



Universidade de Brasília

Tailine Nonato

**ANÁLISE DESCRITIVA DE DADOS:
RESULTADOS DO SAEB (ANEB - PROVA
BRASIL) 2017**

Brasília, DF

Março, 2021

Tailine Nonato

ANÁLISE DESCRITIVA DE DADOS: RESULTADOS DO SAEB (ANEB - PROVA BRASIL) 2017

Relatório de análise de banco de dados apresentado à disciplina Métodos Estatísticos 2 da Universidade de Brasília como objeto de avaliação.

Brasília, DF

Março, 2021

Lista de ilustrações

Figura 1 – Gráfico de colunas da frequência da Região	9
Figura 2 – Gráfico de colunas da frequência do Sexo	10
Figura 3 – Gráfico de barras das Frequências de Idade	11
Figura 4 – Gráfico de colunas da frequência de Raça/Cor	13
Figura 5 – Gráfico de colunas da frequência de Mora com Mãe por Mora com Pai	14
Figura 6 – Histograma de Nota em Língua Portuguesa com Curva de Densidade	16
Figura 7 – Boxplot de Nota em Língua Portuguesa	17
Figura 8 – Histograma de Notas em Matemática com Curva de Densidade	18
Figura 9 – Boxplot de Notas em Matemática	19

Lista de tabelas

Tabela 1 – Classificação das variáveis segundo o tipo	7
Tabela 2 – Frequências da variável Região	10
Tabela 3 – Frequências da variável Sexo	11
Tabela 4 – Frequências da variável Idade	12
Tabela 5 – Frequências da variável Raça/Cor	13
Tabela 6 – Frequências das variáveis Mora_Mãe por Mora_Pai	14
Tabela 7 – Frequências da variável Nota_LP	15
Tabela 8 – Quadro Resumo de Notas em Língua Portuguesa	16
Tabela 9 – Frequências da variável Nota_MT	18
Tabela 10 – Quadro Resumo de Notas em Matemática	19

Lista de abreviaturas e siglas

INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
SAEB	Sistema de Avaliação da Educação Básica

Sumário

	Introdução	6
1	METODOLOGIA	7
2	ANÁLISES E RESULTADOS	9
2.1	Análise da Região	9
2.2	Análise do Sexo	10
2.3	Análise da Idade	11
2.4	Análise de Raça/Cor	12
2.5	Análise de Moradia com Mãe/Pai	13
2.6	Análise da Nota em Língua Portuguesa	15
2.7	Análise da Nota em Matemática	17
3	CONCLUSÃO	21
	REFERÊNCIAS	22
	ANEXO A – CÓDIGO EM R	23

Introdução

Este relatório tem por objetivo analisar características e desempenho de estudantes do 5º ano do Ensino Fundamental de escolas públicas brasileiras. Serão utilizados dados do Sistema de Avaliação da Educação Básica (SAEB) de 2017 divulgado pelo INEP (2018b).

O Saeb é um sistema de avaliação em larga escala realizado periodicamente. Os resultados de aprendizagem dos estudantes apurados nesse sistema, juntamente com as taxas de aprovação, reprovação e abandono apuradas no Censo Escolar, compõem o Índice de Desenvolvimento da Educação Básica (Ideb) INEP (2018a).

O sistema tem como população alvo todos estudantes da educação pública brasileira que estão no 5º ou 9º ano do Ensino Fundamental ou 3º ano do Ensino Médio e uma amostra dos estudantes da rede particular nas mesmas séries já citadas. A população acessível em 2017 foi de cerca de 68% daqueles matriculados no ensino fundamental e médio. As coletas são feitas por meio de testes e questionários, que refletem os níveis de aprendizagem. Nos testes as características levantadas são relacionadas a capacidade de leitura, níveis de alfabetização e resolução de problemas. Já nos questionários possibilitam analisar o nível socioeconômico, os serviços sociais, a infraestrutura, a formação de professores, o material didático e os programas estruturados.

Este estudo será feito a partir de uma amostra de 2000 observações dos resultados do SAEB de 2017, tendo como auxílio teórico o livro Estatística Básica de BUSSAB (2017).

1 Metodologia

Esta pesquisa tem por objetivo fazer análises descritivas, portanto, foram utilizados gráficos de colunas, gráficos de radar, quadros resumo, tabelas de frequência e de contingência.

O banco de dados analisado dispõe de 2000 observações em 19 variáveis que estão listadas, na tabela a seguir, juntamente com seu tipo que pode ser Qualitativa Nominal, Qualitativa Ordinal, Quantitativa Discreta ou Quantitativa Contínua.

Tabela 1 – Classificação das variáveis segundo o tipo

Variável	Tipo
ANO	Qualitativa Nominal
REGIAO	Qualitativa Nominal
UF	Qualitativa Nominal
MUNICIPIO	Qualitativa Nominal
AREA	Qualitativa Nominal
DEPENDENCIA_ADM	Qualitativa Nominal
LOCALIZACAO	Qualitativa Nominal
NOTA_LP	Quantitativa Contínua
NOTA_MT	Quantitativa Contínua
SEXO	Qualitativa Nominal
RACA_COR	Qualitativa Nominal
IDADE	Qualitativa Ordinal
COMPUTADOR	Qualitativa Nominal
MORA_MÃE	Qualitativa Nominal
MORA_PAI	Qualitativa Nominal
REUNIÕES_ESCOLARES	Qualitativa Ordinal
USO_TEMPO_TELAS	Qualitativa Ordinal
AFAZERES_DOM	Qualitativa Ordinal
BIBLIOTECA	Qualitativa Ordinal

Fonte: Elaboração Própria

O banco foi construído de forma a atribuir códigos (como “A”, “B”, “C”) às observações de cada variável. Segue-se então, juntamente com suas descrições, as colunas utilizadas no presente relatório:

- REGIAO: Região de localização da escola do estudante (Norte, Nordeste, Sudeste, Sul, Centro-Oeste)

- NOTA_LP: Proficiência em Língua Portuguesa transformada na escala única do SAEB, com média = 250, desvio = 50 (do SAEB/97)
- NOTA_MT: Proficiência do aluno em Matemática transformada na escala única do SAEB, com média = 250, desvio = 50 (do SAEB/97)
- SEXO: Sexo do estudante (Masculino, Feminino)
- RACA_COR: Raça/cor do estudante (Branca, Preta, Parda, Amarela, Indígena, Não quero declarar)
- IDADE: Idade do estudante (8 anos ou menos, 9 anos, ..., 15 anos ou mais)
- MORA_MÃE: Você mora com sua mãe? (Sim, Não, Não, mas moro com outra mulher responsável por mim)
- MORA_PAI: Você mora com seu pai? (Sim, Não, Não, mas moro com outra homem responsável por mim)

A partir das descrições acima foi feita uma “limpeza” no banco, onde os códigos foram substituídos por suas especificações.

É importante destacar que existe quantidade significativa de observações em branco. Ao decorrer do estudo, essas observações serão descritas em formato de proporção.

Para execução do presente relatório, foi utilizado o Software *RStudio* com *R* versão 4.0.3 e com auxílio dos pacotes *tidyverse*, *dplyr*, *stringr* e *janitor*.

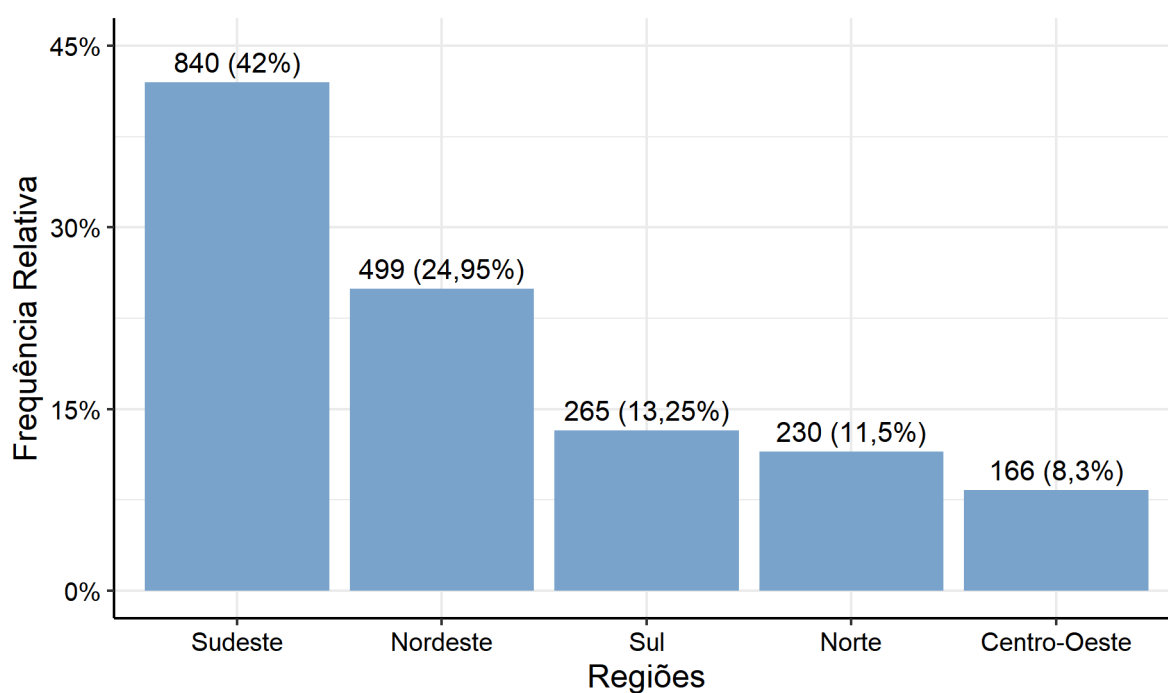
2 Análises e Resultados

O presente capítulo tem por objetivo promover a análise descritiva dos dados disponíveis, utilizando gráficos, tabelas e quadros para dar melhor embasamento aos resultados.

2.1 Análise da Região

A região é um importante fator quando se trata de uma avaliação em todo território nacional, tanto para verificar como se distribui quantidade de participantes quanto para testar possíveis associações.

Figura 1 – Gráfico de colunas da frequência da Região



Fonte: Elaboração própria.

Tabela 2 – Frequências da variável Região

Região	Frequência	Porcentagem
Sudeste	840	42%
Nordeste	499	24,95%
Sul	265	13,25%
Norte	230	11,5%
Centro-Oeste	166	8,3%
Total	2000	100%

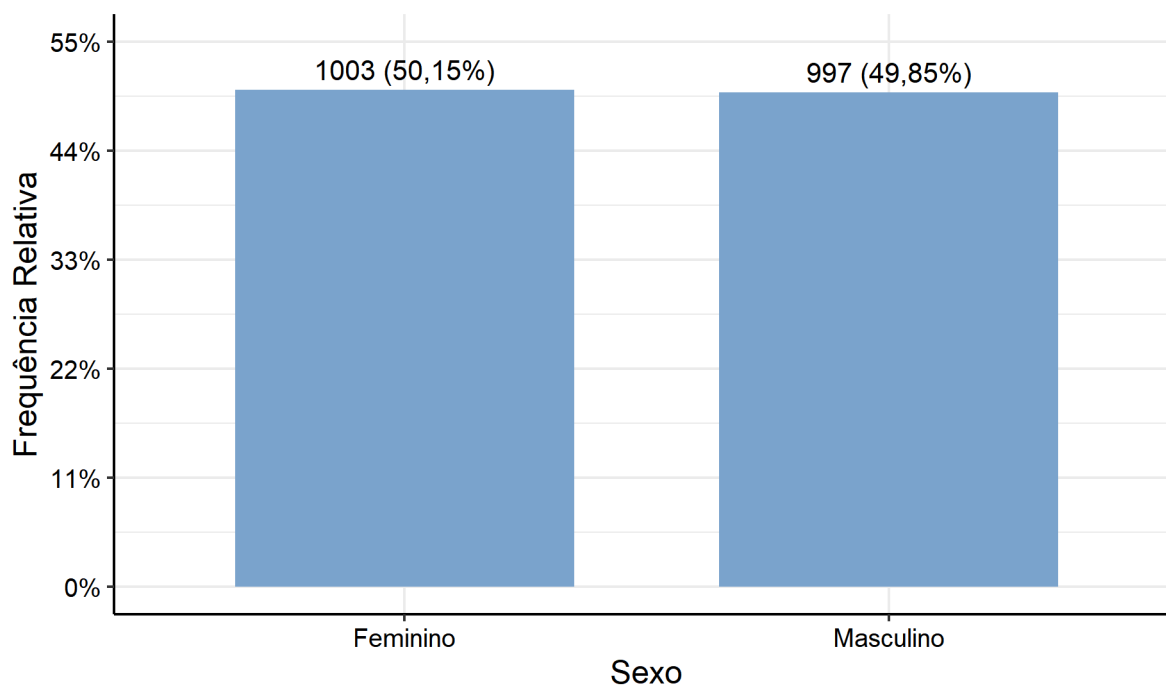
Fonte: INEP

Observa-se, na Figura 1 e na Tabela 2, que grande parte dos indivíduos são estudantes da região Sudeste, representando 42% da amostra e que há pouca representação de estudantes da região Centro-Oeste, com apenas 166 dentre os 2000 observados na amostra.

2.2 Análise do Sexo

A fim de entender como se comporta a distribuição de observações do sexo, segue o Gráfico 2 e a Tabela 3:

Figura 2 – Gráfico de colunas da frequência do Sexo



Fonte: Elaboração própria.

Tabela 3 – Frequências da variável Sexo

Sexo	Frequência	Porcentagem
Feminino	1003	50,15%
Masculino	997	49,85%
Total	2000	100%

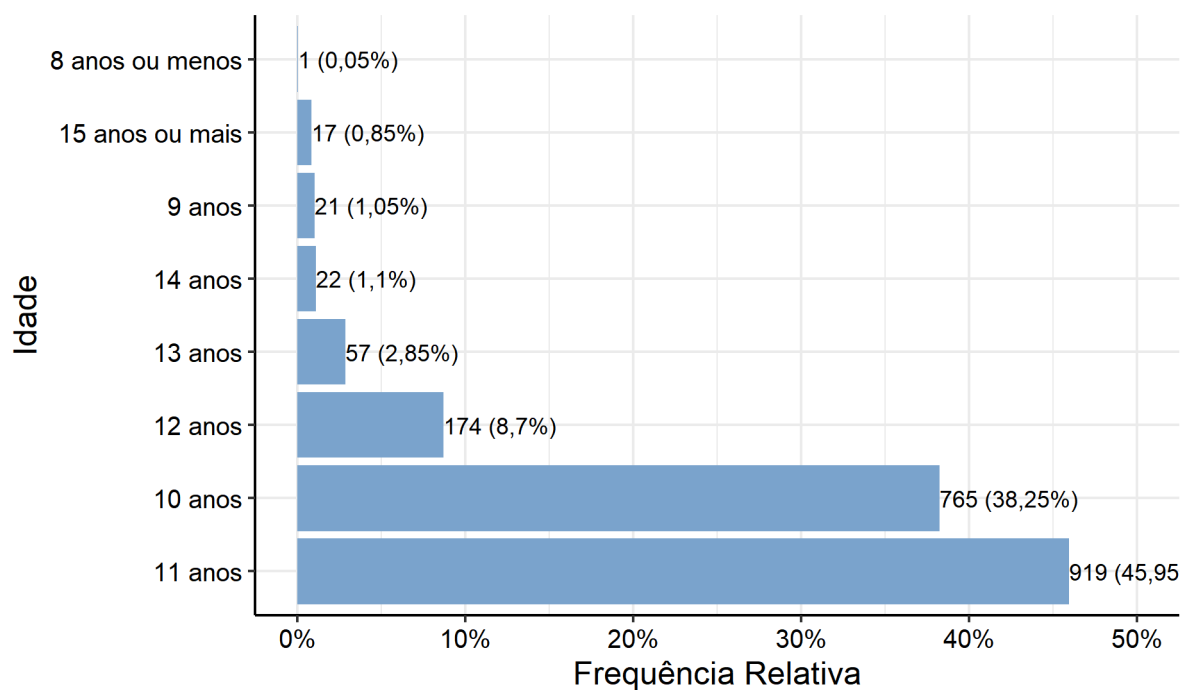
Fonte: INEP

Como mostra a Figura 2 e a Tabela 3, o sexo está semelhantemente distribuído entre feminino e masculino de forma a quase atingir divisão exata.

2.3 Análise da Idade

A presente seção tem por finalidade analisar a idade dos participantes em seu contexto geral. Para idade houve uma proporção de 1,2% de observações em branco. Tais observações foram retiradas da visualização gráfica e da tabela, mas foram contabilizadas nos cálculos de frequências.

Figura 3 – Gráfico de barras das Frequências de Idade



Fonte: Elaboração própria.

Tabela 4 – Frequências da variável Idade

Faixa Etária	Frequência	Porcentagem
8 anos ou menos	1	0,05%
9 anos	21	1,05%
10 anos	765	38,25%
11 anos	919	45,95%
12 anos	174	8,7%
13 anos	57	2,85%
14 anos	22	1,1%
15 anos ou mais	17	0,85%
Total	1976	98,8%

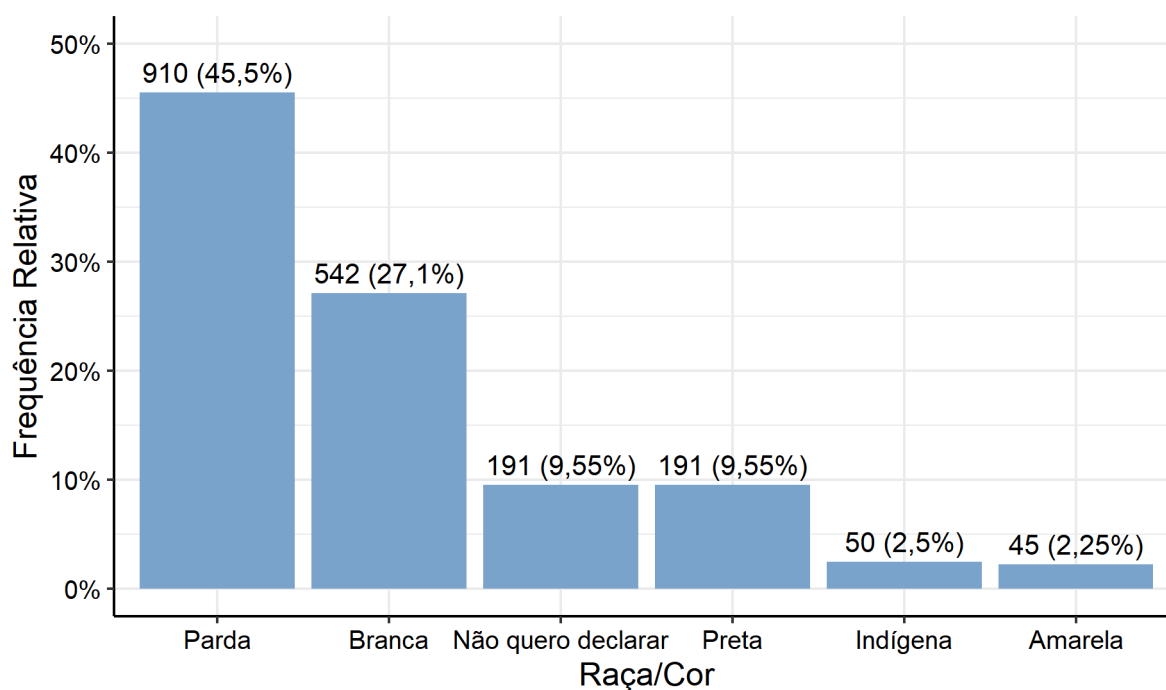
Fonte: INEP

A Figura 3 mostra com clareza e a Tabela 4 complementa que a concentração de idade está entre os 10 anos e os 11 anos, representando juntamente 84,2% de todas as observações. A faixa dos mais novos (8 anos ou menos) comporta-se tipicamente como um outlier no estudo, ou seja, é a idade que mais se diferencia do padrão de idade do estudo.

2.4 Análise de Raça/Cor

Buscando observar como a variável Raça/Cor está distribuída, seguem a Figura 4 e a Tabela 5. Vale ressaltar que a proporção de observações em branco foi 3,55% . Tais observações foram retiradas da visualização gráfica e da tabela, mas foram contabilizadas nos cálculos de frequências.

Figura 4 – Gráfico de colunas da frequência de Raça/Cor



Fonte: Elaboração própria.

Tabela 5 – Frequências da variável Raça/Cor

Raça/Cor	Frequência	Porcentagem
Parda	910	45,5%
Branca	542	27,1%
Não quis declarar	191	9,55%
Preta	191	9,55%
Indígena	50	2,5%
Amarela	45	2,25%
Total	1929	96,45%

Fonte: INEP

Destaque-se que a maioria, 45,5%, declara-se parda. Apenas 9,55% declaram-se pretos, quantidade igualmente dada para proporção dos que não quiseram declarar.

2.5 Análise de Moradia com Mãe/Pai

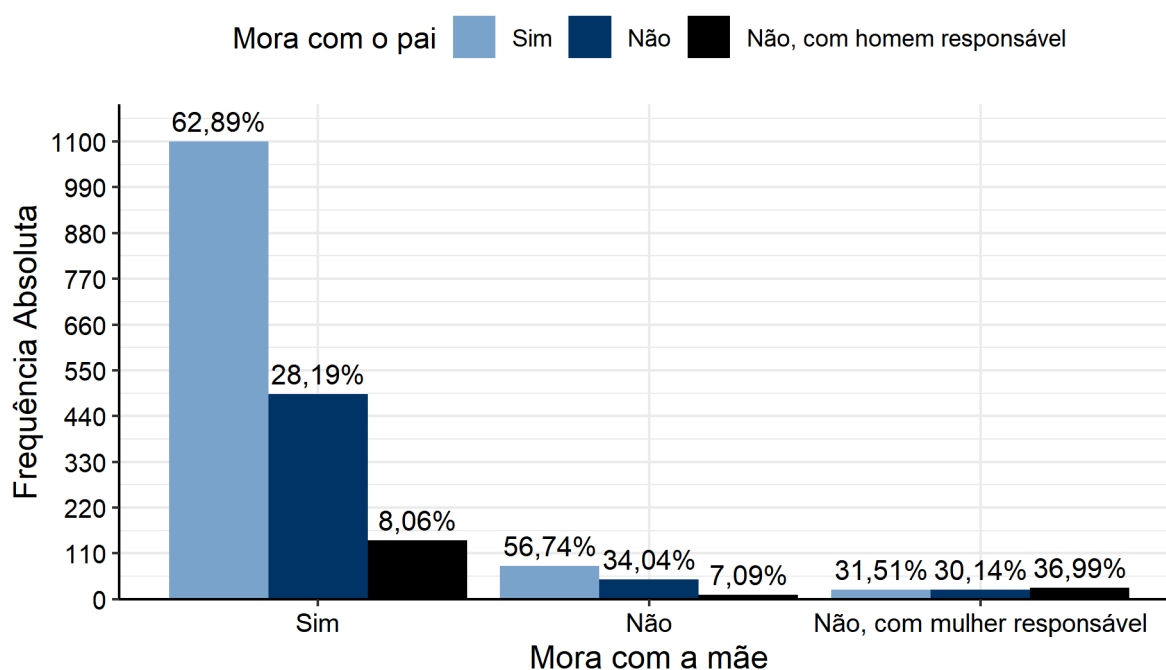
Apesar de existirem muitas estruturas familiares diferentes, existem estudos que afirmam que tal estrutura, juntamente com participação do responsável na escola, pode

ser um fator influenciador no desempenho de um aluno, por isso dadas as informações, é importante analisar o que os dados podem mostrar. Portanto, esta seção tem por objetivo observar a moradia com mãe/pai.

É importante destacar que para moradia com mãe houve uma proporção de 1,85% de observações em branco e para moradia com pai de 1,10%. Quando relacionadas, apresentam uma proporção de 2,8% de observações em branco. Tais observações foram retiradas da visualização gráfica e da tabela, mas foram contabilizadas nos cálculos de frequências.

Para a Figura 5, as frequências relativas foram calculadas de acordo com cada resposta de moradia com pai dado a resposta de moradia com mãe, já na Tabela 6 as frequências relativas foram calculadas para cada uma das variáveis.

Figura 5 – Gráfico de colunas da frequência de Mora com Mãe por Mora com Pai



Fonte: Elaboração própria.

Tabela 6 – Frequências das variáveis Mora_Mãe por Mora_Pai

Mora com mãe	Mora com pai			Porcentagem
	Sim	Não	Não, mas com homem responsável	
Sim	1100	493	141	86,7%
Não	80	48	10	6,9%
Não, mas com mulher responsável	23	22	27	3,6%
Porcentagem	60,15%	28,15%	8,9%	97,2%

Fonte: INEP

A partir da tabela é possível ver com clareza que a maioria mora com a mãe (86,7%) e a maioria mora com o pai (60,15%). Há uma diferença de 26,55% entre os percentuais de moradia com mãe e com pai. Existe uma pequena parte que não mora com a mãe, mas com mulher responsável e a Figura 5 mostra que, dentre esses, há uma distribuição próxima de equânime de moradia com pai. Observa-se também que existem mais estudantes morando homem responsável (8,9%) do que morando com mulher responsável (3,6%).

2.6 Análise da Nota em Língua Portuguesa

Para avaliar da melhor forma o desempenho dos alunos em Língua Portuguesa, a variável foi dividida em 5 intervalos de classe com amplitude de 50 pontos para a Tabela 7 e foi utilizada em sua forma original para Figura 6 e a Tabela 8.

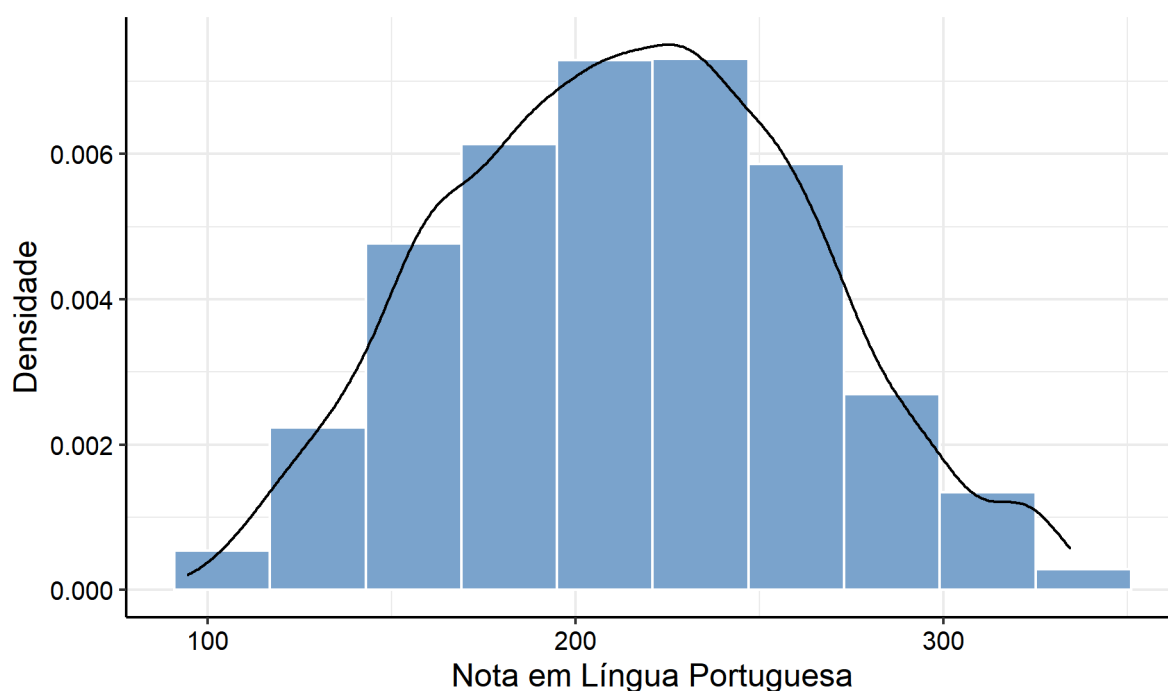
Tabela 7 – Frequências da variável Nota_LP

Nota	Frequência	Porcentagem
Entre 94 e 143	152	7,60%
Entre 144 e 193	548	27,40%
Entre 194 e 243	734	36,70%
Entre 244 e 293	455	22,75%
Entre 294 e 334	111	5,55%
Total	2000	100%

Fonte: INEP

Observa-se que com exceção das faixas extremas, a distribuição de frequências se apresenta em números próximos, porém dá-se maior destaque a terceira, que mostra que 36,7% dos estudantes desta amostra tiraram, em Língua Portuguesa, notas entre 194 e 243. Apenas 5,55% estão incluídos entre as notas mais altas.

Figura 6 – Histograma de Nota em Língua Portuguesa com Curva de Densidade



O histograma, juntamente à curva de densidade, sugere que as notas estão distribuídas de forma simétrica.

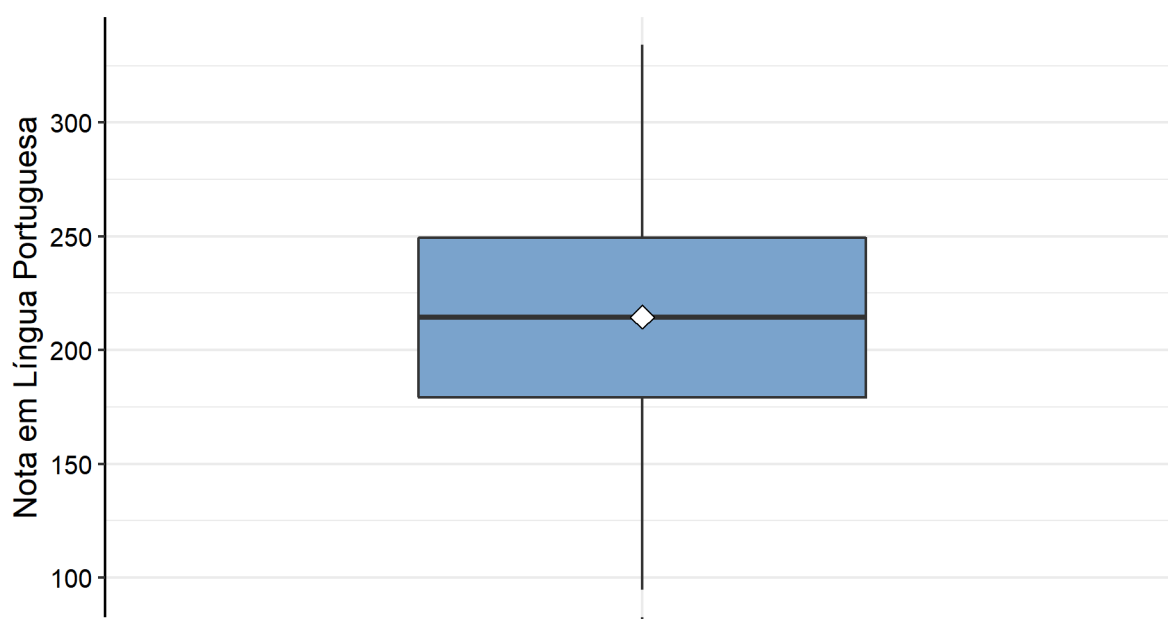
Tabela 8 – Quadro Resumo de Notas em Língua Portuguesa

Estatística	Valores
Média	214,56
Mínimo	94,61
1º Quartil	179,19
Mediana	214,47
3º Quartil	249,36
Máximo	334,23
Variação	2335,34
Assimetria	0,065
Curtose	-0,509

Fonte: Elaboração própria.

No Quadro 8 destaca-se que a média e a mediana estão muito próximas e o valor do Coeficiente de Assimetria está próximo de 0, o que confirma simetria nas Notas em Língua Portuguesa. Além disso, o Coeficiente de Curtose é < 0 , portanto temos uma curva platicúrtica, então a função de distribuição é mais achatada do que a distribuição normal.

Figura 7 – Boxplot de Nota em Língua Portuguesa



Fonte: Elaboração própria.

Assim como mostra o Quadro 8, a Figura 7 mostra que as notas em Língua Portuguesa estão concentradas entre 179 e 250. Analisando dessa forma tem-se uma amplitude maior do que quando a análise é feita através de intervalos de classe.

2.7 Análise da Nota em Matemática

Assim como na seção anterior, para avaliar da melhor forma o desempenho dos alunos, nesta seção, em Matemática, a variável foi dividida em 5 intervalos de classe com amplitude de 49 pontos para a Tabela 7 e foi utilizada em sua forma original para Figura 6 e a Tabela 8.

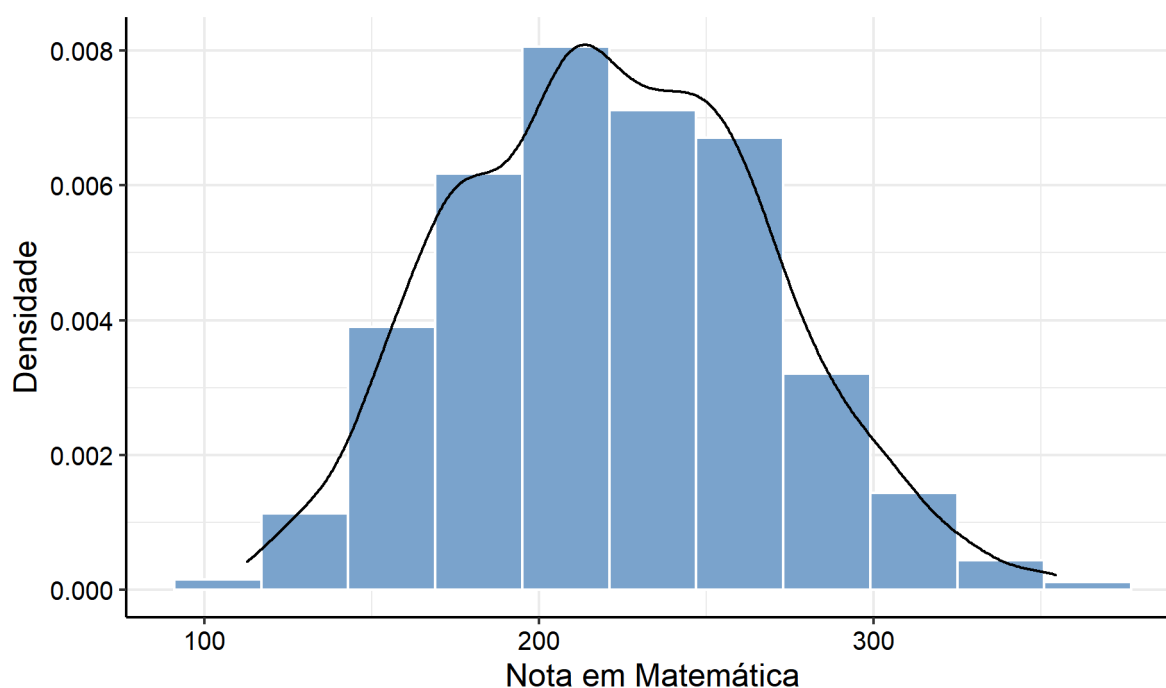
Tabela 9 – Frequências da variável Nota_MT

Nota	Frequência	Porcentagem
Entre 112 e 160	187	9,35%
Entre 161 e 209	639	31,95%
Entre 210 e 258	735	36,75%
Entre 259 e 307	375	18,75%
Entre 308 e 355	64	3,20%
Total	2000	100%

Fonte: INEP

Para as notas em Matemática, a frequência está mais concentrada entre a segunda e a terceira faixa, sendo assim, mais de 2/3 (68,7%) dos estudantes da amostra tiveram a pontuação entre 161 e 258. Apenas 64 dos 2000 estudantes fizeram mais de 308 pontos.

Figura 8 – Histograma de Notas em Matemática com Curva de Densidade



Fonte: Elaboração própria.

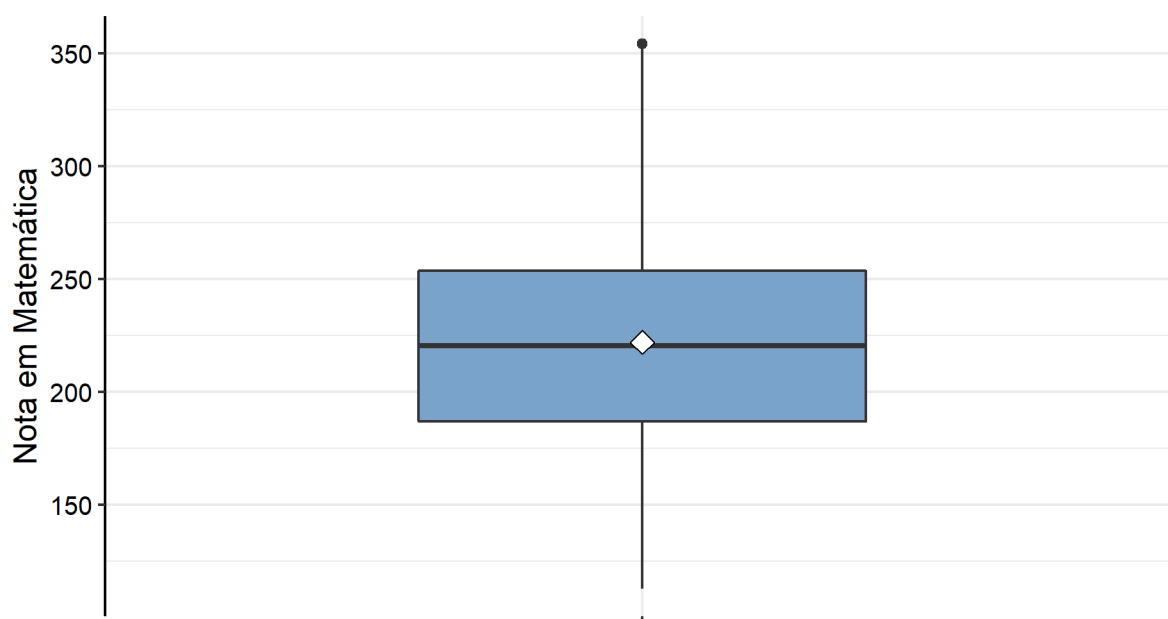
A curva de densidade se apresenta com o pico um pouco mais alto que a curva das notas em Língua Portuguesa. O histograma sugere que as notas estão distribuídas de forma simétrica.

Tabela 10 – Quadro Resumo de Notas em Matemática

Estatística	Valores
Média	222
Mínimo	112,7
1º Quartil	186,9
Mediana	220,3
3º Quartil	253,7
Máximo	354,4
Variação	2119,66
Assimetria	0,137
Curtose	-0,395

Fonte: Elaboração própria.

Figura 9 – Boxplot de Notas em Matemática



Fonte: Elaboração própria.

A Figura 9 mostra que as notas em Matemática estão majoritariamente entre 186 e 254, informações essas que são detalhadas pelo Quadro 10, que mostra também que a média e mediana estão próximas, se distanciando em cerca de 2 pontos. Analisando dessa forma tem-se uma melhor especificação da amplitude do que quando a análise é feita através de intervalos de classe. Além disso, o Coeficiente de Curtose é < 0 , portanto,

apesar de ser maior do que o das notas em Língua Portuguesa, também temos uma curva platicúrtica, então a função de distribuição é mais achatada do que a distribuição normal.

3 Conclusão

A partir de todos os resultados conclui-se alguns pontos principais:

- A maioria dos estudantes são da região Sudeste e a região menos representada é Centro-Oeste.
- O sexo está distribuído quase em 1/2 para Feminino e Masculino.
- A idade dos estudantes está majoritariamente concentrada entre 10 e 11 anos de idade.
- No quesito racial, a maioria declara-se parda.
- A maioria dos estudantes mora com a mãe e a maioria mora com o pai, valendo destacar que existe uma diferença de 26,55% entre mora sim com a mãe e mora sim com o pai.
- As notas em Língua Portuguesa estão concentradas entre 194 e 243. Tais notas seguem uma distribuição simétrica e com curva mais achatada que a distribuição normal.
- Mais de 2/3 dos estudantes tiraram notas entre 161 e 258 em Matemática. As notas nessa disciplina possuem uma distribuição simétrica e com curva mais achatada que a distribuição normal.

Referências

BUSSAB, P. A. M. W. O. *Estatística Básica*. São Paulo: Saraiva, 2017. Citado na página 6.

INEP. *Resultados do Saeb 2017*: O inep apresenta esta narrativa. [S.l.], 2018. 1 p. Disponível em: <<https://medium.com/@inep/resultados-do-saeb-2017-f471ec72168d>>. Citado na página 6.

INEP. *Saeb - Inep*: Sistema de avaliação da educação básica (saeb). [S.l.], 2018. 33 p. Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/saeb>>. Citado na página 6.

ANEXO A – Código em R

Link para o diretório no GitHub:

<<https://github.com/tailineju/SAEB-2017>>