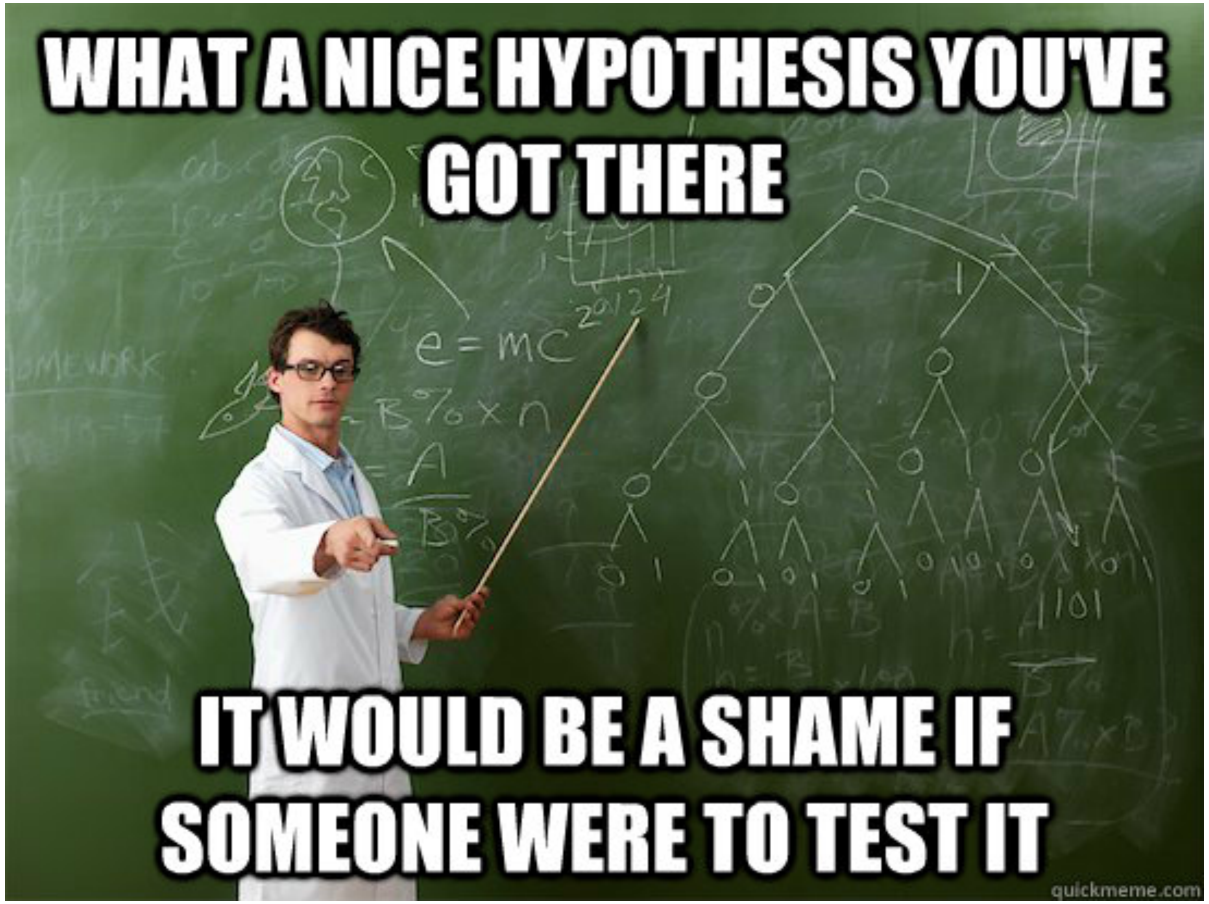
Review of hypothesis tests



Overview

Hypothesis tests for a single proportion

Framework/terminology for hypothesis testing

Hypothesis tests for a single proportion using randomization in R

Hypothesis tests for two means

Randomization tests for comparing two means in R

Overview

Assuming you are familiar with hypothesis tests from Intro Stats

- Particularly parametric hypothesis tests, such as the t-test

Quick review concepts of hypothesis test

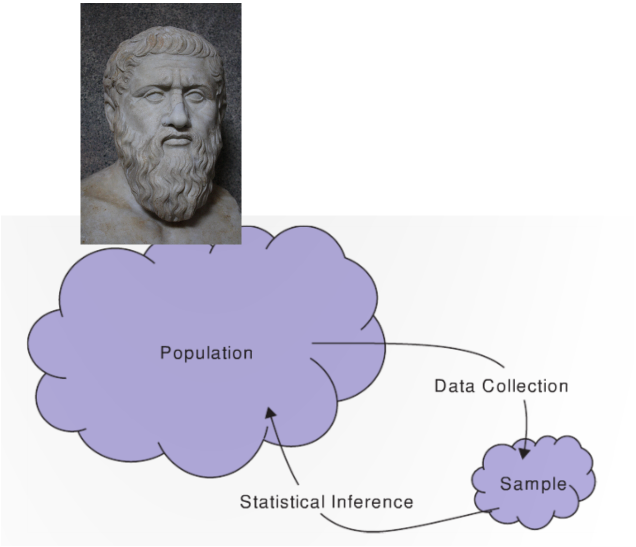
Introduce computational methods for hypothesis tests that use randomization

- These methods make fewer “assumptions” than parametric methods, so they can potentially work in more situations

t-tests	94 respondents	**80** %
confidence intervals	108 respondents	92 %
the bootstrap	18 respondents	15 %
permutation tests	18 respondents	15 %
one-way ANOVA	38 respondents	32 %
multiple regression	42 respondents	36 %
logistic regression	39 respondents	33 %
sampling distributions	66 respondents	56 %
None of the above	6 respondents	5 %
No Answer	1 respondents	1 %

Statistical tests (hypothesis test)

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



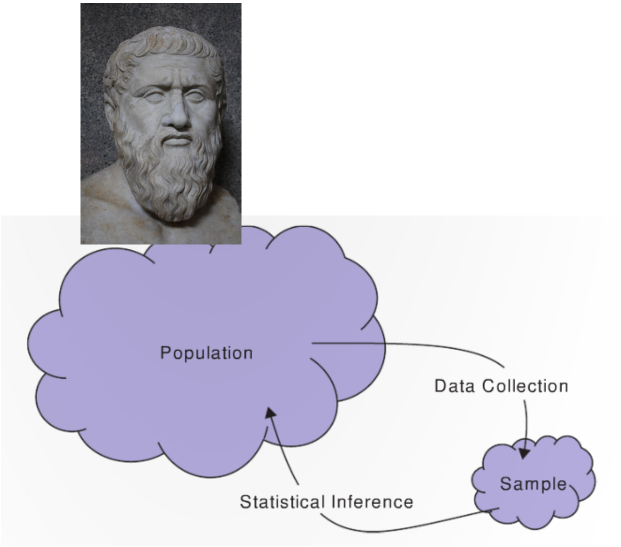
<u>Example 1</u>: we might make the claim that Trump's approval rating for all US citizens is 42%

How can we write this using symbols?

- $\pi = .42$

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?

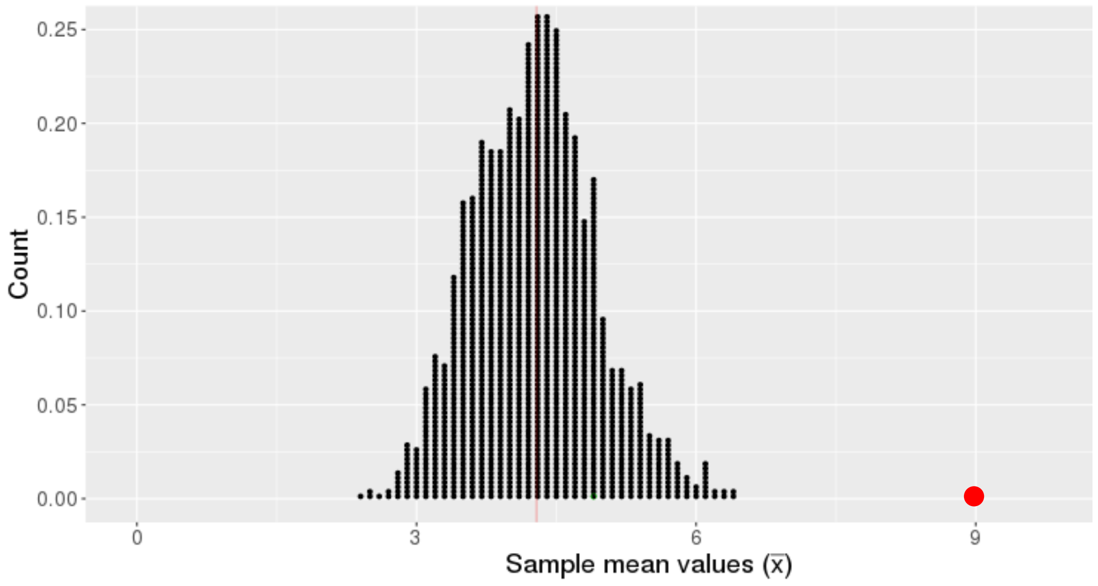
- $\mu = 72$

Basic hypothesis test logic

We start with a claim about a population parameter

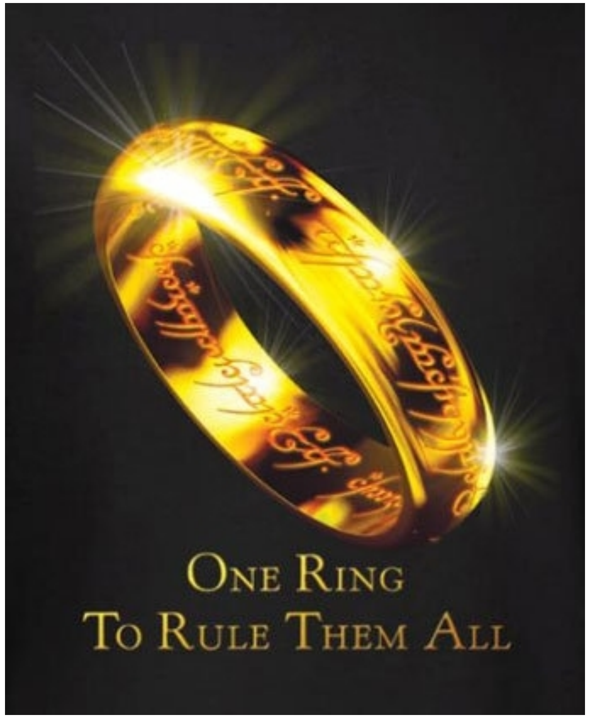
- E.g., μ =

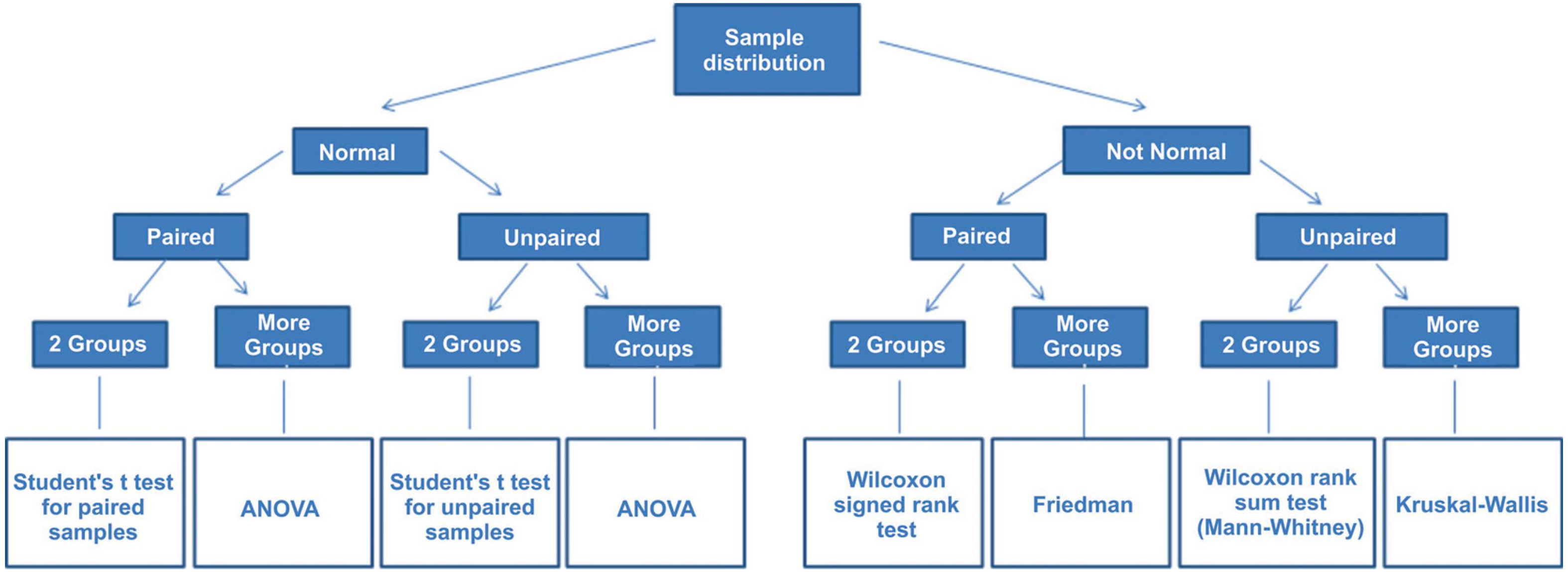
This claim implies we should get a certain distribution of statistics

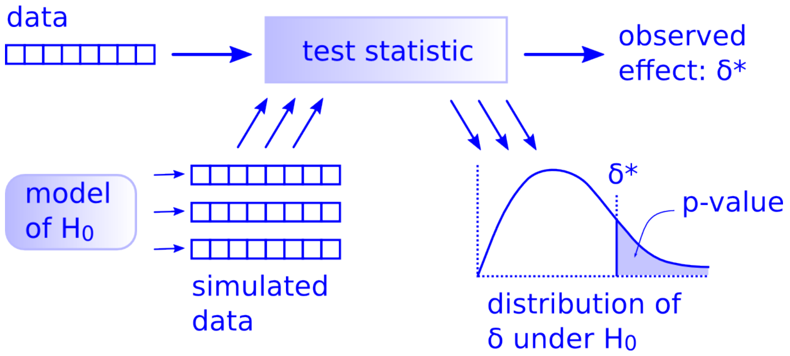


If our observed statistic is highly unlikely, we reject the claim

The big picture: There is only one hypothesis test!







Just need to follow 5 steps!

Example: Is it possible to smell whether someone has Parkinson's disease?

Joy Milne claimed to have the ability to smell whether someone had Parkinson's disease

To test this claim researchers gave Joy 6 shirts that had been worn by people who had Parkinson's disease and 6 shirts by people who did not

Joy identified 11 out of the 12 shirts correctly

Survey questions about the experiment!

1. What are the cases in this experiment?

2-3. What is the variable of interest, and is it categorical or quantitative?

4-5. What is the observed statistic - and what symbols should we use to denote it?

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

7. Do you think the results are due to chance?
    - i.e., do you think Joy got 11 correct answers by guessing?

8. Do you believe Joy can really smell whether someone has Parkinson's disease?

Pause the video and answer the questions on the survey!

Smelling Parkinson’s disease

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

$\pi = 0.5$

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

$\pi > 0.5$

Chance models

How can we assess whether 11 out of 12 correct trials ($\hat{p}$ = .916) is beyond what we would expect by chance?

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

Heads = correct guess

Tails = incorrect guess

We could flip 12 coins and see if we get 11 heads

Chance models

To really be sure, we should repeat flipping a coin 12 many times.

Any ideas how to do this?

Flipping coins in R

We can simulate coin flipping using the rbinom() function

```
flip_simulations <- rbinom(num_sims, size, prob)
```

num_sims: the number of simulations run

- Typically we do around 10,000 repeats

size: the number of trials on each simulation

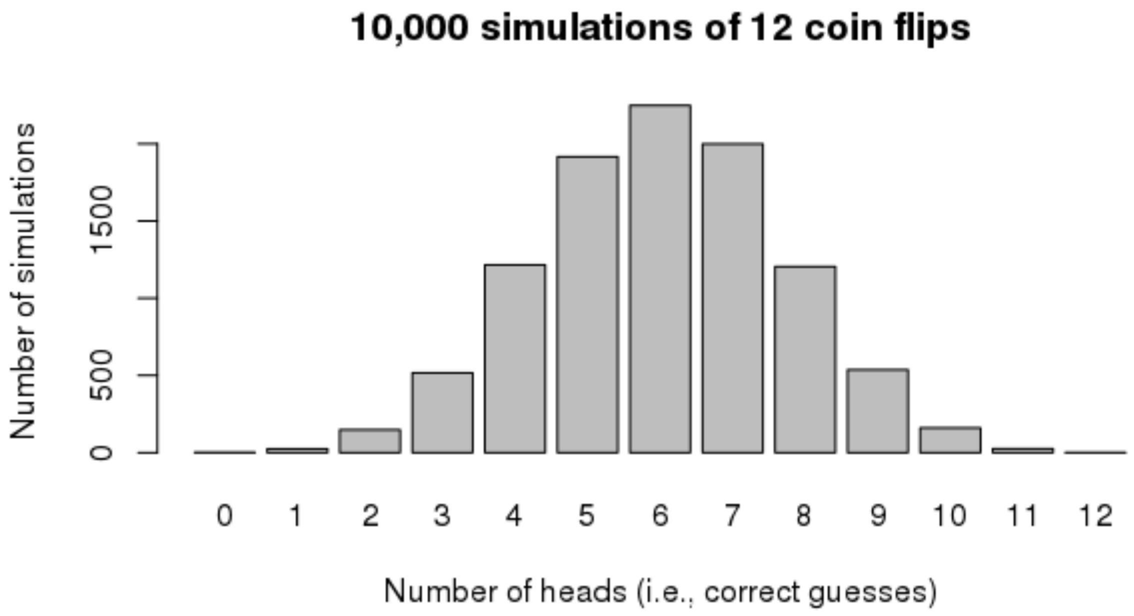
- 12 for simulating Joy's guesses

prob: the probability of success on each trial

- .5 if Joy was guessing

Simulating Flipping 12 coins 10,000

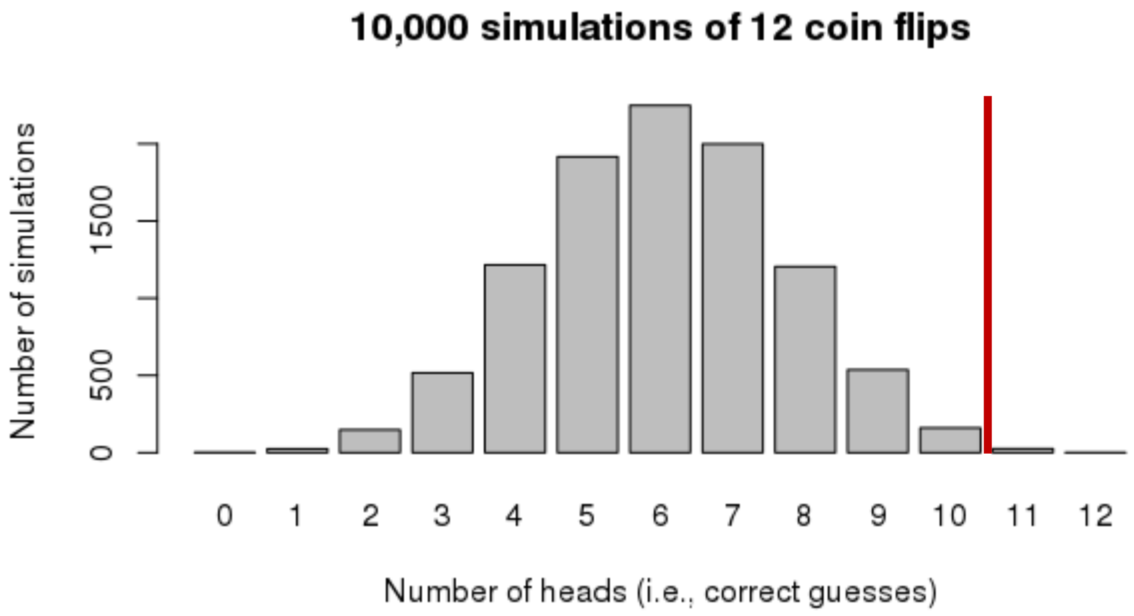
0	2
1	26
2	147
3	558
4	1269
5	1967
6	2310
7	1843
8	1142
9	537
10	162
11	33
12	4



Is it likely that Joy was guessing?

Simulating Flipping 12 coins 10,000

0	2
1	26
2	147
3	558
4	1269
5	1967
6	2310
7	1843
8	1142
9	537
10	162
11	33
12	4



Is it likely that Joy was guessing?

Do you believe Joy can really smell whether someone has Parkinson’s disease?

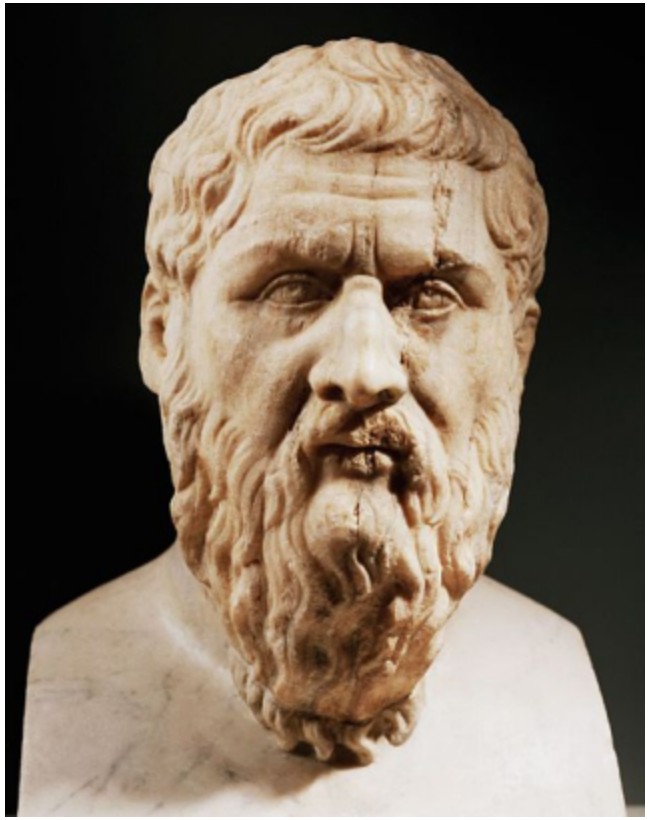
Is it possible to smell whether someone has Parkinson's disease?

We will examine this in R in a minute…

But first, let's review hypothesis testing terminology

Review of terminology and the 5 steps of hypothesis tests

Question: who is this?



A: Gorgias

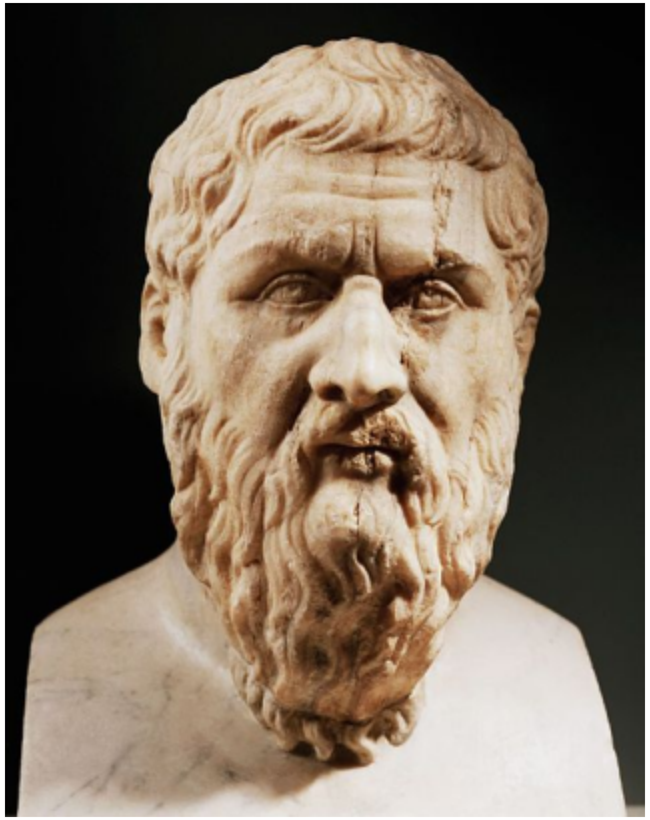
Question: Who is Gorgias?

A: a skeptic

Question: Does Gorgias believe Joy can smell Parkinson's disease?

A: No!

Question: who is this?



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

- that Joy was guessing

How can we write the null hypothesis in symbols?

H_0: $\pi = 0.5$

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

- Joy can smell Parkinson's disease

How can we write the alternative hypothesis in symbols?

H_A: $\pi > 0.5$

Question: who is this?

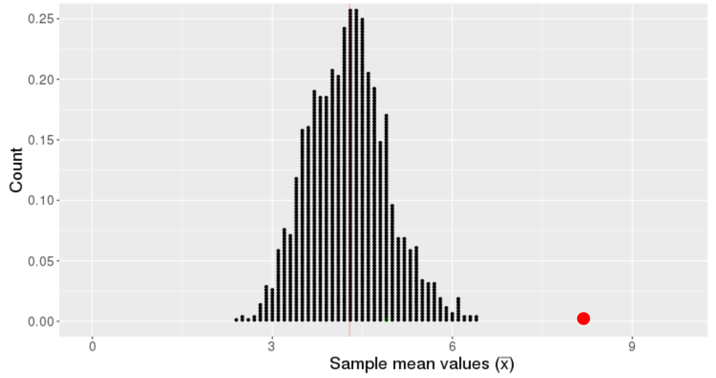
To prove Gorgias wrong, we will start by assuming he is right!

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

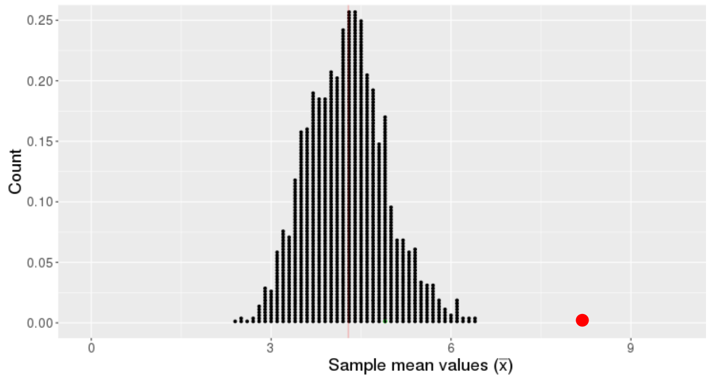


Terminology

Null Hypothesis ( ): 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 Joy smelling Parkinson's disease

1. What is the null hypothesis?

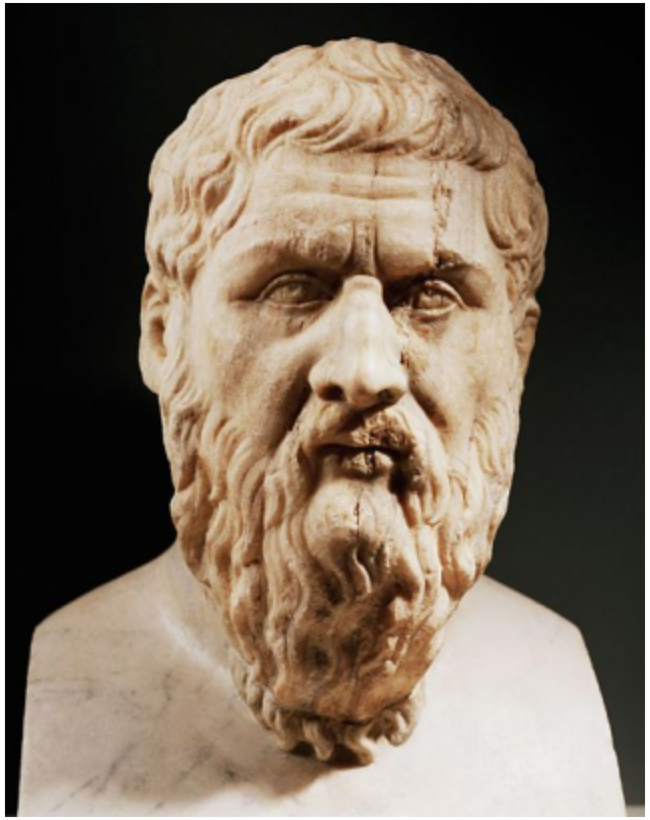
2. We can write this in terms of the population parameter as:

H_0: $\pi = 0.5$

3. What is the alternative hypothesis?

H_A: $\pi > 0.5$

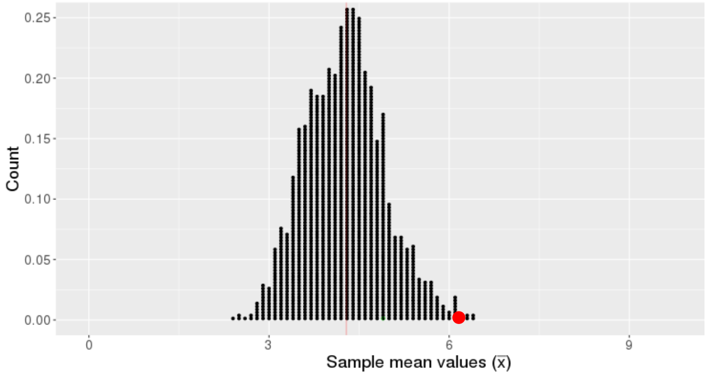
Setting the rules



<u>Life wisdom:</u> If you are going to make a bet with a nihilist, you'd better agree to the rules first!

<u>Rules</u>

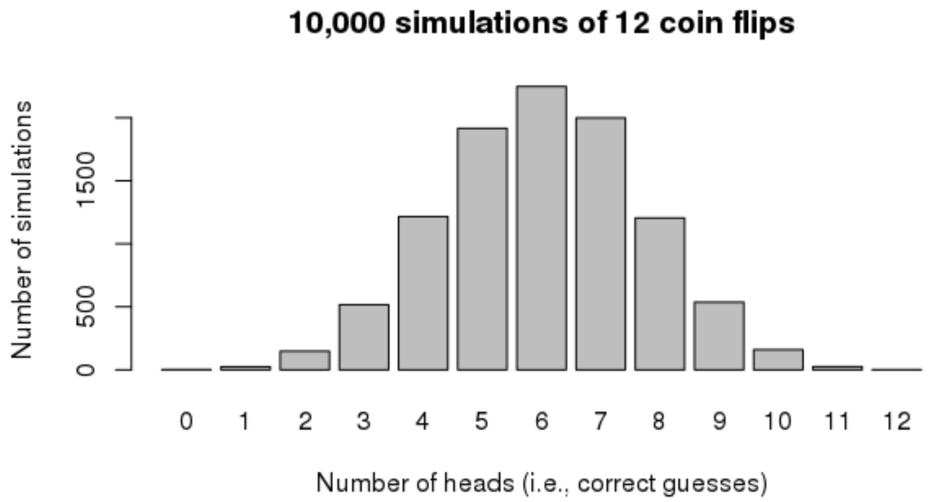
- If there is a less than 5% chance we would get a random statistic as or more extreme than the observed statistic (if H_0 is true) we will reject H_0
  - i.e., Gorgias loses the bet
- In symbols: $\alpha = 0.05$



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



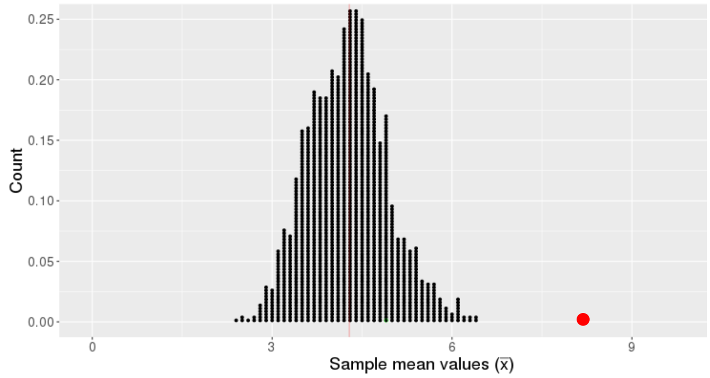
P-values

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 extreme as our observed statistic from the null distribution

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

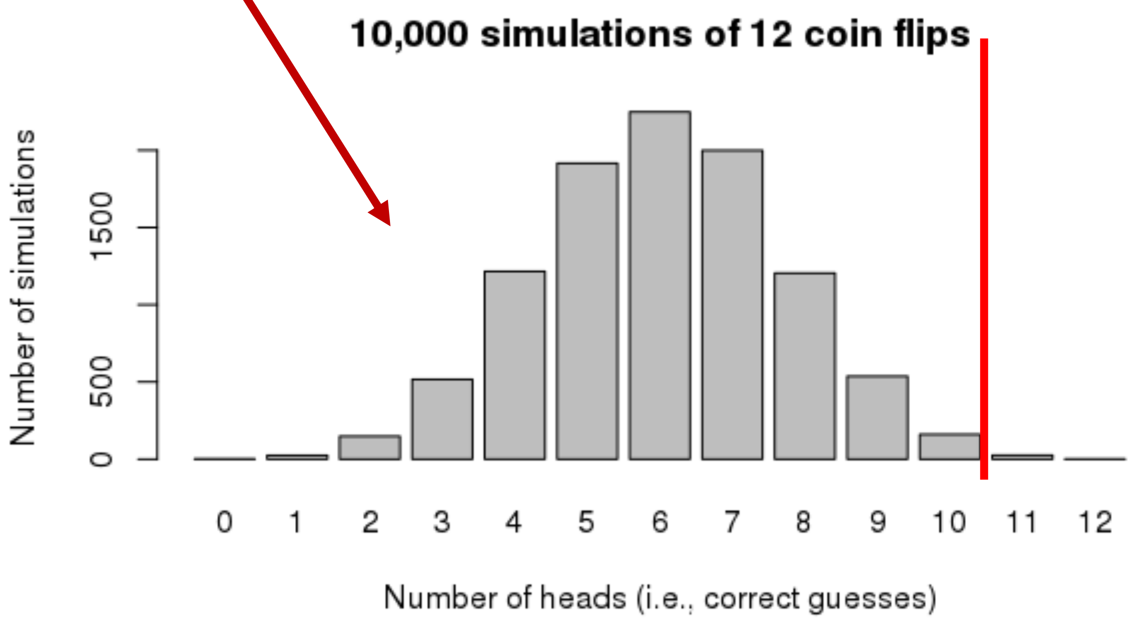
The smaller the p-value, the stronger the statistic evidence is against the null hypothesis



Joy example

0	2
1	26
2	147
3	558
4	1269
5	1967
6	2310
7	1843
8	1142
9	537
10	162
11	33
12	4

Null distribution



p-value = 37/10000 = .0037

Statistical significance

When our observed sample statistic is unlikely to co[illegible] the null distribution, people often say the results are **statistica[illegible]gnificant**

- i.e., our p-value is less than α
- i.e., Gorgias lost the bet!

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

- [The American Statistical Association rejects the phrase 'statistically significant']

5 steps for testing hypotheses

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

2. Calculate the observed statistic (and visualize the data)
   - Joy got 11 out of 12 guesses correct, or $\hat{p} = .917$

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

4. Examine how likely the observed statistic is to come from the null distribution
   - What is the probability that the Joy would guess 11 or more correct?
   - i.e., what is the p-value

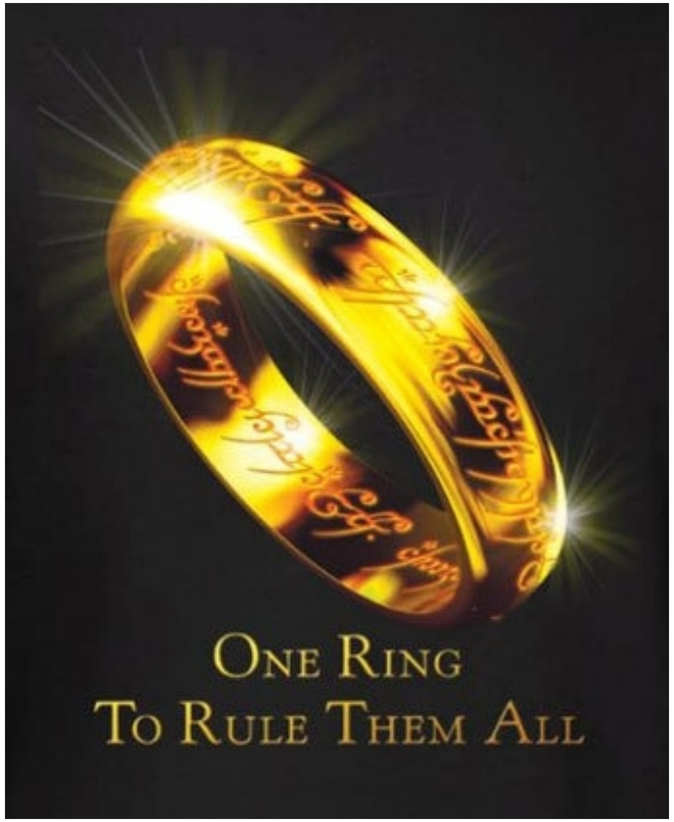
5. Make a judgement
   - If we have a small p-value, this means that $\pi = .5$ is unlikely and so $\pi > .5$
   - i.e., we could say our results are ‘statistically significant’

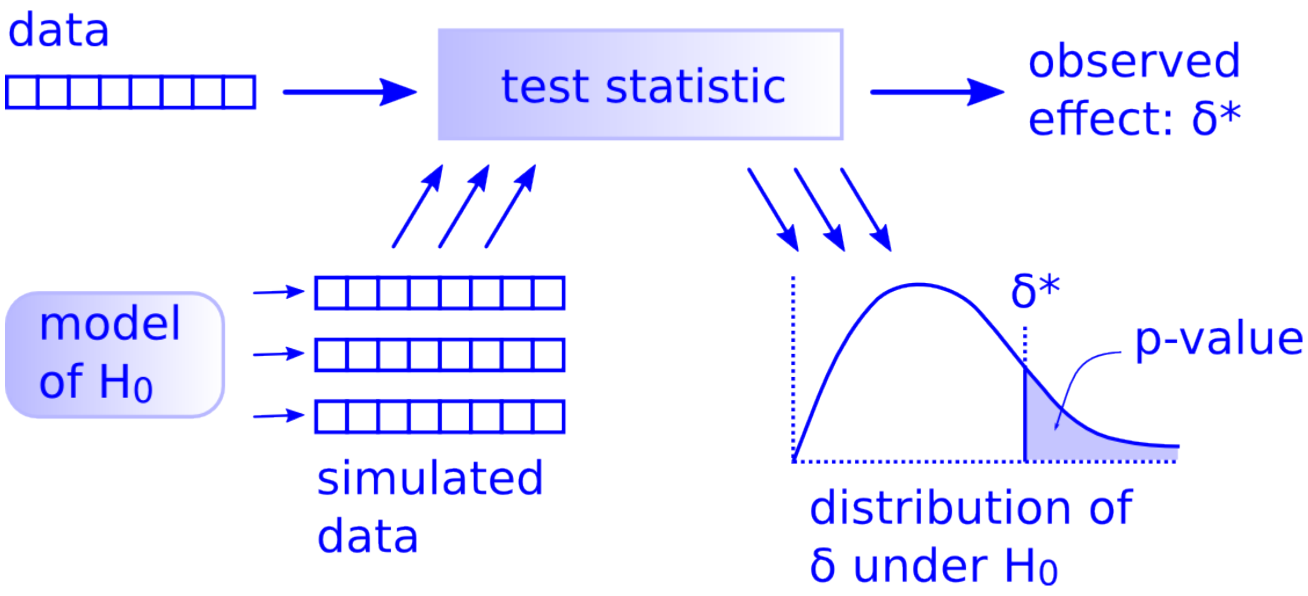
Is it possible to smell whether someone has Parkinson's disease?

Let's examine this in R!

Hypothesis tests comparing 2 means

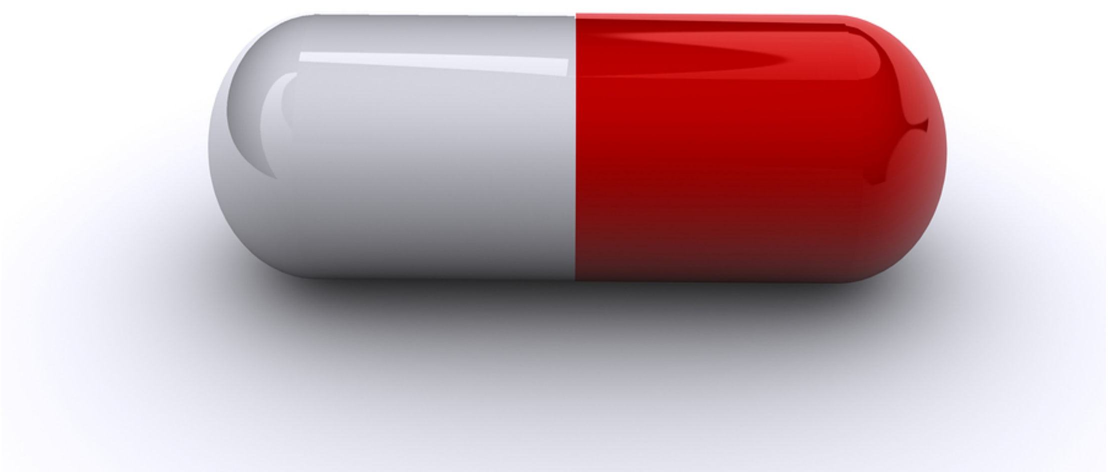
The big picture: There is only one hypothesis test!





Just need to follow 5 steps!

Hypothesis tests for comparing two means



Question: Is this pill effective?

Testing whether a pill is effective

How would we design a study?

What would the cases and variables be?

What would the parameter and statistic of interest be?

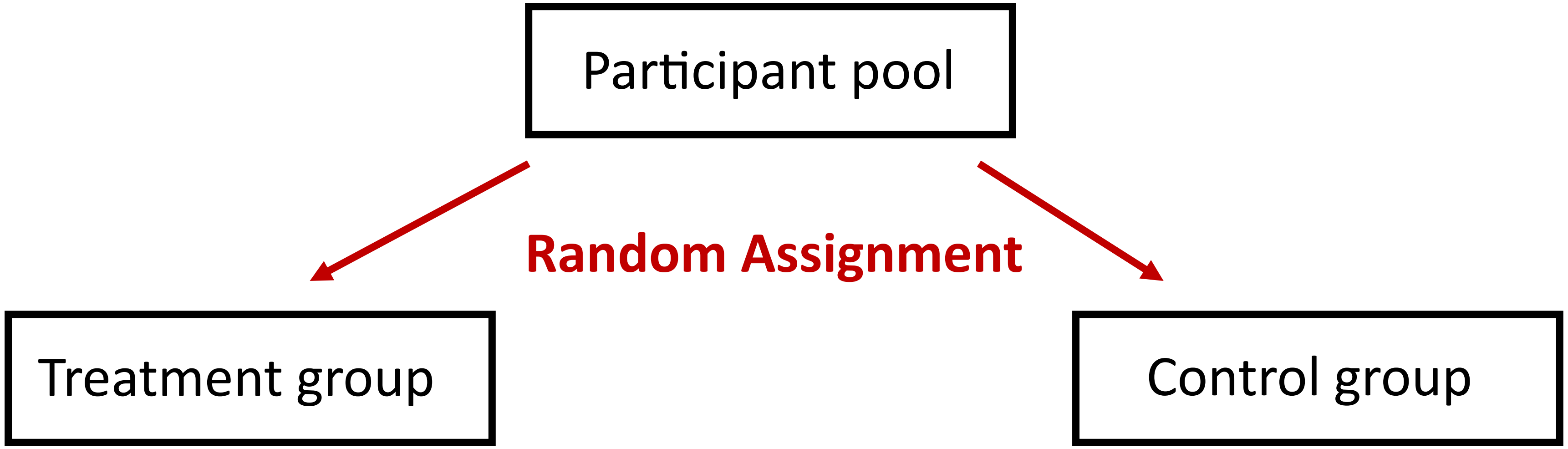
What are the null and alternative hypotheses?

- Assume we are looking for differences in means between the groups

Experimental design

Take a group of participant and ***randomly assign***:

- Half to a *treatment group* where they get the pill
- Half in a *control group* where they get a fake pill (placebo)
- See if there is more improvement in the treatment group compared to the control group



Observational and experimental studies

An **experiment** is a study in which the researcher actively controls one or more of the explanatory variables

- Allows one to get at questions of **causation**!

An **observational study** is a study in which the researcher does not actively control the value of any variable but simply observes the values as they naturally exist

Survey Question: Are the smelling Parkinson's disease and/or drug studies experimental or observational?

Hypothesis tests for differences in two group means

1. State the null and alternative hypothesis
   - H_0: $\mu_{Treatment} = \mu_{Control}$ or $\mu_{Treatment} - \mu_{Control} = 0$
   - H_A: $\mu_{Treatment} > \mu_{Control}$ or $\mu_{Treatment} - \mu_{Control} > 0$

2. Calculate statistic of interest
   - $\bar{x}_{Effect} = \bar{x}_{Treatment} - \bar{x}_{Control}$

Example: Does calcium reduce blood pressure?

A randomized comparative experiment by Lyle et al (1987) investigated whether calcium lowered blood pressure.

- A treatment group of 10 men received a calcium supplement for 12 weeks
- A control group of 11 men received a placebo during the same period

The blood pressure of these men was taken before and after the 12 weeks of the study

1. What are the null and alternative hypotheses?
   - H_0: $\mu_{Treatment} = \mu_{Control}$ or $\mu_{Treatment} - \mu_{Control} = 0$
   - H_A: $\mu_{Treatment} > \mu_{Control}$ or $\mu_{Treatment} - \mu_{Control} > 0$
     - i.e., a <u>greater decrease</u> in blood pressure after taking calcium

Does calcium reduce blood pressure?

Treatment data (n = 10):

Begin	107	110	123	129	112	111	107	112	136	102
End	100	114	105	112	115	116	106	102	125	104
Decrease	**7**	**-4**	**18**	**17**	**-3**	**-5**	**1**	**10**	**11**	**-2**

Control data (n = 11):

Begin	123	109	112	102	98	114	119	112	110	117	130
End	124	97	113	105	95	119	114	114	121	118	133
Decrease	**-1**	**12**	**-1**	**-3**	**3**	**-5**	**5**	**2**	**-11**	**-1**	**-3**

2. What is the observed statistic of interest?
   - $\overline{x}_{\text{Effect}} = 5 - -.2727 = 5.273$

3. What is step 3?

3. Create the null distribution!

How could we create the null distribution?

Need to generate data consistent with H_0: $\mu_{Treatment} - \mu_{Control} = 0$

- i.e., we need fake $\bar{x}_{Effect}$ that are consistent with H_0

Any ideas how we could do this?

3. Create the null distribution!

Treatment group

Control group

Reconstructed participant pool data under H_0

Shuffle data for random assignment consistent with H_0

Shuffled 'treatment group'

Shuffled 'control group'

One null distribution statistic: $\bar{x}_{Shuff\_Treatment} - \bar{x}_{Shuff\_control}$

3. Create a null distribution

1. Combine data from both groups

2. Shuffle data

3. Randomly select 10 points to be the ‘null’ treatment group

4. Take the remaining 11 points to the ‘null’ control group

5. Compute the statistic of interest on these ‘null’ groups

6. Repeat 10,000 times to get a null distribution

Let's try the rest of the hypothesis test in R…