Robust statistics

June 14th, Jyväskylä, Finland

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.

Parametric statistics

Assume gaussian distribution of data

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

Paired

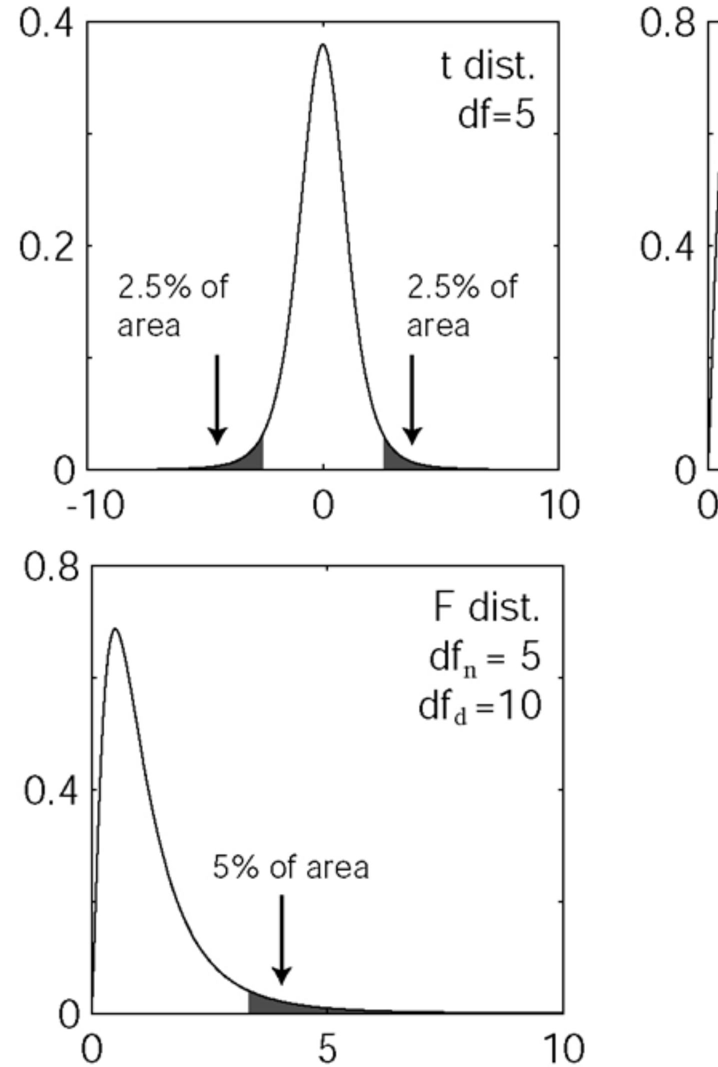
$$t = \frac{Mean\_difference}{Standard\_deviation}\sqrt{N-1}$$

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)

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



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

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

Non-parametric statistics

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

Values Ranks

BOTH ASSUME NORMAL DISTRIBUTIONS

Problems

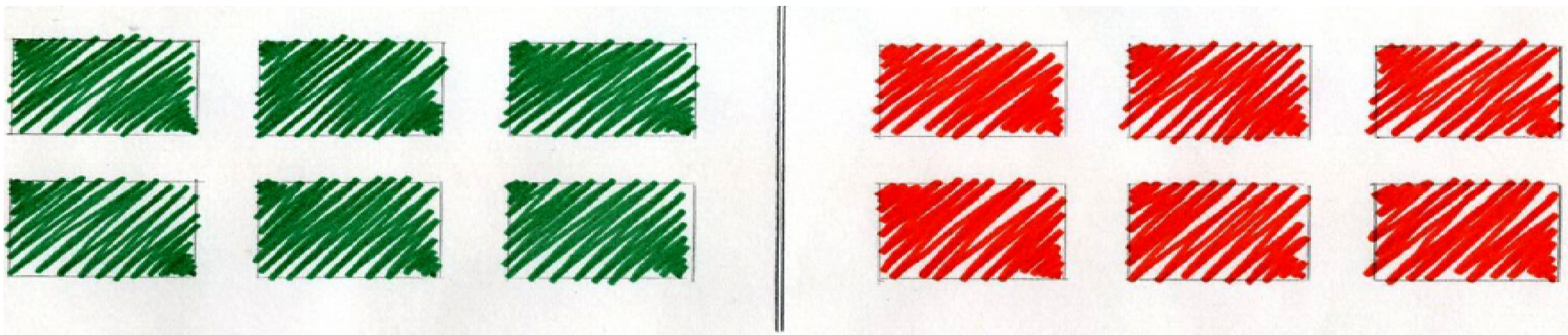
- Not resistant against outliers
- For ANOVA and t-test non-normality is an issue when distributions differ or when variances are not equal.
- Slight departure from normality can have serious consequences

---

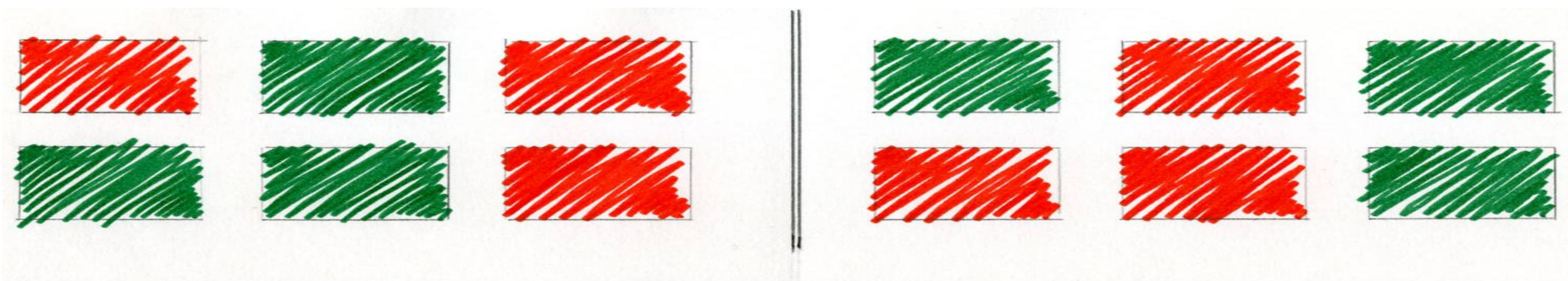
Solutions

1. Randomization approach
2. Bootstrap approach

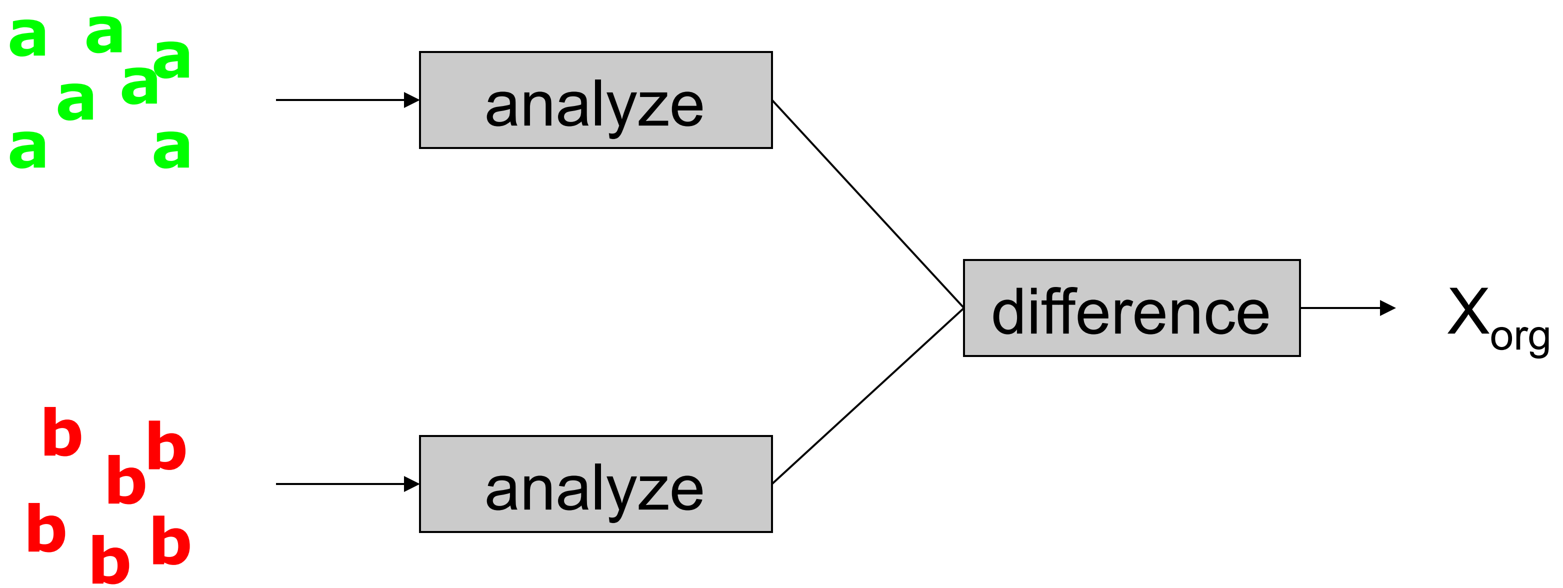
Randomization approach



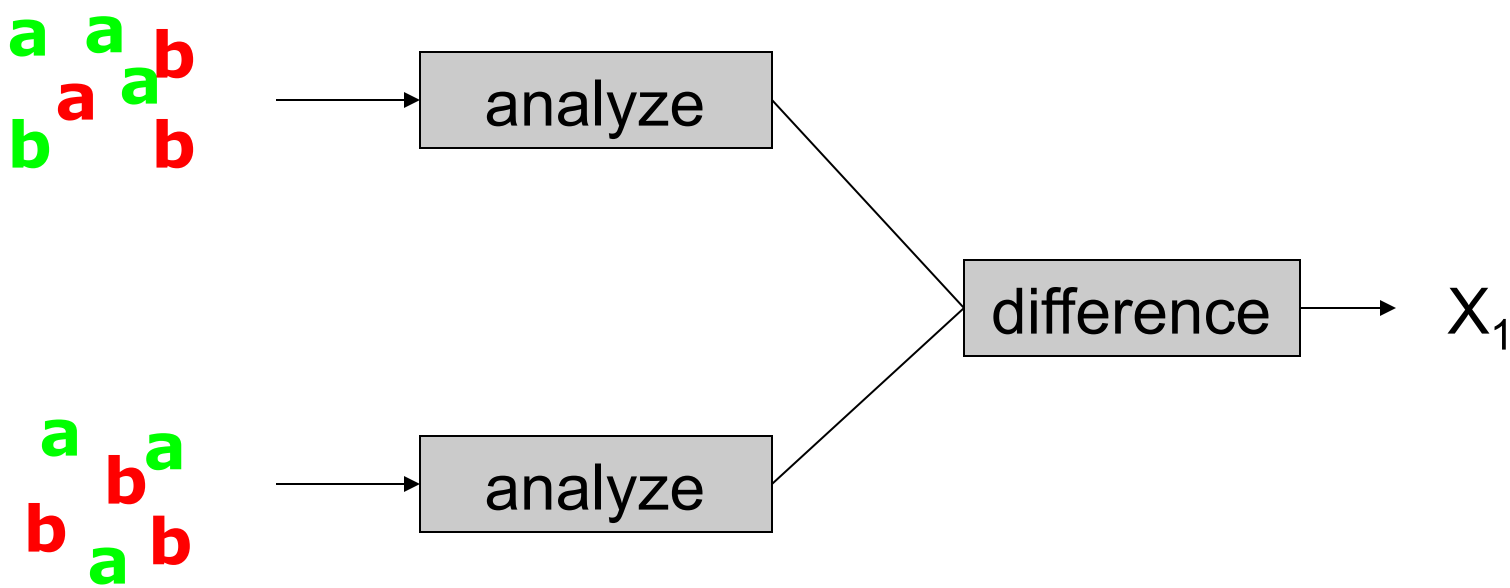
Randomization approach



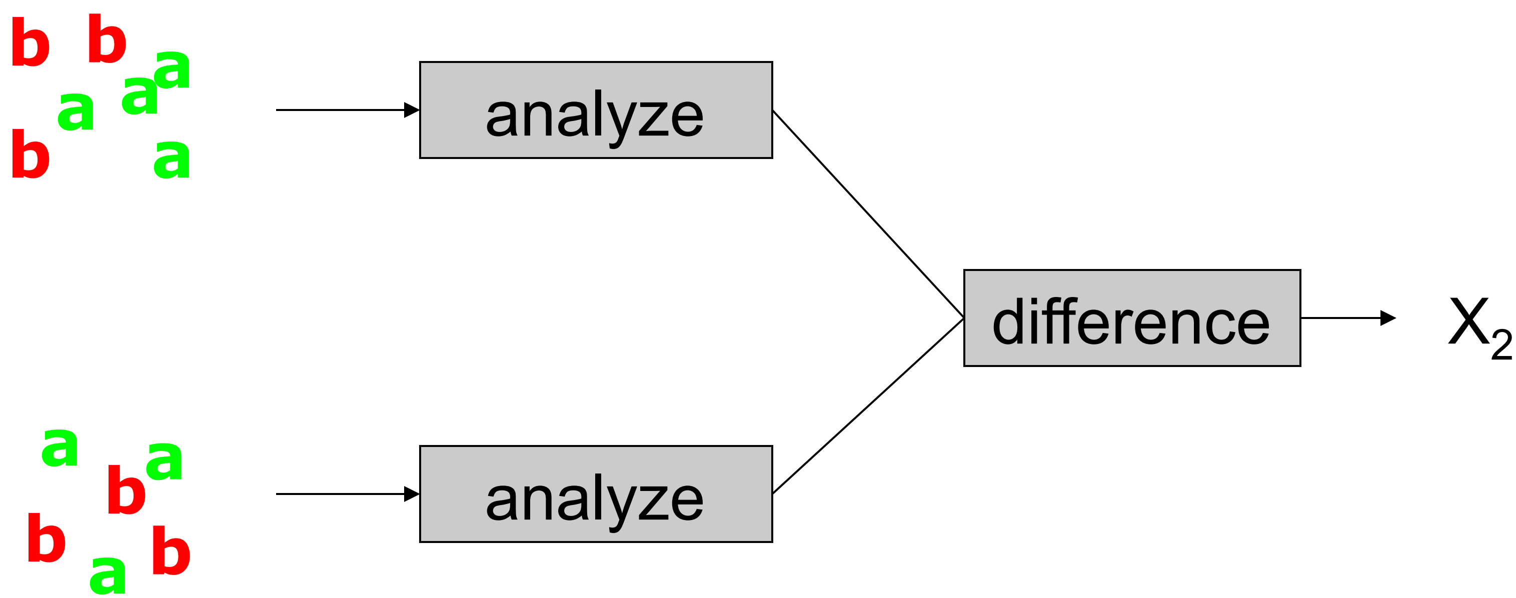
Randomization approach



Randomization approach



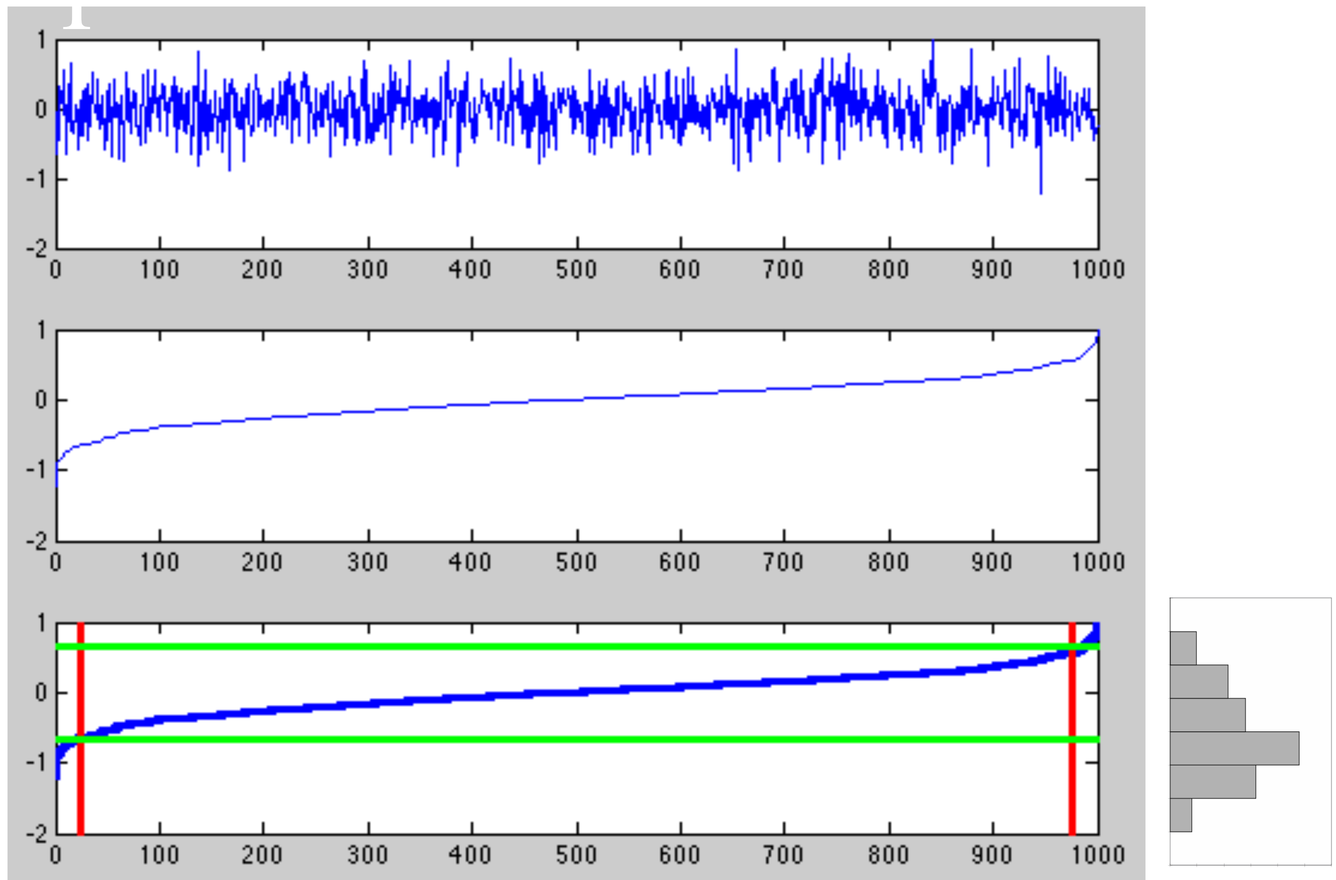
Randomization approach



Permutation
/bootstrap

Sorted values

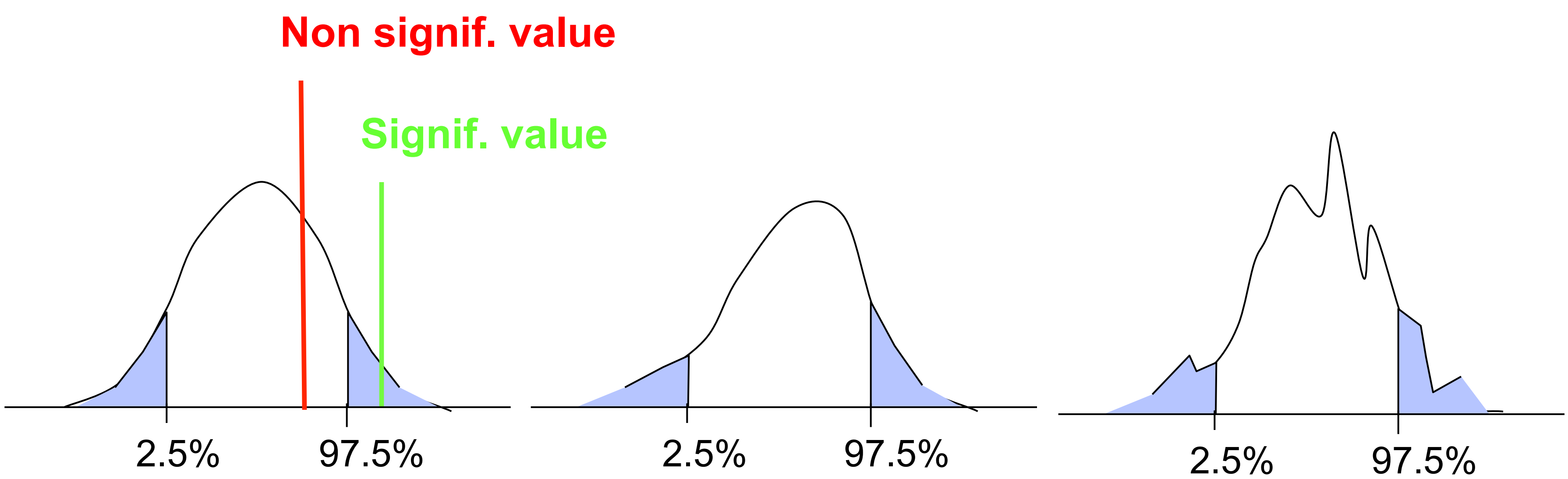
Thresholds



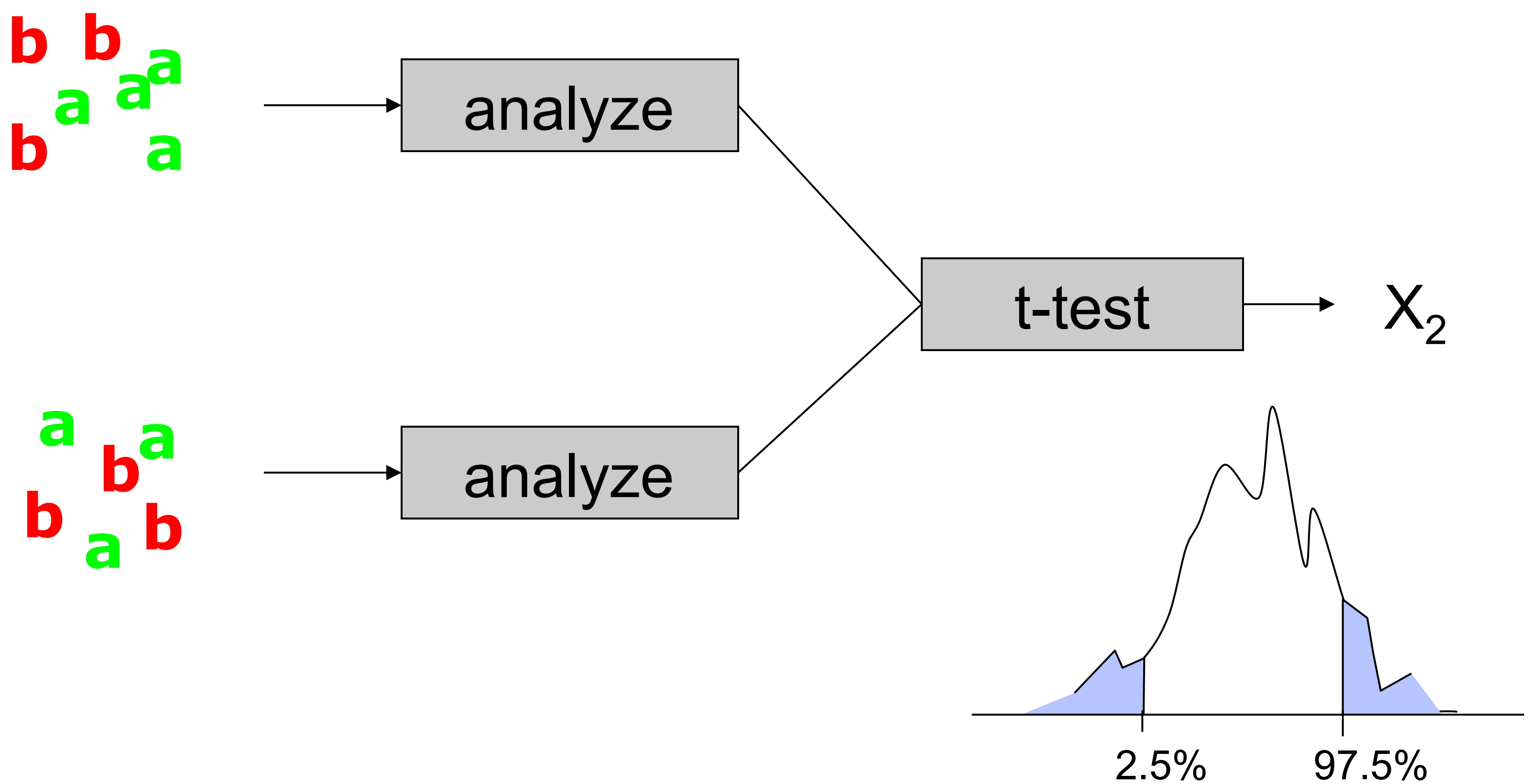
2.5%

97.5%

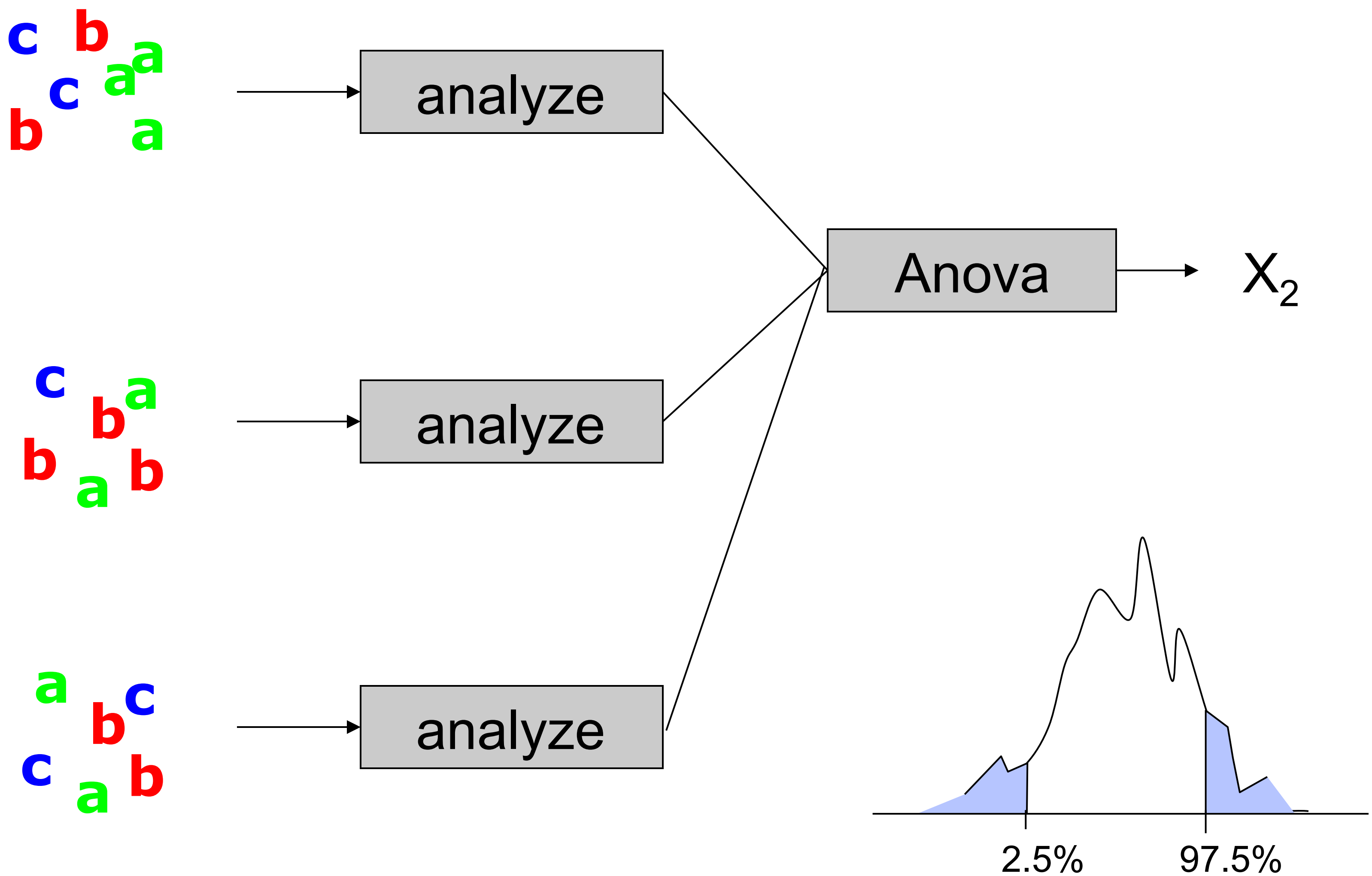
Distribution can take any shape



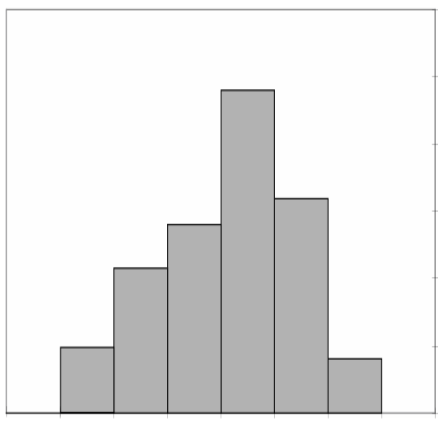
Randomization approach



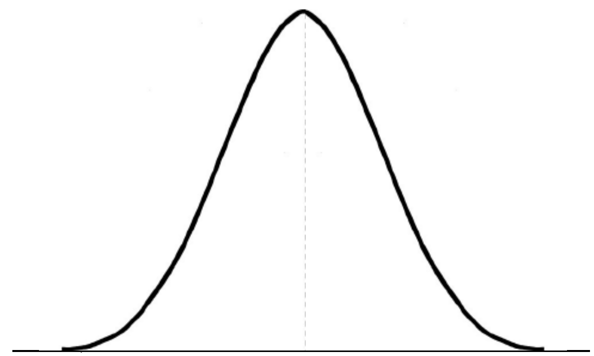
Randomization approach



Sample and population



Sample



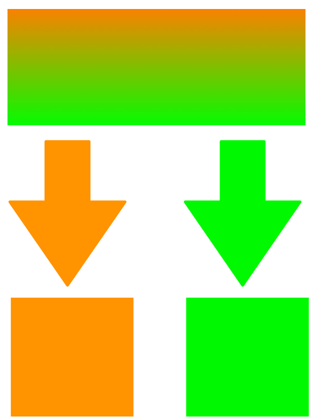
Population

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

Bootstrap versus permutation

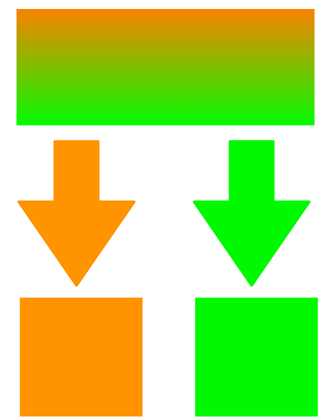
Permutation



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

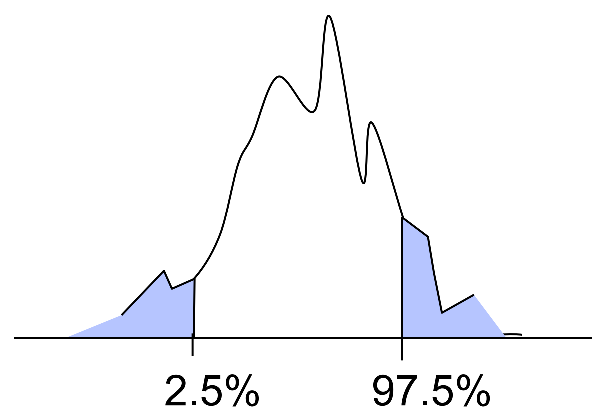
Bootstrap is better!

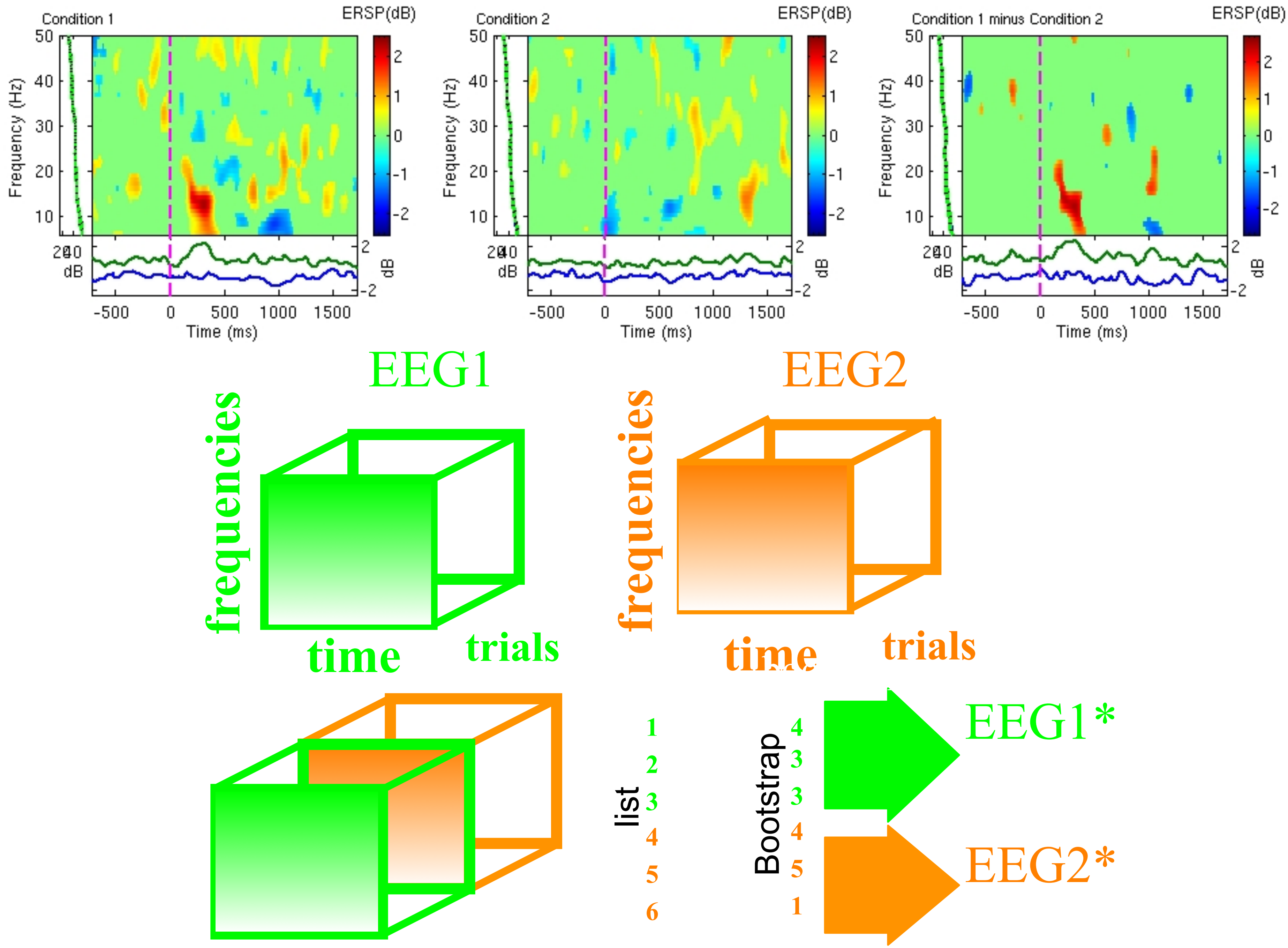
Husband	Wifes
22	25
32	25
50	51
25	25
33	38
27	30
45	60
47	54
30	31
44	54
23	23
39	34
24	25
22	23
16	19
73	71
27	26
36	31
24	26
60	62
26	29
23	31
28	29
36	35

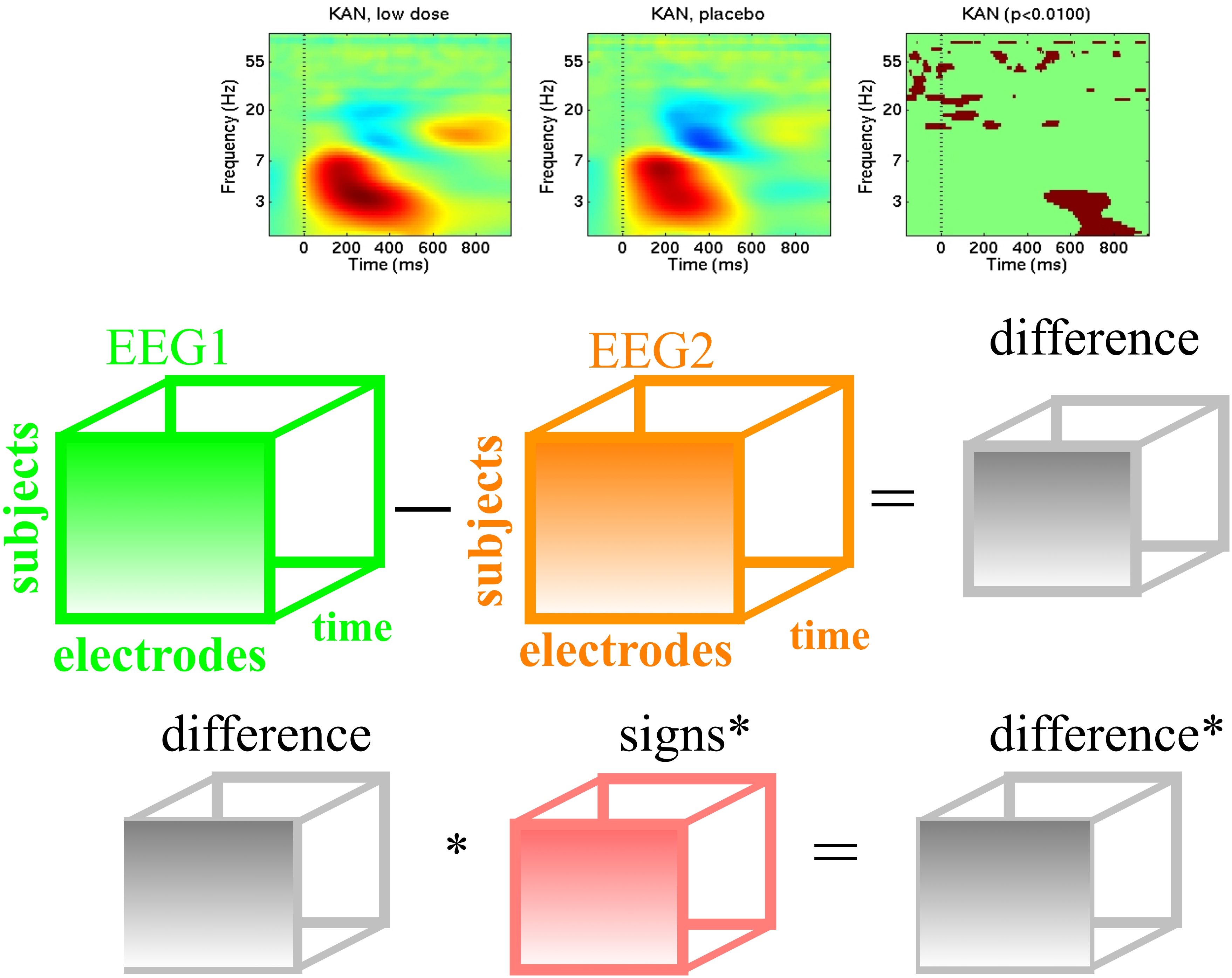
Median

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

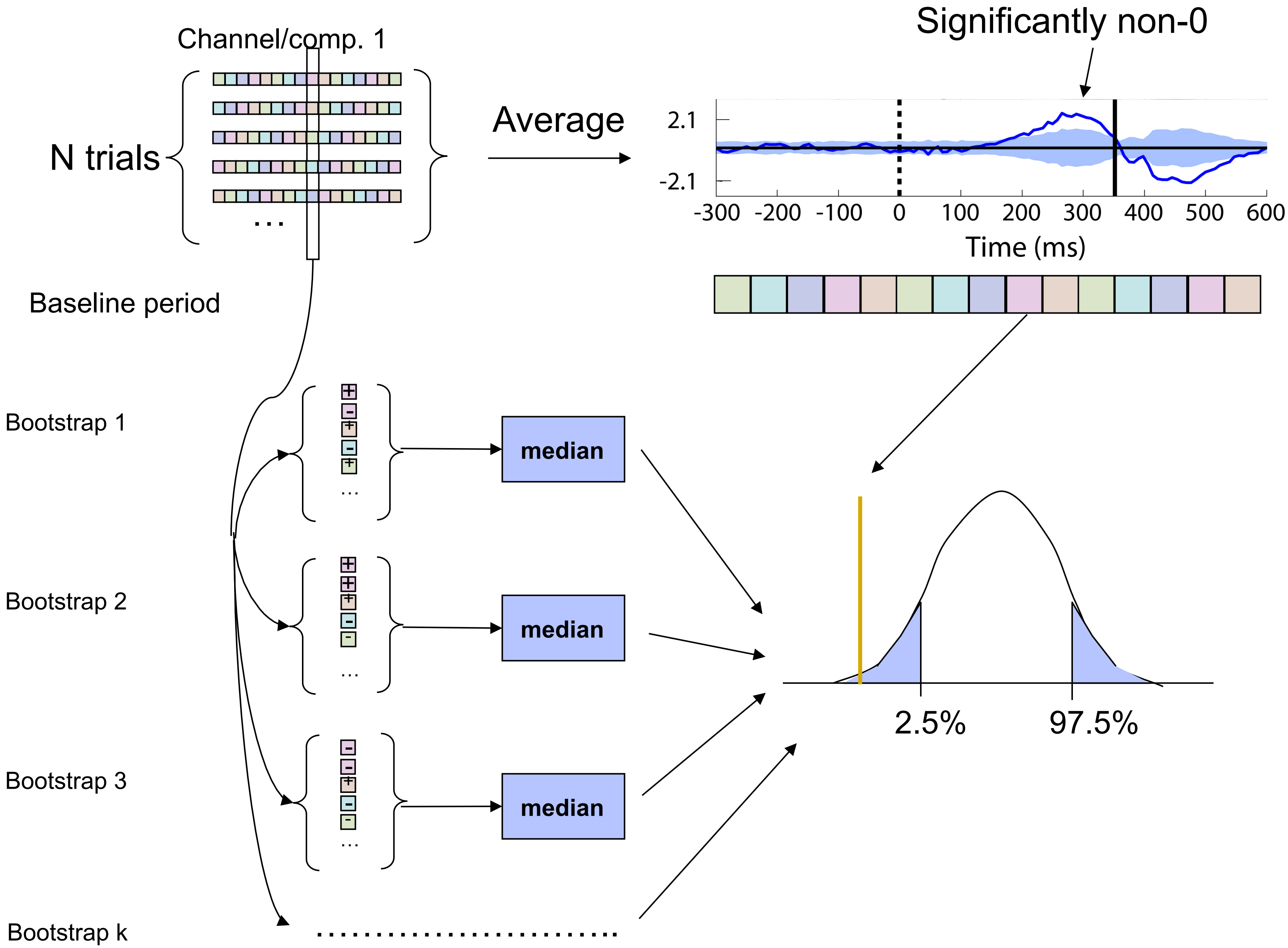
Are husbands older than wifes: that's a paired test. Compute difference between the two and change sign to bootstrap.



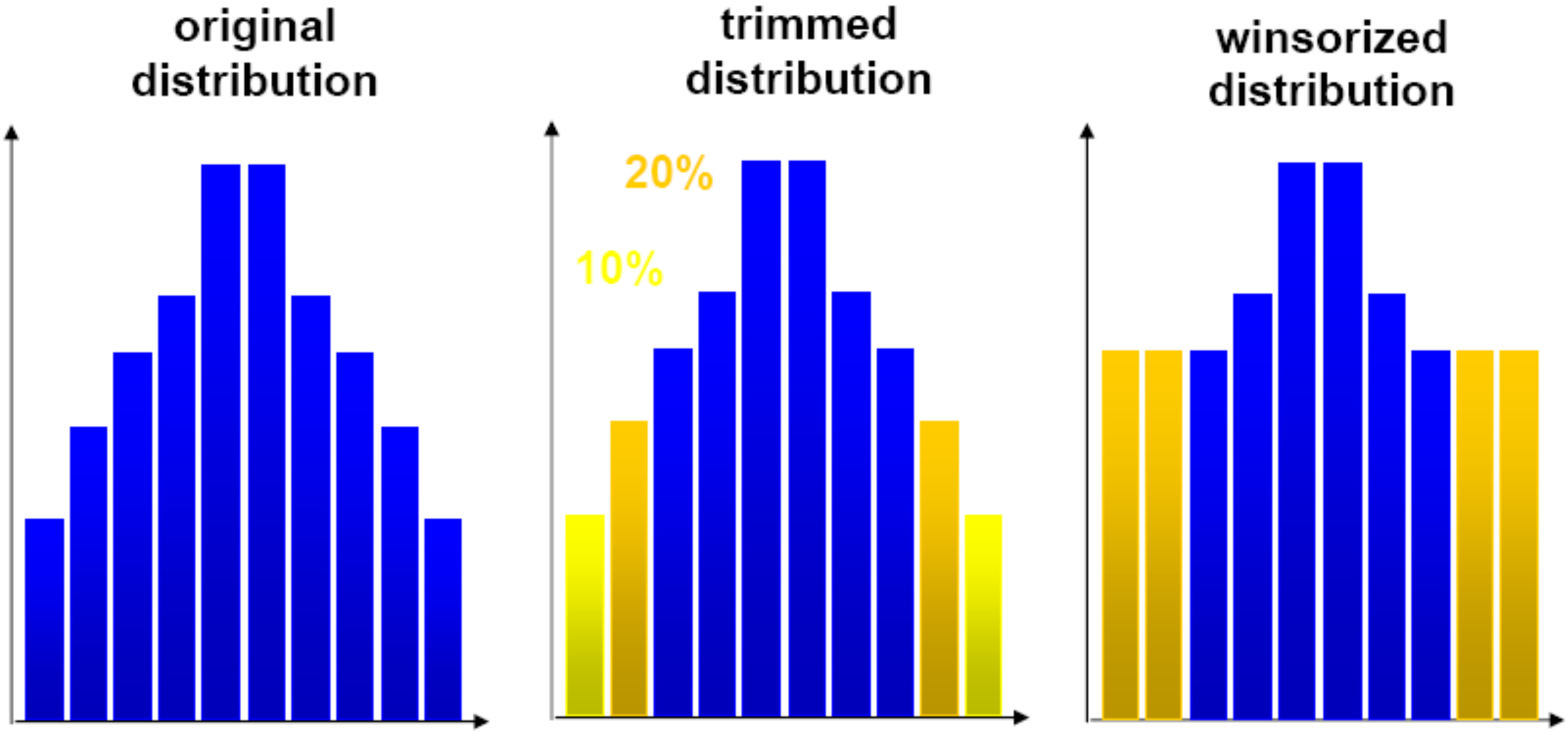

Condition 1
ERSP(dB)
Condition 2
ERSP(dB)
Condition 1 minus Condition 2
ERSP(dB)
Frequency (Hz)
Time (ms)
dB
EEG1
EEG2
frequencies
time
trials
list
1
2
3
4
5
6
Bootstrap
4
3
3
4
5
1
EEG1*
EEG2*


KAN, low dose
KAN, placebo
KAN (p<0.0100)
Frequency (Hz)
55
20
7
3
0
200
400
600
800
Time (ms)
EEG1
EEG2
difference
subjects
electrodes
time
difference
signs*
difference*

Bootstrap for ERPs and time-frequency



Measures of central tendency



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 α/(N-1)

- Max method

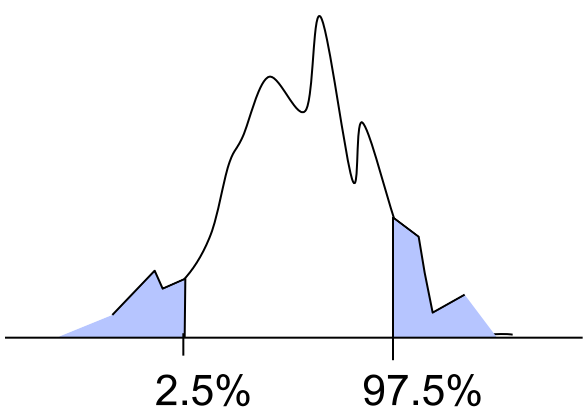
- False detection rate

- 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*

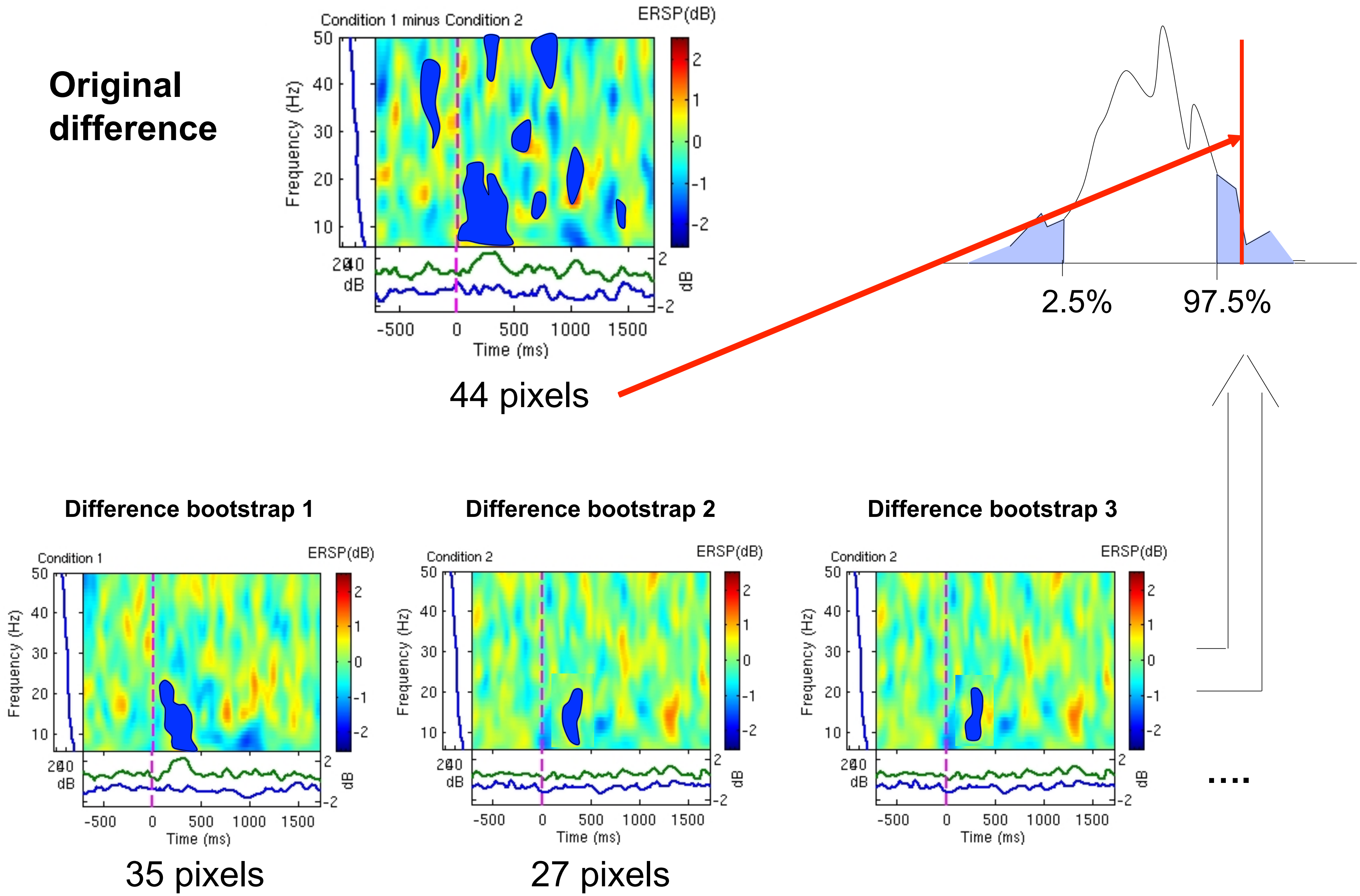
- Subtract column C1 from C2 to build column C3

- 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	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

Cluster correction for multiple comparisons



statcond function in EEGLAB

```
a = { rand(1,10) rand(1,10)+0.5 }; % pseudo 'paired' data vectors

[t df pvals] = statcond(a , 'mode', 'perm'); % perform paired t-test
pvals = 5.2807e-04 % standard t-test probability value

% Note: for different rand() outputs, results will differ.
[t df pvals surog] = statcond(a, 'mode', 'perm', 'naccu', 2000);
pvals = 0.0065 % nonparametric t-test using 2000 permuted data sets

a = { rand(2,11) rand(2,10) rand(2,12)+0.5 };
[F df pvals] = statcond(a , 'mode', 'perm'); % perform an unpaired ANOVA

pvals =
    0.00025 % p-values for difference between columns
    0.00002 % for each data row
```

statcond function in EEGLAB

```
a = { rand(3,4,10) rand(3,4,10) rand(3,4,10); ...
      rand(3,4,10) rand(3,4,10) rand(3,4,10)+0.5 };

% pseudo (2,3)-condition data array, each entry containing
% ten (3,4) data matrices
[F df pvals] = statcond(a , 'mode', 'perm');
                                    % paired 2-way ANOVA

% Output:
pvals{1} % a (3,4) matrix of p-values; effects across columns
pvals{2} % a (3,4) matrix of p-values; effects across rows
pvals{3} % a (3,4) matrix of p-values; interaction effects across
   rows and columns
```

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

Thanks to G. Rousselet