

A Compendium Platform for Reproducible, R-based Research with a focus on Statistics Education

Patrick Wessa

March 30, 2008

This paper discusses a new Compendium Platform (CP) that allows us to create Reproducible Research in R that is easily accessible for anyone who has access to the internet (freestatistics.org). The platform is based on the R Framework (wessa.net) and primarily focuses on ICT-based Statistics Education within a pedagogical paradigm of individual and social constructivism which received a great deal of interest in the academic community (Von Glasersfeld (1987), Erick Smith (1999), Eggen and Kauchak (2001), and Nyaradzo Mvududu (2003)).

The basic idea is to create an environment where students are allowed to interact with each other (and the tutor) about a series of research-related activities (such as assignments or workshops) based on the R language and the R Framework. The novelty about this approach lies in the fact that the newly developed CP empowers students to easily archive, exchange, reproduce, and reuse R computations. The underlying technology facilitates the creation of a learning environment that supports social constructivism which is very similar to the real world of applied statistical research. More importantly, the CP allows us to obtain physical measurements of the actual learning process of students based on detailed information about the use of the statistical software, and the socially constructivist learning activities (based on peer review of statistical analysis in R).

The CP was thoroughly tested in two undergraduate statistics courses with large student populations. During these courses a large number of physical and survey-based measurements were obtained and studied. The preliminary analysis of the relationships between learning attitudes, social interaction (through group work and peer review), learning experiences, software usability, usage of archived R computations, and exam scores (that are related to statistical competences rather than knowledge) is presented.

One of the most interesting results is that social interaction through peer review based on Reproducible Research (which is used as a "learning activity" rather than an "evaluation tool") is very beneficial for the learning experiences of students, and exam scores. Also, there is a strong, positive relationship between the use of the CP and exam performance - even if other important factors are taken into account. Another interesting result is that a large majority of students have a positive perception about the new system as a learning tool and prefer the constructivist approach based on Compendia above traditional

learning methods.

In addition, it is (very) briefly illustrated how Compendia of Reproducible Research can be used to:

- write Compendium-based course materials
- detect plagiarism and free-riding
- quickly identify (and find solutions for) bugs and computing-related problems
- estimate the workload of an assignment
- support new forms of collaboration that lead to improved solutions in R

Finally, some important aspects about the near and distant future of the CP and the underlying R Framework (for the purpose of education, scientific research, and publishing) are illustrated and discussed.

Acknowledgements

This research is funded by the KULeuven Association (OOF 2007/13 Project)