Causality and Ethics

EH6127 – Quantitative Methods

Steven V. Miller

Department of Economic History and International Relations

Stockholm University

Goal(s) for Today

1. Introduce students to the fundamental problem of causal inference
2. Discuss different research design types and their trade-offs.
3. Talk about issues of research ethics, especially in social science settings.
4. Highlight issues of automation, workflow, and replication.

The Problem, in Quotes

- “That correlation is not causation is perhaps the first thing that must be said.” - Barnard, 1982 (p. 387)
- “If statistics cannot relate cause and effect, they add to the rhetoric.” - Smith, 1980 (p. 1000 [stylized by me])

Associational Inference

A set of tools to understand how a response variable corresponds with some attribute. Tools include:

- Probability distributions (conditional, joint)
- Correlation
- Regression

"Associational inference consists of [estimates, tests, posterior distributions, etc.] about the associational parameters relating Y and A [from units in U]. In this sense, associational inference is simply descriptive statistics." - Holland, 1986 (p. 946)

Causal Inference and Rubin's "Potential Outcomes"

The Problem in a Nutshell

An individual (i) who is offered a treatment ($Z_i = 1$) has two potential outcomes:

- An outcome to be revealed if treated ($T_i = 1$): $Y_i(T_i = 1|Z_i = 1)$
- An outcome to be revealed if *un*treated ($T_i = 0$): $Y_i(T_i = 0|Z_i = 1)$

This is a missing data problem of a kind.

- We can only observe one.
- No perfect counterfactuals.
- Unicorns don't exist.

The Solution

For $T_i = 0$ and $T_i = 1$, given both offered treatment ($Z_i = 1$):

$$\text{Individual Treatment Effect for } i = Y_i(T_i = 1|Z_i = 1) - Y_i(T_i = 0|Z_i = 1)$$

Think in terms of population averages.

- Per Rubin, there is an important population parameter to estimate.
- Hence why he referred to it as "effect of the treatment on the treated." (i.e. TOT)
- Also: the "average treatment effect" (i.e. ATE)

The Importance of Random Assignment

Per random assignment: participants assigned to treatment/control must be same on average in the population ("equal in expectation").

- i.e. $E[Y_i(T_i = 0|Z_i = 1)]$ must be equal to $E[Y_i(T_i = 0|Z_i = 0)]$

By substitution:

$$TOT = E[Y_i(T_i = 1|Z_i = 1)] - E[Y_i(T_i = 0|Z_i = 0)]$$

When unbiased, a difference in sample means is sufficient:

$$\hat{TOT} = \frac{\sum_{i=1}^{n_1} Y_i}{n_1} - \frac{\sum_{i=1}^{n_0} Y_i}{n_0}$$

Important Assumptions in This Framework

- Exogeneity (worth reiterating)
- Unit homogeneity (i.e. expected *Y*s are same for same values of *X*)
- Conditional independence (i.e. values of *X* are assigned independently of values of *Y*)
- Stable unit treatment value assumption (SUTVA)
    - This one is a bear: think of it as an unmodeled "spillover."
    - Ideally: an observation responds only to its own treatment status.

Examples of SUTVA Violations

1. Contagion (vaccination effects depend on whether others have been vaccinated)
2. Displacement (cap-and-trade moves around emissions; doesn't curtail them)
3. Communication ("hey control group dude, you gotta try this new medication. It rules!")
4. Social comparison ("I like my housing situation less now that I see this group received new public housing")
5. Signaling (governments that advertise policy interventions are no longer "treating" in that sense)
6. Persistence/memory (respondents respond to need to be consistent)

Observational vs. Experimental Research

Observational research: involves a comparison of units subjected to different treatments.

- More common, more flexible. But difficult to isolate causal effects.

Experimental research: units under study are randomly assigned to treatments.

- satisfies key questions about observational research design

Experiments

Experiments are more effective at addressing causality.

- Want to explain social phenomena like medical researchers testing therapeutic care.
- Satisfies insights from Rubin's potential outcomes framework.
- Researcher control over conditions isolates confounding systematic factors.
- Random assignment isolates systematic differences from random differences.

Types of Experiments

There are numerous ways of assessing causal effects. One typology:

1. "Between subjects": units randomly assigned to distinct treatment/control groups.
2. "Within subjects": units observed before and after receiving a treatment.

External vs. Internal Validity

Internal validity: stimulus faithfully administered, as implemented in the design. Concerns:

- Noncompliance
- Attrition

External validity: results generalizable from the "lab" to the "real world." Concerns:

- Convenience sampling (esp. college students)
- Hawthorne effect

Experiments ideally maximize internal validity, if (possibly) at the expense of external validity.

Types of Experiments

Experiments are super-flexible. Some types you'll encounter:

1. Lab experiments
   - Maximize internal validity, prioritized over external validity
   - Typically prone to convenience sampling.
2. Survey experiments
   - Balance internal/external validity concerns
   - Typically higher *n* with more representativeness
   - Concerns: spillover, less agency over treatment
3. Field experiment
   - Same pros/cons as survey experiments, but with typically less control over treatment administration.
   - Cons (spillover, treatments) even more pronounced
4. Natural experiment
   - i.e. an exogenous shock to a panel design
5. Quasi-experiment
   - Treatments/controls with no randomization, or control over the treatment.

Ethics and Replication

I can't make you do these things in good faith...

- Social science is rife with cases of academic misconduct.
- Publication incentives breed dishonesty; you are compelled to rise above it.

...but I can teach you some tools to help you be honest.

- i.e. this is academic workflow and replication.

Some Ethical Takeaways

Your theoretical model is causal. Your empirical model may not be.

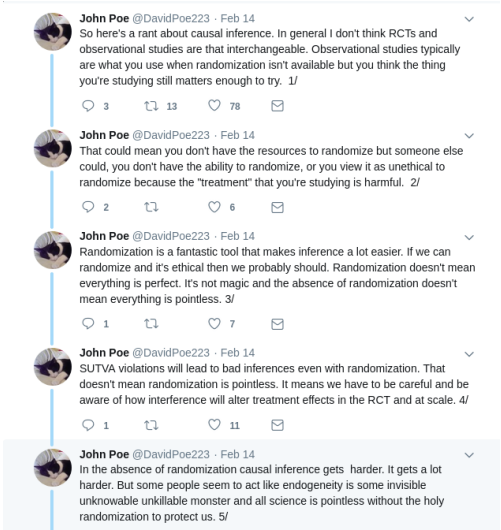
- Remember: everything is a "model."

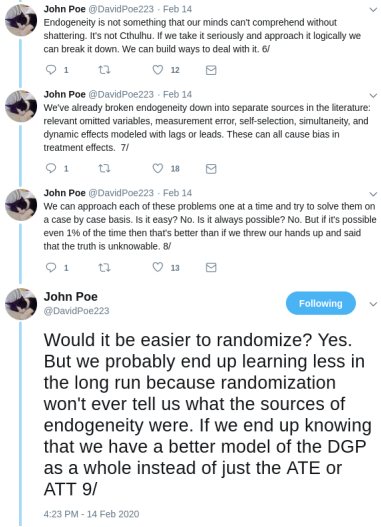
But don't shirk from using causal language!

- Absent a causal drive, the aim of the research is directionless/vague.

There's an unnecessary tension between the RCT people and those doing observational analyses.

- Be forthright, but stand your ground.
- Again: your theoretical model is causal. Your empirical model may not be.


John Poe @DavidPoe223 · Feb 14
So here's a rant about causal inference. In general I don't think RCTs and observational studies are that interchangeable. Observational studies typically are what you use when randomization isn't available but you think the thing you're studying still matters enough to try. 1/
3
13
78
John Poe @DavidPoe223 · Feb 14
That could mean you don't have the resources to randomize but someone else could, you don't have the ability to randomize, or you view it as unethical to randomize because the "treatment" that you're studying is harmful. 2/
2
6
John Poe @DavidPoe223 · Feb 14
Randomization is a fantastic tool that makes inference a lot easier. If we can randomize and it's ethical then we probably should. Randomization doesn't mean everything is perfect. It's not magic and the absence of randomization doesn't mean everything is pointless. 3/
1
7
John Poe @DavidPoe223 · Feb 14
SUTVA violations will lead to bad inferences even with randomization. That doesn't mean randomization is pointless. It means we have to be careful and be aware of how interference will alter treatment effects in the RCT and at scale. 4/
1
11
John Poe @DavidPoe223 · Feb 14
In the absence of randomization causal inference gets harder. It gets a lot harder. But some people seem to act like endogeneity is some invisible unknowable unkillable monster and all science is pointless without the holy randomization to protect us. 5/


John Poe @DavidPoe223 · Feb 14
Endogeneity is not something that our minds can't comprehend without shattering. It's not Cthulhu. If we take it seriously and approach it logically we can break it down. We can build ways to deal with it. 6/
1
12
John Poe @DavidPoe223 · Feb 14
We've already broken endogeneity down into separate sources in the literature: relevant omitted variables, measurement error, self-selection, simultaneity, and dynamic effects modeled with lags or leads. These can all cause bias in treatment effects. 7/
1
18
John Poe @DavidPoe223 · Feb 14
We can approach each of these problems one at a time and try to solve them on a case by case basis. Is it easy? No. Is it always possible? No. But if it's possible even 1% of the time then that's better than if we threw our hands up and said that the truth is unknowable. 8/
1
13
John Poe
@DavidPoe223
Following
Would it be easier to randomize? Yes. But we probably end up learning less in the long run because randomization won't ever tell us what the sources of endogeneity were. If we end up knowing that we have a better model of the DGP as a whole instead of just the ATE or ATT 9/
4:23 PM - 14 Feb 2020

Corrine McConnaughy
@cmMcConnaughy

A problem with social science paper brain applied directly to the real world is that the real world is in fact a tangled endogenous mess and pulling the “one true lever” doesn’t fix everything.

1:54 PM · Sep 30, 2022

17 Retweets **2** Quote Tweets **111** Likes

The C-Word: Scientific Euphemisms Do Not Improve Causal Inference From Observational Data

Causal inference is a core task of science. However, authors and editors often refrain from explicitly acknowledging the causal goal of research projects; they refer to causal effect estimates as associational estimates.

This commentary argues that using the term "causal" is necessary to improve the quality of observational research.

Specifically, being explicit about the causal objective of a study reduces ambiguity in the scientific question, errors in the data analysis, and excesses in the interpretation of the results.
(*Am J Public Health*. 2018;108: 616–619. doi:10.2105/AJPH.2018.304337)

Miguel A. Hernán, MD, DrPH

See also Galea and Vaughan, p. 602; Begg and March, p. 620; Ahern, p. 621; Chiolero, p. 622; Glymour and Hamad, p. 623; Jones and Schooling, p. 624; and Hernán, p. 625.

You know the story:

> Dear author: Your observational study cannot prove causation. Please replace all references to causal effects by references to associations.

Many journal editors request authors to avoid causal language,[1] and many observational researchers, trained in a scientific environment that frowns upon causality claims, spontaneously refrain from mentioning the C-word ("causal") in their work. As a result, "causal effect" and terms with similar meaning ("impact," "benefit," etc.) are routinely avoided in scientific publications that describe nonrandomized studies. Instead, we use terms like

Confusion then ensues at the most basic levels of the scientific process and, inevitably, errors are made.

We need to stop treating "causal" as a dirty word that respectable investigators do not say in public or put in print. It is true that observational studies cannot definitely prove causation, but this statement misses the point, as discussed in this commentary.

OF COURSE "ASSOCIATION IS NOT CAUSATION"

glass of red wine per day versus no alcohol drinking. For simplicity, disregard measurement error and random variability—that is, suppose the 0.8 comes from a very large population so that the 95% confidence interval around it is tiny.

The risk ratio of 0.8 is a measure of the association between wine intake and heart disease. Strictly speaking, it means that drinkers of one glass of wine have, on average, a 20% lower risk of heart disease than individuals who do not drink. The risk ratio of 0.8 does not imply that drinking a glass of wine every day lowers the risk of heart dis-

Some Ethical Takeaways

The proliferation of machine learning/AI/"algorithms" creates more ethical issues.

1. "Treat", don't manipulate.
2. There's no bias-free model; *you are the bias.*
3. Evil is evil, whether intentional or unintentional.

Don't let stupidity transform into evil.

- Good academic workflow can help.

THE WALL STREET JOURNAL.

Blue Feed, Red Feed

See Liberal Facebook and Conservative Facebook, Side by Side

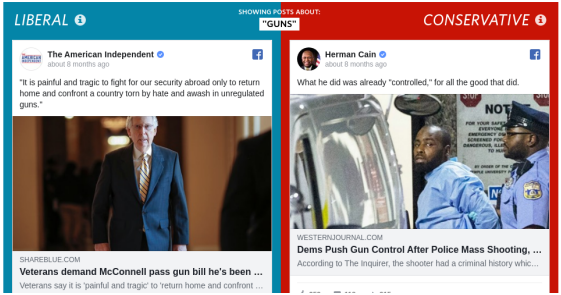
By ***Jon Keegan***

Published May 18, 2016 at 8:00 a.m. ET | Last Updated Aug. 19, 2019

This page has been archived and will no longer be updated.

FILTER ARCHIVED FEEDS BY TOPIC:

PRESIDENT TRUMP | HEALTH CARE | GUNS | ABORTION | ISIS | BUDGET | EXECUTIVE ORDER | IMMIGRATION



NEWS · 24 OCTOBER 2019 · UPDATE 26 OCTOBER 2019

Millions of black people affected by racial bias in health-care algorithms

Study reveals rampant racism in decision-making software used by US hospitals — and highlights ways to correct it.

Heidi Ledford

Numbers don’t always tell the truth

Mark J. Girouard, an employment attorney at Nilan Johnson Lewis, says one of his clients was vetting a company selling a resume screening tool, but didn’t want to make the decision until they knew what the algorithm was prioritizing in a person’s CV.

After an audit of the algorithm, the resume screening company found that the algorithm found two factors to be most indicative of job performance: their name was Jared, and whether they played high school lacrosse. Girouard’s client did not use the tool.

Q. If machine learning is so smart, how come AI models are such racist, sexist homophobes? A. Humans really suck

Our prejudices rub off on our computer pals, sadly

By Katyanna Quach 5 Sep 2019 at 07:02 64 SHARE ▼

Academic Workflow and Replication

Replication crises/academic misconduct are proliferating in social science. Examples:

- Economics: Reinhart and Rogoff's (2010) Excel error
  - *ed. their analysis was **way** more dishonest than the Excel error, but that got the most attention.*
  - See: "Revisiting Reinhart and Rogoff, Ten Years Later" on my website.
- Psychology: too many to list
  - Recurring themes: small-*n*, *p*-hacked experiments, or even fabricated data
- Sociology/criminology: Stewart retractions
- Political science: Lacour and Green (2014) scandal

I'm not going to assign motives (naiveté or something worse) to all these scandals and those involved.

- But, assuming honesty, you can avoid a similar pitfall with good workflow.

Some Tips on Good Workflow/Replication

"Kondo" your projects into sub-directories.

- Keep things tidy/de-cluttered in your project.
- I have my recommendations, but tweak for what works for you.

"Launder" your data; never overwrite them.

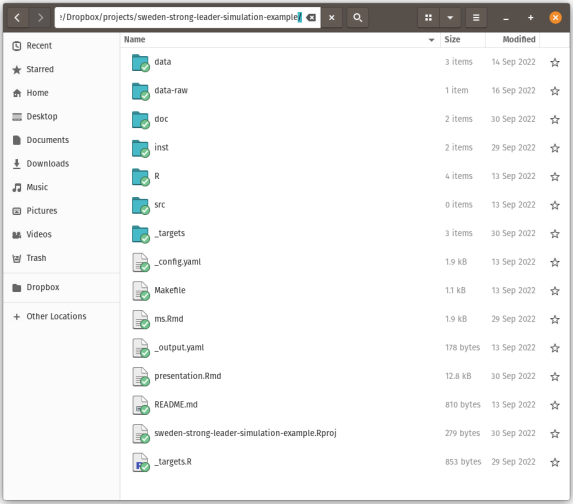
- Never overwrite original columns. Recode into new columns/objects.
- *Definitely* never overwrite raw data.

Related: invest in cloud storage (e.g. Box, Dropbox).

- Create separate folders for raw data (`data`) and your individual projects (`projects`).
- Tongue in cheek: think of "my laptop broke/fried/got stolen" as the 21st century equivalent of "the dog ate my homework."

Learn to automate what you can.

An Example of Sub-Directories



Surprise! You're a Computer Programmer Now

How to improve your relationship with your future self*

Cómo mejorar su relación con su futuro yo

JAKE BOWERS
Universidad de Illinois

MAARTEN VOORS
Wageningen University

ABSTRACT

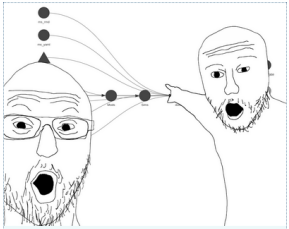
This essay provides practical advice about how to do transparent and reproducible data analysis and writing. We note that doing research in this way today will not only improve the cumulation of knowledge within a discipline, but it will also improve the life of the researcher tomorrow. We organize the argument around a series of homilies that lead to concrete actions. (1) Data analysis is computer programming. (2) No data analyst is an island for long. (3) The territory of data analysis requires maps. (4) Version control prevents clobbering, reconciles history, and helps organize work. (5) Testing minimizes error. (6) Work *can* be reproducible. (7) Research ought to be credible communication.

Key words: research transparency, reproducible research, workflow, methodology

An Example of Automating/Reproducing a Workflow

Automating a Workflow with {steveproj}, {stevetemplates}, and {targets}

Posted on September 13, 2022 by steve in R



I've been racking my brain for some time around the problem of tailoring a project's workflow in a way that optimizes automation, reproducibility, and—depending on the project's scale—speed. A previous stab at this looked to R Markdown as an operating system for a project. This will help link the manuscript side of a project with the analysis side of a project, but has the drawback of asking too much from R Markdown and the researcher. R Markdown really isn't an operating system and the researcher would have to invest some time in learning about the caching and chunk quirks of R Markdown. I think I improved on this greatly with my development of `{steveproj}`. `{steveproj}` (and `{stevetemplates}`) have a lot of parlor tricks for preparing manuscripts, anonymized manuscripts, and title pages for peer review. However, that approach I developed last year has drawbacks in leaning hard on old school Make. It's also a bit inflexible. It implicitly builds in a workflow that separates the analysis from the communication of the analysis, leaving not a lot of room to work interactively with the written report (as the project is still developing). Such a separation of the analysis from the report of the analysis is ideal, but never quite real.

After finally getting (somewhat) settled in Sweden, I dedicated some time to learning more about `{targets}`, a function-oriented Make-like tool for doing statistics in R. I think I finally figured out how to do this well in a way that's automated, reproducible, quick, and flexible. The mechanics here are fairly simple—and my familiarity with `{targets}` is basic—but the approach I outline below should scale nicely. If you're interested, I set up a basic Github repo that you can fork and run on your end to see how this works. I'll describe it in some detail, though.

Conclusion

It was good to talk with you over the semester on these issues. Takeaways for today:

- Causality is exact (and yet multiple). Understand what's at stake.
- Each (obs. or exp.) research design has its own set of trade-offs.
- The whole world's an endogenous mess, but we're all trying.
- I can't make you be honest, but I can give you tips/tools to help.
- Accidental evil is still evil. Don't be evil.

Table of Contents