

Universidade Federal do Paraná  
Setor de Ciências Exatas  
Departamento de Estatística  
Programa de Especialização em *Data Science* e *Big Data*

Tatiani Cristina Ara

## **Desempenho municipal do IDEB: modelagem estatística de indicadores educacionais**

**Curitiba  
2024**

Tatiani Cristina Ara

## **Desempenho municipal do IDEB: modelagem estatística de indicadores educacionais**

Monografia apresentada ao Programa de Especialização em *Data Science* e *Big Data* da Universidade Federal do Paraná como requisito parcial para a obtenção do grau de especialista.

Orientador: Prof. José Luiz Padilha da Silva

Curitiba  
2024

# Desempenho municipal do IDEB: modelagem estatística de indicadores educacionais

Tatiani Cristina Ara<sup>1</sup>

<sup>1</sup>Universidade Federal do Paraná, Curitiba, PR, Brasil\*

O IDEB (Índice de Desenvolvimento da Educação Básica) é o principal indicador nacional Brasileiro referente à qualidade da educação básica. A compreensão dos fatores que afetam seu valor é de suma importância para o entendimento de possíveis caminhos de melhoria da educação e o indicador, conseqüentemente. O objetivo deste trabalho é realizar uma modelagem longitudinal do IDEB ao longo dos anos de 2017, 2019 e 2021, juntamente com variáveis educacionais e geográficas, a fim de compreender possíveis causas da variação do valor do indicador.

**Palavras-chave:** Educação básica, Indicadores, Dados longitudinais

The IDEB (Basic education development index) is Brazil's main national indicator regarding the quality of basic education. Understanding the factors that affect its value is of great importance for proper comprehension of factors that affect public education quality. The aim of this work is to conduct a longitudinal modelling of the IDEB over the years of 2017, 2019, and 2021, along with educational and geographical variables, in order to understand possible causes of variation in the indicator's value.

**Keywords:** Basic Education, Indicators, longitudinal data

## 1. Introdução

Sendo criada em 2005 pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), o Índice de Desenvolvimento da Educação Básica (IDEB) é uma medida criada pelo Governo Federal do Brasil para avaliar a qualidade do ensino nas escolas públicas do país. Ele combina dados de desempenho dos estudantes em exames padronizados (como o SAEB - Sistema de Avaliação da Educação Básica) com taxas de aprovação escolar.

O IDEB é calculado para cada escola, município, estado e para todo o país, sendo uma ferramenta importante para avaliar e acompanhar a qualidade da educação básica no Brasil. O objetivo principal do IDEB é promover melhorias no sistema educacional, incentivando o aumento do desempenho dos alunos e reduzindo as taxas de evasão escolar.

Com escala variando de 0 a 10, o IDEB é calculado a nível municipal, estadual, nacional e por escola. O objetivo do indicador é sumarizar informações que

permitam um entendimento intuitivo da qualidade da educação.

Apesar de sua utilidade inegável, o indicador é alvo de algumas críticas. Uma delas, relevante e investigada no estudo realizado neste trabalho, é a não consideração de diferenças socioeconômicas e regionais que podem afetar seu valor.

Neste trabalho, o foco é na realização de uma investigação, sob o ponto de vista de um modelo longitudinal, sobre o impacto de variáveis geográficas, educacionais e uma variável econômica sobre o IDEB, considerando os três anos de 2017, 2019 e 2021, a fim de verificar se a crítica mencionada anteriormente apresenta bons fundamentos.

Por conta da disponibilidade dos dados, a fim de evitar dados faltantes e ainda assim termos uma quantidade razoável de municípios estudados, a análise foi restrita ao ensino médio em escolas da rede pública em regiões urbanas. Por essa razão, as interpretações dos resultados são aplicáveis exclusivamente a esse contexto.

\*tatiani.ara@gmail.com

## 2. Dados

### 2.1. Descrição dos dados utilizados e contextualização

O objetivo é modelar a nota do IDEB para identificar se aspectos educacionais, geográficos e econômicos possuem associação com o desempenho dos alunos. O IDEB varia de 0 a 10, sendo uma escala contínua onde os valores mais elevados representam uma qualidade mais alta da educação básica.

Visando esse objetivo, foram consideradas algumas covariáveis referentes a indicadores educacionais por município brasileiro, e também a variável de PIB per capita, considerando-a como indicador socioeconômico do modelo. Todas as variáveis utilizadas são registradas ao longo de três anos.

O motivo da restrição a dados do ensino médio da rede pública em regiões urbanas é o excesso de dados faltantes referentes a localidades rurais, e o fato de que a base de dados utilizada possui IDEB apenas para a rede pública. Pela mesma razão, a análise foi restrita a três anos específicos em que o IDEB foi medido: 2017, 2019 e 2021. Dadas as restrições, foram analisados 2437 municípios que possuíam registros de todas as variáveis nos anos analisados.

Os dados utilizados neste trabalho foram coletados em três fontes diferentes. Os indicadores educacionais municipais utilizados na modelagem do IDEB foram obtidos a partir do portal da página de indicadores educacionais do INEP (*Indicadores educacionais - INEP* (2024)). Mais informações sobre o cálculo de cada um dos indicadores podem ser verificados na página mencionada.

Os valores do IDEB, já indexados por ano e código de município, foram obtidos em *Dados Ideb - base-dados* (2024), disponibilizados ano a ano conforme em uma base unificada que atende as necessidades da análise realizada.

Os dados da variável econômica utilizada (PIB per capita) foram obtidos a partir das planilhas disponibilizadas pelo IBGE (*PIB per Capita - IBGE* (2024)), também indexados por ano e município.

A escolha das variáveis em questão se deu por duas principais razões, sendo elas a possível relevância explicativa para o IDEB e disponibilidade. Ao acessar a plataforma de indicadores educacionais, muitas das variáveis disponíveis não possuem registros nos anos analisados, e outras não despertaram interesse na investigação. Além desses dois fatores, algumas variáveis

como taxa de aprovação não são apropriadas pois o próprio cálculo do IDEB depende de seu valor.

### 2.2. Descrição das variáveis utilizadas

Os indicadores educacionais selecionados para serem utilizados no modelo são: Regularidade docente, Média de alunos por turma, Taxa de distorção Idade-série e Adequação da formação docente. As covariáveis utilizadas para modelar o IDEB são listadas a seguir.

- CODMUN: Código do município.
- ANO: Ano do registro observado.
- PIBPERCAPITA: PIB per capita do município, dado em milhares de reais. Na análise exploratória, a variável foi explorada usando a escala logarítmica para facilitar a visualização. A mesma escala foi utilizada na fase de modelagem. O motivo dessa escolha é o fato das observações da variável serem muito concentrados na parte inferior da amplitude de valores (que é consideravelmente grande), o que torna difícil a visualização através de gráficos.
- CLASSE\_RCD: Classe predominante (maior percentual de docentes na classe) de regularidade entre docentes do município. Originalmente, o valor é composto de uma variável para cada categoria, representando o percentual de docentes na classe em questão. Para resumir, os valores foram definidos em uma variável categórica que representa a classe com maior percentual de docentes no município. As categorias são as seguintes:
  - RCD\_BAIXA: Regularidade baixa predominante
  - RCD\_MEDIABAIXA: Regularidade média-baixa predominante
  - RCD\_MEDIAALTA: Regularidade média-alta predominante
  - RCD\_RCDALTA: Regularidade alta predominante
- MAT\_TOTAL\_MEDIO: Média de alunos por turma.
- TDI\_TOTAL\_MEDIO: Taxa de distorção Idade-Série, variando entre 0 e 100.
- AFD\_MED1: Percentual de docentes com formação superior de licenciatura (ou bacharelado com complementação pedagógica) na mesma área da disciplina que leciona, variando entre 0 e 100.

- Região: Região geográfica do país (Centro-Oeste, Nordeste, Norte, Sudeste, Sul) referente ao município da observação.
- UF: Estado do município observado.

Mais informações sobre os indicadores educacionais e sua metodologia de cálculo podem ser encontradas em (*Indicadores educacionais - INEP, 2024*).

### 3. Análise exploratória

Após a extração, limpeza e padronização dos dados, uma análise exploratória ano a ano das covariáveis agrupadas por região geográfica foi conduzida, para que houvesse uma ideia inicial sobre possíveis dependências das covariáveis sobre região, bem como o próprio IDEB.

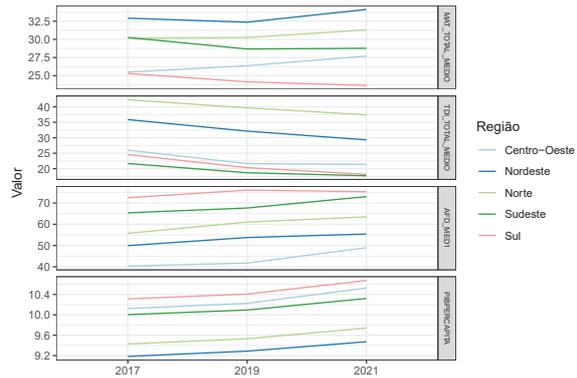
Começando com foco nas diferenças regionais que possivelmente explicam a variação do IDEB, na Tabela 1 é exibida a média anual em cada região do país nos anos em que os dados foram analisados, bem como o número de observações por ano.

Para a análise, foi utilizado o *software* R na versão 4.4.1 (R Core Team, 2024) com funções do pacote (Wickham et al., 2019) para manipulação e visualização dos dados.

**Tabela 1:** Média do IDEB Ano a Ano por região e número de observações por região

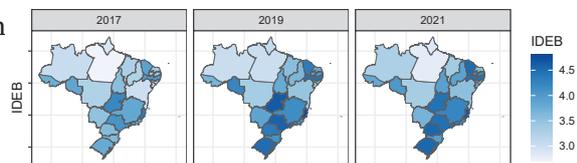
Região	Nº Obs por ano	2017	2019	2021
Centro-Oeste	191	4.04	4.45	4.26
Nordeste	763	3.51	3.92	3.97
Norte	164	3.13	3.32	3.31
Sudeste	884	3.97	4.38	4.36
Sul	435	3.64	4.36	4.46
Médias Anuais	-	3.66	4.09	4.07
Soma Total	2437	-	-	-

Na Figura 1, o objetivo é verificar se há uma diferença clara no valor das covariáveis contínuas utilizadas por região. Essa análise é necessária especialmente para que sejam consideradas interações entre a variável “Região” e as covariáveis contínuas. Pela Figura 1, é razoável supor que as médias das covariáveis divergem de maneira significativa por região geográfica do país.

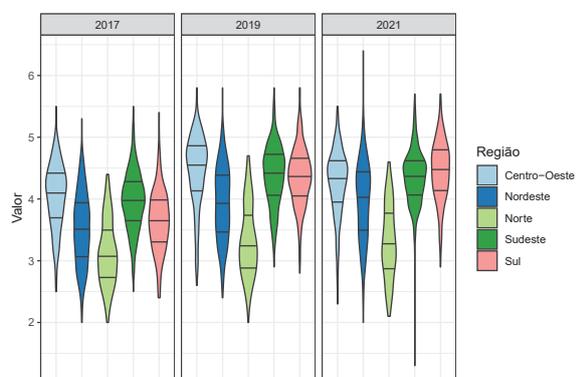


**Figura 1:** Variação das médias das covariáveis contínuas por região ao longo dos anos

O mapa da Figura 2 mostra a diferença entre níveis do IDEB por UF, bem como sua variação ao longo dos anos estudados. Assim como a Figura 1, as diferenças visualizadas fornecem uma oportunidade de explorar as diferenças geográficas no processo de modelagem. Na Figura 3, um gráfico de violino evidencia a diferença na distribuição do IDEB por região geográfica.

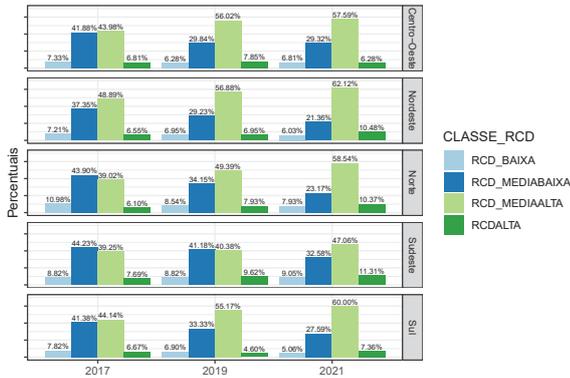


**Figura 2:** IDEB por UF ao longo dos anos



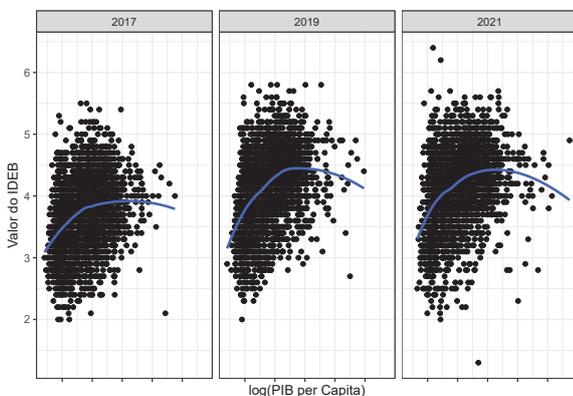
**Figura 3:** Variação do IDEB médio por Região

Na Figura 4 a mesma análise é realizada para verificar alguma diferença na variável CLASSE\_RCD considerando o agrupamento por região. Apesar de algumas diferenças, a distribuição das categorias da variável não parece divergir muito expressivamente de região para região.

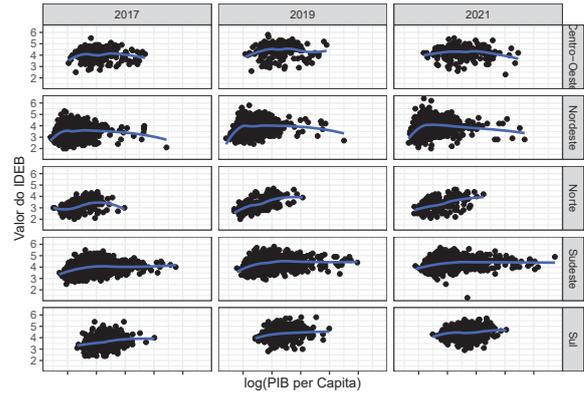


**Figura 4:** Distribuição da regularidade docente mais prevalente por Região

Considerando que existem muitos valores extremos em termos de PIB per capita, a escala log foi utilizada para visualizar a variação da média ano a ano, bem como a média ano a ano por região. A variável resultante da transformação varia entre 8,41 e 13,73. A curva em azul foi ajustada via regressão local (LOESS) através da função `geom_smooth()` para que seja possível visualizar uma relação suavizada entre as variáveis. Na Figura 5, a relação é exibida ano a ano para todos os municípios. Na Figura 6, a relação é dividida por região.



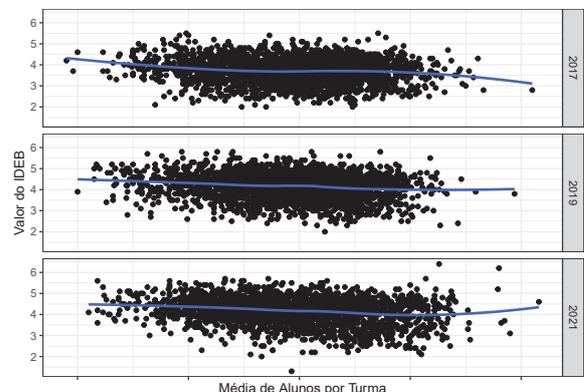
**Figura 5:** Dispersão ano a ano PIB x IDEB



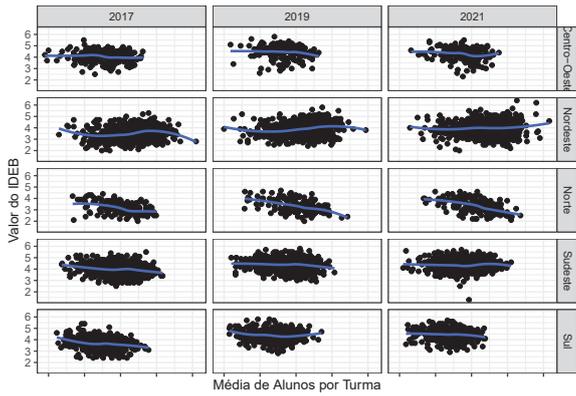
**Figura 6:** Dispersão ano a ano PIB x IDEB (Por região)

Apesar da dispersão grande nos pontos, é visível que existe uma relação positiva entre as duas variáveis indicada pelo ajuste da regressão LOESS em todos os anos. Regionalmente, a relação parece ser mais acentuada nas regiões Norte, Sul e Sudeste. Nas regiões Centro-Oeste e Nordeste, a relação não fica tão clara, de modo que inclusive aparenta ser inversamente proporcional.

Seguindo a mesma ideia dos gráficos referentes a PIB, pode ser visualizada a relação entre média de alunos por turma e IDEB. Na Figura 7, a relação pode ser vista para todos os municípios. Na Figura 8, os gráficos são gerados por região. Assim como na análise sobre PIB Per Capita, os gráficos de dispersão foram suavizados utilizando regressão LOESS para exista uma intuição sobre a inclinação da relação observada.



**Figura 7:** Dispersão ano a ano Média de alunos por turma x IDEB

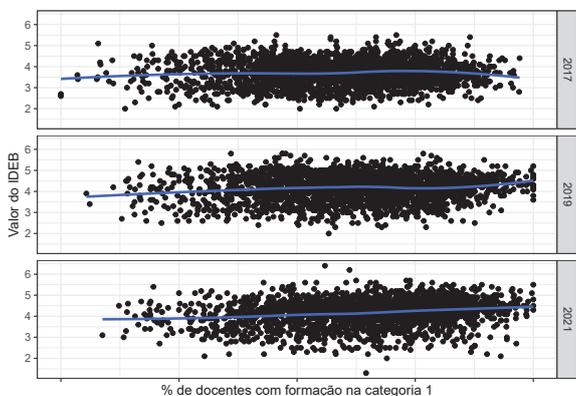


**Figura 8:** Dispersão ano a ano Média de alunos por turma (Por Região) x IDEB

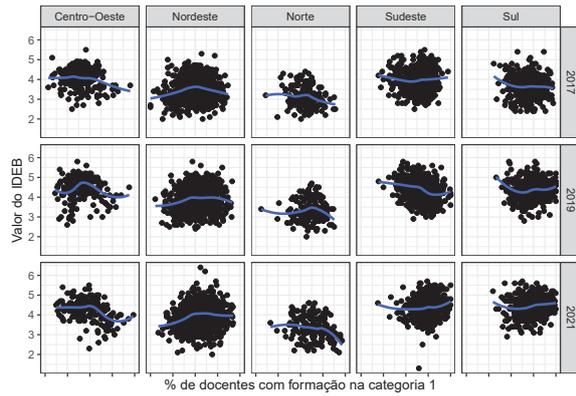
As variáveis parecem ter uma relação negativa muito sutil em relação ao IDEB quando avaliadas apenas ano a ano. Apesar disso, ao observar por região a relação parece ser mais significativa para algumas regiões do que outras, indicando um possível benefício em considerar a interação entre as MAT\_TOTAL\_MEDIO e a variável Região na especificação do modelo.

Com exceção das regiões Nordeste e Centro-Oeste, a relação parece ser bem acentuada de maneira negativa, ou seja, o IDEB parece diminuir conforme a média de alunos por turma do município cresce. A relação parece mais forte na região Norte.

Considerando a variável de adequação de formação docente visualizada de forma geral e por região nas Figuras 9 e 10, não parece haver uma relação positiva ou negativa clara com o IDEB. Sendo assim, a relação da covariável com o IDEB será melhor investigada na etapa de modelagem.



**Figura 9:** Dispersão ano a ano AFD x IDEB



**Figura 10:** Dispersão ano a ano AFD x IDEB (Por Região)

#### 4. Ajustes de modelos de efeitos mistos

Considerando a necessidade de levar em consideração as medidas repetidas, intercepto aleatório por indivíduo e provável necessidade de agrupamento dos indivíduos por estado, a escolha foi utilizar modelos lineares com efeitos mistos Pinheiro and Bates (2000).

##### 4.1. Seleção de modelos e diagnóstico do modelo selecionado

Considerando diversas alternativas de modelagem nesse contexto, foram ajustados, inicialmente, três modelos que diferem da seguinte maneira:

- Modelo 1 (md1): Modelo com covariáveis e intercepto aleatório considerando apenas município.
- Modelo 2 (md2): Modelo com covariáveis e intercepto aleatório dos municípios aninhados por estado.
- Modelo 3 (md3): Modelo com covariáveis e todas as interações com a covariável “Região” e intercepto aleatório dos municípios aninhados por estado.
- Modelo 4 (md4): Modelo com as covariáveis TDI\_TOTAL\_MEDIO e PIBPERCAPITA com a covariável “Região” e intercepto aleatório dos municípios aninhados por estado.

Os coeficientes ajustados, seus respectivos p-valores e comparação na qualidade dos ajustes dado pelo teste anova podem ser vistos nas Tabelas 2, 3 e 4.

Considerando os três modelos, o terceiro parece mais apropriado, com maior valor de log-verossimilhança, bem como apresentando o menor valor de AIC, o que indica que o agrupamento dos interceptos aleatórios por UF e consideração das interações resulta em um modelo melhor. Os valores são indicados na Tabela 5

**Tabela 2:** Coeficientes estimados do modelo 1

	Valor	Desvio-padrão	GL	t-valor	p-valor
(Intercepto)	3.7819	0.1464	4865	25.826	0.0000
TDI_TOTAL_MEDIO	-0.0233	0.0007	4865	-35.542	0.0000
AFD_MED1	-0.0006	0.0004	4865	-1.542	0.1232
MAT_TOTAL_MEDIO	-0.0060	0.0012	4865	-5.074	0.0000
CLASSE_RCDRCDC_MEDIABAIXA	0.0128	0.0177	4865	0.723	0.4700
CLASSE_RCDRCDC_MEDIAALTA	0.0357	0.0177	4865	2.012	0.0442
CLASSE_RCDRCDCDALTA	0.0297	0.0220	4865	1.353	0.1760
PIBPERCAPITA	0.0921	0.0136	4865	6.767	0.0000
RegiãoNordeste	-0.0907	0.0352	2432	-2.574	0.0101
RegiãoNorte	-0.4999	0.0439	2432	-11.394	0.0000
RegiãoSudeste	-0.0517	0.0325	2432	-1.589	0.1122
RegiãoSul	-0.1523	0.0359	2432	-4.237	0.0000
Ano2019	0.3549	0.0093	4865	38.266	0.0000
Ano2021	0.3048	0.0107	4865	28.567	0.0000

**Tabela 3:** Coeficientes estimados do modelo 2

	Valor	Desvio-padrão	GL	t-valor	p-valor
(Intercepto)	3.3957	0.2101	4865	16.160	0.0000
TDI_TOTAL_MEDIO	-0.0202	0.0007	4865	-28.064	0.0000
AFD_MED1	-0.0003	0.0004	4865	-0.804	0.4213
MAT_TOTAL_MEDIO	-0.0075	0.0012	4865	-6.449	0.0000
PIBPERCAPITA	0.0893	0.0124	4865	7.169	0.0000
RegiãoNordeste	0.1925	0.1901	21	1.013	0.3227
RegiãoNorte	-0.1231	0.2052	21	-0.600	0.5549
RegiãoSudeste	0.4495	0.2191	21	2.051	0.0529
RegiãoSul	0.1696	0.2353	21	0.721	0.4789
CLASSE_RCDRCDC_MEDIABAIXA	0.0096	0.0171	4865	0.560	0.5757
CLASSE_RCDRCDC_MEDIAALTA	0.0288	0.0171	4865	1.678	0.0934
CLASSE_RCDRCDCDALTA	0.0270	0.0214	4865	1.266	0.2055
Ano2019	0.3641	0.0093	4865	39.336	0.0000
Ano2021	0.3205	0.0106	4865	30.171	0.0000

O modelo apresenta apenas BIC menos desejável do que o segundo modelo, mas considerando as outras medidas, pode ser visto como o mais apropriado na interpretação dos coeficientes.

Observando o modelo apenas com as covariáveis e interceptos aleatórios por município, as variáveis com coeficientes estimados significativos são TDI\_TOTAL\_MEDIO, MAT\_TOTAL\_MEDIO e PIBPERCAPITA, além das regiões (exceto região sudeste, que não difere significativamente da região Centro-Oeste). Além disso, parece haver uma diferença significativa na média do IDEB em cada um dos três anos. O mesmo padrão se repete para o modelo considerando o aninhamento dos municípios por estado.

No terceiro modelo, com a inclusão da variável de região e todas as possíveis interações com as covariáveis, a interpretação das relações com a variável resposta muda. A variável CLASSE\_RCD continua sendo não significativa na explicação da variação do IDEB, mesmo considerando a interação com todas as regiões. Considerando a região Centro-Oeste como a base de comparação com as outras regiões, a variável AFD\_MED1 apresenta diferença significativa nos coeficientes estimados nas regiões Sul e Nordeste, indicando que essas regiões são as únicas que divergem sig-

**Tabela 4:** Coeficientes estimados do modelo 3

	Valor	Desvio-padrão	GL	t-valor	p-valor
(Intercepto)	4.8275	0.4420	10.922	0.0000	
TDI_TOTAL_MEDIO	-0.0193	0.0027	-7.222	0.0000	
AFD_MED1	-0.0018	0.0014	-1.323	0.1858	
MAT_TOTAL_MEDIO	-0.0060	0.0043	-1.410	0.1585	
PIBPERCAPITA	-0.0360	0.0365	-0.987	0.3239	
CLASSE_RCDRCDC_MEDIABAIXA	-0.0817	0.0616	-1.327	0.1847	
CLASSE_RCDRCDC_MEDIAALTA	-0.0836	0.0613	-1.365	0.1722	
CLASSE_RCDRCDCDALTA	-0.0452	0.0773	-0.585	0.5589	
RegiãoNordeste	-1.4660	0.5051	-2.903	0.0085	
RegiãoNorte	-1.0295	0.7127	-1.445	0.1633	
RegiãoSudeste	-0.9359	0.4998	-1.873	0.0751	
RegiãoSul	-2.9133	0.5961	-4.888	0.0001	
Ano2019	0.3578	0.0090	39.648	0.0000	
Ano2021	0.3093	0.0106	29.286	0.0000	
TDI_TOTAL_MEDIO:RegiãoNordeste	0.0041	0.0029	1.446	0.1482	
TDI_TOTAL_MEDIO:RegiãoNorte	0.0067	0.0036	1.836	0.0664	
TDI_TOTAL_MEDIO:RegiãoSudeste	0.0027	0.0029	0.919	0.3582	
TDI_TOTAL_MEDIO:RegiãoSul	-0.0205	0.0032	-6.483	0.0000	
AFD_MED1:RegiãoNordeste	0.0031	0.0016	1.982	0.0475	
AFD_MED1:RegiãoNorte	-0.0017	0.0020	-0.833	0.4046	
AFD_MED1:RegiãoSudeste	0.0003	0.0015	0.226	0.8210	
AFD_MED1:RegiãoSul	0.0049	0.0018	2.786	0.0054	
MAT_TOTAL_MEDIO:RegiãoNordeste	0.0048	0.0047	1.022	0.3068	
MAT_TOTAL_MEDIO:RegiãoNorte	-0.0012	0.0064	-0.186	0.8524	
MAT_TOTAL_MEDIO:RegiãoSudeste	-0.0015	0.0047	-0.316	0.7521	
MAT_TOTAL_MEDIO:RegiãoSul	-0.0075	0.0052	-1.441	0.1497	
PIBPERCAPITA:RegiãoNordeste	0.0429	0.0067	2.279	0.0227	
PIBPERCAPITA:RegiãoNorte	0.0666	0.0633	1.053	0.2925	
PIBPERCAPITA:RegiãoSudeste	0.1192	0.0399	2.986	0.0028	
PIBPERCAPITA:RegiãoSul	0.3149	0.0485	6.488	0.0000	
CLASSE_RCDRCDC_MEDIABAIXA:RegiãoNordeste	0.0936	0.0693	1.351	0.1769	
CLASSE_RCDRCDC_MEDIAALTA:RegiãoNordeste	0.1388	0.0687	2.020	0.0435	
CLASSE_RCDRCDCDALTA:RegiãoNordeste	0.1018	0.0862	1.181	0.2375	
CLASSE_RCDRCDC_MEDIABAIXA:RegiãoNorte	0.0856	0.0869	0.986	0.3243	
CLASSE_RCDRCDC_MEDIAALTA:RegiãoNorte	0.0414	0.0863	0.480	0.6313	
CLASSE_RCDRCDCDALTA:RegiãoNorte	0.0530	0.1087	0.488	0.6257	
CLASSE_RCDRCDC_MEDIABAIXA:RegiãoSudeste	0.1167	0.0667	1.750	0.0801	
CLASSE_RCDRCDC_MEDIAALTA:RegiãoSudeste	0.1173	0.0665	1.763	0.0779	
CLASSE_RCDRCDCDALTA:RegiãoSudeste	0.0688	0.0837	0.822	0.4111	
CLASSE_RCDRCDC_MEDIABAIXA:RegiãoSul	0.0527	0.0745	0.708	0.4791	
CLASSE_RCDRCDC_MEDIAALTA:RegiãoSul	0.1103	0.0743	1.484	0.1377	
CLASSE_RCDRCDCDALTA:RegiãoSul	0.0885	0.0947	0.935	0.3499	

**Tabela 5:** Medidas de ajuste para comparação dos modelos

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
md1	1	16	7240	7350	-3604		NA	NA
md2	2	17	6443	6561	-3205	1 vs 2	799	0
md3	3	45	6313	6623	-3111	2 vs 3	187	0

nificativamente da região Centro-Oeste. Apesar disso, os coeficientes estimados indicam que a variação do IDEB explicada pela variável é muito pequena.

Quando considerado o efeito das regiões, os coeficientes estimados da variável MAT\_TOTAL\_MEDIO perdem a significância, tanto na interação quanto no efeito individual da variável. Variações positivas no valor da variável TDI\_TOTAL\_MEDIO continuam tendo efeito negativo no IDEB, mas com diferentes níveis de significância nas interações com região, apresentando diferenças significativas em relação à região Centro-Oeste apenas na região Norte e Sul.

É interessante notar que a variável PIBPERCAPITA possui um coeficiente positivamente relacionado com IDEB quando observada isoladamente, e a interação com as regiões indica uma relação inversa nas regiões Nordeste, Sudeste e Sul comparadas à região Centro-Oeste. Isso significa que, tomando a região Centro-

Oeste como referência, todas as outras têm o valor do IDEB menos afetado por PIBPERCAPITA, exceto pela região Norte.

Em suma, as variáveis TDI\_TOTAL\_MEDIO e PIBPERCAPITA parecem ser as mais relevantes na explicação da variação do IDEB, sendo a primeira inversamente relacionada e a segunda com uma relação positiva. Portanto, é justificável ajustar um modelo considerando apenas essas variáveis em interação com as regiões, mantendo a mesma estrutura de efeitos aleatórios, e verificar seu resultado.

**Tabela 6:** Coeficientes do modelo 4

	Valor	Desvio-padrão	GL	t-valor	p-valor
(Intercepto)	4.6228	0.4351	4862	10.623	0.0000
TDI_TOTAL_MEDIO	-0.0183	0.0027	4862	-6.913	0.0000
PIBPERCAPITA	-0.0524	0.0360	4862	-1.456	0.1455
RegiãoNordeste	-1.2488	0.4984	21	-2.506	0.0205
RegiãoNorte	-1.0877	0.6993	21	-1.555	0.1348
RegiãoSudeste	-0.8513	0.4942	21	-1.723	0.0996
RegiãoSul	-2.7852	0.5883	21	-4.735	0.0001
Ano2019	0.3648	0.0088	4862	41.333	0.0000
Ano2021	0.3126	0.0100	4862	31.208	0.0000
TDI_TOTAL_MEDIO:RegiãoNordeste	0.0028	0.0028	4862	0.991	0.3217
TDI_TOTAL_MEDIO:RegiãoNorte	0.0063	0.0036	4862	1.742	0.0815
TDI_TOTAL_MEDIO:RegiãoSudeste	0.0009	0.0029	4862	0.321	0.7482
TDI_TOTAL_MEDIO:RegiãoSul	-0.0242	0.0031	4862	-7.817	0.0000
PIBPERCAPITA:RegiãoNordeste	0.1214	0.0421	4862	2.883	0.0040
PIBPERCAPITA:RegiãoNorte	0.0606	0.0624	4862	0.970	0.3321
PIBPERCAPITA:RegiãoSudeste	0.1205	0.0392	4862	3.069	0.0022
PIBPERCAPITA:RegiãoSul	0.3341	0.0480	4862	6.956	0.0000

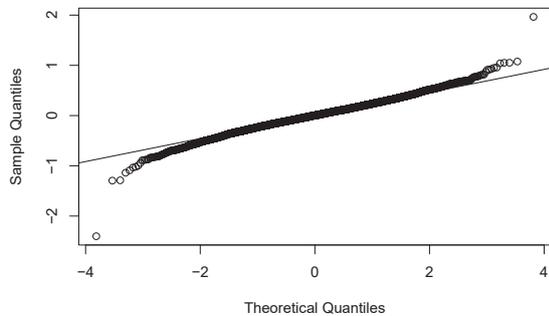
O modelo parece mais razoável do que os outros ajustados, considerando todas as medidas de qualidade de ajuste avaliadas observadas na Tabela 7. Os coeficientes estimados parecem confirmar o que foi analisado anteriormente, apresentando algumas variações dadas por região. A interpretação sobre a variação do PIB per capita considera a escala logarítmica da variável. Essa modificação foi necessária pois a utilização de valores absolutos da variável resultava em problemas na estimação do modelo. Em relação à taxa de distorção de idade, a interpretação é direta sobre o valor da taxa.

**Tabela 7:** Medidas de ajuste para comparação dos modelos

Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
md1	1	16	7240	7350	-3604	NA	NA
md2	2	17	6443	6561	-3205	1 vs 2	798.7
md3	3	45	6313	6623	-3111	2 vs 3	186.6
md4	4	20	6166	6304	-3063	3 vs 4	96.3

### 4.2. Diagnóstico de possíveis problemas no modelo

Para checar a validade da suposição de normalidade da variável resposta, é razoável observar o Gráfico Q-Q do modelo selecionado.



**Figura 11:** Gráfico Q-Q do modelo selecionado (md4)

Na Figura 11, é aparente que os erros padronizados do ajuste apresentam caudas leves nas extremidades dos resíduos, o que pode tornar a suposição de normalidade do modelo inadequada e resultar em um ajuste que não forneça as melhores estimativas possíveis dos coeficientes para municípios com valores de IDEB localizados nas extremidades da distribuição do IDEB. Com exceção dessa observação, os resíduos parecem bem ajustados à suposição de normalidade nas regiões mais centrais da distribuição.

### 4.3. Discussão sobre estimativas

Observando as estimativas do modelo considerado (Tabela 6), é evidente que a variável categórica mais relevante para explicar as diferenças do IDEB entre municípios é a variável Região. As Regiões Norte e Sudeste não apresentam diferenças estatisticamente significativas da região Centro-Oeste, mas as regiões Sul e Nordeste possuem um intercepto significativamente menor.

A Taxa de distorção de idade apresenta, sem considerar as interações, uma interpretação interessante: a cada 1% de aumento na taxa, espera-se uma redução de -0,02 no valor do IDEB. Essa variação é mais acentuada na região Sul, onde pode-se esperar um decréscimo de 0,024 no IDEB a cada aumento de um ponto percentual na taxa. As outras regiões não apresentam coeficientes significativos.

Em relação a PIBPERCAPITA em escala logarítmica, a variável não parece ser relevante para a região Centro-Oeste considerando um nível de confiança de 5%, mas apresenta coeficientes que divergem significativamente

da região Centro-Oeste para a região Nordeste, Sul e Sudeste. A Região onde a variável possui maior impacto, aparentemente, é a região sul, com aumento estimado de 0,3341 para cada aumento de unidade do logaritmo natural de PIBPERCAPITA. Esse aumento foi estimado em 0,1214 e 0,1205 para as regiões Nordeste e Sudeste, respectivamente.

A variável Ano também demonstrou ser significativa a nível de confiança de 5%, sendo as médias dos anos de 2019 e 2021 sendo maiores que a média de 2017 em 0,3648 e 0,3126, respectivamente.

## 5. Conclusões

O trabalho realizado demonstrou, através de uma análise exploratória e posteriormente através de modelagem estatística realizada com um modelo de efeitos mistos, possíveis causas de variação do valor do IDEB nos anos de ensino médio em regiões urbanas do Brasil.

Idealmente, mais variáveis seriam investigadas e levadas em consideração na modelagem para que se obtenha uma visão mais completa sobre fatores que afetam a qualidade da educação no país.

Apesar disso, a análise apresentada foi capaz de apontar evidências estatísticas, para os anos de 2017, 2019 e 2021, de que o contexto socioeconômico, regional e educacional tem efeito sobre os valores observados do IDEB.

## Referências

- Dados Ideb - basedosdados.* (2024). <https://basedosdados.org/dataset/96eab476-5d30-459b-82be-f888d4d0d6b9?table=8873f899-fdca-4ae4-808a-cdaaa1735d6a>.
- Indicadores educacionais - INEP.* (2024). Retrieved from <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/indicadores-educacionais>
- Pib per capita - ibge.* (2024). <https://www.ibge.gov.br/estatisticas/economicas/contas-nacionais/9088-produto-interno-bruto-dos-municipios.html?#t=downloads>.
- Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in s and s-plus*. New York: Springer. doi:
- R Core Team. (2024). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <https://www.R-project.org/>

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. doi: