Robust statistics

Arnaud Delorme

Robust statistics

Parametric & non-parametric statistics: use mean and standard deviation (t-test, ANOVA, …)

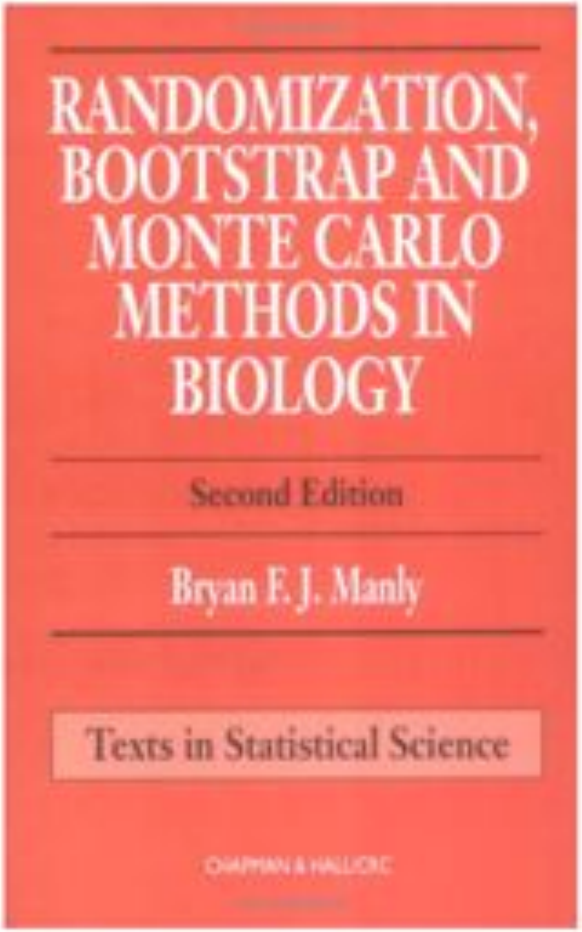
Bootstrap and permutation methods: shuffle/bootstrap data and recompute measure of interest. Use the tail of the distribution to asses significance.

Correction for multiple comparisons: computing statistics on time(/frequency) series requires correction for the number of comparisons performed.

Take-home messages

- *Look at your data! Show your data!*
- *A perfect & universal statistical recipe does not exist*
- *Keep exploring: there are many great options, most of them available in free softwares and toolboxes*

References



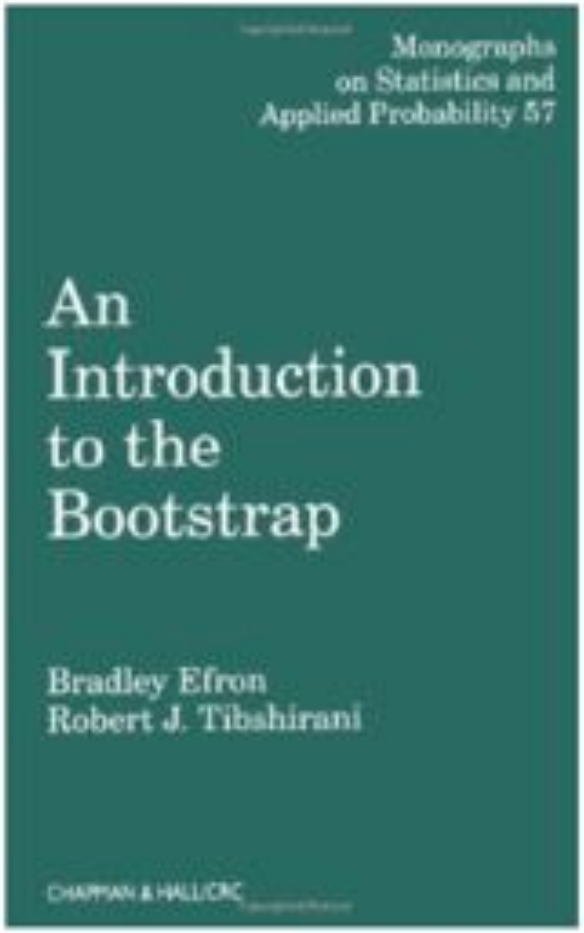
RANDOMIZATION, BOOTSTRAP AND MONTE CARLO METHODS IN BIOLOGY

Second Edition

Bryan F. J. Manly

Texts in Statistical Science

CHAPMAN & HALL/CRC

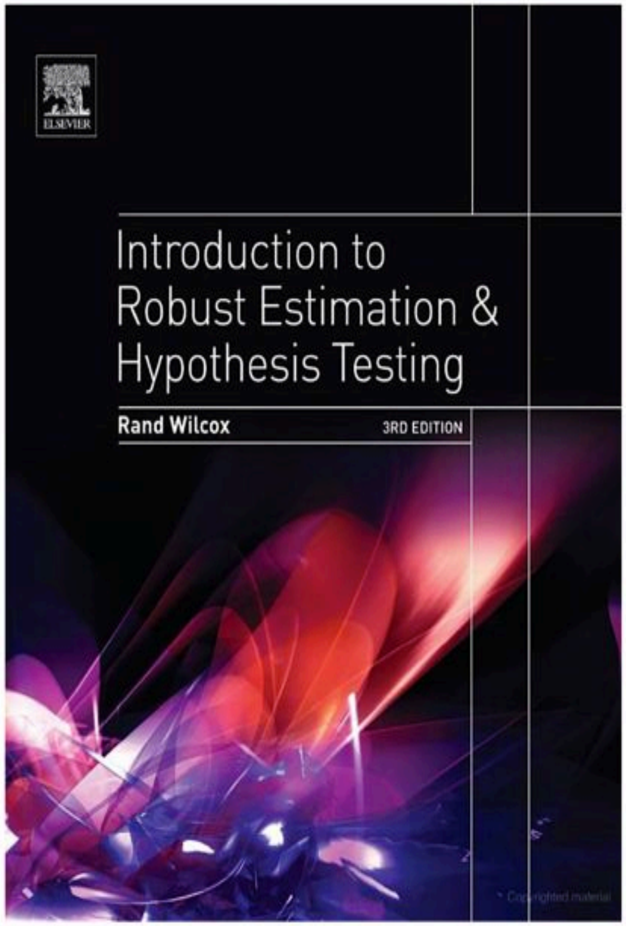


Monographs on Statistics and Applied Probability 57

An Introduction to the Bootstrap

Bradley Efron
Robert J. Tibshirani

CHAPMAN & HALL/CRC



ELSEVIER

Introduction to Robust Estimation & Hypothesis Testing

Rand Wilcox

3RD EDITION

Parametric statistics

Assume gaussian distribution of data

T-test: Compare paired/ unpaired Samples for continuous data. In EEGLAB, used for grand-average ERPs.

Paired

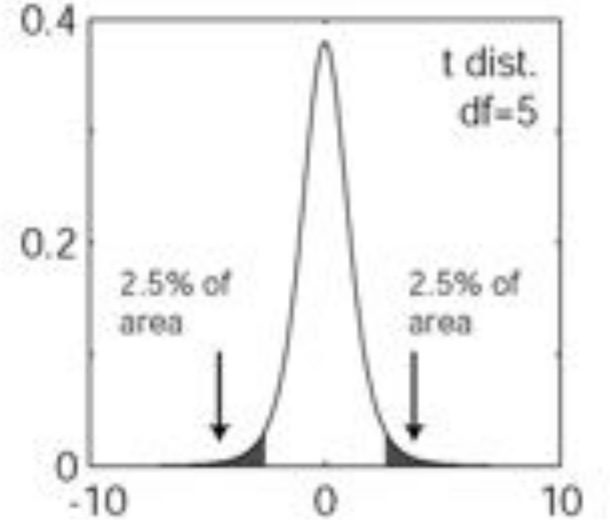
$$t = \frac{Mean\_difference}{Standard\_deviation}\sqrt{N-1}$$

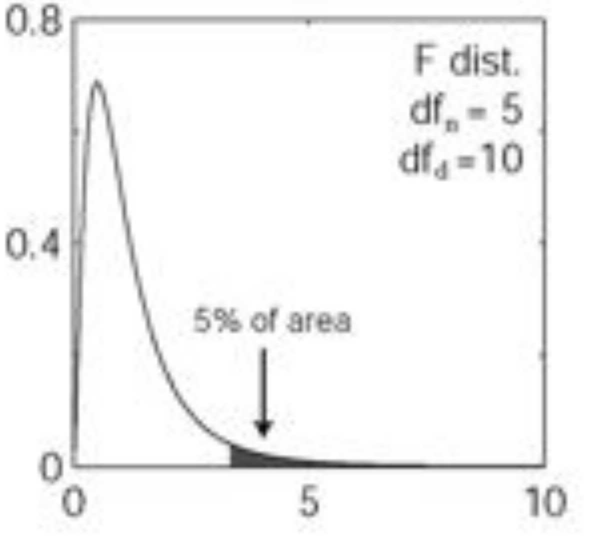
Unpaired

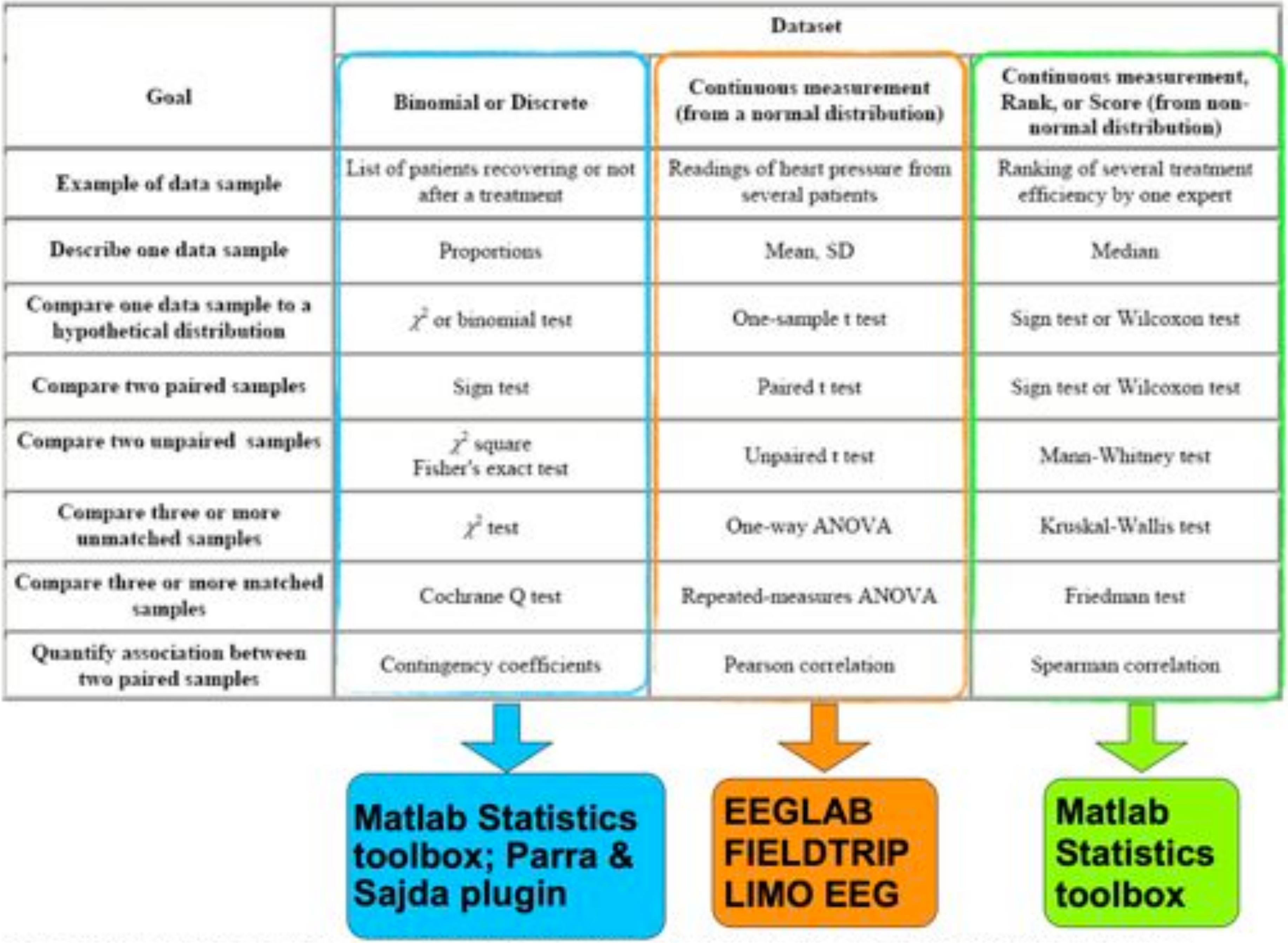
$$t = \sqrt{N}\,\frac{Mean_A - Mean_B}{\sqrt{(SD_A)^2 - (SD_B)^2}}$$

ANOVA: compare several groups (can test interaction between two factors for the repeated measure ANOVA)

$$F = \frac{Variance_{interGroup} \big/ N_{Group} - 1}{Variance_{WithinGroup} \big/ N - N_{Group}}$$







Goal	Dataset		
	Binomial or Discrete	**Continuous measurement (from a normal distribution)**	**Continuous measurement, Rank, or Score (from non-normal distribution)**
Example of data sample	List of patients recovering or not after a treatment	Readings of heart pressure from several patients	Ranking of several treatment efficiency by one expert
Describe one data sample	Proportions	Mean, SD	Median
Compare one data sample to a hypothetical distribution	χ^2 or binomial test	One-sample t test	Sign test or Wilcoxon test
Compare two paired samples	Sign test	Paired t test	Sign test or Wilcoxon test
Compare two unpaired samples	χ^2 square Fisher's exact test	Unpaired t test	Mann-Whitney test
Compare three or more unmatched samples	χ^2 test	One-way ANOVA	Kruskal-Wallis test
Compare three or more matched samples	Cochrane Q test	Repeated-measures ANOVA	Friedman test
Quantify association between two paired samples	Contingency coefficients	Pearson correlation	Spearman correlation

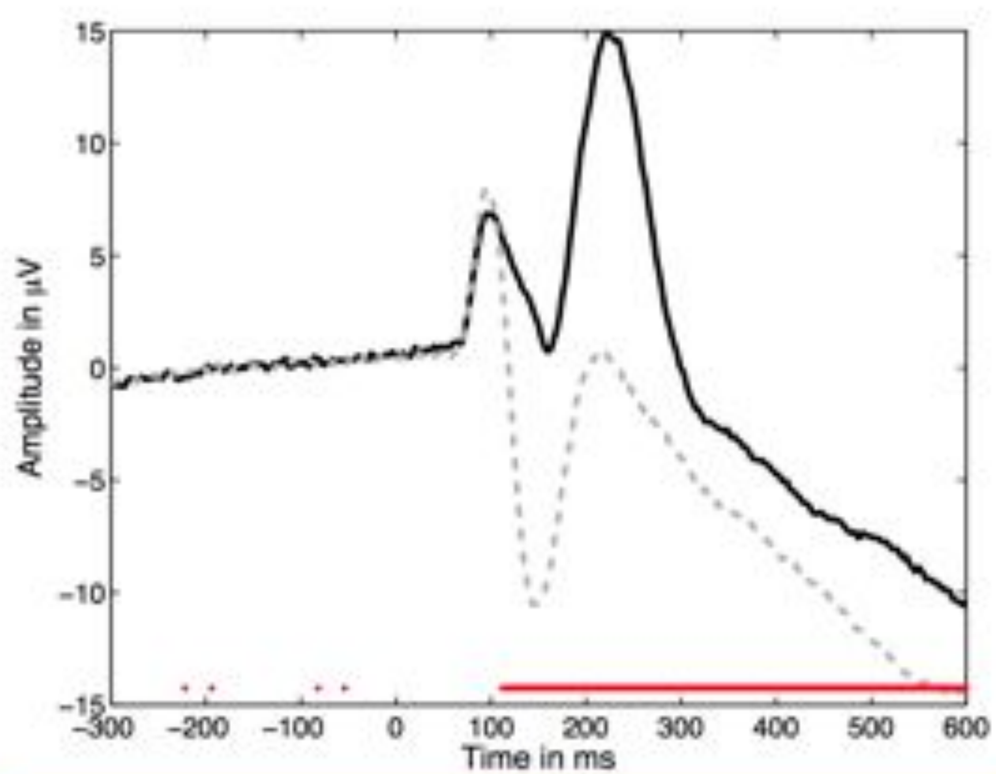
Matlab Statistics toolbox; Parra & Sajda plugin

EEGLAB FIELDTRIP LIMO EEG

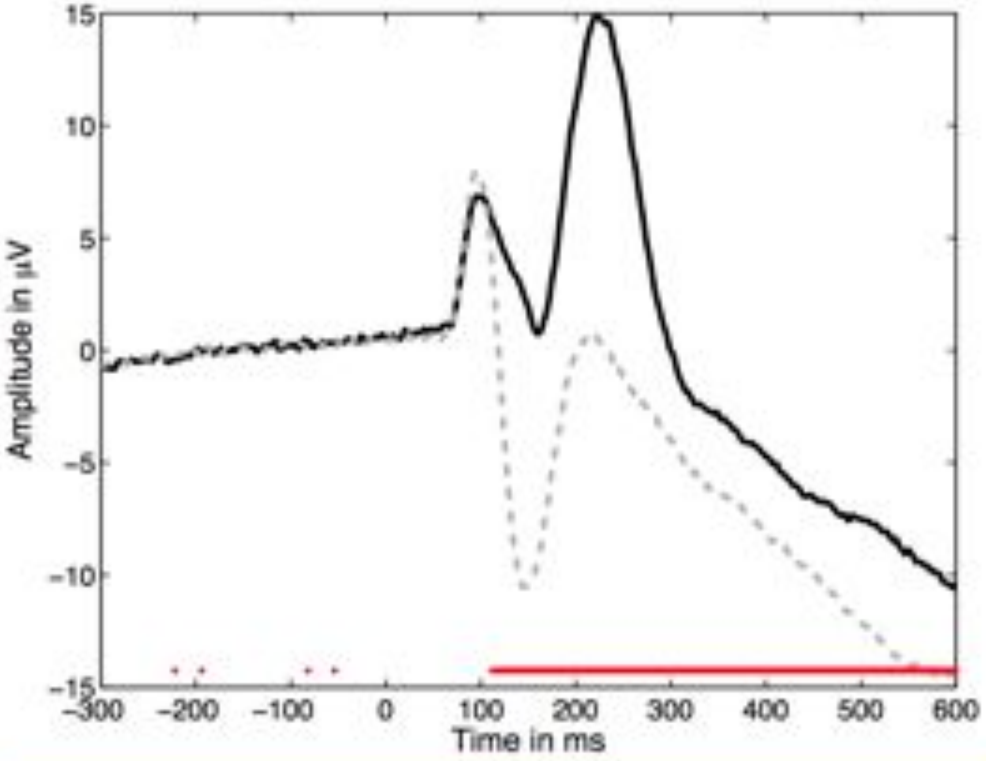
Matlab Statistics toolbox

Delorme, A. (2006) Statistical methods. *Encyclopedia of Medical Device and Instrumentation*, vol 6, pp 240-264. Wiley interscience.

Why the standard figure is not good enough



Why the standard figure is not good enough

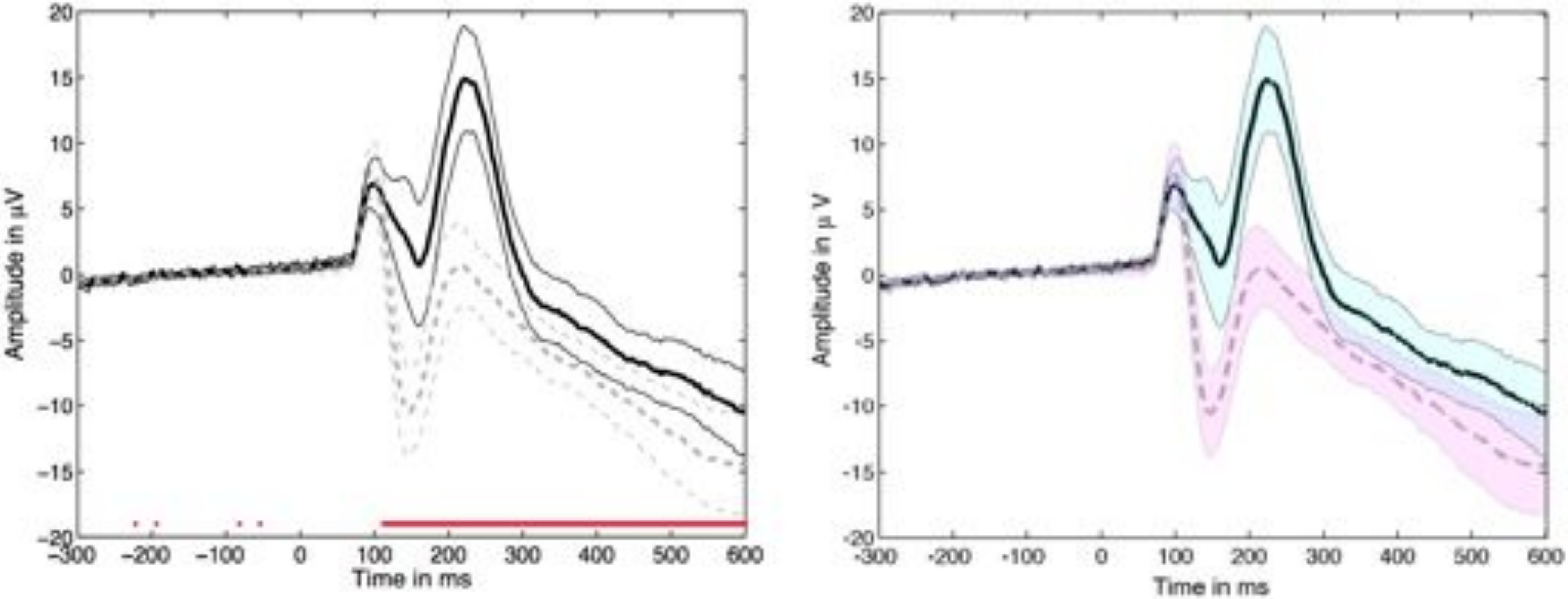


Interpretations should be limited to what was measured: group differences in means

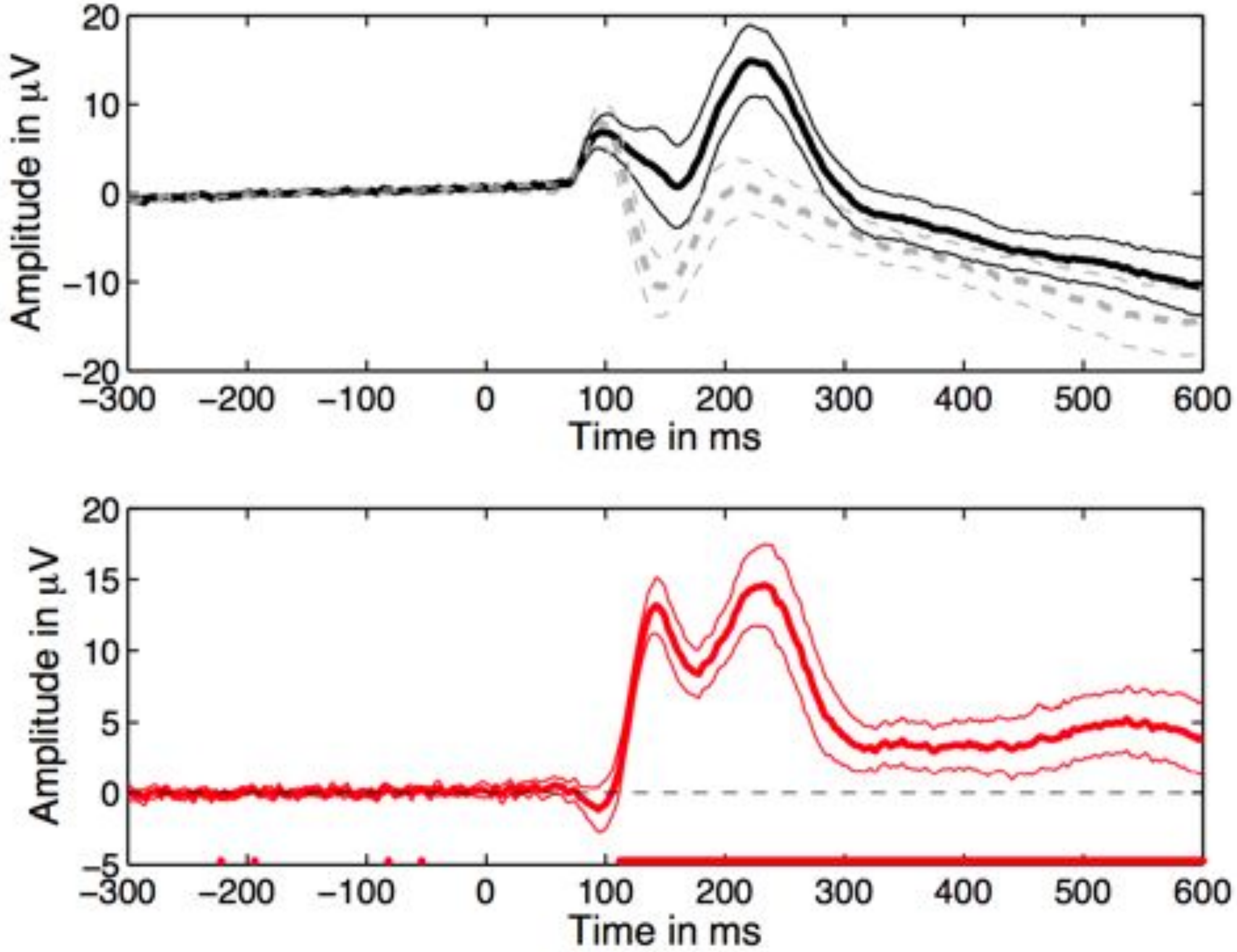
Significant effect?
- interesting?
- how many subjects?
- effect size?

Non-significant effect?
- not there?
- lack of power?
- how many subjects?
- effect size?

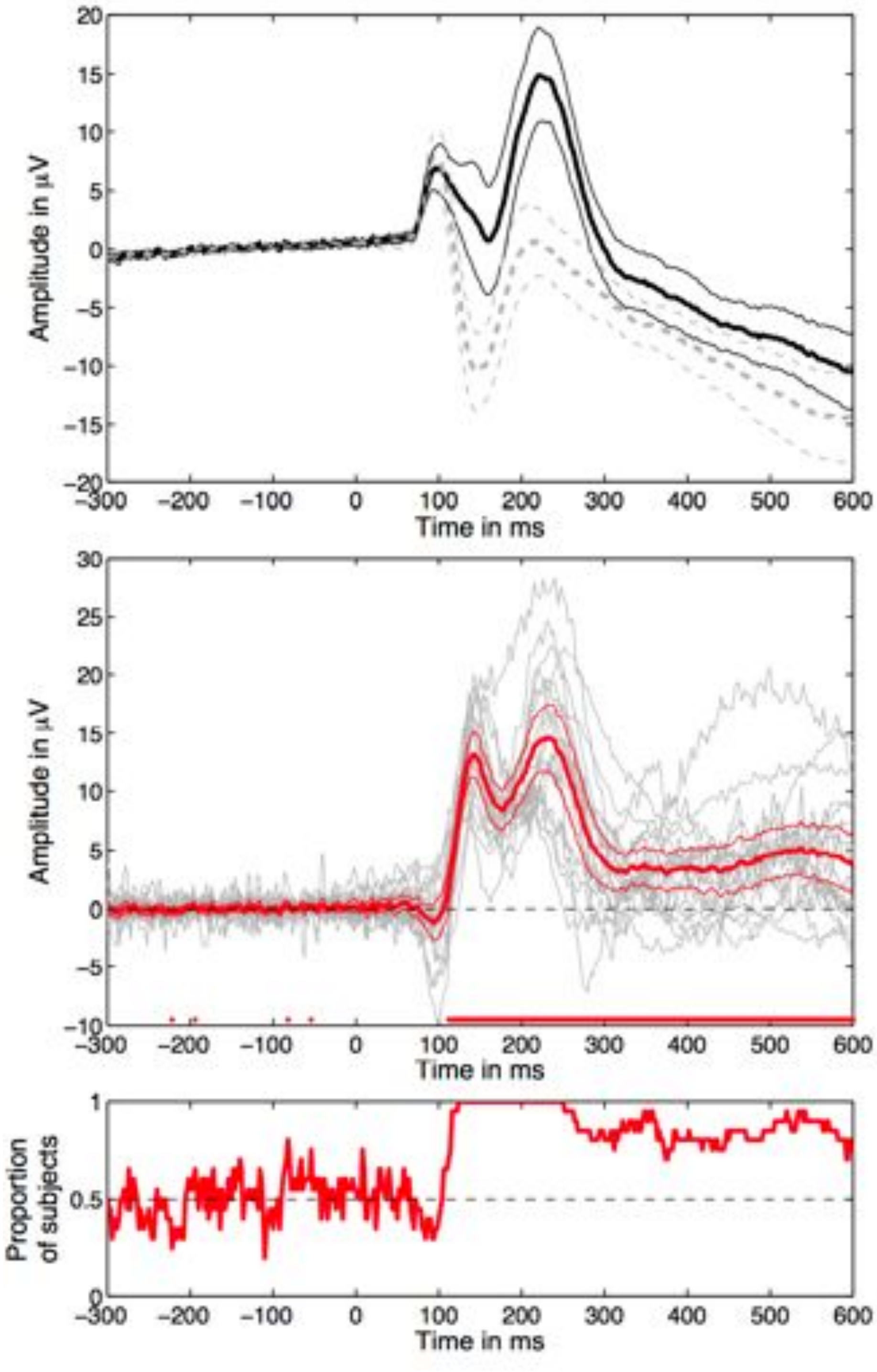
Add confidence intervals



Add plot of the difference



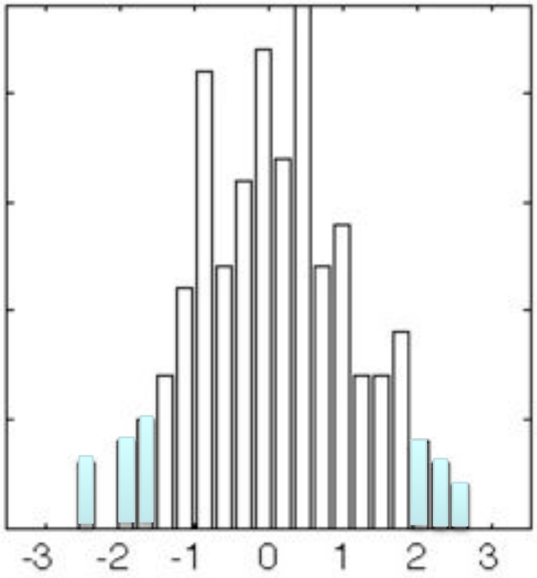
How many subjects show an effect?



Robust measures of central tendency (location)

- Non-robust estimator
  - Mean: mERP = mean(EEG.data,..)
- Robust estimators of central tendency
  - Median: mdERP = median(EEG.data,...)
  - Trimmed mean tmERP = trimmean(EEG.data,...)

Trimmed means



- 20% trimmed means provide high power under normality and high power in the presence of outliers
- Rand Wilcox, 2012, Introduction to Robust Estimation and Hypothesis Testing, Elsevier **ERP application:** Rousselet, Husk, Bennett & Sekuler, 2008, *J. Vis.* + Desjardins 2013

Non-parametric statistics

Paired t-test → Wilcoxon
Unpaired t-test → Mann-Whitney
One way ANOVA → Kruskal Wallis

Values
Ranks

BOTH ASSUME NORMAL DISTRIBUTIONS

Problems

- Not resistant against outliers
- For ANOVA and t-test non-normality is an issue when distributions differ or when variances are not equal.
- Slight departure from normality can have serious consequences

---

Solutions

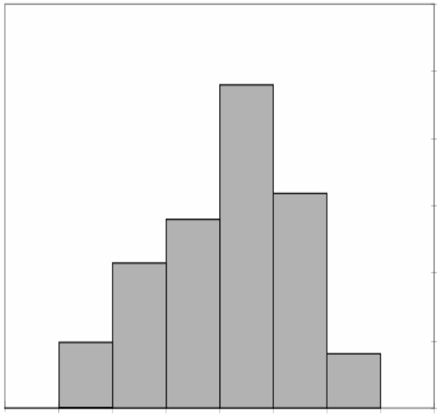
1. Randomization approach
2. Bootstrap approach

Bootstrap: central idea

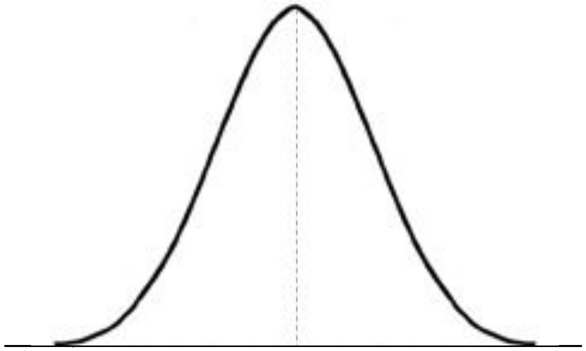
- “The bootstrap is a computer-based method for assigning measures of accuracy to statistical estimates.” Efron & Tibshirani, 1993

- “The central idea is that it may sometimes be better to draw conclusions about the characteristics of a population strictly from the sample at hand, rather than by making perhaps unrealistic assumptions about the population.” Mooney & Duval, 1993

Sample and population



Sample



Population

H0: the mean is not 0 for the population

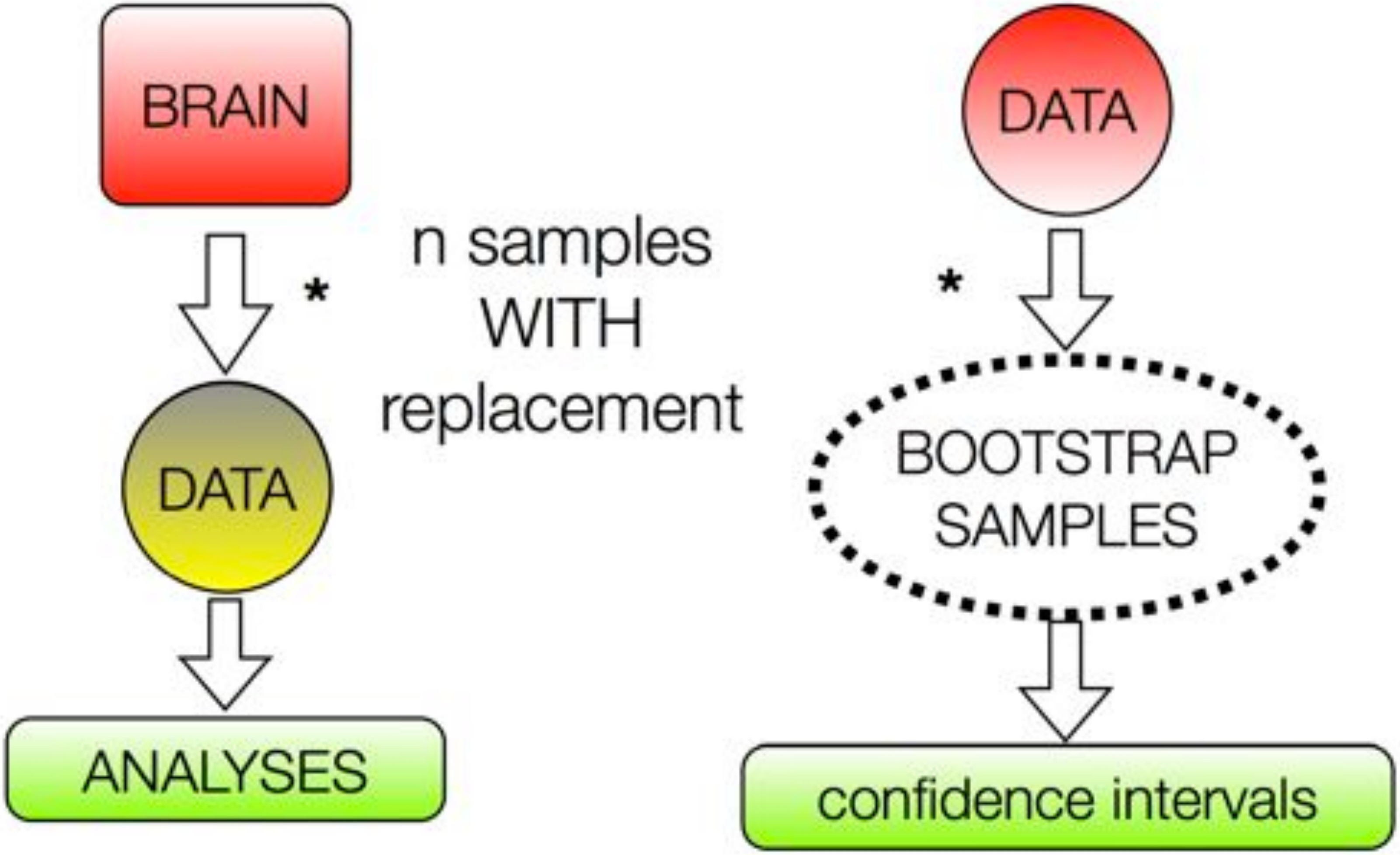
given that we have no other information about the population, the sample is our best single estimate of the population

Percentile bootstrap: general recipe

- sample = X1, ..., Xn
- resample n observations with replacement
- compute estimate
- repeat B times

with B large enough the B estimates provide a good approximation of the distribution of the estimate of the sample

Bootstrap philosophy



BRAIN

*

n samples WITH replacement

DATA

ANALYSES

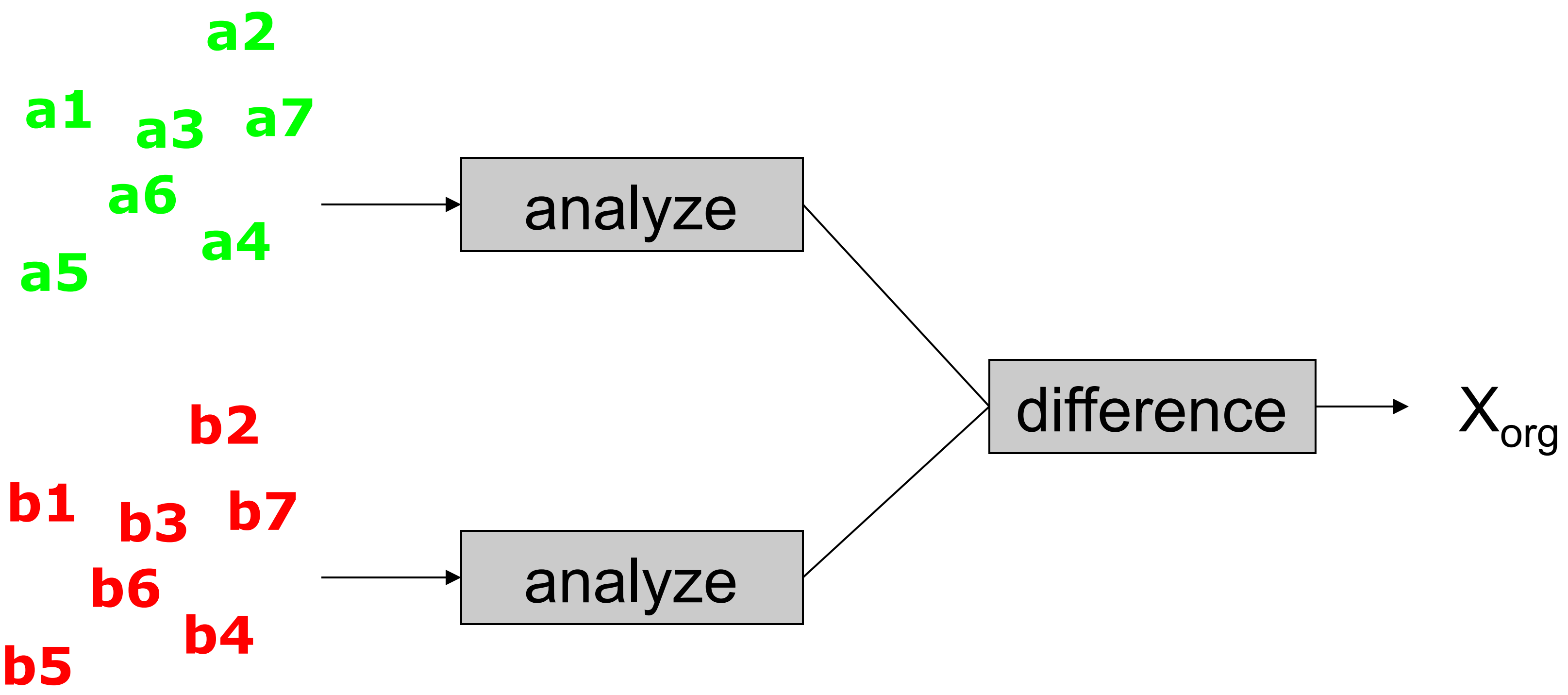
DATA

*

BOOTSTRAP SAMPLES

confidence intervals

Confidence interval for the difference
Bootstrap approach 1



a2
a1 a3 a7
a6
a4
a5

analyze

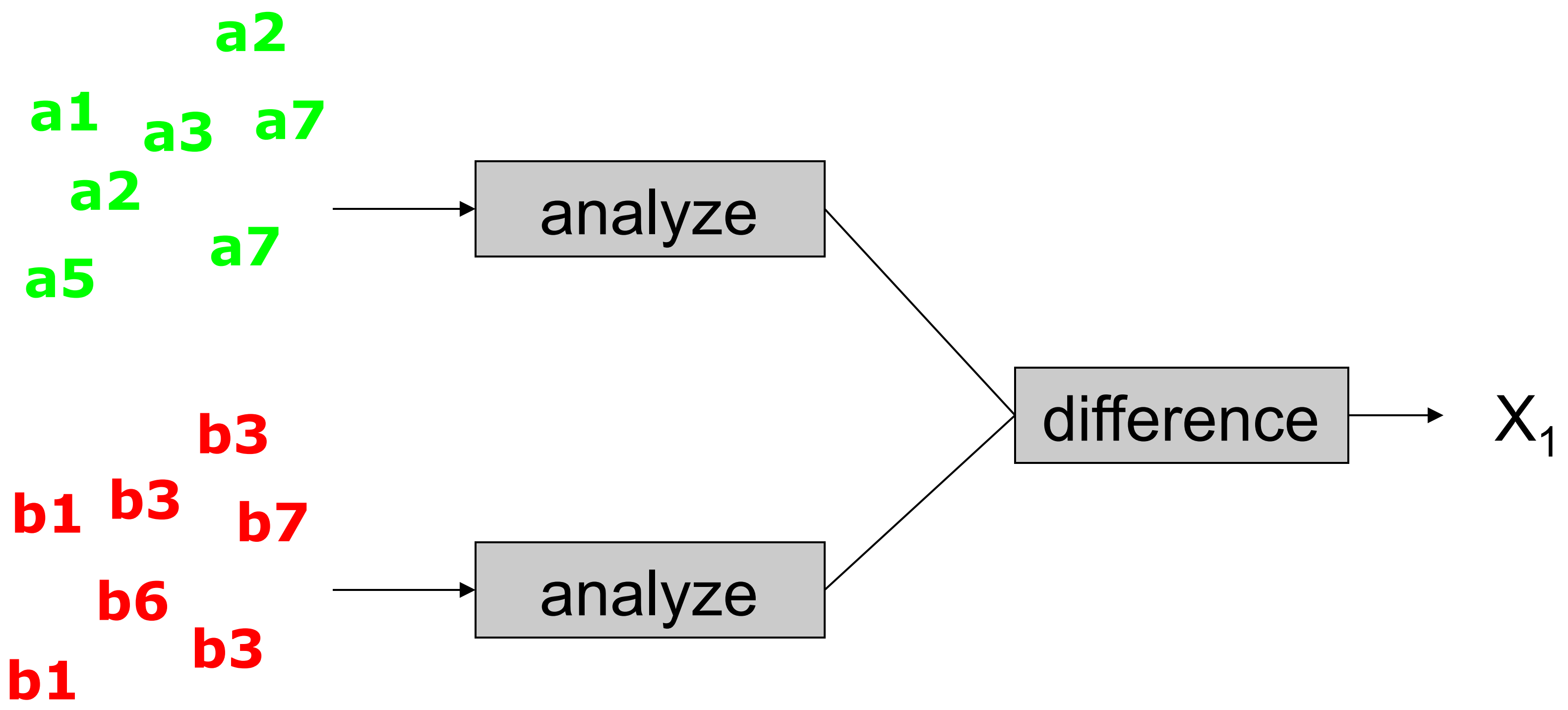
difference

X_{org}

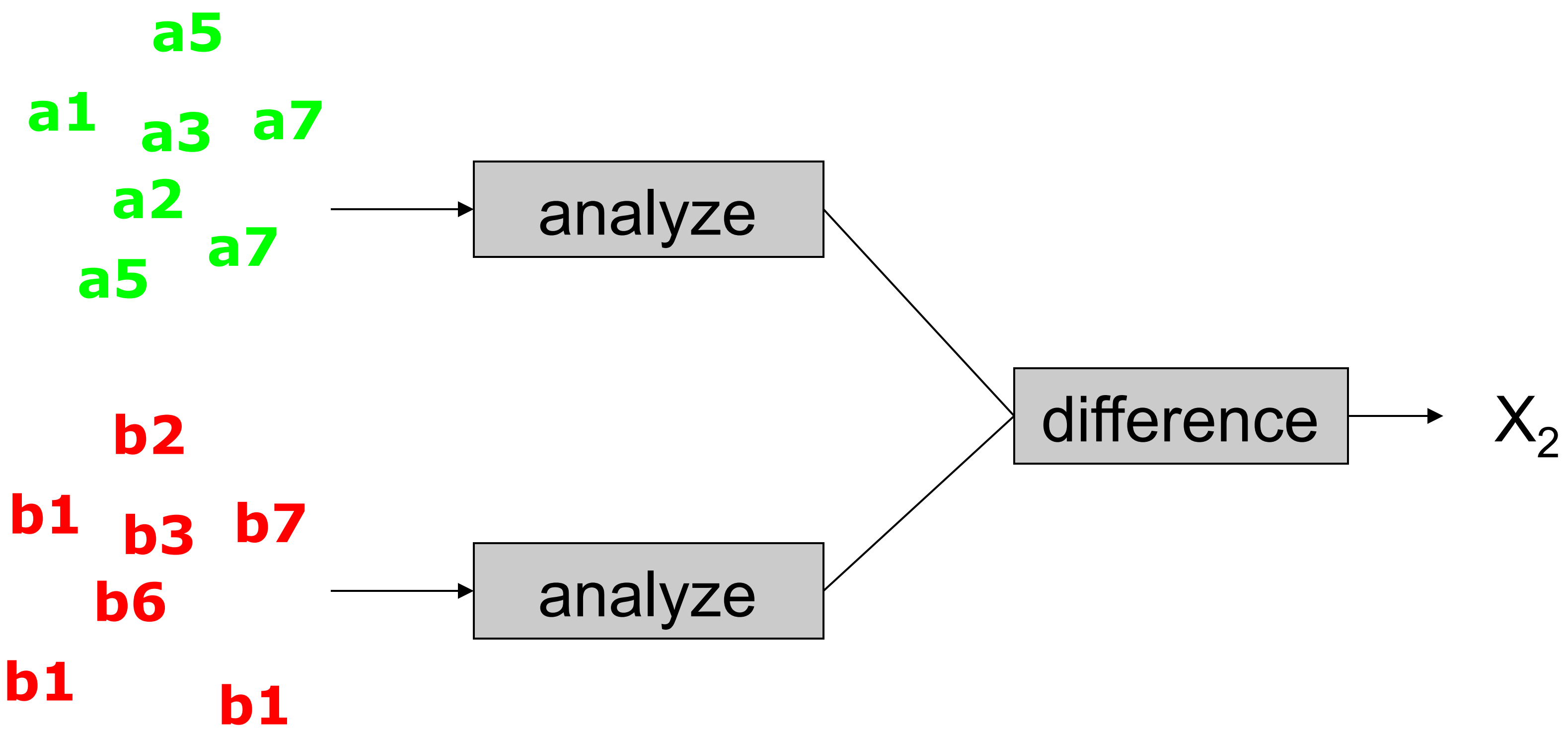
b2
b1 b3 b7
b6
b4
b5

analyze

Confidence interval for the difference
Bootstrap approach 1



Confidence interval for the difference
Bootstrap approach 1



a5 a1 a3 a7 a2 a7 a5

b2 b1 b3 b7 b6 b1 b1

analyze

analyze

difference

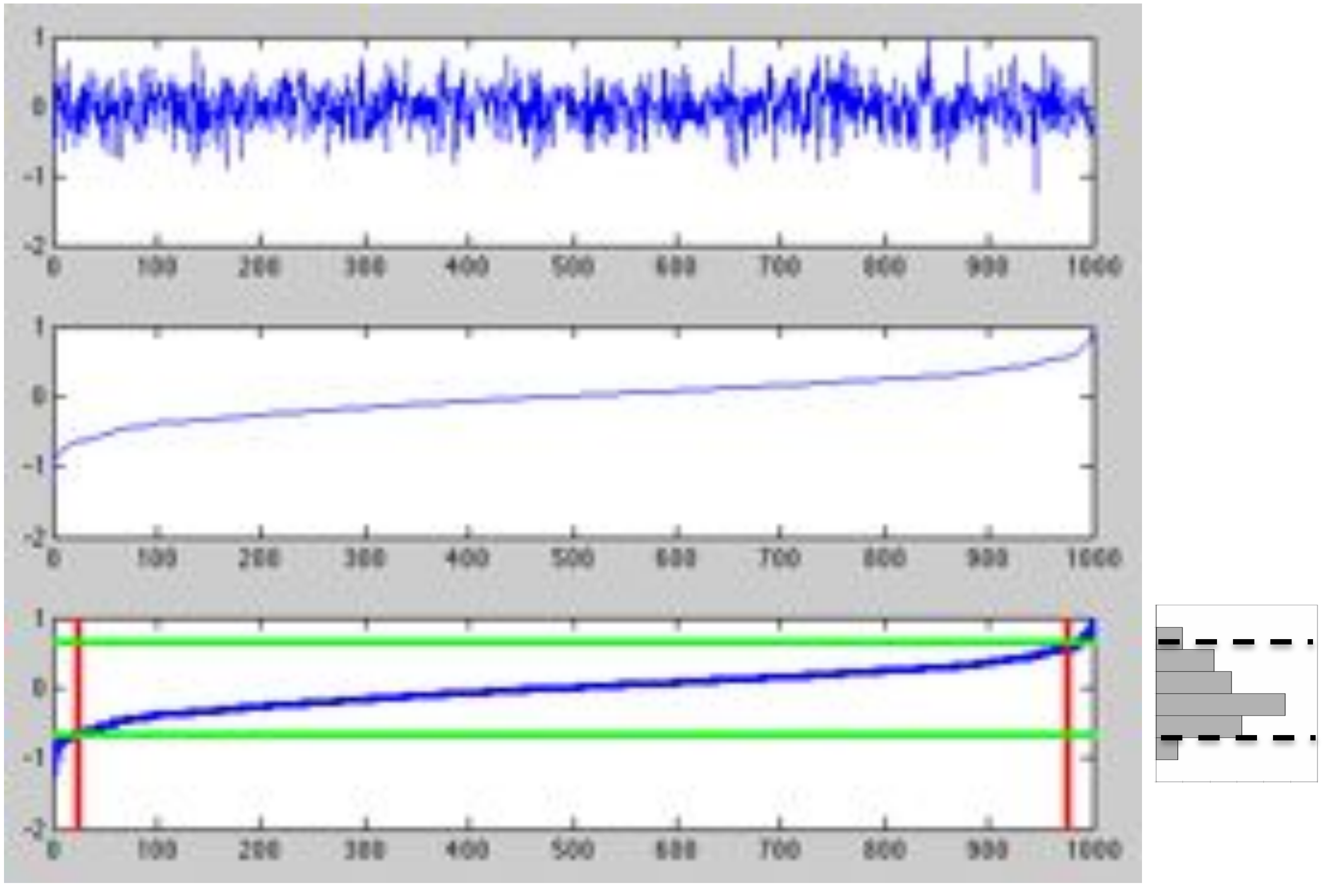
X_2

Confidence interval for the difference
Bootstrap approach 1

Permutation /bootstrap

Sorted values

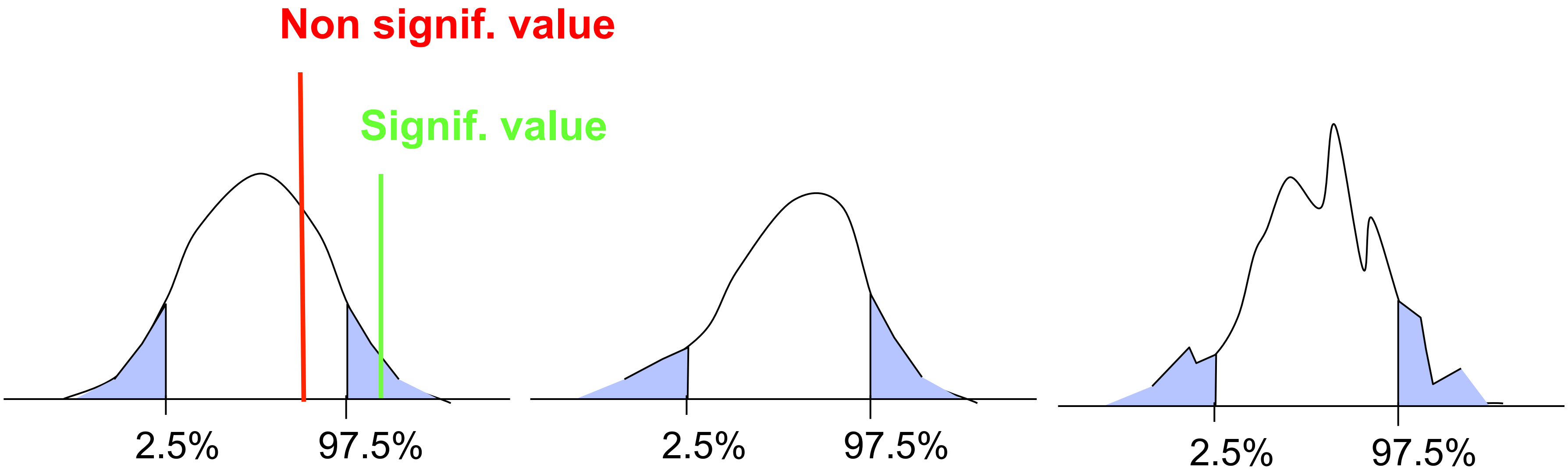
Thresholds



2.5%

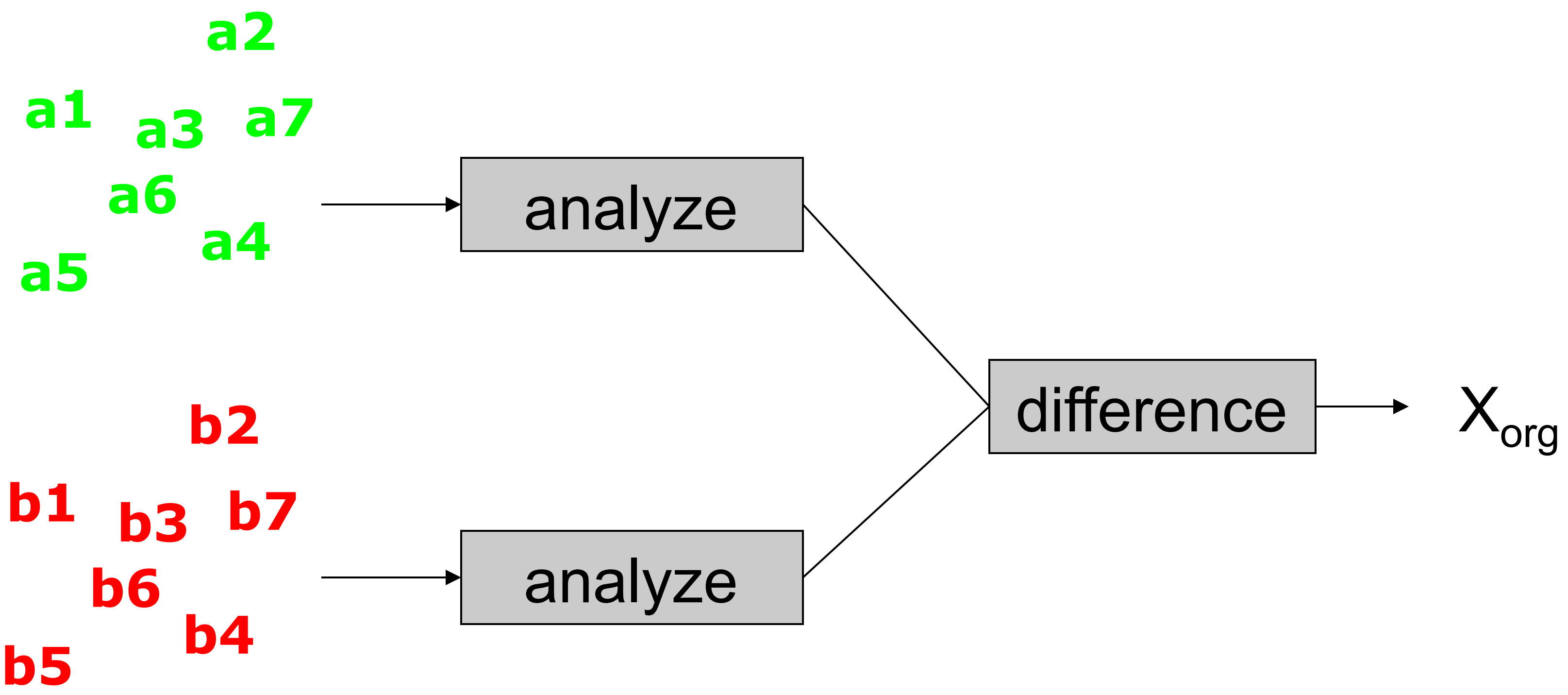
97.5%

Distribution can take any shape

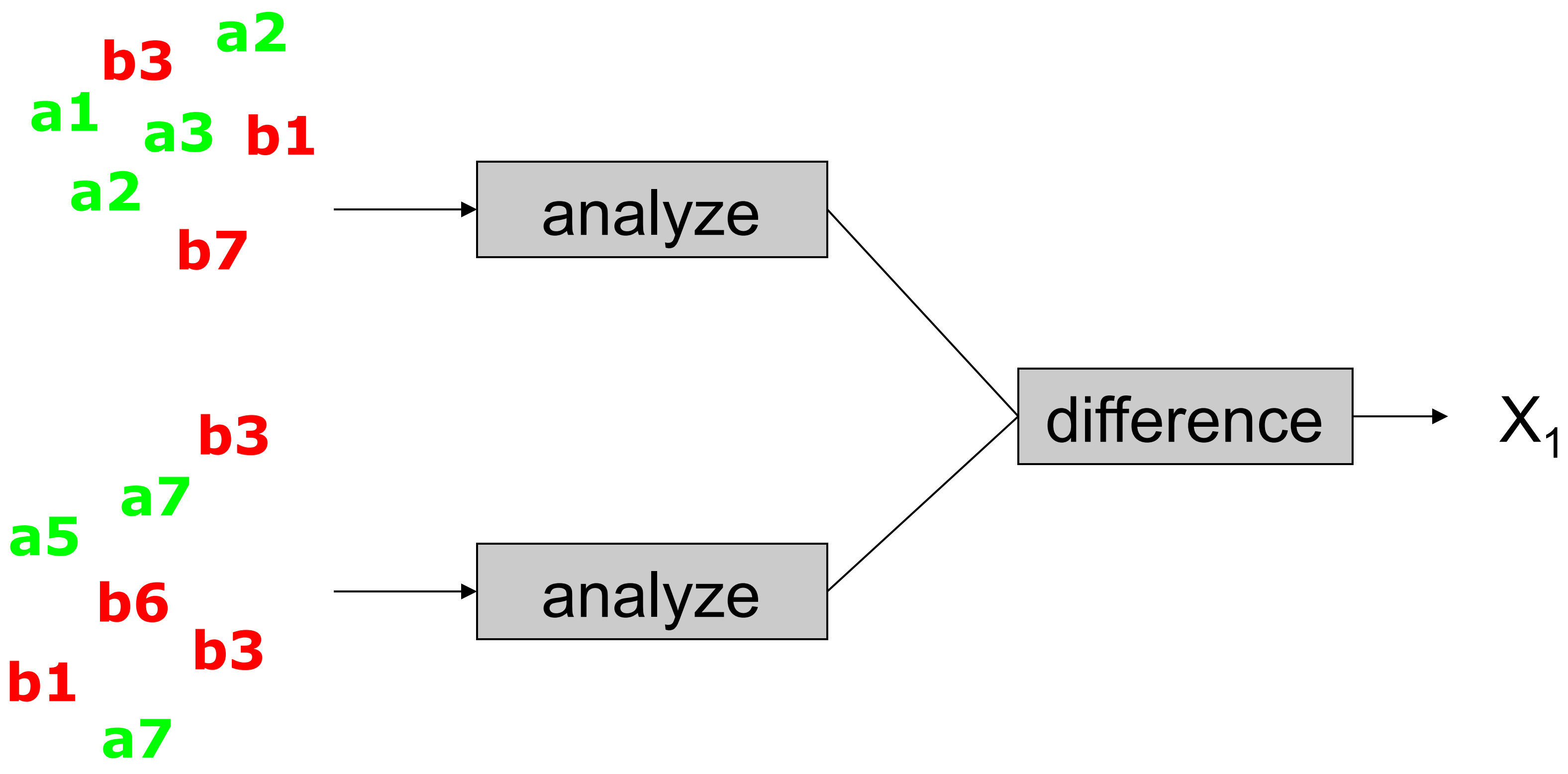


Once you have the 95% confidence interval for the difference: significance only involve assessing if 0 is included in the tails.

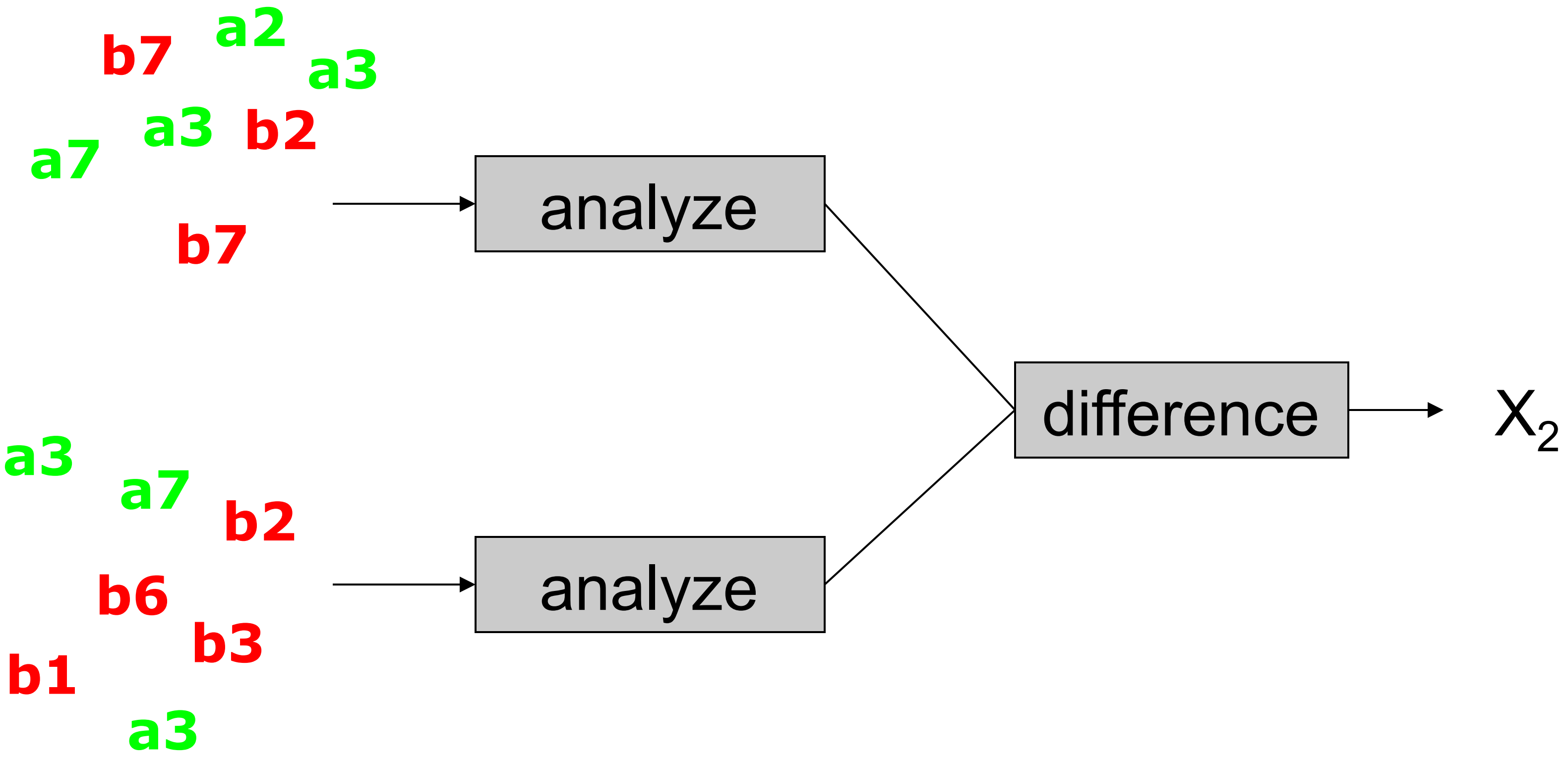
Confidence interval for the difference
Bootstrap approach 2



Confidence interval for the difference
Bootstrap approach 2



Confidence interval for the difference
Bootstrap approach 2

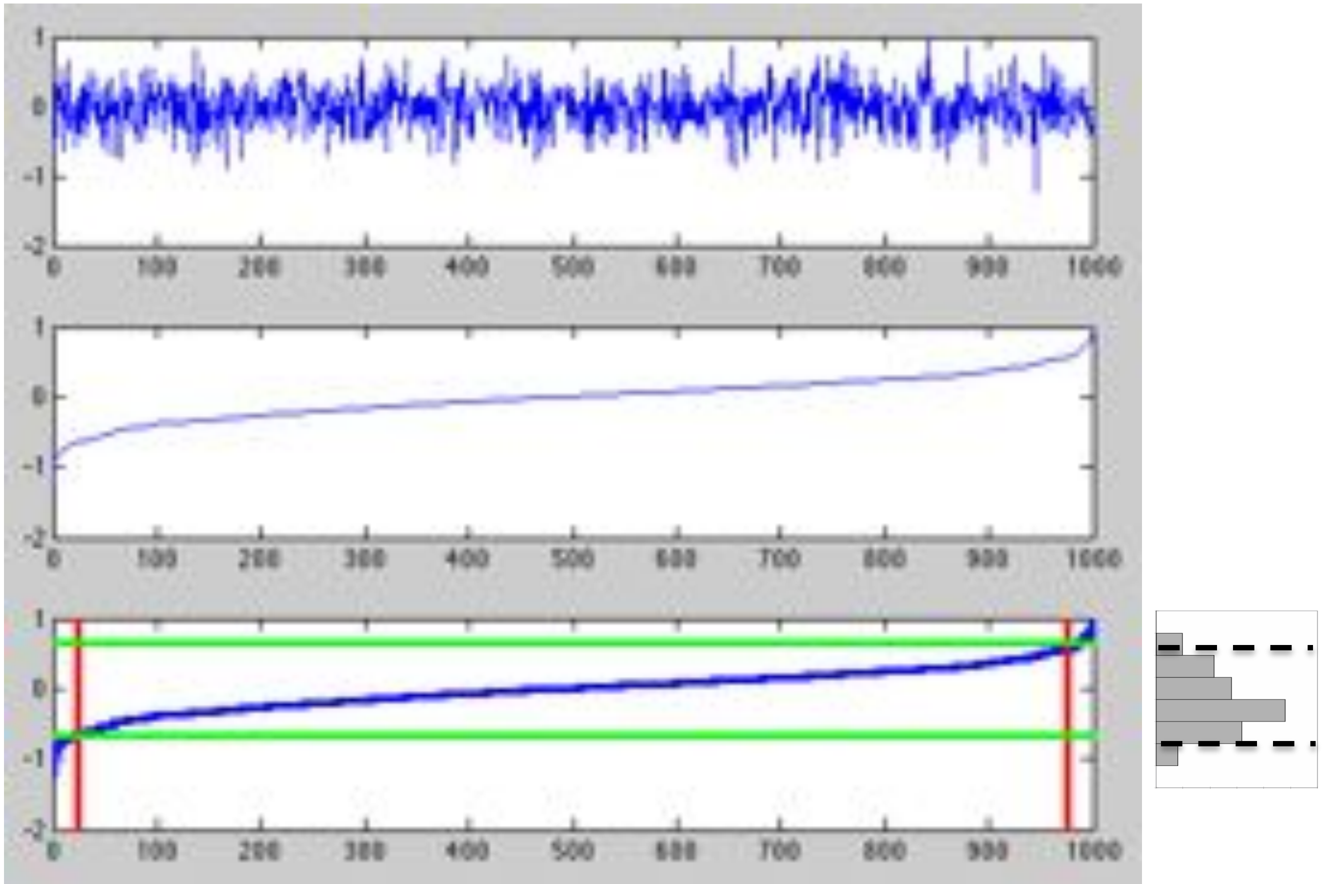


Confidence interval for the difference
Bootstrap approach 2

Permutation /bootstrap

Sorted values

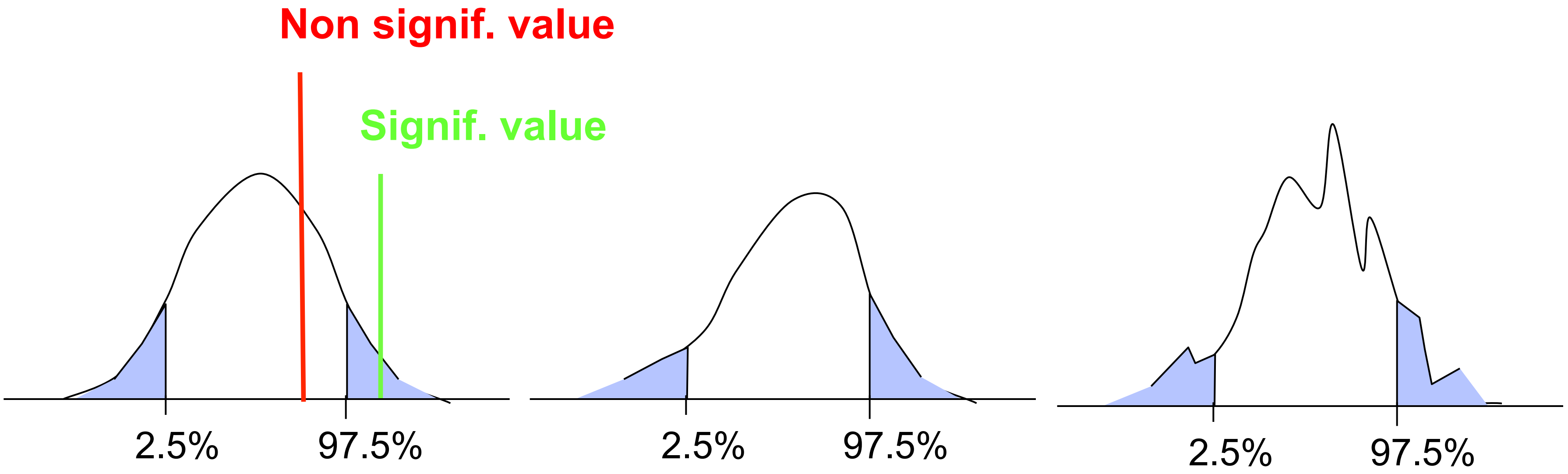
Thresholds



2.5%

97.5%

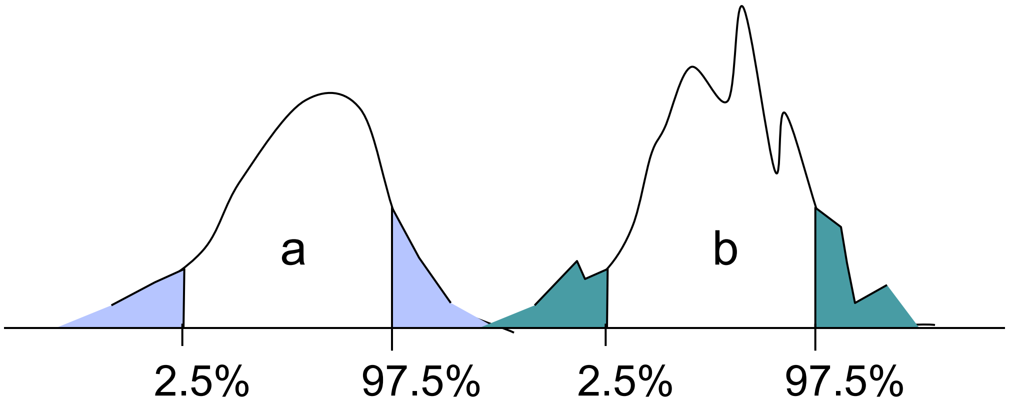
Distribution can take any shape



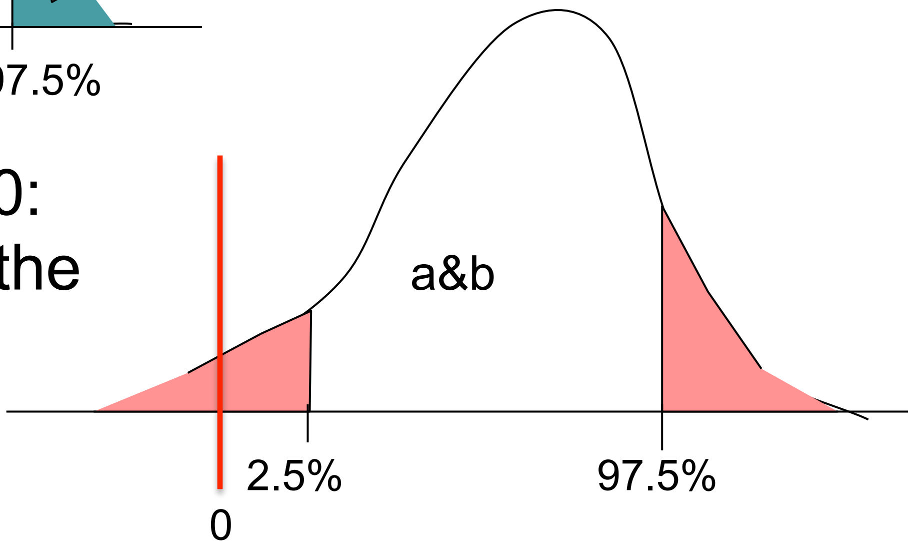
Once you have the 95% confidence interval for the difference: significance only involve assessing if 0 is included in the tails.

Difference between the two bootstrap approches

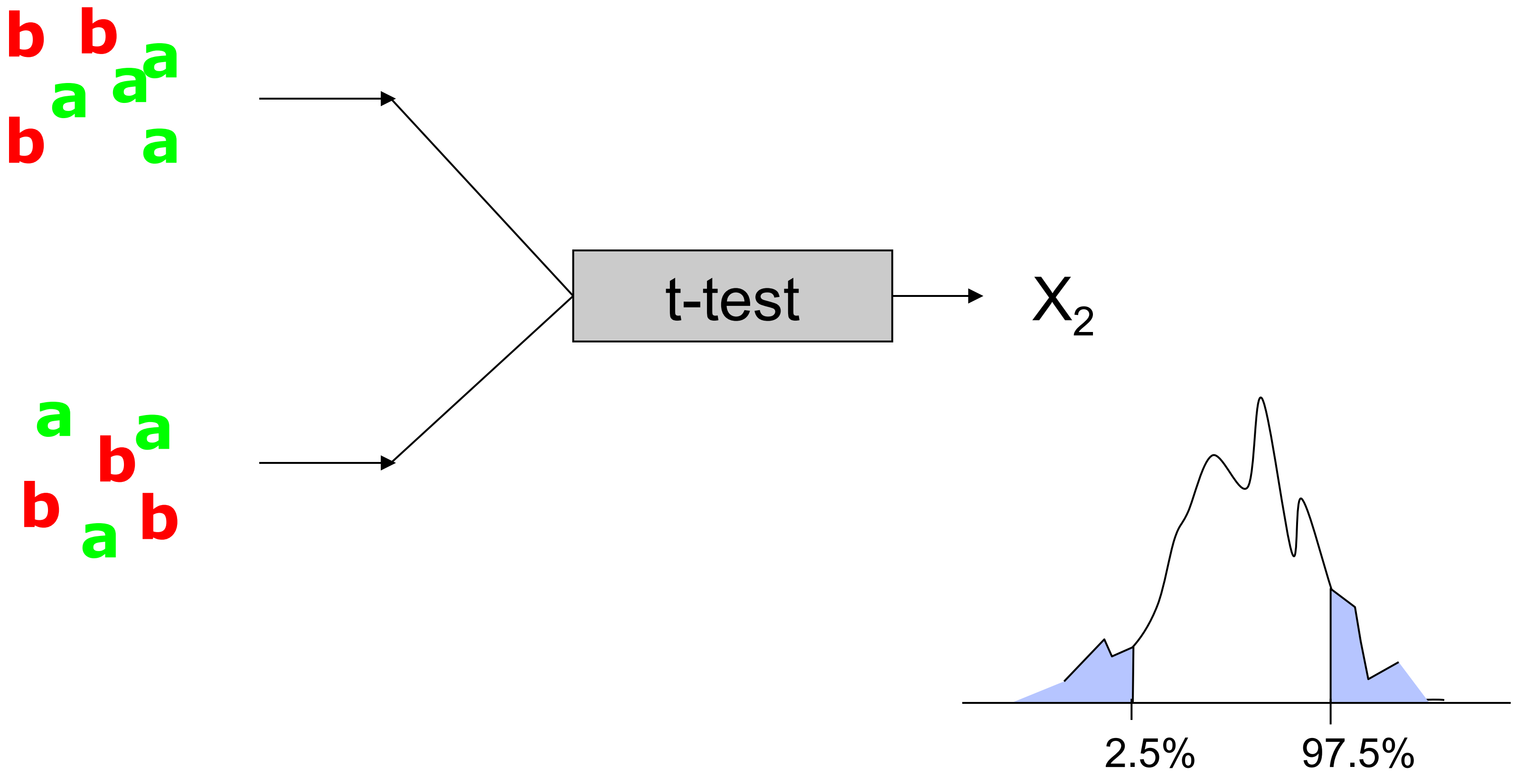
- Bootstrap 1 is testing against H1: the two samples originate from the different distributions.



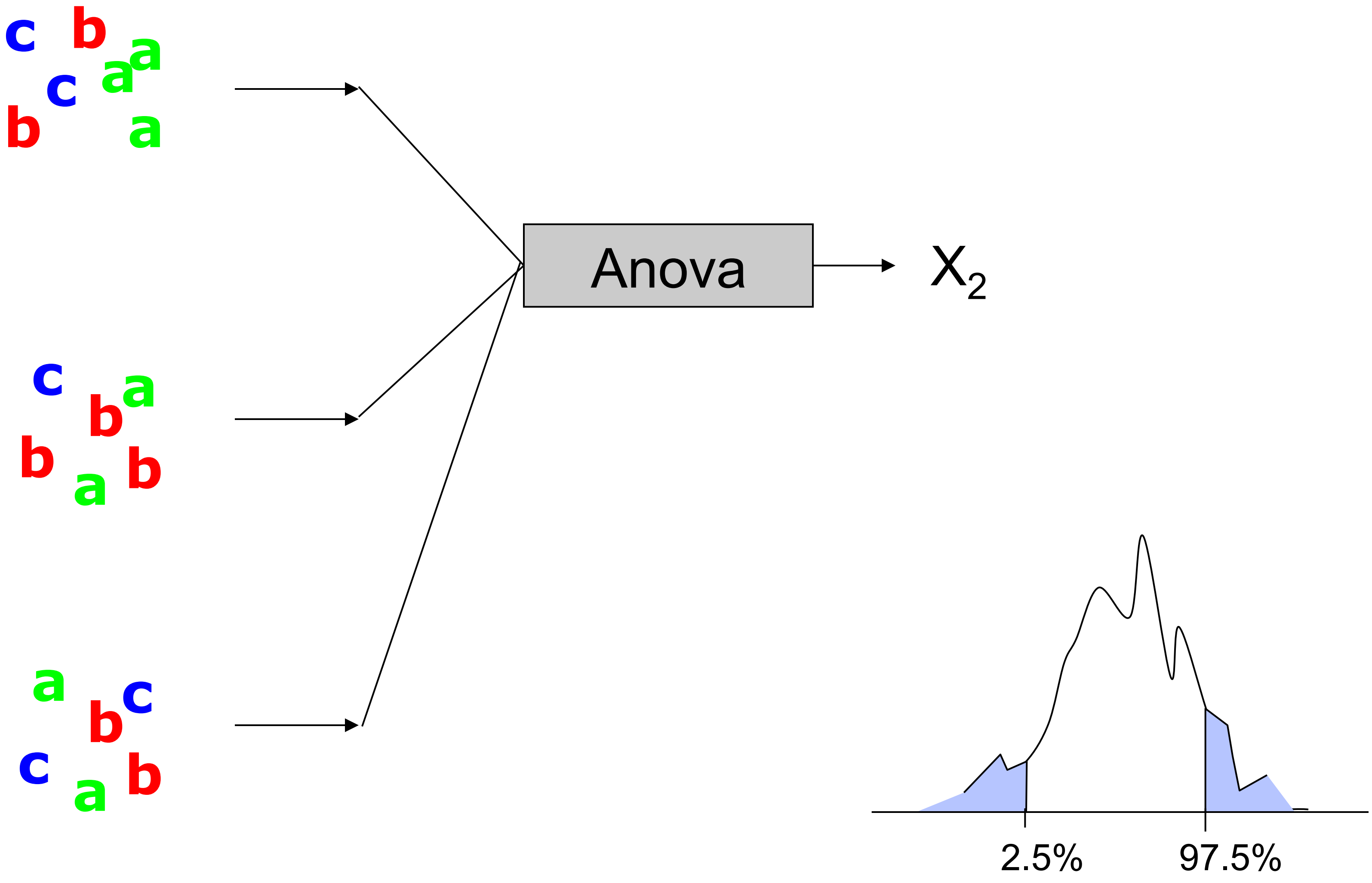
- Bootstrap 2 is testing against H0: the two samples originate from the same distribution.



Measure for the bootstrap

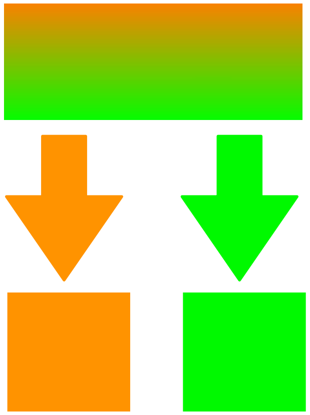


Measure for the bootstrap



Bootstrap versus permutation

Permutation

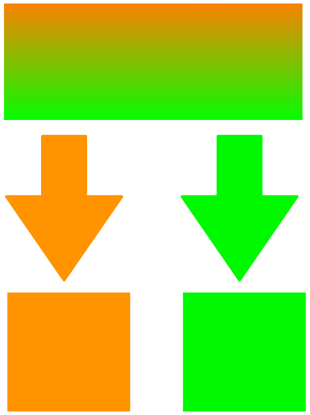


each element only
get picked once

↓

Draws are dependent of each others

Bootstrap



each element can
get picked several
times

↓

Draws are independent of each others

Use bootstrap!

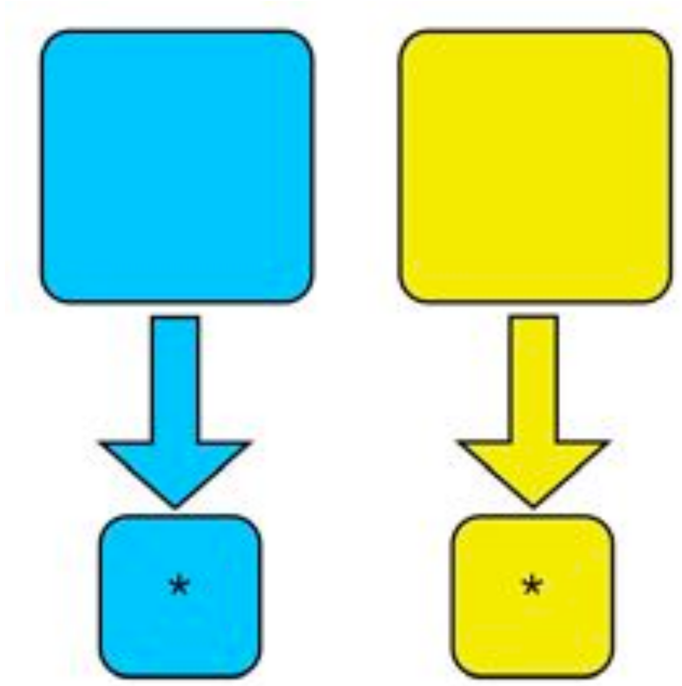
Resampling strategies: follow the data acquisition process

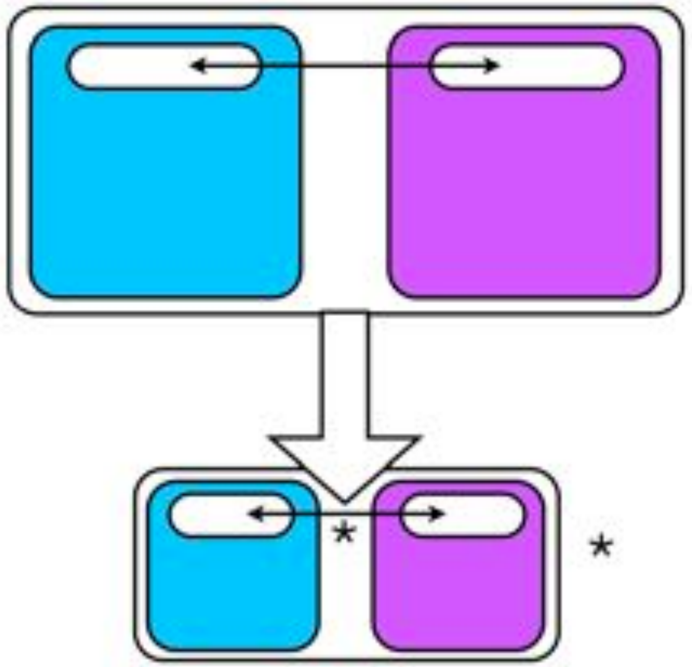
Independent sets:

- 2 conditions in single-subject analyses
- 2 groups of subjects, e.g. patients vs. controls

Dependent sets:

- 2 conditions in group analyses
- Correlations
- Linear regression



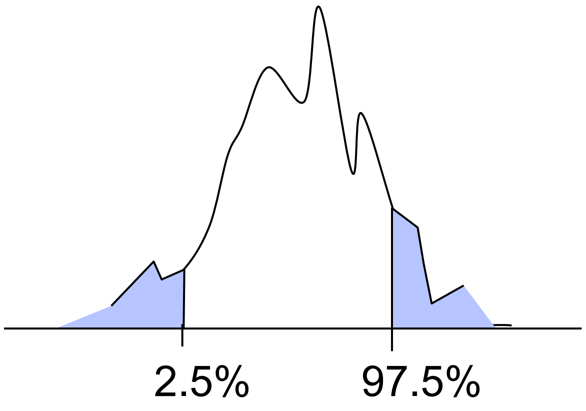


Husband	Wifes
22	**25**
32	**25**
50	**51**
25	**25**
33	**38**
27	**30**
45	**60**
47	**54**
30	**31**
44	**54**
23	**23**
39	**34**
24	**25**
22	**23**
16	**19**
73	**71**
27	**26**
36	**31**
24	**26**
60	**62**
26	**29**
23	**31**
28	**29**
36	**35**

Median

Are the two groups different: that’ s an unpaired test (comparing the mean or median of husband and the mean or median of wife)

Are husbands older than wifes: that’ s a paired test. Compute difference between the two and change sign to bootstrap.



Correcting for multiple comparisons

- Bonferoni correction: divide by the number of comparisons (Bonferroni CE. Sulle medie multiple di potenze. Bollettino dell'Unione Matematica Italiana, 5 third series, 1950; 267-70.)

- Holms correction: sort all p values. Test the first one against α/N, the second one against $\alpha/(N-1)$

- Max method

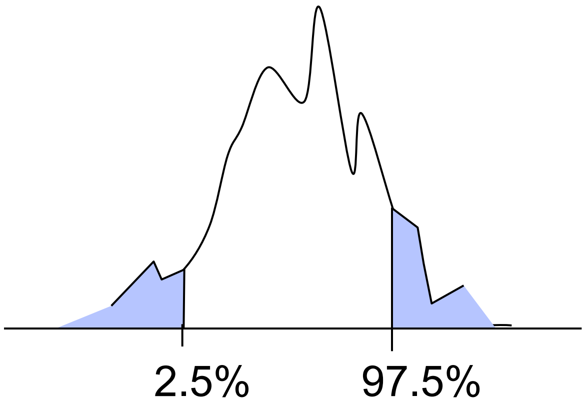
- False detection rate

- Clusters

Max procedure

- for each permutation or bootstrap loop, simply take the MAX of the absolute value of your estimator (e.g. mean difference) across electrodes and/or time frames and/or temporal frequencies.

- compare absolute original difference to this distribution



FDR procedure

Procedure:

- Sort all p values (column C1)

C3

- Create column C2 by computing *j*α/N*

- Subtract column C1 from C2 to build column C3

- Find the highest negative index in C3 and

find the corresponding p-value in C1 *(p_fdr)*

- Reject all null hypothesis whose p-value are less than or equal to *p_fdr*

	C1
Index "j"	Actual
1	0.001
2	0.002
3	0.01
4	0.03
5	0.04
6	0.045
7	0.05
8	0.1
9	0.2
10	0.6

FDR procedure

Procedure:

- Sort all p values (column C1)

C3

- Create column C2 by computing *j*α/N*

- Subtract column C1 from C2 to build column C3

- Find the highest negative index in C3 and find the corresponding p-value in C1 *(p_fdr)*

- Reject all null hypothesis whose p-value are less than or equal to *p_fdr*

	C1	C2
Index "j"	Actual	j*0.05/10
1	0.001	0.005
2	0.002	0.01
3	0.01	0.015
4	0.03	0.02
5	0.04	0.025
6	0.045	0.03
7	0.05	0.035
8	0.1	0.04
9	0.2	0.045
10	0.6	0.05

FDR procedure

Procedure:

- Sort all p values (column C1)
C3

- Create column C2 by computing *j*α/N*

- Subtract column C1 from C2 to build column C3

- Find the highest negative index in C3 and
find the corresponding p-value in C1 (*p_fdr)*

- Reject all null hypothesis whose p-value are less than or equal to *p_fdr*

	C1	C2	C3
Index "j"	Actual	j*0.05/10	C2-C1
1	0.001	0.005	-0.004
2	0.002	0.01	-0.008
3	0.01	0.015	-0.005
4	0.03	0.02	0.01
5	0.04	0.025	0.015
6	0.045	0.03	0.015
7	0.05	0.035	0.015
8	0.1	0.04	0.06
9	0.2	0.045	0.155
10	0.6	0.05	0.55

FDR procedure

Procedure:

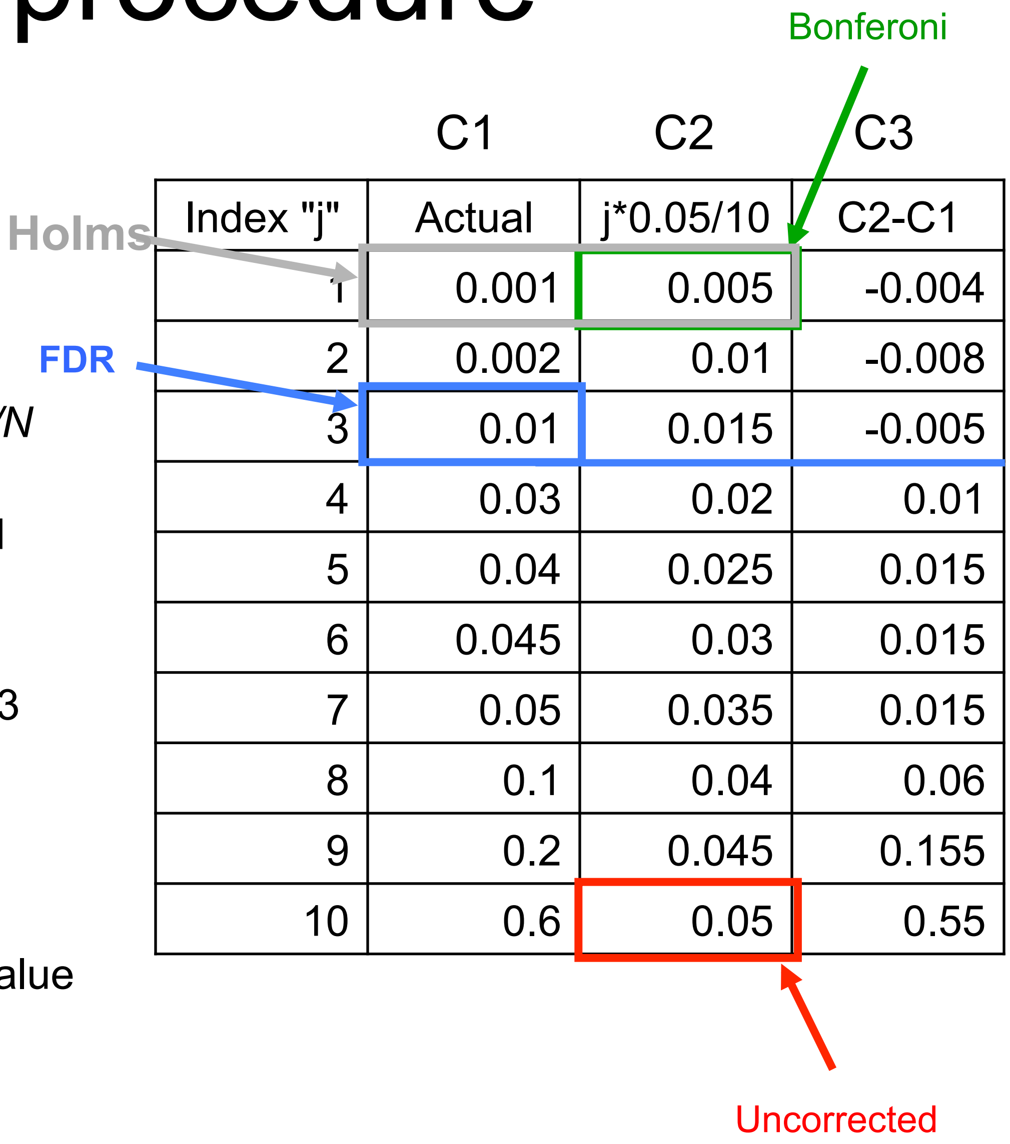
- Sort all p values (column C1)
C3

- Create column C2 by computing *j*α/N*

- Subtract column C1 from C2 to build column C3

- Find the highest negative index in C3
  and
find the corresponding p-value in C1 *(p_fdr)*

- Reject all null hypothesis whose p-value are less than or equal to *p_fdr*



Bonferoni

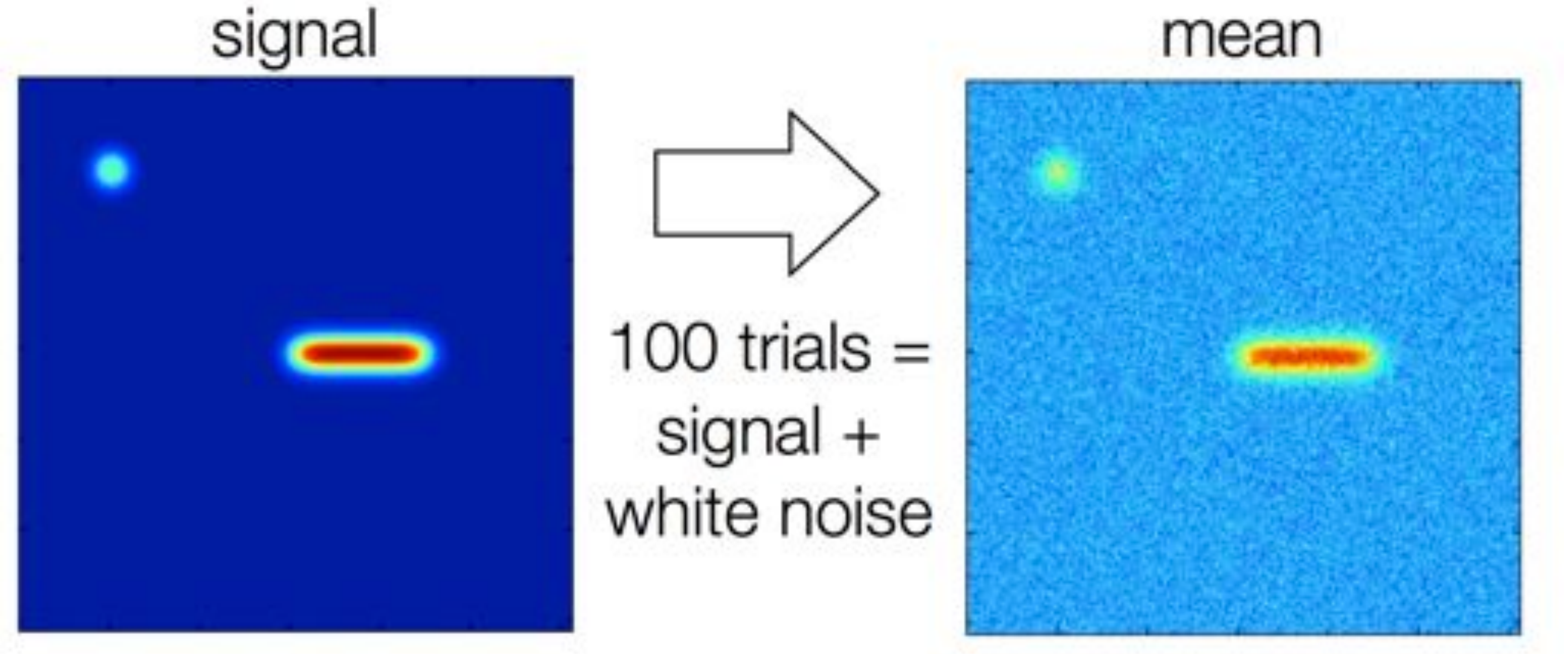
Holms

FDR

	C1	C2	C3
Index "j"	Actual	j*0.05/10	C2-C1
1	0.001	0.005	-0.004
2	0.002	0.01	-0.008
3	0.01	0.015	-0.005
4	0.03	0.02	0.01
5	0.04	0.025	0.015
6	0.045	0.03	0.015
7	0.05	0.035	0.015
8	0.1	0.04	0.06
9	0.2	0.045	0.155
10	0.6	0.05	0.55

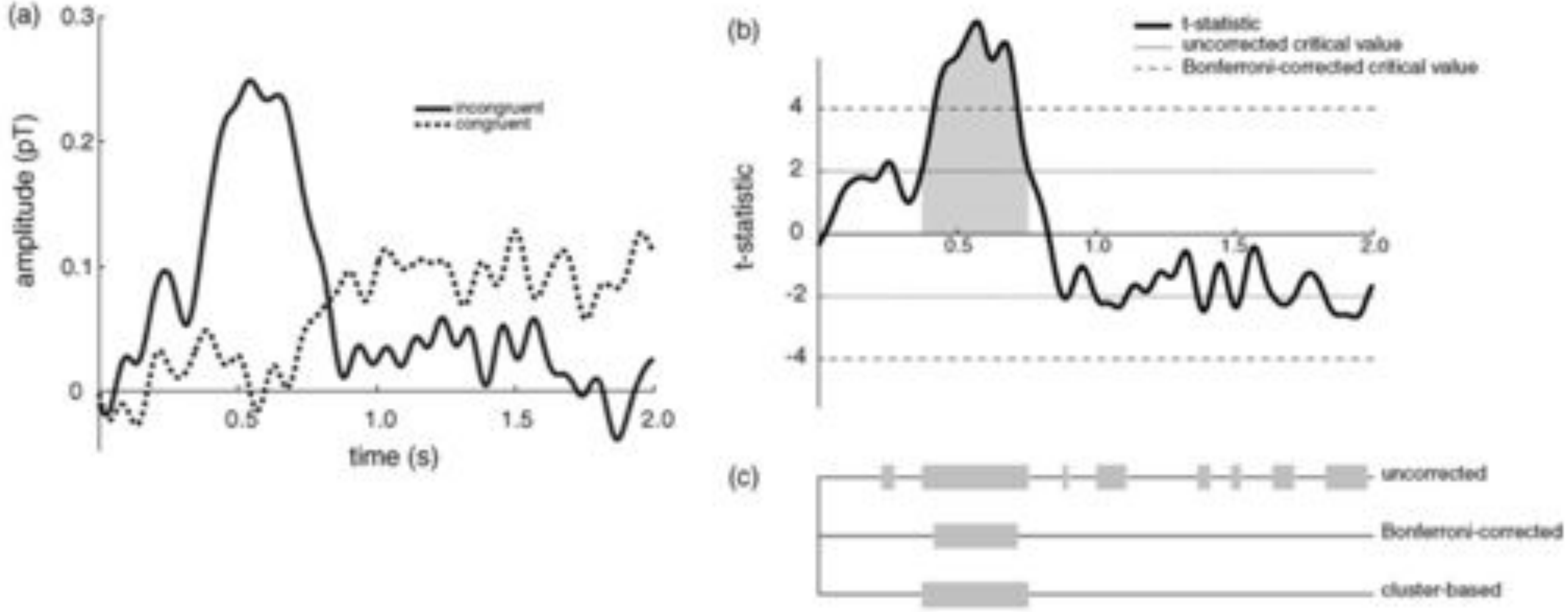
Uncorrected

Control for multiple comparisons cluster method



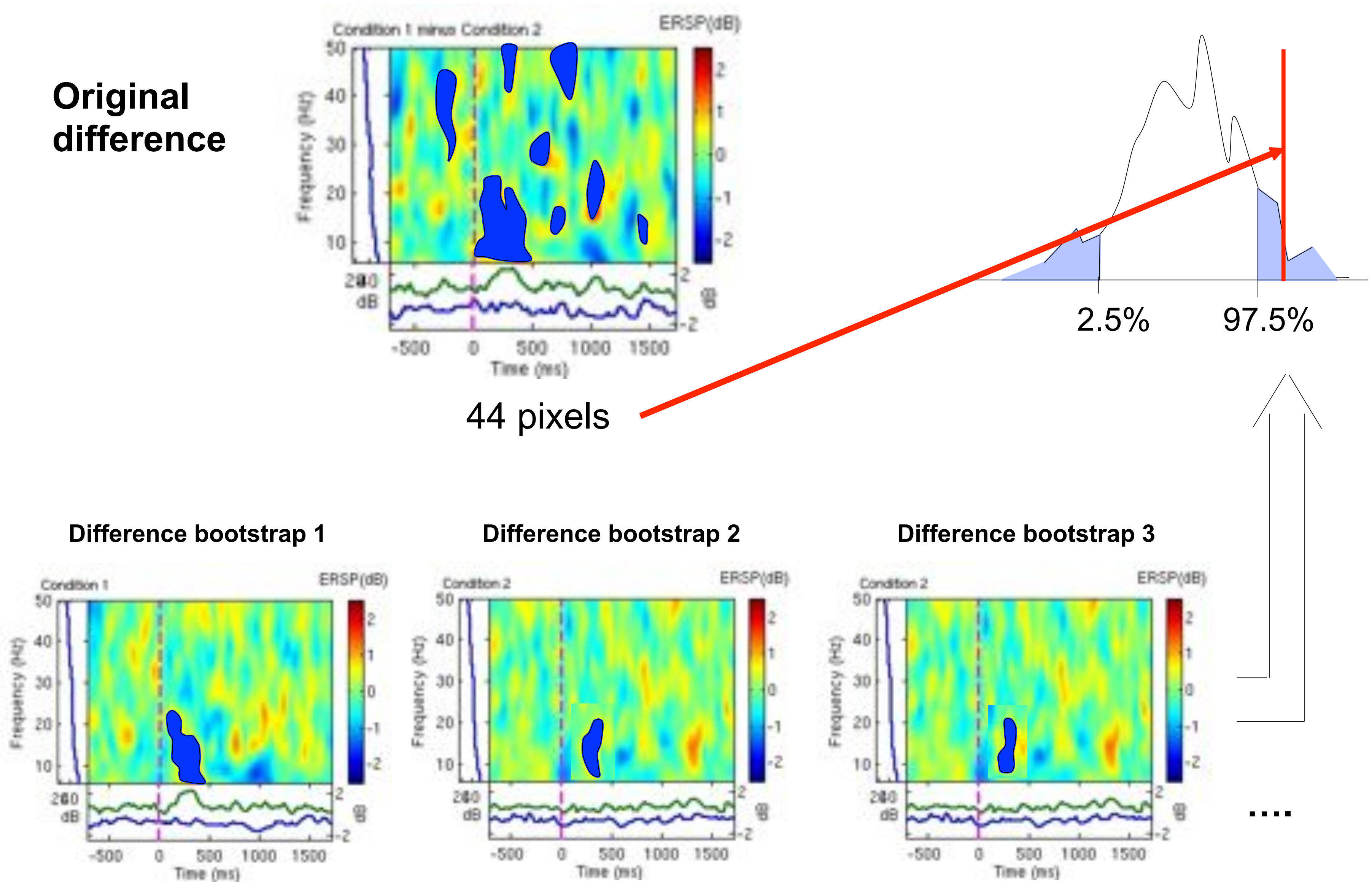
Control for multiple comparisons

Control for multiple comparisons cluster method



Maris & Oostenveld, J. Neurosci. Methods 2007

Cluster correction for multiple comparisons



Original difference

44 pixels

2.5%

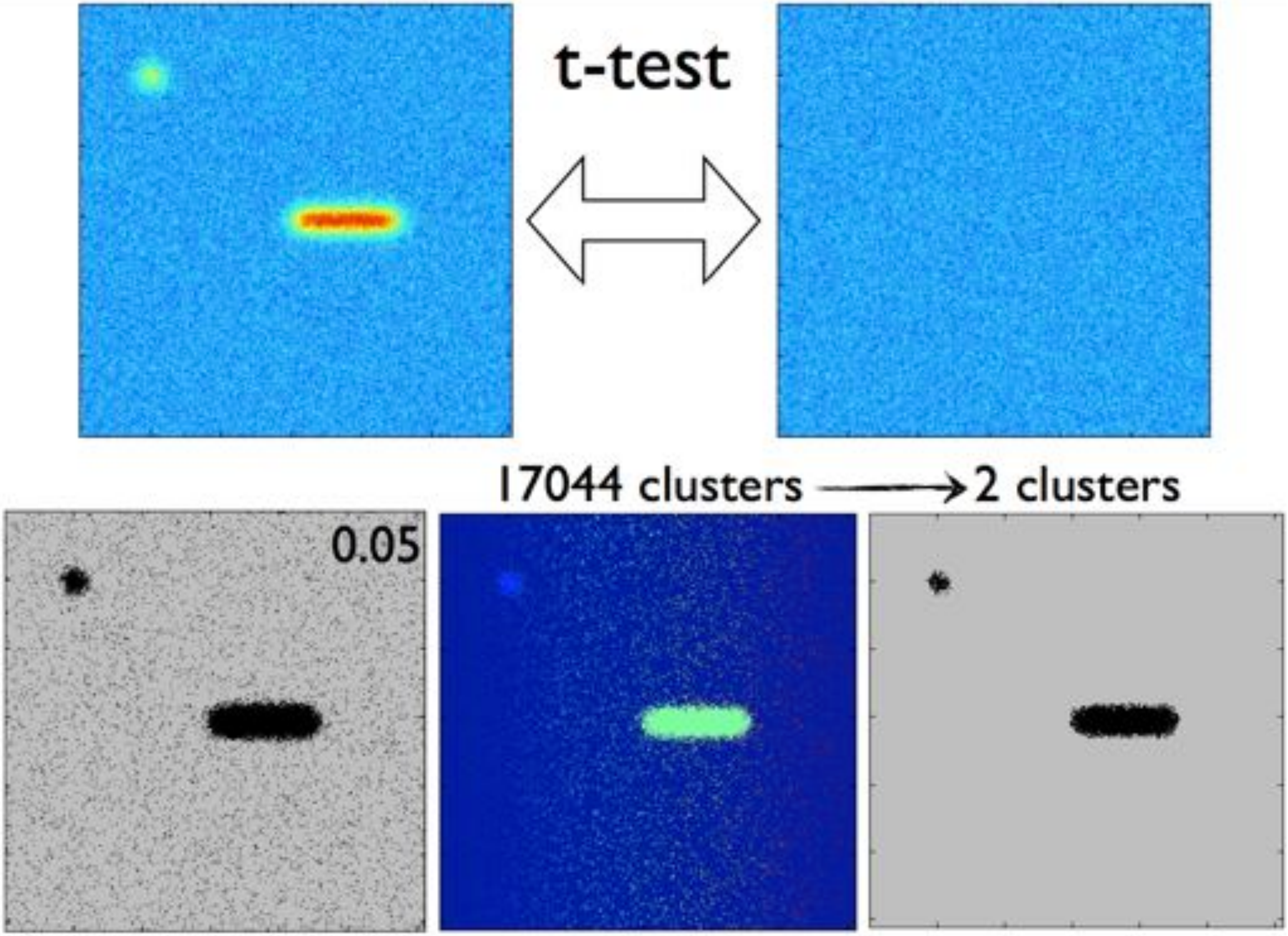
97.5%

Difference bootstrap 1

Difference bootstrap 2

Difference bootstrap 3

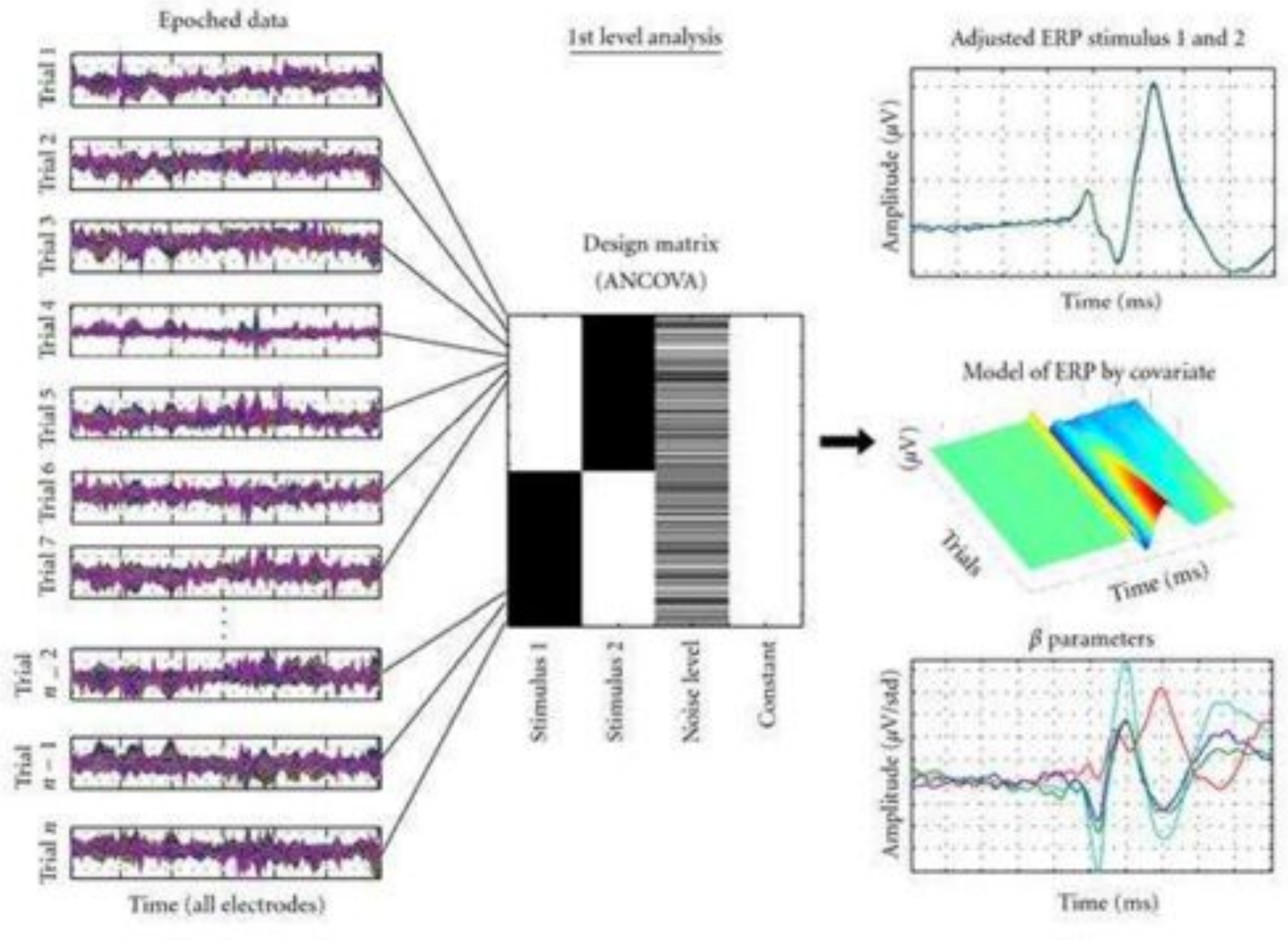
....



t-test

17044 clusters ⟶ 2 clusters

LIMO EEG


Epoched data
Trial 1
Trial 2
Trial 3
Trial 4
Trial 5
Trial 6
Trial 7
Trial n − 2
Trial n − 1
Trial n
Time (all electrodes)
1st level analysis
Design matrix
(ANCOVA)
Stimulus 1
Stimulus 2
Noise level
Constant
Adjusted ERP stimulus 1 and 2
Amplitude (μV)
Time (ms)
Model of ERP by covariate
(μV)
Trials
Time (ms)
β parameters
Amplitude (μV/std)
Time (ms)

References

Delorme, A. 2006. Statistical methods. *Encyclopedia of Medical Device and Instrumentation*, vol 6, pp 240-264. Wiley interscience.

Genovese et al. 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage*, 15: 870-878

Nichols & Hayasaka, 2003. Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical Methods in Medical Research*, 12:419-446

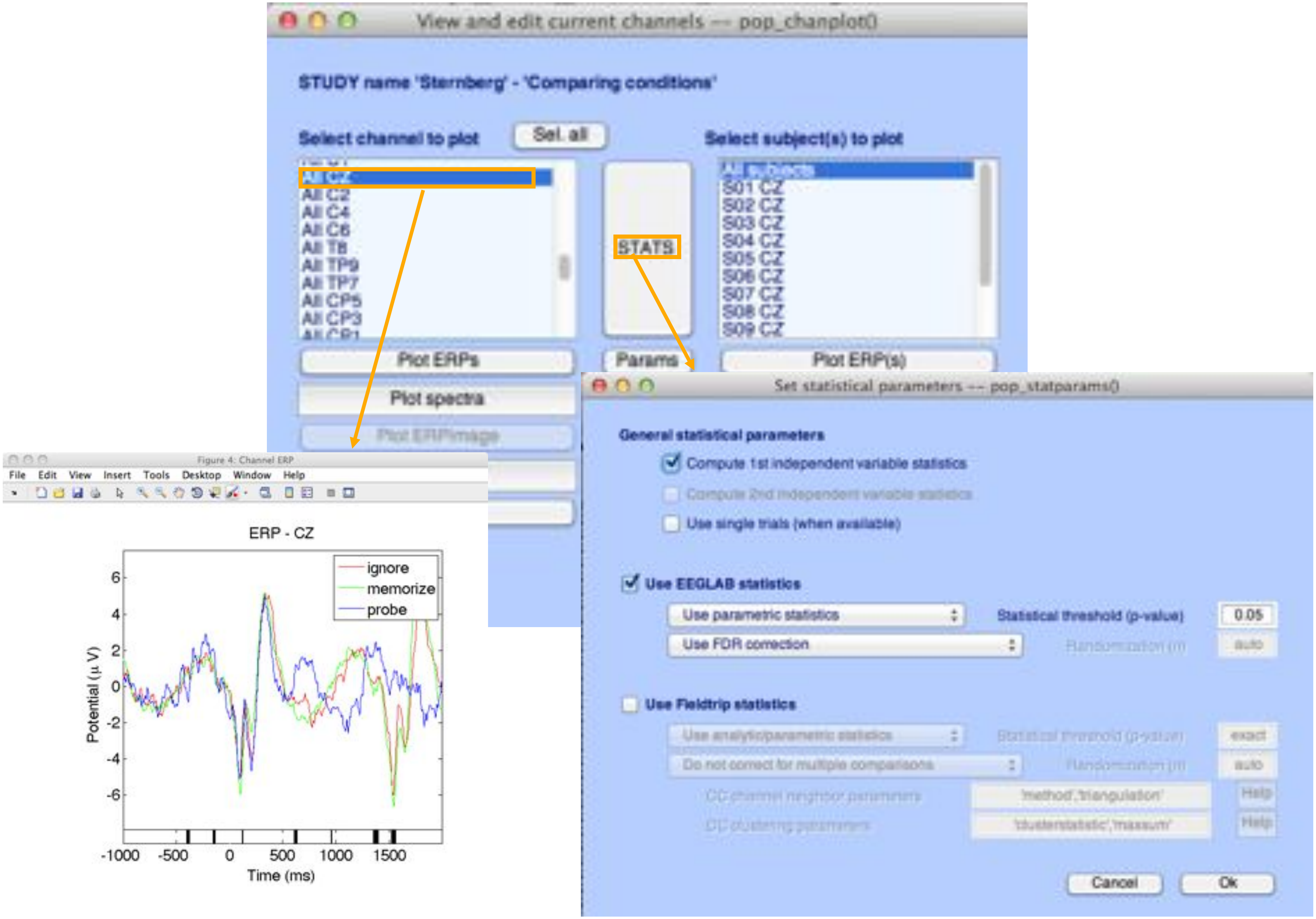
Maris, 2004. Randomization tests for ERP topographies and whole spatiotemporal data matrices. *Psychophysiology*, 41: 142-151

Maris et al. 2007. Nonparametric statistical testing of coherence differences. *Journal of Neuroscience Methods*, 163: 161-175

Groppe, D.M., Urbach, T.P., & Kutas, M. (2011) *Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review*. Psychophysiology, 48(12) pp. 1711-1725.

Thanks to G. Rousselet

EEGLAB and STUDY statistics practicum



View and edit current channels -- pop_chanplot()

STUDY name 'Sternberg' - 'Comparing conditions'

Select channel to plot

Select subject(s) to plot

Set statistical parameters -- pop_statparams()

General statistical parameters

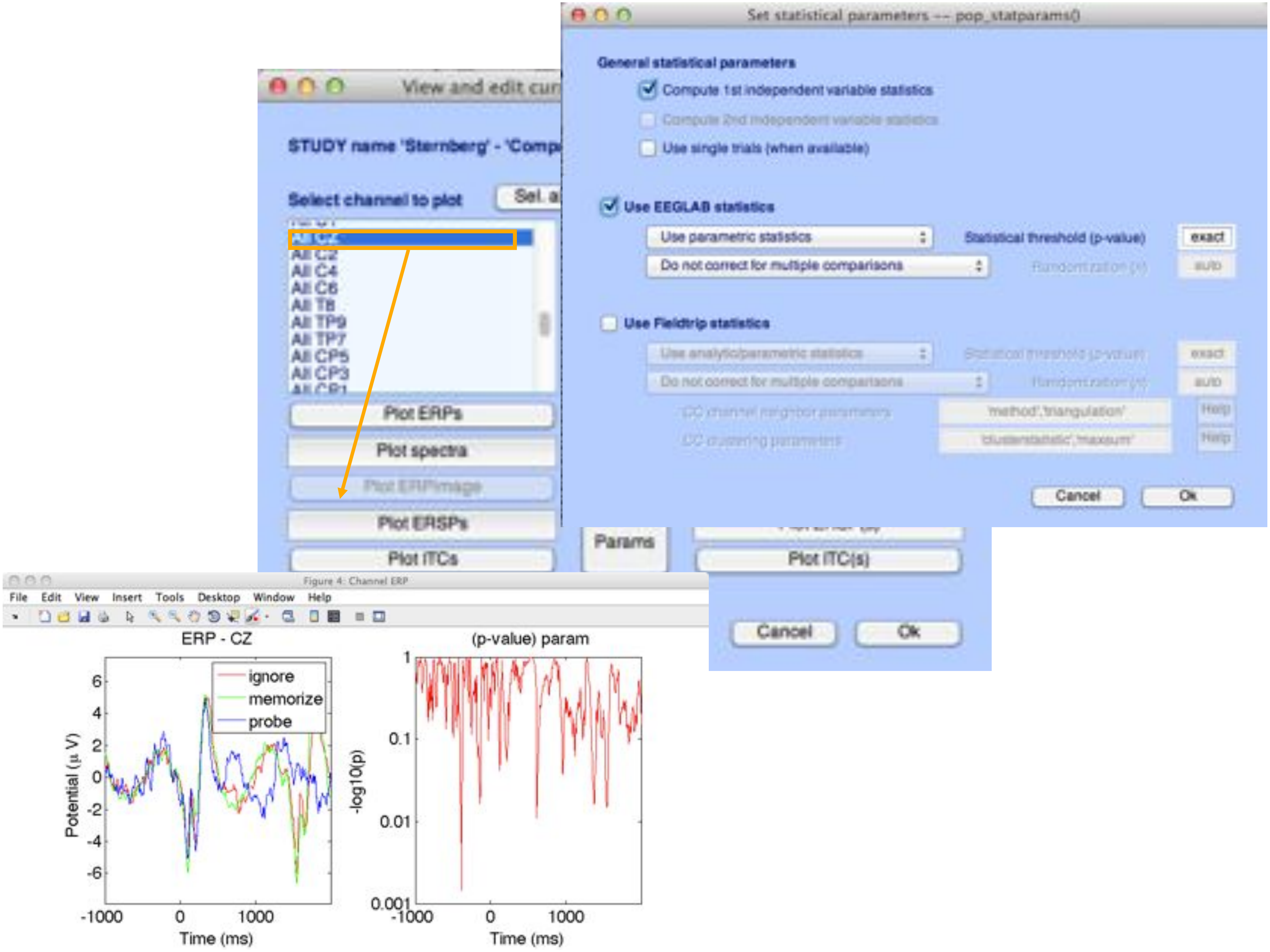
- ☑ Compute 1st independent variable statistics
- ☐ Compute 2nd independent variable statistics
- ☐ Use single trials (when available)

☑ Use EEGLAB statistics

- Use parametric statistics — Statistical threshold (p-value): 0.05
- Use FDR correction — Randomization (n): auto

☐ Use Fieldtrip statistics

- Use analytic/parametric statistics — Statistical threshold (p-value): exact
- Do not correct for multiple comparisons — Randomization (n): auto
- CC channel neighbor parameters: 'method','triangulation'
- CC clustering parameters: 'clusterstatistic','maxsum'



Set statistical parameters -- pop_statparams()

General statistical parameters

☑ Compute 1st independent variable statistics

☐ Compute 2nd independent variable statistics

☐ Use single trials (when available)

☑ **Use EEGLAB statistics**

Use parametric statistics — Statistical threshold (p-value): exact

Do not correct for multiple comparisons — Random ration (n): auto

☐ **Use Fieldtrip statistics**

Use analytic/parametric statistics — Statistical threshold (p-value): exact

Do not correct for multiple comparisons — Randomization (n): auto

CC channel neighbor parameters: 'method','triangulation'

CC clustering parameters: 'clusterstatistic','maxsum'

View and edit cur

STUDY name 'Sternberg' - 'Comp

Select channel to plot

All CZ

All C2

All C4

All C6

All T8

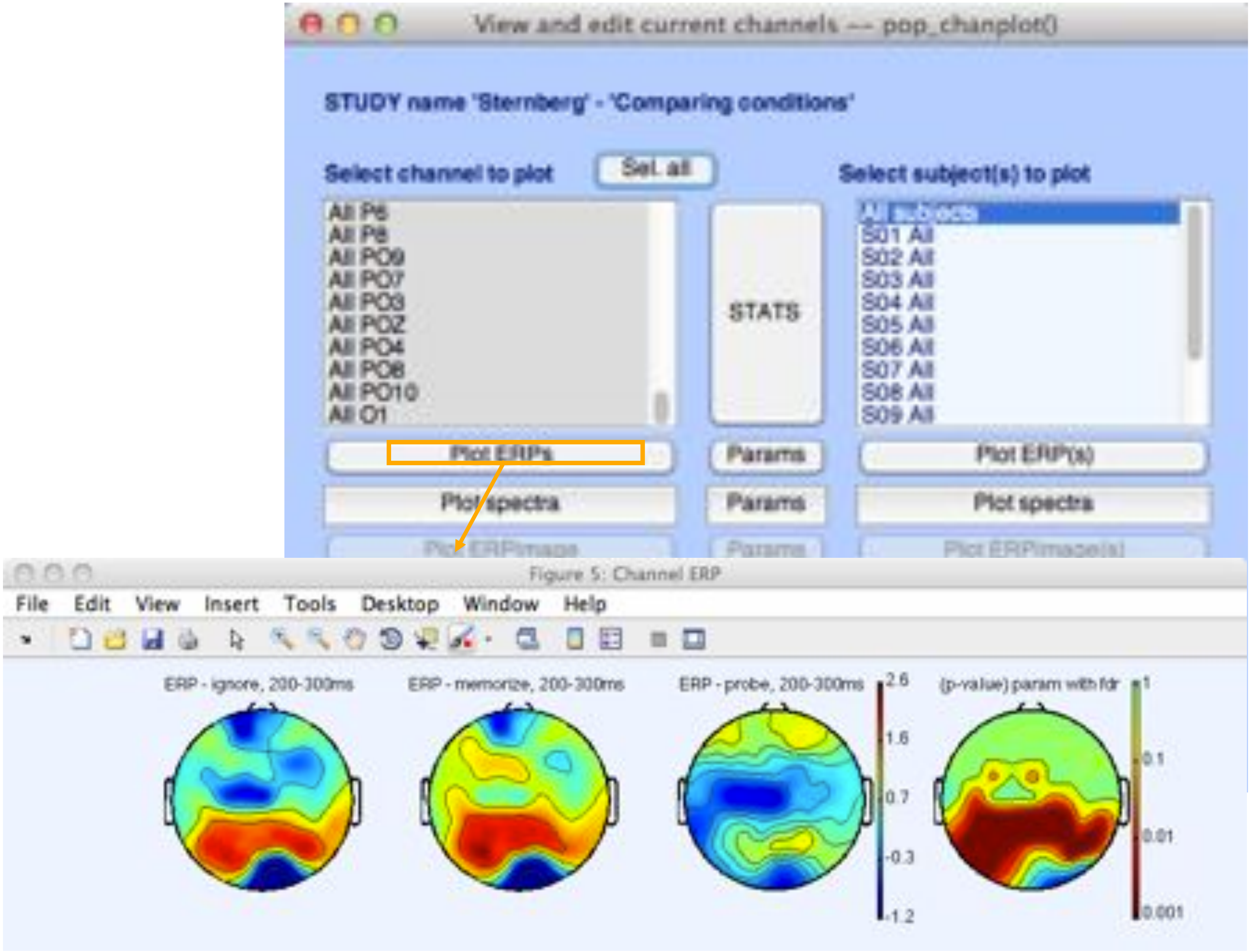
All TP9

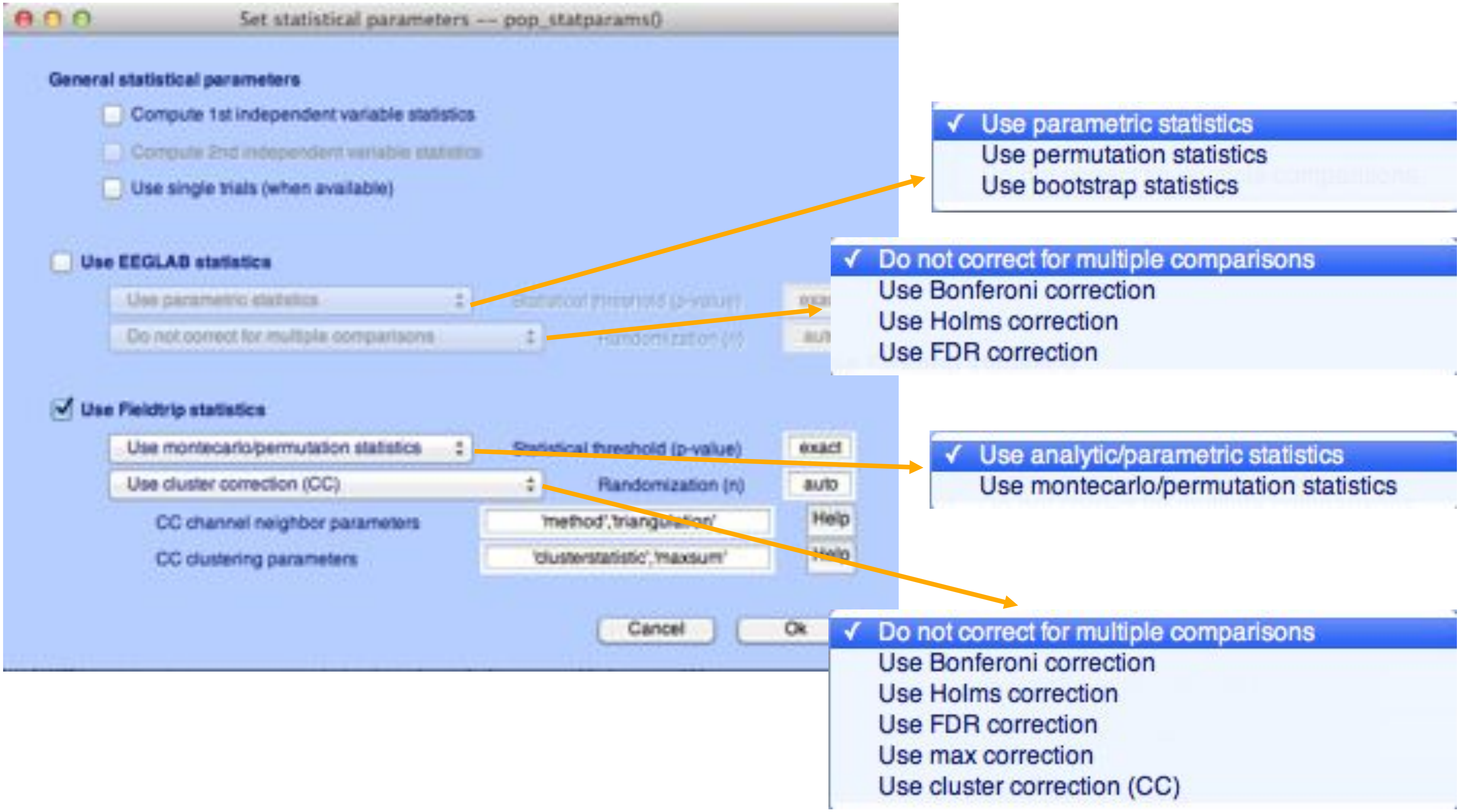
All TP7

All CP5

All CP3

Figure 4: Channel ERP


View and edit current channels -- pop_chanplot()
STUDY name 'Sternberg' - 'Comparing conditions'
Select channel to plot
Sel. all
Select subject(s) to plot
All P6
All P8
All PO9
All PO7
All PO3
All POZ
All PO4
All PO8
All PO10
All O1
STATS
All subjects
S01 All
S02 All
S03 All
S04 All
S05 All
S06 All
S07 All
S08 All
S09 All
Plot ERPs
Params
Plot ERP(s)
Plot spectra
Params
Plot spectra
Plot ERPimage
Params
Plot ERPimage(s)
Figure 5: Channel ERP
File
Edit
View
Insert
Tools
Desktop
Window
Help
ERP - ignore, 200-300ms
ERP - memorize, 200-300ms
ERP - probe, 200-300ms
2.6
1.6
0.7
-0.3
-1.2
(p-value) param with fdr
1
0.1
0.01
0.001



Set statistical parameters -- pop_statparams()

General statistical parameters

☐ Compute 1st independent variable statistics

☐ Compute 2nd independent variable statistics

☐ Use single trials (when available)

☐ Use EEGLAB statistics

Use parametric statistics

Do not correct for multiple comparisons

Statistical threshold (p-value): exact

Randomization (n): auto

☑ Use Fieldtrip statistics

Use montecarlo/permutation statistics

Use cluster correction (CC)

Statistical threshold (p-value): exact

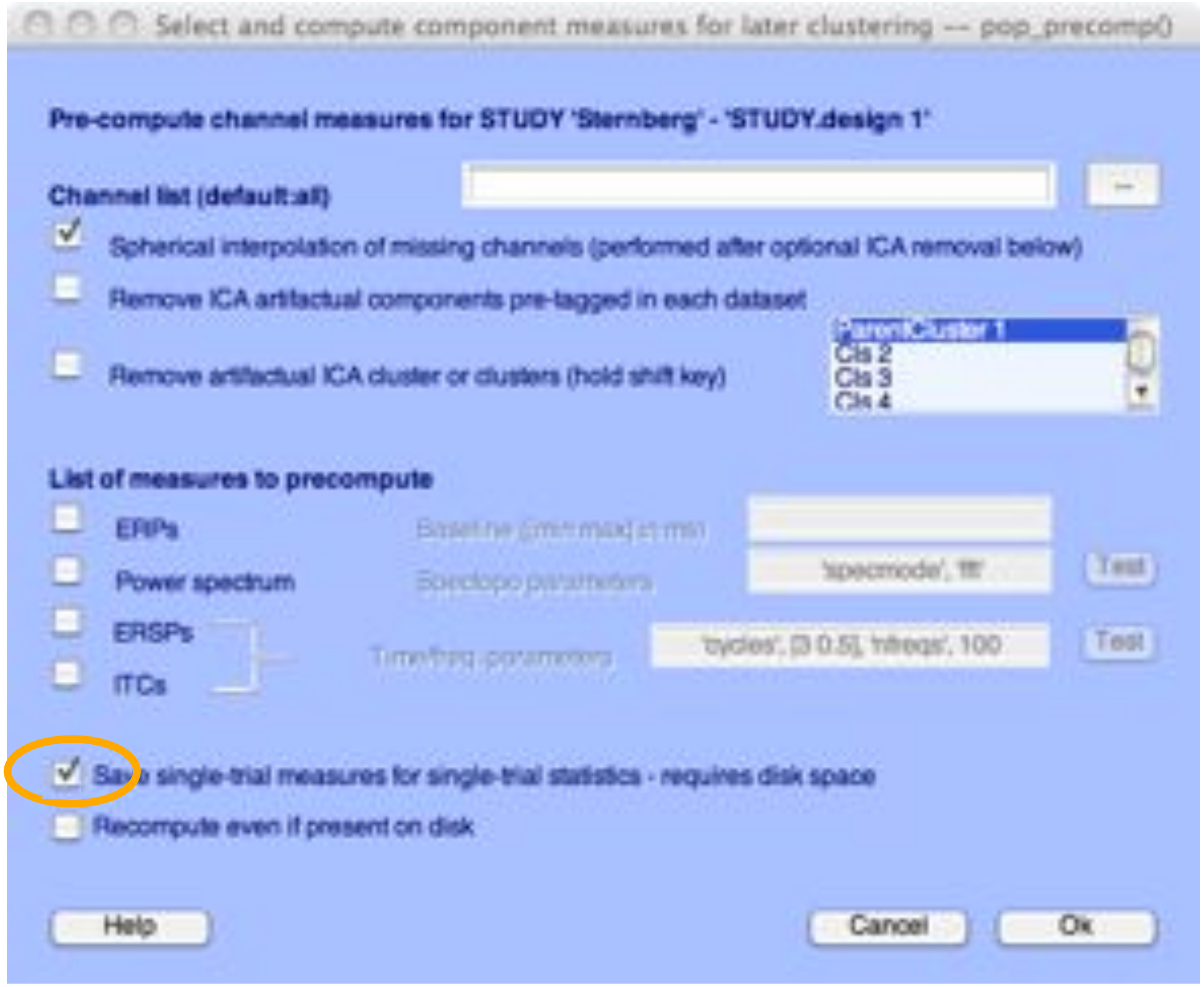
Randomization (n): auto

CC channel neighbor parameters: 'method','triangulation'

CC clustering parameters: 'clusterstatistic','maxsum'

std_stat() function in EEGLAB

Use single trials



Select and compute component measures for later clustering -- pop_precomp()

Pre-compute channel measures for STUDY 'Sternberg' - 'STUDY.design 1'

Channel list (default:all)

☑ Spherical interpolation of missing channels (performed after optional ICA removal below)

☐ Remove ICA artifactual components pre-tagged in each dataset

☐ Remove artifactual ICA cluster or clusters (hold shift key)

ParentCluster 1
Cls 2
Cls 3
Cls 4

List of measures to precompute

☐ ERPs — Baseline ([min max] in ms)

☐ Power spectrum — Spectopo parameters: 'specmode', 'fft'

☐ ERSPs

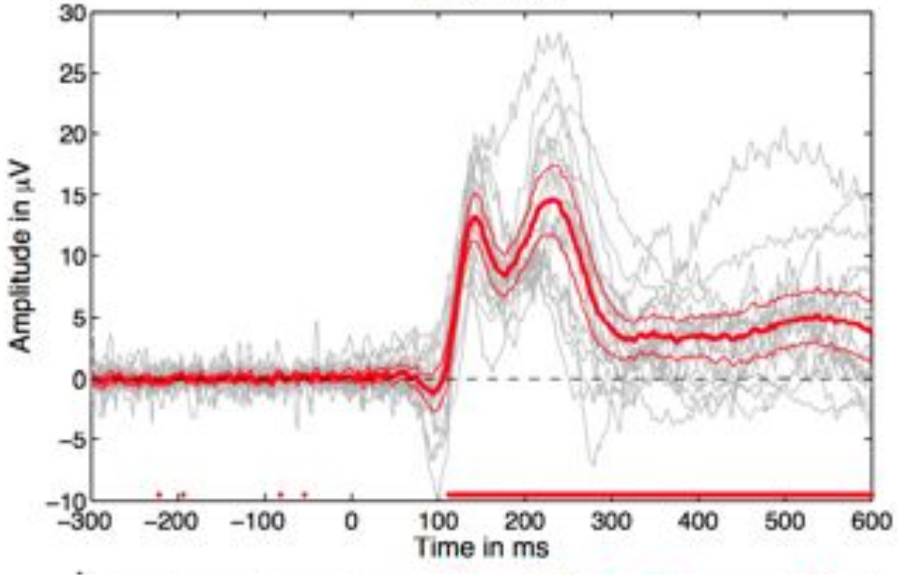
☐ ITCs

Time/freq. parameters: 'cycles', [3 0.5], 'nfreqs', 100

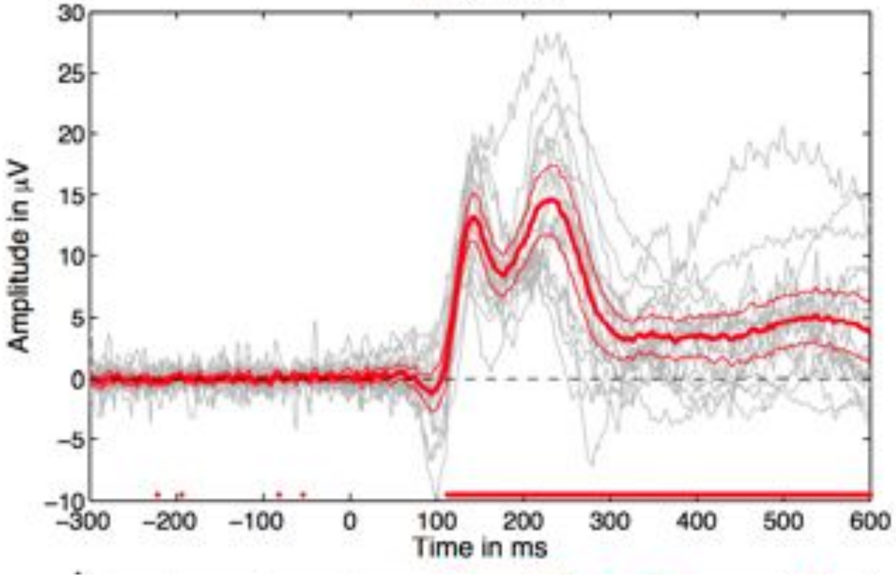
☑ Save single-trial measures for single-trial statistics - requires disk space

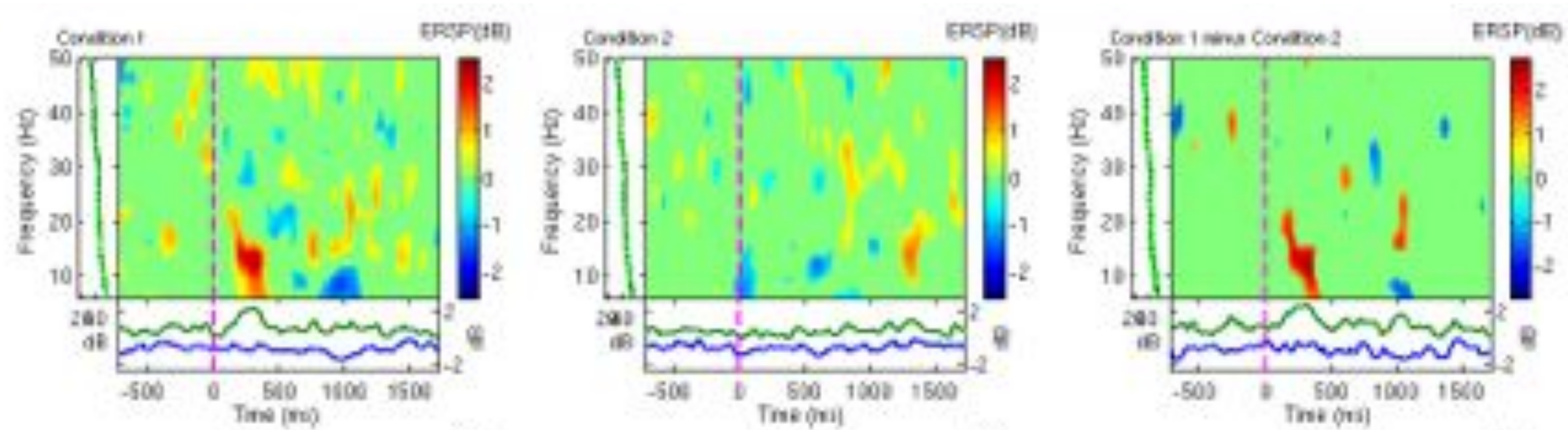
☐ Recompute even if present on disk

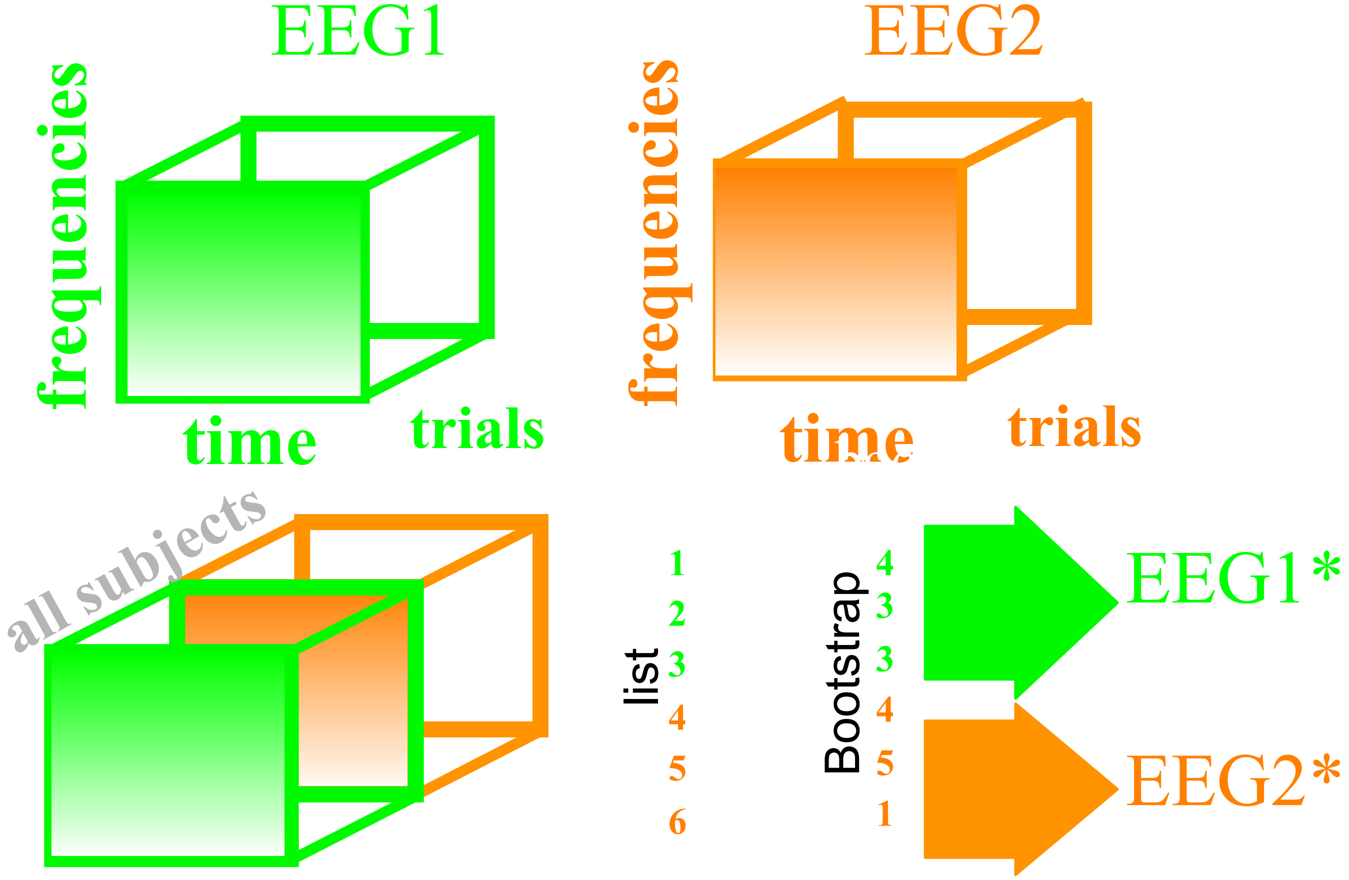
Use subjects’ average



Use single trials







EEG1

frequencies

time

trials

EEG2

frequencies

time

trials

all subjects

list

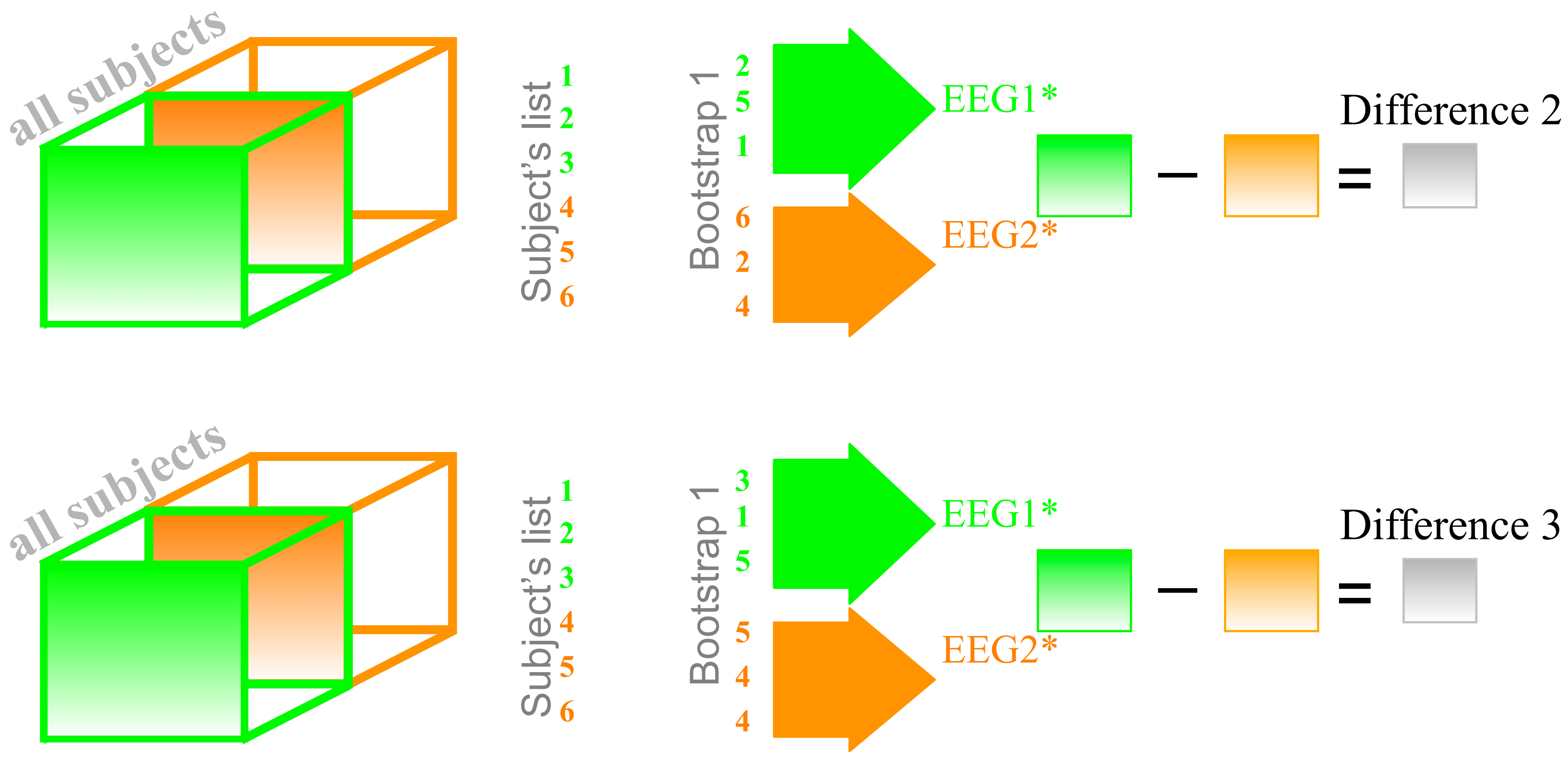
1
2
3
4
5
6

Bootstrap

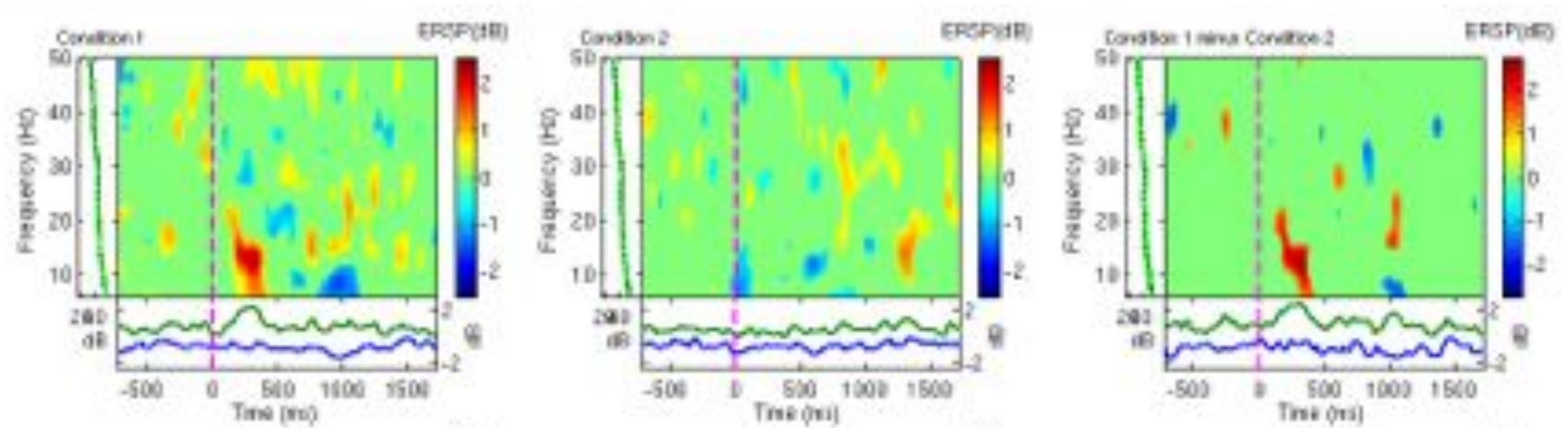
4
3
3
4
5
1

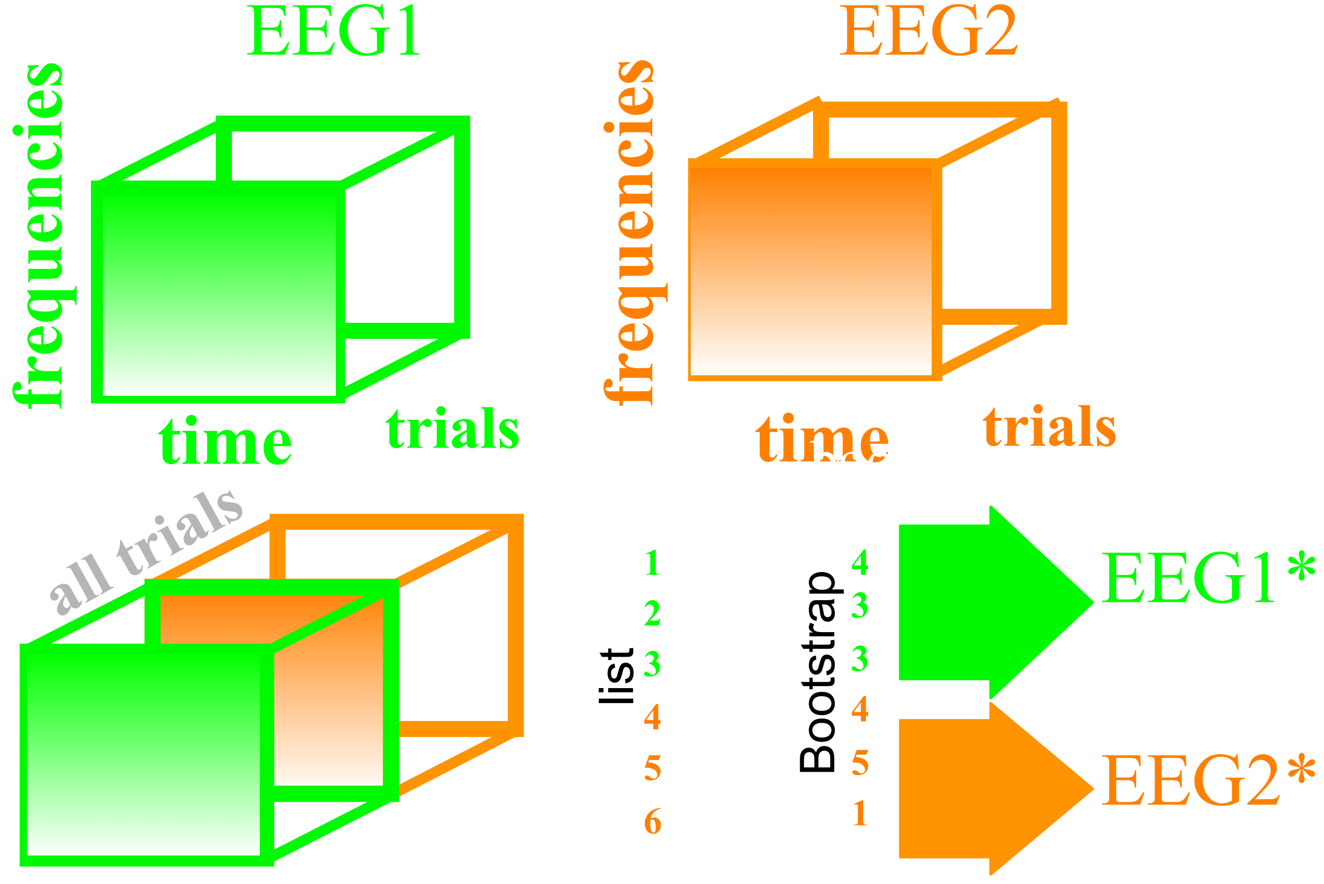
EEG1*

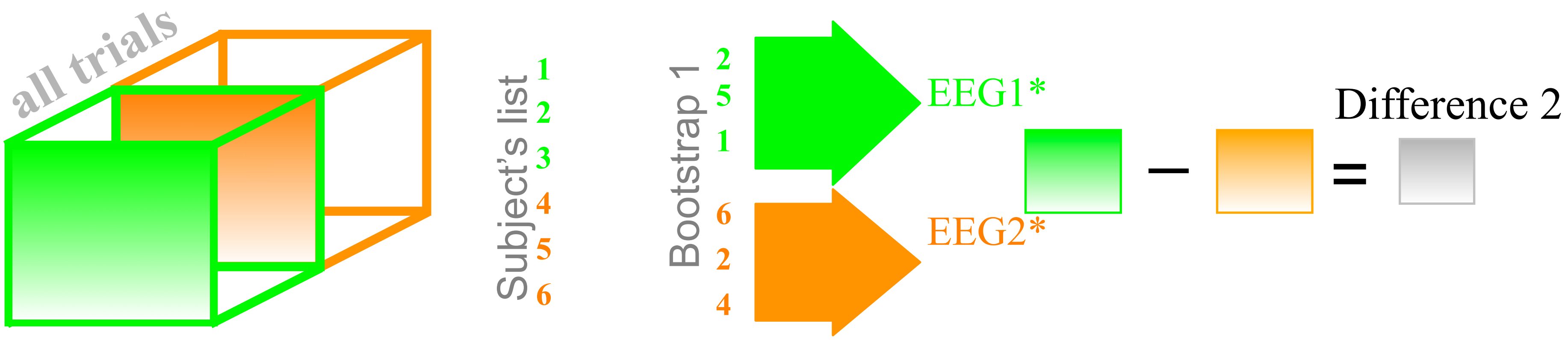
EEG2*

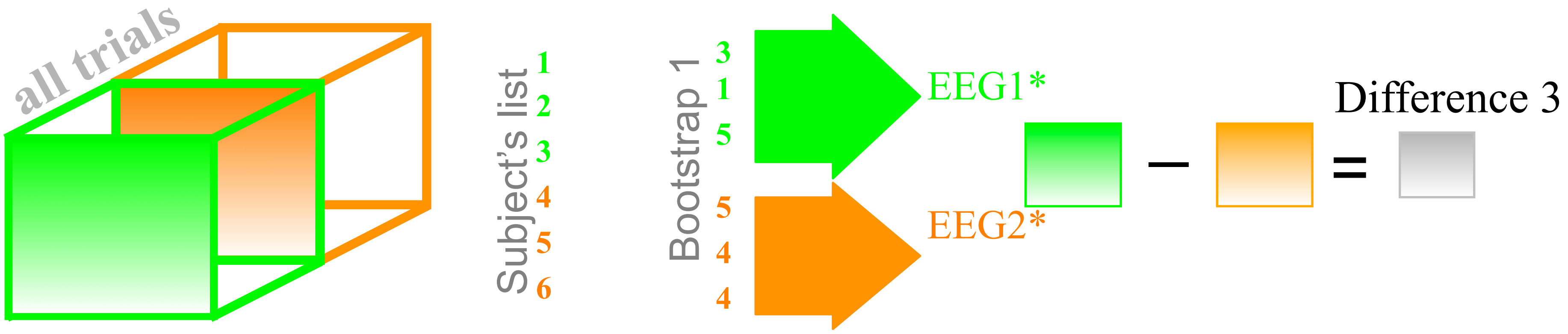

all subjects
Subject's list
1
2
3
4
5
6
Bootstrap 1
2
5
1
6
2
4
EEG1*
EEG2*
–
=
Difference 2
all subjects
Subject's list
1
2
3
4
5
6
Bootstrap 1
3
1
5
5
4
4
EEG1*
EEG2*
–
=
Difference 3

...


Condition 1
ERSP(dB)
Condition 2
ERSP(dB)
Condition 1 minus Condition 2
ERSP(dB)
Frequency (Hz)
50
40
30
20
10
2
1
0
-1
-2
dB
-500
0
500
1000
1500
Time (ms)

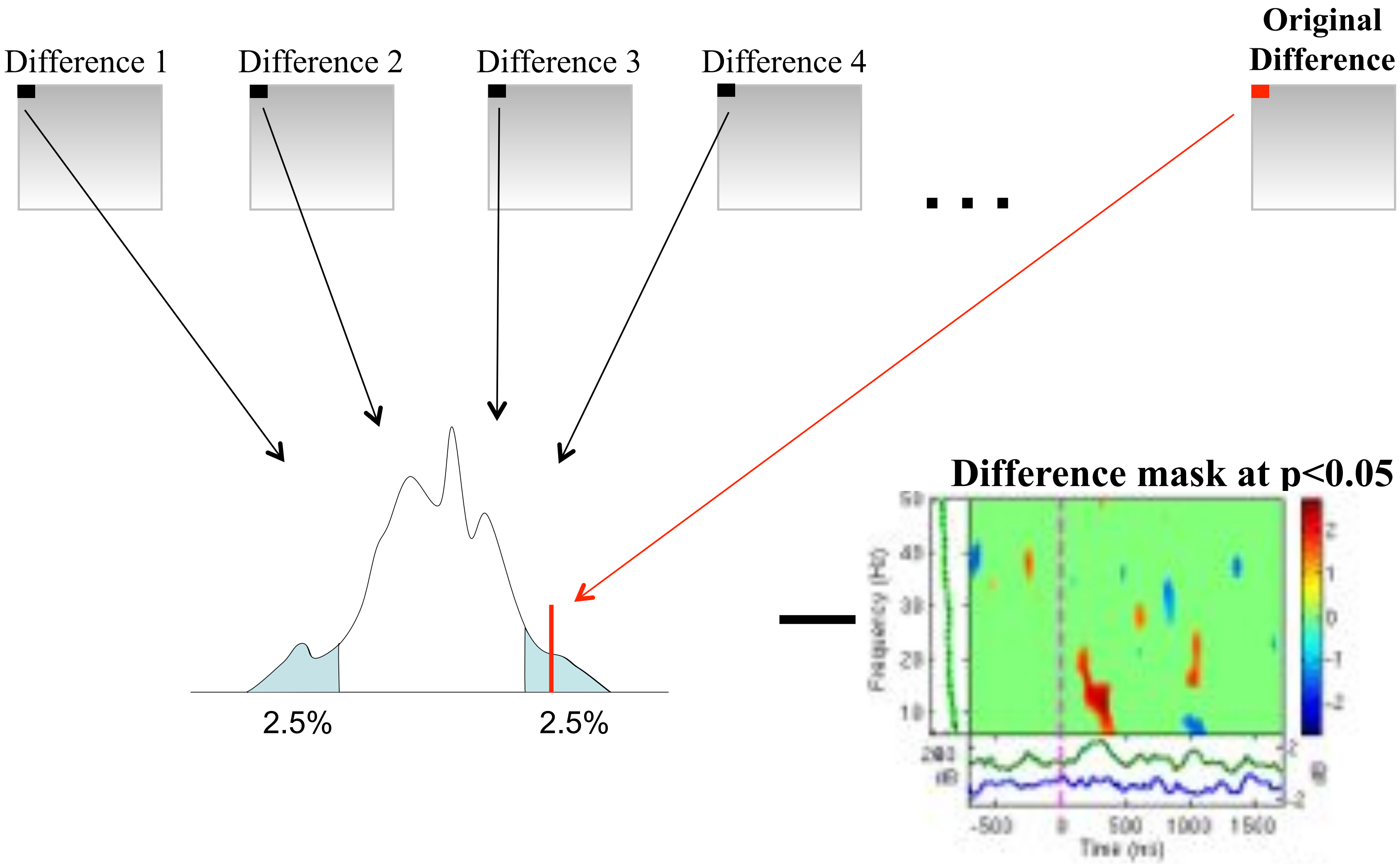

EEG1
frequencies
time
trials
EEG2
frequencies
time
trials
all trials
list
1
2
3
4
5
6
Bootstrap
4
3
3
4
5
1
EEG1*
EEG2*


all trials
Subject's list
1
2
3
4
5
6
Bootstrap 1
2
5
1
6
2
4
EEG1*
EEG2*
Difference 2
–
=


all trials
Subject's list
1
2
3
4
5
6
Bootstrap 1
3
1
5
5
4
4
EEG1*
EEG2*
Difference 3
–
=

...

Assessing significance for single trials



Difference 1

Difference 2

Difference 3

Difference 4

. . .

Original Difference

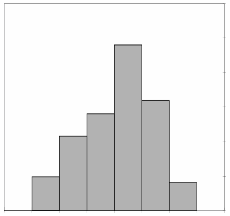
2.5%

2.5%

Difference mask at p<0.05

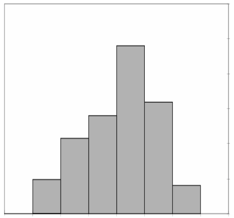
Using more than 1 subject when using single trials statistics

Subjects' values
(1 per subject)



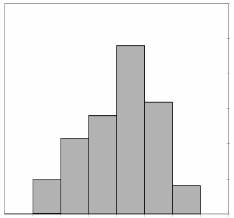
→ **Population of subjects**

Trials' values
(1 subject)



→ **Population of trials**

Trials' values
(n subject)



→ **???**

Bootstrap assumption:
given that we have no other information about the population, the sample is our best single estimate of the population

Exercices

- Experiment with STUDY statistics
  - Load the Stern STUDY
  - Look at significant difference in the first default design in channel Fz (time-frequency plot, ERSP) using the cluster method (Fieldtrip – statistics)
  - Look at the same difference a component cluster of your choice.