

# Criação de um panorama da Taxa de distorção idade-série com base nos dados abertos do INEP

Abílio Nogueira Barros<sup>1</sup>, Andréza Leite de Alencar<sup>1</sup>,  
Rafael Ferreira Mello<sup>1,2</sup>

<sup>1</sup>Departamento de Computação - Universidade Federal Rural de Pernambuco (UFRPE)

<sup>2</sup>Centro de Estudos e Sistemas Avançados do Recife (CESAR)

abilionbarros@gmail.com, andreza.leite, rafael.mello@ufrpe.br

**Abstract.** *This article aims to demonstrate how the Taxa de Distorção Idade-série (TDI) can be used to monitor the Brazilian education scenario using open data. It also presents how these data can be acquired, processed, and classified in order to provide an overview of this rate for Brazilian municipalities. With only the TDI indicator, it is already possible to observe fluctuations in the development of Brazilian education during the years 2018 to 2022, which was the chosen time frame for the development of this work.*

**Resumo.** *Este artigo busca demonstrar como a Taxa de Distorção Idade-série (TDI) pode ser utilizada para acompanhar o cenário da educação brasileira utilizando dados abertos. É apresentado também como esses dados podem ser adquiridos, processados e classificados de modo a fornecer um panorama dessa taxa para municípios brasileiros. Com apenas o indicador TDI é possível já observar flutuações do desenvolvimento da educação brasileira durante os anos 2018 a 2022 que foi a faixa temporal escolhida para o desenvolvimento deste trabalho.*

## 1. Introdução

A educação básica no Brasil enfrenta diversos desafios que impactam diretamente a qualidade do ensino e o desenvolvimento dos estudantes. Entre os principais desafios, destacam-se a desigualdade educacional, a infraestrutura precária das escolas, a formação e valorização dos professores, a deficiência na gestão escolar e a falta de recursos financeiros [Schwartzman and Brock 2005]. Sobre o contexto educacional a análise de dados já é amplamente utilizada como em [Fonseca and Namen 2016] o artigo destaca a utilidade da mineração de dados para explorar e analisar os dados do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), a fim de identificar padrões e relações entre as variáveis. O que foi apresentado no trabalho pode fornecer uma maneira de gerar melhorias no sistema educacional brasileiro e para aprimorar a qualidade dos dados disponíveis para análise.

Dentro da vasta quantidade de dados e indicadores fornecidos pelo INEP, este trabalho busca fornecer um panorama sobre um de seus indicadores, a Taxa de Distorção Idade-série (TDI). A TDI é calculada pelo Ministério da Educação (MEC) como uma medida que indica a proporção de alunos que estão em séries diferentes daquelas adequadas à sua idade cronológica. O cálculo da TDI envolve a comparação entre a idade dos alunos

e a série em que estão matriculados. O MEC utiliza os dados do Censo Escolar, que é realizado anualmente, para obter informações sobre a idade dos alunos e a série em que estão matriculados. Com base nessas informações, é possível calcular a TDI utilizando a seguinte fórmula:

$$\text{TDI} = \frac{\text{Número de alunos com idade superior à recomendada para a série}}{\text{Total de alunos matriculados na série}} \times 100$$

A fórmula permite determinar a proporção de alunos que estão em séries superiores às recomendadas para sua idade. Quanto maior o valor da TDI, maior é a taxa de distorção idade-série. É importante destacar que a TDI é apenas um indicador e não reflete necessariamente a qualidade do ensino.

Outros trabalhos já abordaram o TDI a nível municipal e estadual, destacando o trabalho de [Bezerra 2019], que analisa dados específicos da etapa final do Ensino Fundamental e investiga os principais fatores que contribuem para a ocorrência desse fenômeno, como a idade dos alunos, a série em que estão matriculados, a evasão escolar e a repetência. O objetivo é compreender a magnitude da distorção idade-série na rede municipal de Natal e identificar possíveis estratégias para enfrentar esse desafio educacional.

Diferentemente do trabalho citado, o objetivo deste é fornecer esse panorama a nível de todo o território brasileiro, podendo assim fornecer esse panorama e realizar o acompanhamento em anos importantes para a educação básica.

## 2. Aquisição e Processamento dos dados

O responsável pela publicação desses dados educacionais é o INEP, a entidade fornece os dados a nível de microdados, como o censo da educação básica e o censo da educação superior, como outros indicadores educacionais como taxas de aprovação, reprovação e abandono. Os dados de contexto educacional foram retirados em sua totalidade do site do INEP. Este trabalho busca criar um mapeamento do TDI para assim criar um panorama geral desses resultados levando em consideração a faixa temporal de 2018 a 2022, anos estratégicos que representam períodos pré e pós eventos pandêmicos vivenciados em 2020 e 2021, anos esses que trouxeram novos desafios [Dias et al. 2020] além dos já mencionados no começo desse trabalho, tanto para o momento de vivência da educação remota, quanto para sua retomada a sala de aulas.

Os dados foram adquiridos direto do portal do INEP, no portal é possível selecionar a nível federal, estadual, municipal e escolar. Dado que o primeiro objetivo deste trabalho é analisar o panorama das cidades a nível estadual.

### 2.1. Processamento de dados

Destacamos inicialmente que não é disposto nenhum dicionário de dados para esse tipo de recurso, porém suas colunas são de fácil entendimento e com isso é possível realizar sua utilização. Para essa primeira etapa do projeto foi utilizado a linguagem de programação Python<sup>1</sup> e sua biblioteca para análise de dados Pandas<sup>2</sup> determinando assim que todo o processo pode ser feito com ferramentas totalmente *open source*, trazendo a possibilidade de replicação dessa processo de forma gratuita.

---

<sup>1</sup><https://www.python.org/>

<sup>2</sup><https://pandas.pydata.org/>

As etapas seguidas no processamento dos dados estão ilustradas na figura 1, demonstrando cada etapa do processo que foi seguido para criação da base de dados.

### Extração dos dados

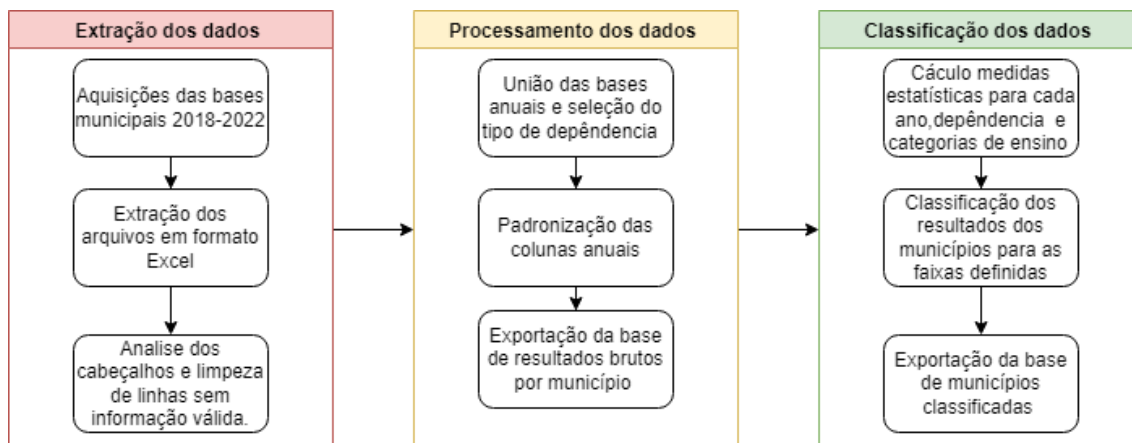


Figura 1. Fluxo de processamento dos dados utilizado.

Os dados foram extraídos diretamente do INEP<sup>3</sup>, onde foram selecionados os anos de interesse, e foi avaliada as colunas que seriam o alvo do estudo bem como a retirada de linhas desnecessárias e outros dados que não tratariam do escopo desse projeto.

### Processamento dos dados

Nessa etapa, foi preciso analisar os dados buscando compreender quais etapas educacionais eram informadas e quais seriam escolhidas para esse trabalho, a dependência administrativa e localidade. Com base na Lei de Diretrizes e Bases da Educação Nacional<sup>4</sup>, é sabido que ao município só compete a estrutura escolar até os anos finais do ensino fundamental, pois a gerência de ensino do ensino médio é de responsabilidade do governo estadual. Além disso, ao criar a base de dados será utilizado o filtro de dependência administrativa para pública, pois o foco do estudo está na rede pública de ensino.

### Classificação dos dados

Os municípios foram classificados individualmente por ano, por etapa escolar (Fundamental anos iniciais, Fundamental anos finais e Ensino médio) e tipo da localidade analisada. Municípios que por algum motivo não tiveram seus resultados apresentados não foram contabilizados.

A classificação é baseada na divisão dos dados em quartis, que é uma técnica estatística para dividir um conjunto de valores em quatro partes iguais, cada uma representando uma porcentagem da distribuição dos dados. Fazendo com que assim os dados sejam sempre calculados baseado naquele ano e com a realidade expressa naquele espaço de tempo, buscando assim levar em conta as flutuações do TDI através dos anos estudados. A classificação foi definida da seguinte forma:

- Valor TDI no intervalo de 0 a 25% (primeiro quartil), será classificado como 1.
- Valor TDI no intervalo de 25% a 50% (segundo quartil), será classificado como 2.
- Valor TDI no intervalo de 50% a 75% (terceiro quartil), será classificado como 3.

<sup>3</sup><https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/indicadores-educacionais/taxas-de-distorcão-idade-serie>

<sup>4</sup>[https://www.planalto.gov.br/ccivil\\_03/leis/l19394.htm](https://www.planalto.gov.br/ccivil_03/leis/l19394.htm)

- Valor TDI acima de 75% (quarto quartil), será classificado como 4.

Formalizando assim a classificação dos municípios em sua determinada faixa educacional em determinado ano. Registrando que quanto menor a faixa em que o município for classificado, menor foi seu TDI, sendo assim tendo um melhor desempenho educacional neste indicador.

Exemplificando a necessidade de gerar as faixas dinamicamente para cada ano, para o ano de 2018 o TDI foi de 2.9 para o ensino fundamental anos iniciais, o primeiro quartil. Então os melhores serão os resultados menores que esse, já para 2020 o primeiro quartil foi de 3.5, onde mais uma vez os melhores que receberam a faixa 1 foram apontados com valores abaixo dessa faixa, garantindo assim que o sistema de classificação irá respeitar as flutuações educacionais que existiram ao decorrer da faixa de tempo analisada.

### 3. Resultados

Após a conclusão da etapa de aquisição e processamento é possível realizar novas agregações buscando obter algumas informações sobre os dados dispostos. Por se tratar da primeira etapa desse estudo, as informações apresentadas a seguir vão ser sobre a mesma localidade *Total*, que corresponde ao resultado do urbano e rural juntos. O objetivo da classificação nos intervalos de quantil é poder dividir em grupos com em média 1380 a 1400 municípios em cada faixa classificatória, e assim podendo observar quais os valores que definiram o limiar de classificação e assim observar sua mudança durante os anos.

#### 3.1. Ensino fundamental anos iniciais e finais

Para os anos iniciais podemos observar na tabela 1 que indica os valores que serviram como base para essa classificação que a variação entre o primeiro quartil, onde estão alocado os municípios com menores índices, teve seu valor reduzido pela metade ao compararmos o ano de 2018 com o de 2022, efeito que pode ser observado para as demais faixas, demonstrando assim que mesmo com as adversidades pode-se reduzir o valor do TDI. Já para os anos finais podemos visualizar ,na tabela 2, que houve uma queda crescente do TDI para o limiar da classificação das faixas. Mesmo que não da mesma magnitude do anterior, vemos uma queda crescente.

Ano	Q1	Q2	Q3
2018	4.3	9.2	17.4
2019	4.0	8.5	16.4
2020	3.6	7.9	15.0
2021	2.7	5.6	10.8
2022	2.3	4.8	9.0

**Tabela 1. Valores dos quartis por ano para o Ensino Fundamental anos iniciais**

Ano	Q1	Q2	Q3
2018	16.7	25.9	37.2
2019	15.5	24.3	35.575
2020	15.4	23.9	34.275
2021	13.2	21.3	32
2022	11.0	18.2	28.2

**Tabela 2. Valores dos quartis por ano para o Ensino Fundamental anos finais**

### 3.2. Estudo de caso do estado de Pernambuco

Outra possibilidade que temos é analisar a nível estadual a taxa TDI, visando mensurar a quantidade de municípios em cada faixa. Selecionando o estado de Pernambuco, todos os seus 185 municípios foram relacionados. As colunas representam a quantidade de municípios em cada faixa. Na tabela 3 podemos observar que o estado, de forma coletiva,

ANO	F1	F2	F3	F4
2018	3	30	99	53
2019	2	35	102	46
2020	3	38	107	37
2021	4	35	108	38
2022	3	36	93	53

**Tabela 3. Quantidade de municípios em cada faixa para Ensino Básico Anos Iniciais**

ANO	F1	F2	F3	F4
2018	0	10	73	102
2019	0	14	79	92
2020	1	20	76	88
2021	3	32	69	81
2022	5	29	60	91

**Tabela 4. Quantidade de municípios em cada faixa para Ensino Básico Anos Finais**

conseguiu manter mais da metade de seus municípios entre as faixas 2 e 3 não conseguindo alavancar seus resultados durante o período de tempo analisado. Já na observação da tabela 4, que reporta os dados relativo aos anos finais, mostra que existiu uma progressão para outras faixas onde após 2020, mesmo sendo um ano pandêmico, se conseguiu avançar na quantidade de municípios em faixas melhores como a 2 e 3, e pela primeira vez nos anos observados no estudo alcançar a faixa 1.

### 4. Conclusão

Em conclusão, este estudo permitiu-nos apresentar uma visão abrangente sobre o TDI e ressaltar a importância do seu acompanhamento para o monitoramento efetivo da educação básica brasileira. Ao compreendermos as nuances e desafios associados ao TDI, podemos desenvolver estratégias mais eficazes para aprimorar o sistema educacional. Ressalta-se que esse não é um indicador direto de qualidade de ensino e sim um indicador para acompanhamento sobre a idade que os estudantes estão sendo matriculados na rede pública de ensino. Existem outros indicadores como Sistema Nacional de Avaliação da Educação Básica (SAEB) que podem atestar a qualidade do ensino das redes de ensino. O acompanhamento do TDI tende a servir de guia para que outros fenômenos escolares, como evasão e reprovações, possam ser investigadas mais de perto. Podendo assim buscar entender qual é a real faixa escolar que aquela localidade está servindo e não apenas a faixa ideal que é relatada na literatura.

O desenvolvimento da base de dados que foi desenvolvida para o desenvolvimento desse panorama pode ser encontrada nesse endereço <https://zenodo.org/record/8105876> podendo ser aplicados em algoritmos inteligentes de classificação ou agrupamento.

Na continuidade desse projeto temos o objetivo de incorporar outros indicadores definidos pelo INEP, tal como as taxas de aproveitamento, como também dados de quantitativos e qualitativos sobre as matrículas disponíveis no Censo educacional da educação básica com o intuito de realizar o cruzamento de todas as informações que compõem o TDI a fim de rastrear quais estão trazendo mais impacto para sua redução de forma positiva e contínua através dos anos.

## Referências

- Bezerra, A. R. d. F. (2019). Distorção idade-série nos anos finais do ensino fundamental: o caso da rede municipal de ensino do natal/rn. Master's thesis, Brasil.
- Dias, G. N., Vogado, G. E. R., Barreto, W. D. L., da Silva Junior, W. L., da Silva Barbosa, E., Rodrigues, A. E., Junior, A. F. S., and Costa, C. A. C. (2020). Retorno às aulas presenciais no sistema educacional do estado do pará-brasil: Obstáculos e desafios durante a epidemia de covid-19 (sars-cov-2). *Brazilian Journal of Development*, 6(6):37906–37924.
- Fonseca, S. O. d. and Namen, A. A. (2016). Mineração em bases de dados do inep: uma análise exploratória para nortear melhorias no sistema educacional brasileiro. *Educação em Revista*, 32:133–157.
- Schwartzman, S. and Brock, C. (2005). Os desafios da educação no brasil. *Rio de Janeiro: Nova Fronteira*, 1320.