

the replication principle

adam okulicz-kozaryn

`adam.okulicz.kozaryn@gmail.com`

this version: Monday 4th September, 2023 18:34

outline

the idea

get code from others!

outline

the idea

get code from others!

bad excel

- ◇ never trust numbers that come from excel
 - no way to find out what happened, there's no code!
[*]http://www.texasoft.com/excel/Should_You_Use_Excel_for_Statistics.pdf
[*]<http://andrewgelman.com/2013/04/17/excel-bashing/>
- ◇ I learned it hard way:
 - my first paper for ecological economics, done in excel
 - reviewers got back after 6mo, i had dozens of excel files
 - couldn't replicate my own results!
- ◇ “Talk is cheap. Show me the code” –Linus

replication, replication

- ◇ replication=write computer code that will do
everything
 - from raw data (eg FED, IMF) to results (eg regression)
- ◇ necessary for science
- ◇ otherwise we don't know what happened
- ◇ how was it calculated? is there a mistake? who knows?
- ◇ pol sci perspective
[*]<http://gking.harvard.edu/files/gking/files/replication.pdf>
- ◇ IT perspective [*]<http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001745>

humans and mistakes

- ◇ a part of human nature is that we make mistakes
 - can't avoid it no matter what's your skills, experience, etc.
 - same pertains to academic research
- ◇ computers, on the other hand, never make mistakes
 - they just do whatever humans tell them to do
 - sometimes they execute our mistakes

rules for everyday practice [revisit/stress later!!]

- ◇ once you have coded everything, double/triple-check it
 - leave it aside and check again
 - show it to other people, post on your website
- ◇ cross-check end output with raw data—e.g. are there the same numbers for randomly chosen data points— does it make sense?
- ◇ check with alt data ? they tell the same story?
 - i always google tables/graphs of what i study
- ◇ everything has been already studied by others
- ◇ like lit rev, its data rev

outline

the idea

get code from others!

the easiest way to do research in 21st century

- ◇ start with code others wrote, and build on their work
- ◇ this is the fastest, most efficient way to do research
- ◇ any research very close to yours, just email author and ask her to share code with you
- ◇ even if it sas or spss etc—you'll be able to figure it out quickly what is going on there and then implement something similar in stata
- ◇ don't reinvent the wheel: almost as if you were to start research without reading literature and had to come up with all theories and ideas on your own!