orgtheory.net

replication and the future of sociology

Consider the following:

- It was discovered that only about a quarter of sociologists are able, permitted, or willing to provide replication materials for third parties.
- Andrew Gelman had major a statistical issue with an article in the *American Sociological Review*. The data bank was unable to release the data to him for an unstated reason. The *ASR* also refused to published a comment, despite positive peer review.
- *On The Run*, one of the most influential ethnographic studies in recent years, is based on research where the field notes were burned, the survey was burned, and the dissertation was embargoed for years.

Sociology, we can do better. Here is what I suggest:

- Dissertation advisers should insist on some sort of storage of data and code for students. For those working with standard data like GSS or Ad Health, this should be easy. For others, some version of the data should accompany the code. There are ways of anonymizing data, or people can sign non-disclosure forms. Perhaps universities can create digital archives of dissertation data, like they have paper copies of dissertations. Secure servers can hold relevant field notes and interview transcripts.
- Journals and book publishers should require quant papers to have replication packages. Qualitative paper authors should be willing to provide complete information for archival work & transcription samples for interview based research. The jury is still out on what ethnographers might provide.
- IRB's should allow all authors to come up with a version of the data that others might read or consult.
- Professional awards should only be given to research that can be replicated in some fashion. E.g., as Phil Cohen has argued – no dissertation awards should be given for dissertations that were not deposited in the library.

Let's try to improve.

50+ chapters of grad skool advice goodness: Grad Skool Rulz ($2!!!!)/From Black Power/Party in the Street

Written by fabiorojas

August 17, 2015 at 12:01 am

Posted in <u>epistemology and methods</u>, <u>fabio</u>, <u>sociology</u>

3 Responses

I'm curious to hear your thoughts on providing raw data. For example, many AER replication files I've seen do not include the various original variables used to construct the key measure/independent variable. Thus, I can run the models and reproduce the tables, but I can't really see how the measure was constructed, and how its construction might influence the results. The original variables would also be necessary to evaluate how other data issues were handled– for example, how the authors dealt with inconsistencies in industry classifications across cases.

Lori

August 17, 2015 at <u>3:42 pm</u>

Lori: I think we can start with some simple rules. If your data is public, you should be able to explain how variables are made and provide raw data. If you are "done" with data (e.g, you collected it and no longer need it to be private) then everything should be up.

<u>fabiorojas</u>

August 17, 2015 at <u>4:02 pm</u>

While I think transparency is important, I think requiring all awards needing to be based upon research that is able to be replicated is not really realistic for research that isn't quantitative or at least quasi-experimental. Ethnographic and phenomenological research really doesn't lend itself to replication. I can't see how William Whytes's street corner research can be replicated – at least replicated in the general sense that it is understood.

While I've always had a strong quantitative approach to research, my dissertation was a qualitative study. In addition to the transparency in my Methods chapter, I kept a long and detailed audit trail to show wherever I made an "executive decision," as well as a coding manual in the appendix.

Paul

August 23, 2015 at <u>10:56 pm</u>

Comments are closed.