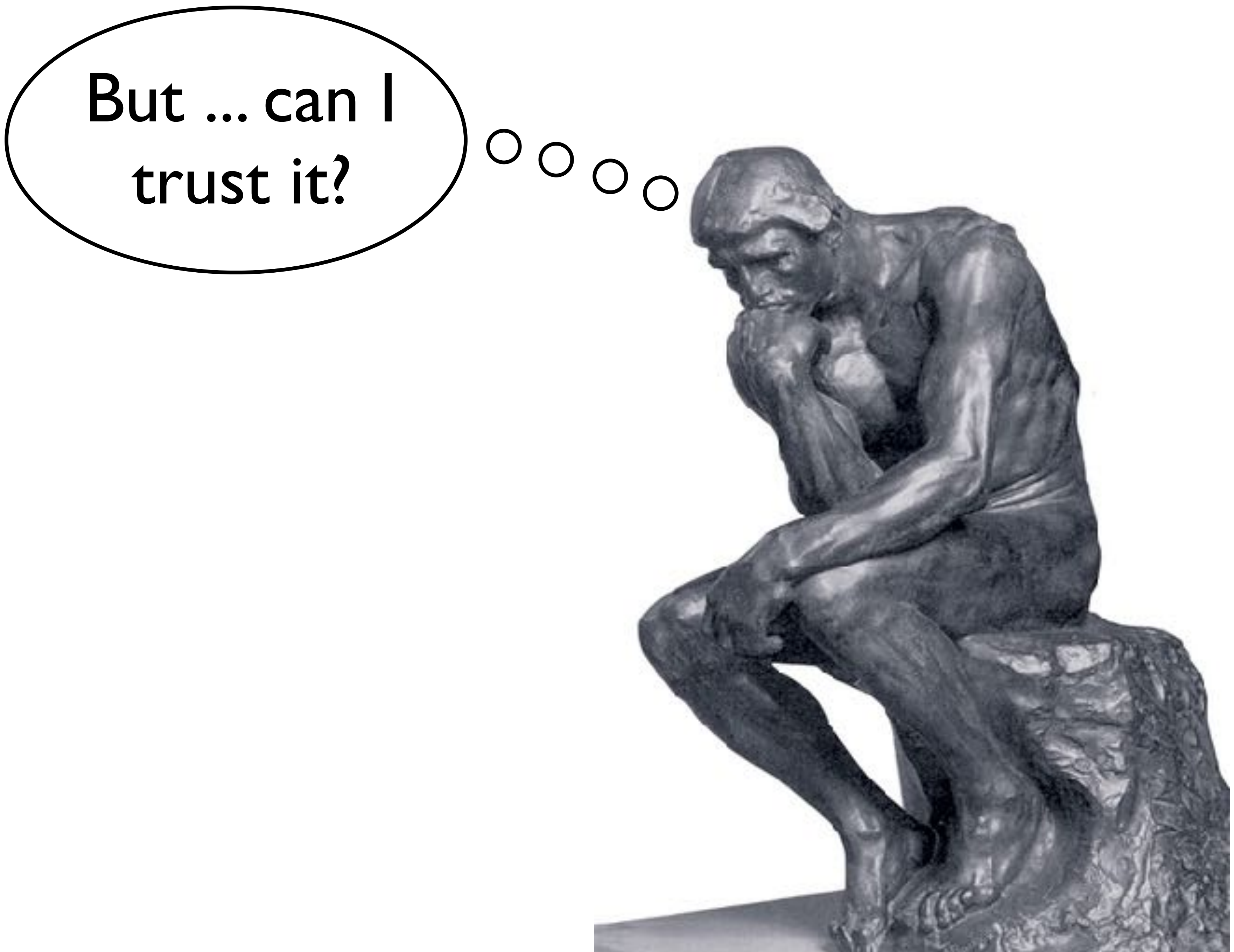
Inference

how surprising is your statistic? (thresholding)



Outline

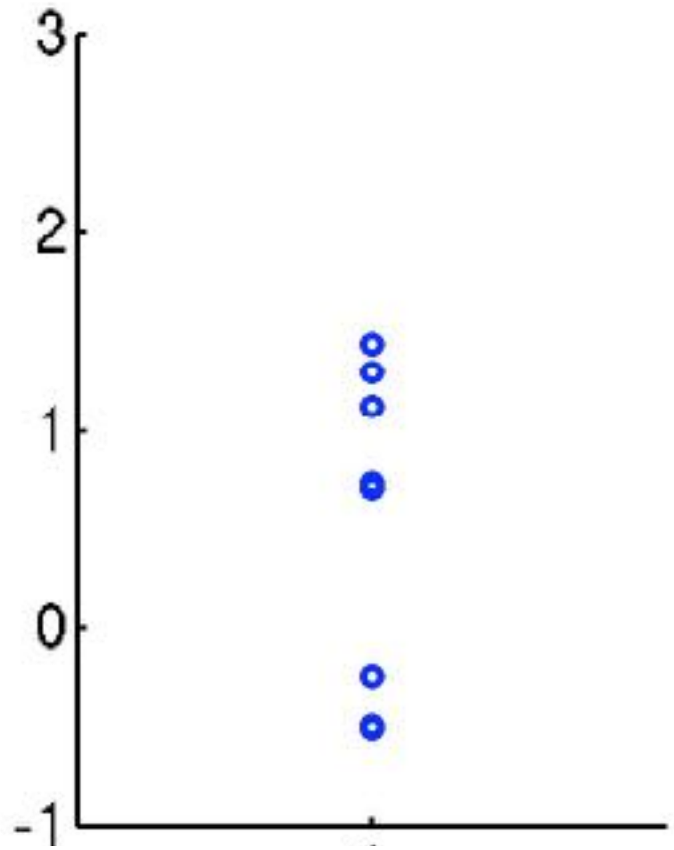
- Null-hypothesis and Null-distribution
- Multiple comparisons and Family-wise error
- Different ways of being surprised
  - Voxel-wise inference (Maximum z)
  - Cluster-wise inference (Maximum size)
- Parametric vs non-parametric tests
- Enhanced clusters
- FDR - False Discovery Rate

Outline

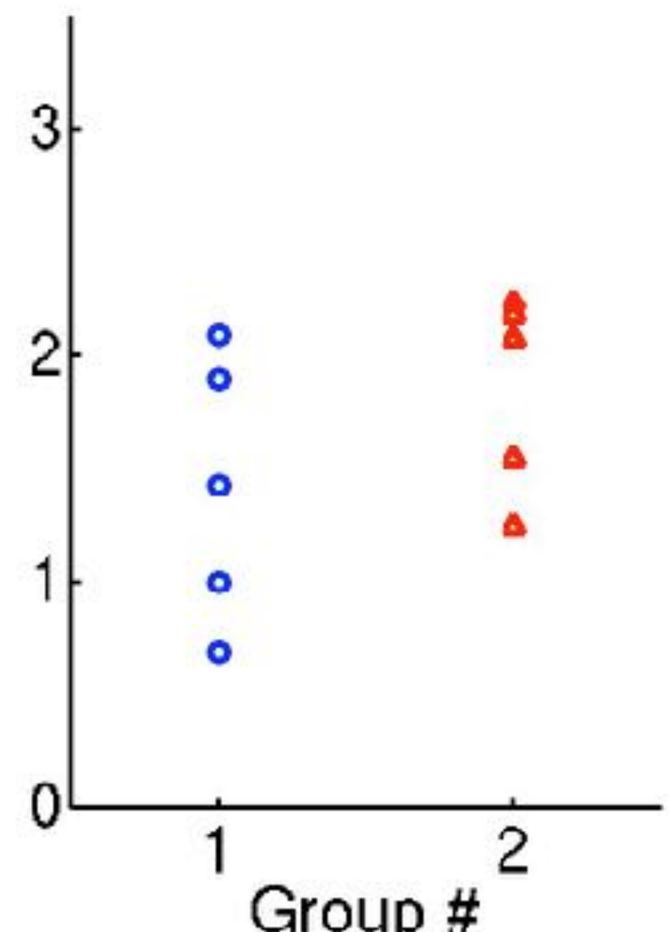
- Null-hypothesis and Null-distribution
- Multiple comparisons and Family-wise error
- Different ways of being surprised
    - Voxel-wise inference (Maximum z)
    - Cluster-wise inference (Maximum size)
- Parametric vs non-parametric tests
- Enhanced clusters
- FDR - False Discovery Rate

The task of classical inference

- Given some data we want to know if (e.g.) a mean is different from zero or if two means are different



> 0 ?



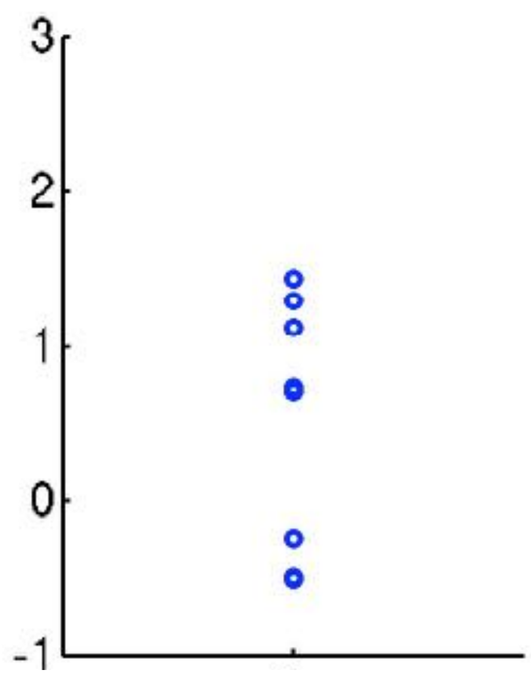
Different?

Tools of classical inference

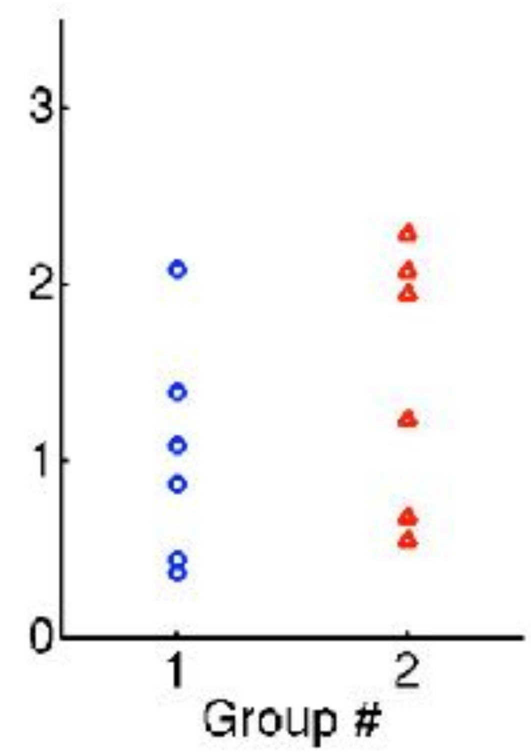
1. A null-hypothesis

Typically the opposite of what we actually "hope", e.g.

There is **no** effect of treatment: $\mu = 0$



There is **no** difference between groups: $\mu_1 = \mu_2$

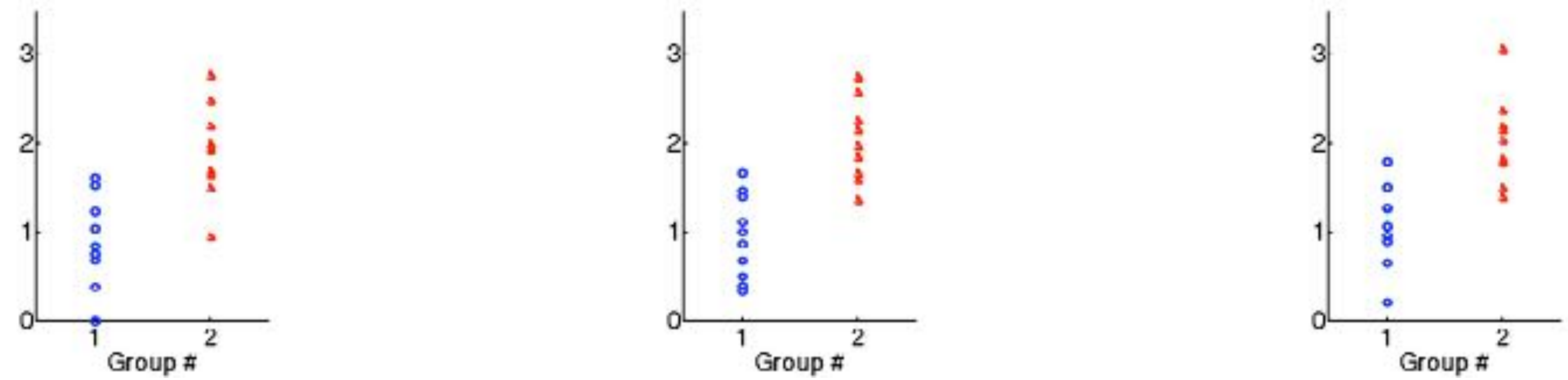


Tools of classical inference

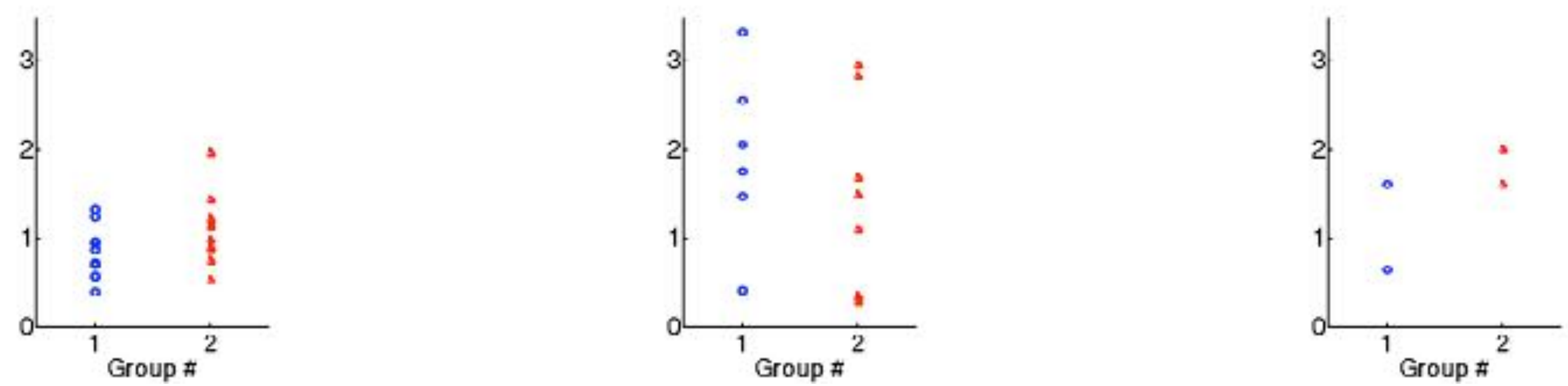
1. A null-hypothesis
2. A test-statistic

Assesses “trustworthiness”

Trustworthy



Dodgy



Tools of classical inference

1. A null-hypothesis
2. A test-statistic

Assesses "trustworthiness"

A *t*-statistic reflects precisely this

Large difference: Trustworthy

$$t = \sqrt{n}\frac{\overline{x_1} - \overline{x_2}}{\sqrt{\sigma^2}}$$

Small variability: Trustworthy

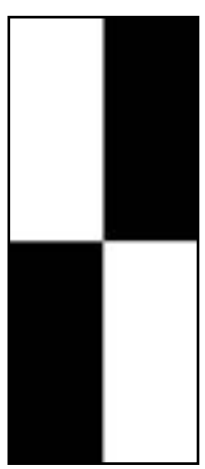
Many measurements: Trustworthy

Tools of classical inference

1. A null-hypothesis
2. A test-statistic

Or expressed in GLM lingo

=



$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} \overline{x}_1 \\ \overline{x}_2 \end{bmatrix}$$

Large difference: Trustworthy

$\overline{x}_1$ - $\overline{x}_2$

$$t = \frac{\mathbf{c}^T\widehat{\boldsymbol{\beta}}}{\sqrt{\sigma^2}\sqrt{\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}}}$$

Small variability: Trustworthy

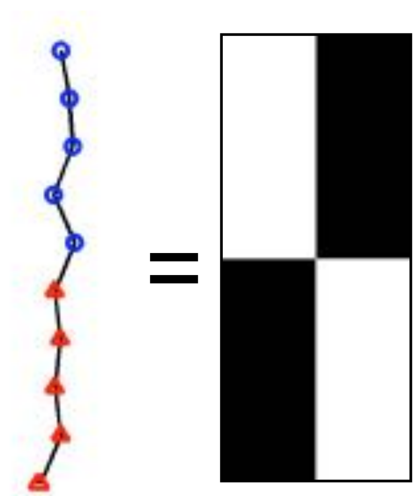
Many measurements: Trustworthy

Tools of classical inference

1. A null-hypothesis
2. A test-statistic
3. A null-distribution

Let us assume there is no difference, i.e. the null-hypothesis is true.

We might then get these data



$$= \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \mathbf{e}$$

$$t = 2.19$$

$$\mathbf{c}^T\widehat{\boldsymbol{\beta}} = 1.17$$

$$t = \frac{\mathbf{c}^T\widehat{\boldsymbol{\beta}}}{\sqrt{\sigma^2}\sqrt{\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}}}$$

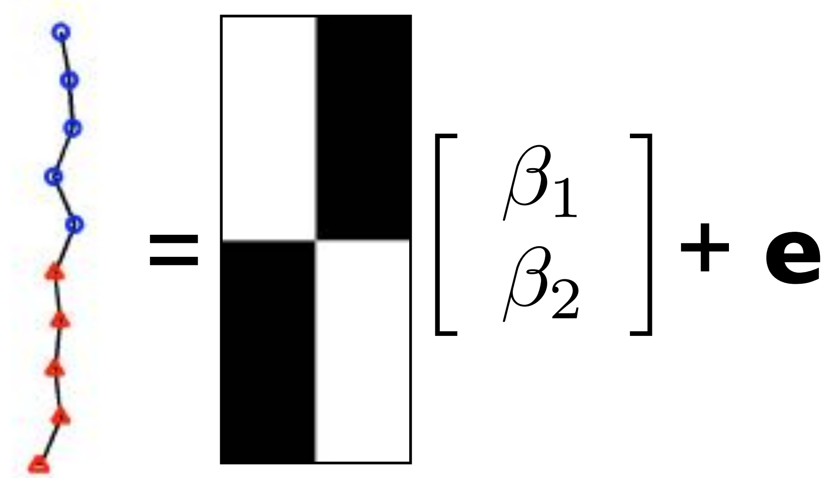
$$\sigma^2 = 0.71$$

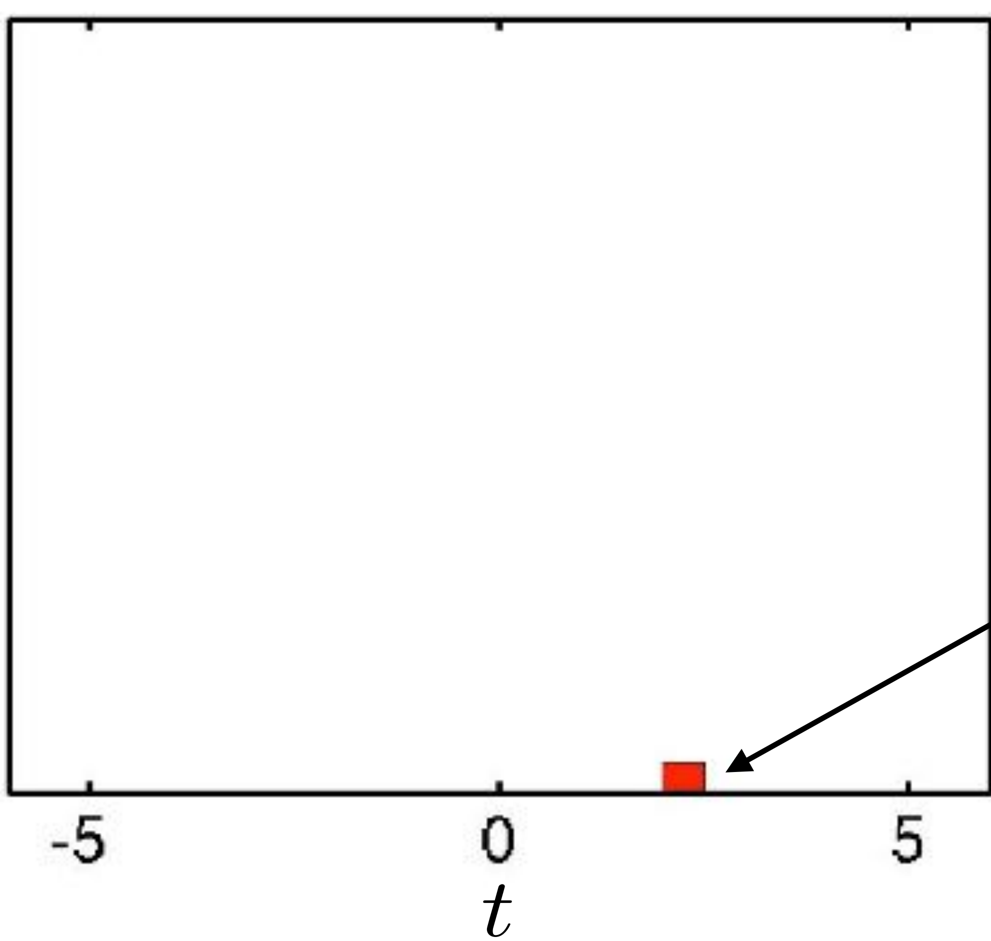
Constant

Tools of classical inference

1. A null-hypothesis
2. A test-statistic
3. A null-distribution

We might then get these data





$t = 2.19$

$\mathbf{c}^T\widehat{\boldsymbol{\beta}} = 1.17$

$$t = \frac{\mathbf{c}^T\widehat{\boldsymbol{\beta}}}{\sqrt{\sigma^2}\sqrt{\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}}}$$

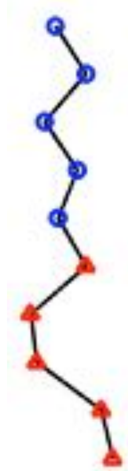
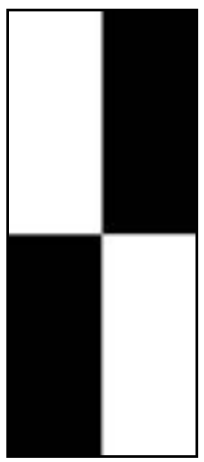
$\sigma^2 = 0.71$

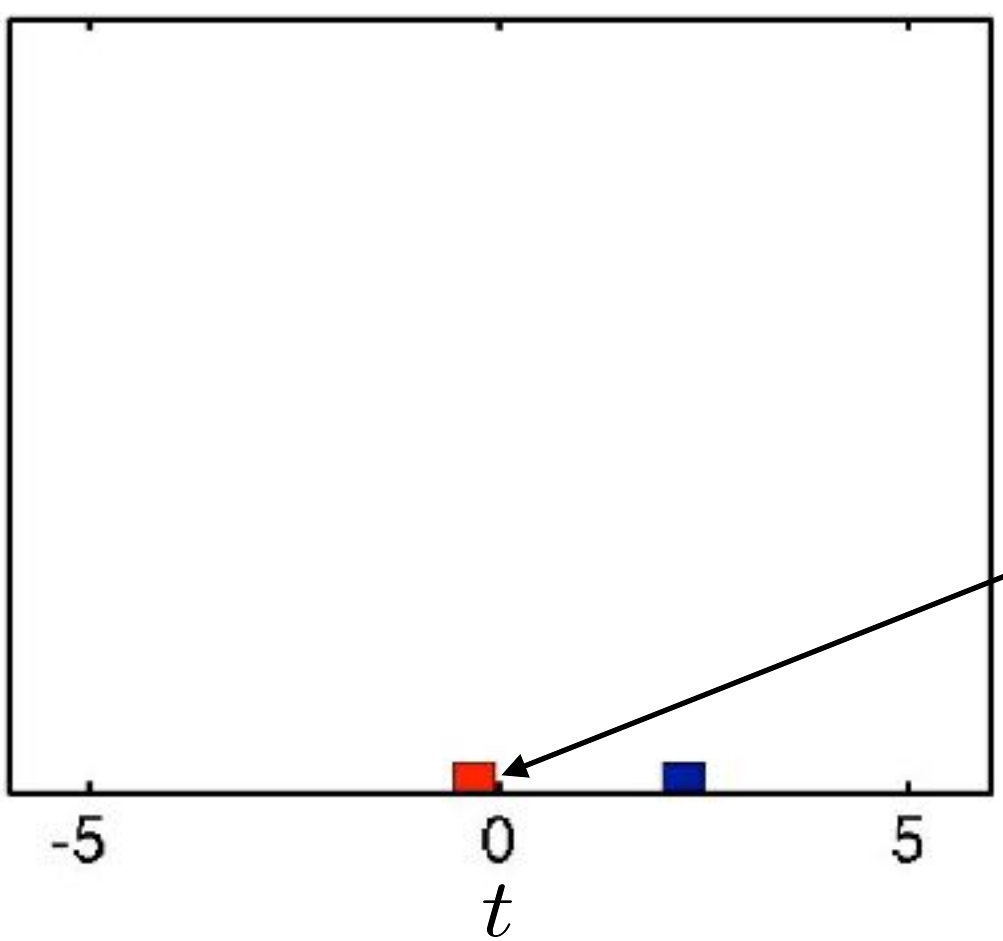
Constant

Tools of classical inference

1. A null-hypothesis
2. A test-statistic
3. A null-distribution

or we could have gotten these

 $=$  $\begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix} + \mathbf{e}$



$t = -0.51$

$\mathbf{c}^T\widehat{\boldsymbol{\beta}} = -0.37$

$$t = \frac{\mathbf{c}^T\widehat{\boldsymbol{\beta}}}{\sqrt{\sigma^2}\sqrt{\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}}}$$

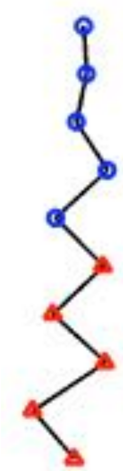
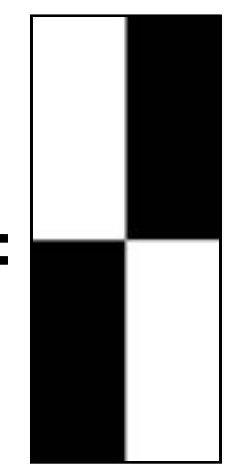
$\sigma^2 = 1.28$

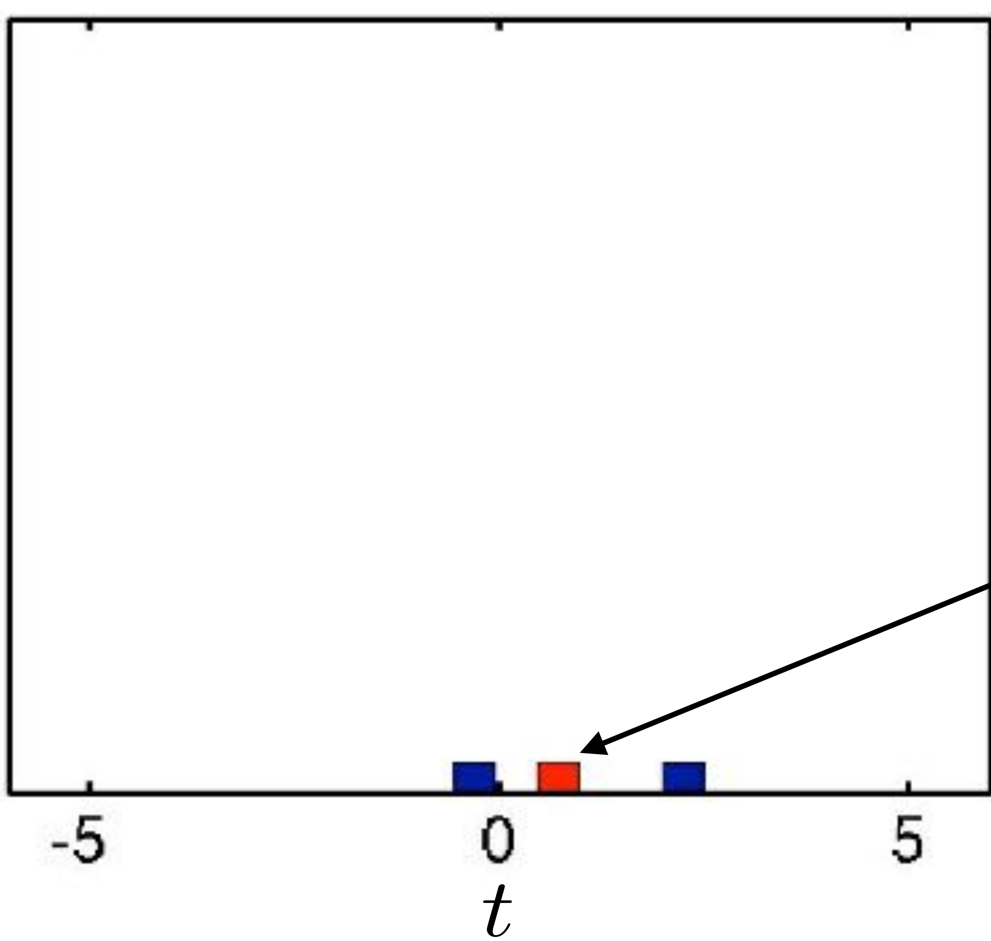
Constant

Tools of classical inference

1. A null-hypothesis
2. A test-statistic
3. A null-distribution

maybe these


=

$\begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix}$ + **e**



$t = 0.49$

$\mathbf{c}^T\widehat{\boldsymbol{\beta}} = 0.31$

$$t = \frac{\mathbf{c}^T\widehat{\boldsymbol{\beta}}}{\sqrt{\sigma^2}\sqrt{\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}}}$$

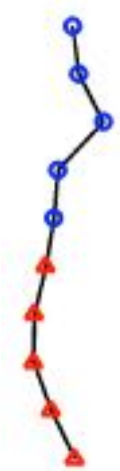
$\sigma^2 = 1.01$

Constant

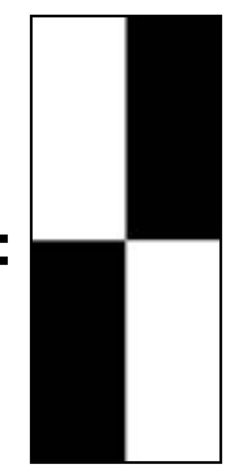
Tools of classical inference

1. A null-hypothesis
2. A test-statistic
3. A null-distribution

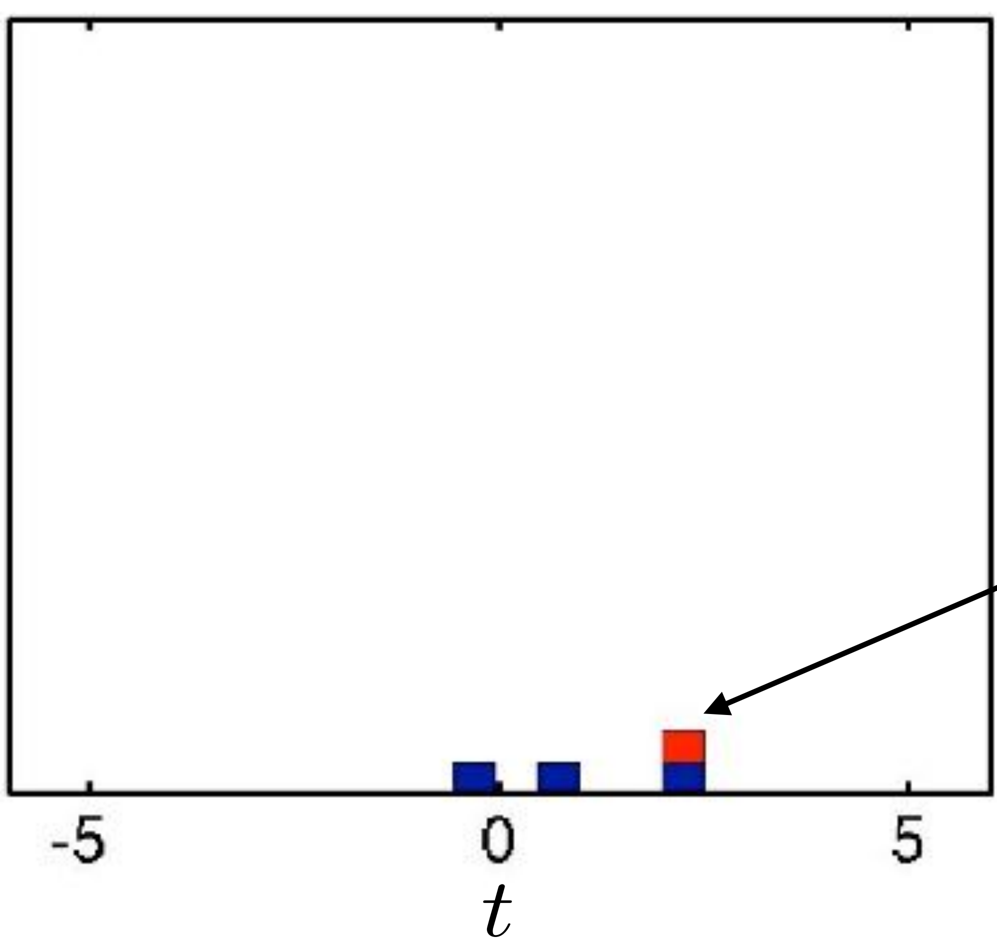
or perhaps these



$=$



$$\begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix} + \mathbf{e}$$



$t = 2.19$

$\mathbf{c}^T\widehat{\boldsymbol{\beta}} = 1.22$

$$t = \frac{\mathbf{c}^T\widehat{\boldsymbol{\beta}}}{\sqrt{\sigma^2}\sqrt{\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}}}$$

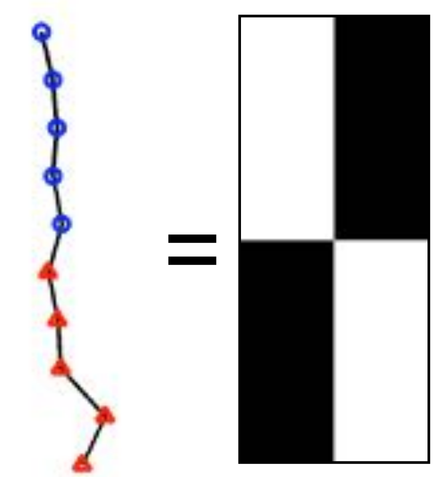
$\sigma^2 = 0.78$

Constant

Tools of classical inference

1. A null-hypothesis
2. A test-statistic
3. A null-distribution

etc



$$= \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \mathbf{e}$$

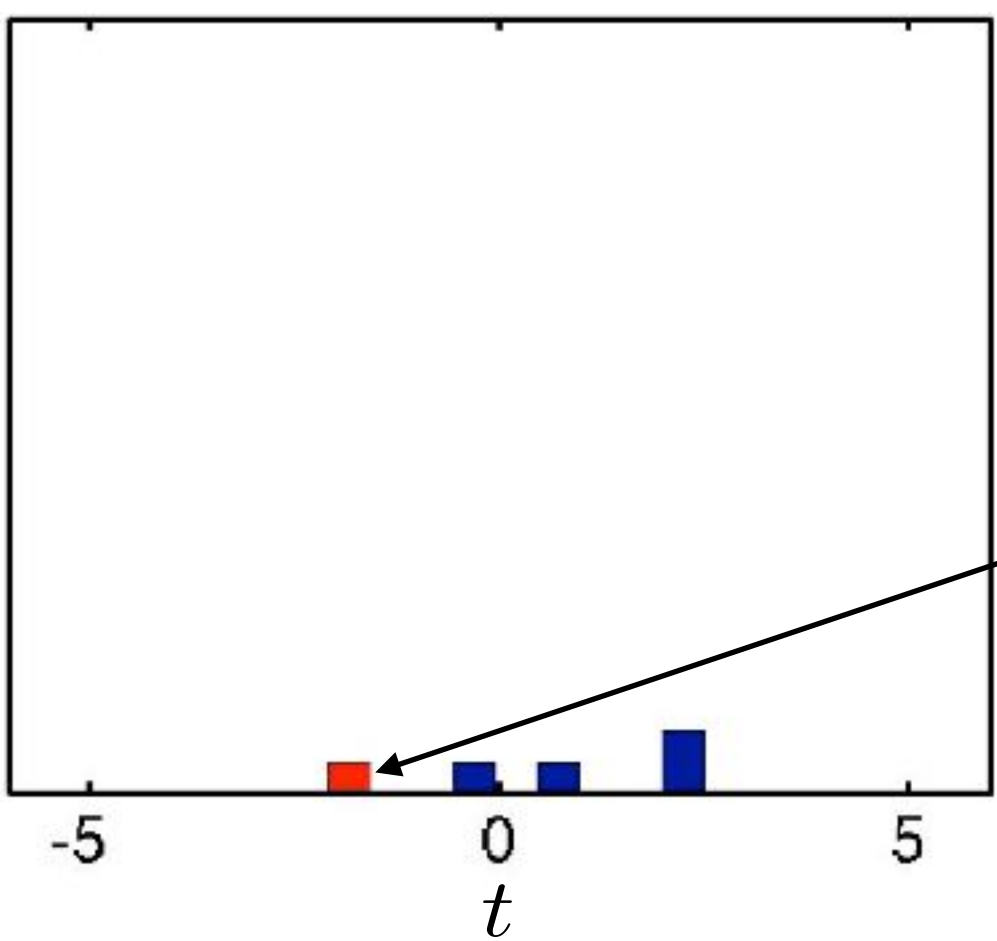
$$\mathbf{c}^T\widehat{\boldsymbol{\beta}} = -0.69$$

$$t = -1.66$$

$$t = \frac{\mathbf{c}^T\widehat{\boldsymbol{\beta}}}{\sqrt{\sigma^2}\sqrt{\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}}}$$

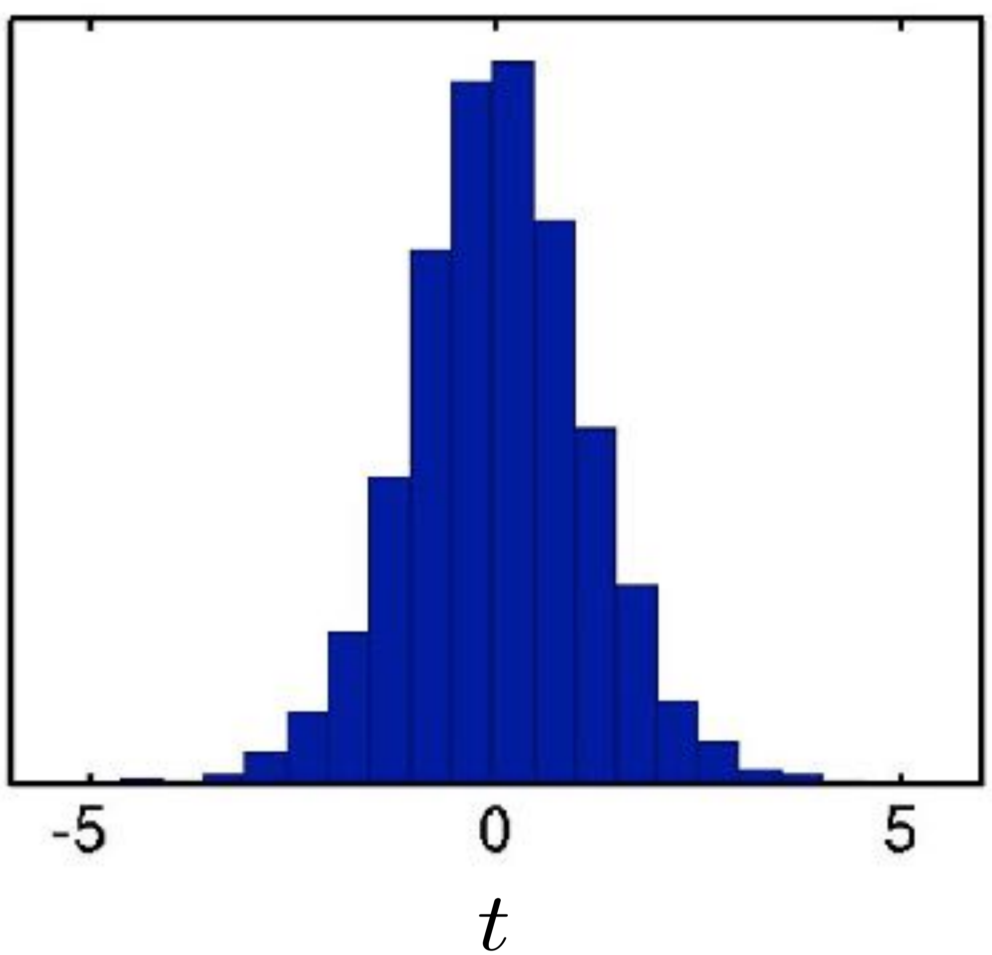
$$\sigma^2 = 0.44$$

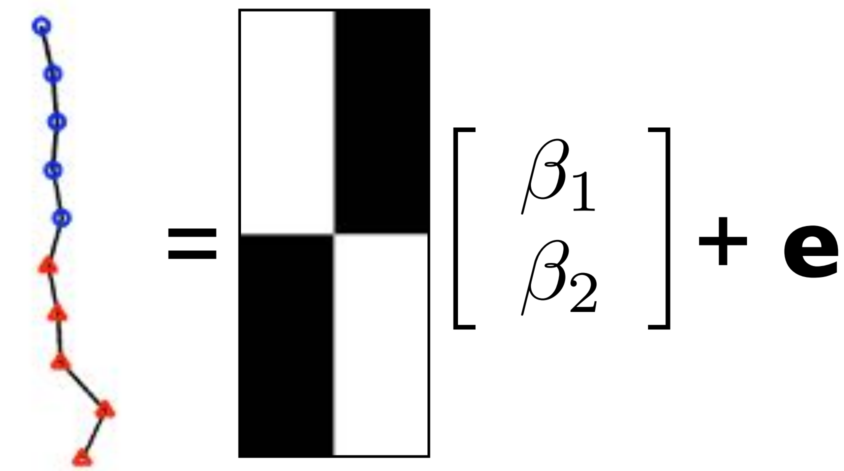
Constant



Tools of classical inference

1. A null-hypothesis
2. A test-statistic
3. A null-distribution

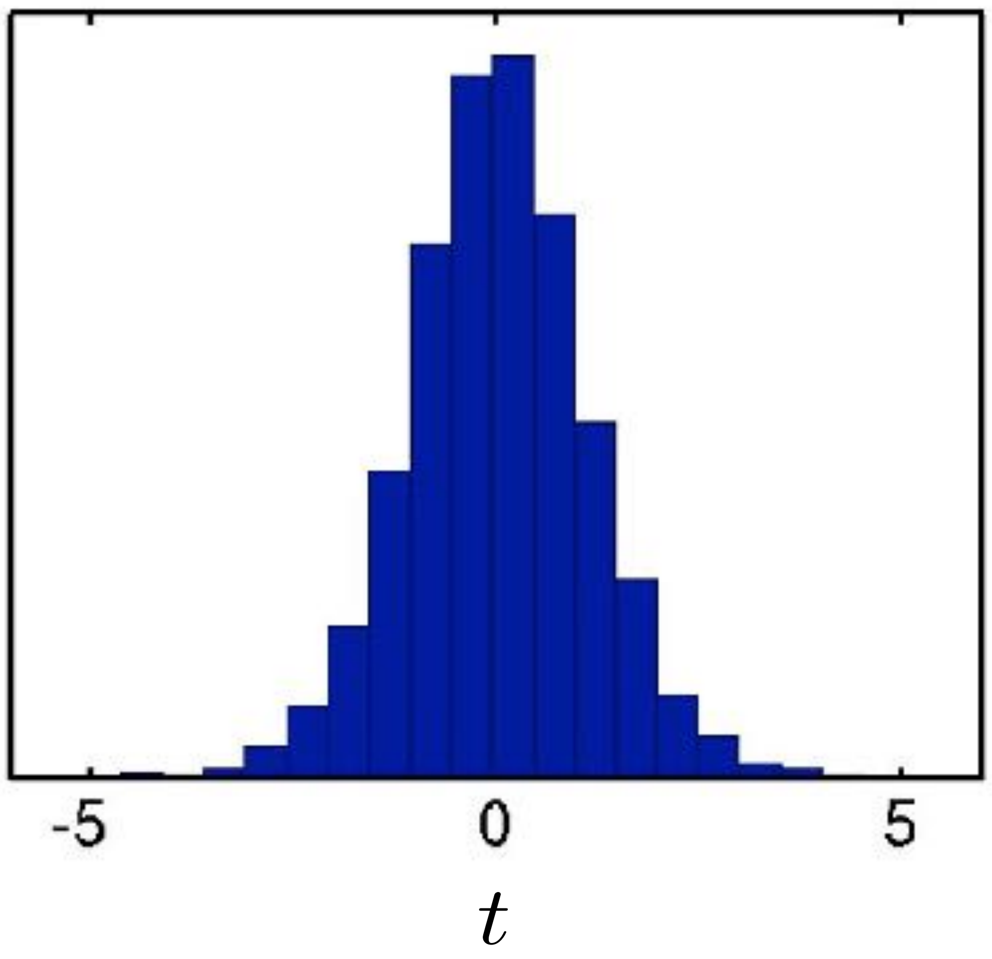




And if we do this til the cows come home

Tools of classical inference

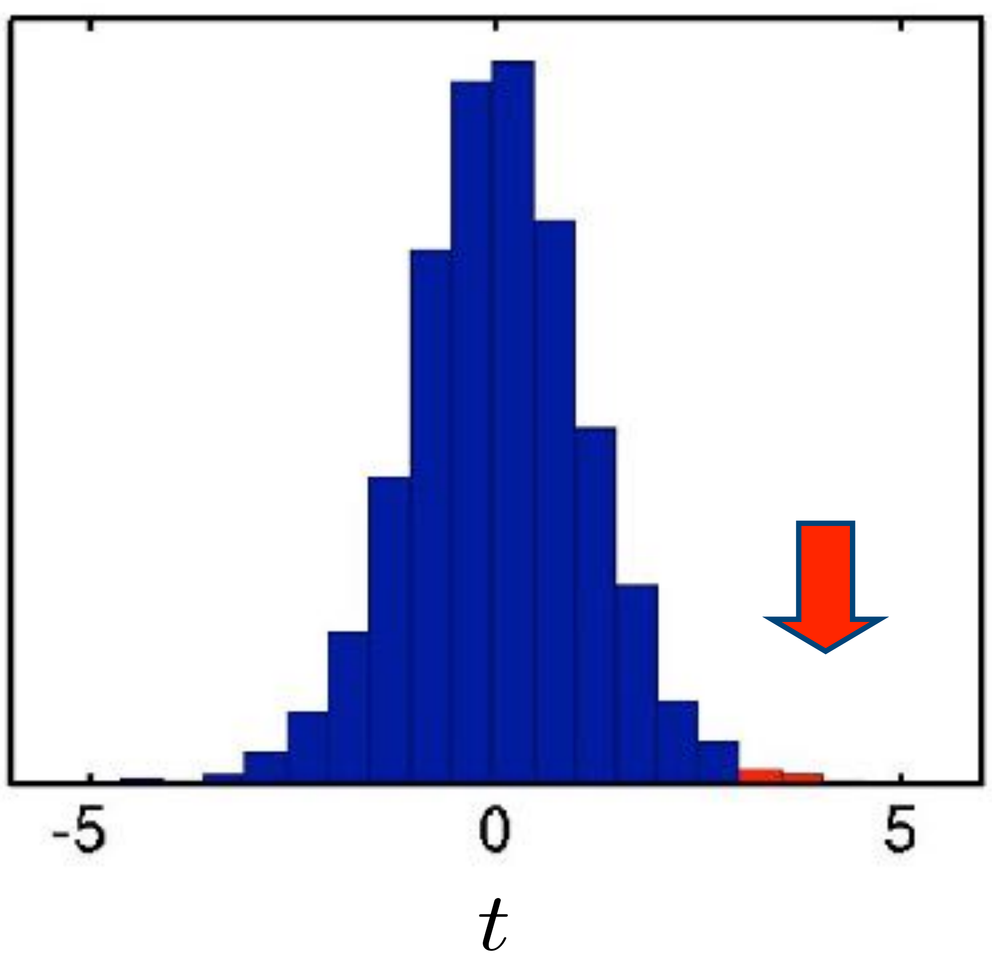
1. A null-hypothesis
2. A test-statistic
3. A null-distribution



So, why is this helpful?

Tools of classical inference

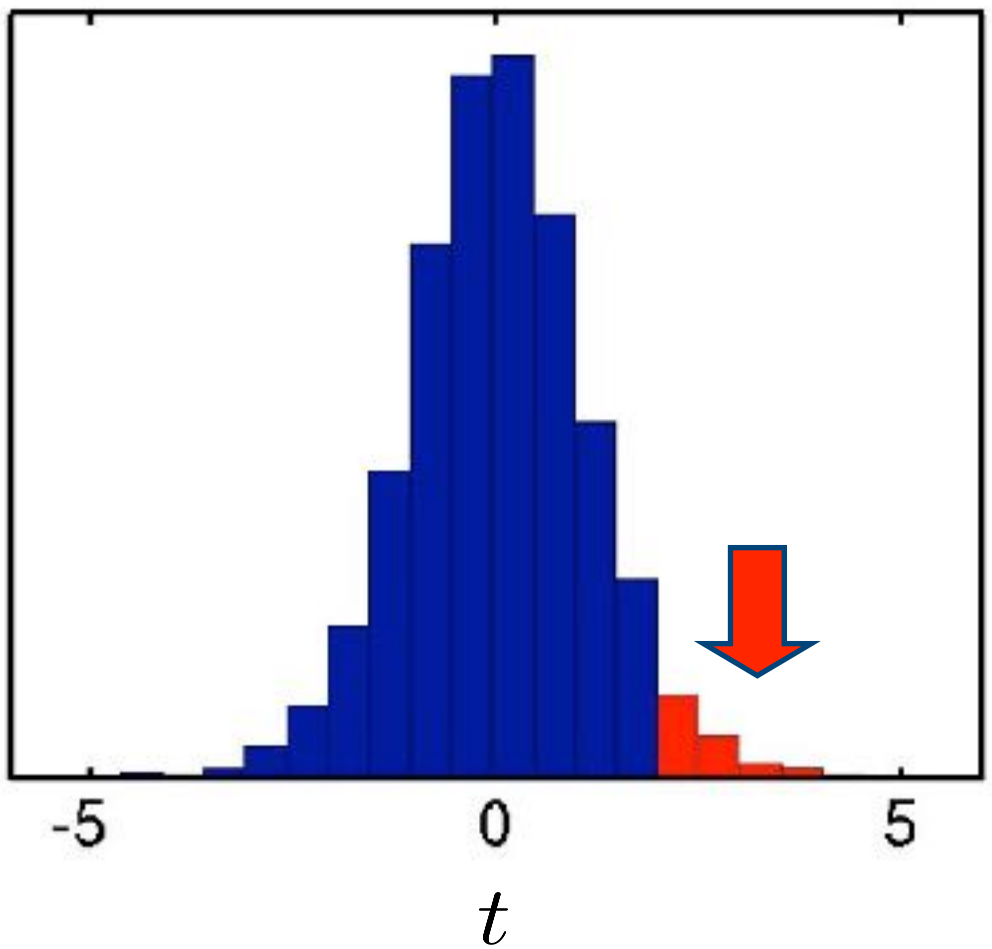
1. A null-hypothesis
2. A test-statistic
3. A null-distribution



Well, it for example tells us that in ~1% of the cases t > 3.00, even when the null-hypothesis is true.

Tools of classical inference

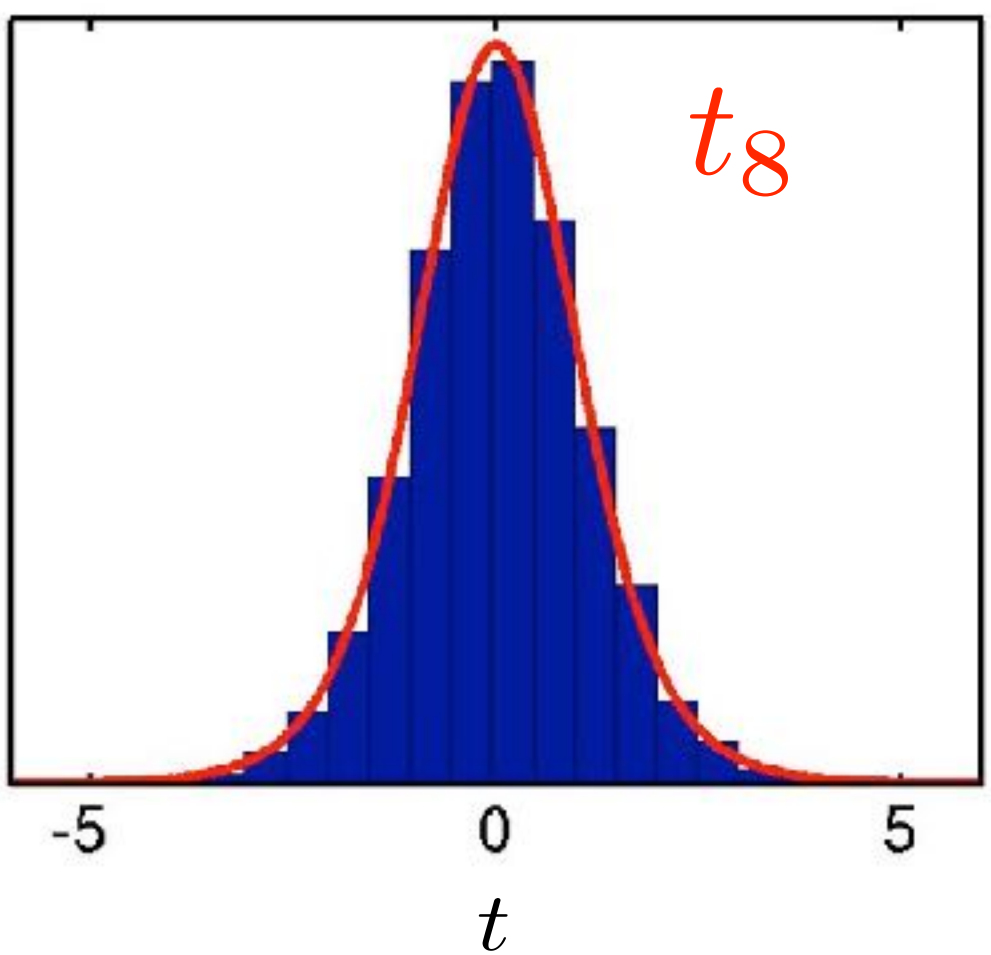
1. A null-hypothesis
2. A test-statistic
3. A null-distribution



Or that in ~5% of the cases t > 1.99.
When the null-hypothesis is true.

Tools of classical inference

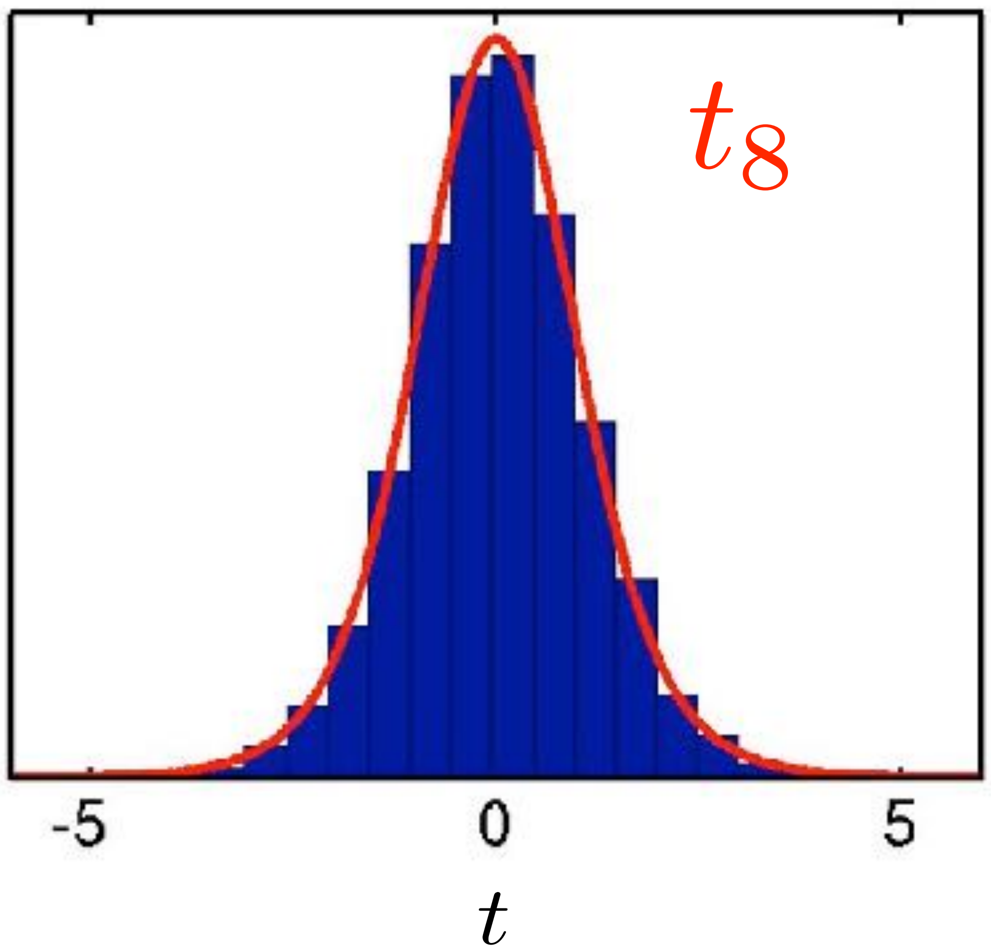
1. A null-hypothesis
2. A test-statistic
3. A null-distribution



And best of all: This distribution is known *i.e.* one can calculate it.
Much as one can calculate sine or cosine

Tools of classical inference

1. A null-hypothesis
2. A test-statistic
3. A null-distribution



And best of all: This distribution is known *i.e.* one can calculate it.
Much as one can calculate sine or cosine

Provided that $\mathbf{e} \sim N(0,\sigma^2)$

An example experiment

1. A null-hypothesis
2. A test-statistic
3. A null-distribution

H_0: $\overline{x}_1 = \overline{x}_2$, H_1: $\overline{x}_1 > \overline{x}_2$

So, with these tools let us do an experiment

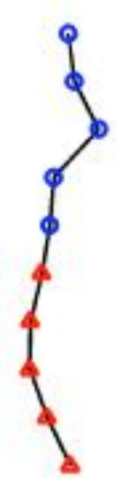
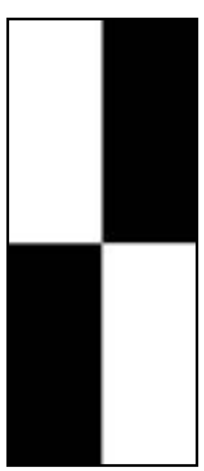
An example experiment

1. A null-hypothesis
2. A test-statistic
3. A null-distribution

H_0: $\overline{x}_1 = \overline{x}_2$, H_1: $\overline{x}_1 > \overline{x}_2$

$t_8 = 2.64$

So, with these tools let us do an experiment

 =  $\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$

$$t = \frac{\mathbf{c}^T \widehat{\boldsymbol{\beta}}}{\sqrt{\sigma^2}\sqrt{\mathbf{c}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{c}}} = \frac{1.53}{\sqrt{0.85}\sqrt{0.4}} = 2.64$$

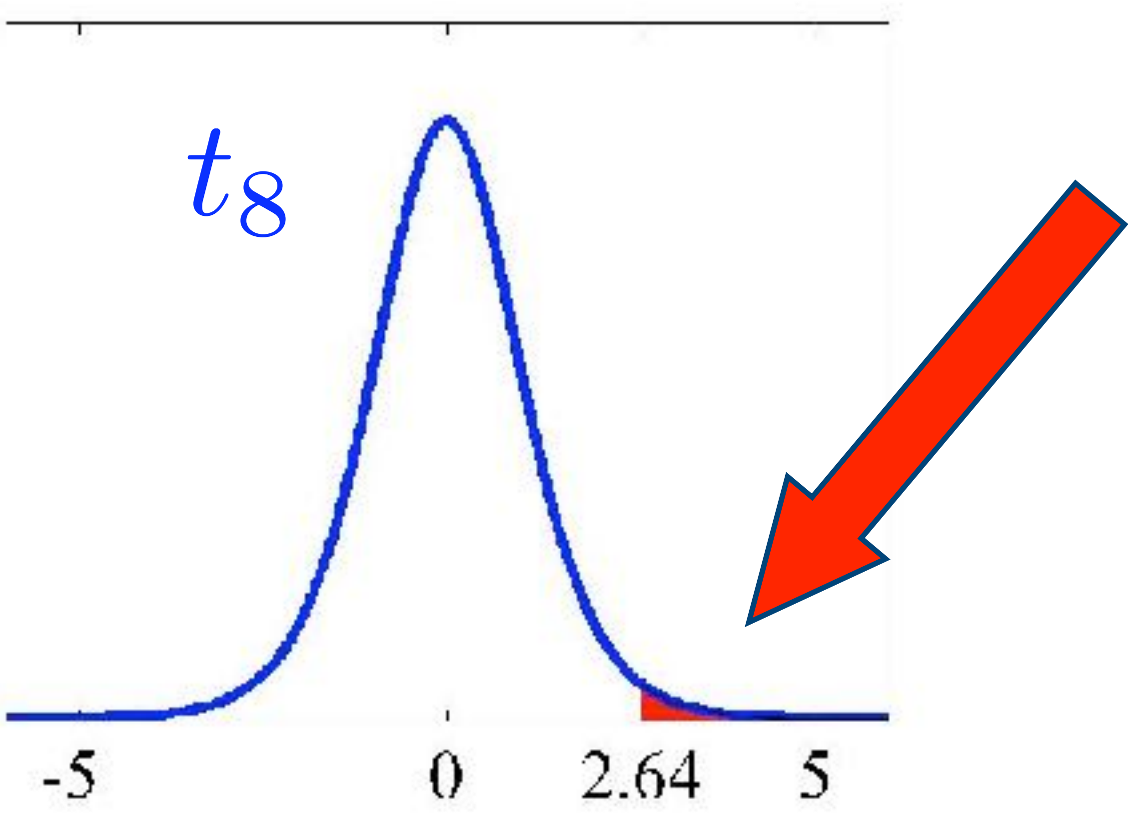
An example experiment

1. A null-hypothesis
2. A test-statistic
3. A null-distribution

H_0: $\overline{x}_1 = \overline{x}_2$, H_1: $\overline{x}_1 > \overline{x}_2$

$t_8 = 2.64$

So, with these tools let us do an experiment

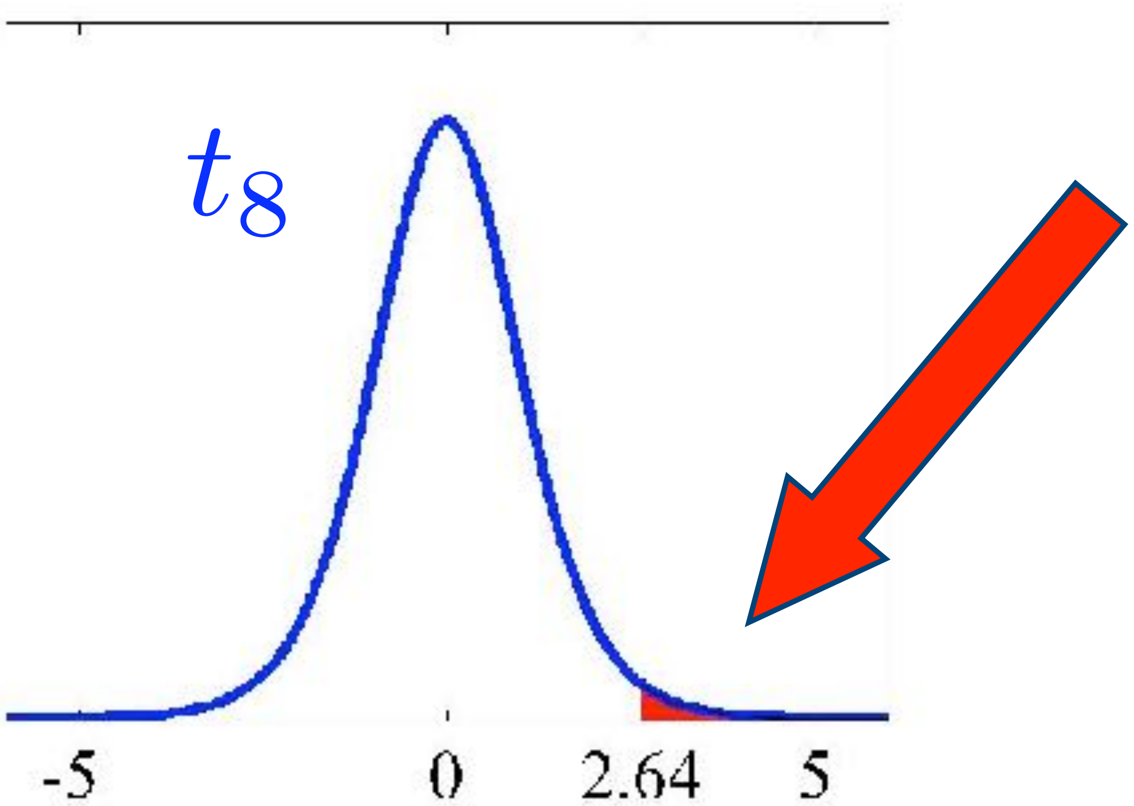


If the null-hypothesis is true, we would expect to have a ~1.46% chance of finding a t-value this large or larger

An example experiment

1. A null-hypothesis — H_0: $\overline{x}_1 = \overline{x}_2$, H_1: $\overline{x}_1 > \overline{x}_2$
2. A test-statistic — $t_8 = 2.64$
3. A null-distribution — $t_8 = 2.64$*

So, with these tools let us do an experiment



There is ~1.46% risk that we reject the null-hypothesis (i.e. claim we found something) when the null is actually true. We can live with that (well, I can).

False positives/negatives

- I am sure you have all heard about "false positives" and "false negatives".
- But what does that actually mean?

False positives/negatives

- I am sure you have all heard about "false positives" and "false negatives".
- But what does that actually mean?
- We want to perform an experiment and as part of that we define a null-hypothesis, e.g. $H_0: \ \mu = 0$
- Now what can happen?

False positives/negatives

- I am sure you have all heard about "false positives" and "false negatives".
- But what does that actually mean?
- We want to perform an experiment and as part of that we define a null-hypothesis, e.g. $H_0: \ \mu = 0$
- Now what can happen?

$$\left.\begin{array}{l} H_0 \text{ is true} \\ H_0 \text{ is false} \end{array}\right\} \text{True state of affairs}$$

False positives/negatives

- I am sure you have all heard about "false positives" and "false negatives".
- But what does that actually mean?
- We want to perform an experiment and as part of that we define a null-hypothesis, e.g. $H_0 : \mu = 0$
- Now what can happen?

H_0 is true
H_0 is false
} True state of affairs

We don't reject H_0
We reject H_0
} Our decision

False positives/negatives

H_0 is true
H_0 is false
True state of affairs

We don't reject H_0
We reject H_0
Our decision

	We don't reject H_0	We reject H_0
H_0 is true		
H_0 is false		

False positives/negatives

H_0 is true
H_0 is false
} True state of affairs

We don't reject H_0
We reject H_0
} Our decision

	We don't reject H_0	We reject H_0
H_0 is true	☺️	
H_0 is false		☺️

False positives/negatives

H_0 is true
H_0 is false } True state of affairs

We don't reject H_0
We reject H_0 } Our decision

	We don't reject H_0	We reject H_0
H_0 is true	☺️	False positive
H_0 is false	False negative	☺️

False positives/negatives

H_0 is true
H_0 is false
True state of affairs

We don't reject H_0
We reject H_0
Our decision

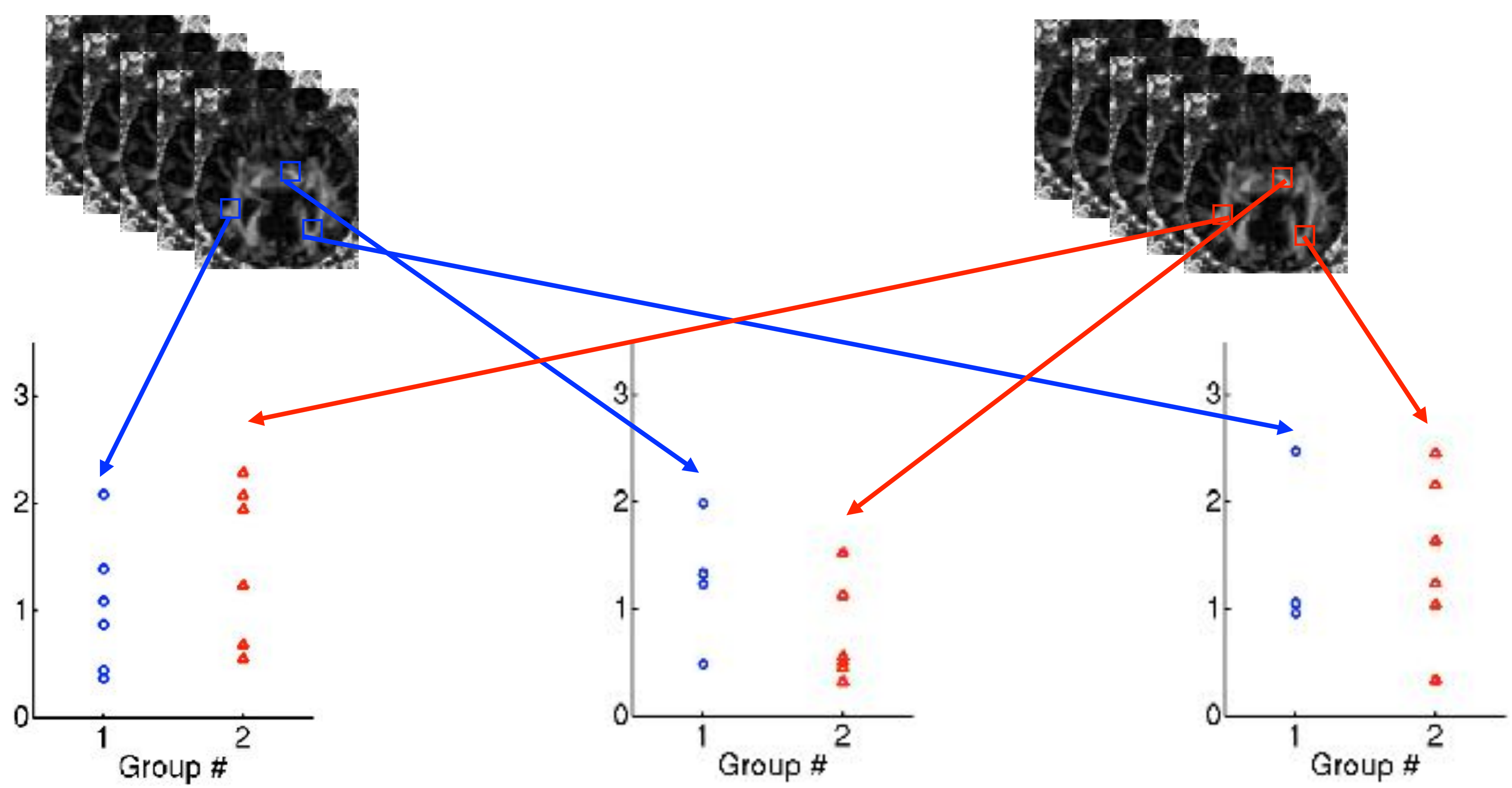
	We don't reject H_0	We reject H_0
H_0 is true	☺️	False positive Type I error
H_0 is false	False negative Type II error	☺️

Outline

- Null-hypothesis and Null-distribution
- Multiple comparisons and Family-wise error
- Different ways of being surprised
    - Voxel-wise inference (Maximum z)
    - Cluster-wise inference (Maximum size)
- Parametric vs non-parametric tests
- Enhanced clusters
- FDR - False Discovery Rate

Multiple Comparisons

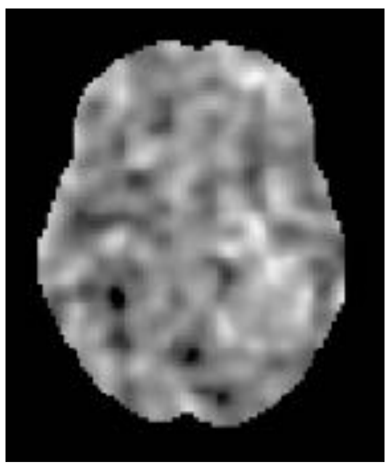
- In neuroimaging we typically perform many tests as part of a study



Different here? Maybe here? Or here? . . .

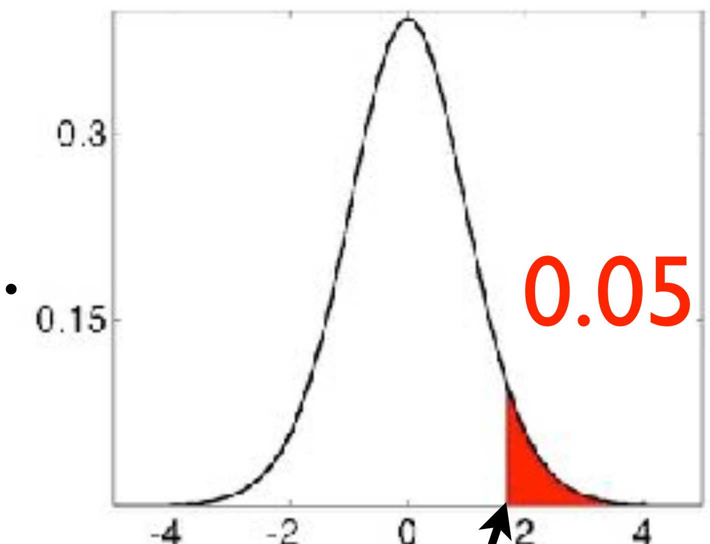
. . .

What happens when we apply this to imaging data?



z-map where each voxel ~*N*.
Null-hypothesis true everywhere, i.e.
NO ACTIVATIONS

z



1.64

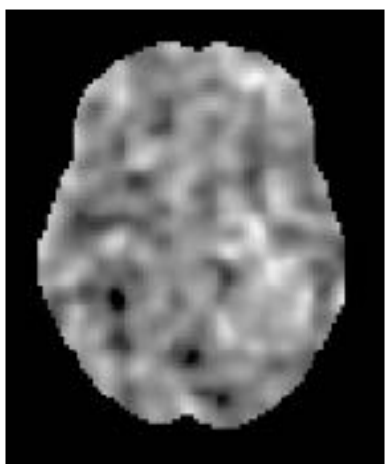
z-map thresholded at 1.64

16 clusters
288 voxels
~5.5% of the voxels

That's a LOT of false positives

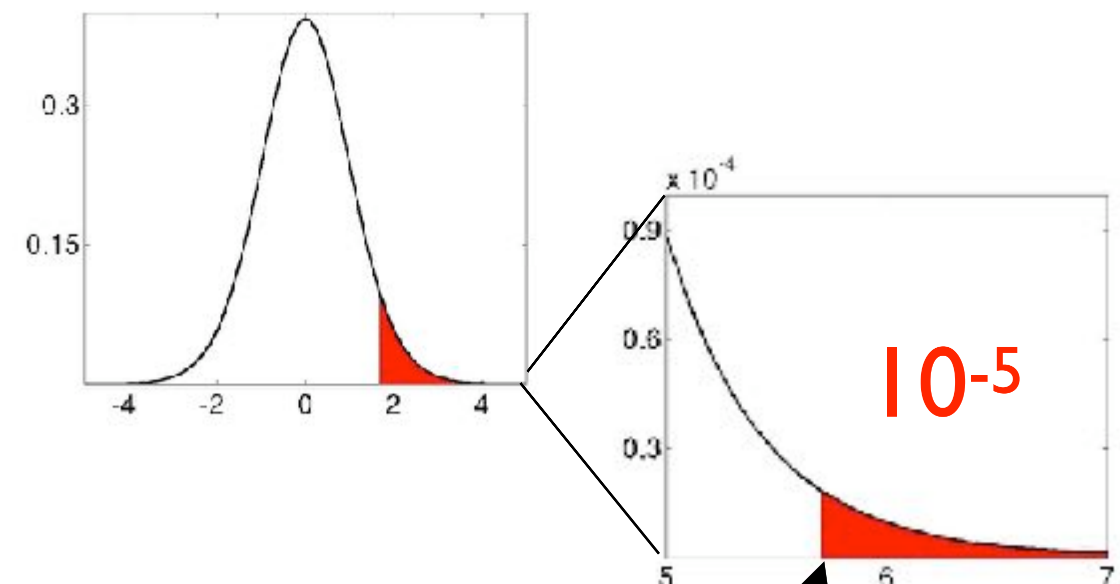
Italians doing maths: The Bonferroni correction

Bonferroni says threshold at α divided by # of tests



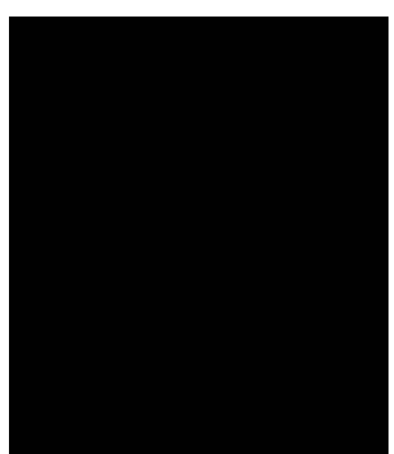
5255 voxels

$0.05/5255 \approx 10^{-5}$



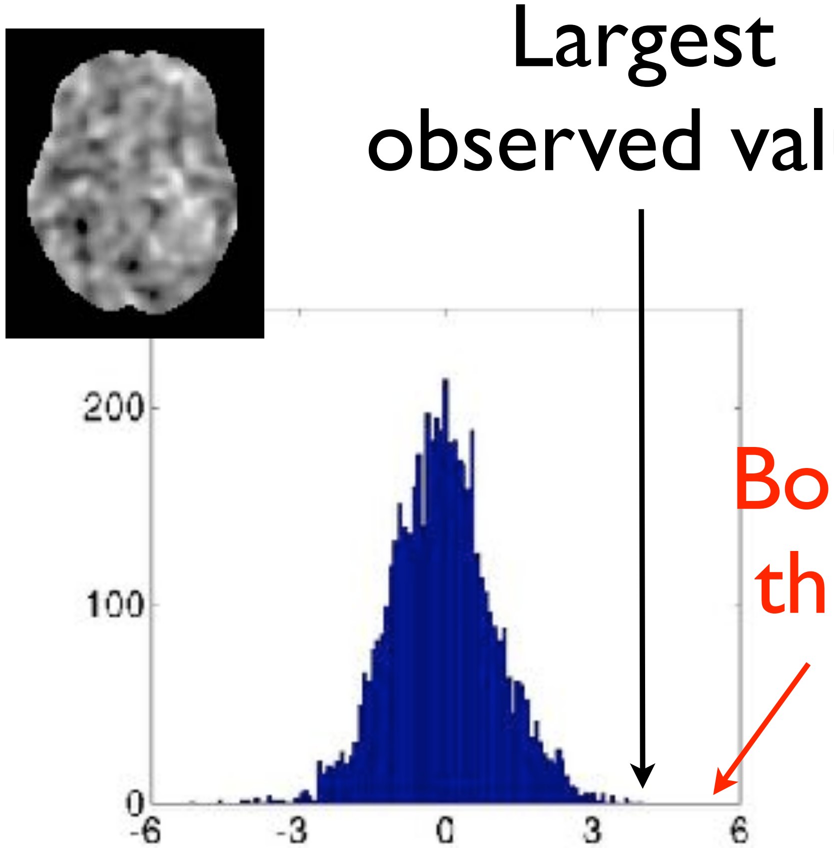
5.65

z-map thresholded at 5.65

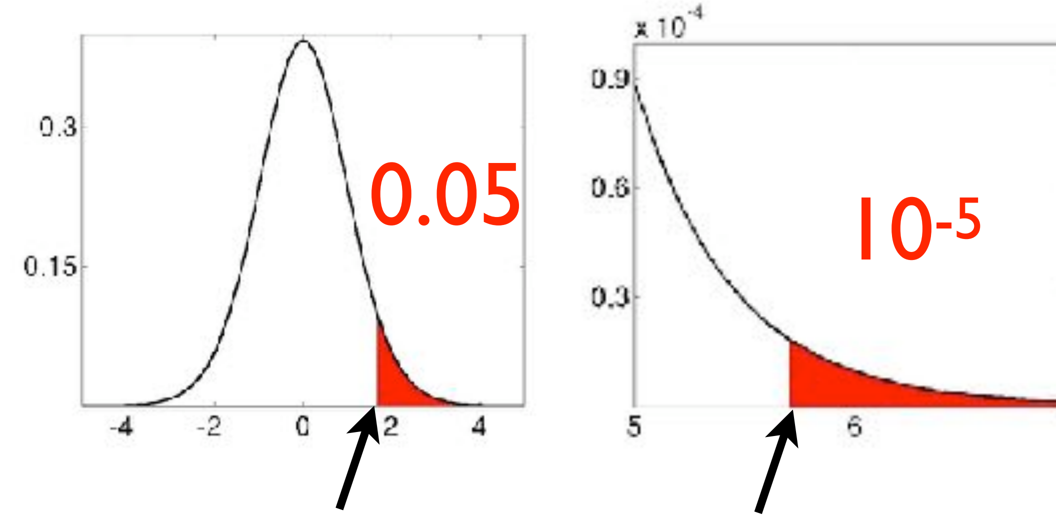


No false positives.
Hurrah for Italy!

But … doesn’t 5.65 sound very high?



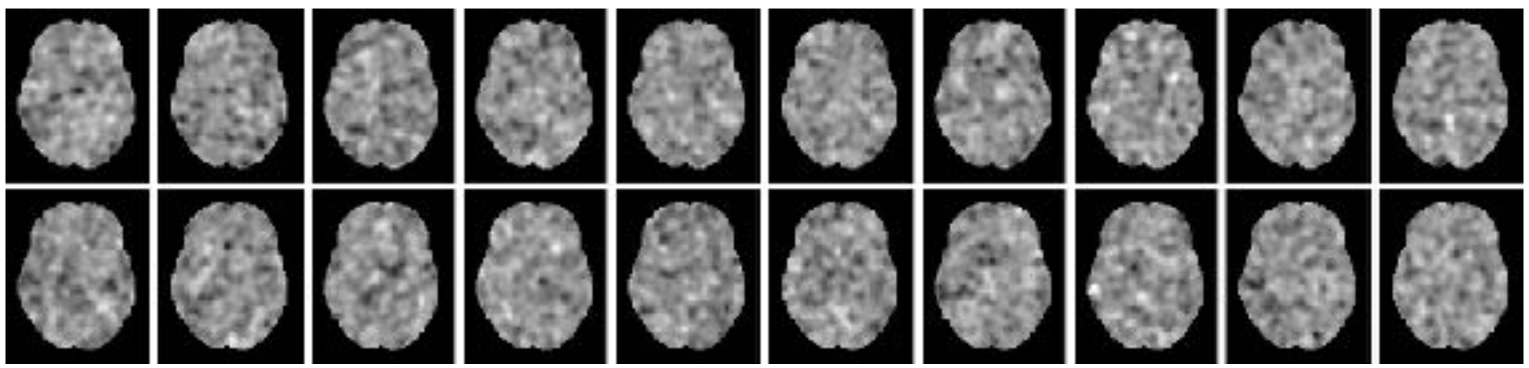
Observed values
in the z-map



So what do we want then?

Family-wise error

Let's say we perform a series of identical studies



Each z-map is the end result of a study

Let us further say that the null-hypothesis is true

We want to threshold the data so that only once in 20 studies do we find a voxel above this threshold

But how do we find such a threshold?