Table of Contents

[Handling missing and non-numeric data 1](#_Toc46249949)

[Exporting SPSS table from SPSS output. 1](#_Toc46249950)

[Something went wrong and the documentation does not help. Now what? 3](#_Toc46249951)

[Downloading error details 3](#_Toc46249952)

[Calculations 4](#_Toc46249953)

[Correlations 4](#_Toc46249954)

[Multiple regression 4](#_Toc46249955)

[T-tests 4](#_Toc46249956)

[Raw input 4](#_Toc46249957)

[Summary data 4](#_Toc46249958)

[Effect sizes 4](#_Toc46249959)

[Multiple tests corrections 5](#_Toc46249960)

# 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 |

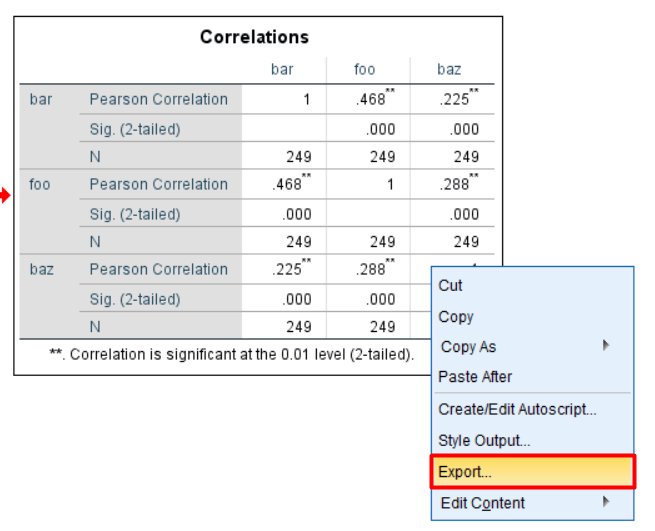
For raw data, you can select to either raise an error message if a column that is expected to contain only numeric data contains non-numeric data, or to ignore the entire row if one value has non-numeric data.

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.

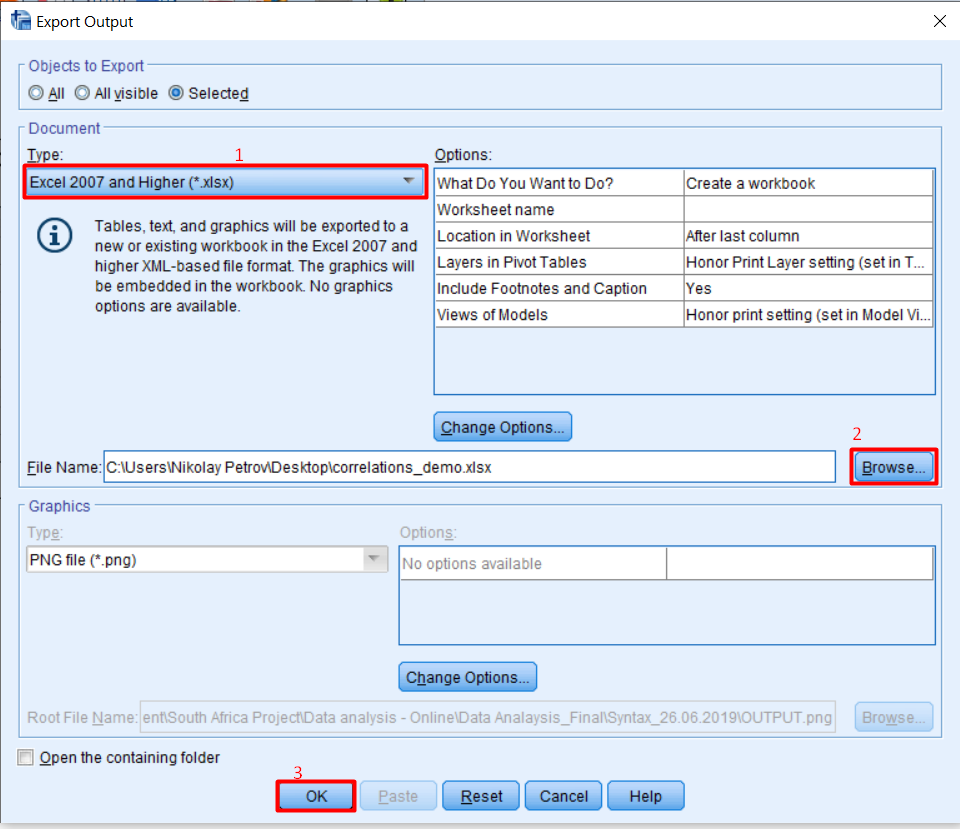
# Exporting SPSS table from SPSS output.

To export any table from 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:



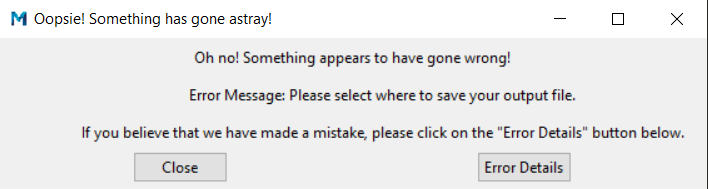
Then under Document, for Type, select “Excel 2007 and Higher (\*.xlsx)”. Then click Browse to select where to export the table. Finally, hit OK:



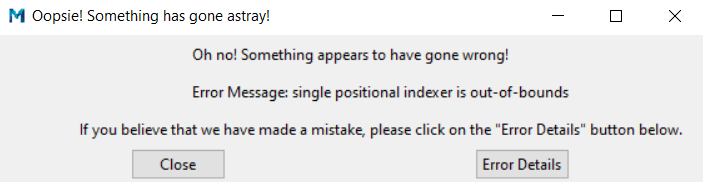
# Something went wrong and the documentation does not help. Now what?

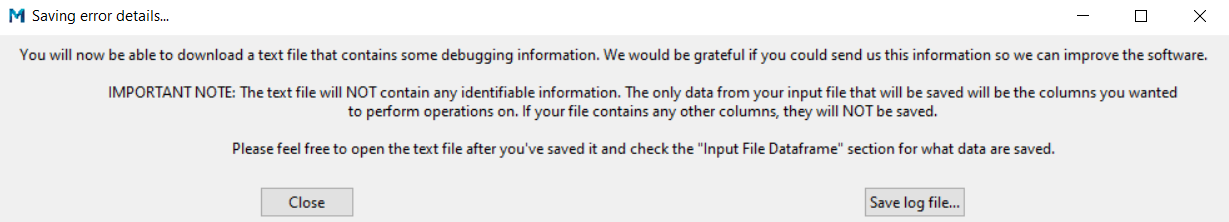
If you are receiving a specific error message when performing some operation, please download the Error Details (see Downloading error details section below) and then send the log file to the leading developer at [nikbpetrov@gmail.com](mailto:nikbpetrov@gmail.com). This will help us improve the software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

# 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, spearmanr or kendalltau functions from scipy.stats

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 MAPAS that mimics the behaviour of the *nan\_policy=omit* argument. See <https://stackoverflow.com/questions/38894488/dropping-nan-with-pearsons-r-in-scipy-pandas> for more information on this.

Confidence intervals are calculated within MAPAS and output the 95% interval. They require a minimum of 4 observations. Algorithm is based on <https://zhiyzuo.github.io/Pearson-Correlation-CI-in-Python/>

## Multiple regression

The model is built via the statsmodels OLS function. To get beta values, the data are transformed to z values after which the same OLS function is applied.

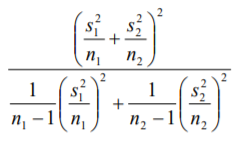
## T-tests

### Raw input

The *researchpy* library is used – documentation here. For independent samples t-test, equality of variance is calculated via the levene 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 function from the scipy.stats library.

### Effect sizes

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

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).

Only Bonferroni and Sidak corrections provide an adjusted alpha threshold level; all other corrections adjust the values based on the alpha.