EEGpal: **Statistics module**

Version 1.0, 16.02.2025

The module 'Statistics' has for purpose to compare statistical differences for each Time Frame (TF) and each electrode of interest. This is usually done during the exploration of the dataset to find period or interest, to study the difference of EEG amplitude between conditions (trace analysis) or perform statistics on the inverse solution map of results (.ris files).

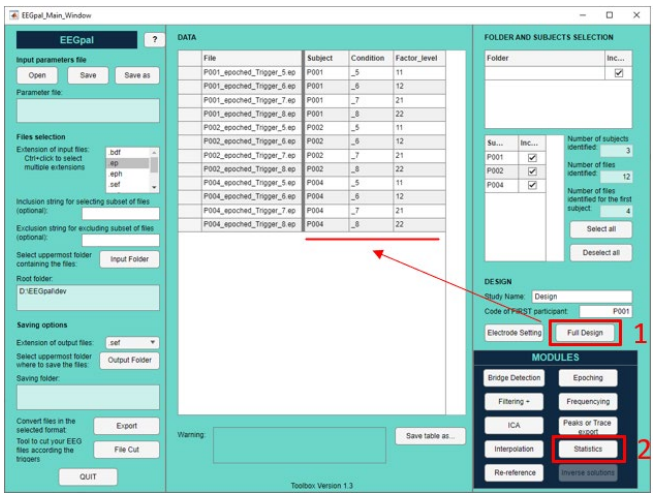
The EEGpal produces outputs similar as what Cartool proposed in its own 'Statistics on tracks' option. However, it permits to perform ANOVA with several factors which is not possible in Cartool. It replaces the STEN toolbox created by Jean-François Knebel (https://zenodo.org/records/1164038).

The Statistics module required the MATLAB **Statistics and Machine Learning Toolbox** add-on package to work.

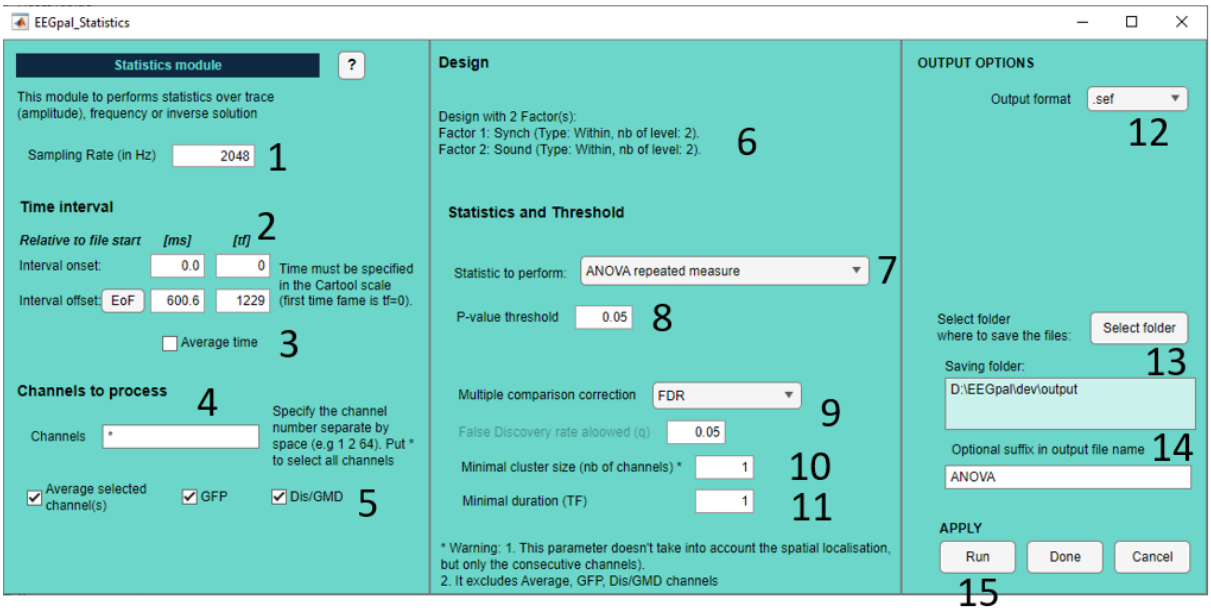
Before to start

The use of the Statistics module depends on the **Full Design** tool of the EEGpal main windows. In order to use the Statistics module, you must first specify the factorial design you wish to study. It will automatically fill the columns *Subject*, *Condition* and Factor_level of the Data table on the main screen. Please refer to the **Full Design** tool manual to understand how it works.

It is also important to note that all EEG file must have the same length, number of electrodes and sampling frequency. This is usually the case after **Epoching**.



Statistics



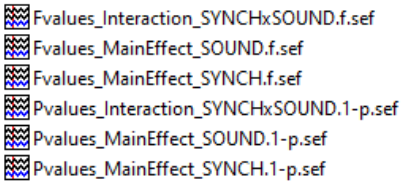
1. Sampling rate of the data. Normally should be automatically transfer from the main windows. You can adjust if it is not the case.
2. Specify the time interval to test.
   WARNING: The value 0 is the beginning of the file. This module doesn't take account about possible .mrk file which could define another origin.
   EoF=End of File.Specify the time interval.
3. Instead of looking at each time point, you can average them into a single time point by selecting this option.
4. Specify the electrode to study, Specify the indices of the electrodes (e.g. 1 2 64) and not the name specified by the coordinate file (e.g A1 A2 B32). A '*' means that all the electrodes will be study.
5. You can study additional channels with the average of electrodes specified in **4**, the global field power (GFP) or the global map dissimilarity (Dis/GMD).
6. This section summarizes the factorial design to be studied according to the elements specified in the **Full Design** tool. In this example we have a 2 within factors with 2 levels each (2x2 factorial design).
7. Specify the statistical test which will be performed. This option will depend of the factorial design specified in the **Full Design** tool:
   a. *One sample T-Test* with one factor and only one condition
   b. *ANOVA repeated measure* or *Paired T-Test* with one within factor and two conditions
   c. *ANOVA between* or *Two Samples T-Test* with one between factor and two conditions
   d. *ANOVA repeated measure* with serval within factors
   e. *ANOVA between* with several between factors
   f. *ANOVA mixed* with a factorial design containing at least one within-factor and one between-factor
8. Specify the P-value statistical threshold of test significance.
9. Correct the statistical threshold specified in **8** for multiple comparison or not. EEGpal allows to use the False Discovery Rate (FDR) correction method proposed in the following paper : *Benjamini, Y. & Yekutieli, D. (2001) The control of the false discovery rate in multiple testing under dependency. The Annals of Statistics. 29(4), 1165-1188.*

You can specify the False Discovery rate (q) you want to apply. Usually, the value of 0.05 is quite standard. See the FAQ for more detail.

10. In addition to statistical significance, you can apply a criterion with a minimum number of contiguous electrodes (clusters) that are significant.
    WARNING1: This parameter doesn't take into account the spatial localization, but only the consecutive channels.
    WARNING2: It excludes Averaged, GFP, Dis/GMD channels.
11. In addition to statistical significance, you can apply a criterion with a minimum duration of significant period (need to be specified in Time Frame and not ms).
12. Select the format for the output files.
13. Select the destination folder where the results files will be saved.
14. You can specify manually a suffix in the output file name.
15. Run the module to generate the output. After completion, you can press on Done or Cancel to close the module.

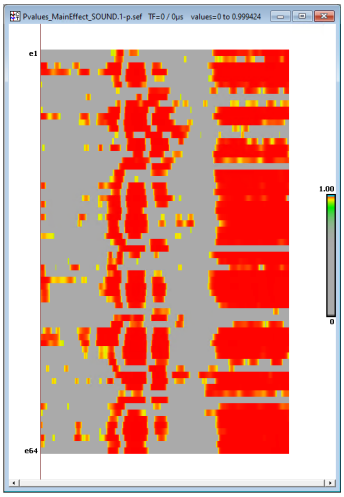
FAQ

What the output looks like?



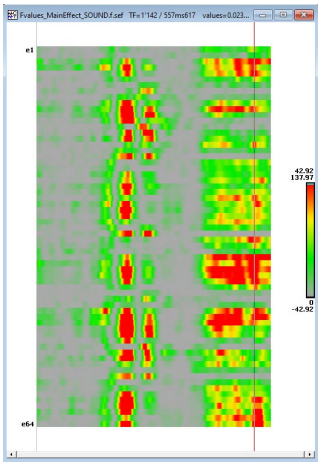
The Statistics module produce to file for each main effect of interaction of the factorial model:

- *Pvalues* file contains these p-values but only for the SIGNIFICANT test according to the threshold specified in the points **8-11**. In fact, these values are 1-p because the significant values are in the interval 0.95-1 (instead of the usual 0-0.05). All zero values mean that the test was not significant according to the criteria specified. You can use the Cartool software to visualize the results:



Yellow and red are the significant results and grey are the non-significant results.

- *Fvalues* or *Tvalues* contains the corresponding test value for every TF or channels tested (and not only for significant one). You can also use Cartool software to visualize it in color code by pressing several times on the button  :



Usually, we use mainly le Pvalues file to determine the period and group of electrodes which shows a significant difference between the experimental conditions.

Which Maltab commands are used to perform the statistical tests?

The commands use will depend on the statistical test performed as explained in point **7**:

- For *One sample T-Test* and *Paired T-Test*, the module uses the function **ttest**
- For *Two Samples T-Test*, the module uses the function **ttest2**
- For *ANOVA repeated measure* and *ANOVA mixed*, the module uses **fitrm** then **rmANOVA** functions
- For *ANOVA between*, the module uses the function **anovan**

All of these functions are part of the MATLAB **Statistics and Machine Learning Toolbox** add-on package. Therefore, the Statistics module will not run if this add-on package is not installed in your Matlab.

Should I use the False Discovery Rate (FDR) correction for multiple comparison and how is it performed?

Yes, it is a good idea to apply a correction for multiple comparisons because you are performing a large number of tests (number of TF * number of channels) which increases the probability of false positives. Therefore, it is highly recommended to apply this correction.

To performed this correction, EEGpal use the script **fdr_bh** of David Groppe (Version 2.3.0.0) download on Matlab central at the following link:
https://www.mathworks.com/matlabcentral/fileexchange/27418-fdr_bh

Complete ref:
David Groppe (2025). fdr_bh (https://www.mathworks.com/matlabcentral/fileexchange/27418-fdr_bh), MATLAB Central File Exchange. Retrieved February 16, 2025.

Is a way to perform non-parametric statistic?

Non-parametric statistics is useful when the number of participants is small, which usually leads to a violation of the normality of the data. Unfortunately, this option is not yet available in the Statistics module. Perhaps in a future release.