Introduction to hypothesis tests

Overview

Last quick review of confidence intervals

A brief note on probability

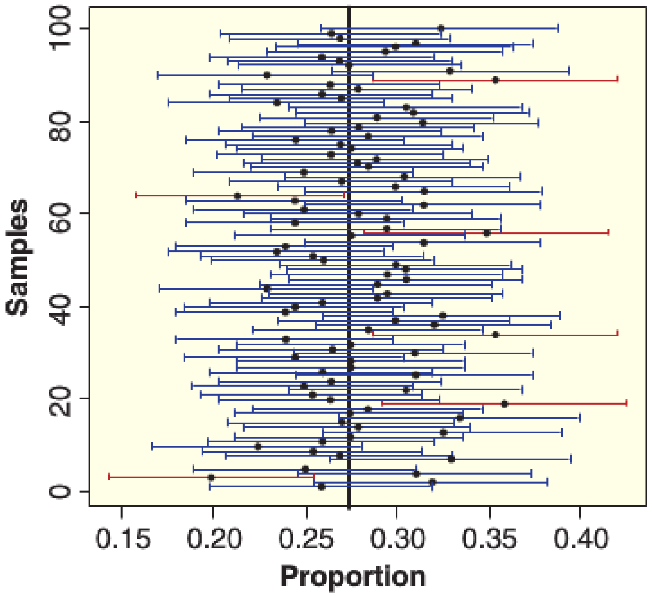
Introduction to hypothesis tests

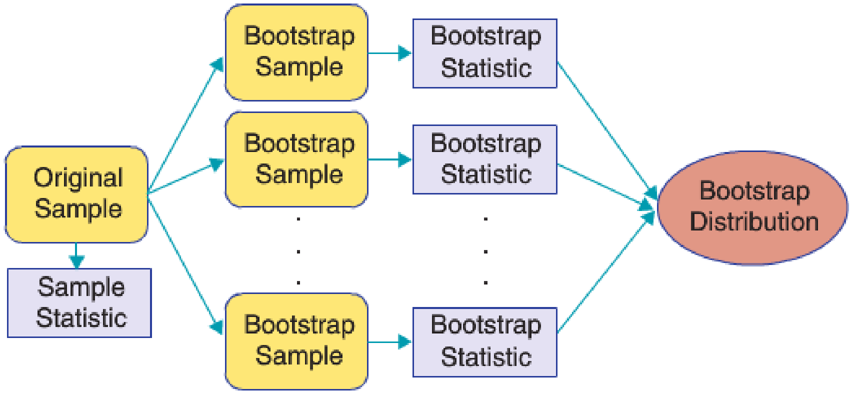
Central concepts and terminology in hypothesis testing

Very quick review of bootstrap confidence intervals

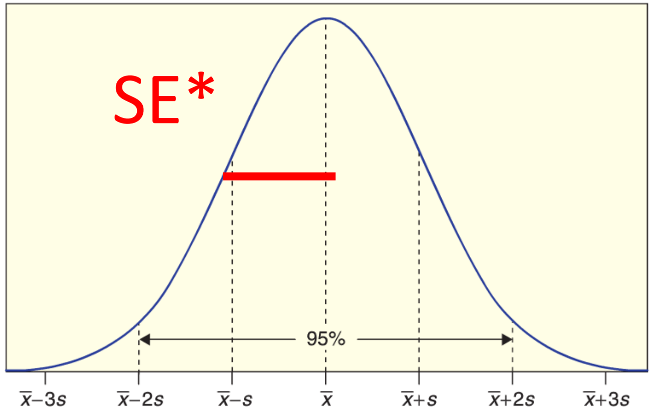
Very quick review of bootstrap confidence intervals

Confidence intervals give range of plausible values for a parameter



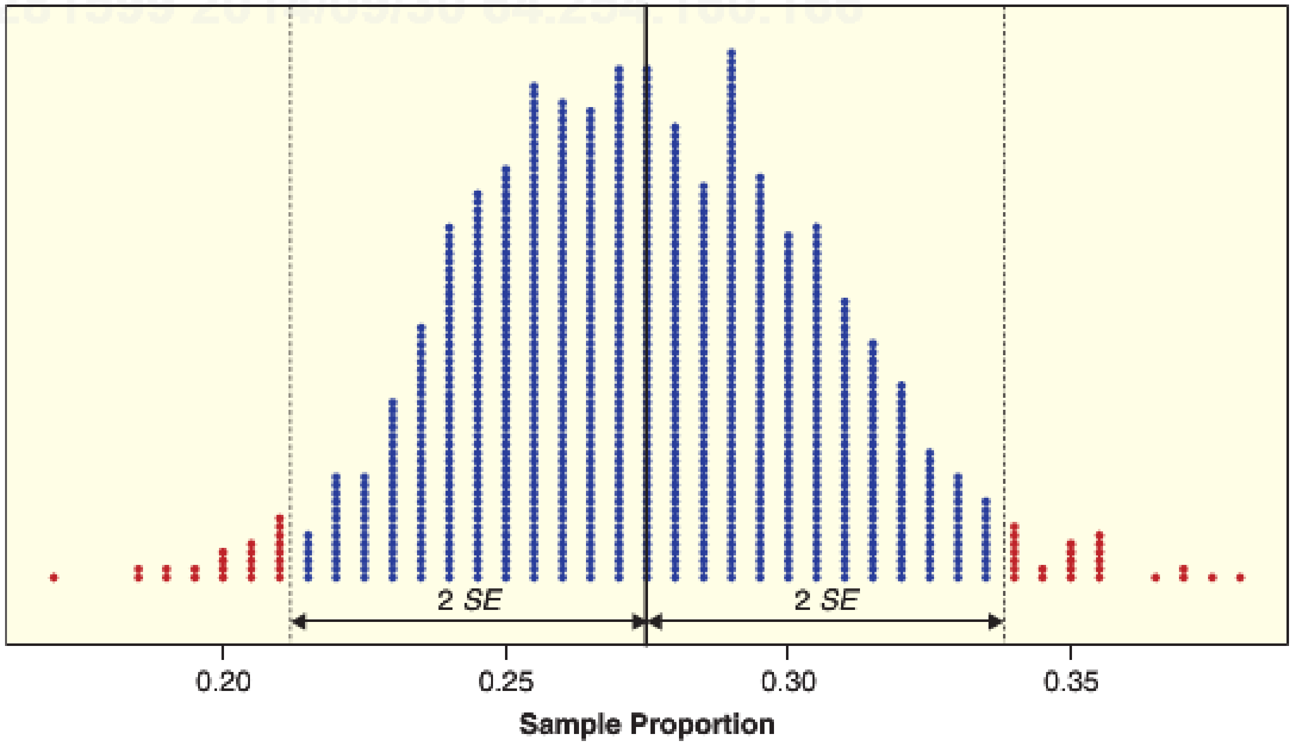


Bootstrap distribution



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

Are we feeling confident about confidence intervals?



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

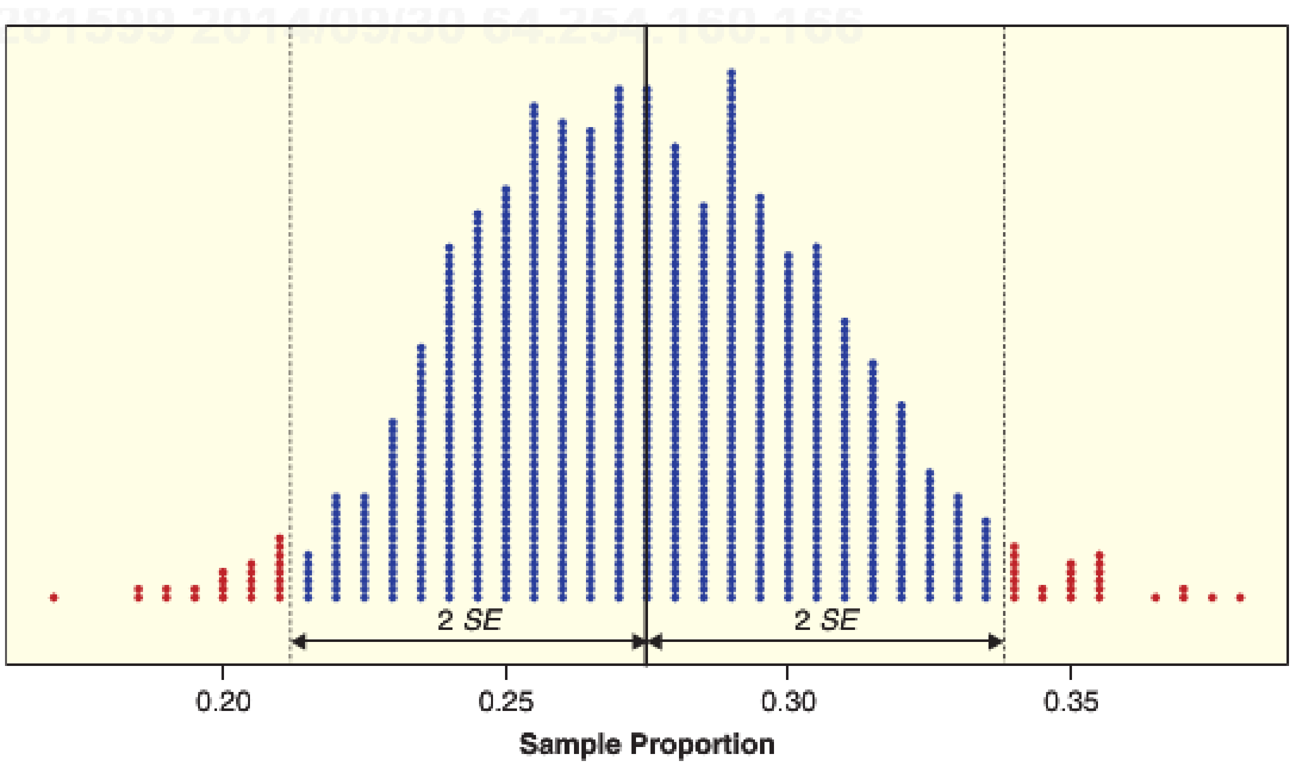
Margin of Error

Half width of confidence interval

Are we feeling confident about confidence intervals?

How far can we slide the center of the confidence interval until we don't capture the parameter?

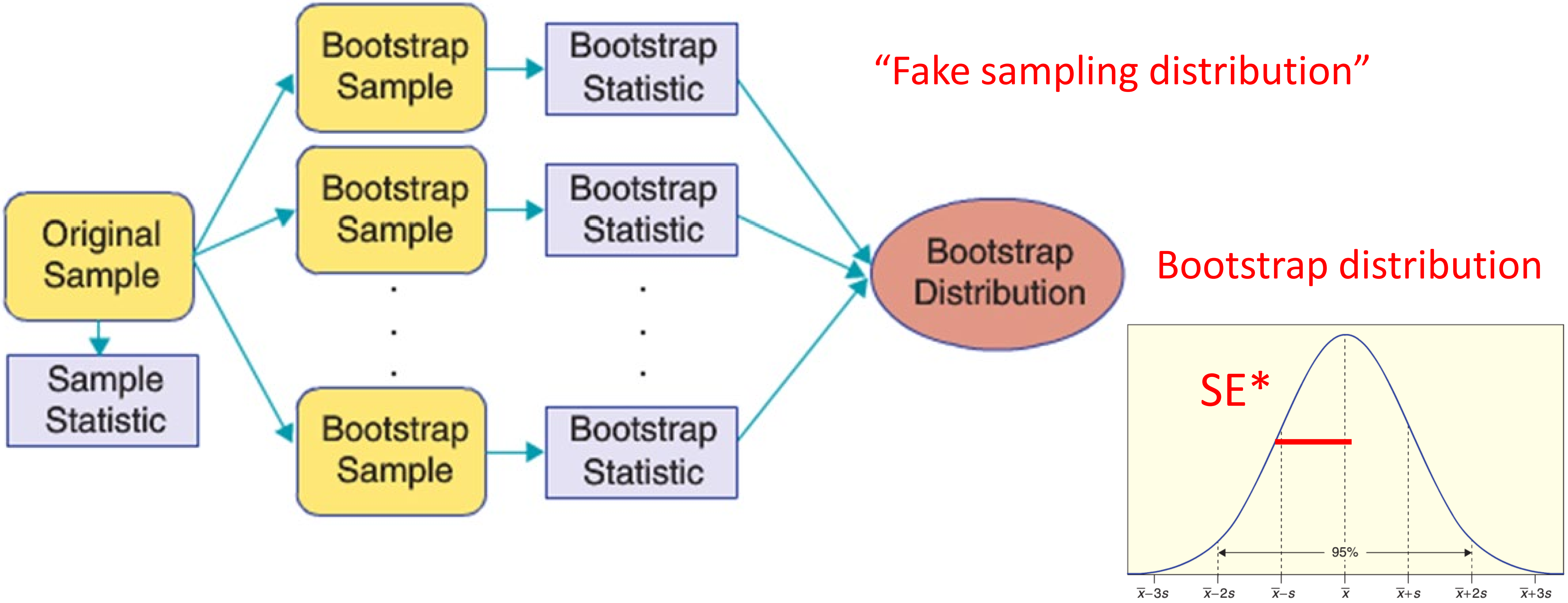
What would happen if we made the margin of error smaller?



Confidence interval

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

Using the bootstrap to estimate the standard error (SE*)



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)
```

Four-step procedure for the bootstrap

Introduction to hypothesis tests

A quick note on probability

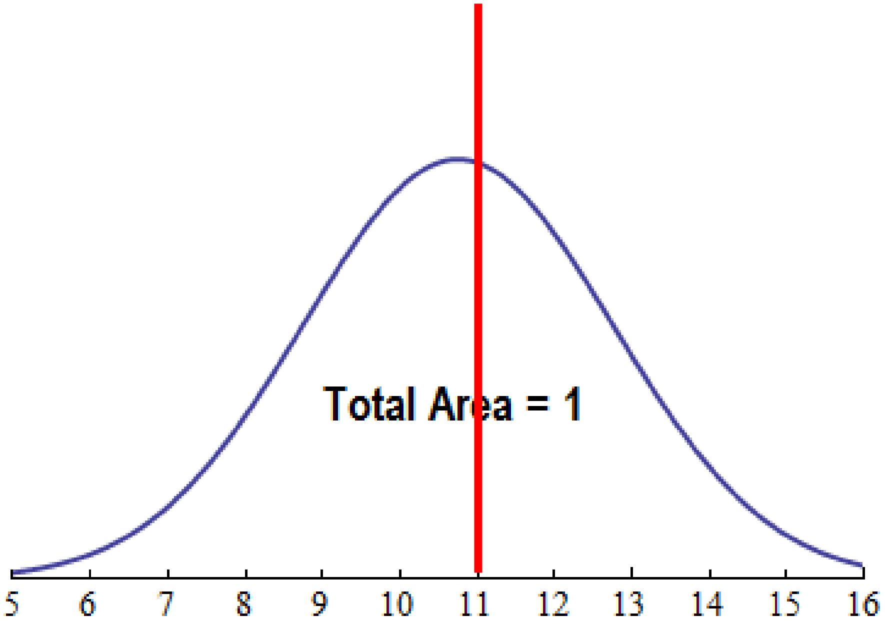
Probability is a way of measuring the likelihood that an event will occur.

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:

- Pr(X < 11) = 0.55
- Pr(X > 20) = 0



Statistical tests (hypothesis test)

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

Example 1: we might make the claim that Biden's approval rating for all US citizens is 54%.

How can we write this using symbols?

- $\pi = .54$

Statistical tests   (hypothesis test)

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

Example 2: we might make the claim that the average height of a baseball player is 72 inches.

How can we write this using symbols?

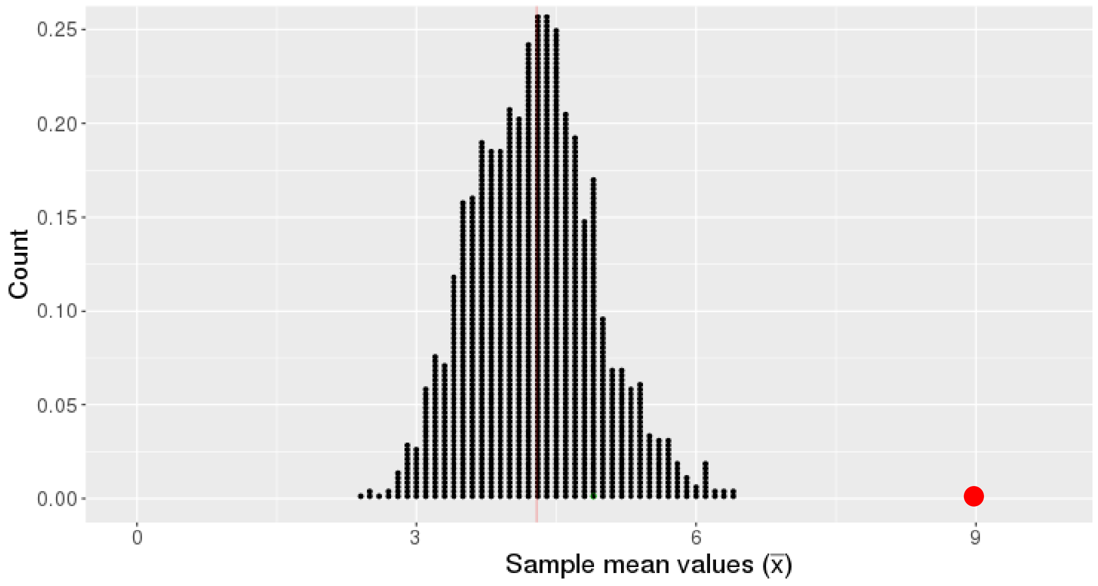
- $\mu = 72$

Basic hypothesis test logic

We start with a claim about a population parameter.

- E.g., $\mu = 4.2$

This claim implies we should get a certain distribution of statistics.



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

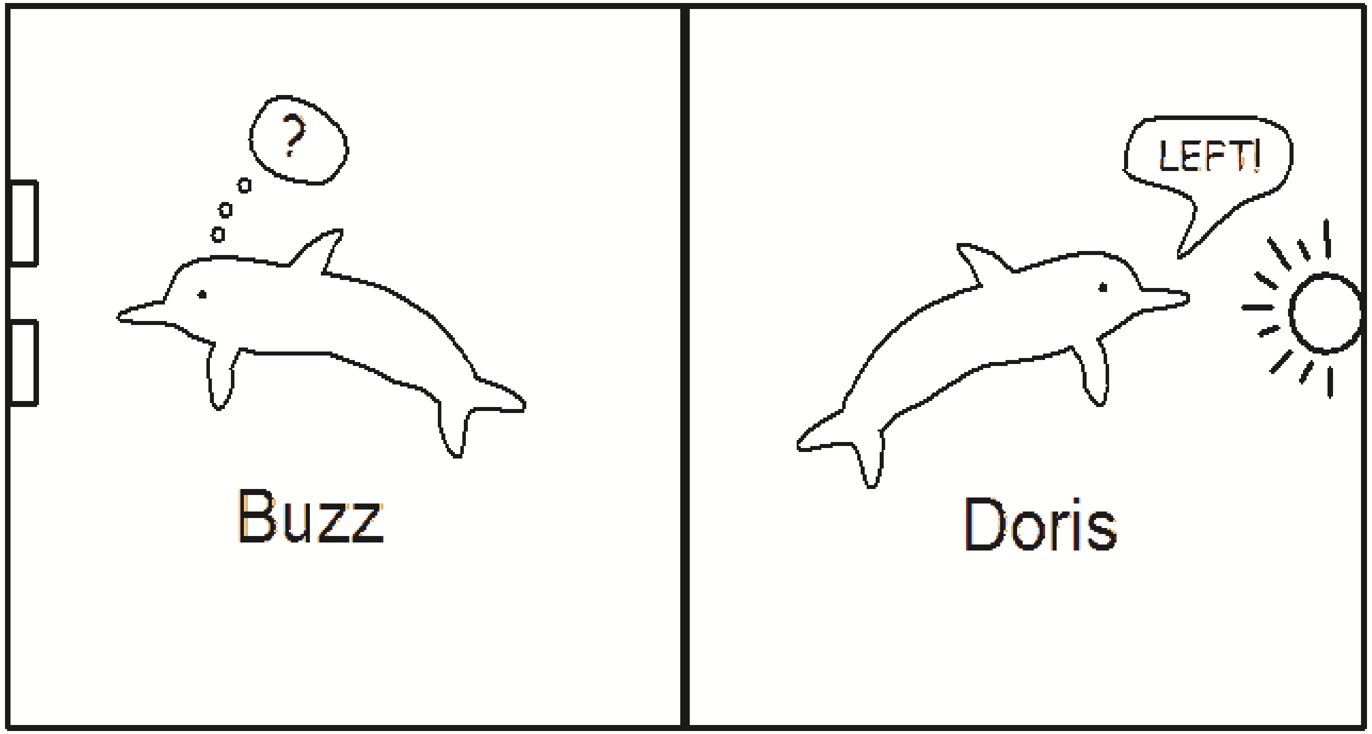
Are dolphins capable of abstract communication?

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.



Left

Right

Buzz got 15 out of 16 trials correct.

On Canvas, answer 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 the results are due to chance?
   - i.e., how many correct answers do you think Buzz would have gotten if he was guessing?

6. Are dolphins capable of abstract communication?

The dolphin communication study

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

$\pi = 0.5$

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

$\pi > 0.5$

Chance models

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

- i.e., beyond what we would expect to see if $\pi = 0.5$?

If buzz was guessing we can model his guesses as a coin flip:

Heads = correct guess

Tails = incorrect guess

Let's flip a coin 16 times and see how many times we get 15 heads

Chance models

To really be sure, how many repetitions of flipping a coin 16 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

- 16 for Doris/Buzz

prob: the probability of success on each trial

- .5 if Buzz was guessing

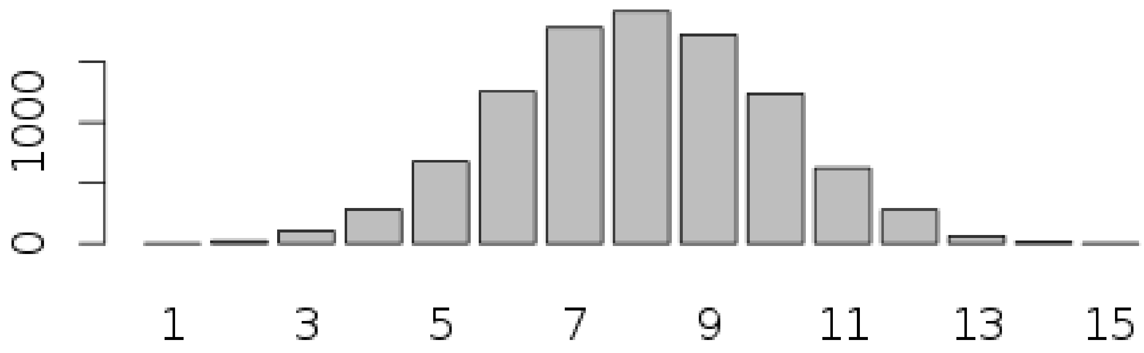
We can repeat flips many times using the do_it() function:

```
library(SDS100)

flip_simulations  <-  do_it(10000) * {
        rflip_count(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

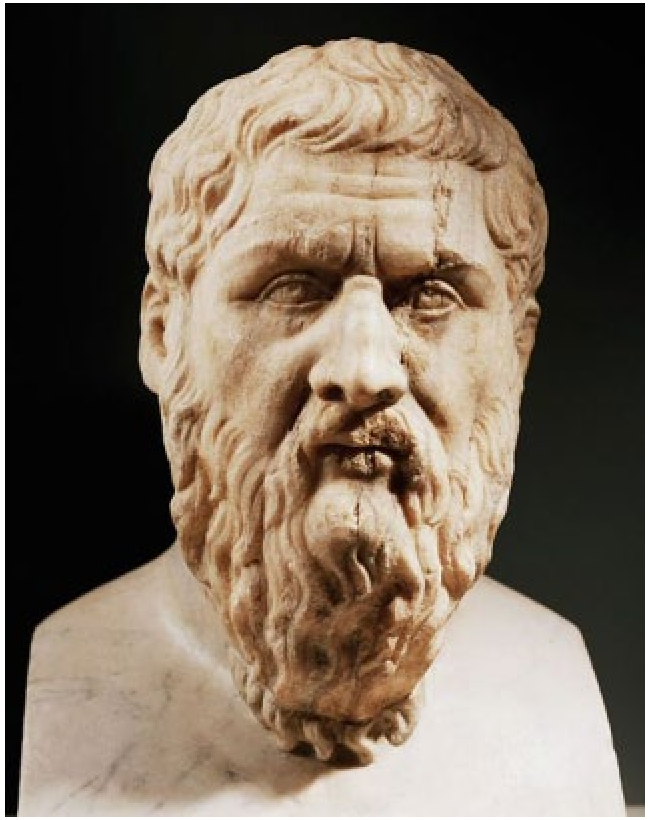


Q10: Is it likely that Buzz was guessing?

Q11: Are dolphins capable of abstract communication?

Hypothesis tests: central ideas and terminology

Question: who is this?



A: Gorgias

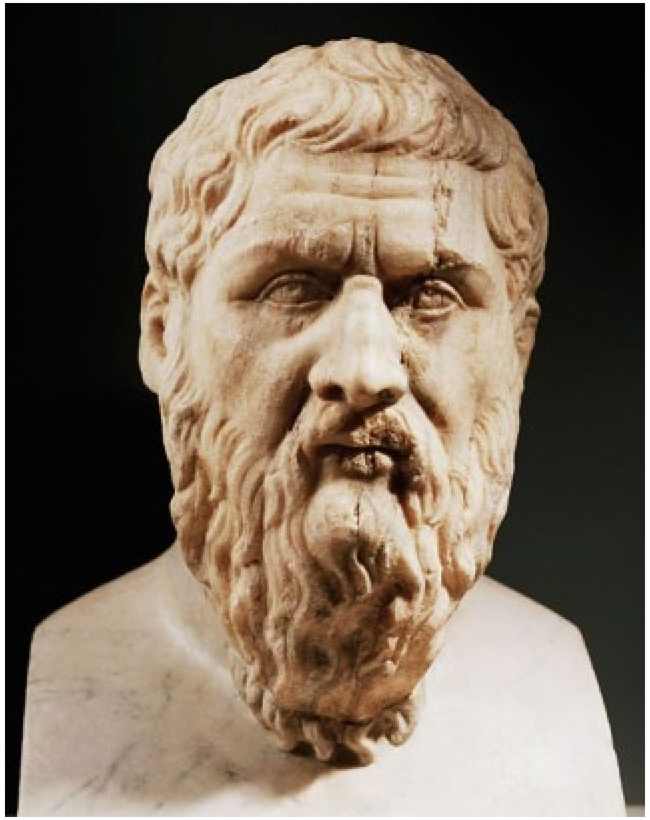
Question: Who is Gorgias?

A: a skeptic/nihilist

Question: Does Gorgias believe Doris and Buzz can communicate?

A: No!

Question: who is this?



Gorgias believes in the ***null hypothesis***

- that Buzz was guessing

How can we write the null hypothesis in symbols?

H_0: $\pi = 0.5$

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

- Doris an Buzz can communicate

How can we write the alternative hypothesis in symbols?

H_A: $\pi > 0.5$

How can we convince Gorgias?

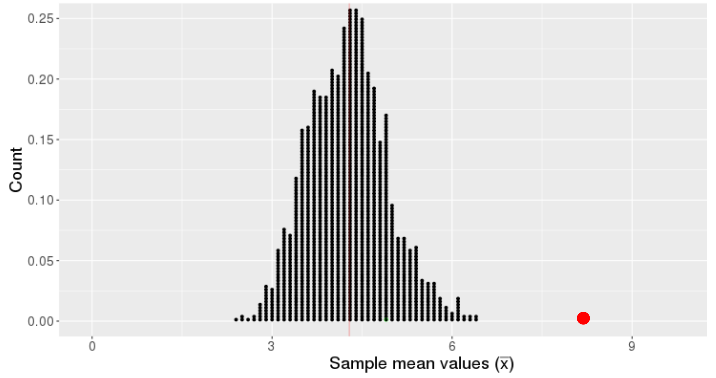
To prove Gorgias wrong, we will start by assuming he is right!

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

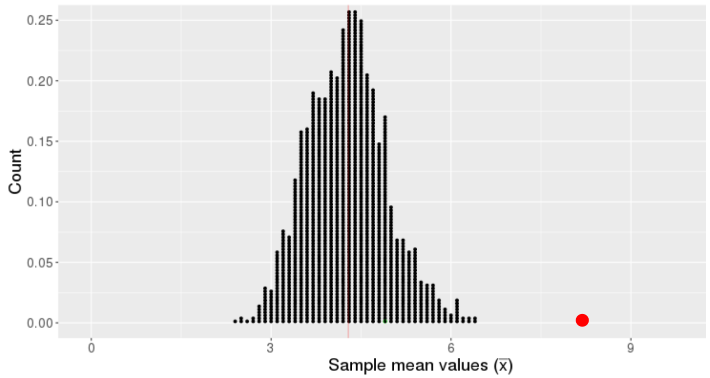


Terminology

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?

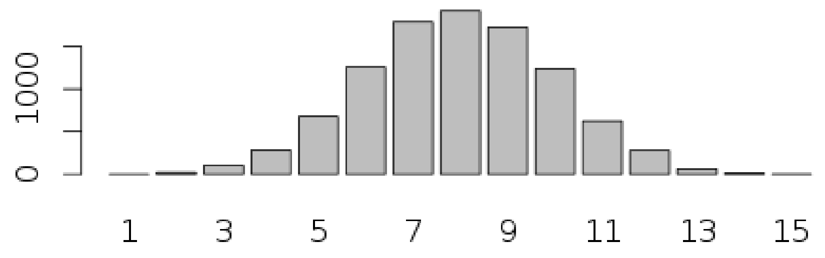
$$H_A: \pi > 0.5$$

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"



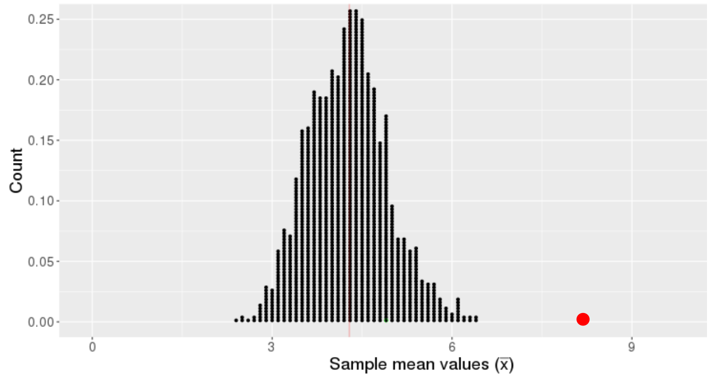
P-values

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>

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

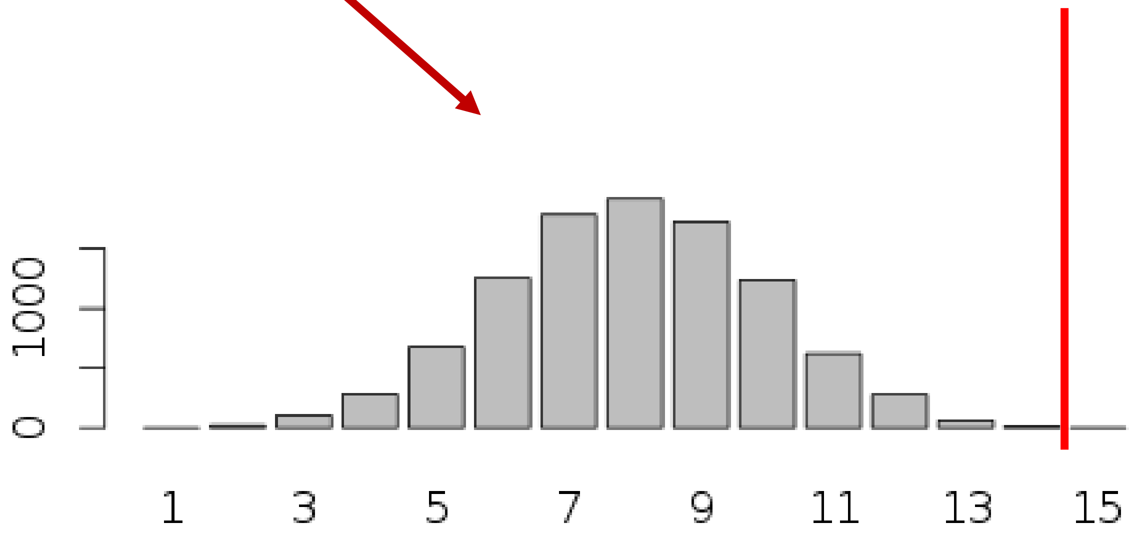
The smaller the p-value, the stronger the statistic evidence is against the null hypothesis



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, we say the sample results are **statistically significant**

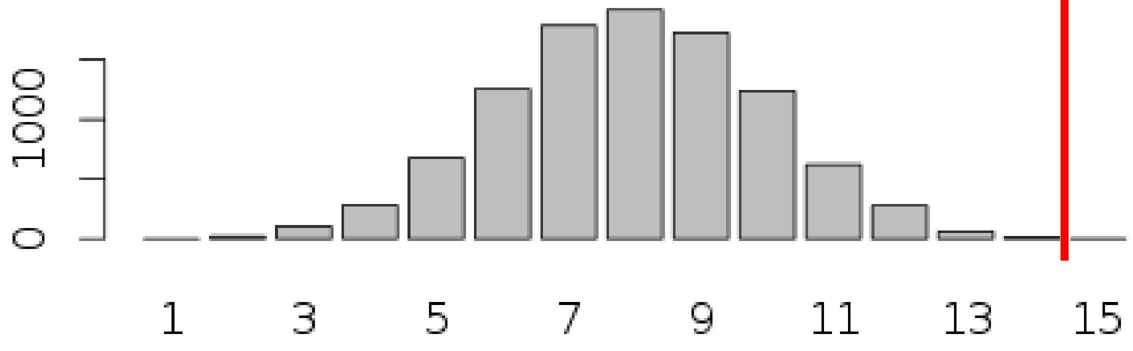
- i.e., we have a small p-value

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

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

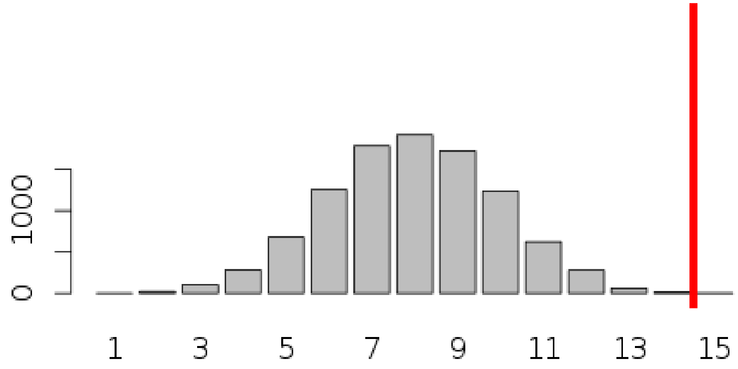


p-value = 3/10000 = 0.0003

Getting p-values using SDS100 functions

Flipping coins many times:

```
flip_simulations <- do_it(10000) * {
    rflip_count(16, 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 <- 15
p_value <- pnull(obs_stat, flip_simulations, lower.tail = FALSE)
```

Key steps hypothesis testing

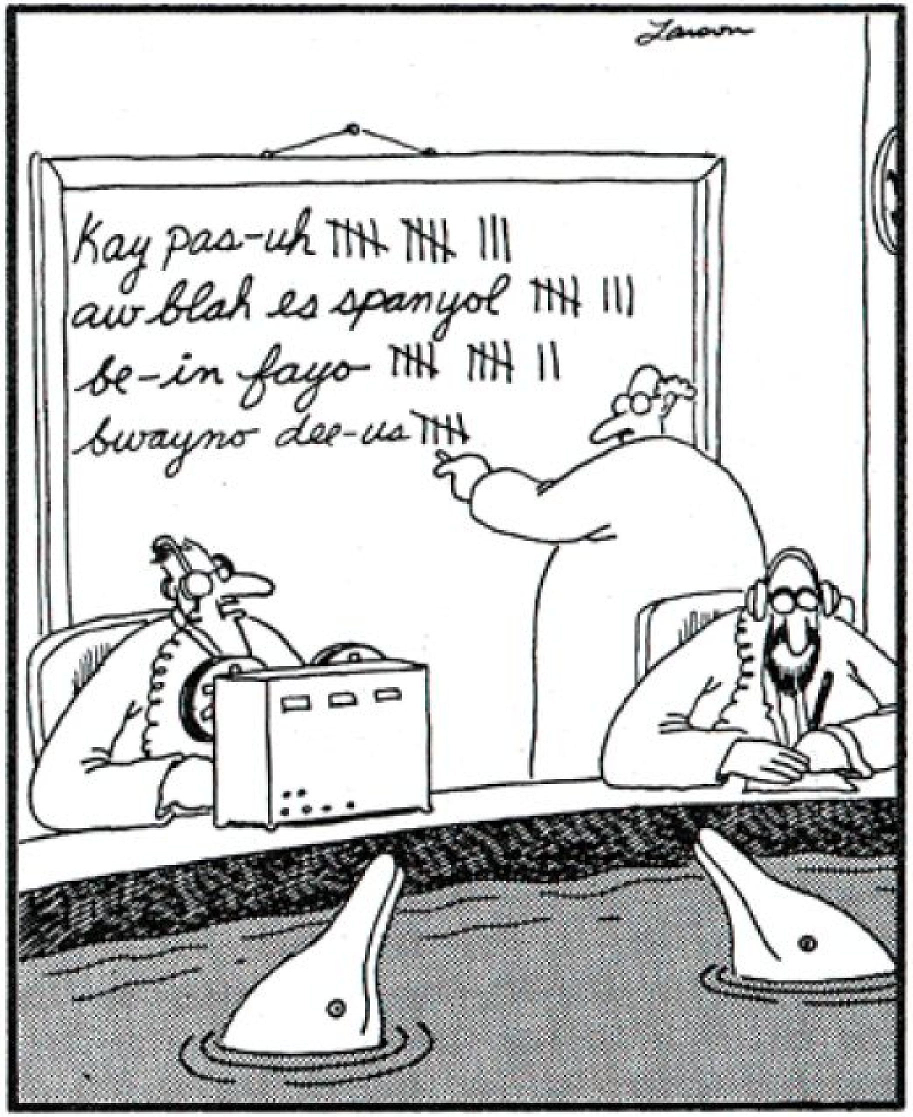
1. State the null hypothesis… and the alternative hypothesis
   - Buzz is just guessing so the results are due to chance: $H_0: \pi = 0.5$
   - Buzz is getting more correct results than expected by chance: $H_A: \pi > 0.5$

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’



"Matthews ... we're getting another one of those strange 'aw blah es span yol' sounds."