

Hypothesis testing for leakage assessment in side channel analysis

Making decisions is easy, making the right decision less so

Ileana Buhan, June 2023

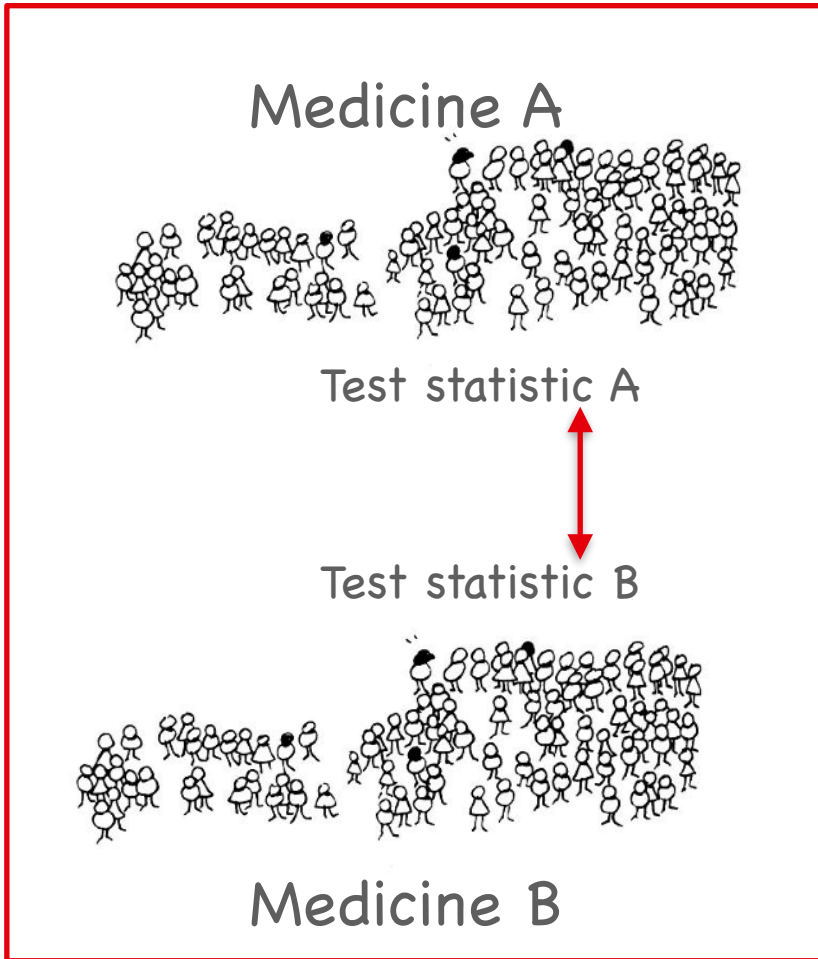
@ileanabuhan



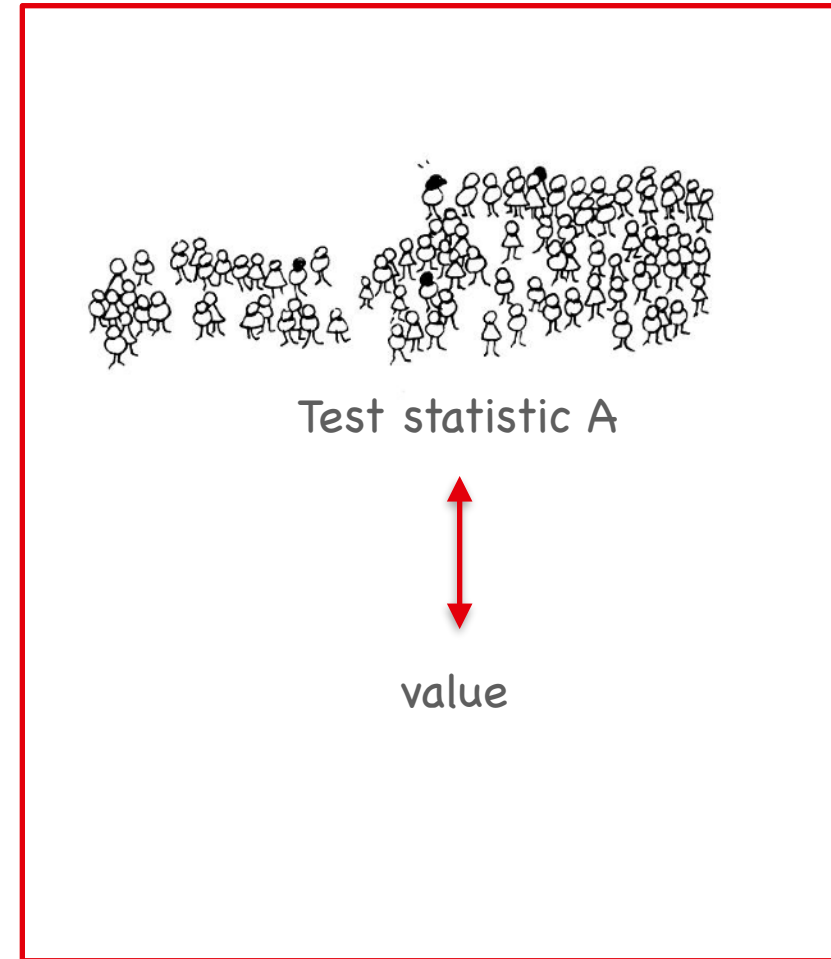
Radboud
University

Null Hypothesis Significance Testing (NHST)

Two types of questions



Two-sample test



One-sample test

What is hypothesis?

A **hypothesis*** is tentative assumption made in order to draw out and test its logical or empirical consequences.

*Source for definition <https://www.merriam-webster.com/dictionary/hypothesis>



Food for brain

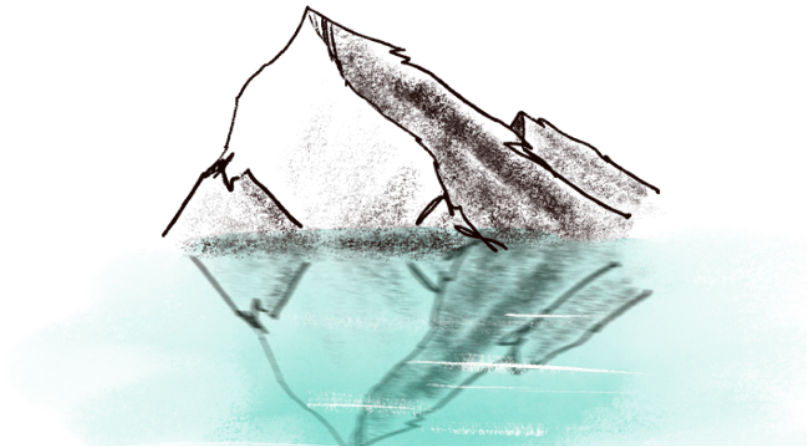
Which of the following statements are hypotheses?

1. The average height of all PhD students is 1.93m;
2. Gas in NL is expensive;
3. Green trolls have an average height of 54cm;
4. The ratio of left- to right handed people in Croatia is equal;
5. Eminem would make a better president for the USA than Justin Bieber;

Answer: 1 and 4.

Population vs sample data

The average concentration of salt for the water in the lake is 3%.



Population

μ, σ

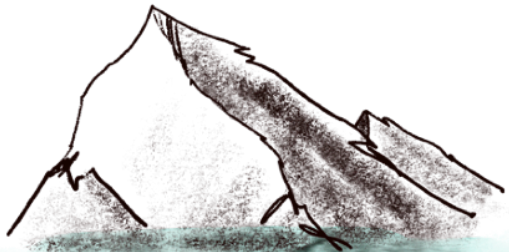


Sample data

\bar{x}, s

Population vs sample data

The average concentration of salt for the water in the lake is 3%.



Is this question:

- (A) Two-sample test
- (B) One-sample test

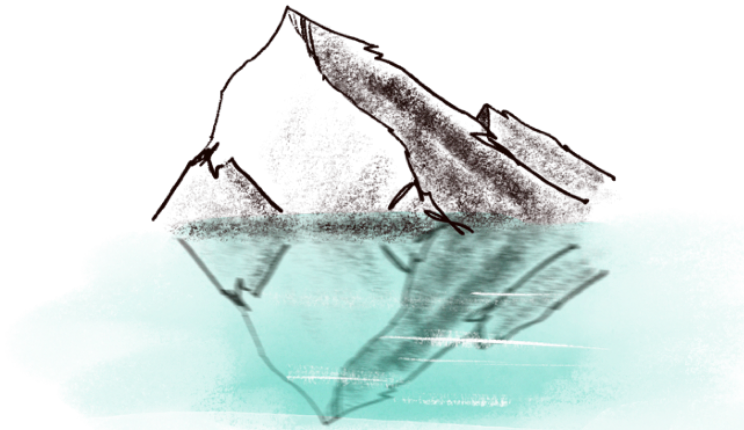
Answer: the B

Population

μ, σ

What is hypothesis testing?

A tool for making decisions about **a population** (lake) given some **sample data** (glass of water).



Hypothesis test

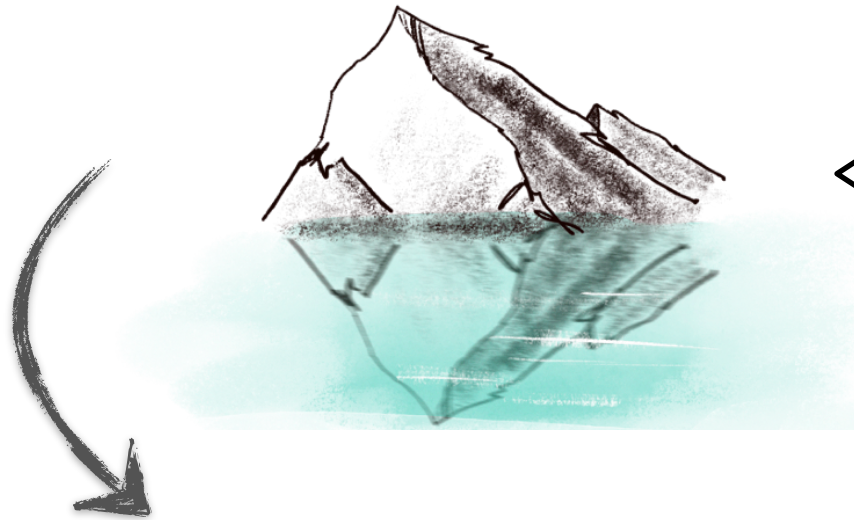


A **hypothesis test** evaluates **two mutually exclusive statements** about a **population** to determine which statement is best supported by the **sample data**.

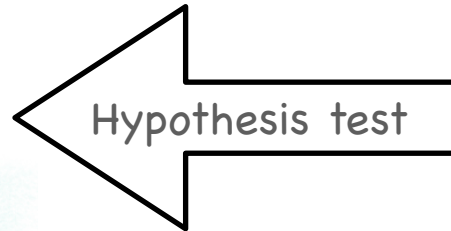
Test statistic

A **test statistic** is a number calculated from **sample data** that is used to evaluate how compatible the experimental results are with the hypothesis test.

population parameter



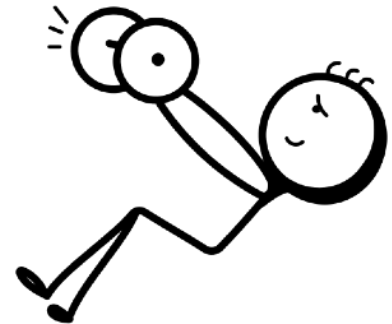
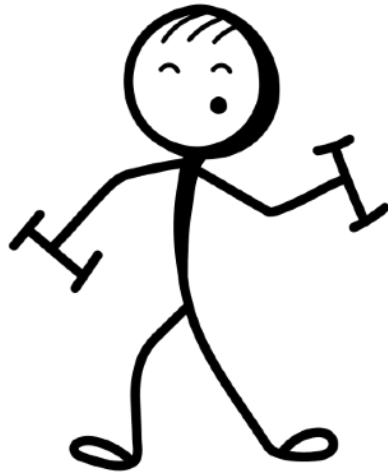
The average concentration of salt for the water in the lake is 3%.



test statistic

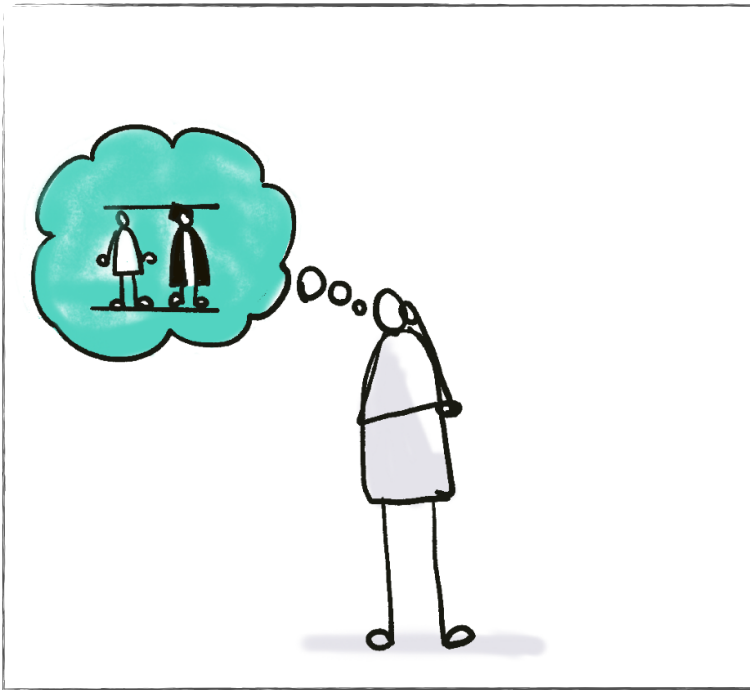
The average concentration of salt for the water in our glass is 2.7%.

Exercise 1 (1,2,3)

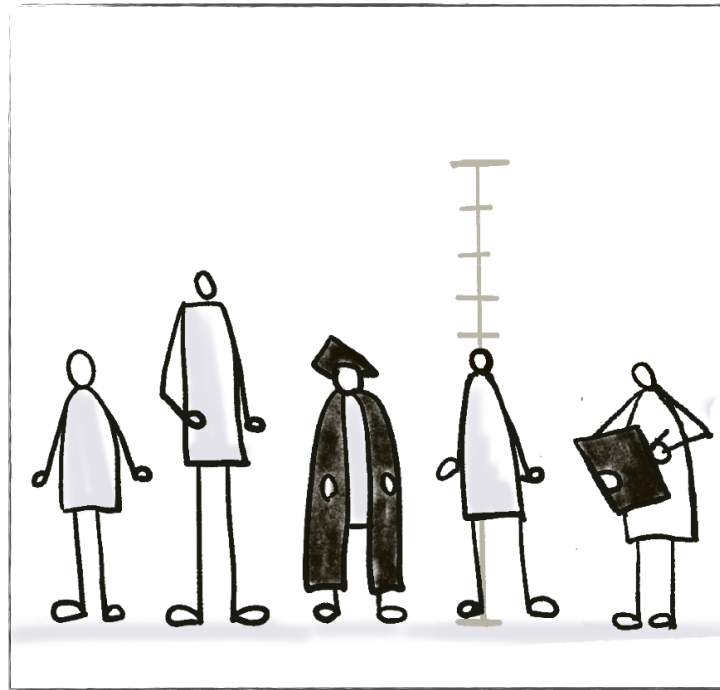


NHST in three steps

1. Select the H_0 , α ;



2. Collect data



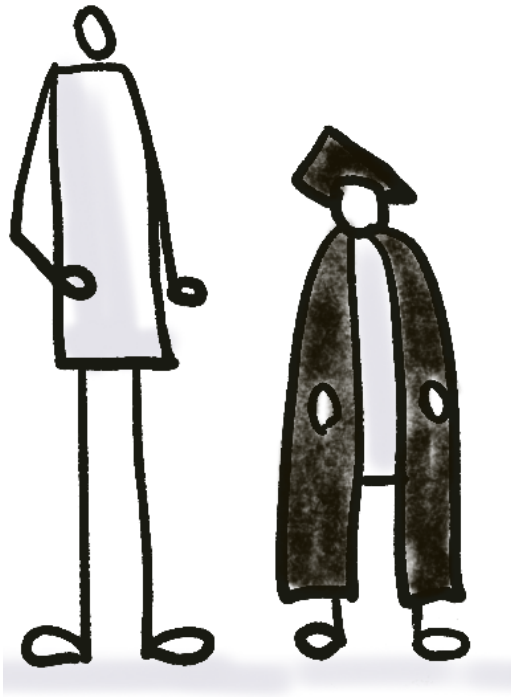
3. Test



Hypothesis testing quantifies how unusual the data is, assuming the null hypothesis to be true.

Lets test!

Are PhD students taller than faculty staff?



Is this question:

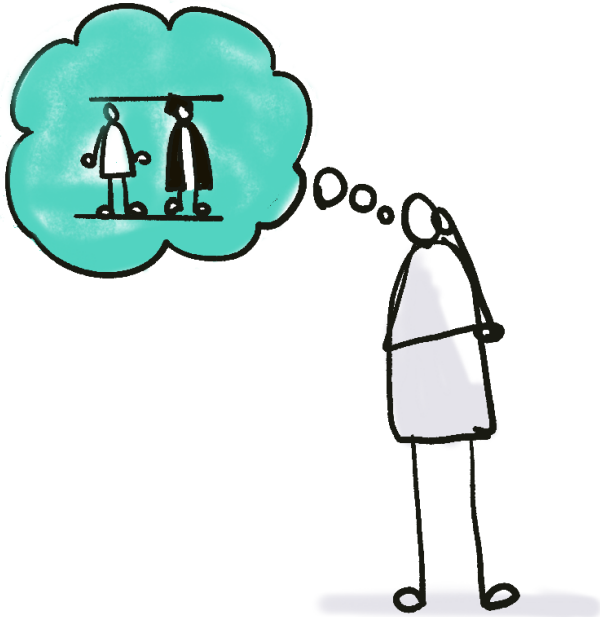
- (A) Two-sample test
- (B) One-sample test

Answer: (A)

We need **two mutually exclusive statements** for testing:

1. the **null-hypothesis** H_0
2. the alternative hypothesis H_a

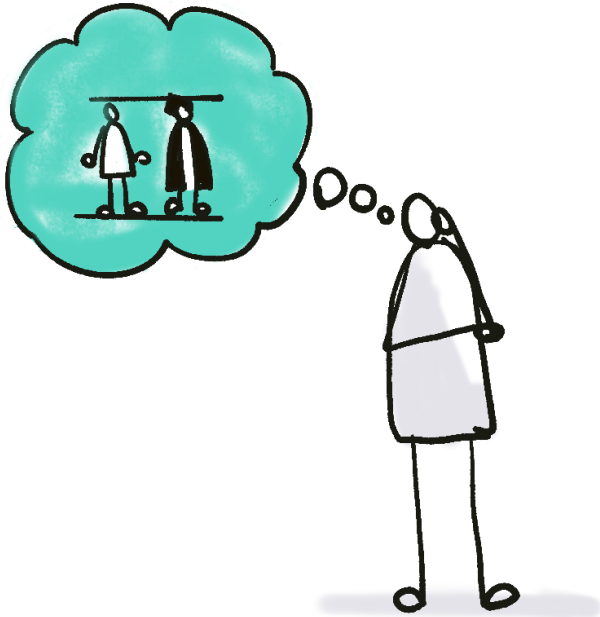
1. Select the null



The **null hypothesis (H_0)** is a specific statement about a population parameter generated **by the researcher** for the purpose of an argument.

A good null-hypothesis is interesting **to reject** and must **be specific**.

Which statement is a good choice for H_0 ?



1. The polio vaccine has no effect on the probability of developing paralytic polio;
2. Adding free gifts does not increase sales;
3. The power consumption of this device does not depend of the processed data;
4. The ratio of left -, right - handed people is equal in the population;

Answer: all.

The null vs the alternative

The two hypothesis are NOT equal. The only hypothesis tested with the data is H_0 .

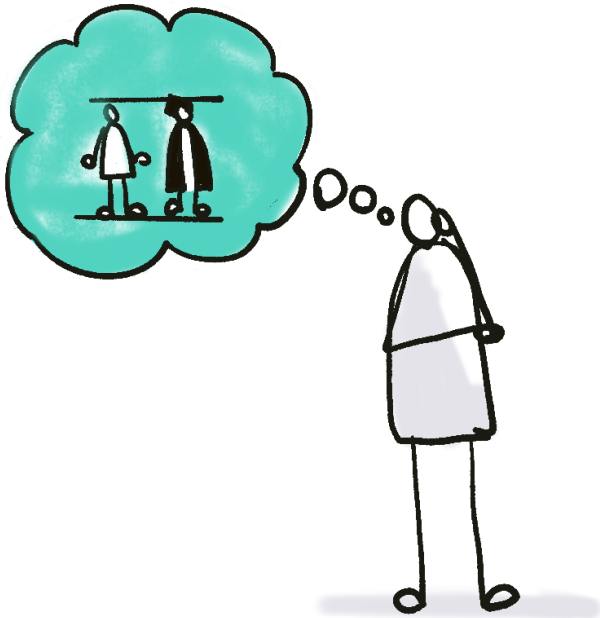


A good null-hypothesis is interesting
to reject and must be specific.



The alternative-hypothesis is not-specific
and contains all other values.

Which statement is more suited as H_0 or H_a :

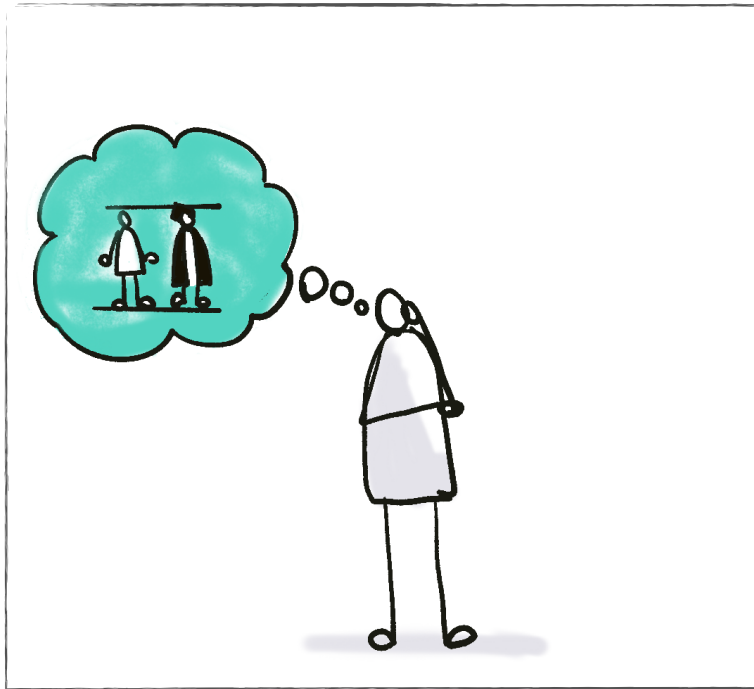


1. The number of hours preschool children spend watching TV affects how they behave at daycare.
2. Social skills influence the number of friends a person has.
3. Smoking influences the risk of allergies.
4. Growth rate of trees are unaffected by increases in carbondioxide levels in the atmosphere.

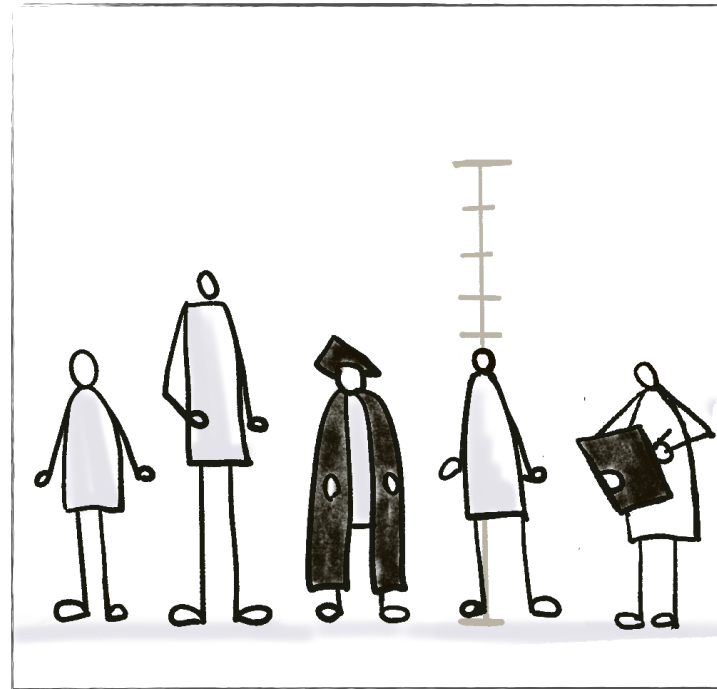
Answer: 1, 2, 3 H_a and 4, H_0 .

NHST in three steps

1. Select the H_0 , α ;



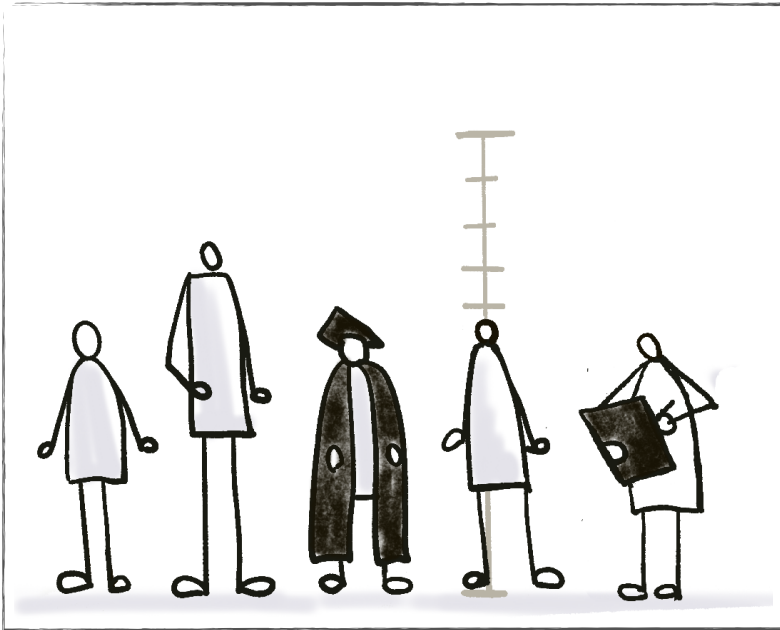
2. Sample data



3. Test



2. Sample data



The sample data must be representative for population.

Known techniques:

- Random sampling
- Counting off
- Convenience sampling
-

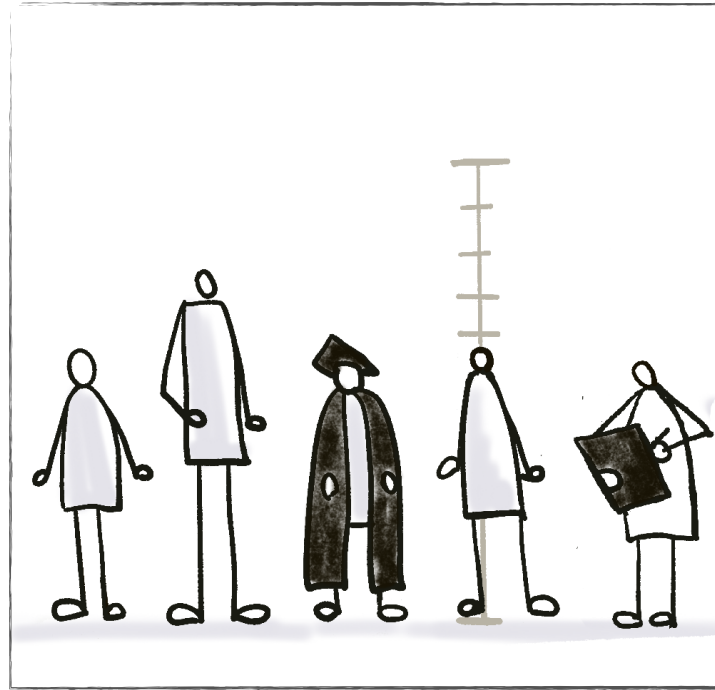
We will revisit the topic when discussing side-channel traces.

NHST in three steps

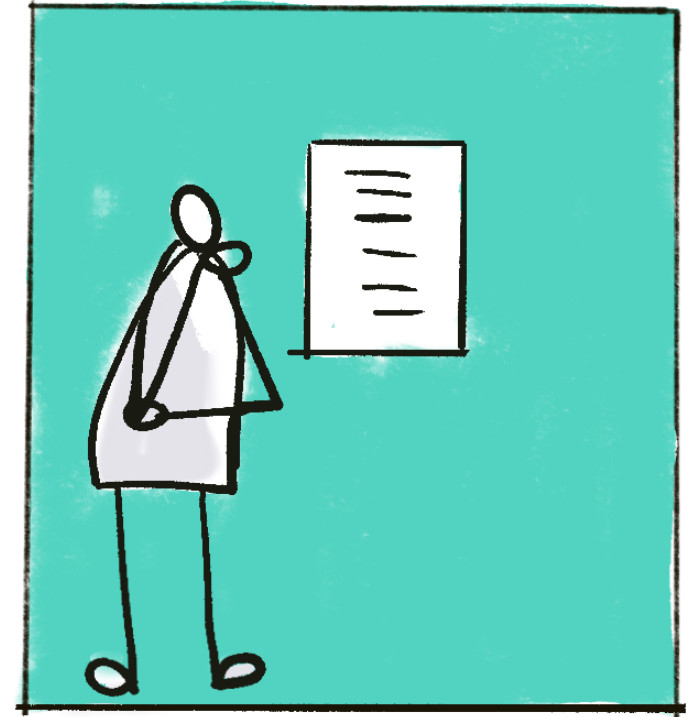
1. Select the null



2. Sample data



3. Test



3. Test the significance



null-distribution

How do we describe the universe where the null-hypothesis is true?

In a universe where the null-hypothesis is true, how surprised are we by the observed data?

How do we measure surprise?

p-values

Significance level α

The null-distribution

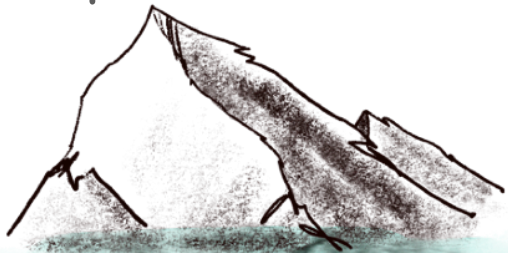


The **null-distribution** is the sampling distribution of the outcomes for a test statistic under the assumption that the null-hypothesis is true.

The null-distribution

The null-distribution is the sampling distribution of **the outcomes** for **a test statistic** under the assumption that the **null hypothesis is true**.

Example 1:*



The average concentration of salt for the water in the lake is 3%.

$$H_0 : \mu_{salt} = 0.3$$

$$\bar{x}_1 = 0.23$$



$$\bar{x}_2 = 0.2$$



$$\bar{x}_3 = 0.19$$



$$\bar{x}_4 = 0.35$$



$$\bar{x}_5 = 0.35$$

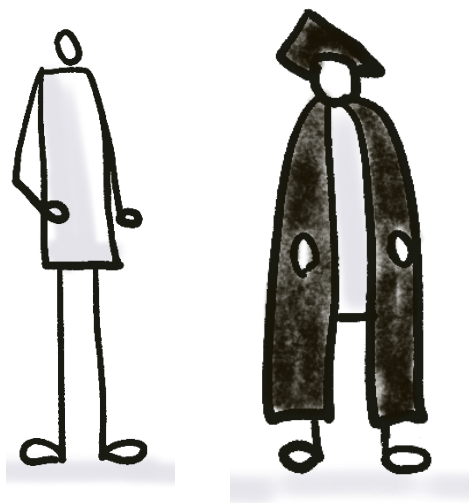


*one-sample t-test

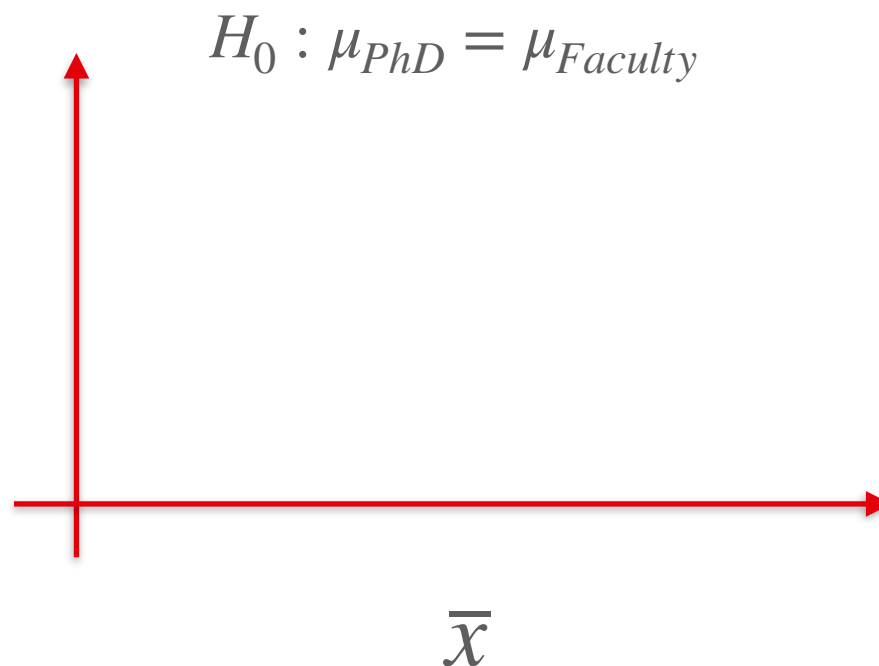
The null-distribution

The null-distribution is the sampling distribution of **the outcomes** for **a test statistic** under the assumption that the **null hypothesis is true**.

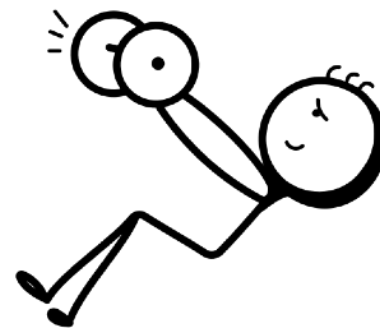
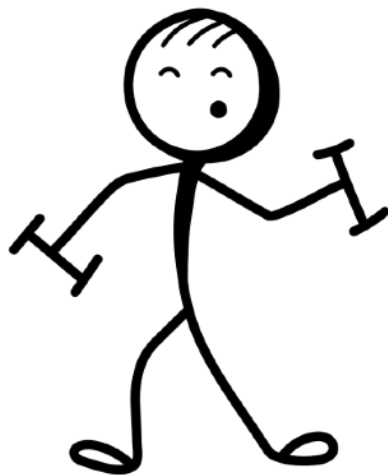
Example 2*



*two-sample t-test



Exercise 2



T-score

$$t = \frac{\bar{x} - \mu}{\sqrt{\frac{s}{n-1}}}$$

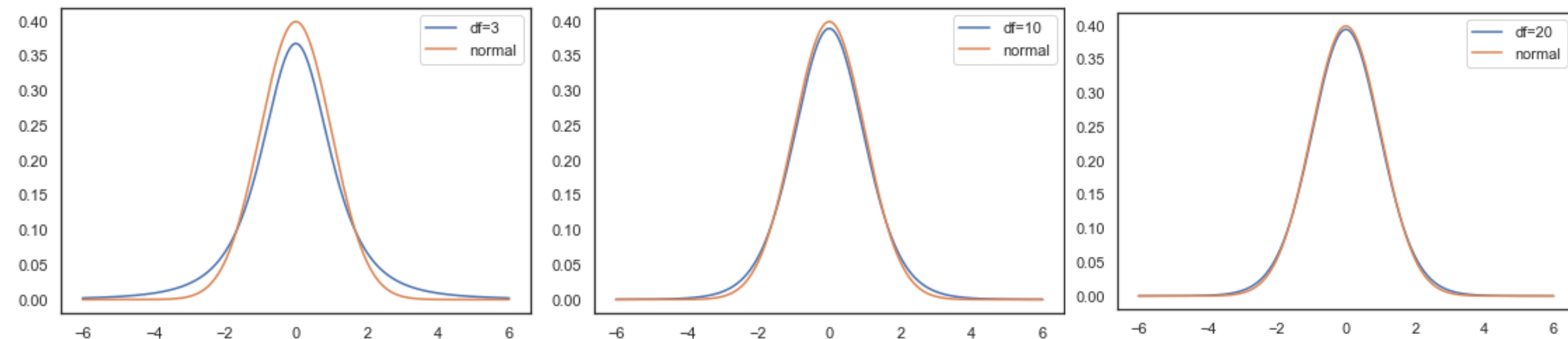
\bar{x}, μ close

\bar{x}, μ different



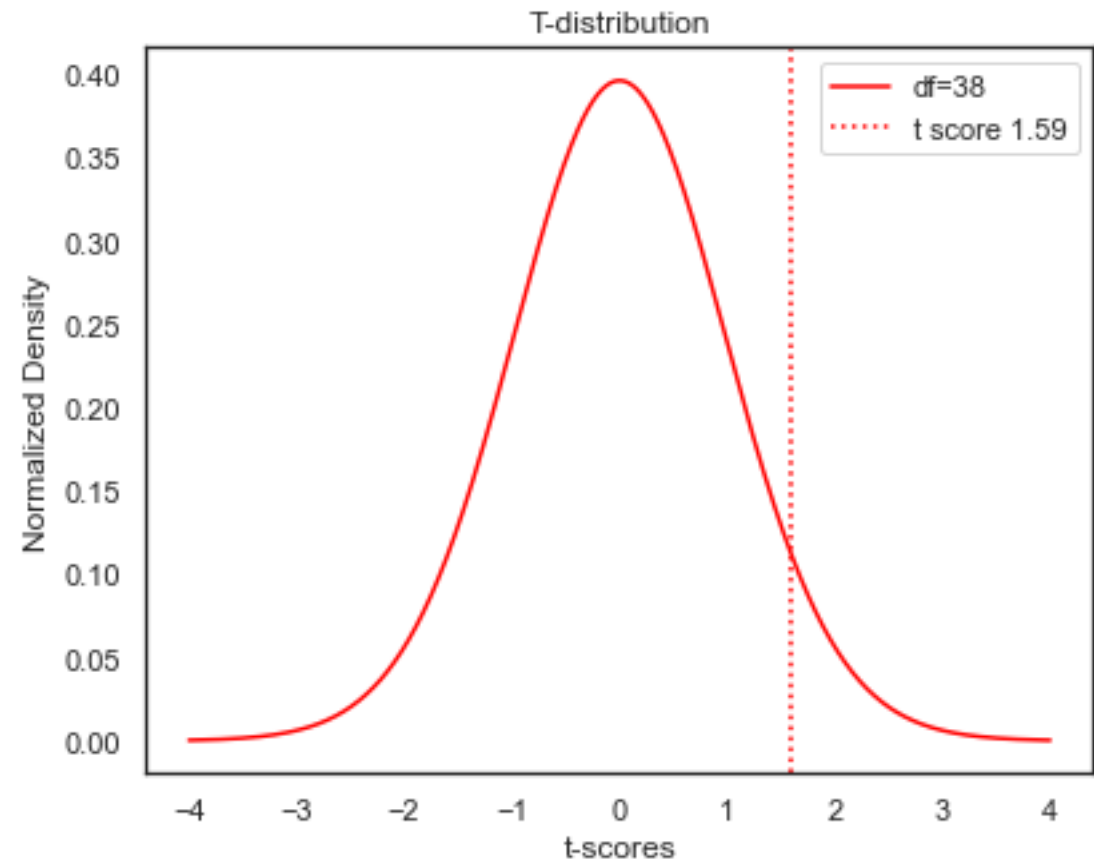
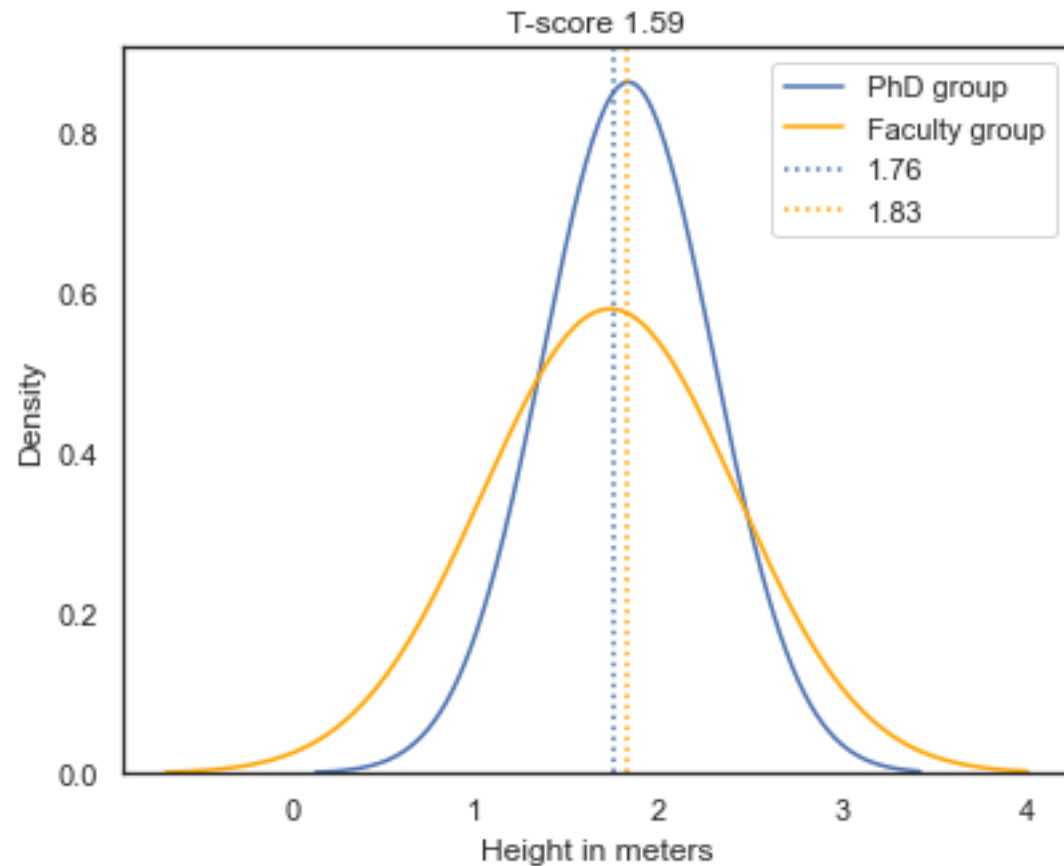
2.2 Normal vs t-distribution

Point a)

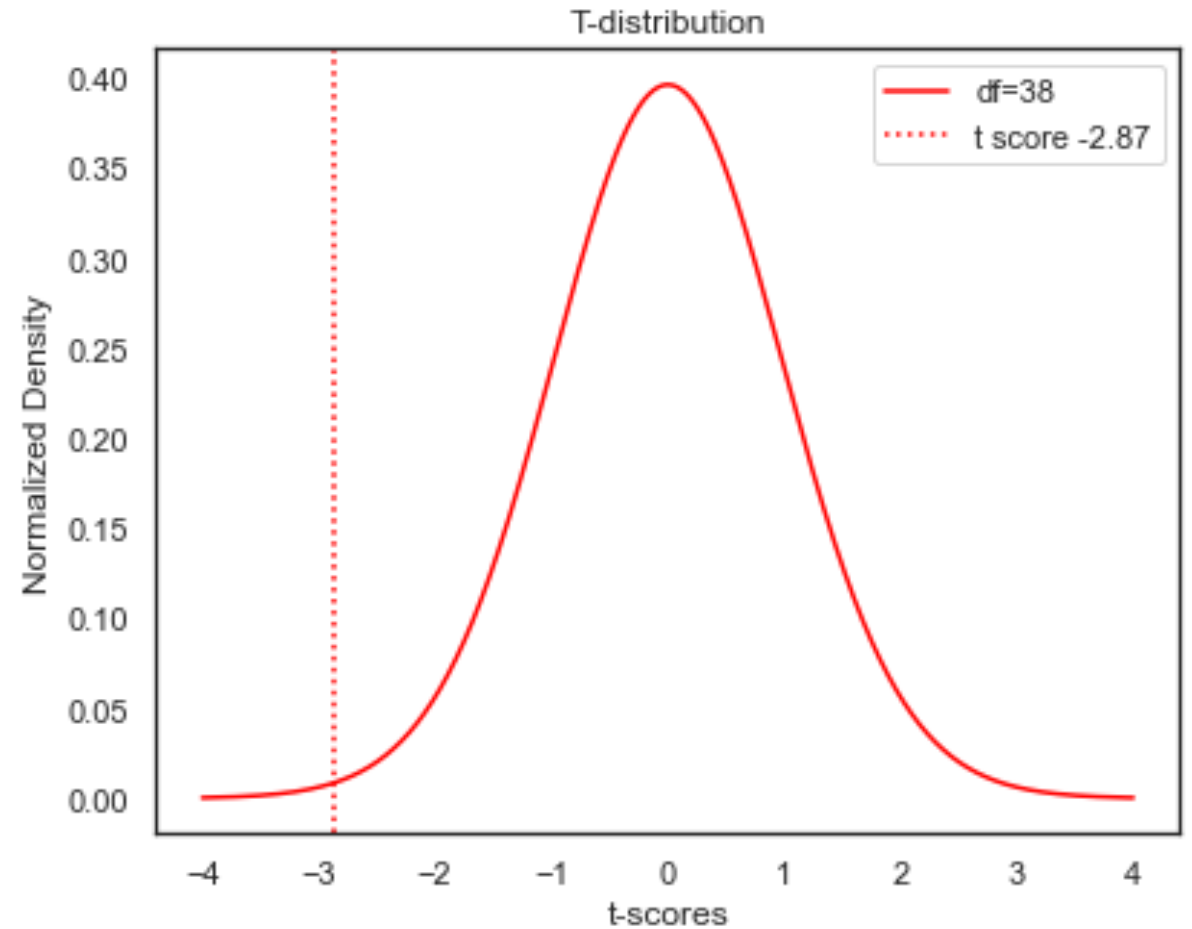
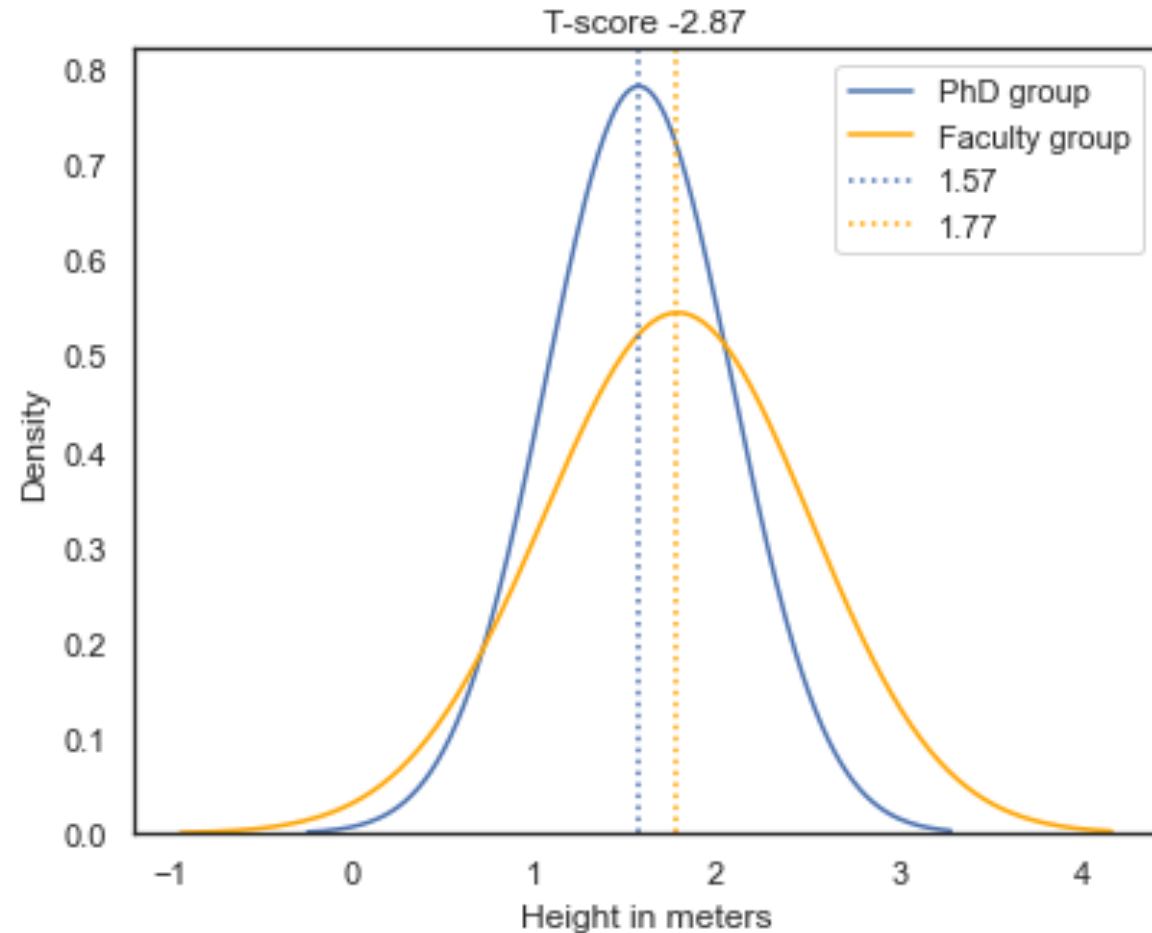


The t-distribution is heavier in the tail, but as the size sample increases (~ 30) it gets closer to a normal distribution.

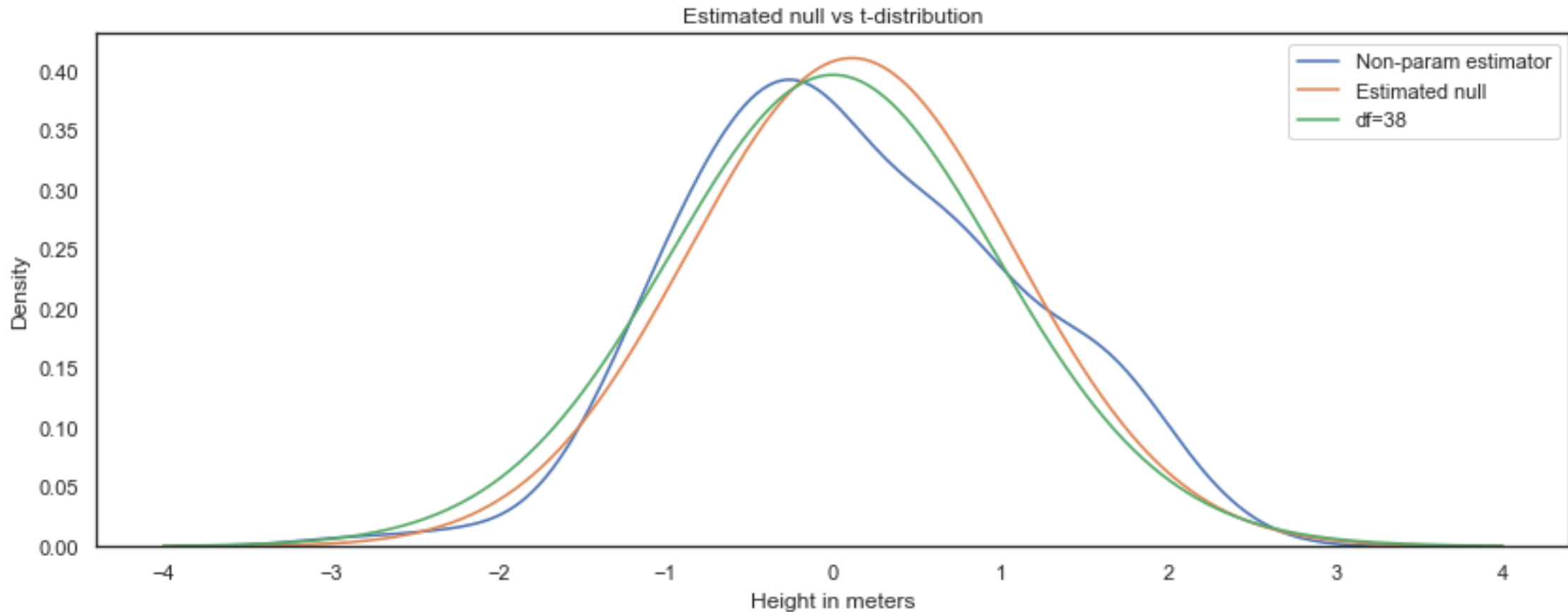
2.3 Two-sample mean test statistic



2.3 Two-sample mean test statistic



2.3 Two-sample mean test statistic

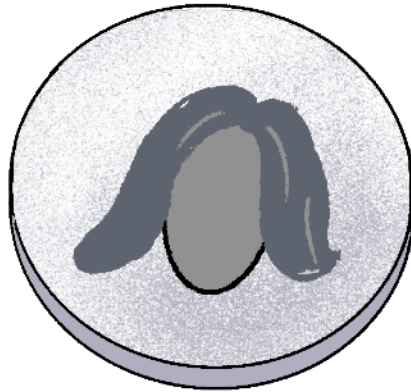


The p-value, a measure of supriz



The p-value, a measure of supriz

How likely* it is that flipping **one coin** will produce head?

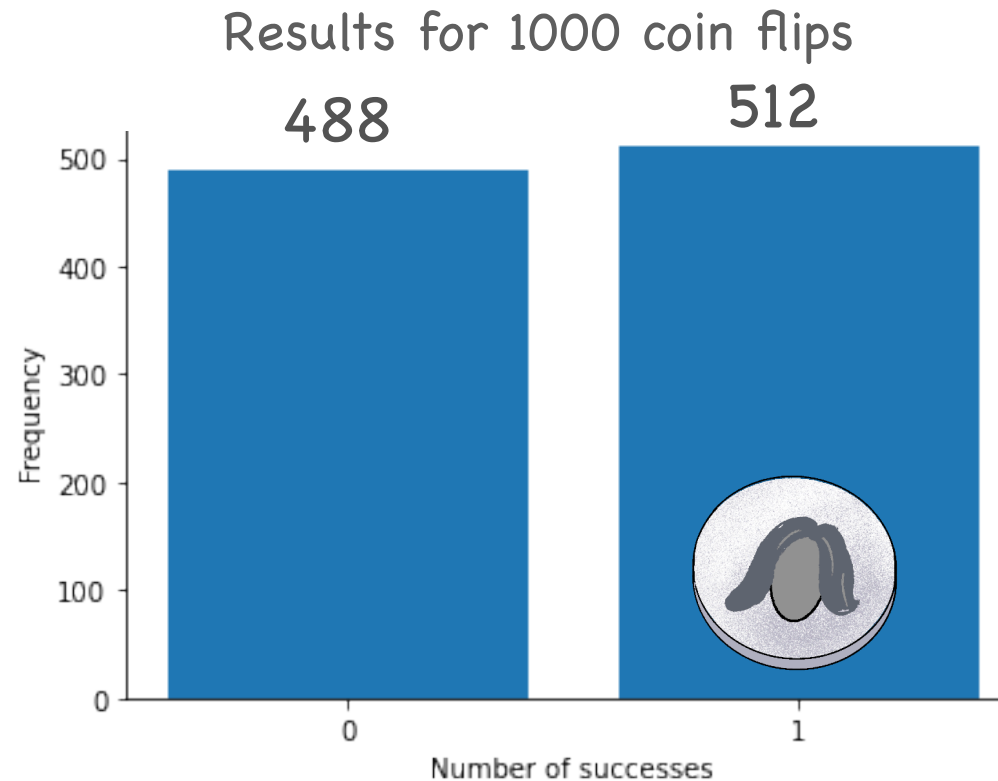


*in a universe where Ileana does not have magic powers



The p-value, a measure of suprise

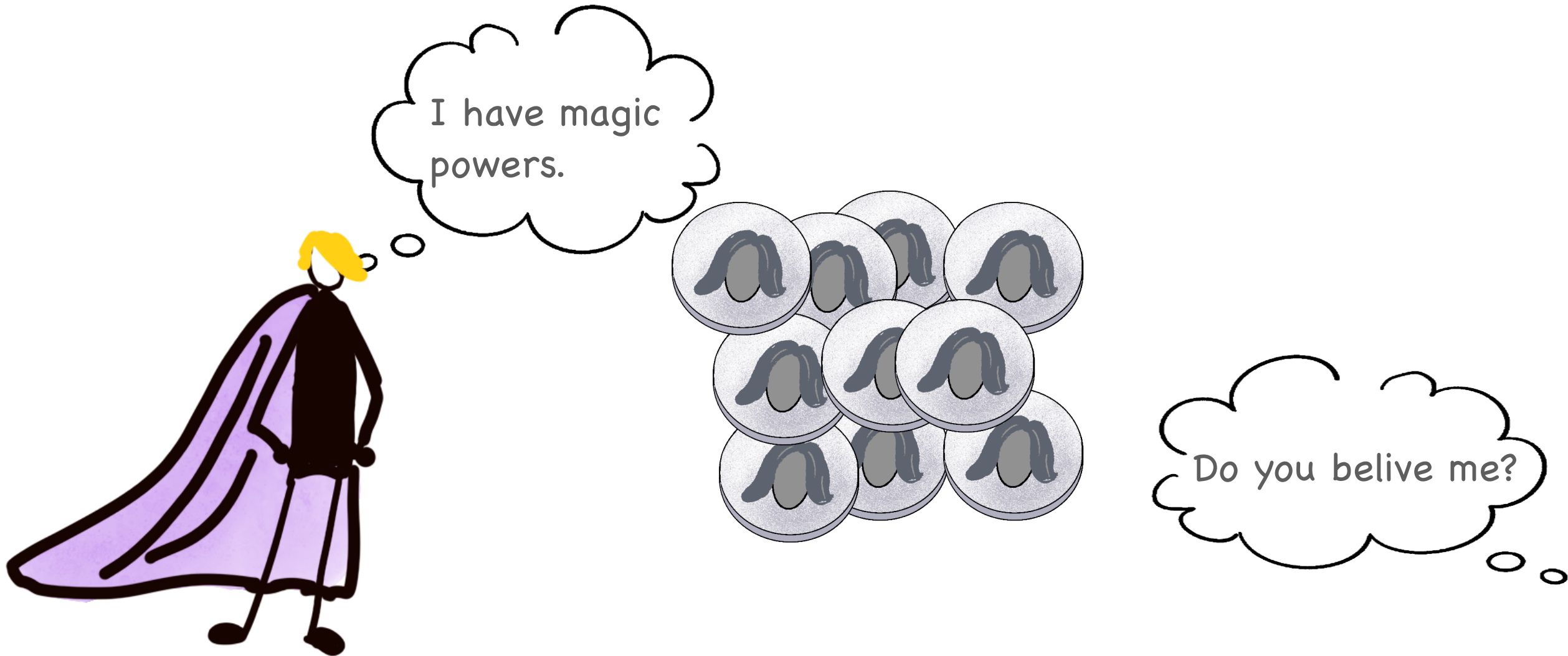
How likely* it is that flipping **one coin** will produce head?



*in a universe where Ileana does not have magic powers

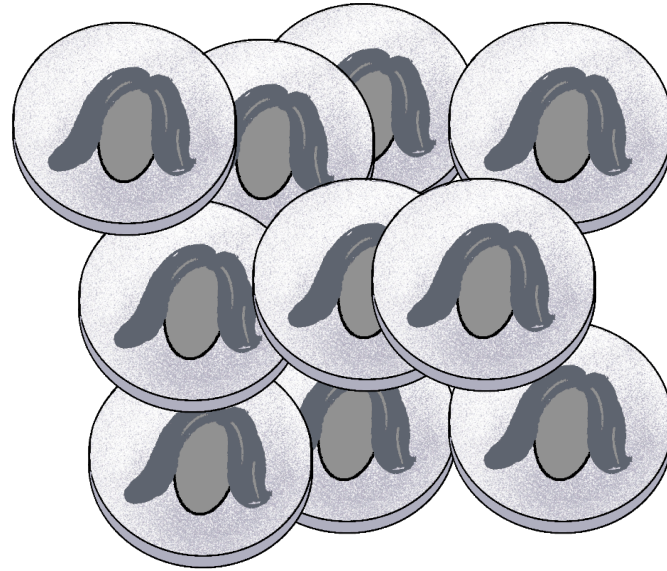


The p-value, a measure of supriz



The p-value, a measure of supriz

How likely* it is that flipping **ten coins** will produce **all heads**?

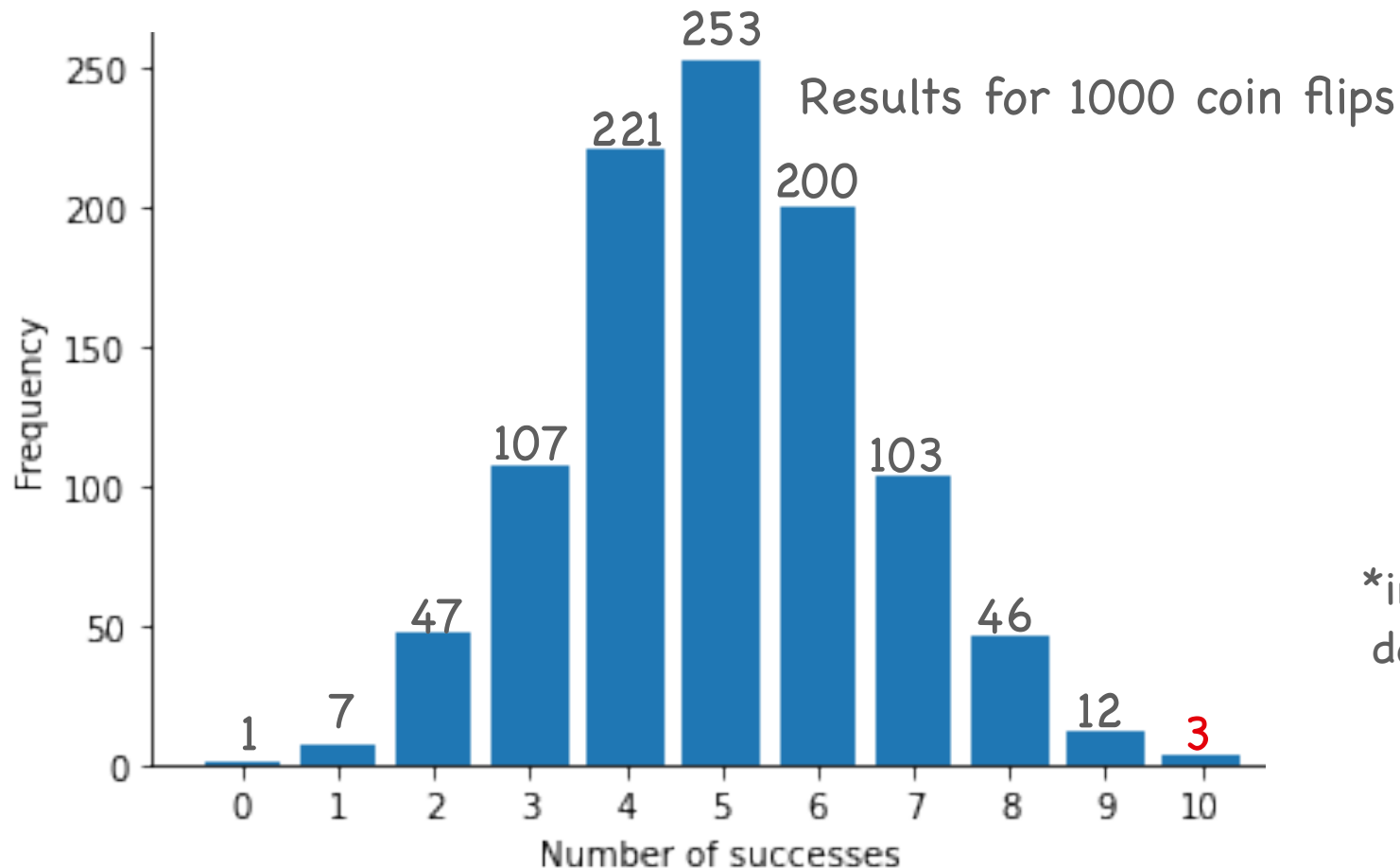


*in a universe where Ileana does not have magic powers



The p-value, a measure of supriz

How likely* it is that flipping **ten coins** will produce **all heads**?

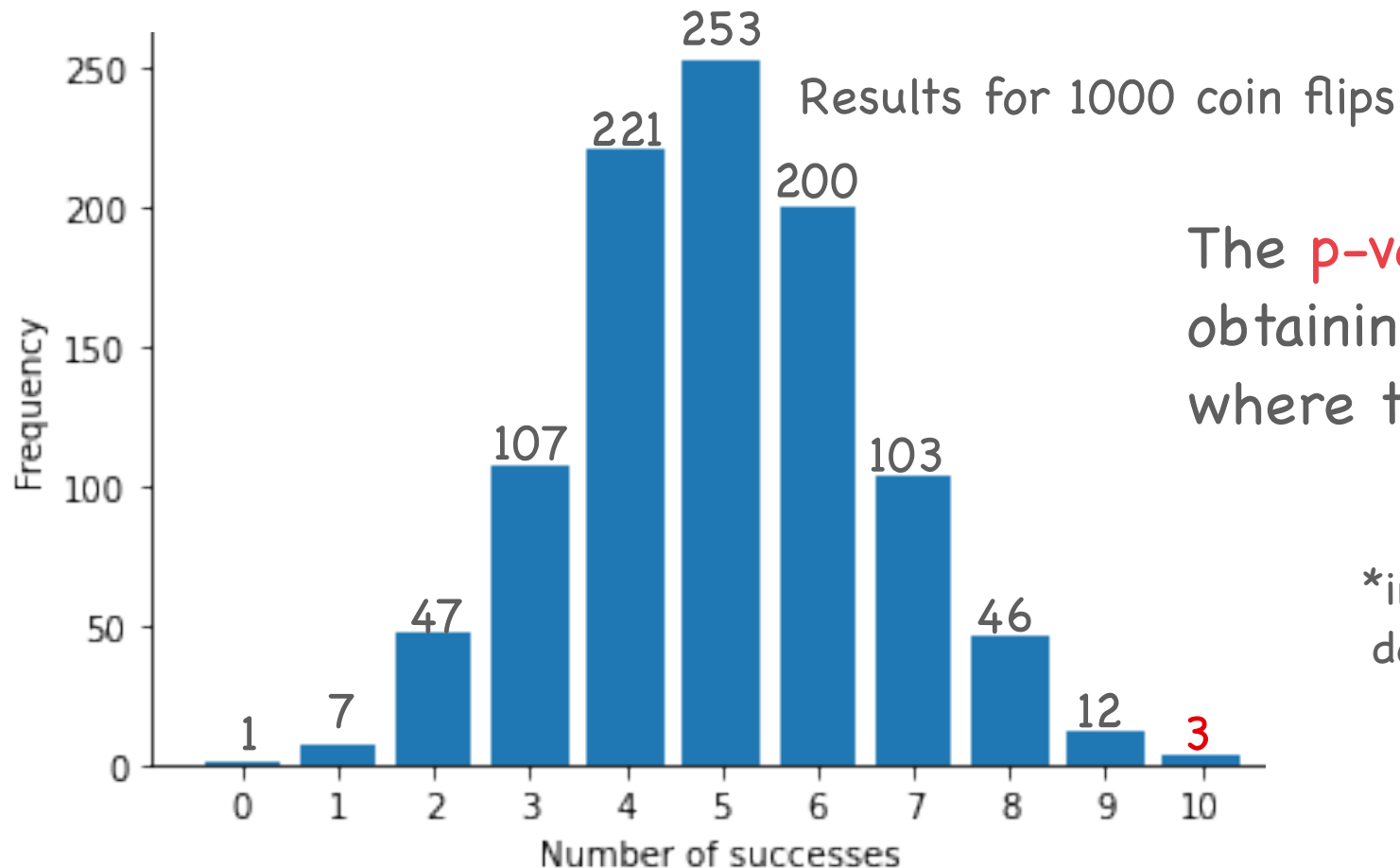


*in a universe where Ileana does not have magic powers



The p-value, a measure of surprize

How likely* it is that flipping **ten coins** will produce **all heads**?



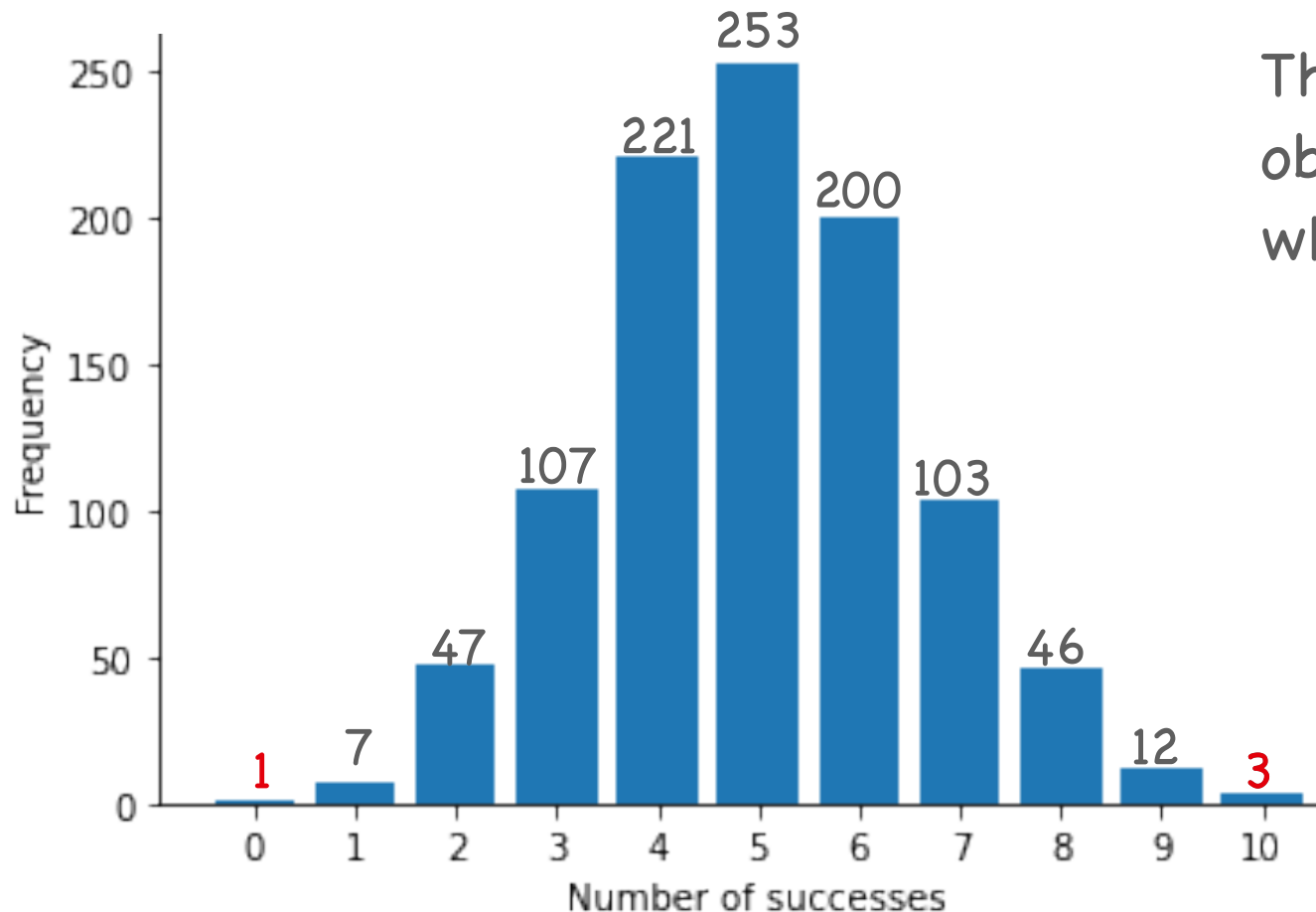
The **p-value** is the probability of obtaining the data* in universe where the null-hypothesis is true.

*in a universe where Ileana does not have magic powers



The p-value, a measure of surprize

Results for 1000 coin flips



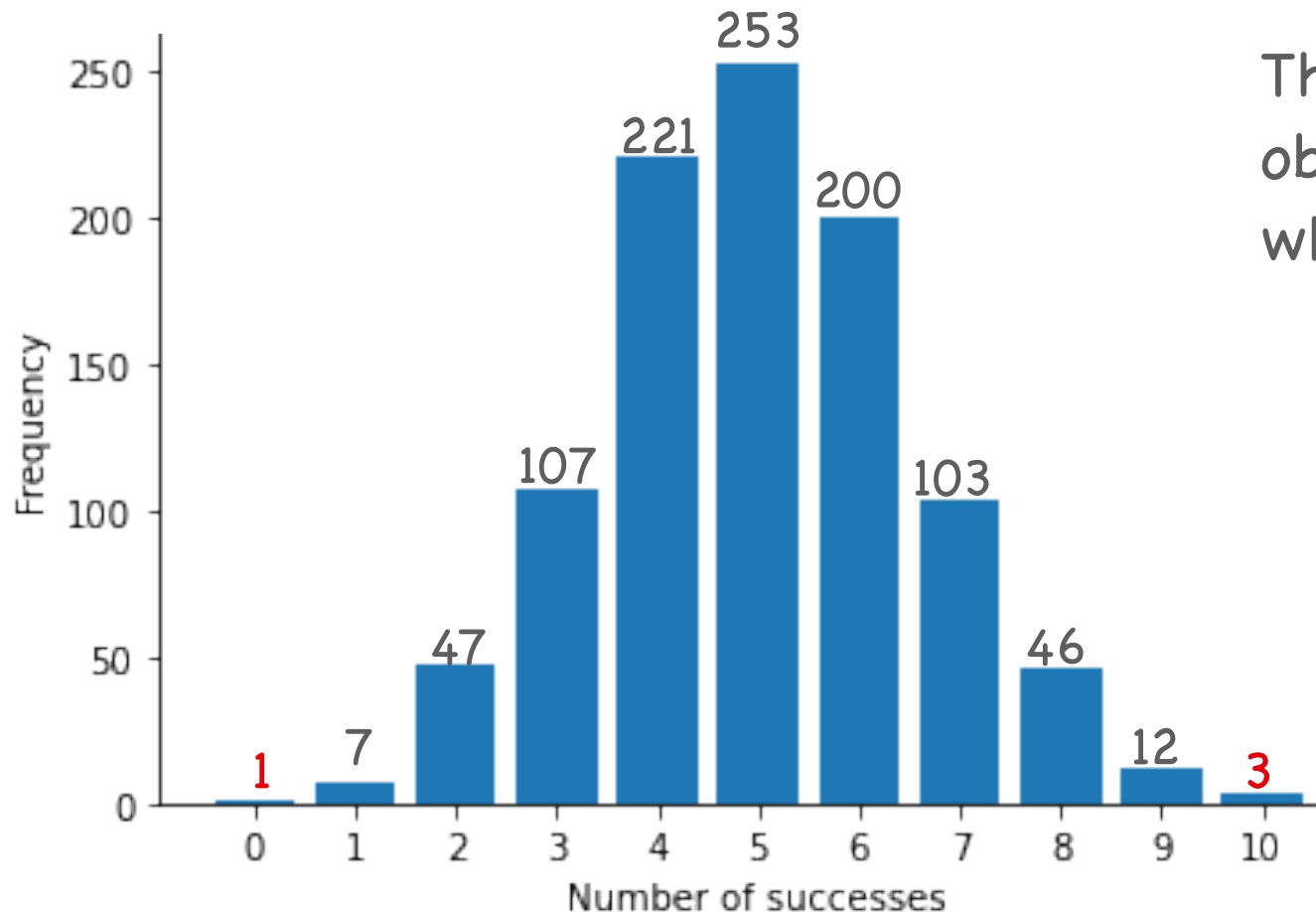
The **p-value** is the probability of obtaining the data* in universe where the null-hypothesis is true.

$$p = \frac{4}{1000} = 0.004$$

*or data showing as great or greater difference than the null

The p-value, a measure of surprize

Results for 1000 coin flips

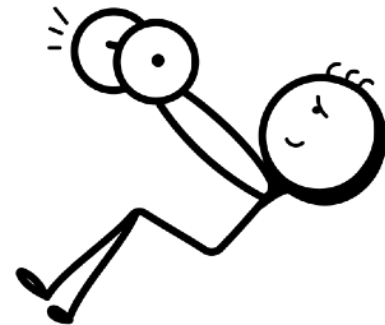
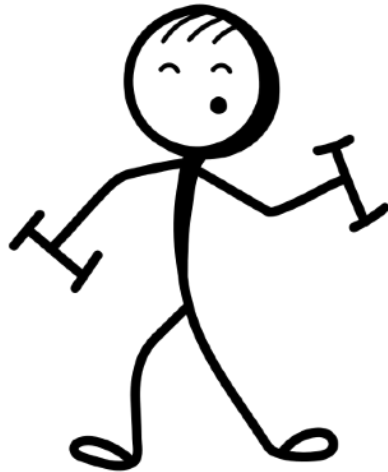
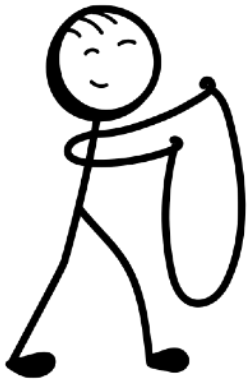


The **p-value** is the probability of obtaining the data* in universe where the null-hypothesis is true.

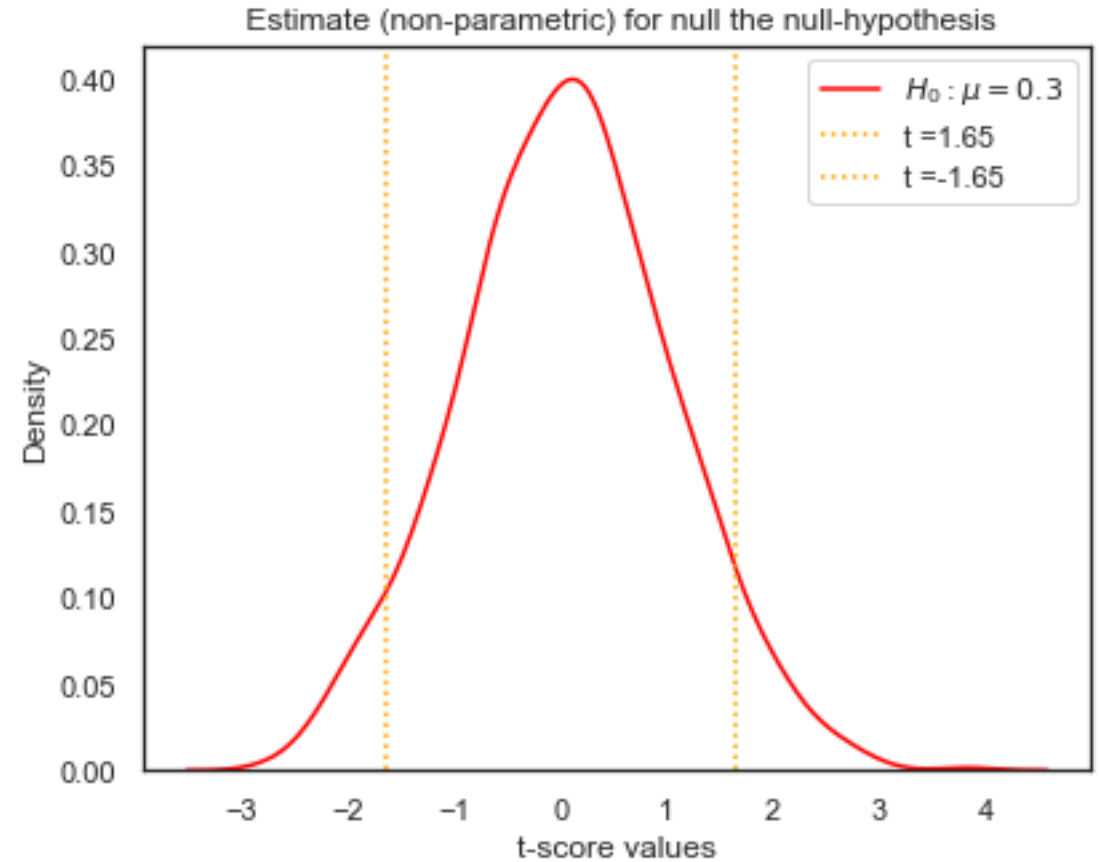
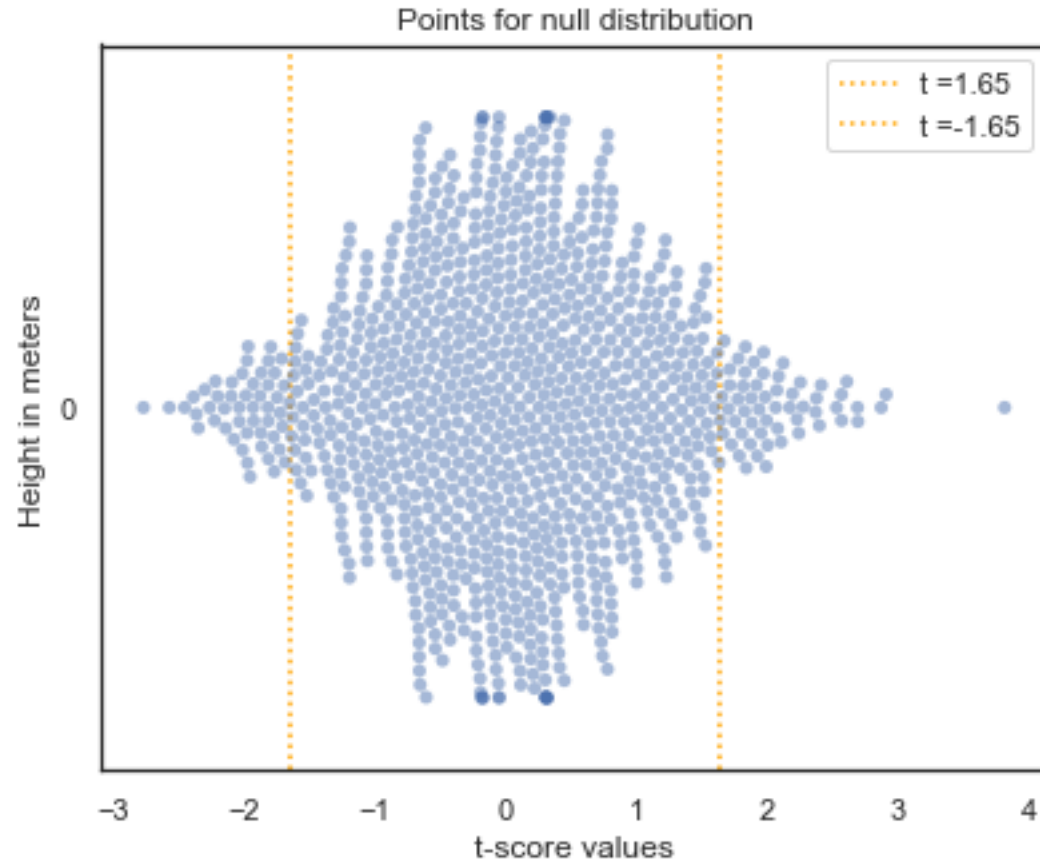
$$p = \frac{4}{1000} = 0.004$$

*or data showing as great or greater difference than the null

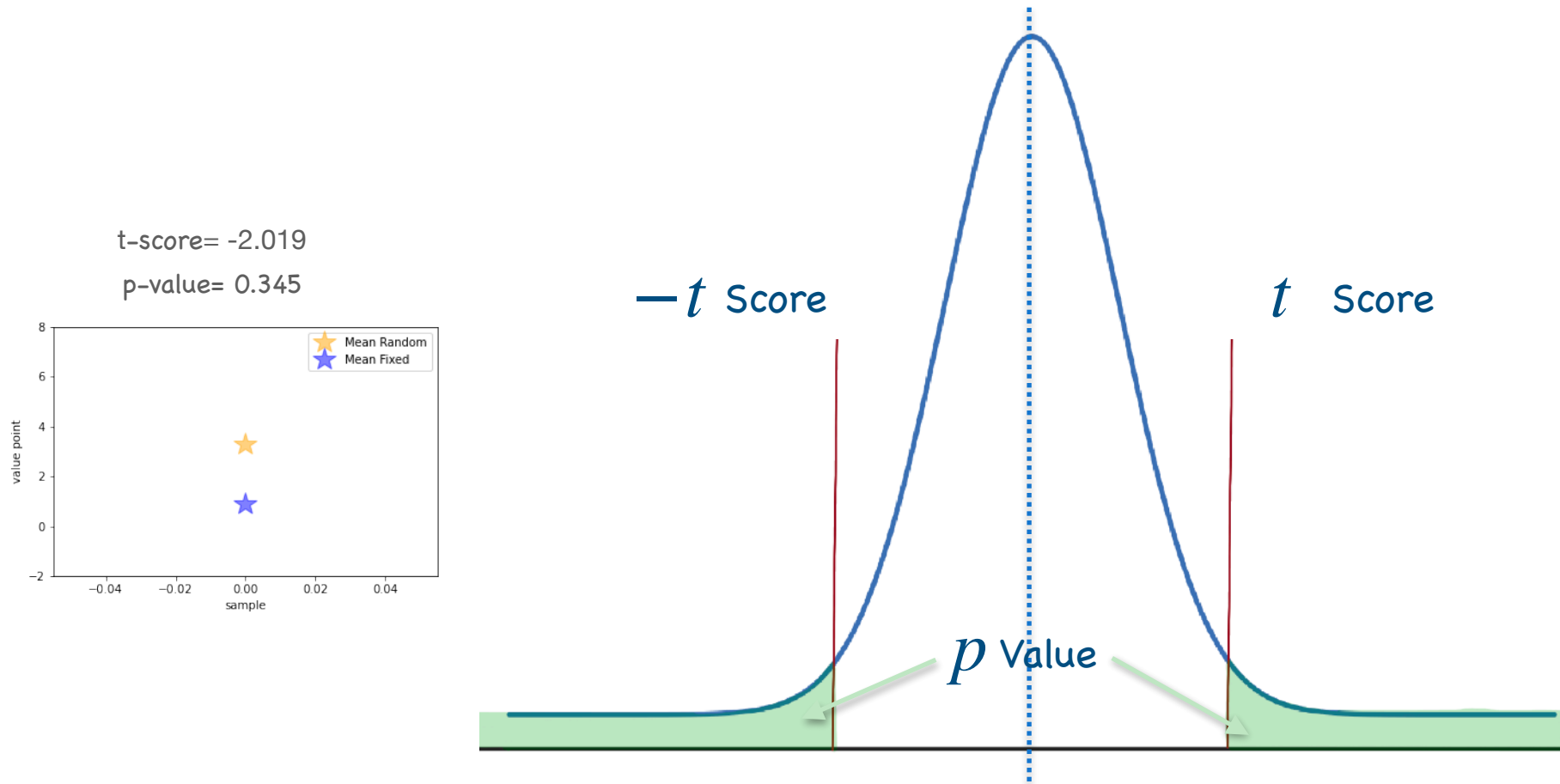
Exercise 3 (p-values)



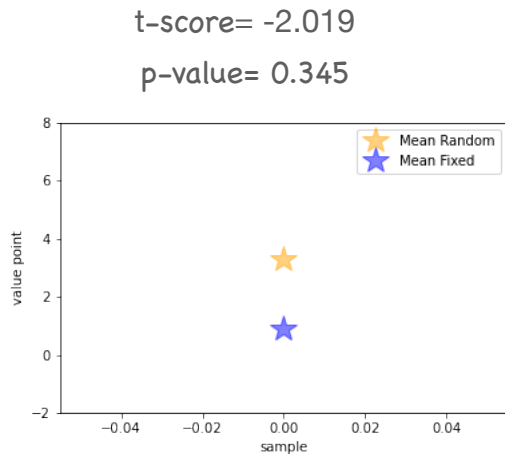
Reporting the results



p-value and α in a nutshell



p-value and α in a nutshell



$$\mu_r = \mu_f$$

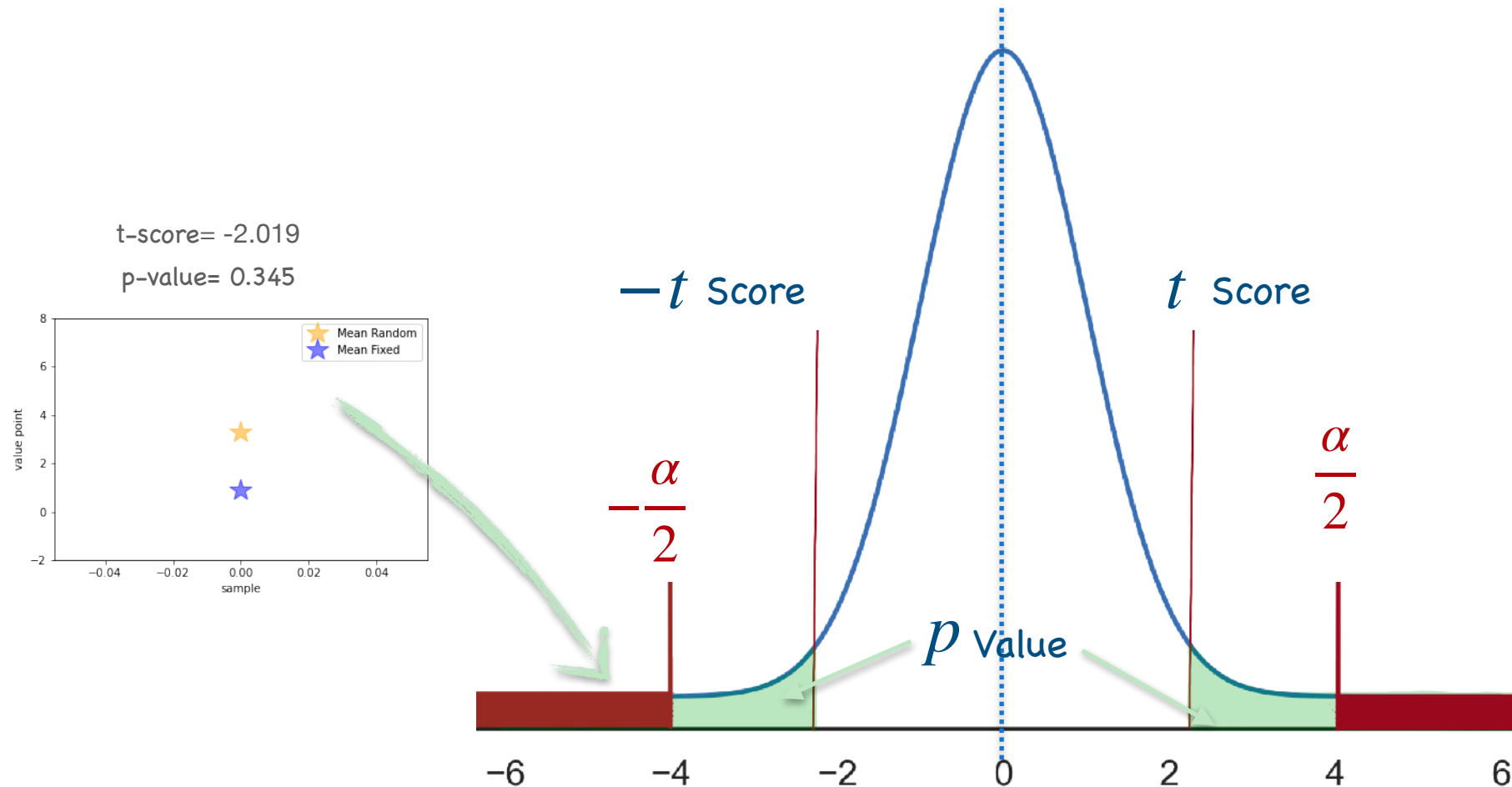
P- value answers the question:

If I live in a universe where H_0 is true, how surprising is to measure a t-score of -2.019?



If p is small the null must go

p-value and α in a nutshell



Reporting the results

When writing up the results you should always include the following information:

- the value of the test statistic
- The sample size
- The p-value

Finally, do your results provide evidence against the null hypothesis?

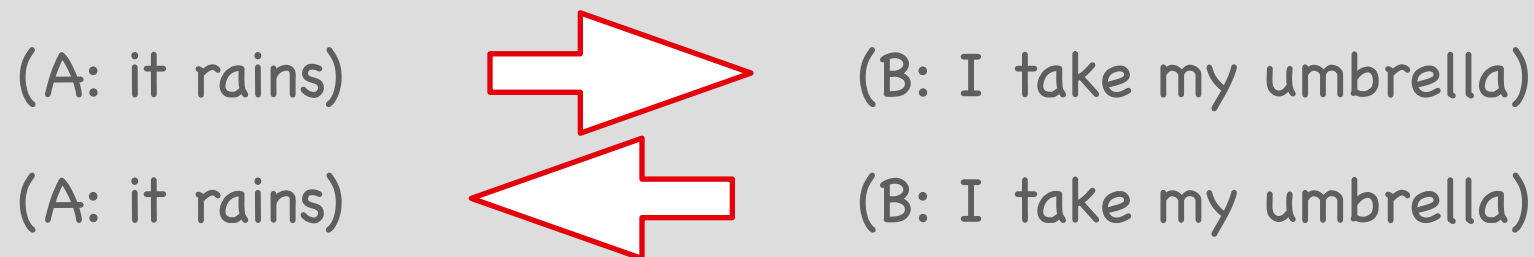
“We reject/fail to reject the null-hypothesis at a significance level of α ”

A note on p-values

The p-value indicates strength of evidence against H_0 . It's the probability of getting the observed results or more extreme results IF H_0 is true.

Red flag: the p-value is NOT the probability of H_0 being true.

Inverse probability fallacy:



Why use p-values?

If its so hard to interpret p-values, why do we use the? Testing means that at each step, we critically assess our data. A high p-value means that maybe there is no pattern in the data, even if it seems to us to be.

Confirmation bias:

After buying a red car, I tend to notice them everywhere.

THE END OF PART 1