Newcastle University

Open Research Case Study

Teaching quantitative methods reproducibly

Chris Moreh

Lecturer in Sociology, School of Geography, Politics and Sociology.

Introduction

Sociology is lagging behind other social science disciplines – such as political science or economics – when it comes to taking replicability standards seriously (1,2). Among the more benign reasons behind this is that quantitative sociologists rarely use experimental methods (3) and mostly rely on large-scale survey data to target ill-defined estimands that generally result in low p-values (and low effect sizes). As many have argued, that's not ideal for the purposes of scientific advancement (4–6). One reason for this, which applies to other (social) sciences as well, is the lack of methodological socialisation in replicability practices. Potential solutions have long been identified and implemented in some trailblazing centres of excellence – like Harvard's Institute for Quantitative Social Science (7) – but have only recently started gaining wider adoption through teaching reproducibility practices as part of applied research methods training at both undergraduate (8,9) and postgraduate levels (10). In this case study I present one of my own contributions to placing replicability practices at the heart of quantitative methodology teaching.

The Context

When I joined Newcastle University in November 2021, I had recently developed an MA-level *Quantitative Research Methods* course at my previous institution but could only teach it once. The idea behind it was simple, albeit unorthodox: take students on a journey across one popular research theme – why not "social trust"? – deconstructing the data and methods underpinning selected articles and incrementally reproducing (parts of) the original analyses.

At Newcastle, I applied this idea in my undergraduate module on Researching Social Life, and later in my Faculty-level PGR course on Quantitative Analysis. I will focus on the former, which, being an introductory module, presented more elementary challenges.

Open Practices

Open research depends on public data, free software, and reproducible code. Researching Social Life introduced all three from the start. I carefully selected articles on 'social trust' that used public secondary survey data. These served as background literature, and in computer workshops students worked on exercises using the original survey data than underpinned them. We used *R* in *RStudio*, aided by the {tidyverse} and {easystats} metapackages that implement more coherent and human-readable code dialects. Students worked in Quarto documents, keeping their code together with narrative interpretations.

I had one conundrum: time for developing data-wrangling skills was limited, *but* pre-processed teaching data obfuscates the link and discrepancies between the raw data and the data presented in publications, even within replication packages.

My solution was to build a GitHub Pages website for the course, with a *Data* page that provides documented code used to shape raw data into usable teaching datasets, as well as additional functionality to explore the codebooks. The datasets themselves are stored on Canvas because the freely available raw data often require user registration with the survey providers for usage

monitoring purposes, but the entire codebase is publicly available on GitHub, allowing anyone to use it to shape the raw survey datasets into a usable format.

Benefits and beneficiaries

Students taking the course gained a clearer understanding of the research process. By replicating bits of published results, they witnessed the mechanics behind he methods, but they also gained confidence in their own abilities. They could learn how basic methods relate to the more complex analyses reported in published research. Many have struggled, but they have struggled with a purpose and were supported. Receiving two student nominations for The Education Awards 2024was an unexpectedly pleasant surprise.

Making the data code publicly available ensures that students can use it in the future too if they choose similar data for their dissertations or job application presentation. But it is an open resource for anyone starting out with data analysis, who wonders what to do with the raw data from commonly used large-scale cross-national datasets.

Challenges and opportunities

Developing statistical understanding, coding skills and reproducible workflows in tandem is a huge challenge. But keeping them apart precludes understanding of applied research and reinforces questionable research practices. There are understandable practical reasons for avoiding teaching introductory-level statistics reproducibly. However, to guarantee the credibility of the next generation of social science research, we must face the challenge.

Lessons and Conclusion

By implementing and expecting open-data and reproducible practices in my teaching I have learned as much about the difficulties of open research as my students have. But it has made me more determined to follow reproducible and public workflows in my own research, and to keep searching for more efficient and painless ways of teaching them as early as possible.

References

1. Freese J. Replication Standards for Quantitative Social Science: Why Not Sociology? Sociol Methods Res. 2007 Nov 1;36(2):153–72.

2. Freese J, Peterson D. Replication for Quantitative Research. In: Elman C, Mahoney J, Gerring J, editors. The Production of Knowledge: Enhancing Progress in Social Science [Internet]. Cambridge: Cambridge University Press; 2020 [cited 2021 Oct 1]. p. 267–83. (Strategies for Social Inquiry). Available from: https://www.cambridge.org/core/books/production-of-knowledge/replication-for-quantitative-research/8E31F842EFAAF65026F446F74DA6662D

3. Gërxhani K, Miller L. Experimental sociology. In: Gërxhani K, de Graaf ND, Raub W, editors. Handbook of sociological science: contribu□ons to rigorous sociology. Cheltenham, UK ; Northmapton, MA, USA: Edward Elgar Publishing; 2022. p. 309–23. (Research handbooks in sociology series).

4. Lundberg I, Johnson R, Stewart BM. What Is Your Estimand? Defining the Target Quantity Connects Statistical Evidence to Theory. Am Sociol Rev. 2021 Jun 1;86(3):532–65.

5. Smaldino PE, McElreath R. The Natural Selection of Bad Science. R Soc Open Sci. 2016 Sep;3(9):160384.

Copyright: The authors (2024). CC BY