EEGpal: **Statistics module**

Version 2.01, 20.11.2025

Video tutorial: https://youtu.be/W8wzVinryh0

The purpose of the **Statistics** module is to compare statistical differences for each time frame (tf) and each electrode of interest. This is typically carried out when exploring the dataset to identify periods of interest, to study differences in EEG amplitude between conditions (trace analysis), or to perform statistics on the inverse solution map of results (.ris files).

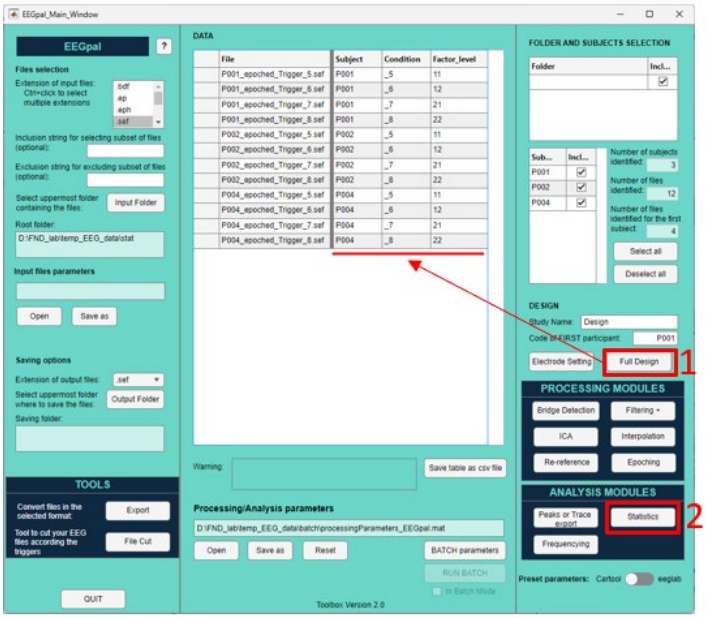
The EEGpal produces outputs similar to those proposed by Cartool in its 'Statistics on tracks' option. In addition, it enables ANOVA with multiple factors, which is not possible in Cartool. It also replaces the STEN toolbox created by Jean-François Knebel (https://zenodo.org/records/1164038).

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

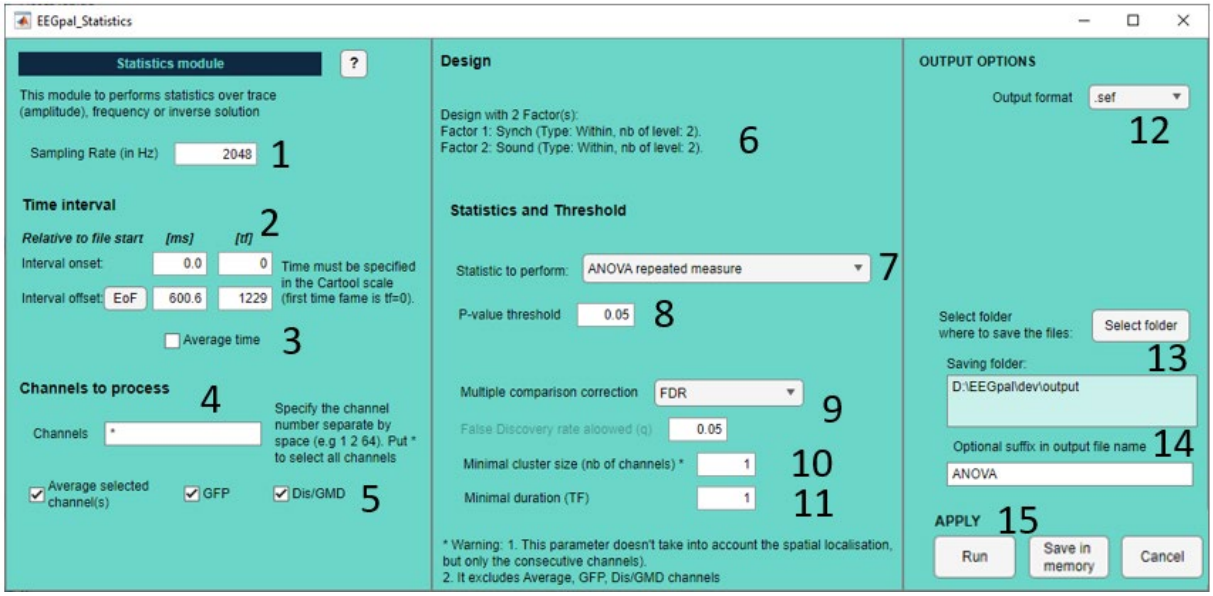
Before to start

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



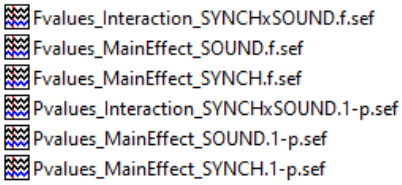
1. Sampling rate of the data. This should normally be transferred automatically from the main window. You can adjust this if necessary.
2. Specify the time interval. Warning: the value 0 marks the start of the file. This module does not take into account a possible .mrk file that could define a different origin.
   EoF=End of File.
3. Instead of looking at each time point individually, you can create an average by selecting this option.
4. Specify the electrode to be studied and the indices of the electrodes (1, 2, 64), rather than the names specified by the coordinate file (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).
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 based on the minimum number of contiguous significant electrodes (clusters).
    WARNING1: This parameter only considers consecutive channels, not spatial localization.
    WARNING2: This is not applied to the Averaged, GFP, Dis channels.
11. n addition to statistical significance, you can apply a criterion with a minimum significant period duration (this needs to be specified in the time frame, not in milliseconds).
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. There are three validation buttons:
    a. The **Run** button will carry out the processing parameterized in the module.
    b. The **Save in memory** button will store all the parameters in memory and close the module without performing the processing.
    c. The button **Cancel** closes the module without processing and without keeping the entered parameters in memory. The same effect will be achieved by closing the module window.

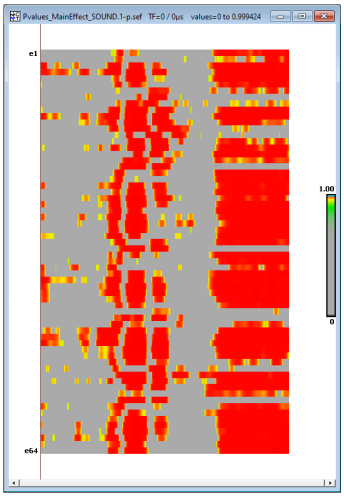
FAQ

What the output looks like?



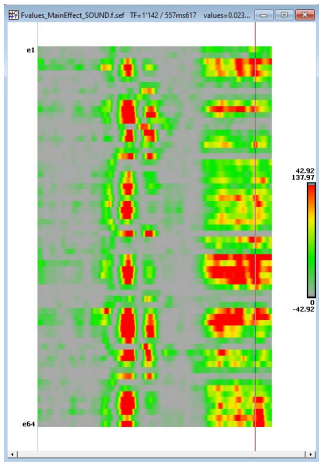
The Statistics module produce to file for each main effect of interaction of the factorial model:

- *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 lie within the 0.95–1 interval (rather than the usual 0–0.05 interval). All zero values indicate that the test was not significant according to the specified criteria. You can use the Cartool software to visualise the results:



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

- *Fvalues* or *Tvalues* files contain the corresponding test values for each tf or channels tested, not just the significant ones. You can also use the Cartool software to visualise them in a colour-coded format by pressing the button  several times:



We usually use the le Pvalues file primarily to determine the period and group of electrodes showing a significant difference between the experimental conditions.

Which Maltab commands are used for statistical testing?

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 apply the false discovery rate (FDR) correction for multiple comparisons, and if so, how is it performed?

Yes, it is a good idea to apply a correction for multiple comparisons, since you are conducting a large number of tests (the number of TFs multiplied by the number of channels), which increases the likelihood of false positives. Therefore, this correction is highly recommended.

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 there a way to perform a non-parametric statistical analysis?

Non-parametric statistics are useful when the number of participants is small, which typically results in the data not being normally distributed. Unfortunately, this option is not yet available in the **Statistics** module. Hopefully it will be included in a future release.