



27

PERSPECTIVE

Reproducible Research in Computational Science

Roger D. Peng
+ See all authors and affiliations

Open Science 2015, 1(1):1-11
DOI: 10.12688/OS.12140.1

Not reproducible ← Gold standard

28

Why do we need to reproduce the computational environment?

Quite often analysis code "breaks" - often in one of two ways:

Code that worked previously now doesn't - maybe a function in an R package was updated (e.g., `lmtest::bptest` became `lmtest::bptest2` so old code using `lmtest::bptest` wouldn't now run).

Code that worked previously still works - but produces a slightly different result or now throws a warning where it didn't previously (e.g., convergence/singular fit warnings in `lm` version 1.1-19 vs. version 1.1-20).

When R moved from version 3.5 to 3.6, the way in which random numbers were generated using the `sample()` function changed - so even with the same randomisation seed, different random numbers were produced.

Much Twitter confusion ensued!

29

Capturing your local computational environment

- You need to capture the versions of your software packages (plus their dependencies incl. system-level ones).
- This may sound trivial but trying running some old analysis code and be amazed at how many things now don't work as they once did!

30

Binder to the rescue!

Turn a Git repo into a collection of interactive notebooks

Have a repository full of Jupyter notebooks? Then Binder, open these notebooks in an executable environment, making your code reproducibly reproducible by anyone, anywhere.

31

The Turing Way

The Turing Way - A handbook for reproducible data science

Developing a handbook for best practice in reproducible data science

<https://www.turingway.org.uk/research-reproducible-research-turing-way-handbook-reproducible-data-science>

32

Summary

- Open data and open code are the future - remember to make your data FAIR and add a license to both your data and code.
- Pre-registration allows you to capture your predictions - time-stamped at a point in time so when you come to write up your work you can use the pre-reg as evidence that you really did make your predictions before data collection commenced.
- In many cases conducting research in a reproducible manner is easy - it requires a bit of planning and organisation up-front, but the pay-off is huge.
- Not only will others be able to reproduce your results, but so will you at some future point in time.
- Working in an open and reproducible manner also makes large-scale collaborations easier - with the extra computational skills that you acquire, you'll be a more effective researcher.

Summary

- Open data and open code are the future - remember to make your data FAIR and add a license to both your data and code.
- Pre-registration allows you to capture your predictions - time-stamped at a point in time so when you come to write up your work you can use the pre-reg as evidence that you really did make your predictions before data collection commenced.
- In many cases conducting research in a reproducible manner is easy - it requires a bit of planning and organisation up-front, but the pay-off is huge.
- Not only will others be able to reproduce your results, but so will you at some future point in time.
- Working in an open and reproducible manner also makes large-scale collaborations easier - with the extra computational skills that you acquire, you'll be a more effective researcher.

Click to add speaker notes

