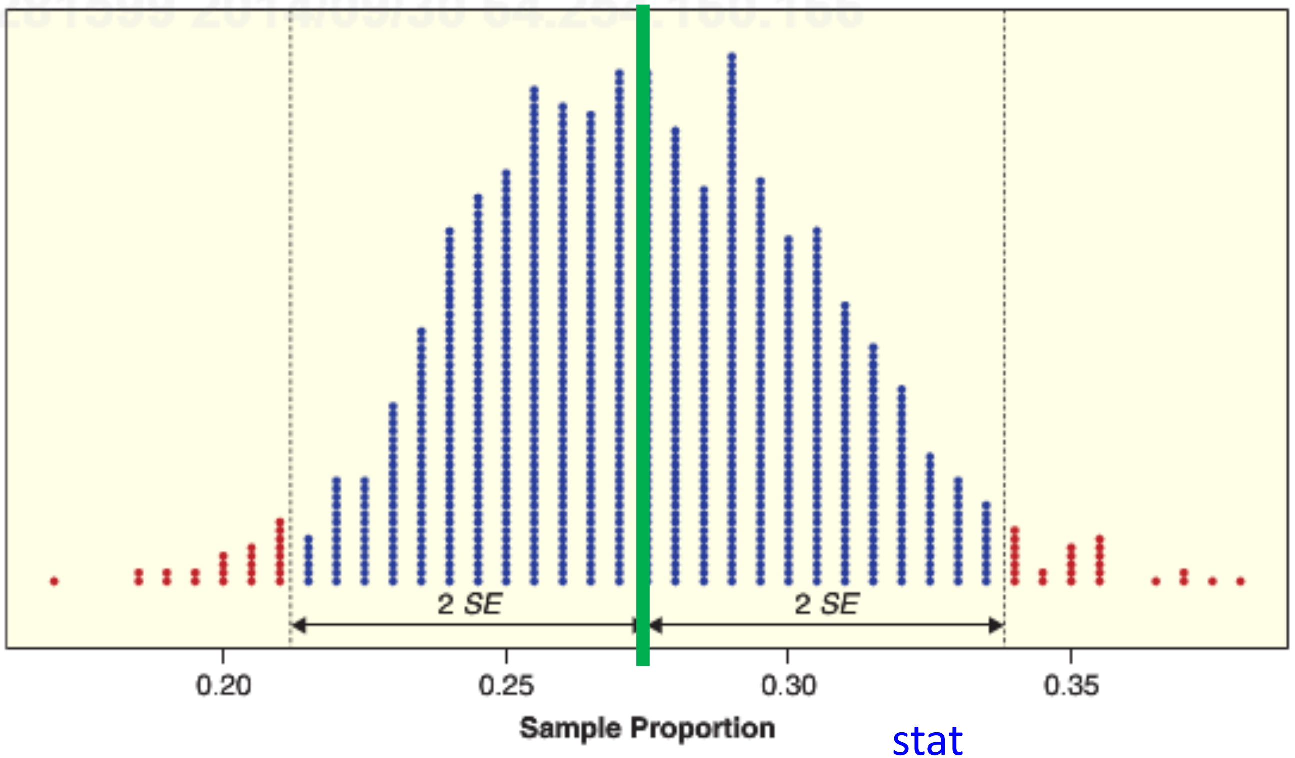
95% confidence interval: stat ± 2 · $\hat{SE}$

Confidence interval is centered at the statistic value

95% of the statistics lie within ± 2 SE of the parameter

- Therefore, the interval captures the parameter 95% of the time

π



stat

2 · SE 2 · SE

Confidence interval

Bootstrap distribution in R for μ

```
my_sample <- c(21, 29, 25, 19, 24, 22, 25, 26, 25, 29)     # n = 10 points here
my_stat <- mean(my_sample)     #  x-bar

bootstrap_dist <-  do_it(10000) * {

        curr_boot <- sample(my_sample , 10, replace = TRUE)
        mean(curr_boot)

}

SE_boot   <-  sd(bootstrap_dist)
CI     <-     c(my_stat    -   2  * SE_boot,        my_stat    +   2  * SE_boot)
```

Any questions about confidence intervals or the bootstrap?

Let's do one last practice example of computing a bootstrap confidence interval for a mean in R

Let's create a confidence interval for the mean gross revenue made by movies

- (in 2013 dollars)

SDS100::download_class_code(11)

Introduction to hypothesis tests

A quick note on probability

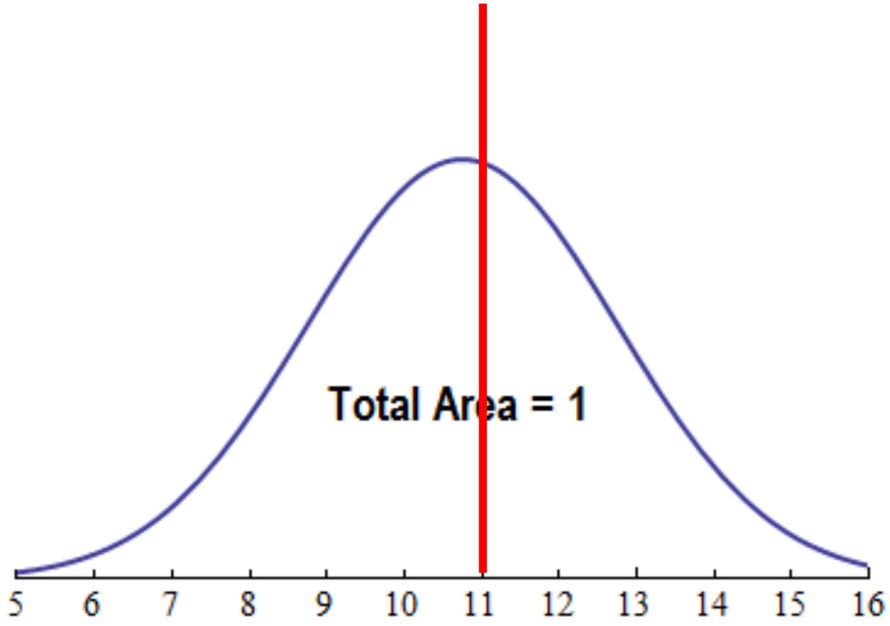
Probability is a way of measuring the likelihood that an event will occur.

Probability models assigns a number between 0 and 1 to the outcome of an event (outcome) occurring.

We can use a probability model to calculate the probability of an event.

For example:

- $P(X < 11) = 0.55$
- $P(X > 20) = 0$



Statistical tests (hypothesis test)

A **statistical test** uses data from a sample to assess a claim about a population.

<u>Example 1</u>: The average body temperature of humans is 98.6˚

How can we write this using symbols?

- $\mu = 98.6$

Statistical tests (hypothesis test)

A **statistical test** uses data from a sample to assess a claim about a population.

<u>Example 2</u>: we might make the claim that the correlation between height and weight is 0.6.

How can we write this using symbols?

- $\rho = 0.6$

Statistical tests (hypothesis test)

A **statistical test** uses data from a sample to assess a claim about a population.

Example 3: [Trump has a lead over Biden in the election]

How can we write this using symbols?

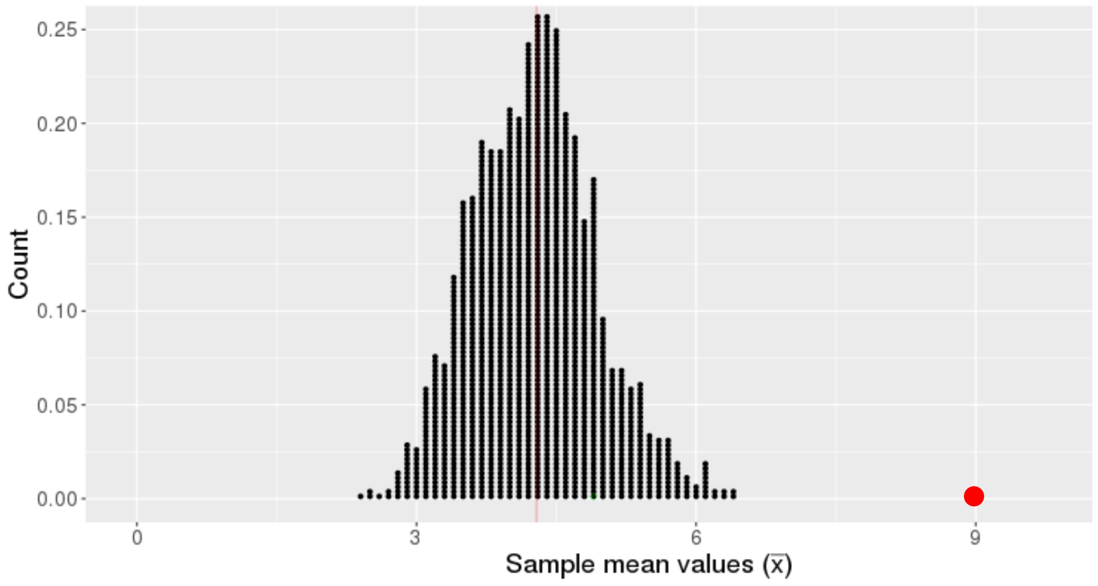
- $\pi_{Trump} > \pi_{Biden}$ or $\pi_{Trump} - \pi_{Biden} > 0$

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

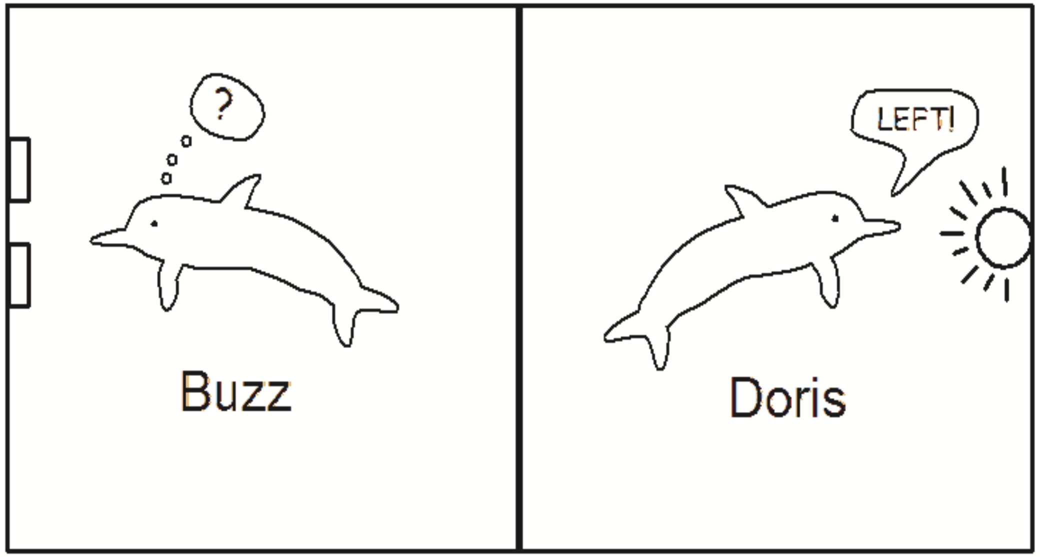
How do we feel about the abstract communication study?

Stead light = push button on right to get food

Flashing = push button on the left to get food

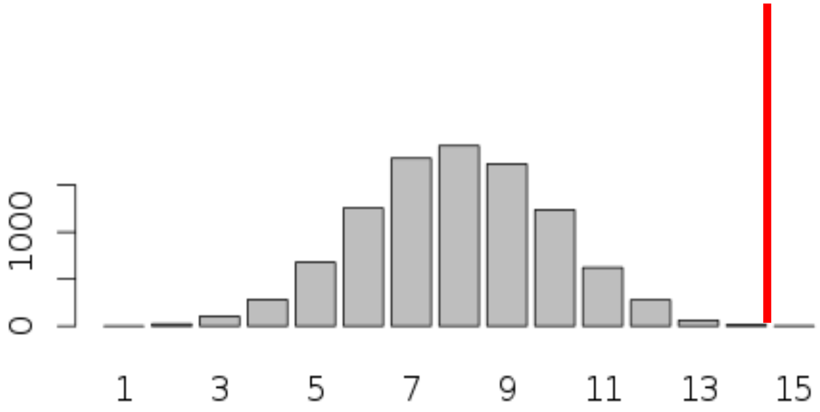
If Buzz was guessing, what % correct would he get?

- 50%
- H_0: $\pi = 0.5$



Right

Left



Buzz got 15 out of 16 trials correct!

- Is Buzz guessing?

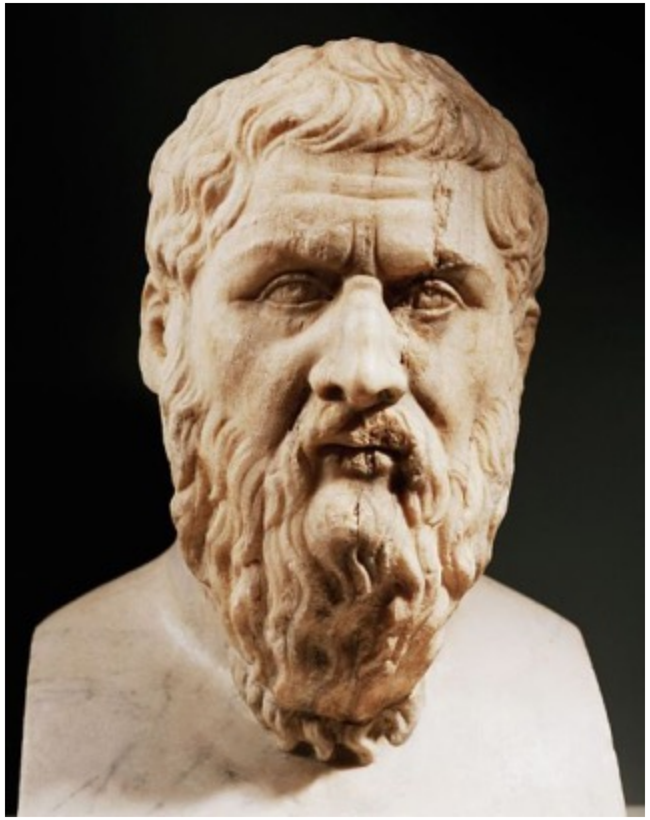
Q: Do fewer than 50% of movies pass the Bechdel test?

If we are being pessimistic, what would we believe?

- We believe fewer than 50% pass the Bechdel test

What does our nihilistic frienemy Gorgias believe?

- He believes (at least) 50% of movies pass the Bechdel test.



Do more 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

As you will recall, there were 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: $\hat{p} = 803/1794 = 0.448$

Chance models

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 $\pi = 0.5$?

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

Heads = passed the Bechdel test

Tails = failed to pass the Bechdel test

Let's flip 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?

Flipping coins using SDS100 functions

rflip_count() returns the number of heads out of num_flip coin flips:

```
rflip_count(num_flips, prob = .5)
```

num_flips: the number of times to flip a coin

- 1794 for Bechdel test

prob: the probability of success on each trial

- .5 if half the movies passed the Bechdel test

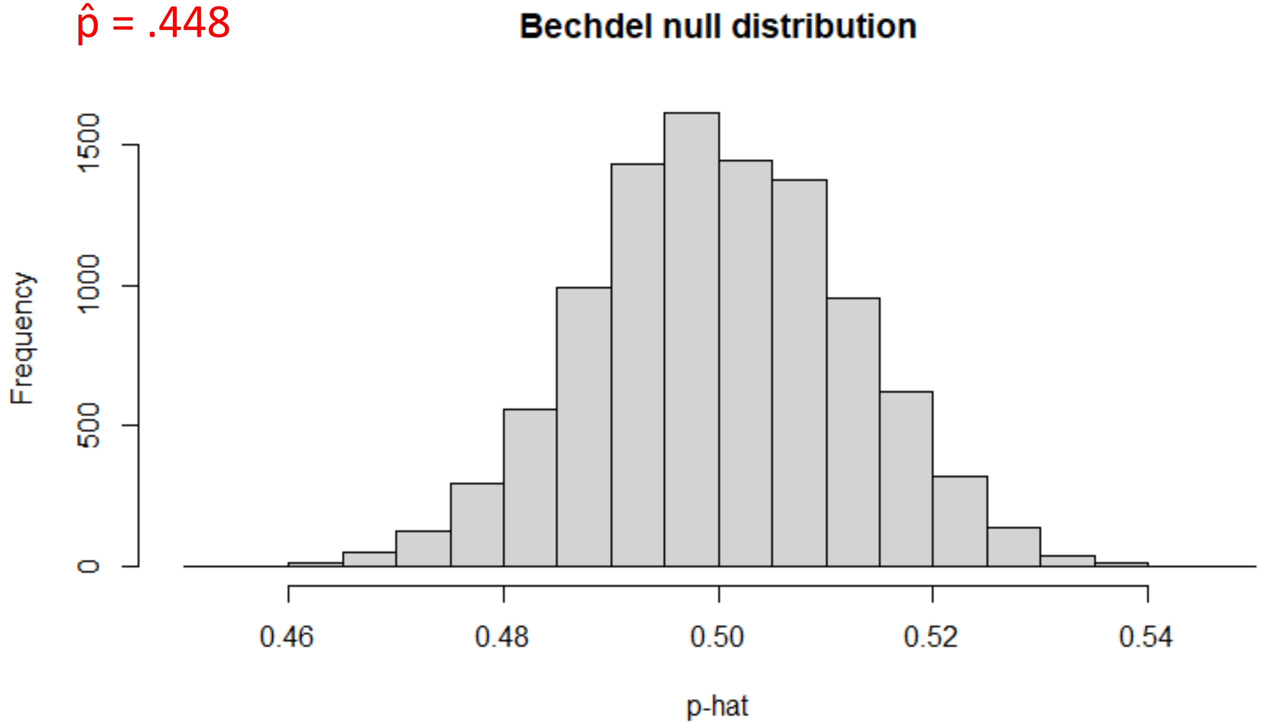
We can repeat flips many times using the do_it() function:

```
library(SDS100)

flip_simulations <- do_it(10000) * {
    rflip_count(1794, prob = .5)
}
```

Simulating Flipping 1794 coins 10,000 times

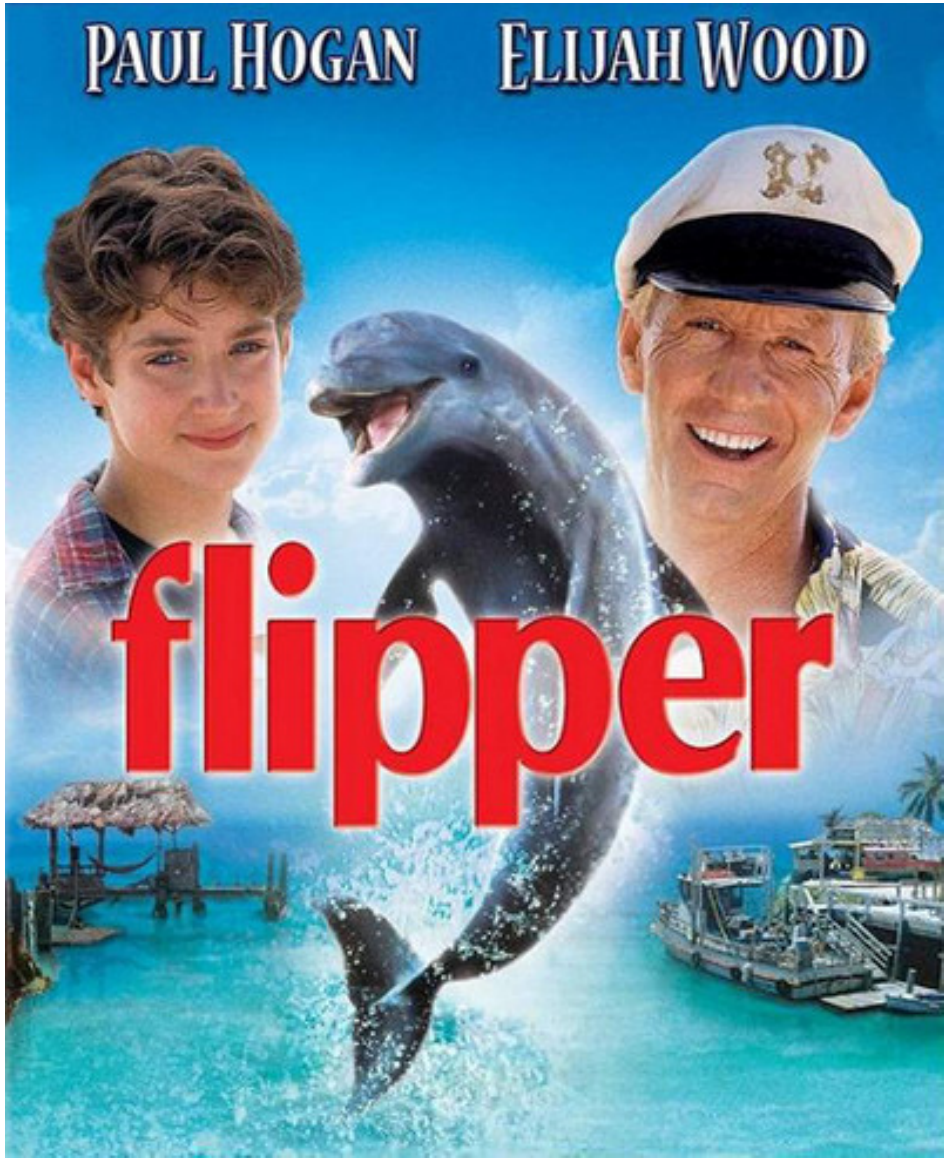
$\hat{p}$ = .448



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

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

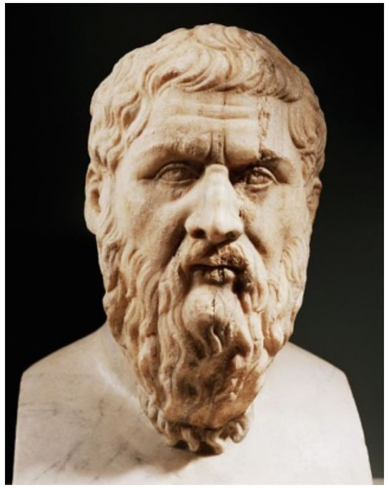
Q: What can be conclude?


PAUL HOGAN
ELIJAH WOOD
flipper

Let's try it in R...

Hypothesis tests: central ideas and terminology

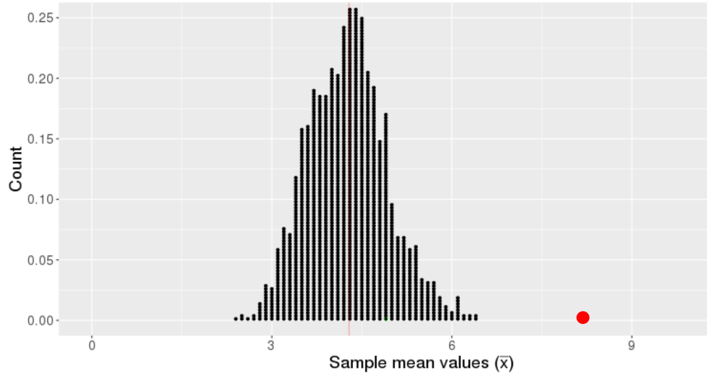
Terminology



Null Hypothesis ( ): Claim that there is no effect or no difference

Alternative Hypothesis (H_A): Claim for which we seek significant evidence

The alternative hypothesis is established <u>by observing evidence that inconsistent with the null hypothesis</u>



Review: the null hypothesis in the Bechdel test?

1. What is the null hypothesis in words?

2. We can write this in terms of the population parameter:

   H_0: $\pi = 0.5$

3. What is the alternative hypothesis?

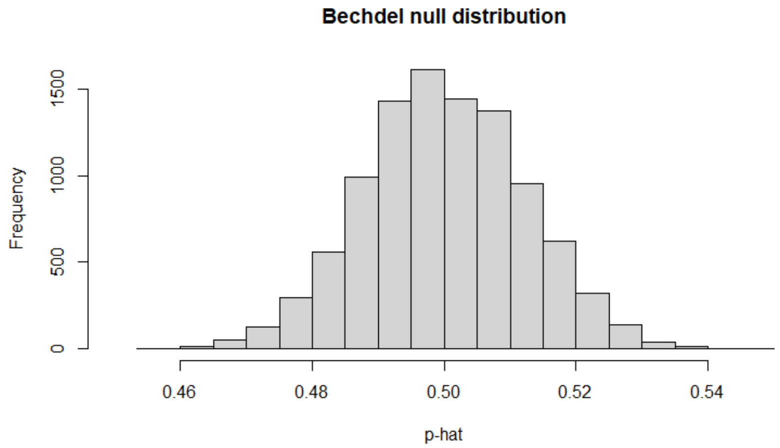
   H_A: $\pi < 0.5$

Null Distribution

A **null distribution** is the distribution of statistics one would expect <u>if the null hypothesis (H_0) was true</u>

i.e., the null distribution is the statistics one would expect to get if nothing interesting was happening

- Note: the Lock5 textbook calls this the "randomization distribution"



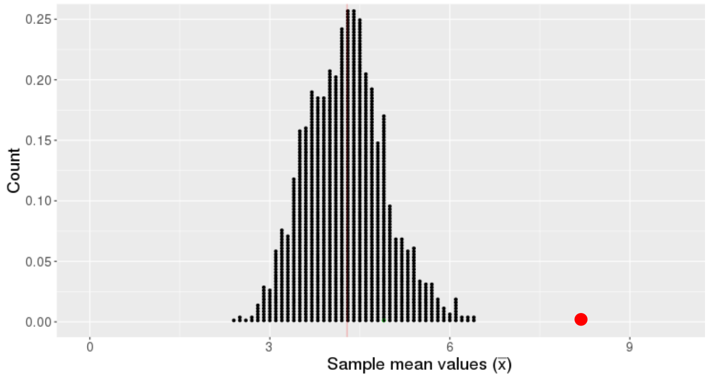
P-values

A **p-value** is the probability, of obtaining a statistic as as (or more) extreme than the observed sample *if the null hypothesis was true*

- i.e., the probability that we would get a statistic as or more extreme as our <u>observed statistic</u> from the <u>null distribution</u>

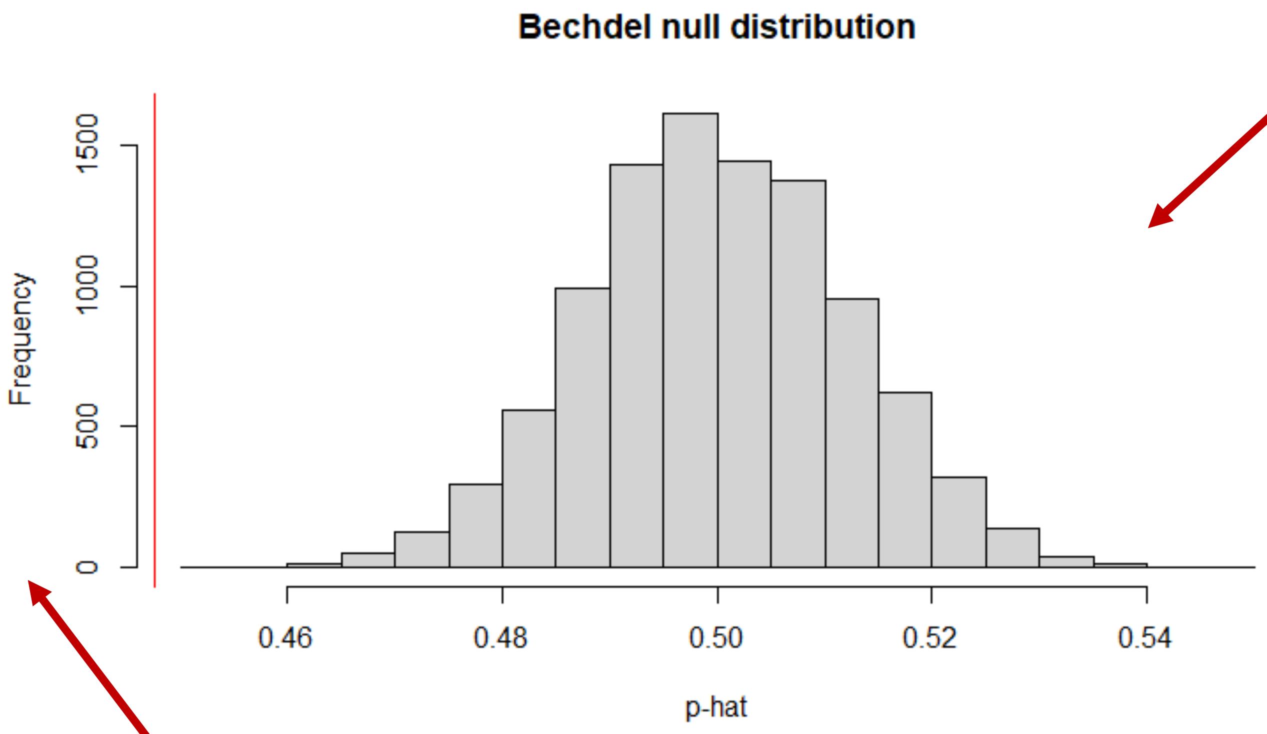
$$P(\text{STAT} \leq \text{ observed statistic } \mid H_0 = \text{True})$$

The smaller the p-value, the stronger the statistic evidence is against the null hypothesis



Bechdel test example

Null distribution



p-value = 0/10000 = 0.0000

Statistical significance

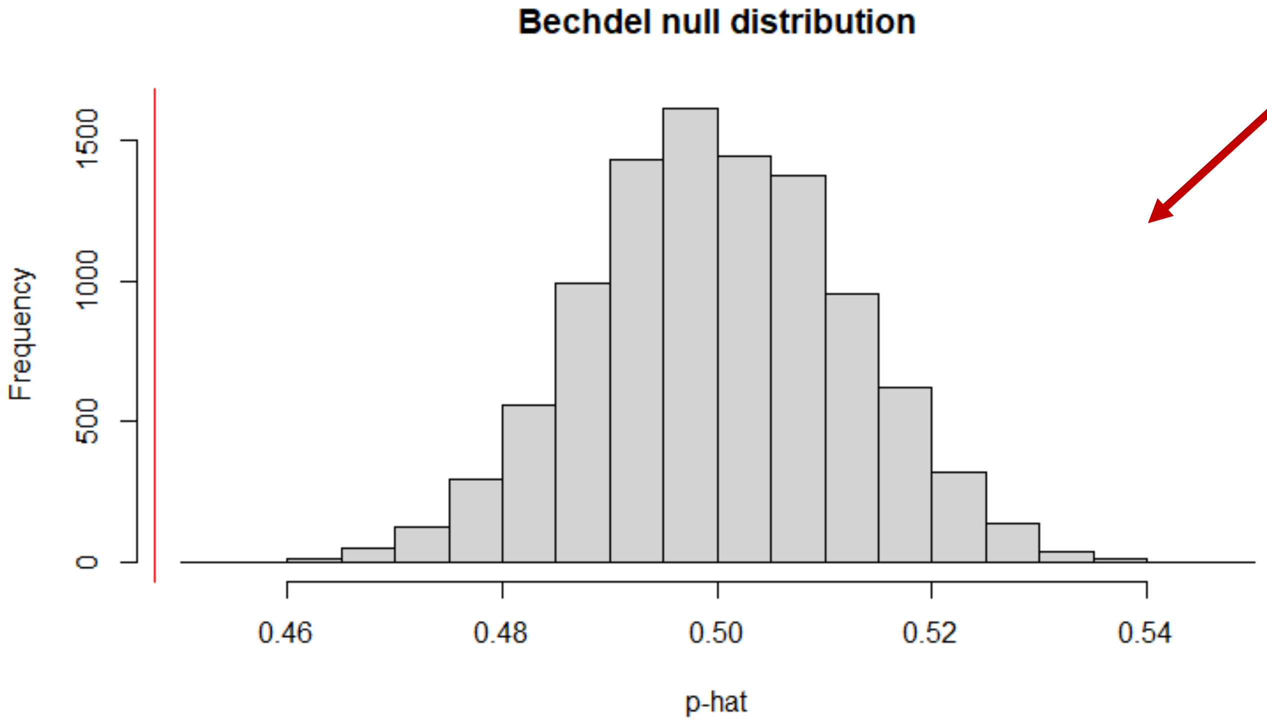
When our observed sample statistic is unlikely to come from the null distribution, we say the sample results are **statistically significant**

- i.e., we have a small p-value

‘Statistically significant’ results mean we have strong evidence against H_0 in favor of H_A

Bechdel test example

Null distribution



p-value = 0/10000 = 0.0000

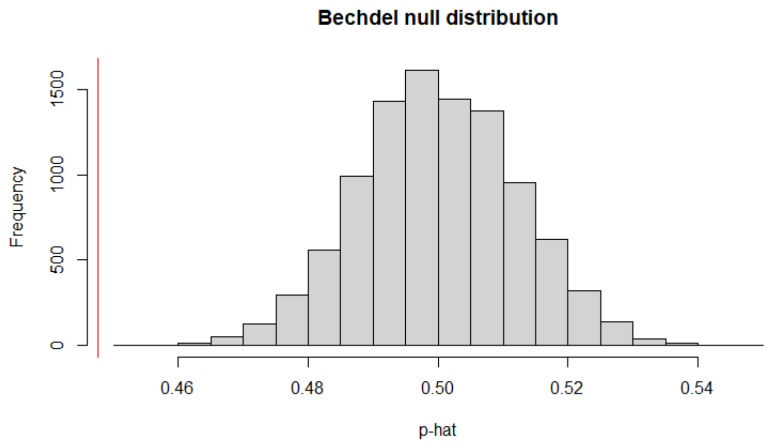
Getting p-values using SDS100 functions

Flipping coins many times:

```
flip_simulations  <-  do_it(10000) * {

     rflip_count(1794,  prob = .5)

}
```



We can get the number of values as or more extreme than an observed statistic (obs_stat) using the pnull() function:

```
obs_stat <-  803
p_value  <- pnull(obs_stat,  flip_simulations,  lower.tail = TRUE)
```

Key steps hypothesis testing

1. State the null hypothesis… and the alternative hypothesis
   - 50% of movie pass the Bechdel test: H_0: $\pi = 0.5$
   - Less than 50% of movies pass the Bechdel test: H_A: $\pi < 0.5$

2. Calculate the observed statistic
   - 803 out of 1793 movies passed the Bechdel test, or $\hat{p} = 0.448$

3. Create a null distribution that is consistent with the null hypothesis
   - i.e., what statistics would we expect if 50% of movies pass the Bechdel test

4. Examine how likely the observed statistic is to come from the null distribution
   - What is the probability that the 803 movies or fewer movies would pass the Bechdel test if the null hypothesis was true (i.e., if 50% of all movies passed the Bechdel test)?
   - i.e., what is the p-value

5. Make a judgement
   - If we have a small p-value, this means that $\pi = .5$ is unlikely and so $\pi < .5$
   - i.e., we say our results are 'statistically significant'