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 assess 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**

\[ t = \frac{\text{Mean}_\text{difference}}{\text{Standard}_\text{deviation}} \sqrt{N - 1} \]

**Unpaired**

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

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

\[ F = \frac{\text{Variance}_{\text{interGroup}}}{\frac{\text{Variance}_{\text{WithinGroup}}}{N - N_{\text{Group}}}} \frac{N_{\text{Group}} - 1}{N - N_{\text{Group}}} \]
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

Median

Mean
Robust measures of ERP

• Non-robust estimator
  – Mean: \( \text{mERP} = \text{mean(EEG.data,...)} \)

• Robust estimator
  - Median: \( \text{mdERP} = \text{median(EEG.data,...)} \)
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
<table>
<thead>
<tr>
<th>Goal</th>
<th>Binomial or Discrete</th>
<th>Continuous measurement (from a normal distribution)</th>
<th>Continuous measurement, Rank, or Score (from non-normal distribution)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example of data sample</td>
<td>List of patients recovering or not after a treatment</td>
<td>Readings of heart pressure from several patients</td>
<td>Ranking of several treatment efficiency by one expert</td>
</tr>
<tr>
<td>Describe one data sample</td>
<td>Proportions</td>
<td>Mean, SD</td>
<td>Median</td>
</tr>
<tr>
<td>Compare one data sample to a hypothetical distribution</td>
<td>$\chi^2$ or binomial test</td>
<td>One-sample t test</td>
<td>Sign test or Wilcoxon test</td>
</tr>
<tr>
<td>Compare two paired samples</td>
<td>Sign test</td>
<td>Paired t test</td>
<td>Sign test or Wilcoxon test</td>
</tr>
<tr>
<td>Compare two unpaired samples</td>
<td>$\chi^2$ square, Fisher's exact test</td>
<td>Unpaired t test</td>
<td>Mann-Whitney test</td>
</tr>
<tr>
<td>Compare three or more unmatched samples</td>
<td>$\chi^2$ test</td>
<td>One-way ANOVA</td>
<td>Kruskal-Wallis test</td>
</tr>
<tr>
<td>Compare three or more matched samples</td>
<td>Cochrane Q test</td>
<td>Repeated-measures ANOVA</td>
<td>Friedman test</td>
</tr>
<tr>
<td>Quantify association between two paired samples</td>
<td>Contingency coefficients</td>
<td>Pearson correlation</td>
<td>Spearman correlation</td>
</tr>
</tbody>
</table>

**Toolboxes and Software**

- Matlab Statistics toolbox; Parra & Sajda plugin
- EEGLAB, FIELDTRIP, LIMO EEG
- Matlab Statistics toolbox

How handle violations of normality?

- Bootstrap
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

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

analyze

difference

analyze

$X_{org}$
Confidence interval for the difference
Bootstrap approach

analyze

difference

X_1

analyze
Confidence interval for the difference
Bootstrap approach

analyze

difference

analyze

X₂
Measure for the bootstrap
Confidence interval for the difference
Bootstrap approach

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 involves assessing if 0 is included in the tails.
Assessing significance

Difference 1  Difference 2  Difference 3  Difference 4

Original Difference

Difference mask at p<0.05

2.5% 2.5%
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.
  - $\rightarrow$ too conservative

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

- **False detection rate (FDR)**

- **Cluster randomization**
FDR

1. For a given $\alpha$, 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 $i = 1, \ldots, k$. 
FDR procedure

1. For a given $\alpha$, 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)

- Create column C2 by computing $k*\alpha/N$

- 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 $p_{fdr}$

<table>
<thead>
<tr>
<th>Index &quot;k&quot;</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.001</td>
</tr>
<tr>
<td>2</td>
<td>0.002</td>
</tr>
<tr>
<td>3</td>
<td>0.01</td>
</tr>
<tr>
<td>4</td>
<td>0.03</td>
</tr>
<tr>
<td>5</td>
<td>0.04</td>
</tr>
<tr>
<td>6</td>
<td>0.045</td>
</tr>
<tr>
<td>7</td>
<td>0.05</td>
</tr>
<tr>
<td>8</td>
<td>0.1</td>
</tr>
<tr>
<td>9</td>
<td>0.2</td>
</tr>
<tr>
<td>10</td>
<td>0.6</td>
</tr>
</tbody>
</table>
FDR procedure

1. For a given $\alpha$, 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)
- Create column C2 by computing $k\alpha/N$
- 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 $p_{fdr}$

<table>
<thead>
<tr>
<th>Index &quot;k&quot;</th>
<th>Actual</th>
<th>$k*0.05/10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.001</td>
<td>0.005</td>
</tr>
<tr>
<td>2</td>
<td>0.002</td>
<td>0.01</td>
</tr>
<tr>
<td>3</td>
<td>0.01</td>
<td>0.015</td>
</tr>
<tr>
<td>4</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>5</td>
<td>0.04</td>
<td>0.025</td>
</tr>
<tr>
<td>6</td>
<td>0.045</td>
<td>0.03</td>
</tr>
<tr>
<td>7</td>
<td>0.05</td>
<td>0.035</td>
</tr>
<tr>
<td>8</td>
<td>0.1</td>
<td>0.04</td>
</tr>
<tr>
<td>9</td>
<td>0.2</td>
<td>0.045</td>
</tr>
<tr>
<td>10</td>
<td>0.6</td>
<td>0.05</td>
</tr>
</tbody>
</table>
FDR procedure

1. For a given $\alpha$, 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)
- Create column C2 by computing $k*\alpha/N$
- 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 $p_{fdr}$

<table>
<thead>
<tr>
<th>Index &quot;k&quot;</th>
<th>Actual</th>
<th>k*0.05/10</th>
<th>C2-C1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.001</td>
<td>0.005</td>
<td>-0.004</td>
</tr>
<tr>
<td>2</td>
<td>0.002</td>
<td>0.01</td>
<td>-0.008</td>
</tr>
<tr>
<td>3</td>
<td>0.01</td>
<td>0.015</td>
<td>-0.005</td>
</tr>
<tr>
<td>4</td>
<td>0.03</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>5</td>
<td>0.04</td>
<td>0.025</td>
<td>0.015</td>
</tr>
<tr>
<td>6</td>
<td>0.045</td>
<td>0.03</td>
<td>0.015</td>
</tr>
<tr>
<td>7</td>
<td>0.05</td>
<td>0.035</td>
<td>0.015</td>
</tr>
<tr>
<td>8</td>
<td>0.1</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>9</td>
<td>0.2</td>
<td>0.045</td>
<td>0.155</td>
</tr>
<tr>
<td>10</td>
<td>0.6</td>
<td>0.05</td>
<td>0.55</td>
</tr>
</tbody>
</table>
FDR procedure

1. For a given $\alpha$, 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)
- Create column C2 by computing $k*\alpha/N$
- 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 $p_{fdr}$

<table>
<thead>
<tr>
<th>Index &quot;k&quot;</th>
<th>Actual</th>
<th>$k*0.05/10$</th>
<th>C2-C1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.001</td>
<td>0.005</td>
<td>-0.004</td>
</tr>
<tr>
<td>2</td>
<td>0.002</td>
<td>0.01</td>
<td>-0.008</td>
</tr>
<tr>
<td>3</td>
<td>0.01</td>
<td>0.015</td>
<td>-0.005</td>
</tr>
<tr>
<td>4</td>
<td>0.03</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>5</td>
<td>0.04</td>
<td>0.025</td>
<td>0.015</td>
</tr>
<tr>
<td>6</td>
<td>0.045</td>
<td>0.03</td>
<td>0.015</td>
</tr>
<tr>
<td>7</td>
<td>0.05</td>
<td>0.035</td>
<td>0.015</td>
</tr>
<tr>
<td>8</td>
<td>0.1</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>9</td>
<td>0.2</td>
<td>0.045</td>
<td>0.155</td>
</tr>
<tr>
<td>10</td>
<td>0.6</td>
<td>0.05</td>
<td>0.55</td>
</tr>
</tbody>
</table>
FDR procedure

1. For a given $\alpha$, 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)
- Create column C2 by computing $k*\alpha/N$
- 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 $p_{fdr}$

<table>
<thead>
<tr>
<th>Index &quot;k&quot;</th>
<th>Actual</th>
<th>$k*0.05/10$</th>
<th>C2-C1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.001</td>
<td>0.005</td>
<td>-0.004</td>
</tr>
<tr>
<td>2</td>
<td>0.002</td>
<td>0.01</td>
<td>-0.008</td>
</tr>
<tr>
<td>3</td>
<td>0.01</td>
<td>0.015</td>
<td>-0.005</td>
</tr>
<tr>
<td>4</td>
<td>0.03</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>5</td>
<td>0.04</td>
<td>0.025</td>
<td>0.015</td>
</tr>
<tr>
<td>6</td>
<td>0.045</td>
<td>0.03</td>
<td>0.015</td>
</tr>
<tr>
<td>7</td>
<td>0.05</td>
<td>0.035</td>
<td>0.015</td>
</tr>
<tr>
<td>8</td>
<td>0.1</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>9</td>
<td>0.2</td>
<td>0.045</td>
<td>0.155</td>
</tr>
<tr>
<td>10</td>
<td>0.6</td>
<td>0.05</td>
<td>0.55</td>
</tr>
</tbody>
</table>
Comparison of different corrections

<table>
<thead>
<tr>
<th>Index &quot;j&quot;</th>
<th>Actual</th>
<th>j*0.05/10</th>
<th>C2-C1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.001</td>
<td>0.005</td>
<td>-0.004</td>
</tr>
<tr>
<td>2</td>
<td>0.002</td>
<td>0.01</td>
<td>-0.008</td>
</tr>
<tr>
<td>3</td>
<td>0.01</td>
<td>0.015</td>
<td>-0.005</td>
</tr>
<tr>
<td>4</td>
<td>0.03</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>5</td>
<td>0.04</td>
<td>0.025</td>
<td>0.015</td>
</tr>
<tr>
<td>6</td>
<td>0.045</td>
<td>0.03</td>
<td>0.015</td>
</tr>
<tr>
<td>7</td>
<td>0.05</td>
<td>0.035</td>
<td>0.015</td>
</tr>
<tr>
<td>8</td>
<td>0.1</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>9</td>
<td>0.2</td>
<td>0.045</td>
<td>0.155</td>
</tr>
<tr>
<td>10</td>
<td>0.6</td>
<td>0.05</td>
<td>0.55</td>
</tr>
</tbody>
</table>

- **Bonferroni**
- **FDR**
- **Uncorrected**
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

2.5% 97.5%
Study GUI

STUDY 'Sternberg' - 'memorize vs ignore' component clusters

Select cluster to plot

- All cluster centroids
- Parentcluster 1 (311 ICs)
- outliers 2 (7 ICs)
- Cls 3 (23 ICs)
- Cls 4 (15 ICs)
- Cls 5 (35 ICs)
- Cls 6 (12 ICs)

Plot scalp maps
- Plot dipoles
- Plot ERP
- Plot spectra
- Plot ERPimage
- Plot ERSPs
- Plot ITCs

Create new cluster
- Rename selected cluster
- Merge clusters

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
- Do not correct for multiple comparisons
  - Bonferroni correction
  - Holms correction
  - FDR correction

Use Fieldtrip statistics
- Use montecarlo/permutation statistics
- Statistical threshold (p-value) = 0.05
- Randomization (n) = auto

CC channel neighbor parameters
- 'method', 'triangulation'
- 'clusterstatistic', 'maxsum'

CC clustering parameters
- 'method', 'triangulation'
- 'clusterstatistic', 'maxsum'

Cancel | Ok
Test between conditions (stern study)
LIMO EEG
References


Thanks to G. Rousselet