

# **AULA EXTRA**

# **Escore de propensão de pareamento**

**Ernesto F. L. Amaral**

**Avaliação de Políticas Públicas (DCP 046)**

# DESENHOS DE EXPERIMENTOS NÃO ALEATÓRIOS

- Há muitos exemplos destes tipos de estudos:
  - Quase-experimentais.
  - Estudos de controle de casos.
  - Análises transversais (*cross-sectional*) repetidas.
- A principal distinção para os experimentos clássicos é que os grupos comparados não são criados com escolha aleatória.
- Indivíduos não são designados para o tratamento por um processo aleatório (não temos um grupo de controle experimental).
- Escolha aleatória nem sempre é possível.

## AVALIAÇÃO NÃO EXPERIMENTAL

- Como nem sempre dados aleatórios são disponíveis, é essencial entender e modelar os processos pelos quais as designações para tratamentos são realizadas.
- Auto-seleção (decisão individual deve ser considerada).
- Seleção do administrador (indivíduos são selecionados para tratamento com base em critérios específicos).
- Combinação de auto-seleção e seleção do administrador.

# VANTAGENS DE EXPERIMENTOS NÃO ALEATÓRIOS

- Custo é muito menor.
- Geralmente demanda menos tempo.
- Pode ser mais generalizável.
- Utiliza grandes amostras (aumenta representatividade, mas também aumenta custo).

# DESVANTAGEM DE EXPERIMENTOS NÃO ALEATÓRIOS

- Principal desvantagem é a suscetibilidade de viés de seletividade.
- Os grupos podem ser auto-selecionados.
- O mecanismo de seleção pode estar relacionado com o resultado final (*outcomes*). Por exemplo, o Programa Bolsa Família é direcionado para pessoas com certo perfil domiciliar e de renda.
- As características da linha de base podem confundir as análises.
- Uma técnica que pode diminuir o viés de seletividade de desenhos não aleatórios é o escore de propensão de pareamento (*propensity score matching - PSM*).

## CONTRAFACTUAL

- A lógica de técnicas de escore de pareamento é baseada na teoria do contrafactual.
- Há o emprego da linguagem de tratamento e controle.
- Podemos observar resultados para aqueles que receberam o tratamento e para os que não receberam.
- Queremos estimar: (1) o que teria acontecido se as pessoas que receberam o tratamento não o tivessem adquirido; (2) o que teria acontecido se o grupo de controle tivesse recebido tratamento.
- Não podemos verdadeiramente analisar o contrafactual.

# PROBLEMA DA AVALIAÇÃO DE IMPACTO

- A = domicílios que recebem o programa.
- B = domicílios que não recebem o programa.
- Y = resultado = % frequência à escola
- ATT = efeito médio do tratamento no tratado (*average treatment effect on the treated*)

Não observado

$$ATT = (Y_A | A \text{ participa}) - (Y_A | A \text{ não participa})$$

$$(Y_A | A \text{ participa}) - (Y_B | B \text{ não participa}) =$$

$$ATT + (\text{Diferença entre A e B}) \text{ Viés de seleção}$$

- Podemos usar B como contrafactual somente se a diferença entre A e B for igual a zero.

## INTUIÇÃO BÁSICA DE DD E PSM

- Tanto o escore de propensão de pareamento (PSM) como o método de diferença em diferenças (DD) tentam corrigir o viés de seleção.
- PSM remove viés associado com **características observáveis** que afetam a designação do tratamento.
- DD remove viés associado com **características constantes no tempo** (*dummy* de tempo), tanto observáveis como não observáveis, que afetam a designação do tratamento.
- A combinação de PSM e DD pode melhorar as estimações.
- Se a designação do tratamento é afetada por características não-observáveis que variam no tempo, é preciso utilizar variáveis instrumentais.



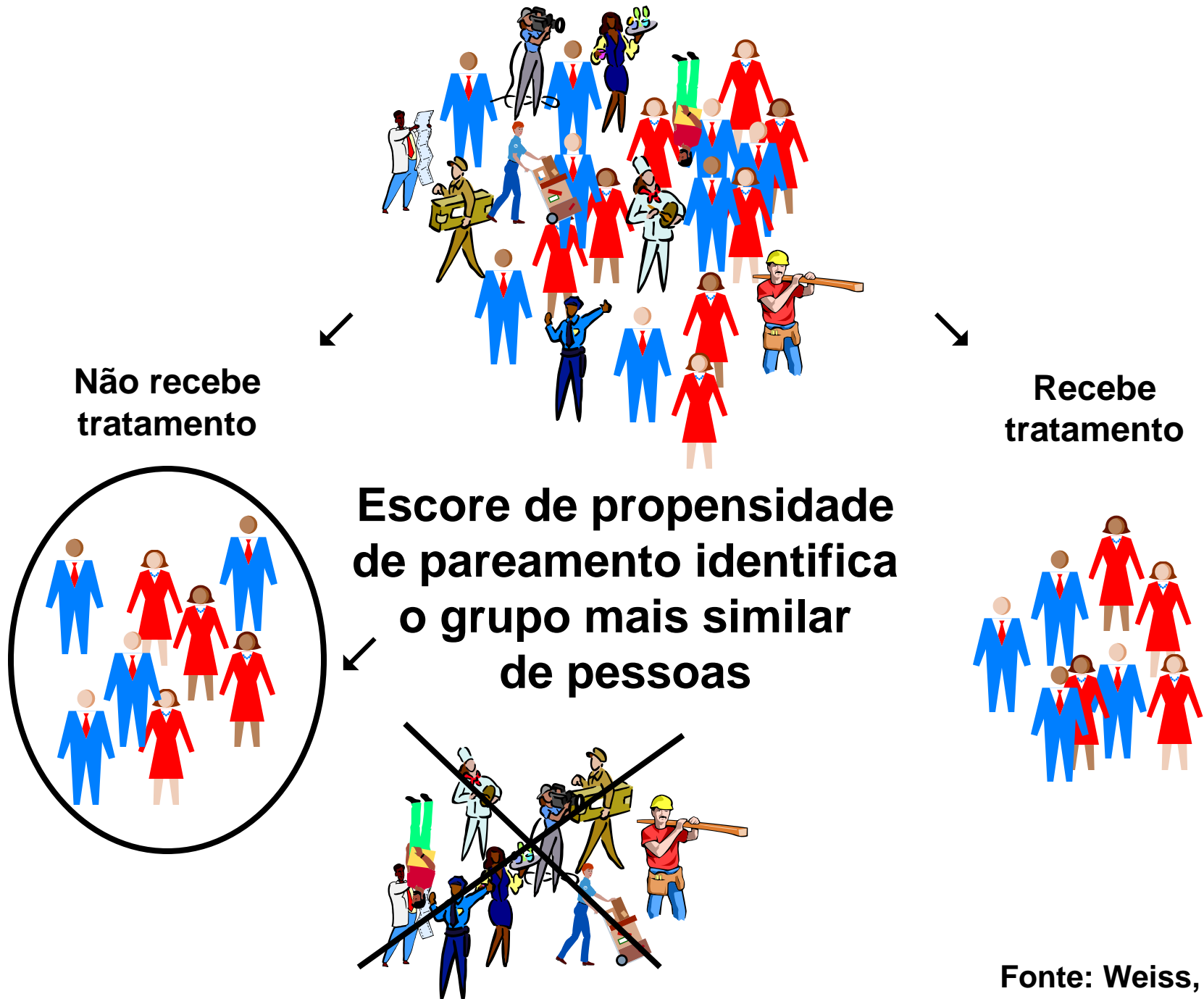
## ESCORE DE PROPENSIDADE DE PAREAMENTO

- Técnicas de escore de propensidade fornecem estratégia de correção que permite a estimação do contrafactual.
- É utilizada a probabilidade predita de pertencimento ao grupo (por exemplo, grupo de tratamento x grupo de controle), baseado em preditores observados (variáveis independentes).
- Geralmente, as probabilidades preditas são obtidas com regressão logística para criar um grupo contrafactual.

## ESCORE DE PROPENSIDADE DE PAREAMENTO

- Portanto, o escore de propensidade é a probabilidade condicional de receber um tratamento, dado um vetor de covariáveis mensuradas (variáveis independentes observadas).
- É uma estratégia particularmente útil para casos em que a escolha aleatória não pode ser empregada.
- Os escores de propensidade pode ser usados para:
  - Realizar o pareamento dos grupos.
  - Incluir como covariáveis em modelos que estimam impacto de políticas (sozinhas; ou com outras variáveis de pareamento; ou com outras covariáveis).





# EXEMPLO DE PAREAMENTO

- A = domicílios que recebem o programa.
- B = domicílios que não recebem o programa.

HH	TREATMENT	EDUCATION	INCOME
1	B	2	60
2	B	3	80
3	B	5	90
4	B	12	200
5	A	5	100
6	A	3	80
7	A	4	90
8	A	2	70

MATCH
[3]
[2]
[2,3]
[1]

INCOME A	INCOME COUNT	DIFF
100	90	10
80	80	0
90	$(80+90)/2=85$	5
70	60	10

ATT	6.25
-----	------

## PRESSUPOSTO DE INDEPENDÊNCIA CONDICIONAL

- Pressuposto de independência condicional (*conditional independence assumption - CIA*) ou condição de não confundimento.
- Dado um conjunto de covariáveis observáveis  $X$ , que não são afetadas pelo tratamento, os potenciais resultados são independentes da designação do tratamento.
- **Na prática**, o pareamento estima o efeito da política no grupo de tratamento, assumindo que, condicionada por características observáveis, a participação é independente dos resultados (efeitos e impactos) da política.

## QUEBRA DO PRESSUPOSTO CIA

- O escore de pareamento remove vieses associados com diferenças pré-política entre grupos de controle e tratamento.
- É útil quando dados pré-política são ricos em características observáveis.
- **Limitação:** se o status de tratamento é influenciado por características não observáveis (quebra do pressuposto de independência condicional), os impactos estimados serão enviesados.

## PRESSUPOSTO DE SOBREPOSIÇÃO

- Escore de propensidade de pareamento: estimação de um “modelo de participação” que reduz o problema de pareamento para uma dimensão única (escore de propensidade).
- Os escores de propensidade são então utilizados para parear os grupos de tratamento e de controle.
- Precisamos ter indivíduos no grupo de controle que tenham características similares aos indivíduos no grupo de tratamento.
- Esta é a condição de sobreposição em PSM.



## PORTANTO, TEMOS ESTES REQUISITOS

- Para que as técnicas de escore de propensidade operem corretamente precisamos de dados com algumas especificidades...
- Grande número de casos na amostra (particularmente para o grupo de controle).
- Grande número de covariáveis usadas para estimar o escore de propensidade.
- Grau razoável de sobreposição entre os grupos de controle e de tratamento.
- Se não há sobreposição, então os dois grupos podem ser muito diferentes para justificar uma comparação.

# CONDIÇÃO DE SOBREPOSIÇÃO

Density

Density of scores for participants

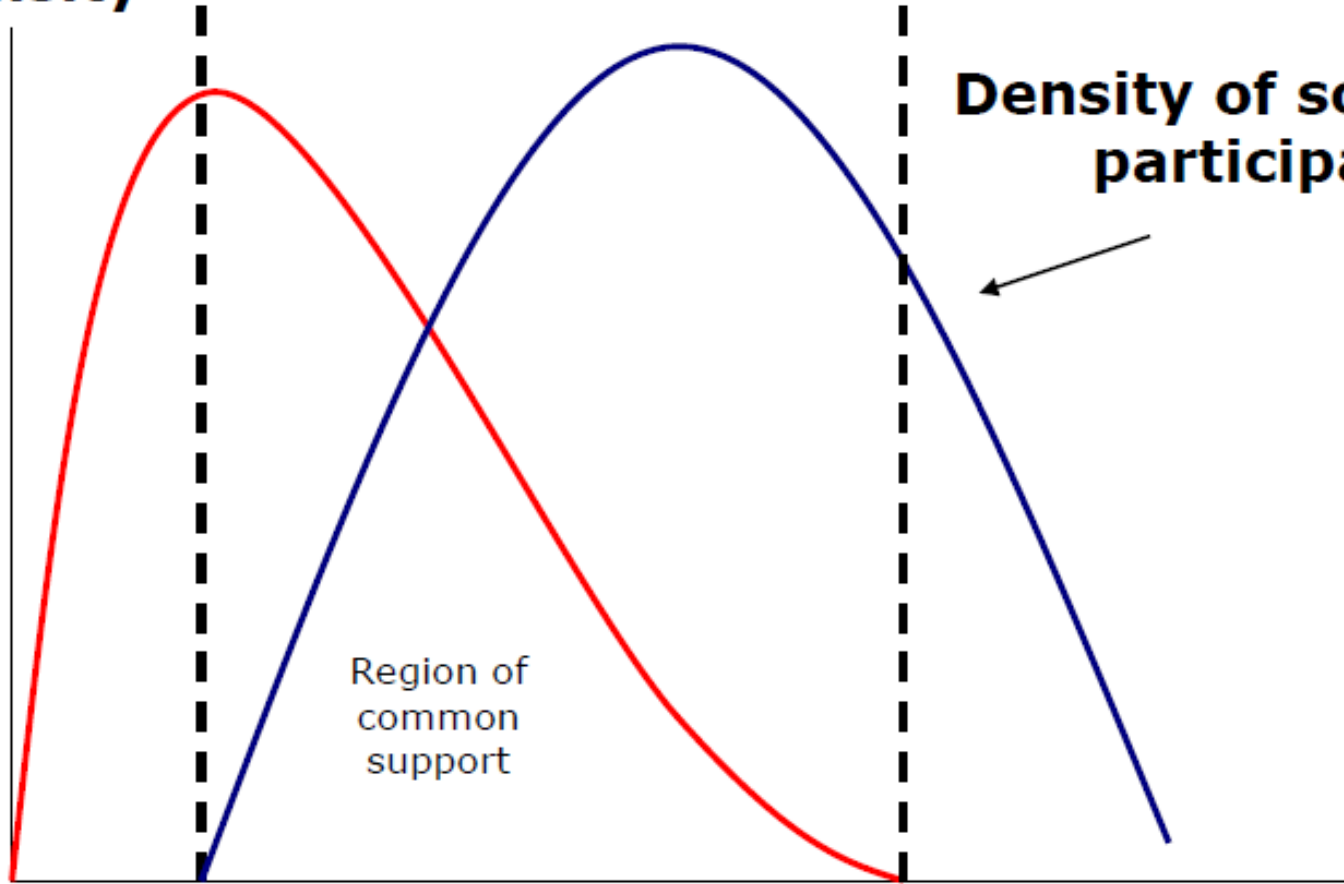
Region of common support

0

Propensity score

1

High probability of participating given X



# IMPLEMENTAÇÃO DO PSM

## **1) Estime modelo de escolha discreta (ex. logístico):**

- Variável dependente:  $Y=1$  se participa;  $Y=0$  se não participa.
- Escolha as variáveis condicionais apropriadas.
- Obtenha o escore de propensidade: probabilidade predita.

## **2) Faça pareamento de cada participante com um ou mais não-participantes, com base no escore de propensidade:**

- Escolha o algoritmo de pareamento.
- Identifique a condição de sobreposição.
- Verifique a qualidade do pareamento.

## **3) Estime o impacto da política com base na nova amostra.**

# ESTIMAÇÃO DO MODELO (ESCORE DE PROPENSIDADE)<sup>20</sup>

- **Escolha o modelo:** modelos logísticos ou probit (multinomial para mais de dois grupos... multi-tratamento).
- **Escolha as covariáveis:** critério básico é escolher variáveis que satisfaçam pressuposto CIA (variáveis que afetam tanto decisão de participação, como o efeito da política).
- Como escolher e saber se temos poucas ou muitas variáveis?... (1) teoria; (2) qualidade do pareamento; (3) especificação parcimoniosa e adição progressiva; (4) estimativa de ajuste do modelo (*goodness-of-fit*).
- Fatores óbvios para incluir na estimação de PSM são os critérios usados para determinar participação no programa.
- Objetivo é pareamento e não estimação de coeficiente.
- Utilize mesmas fontes para controle e tratamento.

## ESCOLHA O ALGORITMO DE PAREAMENTO

- Algoritmos de pareamento são as diferentes alternativas de utilizar o escore de propensidade para parear as unidades de controle com unidades de tratamento. É preciso considerar:
- Parear com ou sem reposição (decisão se membro do grupo de controle pode ser pareado somente uma vez): com reposição é mais eficiente.
- Verificar proximidade do pareamento: melhor usar razões de log do escore  $[\ln(p/1-p)]$ , ao invés de distância absoluta.
- Decidir pelo ponderamento de casos na análise (utilização de pesos) e forma como é realizado: modelo Kernel.
- Determinar o número de unidades de controle pareadas com cada unidade de tratamento: melhor parear vários controles com um tratamento.

# DIFERENTES ALGORITMOS DE PAREAMENTO

Model	Description	Alternatives	Key factor	BIAS	EFF
Nearest Neighbor	Choose the control(s) with the minimum mahalanobis distance	<ul style="list-style-type: none"> <li>No replacement</li> <li>Replacement</li> <li>Oversampling</li> </ul>	<ul style="list-style-type: none"> <li>Order of matching</li> <li>Poorer matches</li> <li>Poorer matches</li> </ul>	<ul style="list-style-type: none"> <li>✓ ✓</li> <li>✓</li> <li>✓</li> </ul>	<ul style="list-style-type: none"> <li>× ×</li> <li>×</li> <li>×</li> </ul>
Caliper and Radius	Choose the control(s) within a certain distance	<ul style="list-style-type: none"> <li>Caliper (NNnR)</li> <li>Radius (NNR)</li> </ul>	<ul style="list-style-type: none"> <li>Tolerance level</li> <li>Radius definition</li> </ul>	<ul style="list-style-type: none"> <li>✓ ✓</li> <li>✓</li> </ul>	<ul style="list-style-type: none"> <li>× ×</li> <li>×</li> </ul>
Stratification	Divide the common support in strata	<ul style="list-style-type: none"> <li>No. of strata</li> </ul>	<ul style="list-style-type: none"> <li>Definition of the No. of strata</li> </ul>	<ul style="list-style-type: none"> <li>✓</li> </ul>	<ul style="list-style-type: none"> <li>×</li> </ul>
Kernel	Use weighted average of all the individual in the common support	<ul style="list-style-type: none"> <li>Kernel functions</li> <li>Bandwith</li> </ul>	<ul style="list-style-type: none"> <li>Proper definition of the common support</li> </ul>	<ul style="list-style-type: none"> <li>×</li> </ul>	<ul style="list-style-type: none"> <li>✓ ✓</li> </ul>

## IDENTIFIQUE A CONDIÇÃO DE SOBREPOSIÇÃO

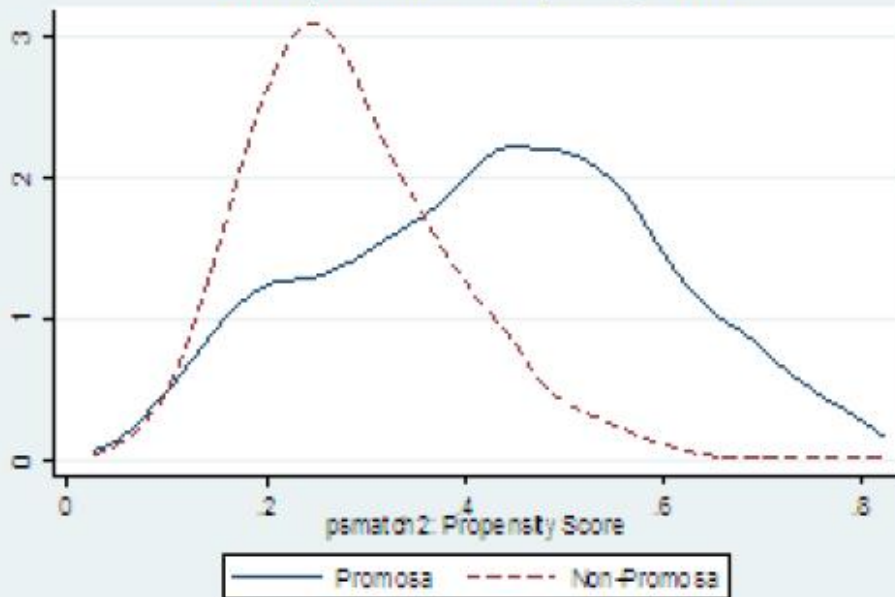
- Diferentes estratégias podem ser usadas para identificar a condição de sobreposição entre os grupos.
- **Análise visual:** faça um gráfico com a distribuição de densidade do escore de propensidade para os dois grupos, com a sobreposição das distribuições (a seguir).
- **Critério mínimo e máximo:** remova todas observações em que o escore de propensidade é menor que o mínimo e maior que o máximo do outro grupo (a seguir).
- **Aparando:** excluir todas observações nas áreas em que uma das duas distribuições do escore de pareamento é zero.
- **Similaridade estatística das distribuições do p-valor:** estime teste de dissimilaridade das 2 distribuições (a seguir).

# ANÁLISE VISUAL

- Sobreposição das curvas de densidade do escore de propensividade dos grupos de controle e tratamento, antes e depois do pareamento.

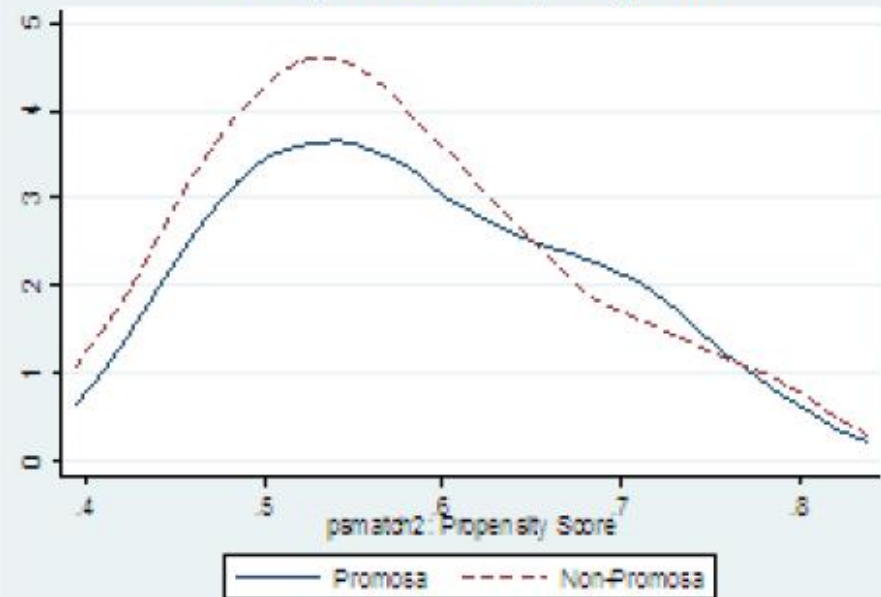
**Antes do pareamento**

Density Function - Propensity Score



**Depois do pareamento**

Density Function - Propensity Score

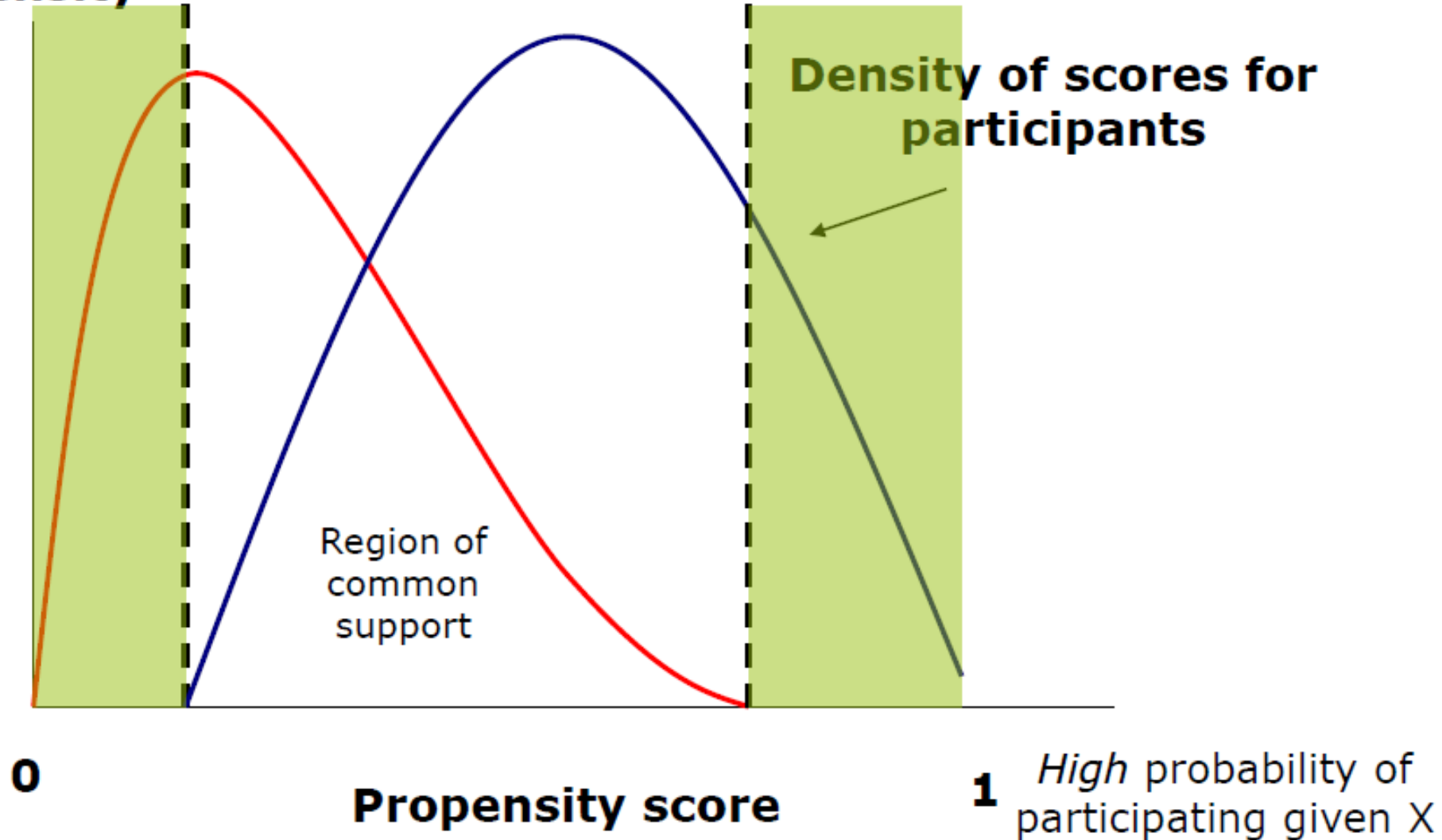


Fonte: Maffioli, 2011.



# CRITÉRIO MÍNIMO E MÁXIMO

Density



# SIMILARIDADE ESTATÍSTICA ENTRE DISTRIBUIÇÕES

- A similaridade estatística das distribuições do p-valor pode ser realizada com o teste de Kolmogorov-Smirnov (ksmirnov).
- Esse é um teste estatístico mais rigoroso, do que a simples análise visual.
- A hipótese nula é que há igualdade entre as distribuições dos escores de pareamento dos grupos de controle e tratamento após pareamento.
- Desejamos que a probabilidade de igualdade entre as distribuições seja maior que 0,05, ou seja, não rejeitaríamos a hipótese de igualdade entre as curvas.

## VERIFIQUE A QUALIDADE DO PAREAMENTO

- Dado o pressuposto de independência condicional, devemos checar se o procedimento de pareamento fez um balanceamento das distribuições das variáveis relevantes, entre os grupos de controle e tratamento.
- Verifique a redução do viés padronizado antes/depois do pareamento (pstest com Stata): pelo menos 5% de redução.
- Teste a igualdade de médias nos grupos de controle e tratamento antes/depois do pareamento (pstest com Stata).
- Verifique a significância conjunta do “modelo de participação”: após pareamento, o pseudo- $R^2$  deve ser baixo.
- Se a qualidade do pareamento não é satisfatória, o pressuposto CIA falhou.

## RESUMINDO

- Quando um avaliador utiliza a metodologia PSM, devemos verificar:
- Quais dados de características observáveis estão disponíveis?
- Quais variáveis são incluídas no “modelo de participação” e porque?
- Qual algoritmo de pareamento é utilizado e porque?
- Qual a condição de sobreposição utilizada?
- Quais são os resultados do balanceamento das covariáveis relevantes (pstest) e da distribuição dos escores de pareamento (visual, ksmirnov...)?

## PRINCIPAIS LIMITAÇÕES DO MÉTODO PSM

- Somente reduz o viés causado por características observáveis.
- Validade interna é limitada pela condição de sobreposição.
- É necessário possuir uma grande quantidade de dados (grande amostra para grupo de controle).

## COMBINANDO PSM E DD

- **Vantagens de combinar os dois métodos:**
- DD melhora PSM, porque controla por heterogeneidade não observada, se for constante no tempo.
- PSM melhora DD, porque pode fazer com que o pressuposto de tendência paralela seja mais aceitável.
- **Como implementar o PSM-DD:**
- PSM tradicional com DD: faça pareamento com base nas características pré-política e estime o impacto com dupla diferença (controle/tratamento & antes/depois).
- Quando houver disponibilidade de dados, faça pareamento das tendências pré-política (mais de um tempo anterior ao tratamento) na variável de interesse.

## PRINCIPAIS REFERÊNCIAS

- Heinrich, Carolyn, Alessandro Maffioli, Gonzalo Vázquez. 2010. “A Primer for Applying Propensity-Score Matching”. Impact-Evaluation Guidelines, Technical Notes, No. IDB-TN-161.
- Maffioli, Alessandro. 2011. “Non-Experimental Methods: Propensity Score Matching and Difference in Difference”. Cuernavaca:IDB.(<http://www.impactevaluation2011.org/forum/wp-content/uploads/2011/06/C5-commissioners-of-evaluations-file-1.pdf>)
- Weiss, Christopher C. 2010. “Quantitative Methodology and Public Policies: New Techniques and Old Problems”. Apresentação no 1º Seminário Internacional de Gestão e Políticas Públicas: Teoria e Prática. Belo Horizonte: PUBLICUS/DGP/FAFICH/UFMG.