Considerações sobre Inferência Causal & LInCa

Marcelo Magalhães Taddeo

marcelo.magalhaes@ufba.br

28 de setembro de 2023

Laboratório de Inferência Causal

- Marcelo M. Taddeo (IME/UFBa)
- Leila Amorim (IME/UFBa)
- Rosemeire L. Fiaccone (IME/UFBa)
- Lilia Costa (IME/UFBa)
- Raydonal Ospina (IME/UFBa)
- Rosana Aquino (ISC/UFBa)
- Vinicius Mendes (FE/UFBa)
- Elzo Júnior (Cidacs/Fiocruz)

+ colaboradores e estudantes de graduação e pós-graduação.

LInCa - Grupo CNPq

Endereço para acessar este espelho: dgp.cnpq.br/dgp/espelhogrupo/9296811361381732

LInCa - Pesquisa

Repercussões dos trabalhos do grupo

O LInCa é um grupo interdisciplinar com enfoque intersetorial composto de pesquisadores com vasta experiência no desenvolvimento de metodologias estatísticas (inferência estatística, análise de dados longitudinais e de sobrevivência, aprendizado de máquina, inferência bayesiana, séries temporais etc.) e com forte componente aplicado, especialmente, em bioestatística, epidemiologia e econometria. Os pesquisadores deste grupo têm avançado na produção científica sobre modelagem estatística e suas aplicações em diversas áreas e estão inseridos em ampla rede de colaborações nacionais e internacionais. O LInCa se propõe a estudar e desenvolver metodologias estatísticas para a inferência causal, especialmente em de estudos observacionais (não aleatorizados), e suas aplicações nas análises de dados reais, e também na formação de recursos humanos para atuar na avaliação do impacto de políticas públicas, facilitando a tomada de decisões.

Linhas de pesquisa

Nome da linha de pesquisa	Quantidade de Estudantes	Quantidade de Pesquisadores
Análise de mediação causal	0	2
Avaliação de impacto em epidemiologia e econometria	4	10
Métodos estatísticos em inferência causal	3	7

LINCA - PESQUISA

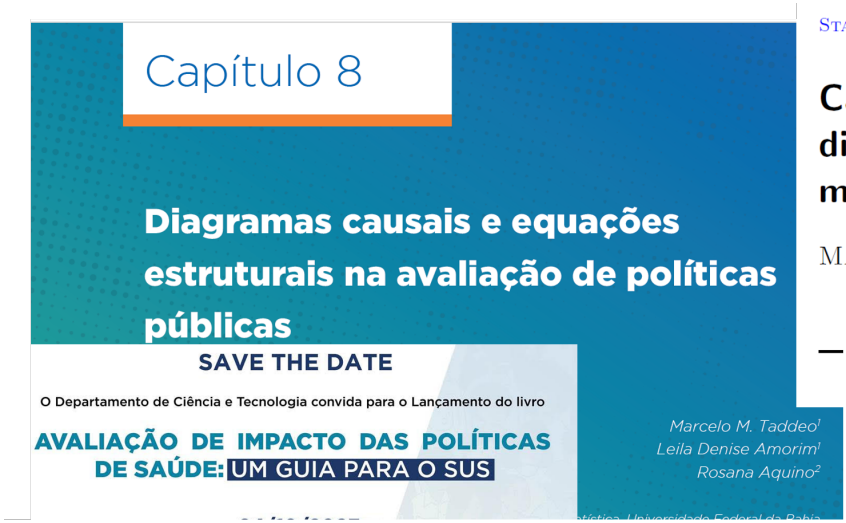
ÁREAS DE INTERESSE:

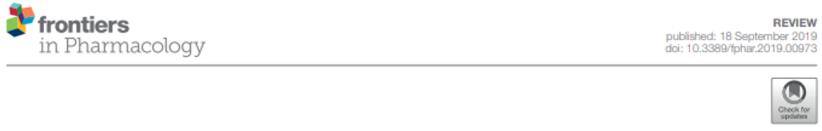
1. Metodologias de identificação causal em diferentes contextos
   (a) Mediação causal
   (b) Dados de sobrevivência
   (c) Modelos com variáveis latentes
   (d) Dados agrupados
   (e) Dados funcionais

2. Metodologias de estimação
   (a) Variações e extensões dos métodos clássicos
   (b) Métodos bayesianos
   (c) Técnicas de machine learning no aprendizado causal

3. Aplicações
   (a) Epidemiologia
   (b) Econometria
   (c) Educação

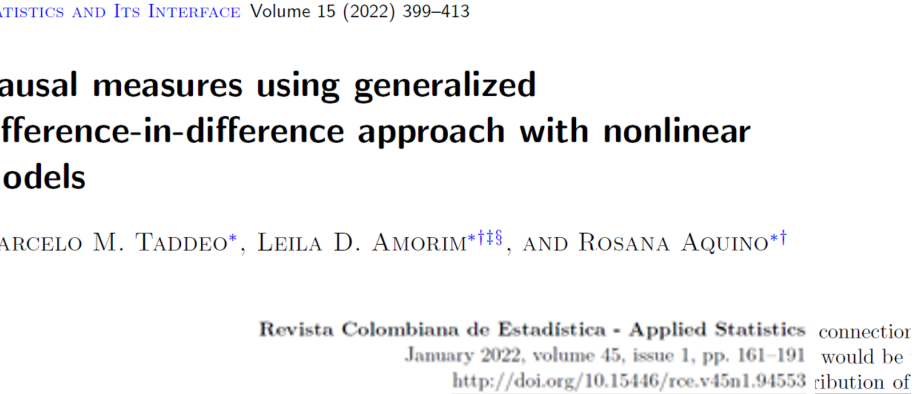
LINCA - PRODUÇÃO





Propensity Score Methods in Health Technology Assessment: Principles, Extended Applications, and Recent Advances

M Sanni Ali, Daniel Prieto-Alhambra, Luciane Cruz Lopes, Dandara Ramos, Nivea Bispo, Maria Y. Ichihara, Julia M. Pescarini, Elizabeth Williamson, Rosemeire L. Fiaccone, Mauricio L. Barreto and Liam Smeeth



Motivação Geral

Por que entender, caracterizar e determinar relações de causalidade?

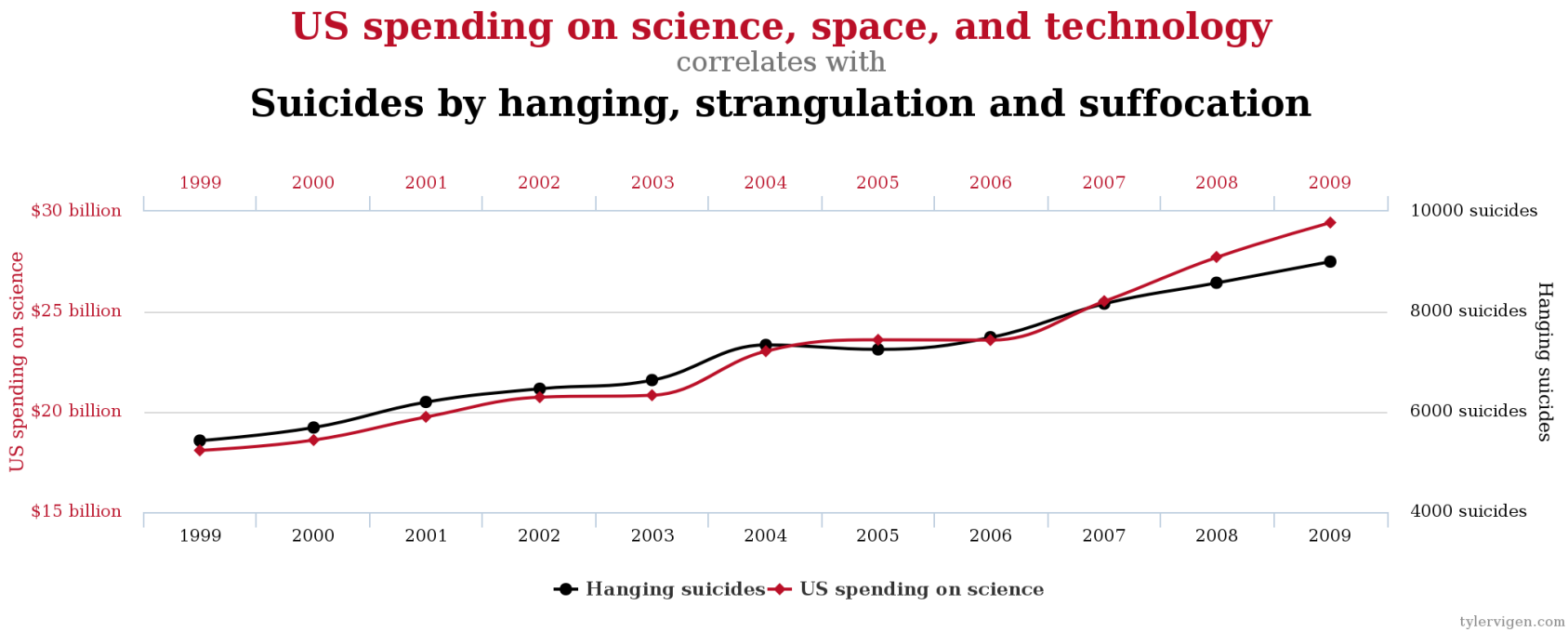
Propósito central da ciência: explicar e prever fenômenos de interesse.

> Popper: o cientista é movido pela busca de explicações causais com o objetivo "de encontrar *teorias explicativas* (...) que descrevam certas propriedades estruturais do mundo e que nos permitam deduzir, com o auxílio de condições iniciais, os efeitos que se pretende explicar."

$\Rightarrow$ encontrar leis (universais) tais que, combinadas com certas condições iniciais (**causas**), permitam explicar determinados acontecimentos (**efeitos**).

$\Rightarrow$ identificar propriedades causais de um conjunto **especificado** de estruturas, variáveis ou circunstâncias.

CORRELAÇÃO E CAUSAÇÃO



Fonte: U.S. Office of Management and Budget and Centers for Disease Control & Prevention

- http://www.tylervigen.com/spurious-correlations

> "The aim of standard **statistical analysis**, typified by regression, estimation, and hypothesis testing techniques, is to assess parameters of a distribution from samples drawn of that distribution. (...)" (Pearl)

TRADUZINDO:

- Inferência estatística é sobre os parâmetros da distribuição populacional...
- ... e não sobre relações causais.

> "The aim of standard **statistical analysis**, typified by regression, estimation, and hypothesis testing techniques, is to assess parameters of a distribution from samples drawn of that distribution. (...)" (Pearl)

TRADUZINDO:

- Inferência estatística é sobre os parâmetros da distribuição populacional...
- ... e não sobre relações causais.

> "**Causal analysis** (...) [aims] to infer not only beliefs or probabilities under static conditions, but also the dynamics of beliefs under **changing conditions**, for example, changes induced by **treatments** or **external interventions**" (Pearl)

> "The aim of standard **statistical analysis**, typified by regression, estimation, and hypothesis testing techniques, is to assess parameters of a distribution from samples drawn of that distribution. (...)" (Pearl)

TRADUZINDO:

- Inferência estatística é sobre os parâmetros da distribuição populacional...
- ... e não sobre relações causais.

> "**Causal analysis** (...) [aims] to infer not only beliefs or probabilities under static conditions, but also the dynamics of beliefs under **changing conditions**, for example, changes induced by **treatments** or **external interventions**" (Pearl)

- Isto não é totalmente verdadeiro... *(por quê?)*

JERZY NEYMAN (1923)

Statistical Science
1990, Vol. 5, No. 4, 465–480

On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9.

Jerzy Splawa-Neyman

Translated and edited by D. M. Dabrowska and T. P. Speed from the Polish original, which appeared in Roczniki Nauk Rolniczych Tom X (1923) 1–51 (Annals of Agricultural Sciences)

m: # áreas / campos cultiváveis (plots)

U_k: rendimentos **observados** nos campos

$$k = 1, \ldots, m$$

n: # variedades de culturas

$\hookrightarrow$ cada campo contém só uma variedade!

Objetivo: comparar os rendimentos das culturas.

- Para cada cultura $i = 1, \ldots, n$, defina

$$U_{i1}, \ldots, U_{im}.$$

- Retorno médio da variedade i:

$$a_i = \frac{1}{m} \sum_{k=1}^{m} U_{ik}.$$

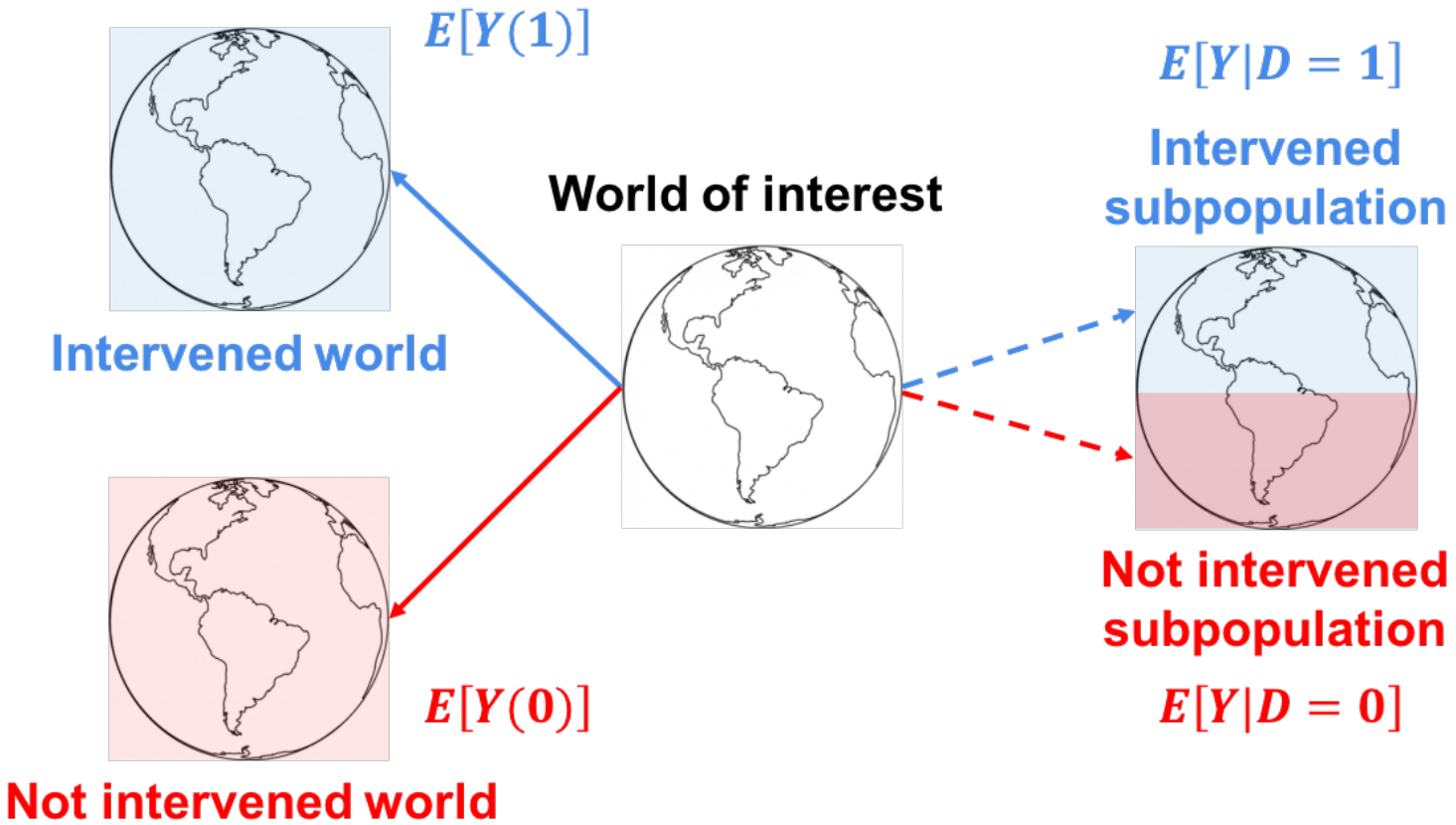
> "The goal of a field experiment which consists of the comparison of v varieties will be regarded as equivalent to the problem of comparing the numbers
>
> $$a_1, \ldots, a_n.$$
>
> or their estimates (...)."

Obs.: para cada i, somente um U_{ik} é observado...

- os outros são contrafactuais!

CONTRAFACTUAIS

[A cause is] *"an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second. Or, in other words, where, if the first object had not been, the second never had existed." (Hume, 1748)*



RUBIN (1974)[1]

A: atribuição de tratamento

- Para Rubin, tratamento $\equiv$ causa

$A = 1 \Rightarrow$ 'Tratamento'

$A = 0 \Rightarrow$ 'Controle'

Y: resposta de interesse

$Y(1)$: resposta quado a unidade é exposta ao tratamento;

$Y(0)$: resposta quado a unidade é exposta ao controle.

Objetivo: avaliar as quantidades $Y(1)$ e $Y(0)$ ou contrastes entre elas em estudos observacionais.

[1] Rubin, D. B. (1974), "Estimating causal effects of treatments in randomized and nonrandomized studies," J. Educ. Psychology, 66, 688–701.

RUBIN (1974)[1]

A: atribuição de tratamento

- Para Rubin, tratamento $\equiv$ causa

$A = 1 \Rightarrow$ 'Tratamento'

$A = 0 \Rightarrow$ 'Controle'

Y: resposta de interesse

$Y(1)$: resposta quado a unidade é exposta ao tratamento;

$Y(0)$: resposta quado a unidade é exposta ao controle.

Objetivo: avaliar as quantidades $Y(1)$ e $Y(0)$ ou contrastes entre elas em estudos observacionais.

PROBLEMA FUNDAMENTAL DA INFERÊNCIA CAUSAL

É IMPOSSÍVEL **OBSERVAR** SIMULTANEAMENTE $Y(1)$ E $Y(0)$! UMA DESSAS QUANTIDADES É OBRIGATORIAMENTE CONTRAFACTUAL.

[1] Rubin, D. B. (1974), "Estimating causal effects of treatments in randomized and nonrandomized studies," J. Educ. Psychology, 66, 688–701.

RUBIN (1974)[1]

A: atribuição de tratamento

- Para Rubin, tratamento $\equiv$ causa

$A = 1 \Rightarrow$ 'Tratamento'

$A = 0 \Rightarrow$ 'Controle'

Y: resposta de interesse

$Y(1)$: resposta quado a unidade é exposta ao tratamento;

$Y(0)$: resposta quado a unidade é exposta ao controle.

Objetivo: avaliar as quantidades $Y(1)$ e $Y(0)$ ou contrastes entre elas em estudos observacionais.

PROBLEMA FUNDAMENTAL DA INFERÊNCIA CAUSAL

É IMPOSSÍVEL **OBSERVAR** SIMULTANEAMENTE $Y(1)$ E $Y(0)$! UMA DESSAS QUANTIDADES É OBRIGATORIAMENTE CONTRAFACTUAL.

PROBLEMA REFORMULADO

QUAIS CONDIÇÕES DEVEM SER SATISFEITAS E QUAIS MÉTODOS PODEM SER USADOS PARA **IDENTIFICAR** AS DISTRIBUIÇÕES DE $Y(1)$ E $Y(0)$ E QUANTIDADES RELACIONADAS EM ESTUDOS OBSERVACIONAIS?

[1] Rubin, D. B. (1974), "Estimating causal effects of treatments in randomized and nonrandomized studies," J. Educ. Psychology, 66, 688–701.

Efeitos Causais

- Def.: qualquer avaliação estatística de $Y(a)$.
- Exemplo:

$$\mathrm{P}(Y = y|\,\mathrm{do}(A = a)).$$

- Tipicamente: qualquer contraste $\tau(Y(0), Y(1); X, A)$ entre $Y(0)$ e $Y(1)$.
- Ainda mais tipicamente: contrastes em termos de médias.
- Average Treatment Effect (ATE):

$$\mathrm{E}[Y(1) - Y(0)],$$

ou sua contraparte para populações finitas

$$\frac{1}{N}\sum_{i=1}^{N}\left(Y_i(1) - Y_i(0)\right).$$

- Inclusão de covariáveis:

$$\mathrm{E}[Y(1) - Y(0)|X = x] \quad \text{ou} \quad \frac{1}{N}\sum_{i:X_i=x}\left(Y_i(1) - Y_i(0)\right).$$

- Caso especial: Average Treatment (Effect) on the Treated (ATT)

$$\mathrm{E}[Y(1) - Y(0)|A = 1] \quad \text{ou} \quad \frac{1}{N_1}\sum_{i:A_i=1}\left(Y_i(1) - Y_i(0)\right).$$

Efeitos Causais — Outras medidas

- Se Y é binária:

$$\mathrm{E}[Y(1) - Y(0)] = \mathrm{P}(Y(1) = 1) - \mathrm{P}(Y(0) = 1),$$

- ou o risco relativo

$$\frac{\mathrm{P}(Y(1) = 1)}{\mathrm{P}(Y(0) = 1)},$$

- ou a razão de chances

$$\frac{\mathrm{P}(Y(1) = 1)/\mathrm{P}(Y(1) = 0)}{\mathrm{P}(Y(0) = 1)/\mathrm{P}(Y(0) = 0)}.$$

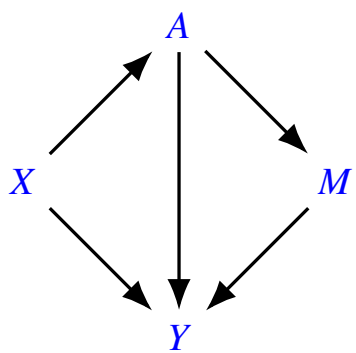
- As mesmas quantidades podem ser avaliadas entre os tratados, e.g.,

$$\frac{\mathrm{P}(Y(1) = 1|A = 1)}{\mathrm{P}(Y(0) = 1|A = 1)}.$$

DIAGRAMAS CAUSAIS — UMA OUTRA PERSPECTIVA

MÉTODO GRÁFICO PARA REPRESENTAR RELAÇÕES CAUSAIS ENTRE VARIÁVEIS.

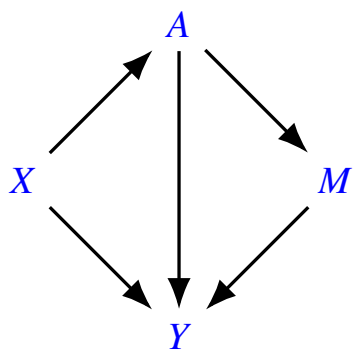
- Um **grafo** $\mathcal{G}$ é um par $(\mathcal{V}, \mathcal{A})$, onde $\mathcal{V}$ é um conjunto finito de **vértices** e $\mathcal{A} \subset \mathcal{V} \times \mathcal{V}$ um conjunto de **arestas** em $\mathcal{G}$.
- Este grafo representa as relações causais entre as variáveis envolvidas no sentido de que $A \longrightarrow B$ significa A **causa** B.



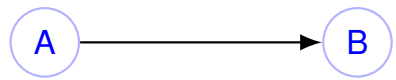
DIAGRAMAS CAUSAIS — UMA OUTRA PERSPECTIVA

MÉTODO GRÁFICO PARA REPRESENTAR RELAÇÕES CAUSAIS ENTRE VARIÁVEIS.

- Um **grafo** $\mathcal{G}$ é um par $(\mathcal{V}, \mathcal{A})$, onde $\mathcal{V}$ é um conjunto finito de **vértices** e $\mathcal{A} \subset \mathcal{V} \times \mathcal{V}$ um conjunto de **arestas** em $\mathcal{G}$.
- Este grafo representa as relações causais entre as variáveis envolvidas no sentido de que $A \longrightarrow B$ significa A **causa** B.



- Uma aresta do tipo

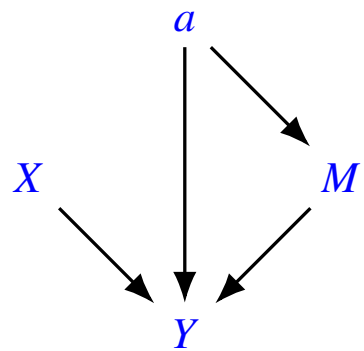


é chamada direcionada.
- Um grafo somente com arestas direcionadas e sem "loops" é chamado grafo acíclico direcionado (DAG).
- Cada estrutura $U \longrightarrow V$ é um mecanismo.
- A totalidade de mecanismos representa o fenômeno de interesse.

- Distribuição conjunta: $\mathrm{P}(a, m, x, y)$
- Decomposição induzida: $\mathrm{P}(a, m, x, y) = \mathrm{P}(y|a, m, x)\,\mathrm{P}(m|a)\,\mathrm{P}(a|x)\,\mathrm{P}(x)$

Diagramas Causais — Uma Outra Perspectiva

- Uma intervenção neste cotexto significa "fixar" uma aresta.
- Por exemplo, se $A \leftarrow a$, temos



- Notação: $(A \leftarrow a) \equiv \text{do}(a)$
- Distribuição "contrafactual": $\mathrm{P}^{\text{do}(a)}(m, x, y)$.

Problema reformulado (grafos causais)

Quais condições devem ser satisfeitas e quais métodos podem ser usados para avaliar $\mathrm{P}^{\text{do}(a)}(m, x, y)$ em estudos observacionais?

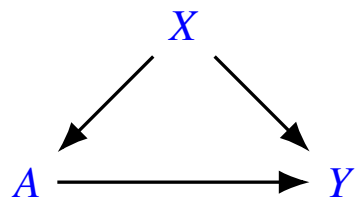
Capítulo 8

Diagramas causais e equações estruturais na avaliação de políticas públicas

Marcelo M. Taddeo[1]
Leila Denise Amorim[1]
Rosana Aquino[2]

[1]*Instituto de Matemática e Estatística, Universidade Federal da Bahia*
[2]*Instituto de Saúde Coletiva, Universidade Federal da Bahia*

IDENTIFICAÇÃO DO EFEITO CAUSAL



- A: Exposição / Tratamento
- X: Variáveis confundidoras (covariáveis)
- Y: Desfecho

PERGUNTA:

- Como estimar o efeito causal médio (ATE):

$$\mathrm{E}\left[Y(1) - Y(0)|X\right]?$$

HIPÓTESES:

Consistência: $Y = AY(1) + (1-A)Y(0)$

Ignorabilidade: $\{Y(0), Y(1)\} \perp\!\!\!\perp A|X$

E O ATE?

- Dadas as condições acima:

$$\begin{aligned}\mathrm{E}[Y(a)|X=x] &\overset{\text{Ign}}{=} \mathrm{E}[Y(a)|A=a, X=x] \\ &\overset{\text{Con}}{=} \mathrm{E}[Y|A=a, X=x].\end{aligned}$$

$$\begin{aligned}\therefore \mathrm{E}\,Y(a) &= \mathrm{E}\,\mathrm{E}[Y(a)|X] \\ &= \mathrm{E}\,\mathrm{E}[Y|A=a, X].\end{aligned}$$

- Ou seja,

$$\begin{aligned}\mathrm{E}\,Y(1) - \mathrm{E}\,Y(0) &= \mathrm{E}\,\mathrm{E}[Y|A=1, X] \\ &\quad - \mathrm{E}\,\mathrm{E}[Y|A=0, X].\end{aligned}$$

ATE $\neq$ ATT!

Sob as hipóteses de identificação (slide anterior):

$$\begin{aligned}
\text{ATT} &\overset{\text{Def}}{=} \text{E}[Y(1)|A=1] - \text{E}[Y(0)|A=1] \\
&= \text{E}[Y|A=1] - \text{E}\,\text{E}[Y(0)|A=1,X] \\
&\overset{\text{Ign}}{=} \text{E}[Y|A=1] - \text{E}\,\text{E}[Y(0)|A=0,X] \\
&\overset{\text{Con}}{=} \text{E}[Y|A=1] - \text{E}\,\text{E}[Y|A=0,X]
\end{aligned}$$

Ilustração:

- Desfecho: $Y = \beta_0 + \tau A + \beta_1 X + \varepsilon$

$\Rightarrow$ $\text{ATE} = \tau$.

- $\text{E}[Y|A=0,X] = \beta_0 + \beta_1 X$

$$\Rightarrow \text{E}\,\text{E}[Y|A=0,X] = \beta_0 + \beta_1\,\text{E}\,X.$$

- $\text{E}[Y|A=1] = \beta_0 + \tau + \beta_1\,\text{E}[X|A=1]$

$\Rightarrow$ $\text{ATT} = (\beta_0 + \tau + \beta_1\,\text{E}[X|A=1]) - (\beta_0 + \beta_1\,\text{E}\,X) = \tau + \beta_1\,\{\text{E}[X|A=1] - \text{E}\,X\}$

- Conclusão:

$$\text{ATT} - \text{ATE} = \beta_1\,\{\text{E}[X|A=1] - \text{E}\,X\}$$

e

$$\text{ATT} = \text{ATE} \quad \Leftrightarrow \quad \text{E}[X|A=1] = \text{E}\,X$$

Viés de Seleção

Sob a condição de consistência:

$$\begin{aligned}
\mathrm{E}[Y|A=1,x] - \mathrm{E}[Y|A=0,x] &= \mathrm{E}[Y(1)|A=1,x] - \mathrm{E}[Y(0)|A=0,x] \\
&= \mathrm{E}[Y(1)|A=1,x] - \mathrm{E}[Y(0)|A=1,x] \\
&\quad + \{\mathrm{E}[Y(0)|A=1,x] - \mathrm{E}[Y(0)|A=0,x]\} \\
&= \mathrm{ATT}(x) + \underbrace{\{\mathrm{E}[Y(0)|A=1,x] - \mathrm{E}[Y(0)|A=0,x]\}}_{\text{Viés de Seleção}}
\end{aligned}$$

- Portanto, a fórmula de identificação vale somente se NÃO houver viés de seleção!
- No caso do ATE, temos algo parecido...

$$\mathrm{E}[Y|A=1,x] - \mathrm{E}[Y|A=0,x] = \underbrace{\mathrm{E}[Y(1)|A=1,x]}_{\neq \mathrm{E}[Y(1)|x]} - \underbrace{\mathrm{E}[Y(0)|A=0,x]}_{\neq \mathrm{E}[Y(0)|x]}$$

Escore de Propensão (ATE)

- Se, além das condições usuais

  (A1) (Consistência) $Y = (1-A)Y(0) + AY(1)$

  (A2) (Ignorabilidade) $Y(0), Y(1) \perp\!\!\!\perp A|X$

  valer

  (A3) (Positividade) Para todo $x \in \mathcal{X}$, vale $0 < \mathrm{P}(A = 1|X = x) < 1$.

Sob as condições (A1) – (A2),

$$\mathrm{ATE} = \mathrm{E}\left[\frac{AY}{p(X)} - \frac{(1-A)Y}{1-p(X)}\right] = \mathrm{E}\left[\frac{A-p(X)}{p(X)(1-p(X))}Y\right].$$

- Estimadores:

(A) Horvitz-Thompson:

$$\widehat{\mathrm{ATE}} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{A_i Y_i}{\pi(X_i|\widehat{\theta})} - \frac{(1-A_i)Y_i}{1-\pi(X_i|\widehat{\theta})}\right\}$$

(B) IPTW: pesos padronizados

$$\omega_i = \frac{A_i}{\pi(X_i|\widehat{\theta})} \Big/ \sum_{j=1}^{n}\frac{A_j}{\pi(X_j|\widehat{\theta})}$$

(no caso dos tratados)

Escore de Propensão (ATT)

Sob as mesmas condições,

$$\begin{aligned}\text{ATT} &= \frac{1}{\text{E}A}\,\text{E}\left[AY - (1-A)\frac{p(X)}{1-p(X)}Y\right] \\ &= \frac{1}{\text{E}A}\,\text{E}\left[\frac{A-p(X)}{1-p(X)}Y\right].\end{aligned}$$

- Estimador: tomando

$$\widehat{\text{E}}A = \widehat{\text{P}}(A=1) = \frac{n_1}{n},$$

temos

$$\widehat{\text{ATT}} = \frac{1}{n_1}\left\{\sum_{i:A_i=1} Y_i - \sum_{i:A_i=0} \frac{\pi(X_i|\widehat{\theta})}{1-\pi(X_i|\widehat{\theta})}Y_i\right\}$$

Obs.: o ATT demanda uma condição de positividade mais fraca do que o ATE,

$$\text{P}(A=0|X=x) > 0,$$

para todo x observável entre os não-tratados.

Métodos de Identificação do Efeito Causal (ATE)

1. Regressão (paramétrica, não paramétrica)
2. Métodos baseados em escores de propensão:
   (a) Ponderação pelo inverso da probabilidade
   (b) Matching / Pareamento
3. Balanceamento de covariáveis (Calibração)
4. Métodos duplo robustos
5. Diferenças em diferenças (diff-in-diff)
6. RDD
7. Variáveis instrumentais
8. Modelos marginais estruturais (MSM) para dados longitudinais
9. Métodos bayesianos