

Allan Victor Almeida Faria

Relatório de Pesquisa

Brasília, DF
Setembro, 2020

Allan Victor Almeida Faria

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
Setembro, 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

Lista de tabelas

Sumário

Introdução

O objetivo deste estudo é avaliar o ensino básico por meio de técnicas estatísticas e 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 de forma geral o ensino básico de forma geral [?]. A intenção é inferir as desigualdades substâncias para o entendimento 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 na obtenção e comparação de valores a ponto de poder criar testes de hipóteses no qual é possível afirmar com um grau de certeza o comportamento e uma relação entre os dados.

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 resumir-los e comparar-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]morettin2017estatistica:

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

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

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

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

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

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

1.2 Testes de comparação

Os testes a seguir de comparação propostos para o estudo, tem como objetivo avaliar a relação de duas populações através do teste do T-student (teste T) e para mais populações utilizando os teste de Kruskal-Wallis (Teste K) e o teste de Fisher (ANOVA), no qual 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 kruskal1952use, 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 fisher1928general, 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 bartlett1954note 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 do pressuposto de normalidade analisado por outros estudos sobre as notas dos alunos do 9º ano de 2017 do Brasil, referente ao banco de dados do SAEB de 2017 divulgada pelo Saeb2017a). Pelo qual esse tem como objetivo 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 efetuado uma amostragem aleatória simples de 500 alunos desta amostra, no qual relacionou as notas em Matemática com as regiões das escolas e as notas em Língua Portuguesa com o tempo de uso de telas (Televisões, Celulares, Computadores e entre outros) com base nestes alunos.

Os testes aplicados utilizam a o pressuposto de normalidade, no qual aplica os testes, de Bartlett (para igualdade de variâncias), de Fisher (para a igualdade na variabilidade das médias) e o teste não paramétrico de Kruskal-Wallis (para a igualdade de posições), com base no fator de correção do P-valor dado por bonferroni1936teoria. O teste T-Student é usado no pareamento dois a dois para a identificação de desigualdade de médias, utilizando a mesma correção do P-valor.

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 500 alunos utiliza ferramentas do SOFTWARE R 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]morettin2017estatistica.

¹ Pacotes externos usados para a manipulação dos dados: *readr*, *tidyverse*, *fBasics*, *reshape2*, *EnvStats*, *goftest*, *nortest*, *MASS* e *ggplo2*

Parte II

Resultados

3 Comparações

Tabela 2 – 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 3 – 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 4 – 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 9^o 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 5 – 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	$\mu_{Raga/Cor}$ 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 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 2000 alunos do 9º ano do SAEB 2017.

Nota: Amostra retiradas de uma amostragem aleatórias simples de tamanho 500.

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**.