

Análise multivariada e sua aplicação em políticas públicas

Parte 01: Randomize Control Trials (RCTs)

Ricardo Ceneviva (UFABC-CECS)

Universidade Federal do ABC
Centro de Engenharias, Modelagem e Ciências Sociais Aplicadas

ricardo.ceneviva@ufabc.edu.br

<https://github.com/rceneviva/>

10 de junho de 2025

Objetivos da Aula

- Discutir a importância da inferência causal nas ciências sociais.
- Compreender os fundamentos dos RCTs e seu papel na Revolução da Credibilidade.
- Analisar evidências empíricas de RCTs aplicados a políticas públicas.
- Refletir sobre os limites e oportunidades do uso de experimentos em políticas públicas.

- A pergunta causal fundamental: *“O que teria acontecido com Y se X não tivesse ocorrido?”*
- Modelo de Neyman-Rubin: cada unidade possui dois potenciais resultados.
- O problema: só observamos um desses resultados. A solução? Comparações válidas.

A Revolução da Credibilidade

- Transição de métodos correlacionais para identificação causal.
- Ênfase em variações exógenas e desenhos de pesquisa robustos.
- Os “Cinco Métricos”: RCT, Regressão, DiD, IV, RD (Angrist e Pischke, 2014).

O que é um RCT?

- Randomização: unidades são alocadas aleatoriamente em tratamento ou controle.
- Elimina viés de seleção e garante balanceamento ex ante.
- Permite estimar o efeito médio de tratamento (ATE) com validade interna elevada.

Exemplo: RCT em Escolas do Quênia

- Estudo de Kremer et al. analisado por Angrist & Pischke (2014).
- Escolas receberam livros didáticos adicionais.
- Objetivo: avaliar impacto sobre desempenho escolar.

Referência: Cap. 1, *Mastering 'Metrics*, p. 10.

- Média geral: sem efeito significativo.
- Efeitos positivos apenas para alunos com bom desempenho inicial.
- Indicação de heterogeneidade no efeito do tratamento.

Ver: Figura 1.1 (p. 11), Tabela 1.1 (p. 12), Tabela 1.2 (p. 13).

- Nem todos os alunos usaram os livros.
- Estima-se o efeito da **oferta** (ITT), não necessariamente do uso.
- Introduz a noção de *compliance* parcial e motivação para IV.

Vantagens dos RCTs

- Inferência causal robusta.
- Transparência metodológica.
- Replicabilidade.
- Utilidade para desenho e avaliação de políticas públicas.

- Problemas de validade externa.
- Questões éticas e logísticas.
- Possibilidade de spillovers e não conformidade.
- Alto custo em alguns contextos.

Referência crítica: Deaton & Cartwright (2018).

- Programas de transferência de renda e incentivos escolares.
- Colaborações universidade-governo (SEAP-MG, FGV, J-PAL LAC).
- Potencial para uso em educação, saúde e administração pública.

- Mesmo em RCTs, uso de regressão pode aumentar precisão.
- Ajuda a controlar pequenas diferenças residuais entre grupos.
- Transição para modelos multivariados nas próximas aulas.

- Resultados do Quênia são generalizáveis?
- Quais os desafios éticos dos RCTs?
- Onde podemos (ou não) fazer RCTs no Brasil?

- Angrist & Pischke (2014), Cap. 1.
- Banerjee & Duflo (2011), *Poor Economics*.
- Glennerster & Takavarasha (2013), *Running Randomized Evaluations*.
- Deaton & Cartwright (2018), *Social Science & Medicine*.
- Glewwe, Kremer & Moulin. 2009. *Many Children Left Behind? Textbooks and Test Scores in Kenya*.

Métodos de Regressão

- Introduzir a regressão linear como ferramenta para inferência causal.
- Mostrar como a regressão complementa os RCTs.
- Discutir interpretações, suposições e aplicações empíricas.

Modelo básico:

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

- Y_i : variável dependente (ex: nota do aluno)
- X_i : variável explicativa (ex: livro = 1)
- β : efeito marginal de X sobre Y
- ε_i : erro (fatores não observados)

- RCT com distribuição aleatória de livros em escolas.
- Regressão usada para ajustar diferenças residuais.
- Resultado: impacto médio pequeno (≈ 0).

Gráfico 2.1: dispersão dos dados e linha de regressão (p. 41)

Interpretação de β

- β representa a mudança esperada em Y para uma unidade de mudança em X .
- Quando X é binária, β é a diferença de médias ajustada.

Exemplo:

"Ter livro" ($X = 1$) versus "Não ter livro" ($X = 0$).

- Permite ajustar para variáveis como renda, gênero, localização.
- **Tabela 2.1 (p. 43):** resultados com e sem controles mostram robustez.
- Ideal quando randomização não é perfeita ou está ausente.

Suposições para Inferência Causal

- ① Linearidade: relação entre X e Y deve ser apropriada.
- ② Exogeneidade: $\mathbb{E}[\varepsilon_i | X_i] = 0$.
- ③ Ausência de variáveis omitidas relevantes.

RCTs asseguram (2), mas em dados observacionais isso é uma hipótese forte.

- Ajuda a aumentar precisão (reduz variância).
- Permite explorar efeitos heterogêneos.
- Ajusta covariáveis desbalanceadas por acaso.

Gráfico 2.2 (p. 45): regressão com ajuste fino da linha de tendência.

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i$$

- Cada β_j : efeito marginal de X_j mantendo os outros constantes.
- Essencial para controlar várias fontes de variação.

- Regressão não elimina viés de seleção sozinha.
- Importante avaliar desenho da pesquisa.
- Evitar "overfitting" com muitas covariáveis irrelevantes.

- Regressão é uma ferramenta essencial da análise causal.
- Deve ser usada com consciência teórica e cuidado técnico.
- Regressão bem usada complementa tanto RCTs quanto dados observacionais.