

Lista 02 - Causalidade e resultados potenciais

Artur Damião

19 de setembro de 2025

Número USP: 10701251. E-mail: arturcardoso@usp.br

1. Escreva, em poucas palavras, o que é considerado o problema fundamental da inferência causal. Procure articular sua resposta com o modelo de resultados potenciais.

Resposta: O **problema fundamental da inferência causal** é que somente um dos resultados potenciais pode ser observado. Isso é, necessariamente, um problema de dados faltantes (dados latentes, ou seja, aquilo que é oculto). Como é impossível a aplicação do controle e do tratamento numa mesma unidade, nós sempre estimaremos o efeito agregado (Aronow e Miller 2019, 236).

2. Considere um estudo observacional que investiga os efeitos de uma determinada intervenção/política (T) sobre um grupo de municípios, comparando-os com outros municípios que não foram contemplados com essa política neste primeiro momento de implementação (C). Explique por que a simples comparação da média dos resultados entre os dois grupos pode não refletir a verdadeira relação causal entre tratamento e resultados. Quais são os principais problemas envolvidos nesse raciocínio?

Resposta: Um dos principais problemas envolvidos neste raciocínio é a presença de viés de seleção, ou seja, os municípios que recebem T podem ser muito diferentes entre si. Além disso, fatores não observados podem enviesar o efeito por meio de variáveis omitidas.

3.

- a) Explique, com suas próprias palavras, o que é viés de seleção.

Resposta: O viés de seleção acontece quando o grupo tratamento e o grupo controle não são comparáveis entre si desde o começo da análise. As diferenças entre os grupos afetam o resultado, superestimando ou subestimando o verdadeiro valor.

- b) Dê um exemplo de uma situação na qual o viés de seleção poderia ocorrer em um estudo que tenta estabelecer relações causais, indicando as possíveis variáveis omitidas. Procure especificar qual é o tratamento e o que corresponderia ao controle (1 se tratado, 0 caso contrário), além de escolher um indicador (Y) como variável dependente/resposta

Resposta: Assumindo que o tratamento (T) seja a implementação de um programa de policiamento comunitário em certos municípios (1 = programa implementado, 0 = não implementado). O resultado (Y) é a taxa de criminalidade no ano seguinte. Se os municípios escolhidos para receber o programa forem justamente os que tinham taxas de criminalidade mais altas no ano anterior, há viés de seleção: a variável omitida é a taxa de criminalidade prévia, que afeta tanto a chance de receber o programa (seleção) quanto o resultado futuro.

4.

a) Escreva, em notação de resultados potenciais, a equação do viés de seleção.

Resposta:

Conforme o capítulo 8.1.1 do livro do professor, a equação seria:

$$\mathbb{E}[Y_i(1) - Y_i(0) \mid B_i = 1] + \underbrace{(\mathbb{E}[Y_i(0) \mid B_i = 1] - \mathbb{E}[Y_i(0) \mid B_i = 0])}_{\text{Viés de seleção}}$$

b) Com base no exemplo descrito na questão anterior, explique em palavras o que a equação está descrevendo neste caso particular do seu desenho de pesquisa¹.

Resposta:

$$\mathbb{E}[Y_i(1) - Y_i(0) \mid B_i = 1] + \underbrace{(\mathbb{E}[Y_i(0) \mid B_i = 1] - \mathbb{E}[Y_i(0) \mid B_i = 0])}_{\text{Viés de seleção}}$$

A equação mostra que a diferença de médias é a soma do efeito causal do programa e da diferença pré-existente entre os grupos.

5. Explique por que um desenho de pesquisa experimental tende a eliminar o viés de seleção.

Resposta: Desenhos experimentais permitem controle sobre o que é tratamento, o que é controle, o método de seleção randômico e a partir disso estimar corretamente o interesse. Desenhos observacionais, ao contrário, não permitem controle sobre os tratamentos ou controle. O experimento é o padrão-ouro da inferência causal porque isola o efeito do tratamento, ou seja, garante que os indivíduos selecionados para o tratamento sejam selecionados por sorte (aleatoriedade), além de garantir que não existam diferenças sistemáticas entre o grupo tratamento e o grupo controle. Esse conjunto de características elimina o viés de seleção.

6. a) Explique, em poucas palavras, o que são estimandos causais.

Resposta: Um estimando causal é uma quantidade teórica de interesse, ou seja, o objetivo da nossa análise. São medidas que quantificam o efeito de um tratamento sobre um resultado, indicando o que muda no desfecho se aplicarmos uma determinada intervenção ao invés de não aplicá-la.

b) Uma ONG lançou um programa de mentorias para jovens de baixa renda, com o objetivo de aumentar a probabilidade de ingresso na universidade. Apenas parte dos jovens interessados participou do programa, e você foi contratado para avaliar o impacto da iniciativa. Para esta avaliação, qual estimando causal é mais apropriado – ATE (Average Treatment Effect) ou ATT (Average Treatment on the Treated)? Justifique sua escolha.

Resposta: O ATT é o estimando causal mais apropriado. Por se tratar de um estudo onde a participação foi voluntária (por pessoas jovens interessadas), estamos medindo o efeito do programa sobre aqueles que efetivamente participaram (ou seja, nos tratados).

7. Como vimos em aula, é possível utilizar a equação de regressão para incluir determinadas variáveis que consideramos que podem estar influenciando tanto sobre nossa variável de interesse, quanto sobre o tratamento. A elas é dado o nome de “variável controle”. Considerando o desenho de pesquisa elaborado para a questão 3 e as variáveis omitidas identificadas, procure explicar o que ocorre quando se inclui uma dessas variáveis na equação. Qual o efeito desse controle para os parâmetros da regressão linear que estamos interessados em medir (intercepto e coeficiente angular)?

¹Para essa pergunta, consultei o ChatGPT para entender o que a equação estaria descrevendo, visto que estou com dificuldade para interpretar a formalização matemática.

Resposta: Variáveis de controle são boas quando, quando adicionadas ao modelo, ajudam a reduzir o viés de variável omitida. “Bad controls”, para usar os termos de Angrist e Pischke (2015), são variáveis que podem aumentar o viés de seleção amostral ou então bloquear o efeito causal. Ou seja, adicionar mais controles ao modelo nem sempre faz com que ele seja mais “robusto”! Considerando a fórmula de regressão múltipla abaixo

$$y_i = \alpha + \beta_1 D_1 + \beta_2 X_2 i + \beta_k X_k i + \epsilon$$

controlar covariáveis é reduzir a variância de D e de β_1 . Se a variância é menor, a precisão do modelo aumenta, porque o intervalo de confiança menor.

Referências

- Angrist, Joshua David, e Jörn-Steffen Pischke. 2015. *Mastering 'metrics: The Path from Cause to Effect*. Princeton; Oxford: Princeton University Press.
- Aronow, Peter M., e Benjamin T. Miller. 2019. *Foundations of Agnostic Statistics*. Cambridge New York, NY Port Melbourne New Delhi Singapore: Cambridge University Press.