



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

Este estudo tem como base uma amostra de 500 alunos do 9º ano de 2017 do Brasil disponibilizada em um banco de dados da Prova Brasil realizada pelo Saeb em 2017. Com o objetivo de avaliar o ensino básico, através das notas dos alunos, realiza-se testes estatísticos para as relações das notas em Matemática com a regiões das escolas e das notas em Língua portuguesa com os tempos de uso de "telas" (Ex.: Televisores, celulares e entre outros), no qual com critérios de confiabilidade, obteve resultados de desigualdades regionais e pontuações nas notas influenciadas por fatores relacionados ao tempo de uso de telas.

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

Lista de ilustrações

| | |
|--|----|
| Figura 1 – Proporção por sexo de tempos de afazeres domésticos por parte dos alunos. | 14 |
| Figura 2 – Distribuições das somas das notas com base na raça/cor dos alunos. . . | 18 |
| Figura 3 – Distribuições empíricas das somas das notas com base nas localizações das das escolas dos alunos. | 19 |
| Figura 4 – Distribuições das somas das notas com base nas escolaridades das mães dos alunos. | 20 |

Lista de tabelas

| | |
|--|----|
| Tabela 1 – Escala de significância de Fisher. | 11 |
| Tabela 2 – Testes de igualdade na variabilidade sobre as relações com o tempo de afazeres domésticos por parte dos alunos. | 15 |
| Tabela 3 – Comparações dois a dois entre as ordens das posições sobre os tempos de afazeres domésticos com base na escolaridade das mães dos alunos. . | 16 |
| Tabela 4 – Testes para as relações com soma das notas dos alunos. | 21 |
| Tabela 5 – Comparações dois a dois entre as médias sobre a soma das notas com base na raça/cor dos alunos. | 22 |
| Tabela 6 – Comparações entre as médias de notas em Matemática e as regiões das escolas dos alunos com base na amostra de tamanho 500. | 23 |

Sumário

| | | |
|------------|-------------------------------------|-----------|
| | Introdução | 6 |
| I | FUNDAMENTAÇÃO | 7 |
| 1 | ESTATÍSTICAS | 8 |
| 1.1 | Medidas-Resumo | 8 |
| 1.2 | Testes de comparação | 9 |
| 1.2.1 | Kruskal-Wallis | 9 |
| 1.2.2 | Fisher | 9 |
| 1.2.2.1 | Bartlett | 9 |
| 1.2.3 | t-Student | 10 |
| 2 | METODOLOGIA | 11 |
| II | RESULTADOS | 12 |
| 3 | COMPARAÇÕES | 13 |
| 3.1 | Tempo de afazeres domésticos | 13 |
| 3.2 | Notas | 17 |
| 4 | CONCLUSÃO | 24 |
| | ANEXOS | 25 |
| | ANEXO A – AMOSTRA | 26 |



Introdução

O objetivo deste estudo é avaliar o ensino básico e, por meio de técnicas estatísticas, testar as hipóteses de diferenças para as relações consideradas às notas dos alunos do 9º ano de 2017, de um estudo realizado pelo Sistema de Avaliação da Educação Básica (SAEB), no intuito de avaliar o ensino básico de forma geral (??). A intenção é inferir as desigualdades substâncias para a compreensão das relações dos alunos do 9º ano e, posteriormente, poder realizar políticas de aprimoramento de ensino.

Parte I

FUNDAMENTAÇÃO



1 Estatísticas

Este estudo usa Métodos estatísticos para obtenção e comparação de valores, de modo que se possa criar hipóteses que serão testadas, para que seja possível afirmar com um grau de certeza o comportamento e uma relação entre as variáveis.

1.1 Medidas-Resumo

As medidas-resumos são a base desse estudo para a compreensão, através de valores, a identidade dos dados, no qual o foco do estudo é obter utilizando métodos e testes a ponto de resumí-los e compará-los para identificar supostas hipóteses com relação aos dados. É possível assemelhar a matemática compreendida na observação em uma visão de possibilidade válida para o comportamento de tal forma a tentar entender a realidade. Formulas dada por ??, cap. 2, 3 e 11):

$$x \in X, \forall x \in \{x_1, \dots, x_n\}$$

$$Média = \bar{x} = \sum_{i=1}^n \frac{x_i}{n} \quad (1.1)$$

$$Mediana = md(X) = \begin{cases} x_{(\frac{n+1}{2})}, & \text{se } n \text{ ímpar} \\ \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}, & \text{se } n \text{ par} \end{cases} \quad (1.2)$$

$$Desvio Padrão Amostral = S = \sqrt{\frac{\sum_{i=1}^n (\frac{x_i}{n} - \bar{x})^2}{n-1}} \quad (1.3)$$

$$1^\circ \text{ Quartil} = q_1 = \sum_{i=1}^k \frac{n_i}{n} = 25\% \quad (1.4)$$

$$3^\circ \text{ Quartil} = q_3 = \sum_{i=1}^k \frac{n_i}{n} = 75\% \quad (1.5)$$



1.2 Testes de comparação

Os testes de comparação a seguir, propostos para o estudo, têm como objetivo avaliar a relação de duas populações através do teste do t-Student (teste T), ou para mais populações utilizando os teste de Kruskal-Wallis (Teste K) e o teste de Fisher (ANOVA), no qual se presume a independência entre categorias propostas para a análise e avalia diferenças substanciais entre as variáveis relacionadas, com base na amostra do banco de dados dos alunos do 9º ano de 2017.

1.2.1 Kruskal-Wallis

O teste proposto por ??), utiliza a ideia de ranqueamento dos valores, no qual usa-se para comparar mais de duas populações sem a confirmação dos dados serem "normais". Estatística do teste dada por:

$$K = \frac{1}{S^2} \sum_{j=1}^r \frac{R_j^2}{n_j} - \frac{n(n+1)^2}{4}$$

1.2.2 Fisher

O famoso teste da análise de variabilidade (ANOVA) dado por ??), aplica para os dados considerados como "normais" e com a mesma variância, a comparação das médias das populações sendo como iguais.

1.2.2.1 Bartlett

O teste de ??) foi proposto para analisar se as variâncias (S^2) das populações são iguais (Homocedasticidade), no qual possibilita a aplicação do teste de Fisher. Estatística dada por:

$$B = \frac{M}{C}$$

onde,

$$M = (n - r) \ln S_e^2 - \sum_{i=1}^r (n_i - 1) \ln S_i^2$$

e



$$C = 1 + \frac{1}{3(r-1)} \left[\sum_{i=1}^r \left(\frac{1}{n_i - 1} \right) - \left(\frac{1}{n - r} \right) \right]$$

1.2.3 t-Student

O T-student (??) tem como base o teste paramétrico, no qual este estudo usa o métodos de comparação de duas populações cujo as variâncias (S^2) são iguais. A estatística de teste é:

$$T = \frac{\bar{X} - \bar{Y}}{S_p \sqrt{1/n_X + 1/n_Y}}$$

onde,

$$S_p^2 = \frac{(n_X - 1)S_X^2 + (n_Y - 1)S_Y^2}{n_X + n_Y - 2}$$

2 Metodologia

Este estudo parte da hipótese de normalidade das notas, analisada por outros estudos referentes ao banco de dados do SAEB de 2017 divulgada pelo ??). O objetivo desse estudo é relacionar variáveis de uma amostra de 2000 alunos deste banco de dados e analisar possíveis diferenças substanciais através de testes estatísticos.

Nas relações propostas pelo estudo, foi feita uma junção de três amostras de tamanho 2.000, e os valores em branco foram omitidos, gerando uma amostra de 5.271 observações.

O P-Valor é a base desse estudo para a decisão da hipótese, no qual a confirmação da hipótese nula (H_0) é avaliada com um grau de significância (ou confiabilidade), considerado na análise. Se este assumir valores menores que um menos a percentagem de confiança considerada, há evidência de recusar a hipótese H_0 (??, pag. 364). A aplicação dos testes de hipóteses para a análise da amostra de 5271 alunos utiliza ferramentas dos softwares R e PYTHON com pacotes¹ para a implementação das análises.

Tabela 1 – Escala de significância de Fisher.

| Evidência | marginal | moderada | substancial | forte | muito forte | fortíssima |
|-----------|----------|----------|-------------|-------|-------------|------------|
| P-valor | 0,10 | 0,05 | 0,025 | 0,01 | 0,005 | 0,001 |

Fonte: ??, p. 364).

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

Parte II

Resultados

3 Comparações

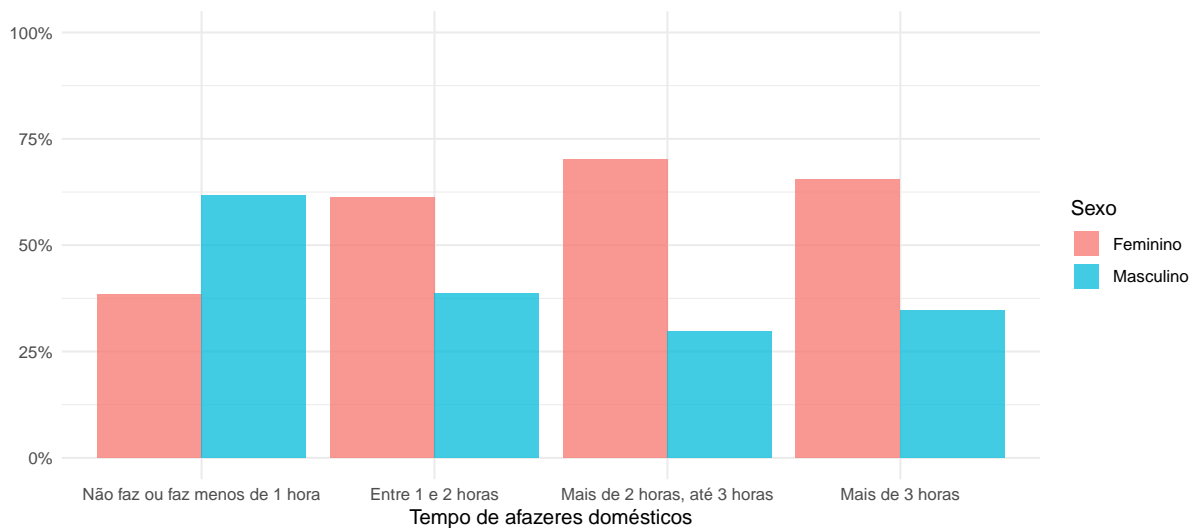
blablabalbalbalbalbalblablabbbbbbblbalbalbalblalbalbalblalbalblaba blablabalbal-
balbalbalblalblabbbbbbblbalbalbalblalbalbalblalbalblaba blablabalbalbalbalbalblalblabbbb-
blbalbalbalblalbalbalblalbalblabal albalblalblalblal

3.1 Tempo de afazeres domésticos

blablalbalbalbalbalblablalbbbbbbblbalbalbalblalbalbalblalbalblaba blablalbalbal-
balbalbalblablalbbbbbbblbalbalbalblalbalbalblalbalblaba blablalbalbalbalbalbalblablalbbbbbb-
blbalbalbalblalbalbalblalbalblabal albalblablalblalbla



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



Fonte: Amostra de 5271 alunos do 9º ano do SAEB 2017.

Nota: Amostra retirada de uma amostragem aleatórias simples.



Tabela 2 – 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 |
| K | $\mu_M = \mu_F$ iguais | Aprox. 0 | Rejeita |

Fonte: Amostra de 5271 alunos do 9º ano do SAEB 2017.

Nota: Amostra retiradas de uma amostragem aleatórias simples.

Anotações: Os subíndices M e F refere-se aos sexos Masculino e Feminino dos alunos respectivamente. O Aprox. 0 refere-se à algum número muito pequeno considerado por este estudo aproximadamente zero.



Tabela 3 – 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 5271 alunos do 9º ano do SAEB 2017.

Nota: Amostra retiradas de uma amostragem aleatórias simples.

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

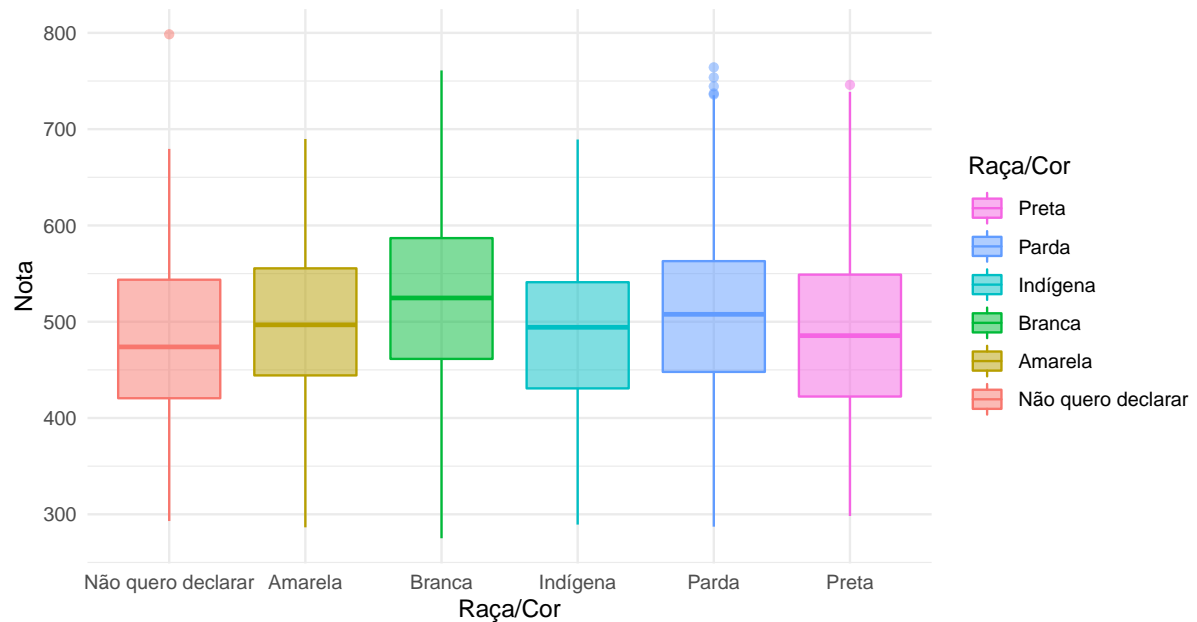


3.2 Notas

blablalbalbalbalbalbalblablalbbbbbbblbalbalbalblalbalbalblalbalblaba blablalbalbal-
balbalbalblablalbbbbbbblbalbalbalblalbalbalblalbalblaba blablalbalbalbalbalbalblablalbbbb-
blbalbalbalblalbalbalblalbalblabal albalblablalblala



Figura 2 – Distribuições das somas das notas com base na raça/cor dos alunos.

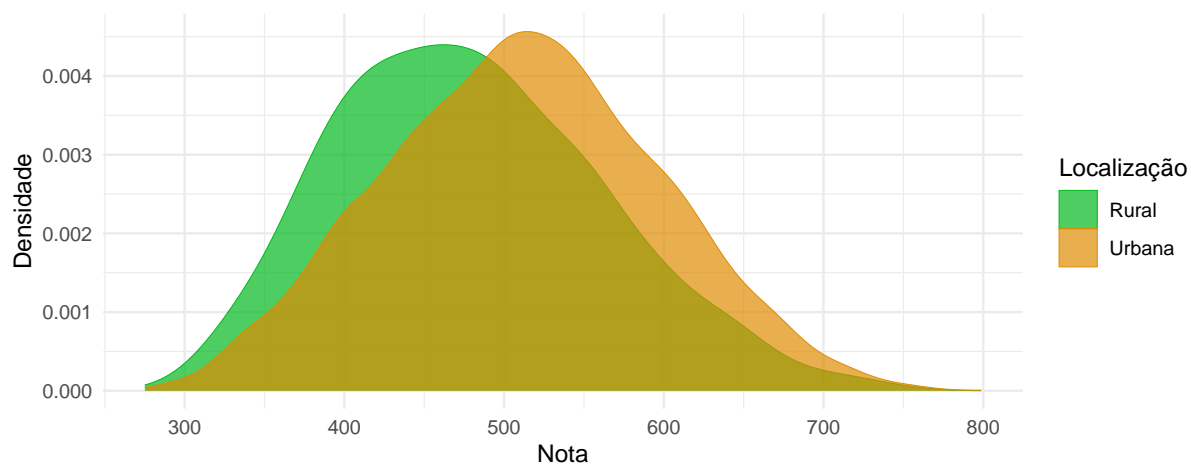


Fonte: Amostra de 5271 alunos do 9º ano do SAEB 2017.

Nota: Amostra retirada de uma amostragem aleatórias simples.



Figura 3 – Distribuições empíricas das somas das notas com base nas localizações das das escolas dos alunos.

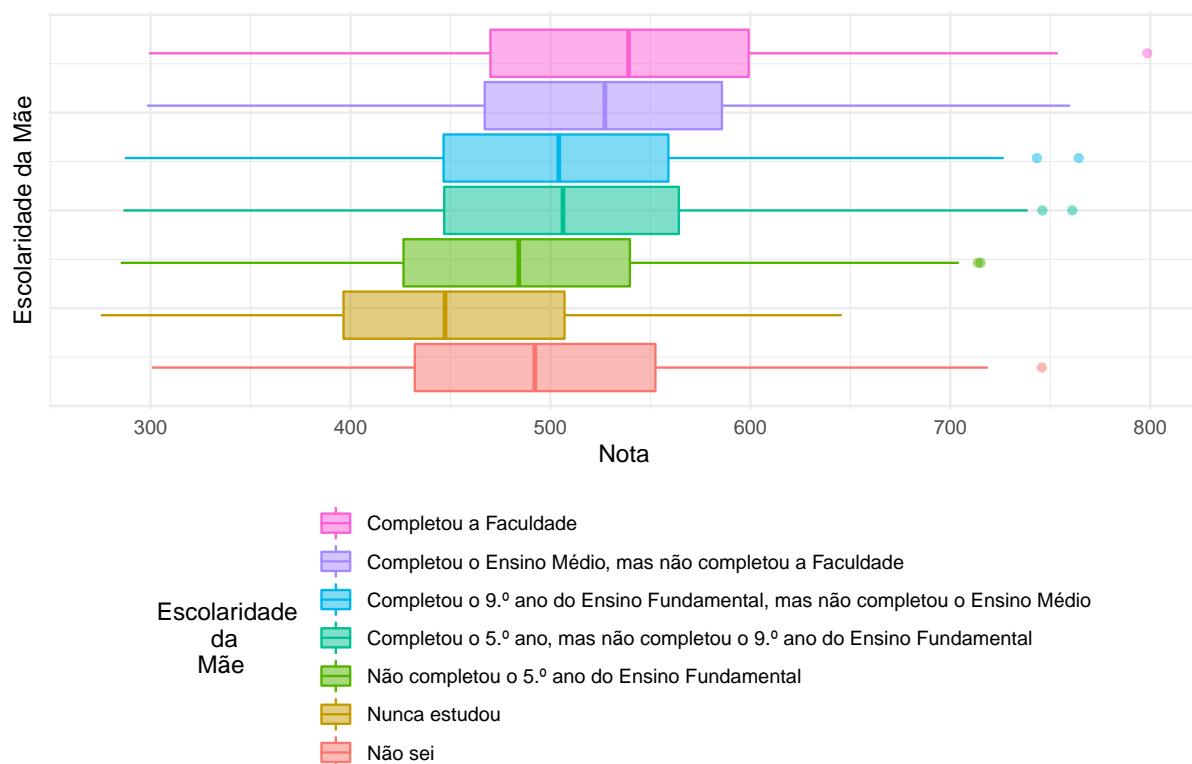


Fonte: Amostra de 5271 alunos do 9º ano do SAEB 2017.

Nota: Amostra retirada de uma amostragem aleatórias simples.



Figura 4 – Distribuições das somas das notas com base nas escolaridades das mães dos alunos.



Fonte: Amostra de 5271 alunos do 9º ano do SAEB 2017.

Nota: Amostra retirada de uma amostragem aleatórias simples.



Tabela 4 – 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 | $\sigma_{Raça/Cor}^2$ iguais | 0.265 | Aceita |
| B | $\sigma_{Esc(mãe)}^2$ iguais | 0.132 | Aceita |
| ANOVA | $\mu_R^2 = \mu_U^2$ iguais | Aprox. 0 | Rejeita |
| ANOVA | $\mu_{Raça/Cor}$ iguais | Aprox. 0 | Rejeita |
| ANOVA | $\mu_{Esc(mãe)}$ iguais | Aprox. 0 | Rejeita |

Fonte: Amostra de 5271 alunos do 9º ano do SAEB 2017.

Nota: Amostra retiradas de uma amostragem aleatórias simples.

Anotações: Os subíndices R e U refere-se as localizações das escolas rurais e urbanas respectivamente. O Aprox. 0 refere-se à algum número muito pequeno considerado por este estudo aproximadamente zero.



Tabela 5 – 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 5271 alunos do 9º ano do SAEB 2017.

Nota: Amostra retiradas de uma amostragem aleatórias simples.

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



Tabela 6 – 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 5271 alunos do 9º ano do SAEB 2017.

Nota: Amostra retiradas de uma amostragem aleatórias simples.

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

4 Conclusão

De acordo com os resultados do estudo, há fatores que comprovam a diferença de qualidade do ensino básico para as turmas do 9º ano de 2017 do Brasil. Sobre as notas em Língua Portuguesa, o fator de tempo de uso de telas por parte dos alunos, há diferenças substanciais para poder afirmar com 95% de confiança, que os alunos que usam menos de 1 hora ou não usam, obtiveram notas menores que comparado com os demais que usa as telas por mais tempo. As notas em Matemática relacionada a região das escolas dos alunos, com a mesma confiança, observou-se que alunos da regiões Sul e Centro-oeste detém de maiores notas que comparados com as regiões Norte e Nordeste, havendo evidências de desigualdade regional por parte do ensino básico em Matemática para o 9º ano, com base nas associações com as notas avaliativas realizados pelos alunos na prova Brasil em 2017.

Anexos



ANEXO A – Amostra

O Banco de dados dos alunos participantes do Saeb de 2017 foi disponibilizado como amostras de 2000 alunos em formato Excel referêntes à cada matrícula para os cursandos de Métodos Estatísticos 2 da Universidade de Brasília (Unb), no qual este documento analisa o arquivo: **amostra_190127180.csv**.