

Pareamento (matching)

Avaliação de Políticas Públicas B

30/abril, 07, 12, 21, 26 e 28/maio e 02/junho de 2025

Leitura básica:

FACURE, Matheus. **Causal inference for the brave and true**. 2022. [Capítulo 10: Matching; Capítulo 11: Propensity score.] Disponível em: <https://matheusfacure.github.io/python-causality-handbook/landing-page.html>

Leitura complementar:

GERTLER, Paul J. et al. **Avaliação de impacto na prática**. Washington: Banco Interamericano de Desenvolvimento; Banco Mundial, 2018. [Capítulo 8: Pareamento.] Disponível em: <https://openknowledge.worldbank.org/bitstream/handle/10986/25030/9781464808890.pdf>

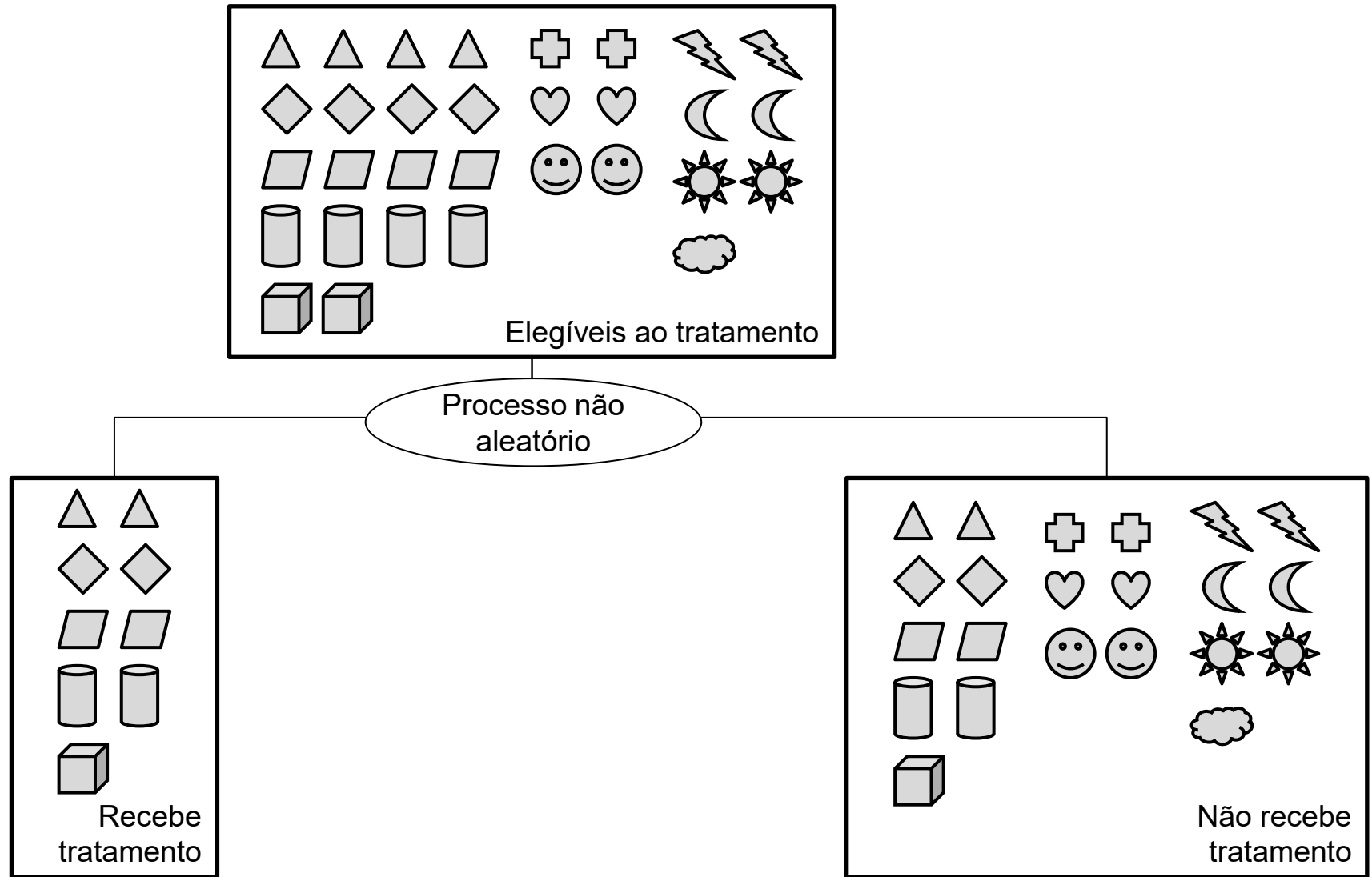
HEINRICH, Carolyn; MAFFIOLI, Alessandro; VÁZQUEZ, Gonzalo. **A primer for applying propensity-score matching**. Impact-Evaluation Guidelines, Technical Notes n. IDB-TN-161. Washington: Banco Interamericano de Desenvolvimento, 2010. <http://dx.doi.org/10.18235/0008567>

HUNTINGTON-KLEIN, Nick. **The effect: an introduction to research design and causality**. 2023. [Capítulo 14: Matching.] Disponível em: <https://theeffectbook.net/>

Em estudos observacionais, usaremos estratégias para lidar com a falta de comparabilidade presumida entre grupos

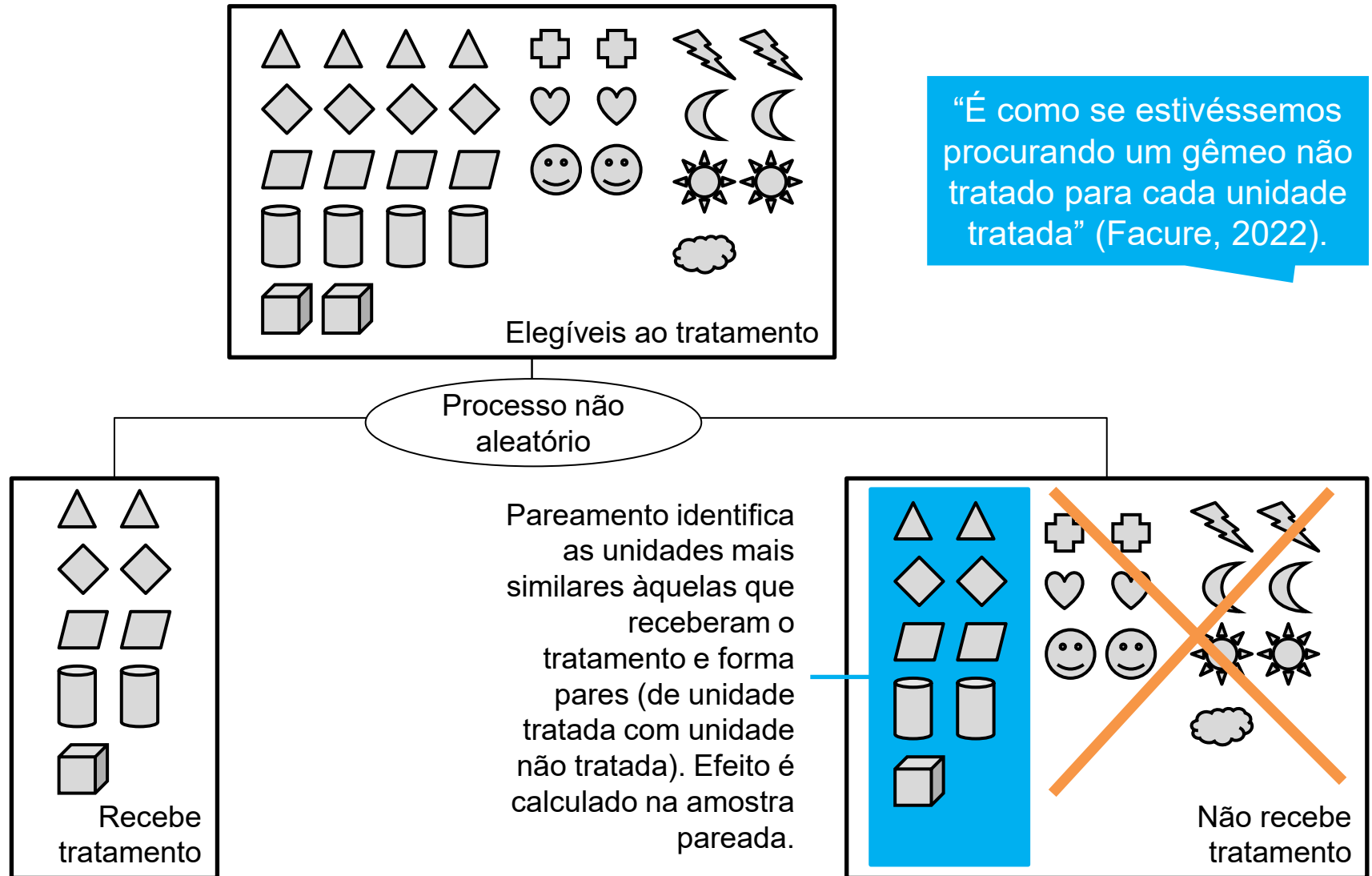
Problema	Estratégia	Exemplo de abordagem
- Em estudos observacionais (i.e., em que não há atribuição aleatória ao tratamento), não há presunção de comparabilidade entre grupo tratado e não tratado (antes da exposição ao tratamento). - De fato, esses grupos podem diferir substancialmente, tanto nas características observáveis como nas não observáveis. Por isso, a associação observada entre T (tratamento) e Y (variável de resultado) não pode ser considerada uma estimativa do efeito causal de T em Y. - Para a estimação de efeito, teremos de usar alguma estratégia para enfrentar a não comparabilidade entre os grupos.	1 Modelar o mecanismo de tratamento (T)	- Variável instrumental - Matching (pode envolver propensity score – PS) - Inverse Probability of Treatment Weighting (IPTW) com probabilidade de tratamento definida como PS
	2 Modelar a variável de resultado (Y)	- Regressão com descontinuidade - Diferença em diferenças - Controle sintético - Efeitos fixos - Controle estatístico (pode envolver PS)
	3 Uma combinação das estratégias anteriores	Estimação do tipo double robust

Pareamento: intuição ilustrada



Pareamento: intuição ilustrada

“É como se estivéssemos procurando um gêmeo não tratado para cada unidade tratada” (Facure, 2022).



Pareamento simples (ou exato): exemplo

Sejam:

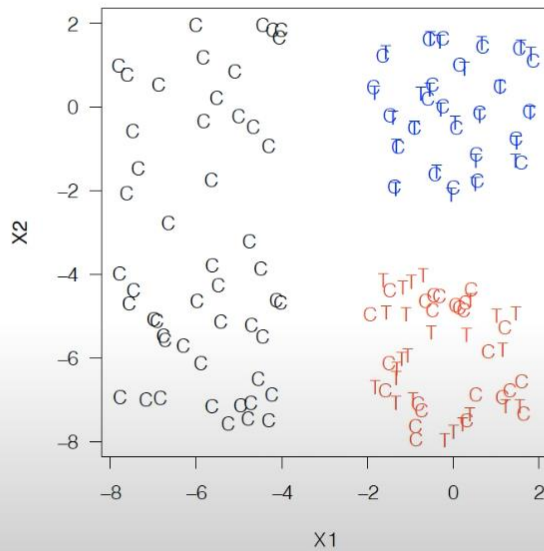
- $D = 1$ se domicílio participou de um programa
- $D = 0$ se domicílio não participou de um programa
- Education = anos de escolaridade do indivíduo com a maior escolaridade do domicílio
- Income = Renda domiciliar per capita, a variável de resultado

i	D	Education	Income	Match	Y_1	Y_0	Difference
1	0	2	60	-	-	-	-
2	0	3	80	-	-	-	-
3	0	5	90	-	-	-	-
4	0	12	200	-	-	-	-
5	1	5	100	[3]	100	90	10
6	1	3	80	[2]	80	80	0
7	1	4	90	[2.3]	90	85	5
8	1	2	70	[1]	70	60	10

Com pareamento, é possível estimar outras quantidades de interesse (ATE e ATC). Porém, normalmente pareamento é utilizado para estimar ATT, e este será o foco nos nossos slides.

ATT (*average treatment effect on the treated*) = $25/4 = 6,25$

Mais de uma dimensão de pareamento



X_1 e X_2 são variáveis de pareamento. <https://youtu.be/tvMyjDi4dyg>

Pareamento: intuição, assunção e requisito

Intuição

- Cria o melhor grupo de comparação possível com base nas características observáveis, identificando no conjunto de **unidades não tratadas** aquelas **que mais se parecem** com as unidades **tratadas**
- Estima o impacto do tratamento com base na **amostra pareada**

Assunção e requisito

- **ASSUNÇÃO:** **Independência condicional (a.k.a. conditional ignorability, selection on observables):** condicionada por características observáveis, a participação (i.e., recebimento do tratamento) é exógena
[A]fter controlling for X [matching variables], the treatment assignment is “as good as random”. (Heinrich; Maffioli; Vázquez, 2010, p. 16)
- **REQUISITO: Sobreposição (a.k.a. common support, positivity):** há no grupo de comparação casos com características similares às aquelas apresentadas pelos casos do grupo de tratamento

Reflexões:

assunção de independência condicional

- Variáveis observáveis nas quais condicionamos a ignorabilidade **afetam tanto a decisão de participar** do programa **quanto o efeito do programa**, mas não são afetadas por ele (i.e., referem-se a um estado anterior à implementação do programa)



*Quando o tratamento afeta as características individuais e usamos essas características para realizar o pareamento, escolhemos um grupo de comparação que se pareça com o grupo tratado por causa do próprio tratamento. Sem o tratamento, essas características pareceriam mais diferentes. Isso viola a exigência básica para uma boa estimativa do contrafactual: **o grupo de comparação deve ser semelhante em todos os aspectos, exceto pelo fato de que o grupo de tratamento recebe o tratamento e o grupo de comparação, não.** (Gertler et al., 2018, p. 161)*

- Independência condicional é um pressuposto **muito exigente** e que **não é testável**; qualidade do pareamento depende das características específicas do tratamento



*Embora seja importante ser capaz de realizar esses pareamentos usando um grande número de características, **ainda mais importante é ser capaz de parear as unidades com base em características que determinam a inscrição no programa.** (Gertler et al., 2018, p. 161)*

- É preciso um **grande número de variáveis** que possam ser utilizadas para o pareamento de forma a **fortalecer chances** de atendimento do pressuposto da **independência condicional**

Reflexões: requisito da sobreposição

Maldição da
dimensionalidade

- Todavia, quanto **maior o número de variáveis de pareamento, maior a dificuldade para se encontrar bons pares** (menor a chance de sobreposição)



*À medida que aumenta o número de características ou dimensões em relação às quais se quer realizar o pareamento das unidades que se inscreveram no programa, é possível se deparar com o que chamamos de **'problema da dimensionalidade'**. Por exemplo, **se forem usadas apenas três características** importantes para identificar o grupo de comparação pareado, como idade, sexo e se o indivíduo tem diploma do ensino médio, provavelmente [...serão encontrados] pares para todos os participantes inscritos no programa no conjunto daqueles que não estão inscritos [...], mas **corre-se o risco de deixar de fora outras características potencialmente importantes. No entanto, se aumentarmos a lista de características** – digamos, para incluir o número de filhos, o número de anos de estudo, o número de meses desempregado, o número de anos de experiência profissional e assim por diante –, **a base de dados pode não conter um bom par para a maioria dos participantes do programa que estão inscritos, a menos que tenha um número muito grande de observações.** (Gertler et al. (2018, p. 160)*

- **Se não houver sobreposição**, os dois grupos podem ser muito diferentes e **não mais se justificaria a comparação** entre eles
- Assim, é preciso um **grande número de casos na amostra** (especialmente no grupo de comparação, para aumentar as chances de uma boa sobreposição)

Tanto para o ATE como
para o ATT, precisamos de
bons pares para tratados

Matching bias (viés de pareamento)

É isso, mesmo! Matching pode produzir viés!

O viés [de pareamento] surge quando as discrepâncias [intrapares] são grandes. Felizmente, é possível ajustar a estimativa de efeito para contornar esse problema.

Facure (2022)

Intuição dessa correção: no cálculo do efeito, dê menos peso aos pares discrepantes.

Mas lembre: esta correção é cosmética, porque você ainda estará baseando sua estimativa em pares inadequadamente formados, o que compromete a assunção de independência condicional. Sempre que possível, evite essas discrepâncias (e.g., se você tiver uma amostra grande, descarte observações tratadas para as quais não exista um bom par entre os não tratados).

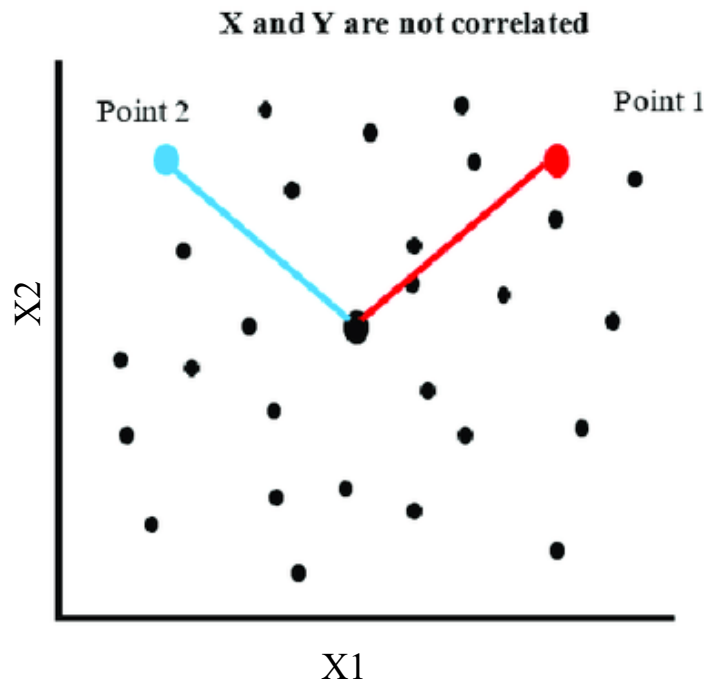
Há muitas formas diferentes de parear, o que implica tomar decisões

NÃO EXAUSTIVO

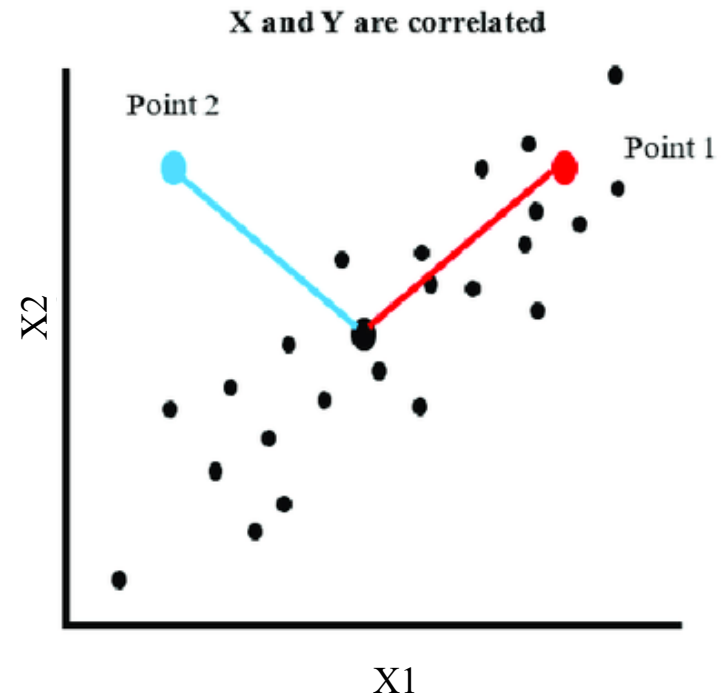
Medida de distância

- Qual medida de distância será usada **para identificar os melhores pares**? Há várias opções, entre elas:
 - Distância (norma) euclidiana
 - Distância de Mahalanobis
 - Genetic matching (Mahalanobis com pesos: importância dada a cada variável de pareamento depende de sua capacidade de gerar grupos balanceados)
 - Propensity score (PS)

Distância de euclidiana vs. Mahalanobis



When X and Y are not correlated, the Euclidean distance from the Centroid can be useful to infer if a point is member of the distribution



Point one and two have the same Euclidean Distance from Centroid but only point one is a member of the distribution. to detect point two as outlier, dist. (point two, centroid) should be much higher than dist. (point one, Centroid)

Mahalanobis distance can be used here instead.

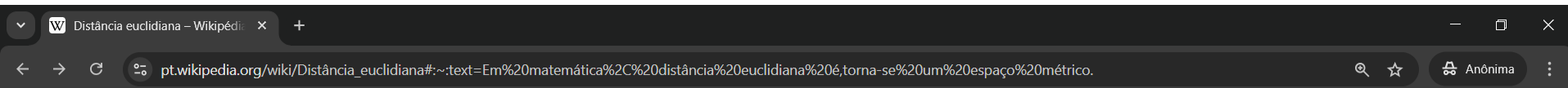
https://www.researchgate.net/figure/The-difference-between-Euclidean-and-Mahalanobis-Distance-and-how-the-Mahalanobis_fig1_363398810

Sugestão de video sobre distâncias:

Euclidean distance and the Mahalanobis distance (and the error ellipse)

<https://youtu.be/xXhLvheEF7o>

Distância euclidiana



Distância euclidiana

[ocultar]

 39 línguas 

Conteúdo 

Início

▼ Definição

[Distância unidimensional](#)

[Distância bidimensional](#)

[Distância tridimensional](#)

[Distância n-dimensional](#)

[Ver também](#)

Artigo [Discussão](#)

[Ler](#) [Editar](#) [Ver histórico](#) [Ferramentas](#) 

Em matemática, **distância euclidiana** é a distância entre dois pontos, que pode ser provada pela aplicação repetida do teorema de Pitágoras. Aplicando essa fórmula como distância, o espaço euclidiano torna-se um espaço métrico.

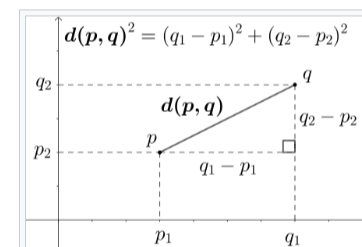
Definição

A distância euclidiana entre os pontos $P = (p_1, p_2, \dots, p_n)$ e $Q = (q_1, q_2, \dots, q_n)$, num [espaço euclidiano n-dimensional](#), é definida como:

$$\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}.$$

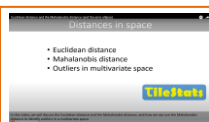
Distância unidimensional

Para pontos unidimensionais, $P = (p_x)$ e $Q = (q_x)$, a distância é computada como:



A distância euclidiana em duas dimensões.

https://pt.wikipedia.org/wiki/Dist%C3%A2ncia_euclidiana



Sugestão de vídeo sobre distâncias:

Euclidean distance and the Mahalanobis distance (and the error ellipse)

<https://youtu.be/xXhLvheEF7o>

Distância de Mahalanobis



Os dados em ambos os grupos a comparar deverão ter o mesmo número de variáveis (ou seja, o mesmo número de colunas) mas não necessariamente o mesmo número de elementos (o número de linhas pode ser diferente).

Conteúdo

ocular

Início

Definição

Exemplo

Explicação intuitiva

Relação com a estatística-alavanca

Aplicações

Bibliografia

Referências

Definição

Formalmente, a distância de Mahalanobis entre um grupo de valores com média $\mu = (\mu_1, \mu_2, \mu_3, \dots, \mu_p)^T$ e **matriz de covariância** S para um vector multivariado $x = (x_1, x_2, x_3, \dots, x_p)^T$ é definida como:

$$D_M(x) = \sqrt{(x - \mu)^T S^{-1} (x - \mu)}.$$

A distância de Mahalanobis pode também definir-se como uma medida de dissimilaridade entre dois **vectores aleatórios** $\vec{x}$ e $\vec{y}$ com a mesma **distribuição** com a matriz de covariância S :

$$d(\vec{x}, \vec{y}) = \sqrt{(\vec{x} - \vec{y})^T S^{-1} (\vec{x} - \vec{y})}.$$

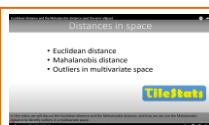
Se a matriz de covariância é a **matriz identidade**, a distância de Mahalanobis coincide com a **distância euclidiana**. Se a matriz de covariância é diagonal, então a medida de distância resultante é chamada *distância euclidiana normalizada*:

$$d(\vec{x}, \vec{y}) = \sqrt{\sum_{i=1}^p \frac{(x_i - y_i)^2}{\sigma_i^2}},$$

onde σ_i é o desvio-padrão de x_i no conjunto amostral.

Exemplo

https://pt.wikipedia.org/wiki/Dist%C3%A2ncia_de_Mahalanobis



Sugestão de video sobre distâncias:

Euclidean distance and the Mahalanobis distance (and the error ellipse)

<https://youtu.be/xXhLvheEF7o>

MDM: Mahalanobis distance matching

16. Matching Methods

Method 1: Mahalanobis Distance Matching

(Approximates Fully Blocked Experiment)

Procedure

1. Preprocess (Matching)

- $\text{Distance}(X_c, X_t) = \sqrt{(X_c - X_t)' S^{-1} (X_c - X_t)}$
- Match each treated unit to the nearest control unit
- Control units: not reused; pruned if unused
- Prune matches if $\text{Distance} > \text{caliper}$
- (Many adjustments available to this basic method)

2. Estimation Difference in means or a model

Interpretation

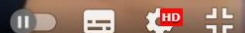
- Quiz: Do you understand the distance trade offs?
- Quiz: Does standardization help?
- Mahalanobis is for methodologists; in applications, use Euclidean!



16. Matching Methods

22:32 / 1:23:16 • Matching: Finding Hidden Randomized Experiments > ▾

22 / 45



<https://youtu.be/tvMyjDi4dyg>

Sugestão de video sobre distâncias:

Euclidean distance and the Mahalanobis distance (and the error ellipse)

<https://youtu.be/xXhLvheEF7o>

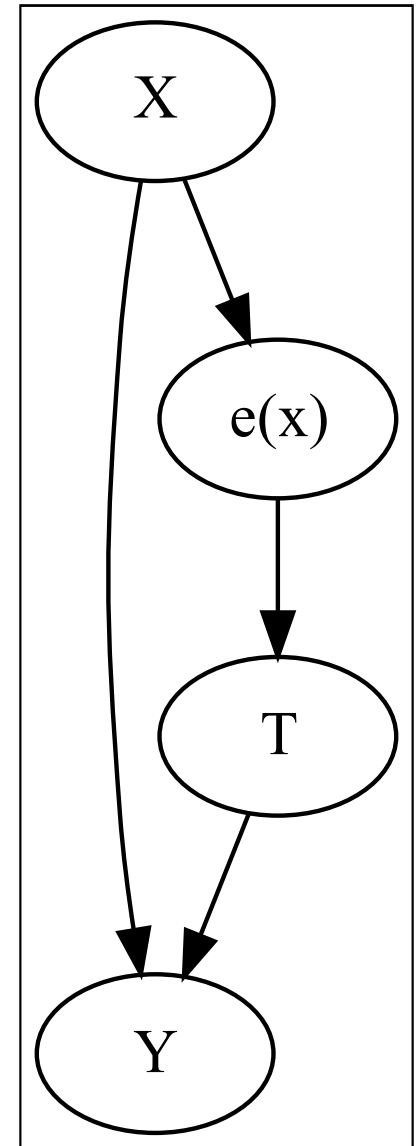
Sobreposição no pareamento baseado em distâncias: distâncias devem se concentrar na vizinhança de zero



Queremos unidades não tratadas com distâncias mínimas em relação às unidades tratadas.

PS: intuição

- Considere os fatores de confusão ($\mathbf{X}$), ou seja, as variáveis que diferem entre grupos no estágio anterior ao tratamento (pré-teste) e que são fonte de espuriedade. $\mathbf{X}$ pode ser um conjunto grande de variáveis.
- Podemos sintetizar o conjunto de variáveis $\mathbf{X}$ em uma única medida: a **probabilidade condicional de tratamento**, $P(T|\mathbf{X})$, que é o **escore de propensão** (propensity score – PS).
- Em Facure (2022), o PS aparece como $e(\mathbf{x})$.



PS: intuição

- **Repetindo:** a ideia é que posso sintetizar um conjunto de variáveis referentes ao pré-teste em um escore que poderá ser usado para formar pares.
- Usaremos o PS para sintetizar as características pré-teste das unidades.
- Buscaremos compor pares com unidades (uma tratada, outra não tratada) que apresentem PS semelhante.

PS: estimação

- O PS é calculado para os grupos de tratamento e comparação a partir de um **modelo de participação** (i.e., **recebimento do tratamento**), no qual a **variável dependente é uma dummy que indica se uma dada unidade participou** ou não do tratamento e as **variáveis independentes são características das unidades no momento anterior ao tratamento** (ou seja, **X**). Estrutura dessa regressão:

$$T = \alpha + \beta(X) + \varepsilon$$

X é um conjunto de variáveis (cada uma multiplicada pelo respectivo β)

- O PS é estimado por um modelo **logit** ou um modelo **probit** (esses são tipos de regressão adaptados para variável dependente binária); logit e probit predizem probabilidades estritamente no intervalo [0,1], o que não é verdade para MQO.

Propensity Score Matching (PSM): intuição básica

O PSM (Propensity Score Matching) utiliza informações de um grupo de unidades que não participam da intervenção para identificar o que teria acontecido com as unidades participantes na ausência da intervenção. [Note como esta descrição refere-se ao ATT.] Ao comparar como os resultados diferem para os participantes em relação a não participantes observacionalmente semelhantes, é possível estimar os efeitos da intervenção.

[...]

Uma das questões críticas na implementação de técnicas de pareamento é definir claramente (e justificar) o que significa “semelhante”. [...] Na prática, para que o processo de pareamento consiga mitigar com sucesso possíveis vieses, ele deve ser feito **considerando toda a gama de covariáveis nas quais as unidades tratadas e de comparação possam diferir**.

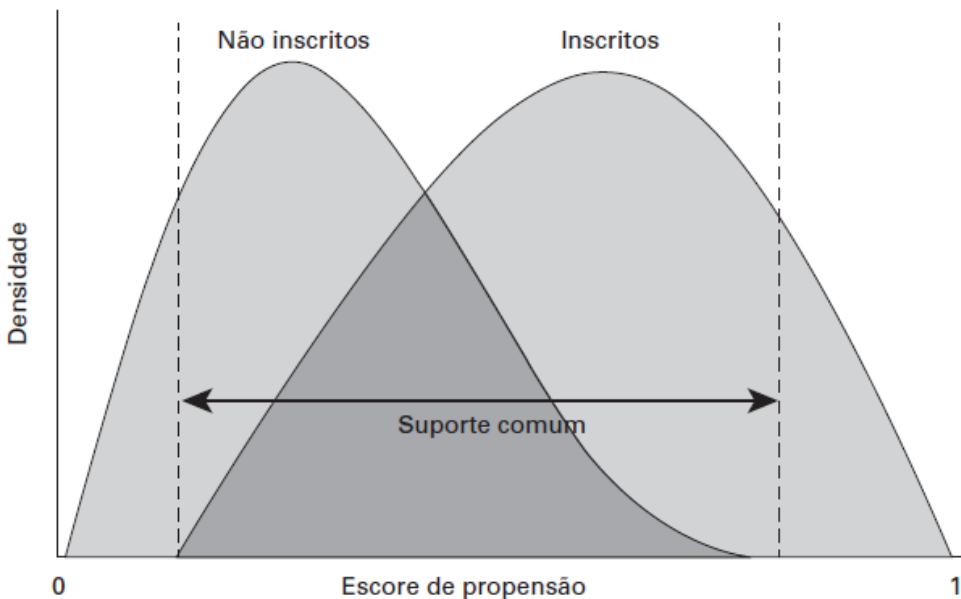
[...]

O escore de propensão é definido como a probabilidade de que uma unidade na amostra combinada de unidades tratadas e não tratadas receba o tratamento, dado um conjunto de variáveis observadas. Se todas as informações relevantes para a participação e os resultados [em conjunto] forem observáveis ao pesquisador, o escore de propensão (ou probabilidade de participação) produzirá pareamentos válidos para estimar o impacto de uma intervenção. Portanto, **em vez de tentar parear com base em todos os valores das variáveis, os casos podem ser comparados apenas com base nos escores de propensão**.

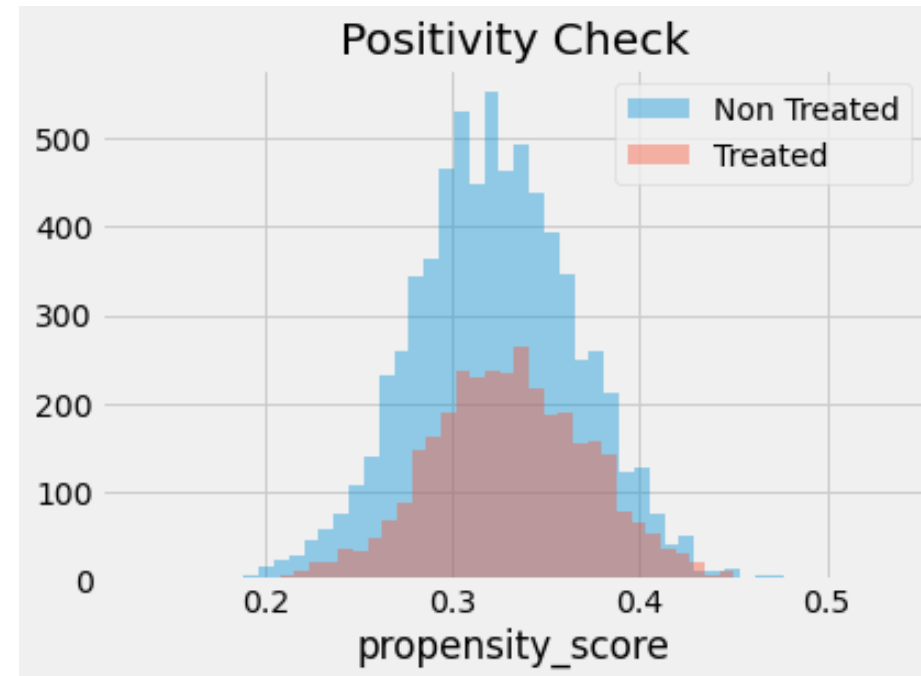
Heinrich, Maffioli, Vázquez (2010, p. 4)

Sobreposição no PSM: grupo de comparação deve ter escores parecidos aos do grupo de tratamento

Figura 8.2 Pareamento por escore de propensão e suporte comum



Gertler et al. (2018, p. 162).



Facure (2022, capítulo 11).

PSM: exemplo

RESUMO

Nesta monografia realizamos a avaliação econômica do Programa Fica Vivo que é o pilar da política de prevenção e controle da criminalidade do Governo do Estado de Minas Gerais. Seu objetivo principal é a redução dos homicídios nas áreas de maior incidência, em geral favelas. A avaliação econômica envolve a apuração de dois elementos: o custo e a efetividade do programa. Os custos são apurados pelo método de contagem através das informações disponibilizadas pela Secretaria Estadual de Defesa Social e pela Polícia Militar de Minas Gerais (PMMG). A efetividade é mensurada pela metodologia Diferenças em Diferenças com Pareamento (*Double Difference Matching*) com base nas ocorrências georeferenciadas registradas pela PMMG e no Censo Demográfico 2000. Consideramos como variável de impacto do programa a taxa de homicídio por cem mil habitantes. Esta metodologia permite a mensuração da qualidade do investimento público através de dois indicadores de eficiência: razões custo-efetividade e custo-benefício. Neste sentido a presente monografia contribui à modernização da gestão governamental através da aplicação do método de avaliação econômica de projetos públicos baseado em registros administrativos e dados oficiais, possibilitando a sua replicação. Além disto, contribui ao desenvolver um método de análise de política na área de segurança pública que é carente de embasamento empírico. Os resultados mostram que o custo de um homicídio evitado pelo programa é de aproximadamente 244,6 mil reais o que implica em uma taxa de retorno do programa de aproximadamente 99%. A comparação destes resultados com avaliações internacionais de programas similares evidencia que o Fica Vivo tem um elevado retorno.

Palavras-chave: Qualidade do Gasto Público, Indicador de Eficiência, Avaliação Econômica do Fica Vivo.



1º Lugar - Qualidade do Gasto Público

Autora: **Betânia Totino Peixoto**
Belo Horizonte/MG

"Avaliação Econômica do Programa Fica Vivo: o caso piloto"

Em agosto de 2002, o programa Fica Vivo foi implantado na área piloto, favela denominada "Morro das Pedras". A escolha desta área como a primeira a receber o programa decorreu do fato desta ser, das seis áreas apontadas pelo diagnóstico, a que exibia maior taxa de homicídio por cem mil habitantes e elevado índice de vulnerabilidade social. Embora esses critérios tenham sido os mais relevantes há que se mencionar que essa área apresentava maior presença de aparelhos públicos locais e iniciativas privadas voltadas para a proteção social (SILVEIRA, 2007). Esse ambiente facilitava a implantação do programa.

Peixoto (2008, p. 27)

PSM: exemplo

O PAREAMENTO

Primeiramente, estimamos a probabilidade de participação no programa dos setores censitários. Esta estimação é realizada através do modelo PROBIT condicionado às características socioeconômicas, demográficas e às taxas de homicídio por cem mil habitantes dos setores censitários antes do programa²⁶. As variáveis utilizadas são ortogonais ao programa. Em seguida o grupo controle é selecionado pela metodologia de Pareamento por Vizinho mais Próximo (*Nearest Neighbor Matching*)²⁷.

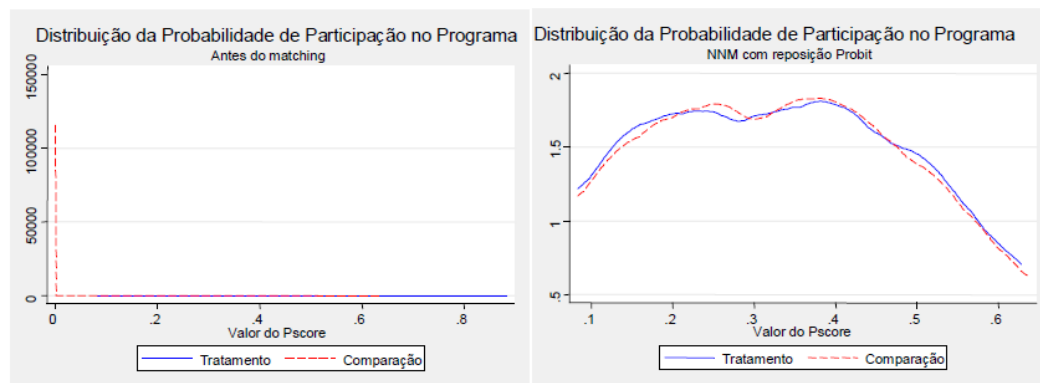
Na Figura 1, mostramos os gráficos da função de densidade da probabilidade de participação no programa para os grupos de tratamento e de comparação, antes e depois do pareamento. Antes do pareamento, a probabilidade de participação, de grande parte do grupo de comparação, está concentrada próxima à zero. Após o pareamento a probabilidade de participação do grupo de comparação passa a ter

²⁶ O modelo Probit estimado é apresentado na tabela B do anexo.

²⁷ Realizamos o pareamento pelas demais metodologias apresentadas no capítulo 1 e os resultados foram similares.

distribuição semelhante ao grupo de controle, sendo a curva de distribuição nos dois grupos quase sobreposta.

Figura 1: Distribuição da probabilidade de participação no programa Fica Vivo, Morro das Pedras e Controle, antes do pareamento e após o Pareamento por Vizinho mais Próximo.



Na tabela 6, apresentamos as médias das covariadas utilizadas na estimação da probabilidade de participação, antes e após o pareamento, entre os grupos de tratamento e comparação. As colunas “Dif-Médias” mostram o resultado do teste de diferenças nas médias das variáveis do grupo de tratamento e comparação. Em outras palavras, indica a semelhança entre as médias das variáveis nos grupos tratamento e controle, antes e após o pareamento. Podemos observar que o pareamento tornou as médias de todas as variáveis estatisticamente iguais. Antes do pareamento, as médias das variáveis socioeconômicas do grupo de tratamento e controle são diferentes.

Peixoto (2008, p. 42)

Tratamento e Comparação antes e após o Pareamento

Variáveis	Antes do <i>Matching</i>			Dif-Médias após o <i>Matching</i>
	Média Tratado	Média Comp.	Dif-Médias	
Txhoms1	26,031	7,647	18,384***	-0,776
Txhoms2	35,745	9,418	26,327***	-41,939
Txhoms3	56,377	9,293	47,084***	-18,459
Txhoms4	35,828	8,268	27,560***	-2,671
Txhoms5	47,691	11,137	36,554***	-34,087
P_1banho	0,804	0,602	0,202***	-0,021
P_2banho	0,083	0,210	-0,127***	0,009
P_3banho	0,038	0,129	-0,091**	0,013
P_4mbanho	0,014	0,039	-0,025	-0,004
P_lixo	0,941	0,984	-0,044***	-0,018
P_homem	0,481	0,470	0,011**	-0,001
p_09aa	0,211	0,152	0,059***	0,001
p_1014aa	0,104	0,082	0,021***	-0,002
p_1519aa	0,115	0,097	0,017***	-0,006
p_2024aa	0,118	0,103	0,015***	0,002
p_2529aa	0,082	0,088	-0,007*	-0,004
p_30maa	0,371	0,477	-0,105***	0,010
P_rend0	0,112	0,069	0,043***	0,007
P_rend_1	0,252	0,112	0,140***	0,000
P_rend1_3	0,438	0,268	0,170***	-0,022
P_rend3_5	0,100	0,148	-0,048***	-0,009
P_rend5_10	0,047	0,188	-0,141***	0,010
População no semestre 1	781,190	879,410	-98,220*	5,180
População no semestre 2	789,610	878,200	-88,590	6,280
População no semestre 3	798,110	877,460	-79,350	7,380
População no semestre 4	807,270	887,250	-79,980	8,140
População no semestre 5	816,540	898,440	-81,900	8,890

Nota: ***significativa a 1%, **significativa a 5%, *significativa 10%.

Peixoto (2008, p. 43)

Há muitas formas diferentes de parear, o que implica tomar decisões

NÃO EXAUSTIVO

Medida de distância

- Qual medida de distância será usada **para identificar os melhores pares**? Há várias opções, entre elas:
 - Distância (norma) euclidiana
 - Distância de Mahalanobis
 - Genetic matching (Mahalanobis com pesos: importância dada a cada variável de pareamento depende de sua capacidade de gerar grupos balanceados)
 - Propensity score (PS)

Seleção de pares

- Quão longe é **longe demais**?
 - Nearest neighbor
 - Caliper/ Radius
 - Estratificação do suporte comum

Número de pares

- Um caso **não tratado** pode ser **pareado com quantos casos tratados**? Em outras palavras, **pode haver reposição**?
- Um **caso tratado** pode ser **pareado com quantos casos não tratados**?

Estimação do efeito

- Com ou sem ajuste por **discrepância** entre médias das variáveis de pareamento nos grupos (bias correction)?

Duas orientações gerais

1

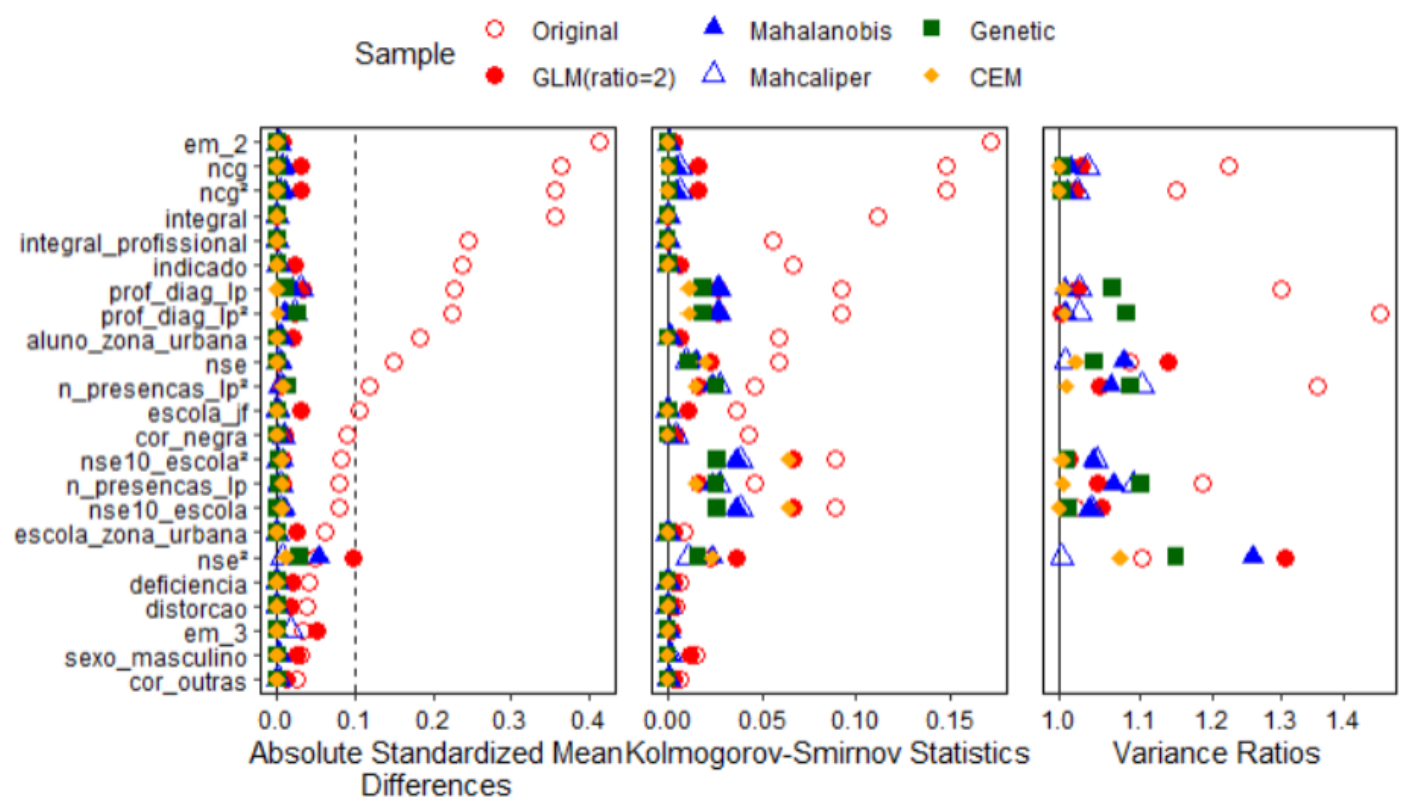
Aplique diferentes abordagens de pareamento e favoreça aquela que oferece melhor balanceamento; reporte várias estimativas

- Entre as **variáveis de pareamento**, inclua o maior número possível de fatores que você acredita **determinarem a participação no tratamento**
- Há vários **testes para balanceamento**; na dúvida, **favoreça testes multivariados**
 - **Testes univariados**: consideram o balanceamento de cada variável de pareamento separadamente – e.g., teste t de diferença de médias
 - **Testes multivariados**: consideram o conjunto de variáveis de pareamento – e.g., teste Kolmogorov-Smirnov (**KS**); **podem não ser viáveis** (“não rodarem”) se houver muitas dimensões de pareamento

Pacote `coba1t`: Love plots para múltiplos tipos de pareamento

<https://cran.r-project.org/web/packages/MatchIt/vignettes/assessing-balance.html#assessing-balance-with-cobalt>

Gráfico 22 – Análise do balanceamento das covariáveis para definição do modelo de *matching*



Fonte: Minas Gerais (2024a, 2024b), Brasil (2015), Soares e Alves (2023). Elaboração própria.

AVALIAÇÃO DA POLÍTICA DE COTAS NA UNIVERSIDADE FEDERAL DE VIÇOSA

Felipe Miranda de Souza Almeida¹

Cristiana Tristão Rodrigues²

Este trabalho visa avaliar a política de cotas na Universidade Federal de Viçosa (UFV), comparando o desempenho escolar e o número de reprovações entre os estudantes beneficiados ou não pela Lei de Cotas nos anos 2013, 2014 e 2015. Os resultados, encontrados por meio da metodologia *propensity score*, mostram que não há diferença significativa entre os dois grupos. Assim, o resultado vai ao encontro do objetivo da lei, que é ampliar o acesso às universidades e aos institutos federais para os jovens das escolas públicas, para os pretos e pardos e para os índios, sem diminuir a qualidade de ensino.

Palavras-chave: políticas de cotas; ensino superior; *propensity score*.

planejamento e políticas públicas | ppp | n. 53 | jul./dez. 2019

www.ipea.gov.br/ppp/index.php/PPP/article/view/868

Apresente múltiplas estimativas: exemplo

Diferentes critérios de *matching* podem ser usados para atribuir participantes a não participantes em função do PS. Isto implica calcular um peso para cada conjunto de participante e não participante. Como discutido a seguir, a escolha de uma técnica de correspondência particular pode, por conseguinte, afetar a estimativa resultante, por meio dos pesos atribuídos (Khandker, Koolwal e Samad, 2010).

- 1) Vizinho mais próximo (NN): é uma das técnicas de pareamento mais utilizadas, em que cada unidade de tratamento é comparada com a unidade de comparação com o PS mais próximo. Pode-se também escolher vizinhos mais próximos e fazer n pareamentos (normalmente é utilizado $n = 5$). O *matching* pode ser feito com ou sem substituição. Com a substituição, por exemplo, significa que o mesmo não participante pode ser utilizado como correspondente para diferentes participantes.
- 2) Calibre e radial: um problema com o *matching* NN é que a diferença no PS para um participante e o seu vizinho mais próximo não participante da política ainda pode ser muito alta. Esta situação resulta em comparações pobres e pode ser evitada mediante imposição de um limite ou “tolerância” na distância PS máxima (calibre). Este procedimento envolve pareamento com substituição apenas entre os PS dentro de um determinado intervalo. Uma maior redução de indivíduos não participantes provavelmente aumenta a chance de viés de amostragem.
- 3) Estratificação e intervalo: este procedimento particiona o suporte comum em diferentes estratos (ou intervalos) e calcula o impacto do programa em cada intervalo. Especificamente, dentro de cada intervalo, o efeito do programa é a diferença média nos resultados entre observações de tratados e de controle. A média ponderada destas estimativas de impacto produz o impacto global do programa, tendo a porcentagem de participantes em cada intervalo como os pesos.
- 4) Kernel e local linear: os algoritmos de pareamento discutidos até agora têm em comum que apenas algumas observações, a partir do grupo de comparação, são usadas para construir o resultado contrafactual de um indivíduo tratado. *Matching* Kernel (KM) e *matching* linear local (LLM) são estimadores não paramétricos de pareamento que utilizam ponderação das médias de todos os indivíduos no grupo de controle para construir o resultado contrafactual. Assim, uma vantagem principal destas abordagens é a mais baixa variância, devido à utilização de mais informação. Uma desvantagem é que, possivelmente, as observações são comparações ruins. Assim, a instituição adequada da condição de suporte comum é de grande importância para KM e LLM.

Avaliação da Política de Cotas na Universidade Federal de Viçosa

371

Na literatura que aborda as diferentes técnicas de pareamento, não há um consenso sobre qual técnica apresentaria melhores resultados, mas argumenta que a consideração conjunta das técnicas pode oferecer um método para avaliar se as estimativas são robustas. Seguindo esta linha, neste trabalho, serão utilizadas as quatro técnicas, visando uma estimativa robusta do possível efeito dos tratamentos.

Assim, o impacto da Lei nº 12.71/2012 sobre o rendimento acadêmico e o número de reprovações dos estudantes que ingressaram pelo sistema de cotas (ATT) foi estimado a partir da comparação entre os que ingressaram na universidade por meio da política e os que ingressaram pelo sistema de ampla concorrência, selecionados por suas características observáveis a partir da estimação do *propensity score* e pareados pelos algoritmos de vizinho mais próximo, radial, estratificação e de Kernel.

Almeida e Rodrigues (2019, p. 370-371)

www.ipea.gov.br/ppp/index.php/PPP/article/view/868

Apresente múltiplas estimativas: exemplo

Avaliação da Política de Cotas na Universidade Federal de Viçosa

375

Nas tabelas 4 e 5, são reportadas as estimativas do valor do efeito do tratamento sobre o coeficiente de rendimento acumulado e o número de reprovações, respectivamente.

TABELA 4
Cálculo do efeito do tratamento sobre o coeficiente de rendimento acumulado (2013-2015)

		NN	Radial	Estratificado	Kernel
2013	ATT	2,164	3,085	3,501	3,336
	Erro-padrão	1,969	1,399	1,566	1,896
	T	1,099	2,205	2,236	1,760
2014	ATT	-1,291	-2,629	-1,005	-1,91
	Erro-padrão	3,470	2,348	5,341	2,336
	T	-0,372	-1,120	-0,188	-0,817
2015	ATT	-2,324	-0,603	-2,118	-1,763
	Erro-padrão	3,100	1,948	1,826	2,160
	T	-0,750	-0,310	-1,160	-0,816

Elaboração dos autores.

Analisando os resultados da tabela 4, temos que, em todos os modelos, o efeito do tratamento sobre a variável rendimento acadêmico acumulado não foi significativo, indicando que não há diferença entre o desempenho acadêmico dos estudantes que ingressaram pelo sistema de cotas e pelo sistema de ampla concorrência.

Os resultados da tabela 5 são similares aos encontrados anteriormente, e são, de certa maneira, implícitos, dado que o cálculo do CRA leva em consideração a nota obtida nas disciplinas e possui relação positiva. Para um aluno ser reprovado, ele deve possuir uma nota inferior ou igual a 59,9 pontos na disciplina. Logo, uma menor nota implica um menor coeficiente de rendimento.

TABELA 5
Cálculo do efeito do tratamento sobre o número de reprovações – NReprovacoes (2013-2015)

		NN	Radial	Estratificado	Kernel
2013	ATT	-0,381	-0,468	-0,4559	-0,438
	Erro-padrão	0,459	0,326	0,351	0,379
	T	-0,829	-1,433	-1,311	-1,154
2014	ATT	-0,317	-0,020	-0,086	0,002
	Erro-padrão	0,844	0,508	0,747	0,521
	T	-0,375	-0,040	0,521	0,003
2015	ATT	-0,481	0,082	0,148	0,175
	Erro-padrão	0,353	0,231	0,244	0,230
	T	1,364	0,354	0,606	0,758

Elaboração dos autores.

Almeida e Rodrigues (2019, p. 375)

www.ipea.gov.br/ppp/index.php/PPP/article/view/868

Duas orientações gerais

1

Aplique diferentes abordagens de pareamento e favoreça aquela que oferece melhor balanceamento; reporte várias estimativas

- Entre as **variáveis de pareamento**, inclua o maior número possível de fatores que você acredita **determinarem a participação no tratamento**
- Há vários **testes para balanceamento**; na dúvida, **favoreça testes multivariados**
 - **Testes univariados**: consideram o balanceamento de cada variável de pareamento separadamente – e.g., teste t de diferença de médias
 - **Testes multivariados**: consideram o conjunto de variáveis de pareamento – e.g., teste Kolmogorov-Smirnov (**KS**); **podem não ser viáveis** (“não rodarem”) se houver muitas dimensões de pareamento

2

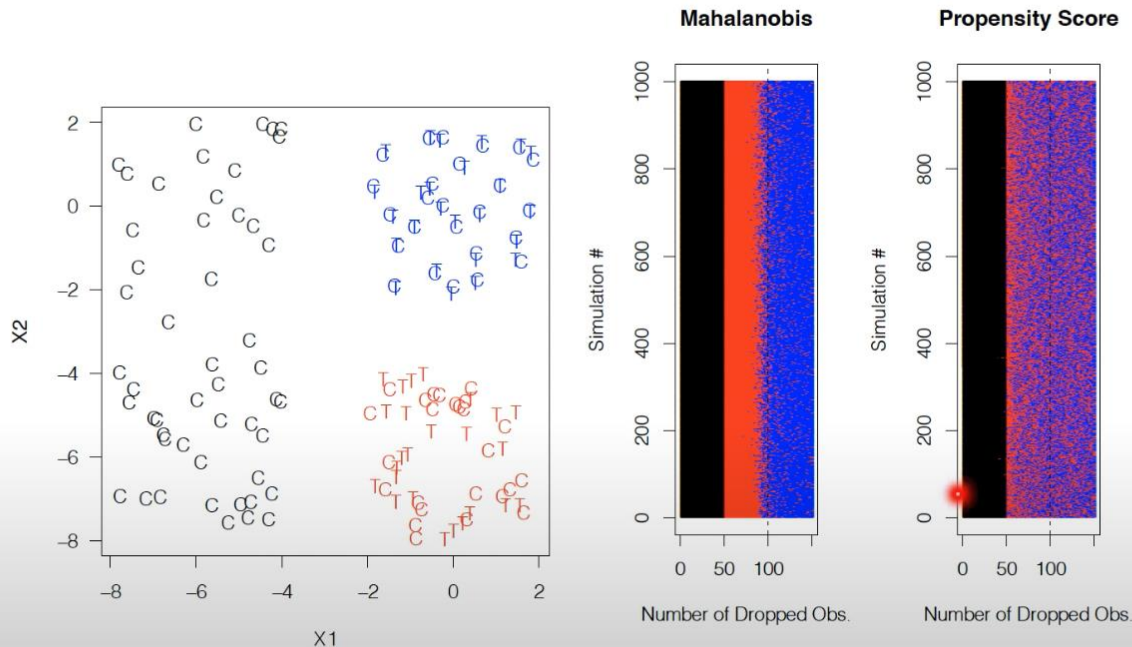
Desconfie do PSM (propensity score matching)

- Abordagem **mais conhecida e utilizada** para pareamento (virou quase sinônimo de pareamento)
- Todavia, sua capacidade de gerar balanceamento é **inferior às alternativas** (e.g., Mahalanobis); em certas condições, é tão “cega” que se aproxima do *random pruning* (remoção aleatória de observações)

PSM vs. MDM

16. Matching Methods

PSM Is Blind Where Other Methods Can See



Os dois gráficos da direita mostram a cor das observações removidas, na ordem em que foram removidas, no pareamento pela distância de Mahalanobis e por propensity score. Esses gráficos reportam 1.000 simulações de pareamento, todas baseadas em valores de X1 e X2 (variáveis de pareamento) gerados aleatoriamente, mas que compartilham o padrão geral do gráfico à esquerda. <https://youtu.be/tvMyjDi4dyg>

Pareamento: limitações

- Somente reduz o viés causado por **características observáveis**: se o status do tratamento for influenciado por características não observáveis (quebra do pressuposto de independência condicional) e que não sejam altamente correlacionadas com as variáveis usadas no pareamento, os impactos **estimados serão enviesados (não por conta de viés e pareamento, mas por conta de viés de seleção na exposição ao tratamento)**
- Compromete validade interna se falta de área de suporte comum causar **descarte de um grupo de observações sistematicamente diferentes** daquelas retidas
- Requer uma grande **quantidade de variáveis de pareamento e de observações**

... quando os dados de linha de base [pré-teste] estão disponíveis, o pareamento baseado nas características socioeconômicas da linha de base pode ser muito útil se for combinado a outras técnicas, como diferença em diferenças, o que nos permitirá corrigir pelas diferenças entre os grupos que são fixas ao longo do tempo. O pareamento também é mais confiável quando a regra de seleção do programa e as variáveis subjacentes são conhecidas e, nesse caso, o pareamento pode ser realizado nessas variáveis.

Gertler et al. (2018, p. 1694)

Pareamento (matching)

Avaliação de Políticas Públicas B

30/abril, 07, 12, 21, 26 e 28/maio e 02/junho de 2025

Leitura básica:

FACURE, Matheus. **Causal inference for the brave and true**. 2022. [Capítulo 10: Matching; Capítulo 11: Propensity score.] Disponível em: <https://matheusfacure.github.io/python-causality-handbook/landing-page.html>

Leitura complementar:

GERTLER, Paul J. et al. **Avaliação de impacto na prática**. Washington: Banco Interamericano de Desenvolvimento; Banco Mundial, 2018. [Capítulo 8: Pareamento.] Disponível em: <https://openknowledge.worldbank.org/bitstream/handle/10986/25030/9781464808890.pdf>

HEINRICH, Carolyn; MAFFIOLI, Alessandro; VÁZQUEZ, Gonzalo. **A primer for applying propensity-score matching**. Impact-Evaluation Guidelines, Technical Notes n. IDB-TN-161. Washington: Banco Interamericano de Desenvolvimento, 2010. <http://dx.doi.org/10.18235/0008567>

HUNTINGTON-KLEIN, Nick. **The effect: an introduction to research design and causality**. 2023. [Capítulo 14: Matching.] Disponível em: <https://theeffectbook.net/>

extra: controle
estatístico com PS

Em estudos observacionais, usaremos estratégias para lidar com a falta de comparabilidade presumida entre grupos

Problema	Estratégia	Exemplo de abordagem
- Em estudos observacionais (i.e., em que não há atribuição aleatória ao tratamento), não há presunção de comparabilidade entre grupo tratado e não tratado (antes da exposição ao tratamento). - De fato, esses grupos podem diferir substancialmente, tanto nas características observáveis como nas não observáveis. Por isso, a associação observada entre T (tratamento) e Y (variável de resultado) não pode ser considerada uma estimativa do efeito causal de T em Y. - Para a estimação de efeito, teremos de usar alguma estratégia para enfrentar a não comparabilidade entre os grupos.	1 Modelar o mecanismo de tratamento (T)	- Variável instrumental - Matching (pode envolver propensity score – PS) - Inverse Probability of Treatment Weighting (IPTW) com probabilidade de tratamento definida como PS
	2 Modelar a variável de resultado (Y)	- Regressão com descontinuidade - Diferença em diferenças - Controle sintético - Efeitos fixos Controle estatístico (pode envolver PS)
	3 Uma combinação das estratégias anteriores	Estimação do tipo double robust

PS como covariável

Causal Inference for the Brave and True

Search this book...

Causal Inference for The Brave and True

PART I - THE YANG

01 - Introduction To Causality

02 - Randomised Experiments

03 - Stats Review: The Most Dangerous Equation

04 - Graphical Causal Models

05 - The Unreasonable Effectiveness of Linear Regression

06 - Grouped and Dummy Regression

07 - Beyond Confounders

08 - Instrumental Variables

09 - Non Compliance and LATE

10 - Matching

11 - Propensity Score

12 - Doubly Robust Estimation

13 - Difference-in-Differences

14 - Panel Data and Fixed Effects

15 - Synthetic Control



Propensity Score Matching

As I've said before, you don't need to control for X when you have the propensity score. It suffices to control for it. As such, you can think of the propensity score as performing a kind of dimensionality reduction on the feature space. It condenses all the features in X into a single treatment assignment dimension. For this reason, we can treat the propensity score as an input feature for other models. Take a regression, model for instance.

```
smf.ols("achievement_score ~ intervention + propensity_score", data=data_ps).fit().summary().ta
```

	coef	std err	t	P> t	[0.025	0.975]
Intercept	-3.0770	0.065	-47.061	0.000	-3.205	-2.949
intervention	0.3930	0.019	20.974	0.000	0.356	0.430
propensity_score	9.0554	0.200	45.314	0.000	8.664	9.447

If we control for the propensity score, we now estimate a ATE of 0.39, which is lower than the 0.47 we got previously with a regression model without controlling for the propensity score. We can also use matching on the propensity score. This time, instead of trying to find matches that are similar in all the X features, we can find matches that just have the same propensity score.

This is a huge improvement on top of the matching estimator, since it deals with the curse of dimensionality. Also, if a feature is unimportant for the treatment assignment, the propensity score model will learn that and give low importance to it when fitting the treatment mechanism. Matching on the features, on the other hand, would still try to find matches where individuals are similar on this unimportant feature.

Contents

The Psychology of Growth

Propensity Score

Propensity Weighting

Propensity Score Estimation

Standard Error

Common Issues with Propensity Score

Propensity Score Matching

Key Ideas

References

Contribute