

Tópicos Especiais de Métodos Computacionais

Roteiro para Sessão Prática de Laboratório #02

Simulação, Regressão Linear Múltipla e Dados Longitudinais em R

Prof. Ricardo Ceneviva

Programa de Pós-graduação em Políticas Públicas

Universidade Federal do ABC

24 de junho de 2025

Introdução

Esta lista foi preparada para a sessão prática em laboratório (90 min). Todos os dados são *simulados* para garantir reprodutibilidade: basta definir `set.seed(12345)` em cada exercício. Trabalhe em grupos de 2 estudantes e entregue: (i) um script `.R` comentado; (ii) um relatório curto (máx. 10 páginas) discutindo resultados.

Objetivos de aprendizagem

- Aplicar princípios de amostragem em *surveys* a dados simulados.
- Estimar e interpretar modelos de regressão linear múltipla com pesos.
- Construir painéis balanceados e comparar estimadores de efeitos fixos e aleatórios.
- Implementar Diferenças-em-Diferenças (DiD) com adoção escalonada e testar a escolha entre FE e RE (Hausman).

Instruções gerais

1. Abra um projeto no RStudio e carregue os pacotes: `tidyverse`, `survey`, `fixest`, `plm`, `did`, `ggplot2`.

2. Inclua `set.seed(12345)` no início de cada exercício.
3. Comente cada bloco de código explicando sua lógica.

Exercício 1 – População sintética e regressão linear múltipla

- 1.1. Gere uma população de $N = 10\,000$ indivíduos com:
 - $sexo \sim \text{Bernoulli}(0,5)$,
 - $escolaridade \sim \mathcal{N}(12, 3^2)$,
 - $experincia \sim \mathcal{N}(10, 5^2)$, truncada em 0,
 - $erro \varepsilon \sim \mathcal{N}(0, 4^2)$,
 - $salrio = 800 + 150 \cdot escolaridade + 60 \cdot experincia - 120 \cdot sexo + \varepsilon$.
- 1.2. Ajuste OLS com *salrio* como dependente e discuta cada coeficiente em relação aos valores “verdadeiros”.
- 1.3. Verifique pressupostos: normalidade dos resíduos, homocedasticidade, multicolinearidade (VIF).

Exercício 2 – Amostragem estratificada e pesos de desenho

- 2.1. A partir da população do Ex. 1, extraia uma amostra estratificada por *sexo* com afiliação proporcional ($n = 1\,200$).
- 2.2. Calcule pesos de amostragem ($w_i = N_h/n_h$).
- 2.3. No pacote **survey**, crie o objeto de desenho e estime a média ponderada de *salrio*.
- 2.4. Compare com a média populacional e discuta o impacto dos pesos.

Exercício 3 – Regressão ponderada para *survey*

- 3.1. Usando a amostra do Ex. 2, estime novamente a regressão linear múltipla ponderada.
- 3.2. Compare coeficientes ponderados versus não-ponderados (Ex. 1).
- 3.3. Diagnostique resíduos ponderados com gráficos apropriados.

Exercício 4 – Painel balanceado, FE, RE e teste de Hausman

- 4.1. **Simulação do painel:** gere $N = 1\,000$ indivíduos (id) observados em $T = 5$ períodos ($t = 1, \dots, 5$).
- Efeito individual $\alpha_i \sim \mathcal{N}(0, 5^2)$.
 - Tendência temporal comum $\beta_t = 5t$.
 - Tratamento $D_{it} = 1$ para 50% dos indivíduos *a partir* de $t \geq 3$.
 - Erro idiossincrático $u_{it} \sim \mathcal{N}(0, 4^2)$.
 - Resultado $Y_{it} = 200 + \alpha_i + \beta_t + 25 \cdot D_{it} + u_{it}$.
- 4.2. **Efeitos fixos (FE):** no `fixest`, estime `feols(Y ~ D | id + t)`. Interprete o coeficiente de D_{it} .
- 4.3. **Efeitos aleatórios (RE):** utilize o `plm` (`model = "random"`) ou `lme4` (`lmer`). Relate variâncias entre e intra-indivíduos.
- 4.4. **Teste de Hausman:** aplique `phtest` (`plm`) para comparar FE vs. RE.
- 4.4.1. Formule H_0 : RE é consistente e eficiente.
- 4.4.2. Interprete valor-p e conclua.
- 4.5. **Curvas de crescimento:** estime um modelo com intercepto e inclinação aleatória (`lmer(Y ~ t + D + (t|id))`) e discuta diferença para FE.

Exercício 5 – Diferenças-em-Diferenças (DiD) com adoção escalonada

- 5.1. Com o painel do Ex. 4, aplique o estimador de Callaway & Sant'Anna (`did::att_gt`). *Grafique* $ATT_{g,t}$ ao longo do tempo (`ggplot2`).
- 5.2. Discuta se o impacto do tratamento é constante ou heterogêneo.

Entrega: enviar até *04/07/2025* via e-mail: ricardo.ceneviva@ufabc.edu.br

▷ `script_grupo_X.R` e `relatorio_grupo_X.pdf`