

# **Introdução à Inferência Causal**

## **Resultados Potenciais**

Fernando Meireles  
IESP-UERJ

Agosto, 2019

# Efeitos e suas causas

Relações de causa e efeito são vistas em todo o lugar. Se eu derrubo meu *laptop* de uma certa altura, ele cai e quebra.

A queda é a causa da quebra.

Obviamente, vocês não precisam de um curso para aprender isso.

# Causas e seus efeitos

Estudar  
causalidade é,  
por definição,  
estudar o que  
não **não**  
**aconteceu**

- Se meu *laptop* não tivesse caído ele estaria intacto?
- Se eu tivesse investido em *bitcoins* há 5 anos, hoje eu estaria rico?
- Se a crise econômica de 2008 nunca tivesse ocorrido, Donald Trump seria hoje o Presidente dos Estados Unidos?

# Precisamos estudar isso?

## As alternativas

Se quisermos entender como fenômenos complexos — desigualdade, violência urbana, corrupção, entre inúmeros outros — surgem ou desaparecem, *precisamos ser capazes de testar afirmações causais.*

# Conhecimento causal

Correlação não é causalidade, mas não é disso que estamos tratando.

---

## Tipos de inferência

---

Preditiva  $\rightsquigarrow$  Sobre o futuro

Descritiva  $\rightsquigarrow$  Sobre estruturas

Classificatórias  $\rightsquigarrow$  Sobre pertencimentos

Causal  $\rightsquigarrow$  Sobre **contrafactuais**

---

# Motivação

---

# Questões causais

A maioria das **questões** que importam para nós são de **natureza causal**

- Corrupção na gestão educacional afeta o ensino dos(as) alunos(as) (e.g., Ferraz, Finan, and Moreira, 2012)?
- Programas de transferência condicionada de renda melhoram a qualidade de vida da população (e.g., Galiani and McEwan, 2013) ?
- Casos de violência em larga escala no passado afetam o compartamentos das pessoas no presente (e.g., Lupu and Peisakhin, 2017) ?
- Guerras promovem a integração de imigrantes (e.g., Lupu and Peisakhin, 2017) ?
- Os discursos de Hitler contribuiriam para a ascensão eleitoral do Partido Nacional Socialista (Selb and Munzert, 2018) ?

# Impacto real

- Mudar geograficamente para locais pobres afeta decisivamente o futuro de uma criança
  - Sabemos disso graças a [Chetty et al. \(2016\)](#);
- Censurar ações coletivas, e não protestos individuais, é a melhor forma de manter um regime autoritário
  - [King et al. \(2013\)](#) mostraram isso.



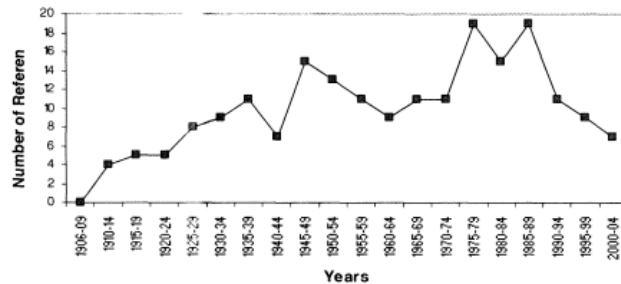
# Revolução de credibilidade

## Em números

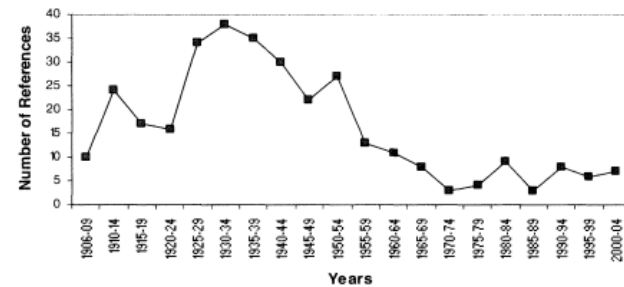
De Druckman et al. (2006):

**FIGURE 1. “Experimental” Trends in the Review**

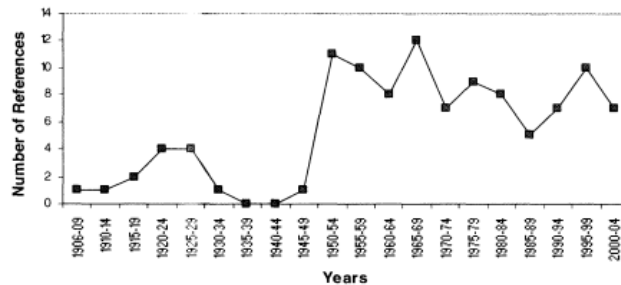
**References to Experiments as Empirical Tests**



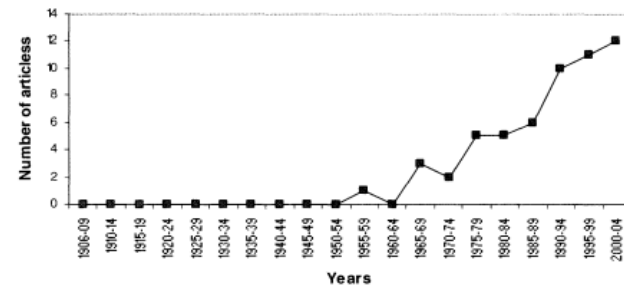
**References to Experiments as Institutional Innovations**



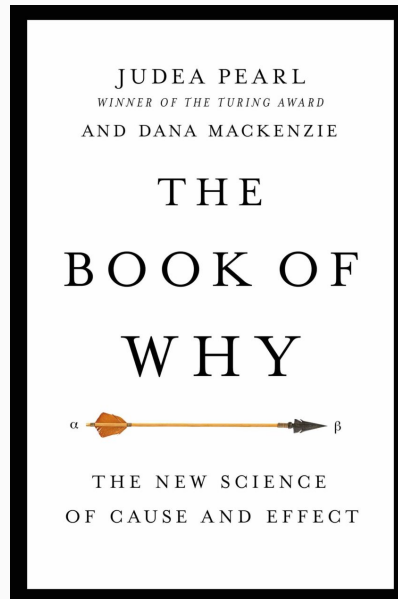
**References to Experiments as Studies with Randomized Trials (excluding experimental articles)**



**Experimental Articles**



## O estudo de causa e efeito se espalhou entre disciplinas



# Objetivo

Nosso objetivo neste curso é precisamente introduzir um conjunto de ferramentas que nos permitem falar, e principalmente analisar, afirmações causais.

# Plano desta aula

## Resultados Potenciais

Precisamos de uma base comum para discutirmos efeitos causais que seja rigorosa e simples.

Isso é o que cobriremos hoje.

Mas, antes, precisamos discutir também:

- *Nosso percurso*: resultados potenciais, DAGS, experimentos, experimentos naturais e outras estratégias comuns;
- *Método*: R, estatística básica, exercícios e tarefas;
- *Expectativas*: Não veremos tudo, mas entraremos no percurso principal.

# Contrafactuais

---

# Dois casos, múltiplos cenários

Obviamente, não podemos observar os dois cenários simultaneamente. Se *Fulano* estudar, e se *Ciclano* não, a comparação direta entre eles leva ao resultado contrário.

Nome	Estudou	Não estudou	Resultado
<i>Fulano</i>	6	?	?
<i>Ciclano</i>	?	7	?

$$6 - 7 = -1$$

# Contrafactualidade

O **problema central** de estudar causalidade é resolver essa falta do **cenário alternativo**, do **contrafactual**

Como só observamos um cenário para cada indivíduo, compará-los diretamente não nos permite saber se estudar causa um aumento nas notas. Precisamos de ferramentas.

# Resultados Potenciais

---



# O *Framework*

Podemos avançar no estudo de contrafactuais colocando o problema em termos mais claros. Usaremos o chamado **Modelo de Resultados Potenciais**, desenvolvido originalmente por Neyman e expandido por D. Rubin, entre outros (e popularizado por Holland).

A ideia central dele é fornecer uma base para *identificar* certas quantias.

## Alguma notação antes de avançarmos

- Digamos que a nota de *Fulano* quando ele estudou seja denotada por  $Y_i(1)$ ;
- E, quando ele não estudou, por  $Y_i(0)$ ;
- Estudar ou não pode ser indicado por  $D_i$  (e.g.,  $D_i = 1$  se estudou);
- $E$  indica um valor esperado e,  $|$ , uma relação condicional.

# Problema

Temos um tratamento  $D_i$ , também chamado de intervenção ou causa, e queremos saber o que acontece em alguma unidade quando ela submetida ao tratamento.

Queremos saber, portanto, quanto é  $Y_i(1) - Y_i(0)$ , ou

$$\tau = [Y_i | D_i = 1] - [Y_i | D_i = 0]$$

$$\tau = E[Y_i | D_i = 1] - E[Y_i | D_i = 0]$$

## A ausência de contrafactuais

Resultados potenciais nada mais são do que os cenários possíveis para  $Y_i$ . Mas, como vimos, só observamos um deles.

Como vimos, entretanto, apenas um pode ser observado.

# Problema fundamental da inferência causal

Não podemos observar diferentes resultados potenciais de uma mesma unidade.

Como contornamos isso?

# Pressupostos de identificação

Uma **quantia é identificada** quando, com dados infinitos, conseguimos estimar um valor fixo para ela. Isso depende **pressupostos**.

- Homogeneidade;
- Ignorabilidade (ou *exchangeability*, em inglês).

# Solução do *framework*

Quando um tratamento  $D$  é independente dos resultados potenciais  $Y$ , uma quantia pode ser identificada comparando indivíduos

- Basta comparar grupos de médias;
- O exemplo de ouro é o experimento, no qual o mecanismo de seleção é aleatório.

# Efeito médio

Enquanto que um efeito causal individual não pode ser calculado, podemos usar esse pressupostos de *ignorability* (mais alguns, que veremos) para estimar diferenças de médias entre tratados e não tratados.



# Vários casos, dois cenários

Vamos voltar ao problema da prova, agora com mais observações.

Nome	Estudou	Não estudou	Resultado
<i>Fulano</i>	6	?	?
<i>Ciclano</i>	?	6	?
<i>Ciclano</i>	?	4	?
<i>Ciclano</i>	8	?	?
<i>Ciclano</i>	?	2	?
<i>Ciclano</i>	4	?	?

$$\tau = \frac{1}{3}(6 + 8 + 4) - \frac{1}{3}(6 + 4 + 2)$$

$$\tau = 6$$

# Identificação

## Uma ciência da identificação

- Em modelagem, a prática é reproduzir o Processo Gerador de Dados (PGD), incluindo especificação correta e tudo o mais;
- Identificação depende, ao contrário, de um desenho, e pode ser implementada com métodos não-paramétricos ou simples testes randomizados.

Modelo  $\rightsquigarrow$  Estimação  $\rightsquigarrow$  Inferência

Desenho  $\rightsquigarrow$  Estimação  $\rightsquigarrow$  Inferência

# Observações

---

# A especificidade da inferência causal

- Na classificação, queremos usar  $X_i$  para descobrir  $Z_i$ ;
- Para fazer previsões, queremos apenas saber o valor de  $Y_i$  com base nos valores de  $X_i$ ;
- Na inferência causal, queremos saber qual valor  $Y_i$  assumirá se alterarmos  $X_i$ .

# Identificação e pressupostos

O importante a reter, enfim, é que **identificação** de efeitos causais depende centralmente de pressupostos, e não de técnicas ou modelos sofisticados.

Quando falamos de **estratégia de identificação**, estamos falando:

- Quais pressupostos, fora estimação, para explorar o mecanismo de seleção ao tratamento, nos permitem fazer afirmações causais com os dados que temos?

# Viés de seleção

Na ausência de **ignorability**, comparar unidades com tratamentos diferentes pode resultar em **viés de seleção**

- Hospital mata pessoas (e.g., se fui parar no hospital, minha saúde não estava boa previamente);
- Dinheiro pode não trazer votos (e.g., se eu tenho votos, consigo dinheiro);
- Educação pode não aumentar renda (e.g., se eu me eduquei, provavelmente tinha tempo livre e dinheiro).

# **Pausa para usar o R**

---

Vamos analisar os dados de (LaLonde, 1986), sobre o efeito de um programa de treinamento profissional sobre os salários de trabalhadores americanos.

- Dados:
  - 445 observações e 12 variáveis;
  - `nsw` é o indicador de tratamento,  $D_i$ ;
  - A variável de interesse é `re87` (salário em 1978, depois da intervenção).



# Referências

Ferraz, C, F. Finan, and D. B. Moreira (2012). "Corrupting learning: Evidence from missing federal education funds in Brazil". In: *Journal of Public Economics* 96.9-10, pp. 712-726.

Galiani, S. and P. J. McEwan (2013). "The heterogeneous impact of conditional cash transfers". In: *Journal of Public Economics* 103, pp. 85-96.

LaLonde, R. J. (1986). "Evaluating the econometric evaluations of training programs with experimental data". In: *The American economic review*, pp. 604-620.

Lupu, N. and L. Peisakhin (2017). "The legacy of political violence across generations". In: *American Journal of Political Science* 61.4, pp. 836-851.

Selb, P. and S. Munzert (2018). "Examining a most likely case for strong campaign effects: Hitler's speeches and the rise of the Nazi party, 1927-1933". In: *American Political Science Review* 112.4, pp. 1050-1066.