

# INTRODUÇÃO À INFERÊNCIA CAUSAL APLICADA À INDÚSTRIA

---

ROBSON TIGRE



Universidade dos Dados



**1.1**

# INTRO

---

Sobre mim e sobre este curso



**Universidade dos Dados**

# SOBRE O INSTRUTOR



## ROBSON TIGRE

PhD em Economia  
Cientista de dados



[linkedin.com/in/robson-tigre](https://www.linkedin.com/in/robson-tigre)



[robson.tigre0@gmail.com](mailto:robson.tigre0@gmail.com)



[github.com/RobsonTigre](https://github.com/RobsonTigre)



[sites.google.com/view/rtigre](https://sites.google.com/view/rtigre)

## Experiência

- Business & Data Science Expert – Mercado Livre
- Data Science Specialist - Stone
- Professor – Universidade Católica de Brasília
- Consultor – World Bank, ILO, CNI

## Formação

- PhD in Economics – Università di Bologna
- Pesq. visitante – UIUC e Toulouse School of Economics
- Mestrado em Economia – UFPE
- Bacharelado em Economia – UFPE



# SOBRE O INSTRUTOR



## ROBSON TIGRE

PhD em Economia  
Cientista de dados



[linkedin.com/in/robson-tigre](https://linkedin.com/in/robson-tigre)



[robson.tigre0@gmail.com](mailto:robson.tigre0@gmail.com)



[github.com/RobsonTigre](https://github.com/RobsonTigre)



[sites.google.com/view/rtigre](https://sites.google.com/view/rtigre)

## Publicações científicas

- "Illegal Markets and Contemporary Slavery: Evidence from the Mahogany Trade in the Amazon", with D. Araujo, Y. Barreto, & Danny Castro. *Journal of Development Economics*, 2024
- "Are Educated Candidates Less Corrupt Bureaucrats? Evidence from Randomized Audits in Brazil". with C. Melo. *Economic Development and Cultural Change*, 2024
- "Corruption and Economic Growth in Latin America and the Caribbean", with R. de Andrade Lima & L. Dutra. *Review of Development Economics*, 2022.
- "The impact of commuting time on youth's school performance", with B. Sampaio & T. Menezes. *Journal of Regional Science*, 2017.
- "Daylight Saving Time and Incidence of Myocardial Infarction: Evidence from a Regression Discontinuity Design", with W. Toro & B. Sampaio. *Economics Letters*, 2015.
- "Commute duration and health: Empirical evidence from Brazil", with R. Oliveira, K. Moura, J. Viana & B. Sampaio. *Transportation Research Part A: Policy and Practice*, 2015.

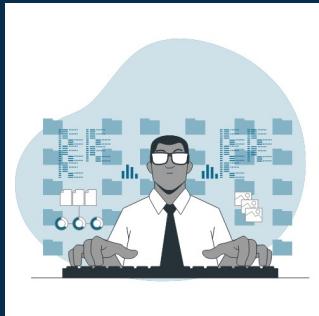


# SOBRE O CURSO

## PÚBLICO ALVO

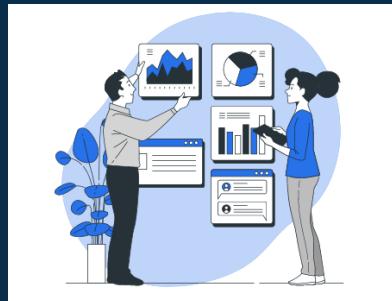
### Áreas de dados

- IC júnior
- IC sênior
- Manager



### Demais áreas

- Marketing
- Produto
- CX/UX
- Outros

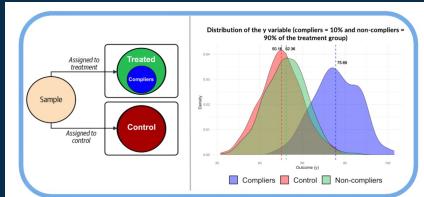


## TÓPICOS

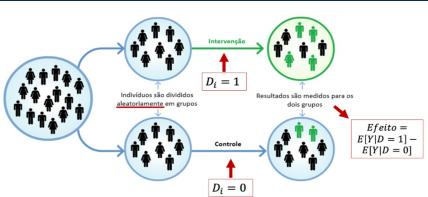
1. Introdução à econometria e viés de variável omitida
2. Experimentos aleatorizados e poder estatístico do experimento
3. Variável instrumental (ITT vs LATE)
4. Desenho de regressão descontínua (RDD)
5. Diferença em diferenças



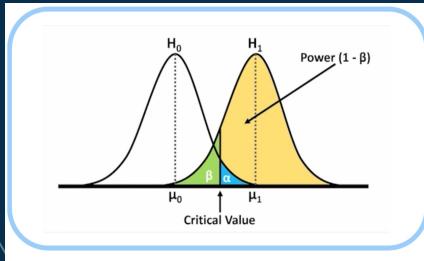
# Leituras recomendadas (curtinhas)



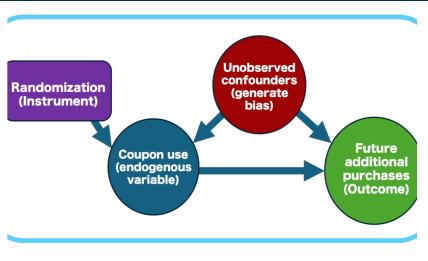
Parte 1: Simplificando a inferência causal para conectar stakeholders e cientistas de dados



Parte 2: Simplificando a inferência causal para conectar stakeholders e cientistas de dados



Parte 3: Simplificando a inferência causal para conectar stakeholders e cientistas de dados



ITT vs LATE: Estimando efeitos causais com IV em experimentos com imperfect compliance



# CAUSALIDADE PARA QUE?

---

*“Os dados podem te dizer que as pessoas que tomaram um medicamento se recuperaram mais rápido do que aquelas que não tomaram, mas eles não podem te dizer por quê. Talvez aqueles que tomaram o medicamento o fizeram porque podiam pagar por ele e teriam se recuperado tão rápido quanto sem ele.”*

Judea Pearl, The Book of Why: The New Science of Cause and Effect



# Inferência causal nos negócios

Mercado livre, Amazon, Stone, Nubank, Uber, etc. tem times para responder perguntas como:

- **Qual é o impacto de lançar uma nova campanha de marketing digital sobre as vendas?**
  - **Causa:** Lançamento da campanha de marketing digital → **Efeito:** Mudança nas vendas
  - **Contrafactual:** Qual seria o nível de vendas se a campanha de marketing digital não fosse lançada?
- **Qual é o efeito do programa de fidelidade sobre o gasto dos clientes ao longo do tempo?**
  - **Causa:** Participação no programa de fidelidade → **Efeito:** Aumento no gasto dos clientes
  - **Contrafactual:** Como seria o gasto dos clientes se eles não participassem do programa de fidelidade?
- **Reducir o preço de um produto em x% aumenta a demanda o suficiente para elevar a receita total?**
  - **Causa:** Redução do preço em x% → **Efeito:** Aumento nas vendas totais e na receita
  - **Contrafactual:** Como seriam as vendas e a receita se o preço não fosse reduzido?



Foco na palavra “**contrafactual**”, ela será o cerne até o fim da nossa jornada sobre inferência causal (IC)



# INFERÊNCIA ≠ PREDIÇÃO

---

## INFERÊNCIA CAUSAL

- Foca em entender a relação de causa e efeito entre variáveis.
- Objetivo: determinar como mudanças em uma variável (a causa) afetam outra variável (o efeito).
- Envolve intervenções e cenários contrafactual para estabelecer causalidade, e não apenas correlações.
- Ex1: "Qual é o efeito de uma nova campanha de marketing digital sobre as vendas?"
- Ex2: "Qual é o impacto do programa de fidelidade na retenção de clientes?"

## PREDIÇÃO/ML

- Foca em prever o valor futuro de uma variável com base em padrões identificados nos dados históricos.
- Objetivo: maximizar a precisão das previsões sem necessariamente entender suas relações causais.
- Se importa com capturar correlações entre variáveis explicativas e resposta
- Ex1: "Quais serão as vendas no próximo mês com base em campanhas de marketing passadas?"
- Ex2: "Qual será a probabilidade de um cliente participar do programa de fidelidade com base no comportamento passado?"



# EXEMPLO

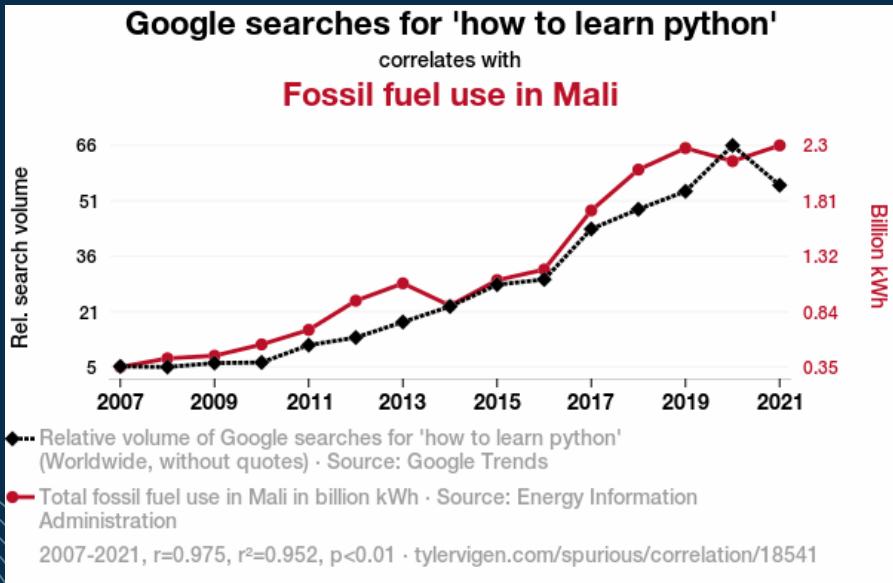
---

*“Em muitas indústrias, preços baixos estão associados a vendas baixas. Por exemplo, na indústria hoteleira, os preços são baixos fora da temporada turística, e os preços são altos quando a demanda é maior e os hotéis estão cheios. Dado esses dados, uma previsão ingênuia poderia sugerir que aumentar o preço levaria à venda de mais quartos.”*

Ajay Agrawal et al., Prediction Machines: The Simple Economics of Artificial Intelligence



# Correlação vs. Causalidade

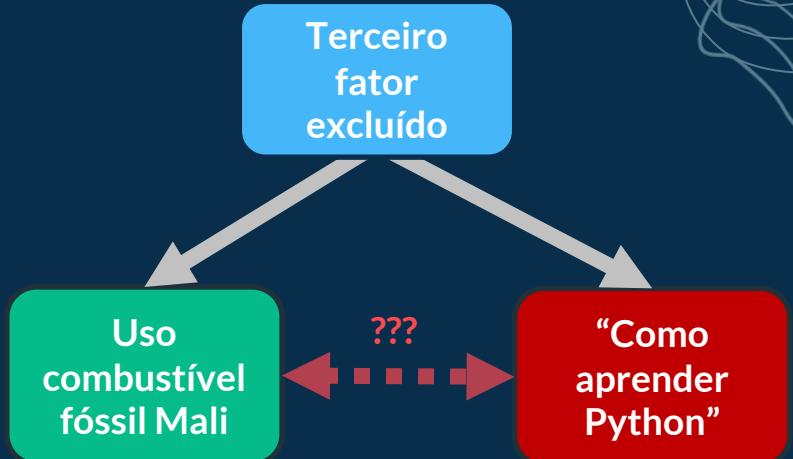
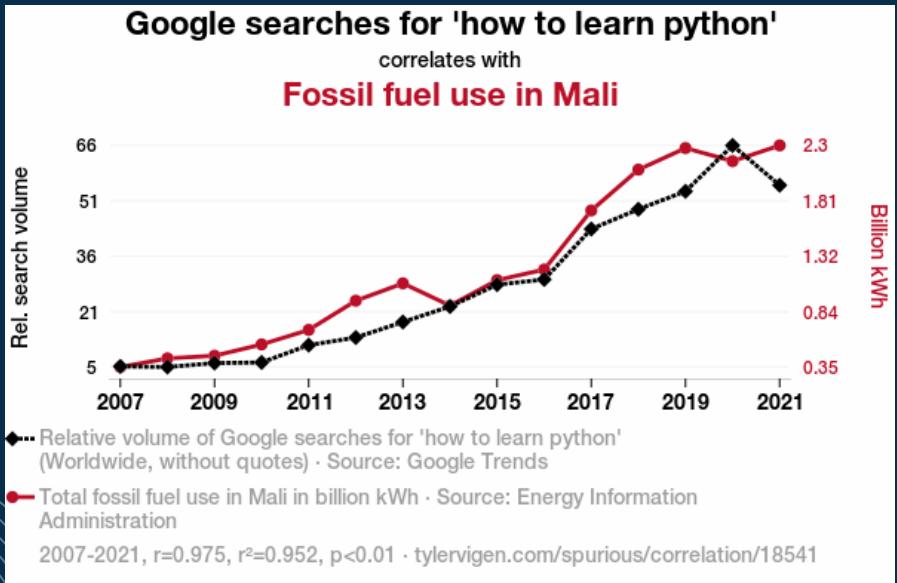


*"Não esperamos ver regularidade produzida por um processo aleatório, e quando detectamos o que parece ser uma regra, rapidamente rejeitamos a ideia de que o processo é realmente aleatório. Processos aleatórios produzem muitas sequências que convencem as pessoas de que o processo não é aleatório afinal."*

Daniel Kahneman, *Thinking, fast and slow*



# Correlação vs. Causalidade



# Terminologias (muito) importantes

Vamos usar esta terminologia durante o curso:

- Econometria: métodos estatísticos para testar teorias e avaliar políticas
- Regressão: relação entre variáveis - causal ou não
- Variável dependente, variável de resposta, regressando ou *outcome*: genericamente, usaremos  $y$  ou  $y_i$
- Variável independente, variável explicativa, regressor ou *variável de controle*: genericamente, usaremos  $x$  ou  $x_i$
- População: universo de indivíduos que fazem do grupo de interesse
- Amostra: uma parte ou subconjunto da *população*



# Terminologias (muito) importantes

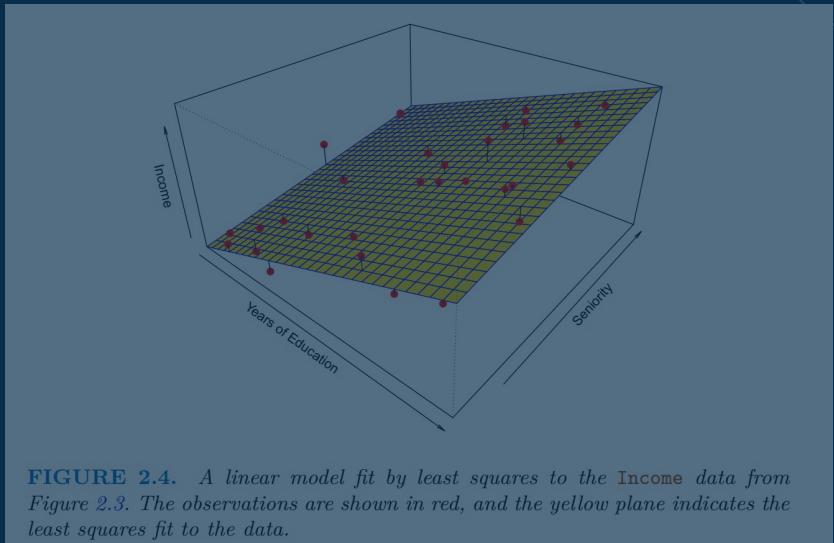
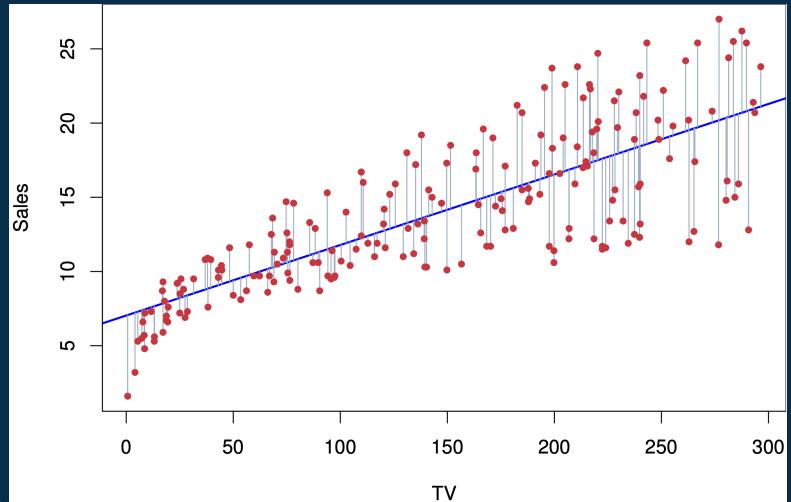
---

(CONTINUAÇÃO) Vamos usar esta terminologia durante o curso:

- Modelo: representação estruturada e simplificada da realidade. Em econometria, de algum fenômeno socioeconômico
- Parâmetro: medida de interesse. Geralmente é desconhecida por nós. Muitas vezes usaremos  $\beta$  para denotar um *parâmetro* da regressão
- Estimativa: aproximação amostral de  $\beta$ . É estimado, por nós, a partir de observações (amostra). Denotaremos por  $\hat{\beta}$  a estimativa de  $\beta$ .
- Inferência estatística: Técnicas para analisar a confiabilidade de relações estatísticas
  - Testes  $t$  e  $F$ , por exemplo, são ferramentas de inferência. Fazemos inferência, por exemplo, quando analisamos o p-valor de um teste  $t$  sobre a hipótese nula de  $\beta = 0$ .



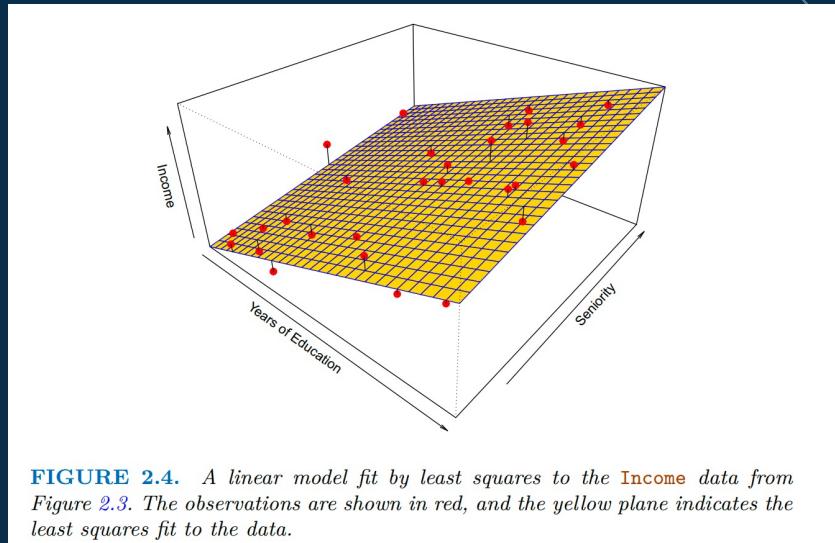
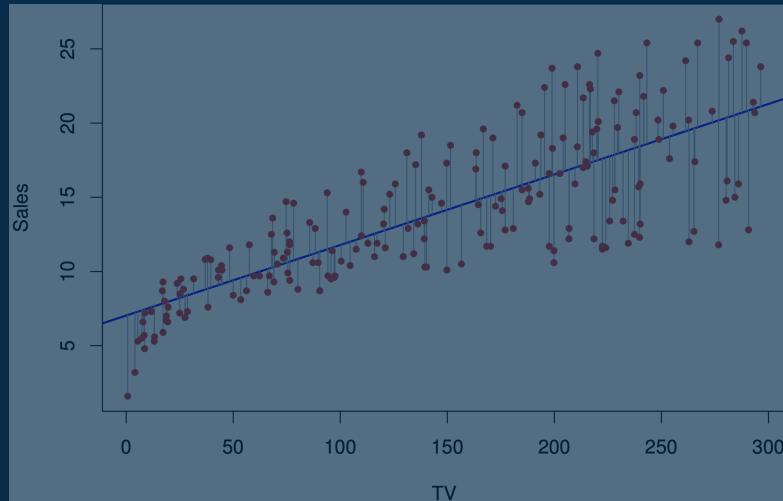
# A regressão linear simples



**FIGURE 2.4.** A linear model fit by least squares to the Income data from Figure 2.3. The observations are shown in red, and the yellow plane indicates the least squares fit to the data.



# A regressão linear múltipla



**FIGURE 2.4.** A linear model fit by least squares to the **Income** data from Figure 2.3. The observations are shown in red, and the yellow plane indicates the least squares fit to the data.



# A regressão linear simples

## Função de regressão populacional

- Suponha que o comportamento de uma determinada variável  $y$  seja sistematicamente explicado pelo comportamento da variável  $x$ , ou seja  $y = f(x)$
- Assumindo que esse função é linear (nos parâmetros), podemos traduzir a relação acima através do modelo linear  $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ , onde  $\varepsilon_i$  é o *termo de erro*, tal que  $E[\varepsilon] = 0$
- Note que  $E[y_i|x_i] = \beta_0 + \beta_1 x_i$ , que é chamada de *função de regressão populacional*
- $E[y_i|x_i = 1] = \beta_0 + \beta_1(\times 1) = \beta_0 + \beta_1$
- $E[y_i|x_i = 0] = \beta_0 + \beta_1(\times 0) = \beta_0$
- $\beta_1 = E[y_i|x_i = 1] - E[y_i|x_i = 0] = (\beta_0 + \beta_1) - \beta_0$
- $\beta_1 = \frac{\Delta E[y|x]}{\Delta x}$ . Em outras palavras,  $\beta_1$  mede a variação esperada ou média em  $y$  associada a uma variação unitária em  $x$ .



# Revisitando teste de hipóteses

Em um teste de hipótese com  $H_0: \beta_1=0$  e  $H_1: \beta_1\neq 0$ , desejamos testar se os dados da nossa amostra fornecem evidência suficiente para que afirmemos que  $\beta_1$  (aprox. por  $\hat{\beta}_1$ ) é  $\neq 0$

Ex: no experimento em que desejamos saber se o efeito de uma campanha de marketing é estatisticamente diferente de zero, analisamos:

$$\text{compras} = \beta_0 + \beta_1 \text{ Recebeu e-mail} + \varepsilon$$

Felizmente, softwares estatísticos já nos dão esse insight, e por isso focamos até agora em olhar o p-valor das estimativas

```
Call:  
lm(formula = compras ~ recebeu_email, data = usuarios)  
  
Residuals:  
    Min      1Q  Median      3Q     Max  
-140.284 -34.275   1.011   32.053 168.139  
  
Coefficients:  
            Estimate Std. Error t value          Pr(>|t|)  
(Intercept) 501.379    2.227 225.14 <0.0000000000000002 ***  
recebeu_email 98.416    3.168 31.06 <0.0000000000000002 ***  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 50.09 on 998 degrees of freedom  
Multiple R-squared:  0.4915,    Adjusted R-squared:  0.491  
F-statistic: 964.8 on 1 and 998 DF,  p-value: < 0.0000000000000002
```

Receber o email causa um aumento médio de  $\hat{\beta}_1=\text{R\$}98.4$ , estatisticamente significante, já que o p-valor é menor que 0,05



**1.2**

# **INTRO: PROBLEMA DO CONTRAFACTUAL**

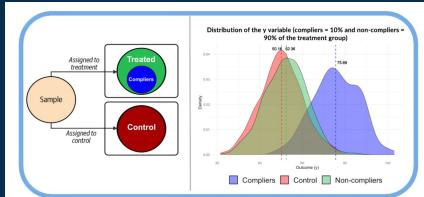
---

O problema fundamental da inferência causal

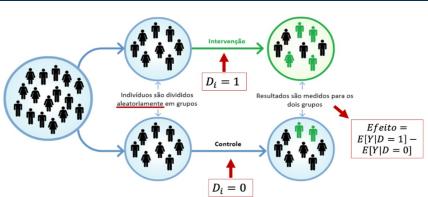


**Universidade dos Dados**

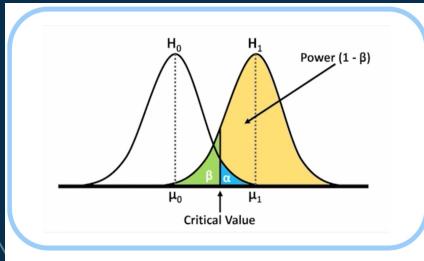
# Leituras recomendadas (curtinhas)



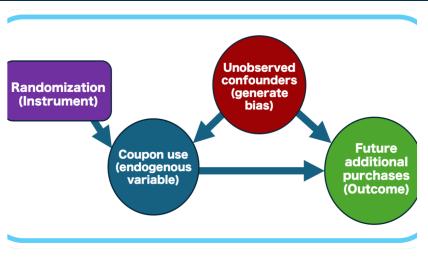
Parte 1: Simplificando a inferência causal para conectar stakeholders e cientistas de dados



Parte 2: Simplificando a inferência causal para conectar stakeholders e cientistas de dados



Parte 3: Simplificando a inferência causal para conectar stakeholders e cientistas de dados



ITT vs LATE: Estimando efeitos causais com IV em experimentos com imperfect compliance



# Modelo de outcomes potenciais

## Vocabulário do modelo

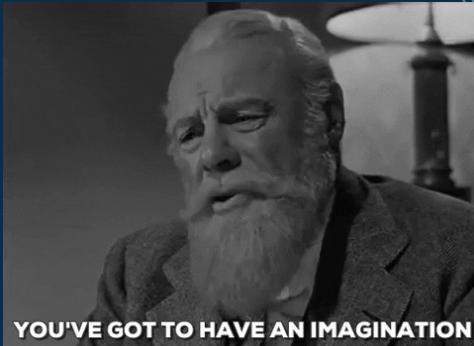
---

- **Tratamento** ( $T_i$  ou  $D_i$ ) → Intervenção recebida ou adotada pelos indivíduos.
- **Grupo de tratamento** ( $D_i = 1$ ) → Grupo de indivíduos que receberam o tratamento.
- **Grupo de controle** ( $D_i = 0$ ) → Grupo de indivíduos que não receberam o tratamento.
- **Resultado potencial** → “*Potential outcome*” é o resultado de  $Y$  que seria atribuído ao indivíduo em cada caso. i.e., ele tendo sido tratado ou não.
- **Contrafactual** → Qual seria o resultado (*outcome*) ou trajetória do indivíduo caso ele tivesse (ou não tivesse) recebido o tratamento.
- **Viés de seleção** → Tipo de viés de variável omitida em que há correlação entre o outcome potencial e algumas características do indivíduo.



# Modelo de resultados potenciais

- Vamos pensar no nosso tratamento  $D_i = \{0, 1\}$  como uma variável aleatória binária
- A pergunta é se o outcome  $Y_i$  é afetado por esse tratamento  $D_i$
- Agora vamos “viajar” um pouco: Imagine que você possa ver o estado do indivíduo  $i$  em cada um dos dois cenários referentes ao recebimento do tratamento  $D_i$ 
  - i.e., observar o indivíduo  $i$  tendo sido tratado e, *simultaneamente*, uma outra versão dele *não* tendo sido tratado - e.g. vendas de um e-commerce por dia após ter feito uma campanha de marketing vs. Esse mesmo e-commerce na ausência da campanha
- É claro que esse exercício é apenas imaginativo...

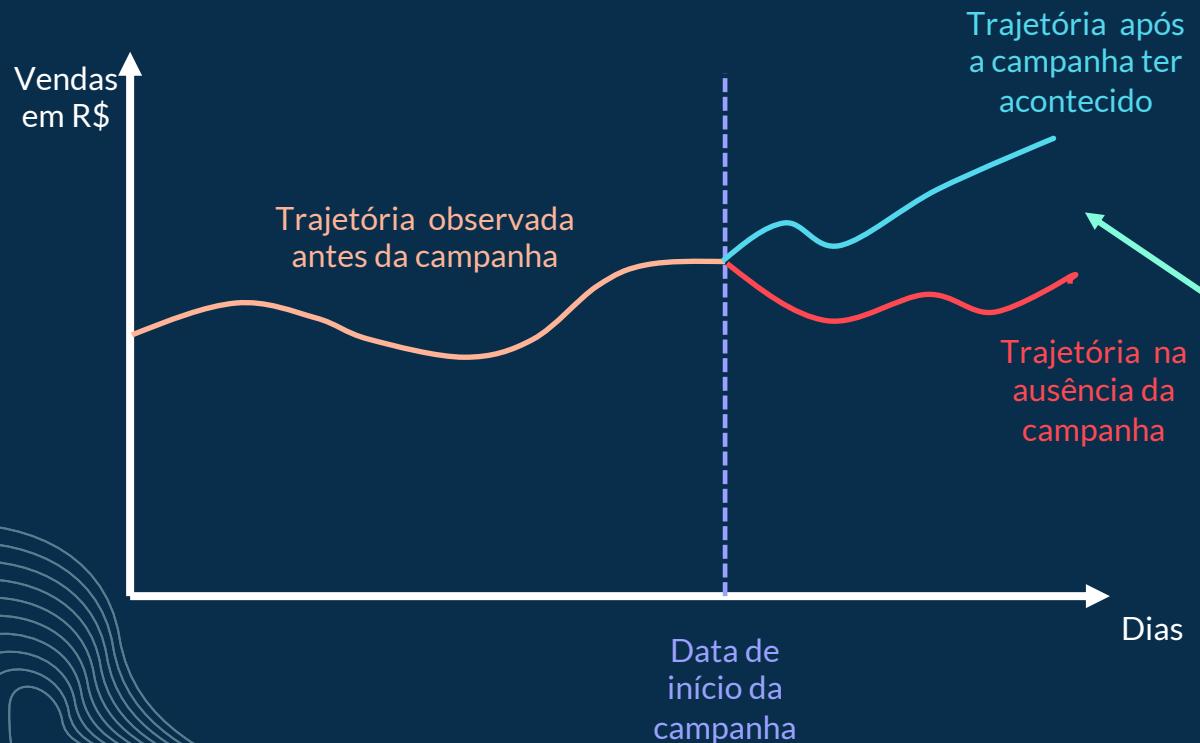


YOU'VE GOT TO HAVE AN IMAGINATION



# Modelo de resultados potenciais

## Contrafactual



Nesse mundo imaginário,  
enxergamos ambos os  
resultados potenciais.

No mundo real, vemos  
apenas um desses  
resultados, enquanto a  
inferência causal nos  
ajuda a estimar o outro.



# Modelo de resultados potenciais

---

- $Y_{1i}$  representa o valor do outcome para o indivíduo  $i$  sendo ele tratado (por isso o 1)
- $Y_{0i}$  representa o valor do outcome para *o mesmo indivíduo i* sendo ele *não* tratado (por isso o 0)
- O efeito causal (*conceitual*) do tratamento para o indivíduo  $i$  seria  $Y_{1i} - Y_{0i}$

$$\text{outcome potencial} = \begin{cases} Y_{1i} & \text{se } D_i = 1 \\ Y_{0i} & \text{se } D_i = 0 \end{cases}$$

- O *problema fundamental da inferência causal* é o simples fato de que não observamos a pessoa  $i$  **simultaneamente** tratada e não tratada

$$\text{outcome observado } Y_i = \begin{cases} Y_{1i} & \text{se } D_i = 1 \\ Y_{0i} & \text{se } D_i = 0 \end{cases} \rightarrow Y_i = Y_{0i} + (Y_{1i} - Y_{0i})D_i$$



# Modelo de resultados potenciais

## Um exemplo

Table: Exemplo de resultados potenciais

|  | João | Maria |
|--|------|-------|
| Resultado observado: $Y_i$               | 4    | 5     |
| Resultado potencial sem seguro: $Y_{0i}$ | 3    | 5     |
| Resultado potencial com seguro: $Y_{1i}$ | 4    | 5     |
| Tratamento (status de seguro): $D_i$     | 1    | 0     |
| Efeito do tratamento: $Y_{1i} - Y_{0i}$  | 1    | 0     |

$$Y_{1,Joao} - Y_{0,Joao} = 4 - 3 = 1$$

$$Y_{1,Maria} - Y_{0,Maria} = 5 - 5 = 0$$



# Modelo de resultados potenciais

## Um exemplo

|  | João | Maria |
|--|------|-------|
| Resultado observado: $Y_i$               | 4    | 5     |
| Resultado potencial sem seguro: $Y_{0i}$ |      | 5     |
| Resultado potencial com seguro: $Y_{1i}$ | 4    |       |
| Tratamento (status de seguro): $D_i$     | 1    | 0     |
| Efeito do tratamento: $Y_{1i} - Y_{0i}$  |      |       |

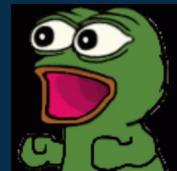
- Na vida real, não observaríamos os valores de  $Y_{0,Joao}$  nem de  $Y_{1,Maria}$
- E se fizéssemos uma **comparação ingênua de medias observadas?**

$$Y_{Joao} - Y_{Maria} = 4 - 5 = -1$$

- Isso é fruto de não termos acesso a uma informação importante: com ou sem plano de saúde, Maria seria mais saudável que João
- João não comprou o plano de saúde por acaso. Ele escolheu fazer parte do grupo de tratamento porque tem saúde invariavelmente mais frágil (**viés de seleção**).



# Exemplos de viés de seleção



Sempre que houver distinção entre os “tratados caso não tivessem sido tratados” e “não tratados de fato”, a diferença “ingênua” de médias trará um viés acoplado ao verdadeiro efeito causal



- Empresas com melhor potencial produtivo buscam com mais frequência determinadas linhas de crédito;
- Indivíduos com maior habilidade ou maior produtividade buscam mais anos de educação formal (e.g., vocês);
- Famílias em piores condições econômicas buscam Bolsa Família e outros programas de transferência de renda;
- Grupos com saúde mais frágil tomam vacina contra COVID primeiro;
- Pessoas com mais dificuldade para se reinserir no mercado de trabalho buscam programas de treinamento e agências de emprego do governo...



# 1.3

## INTRO: EXERCÍCIO ENDOGENEIDADE E VARIÁVEL OMITIDA

---

“Estudo sem prática é entretenimento”  
Heitor Sasaki



Universidade dos Dados

# Viés de variável omitida

---

- Serão disponibilizados scripts com todos os passos a seguir
- Muitas vezes em ciências de dados assumimos que não deixamos nada relevante de fora do nosso modelo.
- Digamos que o modelo verdadeiro seja  $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$  (regressão “longa”), mas que por descuido acabemos estimando  $Y_i = \beta_0 + \beta_1 X_{1i} + \eta_i$  (regressão “curta”)
- O que será que acontece? Isso seria o mesmo que dizer que o termo de erro da regressão curta, na verdade é  $\eta_i = \beta_2 X_{2i} + \varepsilon_i$ .
- É possível mostrar que se  $X_{2i}$  é correlacionado tanto com  $Y_i$  quanto com  $X_{1i}$ , deixá-lo fora da regressão fará com que  $\hat{\beta}_1$  se distancie do verdadeiro valor de  $\beta_1$
- **Isso acontece porque  $\hat{\beta}_1$  na regressão curta acabará absorvendo parte da informação de  $\beta_2$ , graças à omissão de  $X_2$ !**



# Viés de variável omitida e endogeneidade

Analise o “efeito” de escolaridade em rendimento \$?

- $wage_i$  = salário do indivíduo  $i$
- $educ_i$  = anos de estudo do indivíduo  $i$
- $wage_i = \beta_0 + \beta_1 educ_i + u_i$
- $\widehat{wage}_i = 146,95 + 60,21 educ_i$
- Um ano a mais de educação está *associado* a uma variação média no salário de U\$60,21

|             | Coefficients: | Estimate | Std. Error | t value    | Pr(> t ) |
|-------------|---------------|----------|------------|------------|----------|
| (Intercept) | 146.952       | 77.715   | 1.891      | 0.0589 .   |          |
| educ        | 60.214        | 5.695    | 10.573     | <2e-16 *** |          |



# Viés de variável omitida e endogeneidade

Muitas coisas afetam rendimento. Uma delas é habilidade:

- $wage_i$  = salário do indivíduo  $i$
- $educ_i$  = anos de estudo do indivíduo  $i$
- $iq_i$  = quociente de inteligência do indivíduo  $i$
- $wage_i = \beta_0 + \beta_1 educ_i + \beta_2 iq_i + u_i$
- $\widehat{wage}_i = -128,89 + 42,06 educ_i + 5,14 iq_i$
- Um ano a mais de educação está associado a uma variação média no salário de U\$42,06 *mantendo habilidade constante*

|             | Coefficients: | Estimate | Std. Error | t value      | Pr(> t ) |
|-------------|---------------|----------|------------|--------------|----------|
| (Intercept) | -128.8899     | 92.1823  | -1.398     | 0.162        |          |
| educ        | 42.0576       | 6.5498   | 6.421      | 2.15e-10 *** |          |
| iq          | 5.1380        | 0.9558   | 5.375      | 9.66e-08 *** |          |



# Viés de variável omitida e endogeneidade

---

- U\$60,21 (sem QI como controle) contra U\$42,06 (com QI como controle).
  - Entre indivíduos com a mesma habilidade, mais educação está associada a um menor ganho salarial quando comparado à primeira regressão.
- Esse resultado não é uma surpresa, já que *indivíduos com maior habilidade tendem a ter maior nível educacional*.
- Nesse caso, educação pode ser vista como uma *proxy* para habilidade. Na regressão “curta”,  $educ_i$  é correlacionado com  $\varepsilon_i$ , o que viola a hipótese  $E(\varepsilon|X) = 0$ .



# Viés de variável omitida e endogeneidade

Muitas coisas afetam rendimento. Uma delas pode ser tenure?

- $wage_i$  = salário do indivíduo  $i$
- $educ_i$  = anos de estudo do indivíduo  $i$
- $tenure_i$  = tempo do indivíduo  $i$  no emprego
- $wage_i = \beta_0 + \beta_1 educ_i + \beta_2 tenure_i + \varepsilon_i$
- $\widehat{wage}_i = 53,52 + 61,15 educ_i + 11,18 tenure_i$
- Um ano a mais de educação está associado a uma variação média no salário de U\$61,15 *mantendo tenure constante*

|             | Coefficients: | Estimate | Std. Error | t value      | Pr(> t ) |
|-------------|---------------|----------|------------|--------------|----------|
| (Intercept) | 53.519        | 79.557   | 0.673      | 0.501        |          |
| educ        | 61.148        | 5.639    | 10.844     | < 2e-16 ***  |          |
| tenure      | 11.177        | 2.441    | 4.580      | 5.29e-06 *** |          |



# Viés de variável omitida e endogeneidade

- $wage_i = \beta_0 + \beta_1 educ_i + \varepsilon_i$        $\hat{\beta}_1 = 60,21$
- $wage_i = \beta_0 + \beta_1 educ_i + \beta_2 iq_i + \varepsilon_i$        $\hat{\beta}_1 = 42,06$
- $wage_i = \beta_0 + \beta_1 educ_i + \beta_2 tenure_i + \varepsilon_i$        $\hat{\beta}_1 = 61,15$
- Há alta correlação entre educação e iq – habilidade ( $r = 0.51$ ), o que sugere que educação realmente é uma proxy para habilidade.
- O mesmo não vale para tenure, pois há uma correlação baixa ( $r = -0,0362$ )

```
> # Matriz de correlação
> cor(nls80[, c("educ", "iq", "tenure")])
      educ          iq          tenure
educ  1.000000000  0.51569701 -0.03616655
iq    0.51569701  1.00000000  0.04215883
tenure -0.03616655  0.04215883  1.00000000
```



# Viés de variável omitida e endogeneidade

## “Anatomia do viés de variável omitida”

- Modelo verdadeiro:  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$
- Modelo estimado:  $Y = \beta_0 + \beta_1 X_1 + \nu$

$$\begin{aligned}\hat{\beta}_1 &= \frac{Cov(Y, X_1)}{Var(X_1)} = \frac{Cov(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon, X_1)}{Var(X_1)} = \\ &= \frac{Cov(\beta_0, X_1) + \beta_1 Cov(X_1, X_1) + \beta_2 Cov(X_1, X_2) + Cov(\varepsilon, X_1)}{Var(X_1)} =\end{aligned}$$

magnitude da  
“relação” entre  $X_1$  e  
a variável omitida  $X_2$

$$E[\hat{\beta}_1] = \frac{0 + \beta_1 Var(X_1) + \beta_2 Cov(X_1, X_2) + 0}{Var(X_1)} \rightarrow E[\hat{\beta}_1] = \beta_1 + \left[ \beta_2 \times \frac{Cov(X_1, X_2)}{Var(X_1)} \right]$$

↑  
Magnitude da “relação” entre  $Y$  e  
a variável omitida  $X_2$



# Viés de variável omitida e endogeneidade

---

- Modelo verdadeiro:  $wage_i = \beta_0 + \beta_1 educ_i + \beta_2 iq_i + \varepsilon_i$

- Modelo estimado:  $wage_i = \beta_0 + \beta_1 educ_i + v_i$

$$E[\hat{\beta}_1] = \beta_1 + \frac{\beta_2 Cov(X_1, X_2)}{Var(X_1)}$$

- $\hat{\beta}_1 = 42,06$  no modelo verdadeiro e  $\hat{\beta}_1 = 60,21$  no modelo estimado
  - $\beta_2 > 0, Cov(educ, iq) > 0, Var(educ) > 0$ . Resultado  $\rightarrow$  Viés positivo em  $E[\hat{\beta}_1]$



# Viés de variável omitida e endogeneidade

- Modelo verdadeiro:  $wage_i = \beta_0 + \beta_1 educ_i + \beta_2 tenure_i + \varepsilon_i$
- Modelo estimado:  $wage_i = \beta_0 + \beta_1 educ_i + v_i$

$$E[\hat{\beta}_1] = \beta_1 + \frac{\beta_2 Cov(X_1, X_2)}{Var(X_1)}$$

- $\hat{\beta}_1 = 61,15$  no modelo verdadeiro e  $\hat{\beta}_1 = 60,21$  no modelo estimado. Omitir uma variável do modelo *nem sempre* cria viés.
  - $\beta_2 > 0$  ( $\hat{\beta}_2 = 11,18$ ),  $Cov(educ, tenure) \approx 0$  ( $Corr(educ, tenure) = -0,0362$ ,  $Var(educ) > 0 \rightarrow$  Viés quase nulo)
- As condições para haver viés são que: a variável omitida seja correlacionada com  $Y$  e também com a variável explicativa incluída na regressão curta.





# Lição desta primeira parte

- Em resumo, queremos métodos capazes de produzir estimativas do efeito do tratamento que estejam *limpas de viés de variável omitida*
  - O viés de variável omitida ocorre quando omitimos do modelo alguma variável que está correlacionada com  $Y$  e com o  $X$  ou tratamento em questão
- Esses métodos partem do mesmo princípio de exogeneidade que é levado ao extremo no método experimental.
  - Tentam encontrar uma variação na probabilidade de tratamento que não esteja fortemente correlacionada com o termo de erro



# 2.1

## EXPERIMENTOS ALEATORIZADOS

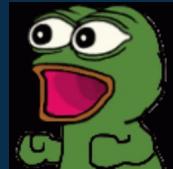
---

Motivação e princípios básicos



Universidade dos Dados

# Exemplos de viés de seleção



Sempre que houver distinção entre os “tratados caso não tivessem sido tratados” e “não tratados de fato”, a diferença “ingênua” de médias trará um viés acoplado ao verdadeiro efeito causal



- Empresas com melhor potencial produtivo buscam com mais frequência determinadas linhas de crédito;
- Indivíduos com maior habilidade ou maior produtividade buscam mais anos de educação formal (e.g., vocês);
- Famílias em piores condições econômicas buscam Bolsa Família e outros programas de transferência de renda;
- Grupos com saúde mais frágil tomam vacina contra COVID primeiro;
- Pessoas com mais dificuldade para se reinserir no mercado de trabalho buscam programas de treinamento e agências de emprego do governo...



# MOTIVAÇÃO

- O método experimental é considerado a maneira mais precisa de obter resultados próximos à “verdade” (Kohavi et al., 2019).
- O impacto estimado usando dados históricos é frequentemente 2 vezes maior do que a estimada com dados experimentais (Gordon et al., 2022).
- Conduzir um experimento poderia nos ajudar a obter resultados mais precisos e confiáveis.

## Close Enough? A Large-Scale Exploration of Non-Experimental Approaches to Advertising Measurement

Brett R. Gordon , Robert Moakler, Florian Zettelmeyer

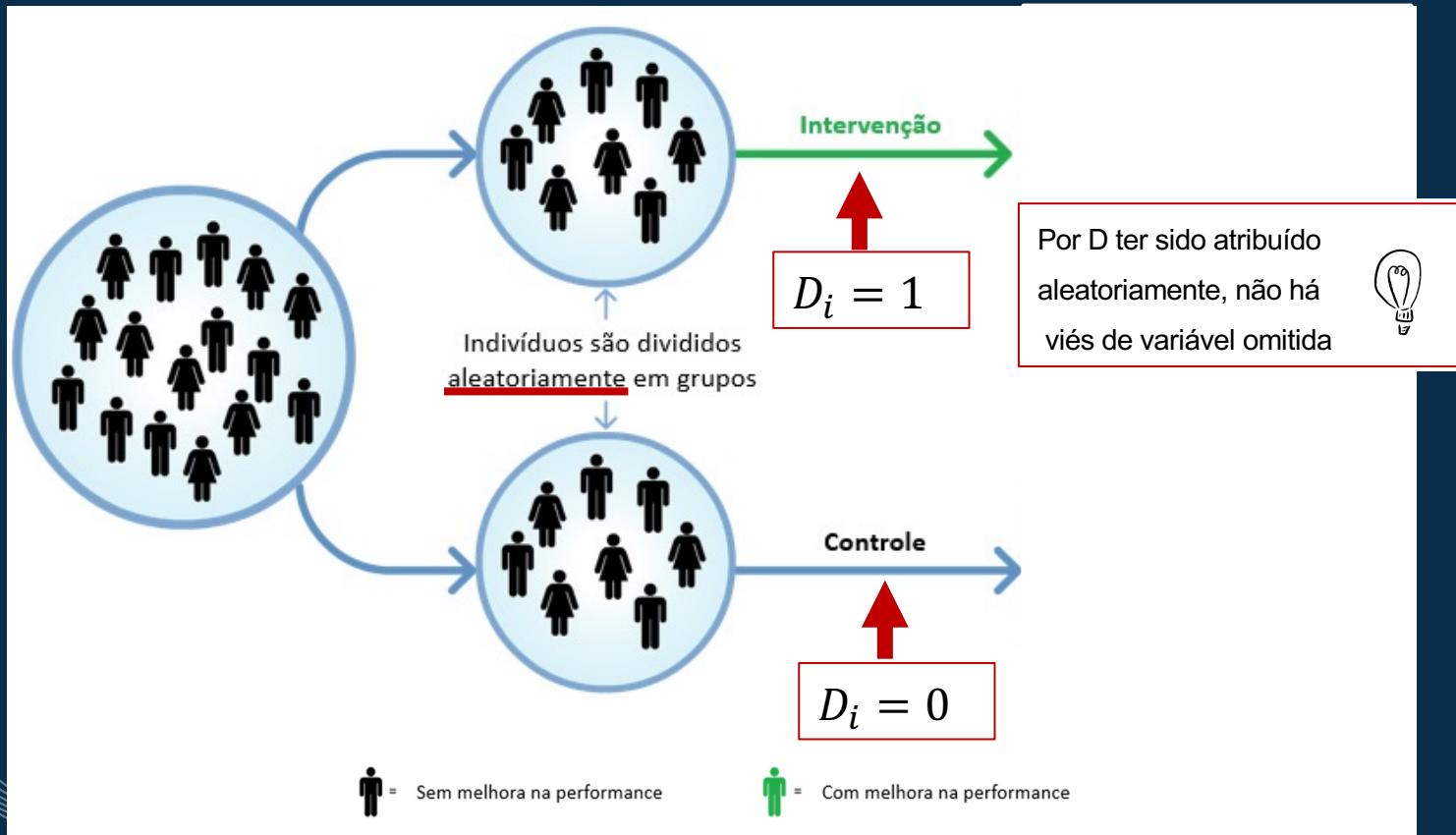
Published Online: 7 Nov 2022 | <https://doi.org/10.1287/mksc.2022.1413>

### Abstract

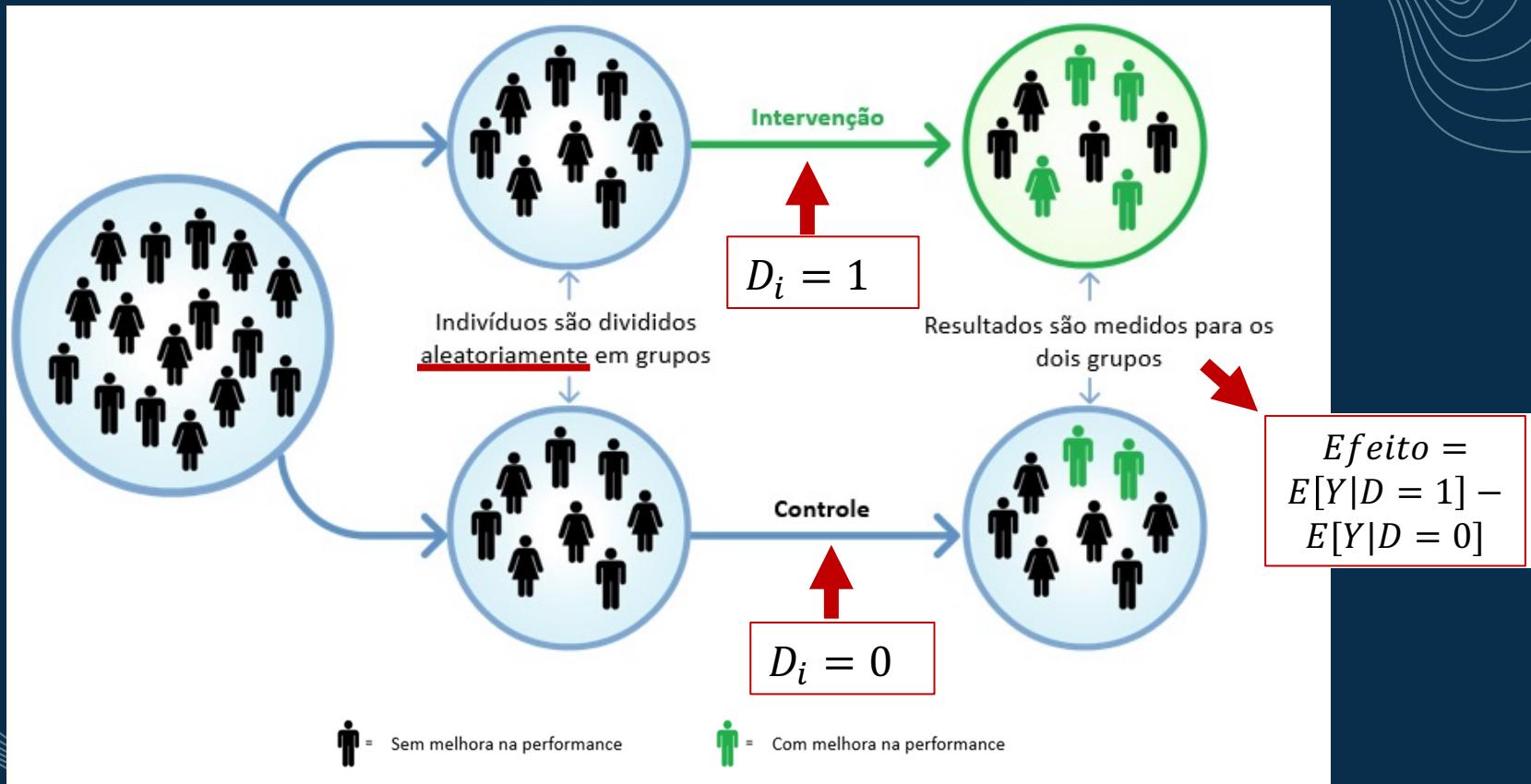
Despite their popularity, randomized controlled trials (RCTs) are not always available for the purposes of advertising measurement. Non-experimental data are thus required. However, Facebook and other ad platforms use complex and evolving processes to select ads for users. Therefore, successful non-experimental approaches need to “undo” this selection. We analyze 663 large-scale experiments at Facebook to investigate whether this is possible with the data typically logged at large ad platforms. With access to over 5,000 user-level features, these data are richer than what most advertisers or their measurement partners can access. We investigate how accurately two non-experimental methods—double/debiased machine learning (DML) and stratified propensity score matching (SPSM)—can recover the experimental effects. Although DML performs better than SPSM, neither method performs well, even using flexible deep learning models to implement the propensity and outcome models. The median RCT lifts are 29%, 18%, and 5% for the upper, middle, and lower funnel outcomes, respectively. Using DML (SPSM), the median lift by funnel is 83% (173%), 58% (176%), and 24% (64%), respectively, indicating significant relative measurement errors. We further characterize the circumstances under which each method performs comparatively better. Overall, despite having access to large-scale experiments and rich user-level data, we are unable to reliably estimate an ad campaign’s causal effect.



# Uma forma de remover o viés de seleção...



# Uma forma de remover o viés de seleção...



## 2.2

# EXPERIMENTOS: EXERCÍCIO

---

Comparação entre resultados  
observacionais e experimentais



Universidade dos Dados

# Assimilando através de um exemplo

Aleatorização é o padrão ouro

- Indivíduos tem probabilidades *conhecidas* de serem tratados (e.g., sorteio).
- Os grupos de **tratamento e controle** serão estatisticamente semelhantes nas mais diversas características (ver tabelas de balanceamento geradas pelo script em R)
- **Diferenças que surjam entre esses dois grupos poderão ser atribuídas à intervenção**

Características

- Os tratamentos devem ser mutuamente excludentes → Cada indivíduo só deve ser exposto a um **único tratamento a cada período**.
- Os efeitos são sempre relativos → O efeito da intervenção com relação aos grupos em outros status de tratamento.
- O efeito do tratamento deve ocorrer *necessariamente* após a exposição da unidade ao tratamento (**checkar usando baseline survey e follow-up survey**)



# Job training program (LaLonde, 1986)

Evaluating the Econometric Evaluations of Training Programs with Experimental Data

## Evaluating the Econometric Evaluations of Training Programs with Experimental Data

By ROBERT J. LALONDE\*

*This paper compares the effect on trainee earnings of an employment program that was run as a field experiment where participants were randomly assigned to treatment and control groups with the estimates that would have been produced by an econometrician. This comparison shows that many of the econometric procedures do not replicate the experimentally determined results, and it suggests that researchers should be aware of the potential for specification errors in other nonexperimental evaluations.*



# Job training program (LaLonde, 1986)

Evaluating the Econometric Evaluations of Training Programs with Experimental Data

| Variável | Significado   |
|----------|---|
| train    | = 1 se participou do treinamento em 1977                        |
| age      | Idade em 1977 (i.e., no ano do treinamento)                     |
| educ     | Anos de educação (no ano do treinamento)                        |
| black    | = 1 se o indivíduo é preto/negro                                |
| hisp     | = 1 se o indivíduo é hispânico                                  |
| re74     | Rendimento do trabalho, <b>em milhares de dólares</b> , em 1974 |
| re75     | Rendimento do trabalho, <b>em milhares de dólares</b> , em 1975 |
| re78     | Rendimento do trabalho, <b>em milhares de dólares</b> , em 1978 |
| unem74   | = 1 se o indivíduo estava desempregado em 1974                  |
| unem75   | = 1 se o indivíduo estava desempregado em 1975                  |
| unem78   | = 1 se o indivíduo estava desempregado em 1978                  |



# Job training program (LaLonde, 1986)

## Evaluating the Econometric Evaluations of Training Programs with Experimental Data

Utilizando os **dados observacionais** para estimar o seguinte modelo

$$re78 = \beta_0 + \beta_1 train + \varepsilon$$

obtemos  $\widehat{re78} = 21.554^{***} - 15.20478^{***} train$

```
> table(jtrain3$train, jtrain3$unem75)
```

|  |   | 0    | 1   | Unem75 = 1 se desempregado em 1975 |
|--|---|------|-----|------------------------------------|
| train = 1 se participou do treinamento em 1977 | 0 | 2241 | 249 |                                    |
|  | 1 | 74   | 111 |                                    |

Apenas 10% dos indivíduos do grupo de controle estavam desempregados em 1975. 60% dos indivíduos do grupo de tratamento estavam desempregados em 1975.



# Job training program (LaLonde, 1986)

## Evaluating the Econometric Evaluations of Training Programs with Experimental Data

- Há “auto seleção”: Trabalhadores em piores condições são aqueles que se “selecionam” para participar do programa. Isso leva a estimativas enviesadas
- Estimando um modelo mais realista, como...

$$re78 = \beta_0 + \beta_1 train + \beta_2 re75 + \beta_3 educ + \beta_4 age + \varepsilon$$

Coefficients:

|             | Estimate | Std. Error | t value | Pr(> t )                |
|-------------|----------|------------|---------|-------------------------|
| (Intercept) | 0.60487  | 1.22266    | 0.495   | 0.6208                  |
| train       | -0.59937 | 0.84109    | -0.713  | 0.4761                  |
| re75        | 0.79532  | 0.01652    | 48.151  | <0.0000000000000002 *** |
| educ        | 0.61154  | 0.07281    | 8.399   | <0.0000000000000002 *** |
| age         | -0.04655 | 0.02063    | -2.257  | 0.0241 *                |
| ---         |          |            |         |                         |

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1



# Job training program (LaLonde, 1986)

## Evaluating the Econometric Evaluations of Training Programs with Experimental Data

E se adicionarmos variáveis que provavelmente explicam salário em 1978 e estão correlacionadas com querer participar do treinamento?

$$re78 = \beta_0 + \underline{\beta_1} train + \beta_2 re75 + \beta_3 re74 + \beta_4 educ + \beta_5 age + \beta_6 black + \beta_7 hisp + \varepsilon$$

| Coefficients:   |          |            |         |                      |              |
|---|----------|------------|---------|----------------------|--------------|
|   | Estimate | Std. Error | t value |                      | Pr(> t )     |
| (Intercept)   | 1.64755  | 1.30093    | 1.266   |                      | 0.205465     |
| train   | 0.21323  | 0.85339    | 0.250   |                      | 0.802716     |
| re75  | 0.56929  | 0.02757    | 20.648  | < 0.0000000000000002 | ***          |
| re74  | 0.28098  | 0.02790    | 10.071  | < 0.0000000000000002 | ***          |
| educ  | 0.52006  | 0.07522    | 6.914   | 0.00000000000589     | ***          |
| age   | -0.07507 | 0.02047    | -3.667  |                      | 0.000251 *** |
| black   | -0.64771 | 0.49193    | -1.317  |                      | 0.188056     |
| hisp  | 2.20261  | 1.09279    | 2.016   |                      | 0.043944 *   |
| ---   |          |            |         |                      |              |
| Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 |          |            |         |                      |              |



# Job training program (LaLonde, 1986)

## Evaluating the Econometric Evaluations of Training Programs with Experimental Data

Utilizando os **dados experimentais** para estimar o seguinte modelo

$$re78 = \beta_0 + \beta_1 train + \varepsilon$$

obtemos  $\widehat{re78} = 4.555^{***} + 1.7943^{***} train$

```
> table(jtrain2$train, jtrain2$unem75)
```

|   | 0  | 1   | Unem75 = 1 se desempregado em 1975 |
|---|----|-----|------------------------------------|
| 0 | 82 | 178 |                                    |
| 1 | 74 | 111 |                                    |

**train** = 1 se participou do  
treinamento em 1977

68% dos indivíduos do grupo de controle estavam desempregados em 1975. 60% dos indivíduos do grupo de tratamento estavam desempregados em 1975.



# Job training program (LaLonde, 1986)

## Evaluating the Econometric Evaluations of Training Programs with Experimental Data

- Aparentemente não há auto seleção. Trabalhadores em piores condições não podem se “selecionar” para participar do programa, já que ele é atribuído de forma aleatória
- Estimando um modelo mais realista, como...

$$re78 = \beta_0 + \underline{\beta_1} train + \beta_2 re75 + \beta_3 educ + \beta_4 age + \varepsilon$$

Coefficients:

|             | Estimate | Std. Error | t value | Pr(> t ) |
|-------------|----------|------------|---------|----------|
| (Intercept) | -0.59130 | 2.08797    | -0.283  | 0.7772   |
| train       | 1.62137  | 0.63160    | 2.567   | 0.0106 * |
| re75        | 0.15689  | 0.09869    | 1.590   | 0.1126   |
| educ        | 0.38730  | 0.17361    | 2.231   | 0.0262 * |
| age         | 0.04151  | 0.04382    | 0.947   | 0.3439   |
| ---         |          |            |         |          |

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1



# Job training program (LaLonde, 1986)

## Evaluating the Econometric Evaluations of Training Programs with Experimental Data

E se adicionarmos variáveis que provavelmente explicam salário em 1978 e estão correlacionadas com *querer* participar do treinamento?

$$re78 = \beta_0 + \underline{\beta_1} train + \beta_2 re75 + \beta_3 re74 + \beta_4 educ + \beta_5 age + \beta_6 black + \beta_7 hisp + \varepsilon$$

Coefficients:

|                | Estimate       | Std. Error | t value | Pr(> t )   |
|----------------|----------------|------------|---------|------------|
| (Intercept)    | 0.67407        | 2.42272    | 0.278   | 0.78097    |
| train          | <u>1.68005</u> | 0.63086    | 2.663   | 0.00803 ** |
| re75           | 0.04677        | 0.13068    | 0.358   | 0.72062    |
| re74           | 0.08331        | 0.07653    | 1.089   | 0.27694    |
| educ           | 0.40360        | 0.17485    | 2.308   | 0.02145 *  |
| age            | 0.05435        | 0.04382    | 1.240   | 0.21560    |
| black          | -2.18007       | 1.15550    | -1.887  | 0.05987 .  |
| hisp           | 0.14356        | 1.54092    | 0.093   | 0.92582    |
| ---            |                |            |         |            |
| Signif. codes: | 0 ***          | 0.001 **   | 0.01 *  | 0.05 .     |
|                | 0.1            | '          | '       | 1          |



## 2.3

# EXPERIMENTOS: SUPOSIÇÕES

---

Hipóteses necessárias para que os resultados sejam causais



Universidade dos Dados

# Considerações

## ✓ Validação interna

- O efeito estimado é uma estimativa não-enviesada dos efeitos médios *na população estudada*.
- Ou seja, o coeficiente estimado corresponde ao efeito real médio *da amostra estimada*.

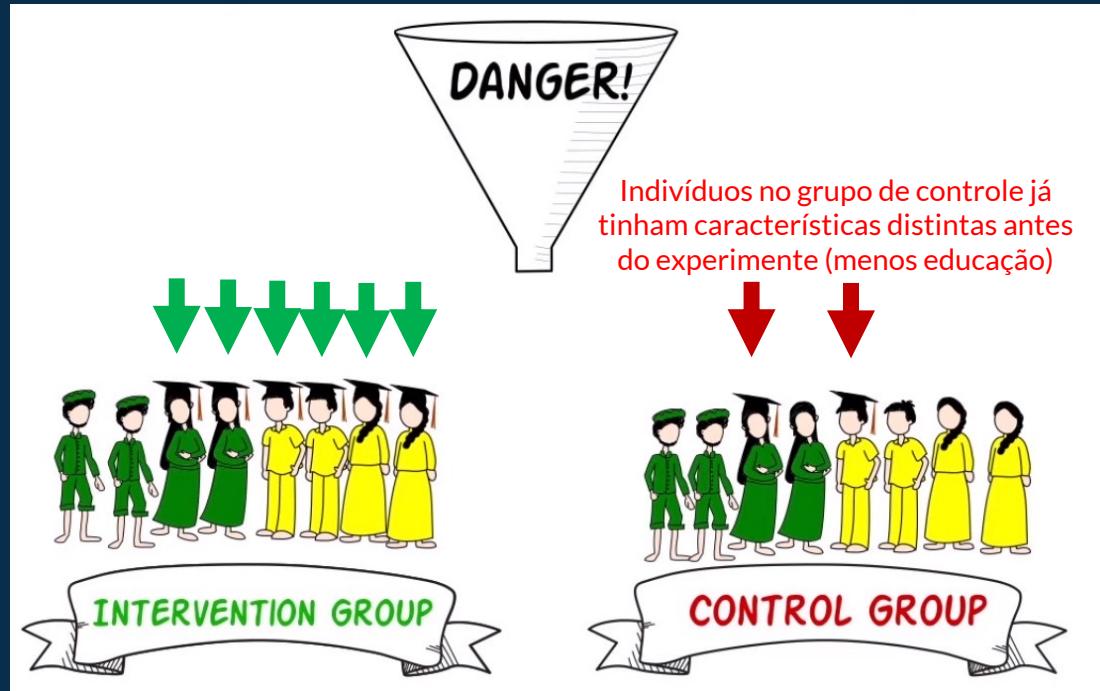
## ⚠ Validação externa

- O coeficiente estimado não necessariamente é generalizável para a população de interesse
- Os efeitos estimados referem-se somente àquela intervenção, naquela amostra, e da forma específica como a intervenção foi implementada



# Principais hipóteses para a identificação de causalidade

## O tratamento realmente é distribuído aleatoriamente

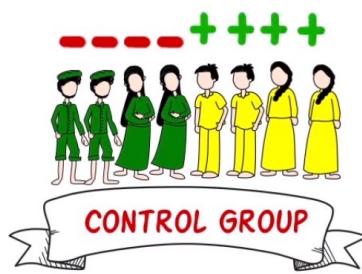


# Principais hipóteses para a identificação de causalidade

Os dados coletados e efeitos estimados referem-se ao que foi planejado

## ATTRITION

Sem atrito, vemos que não haveria diferenças entre os dois grupos. Efeito = 0, mas...



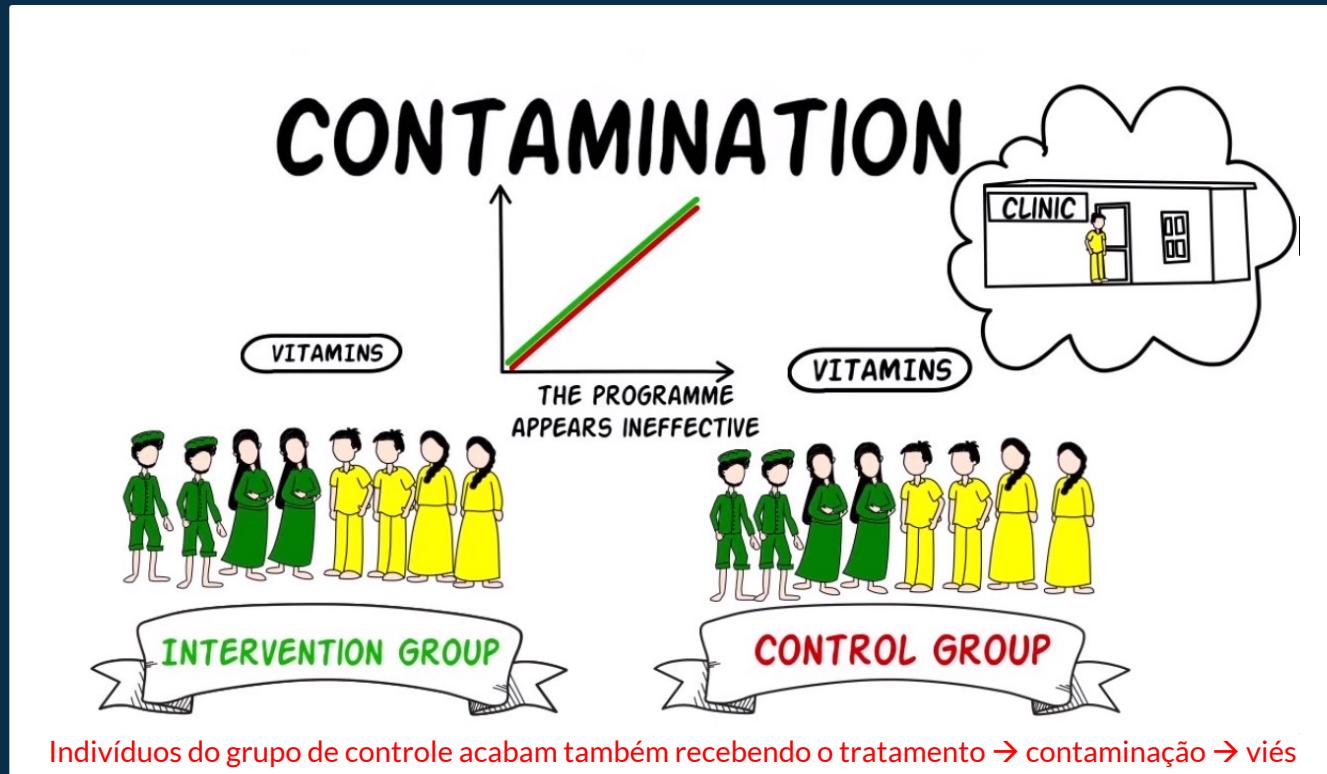
## ATTRITION

... com atrito, os piores indivíduos do grupo de tratamento saíram do estudo antes do término → viés positivo



# Principais hipóteses para a identificação de causalidade

SUTVA – não há contaminação ou transbordamento



# Como estimar o efeito após a aleatorização

- Se a aleatorização for corretamente implementada, os grupos estarão balanceados tanto em características observáveis quanto em características não observáveis
  - Características observáveis → incluídas no vetor de controles  $X$
  - Características não observáveis,  $U$  → relegadas ao termo de erro  $\varepsilon$
- O efeito do tratamento pode ser estimado diretamente por OLS, incluindo controles

$$Y_i = \delta_0 + \delta_1 D_i + X'\beta + \varepsilon_i$$

- Essas covariáveis/controles ajudam a aumentar a precisão da estimativa (eficiência)
- Entretanto, há diferentes medidas de efeito causal a depender de se todos os usuários do grupo de tratamento “consumem” ou não o tratamento. Discutiremos essas medidas na seção de variável instrumental



# 2.4

## EXPERIMENTOS: PODER ESTATÍSTICO DO EXPERIMENTO

---

Análise de poder, tamanho da amostra e clareza dos resultados um experimento



Universidade dos Dados

# Revisitando teste de hipóteses da seção 1.1

Em um teste de hipótese com  $H_0: \beta_1=0$  e  $H_1 : \beta_1 \neq 0$ , desejamos testar se os dados da nossa amostra fornecem evidência suficiente para que afirmemos que  $\beta_1$  (aprox. por  $\hat{\beta}_1$ ) é  $\neq 0$

Ex: no experimento em que desejamos saber se o **efeito de uma campanha** de marketing é estatisticamente diferente de zero, analisamos:

$$\text{compras} = \beta_0 + \beta_1 \text{ Recebeu e-mail} + \varepsilon$$

O p-valor é a probabilidade de observar um efeito tão extremo quanto esse, mas assumindo que na realidade esse efeito não existe.

Portanto, um **p-valor suficientemente baixo** nos traz evidências de que há um efeito estatisticamente diferente de 0, pouco provável de ter acontecido ao acaso.

```
Call:  
lm(formula = compras ~ recebeu_email, data = usuarios)  
  
Residuals:  
    Min      1Q  Median      3Q     Max  
-140.284 -34.275   1.011   32.053 168.139  
  
Coefficients:  
            Estimate Std. Error t value          Pr(>|t|)  
(Intercept) 501.379     2.227 225.14 <0.0000000000000002 ***  
recebeu_email 98.416     3.168 31.06 <0.0000000000000002 ***  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 50.09 on 998 degrees of freedom  
Multiple R-squared:  0.4915,    Adjusted R-squared:  0.491  
F-statistic: 964.8 on 1 and 998 DF,  p-value: < 0.0000000000000002
```

Receber o email causa um aumento médio de  $\hat{\beta}_1 = \text{R\$}98.4$ , estatisticamente significante, já que o p-valor é menor que 0,05

# Como isso se relaciona com poder do experimento?

Experimentar é testar hipóteses!

Se estimarmos a regressão anterior para amostras distintas, provavelmente obteremos que  $\hat{\beta}_1$  e p-valores distintos, então **é possível que cometamos erros.**

- A matriz ao lado mostra os 4 cenários:
  - Dois em que o teste e a realidade estão na mesma direção (**acertos**)
  - dois em que o teste e a realidade apontam em direções distintas (**erros do tipo I e do tipo II**).

|                           |                            | A verdade (“que Deus conhece”)   |  |
|---------------------------|----------------------------|--|--|
|                           |                            | H <sub>0</sub> é falsa   | H <sub>0</sub> é verdadeira  |
| Seu experimento (amostra) | Rejeita H <sub>0</sub>     | <span style="color: green;">✓</span><br>Decisão correta<br>(Poder)<br>(1- $\beta$ )      | <span style="color: red;">✗</span><br>Falso positivo<br>(Erro do tipo I)<br>( $\alpha$ ) |
|                           | Não rejeita H <sub>0</sub> | <span style="color: red;">✗</span><br>Falso negativo<br>(Erro do tipo II)<br>( $\beta$ ) | <span style="color: green;">✓</span><br>Decisão correta                                  |

# Como isso se relaciona com poder do experimento?

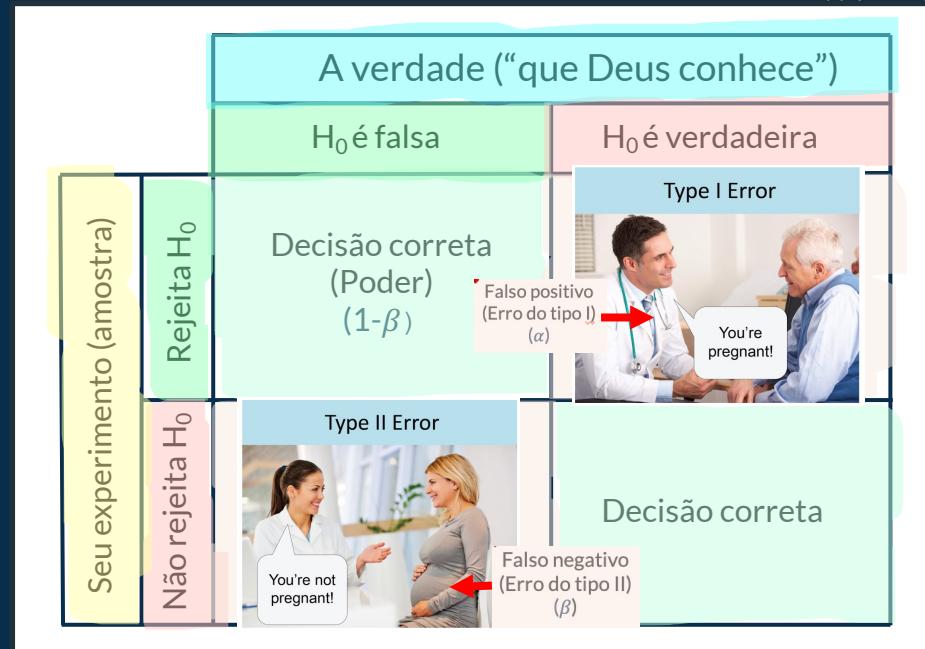
**Poder do Teste ( $1-\beta$ ):** Tipicamente igual a 80%. Isso significa que há uma chance de 20% de *não detectar um efeito real* (Erro Tipo II).

- Para aumentar a chance de detectar um efeito real (ex:  $1-\beta = 95\%$  para evitar um Erro Tipo II) é necessário ter um tamanho de amostra maior.

**Nível de Significância ( $\alpha$ ):** Tipicamente fixada em 5%.

- Aumentar o nível de significância ( $\alpha$ ) torna mais provável rejeitar a hipótese nula verdadeira (erro tipo I). Diminuí-lo, diminui a probabilidade de erro do tipo I e aumenta N necessário para um dado poder

Há um trade-off entre o risco de cometer os dois tipos de erro: ao tornar o teste mais rigoroso (diminuindo  $\alpha$ ), aumentamos a chance de não detectar efeitos reais (erro tipo II), e vice-versa.



# Elementos para a análise de poder

---

Os principais elementos para a análise de poder são:

- O efeito mínimo detectável (MDE), que pode ser traduzido em “tamanho do efeito” (effect size, Cohen’s D, etc)
- O tamanho da amostra (N)
- O Nível de Significância ( $\alpha$ ): probabilidade de cometer erro do tipo I
- O poder do teste ( $1-\beta$ ) é a probabilidade de detectar um efeito real e evitar um erro do tipo II.



Você precisa de três desses elementos para poder encontrar o quarto elemento , como em uma regra de três.

Geralmente escolhemos  $(1-\beta) = 0.8$  ou 80%,  $\alpha = 0.5$  ou 5% e estimamos um MDE baseado em dados passados, benchmark de mercado ou simplesmente uma hipótese conservadora (veja [aqui](#))

**Mas o que é o MDE?**

É o menor *tamanho de efeito* que seu experimento será capaz de detectar com confiança.

- Ex: se você precisa detectar um efeito do tratamento de 2% sobre a variável dependente, o tamanho amostral necessário será muito maior do que se o efeito esperado for de 20%.

# EFEITO MÍNIMO DETECTÁVEL

Se o MDE for muito alto, o estudo pode falhar em detectar efeitos menores, mas ainda relevantes, resultando em um aumento da probabilidade de erro tipo II (falso negativo).

Para podermos reduzir o MDE e detectar efeitos menores, é necessário aumentar o tamanho da amostra, o que pode ser custoso ou inviável.

A definição de um MDE apropriado é, portanto, um equilíbrio entre a sensibilidade estatística e a viabilidade prática do estudo.

**EXEMPLO:** você quer testar se uma nova campanha aumenta as vendas em pelo menos 5%.

O MDE funciona como uma "lupa": se o verdadeiro aumento for menor que 5%, você provavelmente não conseguirá detectá-lo com confiança, mesmo que ele exista.

**Why 5% should be the upper bound of your MDE in A/B tests**



Ron Kohavi

Vice President and Technical Fellow | Data Science,  
Engineering | AI, Machine Learning, Controlled Experiment...



November 6, 2023

# Efeito mínimo detectável e Cohen's D

## Exemplo para variável contínua

Nos softwares estatísticos, o geralmente expressamos o MDE em termos de efeito padronizado, também conhecido como effect size.

Uma fórmula usual para o effect size é o Cohen's D

$$D = \frac{\text{Média y pós tratamento} - \text{Média y pós controle}}{\text{D. P. conjunto}}$$

Onde

$$\text{D. P. conjunto} = \frac{\text{D.P. trat.} * (\text{N}_t - 1) + \text{D.P. contr.} * (\text{N}_c - 1)}{\text{N}_t + \text{N}_c - 2}$$

| <b>Effect size</b> | <b>d</b> | <b>Reference</b>     |
|--------------------|----------|----------------------|
| Very small         | 0.01     | <a href="#">[11]</a> |
| Small              | 0.20     | <a href="#">[10]</a> |
| Medium             | 0.50     | <a href="#">[10]</a> |
| Large              | 0.80     | <a href="#">[10]</a> |
| Very large         | 1.20     | <a href="#">[11]</a> |
| Huge               | 2.0      | <a href="#">[11]</a> |

# Exemplo no R (veja o script)

---

Queremos fazer um teste A/B para medir o impacto de um novo sistema de recomendação de itens sobre valor das compras.

Como dados históricos, sabemos que a média de compras dos clientes que farão parte do experimento é de R\$ 150

Esperamos que o novo sistema de recomendação aumente esse valor em 10%, que levaria os clientes tratados a um patamar médio de compras de R\$165. O desvio padrão combinado é de R\$45.

Calculando o Cohen's D, temos o seguinte resultado. Qual tamanho de amostra precisamos para detectar esse efeito, com poder do teste de 80% e significância de 5?

```
> (effect_size = (165 - 150) / 45)
[1] 0.3333333
> interpret_cohens_d(effect_size, rules = "cohen1988")
[1] "small"
(Rules: cohen1988)
```

# Exemplo no R (veja o script)

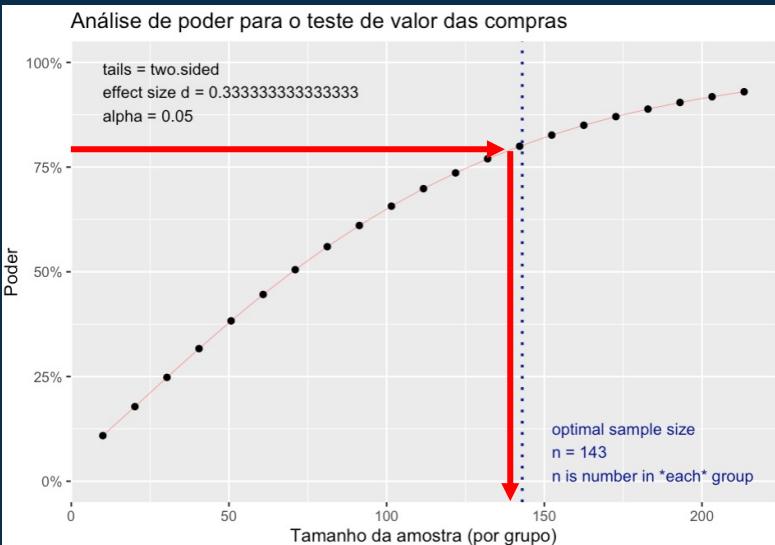
P: Qual tamanho de amostra precisamos para detectar esse efeito, com poder do teste de 80% e significância de 5?

R: precisamos de 143 clientes em cada grupo, totalizando 286 clientes, para ter um poder de 80%

```
effect_size = (165 - 150) / 45
interpret_cohens_d(effect_size, rules = "cohen1988")

# Calculate required sample size
purchase_test <- pwr.t.test(
  d = purchase_effect_size,
  power = 0.80,
  sig.level = 0.05,
  type = "two.sample",
  alternative = "two.sided"
)

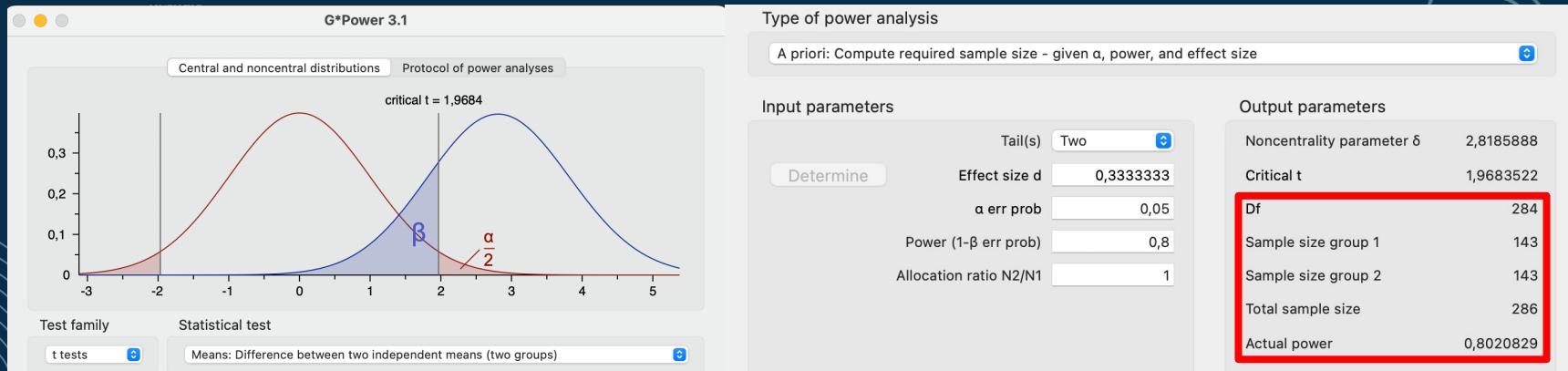
# Plot power curve
plot(
  purchase_test,
  main = "Análise de poder para o teste de valor das compras",
  xlab = "Tamanho da amostra (por grupo)",
  ylab = "Poder"
)
```

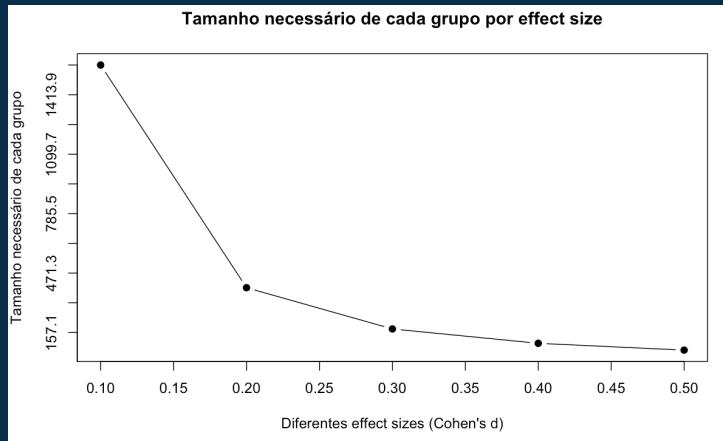


# Exemplo no G\*Power calculadora amostral

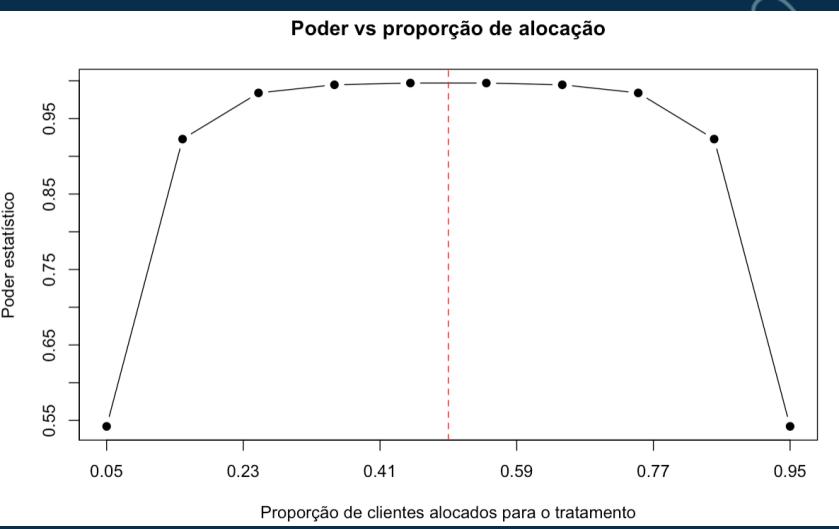
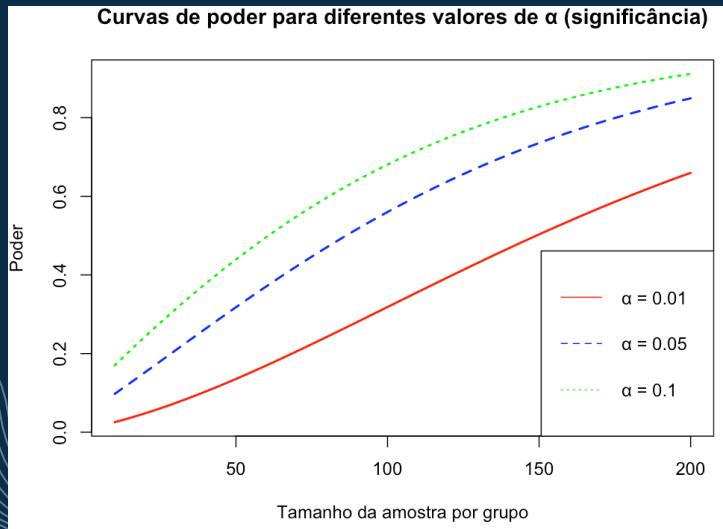
P: Qual tamanho de amostra precisamos para detectar esse efeito, com poder do teste de 80% e significância de 5?

R: precisamos de 143 clientes em cada grupo, totalizando 286 clientes, para ter um poder de 80%



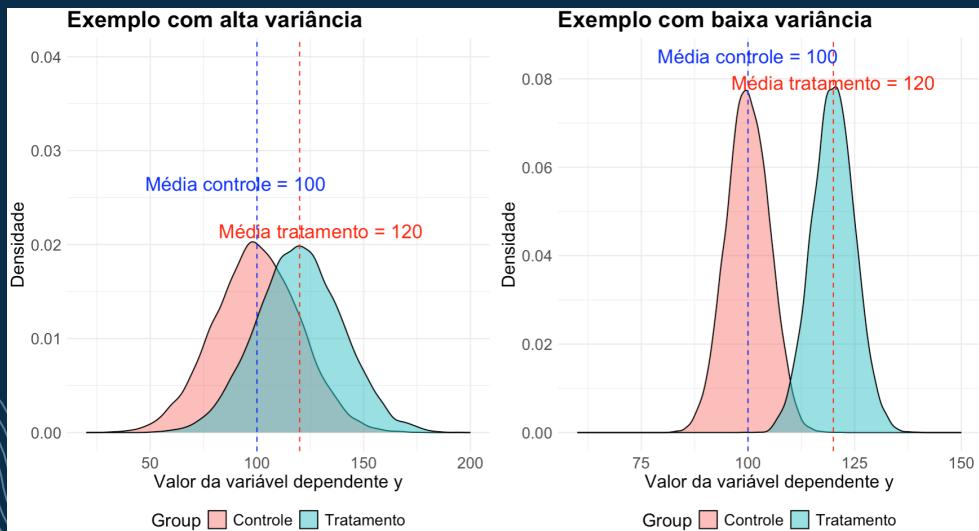


- Quanto maior o effect size, menor N precisamos
- Quanto menor a significância, mais N precisamos
- O poder diminui se uma proporção muito alta ou muito baixa de pessoas for alocada para o tratamento



# Alguns elementos que afetam o poder de um experimento

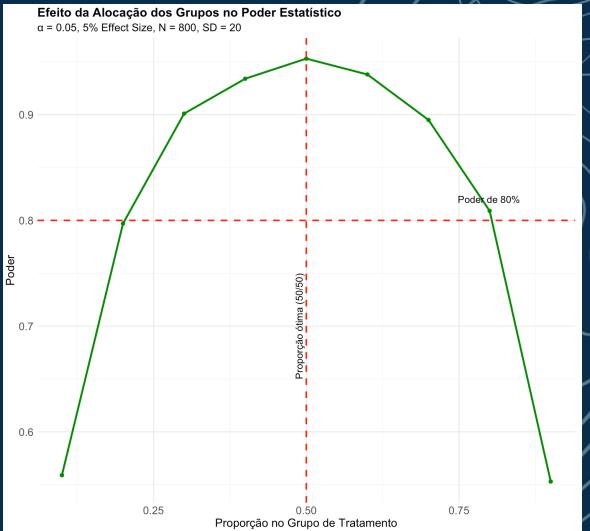
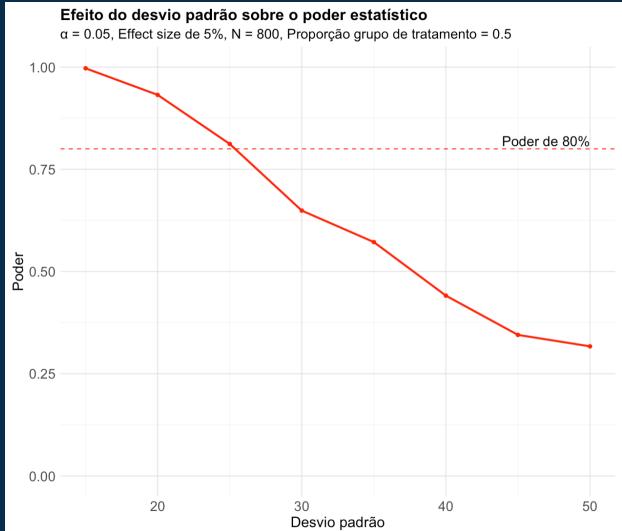
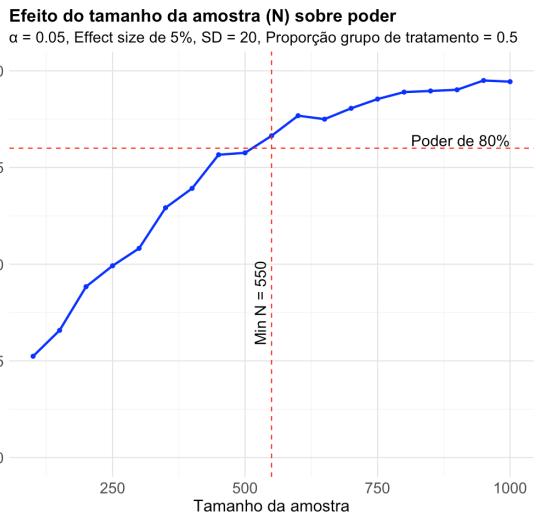
- **Variância da variável dependente:** Refere-se à variabilidade na métrica que você está avaliando (por exemplo, variância do gasto com delivery). Alguns clientes gastam muuuuito e outros gastam bem pouco).
  - Quanto maior a variância, maior será o tamanho da amostra necessário para distinguir entre variabilidade aleatória (ruído) e mudanças causadas pelo tratamento (sinal).



- Na prática, aconselha-se aplicar winsorization ou excluir outliers em experimentos, para evitar que alta variância afete a capacidade de detectar efeitos que realmente existem
- Também é possível usar variáveis com menor variância como “fez compra” (ao invés de R\$ gastos)

# Exemplo com parâmetros fixos – não é assim em todos os casos!

## Veja os scripts e experimente mudar os parâmetros



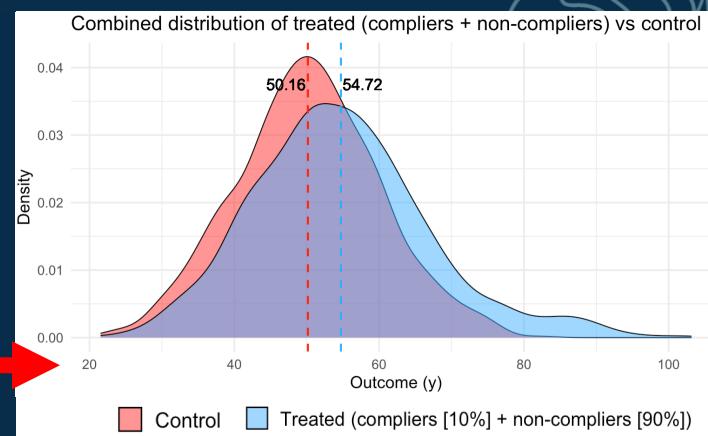
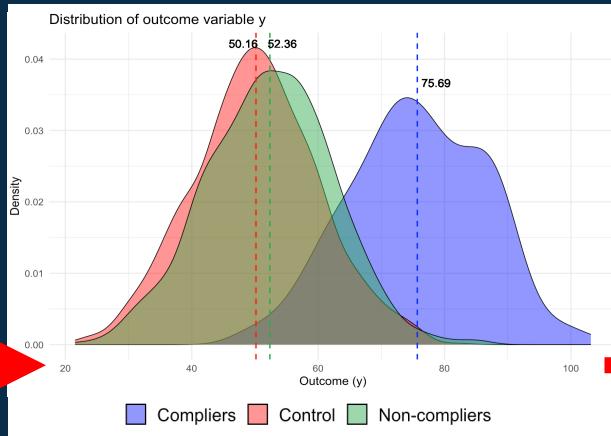
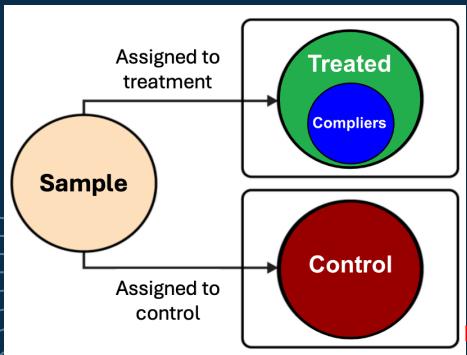
Quanto maior o N, tudo mais constante (SD,  $1-\beta$ ,  $\alpha$ , proporção, etc.) maior tende a ser o poder

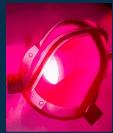
Quanto maior a variância ou SD do Y, tudo mais constante (N,  $1-\beta$ ,  $\alpha$ , proporção, etc) menor tende a ser o poder

O maior poder tende a ser atingido com 50% dos indivíduos no grupo tratamento e 50% no grupo controle, tudo mais constante (N, SD,  $1-\beta$ ,  $\alpha$ , proporção, etc)

# Alguns elementos que afetam o poder de um experimento

- Taxa de adesão (compliance rate): Há participantes que não cumprem ou consomem o tratamento? Uma taxa de adesão mais baixa exige um tamanho de amostra maior. Isso ocorre porque o efeito do tratamento naqueles que usaram será diluído no grupo de tratamento.
- No exemplo, o efeito ocorre apenas para os que aderiram ao tratamento (azul) e pode ser medido pelo efeito LATE. No entanto, com uma adesão baixa (apenas 10%), o efeito não será detectado se compararmos todo o grupo tratado com todo o grupo controle.





# PRINCIPAIS TAKEAWAYS



- Tamanho amostral adequado: Calcular o N necessário via análise de poder; Considerar efeito mínimo detectável de interesse; Quanto menor o efeito esperado, maior a amostra necessária
- Simplificar o design: Minimizar número de grupos de tratamento; Mais grupos = menor poder para cada comparação; Focar nas comparações mais importantes.
- Randomizar no menor nível possível; Usar designs que maximizem compliance (adesão ao tratamento); Minimizar attrition (perda amostral).
- Monitorar implementação continuamente; Documentar desvios do plano original
- Alocação balanceada dos grupos: Divisão igual (50%/50%) maximiza o poder estatístico e ajuda a detectar efeitos pequenos com confiança.

**2.5**

## **EXPERIMENTOS: PASSO A PASSO**

---

Passos gerais de um experimento  
(checklist)



**Universidade dos Dados**

# PASSOS GERAIS DE UM EXPERIMENTO

---

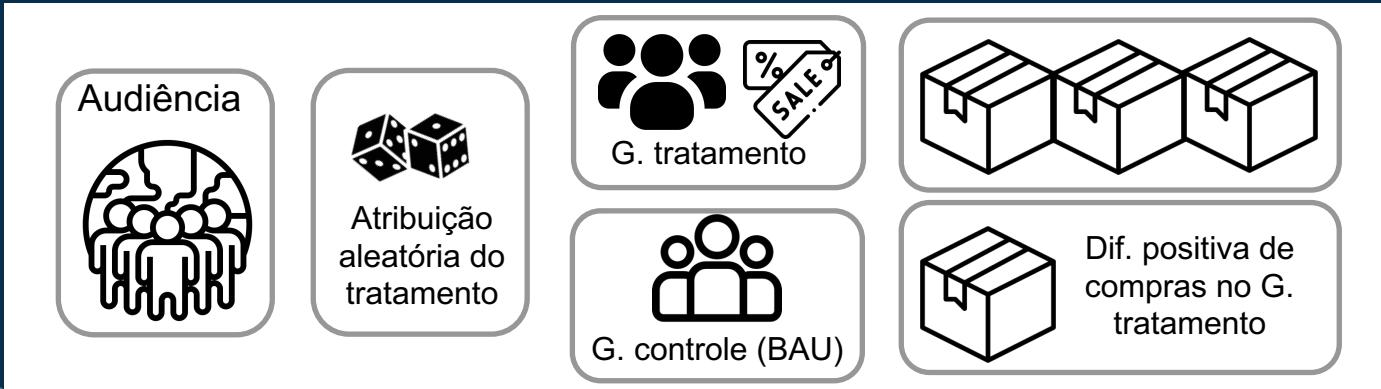
Esta é uma versão simplificada de como conduzir um experimento:

- **Formular a hipótese:** Por exemplo, comprar suplementos nutricionais uma vez aumenta o valor/frequência geral de compras subsequentes em uma rede de farmácias (compra recorrente)?
  - Alguns produtos, como vitaminas, colágeno ou whey protein, podem levar a compras recorrentes devido à formação de hábitos, com efeitos spillover para outros produtos.
  - Avalie se essa **hipótese é respaldada por dados observacionais internos, pesquisas de mercado ou outras evidências**, em vez de ser apenas uma "Opinião da Pessoa Mais Bem Paga" (HiPPO).
  - Também é fundamental **considerar se a hipótese é açãoável** e relevante para o negócio. Pode não fazer sentido testar a eficácia de alavancas que não podemos utilizar.



# PASSOS GERAIS DE UM EXPERIMENTO

- **Identificar as métricas de interesse:** Determine as métricas a serem analisadas, como (1) frequência de compras de outros produtos, (2) valor total das compras mensais, (3) frequência de acessos ao app/site, etc. É importante garantir que essas métricas sejam mensuráveis com precisão e estejam diretamente relacionadas à hipótese.
- **Desenhar a intervenção:** Crie um mecanismo para alterar exogenousamente a variável de interesse, por exemplo, enviando notificações push com ofertas ou descontos em suplementos nutricionais.



# PASSOS GERAIS DE UM EXPERIMENTO

---

Estabeleça as suposições prévias (“priors”):

- Estime a taxa de adesão/compliance - participação esperada de clientes que irão responder à notificação e realizar uma compra. Por exemplo, podemos esperar que 5% dos clientes respondam à notificação e comecem a fazer compras online.
- Estime a magnitude do efeito nas compras para esses clientes (por exemplo, podemos antecipar um aumento de 6% no valor total mensal das compras para os clientes que começaram a fazer compras online devido à nossa campanha).
- Determine o efeito mínimo detectável: Com base nesses parâmetros (taxa de conversão de 5% e efeito nos aderentes de 6%), podemos estabelecer o menor tamanho de efeito que o experimento foi projetado para detectar de maneira estatisticamente significativa.
  - Se o MDE for muito alto, o estudo pode falhar em detectar efeitos menores, mas relevantes, resultando em erro tipo II (falso negativo).



# PASSOS GERAIS DE UM EXPERIMENTO

---

- **Definir nível de significância e poder:** Escolha um nível de significância (a probabilidade de concluir que há um efeito quando na verdade não existe - Erro Tipo I -  $\alpha = 0,05$ ) e um poder estatístico (a probabilidade de rejeitar corretamente uma hipótese nula falsa -  $(1 - \beta) = 0,8$ ).
  - Note que a escolha desses parâmetros deve ser informada pelo contexto do experimento e pelo custo potencial de cada um desses tipos de erro.
- **Calcular o tamanho da amostra:** Esta etapa é importante para que o experimento tenha poder estatístico suficiente para detectar o efeito, caso ele exista. Realize simulações para determinar o tamanho mínimo de amostra necessário, considerando o efeito esperado, o nível de significância e o poder desejado.
- **Selecionar a amostra:** Escolha uma amostra de usuários ativos, possivelmente segmentando clientes específicos (por exemplo, aqueles que já buscaram por suplementos nutricionais).
- **Randomizar a atribuição do tratamento:** Atribua aleatoriamente indivíduos para receber o tratamento ou fazer parte do grupo de controle. Isso é crucial para reduzir vieses e garantir a validade do experimento.



# PASSOS GERAIS DE UM EXPERIMENTO

**Medições de linha de base (baseline):** Coletar dados de linha de base para entender as condições iniciais antes da intervenção, o que será crucial para a análise comparativa.

- Isso é importante para garantir que **ambos os grupos eram semelhantes antes do experimento** e também para aumentar a precisão das estimativas ou até mesmo para controlar diferenças pré-existentes entre os grupos de tratamento e controle após a intervenção.

| Covariate        | Mean<br>Treatment | Mean<br>Control | p-value |
|------------------|-------------------|-----------------|---------|
| Pre-experiment y | 249,51            | 248,92          | 0,53    |
| Age              | 35,24             | 35,43           | 0,891   |
| Income           | 5045,65           | 5015,09         | 0,827   |
| Years of study   | 10,52             | 10,01           | 0,069*  |
| ...              | ...               | ...             | ...     |
| Observations     | 2734              | 2735            |         |



# PASSOS GERAIS DE UM EXPERIMENTO

---

**Fidelidade, análise intermediária e regras de interrupção:** Desenvolver um plano para avaliar a “fidelidade” da implementação da intervenção. Ou seja, certificar-se de que a intervenção seja entregue conforme planejado.

- Use software para acompanhar se os participantes recebem as notificações e se estão navegando nas seções pretendidas do aplicativo (por exemplo, <https://amplitude.com/>).
- Certifique-se de que todos os membros da equipe que interagem com os participantes (como os responsáveis pelo relacionamento com o cliente e suporte ao cliente) estejam treinados para entregar o programa de maneira consistente.
- Defina critérios para análise intermediária e potenciais regras de interrupção em caso de resultados inesperados ou efeitos adversos (pesquise sobre métricas de segurança ou “guardrail metrics”).

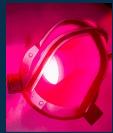


# PASSOS GERAIS DE UM EXPERIMENTO

---

- **Implementar o experimento:** Implemente o experimento com cuidado, evitando qualquer tipo de interferências. Por exemplo, certifique-se de que as mensagens sejam enviadas nos horários planejados e apenas para os participantes definidos, evitando envios acidentais para clientes fora do experimento.
- **Monitorar o experimento:** Utilize ferramentas analíticas, como dashboards, para acompanhar o comportamento dos indivíduos do experimento ao longo do experimento.
- **Avaliação pós-experimento:** Após o experimento, avalie seu sucesso, identifique lições aprendidas e documente recomendações para futuros experimentos. É crucial avaliar não apenas os resultados, mas também o processo do experimento para orientar futuras pesquisas.
- **Lembre-se:** *O conhecimento adquirido é inestimável, mesmo que o experimento "refute" sua hipótese inicial.*





## Lição desta parte

- O viés de variável omitida ocorre quando deixamos de incluir no modelo uma variável que está correlacionada tanto com o resultado ( $Y$ ) quanto com o tratamento ( $X$ ).
- O método experimental usa a exogeneidade (independência entre receber o tratamento e as características individuais) para corrigir o viés de variável omitida.
- Aplicar um experimento por si só não garante causalidade. É necessário atender a certas hipóteses e ter poder estatístico suficiente, caso contrário, podemos obter estimativas enviesadas ou não detectar efeitos reais.
- Em alguns experimentos, parte do grupo de tratamento não utiliza o tratamento. Por isso, precisamos ajustar as estimativas de ITT para LATE (utilizando variável instrumental).



**3.1**

# **VARIÁVEL INSTRUMENTAL**

---

Motivação e princípios básicos



**Universidade dos Dados**

# Introdução

---

Como vimos, uma forma de nos livrarmos do viés de variável omitida é através da aleatorização do tratamento:  $E[\nu|D] = 0$ , na regressão  $Y = \beta_0 + \beta_1 D + \nu$

- Outras formas envolvem encontrar uma fonte de variação em  $D$  que não seja fortemente correlacionada com as variáveis omitidas.
- Nem sempre a aleatorização é uma opção factível (ética, custo, etc.) e nem sempre existe uma regra de atribuição clara para o tratamento.



Ou mesmo na aleatorização, não podemos forçar os indivíduos do grupo de tratamento a consumirem esse tratamento (usar cupons de descontos, free trial, tomar um medicamento, etc)

- Uma saída é encontrar uma terceira variável que seja fortemente correlacionada com  $D$ , mas não com as variáveis omitidas em  $\nu$  ...



# Introdução

---

- ... *Essa variável deve afetar Y somente através de D.* A ela se dá o nome de *variável instrumental (IV)*.
- Um IV pode se encaixar no contexto de experimentos elaborados por um pesquisador (e.g., RCTs) ou no contexto de “experimentos naturais”.
- Vamos começar explicando IV no contexto de experimentos aleatorizados...



# Introdução

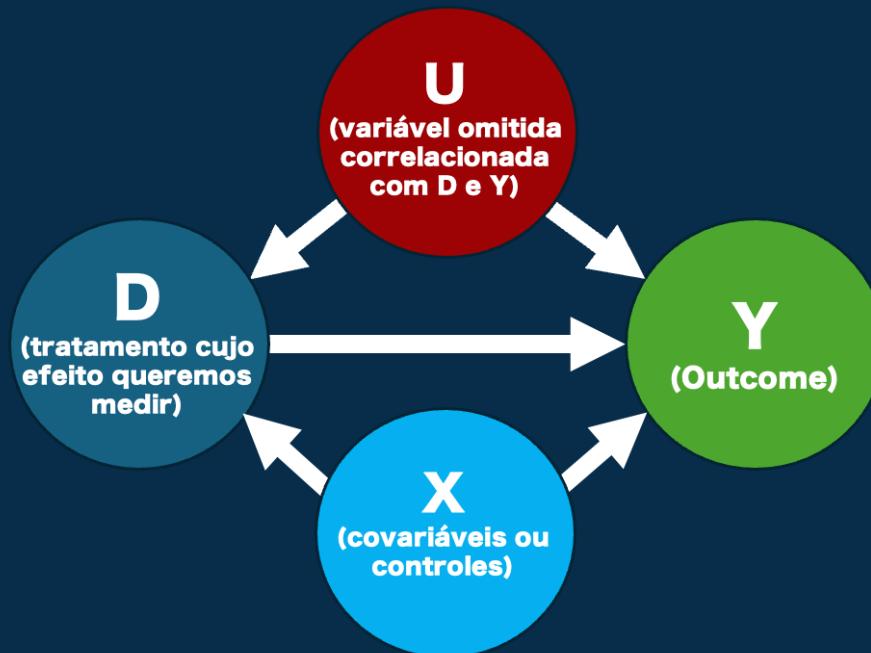
- Em experimentos, assumimos anteriormente que a instituição que implementa o programa consegue designar indivíduos aos grupos de tratamento e controle **de modo que esses indivíduos obedecerão perfeitamente**, como no exemplo de LaLonde da seção passada.
  - i.e., assumimos conformidade perfeita ou “*perfect compliance*”: Todos que recebem oferta para participar do programa assim o farão. Ninguém que não recebe oferta para participar irá participar.
- O método variável instrumental (IV) pode nos ajudar a avaliar programas com *imperfect compliance*, como matrículas voluntárias ou coberturas universais.
- Para estimar o efeito do tratamento, o método IV depende de alguma fonte externa de variação que influencie o status de tratamento.
  - Podemos pensar num IV como uma dimensão que está fora do controle do indivíduo, mas influencia sua probabilidade de participar no programa.



# Introdução

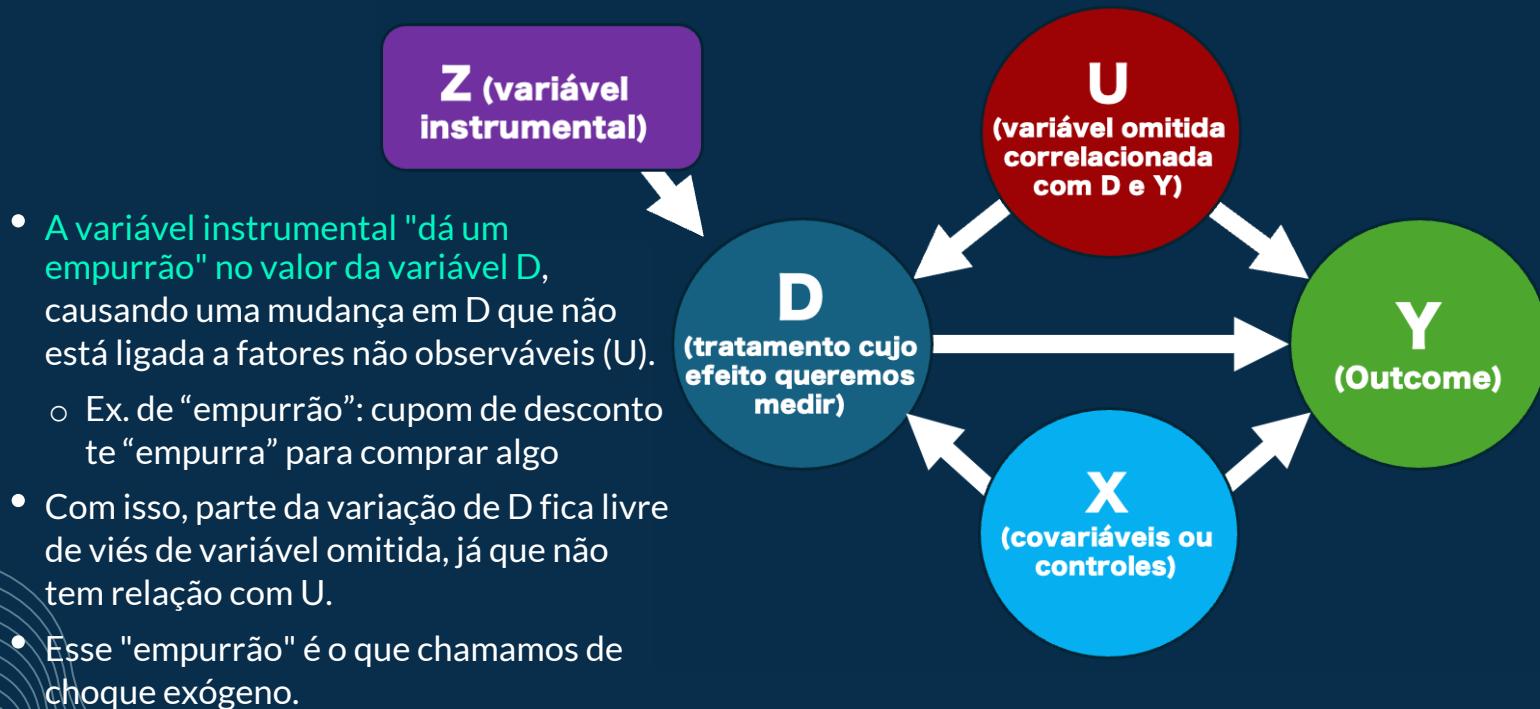
- As setas brancas indicam a direção em que uma variável afeta a outra
- A variável U pode ser vista como um fator que não observamos diretamente (ex: motivação das pessoas)
- Neste caso, considera-se que U está dentro de  $\varepsilon$  no modelo de regressão acima.
- Como U afeta tanto D quanto Y, estimar  $\beta_1$  neste modelo levaria a resultados viesados.

$$y = \beta_0 + \beta_1 D + \beta_2 X + \varepsilon$$



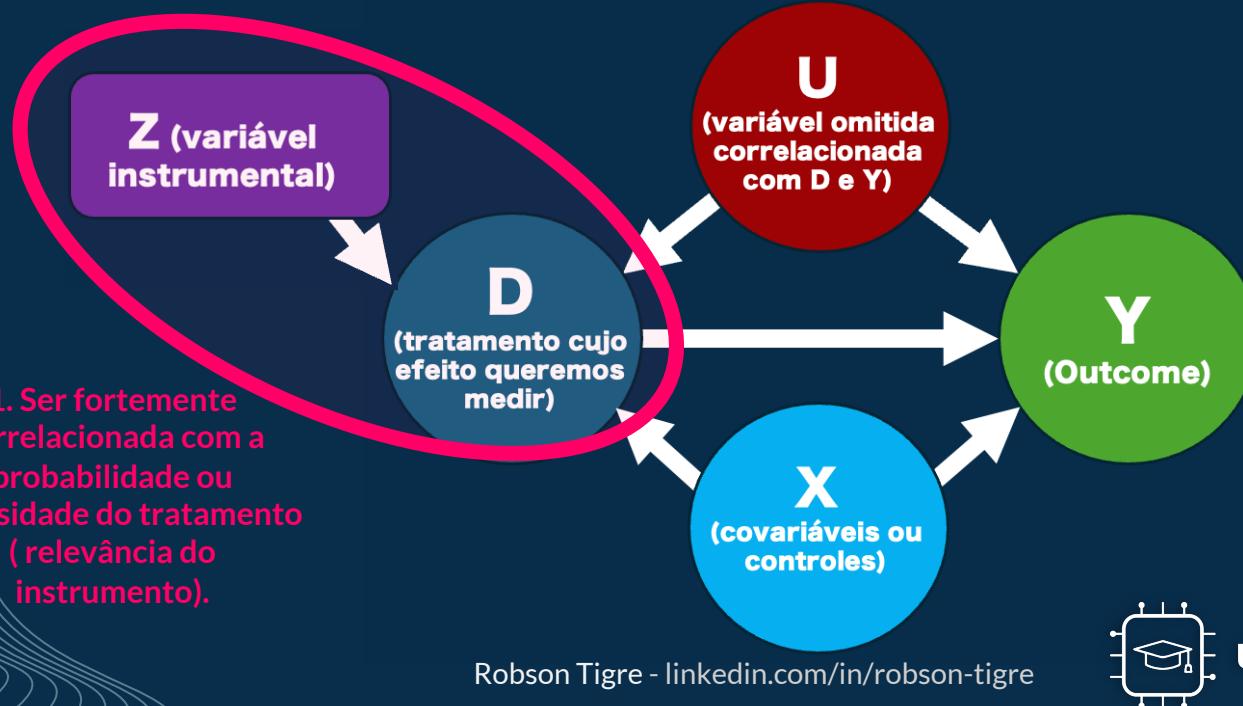
# Introdução

$$y = \beta_0 + \beta_1 D + \beta_2 X + \varepsilon$$



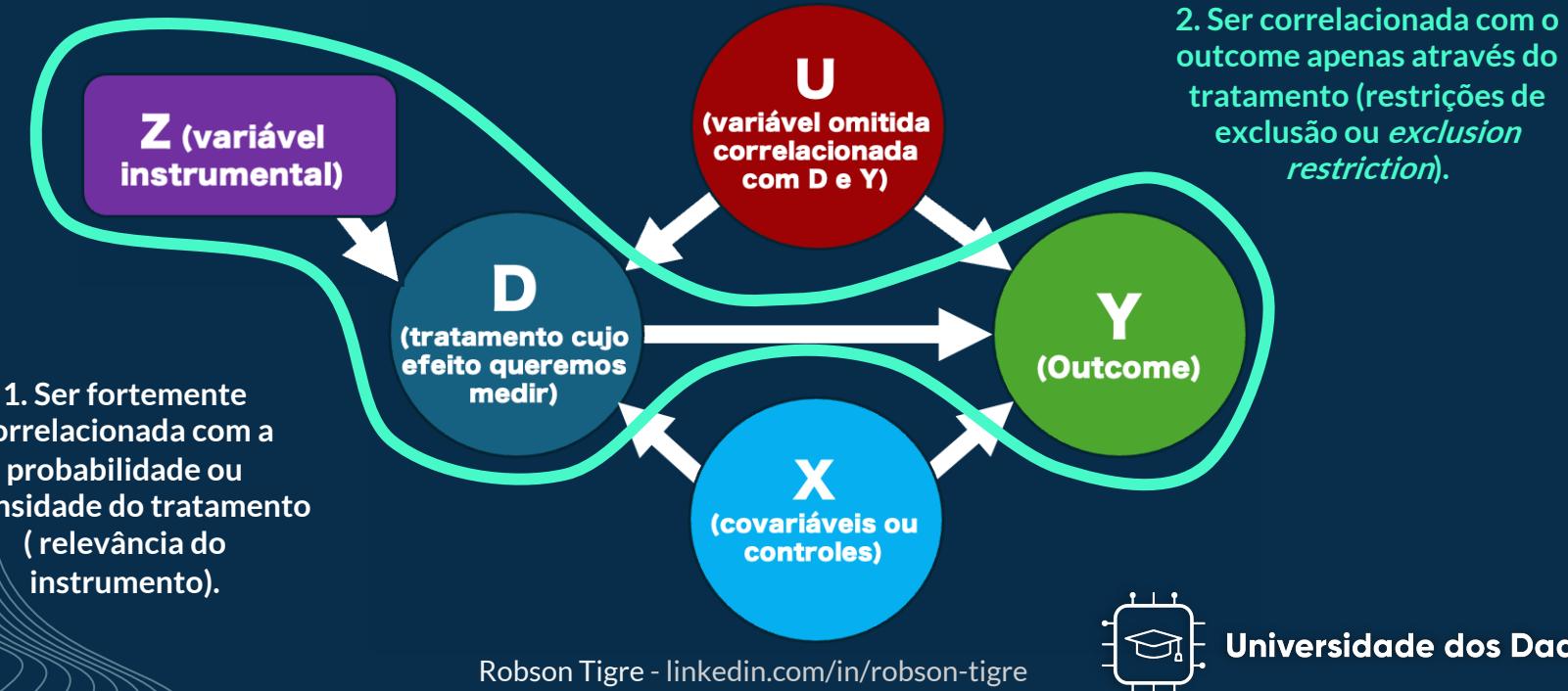
# Premissas para o bom funcionamento da VI

Para produzir estimativas válidas do efeito do tratamento, essa fonte de variação externa deve satisfazer algumas condições:



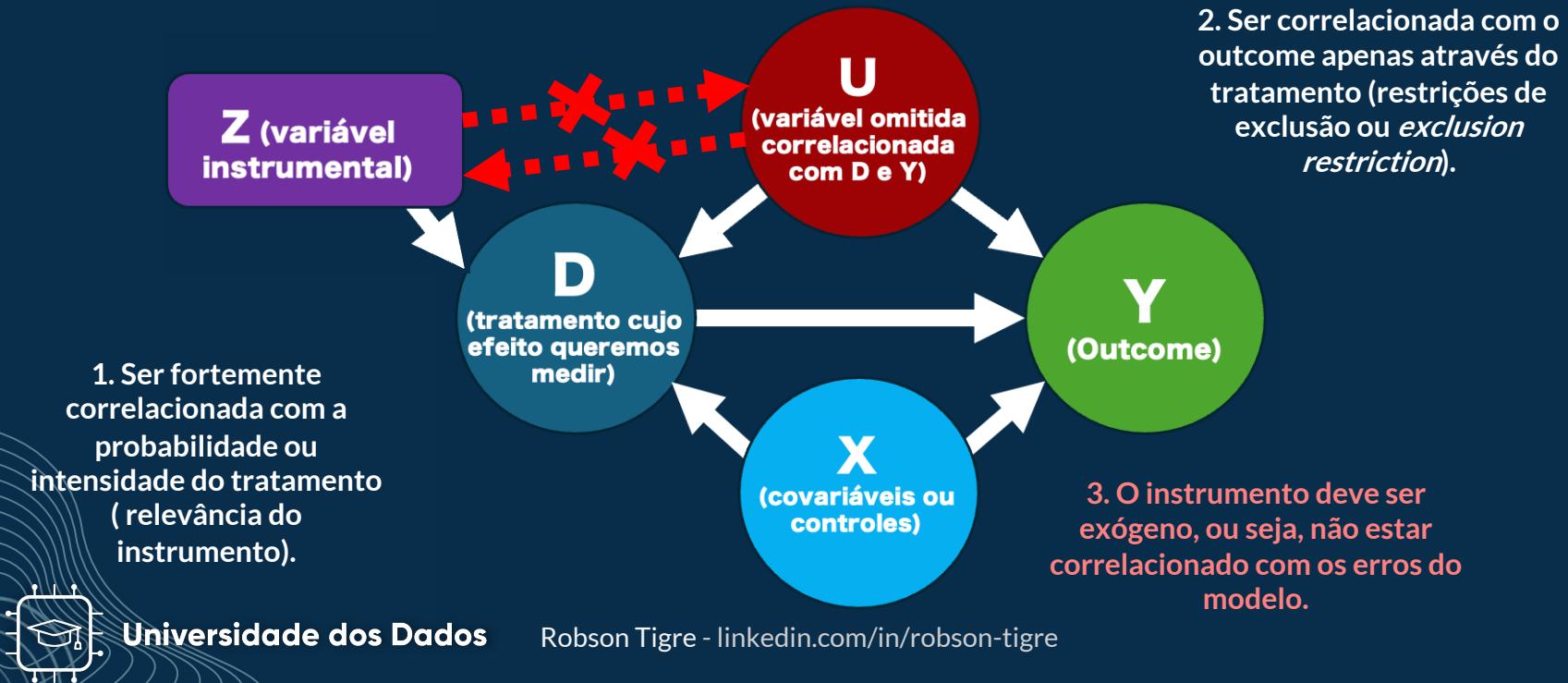
# Premissas para o bom funcionamento da VI

Para produzir estimativas válidas do efeito do tratamento, essa fonte de variação externa deve satisfazer algumas condições:



# Premissas para o bom funcionamento da VI

Para produzir estimativas válidas do efeito do tratamento, essa fonte de variação externa deve satisfazer algumas condições:



# Premissas para o bom funcionamento da VI

- **Pressuposto de monotonicidade:** significa que o instrumento afeta o tratamento D (a variável endógena) sempre na mesma direção para todas as unidades
  - Ou seja, se o instrumento aumenta, ele ou sempre aumenta a probabilidade de tratamento ou sempre diminui, mas não pode aumentar para alguns e diminuir para outros.
  - Se  $Z \uparrow \rightarrow D \uparrow$  para algumas unidades, não pode haver outras onde  $Z \uparrow \rightarrow D \downarrow$
- Consequências da violação:
  - O estimador de variáveis instrumentais (IV) perde consistência
  - O efeito estimado não pode ser interpretado como um efeito causal local médio (LATE)
  - Pode haver viés em qualquer direção
- Na prática, monotonicidade é um pressuposto forte e difícil de testar formalmente. Por isso é importante



## 3.2

# VARIÁVEL INSTRUMENTAL: ILUSTRAÇÃO

Encaixando a aplicação deste método  
em uma contexto específico



Universidade dos Dados

# Revisando duas medidas de efeito causal

💡 Um instrumento ideal é a atribuição aleatorizada do tratamento, que satisfaz muito bem as condições acima.

*Exemplo: "Uma empresa de e-commerce quer avaliar se o uso de cupons de desconto aumenta as compras futuras dos clientes.*

*Para contornar o viés de seleção, cupons foram enviados aleatoriamente para um grupo de clientes, mas nem todos os recipientes os usaram. Além disso, os clientes que não receberam o cupom não tiveram acesso a ele."*

**Intention to treat (ITT):** Mede o efeito de ser designado ao tratamento, *independentemente de os indivíduos seguirem a designação ou não.*

**Local average treatment effect (LATE):** Aqui utilizamos o método de variáveis instrumentais (IV) para estimar o local average treatment effect, que é o efeito causal do tratamento naqueles que o consumiram ("compliers") – ou seja, aqueles que usaram o cupom porque foram designados a recebê-lo.

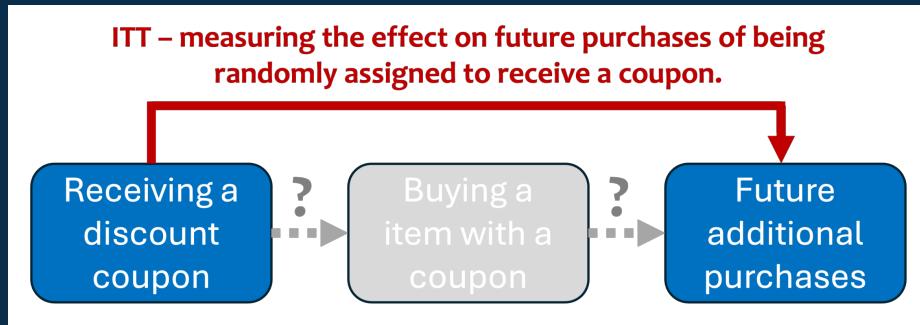


# Efeito da intenção de tratar (não necessita VI)

Conceito importante de adesão ou compliance = % de pessoas no grupo de tratamento que realmente *utilizaram o tratamento*. A adesão imperfeita ou imperfect compliance significa que menos de 100% o fizeram.

**ITT (efeito da intenção de tratar):** Compara todos os indivíduos designados para o grupo de tratamento com todos designados para o grupo de controle, **independentemente da adesão** (ou seja, de comprar ou não usando o cupom).

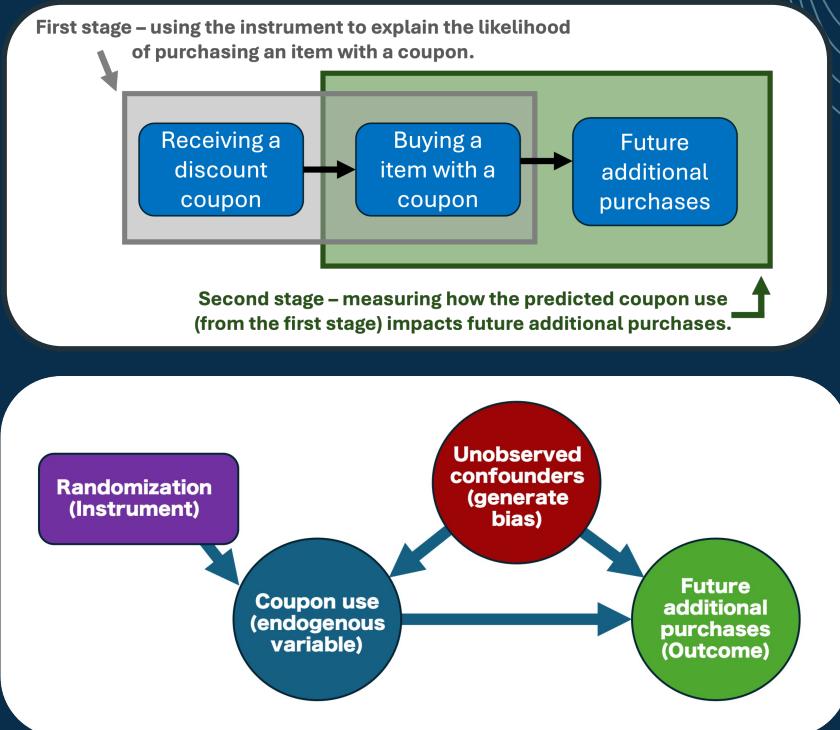
- Este estimador preserva a randomização, mas é conservador, pois **dilui o efeito** sobre aqueles que compraram usando o cupom pelo efeito nulo daqueles que receberam o desconto, mas não compraram usando o cupom.



# Efeito do tratamento em quem usou o cupom

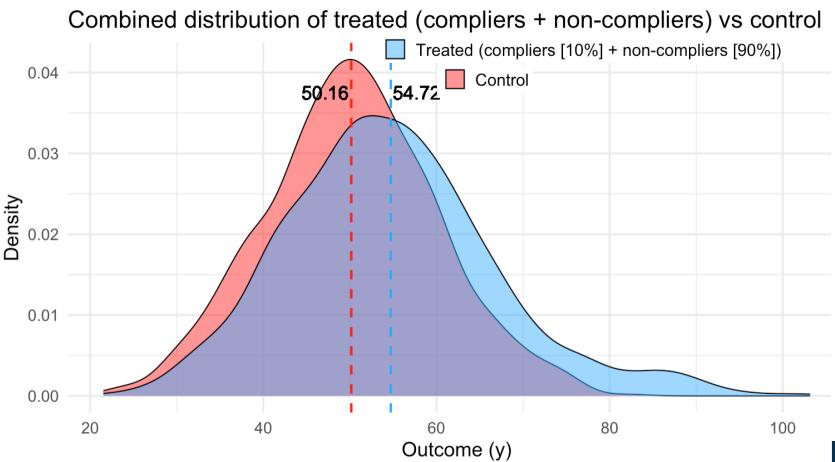
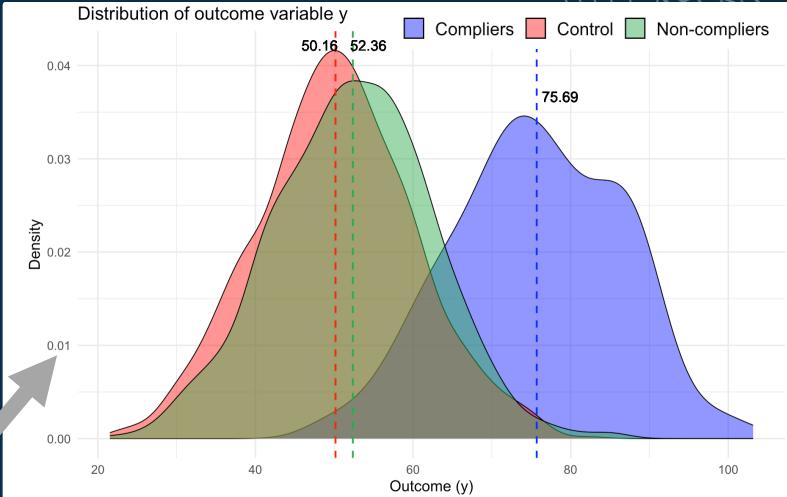
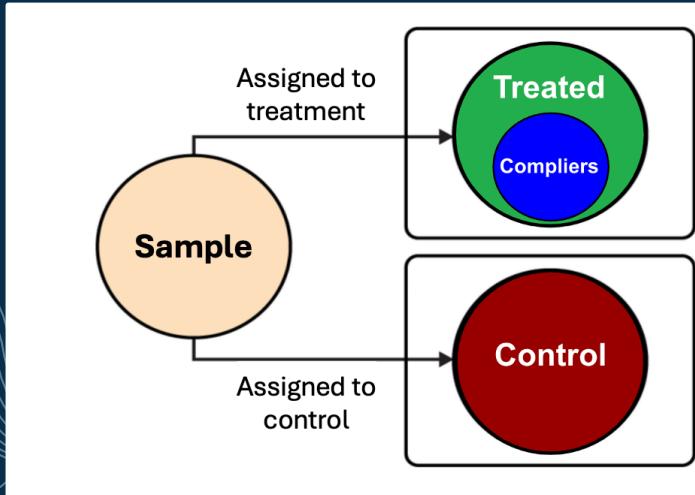
LATE (Efeito médio local do tratamento): Foca no efeito para aqueles que realmente utilizam o tratamento quando ele foi oferecido. É especialmente relevante quando há adesão imperfeita, e você deseja **isolar o efeito apenas no subgrupo que seguiu o tratamento porque ele foi dado exogenamente**.

- No nosso caso, o LATE mediria o impacto sobre gastos futuros após o cliente usar um cupom de desconto, mas somente para os clientes que de fato utilizaram o cupom quando ele foi oferecido. Isso ajuda a isolar o impacto no grupo que respondeu à intervenção.
- Veremos como estimar o LATE usando a aleatorização como variável instrumental



# Observação: a taxa de adesão afeta o poder do experimento

- Neste exemplo o efeito causal existe para os que aderiram ao tratamento (azul) e pode ser detectado estimando o efeito médio local do tratamento (LATE).
- No entanto, devido à baixa adesão (só 10% dos tratados), se considerarmos apenas o efeito de intenção de tratar (ITT), não seremos capazes de detectar nenhum efeito.



# Premissas para o bom funcionamento da VI

O método VI caracteriza uma *reação em cadeia*, partindo do instrumento  $Z$ , passando pelo tratamento  $D$  e chegando até o outcome  $Y$ .

- Primeiro estágio: conecta designação aleatorizada do tratamento, i.e., ter recebido o push com o cupom, à compra com o uso do cupom. O IV afeta exogenousamente o uso tratamento.
- Segundo estágio: conecta a a compra, *exogenamente afetada pelo IV*, ao gasto futuro  $Y$ .
- As hipóteses de independência e *exclusion restriction* nos permitem escrever o efeito da oferta de cupons sobre compras futuras como o produto dessas duas etapas:

$$\begin{aligned} & \text{Efeito de receber o cupom sobre as compras futuras} = \\ & \text{Ef. receber o cupom sobre } \Pr(\text{uso do cupom}) \times \text{Ef. uso do cupom sobre compras futuro} \end{aligned}$$

Rearranjando, temos:

$$\text{Ef. uso do cupom sobre compras futuras} = \frac{\text{Efeito de receber o cupom sobre as compras futuras}}{\text{Ef. receber o cupom sobre } \Pr(\text{uso do cupom})}$$

Chamamos este termo de “efeito de forma reduzida”

↑ Chamamos este termo de “efeito de primeiro estágio”



# Exemplo script - O instrumento é relevante?

Call:

```
lm(formula = coupon_use ~ treatment + income + age, data = data)
```

Residuals:

| Min      | 10       | Median   | 3Q      | Max     |
|----------|----------|----------|---------|---------|
| -0.67690 | -0.24458 | -0.03881 | 0.12899 | 0.94663 |

Coefficients:

|                  | Estimate            | Std. Error          | t value       | Pr(> t )                          |
|------------------|---------------------|---------------------|---------------|-----------------------------------|
| (Intercept)      | -0.0934587523       | 0.0112553243        | -8.304        | <0.0000000000000002 ***           |
| <b>treatment</b> | <b>0.3629912505</b> | <b>0.0044713360</b> | <b>81.182</b> | <b>&lt;0.0000000000000002 ***</b> |
| income           | 0.0000065261        | 0.0000001489        | 43.834        | <0.0000000000000002 ***           |
| age              | -0.0066116804       | 0.0002235674        | -29.574       | <0.0000000000000002 ***           |

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3161 on 19996 degrees of freedom

Multiple R-squared: 0.3201, Adjusted R-squared: 0.32

F-statistic: 3138 on 3 and 19996 DF, p-value: < 0.0000000000000022

A regressão de primeiro estágio explica a relação entre pertencer ao grupo de tratamento e o uso do cupom.

Nessa regressão, o coeficiente para "treatment" foi 0,362, significando que ~36% do grupo de tratamento usou o cupom.

p-valor < 0,01, e valor da estatística t de 81,2 (substancial).



# Exemplo script – exogeneidade e restrição de exclusão

## Without covariates

```
Call:  
ivreg(formula = future_purchases ~ coupon_use | treatment, data = data)  
  
Residuals:  
    Min      1Q   Median      3Q      Max  
-96.6313 -17.4202 -0.1149  17.0029 130.2710  
  
Coefficients:  
            Estimate Std. Error t value     Pr(>|t|)  
(Intercept) 167.4725    0.2530 661.98 <0.0000000000000002 ***  
coupon_use   49.2429    0.9916  49.66 <0.0000000000000002 ***  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 25.49 on 19998 degrees of freedom  
Multiple R-Squared: 0.4252,     Adjusted R-squared: 0.4252  
Wald test: 2466 on 1 and 19998 DF, p-value: < 0.0000000000000022
```

## With covariates

```
Call:  
ivreg(formula = future_purchases ~ coupon_use + income + age |  
      treatment + income + age, data = data)  
  
Residuals:  
    Min      1Q   Median      3Q      Max  
-76.5092 -13.5210 -0.1811  13.3568 85.5493  
  
Coefficients:  
            Estimate Std. Error t value     Pr(>|t|)  
(Intercept) 99.78367684  0.70001900 142.54 <0.0000000000000002 ***  
coupon_use   49.31155169  0.77788990  63.39 <0.0000000000000002 ***  
income        0.00100078  0.00001069   93.64 <0.0000000000000002 ***  
age          0.50669366  0.01504569   33.68 <0.0000000000000002 ***  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 19.96 on 19996 degrees of freedom  
Multiple R-Squared: 0.6476,     Adjusted R-squared: 0.6475  
Wald test: 7990 on 3 and 19996 DF, p-value: < 0.0000000000000022
```

Eles devem produzir resultados não tão distintos, para sustentar a ideia de que nosso instrumento (designação do cupom) afeta o outcome (compras futuras) apenas através da variável endógena (uso do cupom).

# Exemplo script – LATE, juntando os dois estágios

```
first_stage <- lm(coupon_use ~ treatment + income + age,  
data = data)  
second_stage <- ivreg(future_purchases ~ coupon_use +  
income + age | treatment + income + age, data = data)
```

- Os clientes que cumpriram ao usar o cupom viram suas compras futuras aumentarem em cerca de R\$50 em comparação ao que aconteceria sem o cupom.
- *Efeito causal do uso do cupom* nas compras futuras entre os compliers (aqueles que usaram o cupom porque o receberam)

Call:

```
ivreg(formula = future_purchases ~ coupon_use + income + age |  
      treatment + income + age, data = data)
```

Residuals:

| Min      | 1Q       | Median  | 3Q      | Max     |
|----------|----------|---------|---------|---------|
| -76.5092 | -13.5210 | -0.1811 | 13.3568 | 85.5493 |

Coefficients:

|             | Estimate    | Std. Error | t value | Pr(> t )                |
|-------------|-------------|------------|---------|-------------------------|
| (Intercept) | 99.78367684 | 0.70001900 | 142.54  | <0.0000000000000002 *** |
| coupon_use  | 49.31155169 | 0.77788990 | 63.39   | <0.0000000000000002 *** |
| income      | 0.00100078  | 0.00001069 | 93.64   | <0.0000000000000002 *** |
| age         | 0.50669366  | 0.01504569 | 33.68   | <0.0000000000000002 *** |

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.96 on 19996 degrees of freedom  
Multiple R-Squared: 0.6476, Adjusted R-squared: 0.6475  
Wald test: 7990 on 3 and 19996 DF, p-value: < 0.0000000000000022



# Exemplo script - ITT por meio da forma reduzida

```
Call:  
lm(formula = future_purchases ~ treatment + income + age, data = data)  
  
Residuals:  
    Min      1Q  Median      3Q     Max  
-98.973 -17.389 -0.949  16.594 119.399  
  
Coefficients:  
            Estimate Std. Error t value     Pr(>|t|)  
(Intercept) 95.17508074  0.90992762 104.596 <0.000000000000002 ***  
treatment   17.89966181  0.36148155  49.517 <0.000000000000002 ***  
income       0.00132259  0.00001204 109.884 <0.000000000000002 ***  
age          0.18066144  0.01807413   9.996 <0.000000000000002 ***  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 25.56 on 19996 degrees of freedom  
Multiple R-squared:  0.4224,    Adjusted R-squared:  0.4223  
F-statistic: 4875 on 3 and 19996 DF,  p-value: < 0.0000000000000022
```

O ITT mede o **efeito médio que oferecer um cupom** tem sobre as compras futuras dos clientes.

No nosso exemplo, esse efeito foi de R\$17,89 com p-valor < 0,01, enquanto controlamos por idade e renda, o que indica um efeito significativo de oferecer o cupom.

Note que o ITT geralmente é menor do que o LATE devido ao não compliance por uma parte dos indivíduos no grupo de tratamento.



# Outros casos além de non-compliance

## Velocidade de entrega e satisfação do cliente:

**Problema:** Companhia quer saber se **tempo de entrega menor afeta (endógeno)** um índice de satisfação do cliente. Porém, velocidade de entrega é correlacionada com localização do cliente ou flutuações na demanda, que também afetam satisfação.

**Instrumento:** Choques exógenos (fora do controle da empresa ou dos clientes) na cadeia logística, como condições climáticas adversas ou mudanças no algoritmo de otimização de rota usado pela empresa de entregas.

**Primeiro estágio:** Variações aleatórias vindas do choque exógeno afeta o tempo de entrega. Rodamos uma regressão, guardamos o valor previsto de atrasos em decorrência dessa variação.

**Segundo estágio:** Usa essa variação prevista como variável independente numa regressão onde a dependente é o índice de satisfação. O coeficiente dessa variável independente será o efeito causal de velocidade de entrega em satisfação do consumidor.



**4.1**

# **REGRESSÃO DESCONTÍNUA**

---

Motivação e princípios básicos



Universidade dos Dados

# Desenho de regressão descontínua (RDD)

## Regression Discontinuity Design

- Método que requer suposições brandas quando comparadas a de outros métodos observacionais (i.e., não experimentais).
- Muitas suposições são testáveis indiretamente de forma intuitiva.
- Em alguns cenários RDD provê inferência tão boa quanto aquela obtida através de desenho experimental.
  - Isso vale para casos bem específicos. Cuidado para não cair nessa falácia e sair por ai dizendo que RDD é sempre tão bom quanto experimental design.



# Como funciona o RDD?

Gertler et al, Impact Evaluation in Practice (World Bank)

Figure 6.1 Rice Yield, Smaller Farms versus Larger Farms (Baseline)

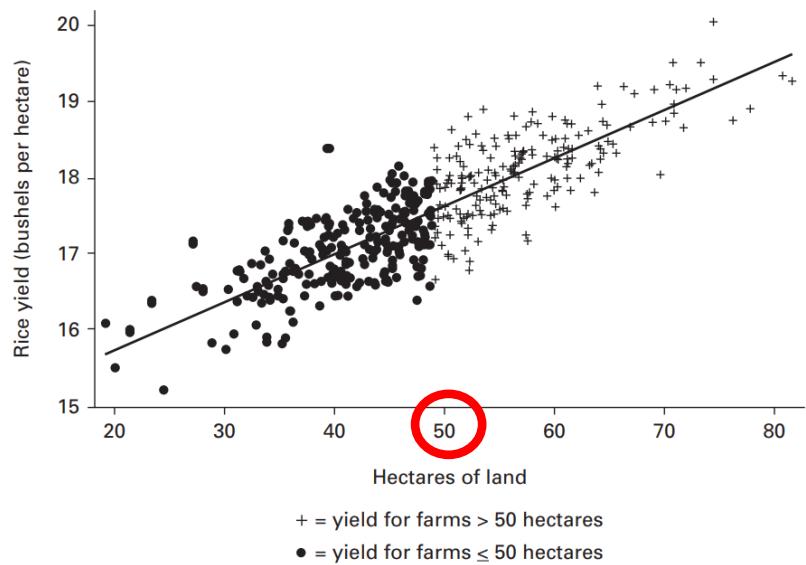
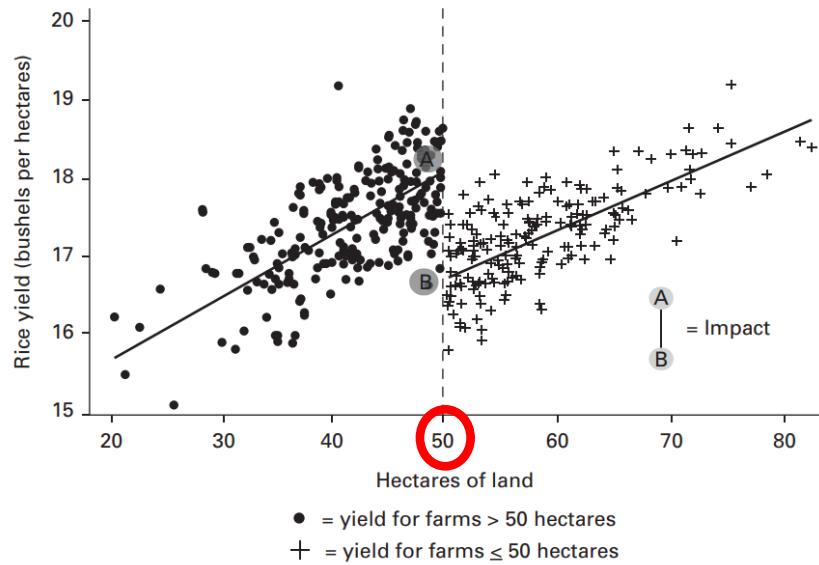


Figure 6.2 Rice Yield, Smaller Farms versus Larger Farms (Follow-Up)



# Proposta do método

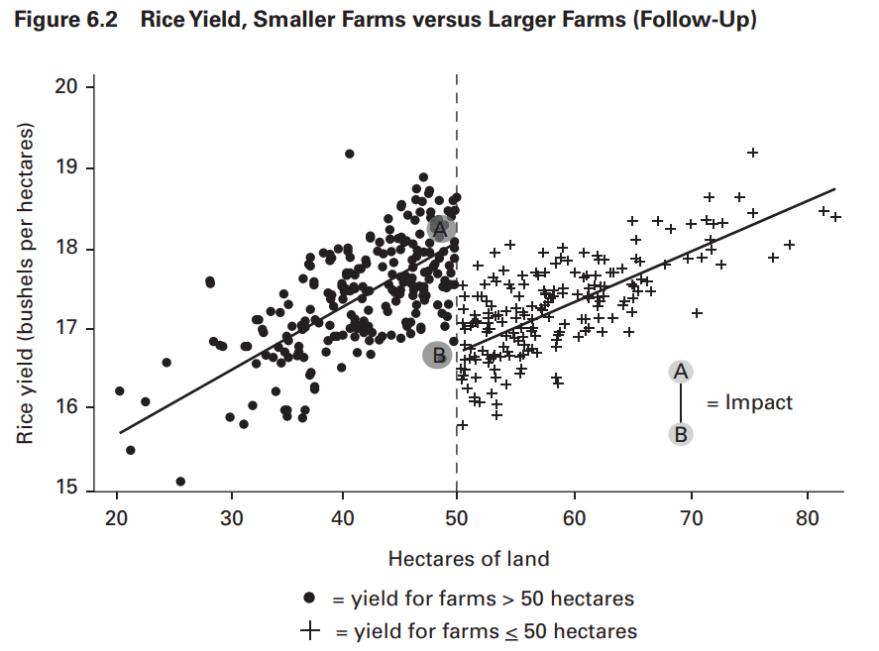
## Qual é a proposta para lidar com endogeneidade?

- A ideia é que indivíduos **imediatamente em torno** de determinado valor de uma variável  $x$ , como hectares = 50, são **relativamente parecidos** em diversas outras características.
  - De modo que qualquer viés de auto seleção seria minimizado quando comparado a uma diferença “ingênua” de médias do tipo  $E[Y|D_i = 1] - E[Y|D_i = 0]$
- Isso depende de algumas hipóteses que discutiremos mais a fundo, uma delas é de que os indivíduos **não conseguem facilmente escolher** de que lado da descontinuidade estariam (*ausência de manipulação*).
- Isso também está relacionado à ideia de que o valor de  $Y$  variaria de forma suave em torno da linha vertical caso os indivíduos não recebessem o tratamento (*continuidade do outcome potencial – não é diretamente testável*).



# Vocabulário do RDD - Anota aí!

- **Assignment rule:** “regra de atribuição” ou regra que define quem recebe o tratamento.
- **Running variable ou forcing variable:** variável em função da qual a regra de atribuição é estipulada.
- **Cut-off ou threshold:** valor da running variable a partir do qual ou abaixo do qual o tratamento é atribuído
- **Bandwidth:** “banda” ou domínio sobre o qual a running variable será *considerada na regressão*



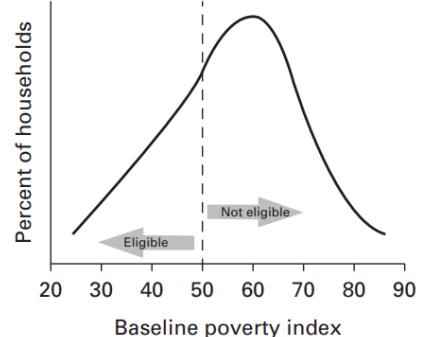
# Hipótese de continuidade

## Testes indiretos – densidade da running variable

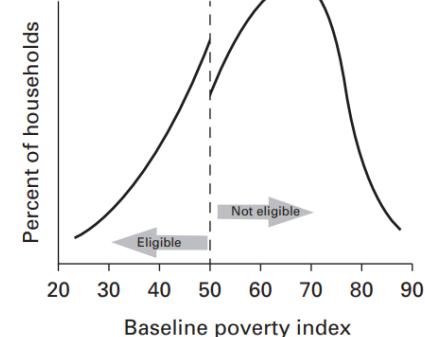
- Um grupo desses testes indiretos das hipóteses necessárias é o de “**testes de manipulação**” ou de continuidade da função de densidade da running variable.
- Esse tipo de teste envolve estimar a densidade de observações em torno do cut-off.
- Se na presença de um efeito do tratamento tudo mais exceto o outcome observado deve variar continuamente em torno do cut-off, então a **densidade da running variable** também deve ser contínua.

Figure 6.4 Manipulation of the Eligibility Index

a. No manipulation



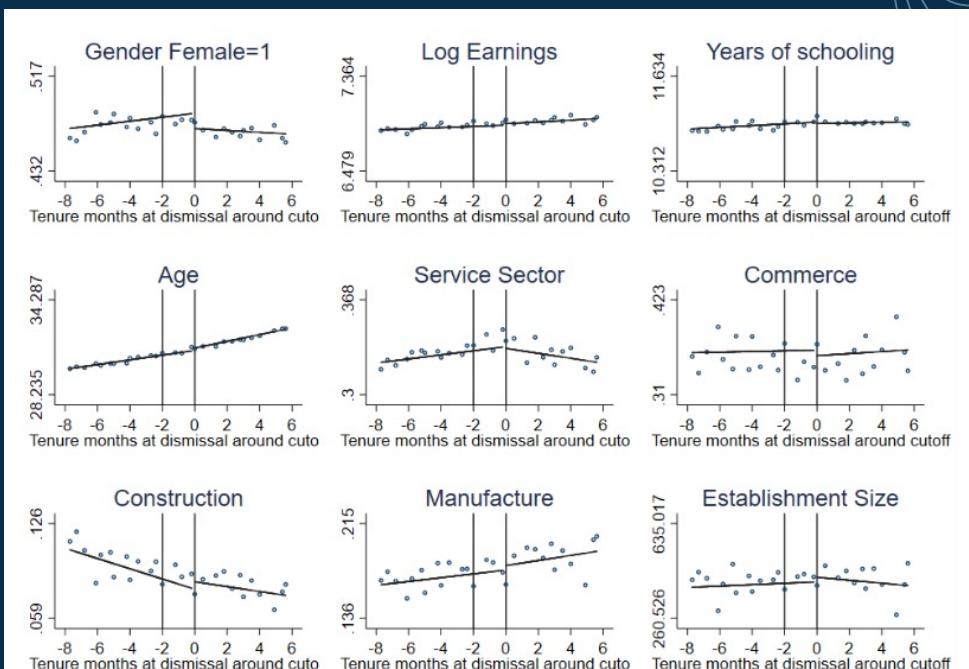
b. Manipulation



# Hipótese de continuidade

## Testes indiretos – continuidade em características observáveis

- Outro teste indireto da hipótese de continuidade da expectativa condicional do outcome potencial é mostrar que não há descontinuidade em características *observáveis* dos indivíduos em torno do cut-off
- Isso equivale a estimar RDDs como os vistos anteriormente trocando a outcome  $Y_i$  em questão por possíveis variáveis de controle  $X_{ij}$ .
- Um resultado confortante é não encontrar resultados estatisticamente significantes nessas regressões



## 4.2

# REGRESSÃO DESCONTÍNUA: ILUSTRAÇÃO DO CASO SHARP

---

Encaixando a aplicação deste método  
em uma contexto específico



Universidade dos Dados

# Exemplo: efeito de frete grátis em compras

---

*Exemplo: “Uma empresa de e-commerce quer avaliar se oferecer frete grátis para clientes com uma pontuação média de lealdade (loyalty ou engajamento) passada igual ou superior a 50 aumenta seus gastos futuros na plataforma.*

*A pontuação de lealdade varia continuamente entre os clientes e é calculada com base em interações passadas, como frequência de compras e engajamento.”*

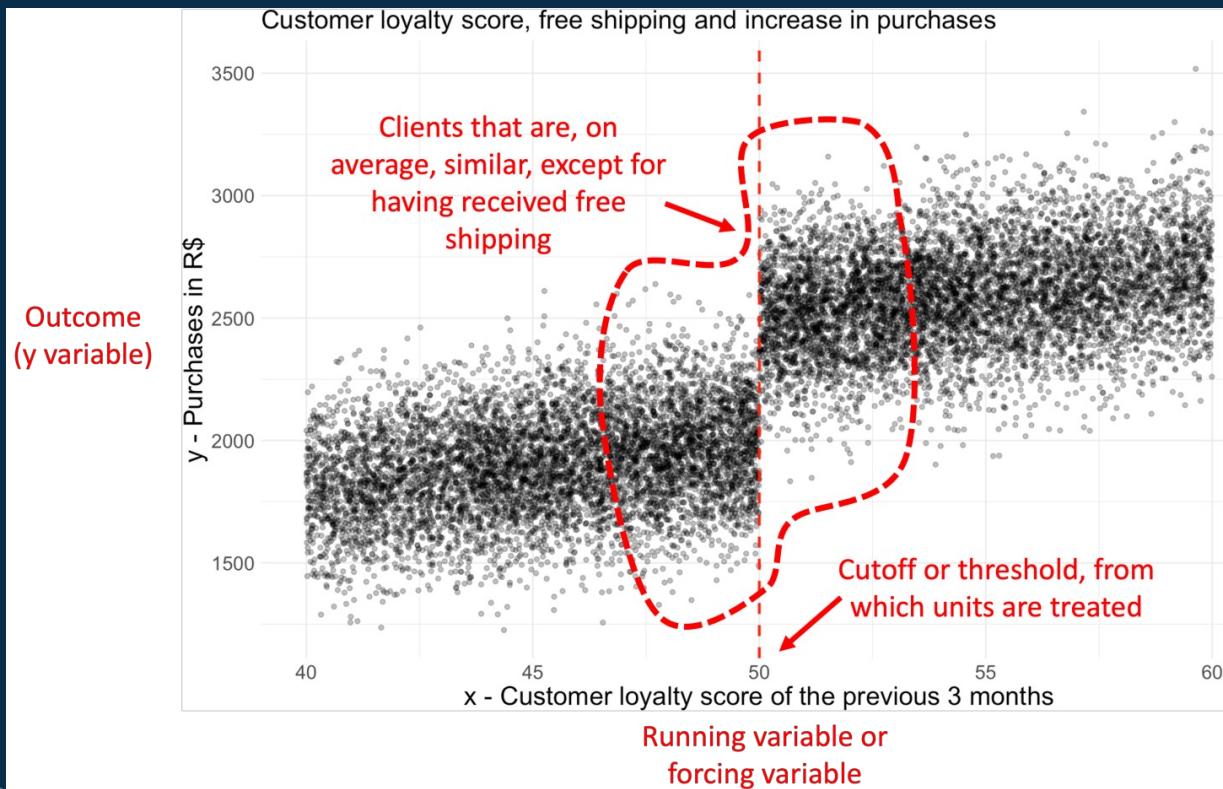
Simulei um conjunto de dados representando a situação descrita:

- **Running variable (x):** Pontuação de lealdade normalmente distribuída em torno de 50.
- **Outcome (y):** Total de gastos futuros em R\$. Introduzi um salto de R\$250 em y para clientes com  $x \geq 50$  como efeito causal do recebimento de frete grátis.
- **Covariáveis:** Idade, tempo como usuário e frequência de compras passadas



# Hipótese a ser testada com o uso do RDD

Dados brutos: Frete grátis aumenta compras?



# Definição do caso com as nomenclaturas que aprendemos

**Regra de atribuição:** Define quem recebe o tratamento com base na variável de corte (neste exemplo, aqueles que tem  $x > 50$  recebem o tratamento).

**Running variable ou forcing variable:** A variável segundo a qual a regra de atribuição é estabelecida (neste caso é  $x$ , a pontuação média de lealdade nos últimos 3 meses).

**Cutoff ou threshold:** O valor de corte que determina a atribuição do tratamento (neste exemplo, a pontuação média de loyalty deve ser  $> 50$  para que o cliente receba frete grátis).

**Bandwidth:** O intervalo ao redor do corte - cutoff - considerado para a análise. Por exemplo, na figura abaixo, as linhas verticais roxas denotam um possível bandwidth em torno do cutoff de 50.



# Hipótese a ser testada com o uso do RDD

Dados brutos: Frete grátis aumenta compras?

Hipótese a ser testada com RDD

Se score  
passado de  
loyalty  $\geq 50$

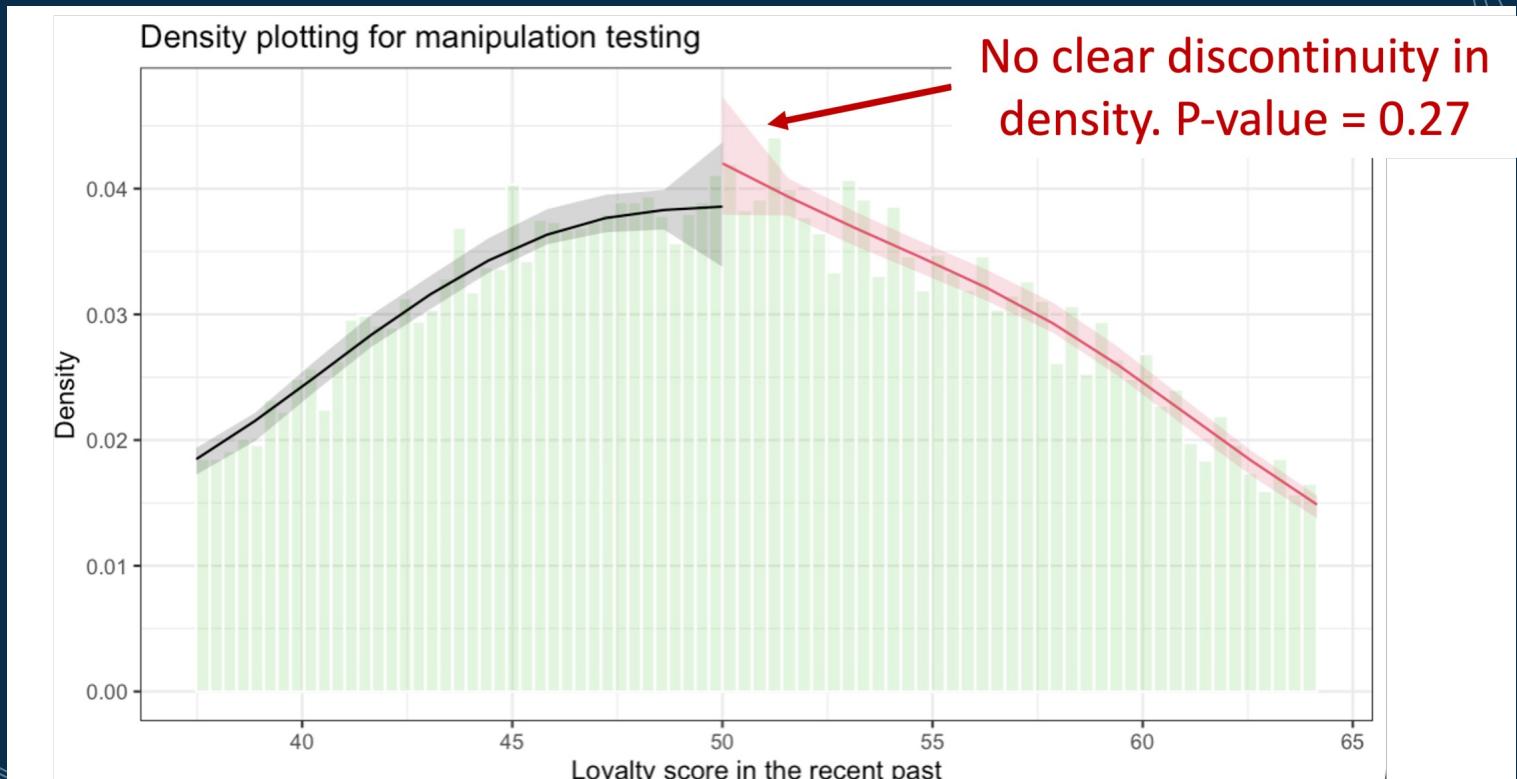
Frete grátis

Mais compras  
no futuro?

Como é um  
"resultado passado",  
não pode ser  
manipulado hoje.



# Teste de manipulação da running variable



# Teste de continuidade das covariáveis

There is no discontinuity in “Age” → p-value = 0.21

| Method       | Coef. | Std. Err. | z     | P> z  | [ 95% C.I. ]     |
|--------------|-------|-----------|-------|-------|------------------|
| Conventional | 0.248 | 0.220     | 1.125 | 0.261 | [-0.184 , 0.680] |
| Robust       | -     | -         | 1.255 | 0.210 | [-0.179 , 0.817] |

There is no discontinuity in “Tenure” → p-value = 0.41

| Method       | Coef. | Std. Err. | z     | P> z  | [ 95% C.I. ]     |
|--------------|-------|-----------|-------|-------|------------------|
| Conventional | 0.086 | 0.124     | 0.688 | 0.491 | [-0.158 , 0.329] |
| Robust       | -     | -         | 0.823 | 0.410 | [-0.166 , 0.406] |

There is no discontinuity in “Past frequency” → p-value = 0.78

| Method       | Coef.  | Std. Err. | z      | P> z  | [ 95% C.I. ]     |
|--------------|--------|-----------|--------|-------|------------------|
| Conventional | -0.036 | 0.089     | -0.406 | 0.684 | [-0.211 , 0.138] |
| Robust       | -      | -         | -0.268 | 0.789 | [-0.236 , 0.179] |



# Estimação do efeito causal LOCAL de frete grátis sobre compras

```
> rdd_result <- rdrobust(data$y, data$x, c = 50)
> summary(rdd_result) # discontinuity in y = Purchases
```

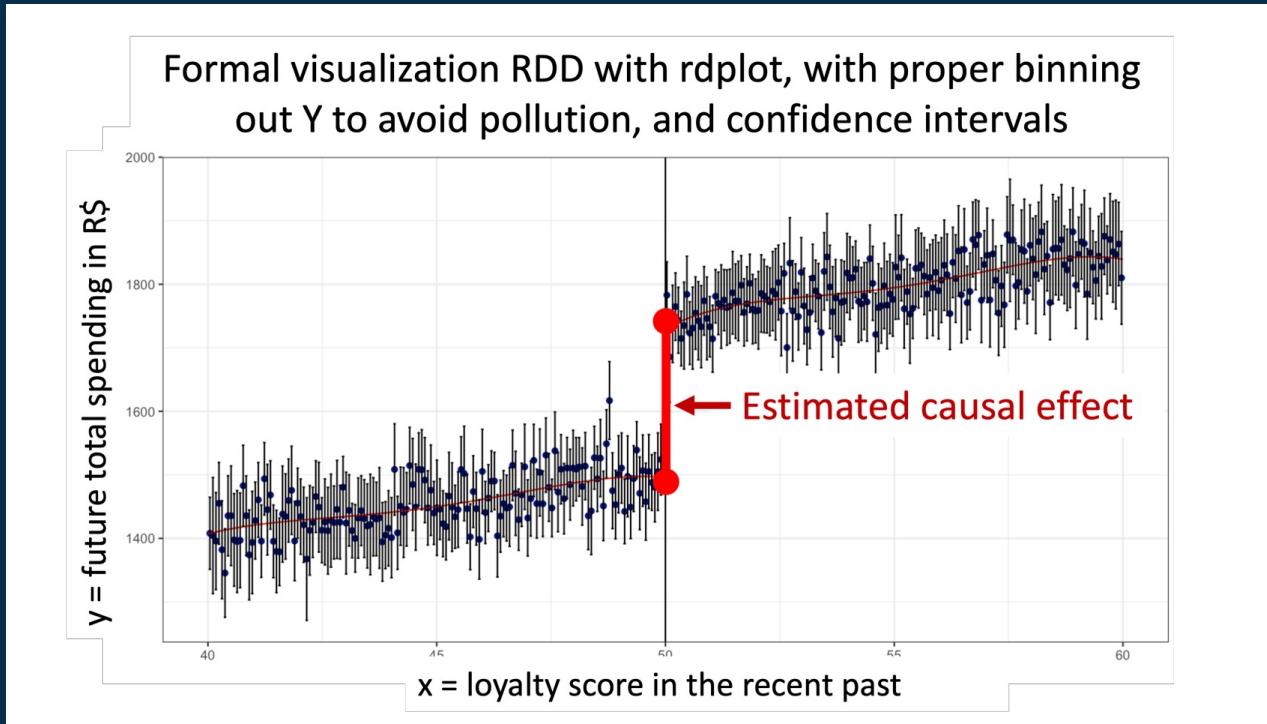
Discontinuity “future purchases” → p-value < 0.01

| Method       | Coef.   | Std. Err. | z      | P> z  | [ 95% C.I. ]        |
|--------------|---------|-----------|--------|-------|---------------------|
| Conventional | 240.439 | 8.203     | 29.310 | 0.000 | [224.360 , 256.517] |
| Robust       | -       | -         | 24.482 | 0.000 | [220.490 , 258.866] |

Efeito causal LOCAL de frete grátis nas compras futuras. p-valor robusto < 0,05 → efeito estatisticamente significativo.

Aumento de R\$240 nos gastos futuros para clientes dentro do bandwidth que receberam frete grátis, sugerindo que a política foi eficaz em incentivar compras adicionais para esse grupo.

# Visualização do efeito causal LOCAL de frete grátis sobre compras



# 5.1

## DIFERENÇA EM DIFERENÇAS (DiD ou Diff-in-Diff)

---

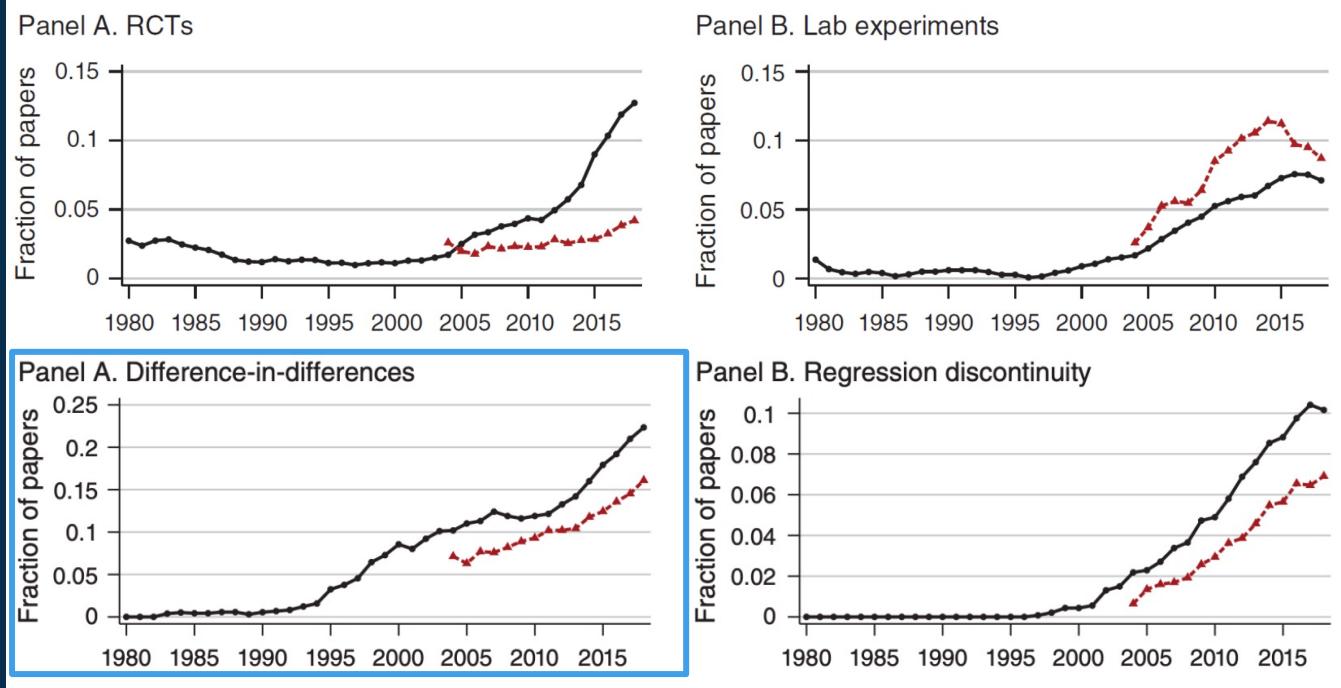
Motivação e princípios básicos



Universidade dos Dados

# MOTIVAÇÃO

Currie et al, 2020 - Technology and big data are changing economics: mining text to track methods



# MOTIVAÇÃO



- **Ex. problema 1:** Observar mudanças *antes e depois* (*variação pré-post ou within*) em compras dos mesmos clientes não irá capturar o efeito causal porque outros fatores influenciam compras ao longo do tempo.
  - Ex. limitação: Não considera tendências e sazonalidades que podem afetar os resultados.
- **Ex problema 2:** Ao mesmo tempo, comparar pessoas que assinaram Amazon Prime durante uma campanha contra aquelas que não assinaram *no mesmo período* (*variação cross section ou between*) é problemático por características não observáveis relacionadas à auto seleção (aula 1).
  - Ex. limitação: Precisamos incluir tudo o que afeta o momento do tratamento e o outcome de interesse.
- No entanto, e se combinássemos os dois métodos e comparássemos as mudanças *antes e depois* nos resultados do grupo que assinou Amazon Prime (tratados) com as mudanças *antes e depois* do grupo que não assinou (controles)?
  - A diferença nos resultados *antes e depois* do grupo de tratamento - a *primeira diferença* - controla pelos fatores que são **constantes ao longo do tempo** nesse grupo, uma vez que estamos comparando o grupo com ele mesmo.
  - A diferença nos resultados *antes e depois* do grupo de controle também “limpa” o efeito de características **fixas no tempo** que podem estar correlacionadas tanto com a *não* adoção do tratamento quanto com o outcome



# MOTIVAÇÃO

---

A abordagem de diferença-em-diferenças (DiD ou diff-in-diff) faz o que seu nome sugere:

Cada uma das **primeiras diferenças** nos permite “limpar” o efeito de variações no outcome que não são relacionadas ao tratamento em si, e sim a *idiossincrasias* de cada grupo (**efeitos fixos**)

A **segunda diferença** nos permite, após ter “limpado” os efeitos fixos de cada grupo ou indivíduo, estimar diferenças da evolução do outcome entre os dois grupos.

Mediante certas hipóteses, que discutiremos mas à frente, essa variação do item anterior pode ser atribuída ao efeito do tratamento.



# MOTIVAÇÃO

**Lembre-se do modelo de resultados potenciais da aula 1**

**Table:** Exemplo de resultados potenciais

|  | João | Maria |
|--|------|-------|
| Resultado observado: $Y_i$               | 4    | 5     |
| Resultado potencial sem seguro: $Y_{0i}$ | 3    | 5     |
| Resultado potencial com seguro: $Y_{1i}$ | 4    | 5     |
| Tratamento (status de seguro): $D_i$     | 1    | 0     |
| Efeito do tratamento: $Y_{1i} - Y_{0i}$  | 1    | 0     |

$$Y_{1,Joao} - Y_{0,Joao} = 4 - 3 = 1$$

$$Y_{1,Maria} - Y_{0,Maria} = 5 - 5 = 0$$



# MOTIVAÇÃO

**Lembre-se do modelo de resultados potenciais da aula 1**

Na vida real, não observaríamos os valores de  $Y_{0,Joao}$  nem de  $Y_{1,Maria}$

E se fizéssemos uma ingênuia comparação de médias observadas?

$$Y_{Joao} - Y_{Maria} = 4 - 5 = -1$$

Viés: Isso é fruto de não termos acesso a uma informação importante: com ou sem plano de saúde, Maria seria mais saudável que João

João não comprou o plano de saúde por acaso. Ele *escolheu* fazer parte do grupo de tratamento porque tem saúde invariavelmente mais frágil.



# MOTIVAÇÃO

E se acompanhássemos ambos ao longo do tempo?

|                                      | João | Maria |
|--------------------------------------|------|-------|
| Periodo $t = 1$ :                    | 3    | 5     |
| Periodo $t = 2$ :                    | 4    | 5     |
| Tratamento (status de seguro): $D_i$ | 1    | 0     |

$$\begin{aligned}\beta_{DD} &= E[Y_{it}|D_i = 1, t = 2] - E[Y_{it}|D_i = 1, t = 1] \\ &\quad - E[Y_{jt}|D_j = 0, t = 2] - E[Y_{jt}|D_j = 0, t = 1]\end{aligned}$$

$$\beta_{DD} = (4 - 3) - (5 - 5) = 1$$



# MOTIVAÇÃO

E se acompanhássemos ambos ao longo do tempo?

|                                      | João | Maria |
|--------------------------------------|------|-------|
| Periodo $t = 1$ :                    | 3    | 5     |
| Periodo $t = 2$ :                    | 4    | 5     |
| Tratamento (status de seguro): $D_i$ | 1    | 0     |

$$\begin{aligned}\beta_{DD} &= E[Y_{it}|D_i = 1, t = 2] - E[Y_{it}|D_i = 1, t = 1] \\ &\quad - E[Y_{jt}|D_j = 0, t = 2] - E[Y_{jt}|D_j = 0, t = 1]\end{aligned}$$

- Primeira diferença  
(within) João
- Primeira diferença  
(within) Maria

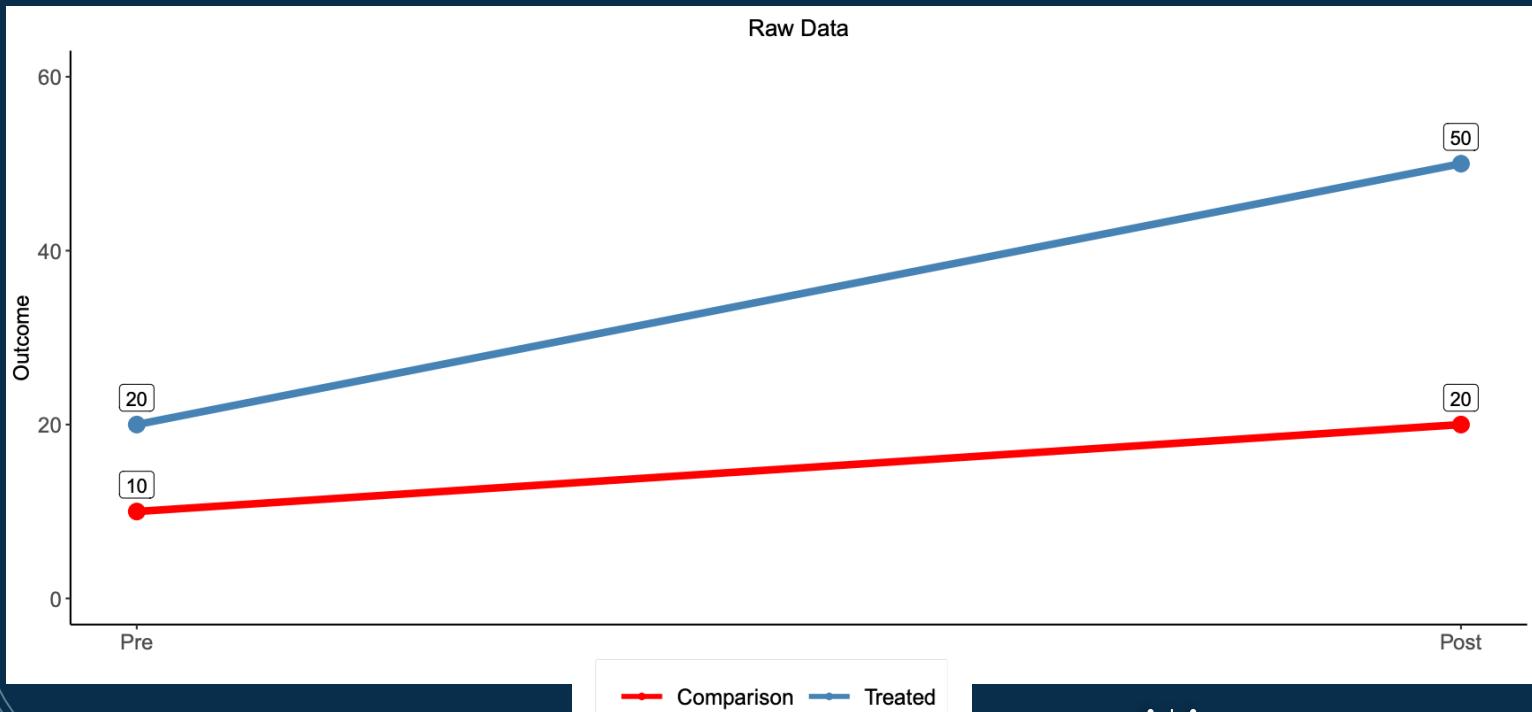
$$\beta_{DD} = (4 - 3) - (5 - 5) = 1$$

Diff in Diff (Diferença das diferenças)



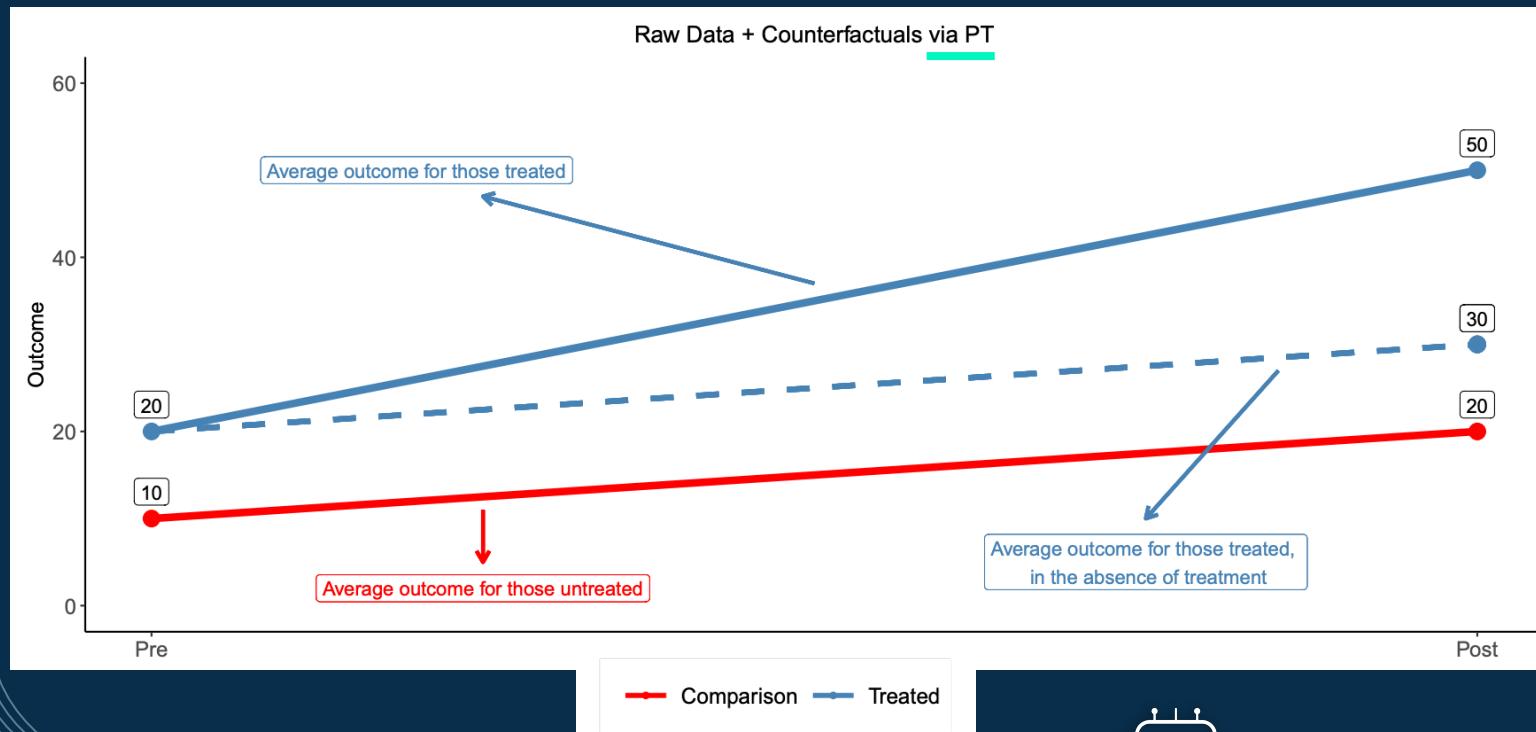
# Exemplo gráfico

Exemplo de Pedro Sant'Anna: <https://psantanna.com/did-resources>



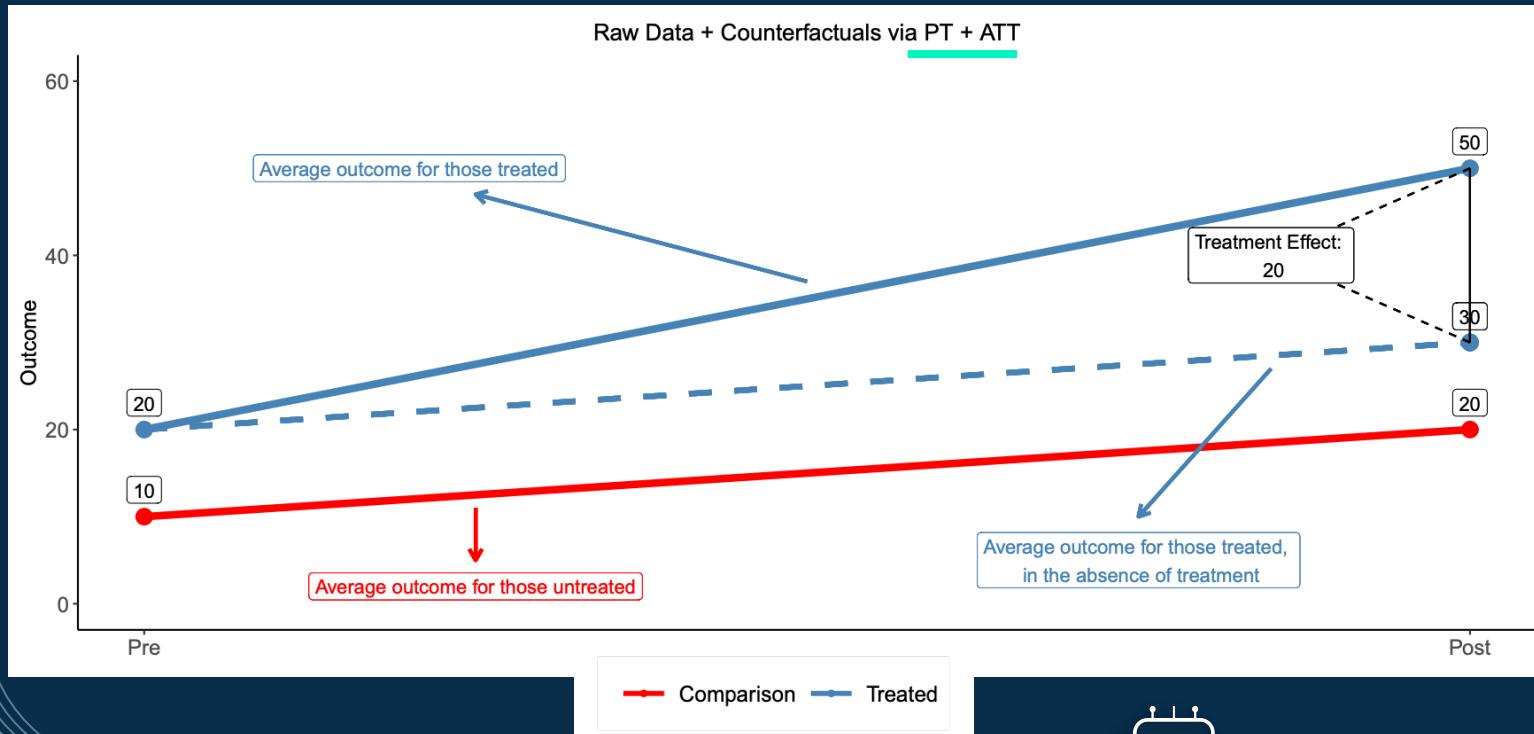
# Exemplo gráfico

Exemplo de Pedro Sant'Anna: <https://psantanna.com/did-resources>



# Exemplo gráfico

Exemplo de Pedro Sant'Anna: <https://psantanna.com/did-resources>



# E se acompanhássemos ambos ao longo do tempo? Para isso, precisamos ter “dados em painel” balanceado (?)

| Indivíduo (i) | Período (t) | Grupo      | Tratado no período (D) | Outcome (y) |
|---------------|-------------|------------|------------------------|-------------|
| 1             | 1           | Tratamento | 0                      | 80          |
|               | 2           | Tratamento | 0                      | 75          |
|               | 3           | Tratamento | 1                      | 100         |
|               | 4           | Tratamento | 1                      | 115         |
| 2             | 1           | Controle   | 0                      | 70          |
|               | 2           | Controle   | 0                      | 82          |
|               | 3           | Controle   | 0                      | 79          |
|               | 4           | Controle   | 0                      | 85          |
| ...           | ...         | ...        | ...                    | ...         |
| N             | 1           | Tratamento | 0                      | 87          |
|               | 2           | Tratamento | 0                      | 82          |
|               | 3           | Tratamento | 1                      | 107         |
|               | 4           | Tratamento | 1                      | 122         |

# E se acompanhássemos ambos ao longo do tempo? Para isso, precisamos ter “dados em painel” balanceado (?)

- A fórmula do DiD não depende necessariamente de acompanharmos os mesmos indivíduos, porque trabalhamos com médias no nível de grupos.
  - Isso significa que amostras aleatórias dos mesmos grupos (painel desbalanceado) ou de grupos semelhantes (repeated cross section) ao longo dos períodos poderiam fornecer resultados válidos.
- Todavia, ter dados de painel balanceado resulta em um maior poder estatístico do que ter apenas repeated cross section (Sant'Anna & Zhao, 2020).
- Além disso, quando você tem acesso apenas a repeated cross sections, é preciso ter cuidado com mudanças compostionais – das características de pessoas em diferentes períodos (Sant'Anna & Xu, 2023).

| Indivíduo (i) | Período (t) | Grupo      | Tratado no período (D) | Outcome (y) |
|---------------|-------------|------------|------------------------|-------------|
| 1             | 1           | Tratamento | 0                      | 80          |
|               | 2           | Tratamento | 0                      | 75          |
|               | 3           | Tratamento | 1                      | 100         |
|               | 4           | Tratamento | 1                      | 115         |
| 2             | 1           | Controle   | 0                      | 70          |
|               | 2           | Controle   | 0                      | 82          |
|               | 3           | Controle   | 0                      | 79          |
|               | 4           | Controle   | 0                      | 85          |
| ...           |             |            |                        |             |
| N             | 1           | Tratamento | 0                      | 87          |
|               | 2           | Tratamento | 0                      | 82          |
|               | 3           | Tratamento | 1                      | 107         |
|               | 4           | Tratamento | 1                      | 122         |

# 5.2

# SUPOSIÇÕES

---

Hipóteses necessárias para que os resultados sejam causais



Universidade dos Dados

# Hipótese de tendência paralela

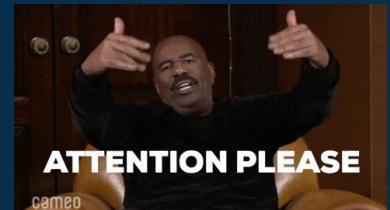
Embora o DID permita “limpar” diferenças que são constantes ao longo do tempo (efeitos fixos) em cada grupo, o método *não nos ajudará* a eliminar as *diferenças que variam com o tempo*.

Uma hipótese crucial para a identificação é que, na ausência do tratamento, o outcome de ambos os grupos deve variar à *mesma taxa*

- I.e., exigimos que os outcomes apresentem *tendências similares na ausência do tratamento*.

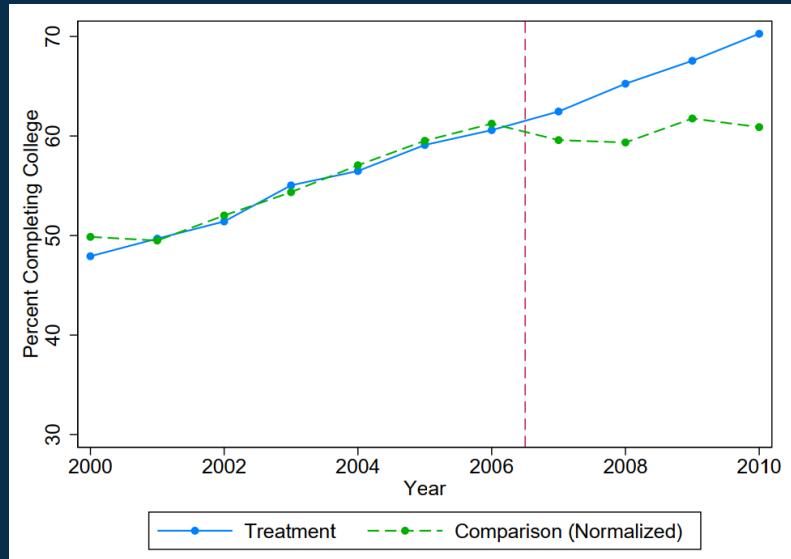
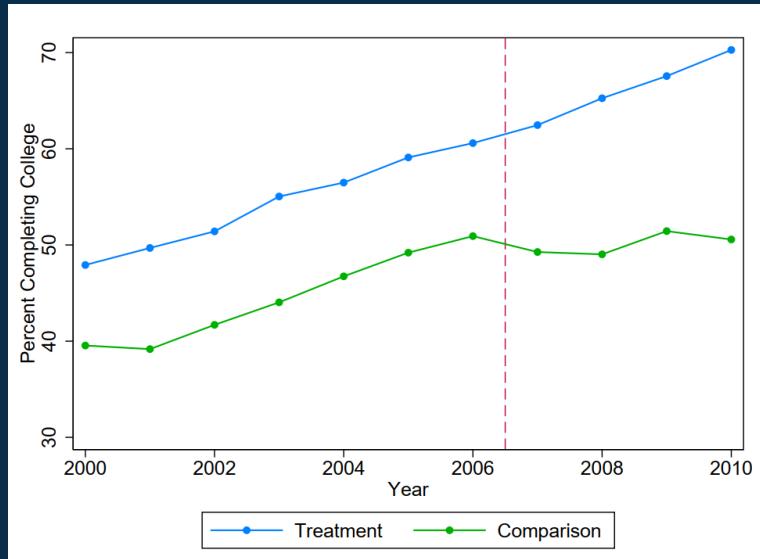
Obviamente isso não é uma hipótese *diretamente* testável ...

... muito menos quando há apenas dois períodos de tempo.

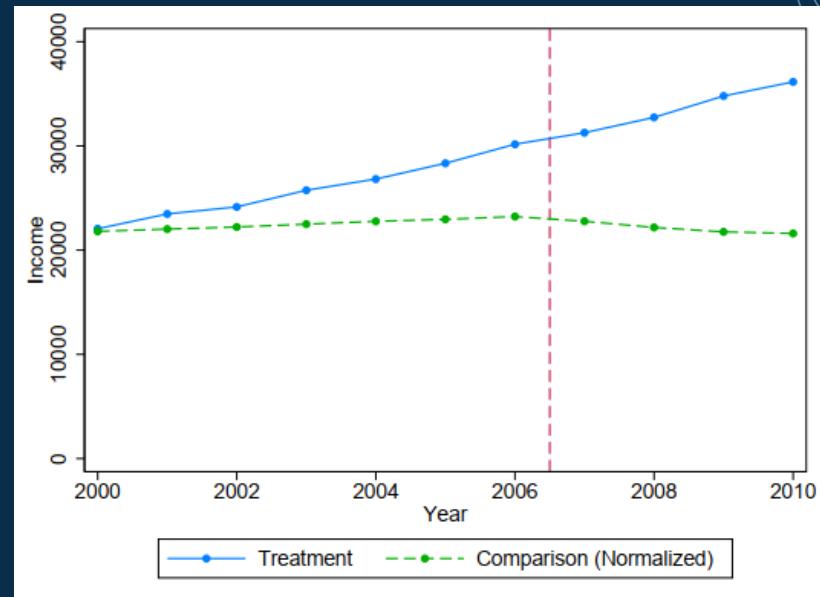
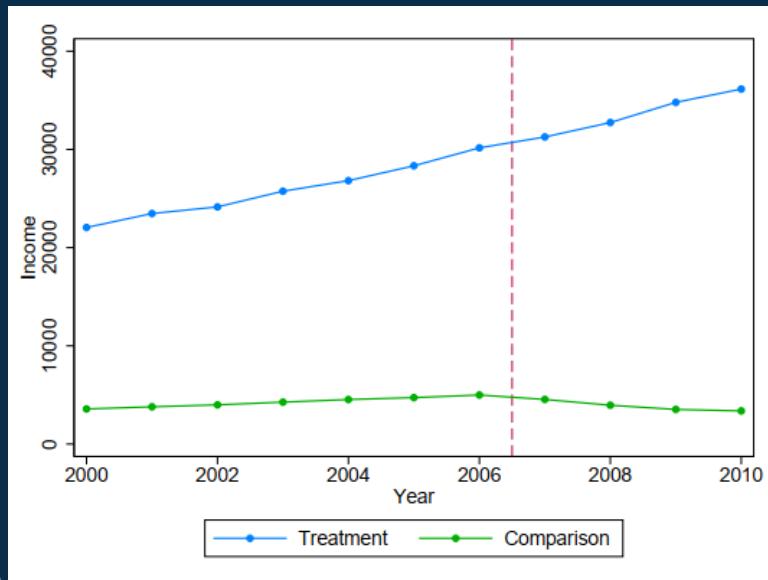


# Intuição por trás da tendência paralela

Um exemplo com cara de dados reais



# Exemplo de ausência de tendência paralela



# Nem tudo está perdido se não houver tendência paralela

Nem tudo está perdido:

A primeira abordagem é incluir variáveis de controle. É possível que a tendência paralela exista uma vez que ajustemos a análise por covariáveis

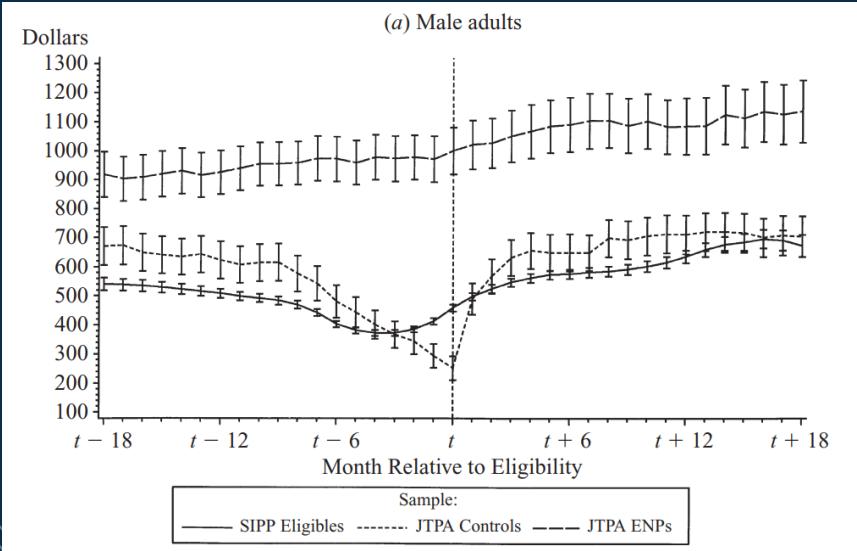
Também é possível transformar os dados. Por exemplo, utilizar  $\ln(\text{income})$  como variável dependente pode tornar a tendência paralela. Nota: Se houver tendência paralela em  $\ln$  não haverá tendência paralela em nível vice-versa (Angrist & Pischke, 2008)

Algumas outras alternativas:

- Abadie (2005) propõe um estimador DID semi-paramétrico baseado em propensity score para balancear os grupos de tratamento e controle em características observáveis.
- Arkhangelsky et al. (2021) apresentam o estimador "Synthetic Difference-in-Differences", que combina o método de controle sintético, usado para construir um grupo de controle comparável, com o método de diferença em diferenças.
- Controlar explicitamente por uma tendência, o que limita a identificação a uma hipótese sobre a forma funcional da tendência (i.e., linear, quadrática, etc.).



# **Queda pré-tratamento: o mergulho de Ashenfelter Heckman & Smith, 1999 – Economic Journal**



- A ausência de tendência paralela pode indicar um problema de autosseleção no tratamento, o que leva inevitavelmente ao viés de seleção discutido em aulas anteriores.
- É como administrar um medicamento apenas aos pacientes em estado mais grave, distorcendo a comparação entre grupos tratados e não tratados.
- Esse fenômeno é conhecido como "Ashenfelter's dip" ou "pre-program dip", conforme descrito por Ashenfelter (1978).



# Outra hipótese importante para a identificação

## Stable Unit Treatment Value assumption (SUTVA)

- Apenas um outcome potencial deve ser observado para cada membro da população. Não deve haver interações relevantes entre membros da população, especialmente dos dois grupos.
- Caso SUTVA seja violada, não observamos nenhum dos outcomes potenciais. Isso acontece caso o tratamento afete tanto indivíduos no grupo de tratados quanto no grupo de controles
- Um desses casos é a existência de externalidades do tratamento ou efeitos spillover (efeito transbordamento).
- **Exemplo de violação do SUTVA:** um grande programa de treinamento de trabalhadores pode mudar o equilíbrio do mercado fazendo com que aqueles que participaram sejam mais valorizados pelas novas habilidades
- Os que não participaram, portanto, terão mais dificuldade de encontrar emprego depois da implementação do programa. Seu outcome não é mais o mesmo que seria em um mundo em que o programa não existisse (efeito conhecido na indústria como “canibalização”).



# Perguntas importantíssimas

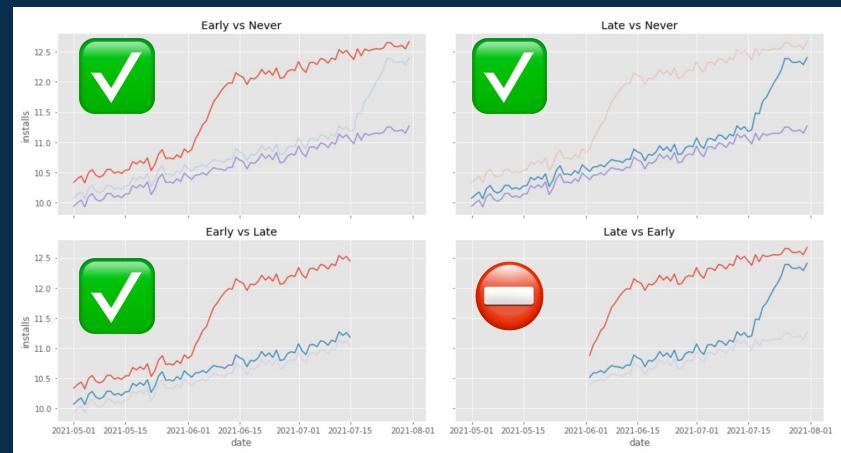
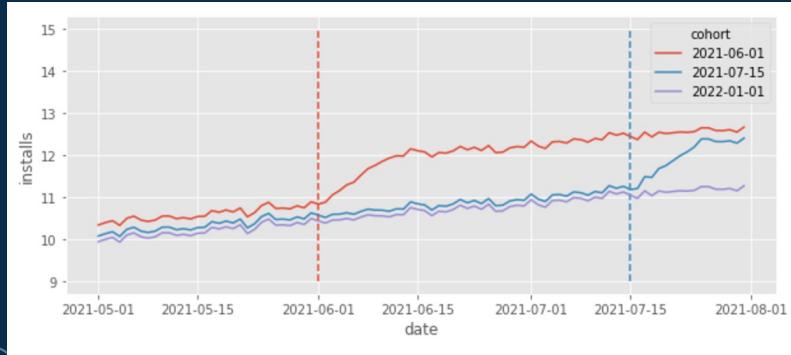


- Há algum motivo específico para os indivíduos tratados estarem recebendo o tratamento? É possível que haja viés de seleção?
  - Mesmo que haja tendência paralela pré tratamento, o tratamento pode estar sendo atribuído aos indivíduos que tenham potencial diferenciado.
- Existe algum outro evento acontecendo na data de tratamento que poderia estar causando um efeito de tratamento composto (*compound treatment* → viés)?
- Todos os indivíduos estão sendo tratados no mesmo período ou é possível que o tratamento esteja ocorrendo em períodos diferentes? O que define o timing do tratamento?
  - A literatura mais recente mostra que Diff-in-Diff com indivíduos sendo tratados em períodos diferentes pode apresentar viés caso haja seleção no timing ou efeitos heterogêneos:
    - 💡 Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of econometrics*, 225(2), 200-230.

# DiD com indivíduos tratados em períodos distintos

## Figuras de “Causal Inference for the Brave and True” – Matheus Facure

A literatura mais recente mostra que Diff-in-Diff com indivíduos sendo tratados em períodos diferentes pode apresentar viés, principalmente caso haja seleção no timing ou efeitos heterogêneos



O modelo padrão de TWFE (efeitos fixos de indivíduo e tempo) não faz apenas comparações entre unidades tratadas e ainda não tratadas...

...Ele também faz comparações "proibidas", como comparar uma unidade por último com uma unidade tratada no começo (caso 🚫)

# Ajustando a análise para correlação serial

- A **correlação serial** ocorre quando a variável explicativa para uma mesma unidade está correlacionada ao longo do tempo, ou seja, o erro em um período está relacionado ao erro em períodos anteriores.
- **Exemplo:** Ao avaliar o impacto de uma promoção sobre TGMV dos clientes, o comportamento persistente dos consumidores pode influenciar períodos subsequentes, gerando dependência entre os dados.
- O **Diff-in-Diff** pressupõe independência dos erros entre observações. Quando há correlação serial, a inferência estatística (ex: teste de hipóteses) é comprometida, resultando em maior taxa de falsos positivos (identificação de efeitos inexistentes).
- Duas hipóteses clássicas do OLS, entretanto, são que  $Cov(e_t, e_s) = 0$  para  $t \neq s$  e  $Cov(e_i, e_j) = 0$  para  $i \neq j$ .
  - Em muitos casos, pela natureza dos outcomes considerados nas estimativas usando Diff-in-Diff, o termo de erro apresentará correlação serial. Em outros casos, pela natureza do fenômeno, o termo de erro será correlacionado entre indivíduos.
- **Viés nos resultados:** A correlação serial afeta a validade do teste de hipóteses, especialmente se os erros estão correlacionados dentro das unidades tratadas ou de controle. Mas não causa viés nos  $\beta$  estimados
- A **clusterização do termo de erro** ajusta os erros padrão para considerar dependências dentro de clusters, como regiões, categorias ou períodos. Exemplo: usar clientes como clusters para capturar correlação em suas compras ao longo do tempo.



# 5.3

# FORMALIZAÇÃO

---

A estrutura dos modelos Diff-in-Diff  
agregado e estudo de evento



Universidade dos Dados

## Avaliação de impacto de uma campanha

- No mês 7 do ano, uma rede de supermercado criou uma campanha Out-of-Home (OOH) na Paraíba mas não em Pernambuco.
- Os consultores avaliam como as vendas de lojas da rede na Paraíba variam com lançamento da campanha utilizando PB como tratado e PE como controle em um diff-in- diff de dois períodos (Antes e Depois)

$$Y_{ist} = \gamma_i + \lambda_t + \delta D_{st} + \varepsilon_{ist}, \text{ com } E[\varepsilon_{ist}|s, t] = 0$$

- $Y_{ist}$  - Vendas na loja  $i$  no estado  $s$  no período  $t$
- $D_{st}$  - Dummy = 1 para o estado com campanha a partir da data de lançamento
- $i$  - Efeito fixo de loja, que poderia ser denotado por uma dummy para cada loja
- $\lambda_t$  - Efeito fixo de mês, que poderia ser denotado por uma dummy para cada mês



# Avaliação de impacto de uma campanha

Calculando a **primeira diferença** do grupo do grupo de **controle**

$$E[Y_{ist}|s = PE, t = DEPOIS] - E[Y_{ist}|s = PE, t = ANTES] = \lambda_{DEPOIS} - \lambda_{ANTES}$$

= Diferença nos  
efeitos fixos de  
tempo

Calculando a **primeira diferença** do grupo do grupo de **tratamento**

$$E[Y_{ist}|s = PB, t = DEPOIS] - E[Y_{ist}|s = PB, t = ANTES] = \lambda_{DEPOIS} - \lambda_{ANTES} + \delta$$

= Diferença nos  
efeitos fixos de  
tempo + efeito  
da campanha

Calculando a diferença das diferenças acima

$$\begin{aligned} &E[Y_{ist}|s = PB, t = DEPOIS] - E[Y_{ist}|s = PB, t = ANTES] \\ &- \{E[Y_{ist}|s = PE, t = DEPOIS] - E[Y_{ist}|s = PE, t = ANTES]\} = \delta \end{aligned}$$

= efeito da  
campanha



# Inclusão de covariáveis na regressão

Falando de covariáveis, é possível incluí-las também no nosso modelo de diff-in-diff. Ainda considerando o exemplo da campanha OOH:

$$Y_{ist} = \gamma_s + \lambda_t + \delta D_{st} + X'_{ist}\beta + \varepsilon_{ist}$$

Onde  $X_{ist}$  pode incluir características individuais que sejam fixas (cuidado!) e/ou variáveis no tempo.

Dois pontos necessitam atenção:

- Caso haja variáveis fixas no tempo em  $X_{ist}$ , essas variáveis devem estar em um nível diferente daquele do efeito fixo de unidade no modelo ou haverá multicolinearidade.
- As variáveis fixas no tempo *podem aumentar* precisão das estimativas, mas são aquelas que variam no tempo que *tem maior potencial* de diminuir um eventual viés de variável omitida



## Estudo de evento (*Event study*)

Ainda considerando a notação que usamos para o exemplo de avaliação da campanha OOH, mas para analisarmos vários períodos - normalizaremos o período de tratamento para ser  $t = 0$ . A especificação para um estudo de evento ou event study seria:

$$Y_{ist} = \gamma_s + \lambda_t + \sum_{\tau=-q}^m \delta_\tau D_{\tau s} + X'_{ist}\beta + \varepsilon_{ist}$$

- $\delta_{-q}, \delta_{-q+1}, \dots, \delta_{-1}$  denotam os  $q$  **efeitos antecipatórios** ou “leads”.
- $\delta_0, \delta_1, \dots, \delta_m$  denotam os  $m + 1$  **efeitos pós-tratamento** ou “lags”.
- Cada um desses  $\delta$  mede o efeito do tratamento no período de referência  $\tau$ , lembrando que não devem existir “efeitos antes das causas”.
- Significância estatística nos **efeitos antecipatórios** ou “leads” é sinal de divergência de tendências pré tratamento.
- Já os **efeitos pós-tratamento** ou “lags” nos dizem sobre a dinâmica do efeito do tratamento *após* o tratamento. Relatam (i) quando o efeito acontece relativamente a  $t = 0$ , (ii) se perdura ou desaparece no curto prazo.



# 5.4

# EXERCÍCIO

---

Diff-in-Diff agregado e estudo de  
evento



Universidade dos Dados

# Exemplo: efeito de campanha OOH

Exemplo: “Uma rede de supermercados deve avaliar o impacto de uma campanha de marketing “Out-of-Home”, realizada no estado da Paraíba, sobre as vendas.

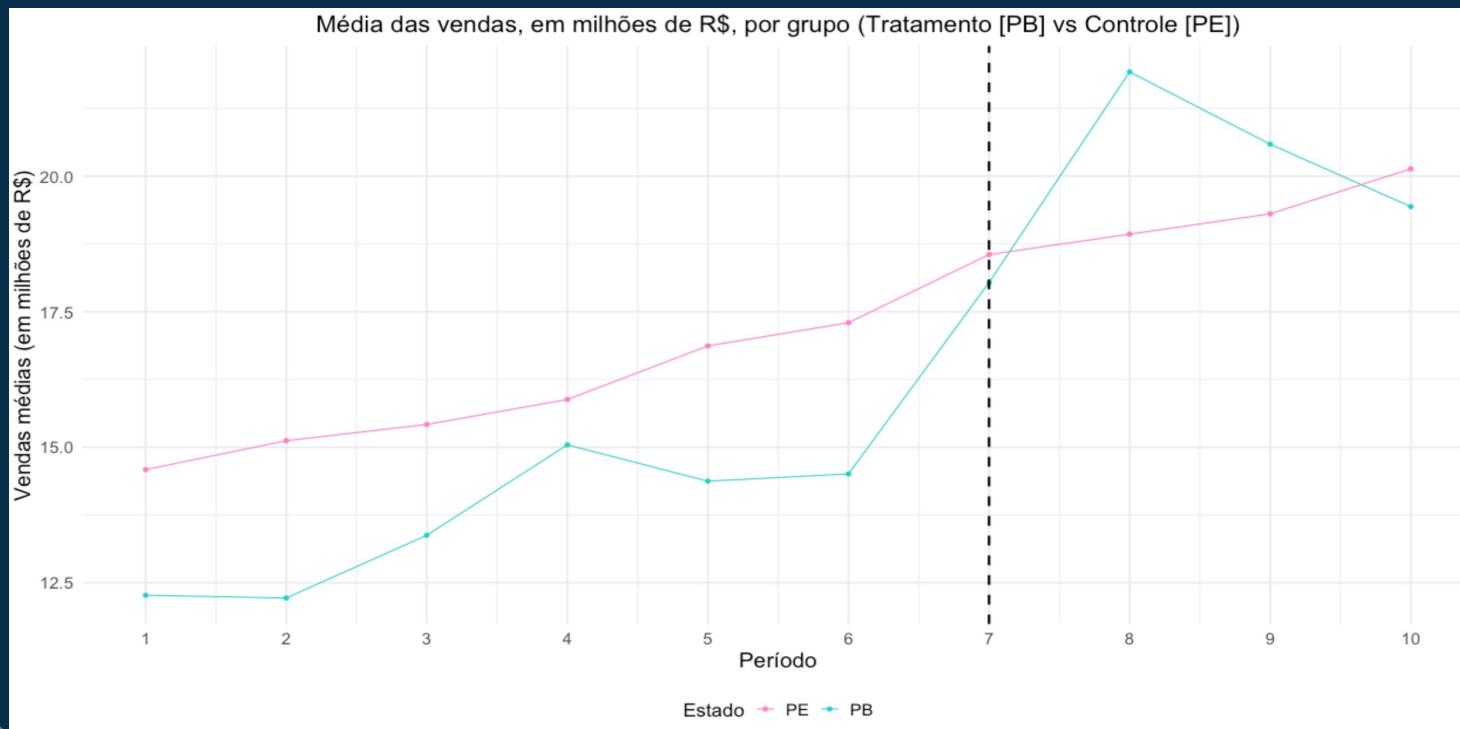
O objetivo é entender se a campanha foi bem-sucedida em aumentar as vendas nas lojas tratadas (localizadas na Paraíba), comparando com as lojas não tratadas (localizadas em Pernambuco).

Simulei um conjunto de dados representando a situação descrita:

- **loja**: Identificação da loja;
- **estado**: Estado onde a loja está localizada (PE ou PB);
- **periodo**: Período de observação;
- **vendas**: Vendas em milhões de reais;
- **tratamento** : Dummy que indica se a loja está no estado tratado ou não
- **tratado\_periodo**: Dummy indicando se a loja está tratada no período;
- **periodo\_relativo**: Distância entre o período de tratamento ( $t = 7$ ) e o período observado



# Exemplo: efeito de campanha OOH



# Exemplo: efeito da campanha OOH

```
# Diff-in-Diff agregado  
modelo_diff <- feols(vendas ~ tratado_periodo | loja + periodo,  
                      cluster = ~loja, data = dados)  
summary(modelo_diff)
```

Efeitos fixos de loja e de período

```
> summary(modelo_diff)  
OLS estimation, Dep. Var.: vendas  
Observations: 1,000  
Fixed-effects: loja: 100, periodo: 10  
Standard-errors: Clustered (loja)  
Estimate Std. Error t value Pr(>|t|)  
tratado_periodo 2.99566 0.639945 4.68112 9.0742e-06 ***  
---  
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
RMSE: 3.42766 Adj. R2: 0.413086  
Within R2: 0.042142
```

Erros clusterizados por loja

As lojas na Paraíba tiveram um aumento médio de  $\hat{\beta}_1$ =R\$2.99MM nas vendas, estatisticamente significante, já que o p-valor é menor que 0,05



# Exemplo: efeito da campanha OOH

```
modelo_event <- feols(vendas ~ i(periodo_relativo, tratamento, ref = "-1") |  
                      loja + periodo, cluster = ~loja, data = dados)  
  
summary(modelo_event)  
  
> summary(modelo_event)  
OLS estimation, Dep. Var.: vendas  
Observations: 1,000  
Fixed-effects: loja: 100, periodo: 10  
Standard-errors: Clustered (loja)  
  
Estimate Std. Error t value Pr(>|t|)  
periodo_relativo:::-6:tratamento 0.470112 1.040988 0.451602 6.5254e-01  
periodo_relativo:::-5:tratamento -0.113125 1.012397 -0.111740 9.1126e-01  
periodo_relativo:::-4:tratamento 0.744677 1.011157 0.736460 4.5515e-01  
periodo_relativo:::-3:tratamento 1.951822 0.945782 2.063713 4.1661e-02 *  
periodo_relativo:::-2:tratamento 0.293964 0.930982 0.315757 7.5285e-01  
periodo_relativo:::0:tratamento 2.271706 0.898038 2.529634 1.2996e-02 *  
periodo_relativo:::1:tratamento 5.779916 1.181574 4.891710 3.8743e-06 ***  
periodo_relativo:::2:tratamento 4.071423 1.093773 3.722365 3.2794e-04 ***  
periodo_relativo:::3:tratamento 2.091223 1.123646 1.861105 6.5697e-02 .  
---  
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
RMSE: 3.38601 Adj. R2: 0.422066  
Within R2: 0.065276
```

Período de referência é o -1

Efeitos fixos de loja e de período  
+ erros clusterizados por loja

Efeitos médios estimados para cada um dos períodos.

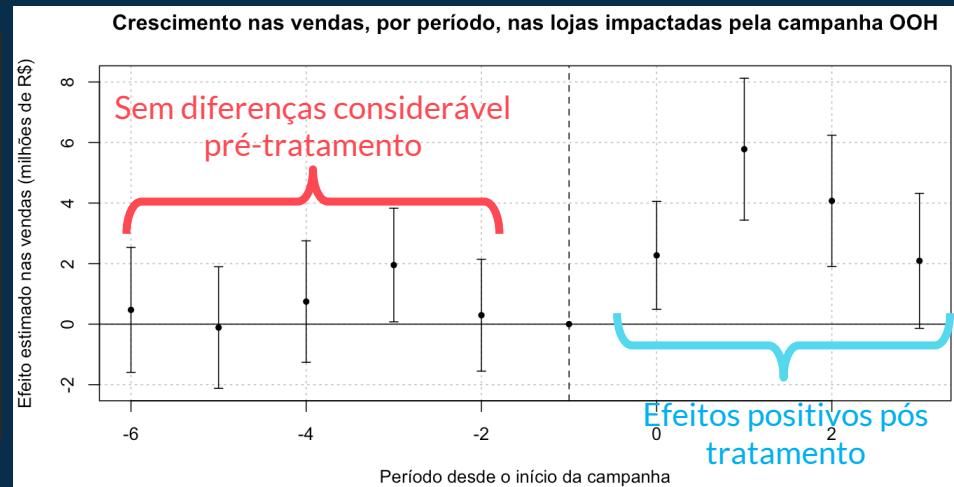
Ex: no mês de início da campanha (período relativo = 0), as lojas na Paraíba tiveram um aumento médio de  $\beta_1$ =R\$2.27MM nas vendas, estatisticamente significante.



Universidade dos Dados

# Exemplo: efeito da campanha OOH

|                                  | Estimate  |
|----------------------------------|-----------|
| periodo_relativo:::-6:tratamento | 0.470112  |
| periodo_relativo:::-5:tratamento | -0.113125 |
| periodo_relativo:::-4:tratamento | 0.744677  |
| periodo_relativo:::-3:tratamento | 1.951822  |
| periodo_relativo:::-2:tratamento | 0.293964  |
| periodo_relativo:::0:tratamento  | 2.271706  |
| periodo_relativo:::1:tratamento  | 5.779916  |
| periodo_relativo:::2:tratamento  | 4.071423  |
| periodo_relativo:::3:tratamento  | 2.091223  |
| ---                              |           |



# É HORA DE DAR TCHAU



Obrigado por estarem aqui e por tudo o que aprendemos juntos

Ao lado, o QR CODE para nos conectarmos no LinkedIn

Para mais aprofundamento nos temas, recomendo os seguintes materiais gratuitos:

- [Causal Inference for the Brave and True, Matheus Facure](#)
- [The Effect: An Introduction to Research Design and Causality, Nick Huntington-Klein](#)
- [Causal Inference - The Mixtape, Scott Cunningham](#)

