

Reproducible Research

Mark Agerton

2019 September 23

New AER Reproducibility Guidelines

- AER's [Data and Code Availability Policy](#)

For econometric, simulation, and experimental papers, the replication materials shall include

- a) the data set(s),
- b) the programs used to create any final and analysis data sets from raw data, c programs used to run the final models, and
- c) description sufficient to allow all programs to be run.

- Social Science Data Editors [Verification guidance](#) and [Replication Template](#)
- “Top 10 list of papers on reproducible research” (Barba [2017](#))

Ten Simple Rules for Reproducible Computational Research

1. For Every Result, Keep Track of How It Was Produced
2. Avoid Manual Data Manipulation Steps
3. Archive the Exact Versions of All External Programs Used
4. Version Control All Custom Scripts
5. Record All Intermediate Results, When Possible in Standardized Formats
6. For Analyses That Include Randomness, Note Underlying Random Seeds
7. Always Store Raw Data behind Plots
8. Generate Hierarchical Analysis Output, Allowing Layers of Increasing Detail to Be Inspected
9. Connect Textual Statements to Underlying Results
10. Provide Public Access to Scripts, Runs, and Results

(Sandve et al. [2013](#))

1. For Every Result, Keep Track of How It Was Produced
 - Notion of “workflow”... from raw data to final table
2. Avoid Manual Data Manipulation Steps

- Excel is evil
- 3. Archive the Exact Versions of All External Programs Used
 - Python virtualenv, R's packrat, Julia Pkg all help here
 - Also hardware
- 4. Version Control All Custom Scripts
 - Git!!!
- 5. Record All Intermediate Results, When Possible in Standardized Formats
 - Show my example from SLURM
- 6. For Analyses That Include Randomness, Note Underlying Random Seeds
 - duh
- 7. Always Store Raw Data behind Plots
 - Plots & tables should be generated from files
- 8. Generate Hierarchical Analysis Output, Allowing Layers of Increasing Detail to Be Inspected
 - ??
- 9. Connect Textual Statements to Underlying Results
 - For each statement, where did it come from? Cite the source or note your data
- 10. Provide Public Access to Scripts, Runs, and Results
 - Github?
 - AER data site

References

- Barba, Lorena A. 2017. "Barba-Group Reproducibility Syllabus." doi:[10.6084/m9.figshare.4879928.v1](https://doi.org/10.6084/m9.figshare.4879928.v1).
- Sandve, Geir Kjetil, Anton Nekrutenko, James Taylor, and Eivind Hovig. 2013. "Ten Simple Rules for Reproducible Computational Research." *PLOS Computational Biology* 9 (10): e1003285. doi:[10.1371/journal.pcbi.1003285](https://doi.org/10.1371/journal.pcbi.1003285).