

Copyright 2013 Biosystems Data Analysis Group - Universiteit van Amsterdam

This is a license to use and modify SOFTWARE & DATA produced by:

THE BIOSYSTEMS DATA ANALYSIS GROUP OF THE UNIVERSITEIT VAN AMSTERDAM

If you use, modify or redistribute the SOFTWARE & DATA and/or your source modifications, you agree:

- i) to use the SOFTWARE & DATA and/or your source modifications solely as part of your research and not in any commercial product;
- ii) that the SOFTWARE & DATA and/or your source modifications will not be distributed for profit;
- iii) all copyright notices and this license note with the SOFTWARE & DATA are retained any redistribution of the SOFTWARE & DATA, or any portion thereof;
- iv) to indemnify, hold harmless, and defend the "Biosystems Data Analysis Group of the Universiteit van Amsterdam" from and against any claims or lawsuits that arise or result from the use of the SOFTWARE & DATA or your source modifications.
- v) to properly reference the SOFTWARE & DATA when used in your research in any publication that may result from that research.

Reserved Rights. The "Biosystems Data Analysis Group of the Universiteit van Amsterdam" retain title and all ownership rights to the SOFTWARE & DATA.

HOW TO RUN A SIMPLIVARIATE ANALYSIS

The zip folder contains 5 Matlab.m files

1. **simply.m**: it is the main routine and contains all the functions needed to run the simplivariate method
2. **CalcW.m**: it is an accessory function needed to generate the reference distribution W needed as input to `simply`. W must be generated before running the `simply` routine. As long as the size of the data set to be analyzed does not change, it is advisable to generate W and save it for further use. `CalcW` takes as input the size of the data set to be analyzed (number of objects x number of variables). Generating W is extremely time consuming due to the large number of iterations needed. The actual number of iterations is set to 1000 but a larger number is recommended, especially in case of very large datasets. For very large data sets parallelization of the code may be a good idea to reduce the running time.
3. **RunSimply.m**: a simple script to run the analysis on the accompanying data set and to save the results in an excel file
4. **Decomp2Excel.m**: a routine to save the results in an excel file. You need to edit it if you look for more than 26 components.
5. **showDecomp.m**: a routine to produce a figure of the components

To run the example unzip the file.zip and all its content in a directory of your choice, make it your Matlab working directory and type **RunSimply** on the Matlab prompt. When **RunSimply** is run on the `DemoColi` data set, it should produce an excel file named **results.xls** containing the same output stored in the file **DemoResults.xls**. When **showDecomp.m** is run it should produce a figure like **FigResults.pdf**.

REFERENCES

[Simplivariate Models: Uncovering the Underlying Biology in Functional Genomics Data](#)

Saccenti E, Westerhuis JA, Smilde AK, van der Werf MJ, Hageman JA, Hendriks MMWB

PloS ONE, Volume 6(6) (2011), article number e20747.

[Tracy-Widom statistic for the largest eigenvalue of autoscaled real matrices](#)

Edoardo Saccenti¹, Age K. Smilde¹, Johan A. Westerhuis¹, Margriet M. W. B. Hendriks

Journal of Chemometrics, Volume 25(12) (2011), pages 644 - 65