

Tutorial 3

Após o período do estágio no Cemif, você é convidada (o) para participar de um projeto no Centro de Excelência e Inovação em Políticas Educacionais (FGV CEIPE). O projeto consiste em examinar o impacto do tamanho da turma no desempenho escolar, usando a base do Sistema de Avaliação de Educação Básica (Saeb) para o estado de Alagoas. Estudos recentes sobre o impacto do tamanho da turma no desempenho escolar no Brasil não chegam a um consenso¹. A Secretária da Educação de Alagoas reconhece que houve avanços recentes nos índices de escolaridade básica no estado², mas há espaço para algumas reformas adicionais.³ O grupo de pesquisa do CEIPE começa a trabalhar nesse projeto em parceria com a Secretaria de Educação de Alagoas. Para isso, a Secretaria disponibiliza os dados da Prova Brasil de 2019 dos estudantes da 4ª série (5º ano) do Ensino Fundamental do estado de Alagoas, encontrados no arquivo **nota-alagoas2019.csv**. O dicionário das variáveis se encontra na próxima página

1. A Secretária gostaria de entender se existe um impacto não linear do tamanho da turma na nota da Prova Brasil em matemática. Para isso você vai usar um modelo de regressão da nota no tamanho da turma, e considerando que cor/raça do estudante, escolaridade da mãe, morar com a mãe, ter internet em casa e se o estudante trabalha, já abandonou a escola ou foi reprovado, pois são variáveis que também podem ter impacto no desempenho escolar (Menezes-Filho, 2007; Matavelli e Menezes-Filho, 2020). Primeiro você usa todas as variáveis de escolaridade da mãe, *educmae1* a *educmae5*. Um outro aluno integrante do grupo olha os resultados preliminares e percebe que aparece NA no coeficiente de *educmae5*. Ele gostaria de entender o motivo.
2. Após corrigir esse problema, você apresenta os resultados preliminares integrantes do CEIPE. Alguém que está assistindo a apresentação sugere que seria melhor checar o **fator de inflação de variância** (VIF – *variance inflation factor* - na sigla em inglês), pois algumas variáveis explicativas podem ter correlação alta entre si. O grupo usa o pacote *car* no RStudio para calcular automaticamente o fator de inflação de variância. O que o resultado indica? Se a multicolinearidade for um problema nesta regressão, o que o grupo pode fazer para corrigir o problema?
3. A Diretora do CEIPE pergunta se você usou *erros-padrão robustos*. Você procura saber do que se trata e descobre que o modelo de regressão pressupõe *homoscedasticidade*, $Var(u|x_1, \dots, x_k) = \sigma^2$. Em outras palavras, o termo de erro

¹ Ver Menezes-Filho, 2007, Oliveira, 2010, e Matavelli e Menezes-Filho, 2020.

² Alagoas deixa histórico de piores resultados do País e supera metas do Ideb em 2019:

<https://www.terra.com.br/noticias/educacao/alagoas-deixa-historico-de-piores-resultados-do-pais-e-supera-metas-do-ideb-em-2019,80fceb5ed6bf7bcb2f95c51b9ea6565erxdns7ud.html>

³ Ensino de matemática e português melhora, mas nível segue baixo, aponta Saeb:

<https://valor.globo.com/brasil/noticia/2020/09/15/ensino-de-matematica-e-portugus-melhora-mas-nivel-ainda-baixo-na-maioria-dos-estados-diz-saeb.ghtml>

possui a mesma variância, dado quaisquer valores para as variáveis explicativas. Se o pressuposto for violado, os erros-padrão tornam-se inválidos, assim como todos os testes de hipóteses. Desta maneira, você estima o mesmo modelo da questão 1, mas usa erros-padrão robustos, comparando com os erros-padrão tradicionais encontrados na primeira questão

4. O grupo mostra os resultados da regressão. A Diretora do CEIPE pergunta se você analisou a hipótese nula de *homoscedasticidade*. Primeiro, ela pede que você faça algum gráfico de resíduos que permita avaliar descritivamente se há ou não presença de homoscedasticidade.
5. Em seguida ela pede para você fazer algum teste de homoscedasticidade. Qual a hipótese nula do teste? Qual seu p-valor e qual a conclusão?

Referências (leitura NÃO obrigatória)

Oliveira, J. M. (2010). Custo-efetividade de políticas de redução do tamanho da classe e ampliação da jornada escolar: Uma aplicação de estimadores de matching.

Bibliografia inicial obrigatória

Gujarati 10.1-10.5; 11.1-11.5 e Wooldridge 2.4; 6.2; 3.4; 8.1; 8.2; 8.3

Aula 4 eclass

Menezes-Filho, N. A. (2007). Os determinantes do desempenho escolar do Brasil.

Matavelli, I. R. e Menezes-Filho, N. (2020). Efeitos de tamanho da sala no desempenho dos alunos: Evidências usando regressões descontínuas no Brasil. Revista Brasileira de Economia, 74(3), 352-401.

Dicionário das variáveis

nota_mt: nota de matemática

tamturma: tamanho da turma

negro: dummy igual a 1 se o estudante é preto ou pardo (0 caso contrário)

educmae1: dummy igual a 1 se a mãe não completou o 5º ano do Ensino Fundamental (0 caso contrário)

educmae2: dummy igual a 1 se a mãe completou Ensino Fundamental, até o 5º ano (0 caso contrário)

educmae3: dummy igual a 1 se a mãe possui Ensino Fundamental completo (0 caso contrário)

educmae4: dummy igual a 1 se a mãe possui Ensino Médio completo (0 caso contrário)

educmae5: dummy igual a 1 se a mãe possui Ensino Superior completo (0 caso contrário)

moramae: dummy igual a 1 se o estudante mora com a mãe (0 caso contrário)

wifi: dummy igual a 1 se o estudante possui internet rede Wi-Fi em casa (0 caso contrário)

trabalha: dummy igual a 1 se o estudante trabalha (0 caso contrário)

abandonou: dummy igual a 1 se o estudante já abandonou a escola (0 caso contrário)

reprovado: dummy igual a 1 se o estudante já foi reprovado (0 caso contrário)