

Programa de Pós-graduação em Políticas Públicas

Universidade Federal do ABC (UFABC)

Lista de Exercícios – Aula #02: RCTs e Regressão

Curso: Métodos Computacionais Aplicados em Políticas Públicas

Professor: Ricardo Ceneviva (PhD): ricardo.ceneviva@ufabc.edu.br

Objetivo: Praticar inferência causal com RCTs, regressão linear e estimadores de tratamento (ATE, ATT, ITT, TOT, LATE).

Pacotes R recomendados: ggplot2, AER, ivpack, sandwich, Matching, stargazer, dplyr, readr.

Parte I – Regressão Linear Simples e Múltipla

Exercício 1 – Regressão Linear Simples com dados simulados

- Simule 200 observações com uma covariável $x \sim N(0,1)$ e resultado $y = 2 + 3x + \epsilon$.
- Estime um modelo $\text{lm}(y \sim x)$ e plote a reta de regressão com ggplot2.
- Perguntas: O coeficiente estimado de x é próximo de 3? Como interpretar o R^2 ?

Exercício 2 – Regressão Múltipla com especificação incorreta

- Simule $y = 1 + 2x_1 + 0x_2 + 3x_3 + \epsilon$.
- Estime os modelos: $y \sim x_1 + x_2$; $y \sim x_1 + x_3$; $y \sim x_1 + x_2 + x_3$.
- Use o pacote stargazer para comparar os modelos.
- Perguntas: A omissão de x_3 gera viés? Qual o melhor modelo?

Exercício 3 – Diagnóstico de regressão

- Simule $y \sim x$ com erro heterocedástico: $\epsilon \sim N(0, x^2)$.
- Plote resíduos vs. valores ajustados. Use bptest() e vcovHC().
- Perguntas: O modelo viola homocedasticidade? Como corrigir?

Parte II – RCTs e Estimadores de Tratamento

Exercício 4 – RCT com dados simulados: ATE e ITT

- Simule 400 unidades: tratamento $\sim rbinom(1, 0.5)$, $Y = 50 + 5 * \text{tratamento} + \epsilon$.
- Estime ATE via diferença de médias e $\text{lm}(Y \sim \text{tratamento})$.
- Use ggplot2 para visualizar com geom_boxplot().
- Perguntas: Qual é o ATE estimado? A variabilidade é igual entre os grupos?

Exercício 5 – Bolsa Família e desempenho escolar (dados reais)

- Fonte: <https://www.kaggle.com/datasets/alexandrejose/brazil-bolsa-familia-education>
- Filtrar crianças 6–14 anos. Estimar $\text{lm}(\text{aprovado} \sim \text{bolsafamilia} + \text{controles})$.

- Use stargazer para comparar modelos com e sem controles.
- Perguntas: O efeito do programa é robusto? Há risco de viés?

Exercício 6 – Não conformidade: ITT vs TOT (simulado)

- Simule Z (incentivo) e D (tratamento real). Estime ITT com $\text{lm}(Y \sim Z)$.
- Estime TOT com $\text{ivreg}(Y \sim D | Z)$ do pacote AER.
- Perguntas: TOT > ITT? Como interpretar o TOT como LATE?

Exercício 7 – ATT com pareamento (dados reais)

- Fonte: <https://rdrr.io/cran/wooldridge/man/wage1.html>
- Use o dataset wage1. Estime ATT para hsgrad usando Matching::Match.
- Compare com estimativa via regressão multivariada.
- Perguntas: ATT difere de ATE? O pareamento melhorou o balanceamento?

Exercício 8 – Efeitos heterogêneos (simulado)

- Simule $Y = 40 + 3 * \text{trat} + 2 * g + 4 * \text{trat} * g + \varepsilon$.
- Estime $\text{lm}(Y \sim \text{trat} * g)$. Visualize com ggplot2.
- Perguntas: Qual o efeito por grupo? Interação significativa?

Exercício 9 – Regressão com instrumento e robustez (dados reais)

- Fonte: <https://economics.mit.edu/files/379>
- Use quarter of birth como instrumento para years of education.
- Compare $\text{lm}(\log(wage) \sim \text{education})$ e $\text{ivreg}(\log(wage) \sim \text{education} | \text{qob})$.
- Use coeftest com vcovHC para erros robustos.
- Perguntas: O instrumento é forte? A robustez altera a significância?

Exercício 10 – Desafio final: interferência (spillovers)

- Simule 500 unidades em clusters. Atribua tratamento por cluster com spillovers.
- Estime ATE, TOT e LATE com erros padrão robustos por cluster (ivpack).
- Compare com modelo que ignora cluster.
- Perguntas: Como spillover afeta os estimadores? Como corrigir?