Hypothesis tests for a single proportion

Overview

Review and continuation of hypothesis tests

Hypothesis test terminology

More practice of hypothesis tests for a single proportion

Announcement

Homework 6 has been posted!

It is due on Gradescope on Sunday March 1st at 11pm

- **Be sure to mark each question on Gradescope!**

Announcement: Midterm exam

Exam is during regular class time on Thursday March 5th

- Exam is on paper

If you have accommodations, please schedule the exam with SAS

A practice exam (last year's exam) has been posted

Midterm exam “cheat sheet”

You are allowed an exam “cheat sheet”

One page, double sided, that contains **<u>only code and equations</u>**

- No code comments allowed

Cheat sheet must be on a regular 8.5 x 11 piece of paper

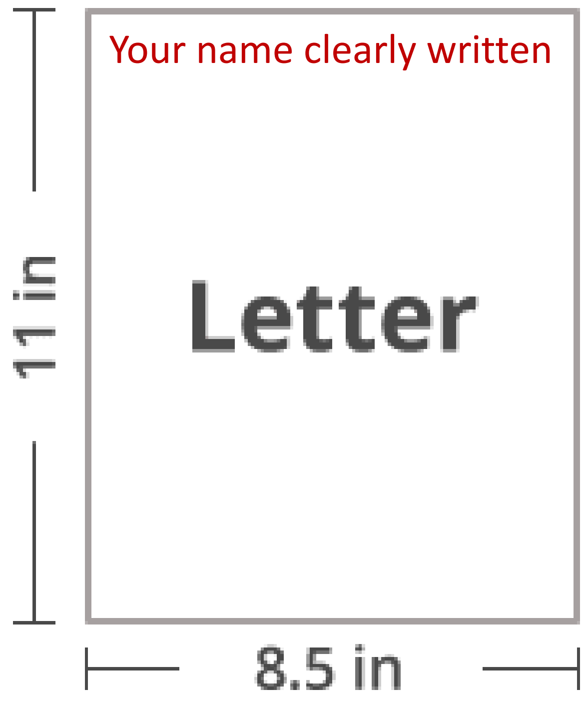
- Your name on the upper left of both sides of the paper

Recommend making a typed list of all functions discussed in class and on the homework

- This will be useful beyond the exam

You must turn in your cheat sheet with the exam

- Failure to do so will result in a 20 point deduction



Continuation of hypothesis tests for a single proportion

Review: A quick note on probability

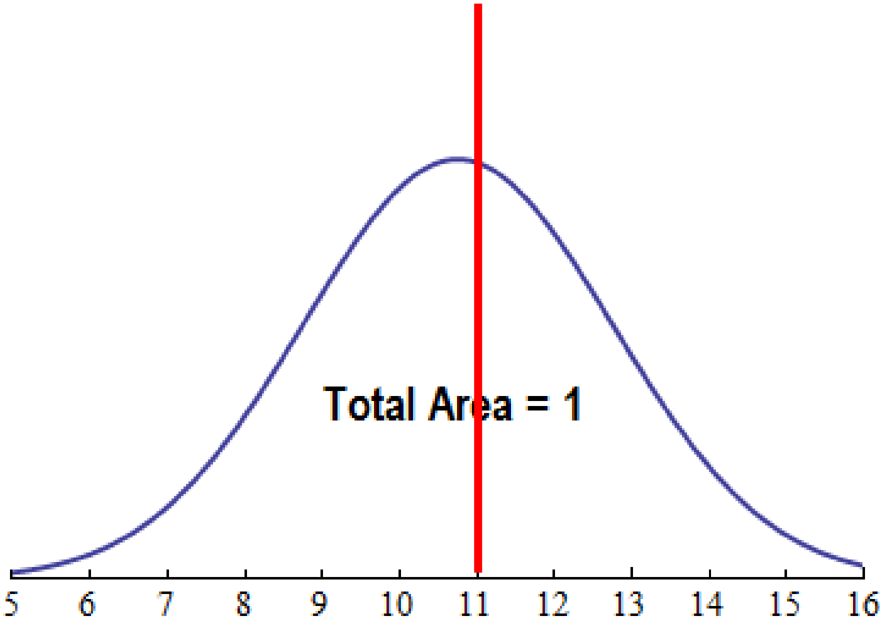
Probability is a way of measuring the likelihood that an event will occur

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:

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



Review: Statistical tests (hypothesis test)

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

<u>Example 1</u>: The correlation between the ages of husbands and wives is 0.85

A: $\rho = 0.85$

<u>Example 2</u>: The variability (standard deviation) in temperatures on February 24th across years in New Haven Connecticut is 12 degrees

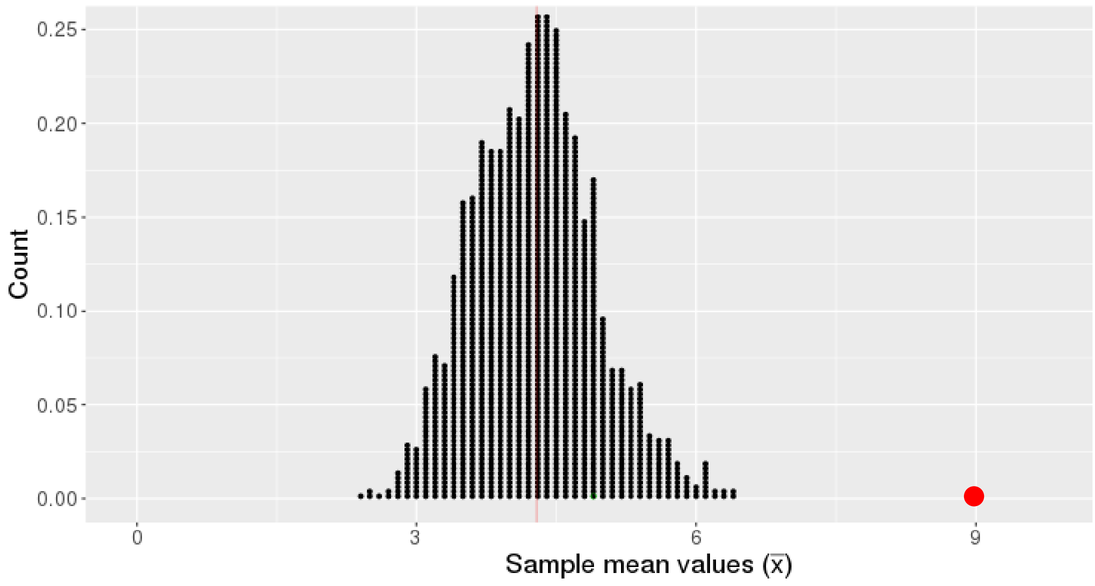
A: $\sigma = 12$

Review: Basic hypothesis test logic

We start with a claim about a population parameter

- E.g., μ =

This claim implies we should get a certain distribution of statistics



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

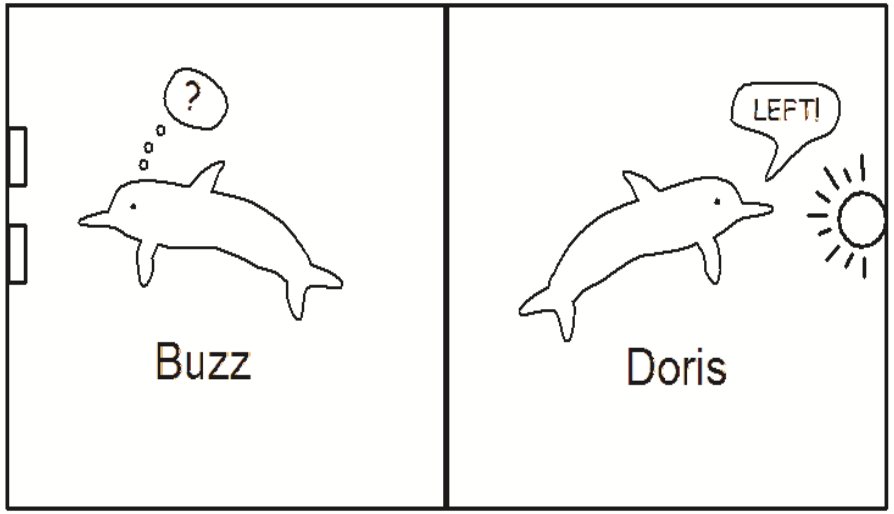
Are dolphins capable of abstract communication?

Dr. Jarvis Bastian is the 1960’s wanted to know whether dolphins are capable of abstract communication

He used an old headlight to communicate with two dolphins (Doris and Buzz)

- **Stead light**: push button on **right** to get food
- **Flashing**: push button on the **left** to get food

A canvas was then put in the middle of the pool with Doris on one side and Buzz on the other



Buzz got 15 out of 16 trials correct

The dolphin communication study

If Buzz was just guessing, what would you expect the value of the parameter to be?

$\pi = 0.5$

If Buzz was not guessing, what would you expect the value of the parameter to be?

$\pi > 0.5$

Chance models

We can assess whether 15 out of 16 correct trials ($\hat{p}$ = .975) is beyond what we would expect to see by chance by:

1. Flipping a fair coin ($\pi_{heads} = 0.5$) 16 times and marking if we get 15 or more heads
   - This simulates the number of times Buzz would have gone to the correct food well if he was merely guessing

2. Repeating this process many times to see what percentage of the time we get 15 or more heads

Flipping coins using SDS1000 functions

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

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

num_flips: the number of times to flip a coin

- 16 for Doris/Buzz

prob: the probability of success on each trial

- .5 if Buzz was guessing

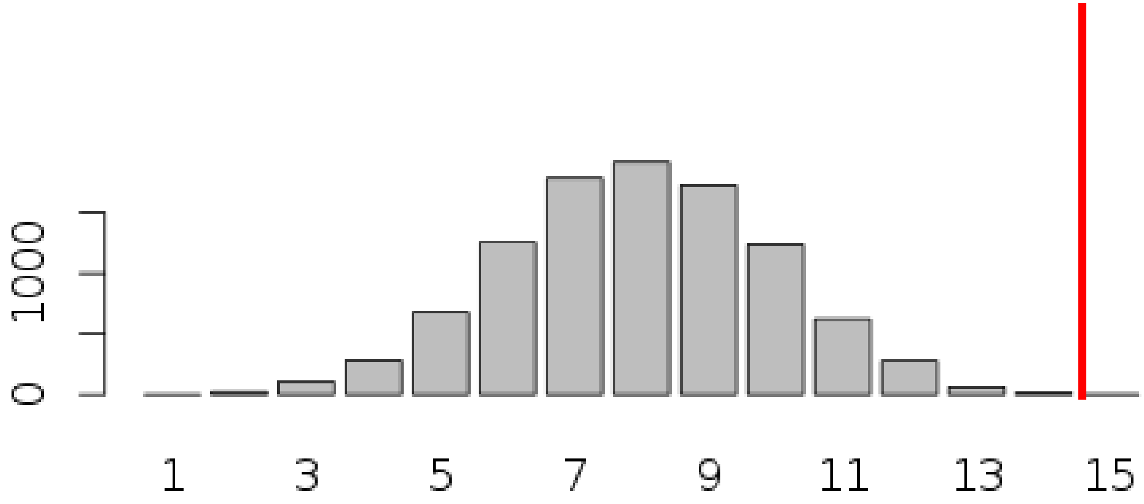
We can repeat flips many times using the do_it() function:

```
library(SDS1000)

flip_simulations  <-  do_it(10000) * {
      rflip(16,  prob = .5)
}
```

Simulating Flipping 16 coins 10,000

0	0
1	1
2	22
3	105
4	283
5	679
6	1257
7	1786
8	1920
9	1726
10	1238
11	623
12	279
13	63
14	15
15	3
16	0

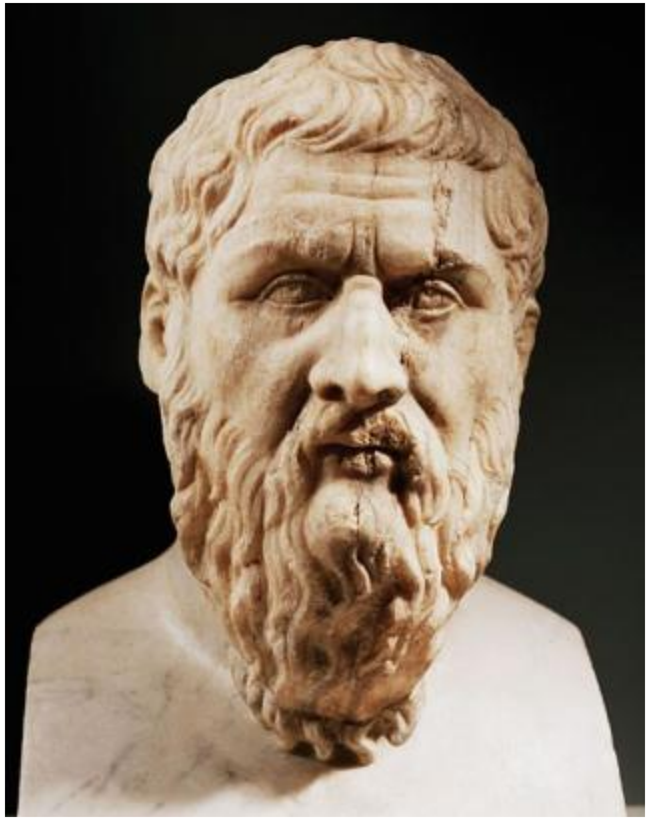


Q: Is it likely that Buzz was guessing?

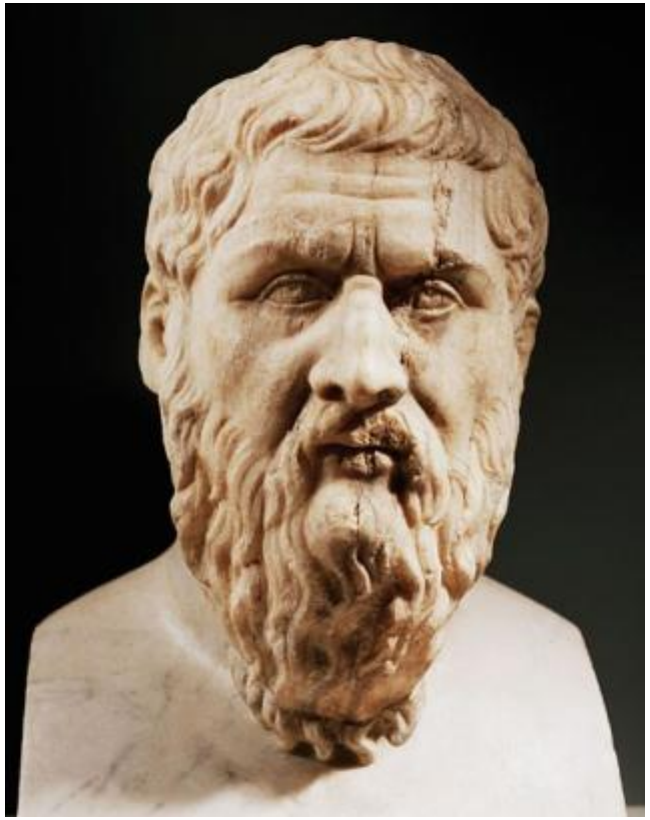
Q: Are dolphins capable of abstract communication?

Hypothesis tests: central ideas and terminology

Question: who is this?



Question: who is this?



How can we write the null hypothesis in symbols?

We believe in the ***alternative hypothesis***

- Doris an Buzz can communicate

How can we write the alternative hypothesis in symbols?

How can we convince Gorgias?

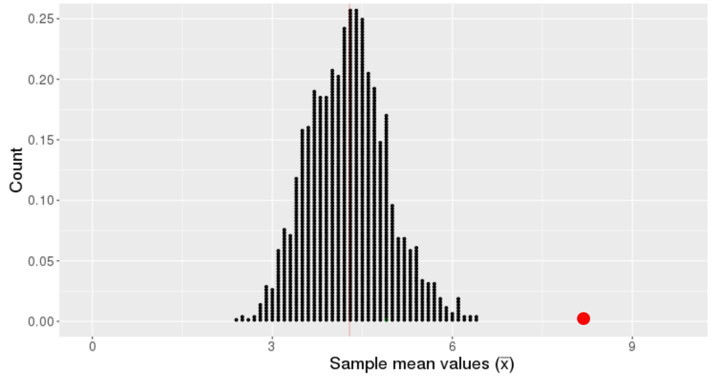
To prove Gorgias wrong, we will start by assuming he is right!

Namely, we will assume H_0 (that $\pi = 0.5$)

We will then generate a number of statistics ($\hat{p}$) that are consistent with H_0

- i.e., we will create a ***null distribution***

If our observed statistic looks very different from the statistics generated under we can reject H_0 and accept H_A

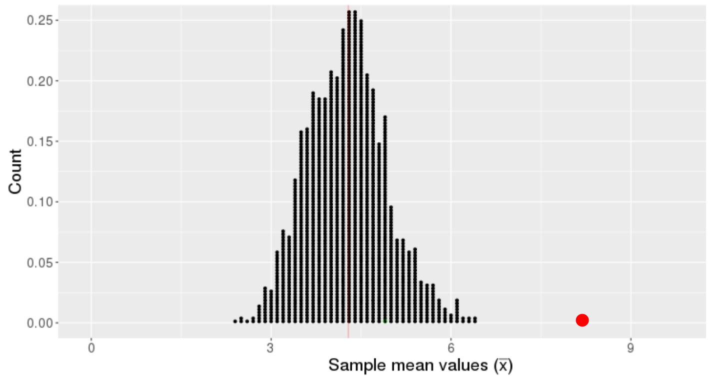


Terminology

Null Hypothesis (~~H_0~~): 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 dolphin communication study

1. What is the null hypothesis in words?

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

H_0: $\pi = 0.5$

3. What is the alternative hypothesis?

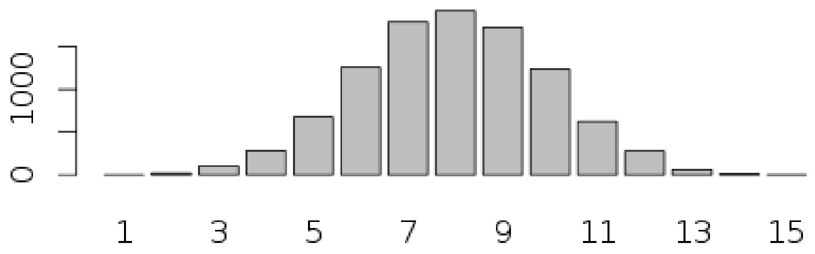
H_A: $\pi > 0.5$

Null Distribution

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

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"



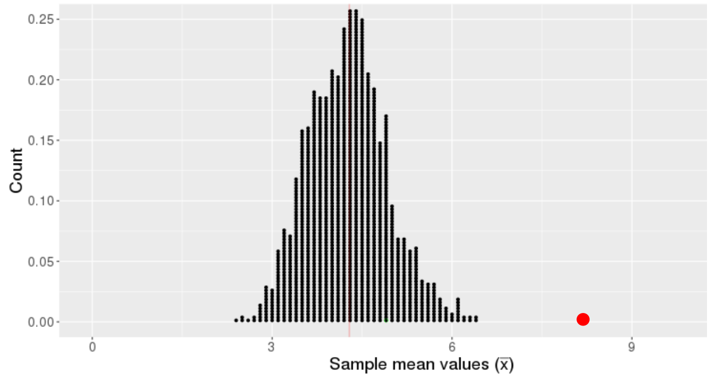
P-values

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 extreme as our <u>observed statistic</u> from the <u>null distribution</u>

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

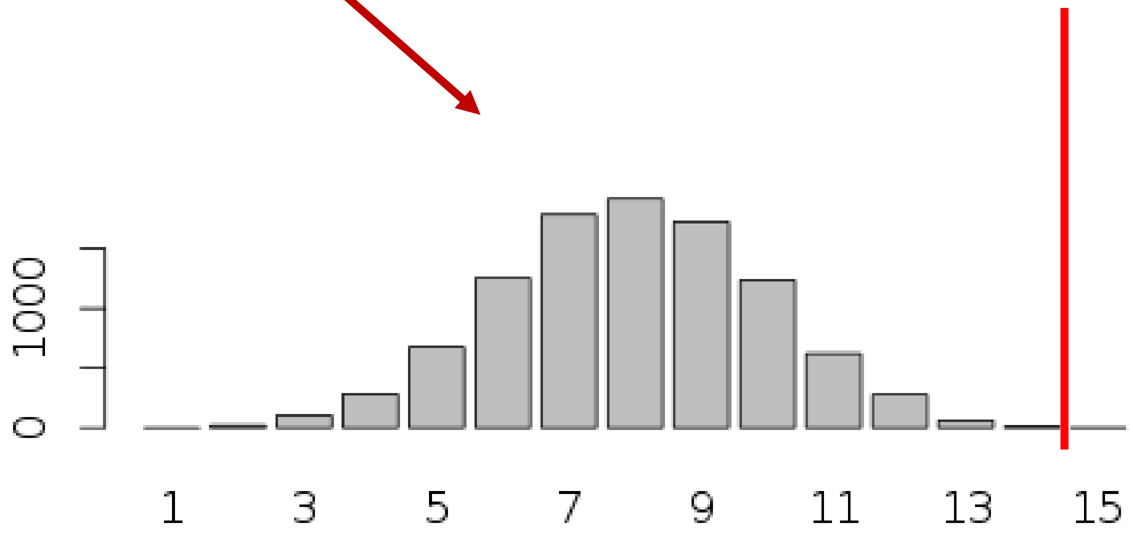
The smaller the p-value, the stronger the statistic evidence is against the null hypothesis



Buzz and Doris example

0	0
1	1
2	22
3	105
4	283
5	679
6	1257
7	1786
8	1920
9	1726
10	1238
11	623
12	279
13	63
14	15
15	3
16	0

Null distribution



p-value = 3/10000 = 0.0003

Statistical significance

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

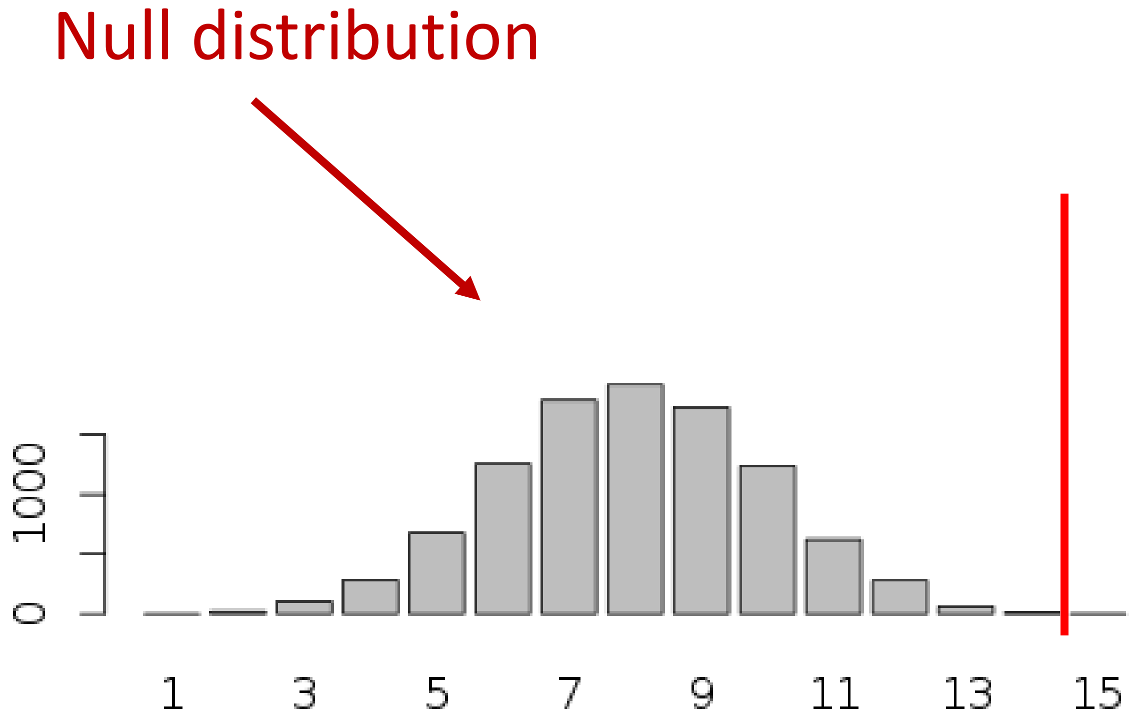
- i.e., we have a small p-value

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

Side note: We will discuss whether we should really use the term "statistical significance" later in the semester

Buzz and Doris example

0	0
1	1
2	22
3	105
4	283
5	679
6	1257
7	1786
8	1920
9	1726
10	1238
11	623
12	279
13	63
14	15
15	3
16	0

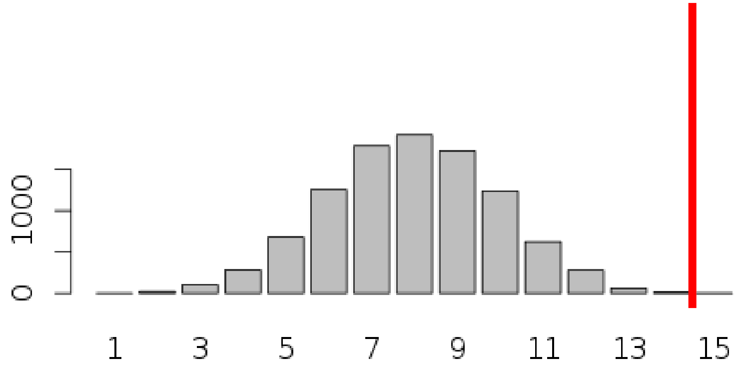


p-value = 3/10000 = 0.0003

Getting p-values using SDS1000 functions

Flipping coins many times:

```
flip_simulations <- do_it(10000) * {
    rflip(16, prob = .5)
}
```



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

```
obs_stat <- 15
p_value <- pnull(obs_stat, flip_simulations, lower.tail = FALSE)
```

Key steps hypothesis testing

1. State the null hypothesis… and the alternative hypothesis
   - Buzz is just guessing so the results are due to chance: H_0: $\pi = 0.5$
   - Buzz is getting more correct results than expected by chance: H_A: $\pi > 0.5$

2. Calculate the observed statistic
   - Buzz got 15 out of 16 guesses correct, or $\hat{p}$ = .973

3. Create a null distribution that is consistent with the null hypothesis
   - i.e., what statistics would we expect if Buzz was just guessing

4. Examine how likely the observed statistic is to come from the null distribution
   - What is the probability that the dolphins would guess 15 or more correct?
   - 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’

Another example of hypothesis tests for a single proportion

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 gathered 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 Hypothesis (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?

Q: What is the alternative hypothesis?

Step 2: Calculate statistic of interest

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

Step 3: Create a null distribution

Q: Please describe what the null distribution is here

Q: How can we create a null distribution?

Step 3: Create a null distribution

Please answer the following questions for the lie detector study

1. How many coins should we flip?

2. What should the probability of heads be on each flip?

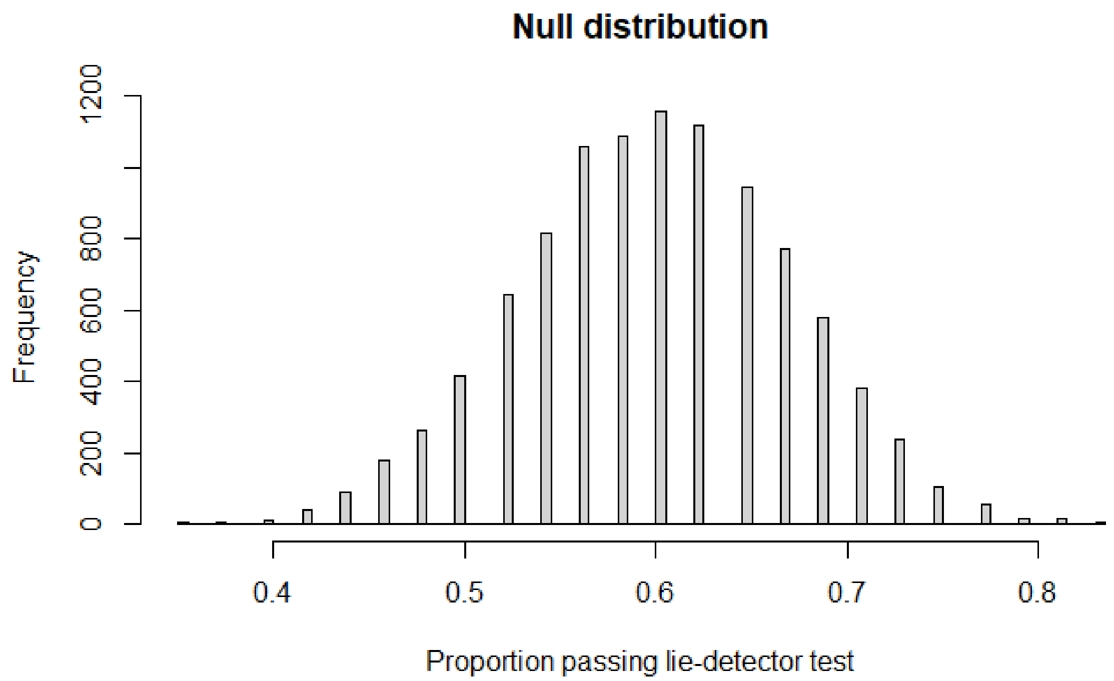
3. How many simulations should we run?

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

```
null_dist <-  do_it(10000) * {
    rflip(48,  prob = .6)
}
```



Step 4: Calculate a p-value

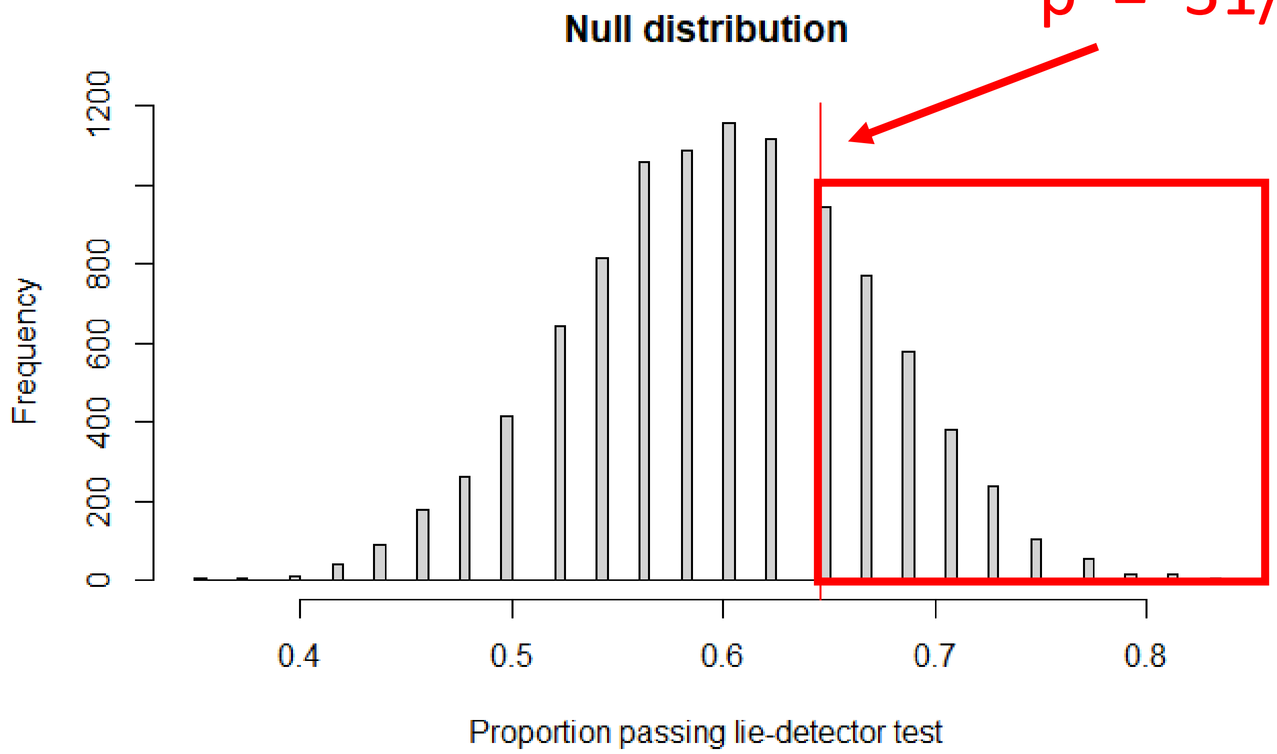
The **p-value** is the probability, when the null hypothesis is true, of obtaining a statistic as extreme or more extreme than the observed statistic

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

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?

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?

Let’s try the lie detector example in R…