Statistical inference with cluster permutation testing

Robert Oostenveld

Donders Institute, Radboud University, Nijmegen, NL

NatMEG, Karolinska Institute, Stockholm, SE

Talk outline

Types of statistics

- descriptive
- inferential

Parametric statistics

Non-parametric randomization test

Clustering-based statistics

What types of statistics do we have?

"Data don't make any sense,
we will have to resort to statistics."

How do large distributions of "something" behave?

- Binomial, Normal, Poisson

How can I describe (or summarize) a distribution?

- Mean, standard deviation, variance, kurtosis

How can I make a decision or draw a conclusion?

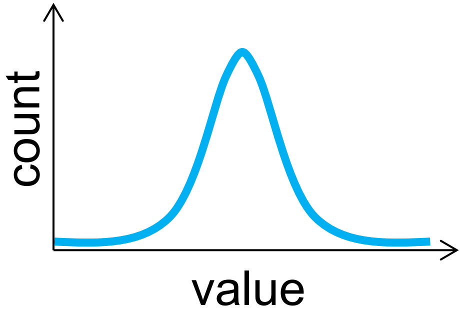
- Inferential statistics, hypothesis testing

Inferential parametric statistics

You make N observation and want to find whether some hypothesis "H1" holds

Step 1: Gathering data

Observation	Value
0	2.5
1	-3.2
.	
.	
.	
N	2.4

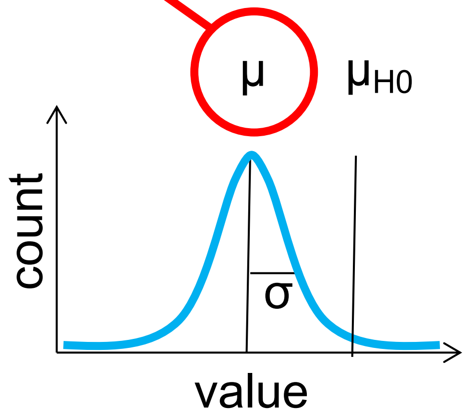


Inferential parametric statistics

You make N observation and want to find whether some hypothesis “H1” holds

Step 2: Statistical testing

Observation	Value
0	2.5
1	-3.2
. . .	
N	2.4



Determine probability of *t* under “H0”

$$t = \frac{\mu - \mu_{H0}}{\sigma/\sqrt{N}}$$

If the observed *t* sufficiently unlikely, reject H0 in favour of H1

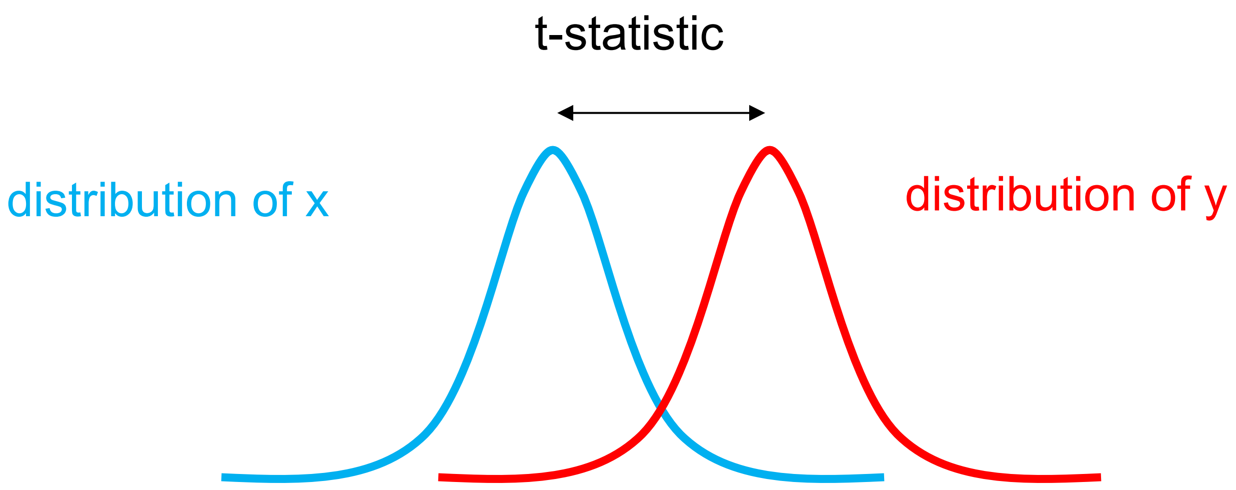
Inferential parametric statistics

Observations in condition 1:

{x1, x2, x3, x4, …}

Observations in condition 2:

{y1, y2, y3, y4, …}



Talk outline

Types of statistics

descriptive

inferential

Parametric statistics

Non-parametric randomization test

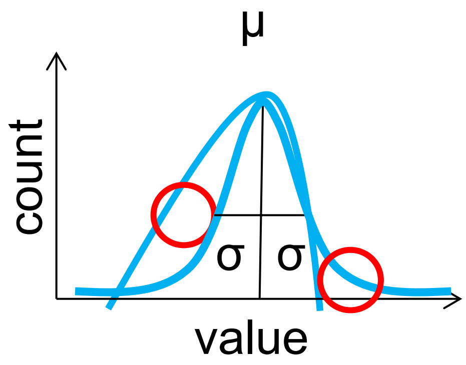
Clustering-based statistics

Problem 1: Distribution of the data and test statistic

You make N observation and want to find whether some hypothesis H1 is true.

The first problem is that this requires a *known distribution* of the test statistic.

Observation	Value
0	2.5
1	-3.2
.	
.	
.	
N	2.4



Problem 2: Multiple comparisons

Typical ERPs

- 64 channels, 250 timepoints
- 16.000 datapoints, repeated over conditions and subjects
- Thousands of parameters and t-values

Chance of false alarm is 5% for every test

With 16.000 data points we expect 800 false alarms in an ERP!

Similar problems for time-frequency ERSPs, connectivity, etc.

Solutions to control the FWER

Bonferroni correction

- Reduce the alpha threshold by a factor N, for example from 5% to 2.5% when N=2.

Use the false discovery rate (FDR)

- Sort the probabilities and adjust the threshold such that the expected proportion of false alarms is controlled
- Slightly less conservative than Bonferroni

Use a Monte Carlo approximation of the randomization distribution of the maximum statistic

Talk outline

Types of statistics

- descriptive
- inferential

Parametric statistics

Non-parametric randomization test

Clustering-based statistics

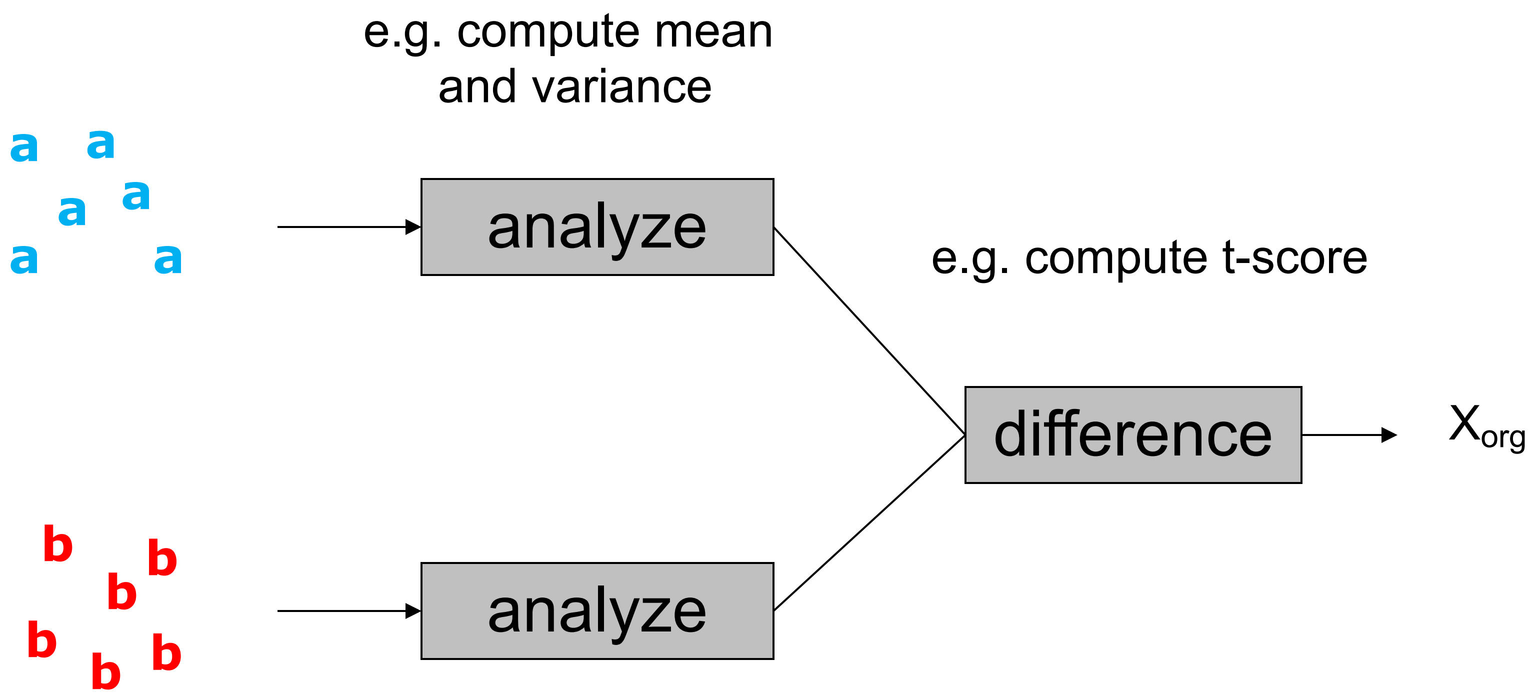
Randomization test: general principle

- *Independent variable: condition*
- *Dependent variable: data*

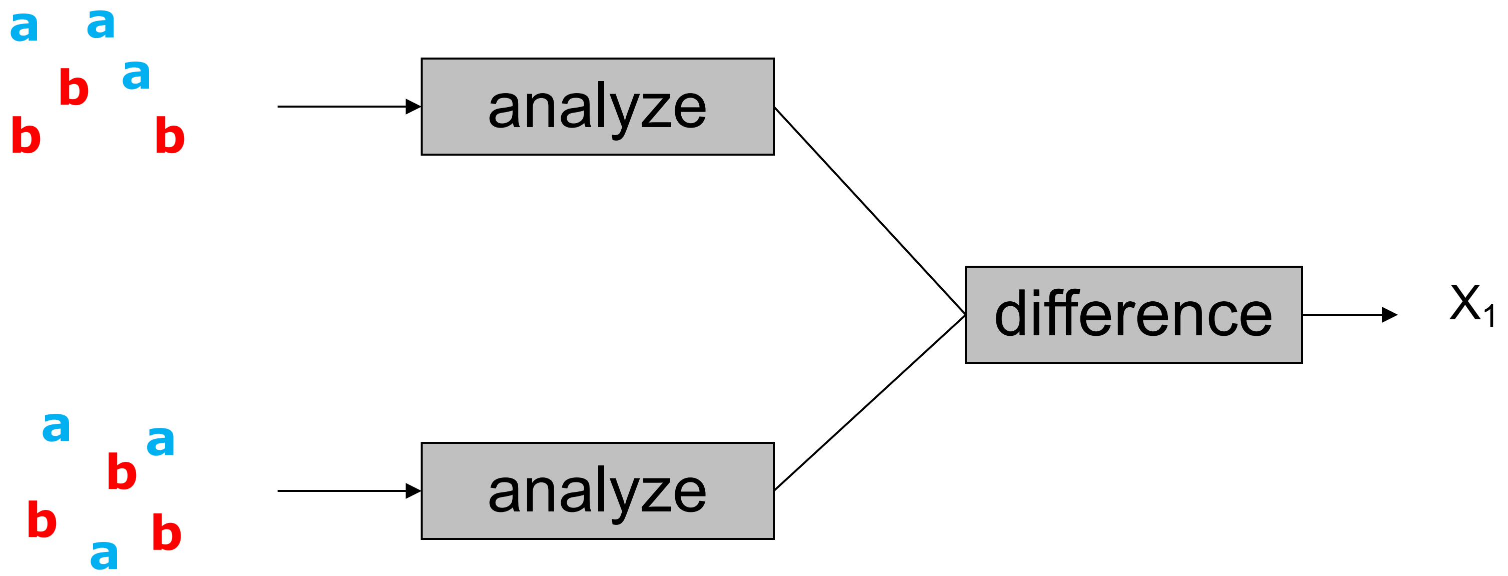
H0: the data is **independent** from the condition in which it was observed

The data in the two conditions is **not** different

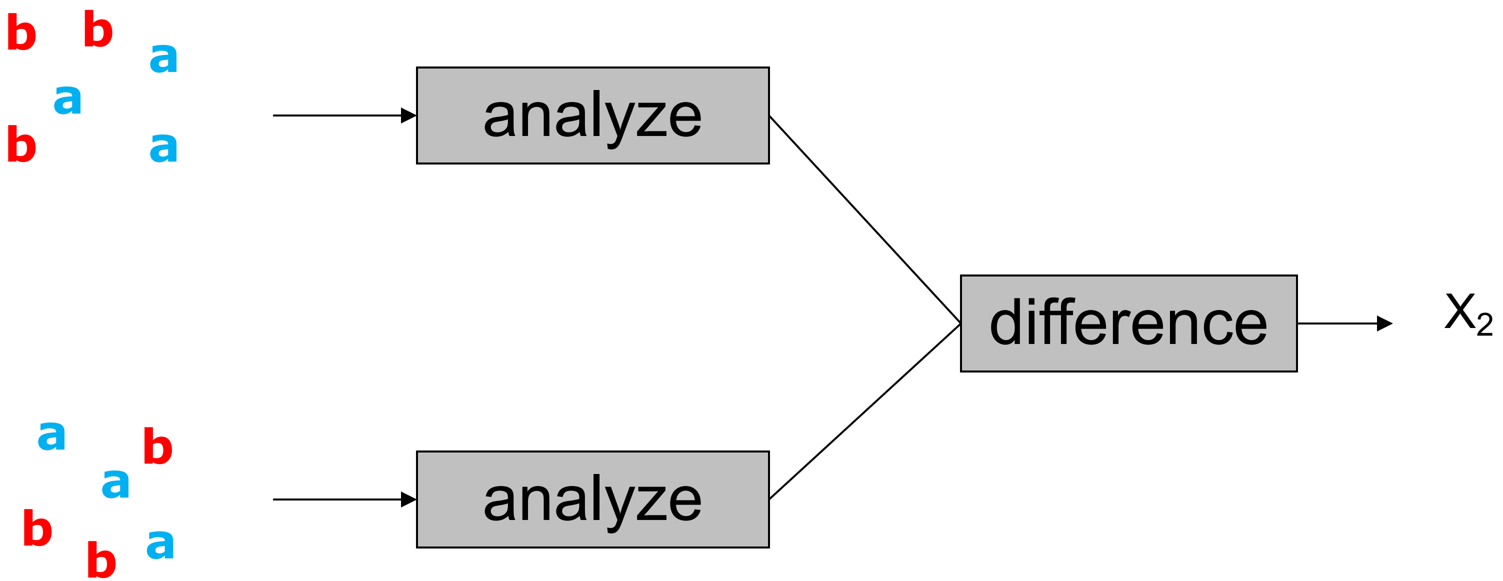
Randomization approach



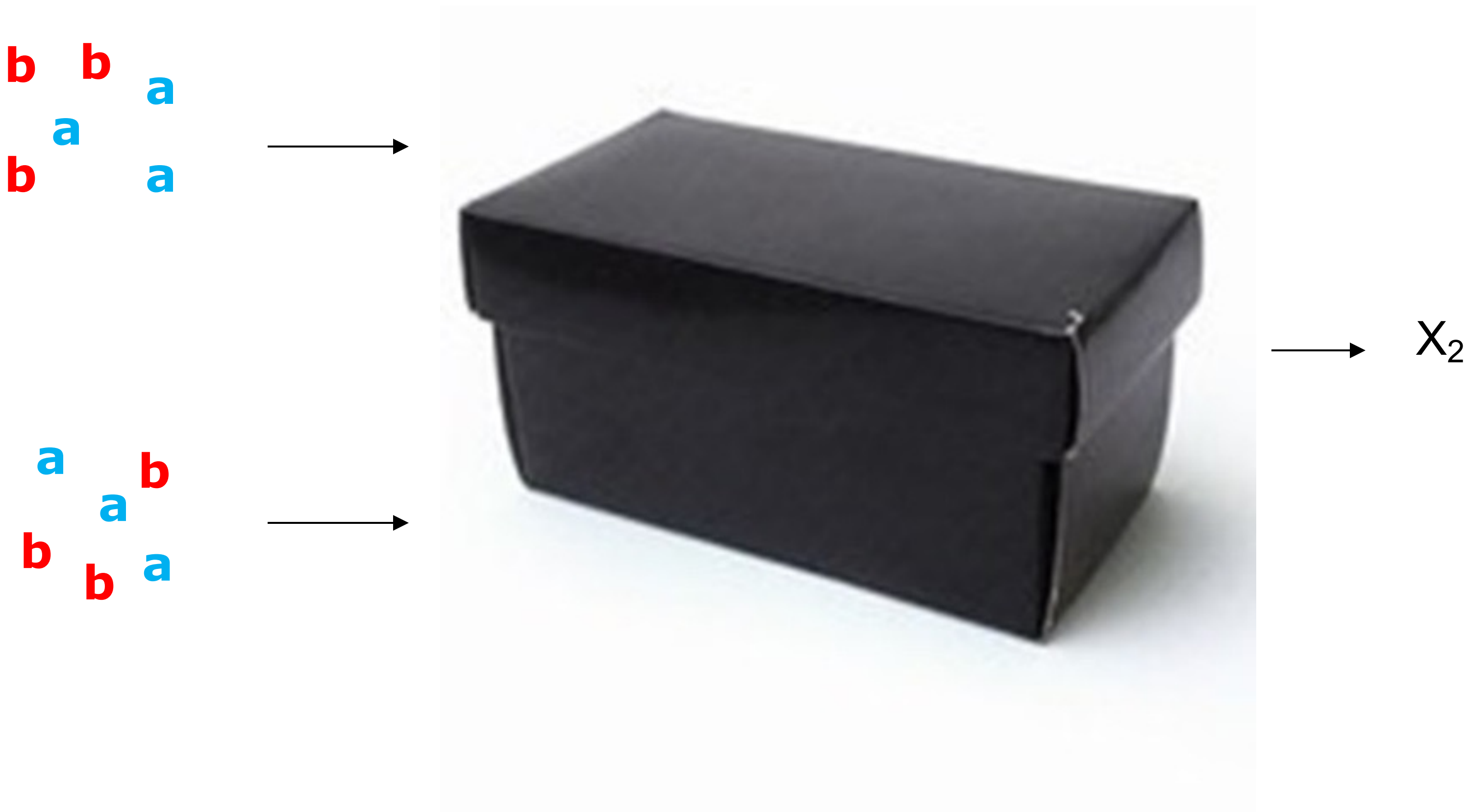
Randomization approach



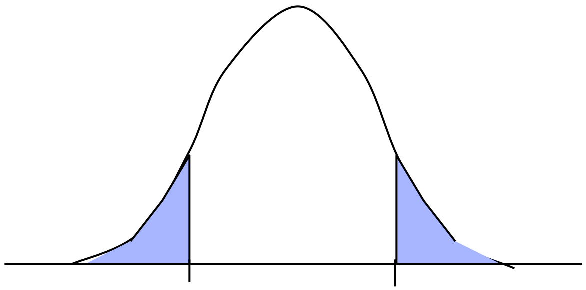
Randomization approach



Randomization approach



Distribution of “x” can take any shape



Non-parametric statistics

Randomization of independent variable

Hypothesis is about data, not about the specific parameter

The distribution of the statistic of interest "x" is approximated using the Monte-Carlo approach, i.e. by random sampling

H0 is tested by comparing the observed statistic against the randomization distribution

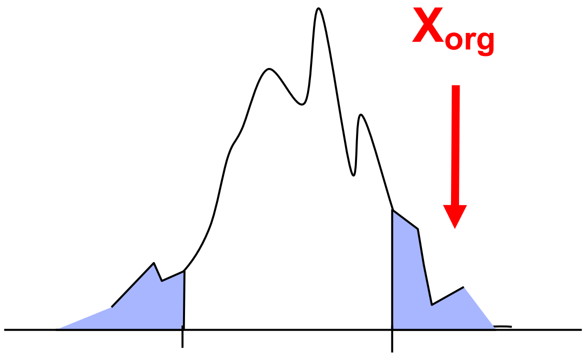
Avoid the multiple comparison problem

The statistic "x" can be anything

Rather than testing everything, only test the most extreme observation (i.e. the max statistic)

Compute the randomization distribution for the most extreme statistic over all channels/times/frequencies

Note that often we compute **two** extrema, one for each tail



Talk outline

Types of statistics

- descriptive
- inferential

Parametric statistics

Non-parametric randomization test

Clustering-based statistics

Increasing the sensitivity

Conventional is univariate parametric

Our approach is to consider the data

- Many channels, timepoints, frequencies
- Massive univariate
- Multiple comparison problem

EEG is relatively blurry over channels, time, and frequency, so there is quite some structure in the data

Increasing the sensitivity

channel/time/frequency points are not independent and are expected to show similar behaviour

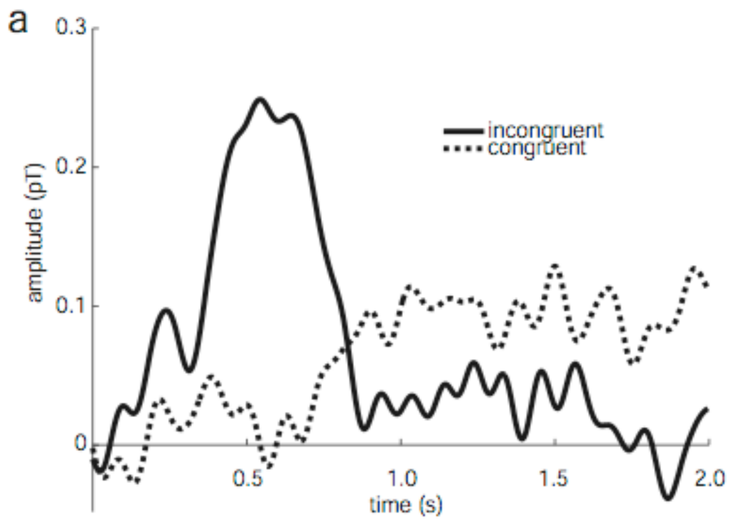
combine neighbouring samples into clusters -> "accumulate the evidence" = cluster-based statistics

avoid the MCP by comparing the largest observed cluster versus the randomization distribution of the largest clusters

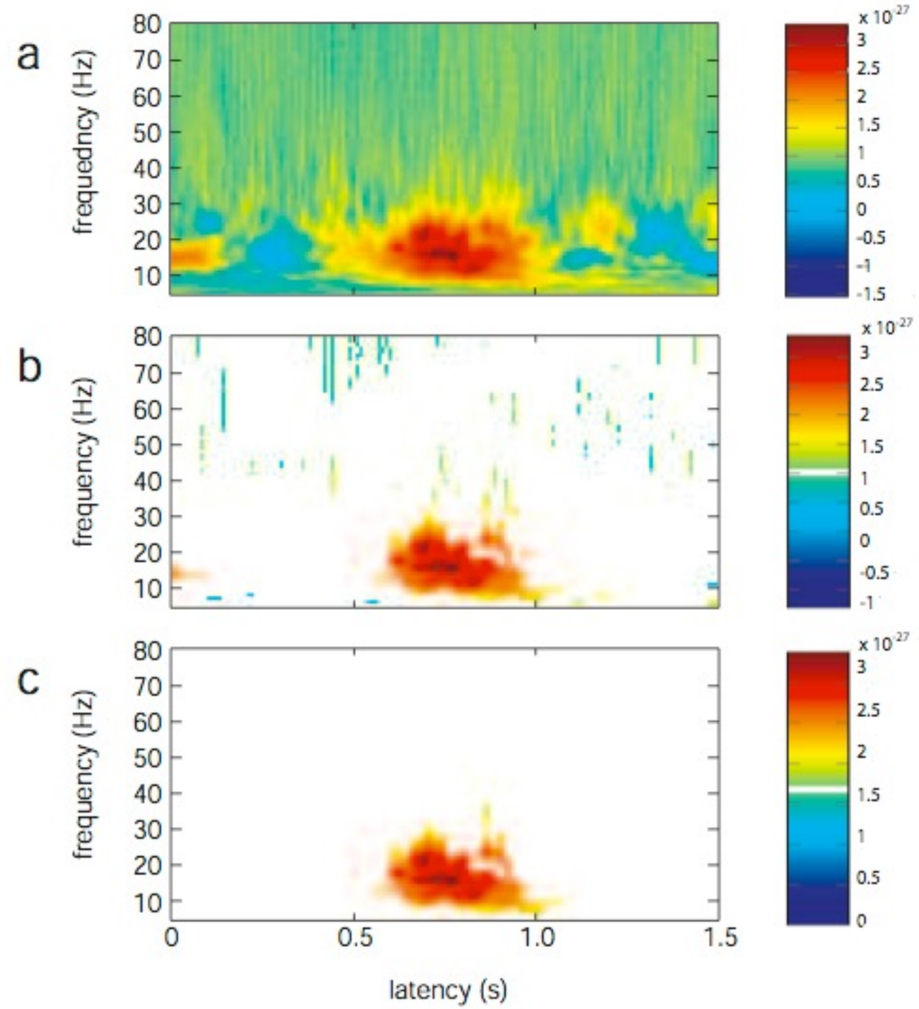
Avoid multiple comparisons

Increase sensitivity

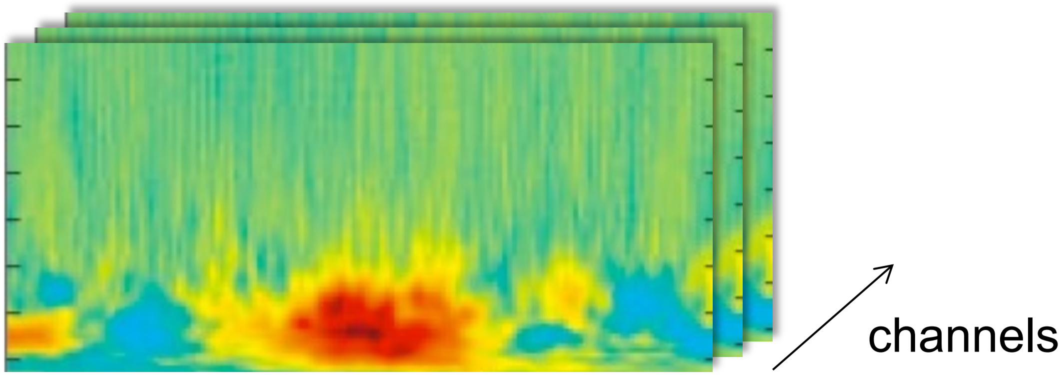
Clustering in time

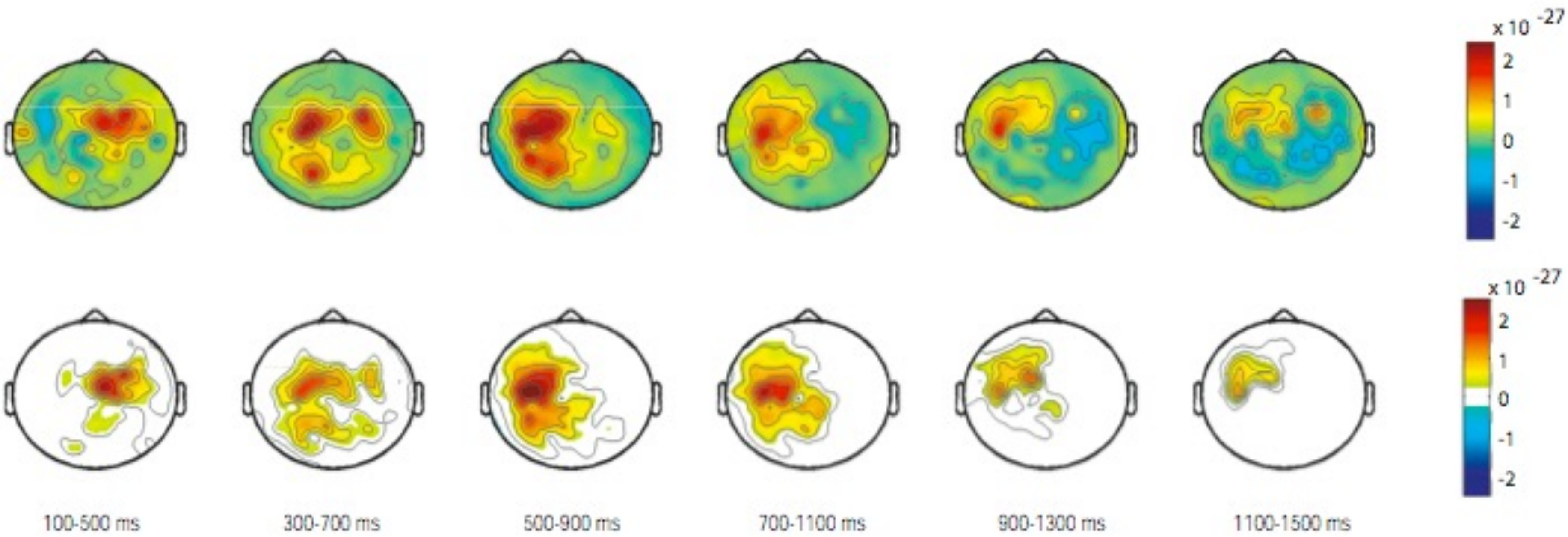


Clustering in time and frequency



Clustering in time, frequency and space





Toy example

Toy example: Original observation

null hypothesis: condition A = condition B

Condition A	Condition B
S1_a	S1_b
S2_a	S2_b
S3_a	S3_b
S4_a	S4_b
S5_a	S5_b
S6_a	S6_b
S7_a	S7_b
S8_a	S8_b
S9_a	S9_b
S10_a	S10_b

Toy example: 1st permutation

null hypothesis: condition A = condition B

Condition A		Condition B
S1_a		S1_b
S2_b	↔	S2_a
S3_a		S3_b
S4_a		S4_b
S5_b	↔	S5_a
S6_b	↔	S6_a
S7_a		S7_b
S8_a		S8_b
S9_a		S9_b
S10_b	↔	S10_a

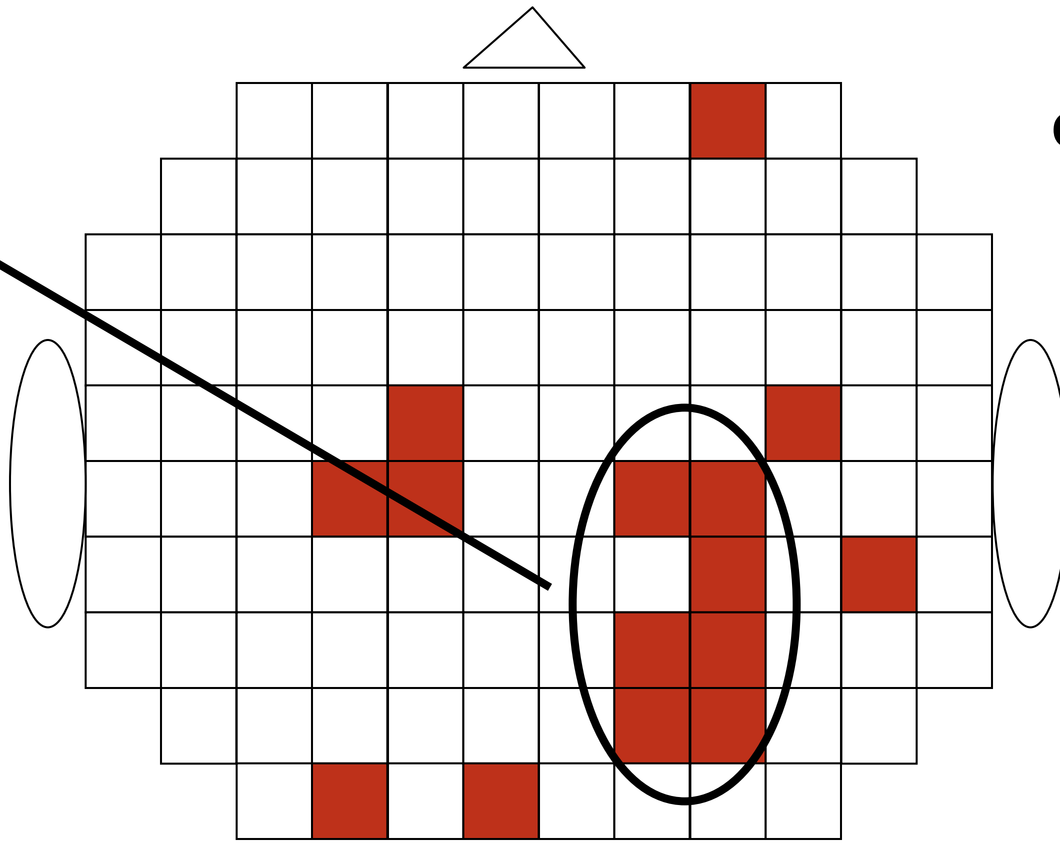
Toy example: 2nd permutation

null hypothesis: condition A = condition B

Condition A	Condition B
S1_b	S1_a
S2_a	S2_b
S3_b	S3_a
S4_a	S4_b
S5_a	S5_b
S6_b	S6_a
S7_a	S7_b
S8_b	S8_a
S9_b	S9_a
S10_a	S10_b

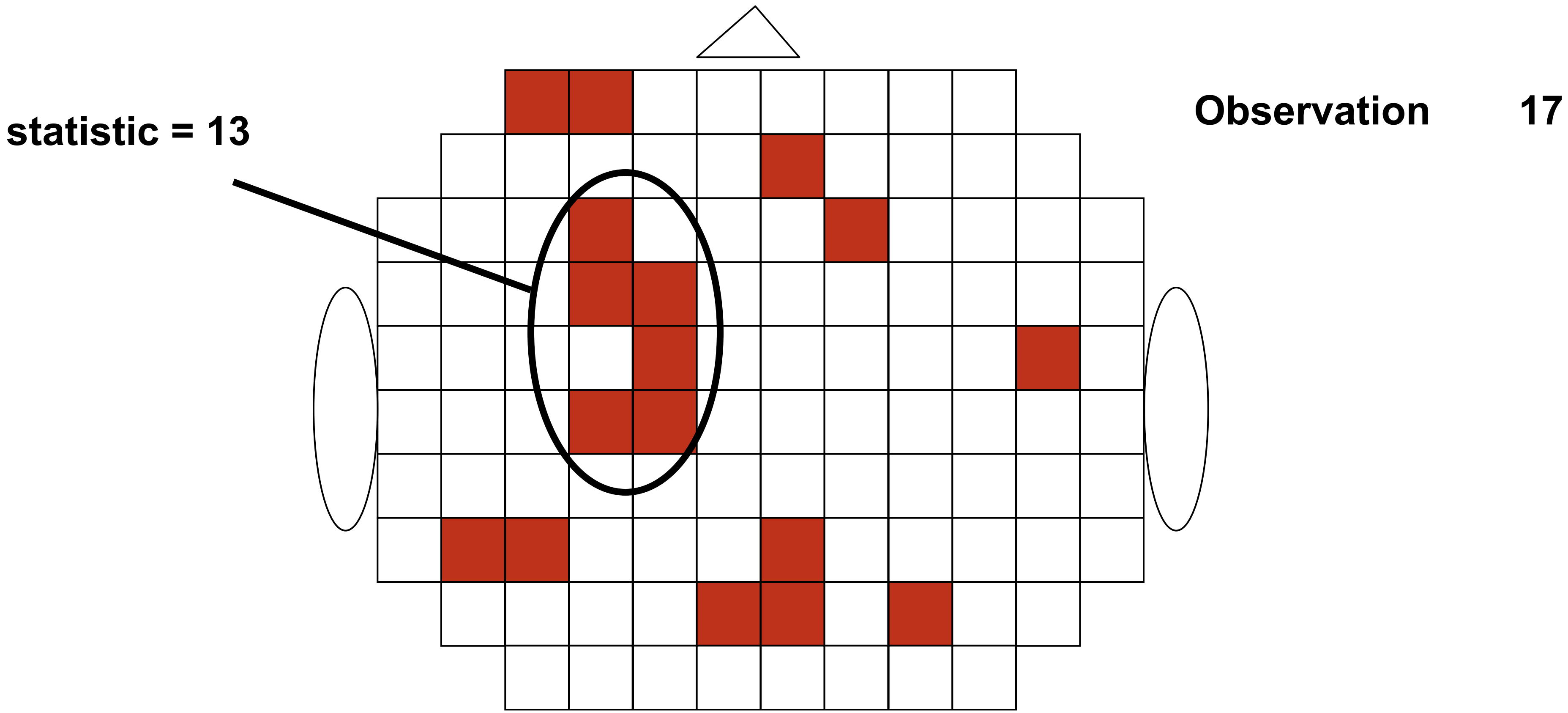
Toy example: Original observation

statistic = 17

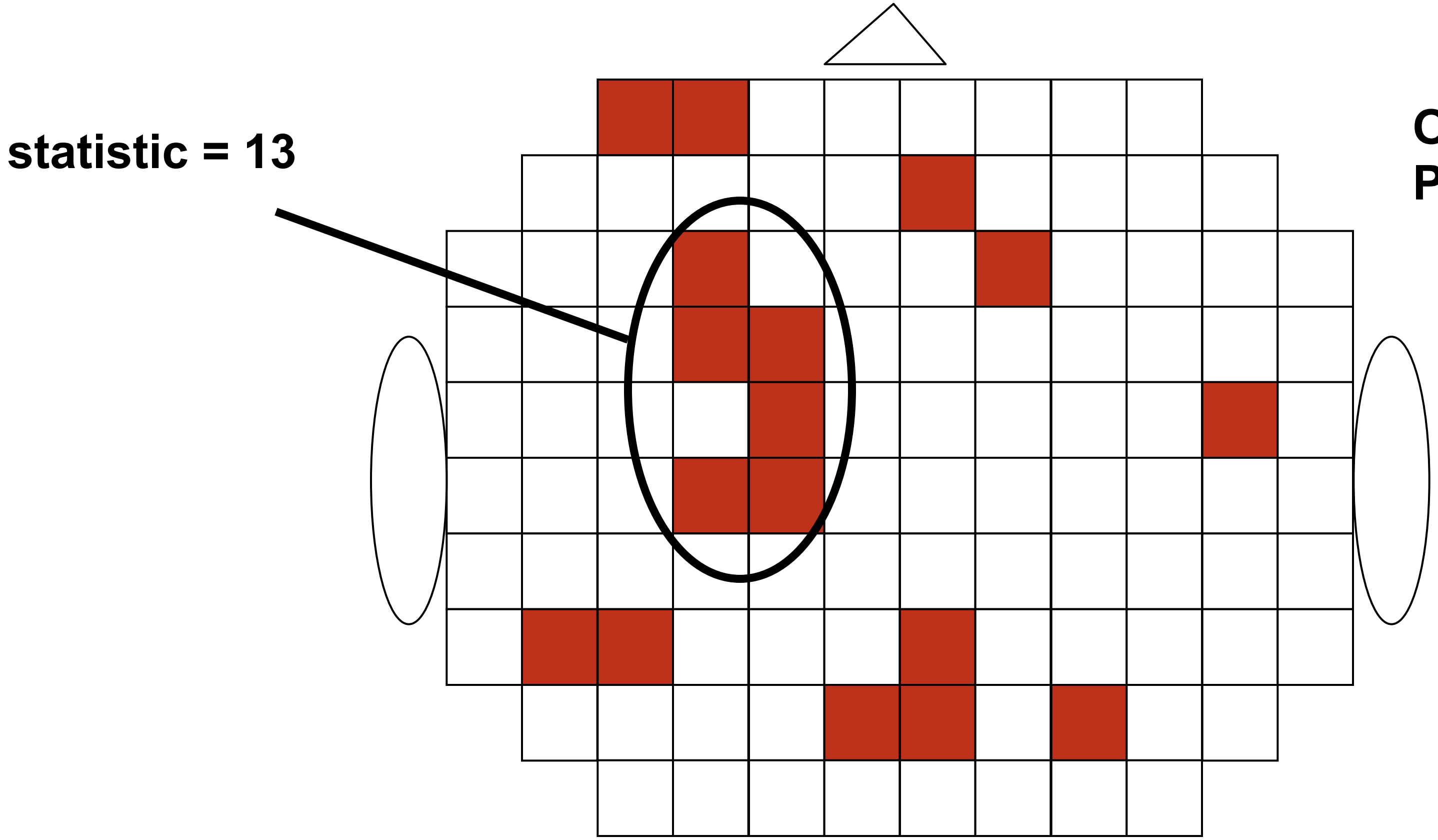


Observation 17

Toy example: 1st permutation

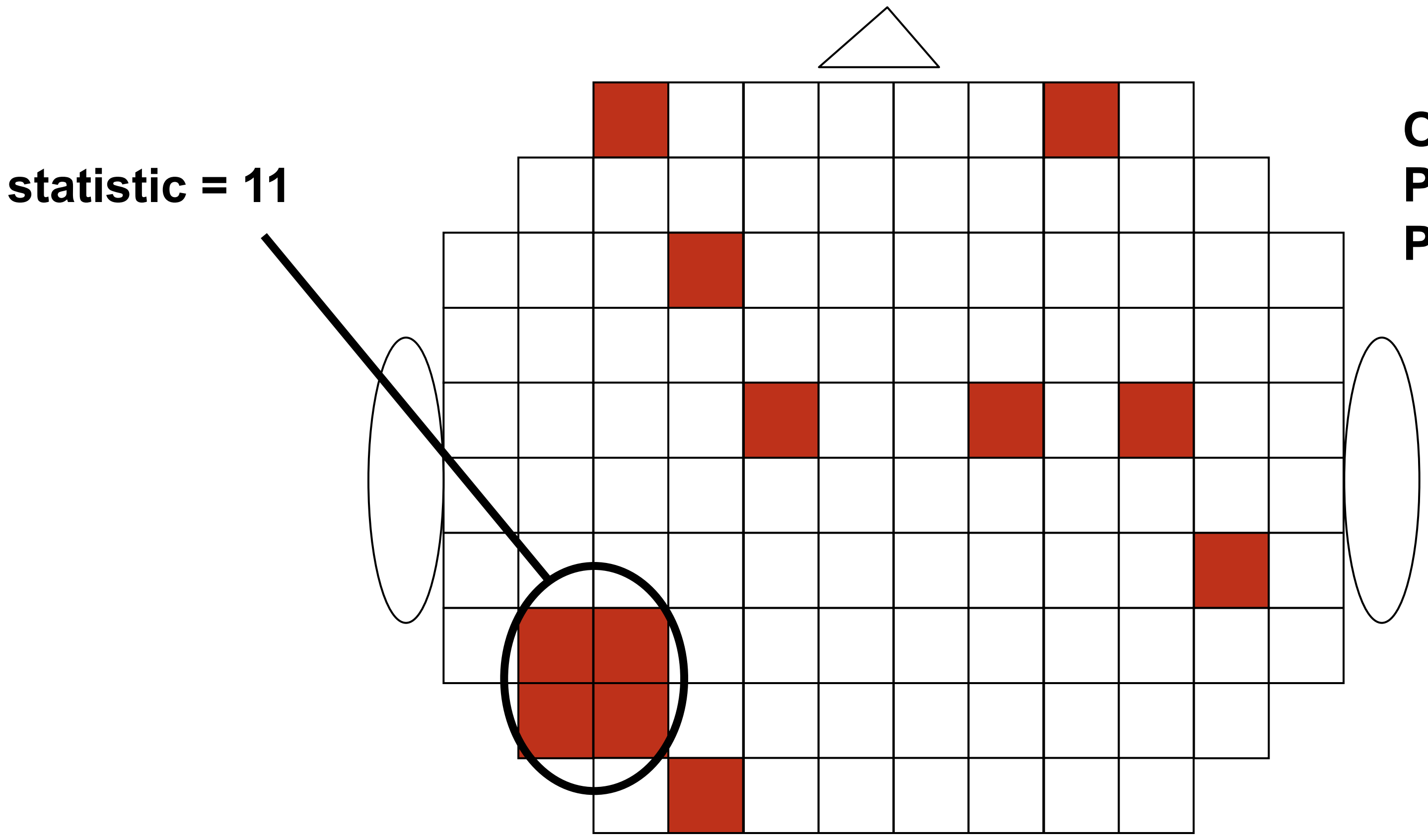


Toy example: 1st permutation



Observation 17
Permutation 1 13

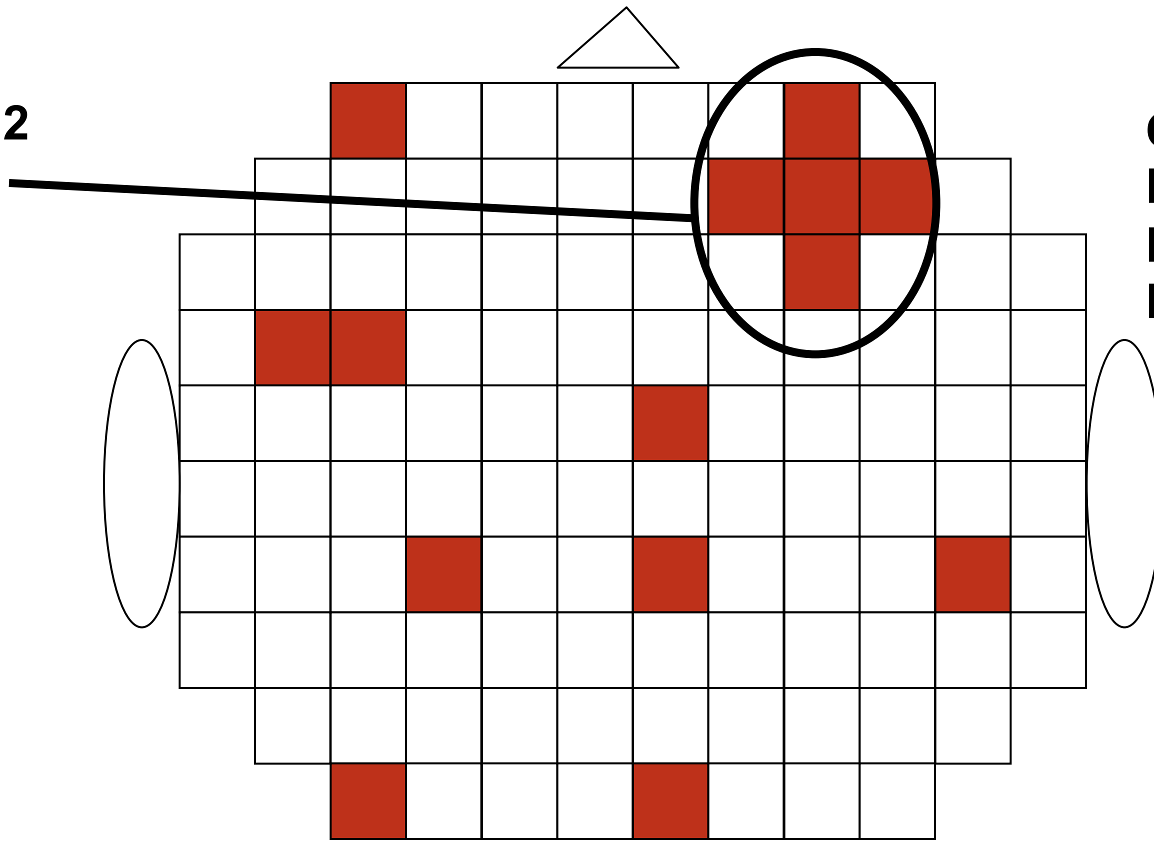
Toy example: 2nd permutation



Observation 17
Permutation 1 13
Permutation 2 11

Toy example: 3rd permutation

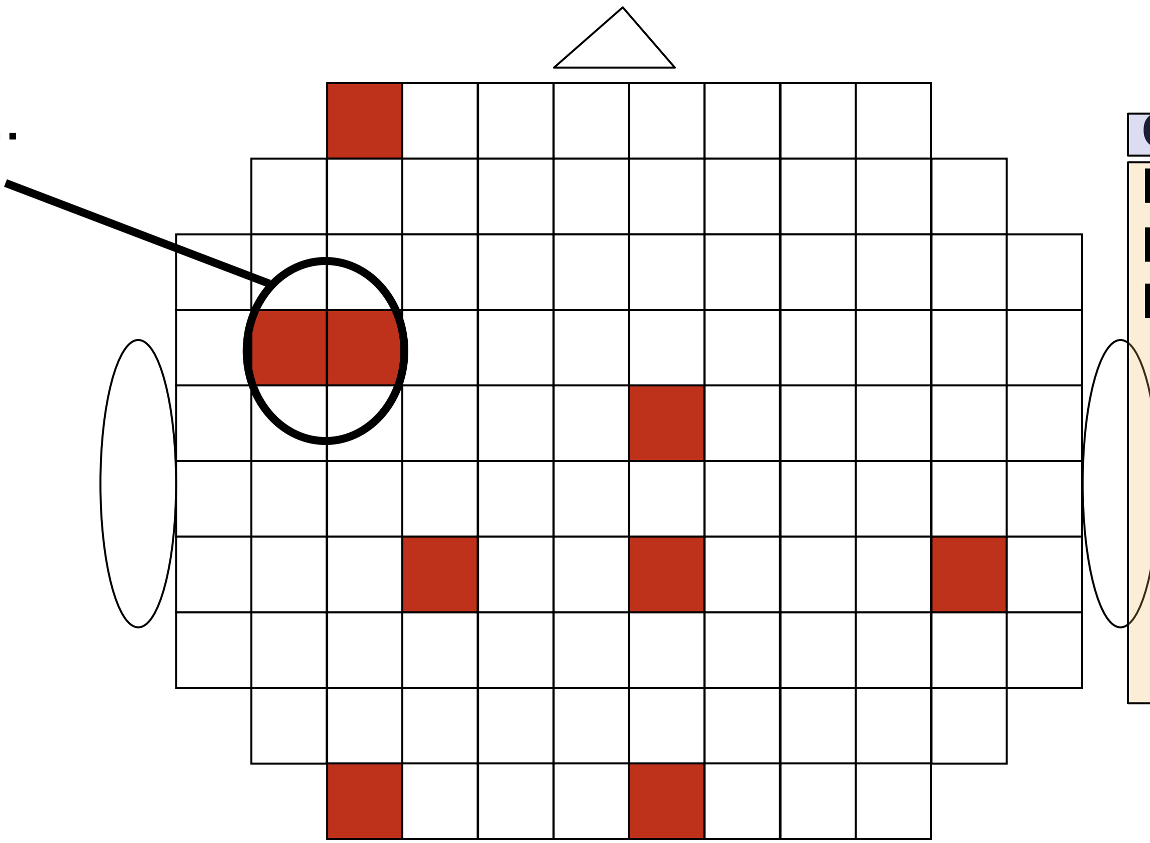
statistic = 12



Observation	**17**
Permutation 1	**13**
Permutation 2	**11**
Permutation 3	**12**

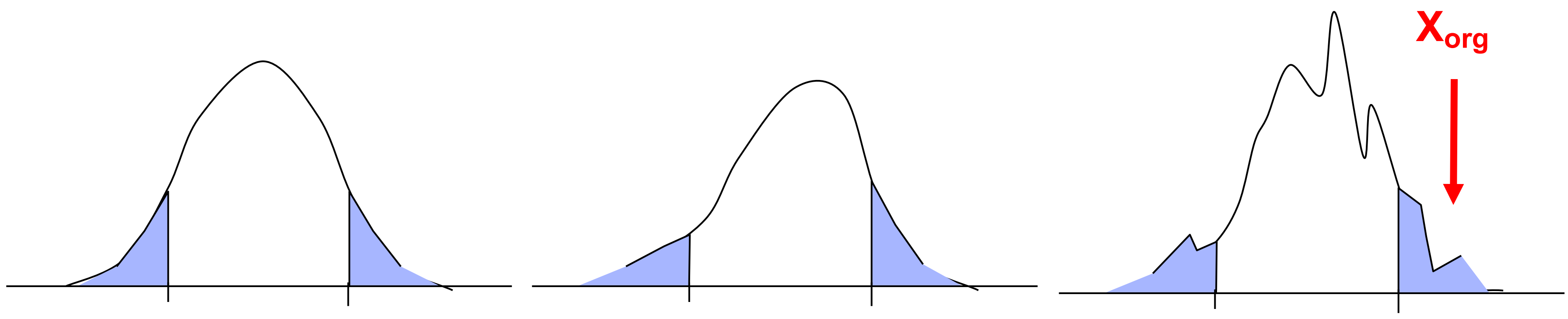
Toy example: N^{th} permutation

statistic = …



Observation	**17**
Permutation 1	**13**
Permutation 2	**11**
Permutation 3	**12**

Assess the likelihood of the *observed max cluster size* given the randomization distribution



General summary

- A formal hypothesis can be tested with randomization test
  - control the chance of false positives
  - reduce the false negative rate
- Multiple comparison problem
  - ERP - one hypothesis per channel-time
  - ERSP - one hypothesis per channel-time-frequency
  - Solution: use one hypothesis for all datapoints
- Increase sensitivity
  - using clusters to capture the structure in the data