

Aline de Almeida Ramos

**Análise descritiva de dados
Resultados do SAEB
(ANEB - Prova Brasil) - 2017**

Brasília

março de 2021

Aline de Almeida Ramos

Análise descritiva de dados
Resultados do SAEB
(ANEB - Prova Brasil) - 2017

Relatório de pesquisa realizado acerca de dados da Prova Brasil, durante o curso da disciplina de Métodos Estatísticos 2.

Universidade de Brasília
Departamento de Estatística

Brasília
março de 2021

Lista de ilustrações

Figura 1 – Percentual de estudantes por sexo	8
Figura 2 – Percentual de estudantes por raça/cor	10
Figura 3 – Percentual de estudantes por área de localização	11
Figura 4 – Percentual de estudantes por quantidade de computador	12
Figura 5 – Percentual de estudantes por uso da biblioteca	13
Figura 6 – Histograma da distribuição das notas de português dos estudantes . . .	15
Figura 7 – Boxplot da distribuição das notas de português	16
Figura 8 – Histograma da distribuição das notas de matemática dos estudantes . .	18
Figura 9 – Boxplot da distribuição das notas de matemática	19
Figura 10 – Boxplots da distribuição das notas de português e matemática	21

Lista de tabelas

Tabela 1	– Distribuição de frequência da variável sexo	8
Tabela 2	– Distribuição de frequência da variável raça/cor	10
Tabela 3	– Distribuição de frequência da variável área de localização da escola . .	11
Tabela 4	– Distribuição de frequência da variável ter computador em casa	13
Tabela 5	– Distribuição de frequência da variável uso da biblioteca ou sala de leitura	14
Tabela 6	– Distribuição de frequência da variável notas de português	16
Tabela 7	– Estatísticas da variável notas de português	17
Tabela 8	– Distribuição de frequência da variável notas de matemática	19
Tabela 9	– Estatísticas da variável notas de matemática	20

Sumário

1	SOBRE O SAEB	5
2	CLASSIFICAÇÃO DAS VARIÁVEIS DA AMOSTRA	6
2.1	Variáveis qualitativas	6
2.2	Variáveis quantitativas	6
3	ANÁLISE DE VARIÁVEIS QUALITATIVAS	8
3.1	Nominais	8
3.1.1	Sexo do estudante	8
3.1.2	Raça/Cor dos estudantes	10
3.1.3	Área de localização da escola do estudante	11
3.2	Ordinais	12
3.2.1	Ter computador em casa	12
3.2.2	Uso da biblioteca ou sala de leitura da escola	13
4	ANÁLISE DE VARIÁVEIS QUANTITATIVAS	15
4.1	Contínuas	15
4.1.1	Notas de português dos estudantes	15
4.1.2	Notas de matemática dos estudantes	18
	REFERÊNCIAS	22

1 Sobre o SAEB

O Sistema de Avaliação da Educação Básica - SAEB tem como objetivo avaliar a qualidade do ensino nas escolas públicas do Brasil. Esse desempenho é medido não somente por meio da aplicação de provas de Língua Portuguesa e de Matemática como também são considerados diversos fatores em que os alunos estão inseridos. A amostra trabalhada no presente relatório corresponde ao desempenho de estudantes do quinto ano do ensino fundamental das escolas públicas brasileiras no ano de 2017.

2 Classificação das variáveis da amostra

A amostra analisada engloba variáveis associadas à localidade da escola do estudante, às particularidades de seu domicílio e às características do aluno. Essas variáveis podem ser qualitativas e quantitativas, as quais podem ser subdivididas em nominais/ordinais e discretas/contínuas, respectivamente.

2.1 Variáveis qualitativas

Morettin e Bussab (2017) as variáveis qualitativas são aquelas que representam atributos. Podem ser ordinais ou nominais, caso exista ou não ordenação nas possíveis realizações, respectivamente.

Qualitativas Nominais

- Região de localização da escola do estudante
- Unidade da Federação de localização da escola do estudante
- Município de localização da escola do estudante
- Área de localização da escola do estudante
- Categoria administrativa da escola do estudante
- Localização da escola do estudante
- Sexo do estudante
- Raça/cor do estudante
- Se o estudante mora com a mãe
- Se o estudante mora com a pai

Qualitativas Ordinais

- Ter computador em casa
- Frequência dos pais, ou responsáveis à reunião de pais
- Tempo gasto assistindo à TV, navegando na internet ou jogando jogos eletrônicos em dias de aula
- Tempo gasto fazendo trabalhos domésticos
- Uso da biblioteca ou sala de leitura da sua escola

2.2 Variáveis quantitativas

As variáveis quantitativas são aquelas em que apresentam como possíveis realizações números resultantes de uma contagem ou mensuração. Podem ser discretas ou contínuas. No

primeiro caso, os possíveis valores formam um conjunto finito ou enumerável de números. No segundo, os possíveis valores pertencem a um intervalo de números reais e resultam de uma mensuração (Morettin e Bussab (2017)).

Quantitativas Discretas

- Ano de realização do SAEB
- Idade do estudante

Quantitativas Contínuas

- Notas de português dos estudantes
- Notas de matemática dos estudantes

3 Análise de variáveis qualitativas

Nesse capítulo são apresentadas as análises descritivas das variáveis qualitativas nominais e ordinais elencadas anteriormente.

3.1 Nominais

3.1.1 Sexo do estudante

O gráfico e a tabela a seguir apresentam a frequência absoluta e relativa dos estudantes da amostra por sexo, cujas realizações possíveis são masculino ou feminino.

Figura 1 – Percentual de estudantes por sexo

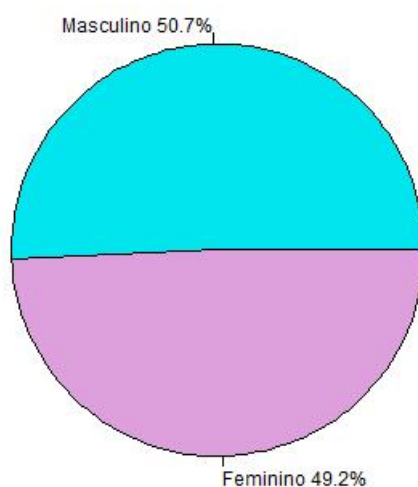


Tabela 1 – Distribuição de frequência da variável sexo

	Frequência Absoluta	Frequência Relativa (%)
Feminino	985	49.25
Masculino	1015	50.75
Total	2000	100

Fonte: INEP.

A partir do gráfico e da tabela é possível observar que há uma pequena diferença entre a porcentagem de estudantes do sexo masculino e feminino. Com um percentual de diferença de apenas 1,5%, o público masculino se torna a maioria da amostra.

3.1.2 Raça/Cor dos estudantes

O gráfico e a tabela a seguir apresentam a distribuição de estudantes por raça/cor, que tem cinco possíveis realizações, quais sejam: amarela, branca, indígena, parda, preta e não quero declarar.

Figura 2 – Percentual de estudantes por raça/cor

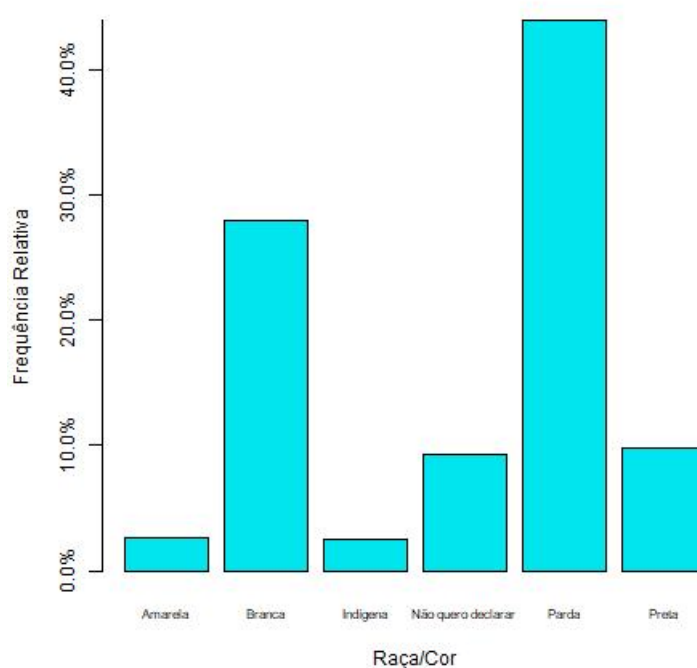


Tabela 2 – Distribuição de frequência da variável raça/cor

	Frequência Absoluta	Frequência Relativa (%)
Amarela	53	2.6
Branca	559	28.0
Indígena	49	2.4
Não quero declarar	185	9.2
Parda	880	44.0
Preta	196	9.8
Total	1922	96.1

Fonte: INEP.

Ao analisar o gráfico e a tabela, é possível concluir que no que tange a variável raça/cor, a maioria se declarou como parda (44%) e branca (28%), respectivamente. As menores porcentagem declaradas foram indígena (2.4%) e amarela (2.6%). Os estudantes que se declararam como pretos correspondem a 9.8%. Cabe destacar que além do percentual

de estudantes que não quiseram declarar (9.2%), há presente na amostra cerca de 3.9% do que são chamados de *missing values*.

3.1.3 Área de localização da escola do estudante

O gráfico e a tabela a seguir apresentam a frequência dos estudantes pela área de localização da escola, que pode ser capital ou interior.

Figura 3 – Percentual de estudantes por área de localização

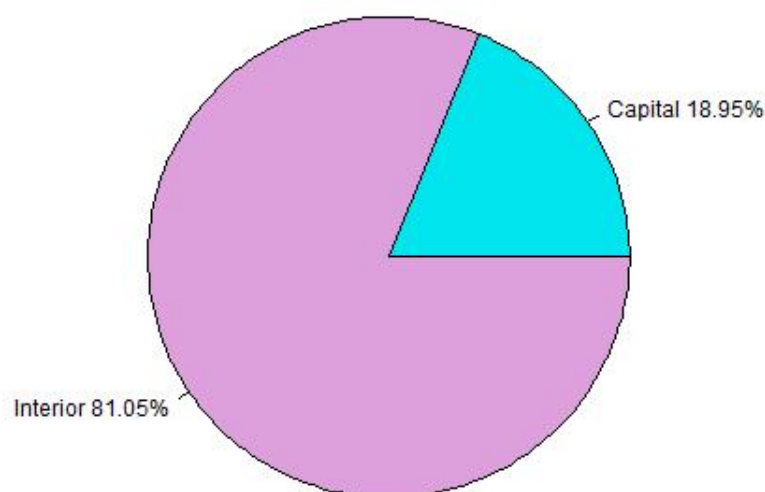


Tabela 3 – Distribuição de frequência da variável área de localização da escola

	Frequência Absoluta	Frequência Relativa(%)
Capital	379	18.95
Interior	1621	81.05
Total	2000	100

Fonte: INEP.

Pelos dados coletados tem-se que a maior parte da amostra engloba alunos cujas escolas estão localizadas no interior (81.05%). Apenas 18.95% das escolas se encontram em capitais.

3.2 Ordinais

3.2.1 Ter computador em casa

O gráfico e a tabela a seguir apresentam a frequência dos estudantes pela quantidade de computadores que possuem em casa. A variável possui cinco realizações, a saber: não tem; sim, um; sim, dois; sim, três ou sim, quatro ou mais.

Figura 4 – Percentual de estudantes por quantidade de computador

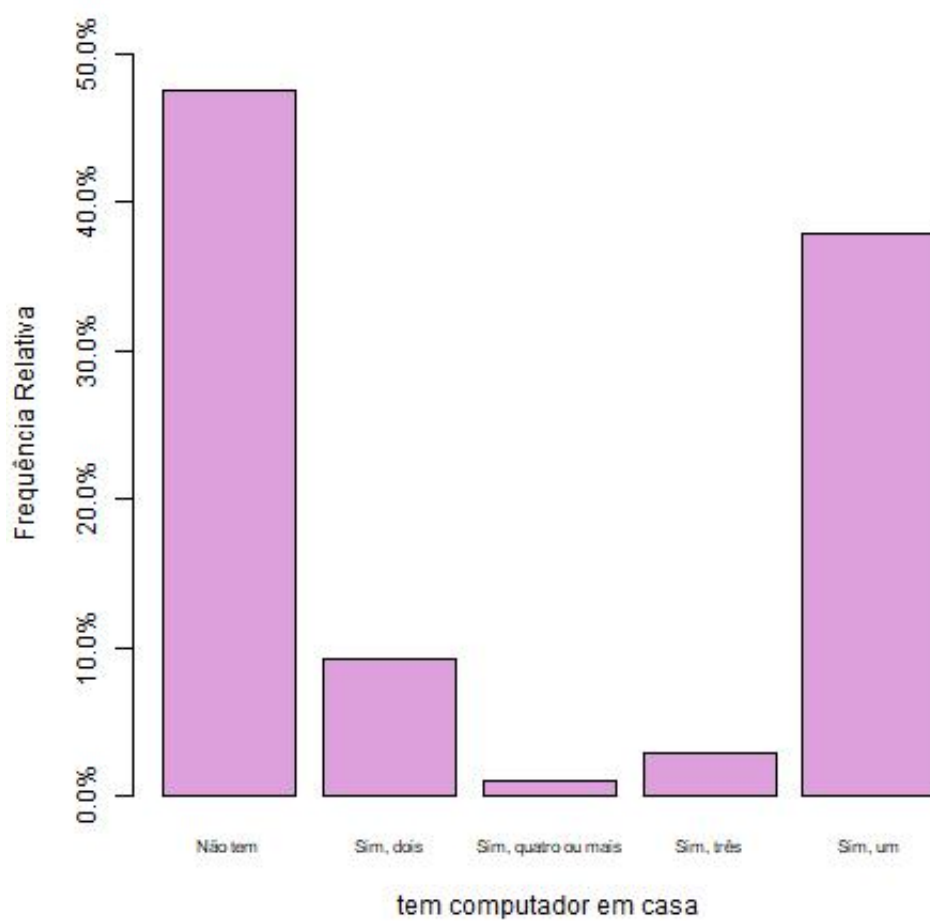


Tabela 4 – Distribuição de frequência da variável ter computador em casa

	Frequência Absoluta	Frequência Relativa (%)
Não tem	949	47.4
Sim, um	758	37.9
Sim, dois	185	9.2
Sim, três	58	2.9
Sim, quatro ou mais	20	1.0
Total	1970	98.5

Fonte: INEP.

A amostra de dados fornece para a variável ter computador em casa conclusões como: a grande maioria dos alunos não possuem computador, com um percentual de quase 50%. Em seguida, nota-se que cerca de 38% dos estudantes possuem um único computador. Uma pequena parte, 1% declararam ter quatro ou mais computadores. Existe ainda uma porcentagem de 1.5% de *missing values*.

3.2.2 Uso da biblioteca ou sala de leitura da escola

O gráfico e a tabela a seguir apresentam a frequência dos estudantes pelo uso da biblioteca ou sala de leitura da escola. Essa variável possui quatro realizações, quais sejam: nunca ou quase nunca; de vez em quando; sempre ou quase sempre; a escola não possui biblioteca.

Figura 5 – Percentual de estudantes por uso da biblioteca

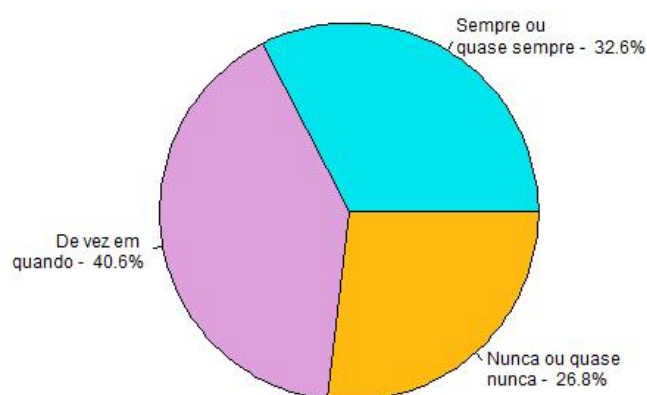


Tabela 5 – Distribuição de frequência da variável uso da biblioteca ou sala de leitura

	Frequência Absoluta	Frequência Relativa(%)
De vez em quando	812	40.6
Nunca ou quase nunca	536	26.8
Sempre ou quase sempre	652	32.6
A escola não possui biblioteca	0	0.0
Total	2000	100

Fonte: INEP.

Por meio do gráfico e tabela, conclui-se que a biblioteca é na maior parte das vezes utilizada de vez em quando pelos estudantes (40.6%). O percentual de alunos que nunca ou quase nunca usam a sala de leitura é de 26.8% enquanto que os alunos que sempre ou quase sempre usam é de 32.6%. Nada foi declarado para a opção "A escola não possui biblioteca", o que resulta em uma porcentagem de 0% para a mesma.

4 Análise de variáveis quantitativas

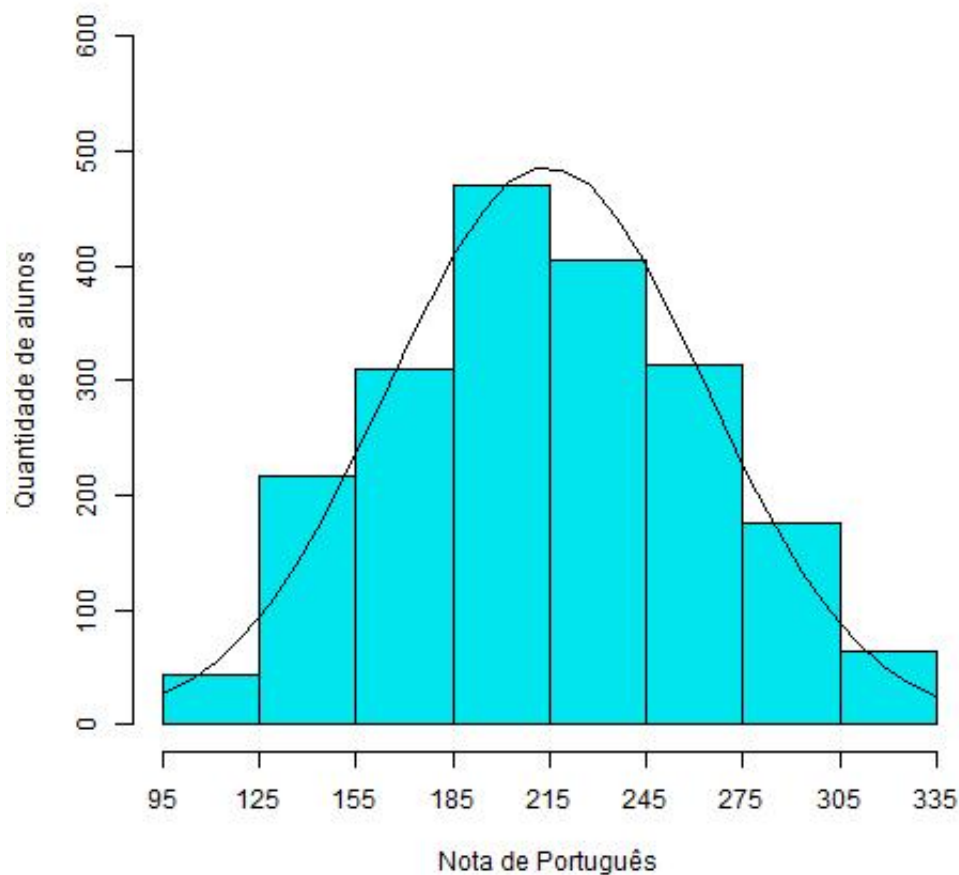
Nesse capítulo são apresentadas as análises descritivas das variáveis quantitativas contínuas elencadas anteriormente.

4.1 Contínuas

4.1.1 Notas de português dos estudantes

O histograma a seguir apresenta a distribuição das notas de português dos estudantes da amostra.

Figura 6 – Histograma da distribuição das notas de português dos estudantes



Ao analisar o histograma é possível observar através das classes há uma simetria dos dados, o que o aproxima de uma normal.

Tabela 6 – Distribuição de frequência da variável notas de português

	Frequência Absoluta	Frequência Relativa(%)
95-125	44	2.2
125-155	217	10.8
155-185	310	15.5
185-215	471	23.6
215-245	405	20.2
245-275	314	15.7
275-305	175	8.8
305-335	64	3.2
Total	2000	100

A partir da tabela de distribuição de frequências nota-se que cerca de 23.6% das notas de português dos estudantes está entre 185 e 215, seguida da classe 215-245 com 20.2%. Os menores percentuais se encontram nos extremos. Cerca de 2.2% dos alunos tem notas entre 95 e 125 enquanto que 3.2% tem notas entre 305 e 335.

Figura 7 – Boxplot da distribuição das notas de português

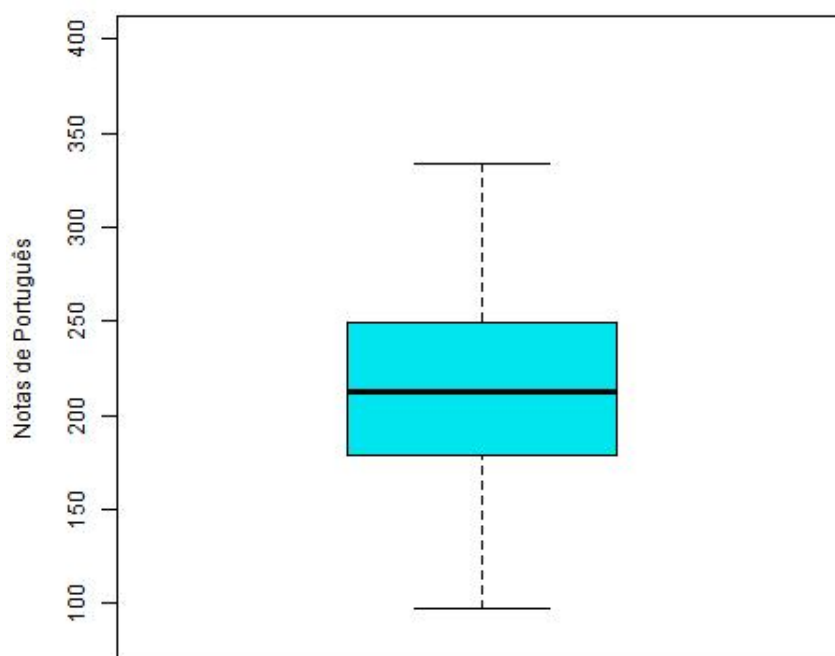


Tabela 7 – Estatísticas da variável notas de português

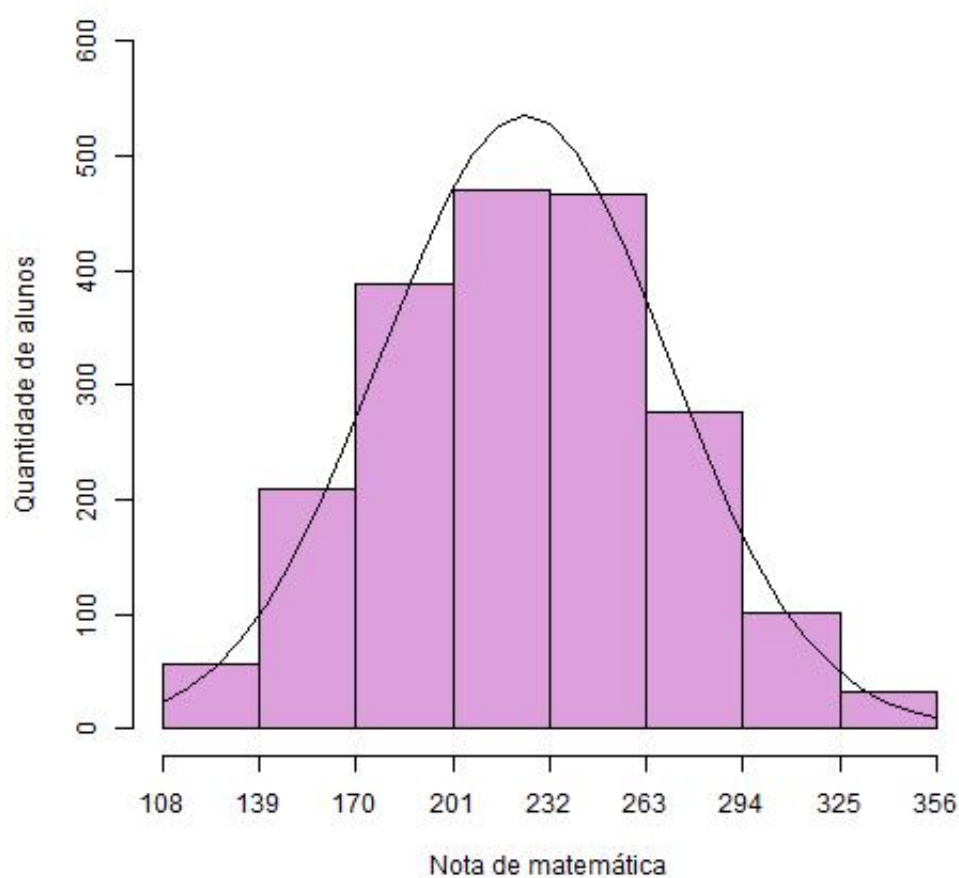
Medida	Valor
Média	214.20
Variância	2420.11
Desvio Padrão	49.19
Mínimo	96.82
Máximo	334.23
Mediana	212.15
Primeiro Quartil	178.26
Terceiro Quartil	249.39
Curtose	2.43
Assimetria	0.12

Ao explorar conclusões que podem ser retiradas a partir da análise do boxplot, é possível observar que todas elas batem com as medidas da tabela 7. Pela figura 7 constata-se que a mediana está um pouco acima do valor 200, o que se torna verossímil com o valor da tabela de 212.15. O terceiro quartil mostra que 75% dos dados estão abaixo de 249.39 enquanto que o primeiro quartil mostra que 25% dos dados estão abaixo de 178.26. A disposição dos dados confirmam uma distribuição simétrica. A partir do valor do coeficiente que indica a curtose averigua-se que a amostra tem uma distribuição leptocúrtica, ou seja, a curva é mais afilada.

4.1.2 Notas de matemática dos estudantes

O histograma a seguir apresenta a distribuição das notas de matemática dos estudantes da amostra.

Figura 8 – Histograma da distribuição das notas de matemática dos estudantes



Assim como no histograma das notas de português nota-se uma simetria dos dados, o que o aproxima de uma normal.

Tabela 8 – Distribuição de frequência da variável notas de matemática

	Frequência Absoluta	Frequência Relativa(%)
108-139	56	2.8
139-170	209	10.4
170-201	389	19.4
201-232	471	23.6
232-263	466	23.3
263-294	276	13.8
294-325	101	5.1
325-356	32	1.6
Total	2000	100

As conclusões para as notas de matemática são semelhantes as das notas de português. Por meio da tabela de distribuição de frequências nota-se que cerca de 23.6% das notas de matemática dos estudantes está entre 201 e 232, seguida da classe 232-263 com 23.3%. Os menores percentuais se encontram nos extremos. Cerca de 2.8% dos alunos possuem notas entre 108 e 139 enquanto que apenas 1.6% tem notas entre 325 e 356.

Figura 9 – Boxplot da distribuição das notas de matemática

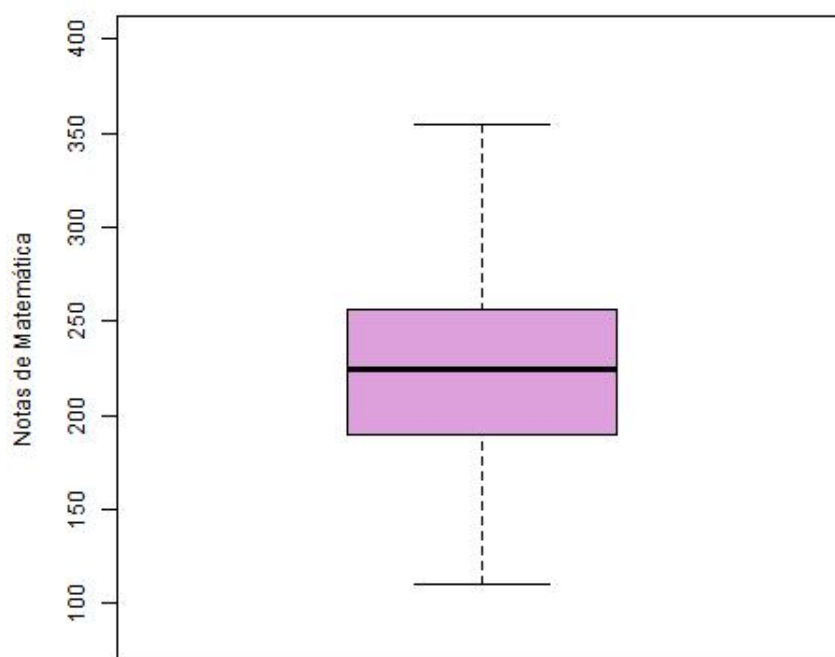


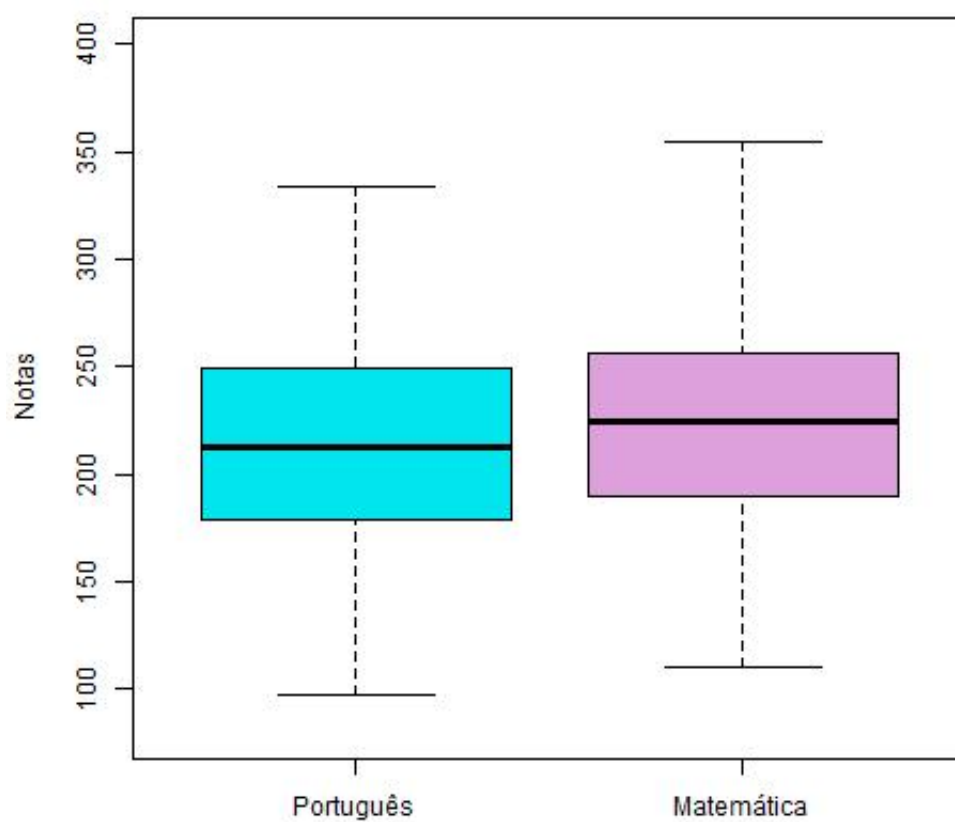
Tabela 9 – Estatísticas da variável notas de matemática

Medida	Valor
Média	223.92
Variância	2134.05
Desvio Padrão	46.20
Mínimo	109.53
Máximo	355.09
Mediana	224.04
Primeiro Quartil	189.68
Terceiro Quartil	255.86
Curtose	2.64
Assimetria	0.11

De modo análogo ao realizado anteriormente, constata-se que as informações que podem ser retiradas a partir da análise do boxplot das notas de matemática se dirigem para a mesma direção que as medidas da tabela 9. Pela figura 9 é possível deduzir que a mediana está um pouco acima do valor 200, o que é verossímil com o valor da tabela de 224.04. O terceiro quartil mostra que 75% dos dados estão abaixo de 255.86 enquanto que o primeiro quartil mostra que 25% dos dados estão abaixo de 189.68. A disposição dos dados confirmam uma distribuição simétrica. Assim como para as notas de português, o valor do coeficiente que indica a curtose define a distribuição como leptocúrtica, ou seja, a curva é mais afilada.

Em geral, tem-se que as notas de matemática são um pouco mais elevadas que as notas de português. A figura a seguir mostra a comparação entre os seus respectivos boxplots.

Figura 10 – Boxplots da distribuição das notas de português e matemática



Referências

MORETTIN, P. A.; BUSSAB, W. O. *Estatística básica*. [S.l.]: Saraiva Educação SA, 2017.
Citado 2 vezes nas páginas 6 e 7.