

UNIVERSIDADE FEDERAL DE ITAJUBÁ  
PROGRAMA DE PÓS GRADUAÇÃO EM  
CIÊNCIA E TECNOLOGIA DA COMPUTAÇÃO

**Ideb, as unidades da federação e o perfil das escolas públicas:  
uma análise exploratória de dados não supervisionada.**

Igor Moreira Alves

Itabira, Novembro de 2022

UNIVERSIDADE FEDERAL DE ITAJUBÁ  
PROGRAMA DE PÓS GRADUAÇÃO EM  
CIÊNCIA E TECNOLOGIA DA COMPUTAÇÃO

Igor Moreira Alves

**Ideb, as unidades da federação e o perfil das escolas públicas:  
uma análise exploratória de dados não supervisionada.**

Dissertação submetida ao Programa de Pós-Graduação em Ciência e Tecnologia da Computação como parte dos requisitos para obtenção do Título de Mestre em Ciência e Tecnologia da Computação

**Área de Concentração:** Matemática Computacional

**Linha de pesquisa:** Inteligência Artificial

**Orientador:** Prof. Dr. Carlos Henrique Silveira

Novembro de 2022  
Itabira - MG

# AGRADECIMENTOS

Agradeço primeiramente a Deus pelos dons fornecidos, a vida, a saúde e a capacidade intelectual. Depois agradeço a minha família que me amparou a todo momento, a minha esposa Karine pela paciência, compreensão e seu amor. A minha filha que me reensinou o que é o AMOR.

Ao meu orientador, agradeço pela paciência e toda ajuda fornecida, agradeço também aos professores João e Rodrigo essenciais para a realização deste trabalho. E agradeço também a todos que me auxiliaram direta ou indiretamente por essa conquista morosa e árdua. A UNIFEI agradeço por me oportunizar a conquista de um título que agregou muito, tanto no quesito conhecimento quanto no quesito amadurecimento.

Agradeço a todos amigos que sentiram minha falta, dizem que quanto mais difícil é a conquista, mais saborosa é a vitória. Por isso, vamos celebrar!

*Necessitamos sempre de ambicionar alguma coisa que, alcançada, não nos torna sem ambição.*  
Carlos Drummond de Andrade

# RESUMO

Essa dissertação busca o uso da ciência dos dados, por meio de análise exploratória não supervisionada, comprovar o que a literatura já sabe e/ou apresentar novas descobertas sobre a influência da infraestrutura sobre a educação básica brasileira. A infraestrutura escolar não se resume apenas à questão arquitetônica das escolas, mas também ao ambiente educativo e administrativo, equipamentos, recursos educacionais, práticas, currículos e processo de ensino e aprendizagem. A coleta de dados foi realizada sobre os dados abertos do Censo Escolar 2019 (Educação Básica) e do Índice de Desenvolvimento da Educação Básica (Ideb), relativos aos anos 2005-2019. A escolha do ano de 2019 foi por se tratar do último ano que as escolas apresentaram resultados antes da influência da pandemia COVID-19. Após vários tratamentos nos dados e a escolha de atendimento apenas do segmento inicial do ensino fundamental, aplicou-se duas metodologias de análise: Correlograma e Análise Fatorial (FA). Para uma clareza nos resultados, criou-se novos atributos referentes aos entes federativos que permitiu identificar quais estados e perfis de escolas estão melhores relacionados ao crescimento e aos bons resultados do Ideb. Para essas correlações optou-se pela Sigma da Cópula Gaussiana que leva em consideração os dados categóricos e contínuos e, ainda, gerou uma matriz positiva definida. O Correlograma gerou uma matriz quadrada que apresentou os relacionamentos dos atributos em um dendrograma de heatmap. Dividido em 4 grandes grupos, cada um apresentou características específicas e relações com os entes federativos. O primeiro grupo apresentou forte relacionamento com as infraestruturas básicas; o segundo grupo, com o Ideb e as infraestruturas mais sofisticadas; o terceiro grupo, apresentou poucas relações entre os atributos; e último grupo forte correlações negativas e conteve maior precariedade nas infraestruturas. Após averiguação de compatibilidade da base de dados para aplicação do FA, foram estimados que 10 fatores seriam adequados a esse estudo. Quatro fatores apresentaram associação com os atributos do Ideb, o foco deste trabalho. Observou-se também três padrões nos atributos que elencaram bons resultados no Ideb com diferentes infraestruturas, políticas e/ou propostas educacionais: o primeiro grupo, norteador por SP, apresentou saneamentos básicos ofertados pelo serviço público, internet de qualidade com uso na aprendizagem e órgãos escolares; o segundo grupo, encabeçado por MG, indica associação a flexibilização no ensino tradicional, com ciclos escolares e salas não seriadas; o terceiro grupo ficou marcado pelas atividades complementares e atendimento especializado, representado pelo CE. Em contradição a esses parâmetros, escolas com a modalidade EJA, principalmente do nordeste, tendem a ter menores resultados no Ideb. Os outros 6 fatores agregaram muitas informações relevantes, inclusive relacionadas às correlações e anti correlações dos entes federativos e atributos específicos. Como visto, a ciência dos dados tem muito a agregar à área da educação. Espera-se com trabalhos futuros acrescentar ainda mais dados, como estudos longitudinais sobre o Ideb e agregar outros índices educacionais como o Ioeb e o nível socioeconômico da população.

# ABSTRACT

This dissertation seeks to use data science, through unsupervised exploratory analysis, to prove what the literature already knows and/or present new discoveries about the influence of infrastructure on Brazilian basic education. School infrastructure is not limited to the architectural issue of schools, but also to the educational and administrative environment, equipment, educational resources, practices, curricula and the teaching and learning process. Data collection was carried out on open data from the 2019 School Census (Basic Education) and the Basic Education Development Index (Ideb), for the years 2005-2019. The choice of the year 2019 was because it was the last year that schools presented results before the influence of the COVID-19 pandemic. After several data treatments and the choice of attending only the initial segment of fundamental education, two analysis methodologies were applied: Correlogram and Factor Analysis (FA). For clarity in the results, new attributes were created referring to the federative entities that allowed identifying which states and school profiles are better related to the growth and good results of the Ideb. For these correlations, the Sigma of the Gaussian Copula was chosen, which takes into account the categorical and continuous data and also generated a definite positive matrix. The Correlogram generated a square matrix that presented the attribute relationships in a Heatmap Dendrogram. Divided into 4 large groups, each one had specific characteristics and relationships with federal entities. The first group had a strong relationship with basic infrastructure; the second group, with IDEB and the most sophisticated infrastructures; the third group showed few relationships between the attributes; and the last group had strong negative correlations and contained greater precariousness in infrastructure. After verifying the compatibility of the database for the application of the FA, it was estimated that 10 factors would be suitable for this study. Four factors were associated with the attributes of the Ideb, the focus of this work. Three patterns were also observed in the attributes that listed good results in the Ideb with different infrastructures, policies and/or educational proposals: the first group, guided by São Paulo state, presented basic sanitation offered by the public service, quality internet for use in learning and institutions schoolchildren; the second group, headed by Minas Gerais state, indicates an association with flexibility in traditional teaching, with school cycles and non-serial classrooms; the third group was marked by complementary activities and specialized care, represented by the Ceará state. In contradiction to these parameters, schools with the EJA modality, mainly in the northeast, tend to have lower results in the Ideb. The other 6 factors added a lot of relevant information, including those related to the correlations and anti-correlations of the federative entities and specific attributes. As seen, data science has a lot to add to the field of education. Future works are expected to add even more data, such as longitudinal studies on the Ideb and to add other educational indices such as the Ioeb and the socioeconomic level of the population.

## LISTA DE FIGURAS

|   |    |
|---|----|
| Figura 1 - Principais entradas de dados .....   | 26 |
| Figura 2 - Principais operações e transformação na tabela SCHOOL2019 .....  | 29 |
| Figura 3 - Principais operações e transformação de dados nas tabelas Ideb .....   | 32 |
| Figura 4 - Principais operações para a junção das tabelas SCHOOLS2019 e Idebs ...   | 33 |
| Figura 5 - Workflow geral descrevendo as fases do tratamento dos dados geral .....  | 44 |
| Figura 6 - Gráfico de densidade das variações e notas Ideb das escolas na base UNI .  | 45 |
| Figura 7a - Gráfico das distribuições de densidade da variação do Ideb por estado ...   | 46 |
| Figura 7b - Gráfico das distribuições de densidade das notas do Ideb por estado .....   | 47 |
| Figura 8 - Gráfico da variância relativa dos autovalores da decomposição PCA .....  | 53 |
| Figura 9 - Gráfico do resultado da PCA sobre a tabela INI_QT .....  | 54 |
| Figura 10 - Gráfico das distâncias ordenadas do k vizinho sobre a base INI-QT .....   | 54 |
| Figura 11 - Gráfico dos atributos QT em função das escolas ordenadas pelas maiores distâncias kNN .....   | 55 |
| Figura 12 - Gráfico comparativo matrizes de correlações: Mista, Pearson e Sigma ....  | 57 |
| Figura 13a - Correlograma em heatmap destacando as correlações Sigma advindas do Gaussian Copula sobre 176 atributos da base INI-FIL - grupos 1 e 2 ..... | 58 |
| Figura 13b - Correlograma em heatmap destacando as correlações Sigma advindas do Gaussian Copula sobre 176 atributos da base INI-FIL - grupos 3 e 4 ..... | 59 |
| Figura 14 - Gráficos para a estimação do número de fatores a serem usados na FA ...   | 62 |
| Figura 15 - Gráfico da variância relativa da variância explicada por cada fator da decomposição FA dos 10 primeiros autovetores .....                     | 65 |
| Figura 16 - Indicação visual dos coeficientes da matriz padrão da decomposição FA da matriz Sigma de INI_FIL .....  | 67 |
| Figura 17 - Projeção das escolas no espaço dos fatores F1 x F2 e F3 x F7 .....  | 76 |
| Figura 18 - Projeção das escolas no espaço dos fatores F3 e F7 - Ceará, Rio de Janeiro e Atividade complementar .....                                     | 77 |
| Figura 19 - FA e Correlações Sigma com IDEB_VAR_INI e IDEB_SCORE_INI .....  | 78 |
| Figura 20 - FA e Correlações Sigma com os estados da federação .....  | 80 |

## LISTA DE TABELAS E QUADROS

|  |    |
|--|----|
| Tabela 1 - Limite superior e inferior das proficiências do Ideb .....                | 22 |
| Quadro 1 - Prefixos adotados nos atributos nas tabelas SCHOOL2019 e Idebs .....      | 27 |
| Tabela 2 - Saída da função describe na tabela SCHOOL2019 .....                       | 27 |
| Tabela 3 - Amostra de registros da tabela IDEB_INI não interpolada .....             | 30 |
| Tabela 4 - Amostra de registros da tabela INTER como uma IDEB_INI interpolada ..     | 30 |
| Tabela 5 - Amostra de registros da tabela DIFF montada da tabela INTER em 4 .....    | 31 |
| Tabela 6 - Amostra da tabela GROW como soma cumulativa da tabela DIFF em 5 ...       | 31 |
| Tabela 7 - Amostra de registros da tabela UNI .....                                  | 34 |
| Tabela 8 - Tipos de valores na variação Ideb .....                                   | 47 |
| Tabela 9 - Quantidade de escolas conforme tipos de valores da nota Ideb de 2019 .... | 48 |
| Tabela 10 - Números absolutos e relativos de NAs nos 51 atributos .....              | 50 |
| Tabela 11 - Atributos dependentes e suas dependências na base SCHOOL2019 .....       | 52 |
| Tabela 12 - Últimos autovalores das correlações Pearson(INI) e Sigma(INI_IMP) ....   | 52 |
| Tabela 13 - Filtragem de valores extremos .....                                      | 56 |
| Tabela 14 - Sumário estatístico básico do KMO aplicado à matriz Sigma de INI-FIL .   | 62 |
| Tabela 15 - Listagem de atributos com communalities menores que 0.495 .....          | 64 |
| Tabela 16 - Análise dos fatores F1 e F2 .....  | 69 |
| Tabela 17 - Análise dos fatores F3, F4, F7 e F9 .....                                | 71 |
| Tabela 18 - Análise dos fatores F5, F6, F8 e F10 .....                               | 74 |

## LISTA DE ABREVIATURAS

|      |  |
|------|--|
| AEE  | Atendimentos Educacionais Especializados                               |
| BNCC | Base Nacional Comum Curricular   |
| Enem | Exame Nacional do Ensino Médio   |
| FA   | Análise Fatorial ( <i>Factor Analysis</i> )                            |
| IA   | Inteligência Artificial  |
| Ideb | Índice de Desenvolvimento da Educação Básica                           |
| Inep | Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira |
| LDB  | Lei de Diretrizes e Bases da Educação Nacional                         |
| LGPD | Lei Geral de Proteção de Dados Pessoais                                |
| MCAR | Missing Completely at Random   |
| NA   | Não disponível ( <i>Not Avalable</i> )                                 |
| SMAE | Scaled Mean Absolute Error   |
| RSL  | Revisão Sistemática de literatura                                      |
| Saeb | Sistema Nacional de Avaliação da Educação Básica                       |
| TIC  | Tecnologia da Informação e Comunicação                                 |



# SUMÁRIO

|  |    |
|--|----|
| <b>1 INTRODUÇÃO</b> .....                                    | 7  |
| 1.1 ESCOPO DA PESQUISA .....                                 | 10 |
| 1.2 JUSTIFICATIVA .....                                      | 11 |
| 1.3 ESTRUTURA DO TRABALHO .....                              | 12 |
| <b>2 OBJETIVOS</b> .....                                     | 13 |
| 2.1 OBJETIVO GERAL .....                                     | 13 |
| 2.2 OBJETIVOS ESPECÍFICOS .....                              | 13 |
| <b>3 REVISÃO DA LITERATURA</b> .....                         | 14 |
| 3.1 CIÊNCIA DOS DADOS NA EDUCAÇÃO .....                      | 14 |
| <b>3.1.1 Bases de dados internacionais</b> .....             | 14 |
| <b>3.1.2 Ciências dos dados na educação brasileira</b> ..... | 16 |
| 3.2 DADOS ABERTOS E MICRODADOS DO INEP .....                 | 17 |
| 3.3 IDEB COMO MÉTRICA NA EDUCAÇÃO .....                      | 19 |
| <b>3.3.1 Como é calculado o Ideb</b> .....                   | 20 |
| 3.4 OUTRAS RSL SOBRE CIÊNCIA DE DADOS APLICADA EM EDUCAÇÃO . | 22 |
| <b>4 METODOLOGIA</b> .....                                   | 24 |
| 4.1 ESCOLHA DOS DADOS .....                                  | 24 |
| 4.2 CONJUNTO DE DADOS ( <i>DATA SET</i> ) .....              | 26 |
| 4.3 PRIMEIRA EXPLORAÇÃO DE DADOS .....                       | 26 |
| 4.4 ADEQUAÇÃO DE DADOS .....                                 | 28 |
| <b>4.4.1 Adequação de dados na tabela SCHOOL2019</b> .....   | 28 |
| <b>4.4.2 Adequação de dados nas tabelas Idebs</b> .....      | 30 |
| 4.5 JUNÇÃO DAS TABELAS SCHOOL2019 E IDEBs .....              | 32 |
| 4.6 FOCO NOS ANOS INICIAIS .....                             | 35 |
| 4.7 PREPARAÇÃO .....   | 35 |
| 4.8 IMPUTAÇÃO .....  | 36 |
| 4.9 COLINEARIDADES .....                                     | 37 |
| 4.10 SEGUNDA EXPLORAÇÃO DE DADOS COM PCA .....               | 37 |
| 4.11 VALORES EXTREMOS ( <i>OUTLIERS</i> ) .....              | 38 |
| 4.12 CORRELAÇÕES .....                                       | 39 |
| 4.13 AGRUPAMENTOS HIERÁRQUICOS .....                         | 39 |

|  |            |
|--|------------|
| 4.14 TERCEIRA EXPLORAÇÃO DE DADOS: REDUÇÃO DIMENSIONAL E SELEÇÃO DE ATRIBUTOS POR FA ..... | 40         |
| 4.15 AFERIÇÃO DA BASE NA FA .....  | 42         |
| 4.16 NÚMERO DE FATORES NA FA .....   | 43         |
| 4.17 OBSERVÁVEIS NA FA .....   | 43         |
| 4.18 WORKFLOW GERAL .....  | 44         |
| <b>5 RESULTADOS E DISCUSSÃO .....</b>  | <b>45</b>  |
| 5.1 DISTRIBUIÇÕES DAS NOTAS DO IDEB .....  | 45         |
| 5.2 IMPUTAÇÃO .....  | 50         |
| 5.3 ANÁLISE DAS COLINEARIDADES .....   | 52         |
| 5.4 PCA SOBRE INI_QT .....   | 53         |
| 5.5 CARACTERIZAÇÃO E ELIMINAÇÃO DE VALORES EXTREMOS POR DBSCAN .....                       | 54         |
| 5.6 CORRELAÇÕES .....  | 57         |
| 5.7 RESULTADOS DA AFERIÇÃO DA BASE PARA A FA .....   | 61         |
| 5.8 ESTIMAÇÃO DO NÚMERO DE FATORES NA FA .....   | 62         |
| 5.9 RESULTADOS E ANÁLISES DA FA .....  | 63         |
| <b>6 CONCLUSÃO .....</b>   | <b>81</b>  |
| 6.1 LIMITAÇÕES E PERSPECTIVAS .....  | 83         |
| <b>REFERÊNCIAS .....</b>   | <b>85</b>  |
| <b>APÊNDICE A - CÁLCULOS DO IDEB .....</b>   | <b>93</b>  |
| <b>APÊNDICE B - ESCALA DE PROFICIÊNCIAS 5ºS E 9ºS .....</b>                                | <b>95</b>  |
| <b>APÊNDICE C - LISTA DE PACOTES R E RESPECTIVAS VERSÕES .....</b>                         | <b>104</b> |
| <b>APÊNDICE D - DESCRIÇÃO DE VARIÁVEIS DO EDUCACENSO .....</b>                             | <b>108</b> |
| <b>APÊNDICE E - REDUÇÃO DE VARIÁVEIS .....</b>   | <b>116</b> |

## 1 INTRODUÇÃO

A educação no Brasil é de responsabilidade do governo federal, estados e municípios. Cada autarquia deve desenvolver e administrar seu próprio sistema de ensino, o que autentica essa informação é a Constituição Federal juntamente com a Lei de Diretrizes e Bases da Educação Nacional (LDB) que diz: “regular os direitos e deveres da política brasileira da educação formal e não-formal” (BRASIL, 2018a).

A LDB afirma que a educação é “dever da família e do Estado”, com objetivo de alcançar “o pleno desenvolvimento do educando, seu preparo para o exercício da cidadania e sua qualificação para o trabalho”. Já a escola, é um dos inúmeros espaços pedagógicos formativos que deve vincular-se às práticas sociais e ao mundo do trabalho (BRASIL, 2018).

Na década de 60, estudos sobre o desempenho escolar começaram a ter significância no âmbito internacional. Se deu pela preocupação social em relação às desigualdades sociais e observou diferentes resultados alcançados nos atendimentos educacionais (COLEMAN, 1966). Essas pesquisas (nos anos 1960: o chamado Relatório Coleman, nos Estados Unidos, e o Relatório Plowden, na Inglaterra) influenciaram as atuais produções teóricas e metodológicas da área de mensuração da educação. Grande parte delas, observava a variação do resultado acadêmico mediante as variações socioeconômicas da comunidade escolar (docentes, discentes e familiares), método de ensino e infraestrutura das instituições de ensino. Os resultados dessas pesquisas apontaram que a infraestrutura da escola e a metodologia aplicada, pouco afetaram no desenvolvimento dos discente, isso quando comparado com o quesito socioeconômico dos alunos e suas famílias (ALBERNAZ; FERREIRA; FRANCO, 2002; BROOKE; SOARES, 2011).

Já a mensuração de dados no Brasil, Brooke e Soares (2011) afirmam:

Quanto à pesquisa em eficácia escolar desenvolvida no Brasil, destaca-se o advento, na década de 1990, do Sistema Nacional de Avaliação da Educação Básica (Saeb), que permitiu ao pesquisador brasileiro conhecer de forma mais sistemática o nível socioeconômico dos alunos, bem como seu desempenho em testes padronizados em larga escala. Destaca-se, também, a criação de alguns sistemas estaduais de avaliação da qualidade da educação. Esses sistemas (o Saeb e sistemas estaduais), de acordo com o texto, constituem a principal base de dados de notáveis estudos em eficácia escolar desenvolvidos entre nós nos últimos anos.

Dados como o clima entre toda a comunidade escolar, uma gestão competente e colaborativa e o esforço dos docentes em relação aos resultados dos discentes resultam em bons resultados nos testes que compõem o Saeb, o principal conceito para diagnosticar a educação no Brasil. É interessante observar que, ao contrário de outros países desenvolvidos que analisam a qualidade do ensino, as limitações físicas das escolas interferem no

desempenho dos discentes brasileiros, visto que os prédios escolares possuem infraestruturas gritantemente distintas (BROOKE; SOARES, 2011).

Quando é dito infraestrutura escolar, usualmente a primeira coisa que vem à cabeça são as paredes, salas e quadros, mas segundo Alves & Xavier (2018),

“O conceito de infraestrutura em educação é polissêmico. O termo compreende a concepção arquitetônica das escolas, seus ambientes educativos e administrativos, os equipamentos e recursos educacionais, mas também as práticas, o currículo, os processos de ensino e aprendizagem e a capacitação dos professores para utilizar os recursos disponíveis.

Pesquisas nacionais como de Palermo, Silva e Novellino (2014), afirmam que “Os fatores que influenciam a proficiência são múltiplos e complexos”.

As informações coletadas atualmente sobre a qualidade de ensino brasileiro, são obtidas através de dados do Censo Escolar, conhecido como Educacenso, Exame Nacional do Ensino Médio (Enem), Prova Brasil entre outros testes, que constituem o sistema de avaliação estatística do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep).

O Inep foi criado em janeiro de 1937 e homologado pelo Decreto-Lei nº 580, de 30 de julho de 1938, e se chamaria de Instituto Nacional de Pedagogia, porém, consideraram o nome de Instituto Nacional de Estudos Pedagógicos mais prudente na época. Em agosto de 2001, o Inep passou a se chamar de Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, homenagem ao educador Anísio Teixeira, que contribuiu imensamente no desenvolvimento da instituição ao longo de duas décadas (SAVIANI, 2012). De acordo com Saviani (2012), ao citar a Lei nº 378 de janeiro de 1937, afirma que o art. 2º do decreto, descrevem as atribuições iniciais do Inep como:

- a) organizar documentação relativa à história e ao estudo atual das doutrinas e das técnicas pedagógicas, bem como das diferentes espécies de instituições educativas;
- b) manter intercâmbio, em matéria de pedagogia, com as instituições educacionais do País e do estrangeiro;
- c) promover inquéritos e pesquisas sobre todos os problemas atinentes à organização do ensino, bem como sobre os vários métodos e processos pedagógicos;
- d) promover investigações no terreno da psicologia aplicada à educação, bem como relativamente ao problema da orientação e seleção profissional;
- e) prestar assistência técnica aos serviços estaduais, municipais e particulares de educação, ministrando-lhes, mediante consulta ou independentemente desta, esclarecimentos e soluções sobre os problemas pedagógicos;
- f) divulgar, pelos diferentes processos de difusão, os conhecimentos relativos à teoria e à prática pedagógica (BRASIL, 1938).

Em 1989, o Ministério da Educação brasileira solicitou ao Inep que desenvolvesse um processo de avaliação nacional, com fim de obter um amplo diagnóstico da educação

brasileira por meio de avaliações (SAVIANI, 2012). Surgiu-se assim o Sistema Nacional de Avaliação da Educação Básica que, posteriormente, removeu o “Nacional” do nome e se tornando apenas Sistema de Avaliação da Educação Básica - Saeb. É um conjunto de avaliações externas (desenvolvidas fora da escola) em larga escala. Atualmente, para o governo nacional brasileiro, o resultado destas avaliações é o indicativo da qualidade do ensino brasileiro (BRASIL, 2021a).

O Índice de Desenvolvimento da Educação Básica (Ideb) é a métrica utilizada pelo Governo do Brasil sobre a educação básica, que combina tanto o rendimento escolar dos alunos (aprovação) computados no Censo Escolar, quanto o desempenho dos alunos nos exames padronizados aplicados nas escolas em anos específicos obtidas no Saeb (BRASIL, 2022d). O resultado do Ideb é dado a cada escola no próximo ano que é realizado o teste do Saeb. Como as avaliações do Saeb são realizadas bianualmente, a nota do Ideb para cada escola também é dada a cada 2 anos.

O Inep disponibiliza em seu site, por meio dos dados abertos, os microdados sobre a educação no Brasil. Esses microdados são considerados dados no menor nível de desmembramento, obtidos por meio de pesquisas, coleta de dados e avaliações realizadas pelo próprio instituto. As informações contidas são alimentadas anualmente e constata que a educação no Brasil tem se desenvolvido positivamente. Essa informação é confirmada no resumo técnico dos Resultados do Ideb de 2021 gerado pelo próprio Inep (BRASIL, 2021a).

A Ciência dos Dados é um campo interdisciplinar de estudo que tem por objetivo a geração de conhecimento, a partir do armazenamento e análise de enormes quantidades de dados, geralmente complexos, de relações intrincadas, heterogêneos e sujeitos a ruídos, como costumam ser os dados coletados do mundo real. Engloba, portanto: captura, armazenamento e gerenciamento de dados; infraestrutura computacional centrada em dados, normalmente acompanhada por processamento intensivo e massivamente paralelo; técnicas avançadas de mineração de dados e inteligência artificial (IA) para detecção de padrões subjacentes aos dados, aprendizagem e análises preditivas (GRUSON, D. *et al.*, 2019). E é com o uso desta Ciência que essa dissertação buscará tratar as bases educacionais de todas as regiões do país para atestar, através de pesquisas, se os resultados alcançados condizem com o que a literatura afirma, identificar possíveis padrões que podem agregar a comunidade escolar e quem sabe apresentar novos resultados para esse campo de estudo.

A mineração de dados é um processo que pode ser utilizado na ciência dos dados, que como o próprio termo sugere, é um processo de extração e filtragem de informações com maior valor agregado aos interesses humanos, criando ou descobrindo relações não triviais

que poderiam estar ou não ocultas tão evidentes nos dados brutos (FIGUEIREDO, E. *et al.*, 2019). Para alguns autores, a mineração de dados é uma etapa importante nos processos mais sofisticados de análise exploratória ofertados pela Inteligência Artificial (IA) (NGUYEN, G. *et al.*, 2019).

A aprendizagem de máquina é um campo da IA que trata de desenvolver programas de computador capazes de aprender e melhorar o seu desempenho em tarefas específicas, através da aquisição de experiência embutida nos dados que recebe. Visa absorver representações estatísticas e matemáticas através da estrutura dos dados de modo a utilizá-las para fazer deduções (NGUYEN, G. *et al.*, 2019).

Observado o esforço do Brasil e de outros países em parametrizar a educação, e visto o avanço tecnológico dos dias atuais, se faz necessário utilizar das técnicas mais avançadas da Tecnologia da Informação e Comunicação (TIC) que possam apresentar fatores que influenciam na educação.

## 1.1 ESCOPO DA PESQUISA

Esse trabalho tem como objetivo identificar de forma não supervisionada (termo utilizado em técnicas computacionais de aprendizagem de máquina - *machine learning*, que exploram padrões, tendências ou comportamentos dos dados e geram um resultado quem vem da aplicação de modelos matemáticos e estatísticos com mínima indução ou expectativa de resultados (TIBCO, 2022)), o que os dados apresentam sobre as infraestrutura escolares, da base de dados microdados do Inep, e o que correlacionam positivamente e/ou negativamente com o crescimento do Ideb. Verificar também se e o quanto as instituições educacionais desenvolveram ao longo dos anos e compará-las com as infraestruturas das escolas. Avaliar se o desenvolvimento das instituições educacionais e das unidades federativas no Ideb é determinante para o ensino-aprendizagem dos discentes. Ressalta-se que os dados do Ideb foram mensurados ao longo dos anos antes mesmo do ano de 2007, ano de sua criação, de 2005 até o ano de 2019, e as infraestruturas serão extraídas da base de microdados do ano 2019.

É de suma importância afirmar que, por se tratar de uma pesquisa exploratória, os atributos se relacionam, mas não se deve afirmar que o relacionamento entre eles é uma casualidade. Quer dizer que, mesmo que se constatar que as escolas que possuem banheiros possuam melhores resultados no Ideb, não se pode afirmar que ter banheiro na escola garante uma boa nota do Ideb.

A principal técnica utilizada na base de dados deste estudo é a Análise Fatorial, em inglês, *Factor Analysis* (FA). Segundo Bandalos & Finney (2018), a análise fatorial é um procedimento que analisa a covariação (covariância positiva: indica que duas variáveis tendem a se mover na mesma direção; covariância negativa: revela que duas variáveis tendem a se mover em direções inversas) entre um conjunto de variáveis (variável pode ser entendida como qualquer quantidade, qualidade, magnitude e outros, de uma característica que pode possuir vários valores numéricos) observadas em função de um ou mais construtos latentes. Construtos são teorias não observáveis, mas passíveis de interpretações, como por exemplo a inteligência, autoeficácia ou criatividade. Essas inferências ou construtos são consideradas latentes por não serem claramente observáveis. Desta forma, o foco do FA é ajudar os pesquisadores a identificar ou melhor compreender a origem desses construtos latentes em relação às variáveis de seu interesse.

A escolha de FA se deu ao fato de o estudo buscar agrupar diversas variáveis correlacionadas, visto que a base de dados do microdados do Inep possui quase 200 variáveis sobre a infraestrutura das escolas. O FA também separa as variâncias compartilhadas (quanto distante os dados estão da média) entre atributos (communalities) das variâncias intrínsecas (uniqueness). Assim, é possível dar destaques às variáveis que têm correlações mais importantes daqueles que se correlacionam pouco.

## 1.2 JUSTIFICATIVA:

Descobrir como melhorar a qualidade da educação deve ser a vontade de todo profissional da área. As primeiras pesquisas oficiais se tornaram mundialmente conhecidas a partir da década de 60 (COLEMAN, 1966). Constatar e/ou encontrar informações, métodos ou características que identifiquem relações relevantes sobre a educação é bem vindo à comunidade e à sociedade. Essa pesquisa pode ajudar as escolas a se desenvolverem e melhorar a reputação da escola; aperfeiçoar a motivação e dedicação dos alunos e dos professores; e contribuir para gerar um ambiente mais propício à aprendizagem.

Mas não basta detectar quais são os padrões de uma educação de qualidade. Somente o termo qualidade já é complexo suficiente. É preciso muito mais que um aspecto para aprimorar a educação como todo. Para a Unesco,

A qualidade se transformou em um conceito dinâmico que deve se adaptar permanentemente a um mundo que experimenta profundas transformações sociais e econômicas. É cada vez mais importante estimular a capacidade de previsão e de antecipação. Os antigos critérios de qualidade já não são suficientes. Apesar das

diferenças de contexto, existem muitos elementos comuns na busca de uma educação de qualidade que deveria capacitar a todos, mulheres e homens, para participarem plenamente da vida comunitária e para serem também cidadãos do mundo (UNESCO, 2001:1 apud GADOTTI, p.2).

Visto que o governo brasileiro entrega seus dados sobre a educação de forma gratuita, percebe-se que essa pesquisa se torna muito viável, pois seu custo se limita apenas ao limite dos computadores que utiliza-se para esse trabalho.

### 1.3 ESTRUTURA DO TRABALHO

Esta dissertação de mestrado está organizada em seis capítulos e 5 apêndices.

- O Capítulo 2: se resume aos objetivos de trabalho.
- O Capítulo 3: apresenta a revisão de literatura sobre o tema abordado.
- O Capítulo 4: são apresentadas as metodologias e métodos aplicados na condução da pesquisa.
- O Capítulo 5: descreve os resultados e as discussões observadas bem como as inferências apresentadas.
- A conclusão no Capítulo 6: apresenta o que o trabalho pode contribuir para o meio acadêmico e as dificuldades e limitações encontradas, finalizando com as propostas de trabalhos futuros.

Após os capítulos são apresentados os apêndices que apresentam:

- Apêndice A: um exemplo de cálculo do Ideb de uma escola fictícia.
- Apêndice B: quadros das proficiências que uma escola pode ter com base nos resultados dos seus alunos.
- Apêndice C: pacotes e versões dos programas utilizados para a prática das metodologias.
- Apêndice D: um detalhamento dos significados de todos atributos do Educacenso.
- Apêndice E: os motivos da eliminação ou mescla de atributos com suas justificativas.



## 2 OBJETIVOS

### 2.1 OBJETIVO GERAL:

Analisar com o uso de técnicas exploratórias não supervisionadas, quais fatores das infraestruturas das escolas públicas brasileiras correlacionam tanto positivamente quanto negativamente com o crescimento do Ideb?

### 2.2 OBJETIVOS ESPECÍFICOS:

- Identificar e coletar as principais bases de dados relacionados à educação brasileira.
- Tratar e selecionar nas bases de dados, informações relevantes à pesquisa, eliminando os dados faltosos, dependentes e extremos.
- Realizar a primeira análise dos dados através de correlogramas e compará-los com os resultados gerados pelo FA.
- Tratar os resultados obtidos a fim de descobrir informações importantes sobre a educação do Brasil e confirmar ou descobrir informações novas sobre o que a literatura diz sobre eles.
- Identificar quais atributos da infraestrutura do Educacenso correlacionam positivamente com o crescimento e os bons resultados do Ideb de 2019.

### 3 REVISÃO DA LITERATURA

É importante identificar quais pesquisas já foram realizadas sobre os dados da educação, preferencialmente, o ensino público brasileiro, e/ou como tratar grandes bases de dados no campo da educação. Por existirem diferentes plataformas digitais em diversas áreas da economia mundial, a comunicação, o armazenamento e a compilação de dados tornam a análise de dados um trabalho demasiado grande (SILVA FILHO, 2017). Obter, reconhecer e analisar pesquisas que trabalham com grande quantidade de dados, principalmente na educação, fez parte deste trabalho.

Com intuito de estabelecer o escopo da pesquisa, definiu critérios que possibilitaram sua realização, assim como acesso aos dados e todo o alcance do estudo. O principal motor de busca deste trabalho, Google Scholar resultou em diversos trabalhos científicos que contribuíram para o estudo desta pesquisa. Foi realizado um levantamento conciso sobre *data science* na educação, estudos sobre Ideb e/ou microdados do Inep e metodologia de avaliação da educação.

Ressalta-se que este trabalho não possui a pretensão de esgotar o debate sobre a análise da infraestrutura escolar com relação aos resultados do Ideb, mas sim, de ampliar o ponto de vista sobre a temática e abrir espaço para novos diálogos sobre o tema.

#### 3.1 CIÊNCIA DOS DADOS NA EDUCAÇÃO

Ao pesquisar sobre ciências de dados na educação, diferentes campos da literatura podem surgir. Existe uma ampla gama de artigos focados no *data science* na educação. Assuntos como dados gerados pela educação a distância, previsibilidade ou aprimoramento no desempenho acadêmico, formas de aprendizagem, comunidades que estudam a mineração de dados nos dados educacionais, entre outros assuntos. (VASIC *et al.*, 2005) (JAVIDI *et al.*, 2017) (SIEMENS & BAKER, 2012) (GRIVOKOSTOPOULOU *et al.*, 2014)

##### 3.1.1 Bases de dados internacionais

Soares (2007) afirma que os governos necessitam saber como andam seus sistemas educacionais com intuito de desenvolver políticas e práticas que venham desenvolver o progresso do seu país. Muitos países desenvolvem sistemas de avaliações internas da educação, mas para haver critérios de comparação em relação a outros países, se faz

necessário um processo de avaliação internacional. Essas avaliações, quando geram dados e são estudadas, colaboram muito para a melhoria da educação nacional.

O Brasil participa de diferentes avaliações e exames nacionais e internacionais. O objetivo é monitorar, comparar e avaliar seu sistema educacional e a aprendizagem dos seus discentes. Com isso é possível criar metas e desenvolver documentos que visam focar nos principais conceitos que devem ser aprendidos pelos alunos em todo país (SOARES, 2007).

Dentre os exames internacionais mais reconhecidos mundialmente que o Brasil realiza, estão o Estudo Regional Comparativo e Explicativo (ERCE), conduzido pelo Laboratório Latino-americano de Avaliação da Qualidade da Educação (LLECE), Progress in International Reading Literacy Study (PIRLS), realizada pela International Association for the Evaluation of Educational Achievement (IEA), adotado pelo Brasil apenas no ano de 2021 e o Programme for International Student Assessment (PISA) realizado pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE). (BRASIL, 2022b)

Existem outras ações internacionais que o Brasil adota, como o Sistema de Acreditação Regional de Cursos de Graduação do Mercosul e Estados Associados (Arcu-Sul) e o *Red Iberoamericana para el aseguramiento de la calidad em la Educacion Superior* (Riaces). Porém, infelizmente existem avaliações externas de grande importância mundial que não são adotadas pelo Brasil como o Trends in International Mathematics and Science Study (TIMSS), dirigido pela Associação Internacional para a Avaliação do Desempenho Educacional (IEA) (BRASIL, 2022b).

Cordero *et al.* (2018) informam que os grandes exames internacionais, Timss, Pisa e Pirls são elaborados para grupos específicos de alunos, definidos por grau de ensino. Normalmente essas avaliações são realizadas entre um período de três a cinco anos, para constatar se houve ou não uma evolução na educação dos seus países adeptos. É muito importante que os países sejam avaliados de forma longitudinal, ou seja, repetidamente para que se constate a evolução ao longo do tempo.

Ainda segundo Cordeiro et al (2018), as avaliações externas internacionais buscam diversificar seus assuntos em cada aplicação. O Pirls que examina o nível de leitura de alunos do 4º ano do ensino fundamental é realizado a cada cinco anos. O Pisa analisa a capacidade de leitura, o conhecimento em matemática e português e o nível de ciências de alunos na faixa de 15 anos. O Timss, aplicado a cada 4 anos, testa o conhecimento da matemática e da ciências dos alunos de 4º e 8º anos do ensino fundamental.

Mais especificamente sobre as avaliações realizadas pela IEA, Mulis *et al.* (2017) explica que o TIMSS está no sétimo ciclo de aplicação e coleta de dados, ou seja, a mais de

três décadas sendo realizada. Sua última avaliação, realizada no ano de 2019, teve cerca de 70 países que utilizaram dos seus resultados como monitoramento da eficácia dos seus programas educacionais.

O TIMSS avalia as disciplinas de ciências e matemática por focar no que é conhecido como STEM, que são as áreas de ciência, tecnologia, engenharia e matemática. Os trabalhadores que atuam na área STEM precisam solucionar problemas mundiais como a fome, crescimento e estabilidade na economia global, entre outras. Enquanto a ciência trabalha a visão do nosso clima, terra e água, fonte de alimentos e combustíveis, a matemática nos ajuda a gerenciar diversas tarefas do dia a dia e é importantíssimo no desenvolvimento da tecnologia, na qual estamos cada dia mais dependentes, como por exemplo o computador, televisão e smartphones. A avaliação internacional PIRLS testa a alfabetização em leitura, e esta é a base para qualquer sucesso acadêmico. A PIRLS é baseada em uma estrutura que abrange os principais objetivos de leitura, que são adquirir experiência literária, para adquirir, pesquisar e usar informações. Sua primeira aplicação foi em 2001, o PIRLS foi aplicado quatro vezes ao longo dos seus 20 anos de execução (BRASIL, 2022b).

Desta forma os países que realizam as avaliações internacionais da IEA, conseguem contemplar as três principais áreas do conhecimento, que são a matemática, ciência e leitura. Com essas avaliações é possível monitorar as tendências de nível de aprendizagem em relação a um contexto global, monitorar possíveis políticas educacionais adotadas pelo país, realizar estudos que comparem a equidade do aprendizado ou até obter dados ricos sobre contextos domésticos (visto que as avaliações possuem questionários socioeconômicos) e escolares sobre a aprendizagem da população do país (MULLIS et al., 2017).

### **3.1.2 Ciências dos dados na educação brasileira**

As pesquisas sobre data science na educação brasileira ainda são bem pouco exploradas. Freneda (2020) buscou descobrir quais fatores poderiam interferir no processo de ensino-aprendizagem e/ou eram predominantes para a melhoria dos discentes. Para sua dissertação utilizou das técnicas de mineração de dados (árvore de decisão) na base de dados do Censo Escolar e do Enem no biênio 2017 e 2018. Com o público alvo de estudantes do Distrito Federal, o autor chegou à conclusão que o aspecto socioeconômico é o que mais interfere no desempenho dos alunos, ou seja, aqueles estudantes que possuíam maior capital social, cultural e econômico se destacavam dentre os demais. Relacionado aos docentes e escolas, Freneda afirma que a capacitação e infraestrutura devem estar alinhados e em

constante evolução para buscar melhorias nos resultados de desempenho escolar, mesmo que menos significativos. Como destaque relacionados à infraestrutura das escolas que apresentaram melhores desempenhos, Freneda citou sala de aulas de Atendimentos Educacionais Especializados (AEE), número de salas de aulas utilizadas, número de computadores e número de funcionários.

Gomes e Regis (2012) utilizaram a técnica da FA no censo escolar de 2009, para verificar como foi o desempenho escolar dos alunos das cidades da região metropolitana e da grande cidade do Rio de Janeiro. A pesquisa buscou identificar se as infraestruturas ofertadas e o grau de conservação dessas nas escolas públicas impactam na aprendizagem. O estudo chegou ao resultado de que enquanto o Índice de Dependências (índice que mensura a quantidade de infraestruturas a escola tem) apresentou impacto negativo no desempenho médio das escolas (municípios com escolas que possuem grandes infraestruturas, ficam em média 2,13 pontos abaixo de municípios que ainda não possuem estruturas escolares complexas), o Índice de Conservação (índice que mensura o quanto as infraestruturas são utilizadas) foi correlacionado positivamente ao aumento no desempenho das instituições de ensino (associado a 1,73 pontos no desempenho médio das escolas). Os autores justificam seus dados em 3 fatores: primeiro fator está ligado à infraestrutura básica já estar presente na maioria das instituições de ensino, sendo assim o que possui a mais pode ser considerado supérfluo; segundo fator está ligado aos municípios que investem na expansão e reforma das escolas não necessariamente investirem em recursos pedagógicos ou em políticas de conservação e uso adequado dos recursos; e por fim o terceiro fator está relacionado aos investimentos na conservação das infraestruturas existentes e no uso da dessas resultam em efeitos positivos na aprendizagem, mesmo que modestamente. Isso é indicado na literatura especializada, que afirma que “os recursos escolares importam e que a gestão adequada dos mesmos é fator determinante para a promoção da aprendizagem.”. No entanto, os autores concluem que é mais importante a gestão agregar uso dos recursos que implementar várias infraestruturas que não serão utilizadas, assim deve-se verificar se os recursos destinados estão sendo utilizados de forma coerente, para assim alcançar os resultados desejados.

### 3.2 DADOS ABERTOS E MICRODADOS DO INEP

Infelizmente os dados abertos sobre a educação brasileira não são amplamente utilizados, mesmo contendo informações de extrema relevância educacional. Apesar da sua pouca utilização, nos últimos anos, sua utilização vem crescendo, mas ainda está aquém de

contribuir significativamente para uma melhoria da educação nacional. Com a revisão da literatura sobre os dados abertos educacionais, foi possível perceber que a partir do ano de 2018, houve um aumento do interesse de pesquisadores sobre a educação para o nível superior, mas não tanto quanto aos dados de nível fundamental e médio.

Com a grande quantidade de dados gerados com os atuais recursos tecnológicos existentes, trabalhar com dados se torna uma habilidade essencial que pode motivar tomadas de decisões que podem mudar a forma de agir de toda uma sociedade (DAVENPORT E PATIL, 2012).

O Inep disponibiliza um grande volume de dados sobre a educação brasileira, segmentado pelos diferentes níveis da educação. O foco é melhorar e agregar valor ao ensino-aprendizagem, dando oportunidades aos educadores para uma melhor tomada de decisão em suas práticas docentes. Para que isso seja possível, Rigotti e Cerqueira (2015) destaca as atualizações anuais da base de dados do Inep e o quanto isso pode e deve auxiliar a educação.

As bases de dados ofertadas pelo Inep são provenientes do Censo Escolar que é um levantamento anual dos dados estatístico-educacionais de todo país. Alimentado pelas escolas públicas e privadas, conta com o apoio das secretarias de educação, tanto estaduais, quanto municipais (BRASIL, 2012). Por ser o principal instrumento nacional de coleta de dados da educação básica, os responsáveis por preencher o Censo Escolar (normalmente o diretor(a) ou o secretário(a) escolar), devem completar um relatório sobre características da escola, funcionários e estudantes (matrículas, movimento e rendimento).

Segundo Brasil (2012), o conteúdo coletado no censo é utilizado como referência para atuação das políticas públicas e programas na área de educação. Isto inclui, verbas públicas para transporte e merenda escolar, implantação de infraestrutura como laboratórios e bibliotecas, entrega de uniforme e livros, e o Fundo de Manutenção e Desenvolvimento da Educação Básica e de Valorização dos Profissionais da Educação (Fundeb). É possível perceber como é importante o preenchimento correto destas informações.

Como complemento, os resultados de mobilidade (abandono) e rendimento (aprovação e reprovação) dos discentes no Censo Escolar, juntamente com o resultado de outras avaliações desenvolvidas pelo Inep, compoem a nota do Ideb da instituição escolar, que é a parametrização e serve como meta para o Plano de Desenvolvimento da Educação (PDE), do Ministério da Educação.

Inicialmente, os microdados eram distribuídos no formato *American Standard Code for Information Interchange* (ASCII) que permitia a interpretação pela maioria dos softwares

que trabalham com pacotes estatísticos comerciais como SAS e SPSS. Atualmente, formatado no formato Comma-Separated Values (CSV), separados por Pipe ( | ), permite que vários outros softwares estatísticos, inclusive os *open source*, consigam carregar e importar a base de dados (BRASIL, 2012).

### 3.3 IDEB COMO MÉTRICA NA EDUCAÇÃO

O governo brasileiro desenvolveu sua própria forma de avaliar a educação básica, e nomeou a métrica utilizada como Ideb. O objetivo do índice é, de forma simplificada, combinar as principais dimensões da qualidade educacional em um índice sintético. As principais dimensões do Ideb são o rendimento escolar dos alunos e o desempenho da aprendizagem, que relaciona-se a aprovação dos discentes e que são medidos através de exames padronizados aplicados nas escolas periodicamente, respectivamente. É coletada a taxa de aprovação das escolas no Censo Escolar, sistema alimentado anualmente por todas as escolas registradas do país. Os exames padronizados, também conhecidos como provas externas, são divididas em três avaliações distintas, Avaliação Nacional da Educação Básica (ANEAB), Avaliação Nacional do Rendimento Escolar (Anresc), conhecida como Prova Brasil, e Avaliação Nacional de Alfabetização (ANA) que compõe o Saeb. O resultado de cada escola é divulgado pelo Inep no ano seguinte à aplicação da prova do Saeb (BRASIL, 2022b).

Existem muitas críticas ao Ideb por diversos pesquisadores (ALVES & XAVIER, 2018 BUCHMANN & NERI, 2008; SILVA, 2012), isso devido a sua forma simplificada de cálculo, cujo peso para seus dois componentes são igualitários. Por exemplo, se uma escola cuja taxa de aprovação da instituição for 0,8 e a média do desempenho escolar das avaliações for 5, o resultado será igual a 4, em uma escala de 0 a 10, calcula-se 0,8 multiplicado por 5 resulta em uma nota 4. Segundo Buchmann e Neri (2008) a proposta de um peso igualitário nos quesitos fluxo escolar e proficiência do Ideb é de tentar mitigar a manipulação dos dados. Se o peso do fluxo escolar (rendimento) for mais valorizado, as escolas poderiam adotar o fluxo automático de todos os alunos, garantindo assim uma melhora no resultado sem esforço. Porém, se a proficiência dos alunos tivessem um peso maior, iria incentivar a escola ao abandono, evasão ou a retenção de alunos com piores desempenhos, assim apenas os melhores alunos seriam avaliados, elevando a nota da escola. Mesmo com objetivo de combater a manipulação dos dados nas escolas, ainda é possível ver algumas ações políticas que adotam regimes de promoção automática ou manipulação dos resultados relacionados ao índice.

Outra crítica levantada contra o Ideb, desta vez pela autora Silva (2012) é que o índice pode causar uma sensação de insegurança aos educadores. Isso se dá ao fato de muitos professores desconhecerem o real objetivo da métrica. Muitas vezes acredita-se pelos docentes que o índice atenda uma prática mais política do que epistemológica. Outra crença identificada pela autora é que os resultados do Ideb não resultam em nenhuma ação efetiva para melhoria da escola. É essencial que a comunidade escolar se aproprie do seu resultado para que ele passe a ser um indicador estratégico, tanto para cobrança da atual política, quanto para desenvolvimento da educação brasileira.

Por mais que existam críticas às avaliações de larga escala, elas são medidas que servem como referência para a comunidade escolar verificar o seu desempenho e parametrizar seu desenvolvimento. Riscal e Luiz (2016) acreditam que existem muitas imperfeições no Ideb se olhar a questão epistemológica ou metodológica, mas reconhecem que é a forma da comunidade escolar acompanhar e servir como parâmetro para cobranças aos educadores e todas as esferas administrativas responsáveis pelo desenvolvimento da educação. O Ideb é extremamente útil para pesquisadores descobrirem questões relacionadas à gestão escolar e às políticas públicas voltadas para a educação. Segundo as descobertas do modelo analítico dos autores, o índice apresentou ter seu maior peso estatístico na característica de localização geográfica e do perfil socioeconômico das escolas.

### 3.3.1 Como é calculado o Ideb

O cálculo geral do Ideb é dada pela Equação 1:

$$Ideb_{ji} = N_{ji} P_{ji}; \quad 0 \leq N_j \leq 10; \quad 0 \leq P_j \leq 1 \text{ e } 0 \leq Ideb_j \leq 10 \quad (1)$$

em que,

i = ano do exame (Saeb e Prova Brasil) e do Censo Escolar;

j = instituição de ensino

$N_{ji}$  = média da proficiência em Língua Portuguesa e Matemática, padronizada entre os indicadores 0 e 10, obtido pelos alunos da instituição de ensino;

$P_{ji}$  = indicador de rendimento baseado na taxa de aprovação da etapa de ensino, obtido pelos alunos da instituição educacional;



Para melhor exemplificar o cálculo do Ideb, na Equação 1 se faz necessário uma melhor explicação sobre o resultado alcançado da média padronizada da escola ( $N_{ji}$ ). Os resultados dos testes das avaliações de larga escala do Inep são análises estatísticas das respostas dos estudantes da escola referente aos itens do teste. Nas primeiras edições do Saeb, os cálculos eram baseados na Teoria Clássica dos Testes (TCT), que apresentavam o percentual de acerto dos estudantes. Assim a proficiência do aluno é os números de acerto do mesmo na avaliação. A partir do ano de 1995, o Inep agregou a sua análise de proficiência a Teoria de Resposta ao Item (TRI), que permitiu a comparação dos resultados das questões ao longo de suas edições. Assim, a proficiência de cada participante do exame é definida em uma escala, que permite a comparação entre os ciclos avaliados e entre os estudantes.

A descrição da escala de proficiência é descrita por BRASIL (2017),

Os resultados de desempenho nos testes do Saeb são expressos por números na escala de proficiência, que variam de 0 a 500 pontos, com média de 250 e desvio padrão de 50. A escala pode ser visualizada como uma régua construída com base nos parâmetros estabelecidos para os itens aplicados nas edições do teste. Em cada ciclo da avaliação, o conjunto de itens aplicados nos testes de desempenho é posicionado na escala de proficiência com base nos parâmetros calculados por intermédio da TRI. Assim, o fato de um item estar alocado, por exemplo, no ponto 250 da escala do Saeb independe do teste do qual ele fez parte, pois se trata de uma característica própria de cada item. A ordenação dos itens na escala é feita com base no valor do parâmetro de dificuldade observada. Quanto maior o valor desse parâmetro, mais difícil considera-se o item.

Assim, segundo a análise do modelo TRI, é possível alocar cada participante da avaliação na mesma escala de proficiência. Isso significa que cada pergunta do teste se encontra em um patamar de dificuldade e cada participante pode aumentar ou diminuir a proficiência da sua escola com base no grau de complexidade da questão.

Para exemplificar melhor a metodologia do cálculo do Ideb, o Apêndice A faz um exemplo fictício de uma escola que atende as séries iniciais do ensino fundamental, seguindo a referência Brasil (2021b).

Ressalta-se que os quadros referentes às notas padronizadas de 2019 serão apresentados no Apêndice B (BRASIL, 2022e).

A Tabela 1 apresenta os limites inferiores e superiores usados nos cálculos do Ideb desde 1997 nos testes de proficiência do Saeb, a partir da Prova Brasil de 2005 para municípios e escolas.

Tabela 1 - Limite superior e inferior das proficiências do Ideb

| Série    | Matemática       |                  | Língua Portuguesa |                  |
|----------|------------------|------------------|-------------------|------------------|
|          | S <sub>inf</sub> | S <sub>sup</sub> | S <sub>inf</sub>  | S <sub>sup</sub> |
| 4ª do EF | 60               | 322              | 49                | 324              |
| 8ª do EF | 100              | 400              | 100               | 400              |
| 3ª do EM | 111              | 467              | 117               | 451              |

Fonte: Nota Técnica Índice de Desenvolvimento da Educação Básica - Ideb (BRASIL, 2021a)

Conforme afirma Soares (2013), o Ideb aparenta ser um cálculo simples, visto apenas pelas publicações das médias. Normalmente confundido com as notas escolares, onde 9 ou mais é um bom valor e abaixo de 5 é uma nota ruim, o Ideb dificilmente chegará a valores extremos. É pouco provável ter uma nota acima de 8 ou abaixo de 2 no Ideb. Isso pelo fato do rendimento da escola (aprovação) afetar diretamente na porcentagem da nota e pela escala de proficiência da prova externa posicionar cada aluno em uma determinada faixa, conforme o apêndice B.

### 3.4 OUTRAS RSL SOBRE CIÊNCIA DE DADOS APLICADA EM EDUCAÇÃO

Apesar de os dados abertos educacionais estarem disponíveis à população, para Pedroso *et al.* (2013) poucos são os cidadãos que possuem conhecimento técnico para manipular essas bases de dados, principalmente, entre os profissionais de educação. Dourado *et al.* (2017) evidencia que é importante gerar mecanismos para facilitar a verificação do desempenho escolar em tempo real e de forma contínua para as escolas.

A revisão sistemática de literatura realizada por Ferreira *et al.* (2021), relata sobre o crescente interesse dos pesquisadores no uso de *data science na educação*. O período de pesquisa, de 2010 a 2020, relatou que apenas os 3 últimos anos citados, apresentaram números significativos (maiores que 10 citações) de artigos sobre os dados abertos educacionais. Dentre estas pesquisas, a maioria analisou as bases de dados do Enem e Enade, e utilizavam principalmente dos softwares WEKA e o R por se tratarem de programas não privados. Os principais algoritmos utilizados nos artigos encontrados foram o J48 e o NaiveBayes e utilizavam as metodologias *Knowledge Discovery in Databases* (KDD) ou *Cross-Industry Standard for Data Mining* (CRISP-DM).

Outra observação importante na pesquisa de Ferreira *et al.* (2021) foi a pouca procura dos pesquisadores para artigos direcionados à educação básica. Como grande parte dos artigos sobre os dados abertos educacionais brasileiros estão focados no Enem e Enade, que prioriza o ensino superior, os dados sobre a rede de ensino fundamental e médio são menos

explorados. Os autores inferem que devido às bases de dados da educação básica serem muito maiores, é uma contradição ser tão pouco explorado, visto que possuem um grande potencial de identificar informações relevantes para a melhoria do ensino.

Já na revisão sistemática de literatura de Soares *et al* (2021), houve bastante ressonância com a pesquisa de Ferreira. Os anos dos artigos encontrados pela RSL que apresentaram maior pesquisas nos dados educacionais brasileiros foram de 2018 a 2020 com destaque ao ano de 2019, a metodologia mais utilizada nos artigos encontrados foram o CRISP-DM seguidos por KDD e as técnicas de análise exploratória. Porém os artigos encontrados por Soares, foram mais voltados para a educação básica, com as bases de dados do Saeb (o mais encontrado em sua revisão), Ideb, Censo escolar e Indicadores Escolares. E as técnicas mais encontradas foram de regressão e análise de correlação.

## 4 METODOLOGIA

É fundamental destacar que os métodos e algoritmos escolhidos tentaram cumprir o objetivo maior de possibilitar uma análise exploratória de dados não supervisionada, no sentido de fazer com que os padrões emergissem dos dados com mínima indução ou expectativa de resultados. Justifica-se assim porque não foram empregadas técnicas como as de regressão, em que já se isola um atributo como variável dependente de demais atributos considerados independentes. Tal escolha configuraria por si só uma forma de interferir na análise. Por conta disso, regressões são consideradas técnicas supervisionadas (JAMES, G. *et al.*, 2021).

Toda a ciência dos dados desenvolvida neste trabalho foi feita em R (R CORE TEAM, 2020), utilizando sempre versões atualizadas, sendo (no momento que este trabalho é escrito) a versão R 4.2.2 (2022-11-10) - "*Innocent and Trusting*". Informações (*session info*) com todos os pacotes e versões encontram-se no Apêndice C. Como plataforma de desenvolvimento, foi utilizada a IDE (*Integrated Development Environment*) RStudio (RSTUDIO TEAM, 2019), versão 2022.07.1 Build 554. O controle de versionamento foi feito no sistema Github (GITHUB, 2020). A escolha da linguagem R se deu devido sua versatilidade e o grande conjunto de pacotes complementares que faz do R uma rica ferramenta para mineração de dados (TORGO, 2011). O RStudio, busca simplificar e facilitar o uso da linguagem sendo um meio termo entre a interface baseada em texto e uma *Graphical User Interface* (GUI) pura (GANDRUD, 2018).

Ao longo do desenvolvimento, optou-se por padronizar todos os nomes utilizados em scripts, tabelas, variáveis, funções etc, em inglês, visando posterior disponibilização pública de todo o código, o que facilitaria também publicações internacionais.

Como a metodologia seguiu um complexo fluxo de trabalho (*workflow*), em que resultados de uma etapa poderiam influenciar etapas seguintes, julgou-se melhor antecipar parte dos resultados na metodologia, de modo a oferecer uma descrição mais compreensível e reprodutível de todo o processo.

### 4.1 ESCOLHA DOS DADOS

Ressalta-se que a escolha dos microdados do censo escolar foi do ano de 2019, por ser o último ano pré-período pandêmico. Assim, os dados não foram influenciados pelos

prejuízos causados à educação pela Covid-19. Como citado pelo Inep (BRASIL, 2021b), “a pandemia do novo coronavírus teve grande impacto nas atividades escolares em 2020 e 2021” que foi apresentado na pesquisa desenvolvida pela própria instituição denominada “Resposta Educacional à Pandemia de Covid-19 no Brasil” em duas edições. Para o enfrentamento do novo Coronavírus no Brasil, foram adotadas medidas de distanciamento social e restrições de atividades presenciais a partir de março de 2020 (SILVA et al., 2020). Dentre as atividades suspensas estavam os eventos públicos, parte das atividades econômicas e as aulas presenciais. Isso gerou uma grande diversidade de medidas pedagógicas, que impactou diretamente o desenvolvimento, a participação e o aprendizado dos alunos, que limitou o interesse desta pesquisa e justificou a seleção das bases de dados que precederam a pandemia.

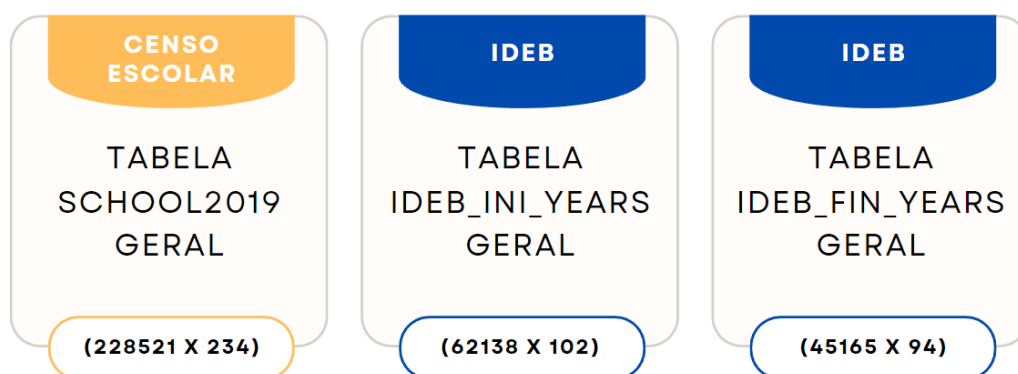
A proposta inicial para essa dissertação era de fazer um estudo longitudinal, que analisaria os dados ao longo dos anos desde sua criação. Mas com as análises feitas nos microdados do censo escolar disponibilizados pelo Inep, foi identificado que os atributos dessa base de dados sofreram diversas alterações ao longo do tempo. Isso dificulta muito o tratamento das informações, visto que existem grande quantidade de atributos, mudanças constantes nas nomenclaturas, descontinuação e inserção nos atributos. Todos os anos, os atributos do censo escolar sofreram algum tipo de alteração, mas entre alguns anos específicos essas alterações foram pouco significativas. As grandes alterações ocorreram entre os anos anteriores a 2007, nos períodos de 2008 a 2011, 2012 a 2014, 2015 a 2018 e a maior mudança aconteceu no ano de 2019, adicionando 103 variáveis, removendo 38 atributos e alterando a nomenclatura de 5 atributos.

Quanto às bases de dados relacionadas ao Ideb, essas sofrem poucas alterações ao longo dos anos, apenas são acrescentadas as colunas referentes a taxa de aprovação, nota do Saeb, a nota do Ideb do biênio correspondente e a projeção do próximo resultado. Para esse trabalho, dois arquivos foram selecionados de nomes: “divulgação dos anos iniciais 2019” e “divulgação dos anos finais 2019”. Com esses arquivos foi possível ver as notas de todas as escolas divulgadas pelo Inep ao longo dos anos de 2005 a 2019.

É importante ressaltar que as bases de dados coletadas para esse trabalho, aconteceram antes do ano de 2019, e precedeu a execução nacional da Lei Geral de Proteção de Dados Pessoais (LGPD), Lei nº 13.709/2018, que visa proteger os direitos fundamentais de liberdade e privacidade e a livre formação da personalidade de cada indivíduo. Sendo assim, ao longo de toda pesquisa, houve uma preocupação em não divulgar nenhum dado pessoal de qualquer indivíduo ou instituição que não esteja compartilhado publicamente no site do Inep após a lei entrar em vigor.

## 4.2 CONJUNTO DE DADOS (*DATA SET*)

Na Figura 1, pode-se observar as principais tabelas que serviram de base para este estudo. A Tabela SCHOOL2019 contém os microdados brutos (*raw data*) com informações gerais das escolas de todo o país coletadas no ano de 2019 (BRASIL, 2022b). Essa tabela contém registros de 228521 escolas caracterizadas por 234 atributos. As tabelas IDEB\_INI\_YEARS e IDEB\_FIN\_YEARS contém dados brutos (*raw data*), envolvendo identificadores das escolas além dos diversos escores das avaliações, coletadas em anos ímpares, de 2005 até 2019 (BRASIL, 2020). A primeira tabela Ideb diz respeito aos anos iniciais, do 1º ao 5º ano do ensino fundamental ou fundamental I. A segunda, do 6º ao 9º ano do ensino fundamental ou fundamental II. Para IDEB\_INI\_YEARS foram 62138 escolas e 102 atributos; para IDEB\_FIN\_YEARS foram 45165 escolas e 94 atributos. Dicionário de dados para essas tabelas podem ser encontrados em Brasil (2022c) e Brasil (2020), respectivamente.



**Figura 1:** Principais entrada de dados, compreendendo as tabelas: SCHOOL2019, montada a partir de microdados brutos (*raw data*) do Inep com informações gerais das escolas de todo o país, coletadas em 2019; IDEB\_INI\_YEARS (anos iniciais, do 1º ao 5º anos do fundamental) e IDEB\_FIN\_YEARS (anos finais, do 6º ao 9º anos do fundamental), com dados brutos das avaliações do Ideb para anos iniciais e finais, respectivamente. Números entre parênteses indicam as quantidades de registros (linhas) e atributos (colunas) das tabelas.

## 4.3 PRIMEIRA EXPLORAÇÃO DE DADOS

Uma vez carregada as tabelas com dados brutos, foi feita uma primeira análise de exploração dos dados (EDA - *Exploration Data Analysis*). Foi constatado que os nomes de atributos na tabela SCHOOL2019 apresentavam um padrão de prefixos conforme mostrado no Quadro 1 e descrito no dicionário de dados do Inep<sup>1</sup> (BRASIL, 2022b).

<sup>1</sup> Localizado no arquivo compactado microdados, disponível em: [https://download.inep.gov.br/dados\\_abertos/microdados\\_censo\\_escolar\\_2019.zip](https://download.inep.gov.br/dados_abertos/microdados_censo_escolar_2019.zip)

Quadro 1 - Prefixos adotados nos nomes dos atributos nas tabelas SCHOOL2019 e Idebs

| PREFIXO | TIPO/SIGNIFICADO  | EXEMPLO                |
|---------|---|------------------------|
| CO      | Variáveis que possuem códigos identificadores.              | CO_MUNICIPIO           |
| NU      | Variável que define o número do ano de aplicação do censo.  | NU_ANO_CENSO           |
| DT      | Variáveis que possuem datas como informação.                | DT_ANO_LETIVO_INICIO   |
| NO      | Variável que possui string como informação.                 | NO_ENTIDADE            |
| TP      | Variáveis categóricas com múltiplas opções.                 | TP_PROPOSTA_PEDAGOGICA |
| IN      | Variáveis categóricas binárias indicando existência ou não. | IN_TABLET_ALUNO        |
| QT      | Variáveis numérico indicando quantidade do item descrito.   | QT_TABLET_ALUNO        |

Fonte: Elaborado pelo Autor (2022)

Para o processo de exploração de dados, foi desenvolvido um conjunto de funções próprias, valendo-se também da função `describe`, presente no pacote `psych` do R (REVELLE, 2022b). Essas funções permitem uma análise estatística das colunas, possibilitando detectar a ausência de valores (NAs ou *Not Available*) e valores anômalos.

Tabela 2 - Saída da função `describe` na tabela SCHOOL2019

|                              | atts   | n     | mean    | sd | min   | max   | range | se |
|------------------------------|--------|-------|---------|----|-------|-------|-------|----|
| IN_TRATAMENTO_LIXO_SEPARACAO | 182468 | 0.54  | 1.62    | 0  | 9     | 9     | 0.00  |    |
| QT_SALAS_UTILIZADAS_DENTRO   | 182468 | 7.98  | 23.63   | 0  | 8888  | 8888  | 0.06  |    |
| QT_TABLET_ALUNO              | 182468 | 17.01 | 1177.19 | 0  | 88888 | 88888 | 2.76  |    |

Fonte: Elaborado pelo Autor (2022)

Um exemplo pode ser visto na Tabela 2. A função `describe` exibe a quantidade de valores não-NAs na coluna “n”, a média, desvio padrão, valores mínimos, máximos, intervalo e o erro padrão dos valores dos atributos. Dado que a quantidade de registros na tabela SCHOOL2019 é de 228521, um “n” de 182468 já indica grande presença de NAs. Valores anômalos também podem ser vistos nos exemplos da Tabela 2, em que é detectada a presença de “9” num atributo categórico binário IN\_TRATAMENTO\_LIXO\_SEPARACAO, cujos valores permitidos deveriam ser 0 e 1. Também são suspeitos os valores 8888 e 88888 para atributos numéricos QT\_SALAS\_UTILIZADAS\_DENTRO e QT\_TABLET\_ALUNO.

Foi verificado que boa parte dos NAs presentes referiam-se a escolas que não estavam mais em funcionamento, controlada pela variável categórica multivariada TP\_SITUACAO\_FUNCIONAMENTO, que assumia os valores: 1 - Em Atividade; 2 - Paralisada; 3 - Extinta (ano censo); 4 - Extinta (anos anteriores).

## 4.4 ADEQUAÇÃO DE DADOS

Uma vez feita a entrada de dados brutos e uma primeira exploração de dados, é momento de operar o que Hadley Wickham (2022) chama de preparação ou adequação dos dados (*data wrangling*), um conjunto de operações ou transformações necessárias para tornar os dados aderentes ao escopo delimitado e mais adequados à visualização e modelagem.

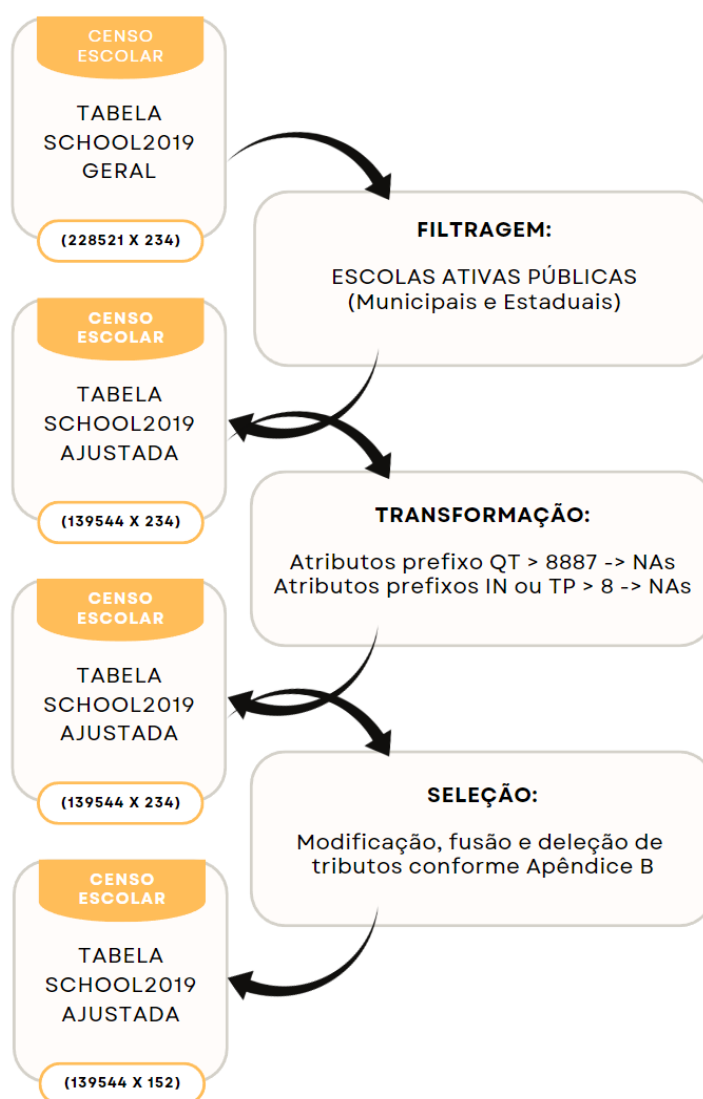
### 4.4.1 Adequação de dados na tabela SCHOOL2019

Na Figura 2, descreve-se as principais operações feitas nas tabela SCHOOL2019, a saber:

- **FILTRAGEM:** tomou-se a decisão de trabalhar apenas com escolas ativas (TP\_SITUACAO\_FUNCIONAMENTO igual a 1), com foco em escolas públicas municipais e estaduais. O atributo que controla essa escolha é a variável categórica TP\_DEPENDENCIA, com valores: 1 - Federal; 2 - Estadual; 3 - Municipal; 4 - Privada.
- **TRANSFORMAÇÃO:** tanto atributos com prefixo QT cujos valores fossem maiores que 8887, quanto com prefixos IN ou TP com valores maiores que 8, todos foram considerados não confiáveis e transformados em NAs, dando chance a valores mais realistas ou de menor impacto nas análises na fase de imputação (vide seção 4.8). Tomou-se o cuidado de verificar se não havia nenhum atributo prefixo IN ou prefixo TP com valores definidos no dicionário de dados maior que 8. Esta transformação consistiu apenas num primeiro controle de valores extremos (*outliers*). Outras ações foram necessárias, conforme se verá mais adiante (vide seção 4.11).
- **SELEÇÃO:** após longa e minuciosa análise de todos os 234 atributos, montou-se uma função no R capaz de ler e operar adequações a partir de instruções definidas numa planilha, envolvendo: modificação, fusão e remoção de atributos. Por exemplo, com o foco em escolas públicas fez-se desnecessário manter atributos atrelados à escolas particulares, como IN\_MANT\_ESCOLA\_PRIVADA\_EMP. Outros, revelaram-se redundantes, como o categórico IN\_TABLET\_ALUNO e o numérico QT\_TABLET\_ALUNO, mostrado no Quadro 1. O primeiro apenas diz se a escola usa tablet para alunos ou não; o segundo, a quantidade. Se uma escola não usa tablet,



ambos são zero; e se usa, ambos são maiores que zero. Detalhamento dos atributos envolvidos com justificativas dessas ações encontra-se no Apêndice D. Dessa forma, os 234 atributos originais ficaram reduzidos à 152. Essa operação redundou em um primeiro processo de seleção de características (*feature selection*), embora outras intervenções ainda se mostraram necessárias em determinadas etapas do desenvolvimento do projeto. Isso será destacado no texto sempre que for o caso.



**Figura 2:** Principais operações iniciais de filtragem e transformação de dados na tabela SCHOOL2019. As operações FILTRAGEM, TRANSFORMAÇÃO e SELEÇÃO modificaram a tabela SCHOOL2019 conforme descrição no texto da metodologia. Resumo das principais ações podem ser vistas à direita. Números entre parênteses indicam as dimensões em linhas e colunas respectivamente.

#### 4.4.2 Adequação de dados nas tabelas Idebs

Na Figura 3, descreve-se as principais operações feitas nas tabelas Ideb tanto INI\_YEARS quanto FIN\_YEARS, a saber:

- **SELEÇÃO:** foram selecionados os atributos identificadores SG\_UF (sigla da unidade da federação ou sigla dos estados); CO\_MUNICIPIO (código do município); NO\_MUNICIPIO (nome do município); ID\_ESCOLA (código da escola); NO\_ESCOLA (nome da escola), seguidos pelos escores normalizados VL\_OBSERVADO de 2005 a 2019. Como a avaliação é feita a cada 2 anos, aparecem nessas colunas apenas os anos ímpares.
- **INTERPOLAÇÃO:** operação necessária para um duplo efeito: 1) imputar com valores os NAs presentes em certos anos para algumas escolas; 2) estimar o escore Ideb para anos pares. Para o primeiro, o último escore real é extrapolado para os anos vizinhos que estejam com NAs. Para o segundo, é feita interpolação linear entre os escores ímpares. Exemplos não interpolados e interpolados são mostrados nas tabelas 3 e 4, respectivamente.

Tabela 3 - Amostra de registros da tabela IDEB\_INI não interpolada.

| SG_UF | NO_MUNICIPIO          | NO_ESCOLA                                | 2005 | 2007 | 2009 | 2011 | 2013 | 2015 | 2017 | 2019 |
|-------|-----------------------|--|------|------|------|------|------|------|------|------|
| RO    | Alta Floresta D'Oeste | EMEIEF BOA ESPERANCA                     | NA   | NA   | NA   | NA   | NA   | 5.40 | 5.60 | 4.60 |
| RO    | Alta Floresta D'Oeste | EEEMTI JUSCELINO KUBITSCHKEK DE OLIVEIRA | 4.50 | 4.7  | 4.7  | 4.40 | 6.00 | 5.9  | NA   | NA   |
| RO    | Alta Floresta D'Oeste | EMEIEF PEDRO ALEIXO                      | NA   | NA   | 4.00 | NA   | NA   | NA   | 5.10 | 5.30 |

**Fonte:** Elaborado pelo Autor (2022)

Tabela 4 - Amostra de registros da tabela INTER como uma IDEB\_INI interpolada.

| SG_UF | NO_MUNICIPIO          | NO_ESCOLA                                | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 |
|-------|-----------------------|--|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| RO    | Alta Floresta D'Oeste | EMEIEF BOA ESPERANCA                     | 5.40 | 5.40 | 5.40 | 5.40 | 5.40 | 5.40 | 5.40 | 5.40 | 5.40 | 5.40 | 5.40 | 5.50 | 5.60 | 5.10 | 4.60 |
| RO    | Alta Floresta D'Oeste | EEEMTI JUSCELINO KUBITSCHKEK DE OLIVEIRA | 4.50 | 4.60 | 4.70 | 4.70 | 4.70 | 4.55 | 4.40 | 5.20 | 6.00 | 5.95 | 5.90 | 5.90 | 5.90 | 5.90 | 5.90 |
| RO    | Alta Floresta D'Oeste | EMEIEF PEDRO ALEIXO                      | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.14 | 4.28 | 4.41 | 4.55 | 4.69 | 4.82 | 4.96 | 5.10 | 5.20 | 5.30 |

**Fonte:** Elaborado pelo Autor (2022)

- **DIFERENÇA:** a partir da tabela INTER sem NAs, esta operação cria a tabela DIFF com a diferença ano a ano entre 2005 e 2019. Essa tabela mostra o quanto uma escola cresceu ou decresceu ao longo dos anos. Exemplo é mostrado na Tabela 5.

Tabela 5 - Amostra de registros da tabela DIFF montada da tabela INTER em 4

| SG_UF | NO_MUNICIPIO          | NO_ESCOLA                                | 2005 | 2006 | 2007 | 2008 | 2009 | 2010  | 2011  | 2012 | 2013 | 2014  | 2015  | 2016 | 2017 | 2018  | 2019  |
|-------|-----------------------|--|------|------|------|------|------|-------|-------|------|------|-------|-------|------|------|-------|-------|
| RO    | Alta Floresta D'Oeste | EMEIEF BOA ESPERANCA                     | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00  | 0.00  | 0.00 | 0.00 | 0.00  | 0.00  | 0.10 | 0.10 | -0.50 | -0.50 |
| RO    | Alta Floresta D'Oeste | EEEMTI JUSCELINO KUBITSCHKEK DE OLIVEIRA | 0.00 | 0.10 | 0.10 | 0.00 | 0.00 | -0.15 | -0.15 | 0.80 | 0.80 | -0.05 | -0.05 | 0.00 | 0.00 | 0.00  | 0.00  |
| RO    | Alta Floresta D'Oeste | EMEIEF PEDRO ALEIXO                      | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.14  | 0.14  | 0.13 | 0.14 | 0.14  | 0.13  | 0.14 | 0.14 | 0.10  | 0.10  |

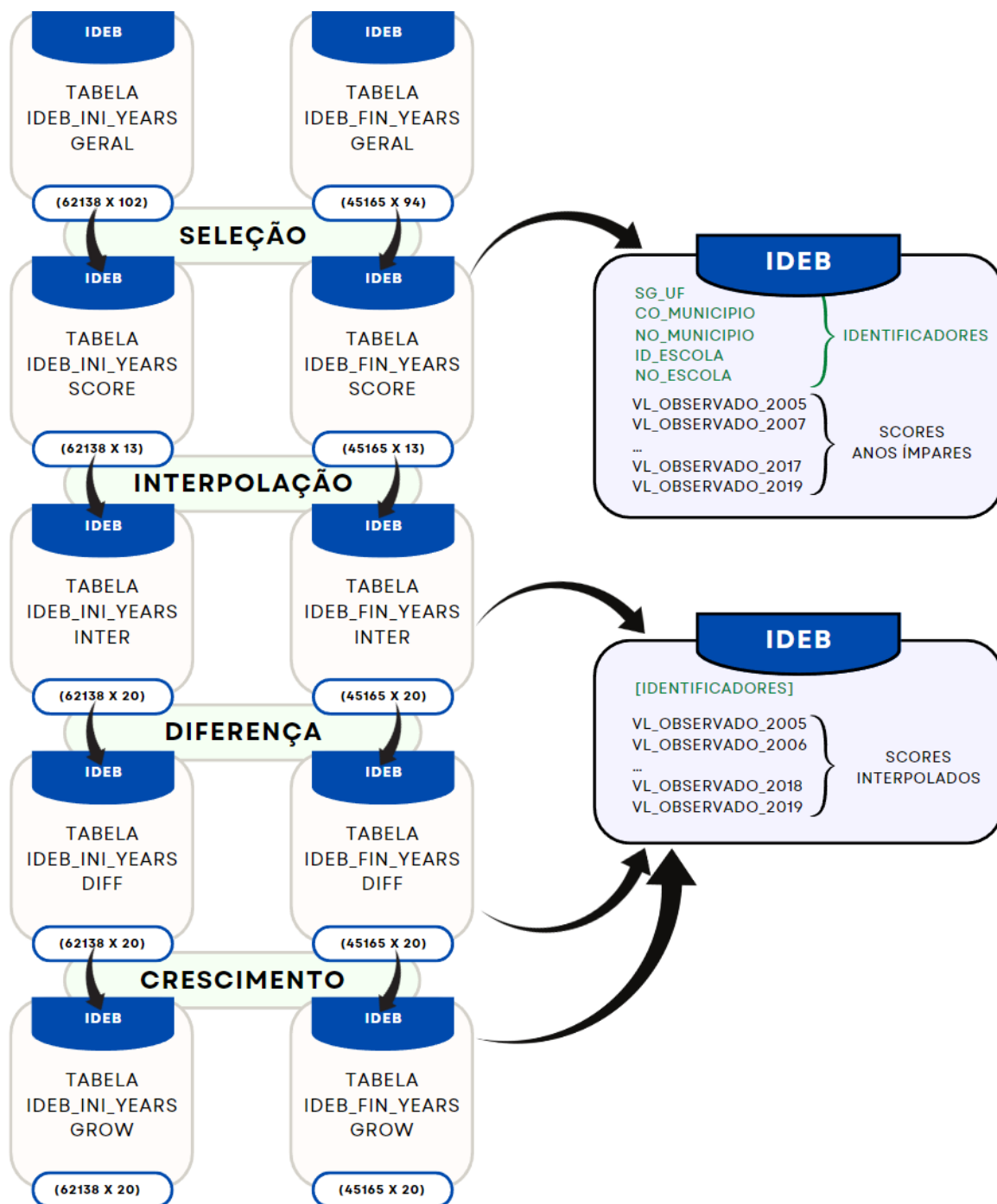
**Fonte:** Elaborado pelo Autor (2022).

- CRESCIMENTO:** esta operação monta a tabela GROW como a soma acumulada das colunas dos anos da tabela DIFF. Essa tabela revela o perfil de evolução do Ideb das escolas ao longo dos anos. Exemplo é mostrado na Tabela 6. Percebe-se na Tabela 6 que a escola EMEIEF BOA ESPERANCA fechou 2019 com decréscimo no Ideb de -0.80 (de 5.40 para 4.60). Já EEEMTI JUSCELINO KUBITSCHKEK DE OLIVEIRA oscilou com crescimentos positivos e negativos ao longo dos anos, até um saldo geral de 1.40 em 2019 (de 4.50 para 5.90). Por fim, EMEIEF PEDRO ALEIXO experimentou taxas monotônicas de crescimento positivo (de 4.00 até 5.30), fechando 2019 em 1.30.

Tabela 6 - Amostra da tabela GROW criada como soma cumulativa da tabela DIFF em 5

| SG_UF | NO_MUNICIPIO          | NO_ESCOLA                                | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011  | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018  | 2019  |
|-------|-----------------------|--|------|------|------|------|------|------|-------|------|------|------|------|------|------|-------|-------|
| RO    | Alta Floresta D'Oeste | EMEIEF BOA ESPERANCA                     | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00  | 0.00 | 0.00 | 0.00 | 0.00 | 0.10 | 0.20 | -0.30 | -0.80 |
| RO    | Alta Floresta D'Oeste | EEEMTI JUSCELINO KUBITSCHKEK DE OLIVEIRA | 0.00 | 0.10 | 0.20 | 0.20 | 0.20 | 0.05 | -0.10 | 0.70 | 1.50 | 1.45 | 1.40 | 1.40 | 1.40 | 1.40  | 1.40  |
| RO    | Alta Floresta D'Oeste | EMEIEF PEDRO ALEIXO                      | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.14 | 0.28  | 0.41 | 0.55 | 0.69 | 0.82 | 0.96 | 1.11 | 1.20  | 1.30  |

**Fonte:** Elaborado pelo Autor (2022).

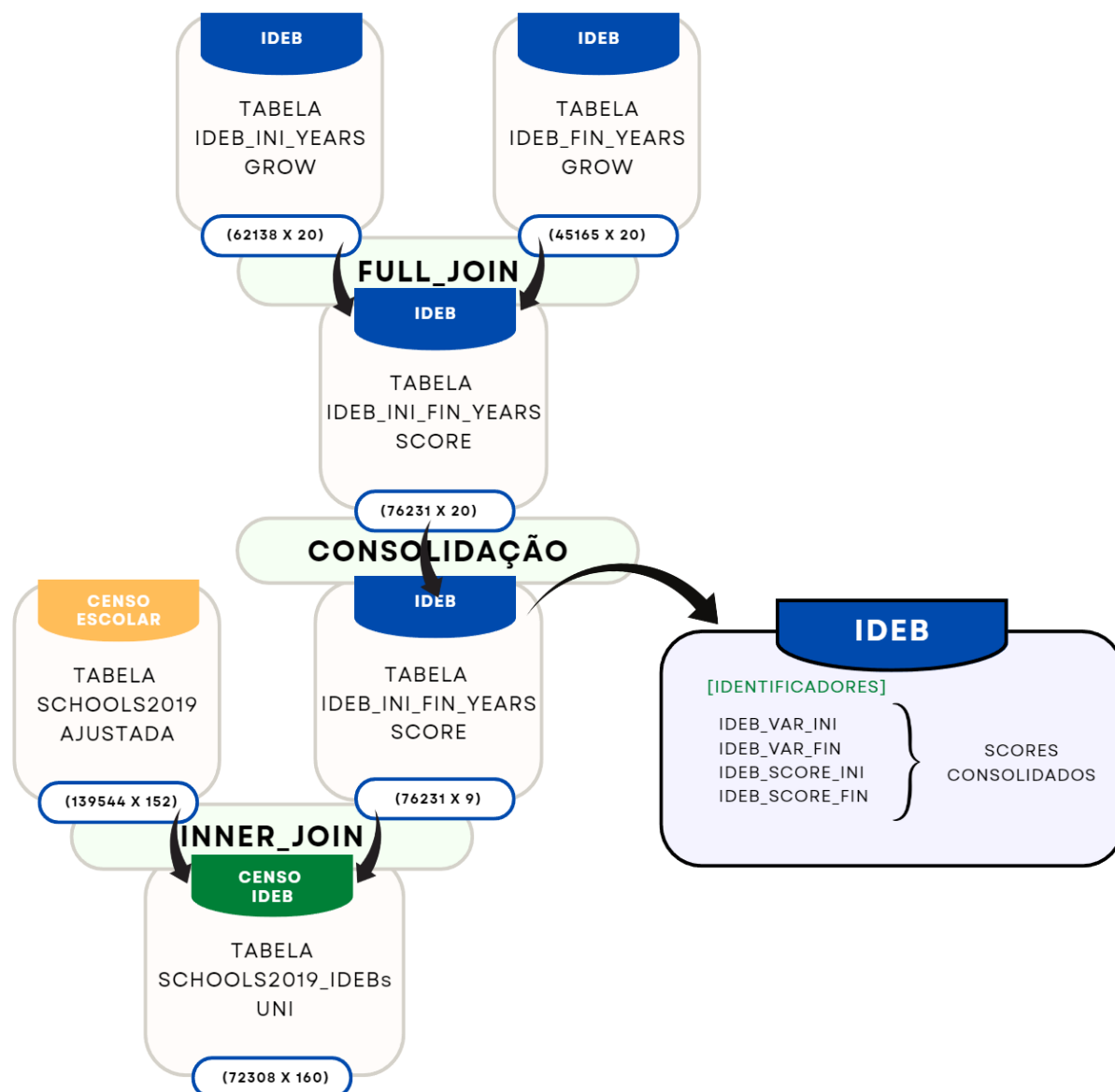


**Figura 3:** Principais operações de filtragem e transformação de dados nas tabelas Ideb. As operações SELEÇÃO, INTERPOLAÇÃO, DIFERENÇA e CRESCIMENTO modificaram as tabelas Ideb GERAIS conforme descrição no texto da metodologia. A estrutura dos campos dessas tabelas podem ser vistas à direita. Números entre parênteses indicam as dimensões em linhas e colunas respectivamente.

#### 4.5 JUNÇÃO DAS TABELAS SCHOOL2019 E IDEBS

As transformações das tabelas SCHOOL2019 e Idebs nas figuras 2 e 3 já evidenciam que há um grande número de escolas sem avaliação do Ideb. A SCHOOL2019 ajustada tem 139544 registros contra 62138 da IDEB\_INI\_YEARS e 45165 da IDEB\_FIN\_YEARS

ajustadas. Como o objetivo maior deste trabalho é mapear padrões de características das escolas que tenham correlação com o Ideb, foram filtradas apenas as informações de escolas em SCHOOL2019 que estavam presentes nas tabelas Idebs, através da junção (*join*) dessas tabelas, conforme descrito na Figura 4.



**Figura 4:** Principais operações para a junção das tabelas SCHOOLS2019 e Idebs. As operações FULL\_JOIN, CONSOLIDAÇÃO e INNER\_JOIN criam uma nova tabela com informações unificadas chamada SCHOOLS2019\_IDEBs\_UNI, doravante UNI, conforme descrição no texto da metodologia. A estrutura dos campos de uma das tabelas pode ser vista à direita. Números entre parênteses indicam as dimensões em linhas e colunas respectivamente.

A junção envolveu as seguintes operações:

- **FULL-JOIN:** foram integradas as tabelas IDEB\_INI\_YEARS com IDEB\_FIN\_YEARS através de um FULL-JOIN, que relembra a operação de união

entre dois conjuntos. Dessa forma, os 62138 registros do primeiro em união com os 45165 registros do segundo redundaram em 76231 registros na tabela IDEB\_INI\_FIN\_YEARS integrando Idebs de anos iniciais e finais.

- **CONSOLIDAÇÃO:** fase em que as colunas das avaliações dos Idebs por ano são consolidadas nos novos campos: IDEB\_VAR\_INI e IDEB\_VAR\_FIN, com a variação ou taxa final do crescimento das escolas de 2005 a 2019 para anos iniciais e finais, respectivamente, a partir das tabelas GROWS; IDEB\_SCORE\_INI e IDEB\_SCORE\_FIN, com o último escore Ideb de 2019 para anos iniciais e finais, respectivamente, a partir das tabelas INTERs.
- **INNER\_JOIN:** finalmente, é feito um INNER\_JOIN entre as tabelas SCHOOL2019 ajustada com IDEB\_INI\_FIN\_YEARS consolidada, operação que relembra a interseção entre dois conjuntos. Isso garante que a nova tabela gerada SCHOOL2019\_IDEBs\_UNI (doravante simplesmente UNI) contenha apenas registros comuns a ambas as tabelas em fusão. Os atributos usados como campos de junção foram ID\_ESCOLA<sup>2</sup> e CO\_MUNICIPIO. Dessa forma, os 139544 registros da primeira tabela em interseção com os 76231 registros da segundo redundaram em 72308 registros na tabela unificada UNI.

Uma amostragem da tabela UNI pode ser vista na Tabela 7. Importante reparar que há escolas sem notas Ideb tanto anos iniciais quanto finais, mas não ambas. Possivelmente: ou não participaram do Ideb para a referida etapa do ensino fundamental, ou a escola não atende tal segmento de ensino, ou era uma etapa recém instituída (ou desativada) e não teve avaliação. Tais hipóteses não foram apuradas.

Tabela 7 - Amostra de registros da tabela UNI

| SG_UF | NO_MUNICIPIO          | NO_ESCOLA                       | IDEB_VAR_INI | IDEB_VAR_FIN | IDEB_SCORE_INI | IDEB_SCORE_FIN | IN_MUNICIPAL | IN_RURAL | ... |
|-------|-----------------------|---------------------------------|--------------|--------------|----------------|----------------|--------------|----------|-----|
| RO    | Alta Floresta D'Oeste | EMEIEF BOA ESPERANCA EEEMTI     | -0.80        | -0.10        | 4.60           | 4.0            | 3            | 2        | ... |
| RO    | Alta Floresta D'Oeste | JUSCELINO KUBITSCHK DE OLIVEIRA | 1.40         | 1.80         | 5.90           | 5.30           | 2            | 1        | ... |
| RO    | Alta Floresta D'Oeste | EMEIEF PEDRO ALEIXO             | 1.30         | 0.90         | 5.30           | 4.70           | 3            | 1        | ... |
| RO    | Alta Floresta D'Oeste | EMEIEF CECILIA MEIRELES         | -1.20        | NA           | 4.50           | NA             | 3            | 1        | ... |
| RO    | Alta Floresta D'Oeste | EMEIEF JOSE BASÍLIO DA GAMA     | NA           | 0.00         | NA             | 3.40           | 3            | 2        | ... |

Fonte: Elaborado pelo Autor (2022)

<sup>2</sup> Cabe destacar que o campo equivalente a ID\_ESCOLA na tabela SCHOOL2019 é CO\_ENTIDADE. Para que a junção fosse possível, CO\_ENTIDADE foi renomeada para ID\_ESCOLA nesta tabela.

#### 4.6 FOCO NOS ANOS INICIAIS

Conforme será discutido mais adiante (vide seção 5.1), ante o maior avanço dos anos iniciais no Ideb frente aos anos finais, optou-se por focar o estudo nessa etapa, no intuito de compreender quais características das escolas poderiam ter exercido influência no crescimento ou decréscimo de suas avaliações. Sendo assim, foram filtradas da base UNI todas as linhas que tinham NAs nos atributos `IDEB_VAR_INI` e `IDEB_SCORE_INI`, bem como retirados os atributos `IDEB_VAR_FIN` e `IDEB_SCORE_FIN`. O resultado foi a tabela `INI`, com 58935 escolas e 158 atributos.

#### 4.7 PREPARAÇÃO

Nesta etapa, para o uso adequado do atributo categórico nominal `SG_UF` (Sigla da Unidade da Federação) nos métodos aqui empregados, foi feita sua “binarização”, transformando em atributo binário (*dummy*) cada unidade da federação (UF), com os nomes `UF_RO`, `UF_AC`, `UF_AM` e assim por diante, sendo 0 (escola de outro estado) e 1 (escola daquele estado). Para evitar questões de colinearidade (quando um atributo é função linear de outros), `UF_RR` (o estado com menor número de escolas na base) foi deletado e escolas de Roraima estariam representadas por todos os atributos UFs dos demais estados sendo zero. Isso fez aumentar a quantidade de atributos de 158 para 184.

Em nome de uma maior padronização numérica, as classes de cada atributo categórico foram forçadas a começarem com valor 0 e seguir em ordem crescente. Assim, por exemplo, `IN_MUNICIPAL` que indicava valores: 2 (estadual) e 3 (municipal), ficou: 0 (estadual) e 1 (municipal). O mesmo para `IN_RURAL` que era 1 (urbana) e 2 (rural), ficou 0 (urbana) e 1 (rural).

Por fim, o atributo categórico `TP_PROPOSTA_PEDAGOGICA` que significava “projeto político pedagógico ou a proposta pedagógica da escola (conforme art. 12 da LDB) atualizado nos últimos 12 meses até a data de referência”, indicando: 0 (sim, mas não atualizado), 1 (sim e atualizado), e 2 (a escola não possui projeto político pedagógico/proposta pedagógica), foi reclassificado para: 0 (a escola não possui projeto político pedagógico/proposta pedagógica), 1 (sim, mas não atualizado), 2 (sim e atualizado). A intenção com isso foi fazer esse atributo ganhar uma semântica de ordem de importância conforme valores crescentes de 0, 1 ou 2. Ou seja, foi transformado numa categoria ordinal.

## 4.8 IMPUTAÇÃO

Um NA (abreviação de *Not Available* ou Não Disponível) indica a ausência de dados numa dada linha e coluna da matriz, evidenciando que determinadas escolas e/ou o próprio sistema Educacenso não conseguiram valorar algum atributo.

Embora muitas pesquisas optem por trabalhar apenas registros completos (sem NAs), a simples eliminação de linhas que contenham dados faltantes podem induzir algum viés na análise, principalmente quando a distribuição de NAs não é aleatória, também chamado MNAR - *Missing Not at Random*) (VAN BUUREN, 2018). Além disso, não parece razoável jogar linhas fora por conta de um ou poucos NAs presentes em alguns atributos. Um procedimento mais confiável tem sido a “imputação” ou estimação de valores para substituição dos NAs. A literatura oferece vários métodos para isso (NGUYEN, 2022).

Foi feita uma análise da distribuição de NAs na base INI, conforme revelado na seção 4.3. Essa análise indicou a necessidade de imputação. O método escolhido foi a imputação por Cópula Gaussiana (GC ou *Gaussian Copula*), implementada por (ZHAO & UDELL, 2020). Estatisticamente, uma cópula é uma função capaz de mapear dependências ou associações entre atributos, tendo em vista as distribuições marginais, que podem assumir distribuições arbitrárias quaisquer. Numa cópula Gaussiana, mapeia-se essas distribuições marginais em Gaussianas (JAWORSKI, 2010). Isso faz com que as GC sejam particularmente interessantes na construção de distribuições multivariadas normais a partir de dados com diferentes tipos de atributos, sejam contínuos ou categóricos (HE, Y. *et al.*, 2021).

Para o ajuste dos parâmetros da distribuição multivariada normal advinda da GC aos dados, a implementação de Zhao & Udell (2020) emprega o método EM (Expectativa Maximização ou *Expectation Maximization*). O EM alterna fases de Expectativa (em que é criada uma função para a expectativa da verossimilhança com os atuais parâmetros da distribuição) e Maximização (em que a função de verossimilhança da etapa anterior é maximizada, gerando novos parâmetros), até que um critério de convergência seja estabelecido (ZAKI & MEIRA JR, 2020).

Entre as vantagens implementação de Zhao & Udell (2020), está a de não demandar a otimização ou afinamento (*tuning*) de hiperparâmetros, como outros métodos. De quebra, retorna uma matriz de associação (que pode ser interpretada como uma correlação) entre atributos, usada no ajuste normal multivariado da GC. Essa matriz de associação/correlação veio a calhar em outros métodos de análise exploratória utilizados neste trabalho (vide seção 5.6). Cabe destacar que o desempenho do método de imputação GC foi superior a outros



considerados estado da arte em experimentos controlados e bases diversas (ZHAO & UDELL, 2020).

Para tanto, foi usado o pacote `gcimputeR` através da sua função `impute_GC` com parâmetro `nlevel = 5` e demais default. Esse parâmetro indica o limite da quantidade de classes ou níveis para uma variável ser considerada categórica ou contínua. Com isso, gerou-se a tabela `INI_IMP`, sem alterações dimensionais.

#### 4.9 COLINEARIDADES

Colinearidades ou dependências lineares entre atributos, quando um pode ser descrito como função linear de outros, podem ter efeitos nocivos em algumas metodologias, por exemplo, naquelas que fazem uso de regressões (DALPIAZ, 2022). Ademais, se um atributo pode ser uma combinação de outros, certamente ele está redundante e desnecessário na base.

Existem várias formas de detectar ou inferir colinearidades em matrizes (IMDADULLAH *et al.*, 2016), mas optou-se por verificar os últimos *eigenvalues* nas correlações das bases em uso. *Eigenvalues* zeros ou muito próximos de zeros podem indicar colinearidades exatas ou quase colinearidades, respectivamente.

Para checagem de colinearidades, foi utilizada a função `eig` do pacote `pracma`, com parâmetros default.

#### 4.10 SEGUNDA EXPLORAÇÃO DE DADOS COM PCA

Tendo gerado a tabela `INI_IMP`, foi feita uma segunda exploração de dados, para uma nova checagem da presença de valores extremos. A existência de tais valores pode enviesar certas análises e métodos estatísticos (AGUINIS *et al.*, 2013). Os valores extremos envolvendo variáveis categóricas (prefixos IN e TP) foram resolvidos no processo de transformação da tabela `SCHOOL2019`, restando apenas categorias bem definidas ou NAs (que foram devidamente tratados na fase de imputação - vide seção 4.8), conforme descrito na Figura 2. Logo, não há mais extremos nessas variáveis categóricas. O foco dessa vez ficou centrado nas variáveis contínuas (prefixo QT)<sup>3</sup>, com intuito de verificar se ainda restavam valores muito dissonantes dos demais.

---

<sup>3</sup> Atributos prefixos QT são contagens inteiras, logo, deveriam ser considerados categóricos ordinais. Mas, quando tendem a assumir um grande número de classes ou valores, assume-se mais pertinente tratá-los como contínuos (KIM, 2017).

Dessa vez, optou-se pelo uso auxiliar da técnica de Análise de Componentes Principais ou PCA (*Principal Component Analysis*) (PEARSON, 1901) sobre o subespaço das variáveis contínuas com prefixo QT da tabela INI\_IMP (doravante, INI\_QT), como forma de dar destaque visual à existência de tais valores extremos em dimensões reduzidas. Cabe destacar que a PCA clássico é uma técnica melhor aplicada a variáveis contínuas (KOLENIKOV & ANGELES, 2004). Foi utilizada a função PCA do pacote FactoMineR (LÊ et al, 2008), com número de componentes iguais aos de atributos e demais parâmetros default.

Lembrando que a PCA reorganiza os dados, reprojando-os em uma nova base de vetores (chamados autovetores) construídos de forma a apontarem na direção de maior variância (que guarda relação com os autovalores) global dos dados. Logo, o autovalor diz o quanto um autovetor teve que se esticar ou encolher para cobrir a variância dos dados. Como a presença de valores extremos afeta a variância, eles podem enviesar a PCA, forçando os autovetores (principalmente os primeiros) a exagerarem-se na direção desses extremos.

#### 4.11 VALORES EXTREMOS (*OUTLIERS*)

Uma forma de encontrar e tratar tais valores extremos é usando um algoritmo de agrupamento (*clustering*) por densidade, como DBSCAN (SCHUBERT *et al.*, 2017) (vide seção 5.5). Resumidamente, o DBSCAN mapeia a vizinhança a cada ponto por meio de dois parâmetros: `minPts` (quantidade mínima de pontos na vizinhança) e `eps` (raio de busca de vizinhos). Dado um ponto, se `minPts` são encontrados num raio `eps` (ou epsilon), então todos são considerados de um mesmo grupo (*cluster*).

Conforme descrito em (SCHUBERT *et al.*, 2017), para estimação do `minPts`, recomenda-se o dobro do número de dimensões da base. Para o raio `eps`, geralmente faz-se uma busca varrendo todos os pontos e anotando a distância do `k` vizinho mais próximo, utilizando o algoritmo kNN (*k-Nearest Neighbor*), com `k` também igual ao dobro de dimensões. Plota-se então um gráfico com as distâncias ordenadas do menor para o maior, estimando o raio `eps` que coincide com o “joelho” da curva.

Para este trabalho, foram utilizadas as funções `kNNdist` e `dbscan` do pacote `dbscan`, com parâmetros `k = 44` e `eps = 20`, sobre bases padronizada em z-score (função `scale` com parâmetros default: média zero e desvio padrão um)

#### 4.12 CORRELAÇÕES

Para uma visão geral das interrelações dos atributos na base INI-FIL, optou-se por utilizar a matriz de associação/correlação (matriz Sigma) advinda da imputação pela metodologia por Cópula Gaussiana. Essa escolha foi orientada pelos seguintes motivos:

- A correlação clássica Pearson geralmente não se mostra adequada para bases mistas, quando há mistura de atributos categóricos e contínuos (HOLGADO–TELLO *et al.*, 2010), como é o caso deste trabalho. Há uma tendência dela subestimar as correlações envolvendo variáveis categóricas.
- Quando há essa mistura, uma opção é operar correlações mistas, customizadas para cada par de atributos conforme o tipo (REVELLE, 2022a): contínuo x contínuo (Pearson); contínuo x categórico (Polisserial ou Biserial); categórico x categórico (Policórica ou Tetracórica).
- Tentou-se montar essa matriz de correlação mista, mas acusou-se a presença de eigenvalues negativos ao aplicar metodologias de redução dimensional (como PCA ou FA), o que implicaria na possibilidade de variâncias negativas e desvios padrões imaginários, algo sem sentido no nível semântico dos dados. Uma matriz assim é chamada de não positiva definida (*Not Positive Definite*), e há formas de tentar torná-la positiva definida, mas não sem controvérsias (LORENZO-SEVA & FERRANDO, 2021).
- Ante tudo isso, optou-se por usar a matriz de associação/correlação (matriz Sigma) advinda da imputação por Cópula Gaussiana. Como defendem os autores dessa proposta (ZHAO & UDELL, 2020), essa matriz concebe correlações otimizadas entre atributos, que levam em consideração o tipo do atributo (contínuo ou categórico), numa metodologia unificada e padronizada para todos eles.

#### 4.13 AGRUPAMENTOS HIERÁRQUICOS

Em certos momentos ao longo deste trabalho fez-se uso de agrupamentos hierárquicos como forma de oferecer uma visualização mais aprimorada do conjunto de correlações entre os atributos. Para tanto, adotou-se agrupamentos aglomerativos hierárquicos ou HAC (de *Hierarchical Agglomerative Clustering*) também conhecidos por AGNES (*Agglomerative Nesting*), em que se inicia cada observação como um cluster unitário, e vão se fundindo ou aglomerando conforme um critério ou métrica de similaridade (MURTAGH &

CONTRERAS, 2017). Dado que na maior parte desta dissertação operou-se com metodologias que destacam as variâncias e covariâncias, optou-se pelas métricas “Ward”, justamente por minimizarem as variâncias intergrupos, ou seja, minimizarem as perdas de informação a cada fusão, produzindo aglomerações mais bem distribuídas ou esféricas (HUSSON et al., 2010).

Neste trabalho, o HAC foi usado principalmente dentro da função `Heatmap` do pacote `ComplexHeatmap`.

#### 4.14 TERCEIRA EXPLORAÇÃO DE DADOS: REDUÇÃO DIMENSIONAL E SELEÇÃO DE ATRIBUTOS POR FA

Após todo o processo de seleção, imputação e filtragem descrito acima, resultou na base INI-FIL com 58920 escolas e 184 atributos, sendo 8 atributos identificadores e 176 atributos de dados, uma matriz com um número razoável de dimensões. Foi o momento de trabalhar num espaço multidimensional mais reduzido e/ou operar novo processo de seleção de atributos (*feature selection*), dessa vez tendo como base a matriz de correlação, no intuito de melhor caracterizar interrelações e/ou irrelevâncias. Essa terceira exploração de dados poderia revelar padrões até então desconhecidos. Lembrando que na etapa de ADEQUAÇÃO DE DADOS, uma primeira leva de seleção de atributos já foi feita, tendo em vista assegurar os objetivos e o foco da dissertação (vide Apêndice E).

A literatura geralmente diferencia a redução dimensional da seleção de atributos, embora ambas possam gerar espaços multidimensionais menores que o original (CUNNINGHAM, 2008). Na redução dimensional, o espaço dos atributos originais é remapeado ou transformado num novo sub-espaço de atributos latentes, de tal forma que seja mínima a perda de informação neste processo. São exemplos de algoritmos de redução dimensional a já vista PCA e a LDA (*Linear Discriminant Analysis*) (FISHER, 1936). Na seleção de atributos, como sugere a expressão, opera-se uma seleção diretamente sobre os atributos originais segundo critérios de relevância, redundância ou indiferença, resultando num subconjunto menor de atributos originais.

Após estudo de várias técnicas de redução dimensional e seleção de atributos, optou-se por uma que pareceu ser capaz de operar ambas ao mesmo tempo (REVELLE, 2009): a FA (*Factor Analysis*) ou Análise Fatorial. Relembrando que FA dá mais atenção às variâncias que PCA (que as considera em sua totalidade, sem diferenciação), separando-as em pelo menos 4 tipos: gerais (g), de grupo (G), únicas (u) e de erro (e), sendo as duas primeiras tratadas como

variâncias comuns (*communalities*) e as duas últimas como variâncias únicas (*uniqueness*). O termo “comum” refere-se à variâncias compartilhadas entre atributos, e o “única” a variâncias isoladas ou próprias de cada atributo. O foco da FA está em estimar as variâncias comuns compartilhadas, de modo a ser capaz de capturar correlações mais “puras” subjacentes a certos conjuntos de atributos.

Logo, atributos que compartilham variâncias comuns (maiores *communalities*) estarão agrupados em fatores latentes, cada fator carregando um significado mais abrangente e que represente todos os atributos associados. Essa representação permite substituir todo o conjunto de atributos com variâncias comuns por um ou poucos fatores, possibilitando a redução dimensional. Por outro lado, atributos com variâncias mais únicas (maiores *uniqueness*) serão aqueles que têm poucas correlações compartilhadas com outros atributos, e provavelmente terão menor importância no estudo em curso. Logo, serão atributos que podem ser descartados, possibilitando a seleção de atributos.

Desde sua criação (SPEARMAN, 1904), a FA encontrou nicho nas ciências sociais, nos estudos psicométricos e análises estatísticas educacionais (HENSON & ROBERTS, 2006). É utilizado principalmente quando se trata de dados gerados por questionários (WILLIAMS *et al.*, 2010). Não deixa de ser um questionário (eletrônico) o sistema Educacenso preenchido anualmente pelas escolas, cujos microdados constituem a base desta dissertação, a partir dos quais também são calculados os índices do Ideb (BRASIL, 2022a). Eis mais um argumento a favor da escolha da técnica FA.

Para a FA, foram usadas diversas funções do pacote `psych`. Também foi adaptada<sup>4</sup> a função `communalities_optimal_solution` do pacote `FactorAssumptions`, tanto para a parte da geração do FA, quanto para a parte de seleção de atributos. Essa função chama recursivamente a função `fa` do `psych`, eliminando o atributo com menor *communality* a cada passo, até que sobrem apenas os atributos com *communalities* maiores que um *cutoff* (no caso, 0.494) (STOROPOLI, 2019). Ou seja, ficarão apenas os atributos que compartilham mais de 50% (inclusive) da sua variância com outros atributos.

Há dois parâmetros importantes a comentar: o método para extração ou estimação dos coeficientes dos fatores; e o método de rotação (REVELLE, 2009). Para o primeiro, dentro os vários métodos disponíveis, optou-se pelo “minres” (*Minimal Residual*). Esse método minimiza no processo de decomposição os resíduos da diferença entre as matrizes estimada para o modelo e da correlação/covariância que deu entrada à função (no caso aqui, a matriz

---

<sup>4</sup> A adaptação constituiu essencialmente na flexibilização de alguns parâmetros fixos, como o *cutoff*, que estava fixo em 0.5.

Sigma), mas desconsiderando as diagonais. Tal estratégia matemática torna o método mais robusto, por não necessitar calcular inversas, o que o faz funcionar até em matrizes não positivas definidas.

Já o método de rotação tem o objetivo maior de tornar a visualização dos resultados mais interpretável, ao procurar uma transformação na matriz dos coeficientes dos fatores que seja mais adequada a esta finalidade. Dentre as várias opções de rotação possíveis (REVELLE, 2022c), considerou-se a “oblimin”. Essa rotação admite soluções tanto ortogonais quanto não ortogonais (obliquas), e é apropriada quando suspeita-se que possam existir correlações entre os fatores (WATKINS, 2018). Cabe salientar que as rotações não interferem nos aspectos básicos resultantes da FA, como as variâncias acumuladas e as *communalities* (COSTELLO & OSBORNE, 2005). Visa apenas facilitar a visualização e interpretação dos resultados.

#### 4.15 AFERIÇÃO DA BASE NA FA

Antes de iniciar uma FA, a literatura recomenda dois testes para verificar a adequação da base para tal metodologia (WATKINS, 2018): teste de Kaiser-Meyer-Olkin (KMO) e teste de Bartlett. O primeiro é uma medida da razão entre correlações totais e parciais, e mede o grau de propensão de um atributo vir a ter variâncias compartilhadas com outros. Quanto maior o KMO, mais adequada a base está para uma FA, pois haverá muitos atributos com variâncias compartilhadas. Já o Bartlett testa se os valores da matriz de correlação são desvios estatisticamente significativos ou não de uma matriz identidade. Se tais desvios não forem significativos (por exemplo,  $p\text{-value} > 0.05$ , num nível de confiança 0.95), a matriz de correlação seria considerada próxima de uma matriz identidade, haveria tantos fatores quanto variáveis, e a FA perderia sentido (SANTOS, 2019).

Nesta dissertação fez-se uso confiável apenas do teste KMO. O teste de Bartlett não é recomendado quando a amostra é muito grande (HENSON, 2006), como é o caso aqui da ordem de 60 mil registros. Alguns testes que fazem uso de  $p\text{-value}$  (como Bartlett) podem tornar-se irrelevantes quando a amostra é muito grande, pois mesmo mínimas diferenças passariam a ser significativas (LIN et al., 2013).

Para o KMO foi utilizada a função `KMO` do pacote `psych`.

#### 4.16 NÚMERO DE FATORES NA FA

Outro ponto a destacar seria quantos fatores extrair para uma FA. Existem vários métodos propostos por Revelle (2009), mas nesta dissertação optou-se por dois deles: o critério MAP (*Minimum Average Partial criterion*) e a Análise Paralela (*Parallel Analysis*). No primeiro método, grosso modo, quanto menor as correlações parciais médias, maior o nível de compartilhamento de variâncias entre atributos. Um ponto de mínimo seria alcançado quando o número de fatores atingisse um ideal, a partir do qual as correlações parciais médias começariam a crescer (WATKINS, 2018). O segundo, como o nome indica, faz uma análise paralela entre os autovetores da base real e de uma base randômica, sendo o número ótimo de fatores aquele que antecede o momento em que o real e randômico passem a ficar indistintos.

Para ambos os critérios foram usadas as funções `VSS` e `fa.parallel` do pacote `psych`, respectivamente.

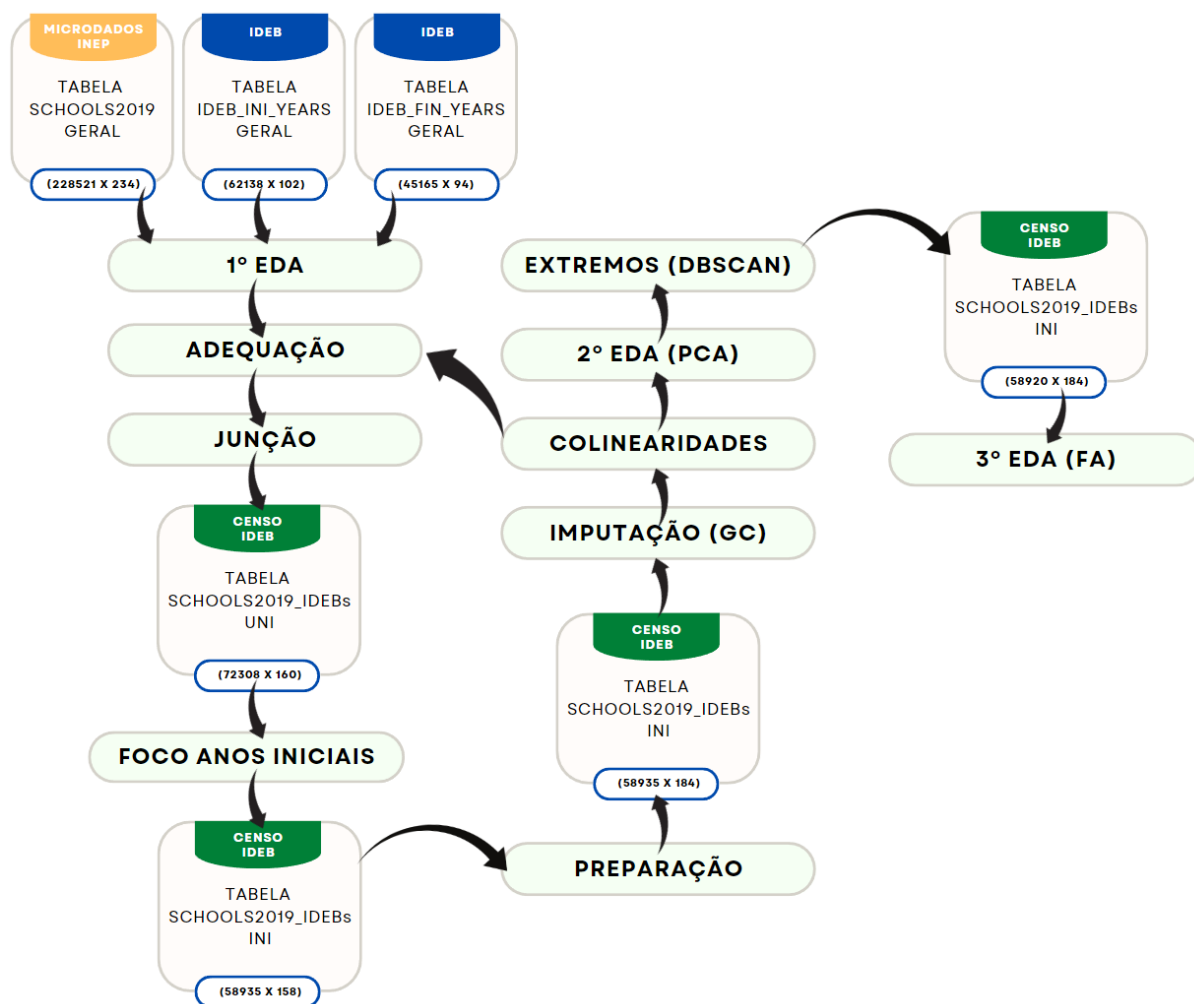
#### 4.17 OBSERVÁVEIS NA FA

As matrizes que dão entrada aos algoritmos de decomposição matricial como PCA ou FA geralmente dispõem nas linhas os indivíduos ou observáveis (aqui, as escolas), e nas colunas, as variáveis ou atributos (as características das escolas). No caso específico da FA, a matriz que a alimenta é uma matriz de correlação entre atributos, de modo que seus principais resultados dizem respeito aos padrões de associações mais relevantes encontrados no espaço dos atributos versus fatores (FABRIGAR, 1999).

Mas, e sobre os mapeamentos entre observáveis e fatores, ou seja, sobre como as escolas individualmente se reorganizariam do espaço dos atributos para o espaço dos fatores? Enquanto na PCA encontrar esse mapeamento entre observáveis e autovetores é possível de forma analítica, na FA só é possível fazer de forma “estimada” (GRICE, 2001). Logo, antecipa-se que a análise dessa estimação deve ser feita com cautela, conferindo se há coerência entre os resultados dos mapeamentos entre atributos x fatores com observáveis x fatores.

Neste trabalho foi utilizado a função `factor.scores` do pacote `psych`, tendo “tenBerge” como método de estimação, por preservar correlações entre fatores tanto em soluções ortogonais quanto oblíquas, dado que foi usada a rotação “oblimin” na decomposição FA.

## 4.18 WORKFLOW GERAL



**Figura 5:** Workflow geral descrevendo as fases do tratamento dos dados geral. Detalhes explicativos encontram-se no texto.

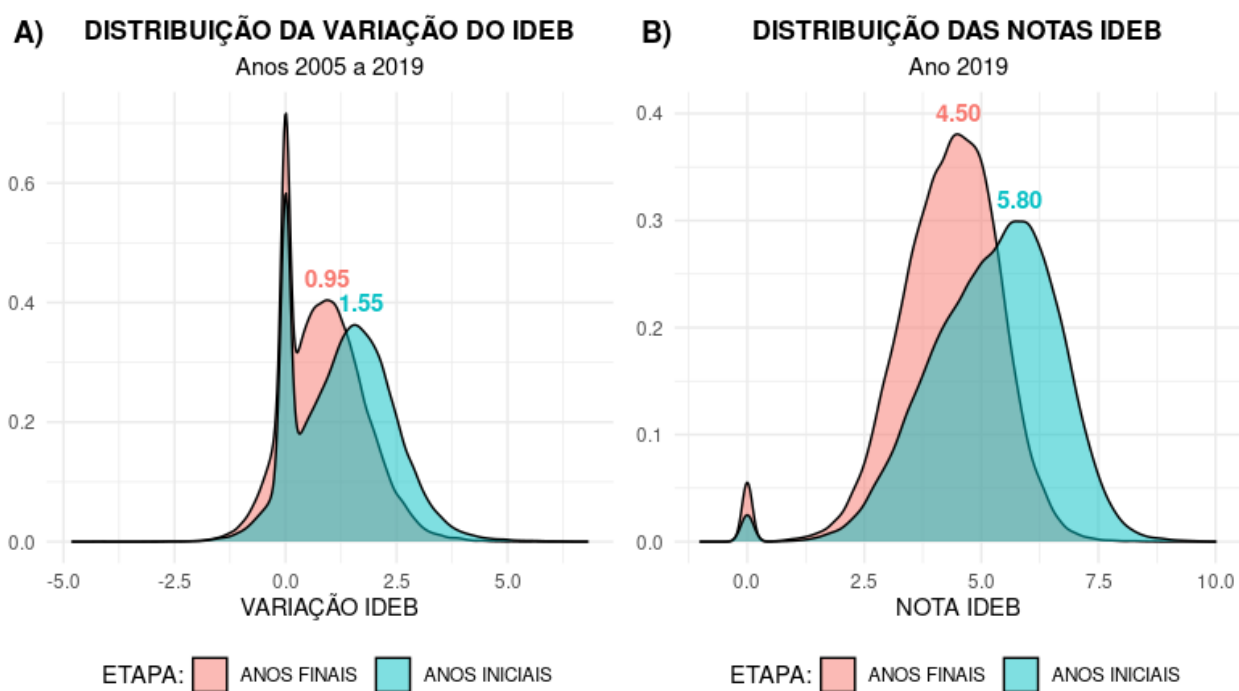


## 5 RESULTADOS E DISCUSSÃO

Nesta seção, em nome de uma maior clareza argumentativa, resultados e discussões são apresentados de forma integrada.

### 5.1 DISTRIBUIÇÕES DAS NOTAS DO IDEB

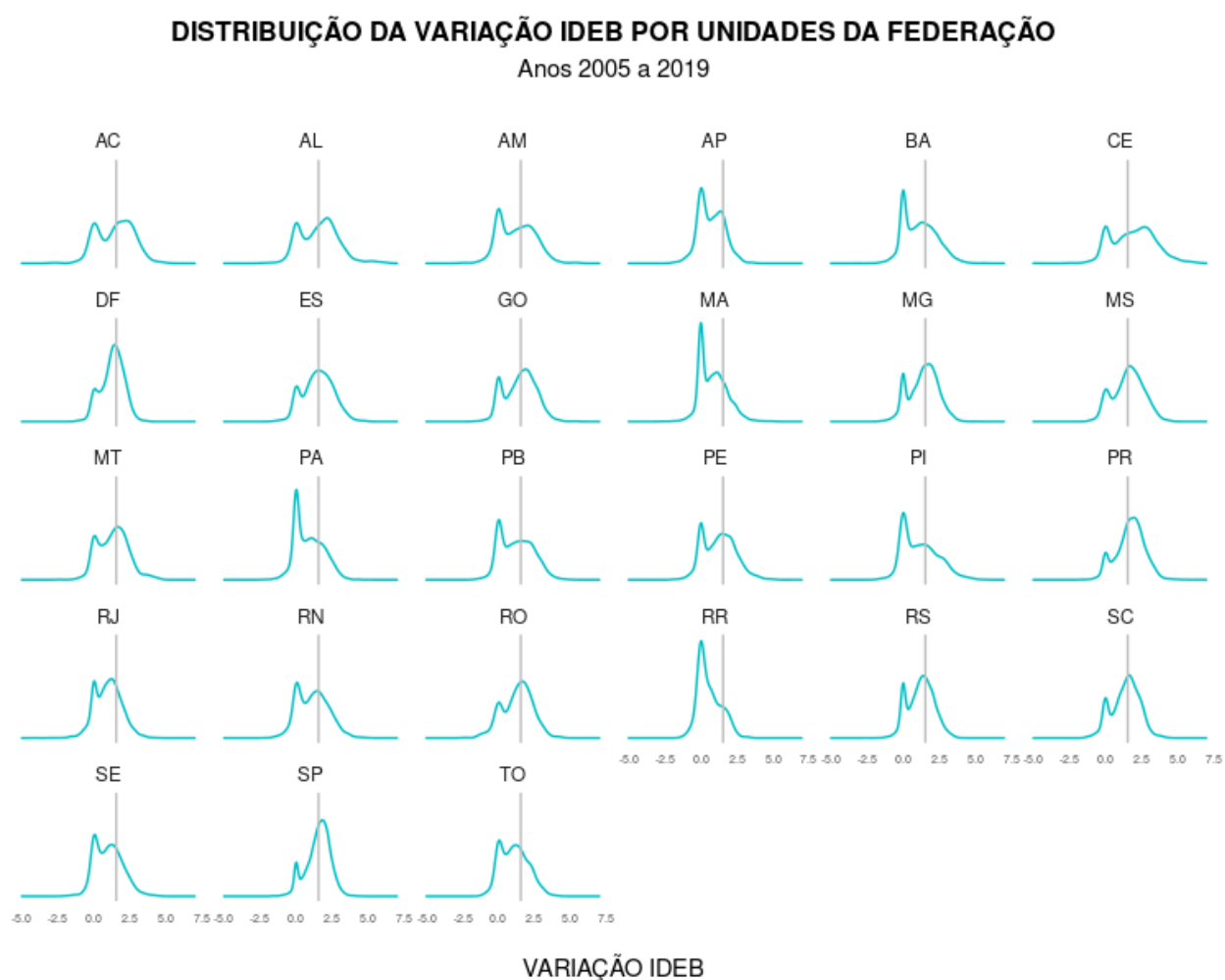
A Figura 6 mostra o perfil das distribuições dos índices do Ideb na base UNI, tanto anos finais quanto iniciais. Em 6.A, temos as distribuições para a variação do Ideb estimada entre os anos 2005 a 2019. Percebe-se a presença de mais de um modal, um com variação nula, independente da etapa, outro em 0.95 e 1.55 para anos finais e iniciais, respectivamente, revelando uma taxa de crescimento do Ideb de 1.63 vezes maior dos anos iniciais frente aos anos finais, fenômeno já conhecido da literatura (MATUCHESKI, 2021).



**Figura 6:** Gráfico das distribuições de densidade das notas Ideb das escolas na base UNI. Números no gráfico indicam os valores modais não zero. **A)** Distribuição da variação do Ideb, de 2005 a 2019, conforme valores estimados pelas variáveis IDEB\_VAR\_INI (azul) e IDEB\_VAR\_FIN (vermelho) para anos iniciais e finais respectivamente. **B)** Distribuição do escore Ideb, em 2019, conforme valores dados pelas variáveis IDEB\_SCORE\_INI (azul) e IDEB\_SCORE\_FIN (vermelho) para anos iniciais e finais respectivamente.

As figuras 7a e 7b mostram as distribuições de densidade para anos iniciais por estado da variação do Ideb (2005 a 2019) e nota do Ideb (2019), respectivamente. As linhas verticais indicam os modais nacionais destacados na Figura 6, de 1.55 para a variação do Ideb e 5.80

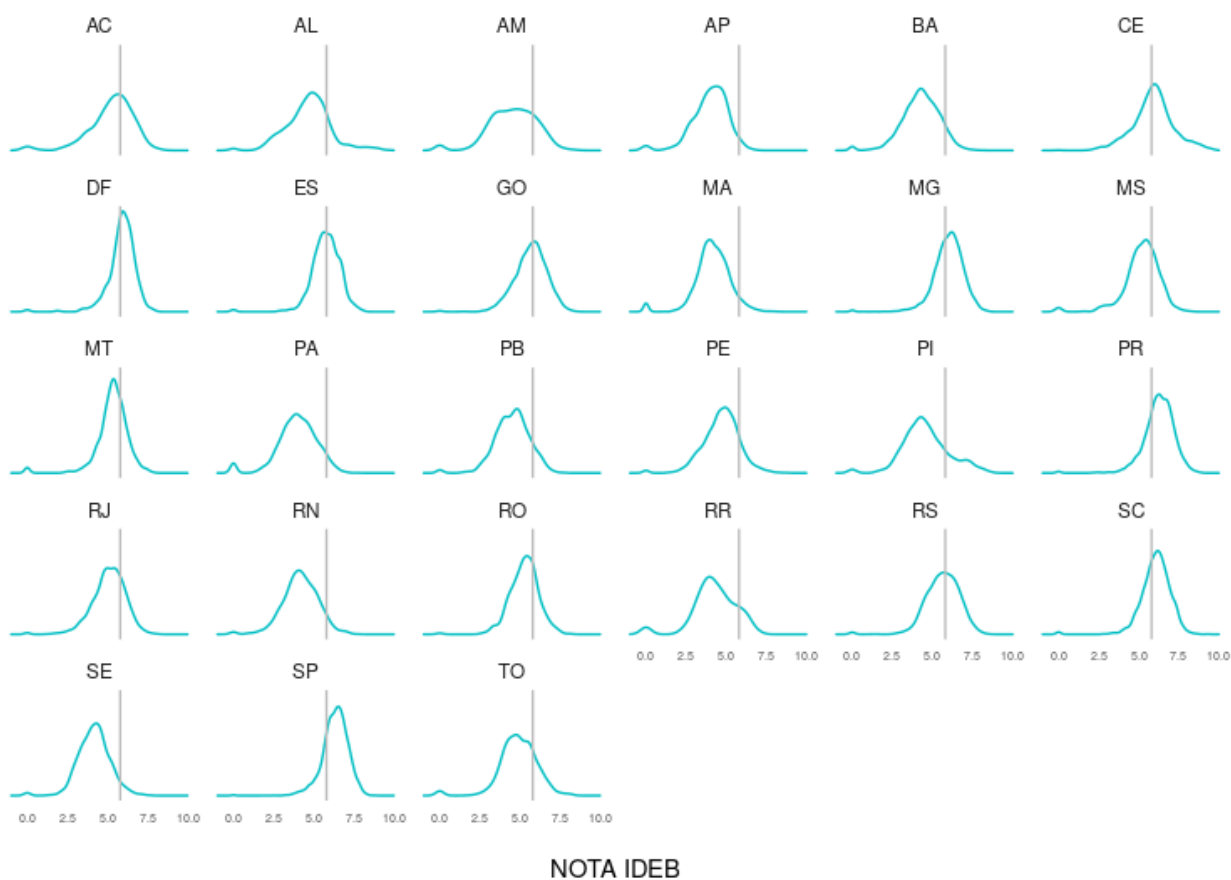
para a nota do 2019 Ideb. Na variação do Ideb, todos os estados apresentam uma distribuição bimodal, sendo o primeiro modal em crescimento zero. No entanto, a intensidade desse modal varia de estado para estado. Por exemplo: MA, PA e RR apresentam o modal zero bem mais agudo que o segundo modal, ao passo que SP, MG, MS, PR, RO o modal zero é menor. Os estados AC, AL, CE, SP e PR são exemplos onde o segundo modal ficou acima do modal nacional. Com respeito à nota Ideb de 2019, as distribuições são essencialmente monomodais, sendo SC, PR, SP e MG alguns estados que se notabilizam por modais acima do modal nacional.



**Figura 7a:** Gráfico das distribuições de densidade da variação do Ideb para anos iniciais por estado da base UNI. As linhas verticais indicam o modal nacional de 1.55.

## DISTRIBUIÇÃO DAS NOTAS IDEB POR UNIDADES DA FEDERAÇÃO

Ano 2019



**Figura 7b:** Gráfico das distribuições de densidade da nota do Ideb em 2019 para anos iniciais por estado da base UNI. As linhas verticais indicam o modal nacional de 5.80.

Pela Tabela 8 nota-se que das 72308 escolas na base, 28598 (39.55%) e 13373 (18.49%) não tinham notas para o cálculo da variação do Ideb, para anos finais e iniciais, respectivamente. O número de escolas com variação nula foi de 6104 (8.44%) e 8162 (11.29%) para anos finais e iniciais, respectivamente. A variação nula indica um grupo de escolas que não melhorou nem piorou seu Ideb ao longo dos 14 anos analisados.

Tabela 8 - Tipos de valores na variação Ideb

| VARIACÃO DO IDEB | ANOS FINAIS |        | ANOS INICIAIS |        |
|------------------|-------------|--------|---------------|--------|
|                  | n           | %      | n             | %      |
| não zeros        | 37606       | 52.01  | 50773         | 70.22  |
| sem nota         | 28598       | 39.55  | 13373         | 18.49  |
| zeros            | 6104        | 8.44   | 8162          | 11.29  |
| Total            | 72308       | 100.00 | 72308         | 100.00 |

**Fonte:** Elaborado pelo Autor (2022)

Em 6.B, as distribuições das notas do ano de 2019, percebe-se novamente a presença

de mais de um modal, um com nota nulo, independente da etapa, outro em 4.50 e 5.80 para anos finais e iniciais, respectivamente. O ano de 2019 fecha com uma nota modal do escore Ideb cerca de 30% maior para anos iniciais que finais, algo também já observado pela literatura (MATUCHESKI, 2021; MINAS GERAIS, 2018; BRASIL, 2018b).

A Tabela 9 tem a mesma organização da Tabela 8, mas agora destacando as quantidades de escolas conforme o escore Ideb de 2019. Interessante notar que o número de escolas sem nota em ambas as tabelas 8 e 9 são iguais tanto para etapas de anos finais quanto iniciais. Isso se deve ao fato de que são escolas que só tinham uma das etapas, e não ambas ao mesmo tempo. Assim, se uma escola tiver só anos iniciais, os valores de escore Ideb para todos os anos (inclusive 2019) de anos finais serão NA na base, gerando uma variação anual do Ideb também NA. Logo, de 72308 escolas, 28598 (39.55%) só tinham anos iniciais, e 13373 (18.49%) só anos finais, indicando que 30337 (41.96%) tinham ambas as etapas. Já o número de escolas com escore 2019 zero foi bem pequeno.

Tabela 9 - Quantidade de escolas conforme tipos de valores da nota Ideb de 2019.

| NOTA DO IDEB 2019 | <u>ANOS FINAIS</u> |        | <u>ANOS INICIAIS</u> |        |
|-------------------|--------------------|--------|----------------------|--------|
|                   | n                  | %      | n                    | %      |
| não zeros         | 43033              | 59.51  | 50773                | 70.22  |
| sem nota          | 28598              | 39.55  | 13373                | 18.49  |
| zeros             | 677                | 0.94   | 8162                 | 11.29  |
| Total             | 72308              | 100.00 | 72308                | 100.00 |

**Fonte:** Elaborado pelo Autor (2022)

A pesquisa quali-quantitativa da dissertação de Vera Matucheski (2021), sugere os motivos dos alunos das séries iniciais conseguirem melhores resultados nas avaliações externas que os alunos das séries finais nos período de 1988 a 2019. Segundo a pesquisa, esses resultados acontecem por cinco fatores descritos em seu trabalho. O primeiro aspecto que interfere no resultado dos alunos está relacionado ao acesso e o hábito de leitura. Os jovens mais velhos se distanciam deste hábito que dificulta a interpretação das perguntas. Através do questionário socioeconômico Saeb, respondido pelos alunos, professores e diretores das escolas que fazem o teste resultante da nota do Ideb, foi identificado que alunos das séries finais percebem menos o hábito de leitura dos responsáveis ou que os responsáveis os incentivam menos a leitura quando comparados com os alunos das séries iniciais. O segundo aspecto observado é a menor realização de dever de casa dos alunos dos 9ºs anos comparado aos dos 5ºs anos. Isso demonstra o menor interesse desses alunos mais velhos aos estudos. É interessante observar que as escolas com o hábito de ofertar dever de casa com

muita frequência e que os professores as corrigirem apresentaram melhores resultados nas avaliações externas que as demais. O terceiro ponto a favor dos estudantes das séries iniciais em comparação ao das séries finais, é o quesito abandono e reprovação. O estudo destaca que muitos jovens abandonam a escola por falta de participação ativa, seja por parte da escola, seja por parte da família ou mesmo pela condição social em que está vivendo. No campo empírico, existe uma participação muito mais ativa da família para com seus filhos de menor faixa etária do que os adolescentes. O quarto motivo do resultado dos alunos de séries finais serem inferiores aos da séries iniciais é o uso de drogas, entorpecentes e a violências verbal ou física. No questionário socioeconômico do Saeb aplicado aos professores, constatou que alunos das séries finais têm mais acesso a entorpecentes e a violência que alunos das séries iniciais. Por fim, o último quesito abordado por Matucheski é que a proficiência dos estudantes de séries iniciais são melhores que os das séries finais, novamente no campo empírico, isso se dá pela menor quantidade de disciplinas ofertadas aos alunos e o número de professores da turma (MATUCHESKI, 2021).

O documento norteador da educação no estado de Minas Gerais (MINAS GERAIS, 2018), denominado Currículo Referência de Minas Gerais, ressalta que a transição dos alunos das séries iniciais para as finais exige um novo patamar de autonomia dos discentes. Isso se dá pelo aumento do número de docentes, os professores especializados (disciplinas escolares) com diferentes metodologias e didáticas, que demandam do aluno maior responsabilidade e organização. Desta forma a rotina escolar se torna mais complexa. Outro aspecto dificultador é que muitas vezes, os docentes das séries finais desconsideram os conhecimentos prévios dos seus alunos da fase anterior, isso dificulta ainda mais a adaptação dos discentes a essa nova etapa da educação. Essas abordagens geram altas taxas de reprovação e abandono nesse período. Já a Base Nacional Comum Curricular (BNCC) ressalta que os estudantes ao entrarem no segundo ciclo do ensino fundamental encontram com uma maior complexidade nas diferentes lógicas de organização do conhecimento (BRASIL, 2018b). E o período em que os alunos entram nas séries finais condiz com a faixa etária de transição entre a infância e adolescência, que ocorrem várias mudanças biológicas, psicológicas, sociais e emocionais. Se não houver uma ressignificação do aprendizado do conteúdo aprendido nas séries iniciais para o contexto das diferentes áreas do ensino nas séries finais, a desmotivação do discente pode levar a falta de vontade ou abandono de suas escolarização.

Conforme dito anteriormente, em função do maior avanço dos anos iniciais no Ideb, este trabalho focou sua análise nessa etapa.

## 5.2 IMPUTAÇÃO

É possível compor a seguinte estatística preliminar sobre a ausência de valores ou NAs na base INI, conforme já visto, escolas com apenas anos iniciais:

1. Das 58935 escolas, 15828 (26.86%) não tinham cadastro completo para todos os 158 atributos;
2. Dos 158 atributos, 51 (32.28%) tinham algum NA.
3. Mas, com relação ao total de células da matriz ( $58935 \times 158 = 9311730$ ), apenas 86345 (0.98%) eram NAs.

Tabela 10 - Mostra os números absolutos e relativos de NAs nos 51 atributos que os tinham.

| <b>NÚMEROS RELATIVOS E ABSOLUTOS DE NAs POR ATRIBUTO</b> |          |          |                            |          |          |
|--|----------|----------|----------------------------|----------|----------|
| <b>Atributos</b>   | <b>n</b> | <b>%</b> | <b>Atributos</b>           | <b>n</b> | <b>%</b> |
| IN_REDE_LOCAL  | 8740     | 10.12    | IN_EXAME_SELECAO           | 726      | 0.84     |
| IN_MATERIAL_PED_MULTIMIDIA                               | 5365     | 6.21     | TP_PROPOSTA_PEDAGOGICA     | 598      | 0.69     |
| IN_MATERIAL_PED_INFANTIL                                 | 5365     | 6.21     | QT_PROF_BIBLIOTECARIO      | 78       | 0.09     |
| IN_MATERIAL_PED_CIENTIFICO                               | 5365     | 6.21     | QT_PROF_SEGURANCA          | 78       | 0.09     |
| IN_MATERIAL_PED_DIFUSAO                                  | 5365     | 6.21     | QT_PROF_SAUDE              | 76       | 0.09     |
| IN_MATERIAL_PED_MUSICAL                                  | 5365     | 6.21     | QT_PROF_ADMINISTRATIVO     | 51       | 0.06     |
| IN_MATERIAL_PED_JOGOS                                    | 5365     | 6.21     | QT_PROF_MONITORES          | 50       | 0.06     |
| IN_MATERIAL_PED_ARTISTICAS                               | 5365     | 6.21     | QT_PROF_PEDAGOGIA          | 32       | 0.04     |
| IN_MATERIAL_PED_DEPORTIVA                                | 5365     | 6.21     | QT_PROF_COORDENADOR        | 27       | 0.03     |
| IN_MATERIAL_PED_INDIGENA                                 | 5365     | 6.21     | QT_PROF_SERVICOS_GERAIS    | 20       | 0.02     |
| IN_MATERIAL_PED_ETNICO                                   | 5365     | 6.21     | IN_SERIE_ANO               | 16       | 0.02     |
| IN_MATERIAL_PED_CAMPO                                    | 5365     | 6.21     | IN_PERIODOS_SEMESTRAIS     | 16       | 0.02     |
| IN_ACESSO_INTERNET_COMPUTADOR                            | 1490     | 1.73     | IN_FUNDAMENTAL_CICLOS      | 16       | 0.02     |
| IN_ACES_INTERNET_DISP_PESSAIS                            | 1490     | 1.73     | IN_GRUPOS_NAO_SERIADOS     | 16       | 0.02     |
| IN_ORGAO_ASS_PAIS  | 1341     | 1.55     | IN_MODULOS                 | 16       | 0.02     |
| IN_ORGAO_ASS_PAIS_MESTRES                                | 1341     | 1.55     | IN_FORMACAO_ALTERNANCIA    | 16       | 0.02     |
| IN_ORGAO_CONSELHO_ESCOLAR                                | 1341     | 1.55     | QT_EQUIP_LOUSA_DIGITAL     | 8        | 0.01     |
| IN_ORGAO_GREMIO_ESTUDANTIL                               | 1341     | 1.55     | QT_PROF_ALIMENTACAO        | 8        | 0.01     |
| IN_ORGAO_OUTROS  | 1341     | 1.55     | QT_COMP_PORTATIL_ALUNO     | 6        | 0.00     |
| IN_TRATAMENTO_LIXO_SEPARACAO                             | 1124     | 1.30     | QT_DESKTOP_ALUNO           | 4        | 0.00     |
| IN_TRATAMENTO_LIXO_REUTILIZA                             | 1124     | 1.30     | QT_PROF_SECRETARIO         | 3        | 0.00     |
| IN_TRATAMENTO_LIXO_RECICLAGEM                            | 1124     | 1.30     | QT_PROF_PSIKOLOGO          | 2        | 0.00     |
| IN_TRATAMENTO_LIXO_INEXISTENTE                           | 1124     | 1.30     | QT_SALAS_UTILIZADAS_DENTRO | 1        | 0.00     |
| IN_REDES_SOCIAIS   | 979      | 1.13     | QT_TABLET_ALUNO            | 1        | 0.00     |
| IN_ESPACO_EQUIPAMENTO                                    | 809      | 0.94     | QT_PROF_NUTRICIONISTA      | 1        | 0.00     |
| IN_ESPACO_ATIVIDADE                                      | 755      | 0.87     | Total                      | 86345    | 100.00   |

Fonte: Elaborado pelo Autor (2022)

Das observações acima, conclui-se que os NAs na base INI como um todo são bem

raros, embora afetem individualmente um número considerável de escolas e atributos. Se fosse feita a opção por trabalhar apenas com registros completos, deletando todas as linhas com algum NA, implicaria num corte de mais de 1/4 das escolas (15828 ou 26.86%).

A Tabela 10 resume a presença de 86345 NAs em 51 atributos da base. Percebe-se que a maior quantidade de NAs estava relacionada às escolas que não indicaram a presença ou não de infraestrutura de rede de dados local (8740 ou 10.12% dos NAs). Escolas que não registraram se usaram ou não material didático de qualquer tipo (5365 ou 6.21% dos NAs) formaram um bloco, bem como se há ou não acesso internet seja computadores ou dispositivos pessoais (1490 ou 1.73% dos NAs), além da presença ou não de algum tipo de órgão ou associação (1341 ou 1.55% dos NAs). Sendo assim, a presença de NAs não parece que foi aleatória ao longo dos atributos (NGUYEN, 2022), constituindo outro argumento contra a simples eliminação de linhas com dados incompletos.

Seja como for, exceção aos atributos Ideb, todos os demais com NAs foram imputados conforme descrito na metodologia (seção 4.8), usando o método de Cópula Gaussiana. Julgou-se melhor não imputar os atributos Ideb pelo fato da presença de NAs nesse campo indicar a ausência da respectiva etapa, entre anos iniciais (sufixo `_INI`) ou anos finais (sufixo `_FIN`). Não faria muito sentido para uma escola que não tem anos finais, por exemplo, imputar-lhe um valor de Ideb para essa etapa.

Cabe destacar que, a fim de fazer uma primeira sondagem da eficiência de imputação da Cópula Gaussiana na base `UNI`, primeiro criou-se uma sub-base `UNI-test` completa, sem os atributos Idebs e sem quaisquer NAs, resultando em uma matriz de 43107 linhas por 150 colunas. A seguir, usando a função `mask_MCAR` (pacote `gcimputer`), foram introduzidos NAs aleatoriamente distribuídos *Missing Completely at Random* (MCAR) numa frequência de 10% por atributo. Escolheu-se esse valor por ser a taxa máxima encontrada na Tabela 10. Foi executado a função `impute_GC` com parâmetros *default*, e então calculado o *Scaled Mean Absolute Error* (SMAE), dada pela fórmula de Zhao & Udell (2020). Um SMAE  $< 1$  indica uma taxa de acerto maior que a simples substituição do NA pela mediana do vetor; SMAE = 1, acerto igual a mediana; SMAE  $> 1$ , pior que a mediana. Foi obtido um valor de 0.85, indicando acertar (em média) 15% a mais o valor NA que a simples substituição pela mediana. Mas, em alguns atributos, o SMAE chegou a 0.15 (ou 85% a mais de acertos que a mediana). Nenhum SMAE foi maior que 1.00.

Certamente que um teste mais rigoroso da imputação seria simular um padrão não-aleatório da distribuição de NAs, conforme Tabela 10. Mas, não se avançou mais no

estudo dos NAs na base UNI, por fugir ao escopo traçado para esta dissertação. Isso será deixado para trabalhos futuros.

### 5.3 ANÁLISE DAS COLINEARIDADES

Os primeiros testes na base SCHOOL2019 evidenciaram as relações de dependência linear descritas na Tabela 11. Todos os atributos da coluna DEPENDENTES foram suprimidos da base ainda na fase de ADEQUAÇÃO DE DADOS, conforme descrito anteriormente (vide seção 4.4.2).

Tabela 11 - Atributos dependentes e suas dependências encontradas na base SCHOOL2019.

| DEPENDENTES                    | DEPENDÊNCIAS  |
|--------------------------------|---|
| IN_LOCAL_FUNC_PRISIONAL_SOCIO: | IN_LOCAL_FUNC_UNID_PRISIONAL + IN_LOCAL_FUNC_SOCIOEDUCATIVO |
| IN_ESGOTO_FOSSA:               | IN_ESGOTO_FOSSA_SEPTICA + IN_ESGOTO_FOSSA_COMUM             |
| IN_ACESSIBILIDADE_INEXISTENTE: | DEMAIS IN_ACESSIBILIDADE_...                                |
| IN_ORGAO_NENHUM:               | DEMAIS IN_ORGAO_...   |
| IN_RESERVA_NENHUMA:            | DEMAIS IN_RESERVA_...                                       |
| QT_SALAS_UTILIZADAS:           | QT_SALAS_UTILIZADAS_DENTRO + QT_SALAS_UTILIZADAS_FORA       |

**Fonte:** Elaborado pelo Autor (2022)

Para a checagem de colinearidades nas bases INI e INI\_IMP, foi utilizada a função `eig` aplicada sobre as correlações: Pearson de INI com linhas NAs removidas, e sobre a Sigma advinda do ajuste normal multivariado em GC na geração da INI\_IMP, conforme Tabela 12. Não se percebe nos últimos autovalores a existência de valores zeros ou muito próximos de zeros, exceção talvez ao 176° em Pearson (INI). Isso poderia indicar ou algum atributo com altas correlações ou com dependências não exatas com outros atributos. Mas, a correlação Sigma retornada da imputação GC não evidencia autovalores tão próximos de zero<sup>5</sup>. Como será a INI\_IMP a base referência, não se aprofundou mais no estudo das colinearidades (dependências lineares).

Tabela 12 - Últimos autovalores das correlações Pearson(INI) e Sigma(INI\_IMP)

| CORRELAÇÃO      | ÚLTIMOS AUTOVALORES |       |       |       |       |       |
|-----------------|---------------------|-------|-------|-------|-------|-------|
|                 | 171                 | 172   | 173   | 174   | 175   | 176   |
| Pearson (INI)   | 0.065               | 0.056 | 0.039 | 0.037 | 0.030 | 0.003 |
| Sigma (INI_IMP) | 0.294               | 0.286 | 0.284 | 0.282 | 0.224 | 0.143 |

**Fonte:** Elaborado pelo Autor (2022)

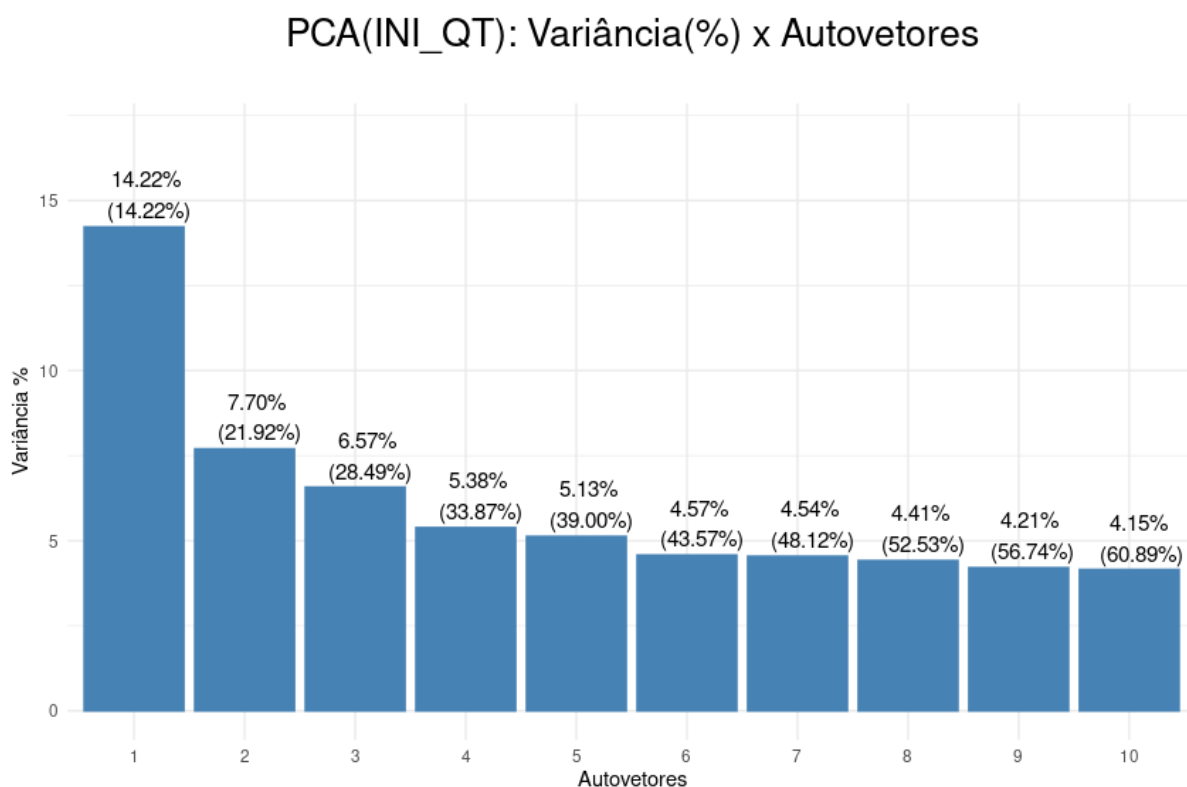
<sup>5</sup> Verificou-se que a função `impute_GC` opera uma regularização caso encontre eigenvalues menores que 0.01 (parâmetro *default*), fazendo este ser o menor eigenvalue.



## 5.4 PCA SOBRE INI\_QT

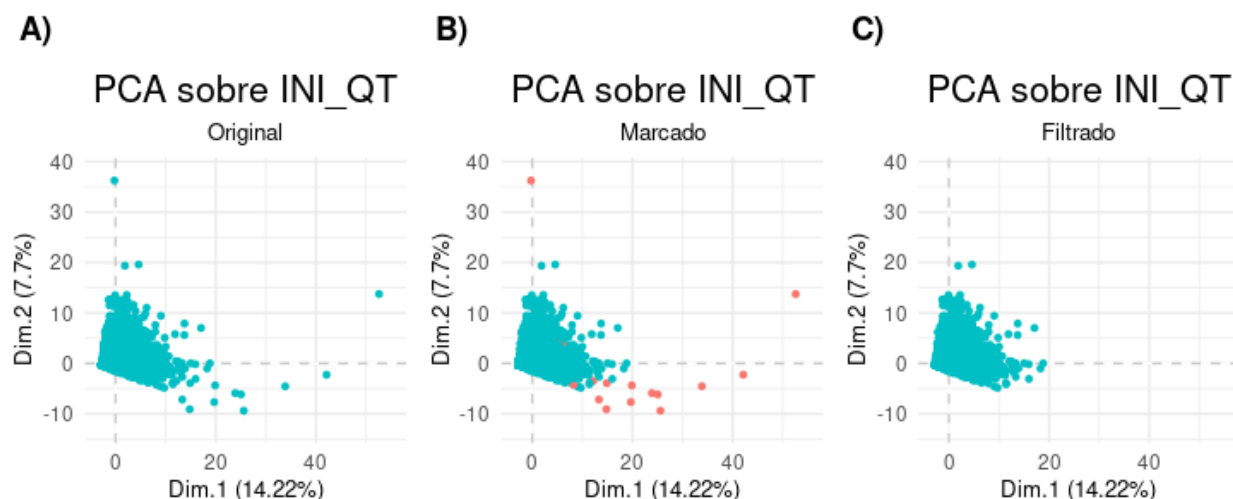
Na seção 4.11, viu-se como valores extremos foram tratados, em especial, o uso da PCA e DBSCAN sobre a base INI\_QT de atributos contínuos.

A Figura 8 mostra os autovalores da PCA sobre a tabela INI\_QT. Percebe-se que os 10 primeiros respondem por mais da metade de toda a variância dos dados (60.89%). O primeiro autovetor tem um autovalor bem maior que os demais, que seguem em um declínio mais suavizado. Isso é sugestivo de que o primeiro autovetor é relevante para destacar uma parte dos valores extremos.



**Figura 8:** Gráfico da variância relativa conforme sequência em ordem decrescente dos autovalores da decomposição PCA sobre tabela INI\_QT para os 10 primeiros autovetores. Entre parênteses destacam-se as variâncias acumuladas.

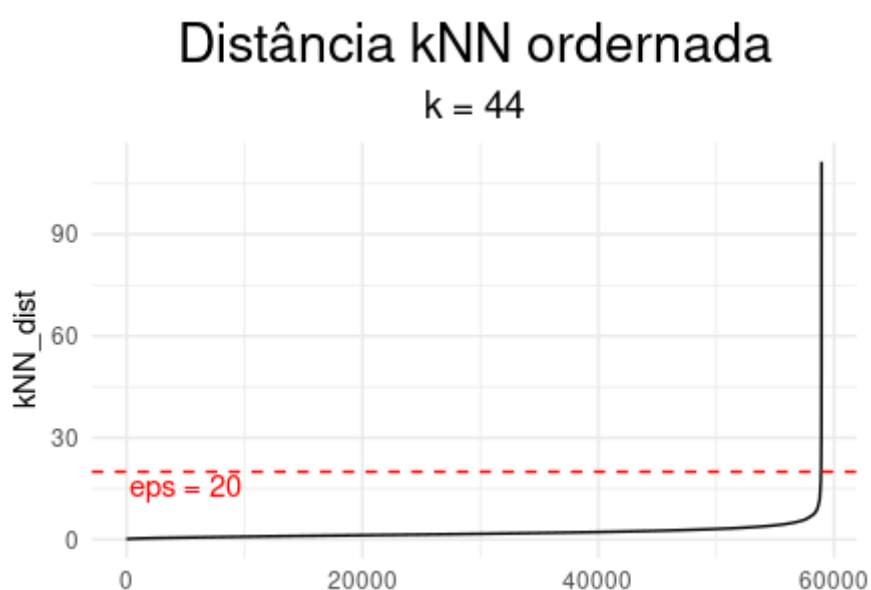
Na Figura 9 pode-se ter uma visualização da decomposição PCA (INI\_QT). Em 9.A, cada ponto representa uma escola, e é possível ver alguns pontos desgarrados nos extremos das coordenadas X e Y, principalmente ao longo do primeiro autovetor. Mas, qual a melhor forma de identificar esses pontos? Optou-se pelo método de clusterização por densidade DBSCAN, como será visto mais adiante.



**Figura 9:** Gráfico do resultado da PCA sobre a tabela INI\_QT. **A)** Cada ponto representa uma das 58935 escolas projetadas sobre os 2 primeiros autovetores (Dim.1 e Dim.2), ou seja, aqueles com os maiores autovalores (maiores variâncias). Percebe-se a presença de pontos com valores extremos ao longo de X e Y. **B)** Em vermelho, 15 pontos extremos identificados pelo DBSCAN e refinamentos do seu agrupamento, conforme descrito no texto. **C)** Distribuição dos pontos após remoção dos pontos vermelhos em B).

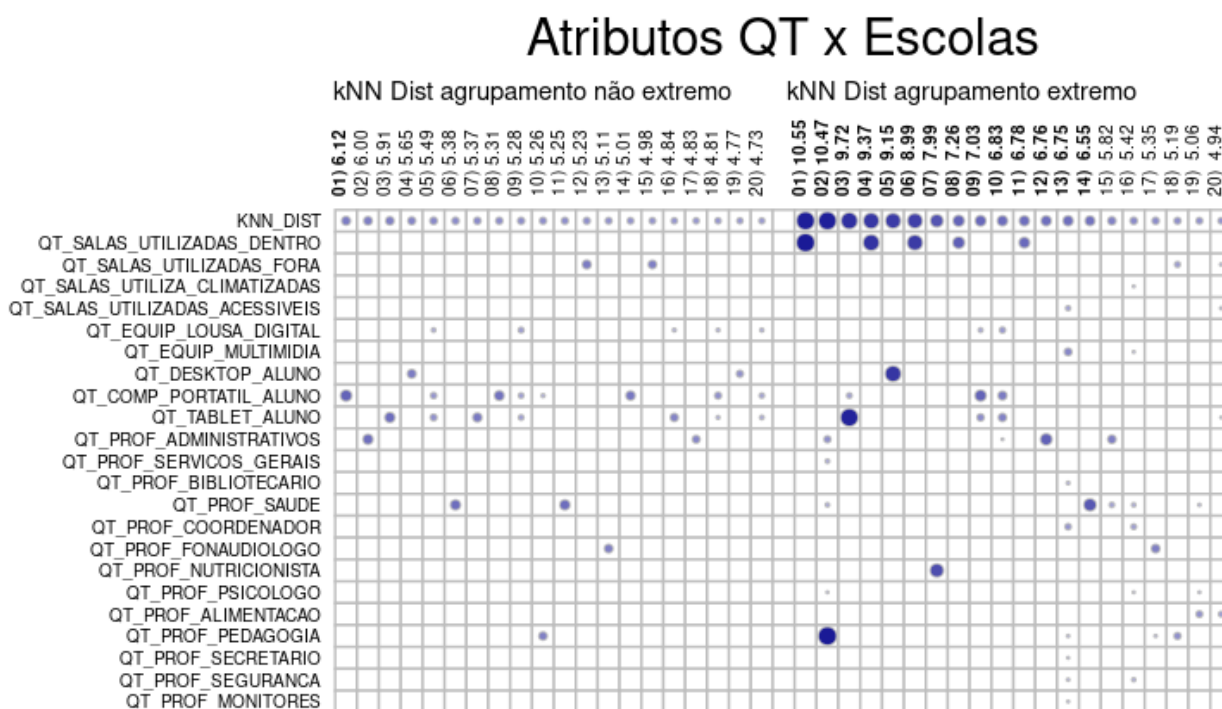
## 5.5 CARACTERIZAÇÃO E ELIMINAÇÃO DE VALORES EXTREMOS POR DBSCAN

A Figura 10 mostra o resultado do kNN ordenado pelas distâncias do vizinho  $k = 44$  sobre a base INI\_QT padronizada em z-score. Após alguns testes empíricos, escolheu-se um raio eps mais conservador de 20, um pouco acima do joelho, por melhor marcar visualmente os pontos extremos (cluster id = 0) no gráfico PCA (Figura 9.B).



**Figura 10:** Gráfico das distâncias ordenadas do k vizinho mais próximo de cada ponto, calculado pela função `kNNdist()` do pacote `dbSCAN` do R, sobre a base INI-QT, com parâmetro k igual ao dobro do número de dimensões dessa base (44), conforme recomendado em [Schubert \(2017\)](#). Nota-se um joelho abaixo da distância 20 (linha tracejada), sugerindo este um primeiro valor para o raio eps a ser usado no DBSCAN.

Foi então rodado um DBSCAN com parâmetros raio eps = 20 e MinPts = 44 sobre a base INI\_QT também padronizada por z-score, resultando em 20 escolas que não se agruparam e receberam cluster id = 0. Escolas que não se agrupam podem ser interpretadas como muito distantes das demais, configurando candidatas a pontos extremos no hiperespaço dos atributos.



**Figura 11:** Gráfico dos atributos QT em função das escolas ordenadas pelas maiores distâncias kNN conforme o agrupamento DBSCAN. À esquerda, 20 escolas do agrupamento considerado não extremo pelo DBSCAN (cluster id > 0). À direita, 20 escolas do agrupamento considerado extremo pelo DBSCAN (cluster id = 0). Os valores nos rótulos das colunas indicam as distâncias kNN por escola, em ordem decrescente. As áreas dos círculos são proporcionais aos valores de cada célula. Pontos nas células com valores acima de 3 foram cortados, a fim de despoluir o gráfico. Na primeira linha, estão os círculos representando as distâncias kNN ordenadas. Nas demais linhas, tem-se as quantidades dos atributos padronizados por z-score. A fim de tornar a visualização mais adequada, os números foram escalados pela raiz quadrada dos valores absolutos. Em negrito, nos rótulos das colunas, as distâncias kNN acima do ponto de corte de 6.00.

O quadro da Figura 11 mostra uma análise prévia comparativa entre as 20 escolas com maiores kNN, conforme o agrupamento DBSCAN em “não extremas” (cluster id > 0) e “extremas” (cluster id = 0). Percebe-se que, como esperado, as escolas em agrupamento extremo tinham as maiores distâncias kNN. Nota-se também que essas escolas apresentavam um ou poucos atributos com valores extremados. Lembrando que a distância kNN em questão indica a distância do k-ésimo vizinho mais distante, no caso, do vizinho 44, dentro do raio eps = 20. Um ponto pode ter seu vizinho 44 distante, mas ainda assim conseguir se enquadrar num grupo (ou seja, ter 44 vizinhos dentro do raio eps = 20), mas isso não é garantido. Dessa forma, para uma mesma distância kNN, um ponto pode conseguir se agrupar ou não,

dependendo do perfil da sua vizinhança.

Feita a identificação e caracterização dos pontos extremos conforme Figura 11, por serem poucos em relação ao tamanho da base, optou-se pela eliminação desses pontos. A fim de operar um refinamento das escolas que seriam eliminadas da base, decidiu-se por um ponto de corte na distância kNN acima de 6.00, estivessem elas agrupadas ou não, totalizando 15 escolas a serem eliminadas, gerando uma base INI filtrada (doravante INI\_FIL) de 58920 linhas por 184 colunas. Esses pontos eliminados podem ser vistos em negrito nos rótulos das colunas na Figura 11. O resultado da eliminação está na Figura 9.C. Cabe dizer que a inspeção visual dos gráficos das figuras 11 e 9.C também orientou a escolha do ponto de corte em 6.00.

Tabela 13 - Filtragem de valores extremos

| ATRIBUTOS                      | ORIGINAL |     |     |     |     |     |     |     | FILTRADO |     |     |     |     |     |     |     |
|--------------------------------|----------|-----|-----|-----|-----|-----|-----|-----|----------|-----|-----|-----|-----|-----|-----|-----|
|                                | 1        | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 1        | 2   | 3   | 4   | 5   | 6   | 7   | 8   |
| QT_SALAS_UTILIZADAS_DENTRO     | 1013     | 808 | 747 | 500 | 400 | 193 | 177 | 143 | 193      | 177 | 143 | 121 | 121 | 116 | 114 | 111 |
| QT_SALAS_UTILIZADAS_FORA       | 90       | 84  | 63  | 57  | 57  | 52  | 50  | 46  | 90       | 84  | 63  | 57  | 57  | 52  | 50  | 46  |
| QT_SALAS_UTILIZA_CLIMATIZADAS  | 90       | 55  | 53  | 47  | 40  | 39  | 39  | 39  | 90       | 55  | 53  | 47  | 40  | 39  | 39  | 39  |
| QT_SALAS_UTILIZADAS_ACESSIVEIS | 80       | 64  | 53  | 50  | 49  | 47  | 47  | 40  | 64       | 53  | 50  | 49  | 47  | 47  | 40  | 39  |
| QT_EQUIP_LOUSA_DIGITAL         | 30       | 28  | 26  | 26  | 24  | 24  | 23  | 23  | 28       | 26  | 26  | 24  | 24  | 23  | 23  | 23  |
| QT_EQUIP_MULTIMIDIA            | 206      | 94  | 88  | 88  | 77  | 73  | 73  | 71  | 94       | 88  | 88  | 77  | 73  | 73  | 71  | 71  |
| QT_DESKTOP_ALUNO               | 922      | 387 | 284 | 252 | 244 | 212 | 203 | 167 | 387      | 284 | 252 | 244 | 212 | 203 | 167 | 150 |
| QT_COMP_PORTATIL_ALUNO         | 780      | 778 | 635 | 589 | 555 | 550 | 513 | 495 | 635      | 589 | 550 | 513 | 495 | 473 | 440 | 440 |
| QT_TABLET_ALUNO                | 1616     | 700 | 600 | 520 | 508 | 450 | 450 | 440 | 700      | 600 | 508 | 450 | 450 | 440 | 423 | 420 |
| QT_PROF_ADMINISTRATIVO         | 137      | 113 | 88  | 79  | 73  | 72  | 68  | 64  | 113      | 88  | 79  | 73  | 72  | 64  | 60  | 51  |
| QT_PROF_SERVICOS_GERAIS        | 71       | 68  | 66  | 61  | 55  | 51  | 51  | 45  | 71       | 66  | 61  | 55  | 51  | 51  | 45  | 44  |
| QT_PROF_BIBLIOTECARIO          | 14       | 14  | 13  | 12  | 11  | 11  | 11  | 10  | 14       | 14  | 12  | 11  | 11  | 11  | 10  | 10  |
| QT_PROF_SAUDE                  | 16       | 12  | 12  | 10  | 10  | 10  | 9   | 8   | 12       | 12  | 10  | 10  | 10  | 9   | 8   | 8   |
| QT_PROF_COORDENADOR            | 36       | 30  | 27  | 26  | 25  | 25  | 24  | 24  | 36       | 27  | 26  | 25  | 25  | 24  | 24  | 23  |
| QT_PROF_FONOAUDIOLOGO          | 6        | 6   | 3   | 3   | 3   | 3   | 2   | 2   | 6        | 6   | 3   | 3   | 3   | 3   | 2   | 2   |
| QT_PROF_NUTRICIONISTA          | 34       | 12  | 11  | 11  | 11  | 11  | 11  | 4   | 12       | 11  | 11  | 11  | 11  | 11  | 4   | 4   |
| QT_PROF_PSIKOLOGO              | 5        | 4   | 4   | 4   | 4   | 4   | 4   | 4   | 5        | 4   | 4   | 4   | 4   | 4   | 4   | 4   |
| QT_PROF_ALIMENTAÇÃO            | 54       | 43  | 34  | 33  | 32  | 30  | 27  | 27  | 54       | 43  | 34  | 33  | 32  | 30  | 27  | 27  |
| QT_PROF_PEDAGOGIA              | 398      | 122 | 96  | 95  | 92  | 76  | 76  | 74  | 122      | 96  | 95  | 92  | 76  | 76  | 74  | 73  |
| QT_PROF_SECRETARIO             | 11       | 11  | 9   | 8   | 8   | 8   | 8   | 7   | 11       | 11  | 9   | 8   | 8   | 8   | 8   | 7   |
| QT_PROF_SEGURANCA              | 17       | 14  | 14  | 13  | 12  | 12  | 12  | 12  | 17       | 14  | 13  | 12  | 12  | 12  | 12  | 12  |
| QT_PROF_MONITORES              | 34       | 33  | 30  | 28  | 27  | 25  | 25  | 25  | 34       | 33  | 30  | 28  | 27  | 25  | 25  | 25  |

**Fonte:** Elaborado pelo Autor (2022)

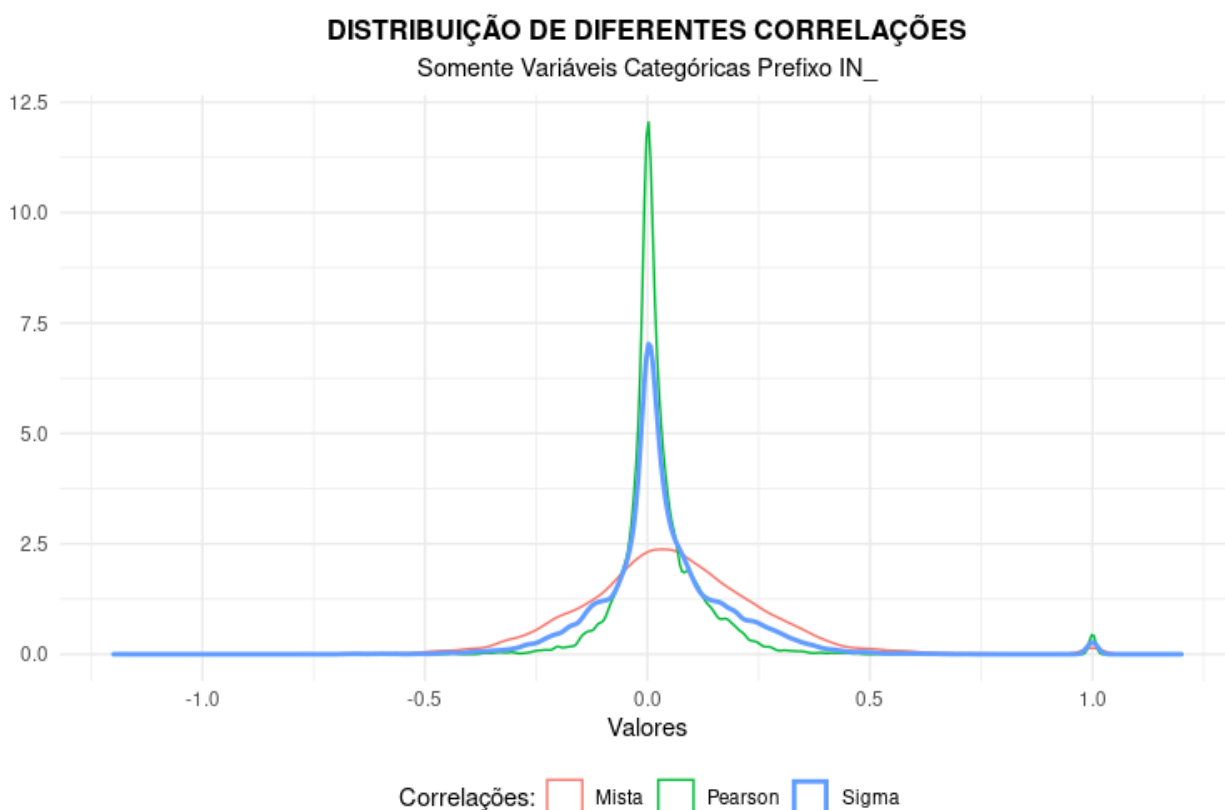
**Nota:** Relação dos 8 maiores valores em ordem decrescente por atributo QT da tabela INI\_QT com valores originais e após filtragem de valores extremos

A Tabela 13 mostra uma relação numérica dos 8 maiores valores mais extremos antes e depois da filtragem do DBSCAN. Observa-se que, após filtragem, houve diminuição considerável da presença de valores discrepantes em relação aos demais.

## 5.6 CORRELAÇÕES

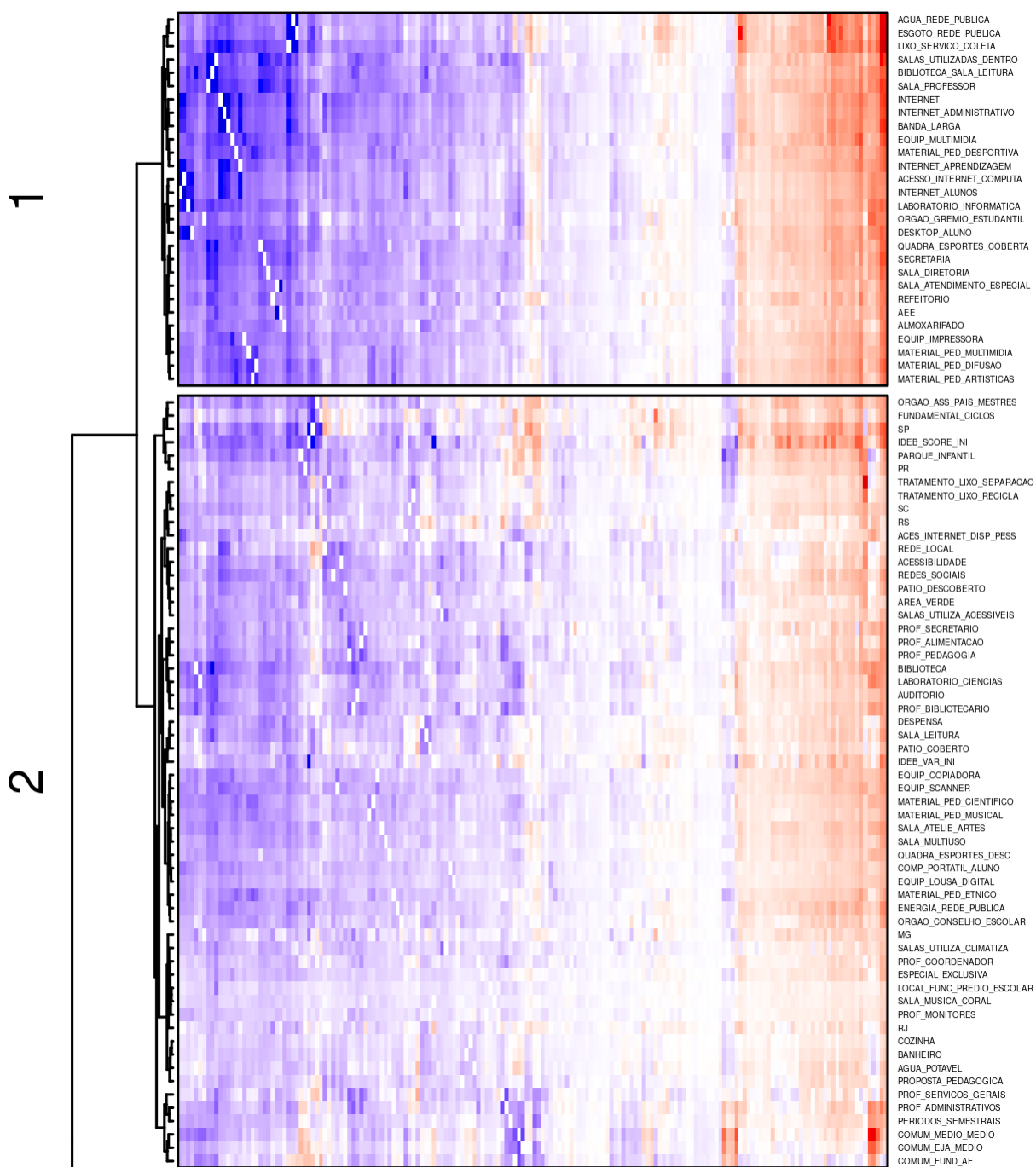
Conforme explicitado na metodologia, entre as correlações Pearson (clássica), Sigma (advinda da imputação por Cópula Gaussiana) e Mista (customizada conforme tipo de cada par de atributos), optou-se pela Sigma. Todas elas apresentam vantagens e desvantagens, sendo que correlações Pearson não são adequadas para dados categóricos, e a Mista pode gerar uma matriz de correlação não positiva definida, que dificulta a interpretação das decomposições matriciais, com soluções polêmicas (LORENZO-SEVA & FERRANDO, 2021). A Sigma leva em conta dados categóricos e gerou uma matriz positiva definida.

Fez-se uma comparação desses três tipos de correlações, com foco apenas em atributos categóricos com prefixo “IN\_”, conforme Figura 12 abaixo. Percebe-se que a matriz de correlação Sigma apresenta uma distribuição de valores intermediária entre Mista e Pearson. Ou seja, não parece subestimar tanto as correlações entre atributos categóricos, ao mesmo tempo que gera uma matriz positiva definida que pode ser devidamente processada pelos algoritmos de decomposição matricial e interpretada.

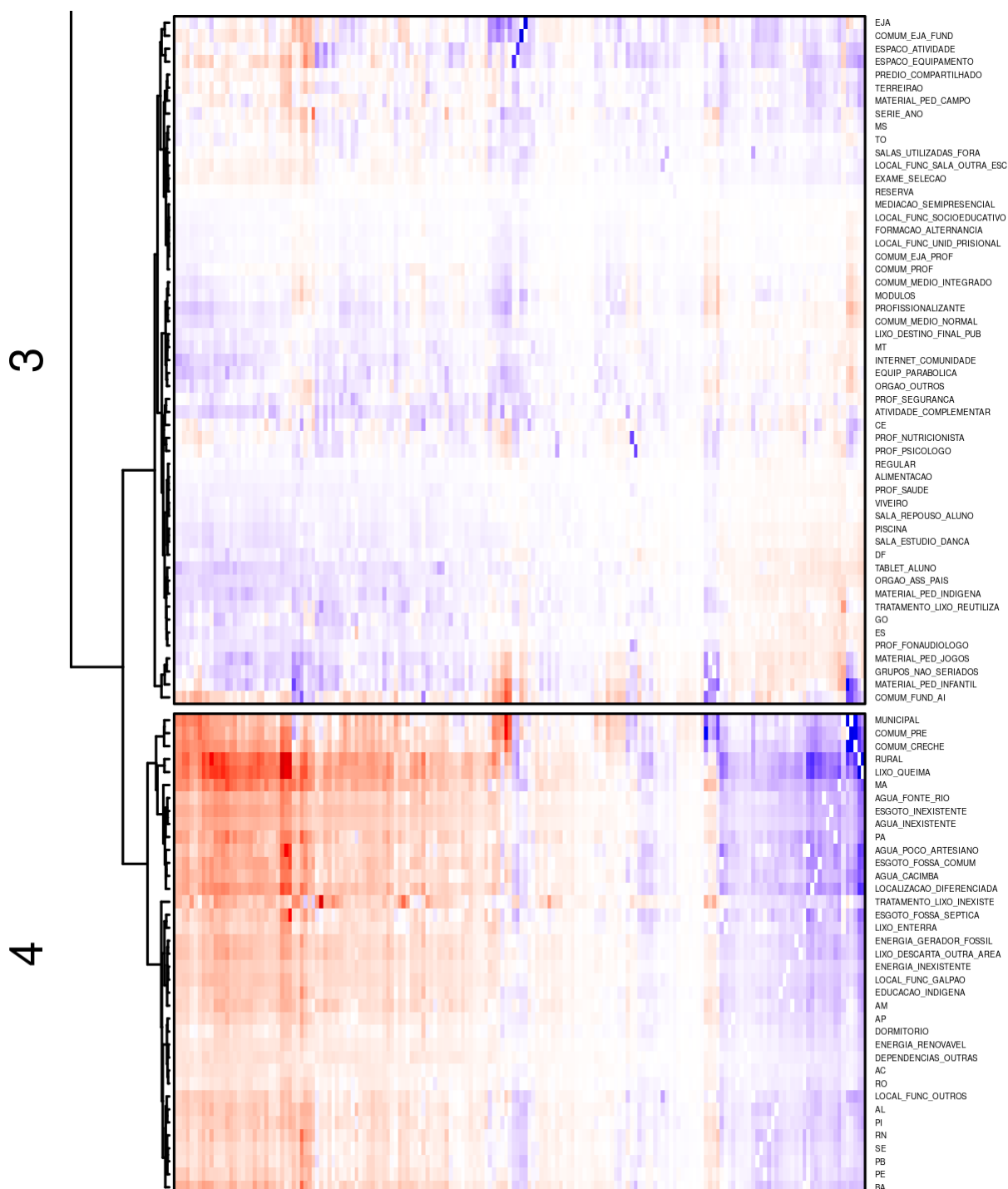


**Figura 12** : Gráfico comparativo da distribuição dos valores de três matrizes de correlações diferentes, entre: Mista, Pearson e Sigma. A comparação levou em conta apenas atributos categóricos com prefixo “IN\_”. A matriz Sigma gerou uma distribuição intermediária entre Mista e Pearson.

## CORRELAÇÕES GC-SIGMA EM INI-FIL



**Figura 13a** : Correlograma em forma de um mapa de calor (heatmap) destacando as correlações Sigma advindas do Gaussian Copula sobre 176 atributos da base INI-FIL. A matriz Sigma é quadra 176 x 176, mas foi deformada na Figura para melhor destaque dos agrupamentos dos atributos em linhas. Os prefixos nos nomes dos atributos foram removidos para maior clareza. As cores indicam os valores das correlações, sendo negativos em vermelho; zeros em branco; e positivos em azul. À esquerda, destaca-se o dendrograma do agrupamento aglomerativo hierárquico dos atributos, em 4 sub-árvores de alto nível, nesta Figura mostrando apenas as 2 primeiras. Figura produzida utilizando a função Heatmap do pacote ComplexHeatmap, com método aglomerativo `ward.D2`, e reordem das sub-árvores do dendrograma por OLO (*Optimal Leaf Ordering*).



**Figura 13b** : Continuação da Figura 13a, mostrando as outras sub-árvores 3 e 4 do dendrograma.

As figuras 13a e 13b mostram uma visão geral da matriz de correlação Sigma advinda do Cópula Gaussiana para a base INI-FIL. Lembrando que, como essa base possui 176 atributos não identificadores, a matriz Sigma será quadrada de 176 x 176, mas nas figuras aparecem deformadas em favor da altura, para melhor visualização dos agrupamentos. Esses agrupamentos foram hierárquicos, usando método aglomerativo Ward, conforme descrito na metodologia (vide seção 4.13). São mostrados apenas os 4 primeiros grandes grupos ou

sub-árvores do dendrograma. O padrão de cores segue uma paleta de tonalidades do azul - branco - vermelho, indicando correlações entre 1, 0 e -1, respectivamente.

Num primeiro olhar panorâmico, é possível notar que há estruturas não aleatórias formando padrões de correlações. Na sub-árvore 1, houve um conjunto de atributos com correlações mais evidentes, dadas as tonalidades mais fortes de azul (positivas) e vermelho (negativas). Na sub-árvore 2, as correlações pareceram mais amenizadas em relação à sub-árvore 1. Já na sub-árvore 3, reuniu-se um conjunto de atributos com correlações próximas de neutras. Finalmente, na sub-árvore 4, agrupou-se atributos com padrões opostos aos das primeiras sub-árvores, com inversão das cores azul por vermelho e vice-versa.

Aprofundando um pouco mais o olhar, percebe-se que a sub-árvore 1 agrupou atributos mais relacionados à infraestrutura básica, como REDE PÚBLICA DE ÁGUA E ESGOTO; várias formas de INTERNET; LABORATÓRIOS E EQUIPAMENTOS DE INFORMÁTICA; QUADRA COBERTA, SECRETARIA, REFEITÓRIO, ALMOXARIFADO; além de alguns tipos de MATERIAL PEDAGÓGICO. Interessante notar que não apareceu nenhuma unidade da federação nesta primeira sub-árvore. Talvez isso indique que nenhum estado contemple esse conjunto de infraestruturas de forma completa, apenas parcialmente, o que ressaltaria a diversidade e regionalização desses atributos.

Na sub-árvore 2, já apareceram algumas unidades da federação, mas todas das regiões Sul e Sudeste. Os atributos de infraestrutura tendem a ser um pouco mais sofisticados, como TRATAMENTO DE LIXO; ACESSIBILIDADE; ÁREA VERDE; SALAS MULTIUSO, CLIMATIZADAS, ATELIÊS, MÚSICA E CORAL; LOUSA DIGITAL; COMPUTADOR PORTÁTIL. Destacou-se também a presença de órgãos como ASSOCIAÇÕES DE PAIS E MESTRES e CONSELHO ESCOLAR, além da existência de PROPOSTA PEDAGÓGICA. Curiosamente, foi nesta sub-árvore que apareceram os atributos relacionados às avaliações do Ideb, tanto a NOTA 2019 quanto a VARIAÇÃO DO Ideb de 2005 a 2019.

Na sub-árvore 3, das correlações tendentes à mais próximas de neutras, surgiram as unidades da federação do Centro-Oeste, além do Ceará e Espírito Santo. Trata-se de uma sub-árvore mais complexa pela diversidade dos atributos reunidos, conseqüentemente dos perfis diferenciados de escolas. Ficaram nesta sub-árvore escolas com certas especificidades, como EJA; as que faziam RESERVA e SELEÇÃO; SOCIOEDUCATIVAS E PRISIONAIS; PROFISSIONALIZANTES; que funcionavam em PRÉDIO COMPARTILHADO ou tinham TERREIRÃO. Ao mesmo tempo também aquelas que tinham mais PROFISSIONAIS DE SAÚDE, NUTRICIONISTA, PSICÓLOGO; PISCINA, ESTÚDIO DANÇA e VIVEIRO. Parece que esta sub-árvore 3 funcionou como um refugio, reunindo todos os atributos que não



se encaixaram bem no padrão das outras sub-árvores.

Por fim, a sub-árvore 4, como uma inversão das demais, capturou o resto das unidades da federação, especialmente Norte e Nordeste (exceção do Ceará). Tendem a ser escolas em situação de maior precariedade ou deficiência de infraestrutura básica. Pra começar, uma boa parte são RURAIS, embora também MUNICIPAIS; QUEIMAM OU ENTERRAM LIXO; NÃO TÊM ÁGUA OU ESGOTO; ou ÁGUA DE POÇO ARTESIANO, CACIMBA, OU DE RIO. O ESGOTO pode ser de FOSSA COMUM OU SÉPTICA. Podem ter DORMITÓRIO, FUNCIONAR EM GALPÃO; ENERGIA INEXISTENTE OU POR GERADOR. Não parece coincidência a presença nesta sub-árvore da EDUCAÇÃO INDÍGENA.

Foram destacados aqui apenas os padrões mais evidentes. Muitas outras análises mais detalhadas em sub-árvores mais profundas poderiam ser feitas. Porém, o agrupamento hierárquico pela matriz de correlação tem algumas limitações (ROKACH & MAIMON, 2005). Uma sempre destacada é que o processo de clusterização não tem retorno ou revisão (*backtracking*), ou seja, uma vez completada uma aglomeração numa determinada etapa, ela é permanente, mesmo que houvesse uma outra possibilidade melhor numa etapa posterior. Dessa forma, o agrupamento hierárquico pode ser capturado por mínimos locais, e não oferecer uma estrutura ótima. Por isso, quanto mais se desce a análise em sub-árvores mais profundas, maiores as chances de se afastar de um resultado ótimo. Em função disso, nas figuras 13a e 13b salientou-se apenas as 4 maiores sub-árvores.

Outra limitação é, sendo o algoritmo um processo hierárquico em árvore, uma folha (no caso, um atributo) tem que pertencer apenas a uma sub-árvore, mesmo que aquele atributo fizesse sentido em mais de uma sub-árvore. Por exemplo, na Figura 13b, pelo padrão de cores, o atributo IN\_COMUM\_FUND\_AI, que ficou na sub-árvore 3, provavelmente poderia ser encaixado também na sub-árvore 4. Nesse caso, deixaria de ser uma estrutura em árvore e passaria a ter que ser tratado como um grafo. Existem muitas formas de clusterizar grafos (FORTUNATO, 2010), mas neste trabalho, dado seu caráter de análise exploratória com foco nas correlações/associações dos atributos de infraestrutura e as avaliações do Ideb, optou-se por outras técnicas como as que serão vistas a seguir.

## 5.7 RESULTADOS DA AFERIÇÃO DA BASE PARA A FA

Na Tabela 14 é possível ver uma estatística geral dos valores do teste KMO sobre todos os atributos da matriz de correlação Sigma de INI-FIL. Segundo os autores do teste (KAISER, 1974), KMO na casa de 0.9 é considerado “maravilhoso”; de 0.8 “meritório”; de

0.7 “mediano”; 0.6 é "mediocre"; de 0.5 “miserável”; e abaixo de 0.5, “inaceitável”<sup>6</sup>. Percebe-se uma mediana e médias próximas de 0.95, um mínimo de 0.70 (QT\_PROF\_PSIKOLOGO) e máximo de 1.00 (IN\_LOCAL\_FUNC\_GALPAO, IN\_ESGOTO\_INEXISTENTE, IN\_AGUA\_INEXISTENTE). Logo, os resultados do KMO indicam que a matriz de correlação Sigma de INI-FIL está satisfatoriamente adequada para a FA, entre “mediano” e “maravilhoso”.

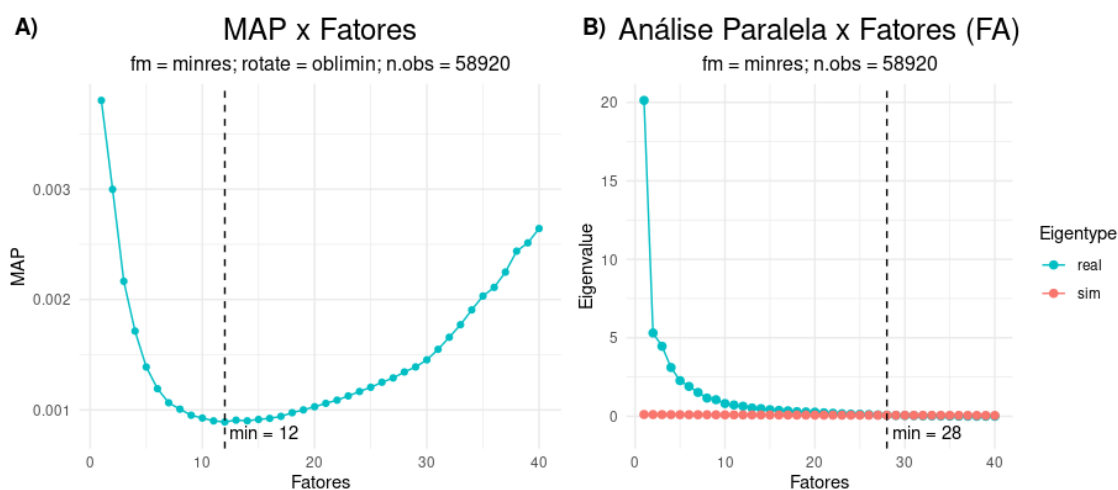
Tabela 14 - Sumário estatístico básico do KMO aplicado à matriz Sigma de INI-FIL

| ESTATÍSTICAS DO TESTE KMO |            |         |       |            |        |
|---------------------------|------------|---------|-------|------------|--------|
| Mínimo                    | 1º Quartil | Mediana | Média | 3º Quartil | Máximo |
| 0.70                      | 0.94       | 0.96    | 0.95  | 0.98       | 1.00   |

Fonte: Elaborado pelo Autor (2022)

## 5.8 ESTIMAÇÃO DO NÚMERO DE FATORES NA FA

A Figura 14 indica a estimação de números de fatores a serem considerados na FA. O método MAP indica algo em torno de 12 fatores e a análise paralela um máximo de 28. A Figura 14B também nos permite explorar visualmente o método do “joelho” (*scree test*), um ponto de “dobra” na curva dos autovalores, sugerindo um número de fatores em torno de 10.



**Figura 14:** Gráficos para a estimação do número de fatores a serem usados na FA. **A)** Variação do critério MAP em função do número de fatores. Observa-se um ponto de mínimo em torno do número 12. **B)** Variação dos eigenvalues com dados reais da base INI-FIL (azul) e dados de uma simulação randômica (vermelho). O número 28 representa o número máximo de fatores em que os eigenvalues reais são maiores que os simulados randomicamente. Gráficos gerados com as funções `VSS` e `fa.parallel` do pacote `psych`, parâmetros destacados nos subtítulos.

<sup>6</sup> Do inglês original: *marvelous*, *meritorious*, *middling*, *mediocre*, *miserable* e *unacceptable*, respectivamente (KAISER, 1974).

Qual número de fatores reter é algo bem debatido e controverso na literatura (VAN DER EIJK & ROSE, 2015). Mas, todos parecem ponderar que qualquer que seja o método haverá prós e contras, e eles podem depender da base de dados e dos objetivos da pesquisa (WATKINS, 2018). Logo, seria prudente contrastar diferentes métodos, deixando claro os critérios adotados na escolha do número de fatores (REVELLE, 2009). Para esta dissertação, contrastou-se três métodos diferentes. E dado o seu carácter exploratório, optou-se por um número mais conservativo de 10 fatores, de modo a evitar superdimensionamentos, e dar maior peso aos primeiros fatores que comumente acumulam maior variância explicada.

## 5.9 RESULTADOS E ANÁLISES DA FA

Segue-se o resultado da metodologia FA recursiva de Storopoli (2019) advinda da matriz de correlação Sigma da GC sobre INI-FIL. Conforme dito anteriormente (vide seção 4.14), tal metodologia permite tanto a seleção de atributos quanto a redução dimensional.

A seleção de atributos é feita com base na remoção dos atributos com *communalities* menores que um ponto de corte ou *cutoff* (no caso, 0.495), até que restem apenas *communalities* maiores que 50% (inclusive). E já que  $uniqueness = 1 - communalities$ , estarão sendo eliminados os atributos com maiores *uniqueness*, aqueles que têm pouca variância compartilhada com outros atributos.

A Tabela 15 elenca 49 atributos que foram eliminados por apresentarem variâncias mais intrínsecas, conseqüentemente sem muita relevância aos objetivos desta dissertação. Logo, a base INI-FIL, considerando atributos de dados, cai de 176 para 127 atributos, uma redução de quase 30%.

Percebe-se que boa parte dos atributos nesta tabela são coincidentes com a análise das correlações exibidas na sub-árvore 3 da Figura 13b, que foi identificada como uma sub-árvore de refugio, condensando atributos que não puderam ser agrupados em outras sub-árvores. Como já dito antes, tais atributos parecem representar escolas com especificidades de perfil, como EJA, PROFISSIONALIZANTES, PRISIONAIS; ou especificidades de infraestrutura, como PISCINA, ESTÚDIO DANÇA, MÚSICA CORAL, VIVEIROS, RESERVA, EXAME DE SELEÇÃO.

Tabela 15 - Listagem de atributos com communalities menores que 0.495

| SELEÇÃO DE ATRIBUTOS |                               |                      |           |                              |                      |
|----------------------|-------------------------------|----------------------|-----------|------------------------------|----------------------|
| <b>Id</b>            | <b>Atributos</b>              | <b>Communalities</b> | <b>Id</b> | <b>Atributos</b>             | <b>Communalities</b> |
| 1                    | IN_RESERVA                    | 0.001                | 26        | UF_AC                        | 0.258                |
| 2                    | IN_MEDIACAO_SEMIPRESENCIAL    | 0.009                | 27        | IN_ORGAO_ASS_PAIS            | 0.273                |
| 3                    | IN_LOCAL_FUNC_SOCIOEDUCATIVO  | 0.023                | 28        | QT_PROF_SEGURANCA            | 0.273                |
| 4                    | IN_ALIMENTACAO                | 0.029                | 29        | IN_PREDIO_COMPARTILHADO      | 0.287                |
| 5                    | IN_FORMACAO_ALTERNANCIA       | 0.030                | 30        | IN_DEPENDENCIAS_OUTRAS       | 0.304                |
| 6                    | IN_COMUM_EJA_PROF             | 0.035                | 31        | UF_RO                        | 0.308                |
| 7                    | IN_VIVEIRO                    | 0.044                | 32        | IN_INTERNET_COMUNIDADE       | 0.309                |
| 8                    | QT_PROF_SAUDE                 | 0.052                | 33        | IN_ENERGIA_RENOVAVEL         | 0.317                |
| 9                    | IN_LOCAL_FUNC_UNID_PRISIONAL  | 0.052                | 34        | QT_TABLET_ALUNO              | 0.324                |
| 10                   | IN_REGULAR                    | 0.067                | 35        | UF_GO                        | 0.335                |
| 11                   | IN_SALA_REPOUSO_ALUNO         | 0.068                | 36        | IN_MODULOS                   | 0.336                |
| 12                   | IN_LIXO_DESTINO_FINAL_PUBLICO | 0.089                | 37        | QT_PROF_FONAUDILOGO          | 0.344                |
| 13                   | IN_EXAME_SELECAO              | 0.114                | 38        | IN_LOCAL_FUNC_PREDIO_ESCOLAR | 0.349                |
| 14                   | IN_COMUM_PROF                 | 0.131                | 39        | IN_SALA_MUSICA_CORAL         | 0.363                |
| 15                   | IN_PISCINA                    | 0.135                | 40        | IN_MATERIAL_PED_INDIGENA     | 0.379                |
| 16                   | UF_TO                         | 0.151                | 41        | IN_ORGAO_OUTROS              | 0.383                |
| 17                   | UF_MT                         | 0.130                | 42        | IN_TERREIRAO                 | 0.389                |
| 18                   | IN_SALA_ESTUDIO_DANCA         | 0.154                | 43        | IN_DORMITORIO                | 0.397                |
| 19                   | IN_LOCAL_FUNC_SALAS_OUTRA_ESC | 0.174                | 44        | IN_TRATAMENTO_LIXO_REUTILIZA | 0.414                |
| 20                   | UF_DF                         | 0.187                | 45        | IN_PROFISSIONALIZANTE        | 0.418                |
| 21                   | UF_ES                         | 0.190                | 46        | IN_EQUIP_PARABOLICA          | 0.422                |
| 22                   | IN_COMUM_MEDIO_NORMAL         | 0.206                | 47        | QT_PROF_MONITORES            | 0.448                |
| 23                   | QT_SALAS_UTILIZADAS_FORA      | 0.206                | 48        | IN_MATERIAL_PED_CAMPO        | 0.460                |
| 24                   | IN_COMUM_MEDIO_INTEGRADO      | 0.227                | 49        | IN_COZINHA                   | 0.476                |
| 25                   | UF_MS                         | 0.229                |           |                              |                      |

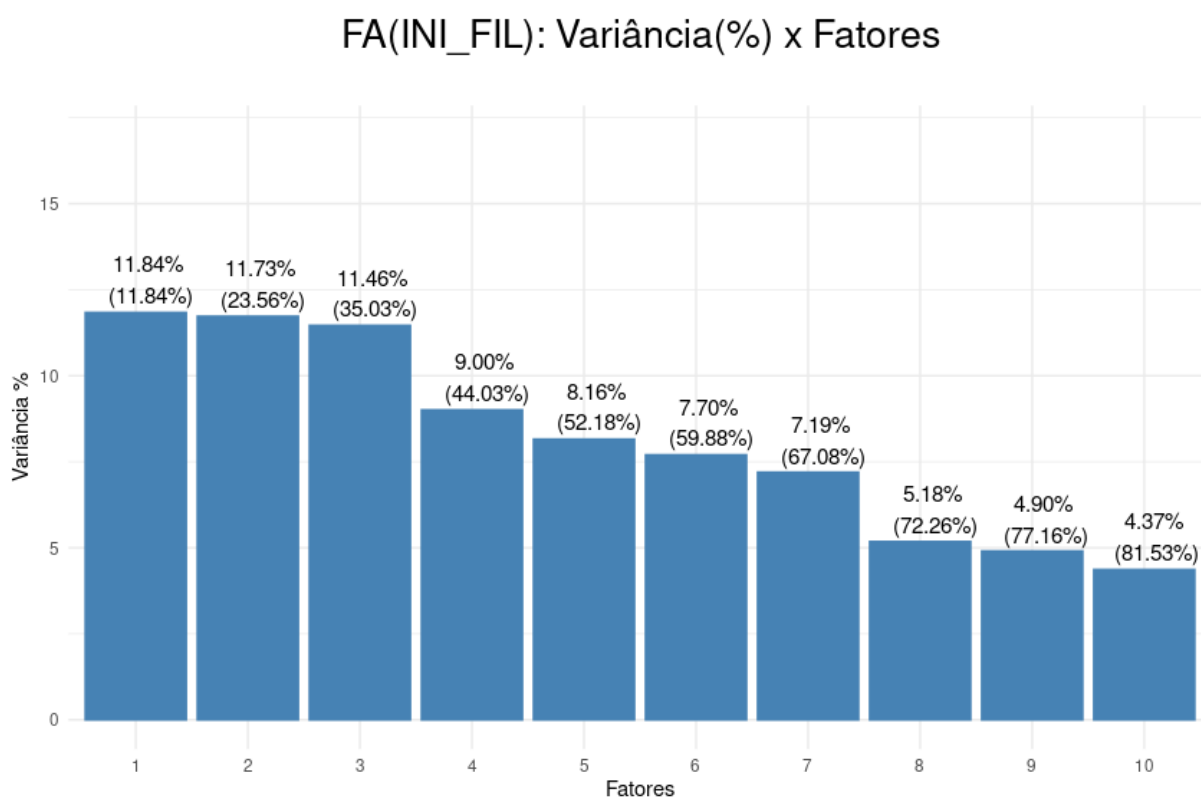
**Fonte:** Elaborado pelo Autor (2022)

Apesar de haver uma correspondência, ela não é total. Por exemplo: o atributo Ceará (UF\_CE) estava na sub-árvore 3, mas não está entre os que têm baixa *communalities*. Por outro lado, o IN\_COZINHA não estava na sub-árvore 3, mas acabou eliminada na FA, ainda que tivesse uma *communality* próxima de 50%. Isso provavelmente se deve à metodologia muito mais sofisticada da FA na estimação de variâncias mais puras e compartilhadas, e não de variâncias totais como é a matriz de correlação Sigma da Figura 13b.

Com relação às unidades da federação eliminadas conforme Tabela 15, curiosamente nota-se que todas da região Centro-Oeste estão presentes: MT, DF, MS, GO. Essas unidades juntamente com ES e TO estavam na sub-árvore 3 da Figura 13b. Importante salientar que a eliminação dessas unidades não deve ser interpretada como entes federativos sem relevância, apenas sugere que não foram encontradas variâncias compartilhadas significativas com os atributos do Educacenso utilizados nesta dissertação. Ou que não bastaram esses atributos

para estabelecer um perfil dessas unidades.

Com relação à redução dimensional, conforme análise na seção anterior, assumiu-se 10 fatores como um subespaço apropriado. A Figura 15 mostra a distribuição das variâncias explicadas por fator e acumulada. Nota-se que os 10 fatores em conjunto conseguiram explicar perto de 80% da variância dos dados, um índice de explicação razoável, e que reforça o acerto na escolha desse número de dimensões.



**Figura 15:** Gráfico da variância relativa conforme sequência em ordem decrescente da variância explicada por cada fator da decomposição FA sobre a matriz Sigma para os 10 primeiros autovetores. Entre parênteses destacam-se as variâncias acumuladas.

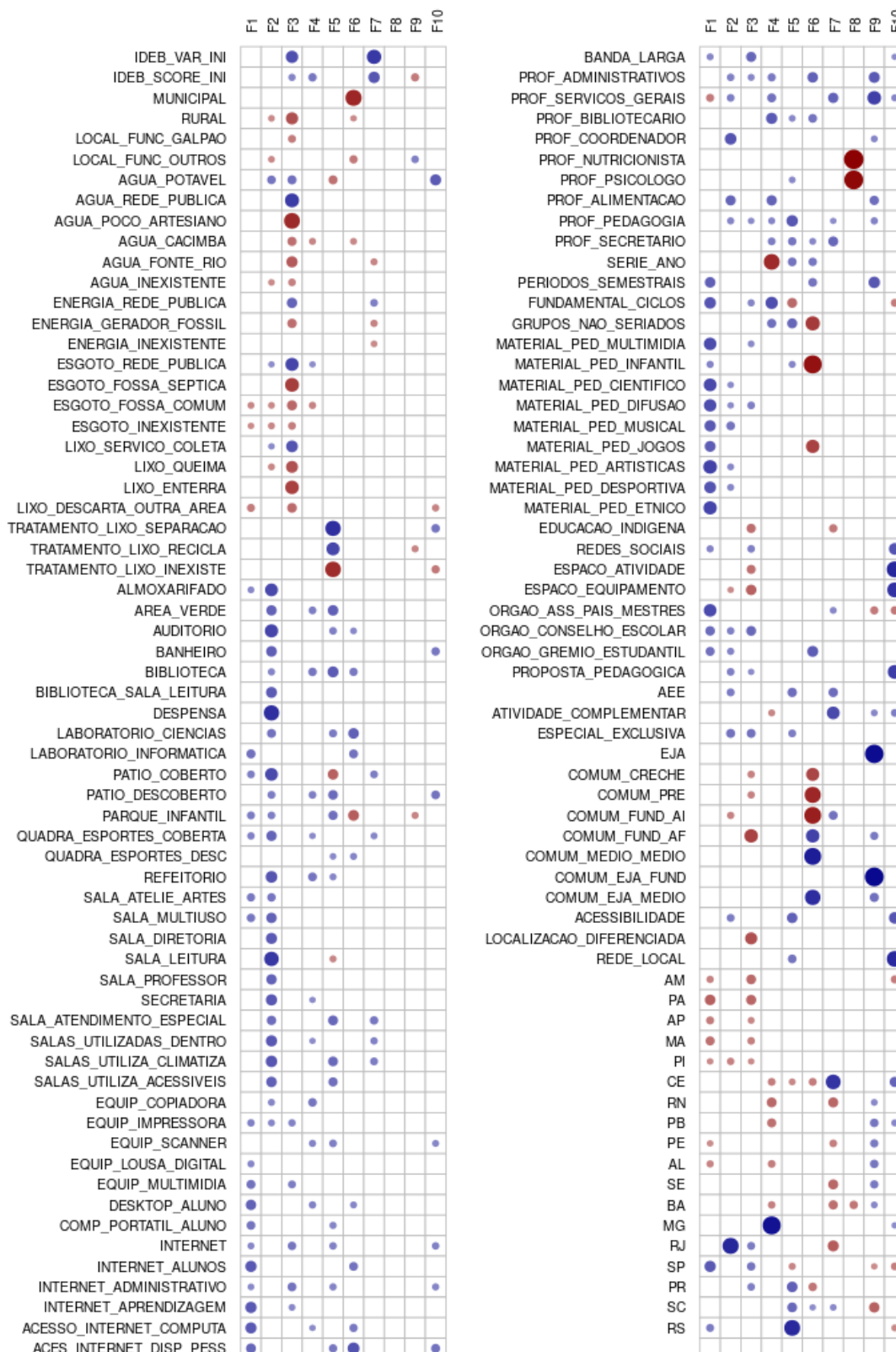
Há um contraste interessante a ser feito entre as figuras 14-B e 15 com relação ao perfil de decaimento das variâncias explicadas, dado que essa variância está associada aos autovalores (ZAKI & MEIRA JR., 2020). Essa diferença de perfis deve-se à rotação, que não foi feita na Figura 14-B pela função `fa.parallel`; mas houve rotação “oblimin” operada na FA recursiva de Storopoli (2019). Ainda que a rotação não interfira na variância total acumulada, ela altera a variância por fator, uma vez que procura direções conforme a densidade de atributos, fazendo com que a carga (*loading*) de cada atributo tenha mais presença nos fatores. Como consequência, cada fator (isoladamente) pode vir a somar uma maior variância explicada, o que auxilia na sua interpretabilidade (RAM, 2022).

A projeção dos 127 atributos com *communalities* acima de 50% (inclusive) pode ser vista na Figura 16. Essa visualização no contexto FA é uma das inovações desta dissertação e permite uma visão holística dessas projeções num espaço multidimensional (denominado provisoriamente de “RnVis”, abreviação de *R<sup>n</sup> Visualization*). Neste RnVis, encontra-se representada a matriz de carga padrão (*pattern loading*) entre atributos e fatores. Quanto maior a projeção de um vetor atributo sobre um vetor fator, maior o seu coeficiente em módulo, e maior a área do círculo. Para efeito de melhor visualização, os coeficientes foram escalados pela raiz quadrada de seus valores absolutos, recuperando o sinal em seguida. E com intuito de despoluir a Figura e destacar os coeficientes mais representativos, são exibidos somente os coeficientes com valor absoluto maior que 0.20. Se a projeção ocorrer no lado positivo do fator, o círculo estará destacado em azul; se sobre o lado negativo, em vermelho. Logo, atributos com mesma cor num mesmo fator, guardam variâncias compartilhadas (correlações), com intensidades relacionadas ao tamanho do círculo.

Uma das principais vantagens da metodologia FA conjuntamente com o RnVis é que, diferente do correlograma atrelado ao dendrograma da Figura 13b, ela permite uma visualização não hierárquica, multidimensional, oferecendo múltiplas combinações ou possibilidades de análise. Como num caleidoscópio, diferentes padrões emergem conforme orientações do olhar seguindo linhas (atributos) ou colunas (fatores).

Nesse sentido, eis alguns pontos a destacar da análise da Figura 16. Primeiro, uma característica típica da FA: seu potencial de interpretabilidade. Cada fator pode ser considerado um atributo latente ou um constructo, como chama a literatura (WILLIAMS *et al.*, 2010), que entrelaça numa semântica convincente vários atributos originais correlacionados.

Por exemplo, olhando os dois primeiros fatores (F1 e F2), percebe-se que eles capturaram diferentes padrões de atributos. O fator F1 entrelaçou atributos relacionados à infraestrutura computacional, incluindo LABORATÓRIO DE INFORMÁTICA, DESKTOP ALUNO, várias modalidades de INTERNET; mas também, diferentes tipos de MATERIAL PEDAGÓGICO (com os maiores pesos em F1); além da presença de ÓRGÃOS entre PAIS E MESTRES, CONSELHO ESCOLAR e GRÊMIO ESTUDANTIL. Curiosamente, foi capaz de atribuir correlações positivas desses atributos com SP e RS (em maior grau o primeiro) e anticorrelações com estados do norte e nordeste, especialmente PA e MA. As anticorrelações também envolviam atributos como ESGOTO INEXISTENTE ou FOSSA COMUM.



**Figura 16:** Um “RnVis”, com indicação visual dos coeficientes (pesos) da matriz padrão (*pattern matrix*) da decomposição FA da matriz Sigma de INI\_FIL, evidenciando as relações entre atributos (linhas) e fatores (colunas). Cores indicam os valores dos coeficientes da matriz normalizados no intervalo  $[-1, 1]$ , sendo azul positivos e vermelho negativos. Logo, se a projeção ocorrer no lado positivo do fator estará destacado em azul; se sobre o lado negativo, em vermelho. Foram exibidos apenas os coeficientes maiores que módulo de 0.20. O tamanho dos círculos é proporcional à raiz quadrada do valor absoluto dos coeficientes (recuperando o sinal, em seguida), apenas para efeito de melhor visualização. Produzido pelo FA recursivo de Storopoli (2019), com parâmetros “fm = minres”, “rotate = oblimin”, “n.obs = 58920”.

Já o fator F2 capturou escolas com padrões diferentes do fator F1, embora houvesse algumas interseções. No fator F2 prevalece a infraestrutura construtiva ou arquitetônica, em termos de ALMOXARIFADO, PÁTIO E QUADRA COBERTAS, ÁREA VERDE, AUDITÓRIO, BANHEIRO, DESPENSA; e diversos tipos de SALAS: MULTIUSO, de LEITURA, da DIRETORIA, dos PROFESSORES, de ATENDIMENTO ESPECIAL, UTILIZADAS DENTRO [das edificações], CLIMATIZADAS, ACESSÍVEIS. Incluem também serviços para alunos ESPECIAIS EXCLUSIVA, PROFISSIONAIS DE COORDENAÇÃO e ALIMENTAÇÃO. Com respeito às unidades federativas, há forte correlação com o RJ (uma carga de 0.718, vide Tabela 16), e anticorrelação com o PI. Outras anticorrelações típicas de F2 incluem escolas RURAIS, ÁGUA INEXISTENTE, LIXO QUEIMA. Naquilo que é interseção entre F1 e F2, pode-se incluir as anticorrelações com ESGOTO INEXISTENTE e FOSSA COMUM, e diversas correlações positivas, como alguns MATERIAIS PEDAGÓGICOS e ÓRGÃOS (CONSELHO ESCOLAR E GRÊMIO). Mas, os pesos dos coeficientes desses atributos comuns tenderam a ser diferentes, privilegiando um ou outro fator.

A Tabela 16 resume os parágrafos anteriores, apresentando uma visão mais detalhada dos atributos e respectivos coeficientes envolvendo os fatores F1 e F2. Revela apenas os atributos cujos fatores têm carga (em módulo) maior que 0.2, com destaque em cores para os com (módulo) maiores que 0.3, sendo azul positivos, vermelho negativos.

Como dito anteriormente, esperava-se que padrões emergissem dos dados de forma espontânea e não supervisionada, com mínimo de indução ou viés, especialmente eventuais correlações entre atributos das escolas com os índices Ideb. Os dois primeiros fatores F1 e F2 mostraram relações que não dependiam desses índices. Mas, algo diferente aconteceu envolvendo os fatores F3, F4, F7 e F9, conforme pode ser visto na Figura 16 e na Tabela 17. Todos esses fatores capturaram, ainda que em graus diferentes, associações com os atributos IDEB\_VAR\_INI (variação do Ideb de 2005 a 2019) e/ou IDEB\_SCORE\_INI (a nota do Ideb em 2019).



Tabela 16 - Análise dos fatores F1 e F2

| ATRIBUTOS                 | F1     | ATRIBUTOS                 | F2     |
|---------------------------|--------|---------------------------|--------|
| ESGOTO_FOSSA_COMUM        | -0,218 | ESGOTO_FOSSA_COMUM        | -0,214 |
| ESGOTO_INEXISTENTE        | -0,205 | ESGOTO_INEXISTENTE        | -0,220 |
| LIXO_DESCARTA_OUTRA_AREA  | -0,263 |                           |        |
| PROF_SERVICOS_GERAIS      | -0,264 | PROF_SERVICOS_GERAIS      | 0,249  |
| AM                        | -0,233 |                           |        |
| PA                        | -0,384 |                           |        |
| AP                        | -0,263 |                           |        |
| MA                        | -0,311 |                           |        |
| PI                        | -0,206 | PI                        | -0,238 |
| PE                        | -0,205 |                           |        |
| AL                        | -0,230 |                           |        |
|                           |        | RURAL                     | -0,211 |
|                           |        | LOCAL_FUNC_OUTROS         | -0,214 |
|                           |        | AGUA_INEXISTENTE          | -0,208 |
|                           |        | LIXO_QUEIMA               | -0,219 |
|                           |        | ESPACO_EQUIPAMENTO        | -0,202 |
|                           |        | COMUM_FUND_AI             | -0,232 |
| ALMOXARIFADO              | 0,213  | ALMOXARIFADO              | 0,507  |
| LABORATORIO_INFORMATICA   | 0,310  |                           |        |
| PATIO_COBERTO             | 0,253  | PATIO_COBERTO             | 0,503  |
| PARQUE_INFANTIL           | 0,261  | PARQUE_INFANTIL           | 0,250  |
| QUADRA_ESPORTE_COBERTA    | 0,240  | QUADRA_ESPORTE_COBERTA    | 0,368  |
| SALA_ATELIE_ARTES         | 0,273  | SALA_ATELIE_ARTES         | 0,283  |
| SALA_MULTIIUSO            | 0,280  | SALA_MULTIIUSO            | 0,374  |
| EQUIP_IMPRESSORA          | 0,240  | EQUIP_IMPRESSORA          | 0,221  |
| EQUIP_LOUSA_DIGITAL       | 0,222  |                           |        |
| EQUIP_MULTIMIDIA          | 0,305  |                           |        |
| DESKTOP_ALUNO             | 0,374  |                           |        |
| COMP_PORTATIL_ALUNO       | 0,295  |                           |        |
| INTERNET                  | 0,217  |                           |        |
| INTERNET_ALUNOS           | 0,419  |                           |        |
| INTERNET_ADMINISTRATIVO   | 0,210  |                           |        |
| INTERNET_APRENDIZAGEM     | 0,421  |                           |        |
| ACESSO_INTERNET_COMPUTA   | 0,414  |                           |        |
| ACES_INTERNET_DISP_PESS   | 0,350  |                           |        |
| BANDA_LARGA               | 0,220  |                           |        |
| PERIODO_SEMESTRAIS        | 0,384  |                           |        |
| FUNDAMENTAL_CICLOS        | 0,435  |                           |        |
| MATERIAL_PED_MULTIMIDIA   | 0,491  |                           |        |
| MATERIAL_PED_INFANTIL     | 0,229  |                           |        |
| MATERIAL_PED_CIENTIFICO   | 0,507  | MATERIAL_PED_CIENTIFICO   | 0,205  |
| MATERIAL_PED_DIFUSAO      | 0,485  | MATERIAL_PED_DIFUSAO      | 0,206  |
| MATERIAL_PED_MUSICAL      | 0,420  | MATERIAL_PED_MUSICAL      | 0,293  |
| MATERIAL_PED_JOGOS        | 0,399  |                           |        |
| MATERIAL_PED_ARTISITICAS  | 0,553  | MATERIAL_PED_ARTISITICAS  | 0,215  |
| MATERIAL_PED_DEPORTIVA    | 0,448  | MATERIAL_PED_DEPORTIVA    | 0,220  |
| MATERIAL_PED_ETNICO       | 0,532  |                           |        |
| REDES_SOCIAIS             | 0,229  |                           |        |
| ORGAO_ASS_PAIS_MESTRES    | 0,499  |                           |        |
| ORGAO_CONSELHO_ESCOLAR    | 0,332  | ORGAO_CONSELHO_ESCOLAR    | 0,240  |
| ORGAO_CONSELHO_ESTUDANTIL | 0,308  | ORGAO_CONSELHO_ESTUDANTIL | 0,229  |
| SP                        | 0,419  |                           |        |
| RS                        | 0,262  |                           |        |
|                           |        | AGUA_POTAVEL              | 0,289  |
|                           |        | ESGOTO_REDE_PUBLICA       | 0,201  |
|                           |        | LIXO_SERVICO_COLETA       | 0,206  |
|                           |        | AREA_VERDE                | 0,371  |
|                           |        | AUDITORIO                 | 0,531  |
|                           |        | BANHEIRO                  | 0,388  |
|                           |        | BIBLIOTECA                | 0,230  |
|                           |        | BIBLIOTECA_SALA_LEITURA   | 0,404  |
|                           |        | DESPENSA                  | 0,662  |
|                           |        | LABORATORIO_CIENCIAS      | 0,303  |
|                           |        | PATIO_DESCOBERTO          | 0,258  |
|                           |        | REFEITORIO                | 0,443  |
|                           |        | SALA_DIRETORA             | 0,413  |
|                           |        | SALA_LEITURA              | 0,607  |
|                           |        | SALA_PROFESSOR            | 0,373  |
|                           |        | SECRETARIA                | 0,419  |
|                           |        | SALA_ATENDIMENTO_ESPECIAL | 0,324  |
|                           |        | SALAS_UTILIZADAS_DENTRO   | 0,419  |
|                           |        | SALAS_UTILIZA_CLIMATIZA   | 0,434  |
|                           |        | SALAS_UTILIZA_ACESSIVEIS  | 0,382  |
|                           |        | EQUIP_COPIADORA           | 0,223  |
|                           |        | PROF_ADMINISTRATIVO       | 0,243  |
|                           |        | PROF_COORDENADOR          | 0,440  |
|                           |        | PROF_ALIEMNTACAO          | 0,362  |
|                           |        | PROF_PEDAGOGIA            | 0,226  |
|                           |        | PROPOSTA_PEDAGOGICA       | 0,250  |
|                           |        | AEE                       | 0,260  |
|                           |        | ESPECIAL_EXCLUSIVA        | 0,306  |
|                           |        | ACESSIBILIDADE            | 0,258  |
|                           |        | RJ                        | 0,718  |

Fonte: Elaborado pelo Autor (2022)

Nota: Análise dos dois primeiros fatores F1 e F2, com pesos maiores que módulo de 0.2. Em destaque, os maiores coeficientes em azul (positivos) e vermelho (negativos), conforme pesos maiores que módulo de 0.3.

O fator F3, que pegou tanto VARIACÃO do Ideb quanto NOTA 2019 do Ideb, ainda que com mais peso o primeiro (0.475) que o segundo (0.231), agregou de forma positiva principalmente os atributos relacionados: aos serviços públicos de ÁGUA (0.580), ENERGIA, ESGOTO (0.525) e LIXO; BANDA LARGA; e presença de CONSELHO ESCOLAR; e negativamente, quase o oposto: ausência dos serviços públicos relatados acima, principalmente POÇO ARTESIANO (-0.698) e FOSSA SÉPTICA (-0.569); RURAIS (-0.471), o que indica que F3 tende a acomodar escolas URBANAS no lado positivo do seu eixo; ao contrário das EDUCAÇÃO INDÍGENA; ESPAÇOS DE ATIVIDADE e EQUIPAMENTOS; ter em COMUM ENSINO FUNDAMENTAL ANOS FINAIS (-0.544); e LOCALIZAÇÃO DIFERENCIADA. Com relação aos estados, correlação positiva: RJ, SP e PR; negativa: AM e PA, e em menor grau AP, MA, e PI.

O fator F4 contemplou com mais evidência apenas NOTA 2019 do Ideb, amarrando positivamente: FUNDAMENTAL CICLOS (0.475); PROFISSIONAIS SERVIÇOS GERAIS, BIBLIOTECÁRIO (0.417), ALIMENTAÇÃO; GRUPOS NÃO SERIADOS; no lado oposto: a modalidade tradicional SÉRIE/ANO (-0.695). Com relação aos estados, um grande peso para MG (0.863), em oposição aos estados de: RN, PB, e em menor grau: CE, AL e BA.

Já o fator F7, costurou ambos VARIACÃO do Ideb (0.600) e NOTA 2019 do Ideb (0.431), atrelando de forma positiva: ATIVIDADE COMPLEMENTAR (0.505); PROFISSIONAIS SERVIÇOS GERAIS, SECRETÁRIO; ATENDIMENTO EDUCACIONAL ESPECIALIZADO (AEE). O principal estado representado positivamente por este fator é CE (0.628), seguido de SC, em oposição a: RJ (-0.408), RN, BA, SE, e em menor grau PE.

Tabela 17 - Análise dos fatores F3, F4, F7 e F9

| ATRIBUTO                 | F3     | ATRIBUTO                | F4     | ATRIBUTO                  | F7     | ATRIBUTO                | F9     |
|--------------------------|--------|-------------------------|--------|---------------------------|--------|-------------------------|--------|
| RURAL                    | -0,471 |                         |        |                           |        |                         |        |
| LOCAL_FUNC_GALPAO        | -0,260 |                         |        |                           |        |                         |        |
| AGUA_POCO_ARTESIANO      | -0,698 |                         |        |                           |        |                         |        |
| AGUA_CACIMBA             | -0,314 | AGUA_CACIMBA            | -0,223 |                           |        |                         |        |
| AGUA_FONTE_RIO           | -0,414 |                         |        | AGUA_FONTE_RIO            | -0,224 |                         |        |
| AGUA_INEXISTENTE         | -0,246 |                         |        |                           |        |                         |        |
| ENERGIA_GERADOR_FOSSIL   | -0,324 |                         |        |                           |        |                         |        |
| ESGOTO_FOSSA_SEPTICA     | -0,569 |                         |        | ENERGIA_GERADOR_FOSSIL    | -0,233 |                         |        |
| ESGOTO_FOSSA_COMUM       | -0,357 | ESGOTO_FOSSA_COMUM      | -0,232 |                           |        |                         |        |
| ESGOTO_INEXISTENTE       | -0,253 |                         |        |                           |        |                         |        |
| LIXO_QUEIMA              | -0,457 |                         |        |                           |        |                         |        |
| LIXO_ENTERRA             | -0,550 |                         |        |                           |        |                         |        |
| LIXO_DESCARTA_OUTRA_AREA | -0,337 |                         |        |                           |        |                         |        |
| EDUCACAO_INDIGENA        | -0,317 |                         |        | EDUCACAO_INDIGENA         | -0,277 |                         |        |
| ESPACO_ATIVIDADE         | -0,308 |                         |        |                           |        |                         |        |
| ESPACO_EQUIPAMENTO       | -0,378 |                         |        |                           |        |                         |        |
| COMUM_CRECHE             | -0,239 |                         |        |                           |        |                         |        |
| COMUM_PRE                | -0,237 |                         |        |                           |        |                         |        |
| COMUM_FUND_AF            | -0,544 |                         |        |                           |        | COMUM_FUND_AF           | 0,271  |
| LOCALIZACAO_DIFERENCIADA | -0,467 |                         |        |                           |        |                         |        |
| AM                       | -0,341 |                         |        |                           |        |                         |        |
| PA                       | -0,358 |                         |        |                           |        |                         |        |
| AP                       | -0,217 |                         |        |                           |        |                         |        |
| MA                       | -0,245 |                         |        |                           |        |                         |        |
| PI                       | -0,203 |                         |        |                           |        |                         |        |
| IDEB_SCORE_INI           | 0,231  | IDEB_SCORE_INI          | 0,278  | IDEB_SCORE_INI            | 0,431  | IDEB_SCORE_INI          | -0,271 |
| RJ                       | 0,262  |                         |        | RJ                        |        |                         |        |
| SP                       | 0,285  |                         |        |                           |        | SP                      | -0,202 |
|                          |        | SERIE_ANO               | -0,695 |                           |        |                         |        |
|                          |        | ATIVIDADE_COMPLEMENTAR  | -0,220 | ATIVIDADE_COMPLEMENTAR    | 0,505  | ATIVIDADE_COMPLEMENTAR  | 0,215  |
|                          |        | CE                      | -0,254 | CE                        | 0,628  |                         |        |
|                          |        | RN                      | -0,347 | RN                        | -0,356 | RN                      | 0,213  |
|                          |        | PB                      | -0,310 |                           |        | PB                      | 0,287  |
|                          |        | AL                      | -0,253 |                           |        | AL                      | 0,284  |
|                          |        | BA                      | -0,243 | BA                        | -0,318 | BA                      | 0,214  |
|                          |        |                         |        | ENERGIA_INEXISTENTE       | -0,209 |                         |        |
|                          |        |                         |        | PE                        | -0,247 | PE                      | 0,270  |
|                          |        |                         |        | SE                        | -0,354 | SE                      | 0,274  |
|                          |        |                         |        | ORGAO_ASS_PAIS_MESTRES    | 0,216  | ORGAO_ASS_PAIS_MESTRES  | -0,269 |
|                          |        |                         |        | SC                        | 0,209  | SC                      | -0,374 |
|                          |        |                         |        |                           |        | TRATAMENTO_LIXO_RECICLA | -0,225 |
|                          |        |                         |        |                           |        | PARQUE_INFANTIL         | -0,227 |
| IDEB_VAR_INI             | 0,475  |                         |        | IDEB_VAR_INI              | 0,600  |                         |        |
| AGUA_POTAVEL             | 0,308  |                         |        |                           |        |                         |        |
| AGUA_REDE_PUBLICA        | 0,580  |                         |        |                           |        |                         |        |
| ENERGIA_REDE_PUBLICA     | 0,369  |                         |        | ENERGIA_REDE_PUBLICA      | 0,249  |                         |        |
| ESGOTO_REDE_PUBLICA      | 0,525  | ESGOTO_REDE_PUBLICA     | 0,201  |                           |        |                         |        |
| LIXO_SERVICO_COLETA      | 0,445  |                         |        |                           |        |                         |        |
| EQUIP_IMPRESSORA         | 0,236  |                         |        |                           |        |                         |        |
| EQUIP_MULTIMIDIA         | 0,257  |                         |        |                           |        |                         |        |
| INTERNET                 | 0,282  |                         |        |                           |        |                         |        |
| INTERNET_ADMINISTRATIVO  | 0,299  |                         |        |                           |        |                         |        |
| INTERNET_APRENDIZAGEM    | 0,214  |                         |        |                           |        |                         |        |
| BANDA_LARGA              | 0,357  |                         |        |                           |        |                         |        |
| PROF_ADMINISTRATIVO      | 0,220  | PROF_ADMINISTRATIVO     | 0,269  | PROF_ADMINISTRATIVO       | 0,408  | PROF_ADMINISTRATIVO     | 0,408  |
| PROF_PEDAGOGIA           | 0,206  | PROF_PEDAGOGIA          | 0,221  | PROF_PEDAGOGIA            | 0,200  | PROF_PEDAGOGIA          | 0,232  |
| FUNDAMENTAL_CICLOS       | 0,227  | FUNDAMENTAL_CICLOS      | 0,475  |                           |        |                         |        |
| MATERIAL_PED_MULTIMIDIA  | 0,211  |                         |        |                           |        |                         |        |
| MATERIAL_PED_DIFUSAO     | 0,249  |                         |        |                           |        |                         |        |
| REDES_SOCIAIS            | 0,246  |                         |        |                           |        |                         |        |
| ORGAO_CONSELHO_ESCOLAR   | 0,346  |                         |        |                           |        |                         |        |
| PROPOSTA_PEDAGOGICA      | 0,200  |                         |        |                           |        |                         |        |
| ESPECIAL_EXCLUSIVA       | 0,304  |                         |        |                           |        |                         |        |
| PR                       | 0,249  |                         |        |                           |        |                         |        |
|                          |        | AREA_VERDE              | 0,253  |                           |        |                         |        |
|                          |        | BIBLIOTECA              | 0,284  |                           |        |                         |        |
|                          |        | PATIO_DESCOBERTO        | 0,250  |                           |        |                         |        |
|                          |        | QUADRA_ESPORTE_COBERTA  | 0,205  | QUADRA_ESPORTE_COBERTA    | 0,217  |                         |        |
|                          |        | REFEITORIO              | 0,297  |                           |        |                         |        |
|                          |        | SECRETARIA              | 0,202  |                           |        |                         |        |
|                          |        | SALAS_UTILIZADAS_DENTRO | 0,200  | SALAS_UTILIZADAS_DENTRO   | 0,236  |                         |        |
|                          |        | EQUIP_COPIADORA         | 0,288  |                           |        |                         |        |
|                          |        | EQUIP_SCANNER           | 0,228  |                           |        |                         |        |
|                          |        | DESKTOP_ALUNO           | 0,236  |                           |        |                         |        |
|                          |        | ACESSO_INTERNET_COMPUTA | 0,202  |                           |        |                         |        |
|                          |        | PROF_SERVICOS_GERAIS    | 0,304  | PROF_SERVICOS_GERAIS      | 0,370  | PROF_SERVICOS_GERAIS    | 0,554  |
|                          |        | PROF_BIBLIOTECARIO      | 0,417  |                           |        |                         |        |
|                          |        | PROF_ALIMENTACAO        | 0,368  |                           |        | PROF_ALIMENTACAO        | 0,322  |
|                          |        | PROF_SECRETARIO         | 0,254  | PROF_SECRETARIO           | 0,350  |                         |        |
|                          |        | GRUPOS_NAO_SERIADO      | 0,311  |                           |        |                         |        |
|                          |        | MG                      | 0,863  |                           |        |                         |        |
|                          |        |                         |        | PATIO_COBERTO             | 0,253  |                         |        |
|                          |        |                         |        | SALA_ATENDIMENTO_ESPECIAL | 0,280  |                         |        |
|                          |        |                         |        | SALAS_UTILIZA_CLIMATIZA   | 0,257  |                         |        |
|                          |        |                         |        | TP_AEE                    | 0,327  |                         |        |
|                          |        |                         |        | COMUM_FUND_AI             | 0,305  |                         |        |
|                          |        |                         |        |                           |        | LOCAL_FUNC_OUTROS       | 0,246  |
|                          |        |                         |        |                           |        | PROF_COORDENADOR        | 0,219  |
|                          |        |                         |        |                           |        | PERIODOS_SEMESTRAIS     | 0,431  |
|                          |        |                         |        |                           |        | EJA                     | 0,903  |
|                          |        |                         |        |                           |        | COMUM_EJA_FUND          | 0,927  |
|                          |        |                         |        |                           |        | COMUM_EJA_MEDIO         | 0,312  |

Fonte: Elaborado pelo Autor (2022)

Nota: Fatores F3, F4, F7 e F9 com pesos maiores que módulo de 0.2. Em destaque, os maiores coeficientes em azul (positivos) e vermelho (negativos), conforme pesos maiores que módulo de 0.3.

Por fim, o fator F9 tem a peculiaridade de uma NOTA 2019 do Ideb negativa (-0.271), e neutro com relação à VARIAÇÃO do Ideb (por isso, não aparece). Congregou em anticorrelação com a NOTA 2019 do Ideb, escolas tendendo a: EDUCAÇÃO DE JOVENS E ADULTOS (EJA), com altíssimo peso (0.903); há alguma presença de ATIVIDADE COMPLEMENTAR; em PERÍODOS SEMESTRAIS; tendem a não ter ASSOCIAÇÃO DE PAIS E MESTRES; NÃO TRATAM/RECICLAM LIXO; e, conseqüentemente, como escola de jovens e adultos, tendem a não possuir PARQUE INFANTIL. Parece uma obviedade, mas é bom lembrar que tratou-se de conclusão automática não supervisionada, e de certa forma, serve para atestar a confiabilidade da metodologia e dos resultados gerados. Entre os estados com correlação positiva com este fator, ficaram amarrados (em grau menor) escolas do nordeste (RN, PB, AL, BA, PE, SE); em oposição a SC e SP, com maior peso ao primeiro.

Cabe destacar como alguns estados se entrelaçaram com os atributos representados por esses fatores, considerando a hipótese de possíveis impactos positivos nos índices do Ideb. Lembrando que as técnicas utilizadas neste estudo não permitem inferir causalidades, apenas associações ou correlações.

SP, RJ e PR, que dominaram o F3 (ainda que com carga abaixo de 0.300), indicaram escolas com boa infraestrutura interna e serviços públicos; Internet de qualidade, utilizada tanto administrativamente quanto em processos de ensino-aprendizagem. Evidencia-se também alguns tipos de materiais didáticos; existência de proposta pedagógica; conselho escolar; atendimento especial exclusivo. Tais conjuntos de fatores, para esses estados, parecem ter tido mais correlação com o crescimento do Ideb do que com a nota final do Ideb em 2019.

Já MG, que tomou conta do F4 com alta carga, a despeito de: boa infraestrutura construtiva e arquitetônica; computadores e acesso à Internet; além de saneamento básico de esgoto; destacou-se mais por um ensino fundamental inicial em ciclos e grupos não seriados (reforça esse destaque a anticorrelação com o atributo SÉRIE/ANO). Curiosamente, ambos são formas de flexibilização da seriação anual mais tradicional. O primeiro possibilita que os currículos sejam trabalhados em tempos mais elásticos, além dos limites anuais (BRASIL, 2018a). O segundo permite a organização de turmas com alunos/alunas em diferentes fases de aprendizagem e/ou desenvolvimento psicomotor e/ou cognitivo, independente de ano/série (BRASIL, 2018a). Importante dizer que SP, RJ e PR também acusaram associação (no F3, em menor grau) com fundamental em ciclos. No entanto, para MG, tais fatores guardam certa associação com a nota final do Ideb de 2019, não sendo revelado associação à variação do Ideb.

CE e em menor grau SC são os estados positivamente ressaltados no F7, sendo este fator o que indicou maior peso tanto na variação quanto nota final do Ideb em 2019. Diferente dos fatores anteriores, no F7 não há destaque para infraestrutura computacional e Internet, nem presença estatisticamente relevante de algum material pedagógico. Em termos construtivos, há indicação de pátio e quadras cobertas, e salas climatizadas. O ponto forte neste fator fica mesmo por conta das atividades complementares, atendimento especial especializado e (em menor grau) associações de pais e mestres. Curiosamente, esses atributos juntamente com o CE mostraram-se em anticorrelação no fator F4 associado à MG, sugerindo que esses estados podem estar em contraposição um ao outro sob certos aspectos. Por outro lado, RJ tem destacada anticorrelação com o F7.

Ante o exposto, entre muitas reflexões possíveis, pode-se perguntar se os estados mencionados não teriam conquistado bom desempenho nos índices do Ideb adotando diferentes infraestruturas, políticas e/ou propostas educacionais. Não parece haver indicação de uma estratégia unificada<sup>7</sup>, mas pelo menos 3 grupos de padrões: capitaneadas por SP, num 1º grupo; MG num 2º grupo; CE num 3º grupo. Se boa infraestrutura interna; serviços públicos de água, energia, esgoto e lixo; internet de qualidade e utilizada no ensino-aprendizagem; escolas tipicamente urbanas (não rurais); caracterizaram o 1º grupo, esses atributos tiveram presença menor no 2º grupo, e quase nenhuma no 3º grupo. Se flexibilizar a seriação foi o que marcou o 2º grupo e em parte o 1º grupo, ela não encontra muita aderência ao 3º grupo, que é bem marcado pela adoção das atividades complementares e atendimento especializado. Em contraposição a essas soluções, escolas tipo EJA, principalmente no nordeste, mas não de SP e SC, tenderam a ter as menores notas no Ideb 2019.

---

<sup>7</sup> O único atributo comum a todos os fatores F3, F4, F7 e F9 foi QUANTIDADE DE PROFISSIONAIS DE PEDAGOGIA.

Tabela 18 - Análise dos fatores F5, F6, F8 e F10

| ATRIBUTO                  | F5     | ATRIBUTO                  | F6     | ATRIBUTO           | F8     | ATRIBUTO                  | F10    |
|---------------------------|--------|---------------------------|--------|--------------------|--------|---------------------------|--------|
| AGUA_POTAVEL              | -0.306 |                           |        |                    |        | AGUA_POTAVEL              | 0.410  |
| TRATAMENTO_LIXO_INEXISTE  | -0.686 |                           |        |                    |        | TRATAMENTO_LIXO_INEXISTE  | -0.275 |
| PATIO_COBERTO             | -0.380 |                           |        |                    |        |                           |        |
| SALA_LEITURA              | -0.223 |                           |        |                    |        |                           |        |
| FUNDAMENTAL_CICLOS        | -0.347 |                           |        |                    |        | FUNDAMENTAL_CICLOS        | -0.262 |
| CE                        | -0.215 | CE                        | -0.259 |                    |        | CE                        | 0.366  |
| SP                        | -0.225 |                           |        |                    |        | SP                        | -0.254 |
| PARQUE_INFANTIL           | 0.319  | PARQUE_INFANTIL           | -0.395 |                    |        |                           |        |
| PROF_PSIKOLOGO            | 0.212  |                           |        | PROF_PSIKOLOGO     | -0.934 |                           |        |
| GRUPOS_NAO_SERIADOS       | 0.353  | GRUPOS_NAO_SERIADOS       | -0.595 |                    |        |                           |        |
| MATERIAL_PED_INFANTIL     | 0.222  | MATERIAL_PED_INFANTIL     | -0.868 |                    |        |                           |        |
| PR                        | 0.386  | PR                        | -0.286 |                    |        |                           |        |
| RS                        | 0.703  |                           |        |                    |        | RS                        | -0.217 |
|                           |        | MUNICIPAL                 | -0.720 |                    |        |                           |        |
|                           |        | RURAL                     | -0.206 |                    |        |                           |        |
|                           |        | LOCAL_FUNC_OUTROS         | -0.273 |                    |        |                           |        |
|                           |        | AGUA_CACIMBA              | -0.217 |                    |        |                           |        |
|                           |        | MATERIAL_PED_JOGOS        | -0.539 |                    |        |                           |        |
|                           |        | COMUM_CRECHE              | -0.523 |                    |        |                           |        |
|                           |        | COMUM_PRE                 | -0.720 |                    |        |                           |        |
|                           |        | COMUM_FUND_AI             | -0.770 |                    |        |                           |        |
|                           |        |                           |        | PROF_NUTRICIONISTA | -0.982 |                           |        |
|                           |        |                           |        | BA                 | -0.281 |                           |        |
|                           |        |                           |        |                    |        | LIXO_DESCARTA_OUTRA_AREA  | -0.234 |
|                           |        |                           |        |                    |        | ORGAO_ASS_PAIS_MESTRES    | -0.276 |
|                           |        |                           |        |                    |        | AM                        | -0.264 |
|                           |        |                           |        |                    |        | TRATAMENTO_LIXO_SEPARACAO | 0.289  |
| TRATAMENTO_LIXO_SEPARACAO | 0.666  |                           |        |                    |        |                           |        |
| TRATAMENTO_LIXO_RECICLA   | 0.521  |                           |        |                    |        |                           |        |
| AREA_VERDE                | 0.382  |                           |        |                    |        |                           |        |
| AUDITORIO                 | 0.247  | AUDITORIO                 | 0.211  |                    |        |                           |        |
| BIBLIOTECA                | 0.404  | BIBLIOTECA                | 0.274  |                    |        |                           |        |
| LABORATORIO_CIENCIAS      | 0.265  | LABORATORIO_CIENCIAS      | 0.383  |                    |        |                           |        |
| PATIO_DESCOBERTO          | 0.336  |                           |        |                    |        | PATIO_DESCOBERTO          | 0.295  |
| QUADRA_ESPORTE_DESCOBERTO | 0.209  | QUADRA_ESPORTE_DESCOBERTO | 0.225  |                    |        |                           |        |
| REFEITORIO                | 0.221  |                           |        |                    |        |                           |        |
| SALA_ATENDIMENTO_ESPECIAL | 0.350  |                           |        |                    |        |                           |        |
| SALAS_UTILIZA_CLIMATIZA   | 0.343  |                           |        |                    |        |                           |        |
| SALAS_UTILIZA_ACESSIVEIS  | 0.316  |                           |        |                    |        |                           |        |
| EQUIP_SCANNER             | 0.249  |                           |        |                    |        | EQUIP_SCANNER             | 0.219  |
| COMP_PORTATIL_ALUNO       | 0.227  |                           |        |                    |        |                           |        |
| INTERNET                  | 0.241  |                           |        |                    |        | INTERNET                  | 0.239  |
| INTERNET_ADMINISTRATIVO   | 0.231  |                           |        |                    |        | INTERNET_ADMINISTRATIVO   | 0.227  |
| ACES_INTERNET_DISP_PESS   | 0.277  | ACES_INTERNET_DISP_PESS   | 0.451  |                    |        | ACES_INTERNET_DISP_PESS   | 0.306  |
| PROF_BIBLIOTECARIO        | 0.215  | PROF_BIBLIOTECARIO        | 0.294  |                    |        |                           |        |
| PROF_PEDAGOGIA            | 0.428  |                           |        |                    |        |                           |        |
| PROF_SECRETARIO           | 0.282  | PROF_SECRETARIO           | 0.216  |                    |        |                           |        |
| SERIE_ANO                 | 0.296  | SERIE_ANO                 | 0.280  |                    |        |                           |        |
| AEE                       | 0.323  |                           |        |                    |        |                           |        |
| ESPECIAL_EXCLUSIVA        | 0.257  |                           |        |                    |        |                           |        |
| ACESSIBILIDADE            | 0.378  |                           |        |                    |        | ACESSIBILIDADE            | 0.439  |
| REDE_LOCAL                | 0.292  |                           |        |                    |        | REDE_LOCAL                | 0.705  |
| SC                        | 0.343  | SC                        | 0.200  |                    |        |                           |        |
|                           |        | LABORATORIO_INFORMATICA   | 0.303  |                    |        |                           |        |
|                           |        | DESKTOP_ALUNO             | 0.218  |                    |        |                           |        |
|                           |        | INTERNET_ALUNOS           | 0.298  |                    |        |                           |        |
|                           |        | ACESSO_INTERNET_COMPUT    | 0.274  |                    |        |                           |        |
|                           |        | PROF_ADMINISTRATIVO       | 0.381  |                    |        |                           |        |
|                           |        | PERIODOS_SEMESTRARIAS     | 0.296  |                    |        |                           |        |
|                           |        | ORGAO_GREMIO_ESTUDANTIL   | 0.402  |                    |        |                           |        |
|                           |        | COMUM_FUND_AF             | 0.543  |                    |        |                           |        |
|                           |        | COMUM_MEDIO_MEDIO         | 0.780  |                    |        |                           |        |
|                           |        | COMUM_EJA_MEDIO           | 0.673  |                    |        |                           |        |
|                           |        |                           |        |                    |        | BANHEIRO                  | 0.283  |
|                           |        |                           |        |                    |        | BANDA_LARGA               | 0.203  |
|                           |        |                           |        |                    |        | PROF_SERVICOS_GERAIS      | 0.223  |
|                           |        |                           |        |                    |        | REDES_SOCIAIS             | 0.447  |
|                           |        |                           |        |                    |        | ESPACO_ATIVIDADE          | 0.717  |
|                           |        |                           |        |                    |        | ESPACO_EQUIPAMENTO        | 0.644  |
|                           |        |                           |        |                    |        | PROPOSTA_PEDAGOGICA       | 0.577  |
|                           |        |                           |        |                    |        | ATIVIDADE_COMPLEMENTAR    | 0.252  |
|                           |        |                           |        |                    |        | PB                        | 0.222  |
|                           |        |                           |        |                    |        | MG                        | 0.204  |

**Fonte:** Elaborado pelo Autor (2022)

**Nota:** Fatores F5, F6, F8 e F10 com pesos maiores que módulo de 0.2. Em destaque, os maiores coeficientes em azul (positivos) e vermelho (negativos), conforme pesos maiores que módulo de 0.3.

A Tabela 18 mostra os demais fatores ainda não analisados: F5, F6, F8 e F10. Como F1 e F2, nenhum deles evidenciou correlação com os índices do Ideb. O fator F5 enlaçou escolas do sul: RS, SC, PR (com maior peso ao primeiro, 0.703), mas não de SP e CE. O grande destaque está na SEPARAÇÃO DE LIXO (0.666) e RECICLAGEM DE LIXO (peso 0.521). No fator F6 predominam escolas do ENSINO MÉDIO (0.780) e MÉDIO COM EJA (0.673), tendo em comum ENSINO FUNDAMENTAL ANOS FINAIS (0.543). Em menor

grau, tenderam a ser de SC, mas não do PR e CE. De forma coerente, nota-se a contraposição ao ENSINO INFANTIL, principalmente MATERIAL PEDAGÓGICO INFANTIL (-0.868). Outro ponto de coerência é a forte anticorrelação com escolas MUNICIPAIS (-0.720), indicando a esperada tendência de ENSINO MÉDIO em escolas estaduais<sup>8</sup>. O fator F8 ficou curiosamente dominado por apenas dois atributos com alto peso: PROFISSIONAIS PSICÓLOGOS (-0.934) e NUTRICIONISTAS (-0.982), e tenderam em menor grau a serem de escolas da BA. O fato desses dois atributos aparecerem isolados num único fator sugere que não guardam correlação relevante com nenhum outro fator, a não ser entre eles mesmos. Enfim, o fator 10 reuniu com maior peso escolas que compartilhavam ESPAÇOS internos para ATIVIDADES comunitárias (0.717) ou uso de ESPAÇOS e EQUIPAMENTOS (0.644) externos em atividades de ensino-aprendizagem; REDES SOCIAIS e presença de infraestrutura de REDE de dados LOCAL (0.705); e PROPOSTA PEDAGÓGICA (0.577). Em maior grau eram do CE; em menor, de MG e PB, mas não de SP e RS.

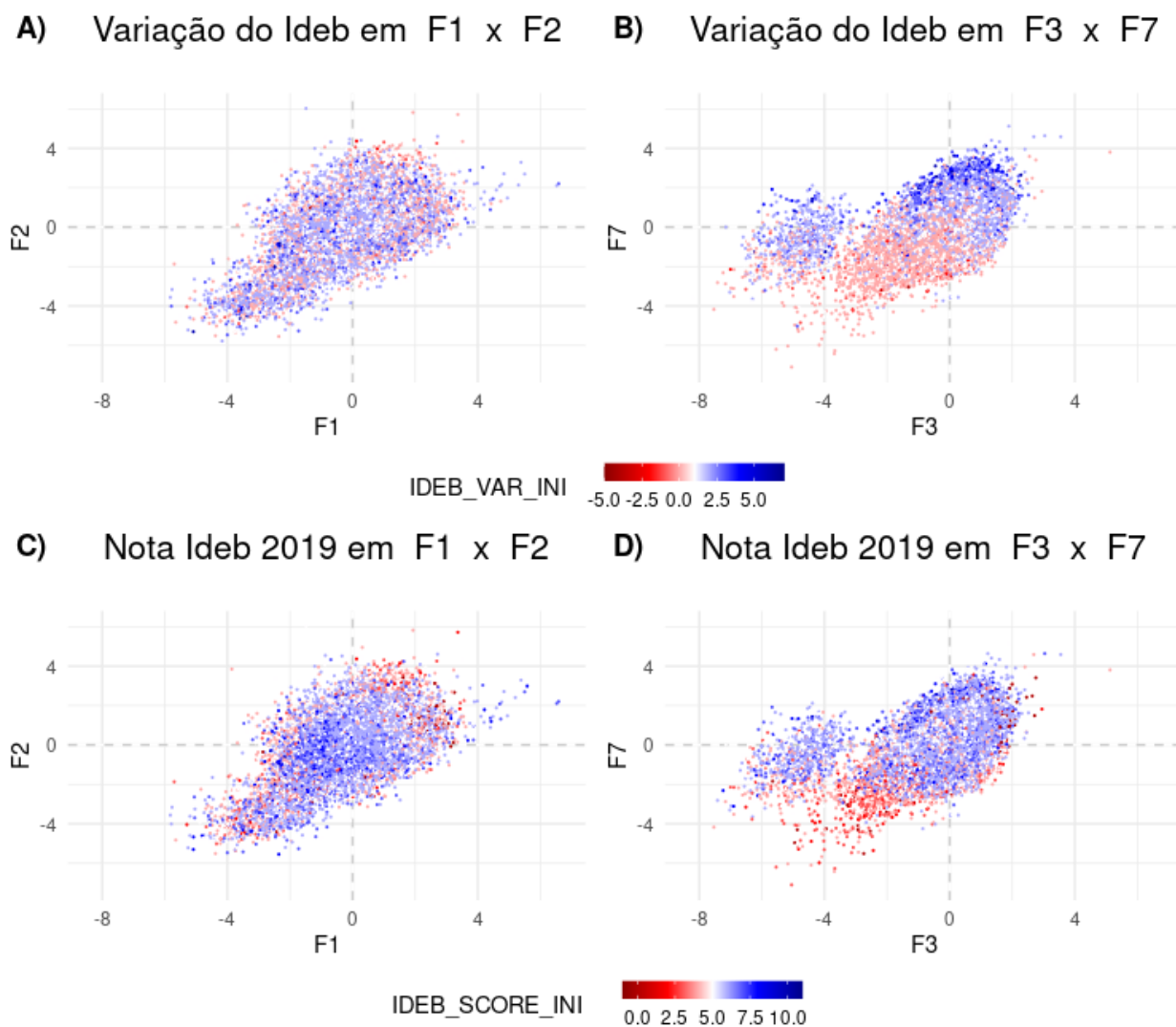
Até aqui pôde-se contemplar o espaço dos atributos e fatores, realçando os padrões mais relevantes das associações entre atributos. Uma outra forma de olhar é o espaço dos observáveis e fatores, ou seja, de como as quase 60 mil escolas ficam projetadas no espaço dos fatores. Esse outro olhar permite dar destaque a novos padrões, além de complementar e expandir os já observados.

A Figura 17 revela a distribuição das 58920 escolas no espaço dos fatores, considerando os pares F1 x F2 e F3 x F7. Cada ponto é uma escola, sendo as marcadas como azuis com índices Ideb acima da mediana, e vermelho abaixo. Pelas Figura 16 e Tabela 16, sabe-se que os índices Ideb não se atrelaram significativamente aos fatores F1 e F2. De fato, na Figura 17-A e 17-C percebe-se uma distribuição de pontos azuis e vermelhos que parece aleatória e sem preferência por direções. O que já não é o caso dos pares F3 x F7, pois há um nítido viés: quanto mais uma escola se posiciona nos valores mais altos dos eixos X e Y do 1º quadrante, maiores as chances dela apresentar índices Ideb acima da média, tanto para a variação da nota entre 2005 e 2019, quanto a nota final em 2019 (Figura 17-B e 17-D). Essa tendência é mais evidente para a variação do Ideb, pois os pesos de IDEB\_VAR\_INI em F3 e F7 (0.475 e 0.600, respectivamente) são maiores que em IDEB\_SCORE\_INI (0.231 e 0.431, respectivamente). Isso porque quanto maiores estes pesos, melhor a caracterização visual da assimetria. Percebe-se, pois, coerência entre as projeções dos observáveis no espaço dos fatores com as já vistas projeções dos atributos no espaço dos fatores

---

<sup>8</sup> O artigo 211 da constituição estabelece a prioridade dos municípios atuarem no ensino fundamental e educação infantil, e dos estados atuarem no ensino fundamental e médio (BRASIL, 2016).

Na verdade, essa assimetria em F3 e F7 pode ter uma interpretação ainda mais sutil. Considerando apenas esses fatores, já revela a possibilidade de três formas de uma escola ter chances de possuir índices Ideb mais altos: apresentar um F3 alto e um F7 baixo; apresentar um F3 baixo, mas F7 alto; e apresentar ambos F3 e F7 altos. E como cada fator tende a amarrar um conjunto de atributos diferentes, pode-se inferir a existência de diferentes padrões de escolas com bom desempenho no Ideb.

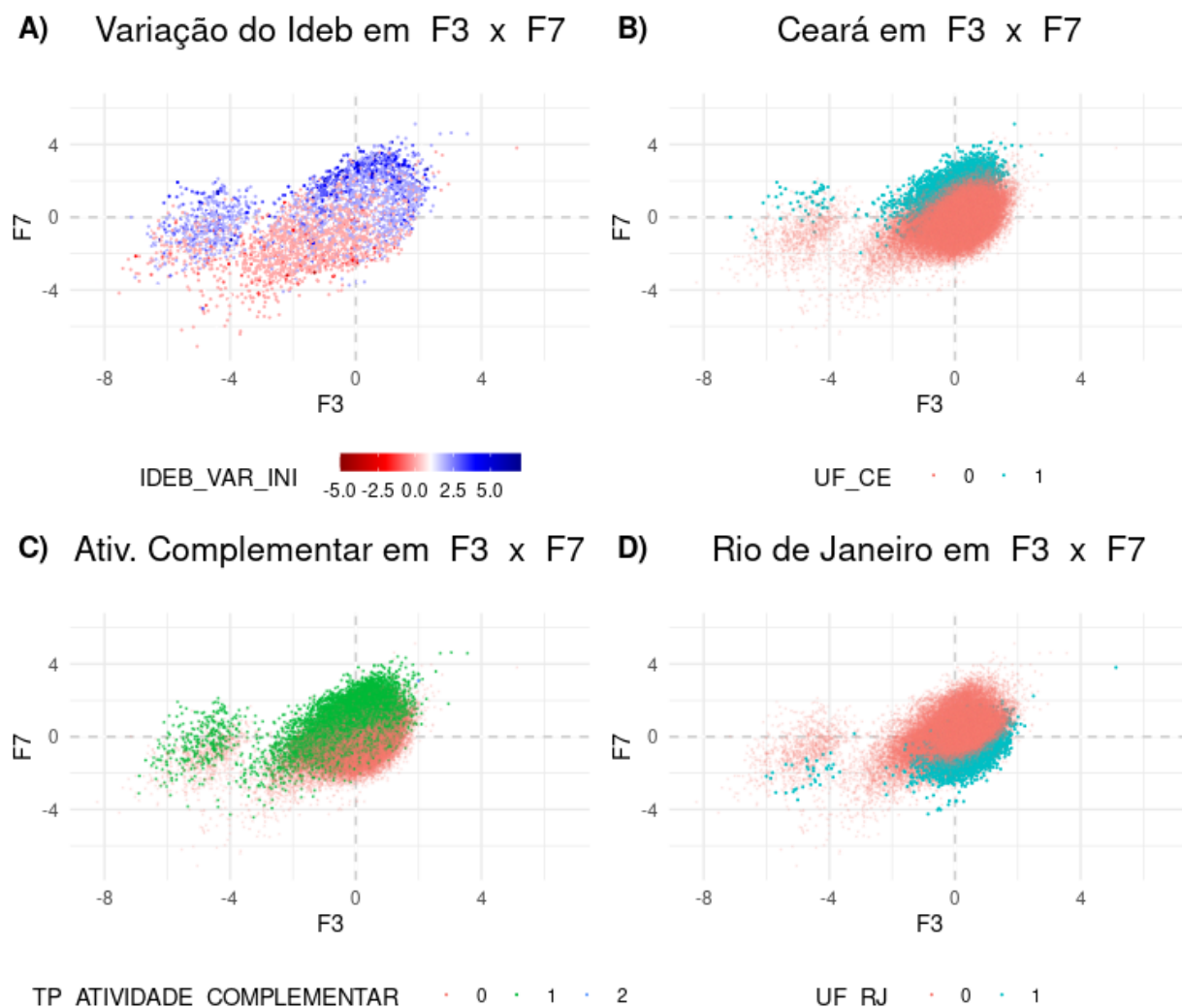


**Figura 17:** Projeção dos observáveis (58920 escolas) no espaço dos fatores. **A)** projeção nos fatores F1 e F2 conforme variação da nota Ideb, sendo pontos em vermelho escolas com variação abaixo da mediana (1.00) e pontos azuis acima. **B)** mesmo que A), mas nos fatores F3 e F7. **C)** projeção nos fatores F1 e F2 conforme a nota 2019 Ideb, sendo pontos em vermelho escolas com nota abaixo da mediana (5.00) e pontos azuis acima. **D)** mesmo que C), mas nos fatores F3 e F7. Nota-se que os fatores F1 e F2 são indiferentes aos índices Ideb, mas não F3 e F7.

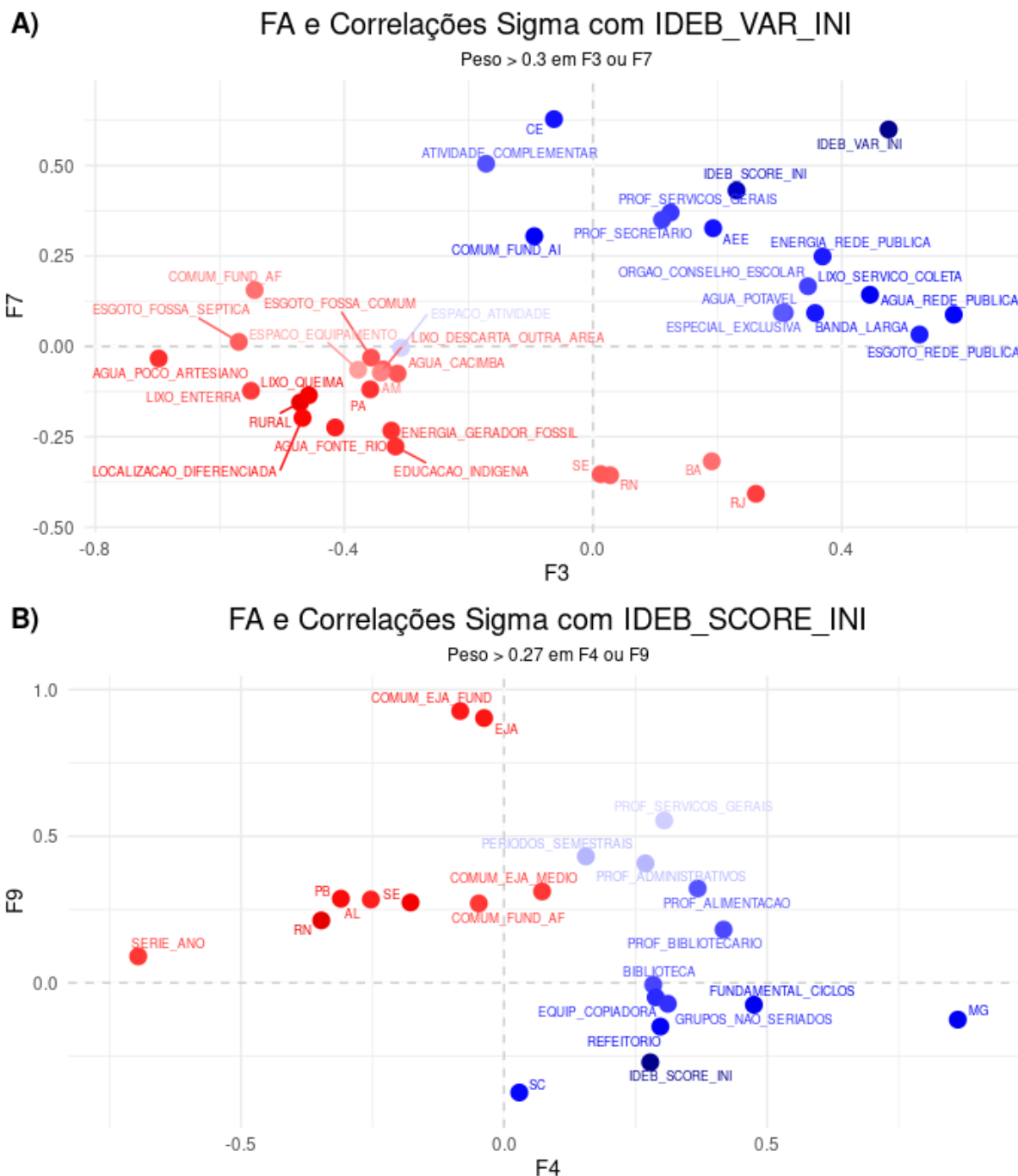
Isso fica mais evidente na Figura 18. Olhando 18-A, 18-B e 18-C, notam-se escolas com variação do Ideb acima da média, o estado do CE e presença de ATIVIDADE COMPLEMENTAR ao longo do lado positivo do eixo Y (F7). Já em 18-D, percebe-se que o



RJ ocupa uma região mais abaixo do CE, em parte no lado positivo de F3 e em parte no negativo de F7.



**Figura 18:** Projeção dos observáveis (58920 escolas) no espaço dos fatores F3 e F7. **A)** Variação da nota Ideb, sendo pontos em vermelho escolas com variação abaixo da mediana (1.00) e pontos azuis acima. **B)** Escolas do Ceará em azul e não do Ceará em vermelho. **C)** Escolas conforme tipo de atividade complementar, sendo: 0 - não oferece (vermelho) ; 1 - oferece não exclusivamente (verde); 2 - oferece exclusivamente (azul). **D)** Escolas do Rio de Janeiro em azul, e não do Rio de Janeiro em vermelho.



**Figura 19** : gráfico cruzando informações dos resultados da FA com as correlações Sigma, sendo positivas em azul, e negativas em vermelho. Cores proporcionais à raiz quadrada do valor absoluto das correlações (recuperando o sinal, em seguida), apenas para efeito de melhor visualização. Os prefixos dos atributos foram suprimidos também para facilitar a visualização. Seleção de pesos e fatores do FA conforme descrição nos subtítulos. **A)** Cores destacam as correlações Sigma dos atributos com a variação do Ideb entre 2005 e 2019. **B)** destaca as correlações Sigma dos atributos com a nota 2019 Ideb.

Isso indica que o RJ exige uma interpretação mais cuidadosa. Para esse estado, aquelas escolas que se posicionam num F3 positivamente mais destacadas tenderam a ter bom desempenho na variação do Ideb, em oposição àquelas posicionadas em F7 mais negativamente. Reforça também que escolas de RJ registraram relativamente menos atividades complementares que escolas do CE. Todas são conclusões coerentes com as

apresentadas na Tabela 17 para os mapeamentos de atributos x fatores.

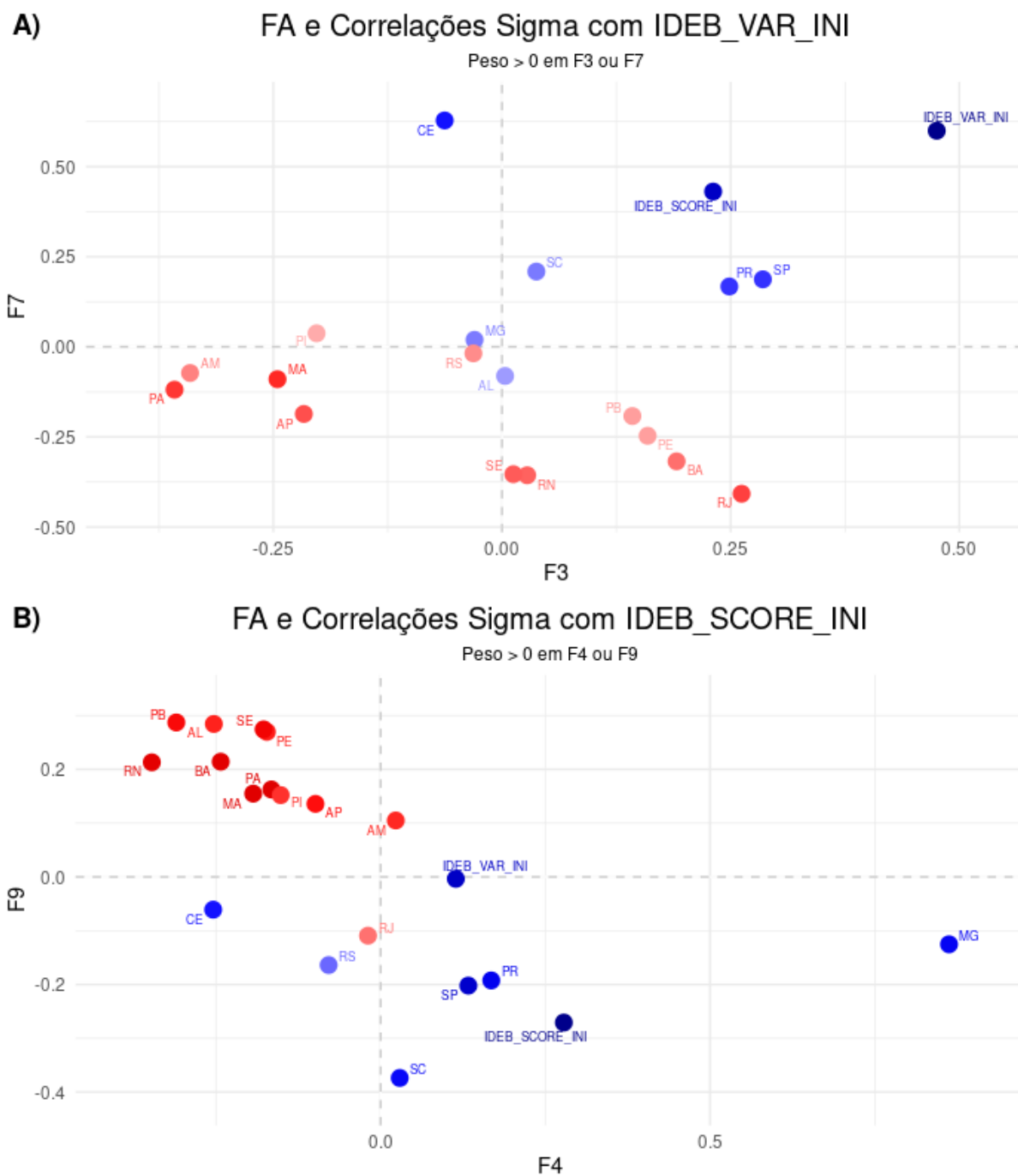
Aliás, esse cuidado com a interpretação dos dados como feito para o RJ também vale para outros estados e atributos. Com relação à MG, por exemplo, é preciso não cair na tentação, ao olhar o fator F4 da Tabela 17, de achar que escolas mineiras com flexibilização da seriação (atributos GRUPOS NÃO SERIADOS e FUNDAMENTAL CICLOS) tenderam a ter melhor desempenho geral na nota 2019 do Ideb que escolas com o tradicional modelo SÉRIE/ANO. O amarramento de atributos ao longo de um fator (como o F4) evidencia a possibilidade de maiores variâncias compartilhadas entre eles, e só permite dizer que escolas que se posicionaram ao longo do eixo desse fator tenderam a acompanhar esse emaranhamento de características. Mas, isso não pode ser generalizado para todas as escolas daquele estado. O espaço dos fatores é um espaço multidimensional e tanto atributos quanto escolas podem ter distribuições complexas nesse hiperespaço. Quando se olha separadamente um ou outro fator, tem-se apenas uma projeção desse espaço multidimensional em um espaço mais reduzido. O que ressalta novamente uma das vantagens do modelo de visualização RnVis, ao permitir um olhar abrangente entre muitas dimensões.

Uma outra forma de visualizar esses resultados está na Figura 19, dessa vez cruzando as correlações totais em Sigma com os fatores da FA. Em 19-A, o foco está nas correlações Sigma da variação do Ideb com os principais atributos em destaque da Tabela 17, conforme os fatores F3 e F7. Já em 19-B, o foco está nas correlações Sigma da nota 2019 do Ideb com os principais atributos em destaque da Tabela 17, conforme os fatores F4 e F9. Vê-se que os pontos que estão mais próximos geometricamente dos atributos Ideb tendem a ter alta correlação Sigma com eles, e os mais distantes, baixa correlação, indicando que o espaço dos Fatores é de fato um espaço relacionado às correlações.

Cabe lembrar que este colorimento diz das correlações totais Sigma. Assim, o estado RJ aparece em vermelho por apresentar correlação geral Sigma negativa (-0.143) com a variação do Ideb. Mas, como destacado antes, a FA permite detectar que as escolas deste estado podem ser desmembradas em grupos, sendo que escolas que se posicionaram ao longo do lado positivo do fator F3 tenderam a apresentar crescimento Ideb positivo, ao contrário das que se posicionaram na parte negativa de F7. Provavelmente, a correlação geral Sigma do RJ ficou negativa por ter mais escolas ao longo de F7 negativo que F3 positivo, conforme realmente parece sugerir a Figura 18-D.

Na Figura 20 foi feito um filtro para destacar as relações dos estados (que apareceram na Tabela 17) com a variação do Ideb e nota 2019 Ideb, conforme os fatores F3 e F7 para o primeiro, e F4 e F9 para o segundo. Nesse espaço tetradimensional, é possível identificar

visualmente melhor grupos já mencionados anteriormente, envolvendo: estados do norte, acrescidos de PI e MA; estados do nordeste, juntamente com o RJ; estados do sul, além de SP; os estados do CE e MG parecem formar grupos mais isolados. Lembrando que os estados que não apareceram: ou foram eliminados por terem baixa *communalities* com os atributos utilizados; ou apresentaram algum coeficiente na FA abaixo de 0.2 nos fatores considerados.



**Figura 20:** gráfico cruzando informações dos resultados da FA com as correlações Sigma para estados da federação, sendo positivas em azul, e negativas em vermelho. Cores proporcionais à raiz quadrada do valor absoluto das correlações (recuperando o sinal, em seguida), apenas para efeito de melhor visualização. Os prefixos dos atributos foram suprimidos também para facilitar a visualização. Seleção de pesos e fatores do FA conforme descrição nos subtítulos. **A)** Cores destacam as correlações Sigma dos atributos com a variação do Ideb entre 2005 e 2019. **B)** destaca as correlações Sigma dos atributos com a nota 2019 Ideb.

## 6 CONCLUSÃO

Esta pesquisa teve como objetivo identificar os dados coletados da educação brasileira utilizando algumas técnicas consagradas na Ciência dos Dados. O objetivo era constatar o que já se sabe sobre o assunto na literatura e apresentar informações novas e relevantes sobre o tema. As bases de dados utilizadas foram coletadas do Inep e constaram tanto os dados do Censo Escolar Nacional - Educacenso do ano de 2019, como os dados do Ideb de sua primeira edição, 2005, até a data de 2019 do ensino fundamental (1º ao 9º ano). Todos os dados foram tratados, conforme a Figura 5 deste trabalho.

Como os resultados dos anos iniciais do ensino fundamental apresentaram melhores resultados que as séries finais, esta pesquisa focou os resultados deste público. Para a obtenção dos resultados, foram utilizadas técnicas de análise exploratória de dados não supervisionadas. Sobre os resultados alcançados, é importante ressaltar que a metodologia permite identificar associações ou correlações entre atributos, mas não causalidades.

Após o tratamento dos dados, foi feita uma análise de correlação dos atributos, do Ideb e dos entes federativos. A técnica de correlação escolhida foi a Sigma da Cópula Gaussiana, visto que ela leva em consideração os dados categóricos e contínuos, e gerou uma matriz positiva definida.

Na análise de correlograma, foi possível perceber 4 sub-árvores. O grupo 1 compilou as infraestruturas básicas e alguns tipos de materiais pedagógicos. Nenhum estado foi associado a esse grupo, o que sugere que talvez nenhum estado contemplasse todas as infraestruturas básicas nas escolas. O grupo 2, apresenta infraestruturas mais aprimoradas (ex: salas climatizadas), as notas do Ideb e sua variação e a presença de órgãos escolares com entes federativos do sul e sudeste. O grupo 3, exibiu as correlações mais neutras, o que significa que são os atributos mais adversos. É composto, por exemplo, pelas escolas com EJA, profissionais sociais e os entes federativos do centro-oeste, dos estados do Ceará e do Espírito Santo. Parece que o grupo 3 é um refugio na análise, apesar de mostrar que o Ceará tem feito práticas diferenciadas em relação aos outros estados do Brasil. O grupo 4 é o grupo com mais correlações negativas e contempla quase todos os estados do norte e nordeste, o que sugere que podem representar escolas com maior precariedade de infraestrutura.

Para a aplicação da técnica do FA, definiu a quantidade de 10 fatores como mais adequados, tendo 82% da variância dos dados. Os fatores 1 e 2, evidenciaram associações positivas a infraestruturas computacionais e arquitetônicas, respectivamente, e negativamente a estruturas precárias e não apresentaram associação ao Ideb. O fator F1 também apresentou

associação aos materiais pedagógicos e órgãos escolares e o F2 serviços ofertados aos alunos. Os estados nesses grupos foram sul e sudeste como positivo e norte e nordeste como negativo.

Já os fatores F3, F4, F7 e F9 apresentaram associação aos atributos IDEB\_VAR\_INI (variação do Ideb de 2005 a 2019) e/ou IDEB\_SCORE\_INI (nota do Ideb 2019). O fator F3, destacou os serviços públicos essenciais, a internet de qualidade e o conselho escolar; e negativamente os serviços básicos precários. Assim, o F3 pode-se deduzir que tendeu a associar as escolas urbanas em seu eixo positivo e escolas rurais ou indígenas no eixo negativo. O fator F4, que contemplou a nota de 2019 do Ideb, apresentou positivamente principalmente o ensino em ciclos e grupos não seriados, em contraposição a modalidade tradicional de ensino série por ano. O estado de MG ganhou um grande peso positivo. O fator F7, de maior relevância na variação do Ideb e a nota 2019, apresentou as atividades complementares, AEE, com grande destaque ao CE. O fator F9, com relação negativa à nota do Ideb de 2019, associou ao EJA, seus períodos semestrais e negativamente aos atributos associação de pais e mestres, parque infantil (faz sentido, se a associação maior é ao EJA) e o não tratamento do lixo. Isso pode indicar que o EJA cresceu menos que o ensino tradicional.

Os fatores restantes F5, F6, F8 e F10 não evidenciaram relação ao Ideb e nem a sua variação. O fator F5 teve destaque no tratamento de lixo tanto na separação quanto na reciclagem. O fator F6 associou-se mais ao ensino médio, EJA e séries finais em contraposição ao ensino infantil e material pedagógico infantil e forte correlação negativa com escolas municipais, visto que o nível de ensino médio normalmente é de responsabilidade das escolas estaduais. O fator F8, com apenas 3 atributos que os profissionais, psicólogos e nutricionistas não se correlacionam com nenhum outro fator, inclinaram-se em menor grau ao estado da BA. E o último fator F10, reuniu em maior peso as escolas que realizam atividades comunitárias dentro da escola, usa o espaço ou equipamentos em atividades externas à escola, possuem redes sociais, rede local e propostas pedagógicas. Destacam-se nesse fator os estados do CE.

Para finalizar este estudo, elaborou a visualização de alguns observáveis (escolas) nos fatores F1 x F2 e F3 x F7. Nos fatores F1 e F2, que não possuem correlação ao Ideb, os pontos parecem aleatórios, já nos fatores F3 e F7 existe uma padronização. Assim, pode-se averiguar que existem 3 formas de alcançar bons resultados no Ideb: apresentar F3 elevado com baixo F7, apresentar F7 elevado e baixo F3 ou apresentar F3 e F7 elevados. Nesse sentido foi curioso perceber como CE está num fator dessa composição, e SP e PR em outra, tendo SC numa posição intermediária, mas todos apresentam-se atrelados a uma variação do Ideb positiva.

## 6.1 LIMITAÇÕES E PERSPECTIVAS

A base de dados contendo somente características de infraestrutura das escolas, conforme o Censo Escolar constitui um primeiro fator limitante. A escolha dessa base mais restrita às características foi intencional no sentido de conferir uma primeira validação confiável da cuidadosa metodologia utilizada neste trabalho frente alguns resultados e padrões já bem conhecidos da literatura. Acredita-se que esse objetivo foi satisfatoriamente cumprido, não só na reprodução de alguns padrões já estabelecidos, mas também na sugestão de novas perspectivas de análises.

Em estudos futuros, pretende-se expandir as bases aqui utilizadas e integrar outros conjuntos de dados e microdados envolvendo o perfil dos docentes, dos gestores, dos alunos e das famílias, aspectos sociais, políticos e econômico-financeiros dos familiares dos discentes, municípios e estados, Índice de Oportunidades da Educação Brasileira (IOEB), investimento público em educação e os outros indicadores educacionais do governo. Em especial, agregar também dados e metadados do Saeb, no geral mais abrangentes que o Censo (ALVES & XAVIER, 2018). Vale também para outros níveis de ensino além do fundamental anos iniciais.

Do ponto de vista da metodologia, embora ela tenha revelado ser capaz de reproduzir alguns padrões já conhecidos e apresentar outros novos nas bases aqui utilizadas, há espaço para estudos mais aprofundados em alguns de seus aspectos. Especificamente sobre os microdados do Educacenso, a metodologia não aprofundou nos estudos dos valores ausentes (NAs) nem de seus valores extremos (*outliers*). Também não ficou muito claro por que todas as unidades da federação da região Centro Oeste apresentaram *uniqueness* tão altas, a ponto de terem sido descartadas na seleção de atributos.

Não ganhou foco neste trabalho uma análise comparativa mais detalhada entre as opções de matrizes de correlações disponíveis, entre Pearson, policóricas, Sigma decorrente da cópula Gaussiana e outras. Com respeito à FA, há uma série de hiperparâmetros passíveis de ajustes, entre: diferentes métodos para extração dos fatores; as diferentes opções de rotações; diferentes alternativas para as projeções dos observáveis no espaço dos fatores. Tudo isso seria passível de análises comparativas mais rigorosas em estudos futuros.

## REFERÊNCIAS

- AGUINIS, H.; GOTTFREDSON, R. K.; JOO, H. Best-Practice Recommendations for Defining, Identifying, and Handling Outliers. **Organizational Research Methods**, [s. l.], v. 16, n. 2, p. 270–301, 2013.
- ALBERNAZ, Â; FERREIRA, F. H. G.; FRANCO, C. Qualidade e equidade na educação fundamental brasileira. **Departamento de Economia, PUC-RIO**. Texto para Discussão, n. 455, Rio de Janeiro, 2002.
- ALVES, M. T. G., & XAVIER, F. P. Indicadores multidimensionais para avaliação da infraestrutura escolar: o ensino fundamental. **Cadernos de Pesquisa**, 48, 708-746, 2018.
- BANDALOS, D. L., & FINNEY, S. J. Factor analysis: Exploratory and confirmatory. In **The reviewer's guide to quantitative methods in the social sciences** (pp. 98-122). Routledge, 2018.
- BRASIL. LDB. Lei de Diretrizes e Bases da Educação Nacional. 2. ed. **Secretaria de Editoração e Publicações** - Coordenação de Edições Técnicas, Senado Federal, Brasília, 2018a.
- BRASIL. BNCC. Ministério da Educação. Secretaria da Educação Básica. **Base nacional comum curricular**. Brasília, 2018b. Disponível em: <http://basenacionalcomum.mec.gov.br/> Acesso em: 29 nov. 2022.
- BRASIL. Constituição (1988). **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidência da República, [2016]. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/constituicao/constituicao.htm](https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm). Acesso em: 29 nov. 2022.
- BRASIL. Inep. **Inep: Microdados do Censo Escolar**, 2012. Disponível em: <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/censo-escolar> Acesso em: 29 nov. 2022
- BRASIL. Inep. **Resultado Índice de Desenvolvimento da Educação Básica (Ideb)**, 2020. Disponível em: <https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/ideb/resultados>. Acesso em: 29 nov. 2022.
- BRASIL. Inep. **Nota Técnica Índice de Desenvolvimento da Educação Básica - Ideb**, 2021a. Disponível em: [https://download.inep.gov.br/educacao\\_basica/portaI\\_deb/o\\_que\\_e\\_o\\_ideb/Nota\\_Tecnica\\_n1\\_concepcaoIDEB.pdf](https://download.inep.gov.br/educacao_basica/portaI_deb/o_que_e_o_ideb/Nota_Tecnica_n1_concepcaoIDEB.pdf) Acesso em: 29 nov. 2022.
- BRASIL. Inep. **Nota Informativa do Ideb 2021**, 2021b. Disponível em: [https://download.inep.gov.br/educacao\\_basica/portaI\\_deb/planilhas\\_para\\_download/2021/nota\\_informativa\\_ideb\\_2021.pdf](https://download.inep.gov.br/educacao_basica/portaI_deb/planilhas_para_download/2021/nota_informativa_ideb_2021.pdf). Acesso em: 29 nov. 2022



BRASIL, Inep. **CENSO ESCOLAR**. [S. l.], 2022a. Disponível em: <https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-escolar>. Acesso em: 1 nov. 2022.

BRASIL. Inep. **Avaliações e Exames Educacionais**, 2022b. Disponível em: <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais>. Acesso em: 08 set. 2022.

BRASIL. Inep. **Microdados do Censo Escolar da Educação Básica**, 2022c. Disponível em: <https://www.gov.br/inep/pt-br/area-de-atuacao/dados-abertos/microdados/censo-escolar>. Acesso em: 29 nov. 2022.

BRASIL. Inep. **Sistema de Avaliação da Educação Básica (Saeb)**, 2022d. Disponível em: <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/saeb>. Acesso em: 08 set. 2022.

BRASIL. Inep. **Relatório de Resultados do Saeb 2019 Volume 1**, 2022e. Disponível em: [https://download.inep.gov.br/educacao\\_basica/saeb/2019/resultados/relatorio\\_de\\_resultados\\_do\\_saeb\\_2019\\_volume\\_1.pdf](https://download.inep.gov.br/educacao_basica/saeb/2019/resultados/relatorio_de_resultados_do_saeb_2019_volume_1.pdf). Acesso em: 29 nov. 2022

BROOKE, N.; SOARES, J. F. **Pesquisa em eficácia escolar: origem e trajetória**. Estudos em Avaliação Educacional. V. 22, n. 50, p. 593-598, 2011.

BUCHMANN, G., & NERI, M. C. **The brazilian education quality index (Ideb): measurement and incentives upgrades**. 2008.

CHAI, C. S., LIN, P. Y., KING, R. B., & JONG, M. S. Y. **Intrinsic Motivation and Sophisticated Epistemic Beliefs Are Promising Pathways to Science Achievement: Evidence from High Achieving Regions in the East and the West**. *Frontiers in Psychology*, p. 448, 2021.

COLEMAN, J. S. **Equality of Educational Opportunity**. Washington: Office of Education, U.S., 1966.

CORDERO, J. M., CRISTOBAL, V., & SANTÍN, D. **Causal inference on education policies: A survey of empirical studies using PISA, TIMSS and PIRLS**. *Journal of Economic Surveys*, 32(3), p. 878-915, 2018.

COSTELLO, A. B.; OSBORNE, J. W. **Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis**. *Practical Assessment, Research and Evaluation*, [s. l.], v. 10, n. 7, p. 1–10, 2005.

CUNNINGHAM, P. Dimension Reduction. In: CORD, M.; CUNNINGHAM, P. (org.). **Machine Learning Techniques for Multimedia**. Cognitive. Berlin, Heidelberg: Springer, 2008. p. 91–112.

DALPIAZ, D. **Applied statistics with R**. [S. l.]: University of Illinois. Urbana-Champaign, 2022. E-book. Disponível em: <https://book.stat420.org/>.

DAVENPORT, T. H.; PATIL, D. J. Data scientist. **Harvard business review**, v. 90, n. 5, p. 70-76, 2012.

DOURADO, R.; RODRIGUES, L. R.; CAVALCANTI, J.; GOMES, A. S. Novas possibilidades de avaliação em larga escala na educação básica através do uso de EDM e Learning Analytics. In: **Anais do VI Workshop de Desafios da Computação aplicada à Educação**. SBC, 2017.

EDM'08. The First International Conference on Educational Data Mining. Montreal, Québec, Canada: Université du Québec à Montréal (UQÀM), 2008.

ERBERT, M. Introdução ao sensoriamento Remoto. Master Tesis, Universidade Federal do Rio Grande do Sul, 2001.

ERIKSSON, K., LINDVALL, J., HELENIUS, O., & RYVE, A. Cultural variation in the effectiveness of feedback on students' mistakes. **Frontiers in Psychology**, 10, p. 3053, 2020a.

ERIKSSON, K., BJÖRNSTJERNA, M., & VARTANOVA, I. The relation between gender egalitarian values and gender differences in academic achievement. **Frontiers in psychology**, 11, p. 236, 2020b.

ERIKSSON, K. Gender Differences in the Interest in Mathematics Schoolwork Across 50 Countries. **Frontiers in Psychology**, p. 11, 2020c.

FABRIGAR, L. R. et al. Evaluating the Use of Exploratory Factor Analysis in Psychological Research. **Psychological Methods**, [s. l.], v. 4, n. 3, p. 272–299, 1999.

FERREIRA, L. A., RODRIGUES, R. L., DE SOUZA, R. N. Dados Abertos Educacionais Brasileiros: Um Mapeamento Sistemático da Literatura. In **Anais do XXXII Simpósio Brasileiro de Informática na Educação** (pp. 1186-1195). SBC, 2021.

FIGUEIREDO, E.; MACEDO, M.; SIQUEIRA, H. V.; SANTANA JR, C. J.; GOKHALE, A.; BASTOS-FILHO, C. J. A. Swarm intelligence for clustering—A systematic review with new perspectives on data mining. **Engineering Applications of Artificial Intelligence**. v. 82, p. 313-329, 2019.

FISHER, R. A. The use of multiple measurements in taxonomic problems. **Annals of eugenics**, [s. l.], v. 7, n. 2, p. 179–188, 1936.

FORTUNATO, S. Community detection in graphs. **Physics Reports**, [s. l.], v. 486, n. 3–5, p. 75–174, 2010.

FRENEDA, F. C. B. **Múltiplos fatores do desempenho escolar: uma análise dos microdados do Inep sobre a educação no Distrito Federal**. 2020.

GADOTTI, M. **Qualidade na educação: uma nova abordagem**. 2010 Disponível em: [http://www.acervo.paulofreire.org:8080/jspui/bitstream/7891/3086/1/FPF\\_PTPF\\_12\\_084.pdf](http://www.acervo.paulofreire.org:8080/jspui/bitstream/7891/3086/1/FPF_PTPF_12_084.pdf) Acesso em: 29 nov. de 2022.

GANDRUD, C. **Reproducible research with R and RStudio**. Chapman and Hall/CRC, 2018.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

GITHUB. **GitHub [Internet]**. 2020. URL <https://github.com/>

GOMES, C. A. **A Educação em Perspectiva Sociológica**. 3ª ed., São Paulo: EPU, 1994.

GOMES, A., & REGIS, A. Desempenho e infraestrutura: mapeamento das escolas públicas da região metropolitana do Rio de Janeiro. In **Congresso Ibero-americano de Política e Administração da Educação** (Vol. 3), 2012.

GRICE, J. W. Computing and evaluating factor scores. **Psychological Methods**, [s. l.], v. 6, n. 3, p. 430–450, 2001.

GRIVOKOSTOPOULOU, F., PERIKOS, I., HATZILYGEROUDIS, I. **Utilizing semantic web technologies and data mining techniques to analyze students learning and predict final performance**. IEEE international conference on teaching, assessment and learning for engineering (TALE) IEEE, 2014.

GRUSON, D., HELLEPUTTE, T., ROUSSEAU, P., GRUSON, D. Data science, artificial intelligence, and machine learning: Opportunities for laboratory medicine and the value of positive regulation. **Clinical biochemistry**, v. 69, p. 1-7, 2019.

HE, Y. *et al.* Online Learning in Variable Feature Spaces with Mixed Data. Proceedings - **IEEE International Conference on Data Mining**, ICDM, [s. l.], v. 2021-December, p. 181–190, 2021.

HENSON, R. K.; ROBERTS, J. K. Use of Exploratory Factor Analysis in Published Research: Common Errors and Some Comment on Improved Practice. **Educational and Psychological Measurement**, [s. l.], v. 66, n. 3, p. 393–416, 2006

HOLGADO–TELLO, F. P. *et al.* Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. **Quality and Quantity**, [s. l.], v. 44, n. 1, p. 153–166, 2010.

HUSSON, F.; JOSSE, J.; PAGÈS, J. Principal component methods - hierarchical clustering - partitional clustering: why would we need to choose for visualizing data? **Technical Report – Agrocampus**. [S. l.: s. n.], 2010.

IMDADULLAH, M.; ASLAM, M.; ALTAF, S. Mctest: An R package for detection of collinearity among regressors. **R Journal**, [s. l.], v. 8, n. 2, p. 499–509, 2016.

JAMES, G. *et al.* An Introduction to Statistical Learning with Applications in R. 2nd. ed. New York: Springer, 2021

JAVIDI, G., RAJABION, L., SHEYBANI, E. **Educational Data Mining and Learning Analytics - Overview of Benefits and Challenges**. International Conference on Computational Science and Computational Intelligence, 2017.

JAWORSKI, P. *et al.* **Copula theory and its applications**. Berlin: Springer, 2010.

JOSSE, J., & HUSSON, F. missMDA: A Package for Handling Missing Values in Multivariate Data Analysis. **Journal of Statistical Software**, 70(1), p. 1–31, 2016.

KAISER, H. F. An index of factorial simplicity. **Psychometrika**, [s. l.], v. 39, n. 1, p. 31–36, 1974.

KIM, Tae Kyun. Practical statistics in pain research. **Korean Journal of Pain**, [s. l.], v. 30, n. 4, p. 243–249, 2017.

KOLENIKOV, S.; ANGELES, G. The Use of Discrete Data in PCA: Theory, Simulations, and Applications to Socioeconomic Indices. Chapel Hill: **Carolina Population Center**, University of North Carolina., [s. l.], n. November 2004, p. 1–59, 2004.

LÊ, S.; JOSSE, J.; HUSSON, F. FactoMineR: An R Package for Multivariate Analysis. **Journal of Statistical Software**, [s. l.], v. 25, n. 1, p. 1–18, 2008.

LIN, M.; LUCAS, H. C.; SHMUELI, G. Too big to fail: Large samples and the p-value problem. **Information Systems Research**, [s. l.], v. 24, n. 4, p. 906–917, 2013.

LORENZO-SEVA, U.; FERRANDO, P. J. Not Positive Definite Correlation Matrices in Exploratory Item Factor Analysis: Causes, Consequences and a Proposed Solution. **Structural Equation Modeling**, [s. l.], v. 28, n. 1, p. 138–147, 2021.

MAĆKIEWICZ, A.; RATAJCZAK W. Principal components analysis (PCA), **Computers & Geosciences**, v. 19, n. 3, 1993

MATUCHESKI, V. L. P. Avaliação em Larga Escala: uma análise da diferença do rendimento do Ideb dos estudantes dos 5º anos comparados aos 9º anos do Ensino Fundamental em Língua Portuguesa. Dissertação (Mestrado em Educação) **Universidade Estadual Do Oeste Do Paraná**, Paraná, p. 127. 2021

MINAS GERAIS. Currículo Referência de Minas Gerais. **Minas Gerais** 2018. Disponível em:  
[http://basenacionalcomum.mec.gov.br/images/implementacao/curriculos\\_estados/documento\\_curricular\\_mg.pdf](http://basenacionalcomum.mec.gov.br/images/implementacao/curriculos_estados/documento_curricular_mg.pdf) Acesso em: 29 nov. 2022

MULLIS, I. V., MARTIN, M. O. TIMSS 2019 Assessment Frameworks. International Association for the Evaluation of Educational Achievement. Herengracht 487, Amsterdam, 1017 BT, The Netherlands, 2017.

MURTAGH, F.; CONTRERAS, P. Algorithms for hierarchical clustering: an overview, II. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, [s. l.], v. 7, n. 6, p. 1–16, 2017.

NGUYEN, G. *et al.* Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey. **Artificial Intelligence Review**. v. 52, n. 1, p. 77-124, 2019.

NGUYEN, M. **A Guide on Data Analysis. [S. l.]: bookdown R package**, 2022. E-book. Disponível em: [https://bookdown.org/mike/data\\_analysis/](https://bookdown.org/mike/data_analysis/).

PALERMO, G. A.; SILVA, D. B. N.; NOVELLINO, M. S. F. Fatores associados ao desempenho escolar: uma análise da proficiência em matemática dos alunos do 5º ano do ensino fundamental da rede municipal do Rio de Janeiro. **Revista Brasileira de Estudos de População**, v. 31, n. 2, p. 367-394, 2014.

PEARSON, K. LIII. On lines and planes of closest fit to systems of points in space. **The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science**, [s. l.], v. 2, n. 11, p. 559–572, 1901.

PEDROSO, L., TANAKA, A., CAPPELLI, C. A Lei de Acesso à Informação brasileira e os desafios tecnológicos dos dados abertos governamentais. In: **SIMPÓSIO BRASILEIRO DE SISTEMAS DE INFORMAÇÃO (SBSI)**, 9., João Pessoa. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, p. 523-528, 2013.

R Core Team **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>, 2020.

RAM, N. **Exploratory Factor Analysis - Basic**. [S. l.], 2022. Disponível em: <https://quantdev.ssri.psu.edu/tutorials/intro-basic-exploratory-factor-analysis>. Acesso em: 1 nov. 2022.

REVELLE, W. **Constructs, Components, and Factor Models**. In: **AN INTRODUCTION TO PSYCHOMETRIC THEORY WITH APPLICATIONS IN R**. [S. l.: s. n.], 2009. p. 1–58. E-book. Disponível em: <http://personality-project.org/r/book/#chapter6>.

REVELLE, W. **How to use the psych package to display item and scale correlations**. [S. l.], 2022a. Disponível em: <https://cloud.r-project.org/web/packages/psychTools/vignettes/mediation.pdf>. Acesso em: 1 nov. 2022.

REVELLE, W. **psych: Procedures for Personality and Psychological Research**, Northwestern University, Evanston, Illinois, USA. Disponível em: <https://CRAN.R-project.org/package=psych>, Version = 2.2.5, 2022b.

REVELLE, W. **Package psych**. [S. l.], 2022c. Disponível em: <https://personality-project.org/r/psych/>. Acesso em: 1 nov. 2022.

RIGOTTI, J. I. R; CERQUEIRA, C. A. **As bases de dados do Inep e os indicadores educacionais: conceitos e aplicações**. Livros, p. 71-88, 2015.

RISCAL, J. R., & LUIZ, M. C. **Gestão democrática e a análise de avaliações em larga escala: o desempenho de escolas públicas no Brasil**. São Carlos: Pixel, 2016.

RODRIGUES, N. **Responsabilidade do estado e da sociedade**. Tecnologia educacional, Rio de Janeiro, v. 20, n. 101, p. 12-19, 1991.

RODRIGUES, R. L., RAMOS, J. L. C., SILVA, J. C. S., & GOMES, A. S. A literatura brasileira sobre mineração de dados educacionais. In: **Anais dos Workshops do Congresso Brasileiro de Informática na Educação**, Vol. 3, No. 1, p. 621, 2014.

ROKACH, L.; MAIMON, O. Clustering Methods. In: **DATA MINING AND KNOWLEDGE DISCOVERY HANDBOOK**. Boston, MA: Springer, 2005. p. 321–352.

RStudio Team **RStudio: Integrated Development for R**. RStudio, Inc. Boston, MA URL <http://www.rstudio.com/>, 2019.

RUIPÉREZ-VALIENTE, J. A., MUÑOZ-MERINO, P. J., & DELGADO KLOOS, C. **Improving the prediction of learning outcomes in educational platforms including higher level interaction indicators**. Expert Systems, e12298, 2018.

SANTOS, R. D. O. **Principal component analysis and factor analysis: Differences and similarities in nutritional epidemiology application**. Revista Brasileira de Epidemiologia, [s. l.], v. 22, p. 1–14, 2019.

SAVIANI, DI. **O Inep, o diagnóstico da educação brasileira e a Rbep**. Revista brasileira de estudos pedagógicos, v. 93, n. 234, 2012. *et al.*

SCHUBERT, E. et al. **DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN**. ACM Transactions on Database Systems, [s. l.], v. 42, n. 3, p. 1–21, 2017.

SIEMENS, G., & BAKER, R. S. D. **Learning analytics and educational data mining: towards communication and collaboration**. In Proceedings of the 2nd international conference on learning analytics and knowledge, 2012.

SILVA FILHO, R. L. C. **Modelo de análise e predição do desempenho dos alunos dos Institutos Federais de Educação usando o Enem como indicador de qualidade escolar.**, 2017.

SILVA, H. A. D. Um Estudo sobre o Índice de Desenvolvimento da Educação Básica (Ideb): Como as Escolas Públicas de Juiz de Fora Melhoraram o seu Desempenho. 126 f. Mestrado Acadêmico em Educação Instituição de Ensino: **Universidade Federal de Juiz de Fora**, Juiz de Fora, 2012.

SILVA, L.L.S.; LIMA, A.F.R.; POLLI, D.A.; RAZIA, P.F.S.; PAVÃO, L.F.A.; CAVALCANTI, M.A.F.H.; *et al.* **Medidas de distanciamento social para o enfrentamento da COVID-19 no Brasil: caracterização e análise epidemiológica por estado**. Cadernos de Saúde Pública. v. 36, n. 9, 2020.

SOARES, J. F., XAVIER, F. P. Pressupostos educacionais e estatísticos do Ideb. Educação & Sociedade [on-line]. 2013, 34(124), 903-923 Disponível em: <https://www.redalyc.org/articulo.oa?id=87328534013> Acesso em: 1 nov. 2022

SOARES, R. C., NETO, N. W., COUTINHO, L. R., SILVA, F. J., SANTOS, D. V., & TELES, A. S. Mineração de dados da educação básica brasileira usando as bases do Inep: Uma revisão sistemática da literatura. **RENOTE**, 19(1), p. 361-370, 2021.

SOARES, S. S. D. **A repetência no contexto internacional: o que dizem os dados de avaliações das quais o Brasil não participa?**, 2007.

SPEARMAN, C. **“General Intelligence,” Objectively Determined and Measured.** The American Journal of Psychology, [s. l.], v. 15, n. 2, p. 201–292, 1904.

STOROPOLI, J. **How to use FactorAssumptions.** [S. l.], 2019. Disponível em: <https://cran.r-project.org/web/packages/FactorAssumptions/vignettes/vignette.html>. Acesso em: 1 nov. 2022.

TIBCO. Site do Tibco, 2022. **O que é aprendizagem não supervisionada?** Disponível em: <https://www.tibco.com/pt-br/reference-center/what-is-unsupervised-learning/>. Acesso em: 1 nov. 2022.

TORGO, L. **Data mining with R: learning with case studies.** chapman and hall/CRC., 2011.

VAN BUUREN, S. **Flexible Imputation of Missing Data.** 2nd. ed. [S. l.]: Chapman & Hall/CRC, 2018.

VAN DER EIJK, C.; ROSE, J. **Risky business: Factor analysis of survey data - Assessing the probability of incorrect dimensionalisation.** PLoS ONE, [s. l.], v. 10, n. 3, p. 1–31, 2015.

VASIC, D., KUNDID, M., PINJUH, A., & SERIC, L. **Predicting student’s learning outcome from Learning management system logs.** 23rd International Conference on Software, Telecommunications and Computer Networks (SoftCOM), 2015.

WATKINS, M. W. **Exploratory Factor Analysis: A Guide to Best Practice.** Journal of Black Psychology, [s. l.], v. 44, n. 3, p. 219–246, 2018.

WICKHAM, H. **R for Data Science.** [S.I.], 2022. Disponível em: <https://r4ds.had.co.nz/index.html>. Acesso em: 20 nov. 2022.

WILLIAMS, B.; ONSMAN, A.; BROWN, T. Exploratory factor analysis: A five-step guide for novices. **Journal of Emergency Primary Health Care**, [s. l.], v. 8, n. 3, p. 1–13, 2010.

ZAKI, M. J.; & MEIRA JR., W. **DATA MINING AND ANALYSIS. 2nd Editioned.** New York: Cambridge University Press, 2020. E-book. Disponível em: [https://dataminingbook.info/book\\_html/](https://dataminingbook.info/book_html/). Acesso em: 1 nov. 2022.

ZHAO, Y; UDELL, M. Missing Value Imputation for Mixed Data via Gaussian Copula. In: , 2020. Proceedings of the **26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.** [S. l.: s. n.], 2020. p. 636–646.

## Apêndice A - Cálculos do Ideb

Exemplo de cálculo do Ideb de uma escola fictícia do ensino fundamental, com as respectivas fórmulas (Equações 2 à 13).

### Nota Padronizada de Matemática - NPM

$$\frac{\text{Nota de Matemática do exame em 2019} - \text{limite inferior da média de proficiência de Matemática no SAEB 1997}}{\text{limite superior da média de proficiência de Matemática no SAEB 1997} - \text{limite inferior da média de proficiência de Matemática no SAEB 1997}} \times 10 \quad (2)$$

$$NPM = \frac{252,4 - 60}{322 - 60} \times 10 = \frac{192,4}{262} \times 10 = 7,344 \quad (3)$$

### Nota Padronizada de Português - NPP

$$\frac{\text{Nota de Português do exame em 2019} - \text{limite inferior da média de proficiência de Português no SAEB 1997}}{\text{limite superior da média de proficiência de Português no SAEB 1997} - \text{limite inferior da média de proficiência de Português no SAEB 1997}} \times 10 \quad (4)$$

$$NPM = \frac{214,6 - 49}{324 - 49} \times 10 = \frac{165,6}{275} \times 10 = 6,022 \quad (5)$$

### Nota Média Padronizada - (N)

$$\frac{\text{Nota Padronizada de Matemática} + \text{Nota Padronizada de Matemática}}{2} \times 10 \quad (6)$$

$$N = \frac{7,34 + 6,02}{2} = \frac{13,36}{2} = 6,683 \quad (7)$$

### Indicador de rendimento - (P)

$$P = \frac{1}{T} \quad (8)$$

Onde,

$$T = \frac{\sum_{r=1}^n \frac{1}{pr}}{n} \quad (9)$$

Sendo,

p = Taxa de aprovação em cada ano dos anos iniciais ou finais do ensino fundamental / 100

r = Ano dos anos iniciais ou finais do ensino fundamental (1º ao 5º ano ou 6º ao 9º ano)

n = Número de anos nos anos iniciais ou finais do ensino fundamental com taxa de aprovação

Assim,

$$\sum_{r=1}^5 \frac{1}{pr} = \frac{1}{1} + \frac{1}{0,987} + \frac{1}{0,873} + \frac{1}{1} + \frac{1}{0,901} = 5,269 \quad (10)$$

Portanto,



$$T = \frac{5,26852451}{5} = 1,054 \quad (11)$$

$$P = \frac{1}{1,05370490} = 0,949 \quad (12)$$

Logo,

$$Ideb = P \times N = 0,949 \times 6,683 = 6,3 \quad (13)$$

Desta forma a escola deste exemplo teria a nota 6,3 no Ideb.

## Apêndice B - Escala de proficiências 5ºs e 9ºs

Quadro B.1 - Escala de proficiência para interpretação dos resultados do 5º ano do ensino fundamental em língua portuguesa no saeb – 2019

| Nível   | DESCRIÇÃO DAS HABILIDADES DESENVOLVIDAS  |
|---|--|
| <p>Nível 1</p> <p>Desempenho maior ou igual a 125 e menor que 150</p> | <p>Os estudantes provavelmente são capazes de: localizar informações explícitas em textos narrativos curtos, informativos e anúncios. Identificar o tema de um texto. Localizar elementos, como o personagem principal. Estabelecer relação entre partes do texto: personagem e ação; ação e tempo; ação e lugar.</p>  |
| <p>Nível 2</p> <p>Desempenho maior ou igual a 150 e menor que 175</p> | <p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: localizar informações explícitas em contos. Identificar o assunto principal e a personagem principal em reportagens e em fábulas. Reconhecer a finalidade de receitas, manuais e regulamentos. Inferir características de personagens em fábulas. Interpretar linguagem verbal e não verbal em tirinhas.</p>  |
| <p>Nível 3</p> <p>Desempenho maior ou igual a 175 e menor que 200</p> | <p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: localizar informação explícita em contos e reportagens. Localizar informação explícita em propagandas com ou sem apoio de recursos gráficos. Reconhecer relação de causa e consequência em poemas, contos e tirinhas. Inferir o sentido de palavra, o sentido de expressão ou o assunto em cartas, contos, tirinhas e histórias em quadrinhos, com o apoio de linguagem verbal e não verbal.</p>  |
| <p>Nível 4</p> <p>Desempenho maior ou igual a 200 e menor que 225</p> | <p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: identificar informação explícita em sinopses e receitas culinárias. Identificar assunto principal e personagem em contos e letras de música. Identificar formas de representação de medida de tempo em reportagens. Identificar assuntos comuns a duas reportagens. Identificar o efeito de humor em piadas. Reconhecer sentido de expressão, elementos da narrativa e opinião em reportagens, contos e poemas. Reconhecer relação de causa e consequência e relação entre pronomes e seus referentes em fábulas, poemas, contos e tirinhas. Inferir sentido decorrente da utilização de sinais de pontuação e sentido de expressões em poemas, fábulas e contos. Inferir efeito de humor em tirinhas e histórias em quadrinhos.</p>  |
| <p>Nível 5</p> <p>Desempenho maior ou igual a 225 e menor que 250</p> | <p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: identificar assunto e opinião em reportagens e contos. Identificar assunto comum a cartas e poemas. Identificar informação explícita em letras de música e contos. Reconhecer assunto em poemas e tirinhas. Reconhecer sentido de conjunções e de locuções adverbiais em verbetes, lendas e contos. Reconhecer finalidade de reportagens e cartazes. Reconhecer relação de causa e consequência e relação entre pronome e seu referente em tirinhas, contos e reportagens. Inferir elementos da narrativa em fábulas, contos e cartas. Inferir finalidade e efeito de sentido decorrente do uso de pontuação e assunto em fábulas. Inferir informação em poemas, reportagens e cartas. Diferenciar opinião de fato em reportagens. Interpretar efeito de humor e sentido de palavra em piadas e tirinhas.</p> |
| <p>Nível 6</p> <p>Desempenho maior ou igual a 250 e menor que 275</p> | <p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: identificar opinião e informação explícita em fábulas, contos, crônicas e reportagens. Identificar informação explícita em reportagens com ou sem o auxílio de recursos gráficos. Reconhecer a finalidade de verbetes, fábulas, charges e reportagens. Reconhecer relação de causa e consequência e relação entre pronomes e seus referentes em poemas, fábulas e contos. Inferir assunto principal e sentido de expressão em poemas, fábulas, contos, crônicas, reportagens e tirinhas. Inferir informação em contos e reportagens. Inferir efeito de humor e moral em piadas e fábulas.</p>   |

|  |  |
|--|--|
| Nível 7<br>Desempenho maior ou igual a 275 e menor que 300 | Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: identificar assunto principal e informações explícitas em poemas, fábulas e letras de música. Identificar opinião em poemas e crônicas. Reconhecer o gênero textual a partir da comparação entre textos e assunto comum a duas reportagens. Reconhecer elementos da narrativa em fábulas. Reconhecer relação de causa e consequência e relação entre pronomes e seus referentes em fábulas, contos e crônicas. Inferir informação e efeito de sentido decorrente do uso de sinais gráficos em reportagens e em letras de música. Interpretar efeito de humor em piadas e contos. Interpretar linguagem verbal e não verbal em histórias em quadrinhos. |
| Nível 8<br>Desempenho maior ou igual a 300 e menor que 325 | Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: identificar assunto principal e opinião em contos e cartas do leitor. Reconhecer sentido de locução adverbial e elementos da narrativa em fábulas e contos. Reconhecer relação de causa e consequência e relação entre pronomes e seus referentes em fábulas e reportagens. Reconhecer assunto comum entre textos de gêneros diferentes. Inferir informações e efeito de sentido decorrente do uso de pontuação em fábulas e piadas.   |
| Nível 9<br>Desempenho maior ou igual a 325                 | Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: identificar opinião em fábulas e reconhecer sentido de advérbios em cartas do leitor.  |

Fonte: Elaborado pelo autor, extraído de Brasil (2022e)

Quadro B.2 - Escala de proficiência para interpretação dos resultados do 5º ano do ensino fundamental em matemática no saeb – 2019

| Nível  | DESCRIÇÃO DAS HABILIDADES DESENVOLVIDAS   |
|--|---|
| Nível 1<br>Desempenho maior ou igual a 125 e menor que 150 | Os estudantes provavelmente são capazes de: Grandezas e medidas – Determinar a área de figuras desenhadas em malhas quadriculadas por meio de contagem.   |
| Nível 2<br>Desempenho maior ou igual a 150 e menor que 175 | Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: Números e operações; álgebra e funções – Resolver problemas do cotidiano envolvendo adição de pequenas quantias de dinheiro. Tratamento de informações – Localizar informações, relativas ao maior ou menor elemento, em tabelas ou gráficos.   |
| Nível 3<br>Desempenho maior ou igual a 175 e menor que 200 | Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: Espaço e forma – Localizar um ponto ou objeto em uma malha quadriculada ou croqui, a partir de duas coordenadas ou duas ou mais referências. Reconhecer, entre um conjunto de polígonos, aquele que possui o maior número de ângulos. Associar figuras geométricas elementares (quadrado, triângulo e círculo) a seus respectivos nomes. Grandezas e medidas – Converter uma quantia, dada na ordem das unidades de real, em seu equivalente em moedas. Determinar o horário final de um evento a partir de seu horário de início e de um intervalo de tempo dado, todos no formato de horas inteiras. Números e operações; álgebra e funções – Associar a fração $\frac{1}{4}$ a uma de suas representações gráficas. Determinar o resultado da subtração de números representados na forma decimal, tendo como contexto o sistema monetário. Tratamento de informações – Reconhecer o maior valor em uma tabela de dupla entrada cujos dados possuem até duas ordens. Reconhecer informações em um gráfico de colunas duplas. |

|   |   |
|---|---|
| <p>Nível 4</p> <p>Desempenho maior ou igual a 200 e menor que 225</p> | <p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: Espaço e forma – Reconhecer retângulos em meio a outros quadriláteros. Reconhecer a planificação de uma pirâmide entre um conjunto de planificações. Grandezas e medidas – Determinar o total de uma quantia a partir da quantidade de moedas de 25 e/ou 50 centavos que a compõe, ou vice-versa. Determinar a duração de um evento cujos horários inicial e final acontecem em minutos diferentes de uma mesma hora dada. Converter uma hora em minutos. Converter mais de uma semana inteira em dias. Interpretar horas em relógios de ponteiros. Números e operações; álgebra e funções – Determinar o resultado da multiplicação de números naturais por valores do sistema monetário nacional, expressos em números de até duas ordens e efetuar adição posterior. Determinar os termos desconhecidos em uma sequência numérica de múltiplos de cinco. Determinar a adição, com reserva, de até três números naturais com até quatro ordens. Determinar a subtração de números naturais usando a noção de completar. Determinar a multiplicação de um número natural de até três ordens por cinco, com reserva. Determinar a divisão exata por números de um algarismo. Reconhecer o princípio do valor posicional do Sistema de Numeração Decimal. Reconhecer uma fração como representação da relação parte-todo, com o apoio de um conjunto de até cinco figuras. Associar a metade de um total a seu equivalente em porcentagem. Associar um número natural à sua decomposição expressa por extenso. Localizar um número em uma reta numérica graduada onde estão expressos números naturais consecutivos e uma subdivisão equivalente à metade do intervalo entre eles. Tratamento de informações – Reconhecer o maior valor em uma tabela cujos dados possuem até oito ordens. Localizar um dado em tabelas de dupla entrada.</p>   |
| <p>Nível 5</p> <p>Desempenho maior ou igual a 225 e menor que 250</p> | <p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: Espaço e forma – Localizar um ponto entre outros dois fixados, apresentados em uma figura composta por vários outros pontos. Reconhecer a planificação de um cubo entre um conjunto de planificações apresentadas. Grandezas e medidas – Determinar a área de um terreno retangular representado em uma malha quadriculada. Determinar o horário final de um evento a partir do horário de início, dado em horas e minutos, e de um intervalo dado em quantidade de minutos superior a uma hora. Converter mais de uma hora inteira em minutos. Converter uma quantia dada em moedas de 5, 25 e 50 centavos e 1 real em cédulas de real. Estimar a altura de um determinado objeto com referência aos dados fornecidos por uma régua graduada em centímetros. Números e operações; álgebra e funções – Determinar o resultado da subtração, com recursos à ordem superior, entre números naturais de até cinco ordens, utilizando as ideias de retirar e comparar. Determinar o resultado da multiplicação de um número inteiro por um número representado na forma decimal, em contexto envolvendo o sistema monetário. Determinar o resultado da divisão de números naturais, com resto, por um número de uma ordem, usando noção de agrupamento. Resolver problemas envolvendo a análise do algoritmo da adição de dois números naturais. Resolver problemas, no sistema monetário nacional, envolvendo adição e subtração de cédulas e moedas. Resolver problemas que envolvam a metade e o triplo de números naturais. Localizar um número em uma reta numérica graduada onde estão expressos o primeiro e o último número representando um intervalo de tempo de dez anos, com dez subdivisões entre eles. Localizar um número racional dado em sua forma decimal em uma reta numérica graduada onde estão expressos diversos números naturais consecutivos, com dez subdivisões entre eles. Reconhecer o valor posicional do algarismo localizado na 4ª ordem de um número natural. Reconhecer uma fração como representação da relação parte-todo, com apoio de um polígono dividido em oito partes ou mais. Associar um número natural às suas ordens e vice-versa.</p> |
| <p>Nível 6</p> <p>Desempenho maior ou igual a 250 e menor que 275</p> | <p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: Espaço e forma – Reconhecer polígonos presentes em um mosaico composto por diversas formas geométricas. Grandezas e medidas – Determinar a duração de um evento a partir dos horários de início e de término, informados em horas e minutos, sem coincidência nas horas ou nos minutos dos dois horários informados. Converter a duração de um intervalo de tempo, dado em horas e minutos, para minutos. Resolver</p>   |

|   |  |
|---|--|
|   | <p>problemas envolvendo intervalos de tempo em meses, inclusive passando pelo final do ano (outubro a janeiro). Reconhecer que entre quatro ladrilhos apresentados, quanto maior o ladrilho, menor a quantidade necessária para cobrir uma dada região. Reconhecer o m<sup>2</sup> como unidade de medida de área. Números e operações; álgebra e funções – Determinar o resultado da diferença entre dois números racionais representados na forma decimal. Determinar o resultado da multiplicação de um número natural de uma ordem por outro de até três ordens, em contexto que envolve o conceito de proporcionalidade. Determinar o resultado da divisão exata entre dois números naturais, com divisor de até quatro, e dividendo de até quatro ordens. Determinar 50% de um número natural com até três ordens. Determinar porcentagens simples (25%, 50%). Associar a metade de um total a algum equivalente, apresentado como fração ou porcentagem. Associar números naturais à quantidade de agrupamentos de 1.000. Reconhecer uma fração como representação da relação parte-todo, sem apoio de figuras. Localizar números em uma reta numérica graduada onde estão expressos diversos números naturais não consecutivos e crescentes, com uma subdivisão entre eles. Resolver problemas por meio da realização de subtrações e divisões, para determinar o valor das prestações de uma compra a prazo (sem incidência de juros). Resolver problemas que envolvam soma e subtração de valores monetários. Resolver problemas que envolvam a composição e a decomposição polinomial de números naturais de até cinco ordens. Resolver problemas que utilizam a multiplicação envolvendo a noção de proporcionalidade. Reconhecer a modificação sofrida no valor de um número quando um algarismo é alterado. Reconhecer que um número não se altera ao multiplicá-lo por 1. Tratamento de informações – Interpretar dados em uma tabela simples. Comparar dados representados pelas alturas de colunas presentes em um gráfico.</p> |
| <p>Nível 7</p> <p>Desempenho maior ou igual a 275 e menor que 300</p> | <p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: Espaço e forma – Interpretar a movimentação de um objeto utilizando referencial diferente do seu. Reconhecer um cubo a partir de uma de suas planificações desenhadas em uma malha quadriculada. Grandezas e medidas – Determinar o perímetro de um retângulo desenhado em malha quadriculada, com as medidas de comprimento e largura explicitados. Converter medidas dadas em toneladas para quilogramas. Converter uma quantia, dada na ordem das dezenas de real, em moedas de 50 centavos. Estimar o comprimento de um objeto a partir de outro, dado como unidade padrão de medida. Resolver problemas envolvendo conversão de quilograma para grama. Resolver problemas envolvendo conversão de litro para mililitro. Resolver problemas sobre intervalos de tempo que envolvam adição e subtração e que passem pela meia-noite. Números e operações; álgebra e funções – Determinar 25% de um número múltiplo de quatro. Determinar a quantidade de dezenas presentes em um número de quatro ordens. Resolver problemas que envolvam a divisão exata ou a multiplicação de números naturais. Associar números naturais à quantidade de agrupamentos menos usuais, como 300 dezenas. Tratamento de informações – Interpretar dados em gráficos de setores.</p>   |
| <p>Nível 8</p> <p>Desempenho maior ou igual a 300 e menor que 325</p> | <p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: Espaço e forma – Reconhecer uma linha paralela a outra, dada como referência em um mapa. Reconhecer os lados paralelos de um trapézio expressos em forma de segmentos de retas. Reconhecer objetos com a forma esférica entre uma lista de objetos do cotidiano. Grandezas e medidas – Determinar a área de um retângulo desenhado numa malha quadriculada, após a modificação de uma de suas dimensões. Determinar a razão entre as áreas de duas figuras desenhadas numa malha quadriculada. Determinar a área de uma figura poligonal não convexa desenhada sobre uma malha quadriculada. Estimar a diferença de altura entre dois objetos, a partir da altura de um deles. Converter medidas lineares de comprimento (m/cm). Resolver problemas que envolvem a conversão entre diferentes unidades de medida de massa. Números e operações; álgebra e funções – Resolver problemas que envolvam grandezas diretamente proporcionais requerendo mais de uma operação. Resolver problemas envolvendo divisão de números naturais com resto. Associar a fração <math>\frac{1}{2}</math> à sua representação na forma decimal. Associar 50% à sua representação na forma de fração. Associar um número natural de seis ordens à sua forma polinomial. Tratamento de informações –</p>   |

|  |  |
|--|--|
|  | Interpretar dados em um gráfico de colunas duplas.   |
| Nível 9<br>Desempenho maior ou igual a 325 e menor que 350 | Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: Espaço e forma – Reconhecer a planificação de uma caixa cilíndrica. Grandezas e medidas – Determinar o perímetro de um polígono não convexo desenhado sobre as linhas de uma malha quadriculada. Resolver problemas que envolvam a conversão entre unidades de medida de tempo (minutos em horas, meses em anos). Resolver problemas que envolvam a conversão entre unidades de medida de comprimento (metros em centímetros). Números e operações; álgebra e funções – Determinar o minuendo de uma subtração entre números naturais, de três ordens, a partir do conhecimento do subtraendo e da diferença. Determinar o resultado da multiplicação entre o número oito e um número de quatro ordens com reserva. Reconhecer frações equivalentes. Resolver problemas envolvendo multiplicação com significado de combinatória. Comparar números racionais com quantidades diferentes de casas decimais. Tratamento de informações – Reconhecer o gráfico de linhas correspondente a uma sequência de valores ao longo do tempo (com valores positivos e negativos). |
| Nível 10<br>Desempenho maior ou igual a 350                | Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: Espaço e forma – Reconhecer entre um conjunto de quadriláteros, aquele que possui lados perpendiculares e com a mesma medida. Grandezas e medidas – Converter uma medida de comprimento, expressando decímetros e centímetros, para milímetros.  |

Fonte: Elaborado pelo autor, extraído de Brasil (2022e)

### Quadro B.3 - Escala de proficiência para interpretação dos resultados do 9º ano do ensino fundamental em língua portuguesa no saeb – 2019

| Nível  | DESCRIÇÃO DAS HABILIDADES DESENVOLVIDAS   |
|--|---|
| Nível 1<br>Desempenho maior ou igual a 200 e menor que 225 | Os estudantes provavelmente são capazes de reconhecer expressões características da linguagem (científica, jornalística etc.) e a relação entre expressão e seu referente em reportagens e artigos de opinião. Inferir o efeito de sentido de expressão e opinião em crônicas e reportagens.  |
| Nível 2<br>Desempenho maior ou igual a 225 e menor que 250 | Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: localizar informações explícitas em fragmentos de romances e crônicas. Identificar tema e assunto em poemas e charges, relacionando elementos verbais e não verbais. Reconhecer o sentido estabelecido pelo uso de expressões, de pontuação, de conjunções em poemas, charges e fragmentos de romances. Reconhecer relações de causa e consequência e características de personagens em lendas e fábulas. Reconhecer recurso argumentativo em artigos de opinião. Inferir efeito de sentido de repetição de expressões em crônicas.   |
| Nível 3<br>Desempenho maior ou igual a 250 e menor que 275 | Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: localizar informações explícitas em crônicas e fábulas. Identificar os elementos da narrativa em letras de música e fábulas. Reconhecer a finalidade de abaixo-assinado e verbetes. Reconhecer relação entre pronomes e seus referentes e relações de causa e consequência em fragmentos de romances, diários, crônicas, reportagens e máximas (provérbios). Interpretar o sentido de conjunções, de advérbios, e as relações entre elementos verbais e não verbais em tirinhas, fragmentos de romances, reportagens e crônicas. Comparar textos de gêneros diferentes que abordem o mesmo tema. Inferir tema e ideia principal em notícias, crônicas e poemas. Inferir o sentido de palavra ou expressão em história em quadrinhos, poemas e fragmentos de romances. |

|   |   |
|---|---|
| <p>Nível 4</p> <p>Desempenho maior ou igual a 275 e menor que 300</p> | <p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: localizar informações explícitas em artigos de opinião e crônicas. Identificar finalidade e elementos da narrativa em fábulas e contos. Reconhecer opiniões distintas sobre o mesmo assunto em reportagens, contos e enquetes. Reconhecer relações de causa e consequência e relações entre pronomes e seus referentes em fragmentos de romances, fábulas, crônicas, artigos de opinião e reportagens. Reconhecer o sentido de expressão e de variantes linguísticas em letras de música, tirinhas, poemas e fragmentos de romances. Inferir tema, tese e ideia principal em contos, letras de música, editoriais, reportagens, crônicas e artigos. Inferir o efeito de sentido de linguagem verbal e não verbal em charges e história em quadrinhos. Inferir informações em fragmentos de romance. Inferir o efeito de sentido da pontuação e da polissemia como recurso para estabelecer humor ou ironia em tirinhas, anedotas e contos.</p> |
| <p>Nível 5</p> <p>Desempenho maior ou igual a 300 e menor que 325</p> | <p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: localizar a informação principal em reportagens. Identificar ideia principal e finalidade em notícias, reportagens e resenhas. Reconhecer características da linguagem (científica, jornalística etc.) em reportagens. Reconhecer elementos da narrativa em crônicas. Reconhecer argumentos e opiniões em notícias, artigos de opinião e fragmentos de romances. Diferenciar abordagem do mesmo tema em textos de gêneros distintos. Inferir informação em contos, crônicas, notícias e charges. Inferir sentido de palavras, da repetição de palavras, de expressões, de linguagem verbal e não verbal e de pontuação em charges, tirinhas, contos, crônicas e fragmentos de romances.</p>  |
| <p>Nível 6</p> <p>Desempenho maior ou igual a 325 e menor que 350</p> | <p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: Identificar ideia principal e elementos da narrativa em reportagens e crônicas. Identificar argumento em reportagens e crônicas. Reconhecer o efeito de sentido da repetição de expressões e palavras, do uso de pontuação, de variantes linguísticas e de figuras de linguagem em poemas, contos e fragmentos de romances. Reconhecer a relação de causa e consequência em contos. Reconhecer diferentes opiniões entre cartas de leitor que abordam o mesmo tema. Reconhecer a relação de sentido estabelecida por conjunções em crônicas, contos e cordéis. Reconhecer o tema comum entre textos de gêneros distintos. Reconhecer o efeito de sentido decorrente do uso de figuras de linguagem e de recursos gráficos em poemas e fragmentos de romances. Diferenciar fato de opinião em artigos e reportagens. Inferir o efeito de sentido de linguagem verbal e não verbal em tirinhas.</p>  |
| <p>Nível 7</p> <p>Desempenho maior ou igual a 350 e menor que 375</p> | <p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: localizar informações explícitas, ideia principal e expressão que causa humor em contos, crônicas e artigos de opinião. Identificar variantes linguísticas em letras de música. Reconhecer a finalidade e a relação de sentido estabelecida por conjunções em lendas e crônicas.</p>   |
| <p>Nível 8</p> <p>Desempenho maior ou igual a 375</p>                 | <p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: localizar ideia principal em manuais, reportagens, artigos e teses. Identificar os elementos da narrativa em contos e crônicas. Diferenciar fatos de opiniões e opiniões diferentes em artigos e notícias. Inferir o sentido de palavras em poemas.</p>  |

Fonte: Elaborado pelo autor, extraído de Brasil (2022e)

Quadro B.4 - Escala de proficiência para interpretação dos resultados do 9º ano do ensino fundamental em matemática no saeb – 2019

| Nível   | DESCRIÇÃO DAS HABILIDADES DESENVOLVIDAS  |
|---|--|
| <p>Nível 1</p> <p>Desempenho maior ou igual a 200 e menor que 225</p> | <p>Os estudantes provavelmente são capazes de: Números e operações; álgebra e funções – Reconhecer o maior ou o menor número em uma coleção de números racionais, representados na forma decimal. Tratamento de informações – Interpretar dados apresentados em tabela e gráfico de colunas.</p>   |
| <p>Nível 2</p> <p>Desempenho maior ou igual a 225 e menor que 250</p> | <p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: Números e operações; álgebra e funções – Reconhecer a fração que corresponde à relação parte-todo entre uma figura e suas partes hachuradas. Associar um número racional que representa uma quantia monetária, escrito por extenso, à sua representação decimal. Determinar uma fração irredutível, equivalente a uma fração dada, a partir da simplificação por três. Tratamento de informações – Interpretar dados apresentados em um gráfico de linhas simples. Associar dados apresentados em gráfico de colunas a uma tabela.</p>  |
| <p>Nível 3</p> <p>Desempenho maior ou igual a 250 e menor que 275</p> | <p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: Espaço e forma – Reconhecer o ângulo de giro que representa a mudança de direção na movimentação de pessoas/objetos; reconhecer a planificação de um sólido simples, dado por um desenho em perspectiva. Localizar um objeto em representação gráfica do tipo planta baixa, utilizando dois critérios: estar mais longe de um referencial e mais perto de outro. Números e operações; álgebra e funções – Determinar uma fração irredutível, equivalente a uma fração dada, a partir da simplificação por sete; determinar a soma, a diferença, o produto ou o quociente de números inteiros em situações-problema. Localizar o valor que representa um número inteiro positivo associado a um ponto indicado em uma reta numérica. Resolver problemas envolvendo grandezas diretamente proporcionais, representadas por números inteiros. Tratamento de informações – Associar dados apresentados em tabela a gráfico de setores. Analisar dados dispostos em uma tabela simples. Analisar dados apresentados em um gráfico de linhas com mais de uma grandeza representada.</p> |
| <p>Nível 4</p> <p>Desempenho maior ou igual a 275 e menor que 300</p> | <p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: Espaço e forma – Localizar um ponto em um plano cartesiano, com o apoio de malha quadriculada, a partir de suas coordenadas. Reconhecer as coordenadas de um ponto dado em um plano cartesiano, com o apoio de malha quadriculada. Interpretar a movimentação de um objeto utilizando referencial diferente do seu. Grandezas e medidas – Converter unidades de medidas de comprimento, de metros para centímetros, na resolução de situação-problema. Reconhecer que a medida do perímetro de um retângulo, em uma malha quadriculada, dobra ou se reduz à metade quando os lados dobram ou são reduzidos à metade. Números e operações; álgebra e funções – Determinar a soma de números racionais em contextos de sistema monetário. Determinar o valor numérico de uma expressão algébrica de 1º grau envolvendo números naturais, em situação-problema. Localizar números inteiros negativos na reta numérica. Localizar números racionais em sua representação decimal. Tratamento de informações – Analisar dados dispostos em uma tabela de dupla entrada.</p>            |
| <p>Nível 5</p> <p>Desempenho maior ou igual a 300 e menor que 325</p> | <p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: Espaço e forma – Reconhecer que o ângulo não se altera em figuras obtidas por ampliação/redução. Localizar dois ou mais pontos em um sistema de coordenadas. Grandezas e medidas – Determinar o perímetro de uma região retangular, com o apoio de figura, na resolução de uma situação-problema. Determinar o volume mediante contagem de blocos. Números e operações; álgebra e funções – Associar uma fração com denominador dez à sua representação decimal. Associar uma situação-problema à sua linguagem algébrica, por meio de equações do 1º grau ou sistemas lineares.</p>  |



|   |   |
|---|---|
|   | <p>Determinar, em situação-problema, a adição e multiplicação entre números racionais, envolvendo divisão por números inteiros. Determinar a porcentagem envolvendo números inteiros. Resolver problema envolvendo grandezas diretamente proporcionais, representadas por números racionais na forma decimal.</p>   |
| <p>Nível 6</p> <p>Desempenho maior ou igual a 325 e menor que 350</p> | <p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: Espaço e forma – Reconhecer a medida do ângulo determinado entre dois deslocamentos, descritos por meio de orientações dadas por pontos cardeais. Reconhecer as coordenadas de pontos representados no primeiro quadrante de um plano cartesiano. Reconhecer a relação entre as medidas de raio e diâmetro de uma circunferência, com o apoio de figura. Reconhecer a corda de uma circunferência, as faces opostas de um cubo, a partir de uma de suas planificações. Comparar as medidas dos lados de um triângulo com base nas medidas de seus respectivos ângulos opostos. Resolver problema utilizando o Teorema de Pitágoras no cálculo da medida da hipotenusa, dadas as medidas dos catetos. Grandezas e medidas – Converter unidades de medida de massa, de quilograma para grama, na resolução de situação-problema. Resolver problema fazendo uso de semelhança de triângulos. Números e operações; álgebra e funções – Reconhecer frações equivalentes. Associar um número racional, escrito por extenso, à sua representação decimal, e vice-versa. Estimar o valor da raiz quadrada de um número inteiro aproximando-o de um número racional em sua representação decimal. Resolver problema envolvendo grandezas diretamente proporcionais, com constante de proporcionalidade não inteira. Determinar o valor numérico de uma expressão algébrica que contenha parênteses, envolvendo números naturais. Determinar um valor monetário obtido por meio de um desconto ou um acréscimo percentual. Determinar o valor de uma expressão numérica, com números irracionais, fazendo uso de uma aproximação racional fornecida. Tratamento de informações – Resolver problemas que requerem a comparação de dois gráficos de colunas.</p>   |
| <p>Nível 7</p> <p>Desempenho maior ou igual a 350 e menor que 375</p> | <p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: Espaço e forma – Reconhecer ângulos agudos, retos ou obtusos de acordo com sua medida em graus. Reconhecer as coordenadas de pontos representados num plano cartesiano localizados em quadrantes que não sejam o primeiro. Determinar a posição final de um objeto, após a realização de rotações em torno de um ponto, de diferentes ângulos, em sentido horário e anti-horário. Resolver problemas envolvendo ângulos, inclusive utilizando a Lei Angular de Tales sobre a soma dos ângulos internos de um triângulo. Resolver problemas envolvendo as propriedades de ângulos internos e externos de triângulos e quadriláteros, com ou sem justaposição ou sobreposição de figuras. Resolver problema utilizando o Teorema de Pitágoras no cálculo da medida de um dos catetos, dadas as medidas da hipotenusa e de um de seus catetos. Grandezas e medidas – Determinar o perímetro de uma região retangular, obtida pela justaposição de dois retângulos, descritos sem o apoio de figuras. Determinar a área de um retângulo em situações-problema. Determinar a área de regiões poligonais desenhadas em malhas quadriculadas. Determinar o volume de um cubo ou de um paralelepípedo retângulo, sem o apoio de figura. Converter unidades de medida de volume, de m<sup>3</sup> para litro, em situações-problema. Reconhecer a relação entre as áreas de figuras semelhantes. Números e operações; álgebra e funções – Determinar o quociente entre números racionais, representados na forma decimal ou fracionária, em situações-problema. Determinar a soma de números racionais dados na forma fracionária e com denominadores diferentes. Determinar o valor numérico de uma expressão algébrica de 2º grau, com coeficientes naturais, envolvendo números inteiros. Determinar o valor de uma expressão numérica envolvendo adição, subtração, multiplicação e/ou potenciação entre números inteiros. Determinar o valor de uma expressão numérica com números inteiros positivos e negativos; determinar o valor de uma expressão numérica com números racionais. Comparar números racionais com diferentes números de casas decimais, usando arredondamento. Localizar na reta numérica um número racional, representado na forma de uma fração imprópria. Associar uma fração à sua representação na forma decimal. Associar uma situação-problema à sua linguagem algébrica, por meio de inequações do 1º grau. Associar a representação gráfica de duas</p> |

|  |   |
|--|---|
|  | retas no plano cartesiano a um sistema de duas equações lineares e vice-versa. Resolver problemas envolvendo equação de 2º grau. Tratamento de informações – Determinar a média aritmética de um conjunto de valores. Estimar quantidades em gráficos de setores. Analisar dados dispostos em uma tabela de três ou mais entradas. Interpretar dados fornecidos em gráficos envolvendo regiões do plano cartesiano. Interpretar gráficos de linhas com duas sequências de valores.  |
| Nível 8<br><br>Desempenho maior ou igual a 375 e menor que 400 | Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: Espaço e forma – Resolver problemas utilizando as propriedades das cevianas (altura, mediana e bissetriz) de um triângulo isósceles, com o apoio de figura. Grandezas e medidas – Converter unidades de medida de capacidade, de mililitro para litro, em situações-problema. Reconhecer que a área de um retângulo quadruplica quando seus lados dobram. Determinar a área de figuras simples (triângulo, paralelogramo, trapézio), inclusive utilizando composição/decomposição. Números e operações; álgebra e funções – Determinar o valor numérico de uma expressão algébrica de 1º grau, com coeficientes racionais, representados na forma decimal. Determinar o valor de uma expressão numérica envolvendo adição, subtração e potenciação entre números racionais, representados na forma decimal. Resolver problemas envolvendo grandezas inversamente proporcionais. |
| Nível 9<br><br>Desempenho maior ou igual a 400                 | Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: Espaço e forma – Resolver problemas utilizando a soma das medidas dos ângulos internos de um polígono. Números e operações; álgebra e funções – Reconhecer a expressão algébrica que expressa uma regularidade existente em uma sequência de números ou de figuras geométricas.   |

Fonte: Elaborado pelo autor, extraído de Brasil (2022e)

## Apêndice C - Lista de Pacotes R e Respektivas Versões

---

```

— Session info —
Setting value
Version R version 4.2.2 Patched (2022-11-10 r83330)
os Ubuntu 20.04.5 LTS
System x86_64, linux-gnu
ui RStudio
Language (EN)
Collate en_US.UTF-8
Ctype pt_BR.UTF-8
tz America/Sao_Paulo
Date 2022-12-08
rstudio 2022.07.2+576 Spotted Wakerobin (desktop)
pandoc 2.19.2 @ /usr/lib/rstudio/bin/quarto/bin/tools/ (via rmarkdown)

```

---

```

— Packages —
package      * version      date (UTC)    lib          source
abind         1.4-5         2016-07-21   [2]          CRAN (R 4.1.3)
assertthat   0.2.1         2019-03-21   [2]          CRAN (R 4.1.3)
backports    1.4.1         2021-12-13   [2]          CRAN (R 4.1.3)
base64enc    0.1-3         2015-07-28   [2]          CRAN (R 4.1.3)
BiocGenerics 0.42.0        2022-04-26   [2]          Bioconductor
bitops       1.0-7         2021-04-24   [2]          CRAN (R 4.2.1)
broom        1.0.1         2022-08-29   [2]          CRAN (R 4.2.1)
car           3.1-0         2022-06-15   [2]          CRAN (R 4.2.1)
carData      3.0-5         2022-01-06   [2]          CRAN (R 4.1.3)
cellranger   1.1.0         2016-07-27   [2]          CRAN (R 4.1.3)
circlize     * 0.4.15       2022-05-10   [2]          CRAN (R 4.2.1)
cli           3.4.0         2022-09-08   [2]          CRAN (R 4.2.1)
clue         0.3-61        2022-05-30   [2]          CRAN (R 4.2.1)
cluster      * 2.1.4         2022-08-22   [4]          CRAN (R 4.2.1)
codetools    0.2-18        2020-11-04   [4]          CRAN (R 4.0.3)
colorspace   2.0-3         2022-02-21   [2]          CRAN (R 4.1.3)
ComplexHeatmap 2.12.1        2022-08-09   [2]          Bioconductor
corrplot     * 0.92          2021-11-18   [2]          CRAN (R 4.1.3)
cowplot      1.1.1         2020-12-30   [2]          CRAN (R 4.1.3)
crayon       1.5.1         2022-03-26   [2]          CRAN (R 4.2.1)
data.table   1.14.2        2021-09-27   [2]          CRAN (R 4.1.3)
DBI           1.1.3         2022-06-18   [2]          CRAN (R 4.2.1)
dbscan       * 1.1-10        2022-01-15   [2]          CRAN (R 4.2.0)
dendextend   * 1.16.0        2022-07-04   [2]          CRAN (R 4.2.1)
DEoptimR     1.0-11        2022-04-03   [2]          CRAN (R 4.2.1)
digest       0.6.29        2021-12-01   [2]          CRAN (R 4.1.3)
doMC         * 1.3.8         2022-02-05   [2]          CRAN (R 4.1.3)
doParallel   1.0.17        2022-02-07   [2]          CRAN (R 4.2.0)
dplyr        * 1.0.10        2022-09-01   [2]          CRAN (R 4.2.1)
DT           0.25          2022-09-12   [2]          CRAN (R 4.2.1)
ellipsis     0.3.2         2021-04-29   [2]          CRAN (R 4.1.3)
emmeans      1.8.1-1       2022-09-08   [2]          CRAN (R 4.2.1)
estimability 1.4.1         2022-08-05   [2]          CRAN (R 4.2.1)
evaluate     0.16          2022-08-09   [2]          CRAN (R 4.2.1)
factoextra   * 1.0.7         2020-04-01   [2]          CRAN (R 4.1.3)
FactoMineR   * 2.6           2022-09-09   [2]          CRAN (R 4.2.1)
FactorAssumptions * 2.0.1         2022-03-08   [2]          CRAN (R 4.2.0)
fans         1.0.3         2022-03-24   [2]          CRAN (R 4.2.1)
farver       2.1.1         2022-07-06   [2]          CRAN (R 4.2.1)

```

|               |            |            |     |                                       |
|---------------|------------|------------|-----|---------------------------------------|
| fastmap       | 1.1.0      | 2021-01-25 | [2] | CRAN (R 4.1.3)                        |
| flashClust    | 1.01-2     | 2012-08-21 | [2] | CRAN (R 4.1.3)                        |
| forcats       | * 0.5.2    | 2022-08-19 | [2] | CRAN (R 4.2.1)                        |
| foreach       | * 1.5.2    | 2022-02-02 | [2] | CRAN (R 4.1.3)                        |
| formattable   | * 0.2.1    | 2021-01-07 | [2] | CRAN (R 4.2.1)                        |
| gapminder     | * 0.3.0    | 2017-10-31 | [2] | CRAN (R 4.1.3)                        |
| gcimputeR     | * 0.1.1    | 2022-10-24 | [2] | Github (udellgroup/gcimputeR@18d97e0) |
| generics      | 0.1.3      | 2022-07-05 | [2] | CRAN (R 4.2.1)                        |
| GetoptLong    | 1.0.5      | 2020-12-15 | [2] | CRAN (R 4.2.1)                        |
| GGally        | * 2.1.2    | 2021-06-21 | [2] | CRAN (R 4.1.3)                        |
| gganimate     | * 1.0.8    | 2022-09-08 | [2] | CRAN (R 4.2.1)                        |
| ggcorrplot    | * 0.1.3    | 2019-05-19 | [2] | CRAN (R 4.1.3)                        |
| ggplot2       | * 3.3.6    | 2022-05-03 | [2] | CRAN (R 4.2.1)                        |
| ggpubr        | * 0.4.0    | 2020-06-27 | [2] | CRAN (R 4.1.3)                        |
| ggrepel       | * 0.9.1    | 2021-01-15 | [2] | CRAN (R 4.1.3)                        |
| ggsignif      | 0.6.3      | 2021-09-09 | [2] | CRAN (R 4.1.3)                        |
| gifski        | * 1.6.6-1  | 2022-04-05 | [2] | CRAN (R 4.2.1)                        |
| glmnet        | * 4.1-4    | 2022-04-15 | [2] | CRAN (R 4.2.1)                        |
| GlobalOptions | 0.1.2      | 2020-06-10 | [2] | CRAN (R 4.2.1)                        |
| glue          | 1.6.2      | 2022-02-24 | [2] | CRAN (R 4.1.3)                        |
| goftest       | 1.2-3      | 2021-10-07 | [2] | CRAN (R 4.2.1)                        |
| gridExtra     | 2.3        | 2017-09-09 | [2] | CRAN (R 4.1.3)                        |
| gtable        | 0.3.1      | 2022-09-01 | [2] | CRAN (R 4.2.1)                        |
| h2o           | * 3.38.0.1 | 2022-09-23 | [2] | CRAN (R 4.2.1)                        |
| hms           | 1.1.2      | 2022-08-19 | [2] | CRAN (R 4.2.1)                        |
| htmltools     | 0.5.3      | 2022-07-18 | [2] | CRAN (R 4.2.1)                        |
| htmlwidgets   | 1.5.4      | 2021-09-08 | [2] | CRAN (R 4.1.3)                        |
| httr          | 1.4.4      | 2022-08-17 | [2] | CRAN (R 4.2.1)                        |
| IRanges       | 2.30.1     | 2022-08-18 | [2] | Bioconductor                          |
| iterators     | * 1.0.14   | 2022-02-05 | [2] | CRAN (R 4.1.3)                        |
| jsonlite      | 1.8.0      | 2022-02-22 | [2] | CRAN (R 4.1.3)                        |
| kableExtra    | * 1.3.4    | 2021-02-20 | [2] | CRAN (R 4.2.1)                        |
| knitr         | 1.40       | 2022-08-24 | [2] | CRAN (R 4.2.1)                        |
| labeling      | 0.4.2      | 2020-10-20 | [2] | CRAN (R 4.1.3)                        |
| lattice       | 0.20-45    | 2021-09-22 | [4] | CRAN (R 4.2.0)                        |
| lazyeval      | 0.2.2      | 2019-03-15 | [2] | CRAN (R 4.1.3)                        |
| leaps         | 3.1        | 2020-01-16 | [2] | CRAN (R 4.1.3)                        |
| lifecycle     | 1.0.2      | 2022-09-09 | [2] | CRAN (R 4.2.1)                        |
| magick        | * 2.7.3    | 2021-08-18 | [2] | CRAN (R 4.2.1)                        |
| magrittr      | * 2.0.3    | 2022-03-30 | [2] | CRAN (R 4.2.1)                        |
| MASS          | * 7.3-58.1 | 2022-08-03 | [4] | CRAN (R 4.2.1)                        |
| matlib        | * 0.9.5    | 2021-08-21 | [2] | CRAN (R 4.2.1)                        |
| Matrix        | * 1.5-1    | 2022-09-13 | [2] | CRAN (R 4.2.1)                        |
| matrixStats   | 0.62.0     | 2022-04-19 | [2] | CRAN (R 4.2.1)                        |
| mclust        | * 5.4.10   | 2022-05-20 | [2] | CRAN (R 4.2.1)                        |
| mice          | * 3.14.0   | 2021-11-24 | [2] | CRAN (R 4.2.1)                        |
| missMDA       | * 1.18     | 2020-12-11 | [2] | CRAN (R 4.2.1)                        |
| mnormt        | 2.1.0      | 2022-06-07 | [2] | CRAN (R 4.2.1)                        |
| multcompView  | 0.1-8      | 2019-12-19 | [2] | CRAN (R 4.2.1)                        |
| multidplyr    | * 0.1.1    | 2021-12-01 | [2] | CRAN (R 4.1.3)                        |
| munsell       | 0.5.0      | 2018-06-12 | [2] | CRAN (R 4.1.3)                        |
| mvtnorm       | 1.1-3      | 2021-10-08 | [2] | CRAN (R 4.2.1)                        |
| nlme          | 3.1-160    | 2022-10-10 | [4] | CRAN (R 4.2.1)                        |
| nortest       | 1.0-4      | 2015-07-30 | [2] | CRAN (R 4.2.1)                        |
| olsrr         | * 0.5.3    | 2020-02-10 | [2] | CRAN (R 4.2.1)                        |
| pheatmap      | * 1.0.12   | 2019-01-04 | [2] | CRAN (R 4.2.1)                        |
| pillar        | 1.8.1      | 2022-08-19 | [2] | CRAN (R 4.2.1)                        |
| pkgconfig     | 2.0.3      | 2019-09-22 | [2] | CRAN (R 4.1.3)                        |

|               |          |            |     |                |
|---------------|----------|------------|-----|----------------|
| plotly        | * 4.10.0 | 2021-10-09 | [2] | CRAN (R 4.1.3) |
| plyr          | 1.8.7    | 2022-03-24 | [2] | CRAN (R 4.2.1) |
| png           | 0.1-7    | 2013-12-03 | [2] | CRAN (R 4.2.1) |
| pracma        | * 2.3.8  | 2022-03-04 | [2] | CRAN (R 4.2.0) |
| prettyunits   | 1.1.1    | 2020-01-24 | [2] | CRAN (R 4.1.3) |
| progress      | 1.2.2    | 2019-05-16 | [2] | CRAN (R 4.1.3) |
| pryr          | * 0.1.5  | 2021-07-26 | [2] | CRAN (R 4.1.3) |
| psych         | * 2.2.5  | 2022-05-10 | [2] | CRAN (R 4.2.0) |
| purrr         | * 0.3.5  | 2022-10-06 | [2] | CRAN (R 4.2.1) |
| pvcclust      | * 2.2-0  | 2019-11-19 | [2] | CRAN (R 4.2.1) |
| qqplotr       | * 0.0.5  | 2021-04-23 | [2] | CRAN (R 4.2.1) |
| R6            | 2.5.1    | 2021-08-19 | [2] | CRAN (R 4.1.3) |
| RColorBrewer  | 1.1-3    | 2022-04-03 | [2] | CRAN (R 4.2.1) |
| Rcpp          | 1.0.9    | 2022-07-08 | [2] | CRAN (R 4.2.1) |
| RCurl         | 1.98-1.9 | 2022-10-03 | [2] | CRAN (R 4.2.1) |
| readr         | * 2.1.2  | 2022-01-30 | [2] | CRAN (R 4.1.3) |
| readxl        | * 1.4.1  | 2022-08-17 | [2] | CRAN (R 4.2.1) |
| registry      | 0.5-1    | 2019-03-05 | [2] | CRAN (R 4.2.1) |
| reshape       | 0.8.9    | 2022-04-12 | [2] | CRAN (R 4.2.1) |
| rgl           | 0.109.6  | 2022-07-08 | [2] | CRAN (R 4.2.1) |
| rjson         | 0.2.21   | 2022-01-09 | [2] | CRAN (R 4.2.1) |
| rlang         | 1.0.6    | 2022-09-24 | [2] | CRAN (R 4.2.1) |
| rmarkdown     | 2.16     | 2022-08-24 | [2] | CRAN (R 4.2.1) |
| robustbase    | 0.95-0   | 2022-04-02 | [2] | CRAN (R 4.2.1) |
| rstatix       | 0.7.0    | 2021-02-13 | [2] | CRAN (R 4.1.3) |
| rstudioapi    | 0.14     | 2022-08-22 | [2] | CRAN (R 4.2.1) |
| rvest         | 1.0.3    | 2022-08-19 | [2] | CRAN (R 4.2.1) |
| S4Vectors     | 0.34.0   | 2022-04-26 | [2] | Bioconductor   |
| scales        | 1.2.1    | 2022-08-20 | [2] | CRAN (R 4.2.1) |
| scatterplot3d | 0.3-42   | 2022-09-08 | [2] | CRAN (R 4.2.1) |
| Seriation     | * 1.3.6  | 2022-07-16 | [2] | CRAN (R 4.2.1) |
| sessioninfo   | 1.2.2    | 2021-12-06 | [2] | CRAN (R 4.2.1) |
| shape         | 1.4.6    | 2021-05-19 | [2] | CRAN (R 4.2.1) |
| stringi       | 1.7.8    | 2022-07-11 | [2] | CRAN (R 4.2.1) |
| stringr       | * 1.4.1  | 2022-08-20 | [2] | CRAN (R 4.2.1) |
| survival      | 3.4-0    | 2022-08-09 | [4] | CRAN (R 4.2.1) |
| svglite       | 2.1.0    | 2022-02-03 | [2] | CRAN (R 4.2.1) |
| systemfonts   | 1.0.4    | 2022-02-11 | [2] | CRAN (R 4.2.1) |
| tibble        | * 3.1.8  | 2022-07-22 | [2] | CRAN (R 4.2.1) |
| tidyr         | * 1.2.1  | 2022-09-08 | [2] | CRAN (R 4.2.1) |
| tidyselect    | * 1.1.2  | 2022-02-21 | [2] | CRAN (R 4.1.3) |
| TSP           | 1.2-1    | 2022-07-14 | [2] | CRAN (R 4.2.1) |
| tweenr        | 2.0.2    | 2022-09-06 | [2] | CRAN (R 4.2.1) |
| tzdb          | 0.3.0    | 2022-03-28 | [2] | CRAN (R 4.2.1) |
| utf8          | 1.2.2    | 2021-07-24 | [2] | CRAN (R 4.1.3) |
| vctrs         | 0.4.1    | 2022-04-13 | [2] | CRAN (R 4.2.1) |
| viridis       | 0.6.2    | 2021-10-13 | [2] | CRAN (R 4.2.1) |
| viridisLite   | 0.4.1    | 2022-08-22 | [2] | CRAN (R 4.2.1) |
| webshot       | 0.5.4    | 2022-09-26 | [2] | CRAN (R 4.2.1) |
| withr         | 2.5.0    | 2022-03-03 | [2] | CRAN (R 4.1.3) |
| xfun          | 0.33     | 2022-09-12 | [2] | CRAN (R 4.2.1) |
| xml2          | 1.3.3    | 2021-11-30 | [2] | CRAN (R 4.1.3) |
| xtable        | 1.8-4    | 2019-04-21 | [2] | CRAN (R 4.2.1) |

[2] /usr/local/lib/R/site-library

[3] /usr/lib/R/site-library

[4] /usr/lib/R/library

## Apêndice D - Descrição de variáveis Educacenso

| Nº | Nome da Variável               | Descrição da Variável  |
|----|--------------------------------|--|
| 1  | NU_ANO_CENSO                   | Ano do Censo   |
| 2  | CO_ENTIDADE                    | Código da Escola   |
| 3  | NO_ENTIDADE                    | Nome da Escola   |
| 4  | CO_ORGAO_REGIONAL              | Código do Órgão Regional de Ensino   |
| 5  | GEO_REF_LATITUDE               | Georreferenciamento da Escola  |
| 6  | GEO_REF_LONGITUDE              | Georreferenciamento da Escola  |
| 7  | TP_SITUACAO_FUNCIONAMENTO      | Situação de funcionamento  |
| 8  | DT_ANO_LETIVO_INICIO           | Início do ano letivo   |
| 9  | DT_ANO_LETIVO_TERMINO          | Término (Previsão) do ano letivo   |
| 10 | CO_REGIAO                      | Código da região geográfica  |
| 11 | CO_MESORREGIAO                 | Código da mesorregião  |
| 12 | CO_MICRORREGIAO                | Código da microrregião   |
| 13 | CO_UF                          | Código da UF   |
| 14 | CO_MUNICIPIO                   | Código do Município  |
| 15 | CO_DISTRITO                    | Código completo do Distrito da escola  |
| 16 | TP_DEPENDENCIA                 | Dependência Administrativa   |
| 17 | TP_LOCALIZACAO                 | Localização  |
| 18 | TP_LOCALIZACAO_DIFERENCIADA    | Localização diferenciada da escola   |
| 19 | IN_VINCULO_SECRETARIA_EDUCACAO | Órgão que a escola pública está vinculada - Secretaria de Educação/Ministério da Educação          |
| 20 | IN_VINCULO_SEGURANCA_PUBLICA   | Órgão que a escola pública está vinculada - Secretaria de Segurança Pública/Forças Armadas/Militar |
| 21 | IN_VINCULO_SECRETARIA_SAUDE    | Órgão que a escola pública está vinculada - Secretaria de Saúde/Ministério da Saúde                |
| 22 | IN_VINCULO_OUTRO_ORGAO         | Órgão que a escola pública está vinculada - Outro órgão da administração pública                   |
| 23 | TP_CATEGORIA_ESCOLA_PRIVADA    | Categoria da escola privada  |
| 24 | IN_CONVENIADA_PP               | Conveniada com o poder público   |
| 25 | TP_CONVENIO_PODER_PUBLICO      | Dependência do convênio com o poder público  |
| 26 | IN_MANT_ESCOLA_PRIVADA         | Mantenedora da escola privada - Empresa ou grupo empresarial do setor privado ou pessoa física     |

|    |                                  |  |
|----|----------------------------------|--|
|    | DA_EMP                           |  |
| 27 | IN_MANT_ESCOLA_PRIVADA_ONG       | Mantenedora da escola privada - Organização não governamental (ONG) - internacional ou nacional.   |
| 28 | IN_MANT_ESCOLA_PRIVADA_OSCIP     | Mantenedora da escola privada - Organização da Sociedade Civil de Interesse Público (Oscip)  |
| 29 | IN_MANT_ESCOLA_PRIVADA_ONG_OSCIP | Mantenedora da escola privada - Organização não governamental (ONG) - internacional ou nacional. Organização da Sociedade Civil de Interesse Público (Oscip) |
| 30 | IN_MANT_ESCOLA_PRIVADA_SIND      | Mantenedora da escola privada - Sindicatos de trabalhadores ou patronais, associações e cooperativas   |
| 31 | IN_MANT_ESCOLA_PRIVADA_SIST_S    | Mantenedora da escola privada - Sistema S (Sesi, Senai, Sesc, Outros)  |
| 32 | IN_MANT_ESCOLA_PRIVADA_S_FINS    | Mantenedora da escola privada - Instituições sem fins lucrativos   |
| 33 | NU_CNPJ_ESCOLA_PRIVADA           | Número do CNPJ da escola privada   |
| 34 | NU_CNPJ_MANTENEDORA              | Número do CNPJ da mantenedora principal da escola privada  |
| 35 | TP_REGULAMENTACAO                | Regulamentação/Autorização no conselho ou órgão municipal, estadual ou federal de educação   |
| 36 | TP_RESPONSAVEL_REGULAMENTACAO    | Esfera administrativa do conselho ou órgão responsável pela Regulamentação/Autorização   |
| 37 | CO_ESCOLA_SEDE_VINCULADA         | Código da escola sede  |
| 38 | CO_IES_OFERTANTE                 | Código da IES vinculada à escola   |
| 39 | IN_LOCAL_FUNC_PREDIO_ESCOLAR     | Local de funcionamento da escola - Prédio Escolar  |
| 40 | TP_OCUPACAO_PREDIO_ESCOLAR       | Forma de ocupação do Prédio escolar  |
| 41 | IN_LOCAL_FUNC_SALAS_EMPRESA      | Local de funcionamento da escola - Salas de empresa  |
| 42 | IN_LOCAL_FUNC_SOCIOEDUCATIVO     | Local de funcionamento da escola - Unidade de Atendimento socioeducativo   |
| 43 | IN_LOCAL_FUNC_UNID_PRISIONAL     | Local de funcionamento da escola - Unidade Prisional   |
| 44 | IN_LOCAL_FUNC_PRISIONAL_SOCIO    | Local de funcionamento da escola - Unidade Prisional ou Unidade de atendimento socioeducativo  |
| 45 | IN_LOCAL_FUNC_TEMPLO_IGREJA      | Local de funcionamento da escola - Templo/Igreja   |
| 46 | IN_LOCAL_FUNC_CASA_PROFESSOR     | Local de funcionamento da escola - Casa do professor   |
| 47 | IN_LOCAL_FUNC_GALPAO             | Local de funcionamento da escola - Galpão/Rancho/Paiol/Barracão  |
| 48 | TP_OCUPACAO_GALPAO               | Forma de ocupação do Galpão/Rancho/Paiol/Barracão  |
| 49 | IN_LOCAL_FUNC_SALAS_OUTRA_ESC    | Local de funcionamento da escola - Salas em outra escola   |

|    |                               |   |
|----|-------------------------------|---|
| 50 | IN_LOCAL_FUNC_OUTROS          | Local de funcionamento da escola - Outros   |
| 51 | IN_PREDIO_COMPARTILHADO       | Prédio compartilhado com outra escola   |
| 52 | IN_AGUA_FILTRADA              | Água consumida pelos alunos   |
| 53 | IN_AGUA_POTAVEL               | Fornece água potável para o consumo humano  |
| 54 | IN_AGUA_REDE_PUBLICA          | Abastecimento de água - Rede pública  |
| 55 | IN_AGUA_POCO_ARTESIANO        | Abastecimento de água - Poço artesiano  |
| 56 | IN_AGUA_CACIMBA               | Abastecimento de água - Cacimba/Cisterna/Poço   |
| 57 | IN_AGUA_FONTE_RIO             | Abastecimento de água - Fonte/Rio/Igarapé/Riacho/Córrego  |
| 58 | IN_AGUA_INEXISTENTE           | Abastecimento de água - Não há abastecimento de água  |
| 59 | IN_ENERGIA_REDE_PUBLICA       | Abastecimento de energia elétrica - Rede pública  |
| 60 | IN_ENERGIA_GERADOR            | Abastecimento de energia elétrica - Gerador   |
| 61 | IN_ENERGIA_GERADOR_FOSSIL     | Abastecimento de energia elétrica - Gerador movido a combustível fóssil   |
| 62 | IN_ENERGIA_OUTROS             | Abastecimento de energia elétrica - Outros (Energia alternativa)  |
| 63 | IN_ENERGIA_RENOVAVEL          | Abastecimento de energia elétrica - Fontes de energia renováveis ou alternativas (gerador a biocombustível e/ou biodigestores, eólica, solar, outras) |
| 64 | IN_ENERGIA_INEXISTENTE        | Abastecimento de energia elétrica - Não há energia elétrica   |
| 65 | IN_ESGOTO_REDE_PUBLICA        | Esgoto sanitário - Rede pública   |
| 66 | IN_ESGOTO_FOSSA_SEPTICA       | Esgoto sanitário - Fossa Séptica  |
| 67 | IN_ESGOTO_FOSSA_COMUM         | Esgoto sanitário - Fossa rudimentar/comum   |
| 68 | IN_ESGOTO_FOSSA               | Esgoto sanitário - Fossa  |
| 69 | IN_ESGOTO_INEXISTENTE         | Esgoto sanitário - Não há esgotamento sanitário   |
| 70 | IN_LIXO_SERVICO_COLETA        | Destinação do lixo - Serviço de coleta  |
| 71 | IN_LIXO_QUEIMA                | Destinação do lixo - Queima   |
| 72 | IN_LIXO_ENTERRA               | Destinação do lixo - Enterra  |
| 73 | IN_LIXO_DESTINO_FINAL_PUBLICO | Destinação do lixo - Leva a uma destinação final financiada pelo poder público  |
| 74 | IN_LIXO_DESCARTA_OUTRA_AREA   | Destinação do lixo - Descarta em outra área   |
| 75 | IN_LIXO_JOGA_OUTRA_AREA       | Destinação do lixo - Joga em outra área   |



|     |                                    |  |
|-----|------------------------------------|--|
| 76  | IN_LIXO_OUTROS                     | Destinação do lixo - Outros  |
| 77  | IN_LIXO_RECICLA                    | Tratamento do lixo/resíduos que a escola realiza - Reciclagem  |
| 78  | IN_TRATAMENTO_LIXO_S<br>EPARACAO   | Tratamento do lixo/resíduos que a escola realiza - Separação do lixo/resíduos  |
| 79  | IN_TRATAMENTO_LIXO_R<br>EUTILIZA   | Tratamento do lixo/resíduos que a escola realiza - Reaproveitamento/reutilização   |
| 80  | IN_TRATAMENTO_LIXO_R<br>ECICLAGEM  | Tratamento do lixo/resíduos que a escola realiza - Reciclagem  |
| 81  | IN_TRATAMENTO_LIXO_I<br>NEXISTENTE | Tratamento do lixo/resíduos que a escola realiza - Não faz tratamento  |
| 82  | IN_ALMOXARIFADO                    | Dependências físicas existentes e utilizadas na escola - Almojarifado  |
| 83  | IN_AREA_VERDE                      | Dependências físicas existentes e utilizadas na escola - Área Verde  |
| 84  | IN_AUDITORIO                       | Dependências físicas existentes e utilizadas na escola - Auditório   |
| 85  | IN_BANHEIRO_FORA_PRE<br>DIO        | Dependências físicas existentes e utilizadas na escola - Banheiro fora do prédio   |
| 86  | IN_BANHEIRO_DENTRO_P<br>REDIO      | Dependências físicas existentes e utilizadas na escola - Banheiro dentro do prédio   |
| 87  | IN_BANHEIRO                        | Dependências físicas existentes e utilizadas na escola - Banheiro  |
| 88  | IN_BANHEIRO_EI                     | Dependências físicas existentes e utilizadas na escola - Banheiro adequado à educação infantil   |
| 89  | IN_BANHEIRO_PNE                    | Dependências físicas existentes e utilizadas na escola - Banheiro acessível, adequado ao uso de pessoas com deficiência ou mobilidade reduzida |
| 90  | IN_BANHEIRO_FUNCIONA<br>RIOS       | Dependências físicas existentes e utilizadas na escola - Banheiro exclusivo para os funcionários   |
| 91  | IN_BANHEIRO_CHUVEIRO               | Dependências físicas existentes e utilizadas na escola - Banheiro ou vestiário com chuveiro  |
| 92  | IN_BERCARIO                        | Dependências físicas existentes e utilizadas na escola - Berçário  |
| 93  | IN_BIBLIOTECA                      | Dependências físicas existentes e utilizadas na escola - Biblioteca  |
| 94  | IN_BIBLIOTECA_SALA_LE<br>ITURA     | Dependências físicas existentes e utilizadas na escola - Biblioteca e/ou Sala de leitura   |
| 95  | IN_COZINHA                         | Dependências físicas existentes e utilizadas na escola - Cozinha   |
| 96  | IN_DESPENSA                        | Dependências físicas existentes e utilizadas na escola - Laboratório de ciências   |
| 97  | IN_DORMITORIO_ALUNO                | Dependências físicas existentes e utilizadas na escola - Laboratório de informática  |
| 98  | IN_DORMITORIO_PROFES<br>SOR        | Dependências físicas existentes e utilizadas na escola - Dormitório de professor (a)   |
| 99  | IN_LABORATORIO_CIENCI<br>AS        | Dependências físicas existentes e utilizadas na escola - Laboratório de ciências   |
| 100 | IN_LABORATORIO_INFOR<br>MÁTICA     | Dependências físicas existentes e utilizadas na escola - Laboratório de informática  |
| 101 | IN_PATIO_COBERTO                   | Dependências físicas existentes e utilizadas na escola - Pátio Coberto   |
| 102 | IN_PATIO_DESCOBERTO                | Dependências físicas existentes e utilizadas na escola - Pátio Descoberto  |

|     |                                |   |
|-----|--------------------------------|---|
| 103 | IN_PARQUE_INFANTIL             | Dependências físicas existentes e utilizadas na escola - Parque infantil  |
| 104 | IN_PISCINA                     | Dependências físicas existentes e utilizadas na escola - Piscina  |
| 105 | IN_QUADRA_ESPORTES             | Dependências físicas existentes e utilizadas na escola - Quadra de esportes coberta ou descoberta   |
| 106 | IN_QUADRA_ESPORTES_COBERTA     | Dependências físicas existentes e utilizadas na escola - Quadra de esportes Coberta   |
| 107 | IN_QUADRA_ESPORTES_DESCOBERTA  | Dependências físicas existentes e utilizadas na escola - Quadra de esportes Descoberta  |
| 108 | IN_REFEITORIO                  | Dependências físicas existentes e utilizadas na escola - Refeitório   |
| 109 | IN_SALA_ATELIE_ARTES           | Dependências físicas existentes e utilizadas na escola - Sala/ateliê de artes   |
| 110 | IN_SALA_MUSICA_CORAL           | Dependências físicas existentes e utilizadas na escola - Sala de música/coral   |
| 111 | IN_SALA_ESTUDIO_DANCA          | Dependências físicas existentes e utilizadas na escola - Sala/estúdio de dança  |
| 112 | IN_SALA_MULTIUSSO              | Dependências físicas existentes e utilizadas na escola - Sala multiuso (música, dança e artes)  |
| 113 | IN_SALA_DIRETORIA              | Dependências físicas existentes e utilizadas na escola - Sala de Diretoria  |
| 114 | IN_SALA_LEITURA                | Dependências físicas existentes e utilizadas na escola - Sala de Leitura  |
| 115 | IN_SALA_PROFESSOR              | Dependências físicas existentes e utilizadas na escola - Sala de professores  |
| 116 | IN_SALA_REPOUSSO_ALUNO         | Dependências físicas existentes e utilizadas na escola - Sala de repouso para aluno(a)  |
| 117 | IN_SECRETARIA                  | Dependências físicas existentes e utilizadas na escola - Sala de secretaria   |
| 118 | IN_SALA_ATENDIMENTO_ESPECIAL   | Dependências físicas existentes e utilizadas na escola - Sala de recursos Multifuncionais para Atendimento Educacional Especializado (AEE)                        |
| 119 | IN_TERREIRAO                   | Dependências físicas existentes e utilizadas na escola - Terreirão (área para prática desportiva e recreação sem cobertura, sem piso e sem edificações)           |
| 120 | IN_VIVEIRO                     | Dependências físicas existentes e utilizadas na escola - Viveiro/criação de animais   |
| 121 | IN_DEPENDENCIAS_PNE            | Dependências físicas existentes e utilizadas na escola - Dependências e vias adequadas a alunos com deficiência ou mobilidade reduzida                            |
| 122 | IN_LAVANDERIA                  | Dependências físicas existentes e utilizadas na escola - Lavanderia   |
| 123 | IN_DEPENDENCIAS_OUTRAS         | Dependências existentes na escola - Nenhuma das dependência relacionadas  |
| 124 | IN_ACESSIBILIDADE_CORRIMAO     | Recursos de acessibilidade para pessoas com deficiência ou mobilidade reduzida nas vias de circulação interna na escola - Corrimão e guarda corpos                |
| 125 | IN_ACESSIBILIDADE_ELEVADOR     | Recursos de acessibilidade para pessoas com deficiência ou mobilidade reduzida nas vias de circulação interna na escola - Elevador                                |
| 126 | IN_ACESSIBILIDADE_PISOS_TATEIS | Recursos de acessibilidade para pessoas com deficiência ou mobilidade reduzida nas vias de circulação interna na escola - Pisos táteis                            |
| 127 | IN_ACESSIBILIDADE_VAO LIVRE    | Recursos de acessibilidade para pessoas com deficiência ou mobilidade reduzida nas vias de circulação interna na escola - Portas com vão livre de no mínimo 80 cm |
| 128 | IN_ACESSIBILIDADE_RAMPAS       | Recursos de acessibilidade para pessoas com deficiência ou mobilidade reduzida nas vias de circulação interna na escola - Rampas                                  |
| 129 | IN_ACESSIBILIDADE_SINAL SONORO | Recursos de acessibilidade para pessoas com deficiência ou mobilidade reduzida nas vias de circulação interna na escola - Sinalização sonora                      |
| 130 | IN_ACESSIBILIDADE_SINAL_TATIL  | Recursos de acessibilidade para pessoas com deficiência ou mobilidade reduzida nas vias de circulação interna na escola - Sinalização tátil (piso/paredes)        |

|     |                                  |  |
|-----|----------------------------------|--|
| 131 | IN_ACESSIBILIDADE_SINAL_VISUAL   | Recursos de acessibilidade para pessoas com deficiência ou mobilidade reduzida nas vias de circulação interna na escola - Sinalização visual (piso/paredes)                          |
| 132 | IN_ACESSIBILIDADE_INEXISTENTE    | Recursos de acessibilidade para pessoas com deficiência ou mobilidade reduzida nas vias de circulação interna na escola - Nenhum dos recursos de acessibilidade listado              |
| 133 | QT_SALAS_EXISTENTES              | Número de salas de aula existentes na escola   |
| 134 | QT_SALAS_UTILIZADAS_DENTRO       | Número de salas de aula utilizadas na escola - Dentro do prédio  |
| 135 | QT_SALAS_UTILIZADAS_FORA         | Número de salas de aula utilizadas na escola - Fora do prédio  |
| 136 | QT_SALAS_UTILIZADAS              | Número de salas de aula utilizadas na escola (dentro e fora do prédio)   |
| 137 | QT_SALAS_UTILIZADAS_CLIMATIZADAS | Condições das salas de aula utilizadas na escola (dentro e fora do prédio escolar) - Número de salas de aula climatizadas  |
| 138 | QT_SALAS_UTILIZADAS_ACESSIVEIS   | Condições das salas de aula utilizadas na escola (dentro e fora do prédio escolar) - Número de salas de aula com acessibilidade para pessoas com deficiência ou mobilidade reduzida. |
| 139 | IN_EQUIP_PARABOLICA              | Equipamentos existentes na escola - Antena parabólica  |
| 140 | IN_COMPUTADOR                    | Equipamentos existentes na escola - Computador   |
| 141 | IN_EQUIP_COPIADORA               | Equipamentos existentes na escola - Copiadora  |
| 142 | IN_EQUIP_IMPRESSORA              | Equipamentos existentes na escola - Impressora   |
| 143 | IN_EQUIP_IMPRESSORA_MULT         | Equipamentos existentes na escola - Impressora Multifuncional  |
| 144 | IN_EQUIP_SCANNER                 | Equipamentos existentes na escola - Scanner  |
| 145 | IN_EQUIP_DVD                     | Equipamentos existentes na escola para o processo ensino aprendizagem - DVD/Blu-ray  |
| 146 | QT_EQUIP_DVD                     | Quantidade de Aparelhos de DVD/Blue-ray  |
| 147 | IN_EQUIP_SOM                     | Equipamentos existentes na escola para o processo ensino aprendizagem - Aparelho de som  |
| 148 | QT_EQUIP_SOM                     | Quantidade de Aparelhos de som   |
| 149 | IN_EQUIP_TV                      | Equipamentos existentes na escola para o processo ensino aprendizagem - Aparelho de televisão  |
| 150 | QT_EQUIP_TV                      | Quantidade de Aparelhos de televisão   |
| 151 | IN_EQUIP_LOUSA_DIGITAL           | Equipamentos existentes na escola para o processo ensino aprendizagem - Lousa digital  |
| 152 | QT_EQUIP_LOUSA_DIGITAL           | Quantidade de Lousas digitais  |
| 153 | IN_EQUIP_MULTIMIDIA              | Equipamentos existentes na escola para o processo ensino aprendizagem - Projetor Multimídia (Datashow)   |
| 154 | QT_EQUIP_MULTIMIDIA              | Quantidade de Projetores Multimídias (Datashow)  |
| 155 | IN_EQUIP_VIDEOCASSETTE           | Equipamentos existentes na escola - Videocassete   |
| 156 | IN_EQUIP_RETROPROJETOR           | Equipamentos existentes na escola - Retroprojetor  |
| 157 | IN_EQUIP_FAX                     | Equipamentos existentes na escola - Fax  |
| 158 | IN_EQUIP_FOTO                    | Equipamentos existentes na escola - Máquina fotográfica/Filmadora  |

|     |                                   |   |
|-----|-----------------------------------|---|
| 159 | QT_EQUIP_VIDEOCASSET<br>E         | Quantidade de Videocassetes   |
| 160 | QT_EQUIP_PARABOLICA               | Quantidade de Antenas parabólicas   |
| 161 | QT_EQUIP_COPIADORA                | Quantidade de Copiadoras  |
| 162 | QT_EQUIP_RETROPROJET<br>OR        | Quantidade de Retroprojetores   |
| 163 | QT_EQUIP_IMPRESSORA               | Quantidade de Impressoras   |
| 164 | QT_EQUIP_IMPRESSORA_<br>MULT      | Quantidade de Impressoras Multifuncionais   |
| 165 | QT_EQUIP_FAX                      | Quantidade de Fax   |
| 166 | QT_EQUIP_FOTO                     | Quantidade de Máquinas Fotográficas/ Filmadoras   |
| 167 | QT_COMP_ALUNO                     | Quantidade de computadores em uso pelos alunos  |
| 168 | IN_DESKTOP_ALUNO                  | Computadores em uso pelos alunos (as) - Computador de mesa (desktop)  |
| 169 | QT_DESKTOP_ALUNO                  | Quantidade de computadores em uso pelos alunos (as) - Computador de mesa (desktop)  |
| 170 | IN_COMP_PORTATIL_ALU<br>NO        | Computadores em uso pelos alunos (as) - Computador portátil   |
| 171 | QT_COMP_PORTATIL_ALU<br>NO        | Quantidade de computadores em uso pelos alunos (as) - Computador portátil   |
| 172 | IN_TABLET_ALUNO                   | Computadores em uso pelos alunos (as) - Tablet  |
| 173 | QT_TABLET_ALUNO                   | Quantidade de computadores em uso pelos alunos (as) - Tablet  |
| 174 | QT_COMPUTADOR                     | Quantidade de computadores na escola  |
| 175 | QT_COMP_ADMINISTRATI<br>VO        | Quantidade de computadores de uso administrativo  |
| 176 | IN_INTERNET                       | Acesso à Internet   |
| 177 | IN_INTERNET_ALUNOS                | Acesso à Internet - Para uso dos alunos   |
| 178 | IN_INTERNET_ADMINISTR<br>ATIVO    | Acesso à Internet - Para uso administrativo   |
| 179 | IN_INTERNET_APRENDIZ<br>AGEM      | Acesso à Internet - Para uso nos processos de ensino e aprendizagem   |
| 180 | IN_INTERNET_COMUNIDA<br>DE        | Acesso à Internet - Para uso da comunidade  |
| 181 | IN_ACESSO_INTERNET_C<br>OMPUTADOR | Equipamentos que os alunos usam para acessar a internet da escola - Computadores de mesa, portáteis e tablets da escola (no laboratório de informática, biblioteca, sala de aula, etc.) |
| 182 | IN_ACES_INTERNET_DISP<br>_PESSOAS | Equipamentos que os alunos usam para acessar a internet da escola - Dispositivos pessoais (computadores portáteis, celulares, tablets, etc.)  |
| 183 | TP_REDE_LOCAL                     | Rede local de interligação de computadores  |
| 184 | IN_BANDA_LARGA                    | Internet Banda Larga  |
| 185 | QT_FUNCIONARIOS                   | Total de funcionários da escola (inclusive profissionais escolares em sala de aula)   |

|     |                            |   |
|-----|----------------------------|---|
| 186 | QT_PROF_ADMINISTRATIVOS    | Total de profissionais que atuam na escola - Auxiliares de secretaria ou auxiliares administrativos, atendentes   |
| 187 | QT_PROF_SERVICOS_GERAIS    | Total de profissionais que atuam na escola - Auxiliar de serviços gerais, porteiro(a), zelador(a), faxineiro(a), horticultor(a), jardineiro(a)  |
| 188 | QT_PROF_BIBLIOTECARIO      | Total de profissionais que atuam na escola - Bibliotecário(a), auxiliar de biblioteca ou monitor(a) da sala de leitura  |
| 189 | QT_PROF_SAUDE              | Total de profissionais que atuam na escola - Bombeiro(a) brigadista, profissionais de assistência a saúde (urgência e emergência), Enfermeiro(a), Técnico(a) de enfermagem e socorrista                                     |
| 190 | QT_PROF_COORDENADOR        | Total de profissionais que atuam na escola - Coordenador(a) de turno/disciplina   |
| 191 | QT_PROF_FONOAUDIOLOGO      | Total de profissionais que atuam na escola - Fonoaudiólogo(a)   |
| 192 | QT_PROF_NUTRICIONISTA      | Total de profissionais que atuam na escola - Nutricionista  |
| 193 | QT_PROF_PSIKOLOGO          | Total de profissionais que atuam na escola - Psicólogo(a) Escolar   |
| 194 | QT_PROF_ALIMENTACAO        | Total de profissionais que atuam na escola - Profissionais de preparação e segurança alimentar, cozinheiro(a), merendeira e auxiliar de cozinha   |
| 195 | QT_PROF_PEDAGOGIA          | Total de profissionais que atuam na escola - Profissionais de apoio e supervisão pedagógica: pedagogo(a), coordenador(a) pedagógico(a), orientador(a) educacional, supervisor(a) escolar e coordenador(a) de área de ensino |
| 196 | QT_PROF_SECRETARIO         | Total de profissionais que atuam na escola - Secretário(a) escolar  |
| 197 | QT_PROF_SEGURANCA          | Total de profissionais que atuam na escola - Seguranças, guarda ou segurança patrimonial  |
| 198 | QT_PROF_MONITORES          | Total de profissionais que atuam na escola - Técnicos(as), monitores(as) ou auxiliares de laboratório(s)  |
| 199 | IN_ALIMENTACAO             | Alimentação escolar para os alunos - PNAE/FNDE  |
| 200 | IN_SERIE_ANO               | Forma de organização do ensino - Série/Ano (séries anuais)  |
| 201 | IN_PERIODOS_SEMESTRAIS     | Forma de organização do ensino - Períodos semestrais  |
| 202 | IN_FUNDAMENTAL_CICLOS      | Forma de organização do ensino - Ciclo(s) do Ensino Fundamental   |
| 203 | IN_GRUPOS_NAO_SERIADOS     | Forma de organização do ensino - Grupos não-seriados com base na idade ou competência (art. 23 LDB)   |
| 204 | IN_MODULOS                 | Forma de organização do ensino - Módulos  |
| 205 | IN_FORMACAO_ALTERNANCIA    | Forma de organização do ensino - Alternância regular de períodos de estudos (proposta pedagógica de formação por alternância com tempo-escola e tempo-comunidade)   |
| 206 | IN_MATERIAL_PED_MULTIMIDIA | Instrumentos, materiais socioculturais e/ou pedagógicos em uso na escola para o desenvolvimento de atividades de ensino aprendizagem - Acervo multimídia  |
| 207 | IN_MATERIAL_PED_INFANTIL   | Instrumentos, materiais socioculturais e/ou pedagógicos em uso na escola para o desenvolvimento de atividades de ensino aprendizagem - Brinquedos para Educação Infantil  |
| 208 | IN_MATERIAL_PED_CIENTIFICO | Instrumentos, materiais socioculturais e/ou pedagógicos em uso na escola para o desenvolvimento de atividades de ensino aprendizagem - Conjunto de materiais científicos  |
| 209 | IN_MATERIAL_PED_DIFUSAO    | Instrumentos, materiais socioculturais e/ou pedagógicos em uso na escola para o desenvolvimento de atividades de ensino aprendizagem - Equipamento para amplificação e difusão de som/áudio                                 |
| 210 | IN_MATERIAL_PED_MUSICAL    | Instrumentos, materiais socioculturais e/ou pedagógicos em uso na escola para o desenvolvimento de atividades de ensino aprendizagem - Instrumentos musicais para conjunto, banda/fanfarras e/ou aulas de                   |

|     |                             |   |
|-----|-----------------------------|---|
|     |                             | música  |
| 211 | IN_MATERIAL_PED_JOGOS       | Instrumentos, materiais socioculturais e/ou pedagógicos em uso na escola para o desenvolvimento de atividades de ensino aprendizagem - Jogos Educativos   |
| 212 | IN_MATERIAL_PED_ARTISTICAS  | Instrumentos, materiais socioculturais e/ou pedagógicos em uso na escola para o desenvolvimento de atividades de ensino aprendizagem - Materiais para atividades culturais e artísticas                   |
| 213 | IN_MATERIAL_PED_DESPORTIVA  | Instrumentos, materiais socioculturais e/ou pedagógicos em uso na escola para o desenvolvimento de atividades de ensino aprendizagem - Materiais para prática desportiva e recreação                      |
| 214 | IN_MATERIAL_PED_INDIGENA    | Instrumentos, materiais socioculturais e/ou pedagógicos em uso na escola para o desenvolvimento de atividades de ensino aprendizagem - Indígena   |
| 215 | IN_MATERIAL_PED_ETNICO      | Instrumentos, materiais socioculturais e/ou pedagógicos em uso na escola para o desenvolvimento de atividades de ensino aprendizagem - Materiais pedagógicos para a educação das Relações Étnicas Raciais |
| 216 | IN_MATERIAL_PED_CAMPO       | Instrumentos, materiais socioculturais e/ou pedagógicos em uso na escola para o desenvolvimento de atividades de ensino aprendizagem - Materiais pedagógicos para a educação do campo                     |
| 217 | IN_MATERIAL_ESP_QUILOMBOLA  | Materiais didáticos específicos para atendimento à diversidade sociocultural - Quilombolas  |
| 218 | IN_MATERIAL_ESP_INDIGENA    | Materiais didáticos específicos para atendimento à diversidade sociocultural - Indígena   |
| 219 | IN_MATERIAL_ESP_NAO_UTILIZA | Materiais didáticos específicos para atendimento à diversidade sociocultural - Não utiliza  |
| 220 | IN_EDUCACAO_INDIGENA        | Educação Escolar Indígena   |
| 221 | TP_INDIGENA_LINGUA          | Educação Indígena - Língua em que o ensino é ministrado   |
| 222 | CO_LINGUA_INDIGENA_1        | Educação Indígena - Língua em que o ensino é ministrado - Língua Indígena - Código da língua Indígena 1   |
| 223 | CO_LINGUA_INDIGENA_2        | Educação Indígena - Língua em que o ensino é ministrado - Língua Indígena - Código da língua Indígena 2   |
| 224 | CO_LINGUA_INDIGENA_3        | Educação Indígena - Língua em que o ensino é ministrado - Língua Indígena - Código da língua Indígena 3   |
| 225 | IN_BRASIL_ALFABETIZADO      | Escola cede espaço para turmas do Programa Brasil Alfabetizado  |
| 226 | IN_FINAL_SEMANA             | Escola abre aos finais de semana para a comunidade  |
| 227 | IN_EXAME_SELECAO            | A escola faz exame de seleção para ingresso de seus alunos (Avaliação por prova e /ou análise curricular)   |
| 228 | IN_RESERVA_PPI              | Reserva de vagas por sistema de cotas para grupos específicos de alunos - Autodeclarado preto, pardo ou indígena (PPI)  |
| 229 | IN_RESERVA_RENDA            | Reserva de vagas por sistema de cotas para grupos específicos de alunos - Condição de Renda   |
| 230 | IN_RESERVA_PUBLICA          | Reserva de vagas por sistema de cotas para grupos específicos de alunos - Oriundo de escola pública   |
| 231 | IN_RESERVA_PCD              | Reserva de vagas por sistema de cotas para grupos específicos de alunos - Pessoa com deficiência (PCD)  |
| 232 | IN_RESERVA_OUTROS           | Reserva de vagas por sistema de cotas para grupos específicos de alunos - Outros grupos   |
| 233 | IN_RESERVA_NENHUMA          | Reserva de vagas por sistema de cotas para grupos específicos de alunos - Sem reservas de vagas para sistema de cotas (ampla concorrência)  |

|     |                            |  |
|-----|----------------------------|--|
| 234 | IN_REDES_SOCIAIS           | A escola possui site ou blog ou página em redes sociais para comunicação institucional   |
| 235 | IN_ESPACO_ATIVIDADE        | A escola compartilha espaços para atividades de integração escola-comunidade   |
| 236 | IN_ESPACO_EQUIPAMENTO      | A escola usa espaços e equipamentos do entorno escolar para atividades regulares com os alunos   |
| 237 | IN_ORGAO_ASS_PAIS          | Órgãos colegiados em funcionamento na escola - Associação de Pais  |
| 238 | IN_ORGAO_ASS_PAIS_MESTRES  | Órgãos colegiados em funcionamento na escola - Associação de Pais e Mestres  |
| 239 | IN_ORGAO_CONSELHO_ESCOLAR  | Órgãos colegiados em funcionamento na escola - Conselho Escolar  |
| 240 | IN_ORGAO_GREMIO_ESTUDANTIL | Órgãos colegiados em funcionamento na escola - Grêmios Estudantis  |
| 241 | IN_ORGAO_OUTROS            | Órgãos colegiados em funcionamento na escola - Outros  |
| 242 | IN_ORGAO_NENHUM            | Órgãos colegiados em funcionamento na escola - Não há órgãos colegiados em funcionamento   |
| 243 | TP_PROPOSTA_PEDAGOGICA     | Projeto político pedagógico ou a proposta pedagógica da escola (conforme art. 12 da LDB) atualizado nos últimos 12 meses até a data de referência  |
| 244 | TP_AEE                     | Atendimento Educacional Especializado (AEE)  |
| 245 | TP_ATIVIDADE_COMPLEMENTAR  | Mediação didático-pedagógica oferecida pela escola - Presencial  |
| 246 | IN_MEDIACAO_PRESENCIAL     | Mediação didático-pedagógica oferecida pela escola - Semipresencial  |
| 247 | IN_MEDIACAO_SEMIPRESENCIAL | Mediação didático-pedagógica oferecida pela escola - Semipresencial  |
| 248 | IN_MEDIACAO_EAD            | Atendimento Educacional Especializado (AEE)  |
| 249 | IN_ESPECIAL_EXCLUSIVA      | Escola possui classe especial exclusiva (ao menos uma turma exclusiva de alunos com deficiência, transtorno global do desenvolvimento ou altas habilidades/superdotação - Classes Especiais)   |
| 250 | IN_REGULAR                 | Modo, maneira ou metodologia de ensino correspondente às turmas com etapas de escolarização consecutivas, Creche ao Ensino Médio. Etapas consideradas (nas antigas modalidades 1 ou 2): TP_ETAPA_ENSINO igual a 1,2,4,5,6,7,8,9,10,11,14, 15,16,17,18,19,20,21,41,25,26,27,28,29,30,31, 32,33,34,35,36,37 ou 38.   |
| 251 | IN_EJA                     | Modo, maneira ou metodologia de ensino correspondente às turmas destinadas a pessoas que não cursaram o ensino fundamental e/ou médio em idade própria. Etapas consideradas (nas antigas modalidades 2 ou 3): TP_ETAPA_ENSINO igual a 65,67,69,70,71,73 ou 74.   |
| 252 | IN_PROFISSIONALIZANTE      | Modo profissionalizante de ensino correspondente às turmas de cursos de formação inicial e continuada ou de qualificação profissional (Cursos FIC) articulados à EJA ou concomitantes; ou de cursos técnicos de nível médio nas formas articulada (integrada ou concomitante) ou subsequente ao ensino médio e de normal/magistério. Etapas consideradas (nas antigas modalidades 1, 2 ou 3): TP_ETAPA_ENSINO igual a 30,31,32,33,34, 35,36,37,38,39,40,65,67,68,73 ou 74. |
| 253 | IN_COMUM_CRECHE            | Escola oferece Creche em classes comuns do ensino regular (TP_ETAPA_ENSINO=1 e IN_ESPECIAL_EXCLUSIVA=0)  |
| 254 | IN_COMUM_PRE               | Escola oferece Pré-escola em classes comuns do ensino regular (TP_ETAPA_ENSINO=2 e IN_ESPECIAL_EXCLUSIVA=0)  |
| 255 | IN_COMUM_FUND_AI           | Escola oferece Anos Iniciais do Ensino Fundamental (8 e 9 anos) em classes comuns do ensino regular (TP_ETAPA_ENSINO igual a 4,5,6,7,14,15,16,17 ou 18 e IN_ESPECIAL_EXCLUSIVA=0)  |
| 256 | IN_COMUM_FUND_AF           | Escola oferece Anos Finais do Ensino Fundamental (8 e 9 anos) em classes comuns do ensino regular (TP_ETAPA_ENSINO igual a 8,9,10,11,19,20,21 ou 41 e IN_ESPECIAL_EXCLUSIVA=0)   |

|     |                                      |  |
|-----|--------------------------------------|--|
| 257 | IN_COMUM_MEDIO_MEDIO                 | Escola oferece Ensino Médio - Médio Propedêutico em classes comuns do ensino regular (TP_ETAPA_ENSINO igual a 25,26,27,28 ou 29 e IN_ESPECIAL_EXCLUSIVA=0)                                       |
| 258 | IN_COMUM_MEDIO_INTEGRADO             | Escola oferece Curso Técnico Integrado ao Ensino Médio em classes comuns do ensino regular (TP_ETAPA_ENSINO igual a 30,31,32,33 ou 34 e IN_ESPECIAL_EXCLUSIVA=0)                                 |
| 259 | IN_COMUM_MEDIO_NORMAL                | Escola oferece Ensino Médio - Normal/ Magistério em classes comuns do ensino regular (TP_ETAPA_ENSINO igual a 35,36,37 ou 38 e IN_ESPECIAL_EXCLUSIVA=0)  |
| 260 | IN_ESP_EXCLUSIVA_CRECHE              | Escola oferece Creche em classes especiais exclusivas do ensino regular (TP_ETAPA_ENSINO=1 e IN_ESPECIAL_EXCLUSIVA=1)  |
| 261 | IN_ESP_EXCLUSIVA_PRE                 | Escola oferece Pré-escola em classes especiais exclusivas do ensino regular (TP_ETAPA_ENSINO=2 e IN_ESPECIAL_EXCLUSIVA=1)  |
| 262 | IN_ESP_EXCLUSIVA_FUNDAMENTAL_INICIAL | Escola oferece Anos Iniciais do Ensino Fundamental (8 e 9 anos) em classes especiais exclusivas do ensino regular (TP_ETAPA_ENSINO igual a 4,5,6,7,14,15,16,17 ou 18 e IN_ESPECIAL_EXCLUSIVA=1)  |
| 263 | IN_ESP_EXCLUSIVA_FUNDAMENTAL_FINAL   | Escola oferece Anos Finais do Ensino Fundamental (8 e 9 anos) em classes especiais exclusivas do ensino regular (TP_ETAPA_ENSINO igual a 8,9,10,11,19,20,21 ou 41 e IN_ESPECIAL_EXCLUSIVA=1)     |
| 264 | IN_ESP_EXCLUSIVA_MEDIO_MEDIO         | Escola oferece Ensino Médio - Médio Propedêutico em classes especiais exclusivas do ensino regular (TP_ETAPA_ENSINO igual a 25,26,27,28 ou 29 e IN_ESPECIAL_EXCLUSIVA=1)                         |
| 265 | IN_ESP_EXCLUSIVA_MEDIO_INTEGRADO     | Escola oferece Curso Técnico Integrado ao Ensino Médio em classes especiais exclusivas do ensino regular (TP_ETAPA_ENSINO igual a 30,31,32,33 ou 34 e IN_ESPECIAL_EXCLUSIVA=1)                   |
| 266 | IN_ESP_EXCLUSIVA_MEDIO_NORMAL        | Escola oferece Ensino Médio - Normal/ Magistério em classes especiais exclusivas do ensino regular (TP_ETAPA_ENSINO igual a 35,36,37 ou 38 e IN_ESPECIAL_EXCLUSIVA=1)                            |
| 267 | IN_COMUM_EJA_FUNDAMENTAL             | Escola oferece Ensino Fundamental em classes comuns da EJA (TP_ETAPA_ENSINO igual a 65,69,70 ou 73 e IN_ESPECIAL_EXCLUSIVA=0)  |
| 268 | IN_COMUM_EJA_MEDIO                   | Escola oferece Ensino Médio em classes comuns da EJA (TP_ETAPA_ENSINO igual a 67,71 ou 74 e IN_ESPECIAL_EXCLUSIVA=0)   |
| 269 | IN_COMUM_EJA_PROF                    | Escola oferece educação profissional integrada em classes comuns da EJA (TP_ETAPA_ENSINO igual a 65,67,73 ou 74 e IN_ESPECIAL_EXCLUSIVA=0)   |
| 270 | IN_ESP_EXCLUSIVA_EJA_FUNDAMENTAL     | Escola oferece Ensino Fundamental em classes especiais exclusivas da EJA (TP_ETAPA_ENSINO igual a 65,69,70 ou 73 e IN_ESPECIAL_EXCLUSIVA=1)  |
| 271 | IN_ESP_EXCLUSIVA_EJA_MEDIO           | Escola oferece Ensino Médio em classes especiais exclusivas da EJA (TP_ETAPA_ENSINO igual a 67,71 ou 74 e IN_ESPECIAL_EXCLUSIVA=1)   |
| 272 | IN_ESP_EXCLUSIVA_EJA_PROF            | Escola oferece educação profissional integrada em classes especiais exclusivas da EJA (TP_ETAPA_ENSINO igual a 65,67,73 ou 74 e IN_ESPECIAL_EXCLUSIVA=1)   |
| 273 | IN_COMUM_PROF                        | Educação Profissional - Curso Técnico Concomitante e/ou Subsequente e/ou Curso FIC Concomitante em classes comuns (TP_ETAPA_ENSINO igual a 39, 40 ou 68 e IN_ESPECIAL_EXCLUSIVA=0)               |
| 274 | IN_ESP_EXCLUSIVA_PROF                | Educação Profissional - Curso Técnico Concomitante e/ou Subsequente e/ou Curso FIC Concomitante em classes especiais exclusivas (TP_ETAPA_ENSINO igual a 39, 40 ou 68 e IN_ESPECIAL_EXCLUSIVA=1) |



## Apêndice E - Redução de variáveis

| cycle | att1                  | att2                           | act    | obs   |
|-------|-----------------------|--------------------------------|--------|---|
| 1     | in_especial_exclusiva | in_esp_exclusiva_creche        | fundir | Junção de variáveis relacionadas às escolas que trabalham com turmas exclusivamente de alunos especiais, independente do nível de ensino. |
| 1     | in_especial_exclusiva | in_esp_exclusiva_pre           | fundir |   |
| 1     | in_especial_exclusiva | in_esp_exclusiva_fund_ai       | fundir |   |
| 1     | in_especial_exclusiva | in_esp_exclusiva_fund_af       | fundir |   |
| 1     | in_especial_exclusiva | in_esp_exclusiva_medio_medio   | fundir |   |
| 1     | in_especial_exclusiva | in_esp_exclusiva_medio_integr  | fundir |   |
| 1     | in_especial_exclusiva | in_esp_exclusiva_medio_normal  | fundir |   |
| 1     | in_especial_exclusiva | in_esp_exclusiva_eja_fund      | fundir |   |
| 1     | in_especial_exclusiva | in_esp_exclusiva_eja_medio     | fundir |   |
| 1     | in_especial_exclusiva | in_esp_exclusiva_eja_prof      | fundir |   |
| 1     | in_especial_exclusiva | in_esp_exclusiva_prof          | fundir |   |
| 1     | in_especial_exclusiva |                                | faz1   |   |
| 1     | in_acessibilidade     | in_acessibilidade_corrimeo     | fundir | Junção de variáveis relacionadas às escolas que possuem algum tipo de infraestrutura de acessibilidade.                                   |
| 1     | in_acessibilidade     | in_acessibilidade_elevador     | fundir |   |
| 1     | in_acessibilidade     | in_acessibilidade_pisos_tateis | fundir |   |
| 1     | in_acessibilidade     | in_acessibilidade_vao_livre    | fundir |   |
| 1     | in_acessibilidade     | in_acessibilidade_rampas       | fundir |   |
| 1     | in_acessibilidade     | in_acessibilidade_sinal_sonoro | fundir |   |
| 1     | in_acessibilidade     | in_acessibilidade_sinal_tatil  | fundir |   |
| 1     | in_acessibilidade     | in_acessibilidade_sinal_visual | fundir |   |
| 1     | in_acessibilidade     |                                | faz1   |   |
| 1     | in_reserva_ppi        |                                | na2val | Valores na destas variáveis foram convertidos para 0, devido às escolas que não realizarem exames de seleção.                             |
| 1     | in_reserva_renda      |                                | na2val |   |
| 1     | in_reserva_publica    |                                | na2val |   |
| 1     | in_reserva_pcd        |                                | na2val |   |
| 1     | in_reserva_outros     |                                | na2val |   |
| 1     | in_reserva            | in_reserva_ppi                 | fundir | Junção de variáveis relacionadas a reserva de vagas para entrada na escola.   |
| 1     | in_reserva            | in_reserva_renda               | fundir |   |
| 1     | in_reserva            | in_reserva_publica             | fundir |   |
| 1     | in_reserva            | in_reserva_pcd                 | fundir |   |
| 1     | in_reserva            | in_reserva_outros              | fundir |   |
| 1     | in_reserva            |                                | faz1   |   |
| 1     | in_banheiro           | in_banheiro_ei                 | fundir | Junção de variáveis relacionadas a banheiros na escola  |
| 1     | in_banheiro           | in_banheiro_pne                | fundir |   |
| 1     | in_banheiro           | in_banheiro_chuveiro           | fundir |   |
| 1     | in_banheiro           | in_banheiro_funcionarios       | fundir |   |

|   |                             |                             |         |  |
|---|-----------------------------|-----------------------------|---------|--|
| 1 | in_banheiro                 |                             | faz1    |  |
| 1 | in_banda_larga              |                             | na2val  | Valores na desta variável foram convertidos para 0, no caso de escolas que não possuem internet.                                     |
| 1 | in_dormitorio               | in_dormitorio_aluno         | fundir  | Junção de variáveis relacionadas à existência de dormitórios na escola   |
| 1 | in_dormitorio               | in_dormitorio_professor     | fundir  |  |
| 1 | in_dormitorio               |                             | faz1    |  |
| 1 | in_equip_imprensa           | in_equip_imprensa_mult      | fundir  | Junção de variáveis relacionadas à existência de impressora na escola  |
| 1 | in_equip_imprensa           |                             | faz1    |  |
| 1 | in_localizacao_diferenciada | tp_localizacao_diferenciada | fundir  | Binarizada apenas para identificar se a escola se encontra em alguma área diferenciada (ex: indígena, quilombola, ...) ou não.       |
| 1 | in_localizacao_diferenciada |                             | faz1    | Junção de variáveis relacionadas à existência de localização diferenciada da escola  |
| 1 | in_rede_local               | tp_rede_local               | fundir  | Variável binarizada para existência ou não de uma rede local   |
| 1 | in_rede_local               |                             | faz1    |  |
| 1 | in_municipal                | tp_dependencia              | fundir  | Variável binarizada para 1 no caso de escola municipal ou 0 para estadual. escolas federais e privadas foram eliminadas da pesquisa. |
| 1 | in_rural                    | tp_localizacao              | fundir  | Binarizada por existir apenas 2 opções, rural ou urbana.   |
| 1 | in_predio_compartilhado     |                             | na2val  | 0; NA para zero, pois todo prédio zero faz prédio compartilhado zero   |
| 1 |                             | in_quadra_esportes          | deletar | Eliminado por existir in_quadra_esportes_coberta e in_quadra_esportes_descoberta   |
| 1 |                             | nu_ano_censo                | deletar | Analisado apenas o ano de 2019   |
| 1 |                             | no_entidade                 | deletar | Irrelevante para a pesquisa  |
| 1 |                             | co_orgao_regional           | deletar |  |
| 1 |                             | dt_ano_letivo_inicio        | deletar |  |
| 1 |                             | dt_ano_letivo_termino       | deletar |  |
| 1 |                             | co_mesorregiao              | deletar |  |
| 1 |                             | co_microrregiao             | deletar |  |
| 1 |                             | co_distrito                 | deletar |  |
| 1 |                             | co_escola_sede_vinculada    | deletar |  |
| 1 |                             | co_ies_ofertante            | deletar |  |
| 1 |                             | co_lingua_indigena_1        | deletar |  |
| 1 |                             | co_lingua_indigena_2        | deletar |  |
| 1 |                             | co_lingua_indigena_3        | deletar |  |
| 1 |                             | tp_indigena_lingua          | deletar |  |
| 1 |                             | tp_categoria_escola_privada | deletar | Escola privada não analisadas  |

|   |                     |                                |         |  |
|---|---------------------|--------------------------------|---------|--|
| 1 |                     | tp_convenio_poder_publico      | deletar | Irrelevante para a pesquisa  |
| 1 |                     | tp_regulamentacao              | deletar |  |
| 1 |                     | tp_ocupacao_galpao             | deletar |  |
| 1 |                     | tp_ocupacao_predio_escolar     | deletar |  |
| 1 |                     | tp_responsavel_regulamentacao  | deletar |  |
| 1 |                     | tp_convenio_poder_publico      | deletar |  |
| 1 |                     | in_conveniada_pp               | deletar | Escola privada não analisadas  |
| 1 |                     | in_mant_escola_privada_emp     | deletar |  |
| 1 |                     | in_mant_escola_privada_ong     | deletar |  |
| 1 |                     | in_mant_escola_privada_sind    | deletar |  |
| 1 |                     | in_mant_escola_privada_sist_s  | deletar |  |
| 1 |                     | in_mant_escola_privada_s_fins  | deletar | Irrelevante para a pesquisa  |
| 1 |                     | in_vinculo_secretaria_educacao | deletar |  |
| 1 |                     | in_vinculo_seguranca_publica   | deletar |  |
| 1 |                     | in_vinculo_secretaria_saude    | deletar |  |
| 1 |                     | in_vinculo_outro_orgao         | deletar | Escola privada não analisadas  |
| 1 |                     | in_mant_escola_privada_oscip   | deletar |  |
| 1 |                     | in_mant_escola_priv_ong_oscip  | deletar | Todas essas variáveis possuem suas quantidades correspondentes em outras variáveis com início "qt", logo deixamos apenas a quantidade e não apenas a existência ou não do equipamento. |
| 1 |                     | in equip_tv                    | deletar |  |
| 1 |                     | in equip_dvd                   | deletar |  |
| 1 |                     | in equip_som                   | deletar |  |
| 1 |                     | in equip_multimidia            | deletar |  |
| 1 |                     | in computador                  | deletar |  |
| 1 |                     | in equip_lousa_digital         | deletar |  |
| 1 |                     | in_desktop_aluno               | deletar |  |
| 1 |                     | in_comp_portatil_aluno         | deletar |  |
| 1 |                     | in_tablet_aluno                | deletar | Analisado apenas escolas presenciais   |
| 1 |                     | in_mediacao_presencial         | deletar |  |
| 1 |                     | in_mediacao_ead                | deletar | Não analisado com ensino ead.  |
| 1 |                     | in_conveniada_pp               | deletar | Irrelevante para a pesquisa  |
| 1 | qt equip_multimidia | qt equip_som                   | fundir  | Junção de variáveis relacionadas a quantidade de equipamentos multimídias na escola.   |
| 1 | qt equip_multimidia | qt equip_tv                    | fundir  |  |
| 1 | qt equip_multimidia | qt equip_dvd                   | fundir  |  |
| 1 |                     | in_orgao_nenhum                | deletar | Dependência  |
| 1 |                     | qt_salas_utilizadas            | deletar | Dependência  |
| 1 |                     | in_local_func_prisional_socio  | deletar | Dependência  |
| 1 |                     | in_acessibilidade_inexistente  | deletar | Dependência  |
| 1 |                     | in_reserva_nenhuma             | deletar | Dependência e NAs  |
| 1 |                     | in_esgoto_fossa                | deletar | Dependência  |

|   |  |                           |         |  |
|---|--|---------------------------|---------|--|
| 2 |  | tp_situacao_funcionamento | deletar |  |
|---|--|---------------------------|---------|--|

**Fonte:** Elaborado pelo Autor (2022)

**Legenda:** As ações: faz1 - mesclam as variáveis, onde se alguma for positiva, a variável se tornará 1; na2val - transforma as variáveis NA para o valor 0.