

Métodos Quantitativos

Aula 02. Introdução à Causalidade

Pedro H. G. Ferreira de Souza

pedro.ferreira@ipea.gov.br

Mestrado Profissional em Políticas Públicas e Desenvolvimento

Instituto de Pesquisa Econômica Aplicada (Ipea)

19 set. 2022

Recapitulação

Introdução

Notação matemática

Modelo de resultados potenciais

Guia de obstáculos à causalidade

Desenhos de pesquisa

Exemplos para o Brasil

Próxima aula

Bibliografia

Recapitulação

Introdução

Notação matemática

Modelo de resultados potenciais

Guia de obstáculos à causalidade

Desenhos de pesquisa

Exemplos para o Brasil

Próxima aula

Bibliografia

Aula passada

1. Rigor de pesquisa para obter validade interna e externa com objetivos preditivos, descritivos ou explicativos
2. Explicações causais podem ser idiográficas (*causas de um efeito*) ou nomotéticas (*efeitos de uma causa*)
3. Fluxograma típico ideal
4. Lógica da inferência vale para pesquisas qualitativas e quantitativas
5. Análises quantitativas → muitos casos, pouco contexto
6. Inferência estatística → usar relações entre variáveis em uma amostra conhecida para inferir algo sobre população não conhecida
7. Também vimos tipos de variáveis, critérios para mensuração, organização de bancos de dados etc

Recapitulação

Introdução

Notação matemática

Modelo de resultados potenciais

Guia de obstáculos à causalidade

Desenhos de pesquisa

Exemplos para o Brasil

Próxima aula

Bibliografia

Motivação

Suponha que vocês já...

formularam suas perguntas de pesquisa,
desenvolveram um arcabouço teórico,
decidiram as definições conceituais e operacionais das variáveis,
e elaboraram hipóteses falsificáveis.

O próximo passo é escolher o **desenho de pesquisa** mais apropriado

- Como obter validade interna e validade externa?

Essa etapa é particularmente importante para estudos sobre causalidade. **Mas o que significa dizer que X causa Y?**

E se meu objetivo for outro?

Causalidade está no coração da pesquisa científica

- Linha entre objetivos às vezes é tênue
- Mesmo trabalhos com outros objetivos muitas vezes possuem implicações causais
- Conhecer limites do que é possível fazer ou afirmar

Inferência causal é questão de pressupostos + modelos + dados

- Importante ter vocabulário comum para explicitar pressupostos
- Até discordâncias dependem de entendimento comum mínimo

Três definições de causalidade (Goldthorpe, 2001)

Causas de efeitos

1 Causalidade como **dependência robusta**

Problema estatístico: causalidade como associação remanescente após “controlar” por todas as variáveis plausíveis

3 Causalidade como **processo gerador**

Formulação teórica de mecanismo subjacente que produz hipóteses testáveis empiricamente

Efeitos de causas

2 Causalidade como **manipulação intencional**

Paradigma experimental: manipulação intencional de um “tratamento” e comparação com contrafactual

Causalidade como manipulação intencional

Busca por **explicações nomotéticas**

- Explicar efeitos causas de poucas variáveis independentes sobre uma variável dependente, *ceteris paribus*
- Mundo é complexo e probabilístico e fenômenos são multicausais → explicações parciais e “na margem”
- **Efeitos de uma causa**: variação exógena que permita identificação
- Redução da ambição explicativa, mas com **validade interna**

Caveat

Contribuição científica depende de reconciliar o **ideal** com as **limitações práticas**

O mais importante não é o que falta em relação ao ideal, mas o progresso em relação ao que já existe

Nenhum estudo é a última palavra sobre um assunto – seguindo Popper, todas as conclusões são provisórias

Recapitulação

Introdução

Notação matemática

Modelo de resultados potenciais

Guia de obstáculos à causalidade

Desenhos de pesquisa

Exemplos para o Brasil

Próxima aula

Bibliografia

O operador de somatório \sum

Definição

$$\sum_{i=j}^n x_i \equiv x_j + x_{j+1} + x_{j+2} + \dots + x_{n-1} + x_n$$

O operador de somatório \sum

Definição

$$\sum_{i=j}^n x_i \equiv x_j + x_{j+1} + x_{j+2} + \dots + x_{n-1} + x_n$$

Propriedades

1. $\sum_{i=1}^n c = n \cdot c$
2. $\sum_{i=j}^n (cx_i + dy_i) = c \cdot \sum_{i=j}^n x_i + d \cdot \sum_{i=j}^n y_i$
3. (**média**) $\bar{x} = \sum_{i=1}^n \frac{x_i}{n} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} (x_1 + x_2 + x_3 + \dots + x_n)$

Exemplo numérico

i	x	y
1	1.76	91.5
2	1.85	71.4
3	1.77	90.7
4	1.71	75.7
5	1.78	76.1
6	1.70	77.0
7	1.68	68.3
8	1.61	71.2
9	1.68	76.7
10	1.75	87.2

Exemplo numérico

$$\sum_{i=4}^5 2 =$$

$$\sum_{i=8}^{10} i =$$

$$\sum_{i=8}^{10} x_i + y_i =$$

$$\bar{z} = \sum_{i=1}^{10} \frac{z_i}{10} = \sum_{i=1}^{10} \frac{y_i/x_i^2}{10} =$$

Exemplos numéricos

$$\sum_{i=4}^5 2 = 2 + 2 = 4$$

Exemplos numéricos

$$\sum_{i=4}^5 2 = 2 + 2 = 4$$

$$\sum_{i=8}^{10} i = 8 + 9 + 10 = 27$$

Exemplos numéricos

$$\sum_{i=4}^5 2 = 2 + 2 = 4$$

$$\sum_{i=8}^{10} i = 8 + 9 + 10 = 27$$

$$\sum_{i=8}^{10} x_i + y_i = \sum_{i=8}^{10} x_i + \sum_{i=8}^{10} y_i = (x_8 + x_9 + x_{10}) + (y_8 + y_9 + y_{10}) = 240.1$$

Exemplos numéricos

$$\sum_{i=4}^5 2 = 2 + 2 = 4$$

$$\sum_{i=8}^{10} i = 8 + 9 + 10 = 27$$

$$\sum_{i=8}^{10} x_i + y_i = \sum_{i=8}^{10} x_i + \sum_{i=8}^{10} y_i = (x_8 + x_9 + x_{10}) + (y_8 + y_9 + y_{10}) = 240.1$$

$$\bar{z} = \sum_{i=1}^{10} \frac{z_i}{10} = \sum_{i=1}^{10} \frac{y_i/x_i^2}{10} = \frac{1}{10} \left(\frac{y_1}{x_1^2} + \dots + \frac{y_{10}}{x_{10}^2} \right) = 26.3$$

O operador de valor esperado \mathbb{E}

Definição

O valor esperado é a “média ponderada” de todos os valores possíveis de uma variável aleatória.

Para variável aleatória discreta X com n possíveis valores com probabilidade p_i de ocorrer:

$$\mathbb{E}(X) = \sum_{i=1}^n x_i \cdot p_i$$

Para variável aleatória contínua com densidade $f(x)$:

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x)dx$$

Exemplos

Qual o valor esperado de um dado não viciado com 6 lados?

Exemplos

Qual o valor esperado de um dado não viciado com 6 lados?

$$\mathbb{E}(X_6) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5$$

Qual o valor esperado de um dado não viciado com n lados?

Exemplos

Qual o valor esperado de um dado não viciado com 6 lados?

$$\mathbb{E}(X_6) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5$$

Qual o valor esperado de um dado não viciado com n lados?

$$\mathbb{E}(X_n) = \frac{1+n}{2}$$

A roleta do cassino tem 38 números. Suponha que o prêmio está em R\$ 18 para R\$ 1 apostado. Vale a pena?

Exemplos

Qual o valor esperado de um dado não viciado com 6 lados?

$$\mathbb{E}(X_6) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5$$

Qual o valor esperado de um dado não viciado com n lados?

$$\mathbb{E}(X_n) = \frac{1+n}{2}$$

A roleta do cassino tem 38 números. Suponha que o prêmio está em R\$ 18 para R\$ 1 apostado. Vale a pena?

$$\mathbb{E}(X) = -1 \cdot \frac{37}{38} + 18 \cdot \frac{1}{38} = \frac{18-37}{38} = \frac{-19}{38} = -R\$0,50$$

Propriedades de $\mathbb{E}(\cdot)$

Linearidade

$$\mathbb{E}(\alpha X + \beta Y + \gamma Z) = \alpha \mathbb{E}(X) + \beta \mathbb{E}(Y) + \gamma \mathbb{E}(Z)$$

$$\mathbb{E}(\alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_n) = \mathbb{E}\left(\sum_{i=1}^n \alpha_i X_i\right) = \sum_{i=1}^n \alpha_i \mathbb{E}(X_i)$$

Propriedades de $\mathbb{E}(\cdot)$

Linearidade

$$\mathbb{E}(\alpha X + \beta Y + \gamma Z) = \alpha \mathbb{E}(X) + \beta \mathbb{E}(Y) + \gamma \mathbb{E}(Z)$$

$$\mathbb{E}(\alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_n) = \mathbb{E}\left(\sum_{i=1}^n \alpha_i X_i\right) = \sum_{i=1}^n \alpha_i \mathbb{E}(X_i)$$

Esperança condicional para $D = \{0, 1\}$

$$\begin{aligned}\mathbb{E}(X) &= \mathbb{E}(\mathbb{E}(X|D)) \\ &= \mathbb{E}(X|D=0) \cdot Pr(D=0) + \mathbb{E}(X|D=1) \cdot Pr(D=1)\end{aligned}$$

Exemplo

Suponha uma nova dieta da moda em que, nos dias pares, você só pode fazer uma refeição; nos dias ímpares, você tem que jogar um dado para determinar o número de refeições daquele dia.

Quais as esperanças condicionais e qual o valor esperado geral?

Exemplo

Suponha uma nova dieta da moda em que, nos dias pares, você só pode fazer uma refeição; nos dias ímpares, você tem que jogar um dado para determinar o número de refeições daquele dia.

Quais as esperanças condicionais e qual o valor esperado geral?

$$\mathbb{E}(X|D = \textit{par}) = 1$$

$$\mathbb{E}(X|D = \textit{impar}) = 3.5$$

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|D)) = 0.5 \cdot 1 + 0.5 \cdot 3.5 = 2.25$$

Recapitulação

Introdução

Notação matemática

Modelo de resultados potenciais

Guia de obstáculos à causalidade

Desenhos de pesquisa

Exemplos para o Brasil

Próxima aula

Bibliografia

Modelo de Rubin

O efeito causal de D sobre Y para um indivíduo i em um momento t é definido pela comparação entre dois estados do mundo, a diferença em Y com e sem o tratamento D :

$$Y_i^1 - Y_i^0$$

Qual o efeito causal do ensino superior sobre salários?

- No momento t , o efeito causal é a diferença entre o salário obtido por i com diploma de graduação e sem diploma

Problema Fundamental da Inferência Causal

É impossível observar os dois estados do mundo para o mesmo indivíduo → problema essencial de **missing data**

Resultados Potenciais

É impossível observar diretamente o efeito causal de D para cada unidade, mas os **contrafactuais** podem ser **identificados** para o nível populacional com a ajuda de pressupostos

Contrafactuais são resultados potenciais não observados

- Resultados potenciais só existem *ex ante*; contrafactuais precisam ser reconstruídos a partir dos dados

Mas como? O modelo de Rubin mostra o poder de **experimentos aleatorizados** para estimar os parâmetros causais de interesse.

ATE, ATT e ATU para $D = \{0, 1\}$

Switching equation

$$Y_i = D_i Y_i^1 + (1 - D_i) Y_i^0$$

ATE, ATT e ATU para $D = \{0, 1\}$

Switching equation

$$Y_i = D_i Y_i^1 + (1 - D_i) Y_i^0$$

Average Treatment Effect → efeito causal médio

$$ATE = \mathbb{E}(Y_i^1 - Y_i^0) = \mathbb{E}(Y_i^1) - \mathbb{E}(Y_i^0) = \pi ATT + (1 - \pi) ATU$$

ATE, ATT e ATU para $D = \{0, 1\}$

Switching equation

$$Y_i = D_i Y_i^1 + (1 - D_i) Y_i^0$$

Average Treatment Effect → efeito causal médio

$$ATE = \mathbb{E}(Y_i^1 - Y_i^0) = \mathbb{E}(Y_i^1) - \mathbb{E}(Y_i^0) = \pi ATT + (1 - \pi) ATU$$

ATE on the Treated → efeito médio sobre os tratados

$$ATT = \mathbb{E}(Y_i^1 - Y_i^0 | D_i = 1) = \mathbb{E}(Y_i^1 | D_i = 1) - \mathbb{E}(Y_i^0 | D_i = 1)$$

ATE, ATT e ATU para $D = \{0, 1\}$

Switching equation

$$Y_i = D_i Y_i^1 + (1 - D_i) Y_i^0$$

Average Treatment Effect → efeito causal médio

$$ATE = \mathbb{E}(Y_i^1 - Y_i^0) = \mathbb{E}(Y_i^1) - \mathbb{E}(Y_i^0) = \pi ATT + (1 - \pi) ATU$$

ATE on the Treated → efeito médio sobre os tratados

$$ATT = \mathbb{E}(Y_i^1 - Y_i^0 | D_i = 1) = \mathbb{E}(Y_i^1 | D_i = 1) - \mathbb{E}(Y_i^0 | D_i = 1)$$

ATE on the Untreated → efeito médio sobre os não tratados

$$ATU = \mathbb{E}(Y_i^1 - Y_i^0 | D_i = 0) = \mathbb{E}(Y_i^1 | D_i = 0) - \mathbb{E}(Y_i^0 | D_i = 0)$$

Identificação

Definições

Parâmetro: o que queremos estimar

Estimativa: aproximação do parâmetro a partir de amostra finita

Estimador: método ou fórmula para produzir uma estimativa para um parâmetro

Identificação

Aplicação de modelo sobre o processo gerador e forma de coleta e mensuração dos dados aos dados observáveis resulta em valores únicos para o parâmetro

- Quais pressupostos precisamos para estimar o parâmetro de forma única?

Cunningham 2021, tabela 11

Table 11. Potential outcomes for ten patients receiving surgery Y^1 or chemo Y^0 .

Patient	Y^1	Y^0	δ
1	7	1	6
2	5	6	-1
3	5	1	4
4	7	8	-1
5	4	2	2
6	10	1	9
7	1	10	-9
8	5	6	-1
9	3	7	-4
10	9	8	1

$$ATE = \mathbb{E}(Y_i^1 - Y_i^0) = \mathbb{E}(\delta_i) = 0.6$$

Cunningham 2021, tabela 12

Para um médico perfeito e onisciente, temos:

Table 12. Post-treatment observed life spans in years for surgery $D = 1$ versus chemotherapy $D = 0$.

Patients	Y	D
1	7	1
2	6	0
3	5	1
4	8	0
5	4	1
6	10	1
7	10	0
8	6	0
9	7	0
10	9	1

O que acontece se fizermos uma simples diferença de médias?

Diferença de médias (SDO)

Estimador populacional

$$SDO = \mathbb{E}(Y^1|D = 1) - \mathbb{E}(Y^0|D = 0)$$

Estimação a partir da amostra

$$SDO = \frac{1}{5} \sum_{i=1}^5 (y_i | d_i = 1) - \frac{1}{5} \sum_{i=1}^5 (y_i | d_i = 0) = 7 - 7.4 = -0.4$$

...então o tratamento é prejudicial?

Como reconciliar isso com o ATE = 0.6?

Reconciliando ATE e SDO

$$\underbrace{\frac{1}{N_T} \sum_{i=1}^n (y_i | d_i = 1) - \frac{1}{N_C} \sum_{i=1}^n (y_i | d_i = 0)}_{\text{Simple Difference in Outcomes}} = \underbrace{E[Y^1] - E[Y^0]}_{\text{Average Treatment Effect}}$$

$$+ \underbrace{E[Y^0 | D = 1] - E[Y^0 | D = 0]}_{\text{Selection bias}}$$

$$+ \underbrace{(1 - \pi)(ATT - ATU)}_{\text{Heterogeneous treatment effect bias}}$$

Diferença de médias é sempre o ATE (que queremos descobrir) + vieses (que não temos como observar diretamente)

Para fixar

Viés de seleção $\rightarrow \mathbb{E}(Y^0|D=1) - \mathbb{E}(Y^0|D=0)$

Viés causado pelas diferenças entre os grupos caso não houvesse tratamento. Para o grupo de controle, observamos Y^0 ; para o grupo de tratamento, Y^0 é contrafactual.

Viés de efeitos heterogêneos do tratamento $\rightarrow (1 - \pi)(ATT - ATU)$

Viés causado pelas diferenças no efeito do tratamento entre grupos caso ambos fossem tratados. Só observamos diretamente π .

*Quais **pressupostos** precisamos para que a diferença entre médias identifique o **ATE**?*

Pressupostos para SDO = ATE

Para não existir **viés de seleção**:

$$\mathbb{E}(Y^0|D = 1) = \mathbb{E}(Y^0|D = 0)$$

Para não existir **viés de efeitos heterogêneos**:

$$ATT = ATU$$

$$\mathbb{E}(Y^1|D = 1) - \mathbb{E}(Y^0|D = 1) = \mathbb{E}(Y^1|D = 0) - \mathbb{E}(Y^0|D = 0)$$

Para evitar **ambos** os vieses:

$$\mathbb{E}(Y^0|D = 1) = \mathbb{E}(Y^0|D = 0)$$

$$\mathbb{E}(Y^1|D = 1) = \mathbb{E}(Y^1|D = 0)$$

Pressupostos para SDO = ATE

Independência

$$(Y^1, Y^0) \perp D$$

Alocação do tratamento sem relação com resultados potenciais

- **Problema:** Humanos escolhem as melhores opções para si com base no futuro antecipado, violando a independência

Como garantir esse pressuposto? **Alocação aleatória** do tratamento é o padrão-ouro, pois elimina as duas fontes de vies.

- Apelo de desenhos de pesquisa experimentais está justamente em estimar os parâmetros de interesse sem vies

Pressupostos

Stable Unit Treatment Value Assumption (SUTVA)

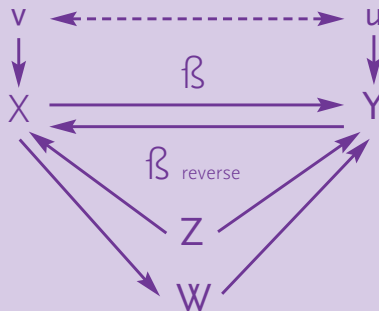
Resultados potenciais de uma unidade não podem ser afetados pela alocação ou não de tratamento às demais unidades.

Conjunto de pressupostos:

1. Doses homogêneas de tratamento para os tratados
2. Sem externalidades ou *spillovers*
3. Sem efeitos de equilíbrio geral

Problemas em desenhos não experimentais

Figure 2:
The Real World



Fonte: Dowd and Town, 2002, p. 4

Recapitulação

Introdução

Notação matemática

Modelo de resultados potenciais

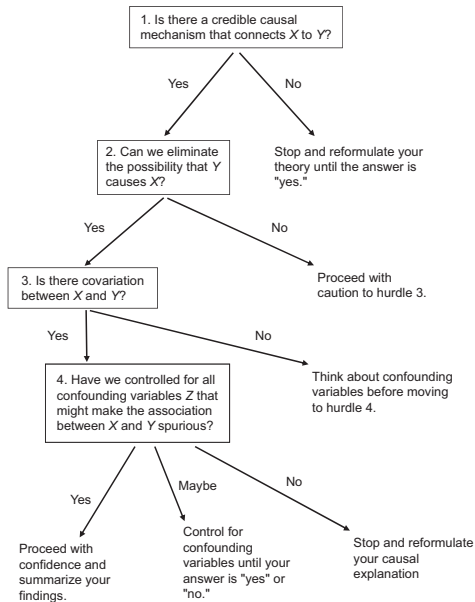
Guia de obstáculos à causalidade

Desenhos de pesquisa

Exemplos para o Brasil

Próxima aula

Bibliografia



(1) Há um mecanismo causal plausível entre X e Y?

Por que e como $X \rightarrow Y$?

Problema teórico que exige de especificação do mecanismo em termos das decisões e ações individuais e/ou coletivas dos atores envolvidos. Conceitos abstratos **não** agem diretamente no mundo!

Filtros

- Teste da risada \rightarrow bom senso vale ouro
- Conhecimento acumulado \rightarrow mecanismo compatível com teoria e evidências produzidas no campo?

...só prossiga se o *sim* for enfático.

(2) Podemos descartar causalidade reversa?

Há risco de $Y \rightarrow X$?

Causalidade reversa pura e simples indica necessidade de reformulação teórica. Para muitas pesquisas, pode ser descartada por critérios lógicos ou pelo acúmulo de evidências.

E se $X \leftrightarrow Y$?

Problema comum e mais complicado ocorre quando fenômenos sociais se retroalimentam. Nesse caso, o desafio do **desenho de pesquisa** será achar alguma variação exógena em X .

...o *sim* facilita; prossiga com cautela se a resposta for *não*.

(3) Há covariação entre X e Y ?

Correlação \neq causalidade, mas é bom ponto de partida

A relação bivariada entre X e Y precisa ter algum tipo de associação objetivamente mensurável.

Quando há correlação bivariada entre X e Y , análise testa como ela se altera quando todos os controles necessários são introduzidos.

Se a relação esperada não existe, é preciso pensar bastante sobre variáveis omitidas. Em última instância, a variação **exógena** em X precisa ter associação com Y .

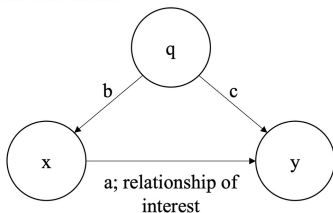
...o *sim* facilita; prossiga com cautela se a resposta for *não*.

(4) Há viés de variáveis omitidas?

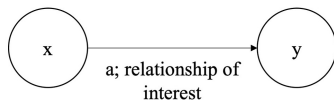
Multicausalidade e correlações espúrias

Todos os fenômenos podem ter muitas causas e/ou consequências, a análise precisa “controlar” – no desenho de pesquisa ou estatisticamente – por todas as outras causas de Y que possam estar correlacionadas a X .

Panel A: True model



Panel B: Specified model



Fonte: Wilms et al., [2021](#)

(4) Há viés de variáveis omitidas?

Estratégia de identificação

Estratégia para isolar o efeito de X sobre Y de todas as outras influências possíveis.

- Em desenhos experimentais, a **alocação aleatória** do tratamento resolve o problema
- Em desenhos observacionais clássicos, nunca podemos garantir que controlamos por todas as variáveis relevantes

Viés de variáveis omitidas no modelo de Rubin

Pode provocar tanto viés de seleção quanto viés de heterogeneidade de efeitos do tratamento.

Erros comuns

- Viés de sobrevivência
- Viés de auto-seleção
- Seleção na variável dependente
- Viés de seleção amostral
- etc

Recapitulação

Introdução

Notação matemática

Modelo de resultados potenciais

Guia de obstáculos à causalidade

Desenhos de pesquisa

Exemplos para o Brasil

Próxima aula

Bibliografia

Desenhos de pesquisa

Experimentais

Controlam a alocação aleatória do tratamento X e os valores de X .

- Manipulação intencional do tratamento
- Comparação entre grupo de tratamento e de controle maximizando validade interna conforme modelo de Rubin

Observacionais

Tentativa de emular experimentos a partir de dados pré-existentes, sem controle sobre alocação e dosagem do tratamento.

- **Quase experimentos** aproveitam manipulação exógena, mas não intencional, do tratamento
- **Estudos correlacionais** tentam isolar $X \rightarrow Y$ controlando estatisticamente por variáveis observadas

Desenhos experimentais

Para amostras com tamanho razoável, alocação aleatória garante grupos de tratamento e controle equivalentes para variáveis observadas e não observadas

- “Padrão-ouro” para contrafactuais para resultados potenciais: ideal para modelo de Rubin
- Intencionalidade e aleatoriedade eliminam obstáculos (2) e (4) da seção anterior
- Alocação aleatória \neq amostra aleatória

Há diversos tipos de experimentos: *field*, *survey*, *lab* etc

Desenhos experimentais

Problemas e limites

- Nem todo X de interesse é manipulável
- Não nos diz se X é a principal causa de Y
- Dilemas éticos
- Duplo risco à validade externa: amostragem e o próprio estímulo

Desenhos observacionais

Comparações a partir de dados pré-existentes, explorando variação no tempo e/ou entre unidades para construir contrafactuais

- **Quase experimentos** costumam combinar fontes de dados para explorar variações (aparentemente) exógenas em X que ocorrem *como se* fossem aleatórias

Problemas e limites

Quase experimentos precisam de estratégias de identificação críveis, o que nem sempre é fácil.

Em **estudos correlacionais**, não temos como garantir que controlamos estatisticamente por *todas* as outras causas possíveis de Y que são correlacionadas a X .

Recapitulação

Introdução

Notação matemática

Modelo de resultados potenciais

Guia de obstáculos à causalidade

Desenhos de pesquisa

Exemplos para o Brasil

Próxima aula

Bibliografia

Carvalhaes et al., 2022

Pergunta de pesquisa

Quais são os efeitos diretos e indiretos da renda e de cor/raça para entrada no ensino superior?

Desenho de pesquisa

Observacional, regressões logísticas.

Dados

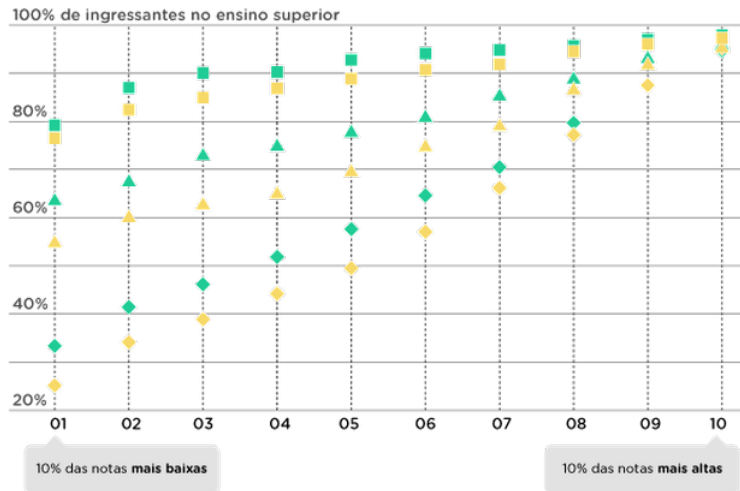
Painel com alunos de 16-22 anos que terminaram Ensino Médio em 2012 e são acompanhados até 2017 para identificar entrada no Ensino Superior. Dados identificados do Censo da Educação Básica, Censo do Ensino Superior e do ENEM.

TAXAS DE TRANSIÇÃO PARA O ENSINO SUPERIOR E DESEMPENHO NO ENEM

Por faixa de renda e raça/cor

	Até 0,25 sm per capita	De 0,75 a 1,0 sm	Maior que 1,75 sm
Branços			
PPI*			

*pretos, pardos e indígenas



Ferraz and Finan, 2008

Pergunta de pesquisa

Disponibilidade de informação aumenta *accountability* político? →
Eleitores punem políticos quando a corrupção é descoberta?

Desenho de pesquisa

Quase experimental, baseado no *timing* da divulgação de auditorias da CGU.

Dados

Relatórios das auditorias em 373 municípios com prefeito em primeiro mandato; resultados eleitorais e características dos prefeitos em 2000 e 2004; Censo 2000 e outras pesquisas do IBGE.

Ferraz and Finan, 2008

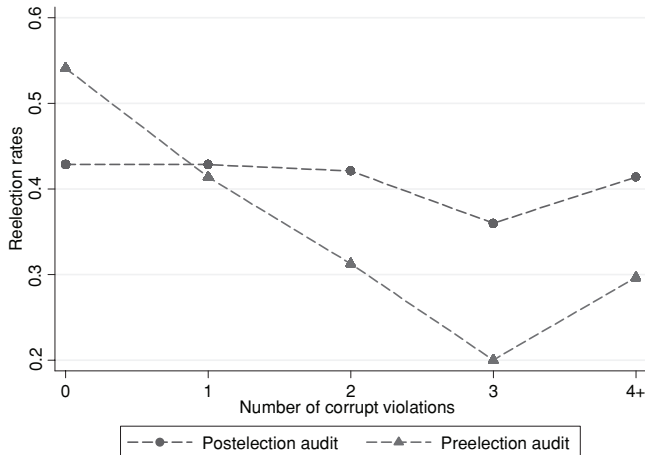


FIGURE III
Relationship between Reelection Rates and Corruption Levels

Fujiwara, 2015

Pergunta de pesquisa

Maior participação política de eleitores pobres aumenta redistribuição? → Inclusão via urna eletrônica aumentou gastos estaduais com saúde?

Desenho de pesquisa

Quase experimental; RDD a partir da introdução da urna eletrônica em 1998

Dados

Dados eleitorais do TSE; Censo 1991; SINASC; FINBRA

Fujiwara, 2015, p. 426

The estimates indicate that the de facto enfranchisement of approximately a tenth of Brazilian voters increased the share of states' budgets spent on health care by 3.4 percentage points (p.p.), raising expenditure by 34% in an eight-year period. It also boosted the proportion of uneducated mothers with more than seven prenatal visits by 7 p.p. and lowered the prevalence of low-weight births by 0.5 p.p. (respectively, a 19% and -6.8% change over sample averages).

Recapitulação

Introdução

Notação matemática

Modelo de resultados potenciais

Guia de obstáculos à causalidade

Desenhos de pesquisa

Exemplos para o Brasil

Próxima aula

Bibliografia

Próxima aula

Atividades

Entrega da atividade #2 no Google Classroom

Leituras obrigatórias

Curso-R, *Ciência de Dados em R*, caps. 1 a 6:

<https://livro.curso-r.com/index.html>

IBPAD, *Ciência de Dados em R – Introdução*, caps. 1 a 5:

<https://cdr.ibpad.com.br/index.html>

Leituras optativas

Roger Peng, *Programming for Data Science*, caps. 3 a 6:

<https://bookdown.org/rdpeng/rprogdatascience/>

Rafael Irizarry, *Introduction to Data Science*, caps. 1 a 6:

<https://rafalab.github.io/dsbook/>

Recapitulação

Introdução

Notação matemática

Modelo de resultados potenciais

Guia de obstáculos à causalidade

Desenhos de pesquisa

Exemplos para o Brasil

Próxima aula

Bibliografia

Bibliografia I

- Carvalhaes, F., Senkevics, A., & Costa Ribeiro, C. (2022). The intersection of family income, race and academic performance in access to Higher Education in Brazil. *Mimeo*.
- Dowd, B., & Town, R. (2002). Does X really cause Y? *Academy Health; Advancing research, policy and practice*.
- Ferraz, C., & Finan, F. (2008). Exposing corrupt politicians: The effect of Brazil's publicly released audits on electoral outcomes. *Quarterly Journal of Economics*, 123(2), 703–745.
- Fujiwara, T. (2015). Voting technology, political responsiveness, and infant health: Evidence from Brazil. *Econometrica*, 83(2), 423–464.

Bibliografia II

Goldthorpe, J. H. (2001). Causation, Statistics, and Sociology.

European Sociological Review, 17(1), 1–20.

Kellstedt, P., & Whitten, G. (2018). *The Fundamentals of Political Science Research*. Cambridge University Press.

Wilms, R., Mäthner, E., Winnen, L., & Lanwehr, R. (2021). Omitted variable bias: A threat to estimating causal relationships. *Methods in Psychology*, 5, 1–10.