Introduction to data science & artificial intelligence (INF7100)

Arthur Charpentier

#122 Counterfactual

été 2020

Maths of Causal Inference

n individuals are either treated $(t_i = 1)$ or not $(t_i = 0)$.
We observe outcome y_i for covariates $\boldsymbol{x}_i$.
We want to study potential outcomes $y_i(1)$ and $y_i(0)$

	turnout				
	$y_i(1)$	$y_i(0)$	t_i	$x_{1,i}$	$x_{2,i}$
1	y_1	?	1	$x_{1,1}$	$x_{2,1}$
2	?	y_2	0	$x_{1,2}$	$x_{2,2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	y_n	?	1	$x_{1,n}$	$x_{2,n}$

The causal effect is $y_i(1) - y_i(0)$

Maths of Causal Inference

- Average Treatment Effect $\mathbb{E}\big[Y(1)-Y(0)\big]$
- Sample Average Treatment Effect $\displaystyle\frac{1}{n}\sum_{i=1}^{n}\big[y_i(1)-y_i(0)\big]$

Assumption : $(Y(1),Y(0))\perp\!\!\!\perp T$

Crude estimator (difference in means),
$\displaystyle\widehat{\tau}=\frac{1}{n_1}\sum_{i:t_i=1}y_i-\frac{1}{n_0}\sum_{i:t_i=0}y_i$, or

$$\widehat{\tau}=\sum_{i=1}^{n}\frac{t_iy_i}{n_1}-\frac{(1-t_i)y_i}{n_0}$$

Then $\mathbb{E}\big[\widehat{\tau}\big]=\mathbb{E}\big[Y(1)-Y(0)\big]$

Maths of Causal Inference

- Local Average Treatment Effect $\mathbb{E}\big[Y(1) - Y(0)|\boldsymbol{X} = \boldsymbol{x}\big]$

The Propensity Score is the probability to receive the treatment
$\pi(\boldsymbol{x}) = \mathbb{P}[T = 1|\boldsymbol{X} = \boldsymbol{x}]$
Assumption: Balancing property $T \perp\!\!\!\perp \boldsymbol{X}|\pi(\boldsymbol{X})$
Assumption: Exogeneity $(Y(1), Y(0)) \perp\!\!\!\perp T|\pi(\boldsymbol{x}), \forall \boldsymbol{x}$
Consider here

$$\widehat{\tau} = \sum_{i=1}^{n} \frac{t_i y_i}{n\widehat{\pi}(\boldsymbol{x}_i)} - \frac{(1-t_i)y_i}{n(1-\widehat{\pi}(\boldsymbol{x}_i))}$$

see The Central Role of the Propensity Score in Observational Studies for Causal Effects

Looking for a Counterfactual

Modal discourse concerns alternative ways things can be, e.g., what might be true, what isn't true but could have been, what should be done (via Stanford Encyclopedia of Philosophy)

Impact of an increase of minimum wage on unemployment (see Card & Krueger (1994), data minwage.csv))

In 1992, New Jersey (NJ) raised the minimum wage from $4.25 to $5.05 (per hour).

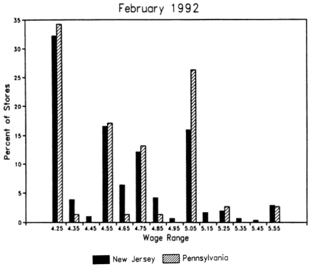
- name of fast-food restaurant chain
- location NJ (central, north, shore, south) & PA
- wage before minimum-wage increase
- wage after minimum-wage increase
- number of full-time employees(before and after)
- number of part-time employees (before and after)

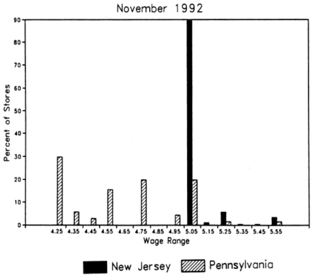
Looking for a Counterfactual

One can look at a counterfactual, with a neighboring state - Pennsylvania (PA) - our control group
cross-section comparison design

	New Jersey, NJ		Pennsylvania, PA	
	mean	below $5.5	mean	below $5.5
before	$4.61	91.06%	$4.65	94.03%
after	$5.08	0.34%	$4.61	95.52%

See Card & Krueger (1994)





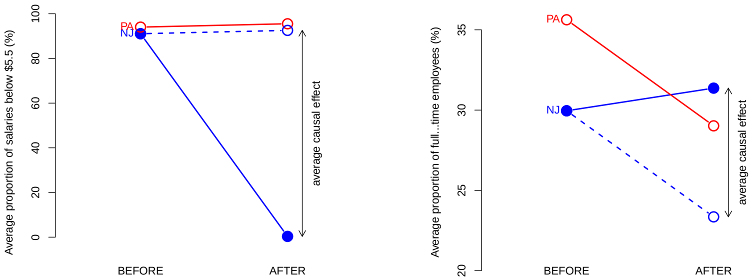
Looking for a Counterfactual

	New Jersey, NJ			Pennsylvania, PA		
	partial	full	proportion	partial	full	proportion
before	5423	2319	29.9%	1291	714	35.6%
after	5351	2446	31.4%	1339	547	29.0%

Difference-in-differences estimate,

$$DID = \underbrace{\overline{y}^{\text{after}}_{t=1} - \overline{y}^{\text{before}}_{t=1}}_{\text{difference in the treatment group}} - \underbrace{\overline{y}^{\text{after}}_{t=0} - \overline{y}^{\text{before}}_{t=0}}_{\text{difference in the control group}}$$

Observation and Experiment



(the counterfactual outcome for the treatment group has a time trend parallel to that of the control group)

Discontinuity

Introduced in Regression-discontinuity analysis: An alternative to the ex post facto experiment , to quantify the effects of college scholarships on later students' achievements

- X is SAT score
- Binary treatment T,

  receipt of scholarship,

  $$T_i = \mathbf{1}(X_i \geq c) = \begin{cases} 1 \text{ if } X_i \geq c \\ 1 \text{ if } X_i < c \end{cases}$$

- Outcome Y

  (e.g. subsequent earnings)

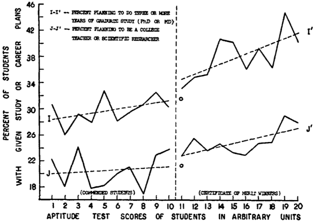


Fig. 3. Regression of study and career plans on exposure determiner.

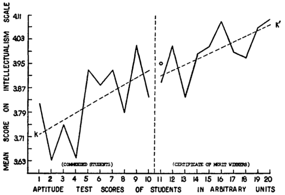
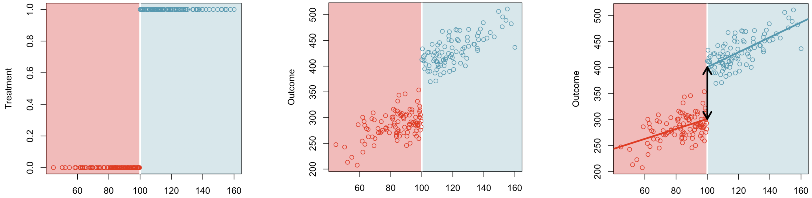


Fig. 4. Regression of attitudes toward intellectualism on exposure determiner.

Discontinuity

D_i and Y_i against X_i, and two regressions,

$$Y_i = \begin{cases} \alpha^+ + \beta^+ X_i \text{ if } X_i \geq c \text{ (i.e. } D_i = 1) \\ \alpha^- + \beta^- X_i \text{ if } X_i < c \text{ (i.e. } D_i = 0) \end{cases}$$



Here the Local Average Treatment Effect is

$$\mathbb{E}[Y(1) - Y(0)|X = c] = (\alpha^+ - \alpha^-) + (\beta^+ - \beta^-) \cdot c$$

Why only linear regression ? See Regression Discontinuity Designs

Discontinuity

One can consider more complex regression (quadratic, nonlinear) but the discontinuity can be artificial...

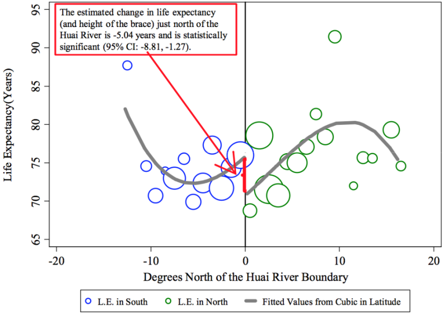


Fig. 3. The plotted line reports the fitted values from a regression of life expectancy on a cubic in latitude using the sample of DSP locations, weighted by the population at each location.

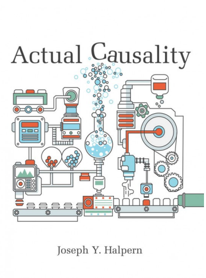
via evidence on the deleterious impact of sustained use of polynomial regression on causal inference.

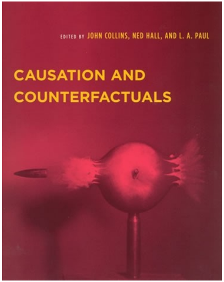
Looking for applications

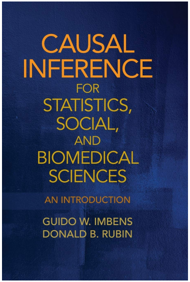
See

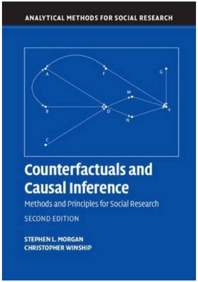
- Are Emily and Greg More Employable than Lakisha and Jamal?
- Attitudes toward Highly Skilled and Low-skilled Immigration
- What Triggers Public Opposition to Immigration?
- Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement
- After Midnight: A Regression Discontinuity Design in Length of Postpartum Hospital Stays

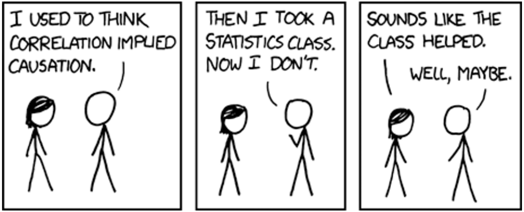
References











Source: https://xkcd.com/552/