MUltiple tests corrections and FOrmatted tables Software (MUFOS) Help Guide and Documentation

Version: 1.0  
Date: XX.XX.XXXX  
Software page: <http://www.nikolaybpetrov.com/mufos>  
Github repository: XXX  
Reference: Petrov, N., Atanasov, V., & Thompson, T. (2020). ...  
Contact: [nikbpetrov@gmail.com](mailto:nikbpetrov@gmail.com)

Table of Contents

[Input types 3](#_Toc46298322)

[Handling missing and non-numeric data 3](#_Toc46298323)

[Exporting SPSS table from the SPSS output 4](#_Toc46298324)

[Downloading error details 5](#_Toc46298325)

[Calculations 6](#_Toc46298326)

[Correlations 6](#_Toc46298327)

[Multiple regression 7](#_Toc46298328)

[T-tests 7](#_Toc46298329)

[Raw input 7](#_Toc46298330)

[Summary data 7](#_Toc46298331)

[Effect sizes 7](#_Toc46298332)

[Multiple tests corrections 7](#_Toc46298333)

[Something went wrong and the documentation does not help. Now what? 8](#_Toc46298334)

# Input types

For example input files, please see the software page at <http://www.nikolaybpetrov.com/mufos>

*Raw data.* This input type requires participants' raw scores on each variable of interest. Each line is one participant and each column is a separate variable.

*Correlations from summary statistics.* This input type requires four columns: one column for variable one, one column for variable 2 and 2 columns for the correlation coefficients and the p-values of each correlation.

*Independent samples t-test from summary statistics*. This input type requires at least 7 and up to 8 columns. There should be one column to describe each of the mean, standard deviation and sample size for each sample. One row represents one test. Equal variances column only accepts TRUE and FALSE values, where TRUE means that equal variance will be assumed and FALSE means that equal variance will not be assumed. If this column is omitted, then the calculation will default to assuming equal variance.

*SPSS table.* An exported SPSS table for either correlations, multiple regression, independent samples t-test or paired samples t-test is expected. For information on how to export an SPSS table, see [Exporting SPSS table from the SPSS output](#_Exporting_SPSS_table)

*p values.* This input type requires one column to contain a list of p values. Other columns in the input file are also permissible. There can be no missing data on the p values column but the other columns can contain missing data. The output file will be the same as the input file with the addition of a adjusted p values column.

# Handling missing and non-numeric data

For all inputs except spss tables, the expectation is that some columns will contain only numeric data. Specifically:

|  |  |
| --- | --- |
| **Input type** | **Expected columns with numeric data** |
| Raw data – correlations | All |
| Raw data – multiple regression | All |
| Raw data – independent t-test | All except the grouping variable |
| Raw data – paired t-test | All |
| Correlations from summary statistics | Correlation coefficient column and pvalues column |
| Independent samples t-test from summary statistics | All except “Variable” column and “Equal variance” column. |
| P values | “pvalues” column |

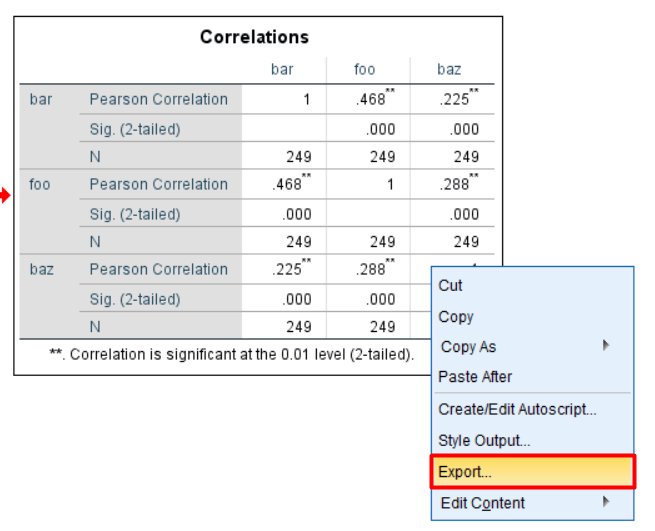
For raw data, you can select to either raise an error message if a column that is expected to contain only numeric data contains non-numeric data, or to ignore missing data pairwise for each test.

For correlations/independent samples t-test from summary statistics and p values input, the software will default to raising errors on non-numeric input where numeric input is expected. For these tests, missing data are not allowed.

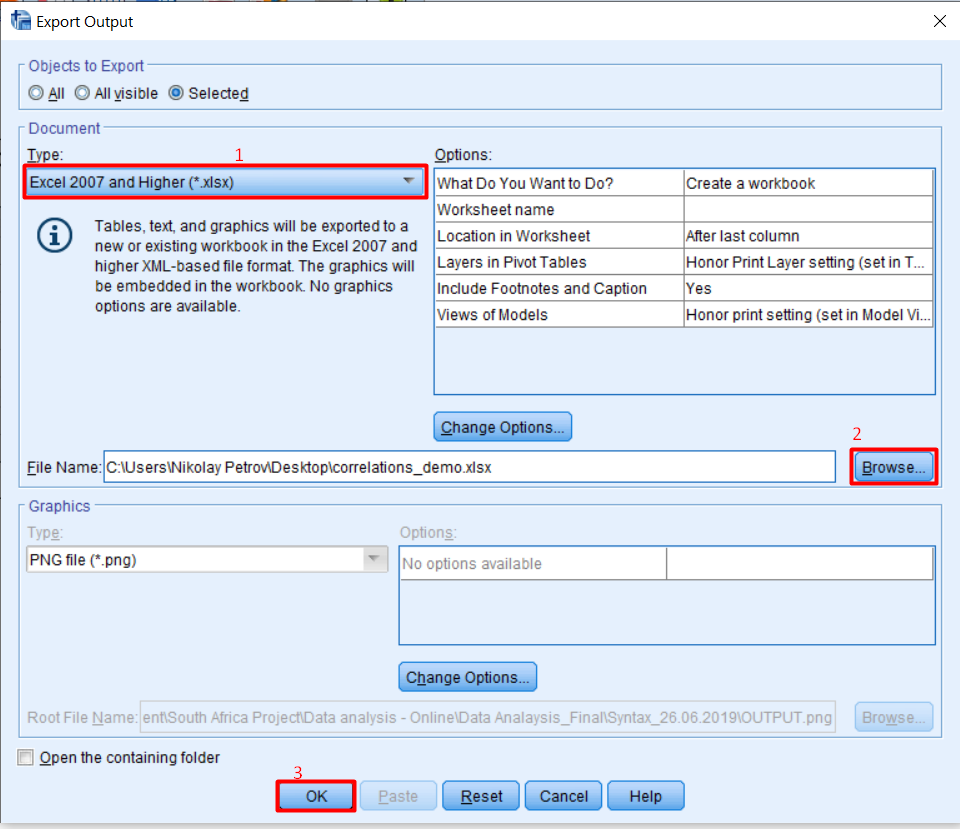
# Exporting SPSS table from the SPSS output

To export any table from the SPSS output, follow the steps below (Correlations table is used as an example but works with any table):

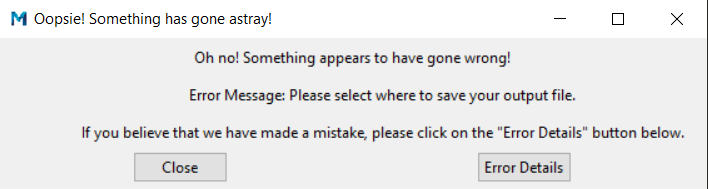
First, right-click on the SPSS table in the SPSS output, then click Export:



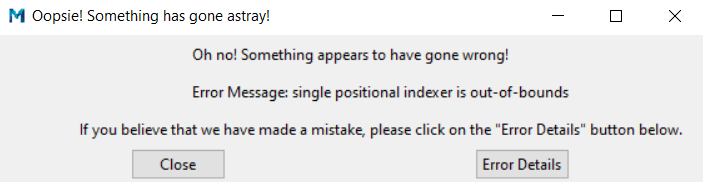
Then under Document, for Type, select “Excel 2007 and Higher (\*.xlsx)”. Then click Browse to select where to export the table. Finally, hit OK:

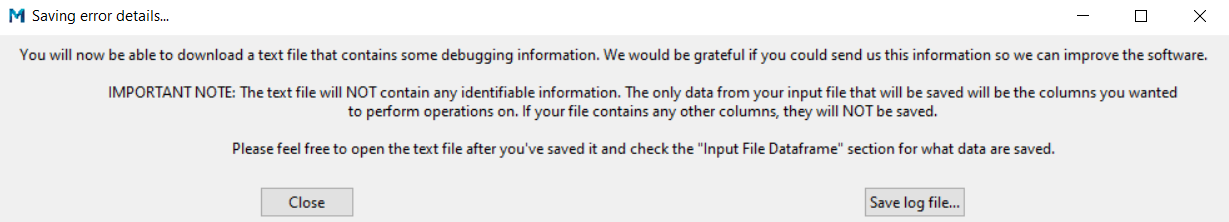


# Downloading error details

Sometimes if something goes wrong, you will get an error message. Most of the time this error message will make sense. For example, in this case the user has not specified an output file:  


In this case all the user needs to do is close this message and specify an output file and everything will run smoothly.

However, in the rare occasion that we have made an error, a user might see gibberish as Error Message. Something like this:  


As this does not make much sense it means that most likely we have made a mistake. In this case, the user is advised to click on the Error Details button to see more information. This will open the following window:  


Then, the user is advised to click on Save log file… button which will bring up a window to allow the user to save a log file, which contains some debugging information. Then the user can send this information to the leading developer at [nikbpetrov@gmail.com](mailto:nikbpetrov@gmail.com) so that we can investigate further.

Notably, if your input file contains a lot of columns that are not used in the calculations, they will NOT be included in this log file to ensure the privacy of your data. Before sending this file, the user is advised to check it.

# Calculations

## Correlations

Correlations coefficients are calculated by omitting missing data pairwise and then applying either [*pearsonr*](https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.pearsonr.html), [*spearmanr*](https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.spearmanr.html#scipy.stats.spearmanr) or [*kendalltau*](https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kendalltau.html#scipy.stats.kendalltau) functions from *scipy* library.

Note: spearman and kendall have an available argument *nan\_policy=omit* which handles the omission of missing data but pearson does not, so a custom algorithm is used in MUFOS that mimics the behaviour of the *nan\_policy=omit* argument. See [here](https://stackoverflow.com/questions/38894488/dropping-nan-with-pearsons-r-in-scipy-pandas) for more information on this.

Confidence intervals are calculated within MUFOS and output the 95% interval. They require a minimum of 4 observations. Algorithm credits [here](https://zhiyzuo.github.io/Pearson-Correlation-CI-in-Python/).

## Multiple regression

The model is built via the *statsmodels* [OLS](https://www.statsmodels.org/dev/example_formulas.html) function. To get beta values, the data are transformed to z values after which the same OLS function is applied.

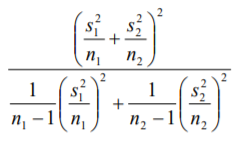
## T-tests

### Raw input

The *researchpy* library is used – documentation [here](https://researchpy.readthedocs.io/en/latest/ttest_documentation.html). For independent samples t-test, equality of variance is calculated via the [*levene*](https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.levene.html) function of the scipy.stats library by dropping any missing data from either pairwise comparison. For paired samples t-test, equality of variance is always assumed to be True.

### Summary data

Degrees of freedom are calculated as the sum of the two group sizes minus 2, if equality of variance is assumed and if not, then the following formula is used:



Test statistics are then calculated via the [*ttest\_ind\_from\_stats*](https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_ind_from_stats.html) function from the *scipy.stats* library.

### Effect sizes

If the t-tests are performed on raw data, then the calculations are done by the [*researchpy*](https://researchpy.readthedocs.io/en/latest/ttest_documentation.html) library.

If the t-tests are performed on summary statistics or SPSS table, then they are calculated as follows:

* Cohen’s d:
* Hedge’s g:

**Note**: Given that Glass's delta require information about means and standard deviations, it can only be performed on raw input data. Glass’s delta is not available on summary statistics due to the difficulty in assigning a control condition.

## Multiple tests corrections

Multiple tests corrections are calculated using the established *statsmodels* library; for details see [here](https://www.statsmodels.org/stable/generated/statsmodels.stats.multitest.multipletests.html). For comparison and performance of these corrections, see [here](http://jpktd.blogspot.com/2013/04/multiple-testing-p-value-corrections-in.html).

# Something went wrong and the documentation does not help. Now what?

If you are receiving a specific error message when performing some operation, please download the Error Details (see [Downloading error details](#_Downloading_error_details) section) and then send the log file to the leading developer at [nikbpetrov@gmail.com](mailto:nikbpetrov@gmail.com). This will also help us improve the software for the community of users!