

Can Quarto Documents and the Checkpoint package help us achieve reproducibility?

Henriette Hansen

Abstract

Reproducibility and replication of studies and experiments are an important approach to gain confidence in conclusions. Quarto Documents and the Checkpoint package helps us achieve reproducibility by sharing knowledge and techniques used in procedures of studies more widespread, which contributes positive for generalizability and robustness in procedures and framework of different studies and experiments.

Introduction

The effect of reproducibility is briefly explained and studied towards the use of a Quarto Document and the Checkpoint package. By looking at replicability the more data and information given about the procedures of an experiment, give better reproducibility. To understand if the Quarto Document and Checkpoint package helps to achieve reproducibility it is discussed in the manner of generalizability, robustness and the Comprehensive R Archive Network (CRAN), which the latter is a part of the Checkpoint package.

The literature review and the discussion gives an unequivocal result of the problem studied.

Reproducibility

Science advances on a foundation of discoveries. To gain confidence in a conclusion, reproducing an experiment is often used. To confirm results and conclusions from one study, the approach of replicability is often used. Replicability is the confirmation of results and conclusions from one study which is obtained independently with new data but the same method. This is considered the scientific gold standard.

Reproducibility methods refers to the provision of enough detail about study and data so the same procedures could, in theory be repeated. Furthermore, obtaining the same results from an independent study whose procedures are as closely replicated to the original experiment as possible.

Robustness and generalizability can be used in lieu of the term reproducibility. Robustness refers to the stability of experimental conclusions to variations in either baseline assumptions or experimental procedures. Generalizability refers to the persistence of an effect in procedures different from outside of an experimental framework.

Quarto document and Checkpoint package

The goal of a Quarto document is to make scientific and technical documents better by creating and collaborating with a writing and publishing environment with integrated tools for technical content.

Quarto aims to make it easier to adopt a reproducible workflow, overarching to replicate research and publications for a confirmation of conclusions obtained independently in different studies. Enhancing reproducibility requires that the code and data to create a procedure is an integrated part of the research.

The `checkpoint()` function aims to solve the problem of package reproducibility in R. Checkpoint allows to install packages as they existed on Comprehensive R Archive Network (CRAN) on a specific date like a CRAN time machine (“Checkpoint: Configures R Session to Use Packages as They Existed on CRAN... In Checkpoint: Install Packages from Snapshots on the Checkpoint Server for Reproducibility” n.d.). CRAN contains an archive of different versions of the R distribution, documentation, and contributed R packages (*The Reproducible r Toolkit (RRT)* n.d.).

To achieve reproducibility, the checkpoint function installs the packages required or called by the project and scripts to a local library exactly as they existed at the specific point in time. Only those packages are available to a specified project, thereby avoiding any package updates that came later which could have altered any results.

Discussion

Research results from one study cannot always be verified whether they will be correct or will remain correct as new information and new research becomes available (Peng 2011). Therefore, reproducibility is key to help con-

tribute to other and future research and verifying results. Providing transparency to the experiment will help others to understand the procedure of the research. Replicating experiments and achieving the same results and conclusions in one study obtained independently in another adds confirmation the conclusion. Furthermore, reproducibility depends on enough details and data to be replicated. Tracking and recording changes during experimentation could thereby have significant value for reproducibility. By adding new datasets, the changes could affect the outcome of a model and results. Therefore, different dataset versions and changes should be logged to achieve better reproducibility.

The Quarto Document’s overarching goal is to make it easier to adopt a reproducible workflow and using the Checkpoint package to this helps to keep track of the different R distribution versions, documentation, and contributed R packages (“Quarto” n.d.).

Reproducibility should be positively affected using Quarto Document and the additional Checkpoint package whereas the CRAN archive contributes by avoiding package updates which could alter research study results, hereby considering replication of experiments and the specific procedures executed to achieve the reported results. In addition, the Quarto Document aims to enhance a collaborating publishing environment by providing tools for technical content which in term could be positive for generalizability of an effect in procedures. Also, the stability of experimental conclusions and variations of procedures could help achieve better reproducibility.

Conclusion

The overarching goal of the Quarto Document and Checkpoint package is to enhance reproducibility by making knowledge and techniques used in procedures of studies more widespread. As discussed, reproducibility depends on, and is better replicated the more details and information given regarding procedures and framework of an experiment.

The conclusion of the problem is therefore that Quarto Document and Checkpoint package does indeed contribute to help achieve reproducibility.

References

“Checkpoint: Configures R Session to Use Packages as They Existed on CRAN... In Checkpoint: Install Packages from Snapshots on the Checkpoint Server for Reproducibility.” n.d. Accessed October 3, 2022. <https://rdr.io/cran/checkpoint/man/checkpoint.html>.

Peng, Roger D. 2011. “Reproducible Research in Computational Science.”
Science 334 (6060): 1226–27. <https://doi.org/10.1126/science.1213847>.
“Quarto.” n.d. Accessed October 3, 2022. <https://quarto.org/>.
The Reproducible r Toolkit (RRT). n.d. Accessed October 3, 2022. <https://cran.r-project.org/web/packages/checkpoint/vignettes/checkpoint.html>.

Appendix

Put in mine `sessionInfo()` since it was missing.

```
sessionInfo()

R version 4.2.1 (2022-06-23)
Platform: aarch64-apple-darwin20 (64-bit)
Running under: macOS Monterey 12.5.1
Matrix products: default
LAPACK: /Library/Frameworks/R.framework/Versions/4.2-arm64/
Resources/lib/libRlapack.dylib
locale:
 [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
attached base packages:
 [1] stats      graphics  grDevices  utils      datasets  methods   base
loaded via a namespace (and not attached):
 [1] compiler_4.2.1 tools_4.2.1  knitr_1.39   xfun_0.32
```