

Métodos de escore de propensão para estimação de efeitos causais em estudos observacionais

Uma breve introdução

Ricardo de Souza Kuchenbecker
rkuchenbecker@hcpa.edu.br

Rodrigo Citton P. dos Reis
citton.padilha@ufrgs.br

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
PROGRAMA DE PÓS GRADUAÇÃO EM EPIDEMIOLOGIA

Porto Alegre, 2020

Introdução

Introdução

- ▶ Ensaios controlados aleatorizados (RCTs) são considerados o **padrão-ouro** para inferir efeitos causais.
 - ▶ Alocação aleatória ao tratamento \Rightarrow ausência de confundimento.
 - ▶ O efeito de tratamento no desfecho pode ser estimado por diretamente comparar desfechos entre indivíduos tratados e não-tratados.
- ▶ Em estudos observacionais, a alocação ao tratamento é geralmente influenciada por características individuais.
 - ▶ Análise ajustada por modelos de regressão para levar em consideração diferenças medidas nas características basais.
 - ▶ Vem crescendo o interesse em métodos baseados no **escore de propensão** para reduzir ou eliminar os efeitos de confundimento quando dados observacionais são utilizados.

RCTs vs Estudos Observacionais

Desfechos potenciais e efeitos causais

- ▶ $Y_i(0)$ e $Y_i(1)$ representam os **desfechos potenciais**.
 - ▶ $Y_i(0)$ é o desfecho que teria sido observado caso o indivíduo i fosse alocado ao grupo controle.
 - ▶ $Y_i(1)$ é o desfecho que teria sido observado caso o indivíduo i fosse alocado ao grupo tratado.
- ▶ Z_i é o tratamento atribuído: $Z_i = 0$, se controle; $Z_i = 1$, se tratado.
- ▶ $Y_i = Z_i \times Y_i(1) + (1 - Z_i) \times Y_i(0)$.
- ▶ O par de desfechos potenciais ($Y_i(0)$, $Y_i(1)$) será descrito como o **efeito causal** (individual).
 - ▶ Chamaremos $Y_i(1) - Y_i(0)$ de o **efeito causal na escala da diferença** (no texto, **efeito de tratamento**).
 - ▶ O **efeito médio de tratamento** $ATE = E[Y_i(1) - Y_i(0)]$.
 - ▶ $ATT = E[Y(1) - Y(0)|Z = 1]$.

Perguntas

- ▶ O que representa o ATT ?
- ▶ ATE vs ATT .

RCTs

- ▶ Em RCTs, $Y(0), Y(1) \perp\!\!\!\perp Z$, e assim

$$\begin{aligned}ATE &= E[Y_i(1) - Y_i(0)] \\&= E[Y_i(1)] - E[Y_i(0)] \\&= E[Y_i(1)|Z_i = 1] - E[Y_i(0)|Z_i = 0] \\&= E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0].\end{aligned}$$

- ▶ No caso em que o desfecho é dicotômico, temos

$$ATE = \Pr(Y_i = 1|Z_i = 1) - \Pr(Y_i = 1|Z_i = 0).$$

- ▶ É possível avaliar o efeito em outras escalas.

Estudos observacionais

Cochran (1965)¹

Uma investigação empírica em que “o objetivo é elucidar **relações de causa e efeito** ... [em que] não é viável o uso de experimentação controlada, no sentido de ser capaz de impor procedimentos ou tratamentos cujo o efeito é desejável descobrir, ou designar indivíduos aleatoriamente para diferentes procedimentos”.

Pergunta

- ▶ O que difere entre RCTs e estudos observacionais?

¹Cochran, W.G. The planning of observational studies of human populations (with discussion). *Journal of the Royal Statistical Society Series A*, 128:134–155, 1965.

Estudos observacionais

Cochran (1965)¹

Uma investigação empírica em que “o objetivo é elucidar **relações de causa e efeito** ... [em que] não é viável o uso de experimentação controlada, no sentido de ser capaz de impor procedimentos ou tratamentos cujo o efeito é desejável descobrir, ou designar indivíduos aleatoriamente para diferentes procedimentos”.

Pergunta

- ▶ O que difere entre RCTs e estudos observacionais?
- ▶ $Y(0), Y(1) \not\perp Z$, e assim

$$ATE = E[Y_i(1) - Y_i(0)] \neq E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0].$$

¹Cochran, W.G. The planning of observational studies of human populations (with discussion). *Journal of the Royal Statistical Society Series A*, 128:134–155, 1965.

Escore de Propensão e Métodos baseados no EP

Score de Propensão

- ▶ Dado um conjunto de covariáveis, quão propenso é o indivíduo de ser (alocado ao grupo) tratado/exposto?
 - ▶ **Score** de propensão²:

$$e_i = \Pr(Z_i = 1|X_i).$$

- ▶ Pode ser estimado através de **regressão logística**, ou por outros modelos de predição e métodos de **aprendizado de máquina**.

Score de Propensão

Ignorabilidade

▶ Ignorabilidade da atribuição ao tratamento:

- (a) $Y(0), Y(1) \perp\!\!\!\perp Z|X$;
- (b) $0 < \Pr(Z_i = 1|X_i) < 1$.

Resultado chave

- ▶ O **escore** de propensão é um escore de balanceamento.
- ▶ Se a atribuição ao tratamento é fortemente ignorável dado X , então esta é fortemente ignorável dado qualquer escore de balanceamento.

Escore de Propensão

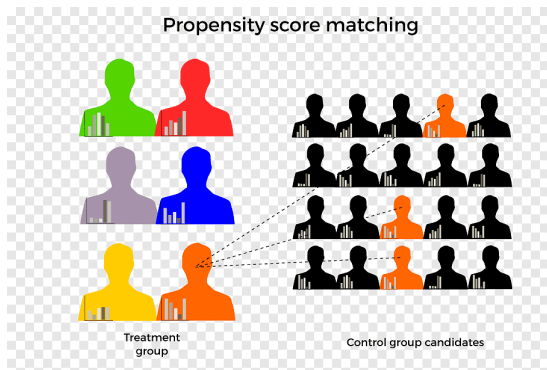
(Ajuste direto)

$$\begin{aligned}
 ATE &= E[Y_i(1) - Y_i(0)] \\
 &= E[Y_i(1)] - E[Y_i(0)] \\
 &= E[E[Y_i(1)|X_i]] - E[E[Y_i(0)|X_i]] \\
 &= E[E[Y_i(1)|Z_i = 1, X_i]] - E[E[Y_i(0)|Z_i = 0, X_i]] \\
 &= E[E[Y_i|Z_i = 1, X_i]] - E[E[Y_i|Z_i = 0, X_i]] \\
 &= E[E[Y_i|Z_i = 1, X_i] - E[Y_i|Z_i = 0, X_i]] \\
 &= \sum_x (E[Y_i|Z_i = 1, X_i = x] - E[Y_i|Z_i = 0, X_i = x]) \Pr(X_i = x).
 \end{aligned}$$

²Rosenbaum, P.R., Rubin, D.B. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70:41–55, 1983.

Pareamento

- Pareamento com base no escore de propensão envolve a formação de conjuntos de indivíduos tratados e não tratados que compartilham um valor semelhante do escore de propensão.



Pareamento

- ▶ **greedy** vs **optimal** *matching*.
- ▶ **nearest neighbor matching** vs **nearest neighbor matching within a specified caliper distance**.
- ▶ ▶ 1:1 vs M:1 *matching*.

Estratificação

- ▶ A estratificação no escore de propensão envolve a estratificação de indivíduos em subconjuntos mutuamente exclusivos, com base no escore de propensão estimado.
- ▶ Uma abordagem comum é dividir os indivíduos em cinco grupos de tamanho igual usando **os quintis** do escore de propensão.
- ▶ Quando o escore de propensão for especificado corretamente, a distribuição das covariáveis de linha de base medidas **será aproximadamente semelhante** entre **tratados** e **não tratados** no mesmo estrato.

IPTW

- ▶ A **ponderação pelo inverso da probabilidade de tratamento** (IPTW; $1/e_i$) usando o escore de propensão usa pesos com base no escore de propensão para criar uma **amostra sintética** na qual a distribuição das covariáveis da linha de base medidas é independente da atribuição do tratamento.
- ▶ As ponderações podem ser definidas como $w_i = \frac{Z_i}{e_i} + \frac{(1-Z_i)}{1-e_i}$.
- ▶ $ATE = E[w_i Y_i | Z_i = 1] - E[w_i Y_i | Z_i = 0]$.

Ajuste de covariável usando Escore de Propensão

► De

$$E[Y|Z = z, X_1, \dots, X_p] = \beta_0 + \beta_Z z + \beta_1 x_1 + \dots + \beta_p x_p.$$

► Para

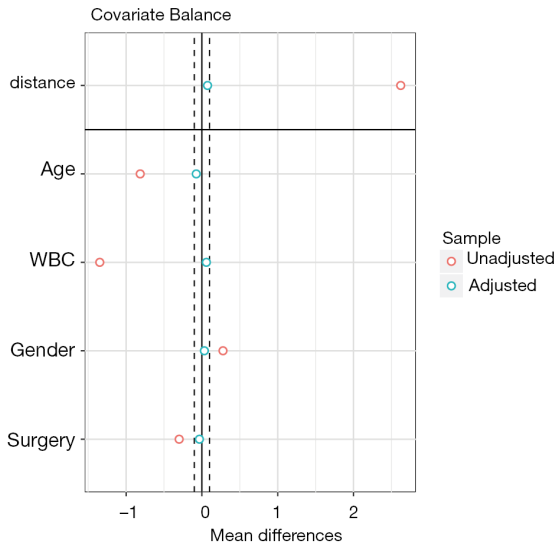
$$E[Y|Z = z, e(X)] = \beta_0 + \beta_Z z + \beta_1 e(X).$$

Comparação dos métodos

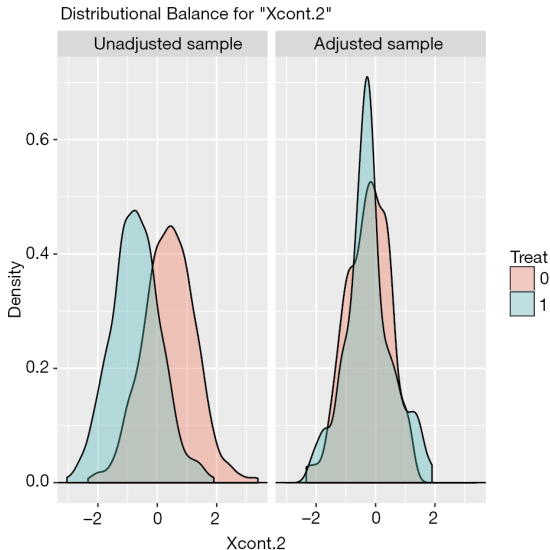
- ▶ Estudos de Monte Carlo compararam o desempenho relativo dos diferentes métodos de escore de propensão para estimar diferenças de risco, riscos relativos e razões de chances marginais e condicionais.
- ▶ Dois desses estudos apresentaram que a estratificação, o pareamento e o ajuste covariável usando o escore de propensão resultaram em estimativas enviesadas.

Diagnóstico do balanceamento

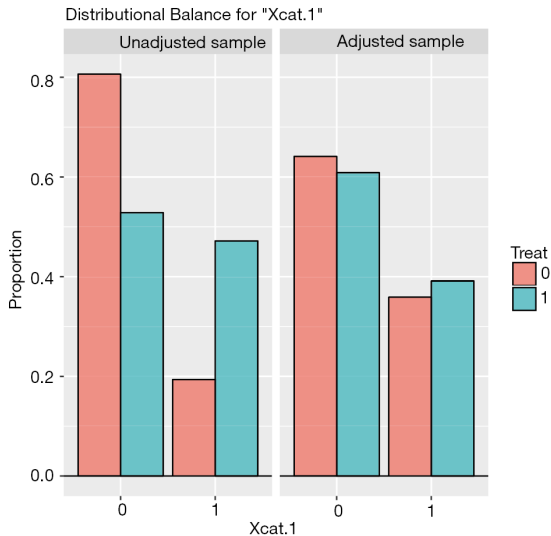
Diagnóstico do balanceamento



Diagnóstico do balanceamento

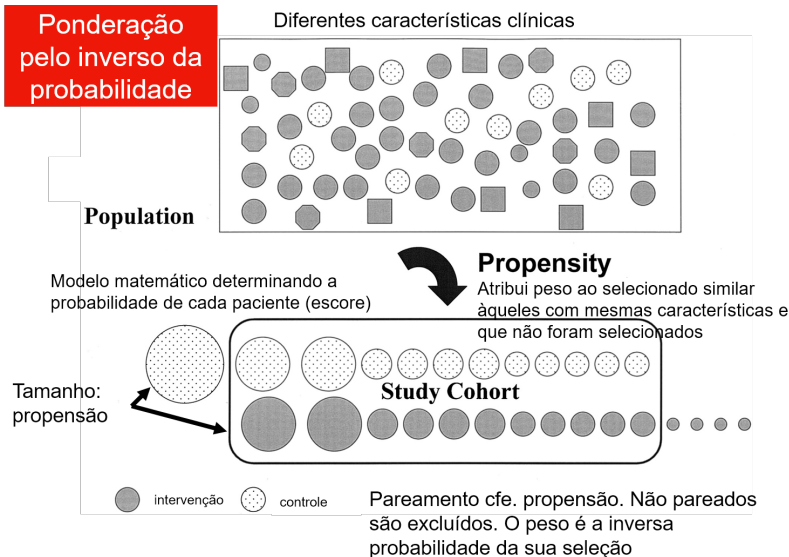


Diagnóstico do balanceamento



Comentários finais

Comentários finais



Comentários finais

- ▶ Seleção da variável para o modelo do Escore de Propensão
- ▶ Métodos baseados no Escore de Propensão vs ajuste de regressão

Por hoje é só!

caffeine causality loop

