Causal Inference Workshop

Day 1: Experiments and Potential Outcomes

Pablo Geraldo (pdgeraldo@ucla.edu)

SICSS - UCLA

June 23, 2021

Why should we study causal inference?

The social sciences are experimenting what some authors have called a "credibility revolution" (Angrist and Pischke, 2010), an "identification revolution" (Morgan, 2016), or simply a "causal revolution" (Pearl and MacKenzie, 2018).

In artificial intelligence/ML, causality have been deemed "the next frontier" and "the next most important thing".

The enormous progress in the last decades has been facilitated by the development of a mathematical framework that provide researchers with tools to handle causal questions: Potential Outcomes and Structural Causal Model.

What should you expect from this workshop?

This workshop is designed as a "Crash Course", so we can obviously focus only on a few things:

- Familiarize yourself with the most widely used CI frameworks
- Understand the role of randomization to tackle causal questions
- Use potential outcomes and the do-operator to formalize causal estimands
- Use directed acyclic graphs (DAGs) to encode qualitative assumptions
- Derive identification results and testable implications from a DAG
- Assess the plausibility of different identification strategies applied to real problems

At the end of our two sessions, I hope you feel better equipped to read and evaluate the applied literature and to design your own studies using appropriate identification strategies (or at least having a clear idea on what to look for and where!).

What are we not covering today?

There is a lot of stuff out there! Some areas that you might be interested in but we don't have enough time to review today:

- Sequential estimation for time-varying treatments (g-methods)
- Estimation in general, including Targeted Maximum Likelihood
- Machine learning for heterogeneous treatment effects
- Do-calculus and PO-calculus for identification
- Causal mediation analysis, causal attribution
- Other graphical models (MCM, SWIGs, the Hypothetical Model)
- Selection diagrams for missing data
- Data fusion (generalization, external validity)
- Causal discovery (going from the data to the DAG)

Intuitions about causality

Have you heard any of these before?

"Correlation does not imply causation"

But can we go from one to the other?

"No causation without manipulation"

Then what about race or gender?

"Causal inference is a missing data problem"

Or is it the other way around?

"For causal inference, design trumps analysis"

But what do we mean by design? And analysis?

What is causal inference about?

Statistics/ML vs Causal Inference

Statistics/ML

- Passive observation of the data generating process
- Estimand: Joint probabilities, CEF

$$P(X, Y)$$

$$E(Y|X)$$

- Focus on asymptotics / out of sample prediction
- Estimation problem: variance-bias tradeoff
- Pearl: "deep learning is just curve fitting"

Causal Inference

- Prediction under interventions on the data generating process
- Estimand: interventional quantities

$$P(Y|do(x))$$

$$E(Y|do(x)) - E(Y|do(x'))$$

$$= E(Y_x) - E(Y_{x'})$$

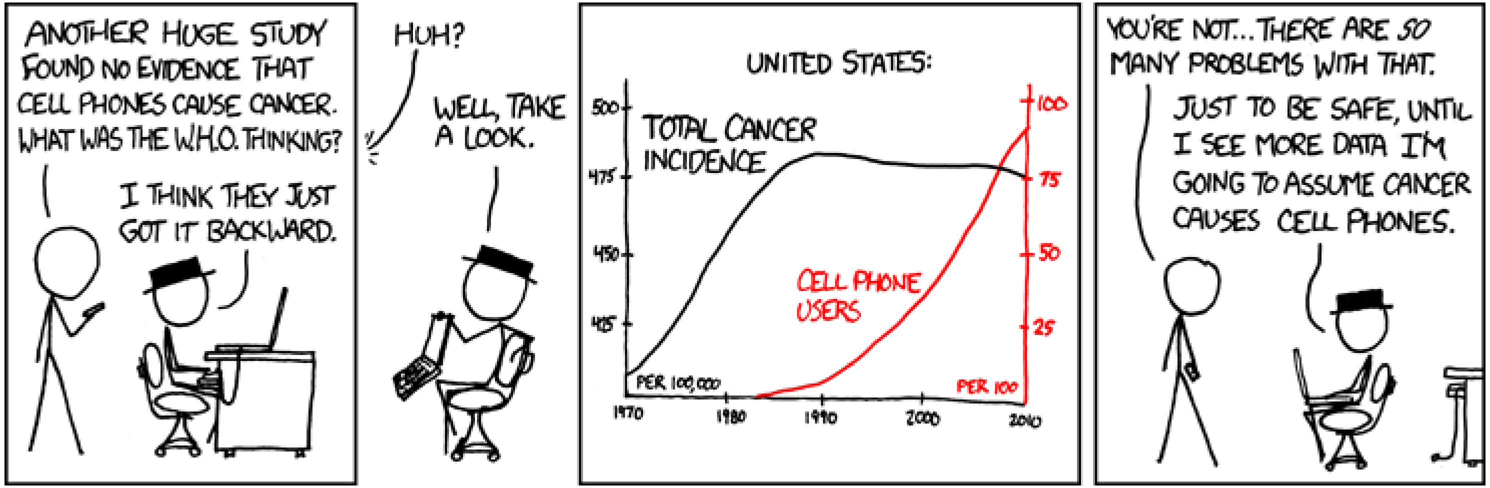
- Identification problem: consistency (infinite sample)
- Estimation problem: in general, focus on bias over variance (but changing)

The ladder of causality

Level	Estimand	Activity	Field/Discipline	Example
Association	$\mathbf{P(Y\|X)}$	Seeing, Observing	Stats, Machine Learning	*What would I believe about Y if I see X?* What is the expected income of a college graduate?
Intervention	$\mathbf{P(Y\|do(x))}$	Doing, Intervening	Experiments, Policy evaluation	*What would happen with Y if I do X?* What would be my income if I graduate from college?
Counterfactual	$\mathbf{P(Y_x\|x',y')}$	Imagining, Retrospecting	Structural Models	*What would have happened with Y have I done X instead of X'? Why?* What would have been my parents' income have they graduated from college, given that they didn't attend?

Pearl and Mackenzie (2018)

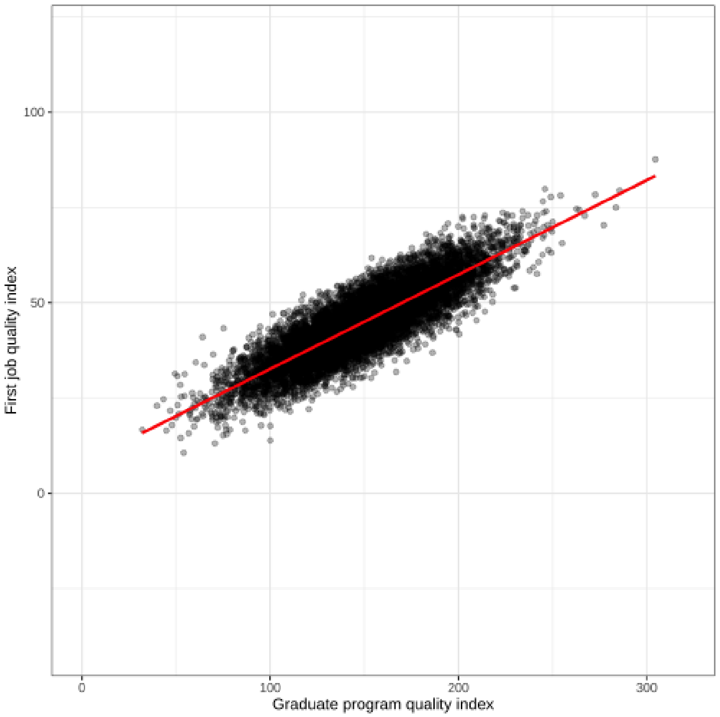
Is doing *really* that different from seeing?


ANOTHER HUGE STUDY FOUND NO EVIDENCE THAT CELL PHONES CAUSE CANCER. WHAT WAS THE W.H.O. THINKING?
I THINK THEY JUST GOT IT BACKWARD.
HUH?
WELL, TAKE A LOOK.
UNITED STATES:
TOTAL CANCER INCIDENCE
CELL PHONE USERS
PER 100,000
PER 100
500
475
450
425
100
75
50
25
1970
1980
1990
2000
2010
YOU'RE NOT... THERE ARE SO MANY PROBLEMS WITH THAT.
JUST TO BE SAFE, UNTIL I SEE MORE DATA I'M GOING TO ASSUME CANCER CAUSES CELL PHONES.

Is doing *really* that different from seeing?

Let's imagine an example: the effect of the graduate program attended on the quality of your first job after graduation

```
set.seed(1988)
# sample size
N <- 10000
# student selectivity
W <- rnorm(N, mean=250, sd=50)
# program selectivity
X <- 0.6*W + rnorm(N, mean=0, sd=10)
# quality of first job
Y <- 0.3*W - 0.2*X + rnorm(N)
data <- tibble(Y=Y, X=X, W=W)
```



Is doing *really* that different from seeing?

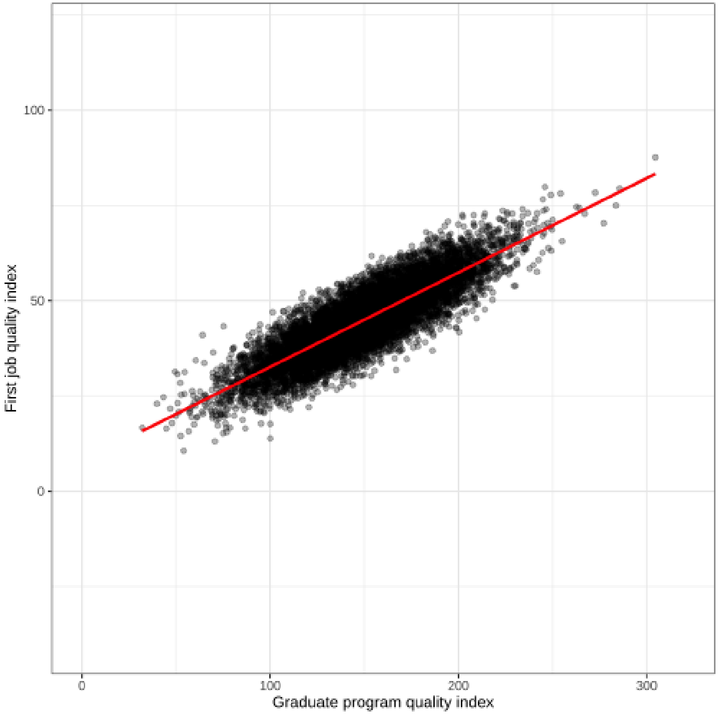
Let's imagine an example: the effect of the graduate program attended on the quality of your first job after graduation

Let's see what a linear regression would tell

```
lm(Y~X, data=data)
```

```
##
## Call:
## lm(formula = Y ~ X, data = data)
##
## Coefficients:
## (Intercept)            X
##      7.8506       0.2476
```

However, we know that the effect is negative!



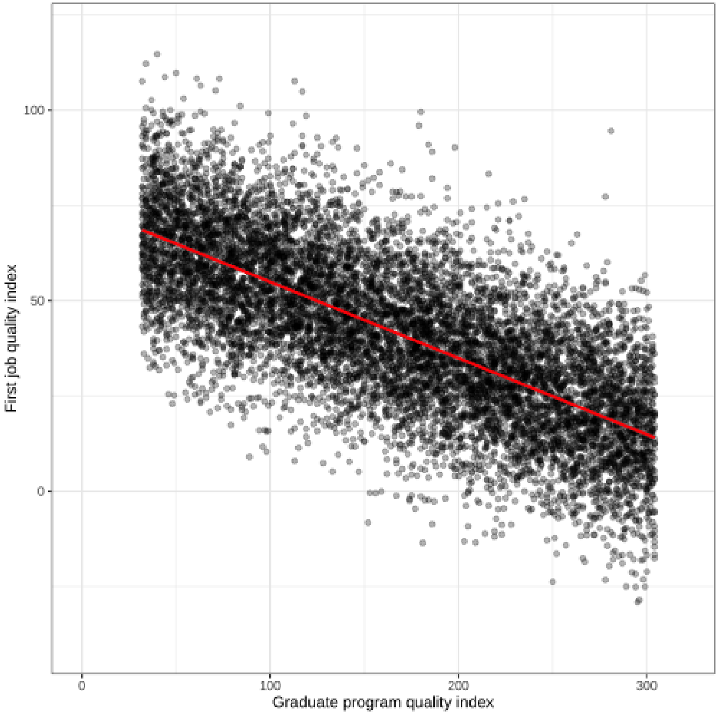
Is doing *really* that different from seeing?

What if we instead randomize students to different programs?

```
# let's randomize students to programs!
X <- sample(min(X):max(X),
            N, replace=TRUE)
# quality of first job
Y <- 0.3*W - 0.2*X + rnorm(N)
data$Xrand <- X
data$Ytrue <- Y

lm(Ytrue~Xrand, data=data)
```

```
##
## Call:
## lm(formula = Ytrue ~ Xrand, data = data)
##
## Coefficients:
## (Intercept)        Xrand
##     74.9410      -0.2001
```

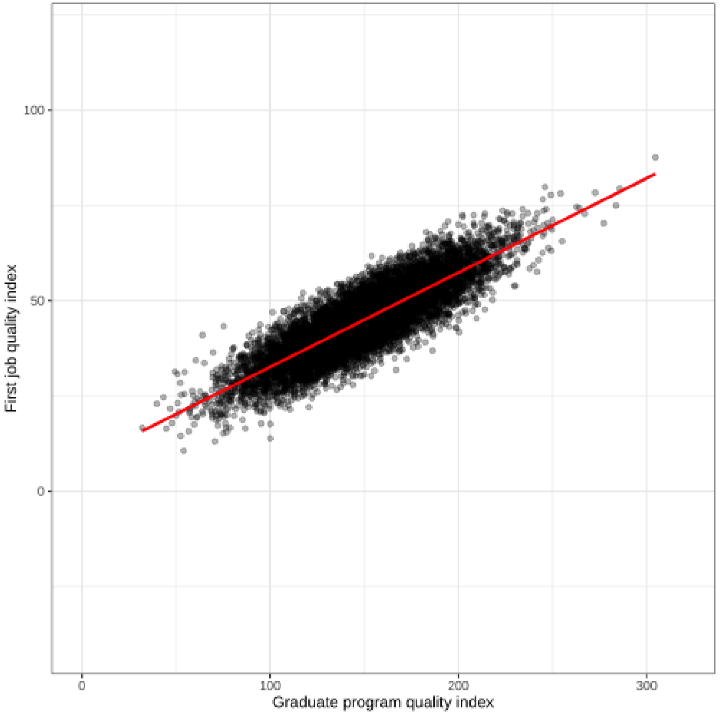


Is doing *really* that different from seeing?

Can we fix the observational data by adjusting for W?

```
# Same regression
# Plus controls
lm(Y~X+W, data=data)
```

```
##
## Call:
## lm(formula = Y ~ X + W, data = data)
##
## Coefficients:
## (Intercept)            X            W
##     0.02049     -0.20011      0.29996
```



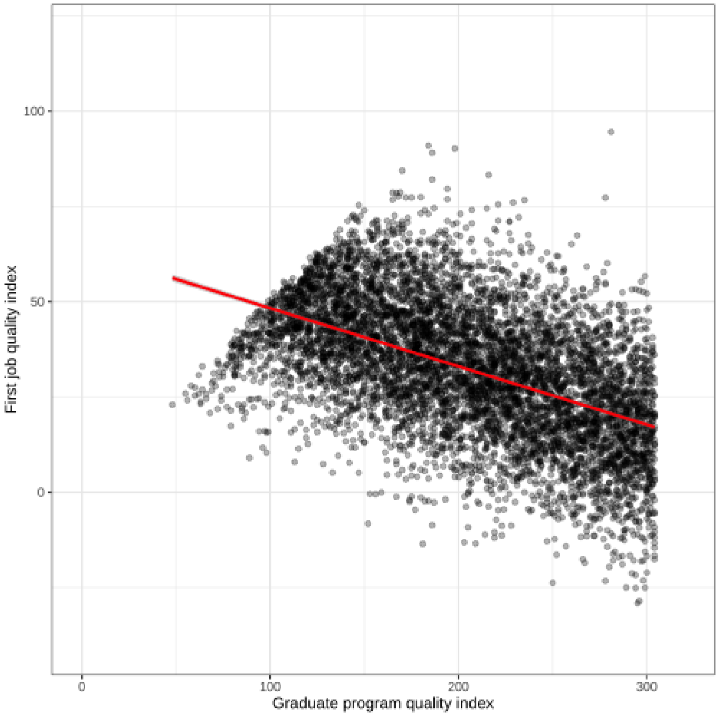
Is doing *really* that different from seeing?

Where are we getting the data from?

```
# Registry of JMC
prob <- 1/(1+exp(-(X-Y*2)))
data$C <- rbinom(N, 1, prob=prob)

lm(Ytrue~Xrand, data=data %>%
     filter(C==1))
```

```
##
## Call:
## lm(formula = Ytrue ~ Xrand, data = data %>%
##
## Coefficients:
## (Intercept)        Xrand
##     63.5425      -0.1526
```



Potential Outcomes

Potential Outcomes

Introduced by Neyman (1923) in the context of experimental design. They remained only used in that context for decades!

- Imported and developed by Donald Rubin for observational studies (c. 1974)
- They are really great to clarify *what do we want to know* (estimand)
- This includes identifying *reasons for discrepancies* between what we observe and our estimand (bias)
- They are great to formalize *what needs to be true* for our estimand to be identified with a given *estimator* (assumptions)
- They are not-so-great to assess if our assumptions are plausible or defensible (more on this soon!)

Potential Outcomes: Notation

Let's start with some definitions:

Y is the outcome variable *as we observe it*

X is the variable whose effect we want to study (treatment, exposure)

Y_x is the potential outcome when we set $X = x$. For example, when $X \in \{0, 1\}$:

- Y_1 is the potential outcome under "treatment"
- Y_0 is the potential outcome under "control"

Potential Outcomes: Notation

You will find a lot of equivalent notations for potential outcomes. It could be confusing, but it is good to get practice working with different variants (at least if you want to read the papers!)

$$Y(x) = Y_x = Y^x$$

Read it like this:

The value that the variable Y would take if we set the variable X to the value x.

Can you ever observe any of those potential outcomes?

Consistency (also known as SUTVA)

$$X = x \to Y = Y_x$$

For the binary treatment case, we have:

$$Y = XY_1 + (1 - X)Y_0$$

What are the assumptions built in this notation? What type of dependence are we ruling out?

Short activity (3 mins)

Researchers, at least in sociology, tend to formalize their effect of interest as regression coefficients (i.e., their hypothesis are formulated within a statistical model)

Potential outcomes offer a way to formalize what do we mean by a causal effect outside any statistical model. This allows us to clearly separate *what do we want* (a certain estimand), the statistical machinery to answer our question (an estimator) and the particular answer we get (our empirical estimate).

Lundberg, Johnson, and Stewart (2021) discuss in great detail this point. Absolutely worth reading! Ungated version here.

Take a moment to think in your own research:

- What causal question is relevant for you to study?
- Can you formulate it using potential outcomes?
- What are the assumptions your are making in this formalization?

Potential Outcomes: The Science Table

One advantage of PO is that we can treat them directly as random variables! So, everything we already know related to probability manipulation still applies here.

Unit	X_i	Y_i	Y_{ai}	Y_{bi}	τ_i
1	A	1	1	1	0
2	A	0	0	0	0
3	A	0	0	1	-1
4	A	1	1	0	1
5	A	1	1	1	0
6	B	0	1	0	1
7	B	0	1	0	1
8	B	0	0	0	0
9	B	1	0	1	-1
10	B	1	1	1	0

The basic calculation device (usually implicit) for that matter is the science table. Basically, the full schedule response of the potential outcomes under different treatment conditions

Potential Outcomes: Average Treatment Effect

One advantage of PO is that we can treat them directly as random variables! So, everything we already know related to probability manipulation still applies here.

Unit	X_i	Y_i	Y_{ai}	Y_{bi}	τ_i
1	A	1	1	1	0
2	A	0	0	0	0
3	A	0	0	1	-1
4	A	1	1	0	1
5	A	1	1	1	0
6	B	0	1	0	1
7	B	0	1	0	1
8	B	0	0	0	0
9	B	1	0	1	-1
10	B	1	1	1	0

$$\text{ATE} = E(Y_{ai}) - E(Y_{bi})$$

$$(6/10) - (5/10) = \mathbf{0.1}$$

Potential Outcomes: difference in means

One advantage of PO is that we can treat them directly as random variables! So, everything we already know related to probability manipulation still applies here.

Unit	X_i	Y_i	Y_{ai}	Y_{bi}	τ_i
1	A	1	1	.	.
2	A	0	0	.	.
3	A	0	0	.	.
4	A	1	1	.	.
5	A	1	1	.	.
6	B	0	.	0	.
7	B	0	.	0	.
8	B	0	.	0	.
9	B	1	.	1	.
10	B	1	.	1	.

$$\text{ATE} = E(Y_{ai}) - E(Y_{bi})$$

$$(6/10) - (5/10) = \mathbf{0.1}$$

$$\text{diff-in-means} = E(Y_i|X=a) - E(Y_i|X=b)$$

$$(3/5) - (2/5) = \mathbf{0.2}$$

$$\text{diff-in-means} \neq \text{ATE}$$

But why?

Potential Outcomes: sources of bias

One advantage of PO is that we can treat them directly as random variables! So, everything we already know related to probability manipulation still applies here.

Unit	X_i	Y_i	Y_{ai}	Y_{bi}	τ_i
1	A	1	1	.	.
2	A	0	0	.	.
3	A	0	0	.	.
4	A	1	1	.	.
5	A	1	1	.	.
6	B	0	.	0	.
7	B	0	.	0	.
8	B	0	.	0	.
9	B	1	.	1	.
10	B	1	.	1	.

$$E(\text{diff-in-means})$$

$$= E(Y_i|X = a) - E(Y_i|X = b)$$

$$= E(Y_a|X = a) - E(Y_b|X = b)$$

$$= ATE+$$

$$(E[Y_b|X = a] - E[Y_b|X = b])+$$

$$(1 - P[X])(ATT - ATC)$$

$$= \mathbf{0.1} + 0.2 + (0.5)(-0.2) = \mathbf{0.2}$$

Potential Outcomes: identification assumptions

One advantage of PO is that we can treat them directly as random variables! So, everything we already know related to probability manipulation still applies here.

Unit	X_i	Y_i	Y_{ai}	Y_{bi}	τ_i
1	A	1	1	1	0
2	A	0	0	0	0
3	A	0	0	1	-1
4	A	1	1	0	1
5	A	1	1	1	0
6	B	0	1	0	1
7	B	0	1	0	1
8	B	0	0	0	0
9	B	1	0	1	-1
10	B	1	1	1	0

We need the following condition to be true:

$$Y_x \perp\!\!\!\perp X$$

Do we meet that condition here? No!

$$(Y_{ai}, Y_{bi}) \not\perp\!\!\!\perp X$$

Because:

$$P(Y_a = y|X = a) \neq P(Y_a = y)$$

$$P(Y_b = y|X = b) \neq P(Y_b = y)$$

Potential Outcomes: identification assumptions

One advantage of PO is that we can treat them directly as random variables! So, everything we already know related to probability manipulation still applies here.

Unit	X_i	Y_i	Y_{ai}	Y_{bi}	τ_i	W_i
1	A	1	1	1	0	Quant
2	A	0	0	0	0	Quant
3	A	0	0	1	-1	Qual
4	A	1	1	0	1	Quant
5	A	1	1	1	0	Qual
6	B	0	1	0	1	Qual
7	B	0	1	0	1	Quant
8	B	0	0	0	0	Qual
9	B	1	0	1	-1	Qual
10	B	1	1	1	0	Quant

What about including another covariate W?

Does the following condition holds?

$$Y_x \perp\!\!\!\perp X|W$$

Not quite either! But still "better" than before, right? Let's define:

$$\text{ATE}_W = E(Y_a - Y_b|W = w)$$

and the estimator

$$\widehat{\text{ATE}}_W =$$

$$E(Y_i|X = a, W = w) - E(Y_i|X = b, W = w)$$

Potential Outcomes: identification assumptions

One advantage of PO is that we can treat them directly as random variables! So, everything we already know related to probability manipulation still applies here.

Unit	X_i	Y_i	Y_{ai}	Y_{bi}	τ_i	W_i
1	A	1	1	1	0	Quant
2	A	0	0	0	0	Quant
3	A	0	0	1	-1	Qual
4	A	1	1	0	1	Quant
5	A	1	1	1	0	Qual
6	B	0	1	0	1	Qual
7	B	0	1	0	1	Quant
8	B	0	0	0	0	Qual
9	B	1	0	1	-1	Qual
10	B	1	1	1	0	Quant

$$\widehat{\text{ATE}}_{W=quant} = 0.16$$

$$\widehat{\text{ATE}}_{W=qual} = 0.16$$

$$\widehat{\text{ATE}} = (0.5)(0.16) + (0.5)(0.16) = 0.16$$

However, look a the true ATE_W:

$$\text{ATE}_{W=quant} = -0.2$$

$$\text{ATE}_{W=qual} = 0.4$$

$$\text{ATE} = (0.5)(-0.2) + (0.5)(0.4) = \mathbf{0.1}$$

Potential Outcomes: how to assess our assumptions?

One advantage of PO is that we can treat them directly as random variables! So, everything we already know related to probability manipulation still applies here.

Unit	X_i	Y_i	Y_{ai}	Y_{bi}	τ_i	W_i
1	A	1	1	.	.	Quant
2	A	0	0	.	.	Quant
3	A	0	0	.	.	Qual
4	A	1	1	.	.	Quant
5	A	1	1	.	.	Qual
6	B	0	.	0	.	Qual
7	B	0	.	0	.	Quant
8	B	0	.	0	.	Qual
9	B	1	.	1	.	Qual
10	B	1	.	1	.	Quant

In general, we rely on extra-statistical assumptions about the data generating process to claim causal identification.

> "No causes in, no cases out"
>
> > Nancy Cartwright

Is there a way to design an study in which we know, by design, that the needed assumptions hold?

Randomized experiments

Why randomization

If we want to predict under interventions, then the best way to do it is interveening!

Random assignment (more specifically, RCTs) has been called the gold standard for causal inference: it guarantees the necessary assumptions for causal inference hold by design.

When unfesible, imagining a hypothetical experiment still offers a useful benchmark to assess the validity of causal claims, and even to clarify what do we mean by a particular causal effect.

Experiments come in many different flavours: lab, field, survey, and even quasi-experimentss!

Here we will only scratch the surface of social science experiments: the idea is to get you interested and point you to the resources out there! Maybe sometime soon you will be running your own experiment! (Or at least someone else's experiment)

Short activity (3 mins)

Sometimes it is hard to imagine an experiment that would be relevant for the type of questions we care about.

Some people even say (and I for sure partially agree!) that experiments tend to emphasize "small" versus "big" questions, promoting incremental/testable policies.

However, there are tons of examples of researchers using experiments to address important, big and difficult questions. Do you know of any example?

Take a moment to check the syllabus of UCLA professor Graeme Blair's Experimental Design class here. He put together a list of experiments conducted by UCLA faculty, and by graduate students.

Any comments?

Why randomization

(more formally)

We already saw that we can identify the causal effect of X on Y if the treatment assignment is independent of the potential outcomes. Formally

$$(Y_x, Y_{x^*}) \perp\!\!\!\perp X$$

Recall the diff-in-means decomposition we review earlier. Given the ignorability of the treatment assignment, we have can further write it as:

By consistency

$$E(Y_i|X = x) - E(Y_i|X = x^*) = E(Y_x|X = x) - E(Y_{x^*}|X = x^*)$$

With some algebra

$$= E(Y_x - Y_{x^*}) + (E[Y_{x^*}|X = x] - E[Y_{x^*}|X = x^*]) + (1 - P[X])(ATT - ATC)$$

By ignorability, this simplifies to

$$E(Y_x - Y_{x^*}) = ATE$$

Forms of validity

Traditionally, researchers argue about the validity of a study's causal conclusion (and, more generally, about the validity of different research designs) based on the potential biases that pose threats to validity. Check this amazing paper by Matthay and Glymour for a review.

We reviewed the bias in the difference-in-means estimator: baseline differences (under the control condition), and differential response to the treatment (under the exposure condition).

But when we randomize an exposure, we know that who ends up in each treatment arm has nothing to do with their potential outcomes!

This is why we generally say that experiments are great for internal validity: among the people that participated in our study, we can rule out systematic sources of bias.

However, this does not imply that our results are externally valid, i.e., that they apply to people outside our study! We need further assumptions to move from one to another.

How to randomize?

Many times, we use randomization not just for identification (ignorability), but also for estimation!

If we assume the potential outcomes are fixed, and the only thing that varies is the treatment assignment scheme, we can derive a permutation distribution and use it for inference.

How much dispersion (i.e. uncertainty) is in our distribution will be affected by the level at which randomization (or, more precisely, the treatment) happens: is it at the individual level? or at a cluster/group level?

- The more the aggregation, the more uncertainty. So why would we want to randomize at the cluster level?

Conditional randomization (i.e., blocking) increase efficiency, when we have variables that are highly predictive of the outcome of interest

- One extreme of this is randomization in matched pairs: for each pair of individual with similar covariates, we randomly assign one to treatment and one to control.

Types of experiments

- Laboratory experiments: Usually conducted with a small sample (of undergraduate psychology students), many times involving games in a computer. Helpful for cognitive/behavioral questions.
- Field experiments: In order to obtain more *externally valid* results, experiments conducted in the field (i.e., under real-world conditions) are the way to go. Definitely more expensive though. Audit studies are a particular type of field experiment.
- Survey experiments: One can randomize treatment conditions *in a survey* to evaluate how participants change their responses based on certain stimulus. Vignettes and list experiments are examples of this approach.
- (Bonus) Quasi-experiments: Researchers usually call quasi-experiments to real-world situations that offer as-if random variation in a treatment of interest. For example, earthquakes, change in laws, date of birth, etc.

Additional Resources

Online learning

- A selected and annotated bibliography on causality here.
- J-PAL research resources here
- EGAP methods guides here

Textbooks

- Gerber and Green (2012) Field Experiments: Design, Analysis, and Interpretation (check here)

Where to conduct survey experiments

(by Natasha Quadlin)

- Time-Sharing Experiments for the Social Sciences: https://tessexperiments.org/

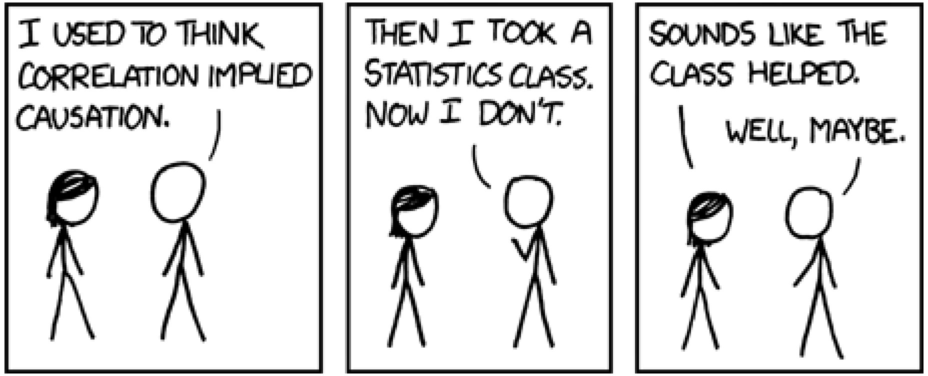
  Proposals and materials for all prior studies: https://tessexperiments.org/previousstudies.html

  (Pros: nationally representative, free, fast; Cons: must submit proposal)
- Qualtrics: https://www.qualtrics.com/

  (Pros: decent sample; Cons: can be data quality issues, expensive)
- Prolific Academic: https://www.prolific.co/

  (Pros: relatively inexpensive, lots of subsamples; Cons: can be data quality issues)
- Amazon Mechanical Turk: https://www.mturk.com/

  (Pros: fast, cheap; Cons: opt-in sampling, can be data quality issues)



I USED TO THINK CORRELATION IMPLIED CAUSATION.
THEN I TOOK A STATISTICS CLASS. NOW I DON'T.
SOUNDS LIKE THE CLASS HELPED.
WELL, MAYBE.

Activity

Afternoon activity

Meet with your project group, and think in a research question that you could possibly address using an experimental design:

- What is your research question?
- What is your estimand? (effect of what? among whom?)
- What type of experiment would you conduct? (lab? field? survey?)
- What would be the level of your randomization? (individual? cluster? why?)

Using your answers to the question above, propose and design a (survey) experiment that could help answering your research questions.