

因果推断：一种思维方式

禹昂

University of Wisconsin-Madison

October 30, 2021

Outline

- 因果推断的几大传统

- 潜在结果 (potential outcomes)

- 有向无环图 (DAG)

- 实例

- 因果视角下的多元回归

- 控制变量：越多越好吗？

- 交互项中的两个变量是对称的吗？

- 2021 年诺奖：工具变量与局部平均效应 (LATE)

- 拓展

- 教科书，从入门到专题

- 两位华人青年因果推断新星

因果推断的传统

代表人物	学科	流派
Rubin, Imbens, Angrist	统计, 经济	potential outcomes
Pearl	计算机	directed acyclic graph
Heckman	经济	parametric control function
Robins	生统, 流病	marginal structural model structural nested mean model g-formula
Rosenbaum, VanderWeele	统计, 流病	sensitivity analysis
Manski	经济	partial identification
Spirites, Glymour	哲学	causal discovery

Potential outcomes

- $Y(D=1)_i; Y(D=0)_i$

假如（通过外在干预）上了清华，收入是多少；假如（通过外在干预）没上清华，收入是多少

- 个体层次的处理效应

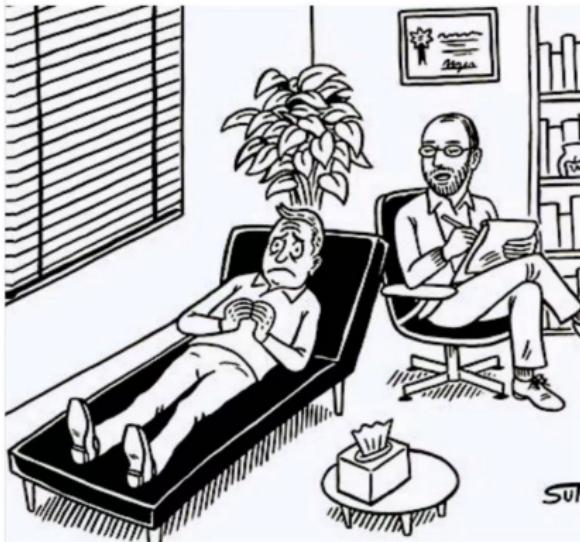
$$\tau_i = Y(D=1)_i - Y(D=0)_i$$

- 因果推断的根本问题：个体层次的处理效应不可观测
- 群体层次上的平均效应：ATE, ATT, ATU

D	$Y(D=0)$	$Y(D=1)$	Effect
0	✓	X	τ_1
0	✓	X	τ_2
1	X	✓	τ_3
1	X	✓	τ_4



Peter Hull @instrumenthull · 14h



And are the potential outcomes in the room with us now?



Anton Strezhnev @a_strezh · 10/15/21

...

Replying to @instrumenthull

Only one.



1



5



用多元回归识别因果效应？

- 实践中社会学家往往将多元回归模型中的系数阐释为（因果）效应
- 变量到底是怎么被“控制”住的？“when other variables are held constant”, “controlling for ...”, “ceteris paribus”, “after accounting for other factors”, “conditional on”
- 多元回归模型只是一种 optimization/inversion 算法。



用多元回归识别因果效应

$$E[Y(D_i = 1)_i] - E[Y(D_i = 0)_i] \quad \text{non-interference}$$

$$= \sum E[Y(D_i = 1)_i | X_i = x] P(X_i = x) - \sum E[Y(D_i = 0)_i | X_i = x] P(X_i = x)$$

$$= \sum E[Y(D_i = 1)_i | X_i = x, D_i = 1] P(X_i = x) - \sum E[Y(D_i = 0)_i | X_i = x, D_i = 0] P(X_i = x) \quad \text{conditional ignorability}$$

$$= \sum E[Y_i | X_i = x, D_i = 1] P(X_i = x) - \sum E[Y_i | X_i = x, D_i = 0] P(X_i = x)$$

positivity **consistency**
functional form

conditional ignorability: $E(Y_i | D_i = 1, X_i = x) = E(Y(D_i = 1)_i | D_i = 0, X_i = x)$

consistency: $E[Y(D_i = d) | D_i = d] = E[Y_i | D_i = d]$

positivity: $0 < P(D_i = d | X_i = x) < 1$

functional form: $E[Y_i | X_i = x, D_i = d] = \alpha + \beta d + \gamma x$

用多元回归识别因果效应：总结

- 用多元回归来识别平均处理效应依赖于多种因果假设和函数形式假设。
- 每种假设都可以被其他假设替代。但无假设不因果。因果推断是关于使用可信的假设推测反事实结果的学科。
- 传统统计是关于上面最后一步的：回归模型是对 *conditional expectation function* 的估计。
- 因果性存在于假设中，并不存在于统计方法中。估计是否因果，取决于假设是否成立。统计方法之间没有谁更因果一说。

用多元回归识别因果效应：总结

- 对任何社会学实证问题，我们应该从估计目标 (estimand) 出发，基于可信的因果假设，获取估计方法 (estimator)。

直接从回归模型 (an estimator) 出发是本末倒置的：回归模型是一个答案，但问题是什么？从问题出发，我们才能灵活地寻找回答问题的方式。

- 不要阐释控制变量的系数。



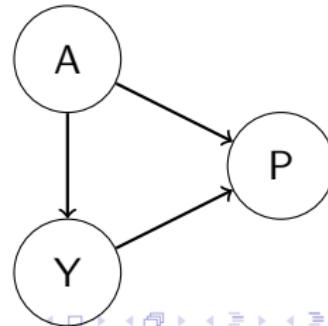
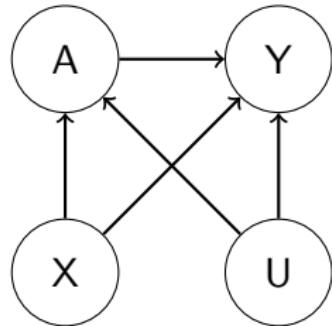
Lundberg, Ian, Rebecca Johnson, and Brandon M. Stewart. "What is your estimand? Defining the target quantity connects statistical evidence to theory." *American Sociological Review* 86.3 (2021): 532-565.



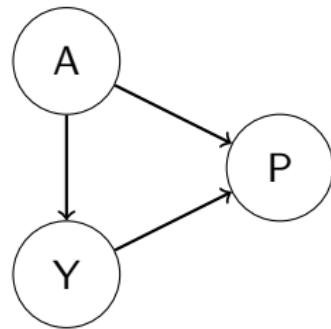
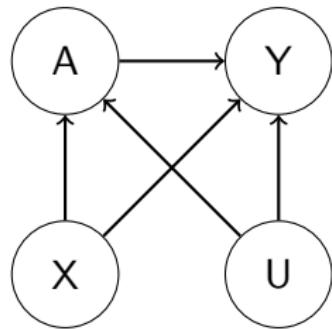
Keele, Luke, Randolph T. Stevenson, and Felix Elwert. "The causal interpretation of estimated associations in regression models." *Political Science Research and Methods* 8.1 (2020): 1-13.

DAG

- 直观告诉我们两个变量间的因果关系（在给定的因果假设下）是否可以识别：图形化的因果假设
- Path: 变量间的任何路径；Causal path: 所有箭头都指向处理变量，指向结果变量
- Collider: 一个变量在一条路径上被两个箭头指向
- Open path: 路径上的非 collider 没有被 condition on, collider 则被 condition on 了
- 因果效应可识别性：当两个变量之间所有的 open path 都是 causal path



控制变量：越多越好吗？



- 控制变量的选择应该基于对变量间因果关系的假设
- 一般而言，要控制发生在处理变量之前的变量
- 一般而言，不要控制发生在处理变量之后的变量

控制变量：越多越好吗？

- 但一般以外，也有很多例外
- 有时候，发生在处理变量之前的变量也可以是 collider.
Hünermund, Paul, Beyers Louw, and Itamar Caspi. "Double Machine Learning and Bad Controls—A Cautionary Tale." arXiv preprint arXiv:2108.11294 (2021).
- 有时候，控制发生在处理变量之后的变量也可以降低估计偏误。
Elwert, Felix, and Fabian T. Pfeffer. "The future strikes back: using future treatments to detect and reduce hidden bias." Sociological Methods & Research (2019).
Groenwold, Rolf HH, Tom M. Palmer, and Kate Tilling. "To Adjust or Not to Adjust? When a “Confounder” Is Only Measured After Exposure." Epidemiology 32.2 (2021).

关于 DAG 入门的推荐阅读

Elwert, Felix. "Graphical causal models." *Handbook of causal analysis for social research*. Springer, Dordrecht, 2013. 245-273.

Elwert, Felix, and Christopher Winship. "Endogenous selection bias: The problem of conditioning on a collider variable." *Annual review of sociology* 40 (2014): 31-53.

交互项的对称性问题

$$Y = \alpha + \beta_1 A + \beta_2 M + \beta_3 AM + \beta_4 X + \epsilon$$

- A 调节 M 和 M 调节 A 是一回事吗？

交互项的对称性问题

$$Y = \alpha + \beta_1 A + \beta_2 M + \beta_3 AM + \beta_4 X + \epsilon$$

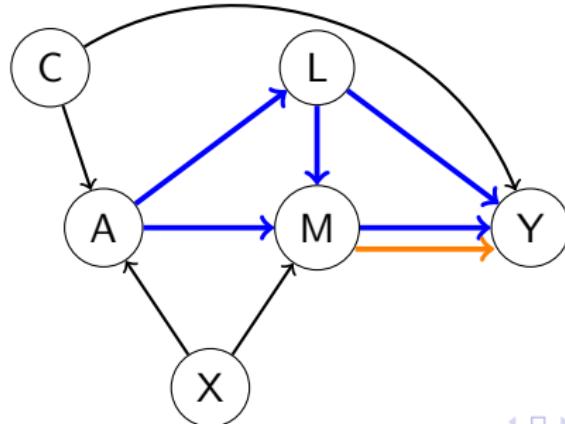
- A 调节 M 和 M 调节 A 是一回事吗？
- 作为一种回归模型中的统计工具，交互项当然是对称的。
- 同一个交互项，可以对应不同的 estimand，进而有着不同的对称性。
- 潜在结果和 DAG 在这一问题上都很有用。

交互项背后的 estimand 可能有哪些？

- Causal interaction effect

$$\underbrace{E[Y(A=1, M=1)] - E[Y(A=0, M=1)]}_{\text{干预使得所有人上清华时农村出身的效应}} \\ - \underbrace{\{E[Y(A=1, M=0)] - E[Y(A=0, M=0)]\}}_{\text{干预使得所有人不上清华时农村出身的效应}}$$

- 两个变量在 estimand 中都是被干预的对象，都是潜在结果的定义的一部分

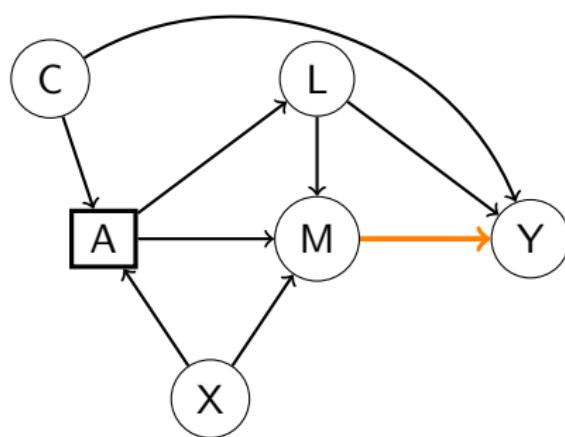


交互项背后的 estimand 可能有哪些？

- Effect modification

$$E[Y(M=1) - Y(M=0)|A=1] - E[Y(M=1) - Y(M=0)|A=0]$$

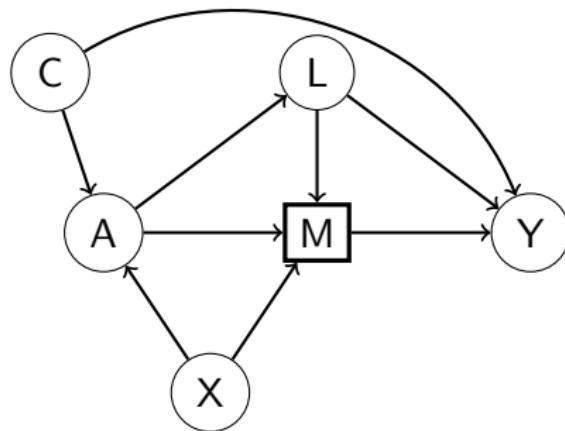
- 只有 M 在 estimand 中是被干预的对象；潜在结果只以 M 来定义：A 调节 M 的效应



交互项背后的 estimand 可能有哪些？

- 可以相反地定义 M 调节 A 的效应吗？

$$E[Y(A=1) - Y(A=0)|M=1] - E[Y(A=1) - Y(A=0)|M=0]$$



交互项背后的 estimand 可能有哪些：总结

- 所以，当 estimand 是因果交互效应时，两个变量的地位是对称的，都是因果变量；当 estimand 是调节效应时，只有其一是因果变量。
- 回归模型里的交互项可以在不同的因果假设下识别不同的 estimand。
- 不同的 estimand 意味着不同的因果假设和估计方法。所以，我们应当从 estimand 出发推出 estimator，而非相反。
- Further readings

VanderWeele, Tyler J. "On the distinction between interaction and effect modification." *Epidemiology* 20.6 (2009): 863-871.

Keele, Luke, and Randolph T. Stevenson. "Causal interaction and effect modification: same model, different concepts." *Political Science Research and Methods* 9.3 (2021): 641-649.

2021 年诺奖：工具变量与局部平均效应 (LATE)

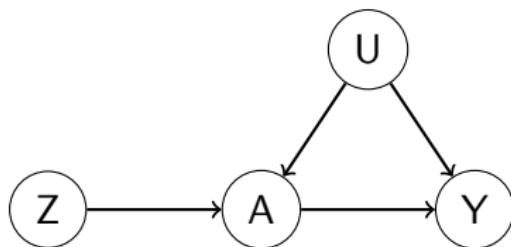
David Card, “for his empirical contributions to labour economics”

Joshua D. Angrist and Guido W. Imbens, “for their methodological contributions to the analysis of causal relationships”



2021 年诺奖：工具变量与局部平均效应 (LATE)

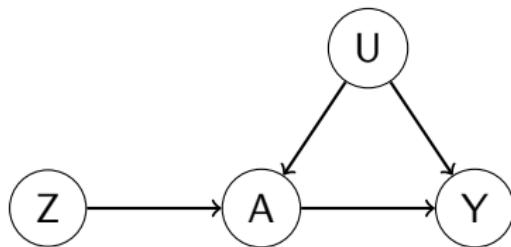
- 类似于我们为多元回归和交互项寻找 estimand, Angrist and Imbens (1994), Angrist, Imbens, and Rubin (1996) 为 IV estimator 找到了一个 estimand。
- IV estimator: $\frac{\text{Cov}(Z, Y)}{\text{Cov}(Z, A)}$
- Local (complier) average treatment effect:
 $E[Y(A = 1) - Y(A = 0)|A(Z = 1) - A(Z = 0) = 1]$



工具变量与局部平均效应 (LATE)

因果假设：

- Consistency
- Random assignment of Z : $A(Z = z) \perp\!\!\!\perp Z$; $Y(Z = z, A = a) \perp\!\!\!\perp Z$ (but see Pearl (2009), p.248)
- Exclusion restriction: $Y(Z = 1, A = a) = Y(Z = 0, A = a)$
- Relevance: $E[A(Z = 1) - A(Z = 0)] > 0$
- Monotonicity: $A_i(Z_i = 1) \geq A_i(Z_i = 0)$



工具变量与局部平均效应 (LATE): 拓展

- 如果处理效应对所有人都是一样的呢？

再次反映出相同的统计工具，基于不同的因果假设，可以识别不同的 estimand

- 可以用 DAG 来表示 LATE 吗？

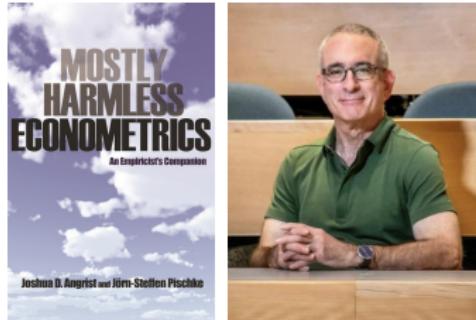
Steiner, Peter M., et al. "Graphical models for quasi-experimental designs." *Sociological methods & research* 46.2 (2017): 155-188.

- 工具变量可以用来估计 ATE 和 ATT 吗？

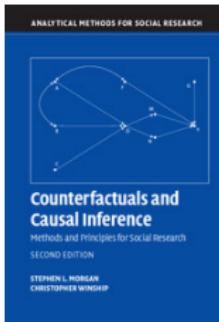
Zhou, Xiang, and Yu Xie. "Heterogeneous treatment effects in the presence of self-selection: a propensity score perspective." *Sociological Methodology* 50.1 (2020): 350-385.

因果推断教科书：从入门到专题

Mostly Harmless Econometrics: An Empiricist's Companion



Counterfactuals and Causal Inference: Methods and Principles for Social Research

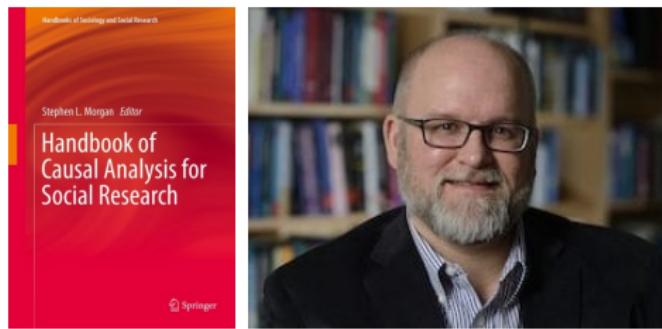


因果推断教科书：从入门到专题

Causal Inference for Statistics, Social, and Biomedical Sciences

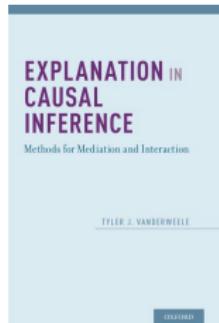


Handbook of Causal Analysis for Social Research

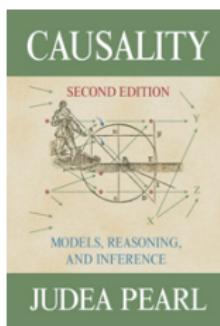


因果推断教科书：从入门到专题

Explanation in causal inference: methods for mediation and interaction



Causality: Models, Reasoning and Inference



因果推断教科书：从入门到专题

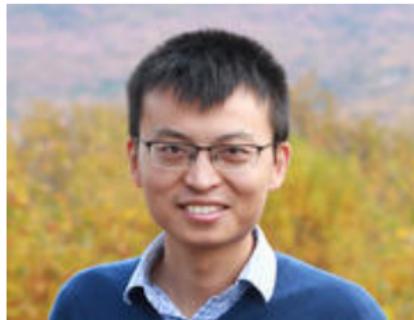
Causal Inference: The Mixtape



Causal Inference: What If



两位华人因果推断新星



- Xiang Zhou
- Causal mediation analysis;
marginal treatment effect



- Weihua An
- Peer effects; treatment
interference

总结

- Estimand 至关重要，从理论到方法的过程即是从 estimand 到 estimator 的过程 (formalized theorization)。
- 因果推断回答关于反事实的问题，为政策干预提供直接的证据。
- 因果推断的理论传统有很多，潜在因果和 DAG 的框架在实践中可以帮助我们厘清因果 estimand，所需的因果假设，以及适用的估计方法。
- 在四个实例中，我们意识到
多元回归用于估计 ATE 依赖于一系列因果假设和函数形式假设。
控制变量的选择依赖于因果假设。
因果交互效应不同于调节效应。
工具变量在一系列因果假设下可以用于估计局部平均处理效应。

总结

谢谢！

@AngYu_soci
ayu33@wisc.edu

