

Avaliação de Políticas Públicas

Prof. Francis Petterini

UFSC

3 de novembro de 2025

Resultados potenciais de uma intervenção

- Pergunta causal: *O que teria acontecido aos tratados sem a intervenção?*
- Ideia central: cada unidade i possui dois resultados potenciais no tempo t :

$$Y_i(1) \text{ se tratado,} \quad Y_i(0) \text{ se não tratado.}$$

- Observamos apenas um:

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

- **Problema fundamental da inferência causal:** o contrafactual é não observado.
https://petterini.ufsc.br/files/2023/07/cara_no_bar.pdf

Definições básicas

- $T_i \in \{0, 1\}$: indicador de **tratamento** (programa na prisão).
- $D_i \in \{0, 1\}$: indicador de **biografia** (experiência pré-prisão).
- X_i : controles.
- Parâmetros de interesse (em relação a T):

$$ATE_T = E[Y(1) - Y(0)], \quad ATT_T = E[Y(1) - Y(0) \mid T = 1], \quad ATU_T = E[Y(1) - Y(0) \mid T = 0],$$

$$CATE_T(x) = E[Y(1) - Y(0) \mid X = x].$$

- Efeito direto controlado de D fixando $T = t^*$:

$$CDE_D(t^*) \equiv E[Y(1, t^*) - Y(0, t^*)],$$

onde $Y(t)$ denota resultados potenciais em relação a T e $Y(d, t)$ em relação ao par (D, T) .

SUTVA – Suposição de Valor de Tratamento de Unidade Estável

- (i) **Ausência de interferência:** o resultado controlado de i depende apenas do seu próprio T_i . (não há, por exemplo, inferência de T_j) – exemplo do Progressa
- (ii) **Ausência de versões ocultas do tratamento:** o tratamento é bem definido. (não há efeito transbordamento, ação confundidora sobreposta ao tratamento etc.) – exemplo do Raízes e Asas

Consistência

Se $T_i = t^*$, então $Y_i = Y_i(t^*)$ (a variável observada coincide com o potencial relevante).

Decomposição do viés de seleção

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

- Sem aleatorização, a diferença amostral mistura efeito causal com seleção em $Y(0)$.
- Objetivo empírico: anular ou modelar esse viés via desenho ou suposições.

Ignorabilidade Condicional e Suporte Comum

Ignorabilidade (Unconfoundedness)

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

Suporte comum (overlap)

$0 < e(X) < 1$, onde $e(X) = \Pr(T = 1 \mid X)$ é o *propensity score*.

- Sob ambas, ATE/ATT podem ser identificados condicionando ou ponderando por X (ou por $e(X)$).
- Diagnóstico: checar balanceamento de X entre grupos após o método escolhido.

Teorema do Propensity Score (Rosenbaum–Rubin)

- Se $\{Y(1), Y(0)\} \perp D \mid X$, então $\{Y(1), Y(0)\} \perp D \mid e(X)$.
- $e(X)$ é estatística de suficiência para o balanceamento.
- Prática: estimar $e(X)$ (logit/probit/ML) e aplicar *matching*, estratificação, ponderação IPW.
- Caliper é um raio máximo de distância aceitável entre tratado e controle ao formar pares. Se não houver controle “perto o bastante”, o tratado fica sem match.
- Trimming é remover observações com escores extremos (muito perto de 0 ou 1) ou fora da interseção de suportes entre tratados e controles.

- Pareamento por X ou por $e(X)$ para aproximar aleatorização local.
- *Nearest neighbor*, *caliper*, *radius*, *kernel*; checar balanceamento pós-pareamento.
- Estratificação: dividir o suporte de $e(X)$ em *bins* e comparar médias dentro de cada estrato.

Ponderação IPW – ideia e fórmulas

Ideia: reponderar a amostra para que T seja independente de X na pseudo-população.

Passo 1: Estime $e(X) = P(T = 1 | X)$ (logit/probit/ML).

Passo 2: Construa pesos e compare médias ponderadas.

ATE (HT):

$$\widehat{E}[Y(1)] = \frac{1}{n} \sum_i \frac{T_i Y_i}{\hat{e}_i}, \quad \widehat{E}[Y(0)] = \frac{1}{n} \sum_i \frac{(1-T_i) Y_i}{1-\hat{e}_i}, \quad \widehat{ATE} = \widehat{E}[Y(1)] - \widehat{E}[Y(0)].$$

ATT: $w_i = 1$ se $T_i=1$; $w_i = \hat{e}_i/(1 - \hat{e}_i)$ se $T_i=0$.

Pesos estabilizados:

$$sw_i = \frac{P(T = 1)}{\hat{e}_i} \mathbb{1}\{T_i = 1\} + \frac{P(T = 0)}{1 - \hat{e}_i} \mathbb{1}\{T_i = 0\}.$$

Condições: ignorabilidade condicional, positividade e SUTVA.

Diagnósticos: balanceamento pós-peso ($SMD < 0.1$), overlap, trimming de escores extremos.

Matching vs IPW

Matching

- Constrói pares/grupos semelhantes em X .
- Intuitivo, evita extrapolação; pode descartar observações.
- Sensível a caliper, multiplicidade de matches e suporte.

Combinação útil: podar regiões sem overlap via matching e depois estimar ATE/ATT com IPW/AIPW no subconjunto.

IPW

- Usa toda a amostra com pesos $1/e(X)$ ou $1/[1 - e(X)]$.
- Direto para ATE/ATT; atenção a pesos extremos.
- Preferir pesos estabilizados e trimming quando necessário.

Diferenças-em-Diferenças (DiD)

- Notação potencial com tempo: $Y_{it}(d)$, $t \in \{0, 1\}$.
- Hipótese de **tendências paralelas**:

$$E[Y_{i1}(0) - Y_{i0}(0) \mid D_i=1, X] = E[Y_{i1}(0) - Y_{i0}(0) \mid D_i=0, X].$$

- Identificação do ATT:

$$\text{ATT}^{DiD} = \{E[\Delta Y \mid T = 1] - E[\Delta Y \mid T = 0]\}.$$

- Extensões: eventos estendidos, adoção escalonada, estimadores compatíveis (*event-study* com cuidados de heterogeneidade temporal).

- Instrumento Z afeta D , mas não Y exceto via D (exclusão), e é independente dos potenciais.
- Tipos de conformidade: *compliers*, *never-takers*, *always-takers*, *defiers*.
- Sob independência, exclusão e monotonicidade (sem *defiers*):

$$\text{LATE} = E[Y(1) - Y(0) \mid \text{compliers}] = \frac{\text{Cov}(Y, Z)}{\text{Cov}(D, Z)}.$$

- Em *fuzzy RDD*, Z pode ser um indicador de ultrapassar o corte.

Descontinuidade de Regressão (RD)

- Atribuição por um escore R com corte c .
- Suponha continuidade de $E[Y(d) \mid R]$ em c .
- **Sharp RD**: salto em D de 0 para 1 no corte identifica efeito local.
- **Fuzzy RD**: usar $\mathbb{1}\{R \geq c\}$ como instrumento para D .
- Natureza local: estimamos um LATE na vizinhança de c .

- **Balanceamento:** diferenças padronizadas $< 0,1$; testes em múltiplos momentos.
- **Suporte:** inspeção de $\hat{e}(X)$; *trimming* em regiões de extrapolação.
- **Incerteza:** *bootstrap* para estimadores complexos (matching, AIPW, RD local).
- **Robustez:** especificações alternativas, *placebos*, falsificações em pré-trends.
- **Sensibilidade a confundidores não observados:** *Rosenbaum bounds*, *Oster Δ* .

- **Validade interna:** credibilidade do contrafactual.
- **Validade externa:** generalização para outras populações/contextos (SATE vs PATE).
- **Transportabilidade:** reponderação por diferenças de composição (X) entre amostra e alvo.

- 1 Defina $Y(1)$, $Y(0)$, D , X e o parâmetro alvo (ATE/ATT/CATE).
- 2 Escolha desenho/estratégia de identificação e explicita as suposições (SUTVA, ignorabilidade, overlap, paralelas, exclusão, monotonicidade...).
- 3 Selecione o estimador (regressão, matching, IPW, AIPW, DiD, IV, RD) e documente o procedimento.
- 4 Faça diagnósticos (balanceamento, suporte, pré-tendências, força do instrumento, largura de banda em RD).
- 5 Quantifique incerteza e faça análises de robustez e sensibilidade.
- 6 Discuta validade externa e implicações de política.

Triangular/recursivo: notação de base

Considere

$$\begin{aligned}T_i^* &= Z_i\gamma + X_i\pi + v_i, & T_i &= \mathbf{1}\{T_i^* > 0\} \\Y_i &= X_i\beta + \alpha T_i + \varepsilon_i, & \text{com } \text{Corr}(\varepsilon_i, v_i) &= \rho \neq 0.\end{aligned}$$

- Z afeta T ; Y depende de T e X ; há correlação de não observáveis entre equações.
- Sem tratamento adequado, α é **enviesado** (viés de seleção/não observáveis).
- **Moral:** supor $\rho \neq 0$ *não* resolve o problema por si só.

Três camadas da intuição

- **1) O que os IPW fazem:** reponderar a amostra para que o tratamento pareça sorteado.
- **2) Por que estabilizar:** reduzir variância e evitar pesos extremos.
- **3) Resultado prático:** pesos mais próximos de 1, melhor tamanho efetivo de amostra.

1) O que os IPW fazem

Queremos comparar “todos tratados” vs “todos não tratados” quando há confundimento.

- Criamos uma **pseudo-amostra** onde o tratamento parece aleatório.
- Cada pessoa recebe peso *inverso* da probabilidade de ter recebido o tratamento que de fato recebeu:

$$w_i = \frac{1}{\Pr(T = t_i | X_i)}.$$

- Quem tinha *pouca* chance de receber o que recebeu ganha peso *grande*; quem tinha *muita* chance, peso *pequeno*.
- Após ponderar, o tratamento tende a ficar “**independente**” dos confundidores observados.

2) Por que estabilizar os pesos

- Pesos crus $1/\Pr(T = t | X)$ podem **explodir** quando $\Pr(T = t | X)$ está perto de 0 ou 1.
- Isso **aumenta a variância** e piora a precisão das estimativas.

Ideia da estabilização: colocar no **numerador** uma probabilidade “mais simples” (marginal ou apenas com baseline), por exemplo $\Pr(T = t)$ ou $\Pr(T = t | X_0)$:

$$w_i^{\text{stab}} = \frac{\Pr(T = t_i \text{ (marginal ou com baseline)})}{\Pr(T = t_i | X_i)}.$$

Intuição: o **denominador** ainda “desconfunde”, enquanto o **numerador** puxa os pesos de volta para perto de 1.

O que se ganha com a estabilização

- **Menos pesos extremos $\Rightarrow$ menor variância.**
- **Maior tamanho efetivo** da amostra (melhor precisão).
- Mantém-se o **balanceamento** entre tratados e controles nos confundidores observados.

Forma prática (tratamento binário, corte transversal)

Passos:

- 1 Estime $\hat{p}_i = \Pr(T = 1 \mid X_i)$ (modelo de escore de propensão).
- 2 Calcule a probabilidade marginal $\hat{p}_T = \frac{1}{n} \sum_i T_i$ (ou condicional em baseline X_0).
- 3 Defina os pesos estabilizados:

$$w_i^{\text{stab}} = \begin{cases} \hat{p}_T / \hat{p}_i, & \text{se } T_i = 1, \\ (1 - \hat{p}_T) / (1 - \hat{p}_i), & \text{se } T_i = 0. \end{cases}$$

Diagnósticos: verificar distribuição dos pesos (caudas curtas) e *balanceamento ponderado* (diferenças padronizadas próximas de 0).

- **Positividade:** evitar regiões onde $\Pr(T = 1 \mid X) \approx 0$ ou ≈ 1 .
- **Trimming/truncagem:** limitar pesos em percentis (p.ex., 1 e 99) quando necessário.
- **Modelo do denominador:** é ele que remove o viés – capricho na especificação.
- **Normalização:** opcional, para que $\bar{w} \approx 1$.

Cenário do exemplo

- 100 pessoas, uma covariável X : **70 jovens** e **30 idosos**.
- Observado:

	Tratados	Controles
Jovens (70)	50	20
Idosos (30)	10	20
Total	60	40

- Jovens têm maior chance de tratar; idosos, menor.
- Queremos comparar tratados vs. controles como se o tratamento tivesse sido sorteado (mesma composição em X).

Propensões por subgrupo

A partir da tabela:

$$\Pr(T = 1 \mid \text{jovem}) = \frac{50}{70} \approx 0,714, \quad \Pr(T = 1 \mid \text{idoso}) = \frac{10}{30} \approx 0,333.$$

Pesos IPW *crus* (por célula):

$$w = \begin{cases} \frac{1}{\Pr(T = 1 \mid X)} & \text{se tratado,} \\ \frac{1}{1 - \Pr(T = 1 \mid X)} & \text{se controle.} \end{cases}$$

Logo:

$$w(\text{trat.}, \text{jovem}) \approx \frac{1}{0,714} = 1,4,$$

$$w(\text{trat.}, \text{idoso}) \approx \frac{1}{0,333} = 3,0,$$

$$w(\text{ctrl.}, \text{jovem}) \approx \frac{1}{1-0,714} = \frac{1}{0,286} = 3,5,$$

$$w(\text{ctrl.}, \text{idoso}) \approx \frac{1}{1-0,333} = \frac{1}{0,667} = 1,5.$$

“Clonar e encolher”: o milagre do balanceamento

Multiplicando contagens pelos pesos IPW *crus*:

	Tratados	Controles
Jovens	$50 \times 1,4 = 70$	$20 \times 3,5 = 70$
Idosos	$10 \times 3,0 = 30$	$20 \times 1,5 = 30$
Total ponderado	100	100

- Após ponderar, **ambos** os grupos têm a **mesma composição** em X : 70 jovens e 30 idosos.
- É isso que um sorteio perfeito faria – comparar “iguais com iguais”.

Pesos *estabilizados*: puxando para perto de 1

Probabilidade marginal de tratamento:

$$\Pr(T = 1) = \frac{60}{100} = 0,6, \quad \Pr(T = 0) = 0,4.$$

Pesos IPW **estabilizados**:

$$w^{\text{stab}} = \begin{cases} \frac{0,6}{\Pr(T = 1 \mid X)} & \text{se tratado,} \\ \frac{0,4}{1 - \Pr(T = 1 \mid X)} & \text{se controle.} \end{cases}$$

Numericamente:

$$w^{\text{stab}}(\text{trat., jovem}) \approx \frac{0,6}{0,714} = 0,84,$$

$$w^{\text{stab}}(\text{trat., idoso}) \approx \frac{0,6}{0,333} = 1,8,$$

$$w^{\text{stab}}(\text{ctrl., jovem}) \approx \frac{0,4}{0,286} = 1,4,$$

$$w^{\text{stab}}(\text{ctrl., idoso}) \approx \frac{0,4}{0,667} = 0,6.$$

O que muda com estabilização

Contagens ponderadas com w^{stab} :

	Tratados	Controles
Jovens	$50 \times 0,84 = 42$	$20 \times 1,4 = 28$
Idosos	$10 \times 1,8 = 18$	$20 \times 0,6 = 12$
Total ponderado	60	40

- A **composição** segue idêntica em X : 70% jovens e 30% idosos em ambos.
- Os pesos ficam **mais próximos de 1** – menor variância, amostra efetiva maior.
- Totais ponderados respeitam as margens observadas (60 tratados, 40 controles).

Resumo: por que funciona

- IPW **equaliza a composição** em X entre tratados e controles.
- O denominador $\Pr(T = t \mid X)$ remove o **confundimento observável**.
- O numerador “simples” $\Pr(T = t)$ (ou $\Pr(T = t \mid X_0)$) **estabiliza** os pesos.
- Depois disso, a diferença de médias ponderada é próxima do que se veria sob **randomização**.

- Binário: $\Pr(Y = 1) = \text{logit}^{-1}(\alpha_0 + \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_1 D_2)$, com `[pweight = sw]`.
- Contínuo: $E(Y) = \alpha_0 + \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_1 D_2$, com `[pweight = sw]` e EP robustos (cluster se painel).
- Interpretação: efeitos marginais por combinação (D_1, D_2) via `margins`.

Exemplo compacto em Stata (pesos e MSM)

```
* 1) D1 | L0 (denominador)
logit d1 x_*
predict ph1_den, pr
gen pr1_den = cond(d1==1, ph1_den, 1 - ph1_den)
* 2) D2 | D1, L0, L1 (denominador)
logit d2 d1 x_* l1_*
predict ph2_den, pr
gen pr2_den = cond(d2==1, ph2_den, 1 - ph2_den)
* 3) Numeradores (estabilização)
logit d1
predict ph1_num, pr
gen pr1_num = cond(d1==1, ph1_num, 1 - ph1_num)
logit d2 d1
predict ph2_num, pr
gen pr2_num = cond(d2==1, ph2_num, 1 - ph2_num)
* 4) Peso estabilizado
gen sw = (pr1_num*pr2_num)/(pr1_den*pr2_den)
```

Exemplo compacto em Stata (pesos e MSM)

```
* 5) Diagnóstico e truncagem
summ sw, detail
_pctile sw, p(1 99)
scalar lo = r(r1)
scalar hi = r(r2)
gen swt = min(max(sw, lo), hi)
* 6) MSM e efeitos marginais
glm y i.d1##i.d2 [pweight=swt], family(binomial) link(logit) vce(robust)
margins d1#d2
```

- Se há perda de acompanhamento, multiplique sw por pesos de censura (*Inverse Probability of Censoring Weighting*).
- Modele $\Pr(C_t = 1 \mid \text{história até } t)$ para permanecer observado; construa pesos análogos (estabilizados) e faça o produto com sw .

- **Consistência:** o resultado observado corresponde ao potencial sob a trajetória realizada.
- **Ignorabilidade sequencial:** dado o histórico antes de cada decisão, a alocação é como-se aleatória.
- **Positividade:** probabilidades de tratamento em todos os estratos relevantes são estritamente entre 0 e 1.
- **Especificação:** modelos de tratamento/censura corretamente especificados (ou suficientemente ricos).

- Use interações e termos flexíveis (splines, polinômios) nos modelos de probabilidade; considere ML (e.g., árvores/boosting) com *cross-fitting*.
- Truncagem moderada (p. ex., 1–99%) costuma estabilizar a variância sem induzir muito viés.
- Generalização para $T > 2$: produto ao longo de $t = 1, \dots, T$ com a mesma lógica sequencial.
- Relate diagnósticos: média/variância dos pesos, share truncado, *overlap*.