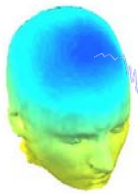
Robust Statistics

EEGLAB Workshop XXIII

AIISH, Mysuru, India

Day 2

Robust statistics



Parametric & non-parametric statistics: Use mean and standard deviation (t-test, ANOVA, …) or rank-based statistics (more robust to outliers), but

Depend on Gaussian assumption.

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

Works for any distribution.

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*

Parametric statistics

Assume gaussian distribution of data

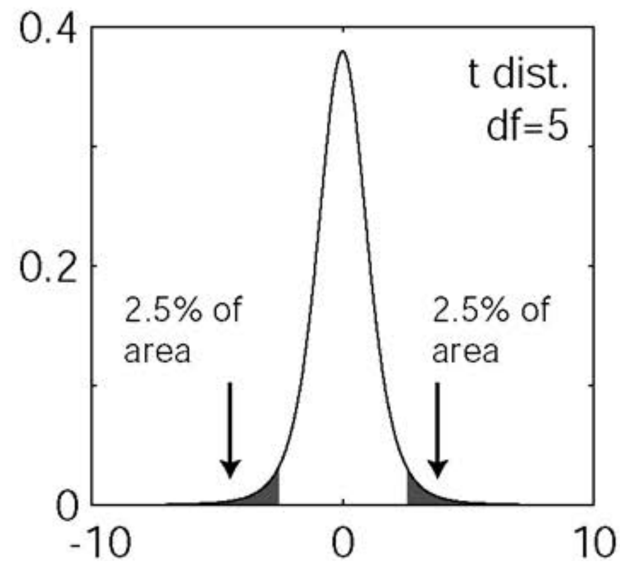
T-test: Compare paired/unpaired Samples for continuous data. In EEGLAB, used for grand-average ERPs.

Paired

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

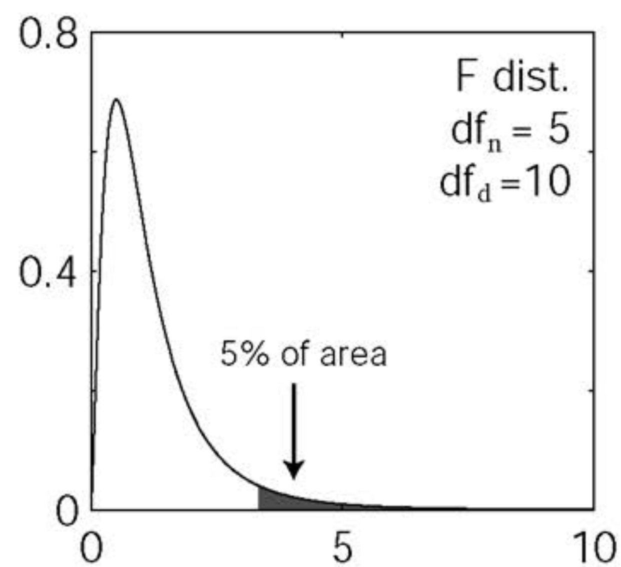
Unpaired

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



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

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



Problems

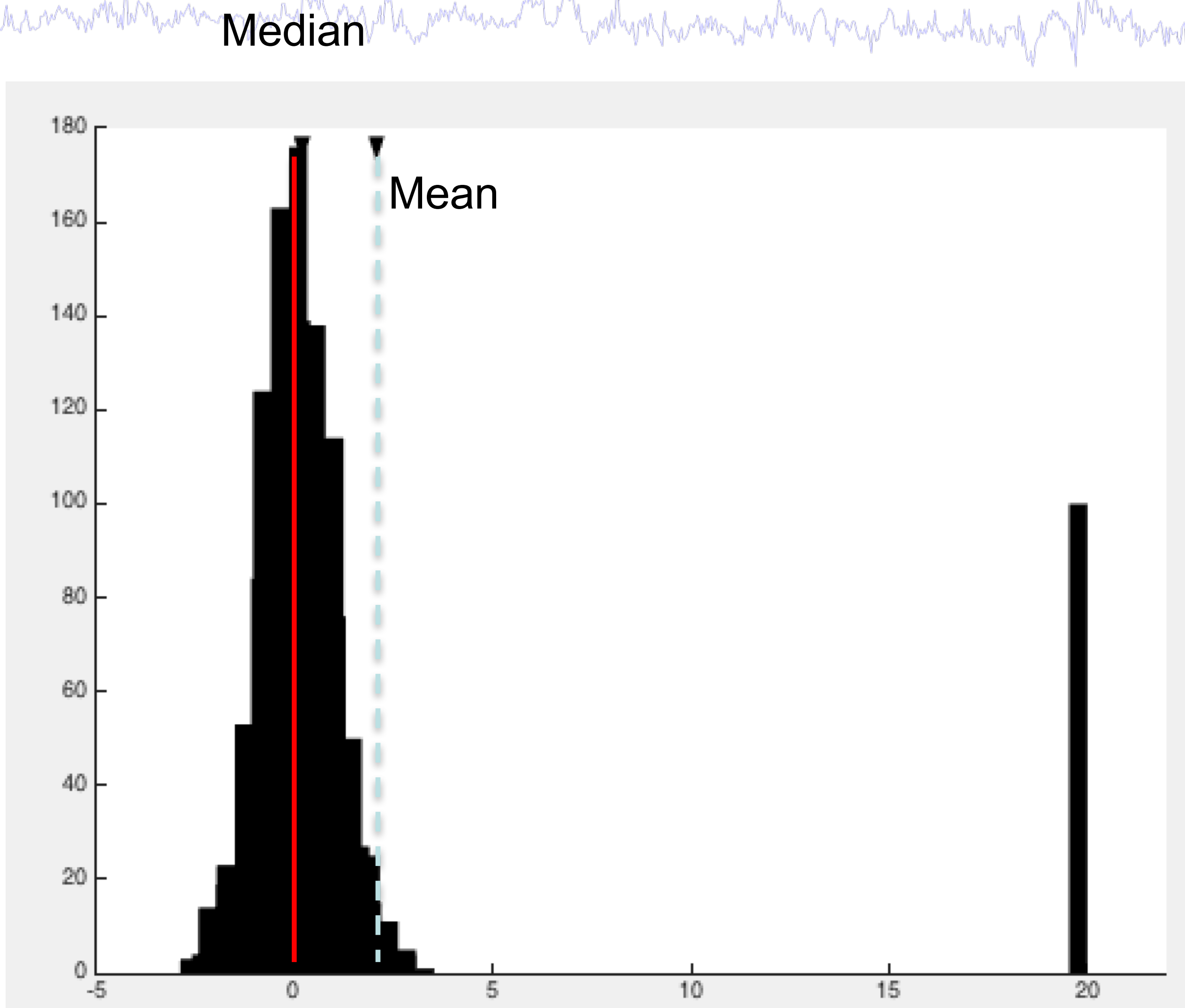
- Not resistant against outliers
- For ANOVA and t-test non-normality is an issue when distributions differ or when variances are not equal.
- Slight departure from normality can have serious consequences

---

Solutions

1. Robust Measures (outliers)
2. Bootstrap approach (non-normality)

Problem of Outliers



Robust measures of ERP

- Non-robust estimator
  – Mean: mERP = mean(EEG.data,..)

- Robust estimator
  - Median: mdERP = median(EEG.data,...)

Non-parametric statistics

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

Values | Ranks

Non-parametric is more robust to outliers

BOTH ASSUME NORMAL DISTRIBUTIONS

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

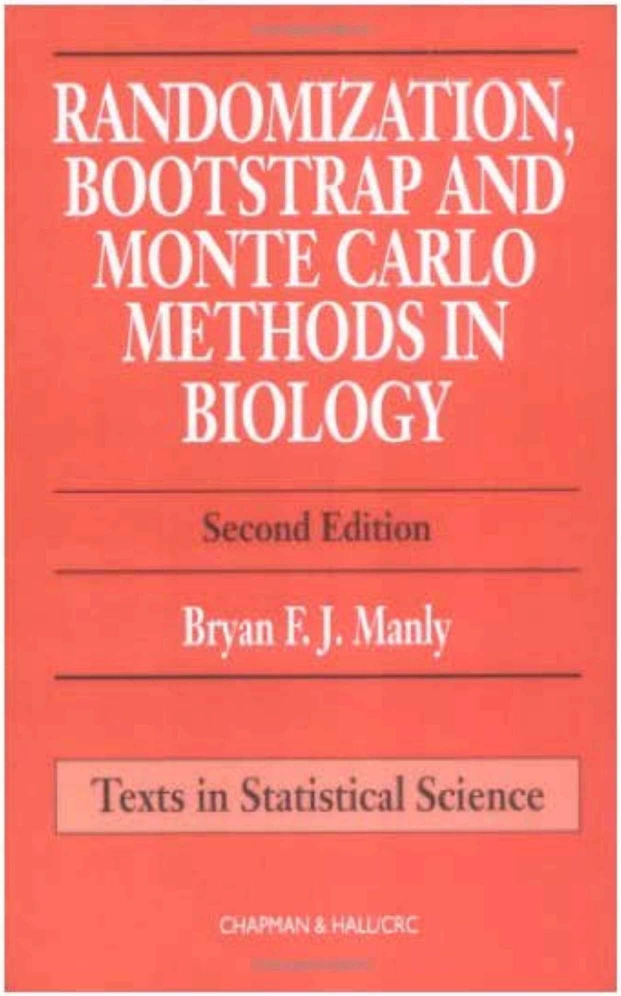
Matlab Statistics toolbox; Parra & Sajda plugin

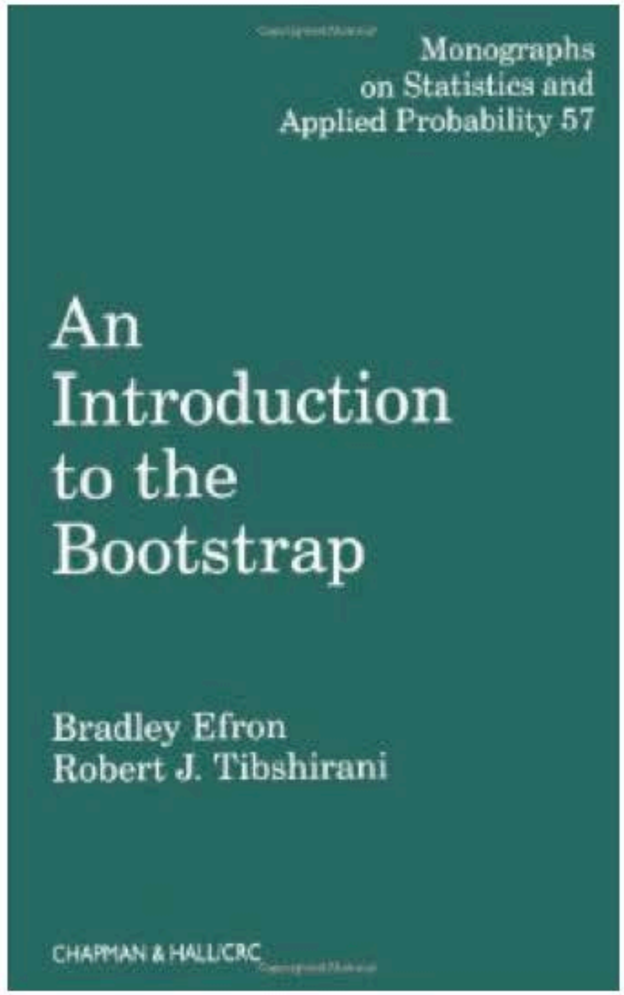
EEGLAB FIELDTRIP LIMO EEG

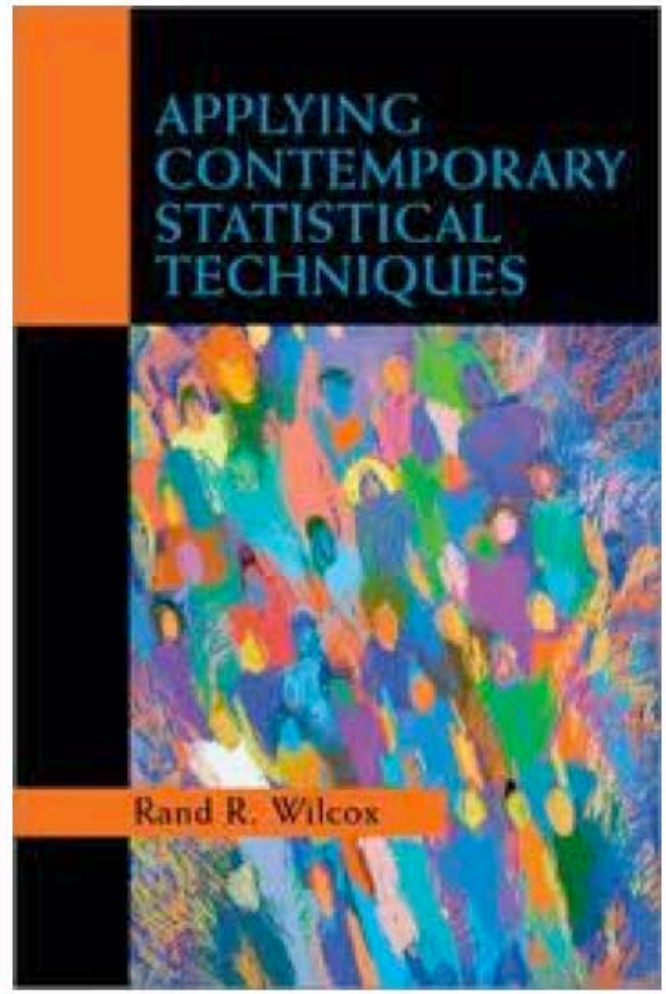
Matlab Statistics toolbox

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

How handle violations of normality?





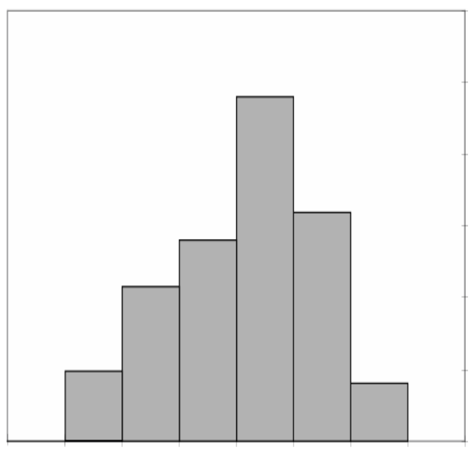


Bootstrap: central idea

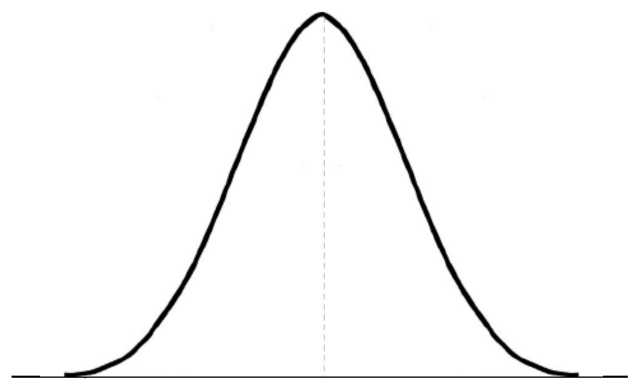
- “The bootstrap is a computer-based method for assigning measures of accuracy to statistical estimates.” Efron & Tibshirani, 1993

- “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

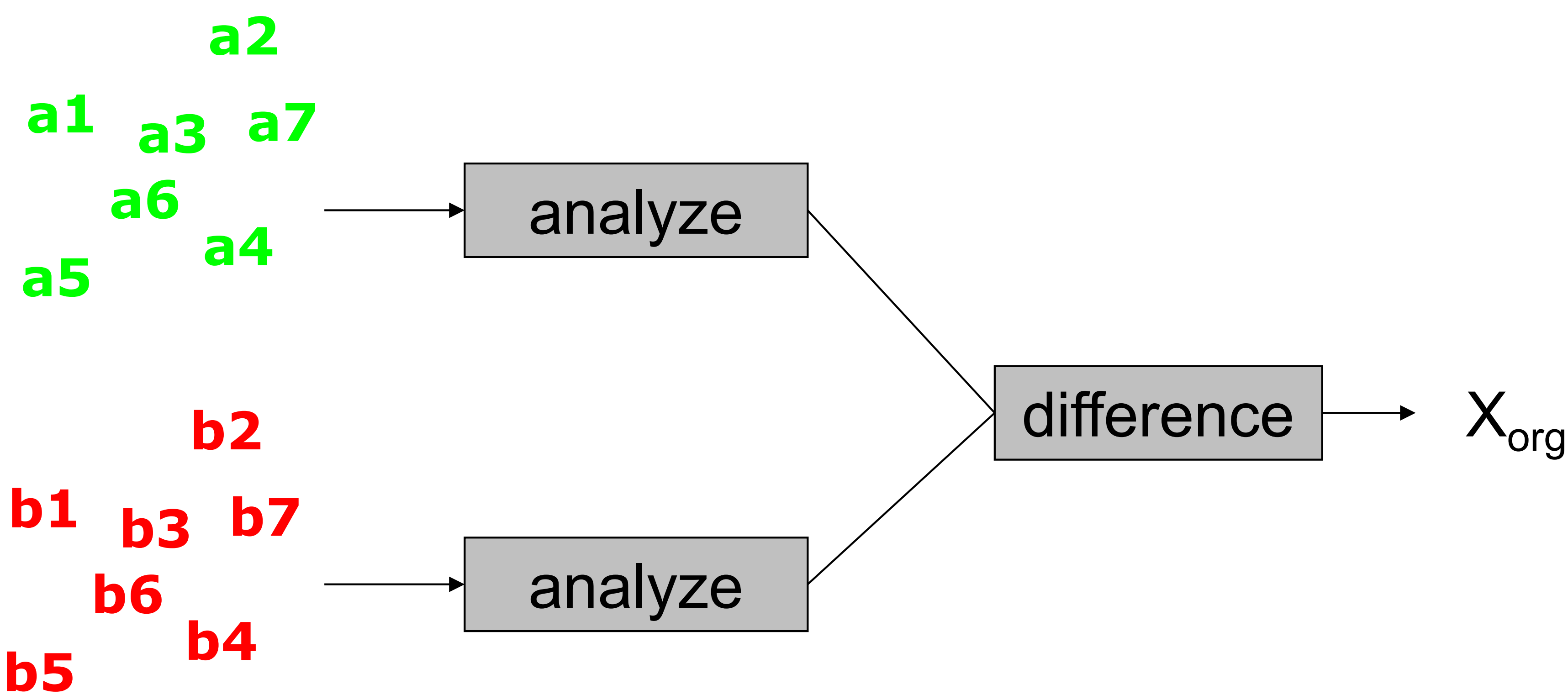


Population

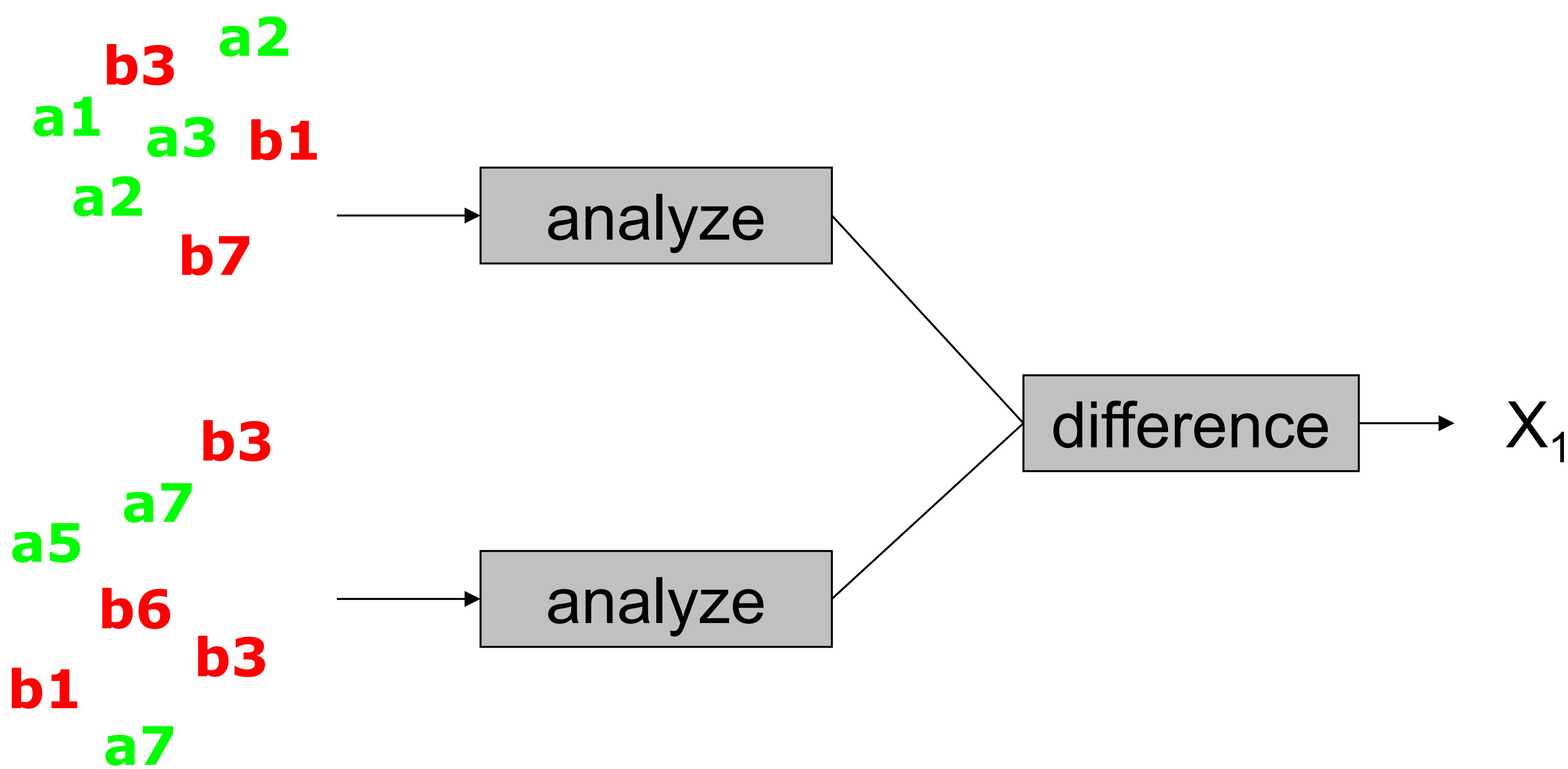
given that we have no other information about the population, the sample is our best single estimate of the population

H0: the mean is not 0 for the population

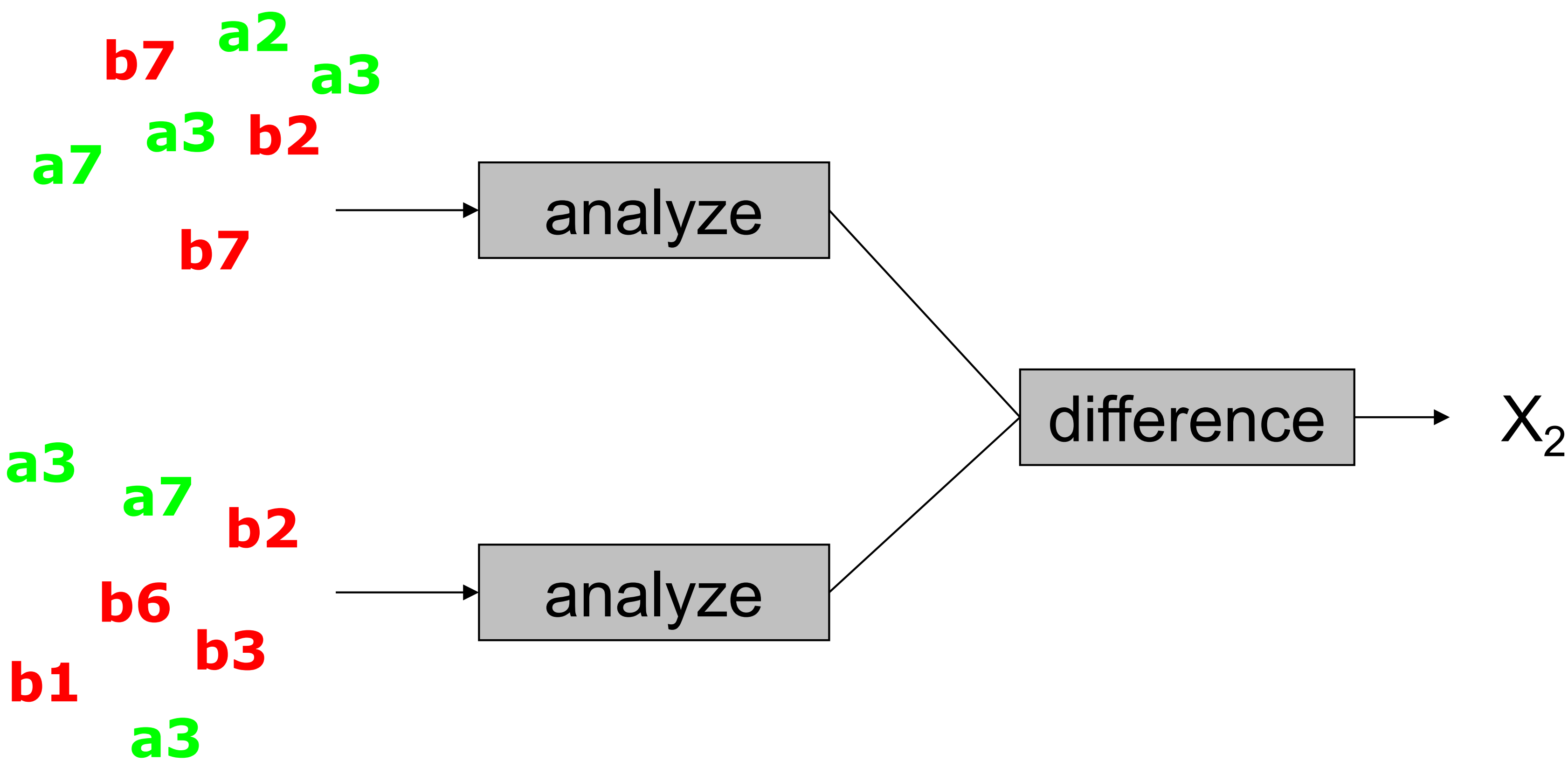
Confidence interval for the difference
Bootstrap approach



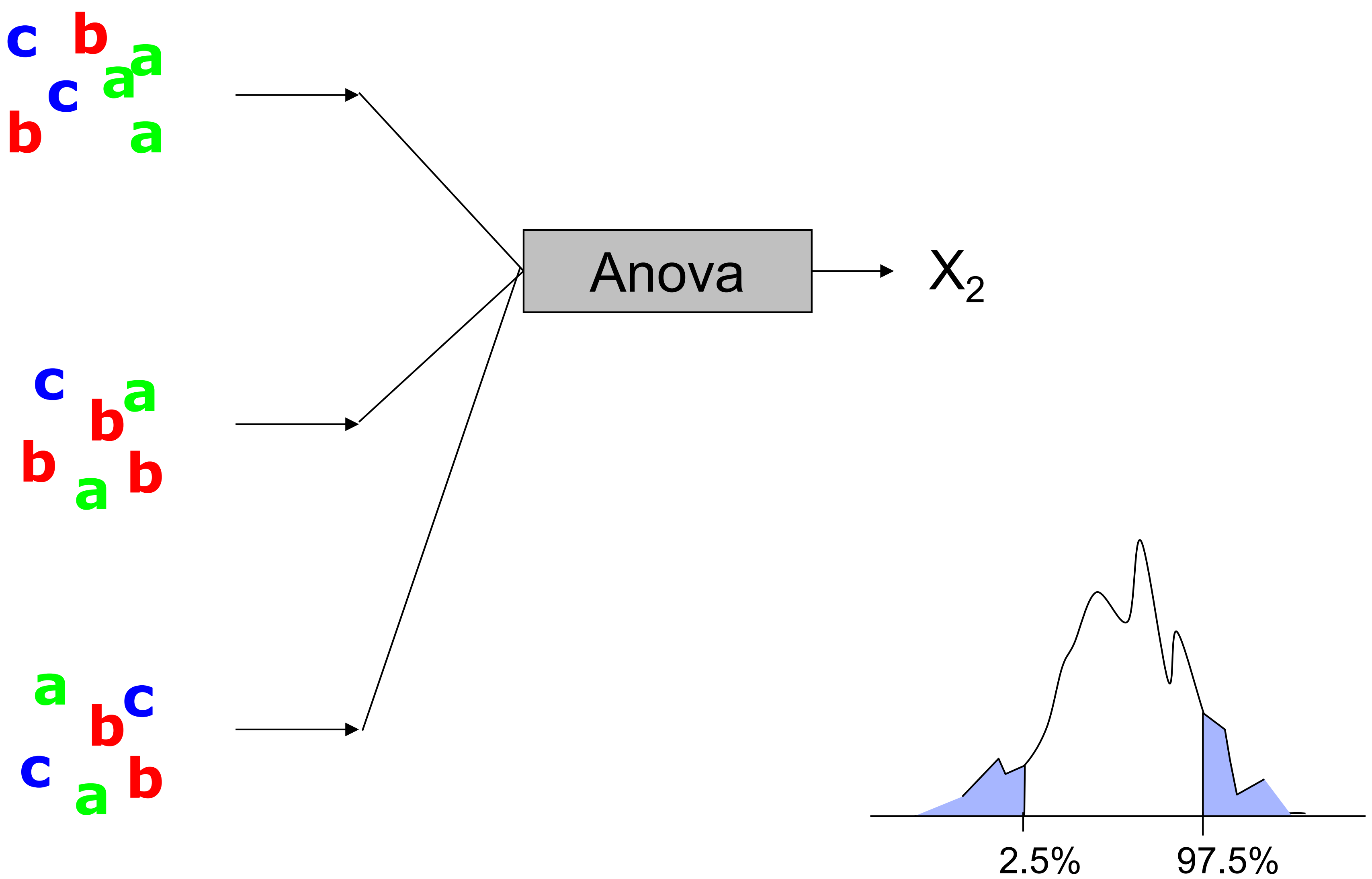
Confidence interval for the difference
Bootstrap approach



Confidence interval for the difference
Bootstrap approach



Measure for the bootstrap

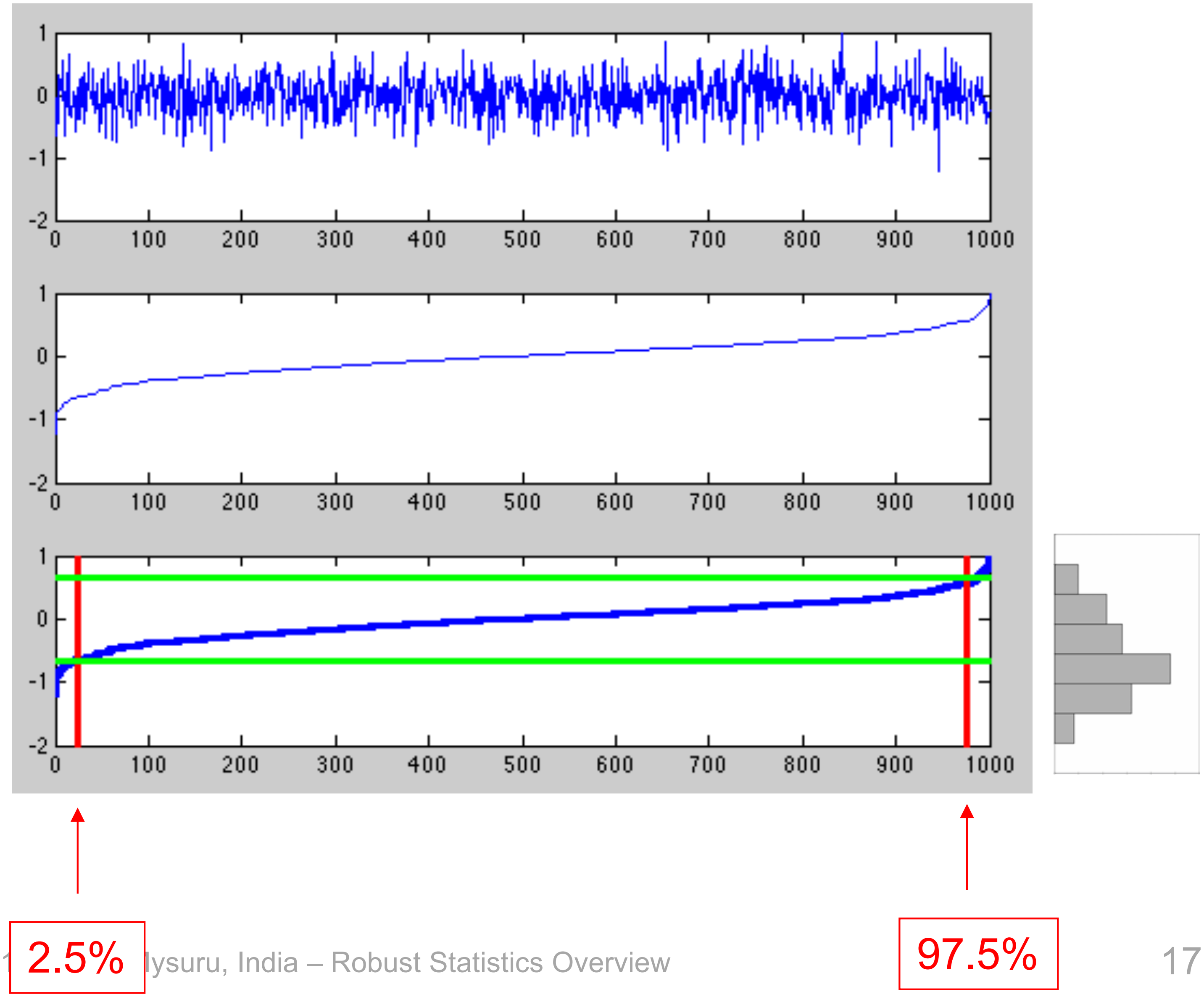


Confidence interval for the difference
Bootstrap approach

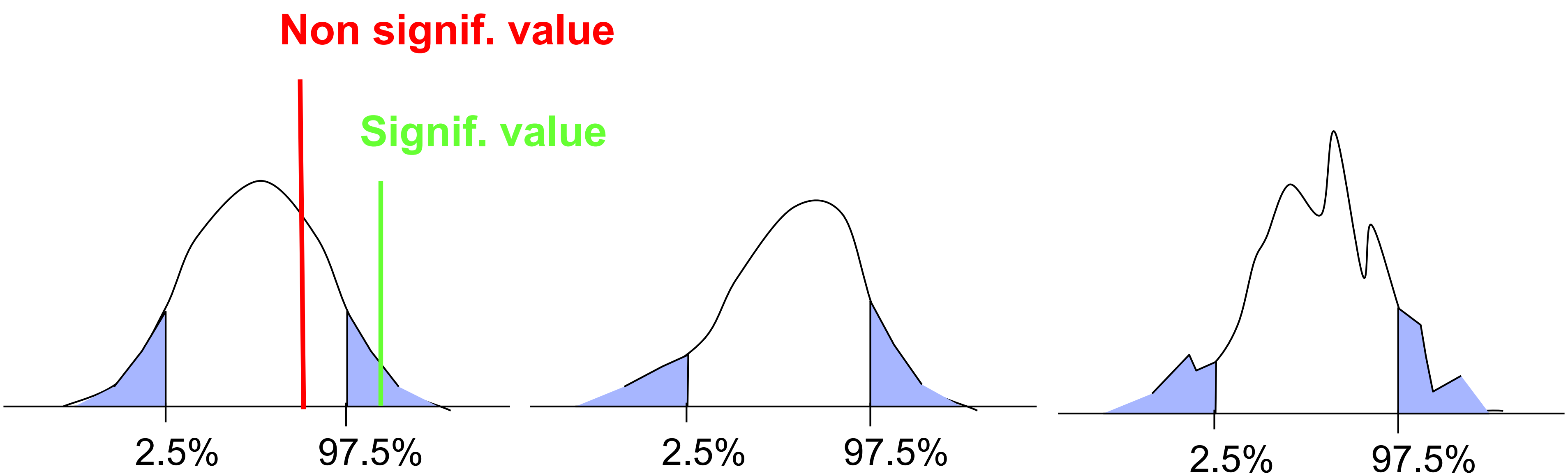
Permutation /bootstrap

Sorted values

Thresholds

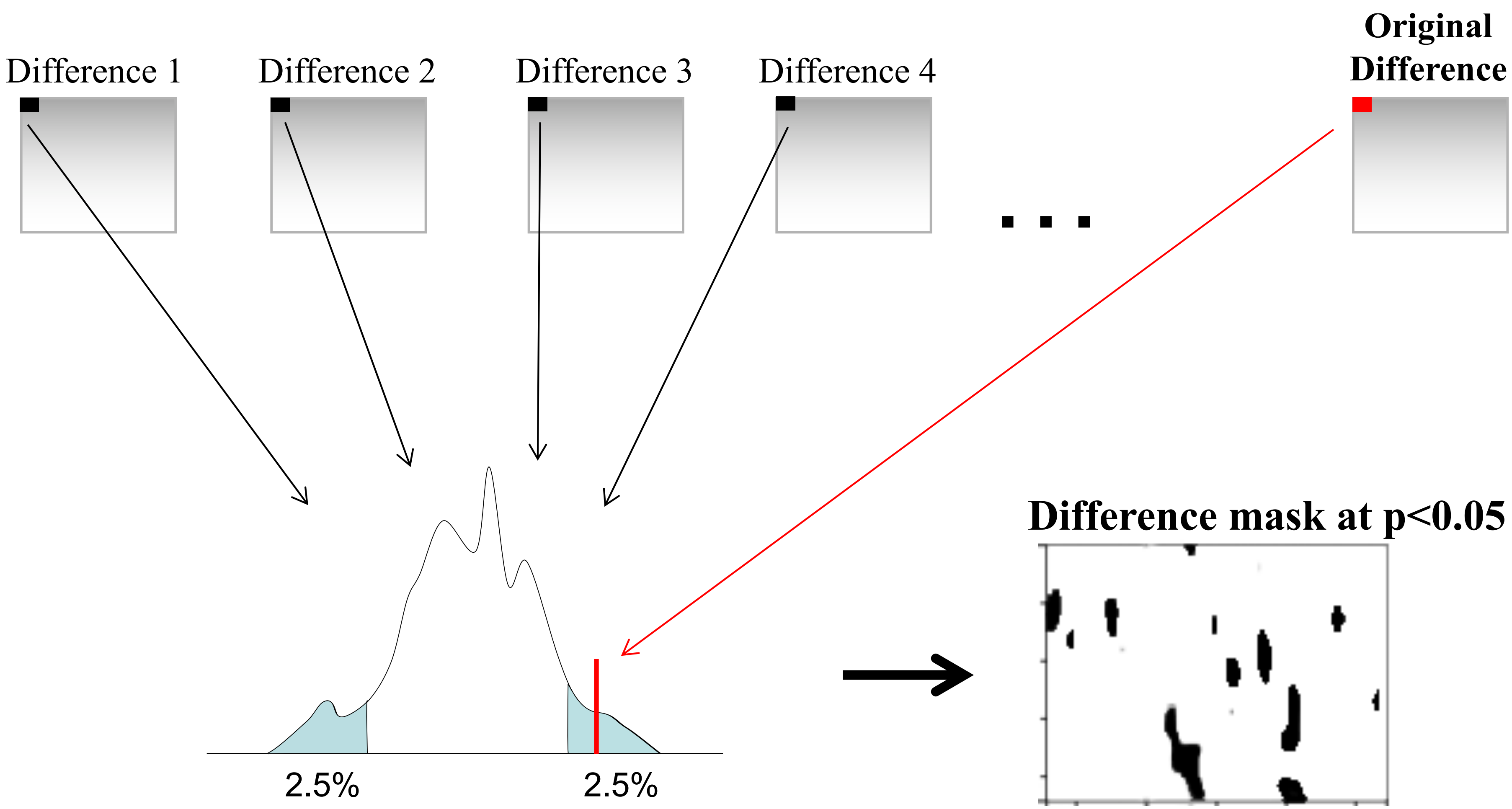


Distribution can take *any* shape



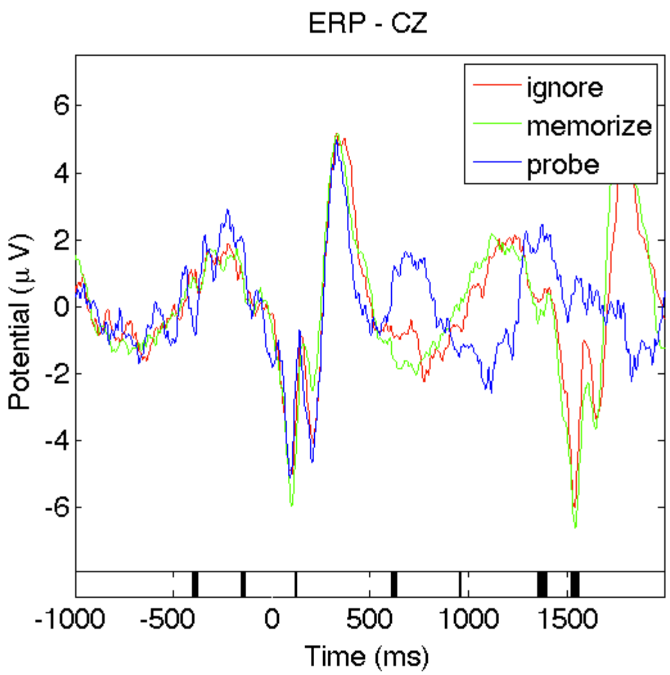
Once you have the 95% confidence interval for the difference: significance only involves assessing if 0 is included in the tails.

Assessing significance

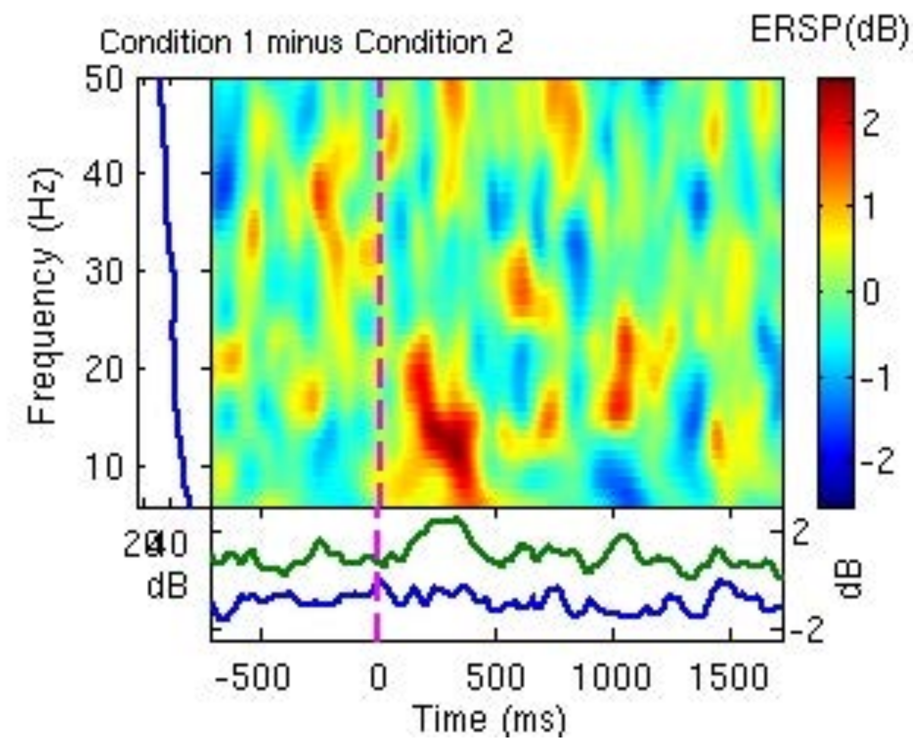


Multiple comparisons

- Problem: Comparison of ERP or ERSP across conditions involves *many* parallel statistical tests
  - ERP: e.g. 3s = 1500 points, so 1500 tests.
  - ERSP: e.g. 50 frequencies x 1000 times = 50,000 tests.



1500 tests at p=0.05: expect *75* points to be significant by chance (150 ms!)



50,000 tests at p=0.05: expect *2500* points to be significant by chance

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.)
  - Correct if every measurement is independent, but this is not the case for biological data, which has many local correlations.
  - → too conservative

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

- False detection rate (FDR)

- Cluster randomization

FDR

1. For a given α, find the largest k such that $P_{(k)} \leq \dfrac{k}{m}\alpha$.
2. Reject the null hypothesis (i.e., declare discoveries) for all $H_{(i)}$ for $i = 1, \ldots, k$.

FDR procedure

1. For a given α, find the largest k such that $P_{(k)} \leq \frac{k}{m}\alpha$.
2. Reject the null hypothesis (i.e., declare discoveries) for all $H_{(i)}$ for $i = 1, \ldots, k$.

Procedure:

- Sort all p values (column C1)
C3

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

- Subtract column C1 from C2 to build column C3

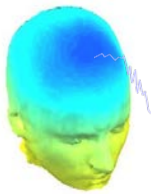
- Find the highest negative value in C3 and find the corresponding p-value in C1 (*p_fdr*)

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

C1 (P_k)

Index "k"	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



1. For a given α, find the largest k such that $P_{(k)} \leq \frac{k}{m}\alpha$.
2. Reject the null hypothesis (i.e., declare discoveries) for all $H_{(i)}$ for $i = 1, \ldots, k$.

Procedure:

- Sort all p values (column C1)

C3

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

- Subtract column C1 from C2 to build column C3

- Find the highest negative value 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 (P_k)	C2
Index "k"	Actual	k*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

1. For a given α, find the largest k such that $P_{(k)} \leq \frac{k}{m}\alpha$.
2. Reject the null hypothesis (i.e., declare discoveries) for all $H_{(i)}$ for $i = 1, \ldots, k$.

Procedure:

- Sort all p values (column C1)

C3

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

- Subtract column C1 from C2 to build column C3

- Find the highest negative value 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 (P_k)	C2	C3
Index "k"	Actual	k*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

1. For a given α, find the largest k such that $P_{(k)} \leq \frac{k}{m}\alpha$.
2. Reject the null hypothesis (i.e., declare discoveries) for all $H_{(i)}$ for $i = 1, \ldots, k$.

Procedure:

- Sort all p values (column C1)
C3

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

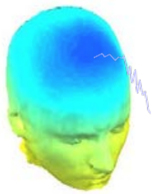
- Subtract column C1 from C2 to build column C3

- Find the highest negative value 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 (P_k)	C2	C3
Index "k"	Actual	k*0.05/10	C2-C1
1	0.001	0.005	-0.004
2	0.002	0.01	-0.008
3	$\boxed{0.01}$	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



1. For a given α, find the largest k such that $P_{(k)} \leq \frac{k}{m}\alpha$.
2. Reject the null hypothesis (i.e., declare discoveries) for all $H_{(i)}$ for $i = 1, \ldots, k$.

Procedure:

- Sort all p values (column C1)
C3

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

- Subtract column C1 from C2 to build column C3

- Find the highest negative value 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 (P_k)	C2	C3
Index "k"	Actual	k*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

Comparison of different corrections

Bonferoni

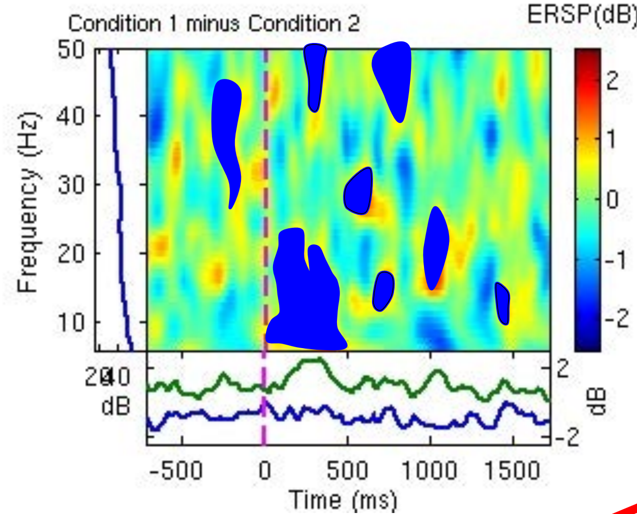
FDR

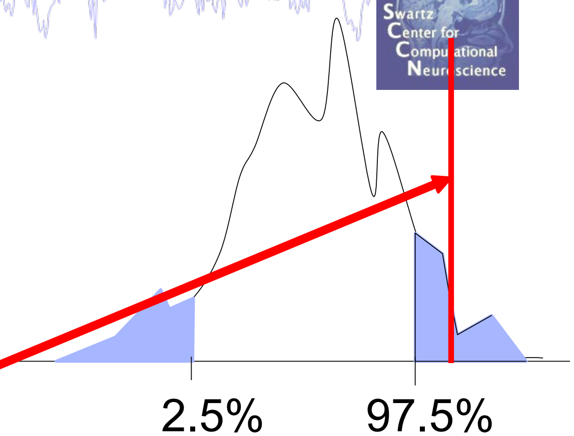
Uncorrected

	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

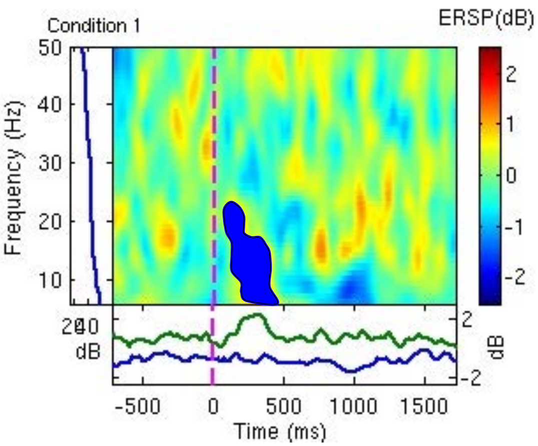
Original difference





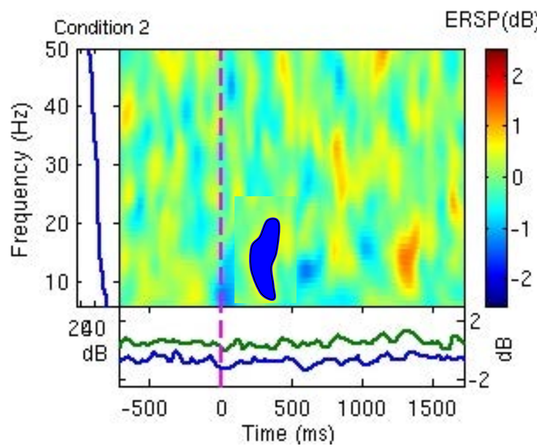
Size of largest sig. cluster: 44 pixels

Difference bootstrap 1



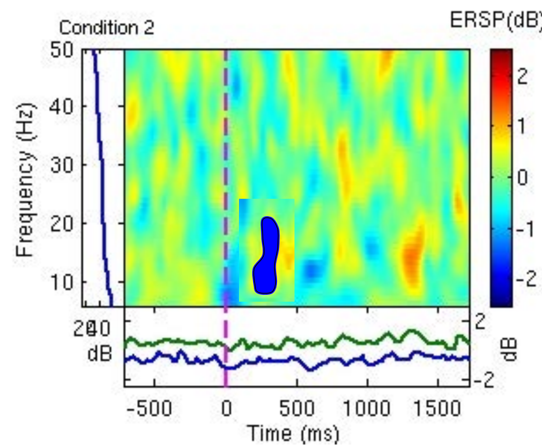
35 pixels

Difference bootstrap 2



27 pixels

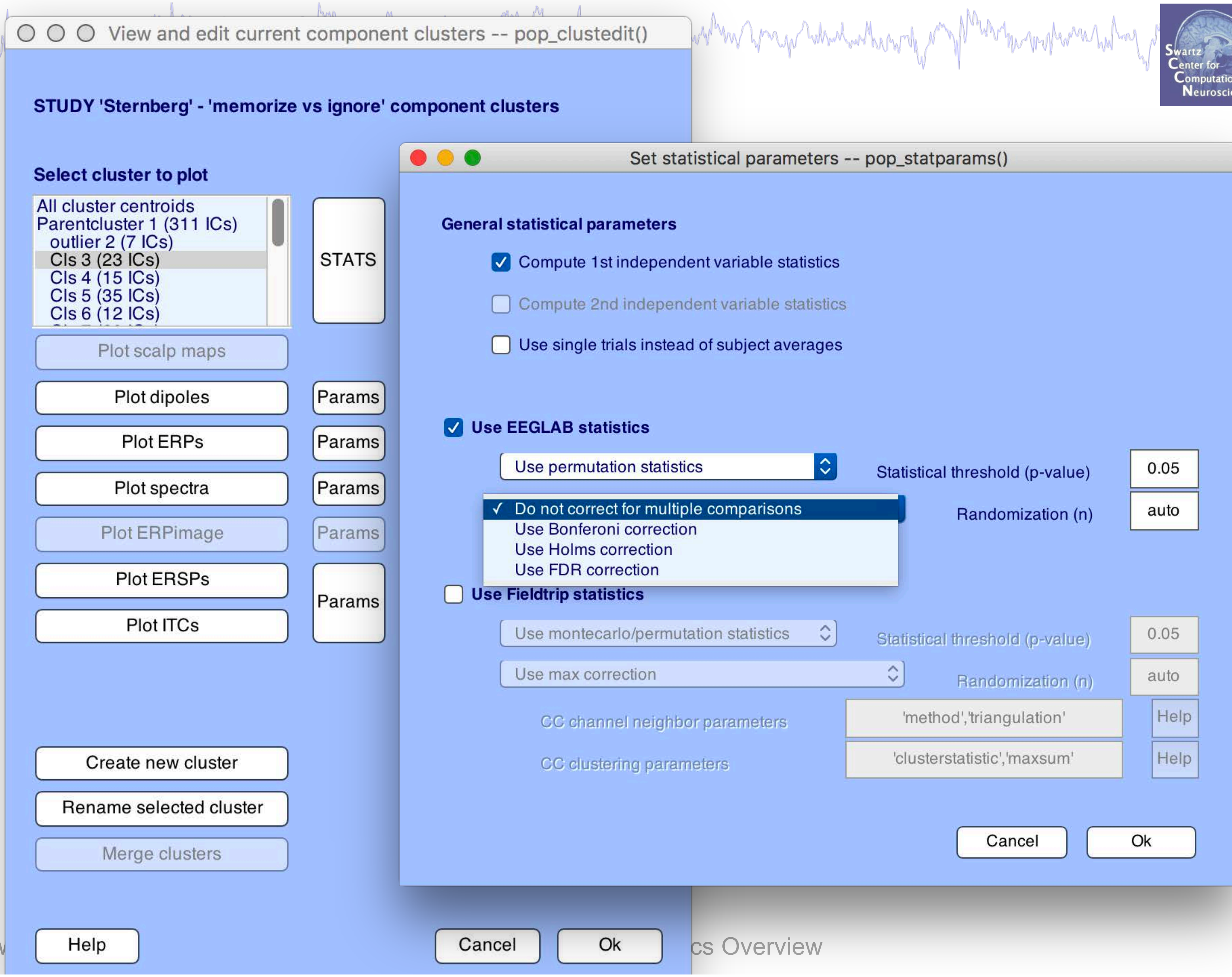
Difference bootstrap 3



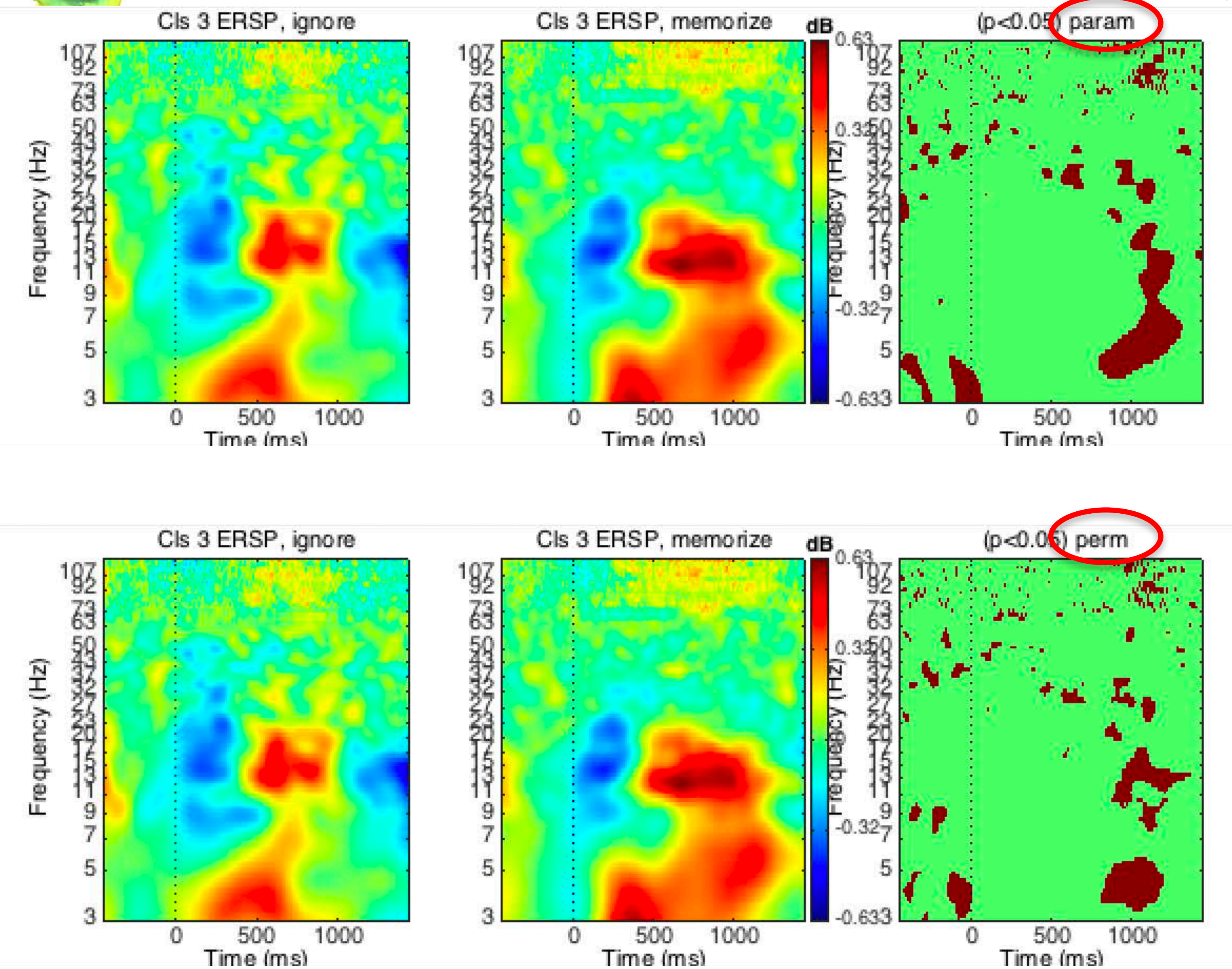
22 pixels

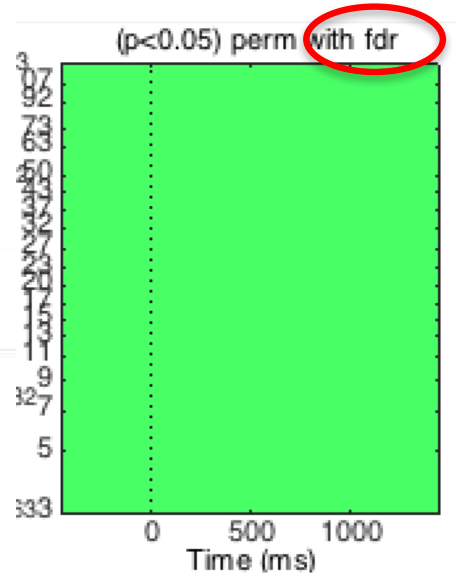
....

Study GUI

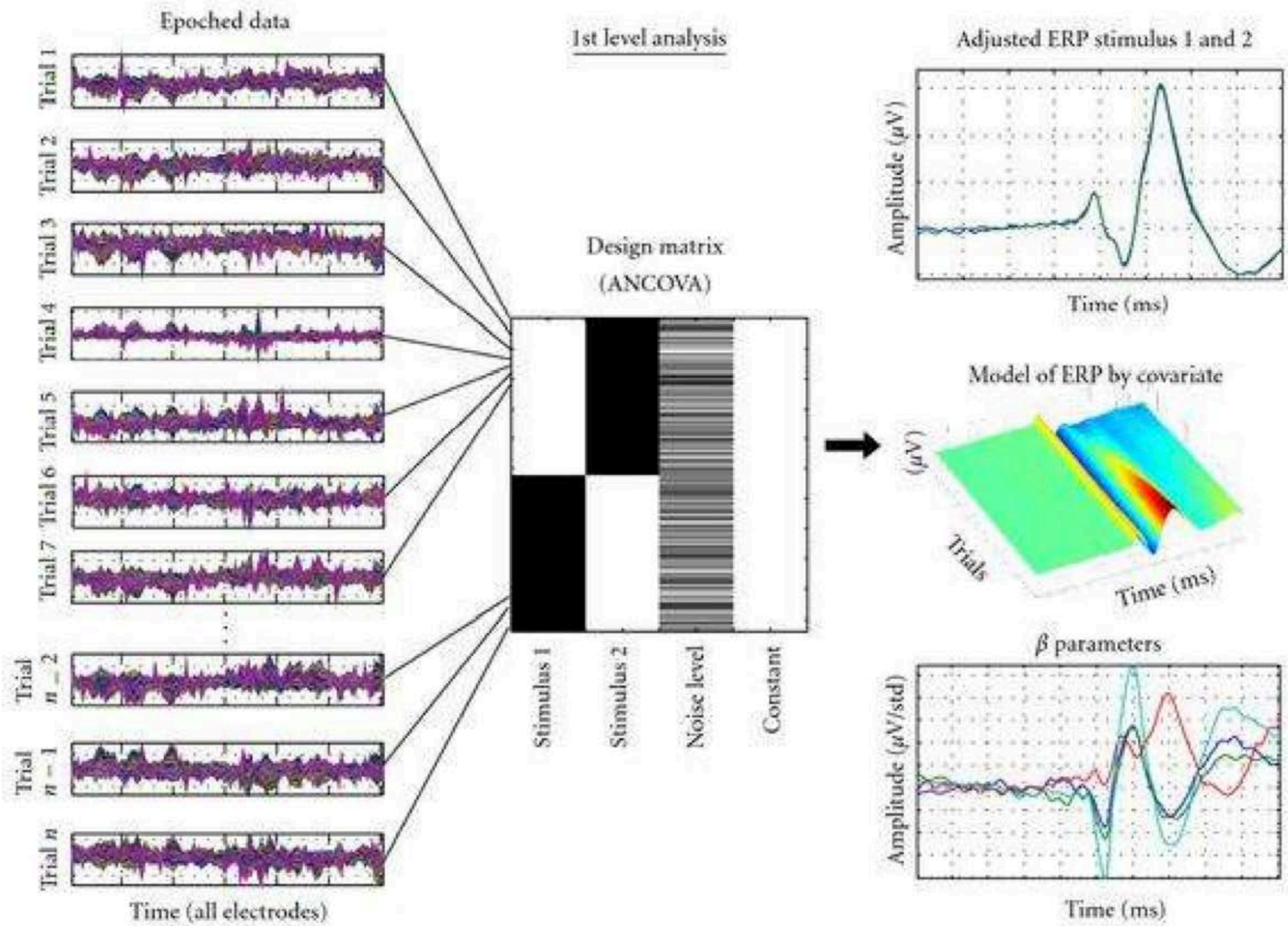

View and edit current component clusters -- pop_clustedit()
STUDY 'Sternberg' - 'memorize vs ignore' component clusters
Select cluster to plot
All cluster centroids
Parentcluster 1 (311 ICs)
outlier 2 (7 ICs)
Cls 3 (23 ICs)
Cls 4 (15 ICs)
Cls 5 (35 ICs)
Cls 6 (12 ICs)
STATS
Plot scalp maps
Plot dipoles
Params
Plot ERPs
Params
Plot spectra
Params
Plot ERPimage
Params
Plot ERSPs
Plot ITCs
Params
Create new cluster
Rename selected cluster
Merge clusters
Help
Cancel
Ok
Set statistical parameters -- pop_statparams()
General statistical parameters
Compute 1st independent variable statistics
Compute 2nd independent variable statistics
Use single trials instead of subject averages
Use EEGLAB statistics
Use permutation statistics
Statistical threshold (p-value)
0.05
Do not correct for multiple comparisons
Use Bonferoni correction
Use Holms correction
Use FDR correction
Randomization (n)
auto
Use Fieldtrip statistics
Use montecarlo/permutation statistics
Statistical threshold (p-value)
0.05
Use max correction
Randomization (n)
auto
CC channel neighbor parameters
'method','triangulation'
Help
CC clustering parameters
'clusterstatistic','maxsum'
Help
Cancel
Ok

Test between conditions (stern study)





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