Minutes of the 2022 SDC User Group meeting

Date and time: 25th of February 2022, online meeting, 10:00-12:30 C.E.T.

Useful web-pages:

- the SDC User Support:

https://sdctools.github.io/UserSupport/ and https://github.com/sdcTools/UserSupport

- the User Group meetings of 2021 and 2022: https://github.com/sdcTools/UserMeetings (slides of presentations and minutes)

Agenda

- 10:00-10:05 Welcome and introduction

- 10:05-10:25 *R sdc tools*, by Johannes Gussenbauer

- 10:25- 10:45 *Argus tools*, by Peter-Paul de Wolf

- 10:45-11:15 Two feedback *polls* (managed by Violeta Calian, commented by Peter-Paul de Wolf) and Break

- 11:15-12:15 *LP Solvers*, by Peter-Paul de Wolf, Sarah Giessing, Marco Stocchi

- 12:15-12:30 *The rtauargus* package, by Julien Jamme, Nathanael Rastout

Participants

There were 36 present participants out of 41 registered, from the Statistical Institutes of:

Austria, Belgium, Croatia, Czech Republic, Eurostat (in Luxembourg), France, Germany, Hungary, Iceland, Ireland, Italy, Latvia, Malta, Netherlands, Norway, Romania, Slovenia, UK.

R - sdc tools

Presentation

Johannes Gussenbauer gave an overview (co-authored with Alexander Kowarik and Bernhard Meindl) of the main R packages implementing SDC methods for both tabular and micro data, i.e. sdcTable, sdcMicro, recordSwapping, cellKey, ptable, sdcHierarchies, simPop, synthpop, sdcSpatial. He described the progress made in the past year and the future developments planned for the coming months.

Questions and comments

- Jeroen Borghuis initiated a discussion about the testing and use of sdcTable while calling tau-argus via the R-package with commercial LP solvers versus open source solvers. Johannes, Peter-Paul and Jeroen concluded that using tau-argus via sdc-table depends on the purpose/use and it requires writing of additional code in most cases. It was pointed out that one could raise an issue on the github repository to suggest extra functionality for this purpose.

Argus tools

Presentation

Peter-Paul de Wolf presented a brief history of the SDC software, its' evolution, main structure and the newest releases of the argus tools. The future plans on mu-argus and tau-argus, were described in detail.

Questions and comments

- Wim Kloek asked for the precise definition of the infeasible tables. Peter-Paul made clear that the term refers to the mathematical solution infeasibility for given tabular data.

- Timothy Linehan asked whether the 64bit update will cover all the main tau-argus solvers. The answer: it covers all LP-solvers, but at the moment there are still some issues with methods like CTA.

- Julien Jamme initiated a discussion about the methods of improving/organizing software testing in the context of sdc tools. Peter Paul pointed out that this is most useful for developers and it is one of the goals of the current meeting, i.e. to try to find users who are willing to become advanced testers (see next item on the agenda). He also referred to the main methods of testing new versions of a given software: a) using a test data set with known sdc-solution (most likely synthetic for obvious reasons) and b) comparing two software versions on same data, including real life instances if possible.

- Aleksandra Bujnowska suggested that developers could guide the testers towards the specific functionalities/features which need to be verified when a new version is released, for a more efficient testing process.

- Wim Kloek suggested that testers themselves might be specialised.

- Øyvind Langsrud pointed out a solution concerning the situation with zero cells and ones as mentioned in "Tau-Argus-future". This is discussed in the easySdcTable-vignettes:

https://cran.r-project.org/web/packages/easySdcTable/vignettes/singletonsAndZeros.html

Polls

Feedback poll (Poll 1) (30 participants answered)

1. Did you receive the SDC tools support you were hoping for, during the past year?

yes-60%; no-0%; NA-40%

2. Did you know that users help each other in the github environment? Select the methods you have already tested for this purpose (one may select more than one answer):

Issues at UserSupport repository - 45%;

Discussions within User Group team -31%;

User support Wiki - 10%;

NA -41%;

3. Would you like to be part of advanced testing of new releases of SDC tools software?

Yes -21%; No - 17%; Undecided - 62%;

Discussion

A most useful discussion was generated by the last question of the poll, made more precise by Karina Dineen who asked about the level of commitment/time would be required. Peter-Paul estimated that one may see rather few new releases per year, about a month available to get testing results of each such release, but also short time invested in the testing process itself.

Several users expressed interest in taking part in such a process and all concluded with the idea to create a new team of users for this purpose. This team has been defined in the few days after the meeting, after a number of emails have been received by the User Support from participants as "@sdcTools/testers" which is a sub-team of the "@sdcTools/user-group" team.

Users' proposals poll (Poll2)(8 participants answered)

1.How could we make it easier for the users to share experience and knowledge ?(Long Answer)

4/8 (50%) answered

- 1. probably using also other tools such as a dedicated chat (e.g "discord" server or reddit) can help
- 2. create test cases that can be shared/reproduced
- Free open forum for discussion. Not requireing registration, logging etc.
- meeting on thematic topics as how to treat non nested hierarchies in order to share experiences and solutions with each other
- Happy with current system

2.Would you like to propose some future developments of the SDC tools? Which?(Long Answer)

7/8 (88%) answered

- having our solution for synthetic data generation / utility and privacy
- sdcTable, adding apriori possibilites so that it's possible to generate apriori files after using sdcTable as a wrapper for tau-argus
- It's already been raised - improved error handling
- NA
- Handling of frequency counts tables in the tabular tools, especially re. suppression of zeros
- Especially saving and reading grc files. Also keeping the status of cells in case of non-additive tabular data. If possible, linked protection of tables with non-nested hierarchies.
- Group attribute disclosure

3.Would you like to propose a specific topic to be addressed during the next User Group meeting?

Which?(Long Answer)

3/8 (38%) answered

- Possible solutions for linked protection of tables with non-nested hierarchies

- NA
- No

The software developers and the user support group will take these suggestions into account and participants are invited to send more ideas at any point in time via the usual discussions/issue github channels of sdc-tools.

LP solvers

Presentation

- Peter-Paul de Wolf reviewed the LP-solver role in the context of solving MILP/LP in tau-argus. He described the experience of Statistics Netherlands regarding commercial and open source solvers, advantages of each type, practical setup at SN and the type of license used.

- Sarah Giessing described the tau-argus production infrastructure at Destatis. She also explained the main LP-solver issues, setup, license types needed and the way all these are solved at Destatis.

- Marco Stocchi presented the setup and LP-solvers technical and license solutions implemented at Eurostat.

Questions and comments

- Janis Jukams asked about the use of VM at Statistics Netherlands, i.e. whether restricted to SDC purposes or more extended, and Peter-Paul confirmed that VM are used for all processes.

- Aleksandra Bujnowska requested a few more details about the batch file running options which were provided.

The rtauargus package

Presentation

Julien Jamme and Nathanael Rastout presented their work co-authored with Clément Guillo, about a new R-package which makes it easy to prepare correctly formatted micro/tab data and the batch file for running tau-argus. The github page of the package is https://github.com/InseeFrLab/rtauargus and it is documented in French. Authors described their plans of moving to CRAN and finalising the English documentation in the near future.

Questions and comments

- Aleksandra Bujnowska suggested that a link to rtauargus from the sdc tools github page could be provided

- Manca Golmajer asked about the tau-argus version used by the current rtauargus package

- Sara Piombo asked whether the authors are planning to work on an rmuargus project. This however is not applicable since muargus does not have batch running.

- Marieke de Vries expressed interest in having the English documentation for this package. She also asked whether it brings new functionality to the standard tau-argus. The answer is that rtuargus can only do what tau-argus is able to, since it is mainly designed to make the R-based production work smoothly with this software. However, in addition to the features of the sdc-table package for instance (which can also run tau-argus via R), rtauargus provides an efficient way to prepare holdings and apriori files.

The meeting ended with hopeful thoughts for future interaction and collaboration.