Contemporary Statistical Methods Useful for EEG Analysis

David Groppe
Marta Kutas’s lab
University of California, San Diego

Arnaud Delorme
Swartz Center for Computational Neuroscience
University of California, San Diego

12th EEGLAB Workshop
Nov. 19, 2010
Presentation Outline

• “Classic” Analytical Inferential Statistics
  - Parametric & non-parametric

• Resampling-Based Inferential Statistics
  - Randomization/permutation tests
  - Bootstrap statistics

• Correcting for Multiple Comparisons
  - Permutation test based control of family-wise error
  - Benjamini methods for control of false discovery rate
  - Evaluating multiple comparison correction on simulated ERP data
Analytic Parametric Statistics:
Assume Data Come from a Particular Distribution

Gaussian Distribution

$\frac{1}{\sigma \sqrt{2\pi}}$
Analytic Parametric Statistics:

Critical Values Analytically Derived

Table B^a
CRITICAL VALUES OF t

<table>
<thead>
<tr>
<th>df</th>
<th>.05*</th>
<th>.01**</th>
<th>.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12.706</td>
<td>63.657</td>
<td>636.62</td>
</tr>
<tr>
<td>2</td>
<td>4.303</td>
<td>9.925</td>
<td>31.598</td>
</tr>
<tr>
<td>3</td>
<td>3.182</td>
<td>5.841</td>
<td>12.924</td>
</tr>
<tr>
<td>4</td>
<td>2.776</td>
<td>4.604</td>
<td>8.610</td>
</tr>
<tr>
<td>5</td>
<td>2.571</td>
<td>4.032</td>
<td>6.869</td>
</tr>
<tr>
<td>6</td>
<td>2.447</td>
<td>3.707</td>
<td>5.959</td>
</tr>
<tr>
<td>7</td>
<td>2.365</td>
<td>3.499</td>
<td>5.408</td>
</tr>
<tr>
<td>8</td>
<td>2.306</td>
<td>3.355</td>
<td>5.041</td>
</tr>
<tr>
<td>9</td>
<td>2.262</td>
<td>3.256</td>
<td>4.781</td>
</tr>
</tbody>
</table>

Two-tailed or Nondirectional Test
LEVEL OF SIGNIFICANCE
(p-value in color)

<table>
<thead>
<tr>
<th>df</th>
<th>.5</th>
<th>.05</th>
<th>.01</th>
<th>.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.314</td>
<td>31.821</td>
<td>318.31</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2.920</td>
<td>6.965</td>
<td>22.326</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2.353</td>
<td>4.541</td>
<td>10.213</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2.132</td>
<td>3.747</td>
<td>7.173</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2.015</td>
<td>3.365</td>
<td>5.893</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1.943</td>
<td>3.143</td>
<td>5.208</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1.895</td>
<td>2.998</td>
<td>4.785</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1.860</td>
<td>2.896</td>
<td>4.501</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1.833</td>
<td>2.797</td>
<td>4.292</td>
<td></td>
</tr>
</tbody>
</table>
Analytic Parametric Statistics:

**Popular Parametric Tests**

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

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

**Paired**

$t = \frac{\text{Mean}_{\text{difference}}}{\text{Standard\_deviation}} \sqrt{N - 1}$

**Unpaired**

$t = \sqrt{N} \frac{\text{Mean}_A - \text{Mean}_B}{\sqrt{(SD_A)^2 - (SD_B)^2}}$

**F test**

$F = \frac{\text{Variance}_{\text{interGroup}}}{N_{\text{Group}} - 1} \frac{N_{\text{Group}}}{\text{Variance}_{\text{WithinGroup}}} \frac{N - N_{\text{Group}}}{N}$
Analytic Non-Parametric Statistics:

Minimal Distribution Assumptions

Mann-Whitney U Test: Null hypothesis is that the distribution of Population A and B are the same.
Analytic Non-Parametric Statistics:

<table>
<thead>
<tr>
<th>Parametric</th>
<th>Non-Parametric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paired t-test</td>
<td>Wilcoxon</td>
</tr>
<tr>
<td>Unpaired t-test</td>
<td>Mann-Whitney</td>
</tr>
<tr>
<td>One way ANOVA</td>
<td>Kruskal Wallis</td>
</tr>
<tr>
<td>Values</td>
<td>Ranks</td>
</tr>
</tbody>
</table>
Problems with Analytic Statistics:

1. No analytic solution for some situations (e.g., comparing the mean of two groups that differ in variance)

2. Often, data don’t fit parametric assumptions

3. Non-parametric tests may lack power and rank transformation can make it tricky to do things like derive confidence intervals
Presentation Outline

• “Classic” Analytical Inferential Statistics
  - Parametric & non-parametric

• Resampling-Based Inferential Statistics
  - Randomization/permutation tests
  - Bootstrap statistics

• Correcting for Multiple Comparisons
  - Permutation test based control of family-wise error
  - Benjamini methods for control of false discovery rate
  - Evaluating multiple comparison correction on simulated ERP data
Resampling-Based Statistics:

Inferential statistics based on “simulating” an experiment a large number of times with the observed data

**Observed Data**

<table>
<thead>
<tr>
<th>Group A</th>
<th>Group B</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>
Resampling-Based Statistics:

Inferential statistics based on “simulating” an experiment a large number of times with the observed data.

<table>
<thead>
<tr>
<th>Observed Data</th>
<th>“Simulated Replication”</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group A</strong></td>
<td><strong>Group B</strong></td>
</tr>
<tr>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>

Resample
Resampling-Based Statistics:

Inferential statistics based on “simulating” an experiment a large number of times with the observed data.

<table>
<thead>
<tr>
<th>Observed Data</th>
<th>“Simulated Replication”</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group A</strong></td>
<td><strong>Group B</strong></td>
</tr>
<tr>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>

Resample
Resampling-Based Statistics:

Inferential statistics based on “simulating” an experiment a large number of times with the observed data.
Resampling-Based Statistics: Two Popular Resampling Methods

1. Permutation Tests (also called “Randomization Tests”)

2. Bootstrap Statistics
Advantages of Permutation Tests & Bootstrap Statistics

1. Non-parametric (i.e., make minimal assumptions about population distributions)

2. Can be used in situations for which there is no analytic solution

3. Simple to use and easily provide confidence intervals

4. Useful for multiple comparison correction
Resampling-Based Statistics: Two Popular Resampling Methods

1. Permutation Tests (also called “Randomization Tests”)

2. Bootstrap Statistics
Permutation Tests

1. Old idea (Neyman, 1923; Fisher, 1935) but too computationally intensive to be widely used until relatively recently

2. Test the null hypothesis that the observations in multiple groups of data are exchangeable (i.e., they were just as likely to occur in one condition/group as any other)
Hypothetical Experiment #1

• Two conditions: A & B
• Within-subject design
• Three subjects

Observed Data from: Blair & Karniski (1993) *Psychophysiology*

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>t value</td>
<td></td>
<td>3.46</td>
</tr>
</tbody>
</table>
Null Hypothesis

- Observations in Condition A could have just as likely come from Condition B (and vice-versa).
- Each possible permutation of observations equally likely.

<table>
<thead>
<tr>
<th></th>
<th>Sub1</th>
<th>Sub2</th>
<th>Sub3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>orig</td>
<td>orig</td>
<td>orig</td>
</tr>
<tr>
<td></td>
<td>orig</td>
<td>flip</td>
<td>orig</td>
</tr>
<tr>
<td></td>
<td>flip</td>
<td>orig</td>
<td>flip</td>
</tr>
<tr>
<td></td>
<td>orig</td>
<td>flip</td>
<td>orig</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8 5</td>
<td>8 5</td>
<td>8 5</td>
</tr>
<tr>
<td></td>
<td>4 3</td>
<td>3 4</td>
<td>4 6</td>
</tr>
<tr>
<td></td>
<td>6 4</td>
<td>4 6</td>
<td>6 4</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.46</td>
<td>1.11</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>0.0</td>
<td>-1.11</td>
</tr>
<tr>
<td></td>
<td>-0.46</td>
<td>-0.46</td>
<td>-3.46</td>
</tr>
</tbody>
</table>

2^n possible permutations
### Null Hypothesis

**Observed Data**

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>

$t\text{ value} = 3.46$

**Remaining Possible Permutations**

<table>
<thead>
<tr>
<th></th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>B</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>A</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>B</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

$t\text{ dist, df}=5$

2.5% of area

2.5% of area
### Null Hypothesis

<table>
<thead>
<tr>
<th>Permutation</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>-3.46</td>
<td>-1.11</td>
<td>-0.46</td>
<td>0</td>
<td>0</td>
<td>0.46</td>
<td>1.11</td>
<td>3.46</td>
</tr>
</tbody>
</table>
**Null Hypothesis**

<table>
<thead>
<tr>
<th>Permutation</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t$</td>
<td>-3.46</td>
<td>-1.11</td>
<td>-0.46</td>
<td>0</td>
<td>0</td>
<td>0.46</td>
<td>1.11</td>
<td>3.46</td>
</tr>
</tbody>
</table>

**Decision Rule:** If observed difference is the most positive permutation, reject null hypothesis (upper tailed test).

$\alpha = \frac{1}{8} = 0.125$
<table>
<thead>
<tr>
<th>Permutation</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t$</td>
<td>-3.46</td>
<td>-1.11</td>
<td>-0.46</td>
<td>0</td>
<td>0</td>
<td>0.46</td>
<td>1.11</td>
<td>3.46</td>
</tr>
</tbody>
</table>

**Null Hypothesis**

Decision Rule: If observed difference is the most positive or negative, reject null hypothesis (two tailed test).

$\alpha = \frac{2}{8} = 0.25$
Hypothetical Experiment #2

- Two conditions: A & B
- Within-subject design
- 25 subjects

\[2^{25} \text{ (i.e., 33,554,432) permutations}\]

Approximate distribution of null hypothesis with thousands of random permutations.
### Hypothetical Experiment #3
- Two groups: A & B
- Between-subject design
- 3 “A” subjects, 2 “B” subjects

<table>
<thead>
<tr>
<th>Group</th>
<th>Observed Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5</td>
</tr>
<tr>
<td>A</td>
<td>18</td>
</tr>
<tr>
<td>A</td>
<td>-23</td>
</tr>
<tr>
<td>B</td>
<td>9</td>
</tr>
<tr>
<td>B</td>
<td>3</td>
</tr>
</tbody>
</table>
# Null Hypothesis

- Observations in Group A could have just as likely come from Group B (and vice-versa).
- Each possible permutation of observations equally likely.

<table>
<thead>
<tr>
<th>Group</th>
<th>Observed Data</th>
<th>Perm 2</th>
<th>Perm 3</th>
<th>Perm 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>A</td>
<td>18</td>
<td>9</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>A</td>
<td>-23</td>
<td>3</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>B</td>
<td>9</td>
<td>-23</td>
<td>-23</td>
<td>-23</td>
</tr>
<tr>
<td>B</td>
<td>3</td>
<td>18</td>
<td>3</td>
<td>9</td>
</tr>
</tbody>
</table>

Possible Permutations:

\[
\binom{5}{3} = \frac{5!}{3!(5-2)!} = 10
\]
Resampling-Based Statistics: Two Popular Resampling Methods

1. Permutation Tests (also called “Randomization Tests”)

2. Bootstrap Statistics
Sample and Population

What we observed

What we sampled from

Sample

Population

Bootstrap Statistics: Treat the sample as if it is the population
Hypothetical Experiment #4

- Two conditions: A & B
- Within-subject design
- Three subjects

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>A-B</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td></td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

Mean Difference: 2
**Hypothetical Experiment #4**

- Two conditions: A & B
- Within-subject design
- Three subjects

---

<table>
<thead>
<tr>
<th>A-B</th>
<th>A-B*</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

Make a “bootstrap” sample by randomly selecting one of the difference values three times.

**Mean Difference:** 2
Hypothetical Experiment #4

- Two conditions: A & B
- Within-subject design
- Three subjects

Mean Difference:

<table>
<thead>
<tr>
<th>A-B</th>
<th>A-B*</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Make a “bootstrap” sample by randomly selecting one of the difference values three times.
Bootstrap versus Permutation

**Permutation**
- Each data point gets picked exactly once

**Bootstrap**
- Each data point can be picked zero, one, or multiple times
**Hypothetical Experiment #4**

- Two conditions: A & B
- Within-subject design
- Three subjects

<table>
<thead>
<tr>
<th>A-B</th>
<th>A-B*</th>
<th>A-B*</th>
<th>A-B*</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Mean Difference: 2 2.7 1.7 2.3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Distribution of Mean of 10,000 Bootstrap Samples
Distribution of Mean of 10,000 Bootstrap Samples

“Percentile Bootstrap” Confidence Intervals
Presentation Outline

• "Classic" Analytical Inferential Statistics
  - Parametric & non-parametric

• Resampling-Based Inferential Statistics
  - Randomization/permutation tests
  - Bootstrap statistics

• Correcting for Multiple Comparisons
  - Permutation test based control of family-wise error
  - Benjamini methods for control of false discovery rate
  - Evaluating multiple comparison correction on simulated ERP data

Summary:
Advantages of Permutation Tests & Bootstrap Statistics

1. Non-parametric (i.e., make minimal assumptions about population distributions)
2. Can be used in situations for which there is no analytic solution
3. Simple to use and easily provide confidence intervals
4. Useful for multiple comparison correction

Coming up next!
Disadvantages of Permutation Tests & Bootstrap Statistics

1. Poor performance with small sample sizes
   • Might be inaccurate

What we observed  What we sampled from

Sample  Population
<table>
<thead>
<tr>
<th>Simple Analyses (e.g., t-tests, correlation)</th>
<th>Permutation</th>
<th>Bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complex Analyses (e.g., multifactor ANOVAS)</td>
<td>✓ Always Accurate</td>
<td>Asymptotically Accurate</td>
</tr>
<tr>
<td></td>
<td>Asymptotically Accurate or Not Applicable</td>
<td>✓ Asymptotically Accurate</td>
</tr>
</tbody>
</table>
Disadvantages of Permutation Tests & Bootstrap Statistics

I. Poor performance with small sample sizes
   • Might be inaccurate
   • Limited set of possible $p$-values
Disadvantages of Permutation Tests & Bootstrap Statistics

1. Poor performance with small sample sizes
   • Might be inaccurate
   • Limited set of possible $p$-values

2. Not practical for computationally intensive analyses (e.g., non-linear regression via gradient descent)
Presentation Outline

• “Classic” Analytical Inferential Statistics
  - Parametric & non-parametric

• Resampling-Based Inferential Statistics
  - Randomization/permutation tests
  - Bootstrap statistics

• Correcting for Multiple Comparisons
  - Permutation test based control of family-wise error
  - Benjamini methods for control of false discovery rate
  - Evaluating multiple comparison correction on simulated ERP data
Potentially Lots of Possible Statistical Tests

Conventional ERP study:
- 2 conditions
- 26 electrodes
- 218 time points (50-920 ms)
- 5,668 dependent variables
Potentially Lots of Possible Statistical Tests

Even more dependent variables with time-frequency analyses!!

DANGER: Lots of statistical tests means a high likelihood of false discoveries!!

Even more dependent variables with time-frequency analyses!!
## Hypothetical Experiment #4

- Two conditions: A & B
- Within-subject design
- Three subjects
- Two dependent variables: X & Y

### Table

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>A-B</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sub1</strong></td>
<td>-4</td>
<td>28</td>
<td>-32</td>
</tr>
<tr>
<td><strong>Sub2</strong></td>
<td>3</td>
<td>-13</td>
<td>16</td>
</tr>
<tr>
<td><strong>Sub3</strong></td>
<td>36</td>
<td>30</td>
<td>6</td>
</tr>
</tbody>
</table>

### Table Y

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>A-B</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sub1</strong></td>
<td>141</td>
<td>-121</td>
<td>262</td>
</tr>
<tr>
<td><strong>Sub2</strong></td>
<td>142</td>
<td>72</td>
<td>70</td>
</tr>
<tr>
<td><strong>Sub3</strong></td>
<td>67</td>
<td>163</td>
<td>-96</td>
</tr>
</tbody>
</table>

\[ t_x = -0.23 \]
\[ t_y = 0.76 \]
Control of Family-Wise Error Rate (FWER)

\[ FWER = P(R_F > 0) = \alpha_{fam} \]

\( R_F \) = number of false discoveries in the family of tests
Control of Family-Wise Error Rate (FWER)

\[ FWER = P(R_F > 0) = \alpha_{fam} \]

\( R_F \) = number of false discoveries in the family of tests

This “family” consists of two tests:

\[ t_x = -0.23 \]

\[ t_y = 0.76 \]
Control of Family-Wise Error Rate (FWER)

\[ FWER = P(R_F > 0) = \alpha_{fam} \]

\[ R_F = \text{number of false discoveries in the family of tests} \]

**Bonferroni Correction:**

Desired "family-wise alpha" = Desired \( \alpha_{fam} = 0.05 \)

Bonferroni "test-wise alpha" = \( \alpha_{test} = \frac{\text{Desired } \alpha_{fam}}{\# \text{ of comparisons}} = \frac{0.05}{2} = 0.025 \)

True \( \alpha_{fam} \leq \text{Desired } \alpha_{fam} \)
Control of Family-Wise Error Rate (FWER)

\[ FWER = P(R_F > 0) = \alpha_{\text{fam}} \]

\[ R_F = \text{number of false discoveries in the family of tests} \]

**Bonferroni Correction:**

Desired "family - wise alpha" = Desired \( \alpha_{\text{fam}} = 0.05 \)

Bonferroni "test - wise alpha" = \( \alpha_{\text{test}} = \frac{\text{Desired } \alpha_{\text{fam}}}{\# \text{ of comparisons}} = \frac{0.05}{2} = 0.025 \)

True \( \alpha_{\text{fam}} \leq \text{Desired } \alpha_{\text{fam}} \rightleftharpoons \text{Might be overly conservative} \)
Bonferroni Correction

- Desired $\alpha_{\text{fam}}$: 5%
- Bonferroni $\alpha_{\text{test}}$: 2.5%

**Independent**

Variable X vs Variable Y

**Perfectly Correlated**

Variable X vs Variable Y

**Imperfectly Correlated (rho=0.81)**

Variable X vs Variable Y

- Estimated true family-wise $\alpha$ level (95% Confidence Intervals)
  - 4.9(±0.3)%
  - 2.3(±0.3)%
  - 4.1(±0.3)%
Bonferroni Correction

• Desired $\alpha_{\text{fam}}$: 5%
• Bonferroni $\alpha_{\text{test}}$: 2.5%

PERMUTATION TESTS CAN DO BETTER!!

Estimated true family-wise $\alpha$ level (95% Confidence Intervals)

- Independent: 4.9(±0.3)%
- Perfectly Correlated: 2.3(±0.3)%
- Imperfectly Correlated (rho=0.81): 4.1(±0.3)%

Bonferroni Correction

- Desired $\alpha_{\text{fam}}$: 5%
- Bonferroni $\alpha_{\text{test}}$: 2.5%
## Permutation Test

### Observed Values (Permutation #1)

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>A-B</td>
</tr>
<tr>
<td>Sub1</td>
<td>-4</td>
<td>28</td>
<td>-32</td>
</tr>
<tr>
<td>Sub2</td>
<td>3</td>
<td>-13</td>
<td>16</td>
</tr>
<tr>
<td>Sub3</td>
<td>36</td>
<td>30</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Y</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>A-B</td>
</tr>
<tr>
<td></td>
<td>141</td>
<td>-121</td>
<td>262</td>
</tr>
<tr>
<td></td>
<td>142</td>
<td>72</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>67</td>
<td>163</td>
<td>-96</td>
</tr>
</tbody>
</table>

\[ t_x = -0.23 \]

\[ t_y = 0.76 \]

\[ t_{\text{max}} = \text{most extreme } t\text{-score} = 0.76 \]
### Permutation Test

#### Permutation #2

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>A-B</td>
</tr>
<tr>
<td>Sub1</td>
<td>28</td>
<td>-4</td>
<td>32</td>
</tr>
<tr>
<td>Sub2</td>
<td>3</td>
<td>-13</td>
<td>16</td>
</tr>
<tr>
<td>Sub3</td>
<td>36</td>
<td>30</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Y</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>A-B</td>
</tr>
<tr>
<td></td>
<td>-121</td>
<td>141</td>
<td>-262</td>
</tr>
<tr>
<td></td>
<td>142</td>
<td>72</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>67</td>
<td>163</td>
<td>-96</td>
</tr>
</tbody>
</table>

\[
t_x = 2.38
\]

\[
t_y = -1.00
\]

\[
t_{max} = \text{most extreme } t\text{-score} = 2.38
\]
**Null Hypothesis**

<table>
<thead>
<tr>
<th>Permutation</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_{\text{max}}$</td>
<td>-2.377</td>
<td>-2.372</td>
<td>-1.27</td>
<td>-0.76</td>
<td>0.76</td>
<td>1.27</td>
<td>2.372</td>
<td>2.377</td>
</tr>
</tbody>
</table>

**Decision Rule:** If observed difference is most positive or negative, reject null hypothesis (two tailed test).

Critical $t=±2.377$
## Null Hypothesis

<table>
<thead>
<tr>
<th>Permutation</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t_{\text{max}} )</td>
<td>-2.377</td>
<td>-2.372</td>
<td>-1.27</td>
<td>-0.76</td>
<td>0.76</td>
<td>1.27</td>
<td>2.372</td>
<td>2.377</td>
</tr>
</tbody>
</table>

### Decision Rule:
If observed difference is most positive or negative, reject null hypothesis (two tailed test).

\[
\text{Critical } t = \pm 2.377
\]

\[
\alpha_{\text{fam}} = \frac{2}{8} = 0.25
\]
### Permutation Test

**Observed Values (Permutation #1)**

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Sub1</td>
<td>-4</td>
<td>28</td>
</tr>
<tr>
<td>Sub2</td>
<td>3</td>
<td>-13</td>
</tr>
<tr>
<td>Sub3</td>
<td>36</td>
<td>30</td>
</tr>
</tbody>
</table>

\[ t_x = -0.23 \]

\[ t_y = 0.76 \]

Perm Test Critical \( t = \pm 2.377 \)
### Permutation Test

#### Observed Values (Permutation #1)

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Sub1</td>
<td>-4</td>
<td>28</td>
</tr>
<tr>
<td>Sub2</td>
<td>3</td>
<td>-13</td>
</tr>
<tr>
<td>Sub3</td>
<td>36</td>
<td>30</td>
</tr>
</tbody>
</table>

\[ t_x = -0.23 \]
\[ t_y = 0.76 \]

Perm Test Critical \( t = \pm 2.377 \)

Retain null hypothesis
(i.e., neither X nor Y significantly differ across A & B)
Corrects for Multiple Comparisons by Raising Critical $t$

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th></th>
<th></th>
<th>Y</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>A-B</td>
<td>A</td>
<td>B</td>
<td>A-B</td>
</tr>
<tr>
<td>Sub1</td>
<td>-4</td>
<td>28</td>
<td>-32</td>
<td>141</td>
<td>-121</td>
<td>262</td>
</tr>
<tr>
<td>Sub2</td>
<td>3</td>
<td>-13</td>
<td>16</td>
<td>142</td>
<td>72</td>
<td>70</td>
</tr>
<tr>
<td>Sub3</td>
<td>36</td>
<td>30</td>
<td>6</td>
<td>67</td>
<td>163</td>
<td>-96</td>
</tr>
</tbody>
</table>

$t_x = -0.23$

Perm Test Critical $t = \pm 2.377$

Repeated Measures $t$-test Critical $t$

(no correction for two comparisons) $= \pm 2.353$

$t_y = 0.76$
\textbf{$t_{\text{max}}$ Permutation Test}

- Desired $\alpha_{\text{fam}}$: 5%

\begin{align*}
\text{Independent} & \quad 4.9(\pm 0.3)\% \\
\text{Perfectly Correlated} & \quad 4.8(\pm 0.3)\% \\
\text{Imperfectly Correlated (rho=0.81)} & \quad 5.1(\pm 0.3)\%
\end{align*}

Estimated true family-wise $\alpha$ level (95% Confidence Intervals)
**t_{max} Permutation Test**

- Desired $\alpha_{fam}$: 5%

---

**PERMUTATION TESTS DID BETTER!!**

<table>
<thead>
<tr>
<th>Correlation Type</th>
<th>$\alpha_{fam}$ (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent</td>
<td>4.9(±0.3)%</td>
</tr>
<tr>
<td>Perfectly Correlated</td>
<td>4.8(±0.3)%</td>
</tr>
<tr>
<td>Imperfectly Correlated (ρ=0.81)</td>
<td>5.1(±0.3)%</td>
</tr>
</tbody>
</table>

Estimated true family-wise $\alpha$ level (95% Confidence Intervals)
Figure 1. Averaged frequent and rare waveforms obtained from 13 subjects in a study of P3 (top); average difference potential waveform obtained by subtracting frequent from rare waveforms (middle); plot of paired-samples t statistics computed at each time point (bottom).
Target-Standard Difference Wave (26 electrodes)

- Slow Wave
- P2 effect
- P3b effect

Test Window
Target-Standard Difference Wave (26 electrodes)

- Slow Wave
- P2 effect
- P3b effect
- Test Window
$t_{\text{max}}$ Distribution from 5000 Permutations

Critical Regions

2.9%  

Relative Frequency

$t_{\text{max}}$
Target-Standard Difference Wave (26 electrodes)

Permutation Test Critical Values ($\alpha_{\text{fam}} = 0.057$)

- T-score
- Slow Wave
- P2 effect
- P3b effect
- Test Window
Target-Standard Difference Wave (26 electrodes)

Bonferroni Critical Values ($\alpha_{fam} = .057$)

Test-Wise $\alpha_{test} = .00001$

Test Window

Slow Wave

P2 effect

P3b effect
Target-Standard Difference Wave (26 electrodes)

Permutation Test Critical Values \( (\alpha_{\text{fam}}=0.057) \)

Test-Wise \( \alpha_{\text{test}}=0.00005 \)

Slow Wave

P2 effect

P3b effect

Test Window
Permutation Tests: Some Pros

1. FWER control provides the same degree of certainty as more selective a priori tests

2. Guaranteed accuracy for simple tests (e.g., t-tests, correlation)

3. Relatively powerful when dependent variables are highly correlated (like EEG)
Permutation Tests: Some Cons

1. For more complicated tests (e.g., two factor ANOVAs) the results are only “asymptotically exact” (like bootstrapping).

2. Power can still be rather weak with a larger number of comparisons.
Presentation Outline

• “Classic” Analytical Inferential Statistics
  - Parametric & non-parametric

• Resampling-Based Inferential Statistics
  - Randomization/permutation tests
  - Bootstrap statistics

• Correcting for Multiple Comparisons
  - Permutation test based control of family-wise error
  - Benjamini methods for control of false discovery rate
  - Evaluating multiple comparison correction on simulated ERP data
Control of Family-Wise Error Rate (FWER)

$$FWER = P(R_F > 0) = \alpha$$

$$R_F = \text{number of false discoveries in the family of tests}$$

If FWER=5%, you have a 5% chance that one or more of your significant $p$-values is a mistake.
Control of Family-Wise Error (FWER)

\[ FWER = P(R_F > 0) = \alpha \]

\[ R_F = \text{number of false discoveries in the family of tests} \]

If FWER=5%, you have a 5% chance that one or more of your significant \( p \)-values is a mistake.

Control of False Discovery Rate (FDR)

False Discovery Proportion = \( FDP = \begin{cases} \frac{R_F}{R} & \text{if } R > 0 \\ 0 & \text{if } R = 0 \end{cases} \)

\[ R = \text{number of rejected null hypotheses} \]

\[ FDR = E(FDP) = \alpha \]

If FDR=5%, on average, 5% of your significant \( p \)-values are mistakes.
Imagine you replicate an experiment thousands of times

- FDR \( \leq 5\% \)

Diagram:
- t-score vs. Time (ms)
- Test Window

\[ \text{t-score} \]
\[ \text{Time (ms)} \]
Imagine you replicate an experiment thousands of times

- FDR ≤ 5%
- 100 significant t-tests, on average

Critical Values

Test Window
Imagine you replicate an experiment thousands of times

- $FDR \leq 5\%$
- 100 significant $t$-tests, on average
- 5 $t$-tests or less will be false discoveries, on average
Imagine you replicate an experiment thousands of times

Some False Discoveries OK!!

- FDR ≤ 5%
- 100 significant t-tests, on average
- 5 t-tests or less will be false discoveries, on average

Critical Values

Test Window
**Most Popular FDR Control Algorithm**

**Benjamini & Hochberg (1995)**

1. Sort the $p$-values from the entire family of $m$ tests (i.e., $m$ is the total number of hypothesis tests) in order of smallest to largest. $p_i$ refers to the $i$th largest $p$-value.

2. Define $k$, as the largest value of $i$ for which the following is true:

   $$ p_i \leq \left( \frac{i}{m} \right) \alpha $$

3. If at least one value of $i$ satisfies that relationship, then hypotheses 1 through $k$ are rejected. If not, no hypotheses are rejected.
Sorted $p$-Values

\[
p_i \leq \left( \frac{i}{m} \right) \alpha
\]
Sorted p-Values

Greatest Crossing Point

FDR Criterion

Significant p-values
Most Popular FDR Control Algorithm

Benjamini & Hochberg (1995)

1. If the dependent variables are independent or exhibit positive regression dependency, the BH algorithm guarantees:

\[ FDR \leq \left( \frac{m_0}{m} \right) \alpha \]

where \( m_0 \) equals the number of null hypotheses that are true and \( m \) equals the total number of null hypotheses.

2. If the dependent variables are Gaussian, then positive regression dependency means that none of the variables are negatively correlated.

Most Popular FDR Control Algorithm

Benjamini & Hochberg (1995)

Problem

1. If the dependent variables are independent or exhibit positive regression dependency, the BH algorithm guarantees:

\[ FDR \leq \left( \frac{m_0}{m} \right) \alpha \]

where \( m_0 \) equals the number of null hypotheses that are true and \( m \) equals the total number of null hypotheses.

2. If the dependent variables are Gaussian, then positive regression dependency means that none of the variables are negatively correlated.

More General Variant of BH FDR Control Algorithm

Benjamini & Yekutieli (2001)

1. Sort the $p$-values from the entire family of $m$ tests (i.e., $m$ is the total number of hypothesis tests) in order of smallest to largest. $p_i$ refers to the $i$th largest $p$-value.

2. Define $k$, as the largest value of $i$ for which the following is true:

New BY Criterion: 
$$p_i \leq \left( \frac{i}{m \sum_{j=1}^{m} \frac{1}{j}} \right) \alpha$$

Original BH Criterion: 
$$p_i \leq \left( \frac{i}{m} \right) \alpha$$

3. If at least one value of $i$ satisfies that relationship, then hypotheses 1 though $k$ are rejected. If not, no hypotheses are rejected.
More General Variant of BH FDR Control Algorithm

Benjamini & Yekutieli (2001)

1. Regardless of dependent variable dependency structure, BY algorithm guarantees:

\[
FDR \leq \left( \frac{m_0}{m} \right) \alpha
\]

where \(m_0\) equals the number of null hypotheses that are true and \(m\) equals the total number of null hypotheses.
Target-Standard Difference Wave (26 electrodes)

Permutation Test Critical Values ($\alpha_{fam} = .057$)

Test-Wise $\alpha_{test} = .00005$

Slow Wave

P2 effect

P3b effect

Test Window
Target-Standard Difference Wave (26 electrodes)

Test-Wise $\alpha_{\text{test}} = .02$

BH Critical Values ($FDR \leq .05$)

Test Window
Target-Standard Difference Wave (26 electrodes)

Test-Wise \( \alpha_{test} = .0008 \)

Critical Values \((FDR \leq .05)\)

Test Window
Target-Standard Difference Wave (26 electrodes)

EEGLAB uses BH FDR method

Test-Wise $\alpha_{test} = .02$

Test Window

BH Critical Values ($FDR \leq .05$)
FDR Control: **Pros**

1. With a large number of comparisons, FDR is generally more powerful than FWER control (especially if an appreciable proportion of null hypotheses are false).

2. If all null hypotheses are true, FDR control=FWER control. Thus, if you find effects with FDR control you can be $1 - \alpha$ confident that some effect is present.

3. Benjamini procedures can be used with any hypothesis test (simply requires test $p$-values).
1. FDR control may lead to a high proportion of false positives with some frequency.

When applied to simulated data and an $\alpha$-level of 10%, Korn et al. (2004) found that the BH algorithm produces 29% or more false discoveries 10% of the time.
FDR Control: Cons

1. FDR control may lead to a high proportion of false positives with some frequency

2. FDR can be difficult to interpret as effects may disappear when analyses become more selective
Target-Standard Difference Wave (26 electrodes)

- BH Critical Values ($FDR \leq 0.05$)
- Test-Wise $\alpha_{test} = 0.02$
- Test Window
Target-Standard Difference Wave (26 electrodes)

Test-Wise
\[ \alpha_{\text{test}} = .005 \]

BH Critical Values
\( \text{FDR} \leq .05 \)

Test Window
FDR Control: **Cons**

1. FDR control may lead to a high proportion of false positives with some frequency

2. FDR can be difficult to interpret as effects may disappear when analyses become more selective

3. More powerful and popular FDR control algorithm (BH) is not guaranteed to work for data with negatively correlated variables.
FDR Control: **Cons**

1. FDR control may lead to a high proportion of false positives with some frequency

2. FDR can be difficult to interpret as effects may disappear when analyses become more selective

3. More powerful and popular FDR control algorithm (BH) is not guaranteed to work for data with negatively correlated variables

   - **However**, recent work by Clarke & Hall (2009) shows that for light tailed data (e.g., Gaussian) multiple comparison correction procedures will behave as if the data were independent if the number of variables is large enough
Presentation Outline

• “Classic” Analytical Inferential Statistics
  - Parametric & non-parametric

• Resampling-Based Inferential Statistics
  - Randomization/permutation tests
  - Bootstrap statistics

• Correcting for Multiple Comparisons
  - Permutation test based control of family-wise error
  - Benjamini methods for control of false discovery rate
  - Evaluating multiple comparison correction on simulated ERP data
ERP Simulations

• Simulation Parameters
  - Simulated ERP noise estimated from ERP noise in a real ERP study
  - 26 electrodes, 201 time points (100-900 ms)
  - Average & bimastoid reference
  - Negatively correlated dependent variables ranged from 13-51%

• ERP Effects
  1. Null effect: 0% of comparisons differ from 0
  2. Focal effect ("N170"): 0.2% of comparisons differ from 0
  3. Broad effect ("P300"): 18.9% of comparisons differ from 0
  4. Combined focal & broad effect: 19.1% of comparisons differ from 0

Groppe, Urbach, & Kutas (in prep)
$Perm = t_{\text{max}}$ permutation test FWER control; $BH =$ Benjamini & Hochberg FDR control; $BY =$ Benjamini & Yekutieli FDR control
Perm = $t_{\text{max}}$ permutation test FWER control; BH = Benjaminin & Hochberg FDR control; BY = Benjamini & Yekutieli FDR control.
Probability of 20% or More False Discovery Proportion

Perm = \( t_{\text{max}} \) permutation test FWER control; BH = Benjamini & Hochberg FDR control; BY = Benjamini & Yekutieli FDR control
**Mean Proportion of Effects Detected**

**Bimastoid Reference**

- Sensitivity

- **Perm** = $t_{\text{max}}$ permutation test FWER control;
- **BH** = Benjamini & Hochberg FDR control;
- **BY** = Benjamini & Yekutieli FDR control

**Average Reference**

- Sensitivity

- **Perm** = $t_{\text{max}}$ permutation test FWER control;
- **BH** = Benjamini & Hochberg FDR control;
- **BY** = Benjamini & Yekutieli FDR control
Presentation Outline

• “Classic” Analytical Inferential Statistics
  - Parametric & non-parametric

• Resampling-Based Inferential Statistics
  - Randomization/permutation tests
  - Bootstrap statistics

• Correcting for Multiple Comparisons
  - Permutation test based control of family-wise error
  - Benjamini methods for control of false discovery rate
  - Evaluating multiple comparison correction on simulated ERP data

Summary:
Summary

1. FWER control via permutation tests:

   • **Pros:**
     - Relatively powerful because EEG is highly correlated
     - Same degree of error control as a priori analyses

   • **Cons:**
     - May sacrifice considerable power when applied to large numbers of comparisons
     - Only guaranteed to work for simple analyses
Summary

2. FDR control via BH & BY procedures:

- **Pros:**
  - Relatively powerful because of less conservative error measure
  - More general than permutation test procedures and often more powerful

- **Cons:**
  - Can be difficult to interpret due to invalid statistical assumptions, potentially high proportions of false discoveries, and interactions between variables
  - Simulations found no evidence that these FDR procedures are prone to the former two problems when applied to ERPs
Yet More Multiple Comparison Correction Procedures

1. Control of False Discovery Exceedance (FDX) (also called control of FDP)

\[ FDX = P(FDP > c) \]

\[ FDP = \begin{cases} \frac{R_F}{R} & \text{if } R > 0 \\ 0 & \text{if } R = 0 \end{cases} \]

2. Control of Generalized Family-Wise Error Rate (GFWER)

\[ GFWER = P(R_F > u) \]

\[ u = \text{an acceptable number of false discoveries} \]

3. Control of Local False Discovery Rate:
   Bootstrap based control of FDR (Efron, 2004)
Cluster correction for multiple comparisons

Original difference

Difference bootstrap 1

Difference bootstrap 2

Difference bootstrap 3

Maris & Oostenveld (2007) *Jnl of Neuro Methods*
Presentation Outline

• “Classic” Analytical Inferential Statistics
  - Parametric & non-parametric

• Resampling-Based Inferential Statistics
  - Randomization/permutation tests
  - Bootstrap statistics

• Correcting for Multiple Comparisons
  - Permutation test based control of family-wise error
  - Benjamini methods for control of false discovery rate
  - Evaluating multiple comparison correction on simulated ERP data
Recommended Textbooks

- **Randomization, Bootstrap and Monte Carlo Methods in Biology**
  - Second Edition
  - Bryan F. J. Manly

- **An Introduction to the Bootstrap**
  - Bradley Efron
  - Robert J. Tibshirani

- **Applying Contemporary Statistical Techniques**
  - Rand R. Wilcox
Recommended Papers


Thanks to G. Rousselet
Thanks!

EEGLAB Compatible Software for ERP Analysis

Questions:
dgroppe@cogsci.ucsd.edu

http://openwetware.org/wiki/Mass_Univariate_ERP_Toolbox
statcond function in EEGLAB

```matlab
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 surogs] = 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

\[ a = \{ \text{rand}(3,4,10) \text{ rand}(3,4,10) \text{ rand}(3,4,10); \ldots \text{ rand}(3,4,10) \text{ rand}(3,4,10) \text{ rand}(3,4,10)+0.5 \}; \]

% pseudo (2,3)-condition data array, each entry containing
% ten (3,4) data matrices
\[ [F \ df \ pvals] = \text{statcond}(a, 'mode', 'perm'); \]
% 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
Non-parametric statistics

Do not assume a distribution for the data

χ² is used to compare 2 or more unpaired samples

\[
\chi^2 = \sum_{i,j} (\text{Observed}_{i,j} - \text{expected}_{i,j})^2 / \text{expected}_{i,j}
\]

Signal A

Signal B

Observed A

Observed B

expected

χ² dist. df=5

5% of area
Bootstrap for ERPs and time-frequency

N trials

Baseline period

Channel/comp. 1

Average

Bootstrap for ERPs and time-frequency

Significantly non-0

Bootstrap 1

Bootstrap 2

Bootstrap 3

Bootstrap k

median

median

median

2.5% 97.5%

-300 -200 -100 0 100 200 300 400 500 600

Time (ms)