# V-2. STATISTICAL ANALYSES 2

EEG-TRAINING

ISABELLE HOXHA - ISABELLEHOXHA@GMAIL.COM

10/05/2021

# OUTLINE

1. Recap of part 1
2. Cluster-based analysis
    1. Definition
    2. How to
    3. Advantages and drawbacks
3. What to shuffle

# 1) RECAP OF PART 1

- All statistical analyses have key assumptions. Parametric test assume a certain (identified) distribution of the data (e.g.: Gaussian), which is rarely the case with EEG data
- We often deal with high-dimensional feature spaces (range of 100,000 features), which requires multiple comparisons, and adequate correction

# 2) WHAT IS CLUSTER-BASED ANALYSIS?

- With permutation testing, we assumed that features were independent. With EEG data, this is not necessarily the case, e.g.: signals at neighboring electrodes are correlated
- While permutation testing alone will list features that are significantly different from one condition to the other, cluster-based analysis will group features in neighborhoods and determine whether these neighborhoods are important in discrimating the two conditions

# 2) HOW TO DO CLUSTER-BASED ANALYSIS?

Recipe adapted from Maris and Oostenveld, « Nonparametric Statistical Testing of EEG- and MEG-Data ».

1. For every feature, compare the EEG-signal on the two conditions (for example: compute the $t$ −values)
2. Select all samples whose *t*-value is larger than some threshold.
3. Define your neighborhoods (in time, in frequency, and in space). These neighborhoods must be connected (ie you cannot skip an adjacent feature)
4. Calculate cluster-level statistics by taking the sum of the *t*-values within a cluster.
5. Take the largest of the cluster-level statistics

Alternative to steps 4. and 5.: identify blobs of above-threshold and threshold them again, or test each blob independently and correct for multiple comparisons (see part 1)

# 2) ADVANTAGES AND DRAWBACKS

**Advantages:**

- We no longer assume that features are independent
- May identify more physiologically plausible differences

**Drawbacks:**

- Definition of a neighborhood may be subjective (in particular spatially)
- Setting of the threshold is subjective

# 3) WHAT TO SHUFFLE?

In order to determine what to shuffle, it is important to be clear about the hypotheses. Some examples:

- We want to test the difference between the signal at two locations (« Signal at frequency A is bigger than signal at frequency B »):
  - ⇒ for each trial, permute the location of the signal randomly
- We want to test the difference between the signal for two different conditions (« Brain activity differs when participants are doing task A compared to when they do task B »)
  - ⇒ Asign each trial to either task A or task B randomly (keeping the same number of As and Bs as in the original data)

# CONCLUSION

- This last part concludes our introduction to EEG data analysis
- We provided methods for qualitative visualization and quantitative assessment of data
- It is important to note that all the techniques introduced are also applicable to MEG data
- The best way to learn the content is by doing: we encourage you to go through the tutorials

# REFERENCES


- Maris, Eric, and Robert Oostenveld. « Nonparametric Statistical Testing of EEG- and MEG-Data ». *Journal of Neuroscience Methods* 164, n° 1 (August 2007): 177-90. https://doi.org/10.1016/j.jneumeth.2007.03.024.