

Manual

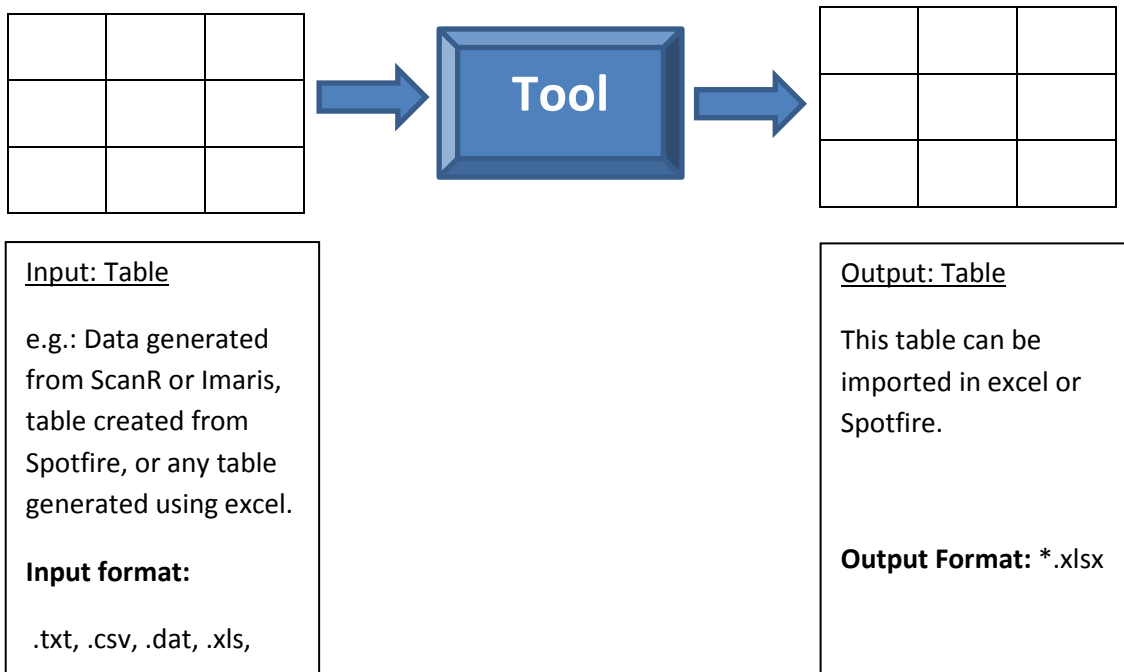
(Sample size equalization, data normalization, and outlier removal tool)

By

Gopal Karemore

What is this tool about?

This tool act as a plugin for data analyzed from ScanR, Spotfire or any Image-data analysis software. It can also work on excel sheet generated by the user.



Front-end of the tool:

Data Normalization and Group Equalization Software (gopal.karemore@cpr.ku.dk)

Data Normalization and Equalization Software

Load file User selected P:\CPRIEqualize_SampleSize\ParameterData_Main.txt

Step 1

Initial filter by Gates

No Gate set	Logical operation
No Gate set	None
No Gate set	None
No Gate set	None

Step 2

Select feature to be used for equalizing

None

Predefined number (optional)

☐ Restrict number of events to or less 100

Virtual replicates

How many replicates? 1

Step 3

Normalization

None

Step 4

Outlier Removal (Grubbs's test)

No

Output file name Processed

Run

Developed by Gopal Karemore (Copenhagen Univ.), 2017 (Academic use only)

Step 1: Load the file that you want to work with by clicking “Load file” button

Load file User selected P:\CPRIEqualize_SampleSize\ParameterData_Main.txt

It will take some time depending on how big is the file.

Step 2: Initial filter by Gates

This step is optional and made specific to data generated from “ScanR” or “Imaris” where user might be interested in applying the Gates and logical combination of Gates.

The screenshot shows a software interface for Step 1, titled "Initial filter by Gates". It contains four dropdown menus for selecting gates: "Nuclei", "G1", "G2", and "No Gate set". To the right of these dropdowns is a section labeled "Logical operation" with three dropdown options: "AND", "OR", and "None". The "AND" option is currently selected.

For example above, user want to apply Gate “Nuclei” which he might have made during ScanR gating to only (AND Operation) select nuclei/cell of his interest. Later he wants to include “G1” and “G2” phase cells be added (OR operation) in the final population.

By default: if you do not need any gating select “None”.

Step 3: Sample size equalization

The screenshot shows a software interface for Step 2, titled "Select feature to be used for equalizing". It features a dropdown menu with "Well" selected, which is highlighted by a red rectangle. Below this dropdown is a section labeled "Predefined number (optional)" containing a checkbox labeled "Restrict number of events to or less" and a text input field with the value "100". At the bottom, there is a section labeled "Virtual replicates" with a text input field labeled "How many replicates?" containing the value "1".

Here user has to select feature by which he wants to equalize his dataset.

For example: in above illustration user selected “Well” as a feature. i.e. Tool will find the well with smallest number of cells within the whole dataset or plate data and

ensure all wells have the equal number of cells. Tool will randomly select cells from wells which have more sample size than the smallest sample size well.

Option 1: As shown below, it is also possible to enforce a predefined sample size to be used for all wells. Here user used 100. i.e. ensuring 100 cells in each well.

Step 2

Select feature to be used for equalizing

Well

Predefined number (optional)

☐ Restrict number of events to or less 100

Virtual replicates

How many replicates? 1

Option 2: As shown below, it also possible to create “Virtual replicates” of equalized samples. i.e. there are multiple ways one can randomly select particular number of cells from large sample size well. If user select more than 1 replicates. Tool will create 2 replicates of same sample size but taken different random cells each time. With this user can check the variability among different wells or conditions.

Step 2

Select feature to be used for equalizing

Well

Predefined number (optional)

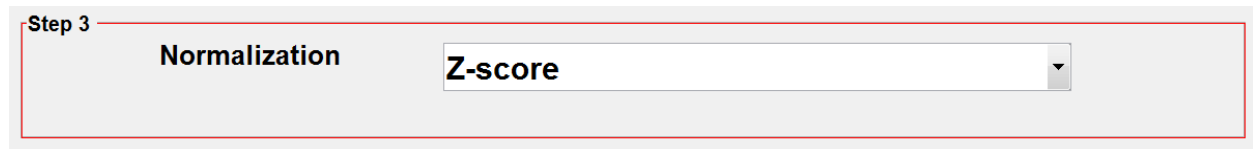
☐ Restrict number of events to or less 100

Virtual replicates

How many replicates? 1

File format support: .txt, .csv, .dat, .xls, .xlsx

Step 4: Data Normalization

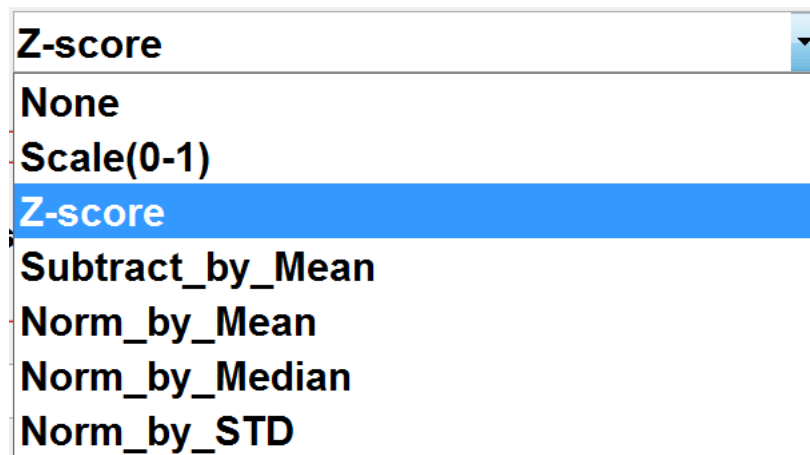


Step 3

Normalization

Z-score

After equalizing dataset, user can normalize the data by one of the following method



Z-score

None

Scale(0-1)

Z-score

Subtract_by_Mean

Norm_by_Mean

Norm_by_Median

Norm_by_STD

If user does not wish to perform normalization select “None” from the pop-up menu bar.

Step 5: Outlier removal

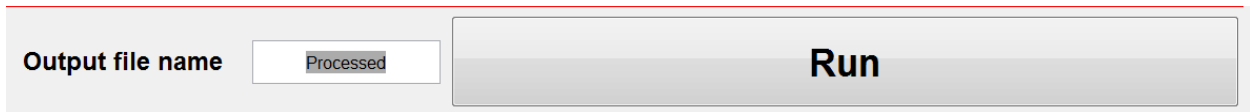
User can remove outlier in the dataset (per feature). This is done by “**Grubb’s test**” with level of significance $\alpha=0.05$.

https://en.wikipedia.org/wiki/Grubbs%27_test_for_outliers

All data entry detected as an outlier will be replaced by Empty cells. In Spotfire or excel these values will be ignored for statistical calculations.

Step 6: Run the tool

Click on “Run” button to process your data. User can give “filename e.g. processed” which will be appended to the original file and output file will be saved in the input file location.



The screenshot shows a web interface with a light gray background. On the left, the text "Output file name" is displayed in a dark gray font. To its right is a white input field with a thin gray border, containing the text "Processed". Further to the right is a large, rectangular button with a light gray gradient and a thin black border, labeled "Run" in a bold, dark gray font.

If you have any question/suggestion about the tool, please contact me at gopal.karemore@cpr.ku.dk

Sincerely,

Gopal

16/06/2017