



Allan Victor Almeida Faria, Hermes Jasper Winarski, Kassyano Kevyn Andrade  
de Souza

## **Relatório de Pesquisa**

Brasília, DF

Dezembro, 2020



Allan Victor Almeida Faria, Hermes Jasper Winarski, Kassyano Kevyn Andrade  
de Souza

## **Relatório de Pesquisa**

Relatório de Pesquisa do projeto Saeb 2017.

Universidade de Brasília (UnB)  
Instituto de Ciências Exatas (IE)  
Departamento de Estatística (DE)

Brasília, DF  
Dezembro, 2020

# Resumo

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Nisl tincidunt eget nullam non nisi est sit amet. Turpis nunc eget lorem dolor. Posuere urna nec tincidunt praesent semper feugiat nibh. Platea dictumst quisque sagittis purus sit amet volutpat consequat mauris. Velit scelerisque in dictum non consectetur a. Lectus proin nibh nisl condimentum. Nulla pharetra diam sit amet.

Risus in hendrerit gravida rutrum quisque non tellus. Purus gravida quis blandit turpis cursus in hac habitasse. Et malesuada fames ac turpis egestas maecenas pharetra convallis posuere. Id donec ultrices tincidunt arcu non sodales neque. Suspendisse faucibus interdum posuere lorem ipsum dolor sit amet consectetur. Amet purus gravida quis blandit. Sit amet nisl purus in mollis nunc sed. Rutrum quisque non tellus orci. Pellentesque eu tincidunt tortor aliquam nulla facilisi cras fermentum. Lacus suspendisse faucibus interdum posuere lorem ipsum dolor sit amet.

A iaculis at erat pellentesque adipiscing commodo elit. Tellus rutrum tellus pellentesque eu tincidunt tortor aliquam nulla. Eros donec ac odio tempor orci dapibus ultrices. Vitae ultricies leo integer malesuada. Aliquam eleifend mi in nulla posuere sollicitudin aliquam. Vitae purus faucibus ornare suspendisse sed nisi lacus sed.

Consequat interdum varius sit amet mattis. Elementum pulvinar etiam non quam lacus suspendisse faucibus. Morbi non arcu risus quis varius quam quisque id diam. Est ante in nibh mauris cursus mattis molestie a iaculis. Odio aenean sed adipiscing diam donec adipiscing tristique risus nec. Mi tempus imperdiet nulla malesuada pellentesque. Pellentesque habitant morbi tristique senectus et netus et. Luctus venenatis lectus magna fringilla urna porttitor rhoncus. Integer enim neque volutpat ac tincidunt vitae. Viverra ipsum nunc aliquet bibendum.

**Palavras-chaves:** 1. Saeb 2017. 2. Amostra. 3. Pesquisa.

# Lista de ilustrações

Figura 1 – Proporção por sexo de tempo em afazeres domésticos por parte dos alunos. . . . .	12
Figura 2 – Proporção total por nível de escolaridade da mãe com base no tempo de afazeres dos alunos. . . . .	13
Figura 3 – Distribuições das somas das notas com base na raça/cor dos alunos. . .	17
Figura 4 – Distribuições empíricas das somas das notas com base nas localizações das das escolas dos alunos. . . . .	18
Figura 5 – Distribuições das somas das notas com base nas escolaridades das mães dos alunos. . . . .	19

# Lista de tabelas

Tabela 1 – Testes de igualdade na variabilidade sobre as relações com o tempo de afazeres domésticos por parte dos alunos. . . . .	14
Tabela 2 – Comparações dois a dois entre as ordens das posições sobre os tempos de afazeres domésticos com base na escolaridade das mães dos alunos. .	15
Tabela 3 – Testes para as relações com soma das notas dos alunos. . . . .	20
Tabela 4 – Comparações dois a dois entre as médias sobre a soma das notas com base na raça/cor dos alunos. . . . .	21
Tabela 5 – Comparações entre as médias de notas em Matemática e as regiões das escolas dos alunos com base na amostra de tamanho 500. . . . .	22

# Lista de abreviaturas e siglas

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

# Lista de símbolos

$\sigma$  Letra grega minúscula sigma

$\mu$  Letra grega minúscula mu

# Sumário

	<b>Introdução</b> . . . . .	<b>8</b>
<b>I</b>	<b>METODOLOGIA</b>	<b>9</b>
<b>1</b>	<b>METODOLOGIA</b> . . . . .	<b>10</b>
<b>II</b>	<b>RESULTADOS</b>	<b>11</b>
<b>2</b>	<b>COMPARAÇÕES</b> . . . . .	<b>12</b>
<b>2.1</b>	<b>Tempo em afazeres domésticos</b> . . . . .	<b>12</b>
<b>2.2</b>	<b>Notas</b> . . . . .	<b>16</b>
<b>3</b>	<b>CONCLUSÃO</b> . . . . .	<b>23</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>24</b>
	<b>ANEXOS</b>	<b>25</b>
	<b>ANEXO A – AMOSTRA</b> . . . . .	<b>26</b>





## Introdução

Este documento tem como objetivo analisar os fatores sociais de alunos brasileiros do 9º ano de 2017, sobre os quais avalia possibilidades de diferenças no desempenho do aprendizado básico destes. Para a análise, serão utilizados dados do Sistema de Avaliação da Educação Básica (SAEB) de 2017 divulgado pelo [INEP \(2018\)](#). O banco de dados utilizado no estudo é constituído por dados de estudantes do 9º ano que realizaram a Prova Brasil no ano de 2017.

O SAEB avalia o desempenho em Matemática e Língua Portuguesa por meio de testes e questionários sobre fatores sociais de alunos, aplicados bienalmente em redes públicas e amostras de escolas privadas sobre turmas do 5º e 9º anos do ensino fundamental e 3º ano do ensino médio.

Esta análise parte de uma amostragem aleatória simples de 5.271 alunos deste banco de dados, e relaciona fatores como raça/cor, sexo dos alunos, localizações das escolas e escolaridade da mãe destes, com as variáveis a serem explicadas como a soma destas notas e o tempo de afazeres domésticos realizados por dia com base nos alunos.

A aplicação dos testes de hipóteses para a análise da amostra, bem como dos gráficos (para a análise descritiva), utiliza ferramentas do software R nas versões 3.6.3 e 4.0.3. Parte do processamento dos dados foi feito com o software PYTHON versão 3.7.3.<sup>1</sup>

---

<sup>1</sup> Pacotes externos usados para a manipulação dos dados:  
R: *tidyverse*, *data.table*, *reshape2*, *patchwork*, *EnvStats*, *PMCMR*;  
Python: *pandas*

# Parte I

## Metodologia



# 1 Metodologia

Como variáveis explicativas para a soma das notas, o estudo relaciona o sexo do aluno, a raça/cor, a escolaridade da mãe e as localizações das escolas, e avalia o quanto essas variáveis influenciam no desempenho na Prova Brasil. Para avaliar estas relações, um estudo prévio foi realizado com amostras de tamanhos 30, 50 e 100 dos 5.271 alunos em que, com os testes [Anderson e Darling \(1954\)](#), [Shapiro e Wilk \(1965\)](#) e [Shapiro e Francia \(1972\)](#), se concluiu que as distribuições das notas são normais. Este resultado será utilizado para atender os testes paramétricos que pressupõem normalidade dos dados. Os testes a serem aplicados serão o teste ANOVA Fisher por [Fisher \(1928\)](#) para avaliar os fatores com mais de 2 categorias e o teste t-Student para as comparações dois a dois ([O'CONNOR; ROBERTSON; EDMUND, 1908](#)). Estes testes utilizam das médias ( $\bar{x}$ ) de cada grupo para avaliar se as distâncias são significativas entre eles, no qual previamente avalia se as variâncias ( $S^2$ ) destes são iguais (Teste B), através do teste proposto por [Bartlett \(1954\)](#).

Quanto ao tempo gasto com afazeres domésticos, estas mesmas variáveis são exploradas, com exceção das localizações das escolas, para observar se há indícios de diferenças sociais entre a disponibilidade de tempo em casa para outras possíveis tarefas na formação do aprendizado básico. Para estas relações, os testes estatísticos não paramétricos são apropriados, e utiliza-se o teste de [Kruskal e Wallis \(1952\)](#) para fatores com mais de duas categorias (Teste K), e o teste de [Mann e Whitney \(1947\)](#) para a comparação dois a dois das categorias (Teste W). Estes testes avaliam as distribuições das informações com base na posição, para verificar se as distâncias entre as categorias são significativas.

Para avaliar os resultados dos testes, foi proposto o uso da correção de [Bonferroni \(1936\)](#) para os testes com mais de duas categorias. Se houver evidências para rejeitar igualdade destas, a comparação dois a dois é efetuada e a correção é utilizada. Esta correção é sobre o p-valor, que é avaliado em uma escala de significância, adotado por este estudo como uma confiança de 95%, no que diz respeito a aceitar a hipótese nula  $H_0$ .

Parte II

Resultados

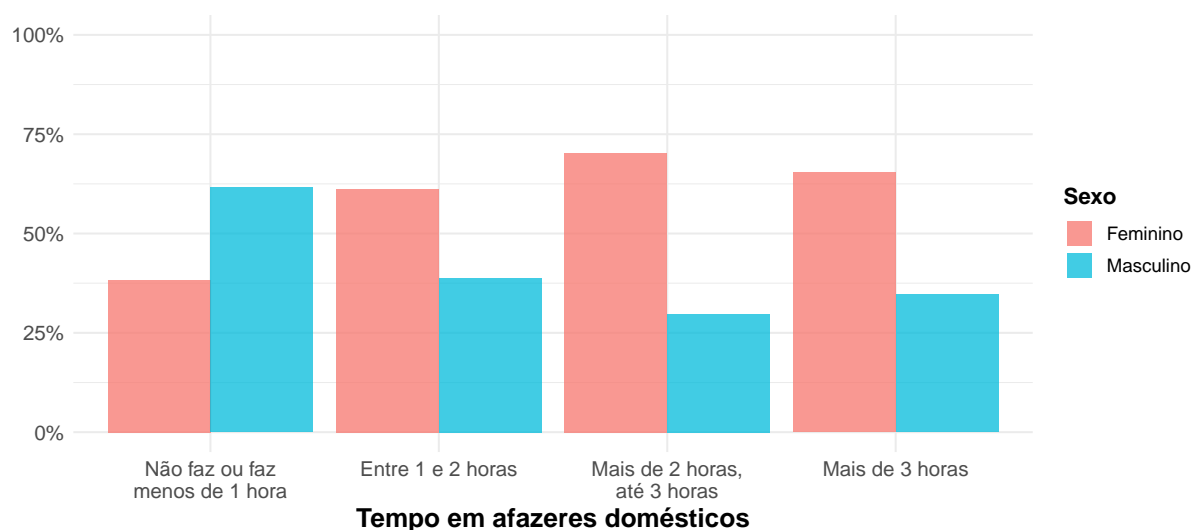
## 2 Comparações

### 2.1 Tempo em afazeres domésticos

O tempo gasto com afazeres domésticos pode privar o aluno de exercer o estudo do ensino básico. Foram feitas análises sociais com base nas diferenças entre os períodos de tempo gastos diariamente nestas atividades.

A [Figura 1](#) apresenta como se distribuem estes períodos em relação ao sexo do aluno, no qual estudantes do sexo feminino tendem a gastar mais tempo com atividades domésticas. O único período em que o sexo feminino teve menos representatividade foi a categoria que menos tempo diário é gasto com atividades domésticas, "Não faz ou faz menos de 1 hora". Para os períodos em que pelo menos uma hora por dia é gasta, o sexo feminino representa no mínimo 60% de todos os indivíduos, o que pode indicar que a proporção de estudantes do sexo masculino possui mais disponibilidade de tempo em casa para estudos.

Figura 1 – Proporção por sexo de tempo em afazeres domésticos por parte dos alunos.



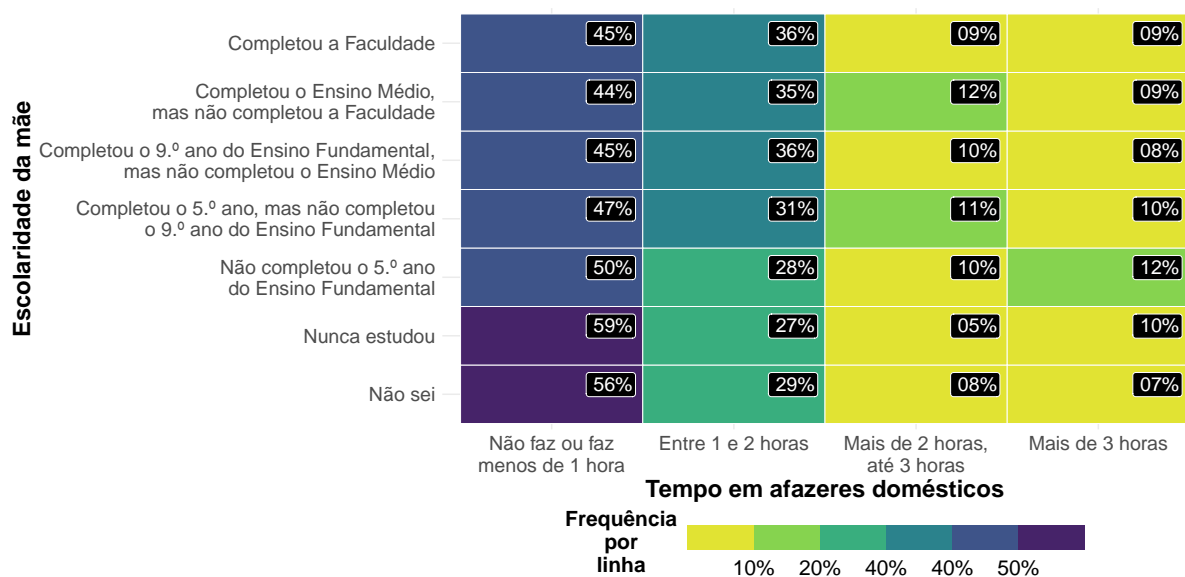
Fonte: Amostra de 5.271 alunos do 9º ano do SAEB 2017.

Nota: Amostra retirada de uma amostragem aleatórias simples.



Ao comparar os períodos de tempo nestes afazeres com o nível escolaridade da mãe (Figura 2), a proporção de alunos que exercem estas atividades, para cada nível escolar da mãe, apresenta pelo menos 70% localizado entre os que não fazem ou fazem até 2 horas.

Figura 2 – Proporção total por nível de escolaridade da mãe com base no tempo de afazeres dos alunos.



Fonte: Amostra de 5.271 alunos do 9º ano do SAEB 2017.

Nota: Amostra retirada de uma amostragem aleatórias simples.

Ao realizar testes estatísticos com base nesses afazeres (Tabela 1) e com a confiança de 95%, a variável sexo obteve evidências significativas para afirmar que o tempo médio não é igual entre os sexos, no qual o sexo feminino possui proporções superiores em períodos de tempo maiores nestas atividades. Ao testar a variável raça/cor dos alunos, não foram obtidas evidências significativas, apontando que não há indícios de diferenças, em média, entre as raças/cores, no tempo gasto nestes afazeres.



Tabela 1 – Testes de igualdade na variabilidade sobre as relações com o tempo de afazeres domésticos por parte dos alunos.

Teste	$H_0$	P-valor	Decisão de $H_0$ (95%)
K	$\mu_{Raça/Cor}$ iguais	0.369	Aceita
K	$\mu_{Esc(mãe)}$ iguais	Aprox. 0	Rejeita
W	$\mu_M = \mu_F$ iguais	Aprox. 0	Rejeita

Fonte: Amostra de 5.271 alunos do 9º ano do SAEB 2017.

Nota: Amostra retirada de uma amostragem aleatória simples.

Anotações: Os subíndices M e F referem-se, respectivamente, aos sexos Masculino e Feminino dos alunos. Esc (mãe) diz respeito à escolaridade da mãe destes. O Aprox. 0 refere-se a um número muito pequeno considerado por este estudo aproximadamente zero.

O teste realizado com base na escolaridade da mãe sobre o tempo em afazeres domésticos, ao mesmo nível de confiança, obteve evidências significativas sobre existência de diferença no tempo gasto com afazeres domésticos. Ao efetuar testes pareados para cada nível escolar da mãe com a confiança de 95% ([Tabela 2](#)), a proporção de alunos que gastam tempo nestes afazeres, para as mães que nunca estudaram se difere com todos os outros níveis escolares que, em grande parte gastam menos de uma hora. O mesmo se observa para as mães com o 5º ano incompleto, no qual o único grau que não se difere é dos alunos que responderam que não sabiam sobre o nível educacional da mãe. As mães que completaram o 9º obtiveram igualdade apenas com as mães do 5º ano, e aqueles que responderam que não sabiam. Para aquelas mães que completaram a faculdade, há igualdade entre as que completaram o ensino médio, onde o tempo de afazeres por parte dos alunos para as mães com o nível de escolaridade do 5º ano ou superior possui maiores proporções entre 1 e 2 horas que entre as escolaridades abaixo desta.



Tabela 2 – Comparações dois a dois entre as ordens das posições sobre os tempos de afazeres domésticos com base na escolaridade das mães dos alunos.

Comparações	P-valor	Evidência (RA 95%)
Não sabe = Nunca estudou	Aprox. 0	Desiguais
Não sabe = Incompleto 5.º ano do EF	1.0000	Iguais
Não sabe = Completou 5.º ano do EF	0.0084	Desiguais
Não sabe = Completou 9.º ano do EF	0.1927	Iguais
Não sabe = Completou EM	Aprox. 0	Desiguais
Não sabe = Completou Faculdade	Aprox. 0	Desiguais
Nunca estudou = Incompleto 5.º ano do EF	0.0038	Desiguais
Nunca estudou = Completou 5.º ano do EF	Aprox. 0	Desiguais
Nunca estudou = Completou 9.º ano do EF	Aprox. 0	Desiguais
Nunca estudou = Completou EM	Aprox. 0	Desiguais
Nunca estudou = Completou Faculdade	Aprox. 0	Desiguais
Incompleto 5.º ano do EF = Completou 5.º ano do EF	0.0002	Desiguais
Incompleto 5.º ano do EF = Completou 9.º ano do EF	0.0048	Desiguais
Incompleto 5.º ano do EF = Completou EM	Aprox. 0	Desiguais
Incompleto 5.º ano do EF = Completou Faculdade	Aprox. 0	Desiguais
Completo 5.º ano do EF = Completou 9.º ano do EF	1	Iguais
Completo 5.º ano do EF = Completou EM	Aprox. 0	Desiguais
Completo 5.º ano do EF = Completou Faculdade	Aprox. 0	Desiguais
Completo 9.º ano do EF = Completou EM	Aprox. 0	Desiguais
Completo 9.º ano do EF = Completou Faculdade	Aprox. 0	Desiguais
Completou EM = Completou Faculdade	1.0000	Iguais

Fonte: Amostra de 5.271 alunos do 9º ano do SAEB 2017.

Nota: Amostra retirada de uma amostragem aleatória simples.

Anotações: Aprox. 0 refere-se à algum número muito pequeno considerando aproximadamente zero.

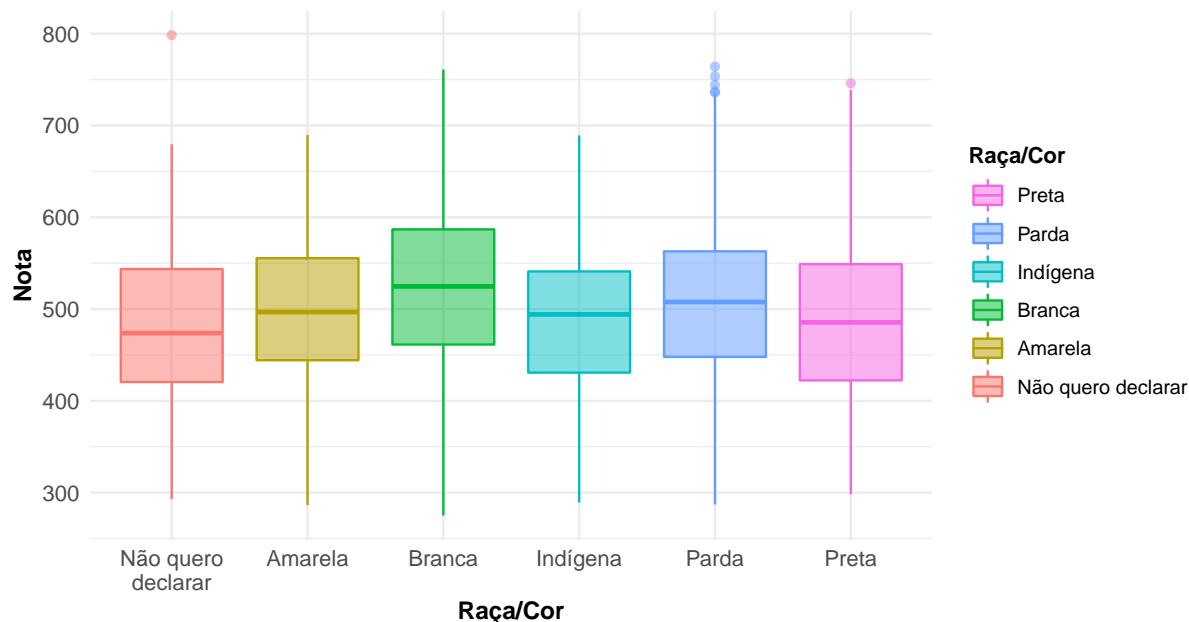




## 2.2 Notas



Figura 3 – Distribuições das somas das notas com base na raça/cor dos alunos.



Fonte: Amostra de 5.271 alunos do 9º ano do SAEB 2017.

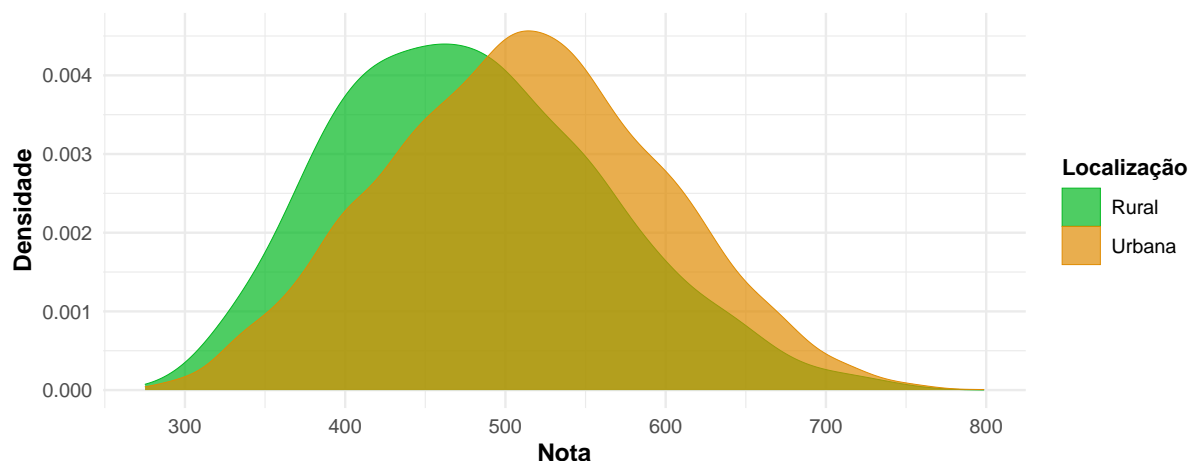
Nota: Amostra retirada de uma amostragem aleatórias simples.

Ao se observar a [Figura 3](#), nota-se que a soma das notas dos grupos tem uma tendência de medida central em 500, que varia pouco de acordo com a raça, sendo o fenótipo Branco aquele que possui as maiores medidas de posição.

A raça Parda é a que possui o maior número de outliers entre os estudantes, mas a maior nota observada esteve presente no grupo que não quis declarar raça/cor.

A [Figura 4](#) mostra as distribuições das somas das notas por localização.

Figura 4 – Distribuições empíricas das somas das notas com base nas localizações das das escolas dos alunos.



Fonte: Amostra de 5.271 alunos do 9º ano do SAEB 2017.

Nota: Amostra retirada de uma amostragem aleatórias simples.

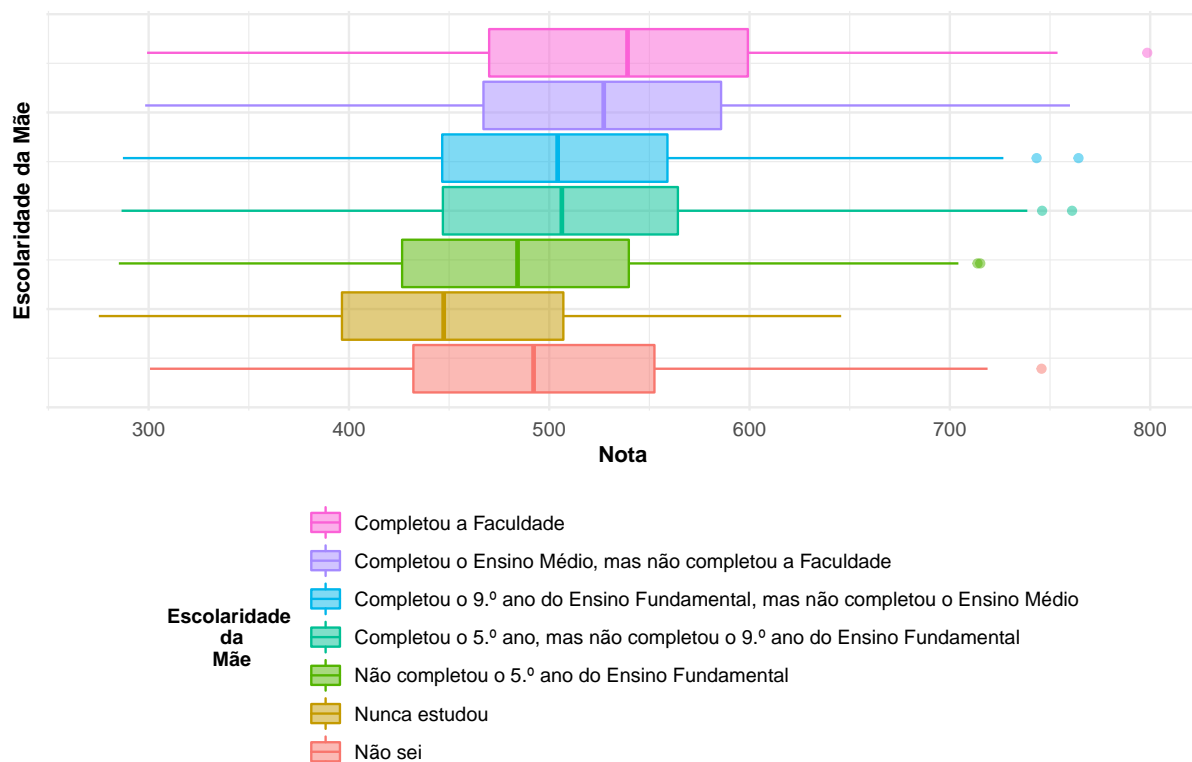
É possível observar um grau de assimetria um pouco maior nas notas para a região rural em relação à região urbana. A zona rural teve sua distribuição um pouco mais inclinada para a esquerda, enquanto a zona urbana foi mais centralizada. É possível observar que a moda das notas foi inferior na região rural em relação à região urbana.

A área rural apresentou uma nota média de 479, que também é inferior à nota média da área urbana, que foi de 512.



A Figura 5 mostra a distribuição das notas em relação à escolaridade das mães.

Figura 5 – Distribuições das somas das notas com base nas escolaridades das mães dos alunos.



Fonte: Amostra de 5.271 alunos do 9º ano do SAEB 2017.

Nota: Amostra retirada de uma amostragem aleatórias simples.

É possível observar um crescimento das notas no geral à medida que o grau de escolaridade das mães é maior, de modo que filhos cujas mães têm grau de escolaridade alto têm tendência a terem melhor desempenho em provas. Esse comportamento é observado mesmo entre os outliers, onde a maior nota registrada vem da parte de um aluno cuja mãe completou a faculdade.



Tabela 3 – Testes para as relações com soma das notas dos alunos.

Teste	$H_0$	P-valor	Decisão de $H_0$ (95%)
B	$\sigma_R^2 = \sigma_U^2$	0.503	Aceita
B	$\sigma_M^2 = \sigma_F^2$	0.002	Rejeita
B	$\sigma_{Raça/Cor}^2$ iguais	0.265	Aceita
B	$\sigma_{Esc(mãe)}^2$ iguais	0.132	Aceita
T	$\mu_R = \mu_U$	Aprox. 0	Rejeita
T	$\mu_M = \mu_F$	0.905	Aceita
ANOVA	$\mu_{Raça/Cor}$ iguais	Aprox. 0	Rejeita
ANOVA	$\mu_{Esc(mãe)}$ iguais	Aprox. 0	Rejeita

Fonte: Amostra de 5.271 alunos do 9º ano do SAEB 2017.

Nota: Amostra retirada de uma amostragem aleatória simples.

Anotações: Os subíndices, com base nos alunos,  $R$  e  $U$  refere-se as localizações das escolas rurais e urbanas,  $M$  e  $F$  sobre os sexos Masculino e Feminino respectivamente e  $Esc(mãe)$  diz respeito a escolaridade da mãe. O Aprox. 0 refere-se à algum número muito pequeno considerado por este estudo aproximadamente zero.



A [Tabela 4](#) mostra os resultados da comparação dois a dois entre as notas por raça/cor.

Tabela 4 – Comparações dois a dois entre as médias sobre a soma das notas com base na raça/cor dos alunos.

Comparações	P-valor	Evidência (RA 95%)
Amarela = Não quero declarar	0.4113	Iguais
Amarela = Branca	0.0005	Desiguais
Amarela = Indígena	1.0000	Iguais
Amarela = Parda	1.0000	Iguais
Amarela = Preta	1.0000	Iguais
Branca = Não quero declarar	Aprox. 0	Desiguais
Branca = Indígena	0.0010	Desiguais
Branca = Parda	Aprox. 0	Desiguais
Branca = Preta	Aprox. 0	Desiguais
Indígena = Não quero declarar	1.0000	Iguais
Indígena = Parda	0.7758	Iguais
Indígena = Preta	1.0000	Iguais
Parda = Não quero declarar	Aprox. 0	Desiguais
Parda = Preta	Aprox. 0	Desiguais

Fonte: Amostra de 5.271 alunos do 9º ano do SAEB 2017.

Nota: Amostra retirada de uma amostragem aleatória simples.

Anotações: Aprox. 0 refere-se à algum número muito pequeno considerando aproximadamente zero.

A raça branca apresentou diferença nos testes par a par entre todas as outras raças, confirmando o que a [Figura 3](#) apontava. Há uma clara desigualdade entre as notas por raça quando se compara a branca com as demais, e quase todas as desigualdades derivaram dela.

As únicas desigualdades que não foram entre comparações de indivíduos de cor branca com outros foram a comparação entre indivíduos pardos e pretos, e também a comparação entre indivíduos pardos que não quiseram declarar.

Todos os outros testes não só levam à aceitação da hipótese nula, como apresentaram um p-valor significativamente alto.



Tabela 5 – Comparações entre as médias de notas em Matemática e as regiões das escolas dos alunos com base na amostra de tamanho 500.

Comparações	P-valor	Evidência (RA 95%)
Não sabe = Nunca estudou	1.0000	Iguais
Não sabe = Incompleto 5.º ano do EF	0.0078	Desiguais
Não sabe = Completou 5.º ano do EF	0.0005	Desiguais
Não sabe = Completou 9.º ano do EF	0.0001	Desiguais
Não sabe = Completou EM	Aprox. 0	Desiguais
Não sabe = Completou Faculdade	0.0011	Desiguais
Nunca estudou = Incompleto 5.º ano do EF	1.0000	Iguais
Nunca estudou = Completou 5.º ano do EF	0.5598	Iguais
Nunca estudou = Completou 9.º ano do EF	0.4165	Iguais
Nunca estudou = Completou EM	0.1114	Iguais
Nunca estudou = Completou Faculdade	0.4707	Iguais
Incompleto 5.º ano do EF = Completou 5.º ano do EF	1.0000	Iguais
Incompleto 5.º ano do EF = Completou 9.º ano do EF	1.0000	Iguais
Incompleto 5.º ano do EF = Completou EM	1.0000	Iguais
Incompleto 5.º ano do EF = Completou Faculdade	1.0000	Iguais
Completo 5.º ano do EF = Completou 9.º ano do EF	1.0000	Iguais
Completo 5.º ano do EF = Completou EM	1.0000	Iguais
Completo 5.º ano do EF = Completou Faculdade	1.0000	Iguais
Completo 9.º ano do EF = Completou EM	1.0000	Iguais
Completo 9.º ano do EF = Completou Faculdade	1.0000	Iguais
Completou EM = Completou Faculdade	1.0000	Iguais

Fonte: Amostra de 5.271 alunos do 9º ano do SAEB 2017.

Nota: Amostra retirada de uma amostragem aleatória simples.

Anotações: Aprox. 0 refere-se à algum número muito pequeno considerando aproximadamente zero.



## 3 Conclusão

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Nisl tincidunt eget nullam non nisi est sit amet. Turpis nunc eget lorem dolor. Posuere urna nec tincidunt praesent semper feugiat nibh. Platea dictumst quisque sagittis purus sit amet volutpat consequat mauris. Velit scelerisque in dictum non consectetur a. Lectus proin nibh nisl condimentum. Nulla pharetra diam sit amet.

Risus in hendrerit gravida rutrum quisque non tellus. Purus gravida quis blandit turpis cursus in hac habitasse. Et malesuada fames ac turpis egestas maecenas pharetra convallis posuere. Id donec ultrices tincidunt arcu non sodales neque. Suspendisse faucibus interdum posuere lorem ipsum dolor sit amet consectetur. Amet purus gravida quis blandit. Sit amet nisl purus in mollis nunc sed. Rutrum quisque non tellus orci. Pellentesque eu tincidunt tortor aliquam nulla facilisi cras fermentum. Lacus suspendisse faucibus interdum posuere lorem ipsum dolor sit amet.

A iaculis at erat pellentesque adipiscing commodo elit. Tellus rutrum tellus pellentesque eu tincidunt tortor aliquam nulla. Eros donec ac odio tempor orci dapibus ultrices. Vitae ultricies leo integer malesuada. Aliquam eleifend mi in nulla posuere sollicitudin aliquam. Vitae purus faucibus ornare suspendisse sed nisi lacus sed.

Consequat interdum varius sit amet mattis. Elementum pulvinar etiam non quam lacus suspendisse faucibus. Morbi non arcu risus quis varius quam quisque id diam. Est ante in nibh mauris cursus mattis molestie a iaculis. Odio aenean sed adipiscing diam donec adipiscing tristique risus nec. Mi tempus imperdiet nulla malesuada pellentesque. Pellentesque habitant morbi tristique senectus et netus et. Luctus venenatis lectus magna fringilla urna porttitor rhoncus. Integer enim neque volutpat ac tincidunt vitae. Viverra ipsum nunc aliquet bibendum.





## Referências

ANDERSON, T. W.; DARLING, D. A. A test of goodness of fit. *Journal of the American statistical association*, Taylor & Francis, v. 49, n. 268, p. 765–769, 1954. Citado na página 10.

BARTLETT, M. S. A note on the multiplying factors for various  $\chi^2$  approximations. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 296–298, 1954. Citado na página 10.

BONFERRONI, C. Teoria statistica delle classi e calcolo delle probabilit . *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, v. 8, p. 3–62, 1936. Citado na p gina 10.

FISHER, R. A. The general sampling distribution of the multiple correlation coefficient. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, The Royal Society London, v. 121, n. 788, p. 654–673, 1928. Citado na p gina 10.

INEP. *SAEB 2017 MICRODADOS DA ANEB E DA ANRESC (PROVA BRASIL)*: Saeb (aneb/prova brasil 2017). [S.l.], 2018. 33 p. Dispon vel em: <<http://portal.inep.gov.br/web/guest/microdados>>. Citado na p gina 8.

KRUSKAL, W. H.; WALLIS, W. A. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, Taylor & Francis Group, v. 47, n. 260, p. 583–621, 1952. Citado na p gina 10.

MANN, H. B.; WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, JSTOR, p. 50–60, 1947. Citado na p gina 10.

O’CONNOR, J. J.; ROBERTSON, E. F.; EDMUND, F. Student’s t-test. *MacTutor History of Mathematics archive*, 1908. Citado na p gina 10.

SHAPIRO, S. S.; FRANCIA, R. An approximate analysis of variance test for normality. *Journal of the American Statistical Association*, Taylor & Francis Group, v. 67, n. 337, p. 215–216, 1972. Citado na p gina 10.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). *Biometrika*, JSTOR, v. 52, n. 3/4, p. 591–611, 1965. Citado na p gina 10.

# Anexos



## ANEXO A – Amostra

Os bancos de dados dos alunos participantes do SAEB de 2017 foram disponibilizados como amostras de tamanho 2.000 em formato CSV. Cada banco de dados era uma amostra aleatória diferente do banco de dados original do SAEB, para cada aluno da Turma A de Métodos Estatísticos 2 do semestre 1/2020 da Universidade de Brasília (UnB).

Os dados utilizados desse documento foram a junção dos três bancos de dados dos integrantes, com os seguintes nomes: **amostra\_190015853.csv**, **amostra\_190029498.csv** e **amostra\_190127180.csv**