

# the replication principle

adam okulicz-kozaryn

`adam.okulicz.kozaryn@gmail.com`

this version: Tuesday 21<sup>st</sup> January, 2025 16:32

## 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://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

- ◇ we make mistakes, its human nature, can't avoid it!
  - no mater your skills, experience, etc
  - same pertains to academic research
- ◇ computers never make mistakes

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

- ◇ once 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—eg randomly chosen data points; does it make sense?
- ◇ check with alt data? same story?
  - google tables/graphs of what you 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
- ◇ fastest most efficient way to do research
- ◇ just email author and ask for code
- ◇ even if sas, spss, etc—easy to figure out what's going on and do sth something similar in python