

gosset: Tools for Data Analysis in Experimental Agriculture

Kauê de Sousa^{1,2}, Jacob van Etten¹, David Brown^{3,4}, Jonathan Steinke^{1,5}

Abstract

Appropriate data management and analysis are necessary to produce practical information from agronomy and crop science experimental data. There is also an ongoing trend advocating for programmatic tools that supports reproducible workflows in scientific research. Recently developed approaches for data analysis and synthesis based on ranking data lack of customized analytical software to facilitate data science workflows. We present the R package gosset, which provides functions to support data preparation, modelling, validation, and results presentation with rank-based models, such as Plackett-Luce and Bradley-Terry. We demonstrate the functionality of the package with a case of on-farm evaluations of common bean (*Phaseolus vulgaris* L.) varieties in Nicaragua.

data-driven agriculture, Plackett-Luce model

Required Metadata

Current code version

Ancillary data table required for subversion of the codebase. Kindly replace examples in right column with the correct information about your current code, and leave the left column as it is.

Table 1: *Code metadata (mandatory)*

Nr.	Code metadata description	Please fill in this column
C1	Current code version	0.4.003
C2	Permanent link to code/repository used for this code version	https://github.com/AgrDataSci/gosset
C3	Code Ocean compute capsule	
C4	Legal Code License	MIT
C5	Code versioning system used	Git
C6	Software code languages, tools, and services used	R
C7	Compilation requirements, operating environments & dependencies	
C8	If available Link to developer documentation/manual	https://agrdatasci.github.io/gosset/
C9	Support email for questions	k.desousa@cgiar.org

The permanent link to code/repository or the zip archive should include the following requirements: README.txt and LICENSE.txt.

Source code in a src/ directory, not the root of the repository.

Tag corresponding with the version of the software that is reviewed.

Documentation in the repository in a docs/ directory, and/or READMEs, as appropriate.

Motivation and significance

Introduce the scientific background and the motivation for developing the software.

Explain why the software is important, and describe the exact (scientific) problem(s) it solves.

Indicate in what way the software has contributed (or how it will contribute in the future) to the process of scientific discovery; if available, this is to be supported by citing a research paper using the software.

Provide a description of the experimental setting (how does the user use the software?).

Introduce related work in literature (cite or list algorithms used, other software etc.).

Software description

Describe the software in as much as is necessary to establish a vocabulary needed to explain its impact.

The R package gosset provides functionality supporting the data analysis workflows in experimental agriculture, especially with rank-based approaches in on-farm experimental agriculture. Typically, this data analysis workflow includes (1) data management and preparation, (2) modelling and (3) results visualization and presentation.

Software Architecture

Give a short overview of the overall software architecture; provide a pictorial component overview or similar (if possible). If necessary provide implementation details.

Software Functionalities

Present the major functionalities of the software.

Data management and preparation

Sample code snippets analysis (optional)

Illustrative Examples

Provide at least one illustrative example to demonstrate the major functions.

Optional: you may include one explanatory video that will appear next to your article, in the right hand side panel. (Please upload any video as a single supplementary file with your article. Only one MP4 formatted, with 50MB maximum size, video is possible per article. Recommended video dimensions are 640 x 480 at a maximum of 30 frames/second. Prior to submission please test and validate your .mp4 file at [http : //elsevier – apps.sciverse.com/GadgetVideoPodcastPlayerWeb/verification](http://elsevier-apps.sciverse.com/GadgetVideoPodcastPlayerWeb/verification). This tool will display your video exactly in the same way as it will appear on ScienceDirect.).

Impact

This is the main section of the article and the reviewers weight the description here appropriately

Indicate in what way new research questions can be pursued as a result of the software (if any).

Indicate in what way, and to what extent, the pursuit of existing research questions is improved (if so).

Indicate in what way the software has changed the daily practice of its users (if so).

Indicate how widespread the use of the software is within and outside the intended user group.

Indicate in what way the software is used in commercial settings and/or how it led to the creation of spin-off companies (if so).

Conclusions

Set out the conclusion of this original software publication.

Conflict of Interest

No conflict of interest exists: We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Acknowledgements

00

Please add the reference to the software repository if DOI for software is available.

Current executable software version

Ancillary data table required for sub version of the executable software: (x.1, x.2 etc.) kindly replace examples in right column with the correct information about your executables, and leave the left column as it is.

Table 2: Software metadata (optional)

Nr.	(Executable) software metadata description	Please fill in this column
S1	Current software version	For example 1.1, 2.4 etc.
S2	Permanent link to executables of this version	For example: <i>https : //github.com/combogenomics/ DuctApe/releases/tag/DuctApe – 0.16.4</i>
S3	Legal Software License	List one of the approved licenses
S4	Computing platforms/Operating Systems	For example Android, BSD, iOS, Linux, OS X, Microsoft Windows, Unix-like , IBM z/OS, distributed/web based etc.
S5	Installation requirements & dependencies	

Nr.	(Executable) software metadata description	Please fill in this column
S6	If available, link to user manual - if formally published include a reference to the publication in the reference list	For example: <i>http : //mozart.github.io/documentation/</i>
S7	Support email for questions	