Causalidade e Resultados Potenciais

Davi Ferreira Veronese (monitor)

Universidade de São Paulo (USP)
Métodos Quantitativos II
Causalidade

Causalidade: noções introdutórias

- Influência de uma variável (T) sobre a outra (Y)
- Existência de uma ação (tratamento ou intervenção) aplicada sobre uma unidade (i)
- Cada unidade i pode ser exposta a um conjunto de ações
- Cada par ação-unidade está associado a um determinado resultado (realização de Y)

Exemplo

- Suponha que desejemos estimar o efeito do Programa de Aquisição de Alimentos (PAA) sobre determinado índice de segurança alimentar
- Unidade de análise (i): município
- Determinados municípios são expostos à política ($T_i = 1$), enquanto outros municípios não são ($T_i = 0$)

Resultados Potenciais

- Assuma que, em uma população de tamanho N, a variável de tratamento T assume 0 para o grupo de controle e 1 para o grupo de tratamento
- Para cada indivíduo i, a variável de interesse (dependente, explicada, resultado etc) assume $Y_i(0)$ ou $Y_i(1)$

$$Y_i^{\text{obs}} = Y_i(T_i) = \begin{cases} Y_i(0) & \text{if } T_i = 0, \\ Y_i(1) & \text{if } T_i = 1. \end{cases} \tag{1}$$

$$Y_i^{\text{mis}} = Y_i(1 - T_i) = \begin{cases} Y_i(0) & \text{if } T_i = 1, \\ Y_i(1) & \text{if } T_i = 0. \end{cases} \tag{2}$$

$$Y_i^{\text{obs}} = Y_i(1).T + Y_i(0).(1 - T) \tag{3}$$

Efeito Causal e Problema Fundamental da Inferência Causal

- O efeito causal é a comparação (a diferença) entre os resultados potenciais, **para a mesma unidade**, no mesmo momento pós-tratamento
- Efeito causal (treatment effect) = $Y_i(1) - Y_i(0)$
- O efeito causal **não é** uma comparação de um mesmo indivíduo (unidade) antes e depois de receber o tratamento
- ”Problema fundamental da inferência causal”: apenas um resultado potencial se realiza (problema de missing data)
- Necessidade de múltiplas unidades para a inferência causal

Quantidades de Interesse (Causal Estimands)

- Average treatment effect (ATE):

$$\mathbb{E}[Y_i(1) - Y_i(0)] \tag{4}$$

- Average treatment effect on the treated (ATT):

$$\mathbb{E}[Y_i(1) - Y_i(0) | T_i = 1] \tag{5}$$

Decomposição do Efeito Causal

$$
\begin{aligned}
&\mathbb{E}[Y_i^{\text{obs}}|T = 1] - \mathbb{E}[Y_i^{\text{obs}}|T = 0] = \\
&\mathbb{E}[Y_i(1).T + Y_i(0).(1 - T)|T = 1] - \mathbb{E}[Y_i(1).T + Y_i(0).(1 - T)|T = 0] = \\
&\mathbb{E}[Y_i(1)|T = 1] - \mathbb{E}[Y_i(0)|T = 0] = \\
&\mathbb{E}[Y_i(1)|T = 1] {\color{red}- \mathbb{E}[Y_i(0)|T = 1] + \mathbb{E}[Y_i(0)|T = 1]} - \mathbb{E}[Y_i(0)|T = 0] = \\
&{\color{orange}\mathbb{E}[Y_i(1) - Y_i(0)|T = 1]} + {\color{blue}\mathbb{E}[Y_i(0)|T = 1] - \mathbb{E}[Y_i(0)|T = 0]} = \\
&{\color{orange}ATT} + {\color{blue}\text{selection bias}}
\end{aligned}
\tag{6}
$$

Regressão e Causalidade

$$Y_i^{\text{obs}} = \alpha + \beta . T_i + \epsilon_i \tag{7}$$

- Se o mecanismo de atribuição ao tratamento (T_i) é independente da variável dependente (Y_i), os grupos de tratamento e controle tendem a ser, na média, iguais, exceto pela exposição ao tratamento
- Nesse caso, $\hat{\beta}$ é um estimador não viesado do efeito causal β (i. e., $\mathbb{E}[\hat{\beta}] = \beta$)

$$\hat{\beta}_{ols} = \frac{1}{\sum_{i=1}^{n}(T_i - \bar{T})^2} \sum_{i=1}^{n}(Y_i^{\text{obs}} - \bar{Y}^{\text{obs}})(T_i - \bar{T}) = \bar{Y}_t^{\text{obs}} - \bar{Y}_c^{\text{obs}} \tag{8}$$

- Por causa da randomização, podemos afirmar (ver Athey & Imbens):

$$\beta = \mathbb{E}[Y_i(1) - Y_i(0)] \tag{9}$$

$$\alpha = \mathbb{E}[Y_i(0)] \tag{10}$$

$$\mathbb{E}[\epsilon_i | T_i] = 0 \tag{11}$$

- Intuição: como o mecanismo de seleção (atribuição) ao tratamento é aleatório, não há diferenças sistemáticas entre tratados e controles (não há variáveis omitidas determinando a atribuição ao tratamento), de modo que a diferença pós-tratamento entre tratados e controles deve-se apenas ao tratamento (e não ao viés de seleção)
- Em outras palavras, o viés de seleção é igual a zero

Contextos mais gerais

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \epsilon_i \tag{12}$$

- Suponha que estamos interessados no efeito (causal) de X_1 sobre Y, e as demais variáveis ($X_2 \ldots X_k$) são controles
- A ausência de viés de seleção ocorre quando a **hipótese de média condicional zero** é satisfeita

$$\mathbb{E}[\epsilon | X_1, X_2 \ldots X_k] = 0 \tag{13}$$

- Se essa hipótese não for satisfeita, não podemos conferir interpretação causal aos estimadores da regressão

Referências

- Imbens, G. & Rubin, D. **Causal Inference for Statistics, Social and Biomedical: an Introduction**. Cambridge University Press, 2015.
- Athey, S. & Imbens, G. The Econometrics of Randomized Experiments.

Obrigado!