Hypothesis tests for a single proportion continued…

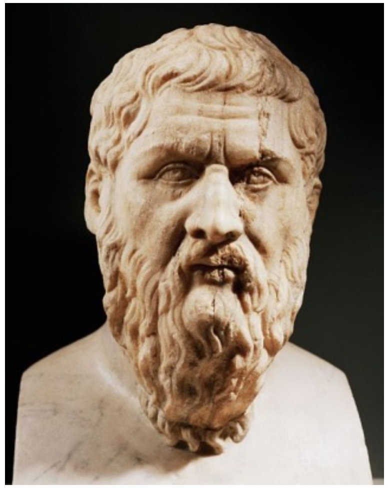
Overview

Review and more practice of hypothesis tests for a single proportion

One-sided vs. two-sided tests

Hypothesis tests: central ideas and terminology

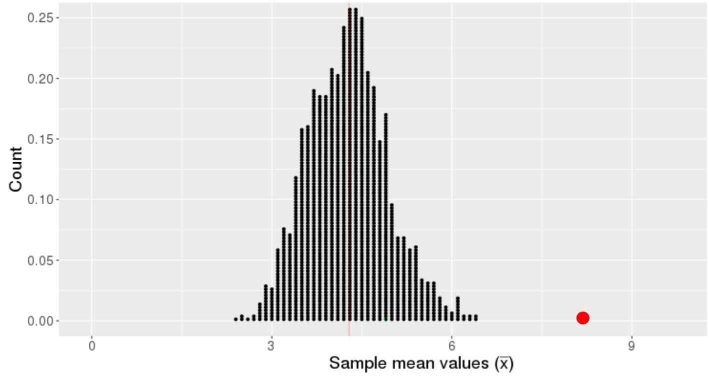
Terminology



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?

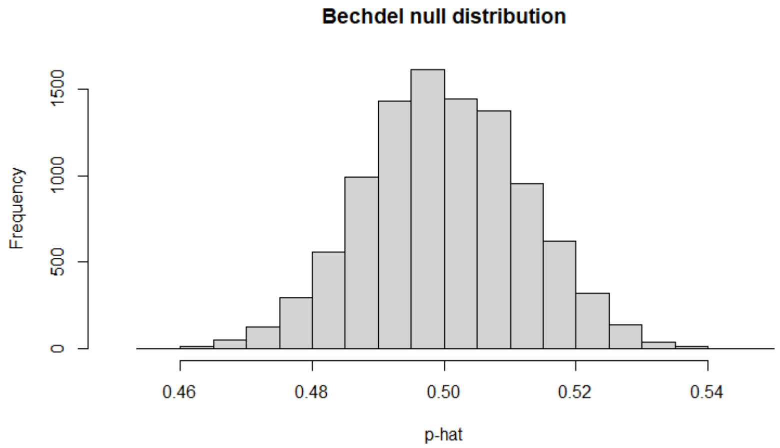
H_A: $\pi < 0.5$

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"



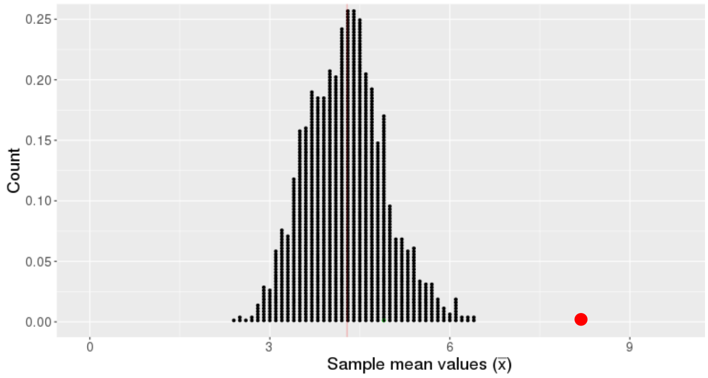
P-values

A **p-value** is the probability, of obtaining a statistic as (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>

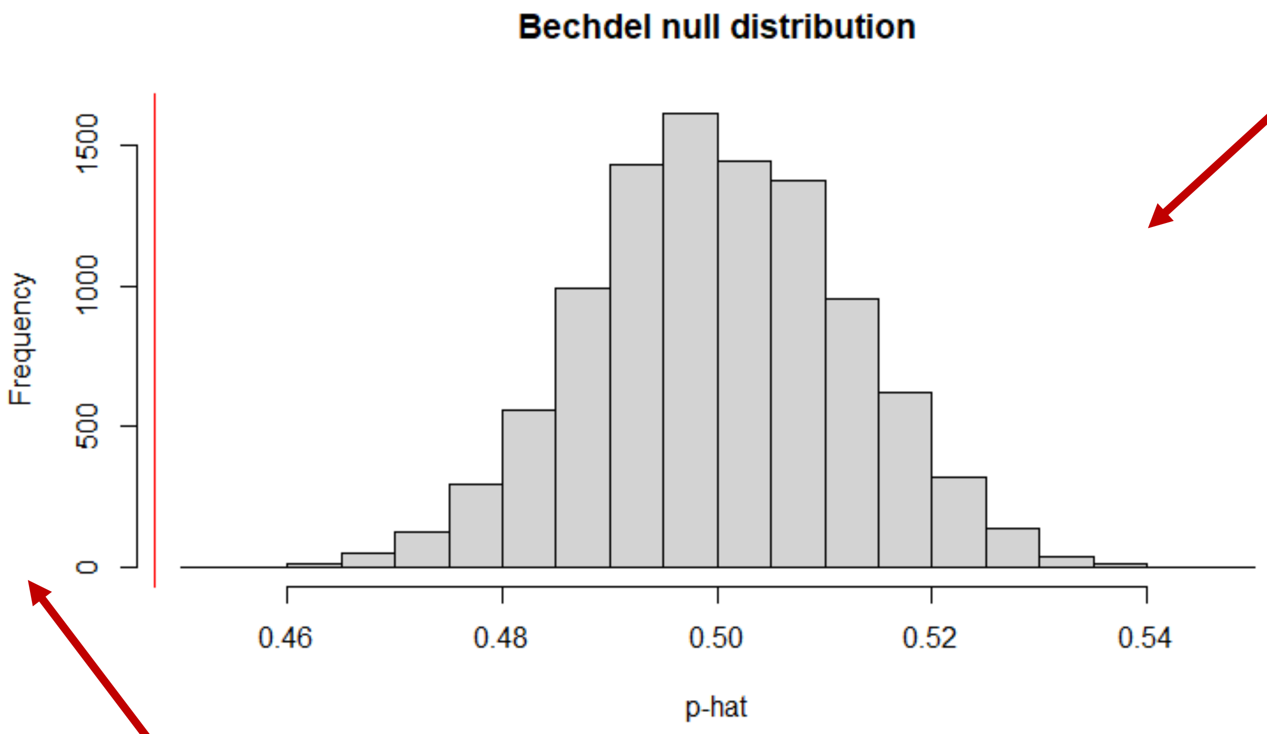
$$P(\text{STAT} \leq \text{ observed statistic } \mid H_0 = \text{True})$$

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

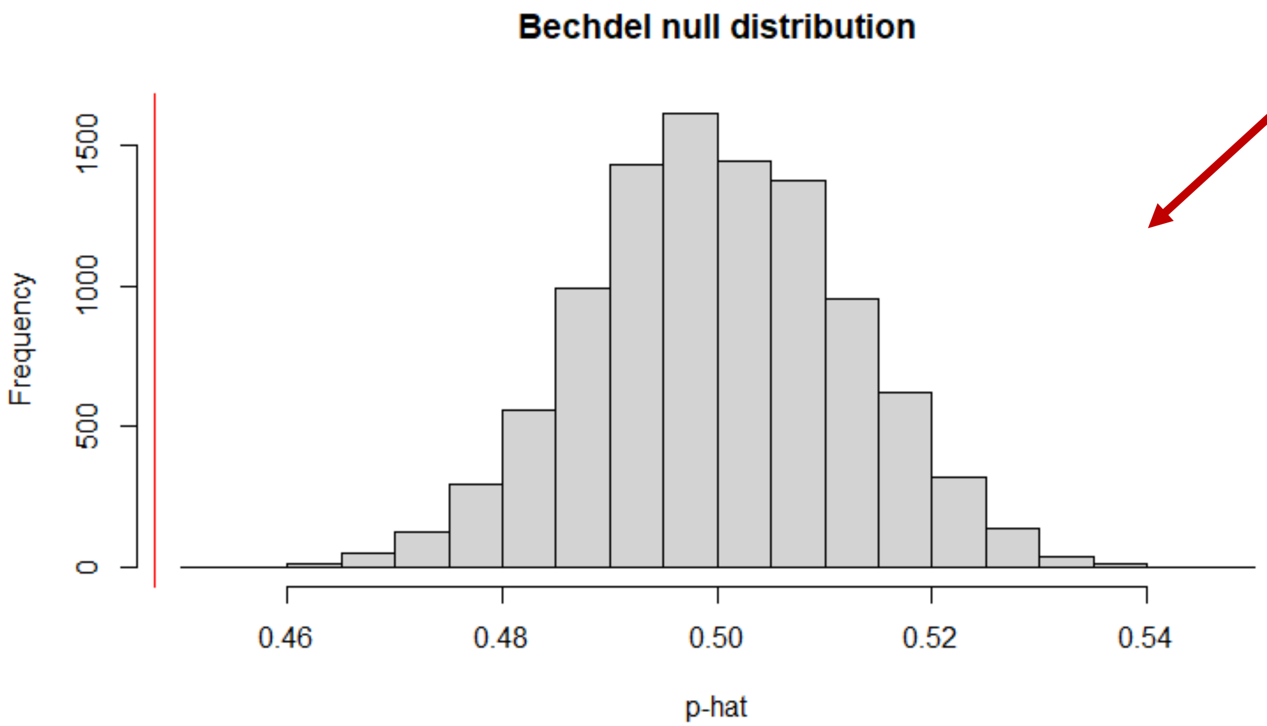
When our observed sample statistic is unlikely to come from the null distribution, we say the sample results are **statistically significant**

- 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

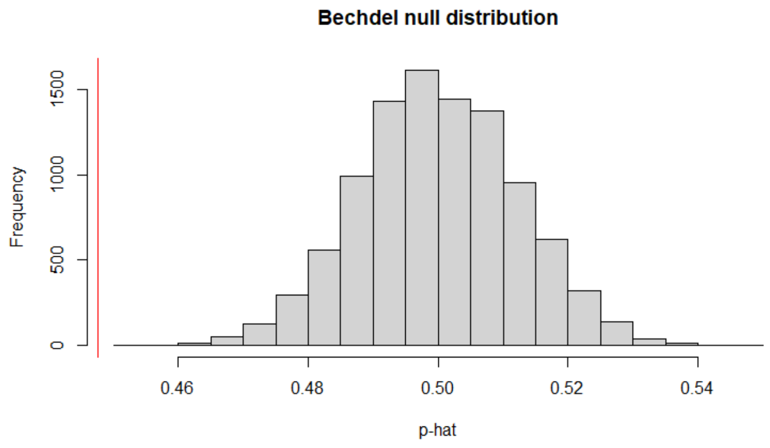


p-value = 0/10000 = 0.0000

Getting p-values using SDS100 functions

Flipping coins many times:

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



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

```
obs_stat <-  803/1794
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’

Are lie detectors more than 60% accurate?

A study by Hollien, Harnsberger, Martin and Hollien (2010) tried to assess the accuracy of lie detection software.

A sample of 48 participants were gather and attached to a lie detection device. They were asked to read deceptive (lying) material out loud.

The lie detector correctly reported that 31 out of the 48 participants were lying.

Does this provide evidence that lie detectors are more than 60% accurate?

Write down/discuss answers to the following questions

1. What are the cases here?

2. What is the variable of interest and is it categorical or quantitative?

3. What is the observed statistic - and what symbols should we use to denote it?

4. What is the population parameter we are trying to estimate - and what symbol should we use to denote it?

5. Do you think that this provides evidence that lie detector tests are more than 60% accurate?

5 steps to null-hypothesis significance testing (NHST)

5 steps of hypothesis testing:

1. State null and alternative hypotheses
2. Calculate statistic of interest
3. Create a null distribution
4. Calculate a p-value
5. Assess if there is convincing evidence to reject the null hypothesis

Let's go through these 5 steps now!

Step 1: State the null and alternative hypotheses

Null Hypotheis (H_0): Claim that there is no effect or no difference

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

Lie detector study

Q: What is the null hypothesis? (please state it using words)

Q: How would you write it in terms of the population parameter?

H_0: $\pi = 0.60$

Q: What is the alternative hypothesis?

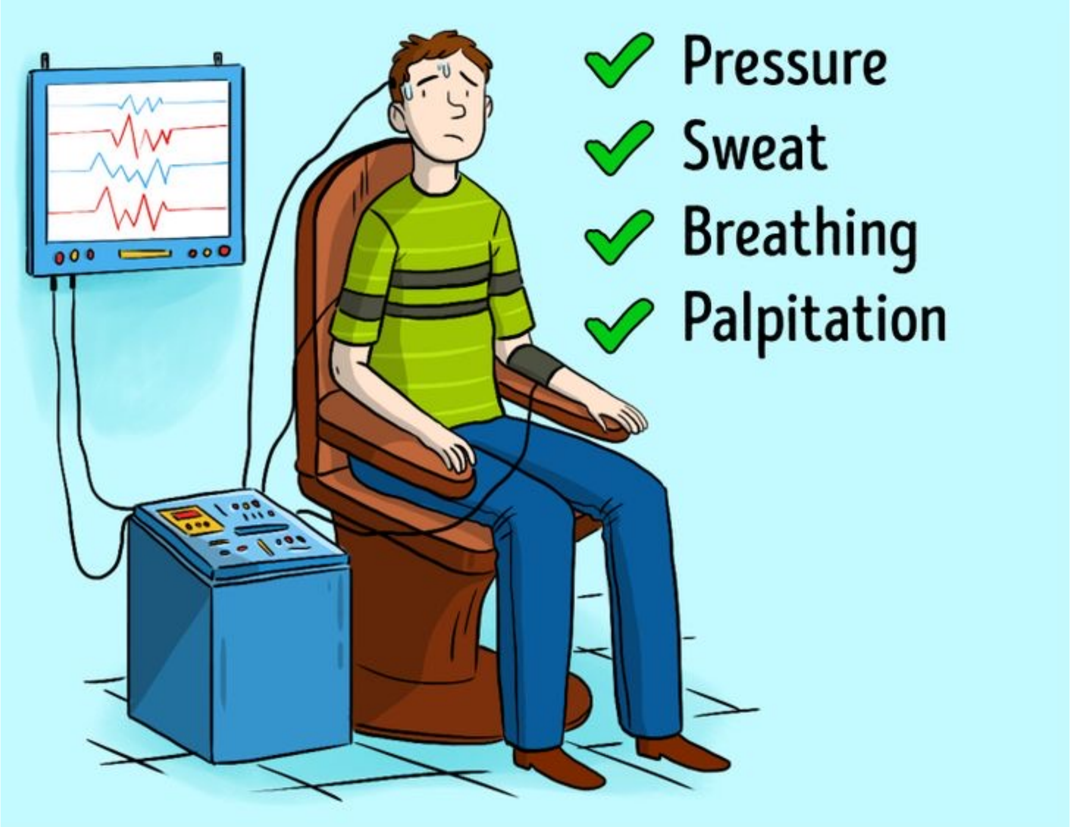
H_A: $\pi > 0.60$

Step 2: Calculate statistic of interest

For the lie detector study, what was the observed statistic?

31/48 percent of participants were detected as lying.

$\hat{p} \quad = \quad 31/48 \quad = \quad 0.645$



Step 3: Create a null distribution

Q: Please describe what the null distribution is here

Answer: A distribution of statistics ($\hat{p}$'s) consistent with the null hypothesis (H_0: $\pi = 0.60$)

Q: How can we create a null distribution?

Answer: when making inferences on *proportions* we can simulate flipping coins

Step 3: Create a null distribution

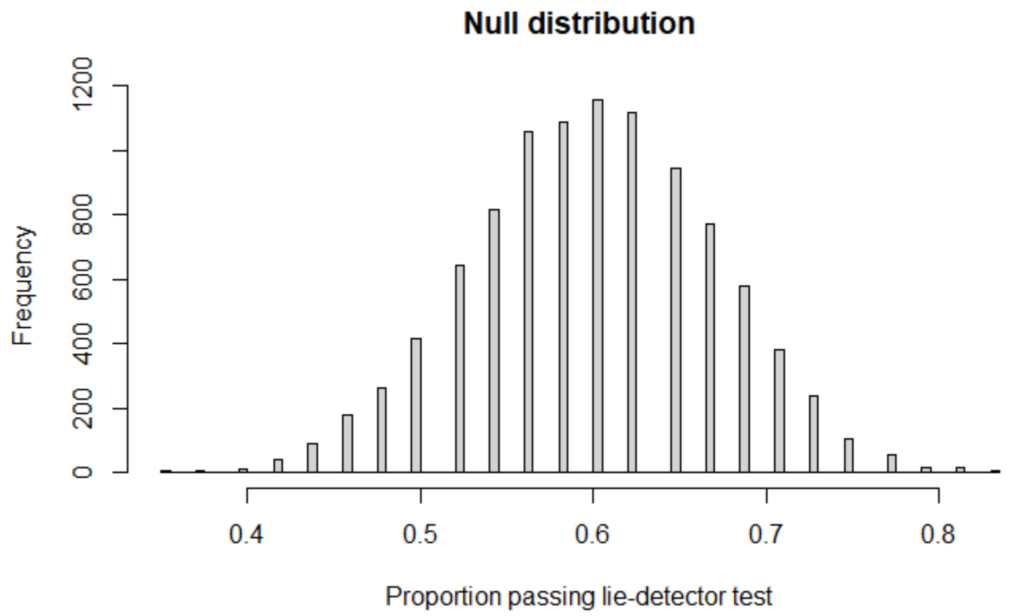
Please answer the following questions for the lie detector study

1. How many coins should we flip?
   - A: Data from 48 people was collected, so we want to flip 48 coins to simulate whether the lie detector got the correct answer for each person

2. What should the probability of heads be on each flip?
   - A: Our null parameter is $\pi = 0.60$ so to be consistent with the null hypothesis the probability of heads on each flip should be 0.60

3. How many simulations should we run?
   - A: 10,000 simulations should be enough to give us a good sense of the statistics we would get if the null hypothesis was true

Step 3: Create a null distribution



A null distribution (p̂'s) based on:

- ***10,000 simulations***
- Each simulation consists of flipping 48 coins
- With the probability of getting a head on each flip of 0.60

Step 4: Calculate a p-value

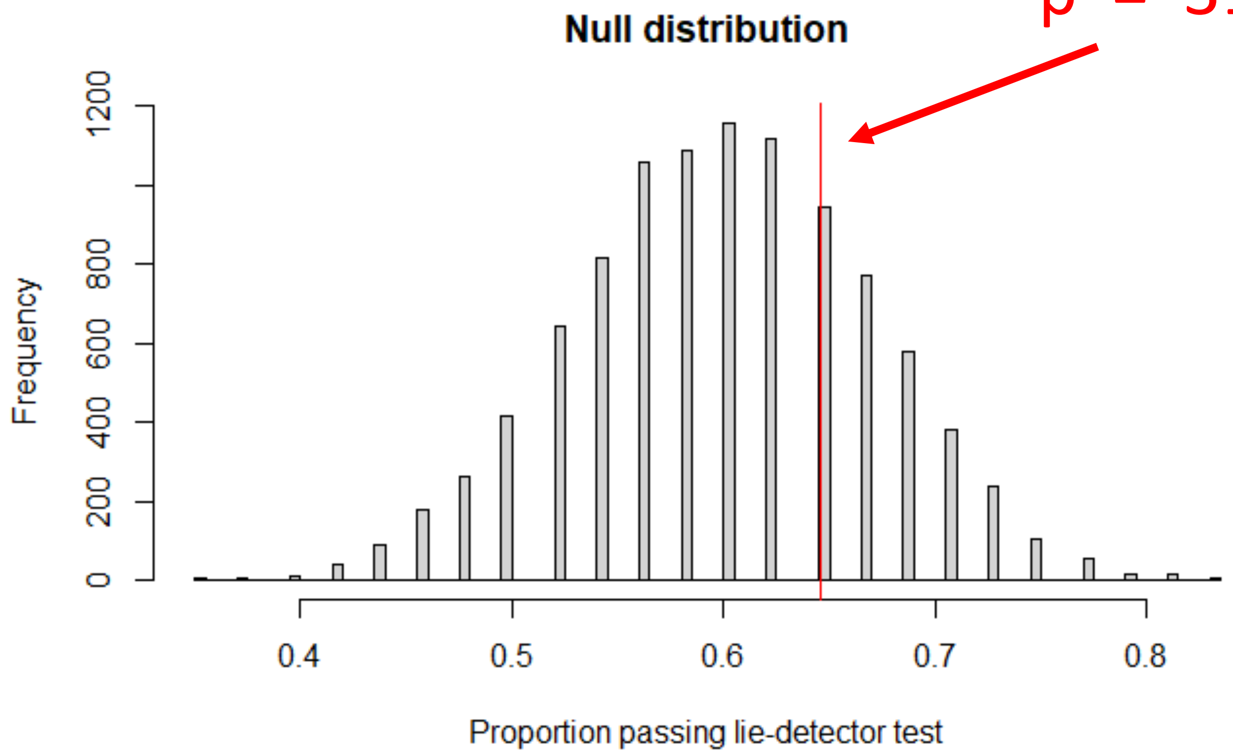
The **p-value** is the probability, when the null hypothesis is true, of obtaining a statistic as extreme or more extreme than the observed statistic

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

The smaller the p-value, the stronger the statistic evidence is against the null hypothesis and in favor of the alternative

Step 4: Calculate a p-value

$\hat{p}$ = 31/48 = 0.646



What is the p-value here?

- A: The p-value is 0.311

Step 5a: Assess if results are statistically significant

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 convincing evidence against H_0 in favor of H_A

Step 5b: Make a decision

Are the results seem statistically significant?


You won't get anything out of me. I can pass ANY lie detector test
BEEEEEEP
...Okay! I'll tell you everything!
MACHINE THAT BEEPS
THEJENKINSCOMIC

Let’s try the lie detector example in R…

One-sided vs. two-sided

In the examples we have seen, we were just interested if the parameter was <u>**greater**</u> (or less) than a hypothesized value

H_0: $\pi = 0.60$ H_A: $\pi > 0.60$

In other cases we might not have a directional alternative hypothesis

Testing whether a lie detector is not 60% accurate

Suppose we wanted to test what whether the lie detector was correct more ***or less*** than 60% of the time

- i.e., we are testing whether the lie detector is **not** 60% accurate

Step 1: Write down the null and alternative hypotheses

H_0: $\pi = 0.60$

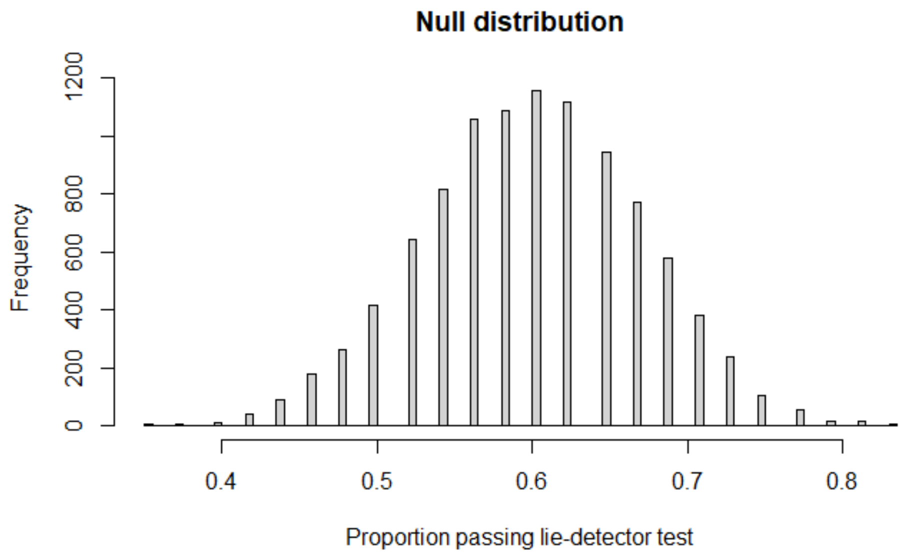
H_A: $\pi \neq 0.60$

Testing whether a lie detector is not 60% accurate

Step 2: Would the statement of hypotheses affect the observed statistic value?

A: No!

$$\hat{p} \quad = \quad 31/48 \quad = \quad 0.645$$



Step 3: Would this change in statement of hypotheses affect the null distribution?

A: No!

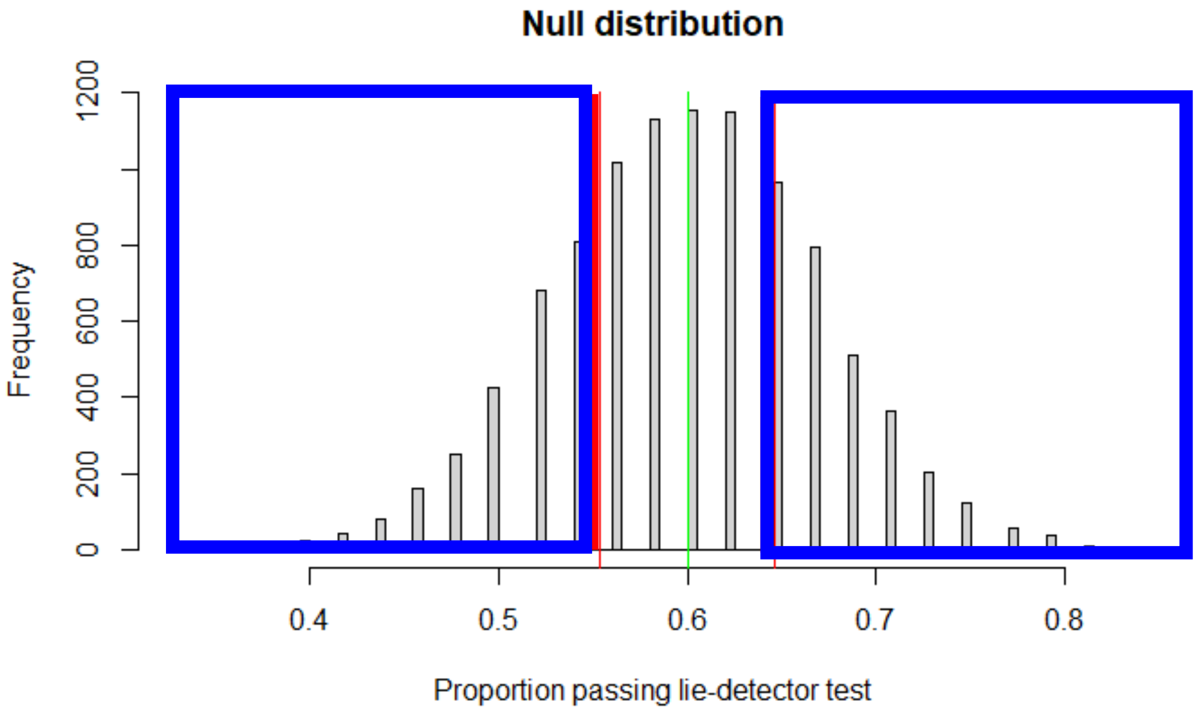
Testing whether a lie detector is not 60% accurate

Step 4: Would the statement of hypotheses affect p-value?

A: Yes!

We need to look for values **more extreme** than the observed statistic

Thus, the p-value for a two-sided test is about twice as large



Statement of alternative hypothesis is important

We need to state what you expect before analyzing the data

Our expectation (hypothesis statement) can change the p-value!

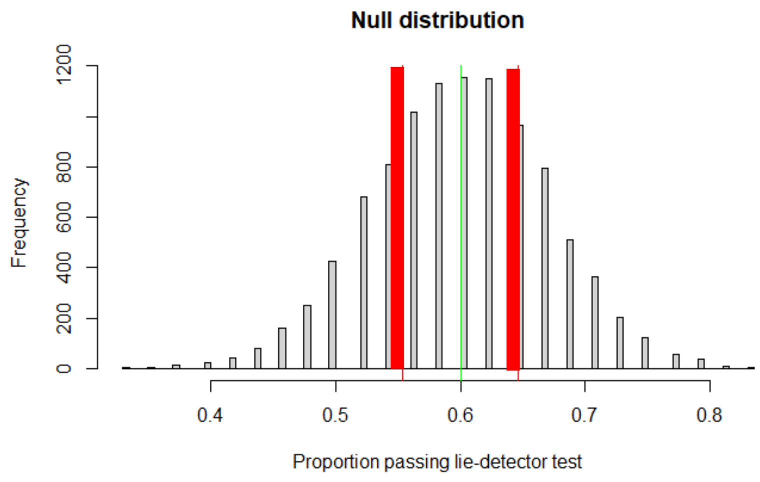
Estimating a p-value from a null distribution

For a one tailed alternative: Find the proportion of statistics in the null distribution that equal or exceed the original statistic in the direction (tail) indicated by the alternative hypothesis

For a two-tailed alternative: Find the proportion of statistics in the null distribution beyond the deviation of the observed statistic from the parameter value in both tails

- Alternatively, find the proportion of statistics in the null distribution beyond the original statistic in one of the tails, and then double the proportion to account for the other tail

How to estimate two sided p-values in R?




```
null_distribution <-  do_it(10000) * {
        rflip_count(48,  prob = .6)/48
}

pval_right_tail  <- pnull(lie_phat, null_dist, lower.tail = FALSE)
pval_left_tail  <- pnull(.6 + (.6 - lie_phat), null_dist, lower.tail = TRUE)

p_value <-  p_right_tail + p_left_tail
```