Avaliação de Políticas Públicas II: Métodos não experimentais

Paula Pereda (USP)

October 30, 2020

Aula 4

voltar

Resumo da Aula 3

1. **Estimação** do propensity score: Especificação, variáveis e estimação.

2. **Propensity Score Matching**: Métricas e mecânica de estimação (matching local), elimina observações não pareadas.

3. **Procedimento de *Trimming***: Fácil de implementar. Recurso para obter suporte comum. Checar quantas unidades são excluídas (o quanto afeta análise).

Inferência: PS Matching

- Questão relevante: Uso de $\hat{e}(X)$ no primeiro estágio afeta distribuição do efeito do tratamento (Abadie e Imbens, 2016)
- Abadie e Imbens (2006) derivam distribuição (implementada nos softwares)
- Abadie e Imbens (2016) propõe distribuição assintótica
- Standard bootstrap pode ser problemático em alguns casos: Bodory et al (2020) mostram os casos em que são possíveis usar (recomendam o uso do wild bootstrap em alguns casos, veremos mais a frente no curso)
- Outros: Para PSM e amostras complexas, ver Austin et al (2016) e Bodory et al (2020)

Pareamento Global: Optimal Matching

- Sobrepõe dificuldades do matching local (“greedy”).
- Revisão: Rosenbaum (2002); Implementação no R: Hansen (2007).
- Sejam 2 grupos de participantes A (tratados) e B (não-tratados) e vale $A \cap B = \emptyset$,
- $|A| = \alpha$ e $|B| = \beta$ são o número de participantes de cada grupo .
- Temos que existe a distância $d_{ab}, \forall a \in A, b \in B$ t.q. $0 \leq d_{ab} < \infty$. d_{ab} é a distância tomada em termos de covariadas ou do propensity score.
- Um matching consiste em produzir s estratos, $(A_1, \ldots, A_s, B_1, \ldots, B_s)$, com as seguintes propriedades:
  1. $|A_t|, |B_t| \geq 1, \forall t$;
  2. $A_t \cap A_{t'} = \emptyset, \forall t \neq t'$;
  3. $B_t \cap B_{t'} = \emptyset, \forall t \neq t'$;
  4. $A_1 \cap \ldots \cap A_s = A_S \subseteq A$; e
  5. $B_1 \cap \ldots \cap B_s = B_S \subseteq B$.

Pareamento Global: Optimal Matching

- Um matching produz S pares de conjuntos cada um contendo elementos similares em termos de propensity scores: $|A_1|e|B_1|,...,|A_S|e|B_S|$
- Esses matchings podem ser classificados em 3 tipos:
  1. Pair Matching: cada tratado é pareado com apenas um controle ou estratificação, i.e., $|A_t| = |B_t| = 1, \forall t$;
  2. Variable Matching: cada tratado pareia com ao menos 1 e no máximo 4 controles. Formalmente, a razão entre $|A_t|$ e $|B_t|$ varia; e
  3. Full Matching: cada tratado é pareado com 1 ou mais controles e cada controle pode ser pareado com 1 ou mais controles de tal forma que $Min(|A_t|, |B_t|) = 1, \forall t$.
- A escolha entre matchings 1–1 , m–1, 1–n, m–n impacta o viés e a eficiência dos estimadores resultantes (veremos)

Pareamento Global: Optimal Matching

- Seja Δ a distância total dada por

$$\Delta = \sum_{i=1}^{S} w(A_i, B_i) d(A_i, B_i),$$

- em que $w(A_i, B_i)$ é uma função de ponderação. RR 2002 propõe (i) $w(A_i, B_i) = |A_i|/\alpha$, (ii) $w(A_i, B_i) = |B_i|/\beta$, ou (iii) $w(A_i, B_i) = (|A_i| + |B_i|/(\alpha + \beta)$.
- O matching ótimo consiste em encontrar conjuntos $(A_1, \ldots, A_s, B_1, \ldots, B_s)$ de tamanho (α, β) tais que a distância amostral total dos PS é minimizada.
- Exemplo: Lousa
- Como encontrar S? Abordagem de *network flows*.

Pareamento Global: Optimal Matching

- **Análise dos Resultados**: ATE será a média ponderada da diferença de médias entre tratados e controle em todos os conjuntos pareados

$$ATE = \hat{\tau} = \sum_{i=1}^{S} \frac{|A_i| + |B_i|}{N} [\bar{y}_i(1) - \bar{y}_i(0)],$$

  em que N é o total de participantes da amostra $(\alpha + \beta)$, $|A_i|$ e $|B_i|$ são, respectivamente, os números de tratados e de controles do estrato i e $\bar{y}_i(1)$ e $\bar{y}_i(0)$ são as médias da variável de resposta para os tratados e os controles do estrato i.
- Inferência usando *bootstrap* (Haviland 2007 e Lehman 2008).
- É possível fazer ajuste de covariadas usando regressão (veremos mais sobre isso no PS estratificado)

Pareamento Global: Optimal Matching

- Escolha entre 1–1 , m–1, 1–n, m–n, de acordo com os resultados de simulações
  - Redução de viés ao descartar alguns controles (pouca perda de eficiência ao retirá-los)
  - Redução de viés ao permitir diferentes números de controles nos grupos e variar o número de controles pareados entre os conjuntos dos estratos não prejudica eficiência (Haviland et al 2007)
  - ter grande número de controles gera ganhos inexpressivos de eficiência
- Decisão depende da estrutura dos dados: Se tratados e controles são 50% da amostra, usa-se o matching 1–1. Se há mais controles que tratamento, consideram-se outras estruturas.
- Checar balanceamento pós-matching (ver Haviland et al 2007 para testes no R)
- Matching local gera bom balanceamento e é localmente suficiente (Gu e Rosembaum 93)

Propensity Score Estratificado ou Subclassification

- PS Matching pode comprometer o tamanho da amostra.
- Vantagem do PS Subclassification: retém boa parte da amostra.
- Ideia: Subpopulações com mesmo $b(X)$ tem mesma distribuição de covariadas X entre tratados e não-tratados.
- Estratificação pode ser feita em X, mas conforme $k \rightarrow \infty$, número de estratos/blocos também cresce
- Procedimento:
  1. estratifica-se em J estratos de $e(X)$, dentro de cada j, $e(X)$ é constante;
  2. estima-se $\hat{\tau}_j$ usando nos J estratos.
  3. calcula-se $\hat{\tau}$ (média ponderada dos $\hat{\tau}_j$)
- Desafios: (1) a escolha do número de estratos e (2) a escolha das fronteiras dos estratos.

Propensity Score Estratificado ou Subclassification

- Sejam J estratos/blocos com intervalos $(b_{j-1}; b_j]$
- $b_0 = 0$ e $b_J = 1$
- $\cup_{j=1}^{J} [b_{j-1}; b_j] = (0, 1]$.
- Seja B uma função indicadora tal que

$$B_{ij} = \begin{cases} 1, & \text{se } b_{j-1} < e(X) \leq b_j \\ 0, & \text{c.c.} \end{cases}.$$

- $N_{wj} = \sum_{i: w_i = w} B_{ij}$ para $w = 0, 1$ e $j \in \{1, \ldots, J\}$ é o número de unidades de cada tipo de tratamento em cada estrato
- $N_j = N_{cj} + N_{tj}$.
- Sejam ainda $p_j = \frac{N_{1j}}{N_j}$ a proporção de unidades tratadas no estrato j e $q_j = \frac{N_j}{N}$ a proporção de unidades no estrato j.

Propensity Score Estratificado ou Subclassification

Encontrar fronteiras (processo iterativo):

1. Inicia com J = 1 ($b_0 = 0$ e $b_1 = 1$). Sob $H_0 : e_{1j}(X) = e_{0j}(X)$ ou $H_0 : l_{1j}(X) = l_{0j}(X)$ (Log das odds ratios), calcula-se a estatística[1]

$$t = \frac{\bar{l}_{1j} - \bar{l}_{0j}}{\sqrt{S_{lj}^2 \left(\frac{1}{N_{0j}} + \frac{1}{N_{1j}} \right)}}$$

2. Definem-se 3 parâmetros: (i) Valor crítico para t (t = 1,96); (ii) Número mínimo de unidades de tratamento ou controle em um estrato: $\{N_{1j}, N_{0j}\} \geq N_{min}^{*}, \forall j$; e (iii) Número mínimo de unid. no novo estrato $N_j \geq N_{min}^{**}$.
3. Aumenta J (ex. quantis) até H_0 não ser rejeitado
4. Aplicar redução de viés caso ainda haja desbalanceamento

[1] $\bar{l}_{wj} = \frac{1}{N_{wj}} \sum_{i:W_i=w} B_{ij} \hat{l}(X_i)$ e $S_{lj}^2 = \frac{1}{N_{1j}+N_{0j}-2} \left[\sum_{i:W_i=0} B_{ij} \left(\hat{l}(X_i) - \bar{l}_{0j} \right)^2 + \sum_{i:W_i=1} B_{ij} \left(\hat{l}(X_i) - \bar{l}_{1j} \right)^2 \right]$

Propensity Score Estratificado ou Subclassification

Análise dos Resultados: Dados J estratos, calculamos $\tilde{\tau}_j$

$$\tilde{\tau}_j = (\bar{y}_{1j} - \bar{y}_{0j}) = \frac{1}{N_{1j}} \sum_{i:w_i=1} B_{ij} y_i - \frac{1}{N_{0j}} \sum_{i:w_i=0} B_{ij} y_i$$

$$\tilde{\tau}_{PSE} = \sum_{j=1}^{J} q_j \hat{\tau}_j$$

Como o PS é aproximadamente constante dentro dos estratos, podemos reduzir o viés remanescente

PS Estratificado: Redução de Viés

Supondo linearidade:

- $E[y_i(w)|X_i = x] = \alpha w + x\beta = \alpha_0 + \tau w + x\beta$.
- Assim, $\tau = \alpha_1 - \alpha_0$.
- O efeito médio da amostra toda (s/ estratificação) é $\hat{\tau}^* = \bar{y}_1 - \bar{y}_0 = \frac{1}{N_1} \sum_{i:w_i=1} y_i - \frac{1}{N_0} \sum_{i:w_i=0} y_i$.
- Portanto, o viés tem dois componentes:

$$
\begin{aligned}
E[\hat{\tau}^*|X, W] - \tau = {} & \frac{N_0}{N} \underbrace{\{E[y_i(1)|w_i = 1] - E[y_i(1)|w_i = 0]\}}_{\text{tratado se não fosse tratado}} \\
& - \frac{N_1}{N} \underbrace{\{E[y_i(0)|w_i = 1] - E[y_i(0)|w_i = 0]\}}_{\text{controle se fosse tratado}} . \qquad (4)
\end{aligned}
$$

- Tentar eliminar viéses remanescentes (atribuídos às diferenças em X_k's).

PS Estratificado: Redução de Viés

- Sejam $\bar{X}_0 = \frac{\sum_{i:w_i=0} X_i}{N_0}$, $\bar{X}_1 = \frac{\sum_{i:w_i=1} X_i}{N_1}$, $\bar{X} = \frac{\sum X_i}{N} = \frac{N_0}{N}\bar{X}_0 + \frac{N_1}{N}\bar{X}_1$.
- Temos que

$$E[\hat{\tau^*}|X,W] - \tau = \frac{N_0}{N}(\bar{X}_1 - \bar{X}_0)\beta - \frac{N_1}{N}(\bar{X}_0 - \bar{X}_1)\beta = (\bar{X}_1 - \bar{X}_0)\beta.$$

- A mesma conta por estratos fica:
- Sejam as médias por estratos j: $\bar{X}_{0j} = \frac{\sum_{i:w_i=0} B_{ij}X_i}{N_{0j}}$, $\bar{X}_{1j} = \frac{\sum_{i:w_i=1} B_{ij}X_i}{N_{1j}}$, $\bar{X}_j = \frac{\sum B_{ij}X_i}{N_j}$.

$$E[\hat{\tau_j^*}|X,W] - \tau_j = (\bar{X}_{1j} - \bar{X}_{0j})\beta$$

e

$$E[\hat{\tau^*}|X,W] - \tau = \left(\sum q_j(\bar{X}_{1j} - \bar{X}_{0j})\right)\beta.$$

- Outra opção: regressão dentro dos estratos, quando houver unidades suficientes dentro dos estratos.

PS Estratificado

Inferência (Imbens e Rubin, 2014):

- Sucesso no balanceamento: e(X) constante dentro de j $\forall j$
- Normalidade de $\hat{\tau}$
- *Cutoffs* de B_{ij} são fixos
- Assume *overlap* de $\hat{e}(X)$ entre tratamento e controle: não precisa de *trimming*.