Reproducibility up front

Matthew Brett

The cast

- Jarrod Millman
- Jean-Baptiste Poline
- Stéfan van der Walt
- Paul Ivanov

Replication and reproducibility

Computational reproducibility:

> *An article about computational science in a scientific publication is not the scholarship itself, it is merely* ***advertising*** *of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures.*

Buckheit and Donoho (1995) "WaveLab and Repro ducible Research".

Replication - where are we? Position 1

Letter | Published: 27 August 2018

Evaluating the replicability of social science experiments in *Nature* and *Science* between 2010 and 2015

Colin F. Camerer, Anna Dreber, Felix Holzmeister, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Gideon Nave, Brian A. Nosek, Thomas Pfeiffer, Adam Altmejd, Nick Buttrick, Taizan Chan, Yiling Chen, Eskil Forsell, Anup Gampa, Emma Heikensten, Lily Hummer, Taisuke Imai, Siri Isaksson, Dylan Manfredi, Julia Rose, Eric-Jan Wagenmakers & Hang Wu

Nature Human Behaviour **2**, 637–644 (2018) | Download Citation

Replication - where are we? Position 2

TABLE 1. Large-scale Efforts to Massively Replicate Reported Candidate-gene Associations[a]

First Author	Disease/Phenotype	Gene Loci Tested	Sample Size (Design)	Replicated Gene Loci[b]
Bosker et al[15]	Major depressive disorder	57	3540 (case-control)	1
Caporaso et al[16]	Smoking (7 phenotypes)	359	4611 (cohort[c])	1
Morgan et al[17]	Acute coronary syndrome	70	1461 (case-control)	0
Richards et al[18]	Osteoporosis (2 phenotypes)	150	19,195 (cohort[d])	3[e]:9[f]
Samani et al[19]	Coronary artery disease	55	4864; 2519 (case-control)	1[g]
Scuteri et al[20]	Obesity (3 phenotypes)	74	6148 (cohort)	0
Sõber et al[21]	Blood pressure	149	1644; 8023 (cohort[h])	0
Wu et al[22]	Childhood asthma	237	1476 (triads[i])	1

Ioannidis *et al* (2011) "The False-positive to False-negative Ratio in Epidemiologic Studies" Epidemiology 22(4) p450-6

But really, where are we?

Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

PLoS 2005.

Also see http://matthew-brett.github.io/teaching/ioannidis_2005.html

Where are we? Guesses.

Corollary 1: The smaller the studies conducted in a scientific field, the less likely the research findings are to be true. Small sample size means smaller

Where are we? Guesses.

Corollary 2: The smaller the effect sizes in a scientific field, the less likely the research findings are to be true.

Where are we? Guesses.

Corollary 3: The greater the number and the lesser the selection of tested relationships in a scientific field, the less likely the research findings are to be true. As shown above, the post-study

Where are we? Guesses.

Corollary 4: The greater the flexibility in designs, definitions, outcomes, and analytical modes in a scientific field, the less likely the research findings are to be true.

Where are we? Guesses.

Corollary 5: The greater the financial and other interests and prejudices in a scientific field, the less likely the research findings are to be true.

Where are we? Guesses.

Corollary 6: The hotter a scientific field (with more scientific teams involved), the less likely the research findings are to be true.

Computational reproducibility in neuroimaging practice

To various degrees:

- GUIs
- Ad-hoc data archiving and directory structure
- Poor transfer of metadata
- "Wisdom of the ancients" analysis scripts using various batch mechanisms.
- "Makes sense" epistemology for the statistics.
- Tests rare and optional.
- Stream of consciousness exploration.

The fix in current practice

- Proceed as above.
- When you get to something you like, go back and clean up.
- Package for publication.

Teaching the fix

- Reproduciblity workshops and bootcamps.
- Ad-hoc transmission between post-docs and grad students.

The philosophy behind the fix

- "You don't need to know how a car works, to drive a car": a feeling that you get the general idea is usually enough.
- "A scientist does not need to be a good programmer".
- Errors are probably rare.
- If they are not rare, they are not important.
- Reducing error makes research less efficient.

Philosophy continued

From a discussion after a lab meeting on a Ioannidis paper.

> *Scientists range from automatons to revolutionaries . . . it is critical that "automatons" understand and be able to apply experimental methods rigorously and that "revolutionaries" be able to set aside details and step outside the box.*

We need surgery, not reassurance

"Rigor is hard to retrofit"

- First: coding
- Second: reliable working practice
- Third: deep understanding by building

Coding - first

The trained programmer:

- is more efficient;
- makes fewer mistakes;
- has a accurate and high estimate of the probability of mistakes;
- continues to learn.

https://blog.udacity.com/2014/01/peter-norvig-teach-yourself-programming.html

Reliable working practice

- version control
- testing
- continuous integration
- replicate and extend.

Deep understanding by building

"What I cannot create, I do not understand"

Found on Richard Feynman's blackboard after his death.

Can it be done?

"Reproducible and Collaborative Statistical Data Science", UC Berkeley, fall 2015.

Main course page at http://www.jarrodmillman.com/stat159-fall2015/

Our students

- Statistics undergraduates / masters students
- Various other disciplines, including neuroscience, architecture.
- Some background in probability, statistics, basic R programming.
- A couple of 40 students had experience of FMRI analysis.

What we covered

- Unix shell
- Version control
- Coding in Python
- Testing and continuous integration
- Images and arrays
- Basic statistical analysis of FMRI data.
- Group project replicating and / or extending OpenFMRI data analysis.

How we did it

- Lecture, exercise, reinforce, add new element, repeat.
- Using code to explain the ideas.
- Heavy emphasis on project, with multiple review points.
- All group projects public.
- All project work using public Git and Github mechanisms
- No FMRI analysis packages allowed.
- We used no Jupyter Notebooks.

Code for building and understanding

If we have time:

- http://matthew-brett.github.io/teaching/glm_intro.html

How well did it work?

[U]nlike most group projects (which last for maybe a few weeks tops or could conceivably be pulled off by one very dedicated person), this one will dominate the entire semester. . . . Try to stay organized for the project and create lots of little goals and checkpoints. You should always be working on something for the project, whether that's coding, reviewing, writing, etc. Ask lots of questions and ask them early!

Projects

All at https://github.com/berkeley-stat159/

A simple replication / exploration

https://github.com/berkeley-stat159/project-epsilon

- dataset ds000005—the “Mixed-gambles task”
- outlier detection (from class)
- logistic regression models of behavioral data.
- image smmoothing (Scipy)
- FMRI drift models, including PCA
- GLM at each voxel
- Bonferroni correction

Extension

https://github.com/berkeley-stat159/project-lambda

- ds000113—a high-resolution FMRI dataset of Forrest Gump;
- Very large images
- Cross voxel time course correlations
- Found image artefacts, error in original paper
- Replicated image correlations
- Extended to Random Forest Model to predict indoor, outdoor.

  We strongly encourage running on a machine with 120 GBs of accessible RAM to emulate development environment.

Git / github

Project	Commits	Issues	PRs	Comments	Words/comment	LoC	% covered
Alpha	787	23	190	379	24.7	3293	3.7
Beta	534	7	147	105	20.1	1753	2.0
Delta	571	31	121	117	35.1	996	21.3
Epsilon	593	26	310	79	40.9	1809	19.5
Eta	259	11	89	44	21.7	588	12.4
Gamma	329	4	79	35	22.3	1040	37.6
Iota	414	26	113	144	24.3	928	8.4
Kappa	337	30	99	86	16.6	1157	3.5
Lambda	365	22	67	82	17.6	732	91.5
Theta	547	25	133	450	22.1	1186	23.1
Zeta	344	3	49	21	20.8	8287	6.2

How we followed up

https://bic-berkeley.github.io/psych-214-fall-2016/

What y'all should do

- Ask your mentor about reproducibility-first teaching.
- Learn enough to know what you need to learn.
- Discover new things about the world.

Is this the end?

Yes, it's the end of the talk.

All material for this talk at:
https://github.com/matthew-brett/nss-reproduce