

Causalidade e Estratégias de Identificação

Do modelo de Neyman–Rubin aos principais desenhos observacionais

Francis Petterini

UFSC

Roteiro da aula — 80 minutos

Parte I — Fundamentos (20 min)

- resultados potenciais;
- ATT e o problema do contrafactual;
- viés de seleção.

Parte II — Identificação (15 min)

- aleatorização;
- endogeneidade;
- instrumentos e 2SLS.

Parte III — Sem instrumento (35 min)

- PSM;
- Diferenças-em-Diferenças;
- RDD;
- Controle Sintético.

Parte IV — Fronteira aplicada (10 min)

- novos estimadores;
- heterogeneidade;
- transparência de desenho.

A pergunta causal

Considere um tratamento binário $T_i \in \{0, 1\}$.

Para cada unidade i , existem dois resultados potenciais:

$$Y_i(1) \text{ e } Y_i(0).$$

Efeito causal individual

$$\tau_i = Y_i(1) - Y_i(0).$$

O problema: para a mesma unidade, no mesmo momento, só observamos um dos dois resultados potenciais.

O problema fundamental da inferência causal

O resultado observado é

$$Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0).$$

T_i	observado	não observado
1	$Y_i(1)$	$Y_i(0)$
0	$Y_i(0)$	$Y_i(1)$

Contrafactual

A inferência causal é, essencialmente, um problema de construir uma comparação válida para aquilo que não foi observado.

Parâmetros causais

Efeito médio do tratamento

$$ATE = \mathbb{E}[Y(1) - Y(0)].$$

Efeito médio sobre os tratados

$$ATT = \mathbb{E}[Y(1) - Y(0) \mid T = 1].$$

Para avaliação de políticas, o ATT costuma ser particularmente natural:

qual foi o efeito da política sobre quem efetivamente recebeu a política?

A equação central do ATT

Partindo de

$$ATT = \mathbb{E}[Y(1) \mid T = 1] - \mathbb{E}[Y(0) \mid T = 1],$$

observamos diretamente

$$\mathbb{E}[Y(1) \mid T = 1],$$

mas não observamos

$$\boxed{\mathbb{E}[Y(0) \mid T = 1]}.$$

Toda estratégia de identificação precisa resolver isto

Como recuperar o resultado médio que os tratados teriam apresentado na ausência do tratamento?

Diferença de médias: efeito causal + viés

A diferença observada é

$$\mathbb{E}[Y \mid T = 1] - \mathbb{E}[Y \mid T = 0].$$

Somando e subtraindo $\mathbb{E}[Y(0) \mid T = 1]$:

$$\mathbb{E}[Y \mid T = 1] - \mathbb{E}[Y \mid T = 0] = \text{ATT} + \underbrace{\mathbb{E}[Y(0) \mid T = 1] - \mathbb{E}[Y(0) \mid T = 0]}_{\text{viés de seleção}}$$

A comparação simples entre tratados e não tratados é causal somente quando o segundo termo é zero.

O que é o viés de auto-seleção?

$$\text{Viés} = \mathbb{E}[Y(0) \mid T = 1] - \mathbb{E}[Y(0) \mid T = 0].$$

Ele aparece quando as unidades que recebem tratamento seriam diferentes das demais **mesmo sem tratamento**.

Exemplos:

- treinamento profissional escolhido por trabalhadores mais motivados;
- escola de tempo integral implantada em áreas mais vulneráveis;
- crédito subsidiado demandado por firmas com projetos de expansão.

Ponto-chave

Seleção para o tratamento mistura efeito causal com diferenças pré-existentes.

Por que a aleatorização resolve o problema?

Sob atribuição aleatória,

$$T \perp\!\!\!\perp \{Y(0), Y(1)\}.$$

Logo,

$$\mathbb{E}[Y(0) \mid T = 1] = \mathbb{E}[Y(0) \mid T = 0].$$

Portanto,

$$\mathbb{E}[Y \mid T = 1] - \mathbb{E}[Y \mid T = 0] = \text{ATT}$$

(e, sob o desenho adequado, também identifica o ATE).

A aleatorização constrói o contrafactual

O grupo de controle reproduz, em média, aquilo que teria ocorrido com os tratados sem a intervenção.

Quando não podemos aleatorizar

Em dados observacionais, frequentemente

$$\mathbb{E}[Y(0) \mid T = 1] \neq \mathbb{E}[Y(0) \mid T = 0].$$

Na linguagem de regressão, considere

$$Y_i = \alpha + \beta T_i + u_i.$$

Se fatores não observados que afetam Y_i também afetam T_i , então

$$\text{Cov}(T_i, u_i) \neq 0.$$

Tradução econométrica

O problema causal vira um problema de **endogeneidade do tratamento**.

Se existe um instrumento

Um instrumento Z deve deslocar o tratamento, mas não afetar diretamente o resultado.

Primeiro estágio:

$$T_i = \pi_0 + \pi_1 Z_i + X_i' \pi + v_i.$$

Segundo estágio:

$$Y_i = \alpha + \beta \hat{T}_i + X_i' \gamma + u_i.$$

Cuidado com a interpretação

Com heterogeneidade de efeitos, IV/2SLS tende a identificar um efeito local para os *compliers* (LATE), não necessariamente o ATT.

E sem instrumento?

Quatro estratégias clássicas cobrem grande parte da avaliação aplicada:

- 1 **Matching / Propensity Score**: construir comparabilidade em observáveis;
- 2 **Diferenças-em-Diferenças**: usar evolução temporal de um grupo de comparação;
- 3 **Regression Discontinuity**: explorar uma regra de corte;
- 4 **Controle Sintético**: construir o contrafactual como combinação de unidades não tratadas.

Unidade conceitual

Todos tentam recuperar, por caminhos diferentes,

$$\mathbb{E}[Y(0) \mid T = 1].$$

1. Rosenbaum–Rubin: seleção em observáveis

A hipótese central é a independência condicional:

$$\{Y(0), Y(1)\} \perp\!\!\!\perp T \mid X.$$

Ou, para o ATT, basta uma versão mais fraca envolvendo $Y(0)$.

Se tratamos unidades comparáveis em X , então o grupo não tratado pode representar o contrafactual dos tratados.

Mas X pode ser multidimensional

Rosenbaum e Rubin (1983) mostram que podemos condensar X no **propensity score**

$$p(X) = \mathbb{P}(T = 1 \mid X).$$

Propensity Score Matching (PSM)

Fluxo básico:

- 1 estimar $p(X) = \mathbb{P}(T = 1 \mid X)$;
- 2 encontrar tratados e controles com propensity scores semelhantes;
- 3 verificar suporte comum e balanceamento;
- 4 comparar os resultados entre unidades pareadas.

Para o ATT:

$$ATT = \mathbb{E}[Y(1) - \mathbb{E}\{Y(0) \mid p(X), T = 0\} \mid T = 1].$$

Quando PSM é convincente?

Contexto favorável

- base rica em covariadas pré-tratamento;
- boa sobreposição entre grupos;
- processo de seleção plausivelmente observável.

Limitação fundamental

- não resolve seleção em não observáveis;
- matching não cria exogeneidade por si só;
- resultado depende do desenho do grupo de comparação.

Mensagem

PSM é uma estratégia de comparabilidade, não uma máquina automática de causalidade.

2. Diferenças-em-Diferenças (DiD)

Agora temos grupos e tempo.

	Antes	Depois
Tratados	$\mathbb{E}[Y \mid T = 1, t = 0]$	$\mathbb{E}[Y \mid T = 1, t = 1]$
Controles	$\mathbb{E}[Y \mid T = 0, t = 0]$	$\mathbb{E}[Y \mid T = 0, t = 1]$

$$\hat{\tau}_{DiD} = (\bar{Y}_{1,post} - \bar{Y}_{1,pre}) - (\bar{Y}_{0,post} - \bar{Y}_{0,pre}).$$

A segunda diferença tenta remover a evolução que ocorreria mesmo sem tratamento.

O caso clássico: Card e Krueger

Card e Krueger (1994) estudam o aumento do salário mínimo em **New Jersey**.

Estrutura do desenho:

- tratamento: restaurantes fast-food de New Jersey;
- comparação: restaurantes da Pennsylvania;
- observações antes e depois da mudança no salário mínimo;
- resultado: emprego.

A ideia

Pennsylvania aproxima a trajetória contrafactual de New Jersey na ausência da política.

A hipótese decisiva do DiD

Tendências paralelas:

$$\mathbb{E}[Y_{post}(0) - Y_{pre}(0) \mid T = 1] = \mathbb{E}[Y_{post}(0) - Y_{pre}(0) \mid T = 0].$$

Não é necessário que os grupos tenham o mesmo **nível** inicial.

É necessário que, sem tratamento, tenham a mesma **evolução média**.

Pergunta prática

O grupo de controle reproduz uma trajetória contrafactual plausível para os tratados?

DiD em regressão

No caso 2×2 :

$$Y_{it} = \alpha + \gamma T_i + \lambda Post_t + \beta(T_i \times Post_t) + u_{it}.$$

O coeficiente de interesse é

$$\boxed{\beta = \hat{\tau}_{DiD}}.$$

Em aplicações modernas, é comum ter múltiplos períodos, entrada escalonada no tratamento e efeitos heterogêneos — situação em que o TWFE tradicional exige cuidado.

3. Regression Discontinuity Design (RDD)

Suponha que o tratamento dependa de uma variável de ordenamento R :

$$T_i = 1\{R_i \geq c\}.$$

Exemplos:

- nota mínima para receber uma bolsa;
- idade mínima para elegibilidade;
- índice municipal acima de um cutoff administrativo.

Perto de c , unidades logo abaixo e logo acima do corte podem ser altamente comparáveis.

O estimando no RDD

No **sharp RDD**,

$$\tau_{RDD} = \lim_{r \downarrow c} \mathbb{E}[Y \mid R = r] - \lim_{r \uparrow c} \mathbb{E}[Y \mid R = r].$$

Identificação

Se os resultados potenciais evoluem suavemente em torno do cutoff, um salto em Y no ponto c é atribuído ao tratamento.

Interpretação

O efeito é **local**: refere-se às unidades próximas ao cutoff.

RDD: o que precisa ser investigado

- a regra de tratamento realmente muda no cutoff?
- existe manipulação precisa da variável de ordenamento?
- covariadas pré-tratamento são contínuas no cutoff?
- resultados são robustos à escolha de bandwidth e especificação local?

No **fuzzy RDD**, o cutoff altera a probabilidade de tratamento, mas não determina T perfeitamente.

cutoff $\longrightarrow$ instrumento local para T .

4. Controle Sintético

Problema típico: uma unidade agregada é tratada e existem várias unidades potenciais de comparação.

Construímos

$$Y_{0t}^{SC} = \sum_{j=1}^J w_j Y_{jt}, \quad w_j \geq 0, \quad \sum_j w_j = 1.$$

Os pesos são escolhidos para reproduzir, antes do tratamento, as características e a trajetória da unidade tratada.

A lógica do Controle Sintético

Depois da intervenção:

$$\hat{\tau}_t = Y_{1t} - Y_{0t}^{SC}.$$

Interpretação

O controle sintético é uma estimativa explícita da trajetória contrafactual da unidade tratada.

Aplicações típicas:

- políticas estaduais ou municipais;
- reformas institucionais;
- desastres, conflitos e grandes choques;
- intervenções que afetam uma ou poucas unidades agregadas.

Os quatro métodos sob a mesma pergunta

Método	Como constrói $Y(0)$?	Hipótese-chave
PSM	controles semelhantes em X ou $p(X)$	seleção em observáveis
DiD	evolução temporal do grupo de comparação	tendências paralelas
RDD	unidades imediatamente do outro lado do cutoff	continuidade local
Controle sintético	média ponderada de unidades não tratadas	bom ajuste pré-tratamento / estrutura contrafactual estável

O método muda; o problema contrafactual é o mesmo.

O que mudou na literatura recente?

Hoje, a discussão está menos centrada em “rodar um estimador” e mais em **credibilidade do desenho de pesquisa**.

- DiD com tratamento escalonado e efeitos heterogêneos;
- event studies com estimadores adequados à heterogeneidade;
- métodos doubly robust;
- causal machine learning e Double/Debiased ML;
- Synthetic DiD e extensões do controle sintético;
- inferência e sensibilidade mais transparentes;
- maior ênfase em pré-análise, falsificações e robustez.

Um mapa mental para escolher o desenho

- 1 PODE aleatorizar? Use um experimento.
- 2 Existe uma fonte plausível de variação exógena? Considere IV.
- 3 Seleção é plausivelmente observável? Matching / weighting / ajuste por covariadas.
- 4 Há antes e depois + grupo de comparação? DiD.
- 5 Existe regra de elegibilidade com cutoff? RDD.
- 6 Uma unidade tratada + bons doadores? Controle sintético.

Mais importante que o nome do estimador

Qual é a fonte de identificação? Por que o contrafactual é crível?

Referências essenciais

- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*.
- Rosenbaum, P. R.; Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*.
- Card, D.; Krueger, A. B. (1994). Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review*.
- Thistlethwaite, D. L.; Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*.
- Abadie, A.; Diamond, A.; Hainmueller, J. (2010). Synthetic control methods for comparative case studies. *Journal of the American Statistical Association*.
- Callaway, B.; Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*.
- Sun, L.; Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*.
- Chernozhukov, V. et al. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*.