

Matching

Manoel Galdino

2024-03-07

Introdução

Na aula de hoje, iremos aprender sobre a principal estratégia de “seleção em não-observável”, que é matching. Mas antes, vamos falar de subclassificação, que é uma técnica mais simples e é útil para introduzir a ideia de matching.

Subclassificação

Subclassificação é um método para para cumprir os requisitos de ajustamento de “backdoor” de um DAG, por meio de estratificação e uso dos pesos de cada estrato. A ideia é que em vez de comparar a diferença na média entre o tratamento e controle em um estudo observacional, iremos comparar a diferença na média por estratos, responderando pelos pesos dos estratos. Este método consegue produzir equilíbrio (balancing) entre tratamento eo controle em termos do controle observável.

O artigo clássico que é a referência no método é de Cochran (1968). O exemplo que ele usa para ilustrar o procedimento é a investigação sobre se cigarro causa câncer de pulmão. À época, ainda era um assunto bastante debatido se cigarro causava câncer. Fisher e Neyman, por exemplo, eram críticos das evidências de que cigarro causaria câncer.

Suposição de Independência Condicional

A suposição de independência condicional (CIA, na sigla em inglês) diz que, condicional a um vetor de (uma ou mais) covariáveis X , resultados potenciais são independentes do mecanismo de alocação de tratamenhto e controle. Formalmente, escrevemos:

$$(Y^1, Y^0) \perp\!\!\!\perp D | X$$

O símbolo $\perp\!\!\!\perp$ é nossa notação para independência estatística. E implica que a esperança dos resultados potenciais de tratamento e controle são iguais no tratamento e controle para cada valor de x . Formalmente, escrevemos:

$$\begin{aligned}\mathbb{E}[Y^1 | D = 1, X] &= \mathbb{E}[Y^1 | D = 0, X] \\ \mathbb{E}[Y^0 | D = 1, X] &= \mathbb{E}[Y^0 | D = 0, X]\end{aligned}$$

Portanto, se CIA for crível, podemos usar subclassificação. E, quando a alocação do tratamento é condicional a variáveis observáveis, é uma estratégia de identificação baseada em observáveis.

Cigarro e Câncer de Pulmão

Vamos ver um exemplo para entender como funciona a subclassificação.

Table 1: Death rates per 1,000 person-years (placeholder values)

Group	Canada	UK	US
Non-smokers	20.2	NA	NA
Cigar/pipe smokers	35.5	NA	NA
Cigarette smokers	NA	NA	NA

A tabela 1 mostra que a taxa de mortalidade é maior para fumantes de cachimbos e charutos do que cigarros. Isso não quer dizer que o efeito causal de cachimbos e charutos é maior do que cigarros, até porque o mecanismo sugerido para câncer de pulmão era a inalação da fumaça, que depositava alcatrão nos pulmões. Para que isso seja resultado do efeito causal, precisamos acreditar que:

$$\begin{aligned}\mathbb{E}[Y^1|D = \text{"cigarro"}] &= \mathbb{E}[Y^1|D = \text{"cigarro"}] \\ \mathbb{E}[Y^0|D = \text{"cigarro"}] &= \mathbb{E}[Y^0|D = \text{"cigarro"}]\end{aligned}$$

Cochran, W. G. 1968. “The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies.” *Biometrics* 24 (2): 295–313.