



Universidade de Brasília

Tailine Nonato

**COMPARAÇÃO ENTRE DUAS
POPULAÇÕES: RESULTADOS DO SAEB
(ANEB - PROVA BRASIL) 2017**

Brasília, DF

Abril, 2021

Tailine Nonato

**COMPARAÇÃO ENTRE DUAS POPULAÇÕES:
RESULTADOS DO SAEB (ANEB - PROVA BRASIL)
2017**

Relatório de comparação entre duas populações apresentado à disciplina Métodos Estatísticos 2 da Universidade de Brasília como objeto de avaliação.

Brasília, DF

Abril, 2021

Lista de ilustrações

Figura 1 – Boxplot de Notas em Matemática por Localização	8
Figura 2 – Boxplot de Notas em Língua Portuguesa por Dependência Administrativa	10
Figura 3 – Boxplot de Notas em Matemática por Localização	12
Figura 4 – Boxplot de Notas em Língua Portuguesa por Dependência Administrativa	14

Lista de tabelas

Tabela 1 – Testes para Notas em Matemática por Localização	9
Tabela 2 – Testes para Notas em Língua Portuguesa por Dependência Administrativa	11
Tabela 3 – Teste para Notas em Matemática por Localização	13
Tabela 4 – Teste para Notas em Língua Portuguesa por Dependência Administrativa	15

Sumário

	Introdução	5
1	METODOLOGIA	6
2	RESULTADOS	7
2.1	Testes para normalidade e homocedasticidade	7
2.2	Amostra de tamanho 100	8
2.2.1	Associação entre Notas em Matemática e Localização	8
2.2.2	Associação entre Notas em Língua Portuguesa e Dependência Administrativa	10
2.3	Amostras de tamanho 30	12
2.3.1	Associação entre Notas em Matemática e Localização	12
2.3.2	Associação entre Notas em Língua Portuguesa e Dependência Administrativa	14
3	CONCLUSÃO	16
	REFERÊNCIAS	17
	ANEXO A – CÓDIGO EM R	18

Introdução

Utilizando os dados do Sistema de Avaliação da Educação Básica (SAEB) de 2017 divulgado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP (2018)), esta pesquisa tem por objetivo associação entre variáveis e se há diferença entre as notas em Língua Portuguesa e em Matemática.

1 Metodologia

O banco de dados em questão dispõe de 2000 observações em 19 variáveis. No relatório anterior (Atividade 2) foram retiradas 50 amostras aleatórias de tamanho 30 e 50 amostras aleatórias de tamanho 100, assim, para este relatório serão utilizadas a primeira amostra de tamanho 30 e a primeira amostra de tamanho 100 retiradas anteriormente.

As variáveis analisadas serão:

- NOTA_LP: Proficiência em Língua Portuguesa transformada na escala única do SAEB;
- NOTA_MT: Proficiência do aluno em Matemática transformada na escala única do SAEB;
- LOCALIZAÇÃO: Localização da escola do estudante(Urbana ou Rural);
- DEPENDENCIA_ADM: Categoria administrativa da escola do estudante (Federal, Estadual, Municipal ou Privada).

O nível de significância considerado foi de 5%.

Para execução das análises, foi utilizado o Software *RStudio* com *R* versão 4.0.3 e com auxílio dos pacotes *tidyverse*.

2 Resultados

A fim de explorar associação entre variáveis e se há diferença entre as notas em Língua Portuguesa e em Matemática, as análises e resultados estão divididos em três subseções: testes para normalidade e homocedasticidade, amostras de 100 observações e amostras de 30 observações.

2.1 Testes para normalidade e homocedasticidade

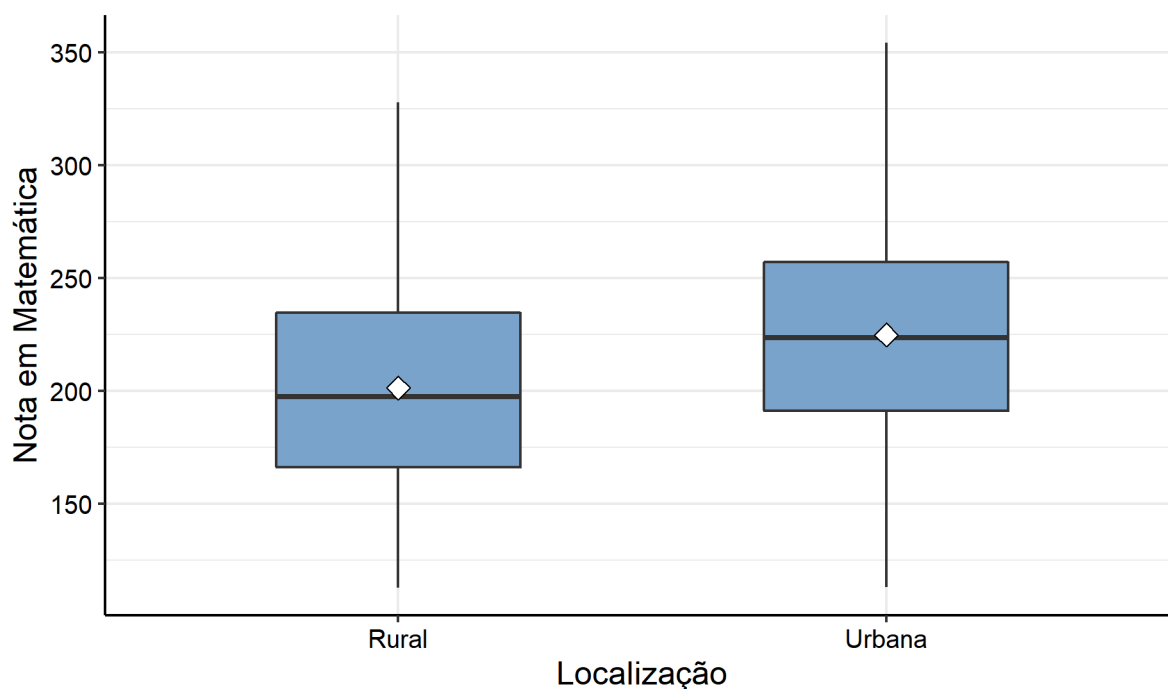
Para prosseguir com os testes de associação entre variáveis é necessário entender o comportamento das mesmas, assim, pela *Atividade 3.2* tem-se que:

- Amostra de tamanho 100
 - Notas em Matemática: Não seguem distribuição normal.
- Amostra de tamanho 30
 - Notas em Matemática: Seguem distribuição normal.
 - Notas em Língua Portuguesa: Seguem distribuição normal.

2.2 Amostra de tamanho 100

2.2.1 Associação entre Notas em Matemática e Localização

Figura 1 – Boxplot de Notas em Matemática por Localização



Fonte: INEP, SAEB, 2017.

Quadro 2.1 – Quadro Resumo de Notas em Matemática por Localização

Estatística	Urbana	Rural
Média	224,7	201,3
Mínimo	112,9	112,7
1º Quartil	191,1	166,3
Mediana	223,6	197,3
3º Quartil	257,1	234,6
Máximo	354,4	328
Desvio padrão	45,94	44,21

Fonte: INEP, SAEB, 2017.

Pela visualização do Gráfico 1 e pelos dados do Quadro 2.1, pode-se observar que as Notas em Matemáticas são um pouco maiores para estudantes de localização urbana. Tal fato indica possível associação. Assim, para testar se há associação entre as notas em Matemática e localização dos estudantes, tem-se as seguintes hipóteses:

$$\begin{cases} H_0 : \text{As variáveis Notas em Matemática e Localização estão associadas.} \\ H_a : \text{Não há associação entre Notas em Matemática e Localização.} \end{cases}$$

Para analisar as hipóteses serão feitos a seguir os testes T de Student, Wilcoxon-Mann-Whitney e Kolmogorov-Smirnov.

Tabela 1 – Testes para Notas em Matemática por Localização

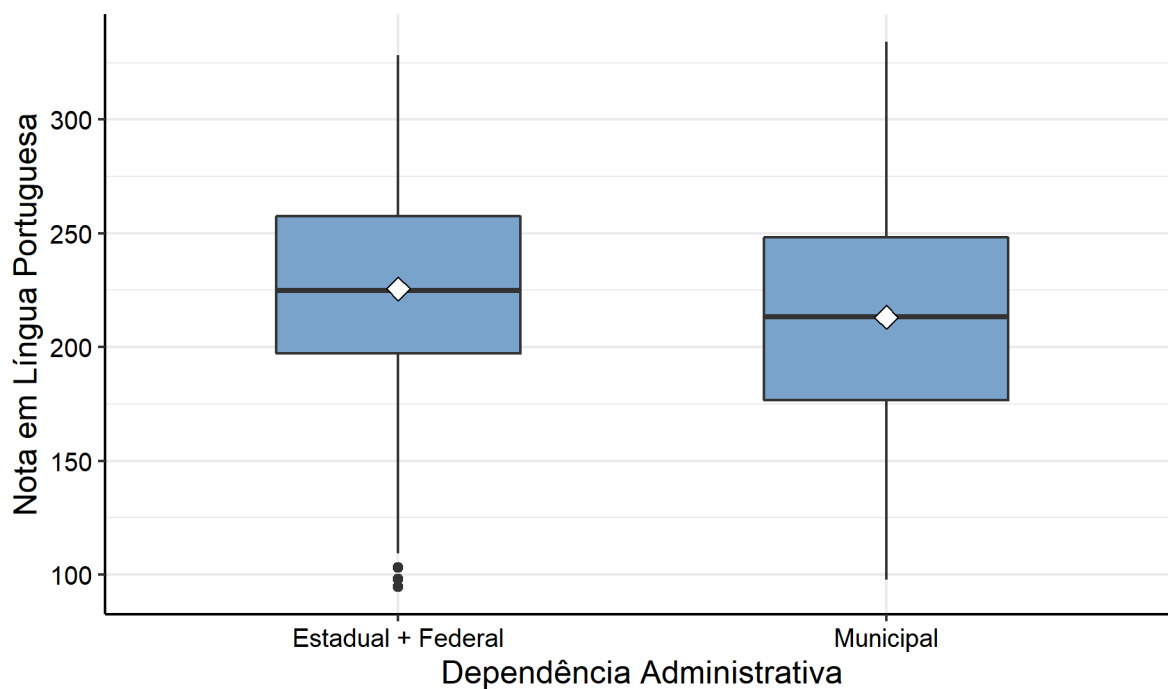
Teste	P-valor	Decisão do teste
T de Student	$< 2,2 * 10^{-16}$	Rejeita-se H_0
Wilcoxon-Mann-Whitney	$< 2,2 * 10^{-16}$	Rejeita-se H_0
Kolmogorov-Smirnov	$< 2,2 * 10^{-16}$	Rejeita-se H_0

Fonte: INEP, SAEB, 2017.

Para todos os testes obteve-se o mesmo p-valor que, por sua vez, é menor que o nível de significância (5%). Assim, há evidências para dizer que as notas em Matemática e a localização do estudantes não estão associadas.

2.2.2 Associação entre Notas em Língua Portuguesa e Dependência Administrativa

Figura 2 – Boxplot de Notas em Língua Portuguesa por Dependência Administrativa



Fonte: INEP, SAEB, 2017.

Quadro 2.2 – Quadro Resumo de Notas em Língua Portuguesa por Dependência Administrativa

Estatística	Estadual+Federal	Municipal
Média	225,66	213,14
Mínimo	94,61	97,85
1º Quartil	197,34	176,69
Mediana	224,72	213,27
3º Quartil	257,41	248,12
Máximo	328,3	334,23
Desvio padrão	46,43	48,43

Fonte: INEP, SAEB, 2017.

Analisando o recurso gráfico e o Quadro, podemos observar que para dependência Estadual+Federal, tem-se menor variabilidade dos dados, resultados em notas mais concentradas em um certo intervalo, que demonstra incluir notas maiores do que as dos estudantes da dependência Municipal. Vale destacar também que ainda na dependência

Estadual+Federal houve registros de *outliers*, ou seja, estudantes que se destacam significativamente dos outros e que, neste caso, se destacaram negativamente já que tiraram notas discrepavelmente mais baixas.

Assim, para testar se há associação entre as notas em Língua Portuguesa e a dependência administrativa dos estudantes, tem-se as seguintes hipóteses:

$$\begin{cases} H_0 : \text{As variáveis Notas em Língua Portuguesa e Dependência Administrativa estão associadas.} \\ H_a : \text{Não há associação entre Notas em Língua Portuguesa e Dependência Administrativa.} \end{cases}$$

Para analisar as hipóteses serão feitos a seguir os testes T de Student, Wilcoxon-Mann-Whitney e Kolmogorov-Smirnov.

Tabela 2 – Testes para Notas em Língua Portuguesa por Dependência Administrativa

Teste	P-valor	Decisão do teste
T de Student	$4,977 * 10^{-14}$	Rejeita-se H_0
Wilcoxon-Mann-Whitney	$1,411 * 10^{-13}$	Rejeita-se H_0
Kolmogorov-Smirnov	$< 2,2 * 10^{-16}$	Rejeita-se H_0

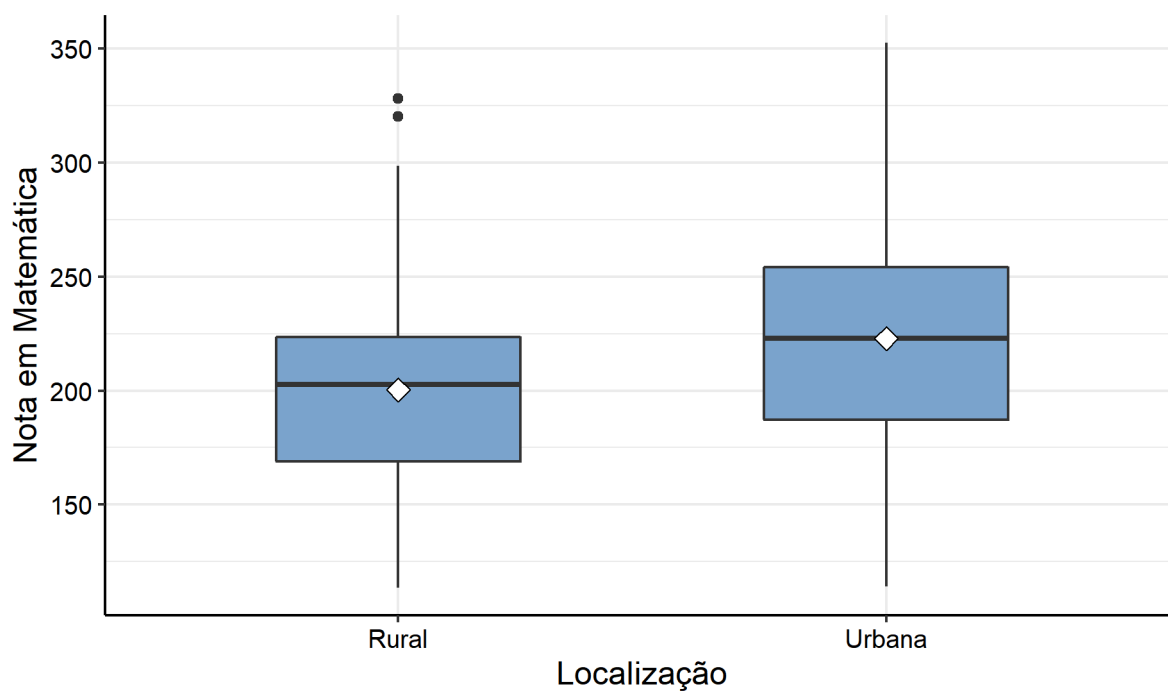
Fonte: INEP, SAEB, 2017.

Para todos os testes obteve-se o mesmo p-valor que, por sua vez, é menor que o nível de significância (5%). Assim, há evidências para dizer que as notas em Língua Portuguesa e a dependência administrativa do estudantes não estão associadas.

2.3 Amostras de tamanho 30

2.3.1 Associação entre Notas em Matemática e Localização

Figura 3 – Boxplot de Notas em Matemática por Localização



Fonte: INEP, SAEB, 2017.

Quadro 2.3 – Quadro Resumo de Notas em Matemática por Localização

Estatística	Urbana	Rural
Média	222,8	200,3
Mínimo	114,2	113,5
1º Quartil	187,2	168,9
Mediana	222,9	202,7
3º Quartil	254,2	223,5
Máximo	352,6	328
Desvio padrão	46,4	43,3

Fonte: INEP, SAEB, 2017.

Assim, como na amostra de tamanho 100, observa-se pelo Gráfico 3 e pelo Quadro 2.3 que as notas em Matemática de estudantes com localização urbana são consideravelmente maiores. Pode-se observar também que há *outliers* de notas significativamente mais altas

que as outras de estudantes de localização rural, mas estes ainda obtiveram notas menores que os de localização urbana.

As informações sugerem que existe associação entre as variáveis analisadas, assim, para testar se há associação entre as notas em Matemática e localização dos estudantes, tem-se as seguintes hipóteses:

$$\begin{cases} H_0 : \text{As variáveis Notas em Matemática e Localização estão associadas.} \\ H_a : \text{Não há associação entre Notas em Matemática e Localização} \end{cases}$$

Para analisar as hipóteses serão feitos a seguir os testes T de Student, Wilcoxon-Mann-Whitney e Kolmogorov-Smirnov.

Tabela 3 – Teste para Notas em Matemática por Localização

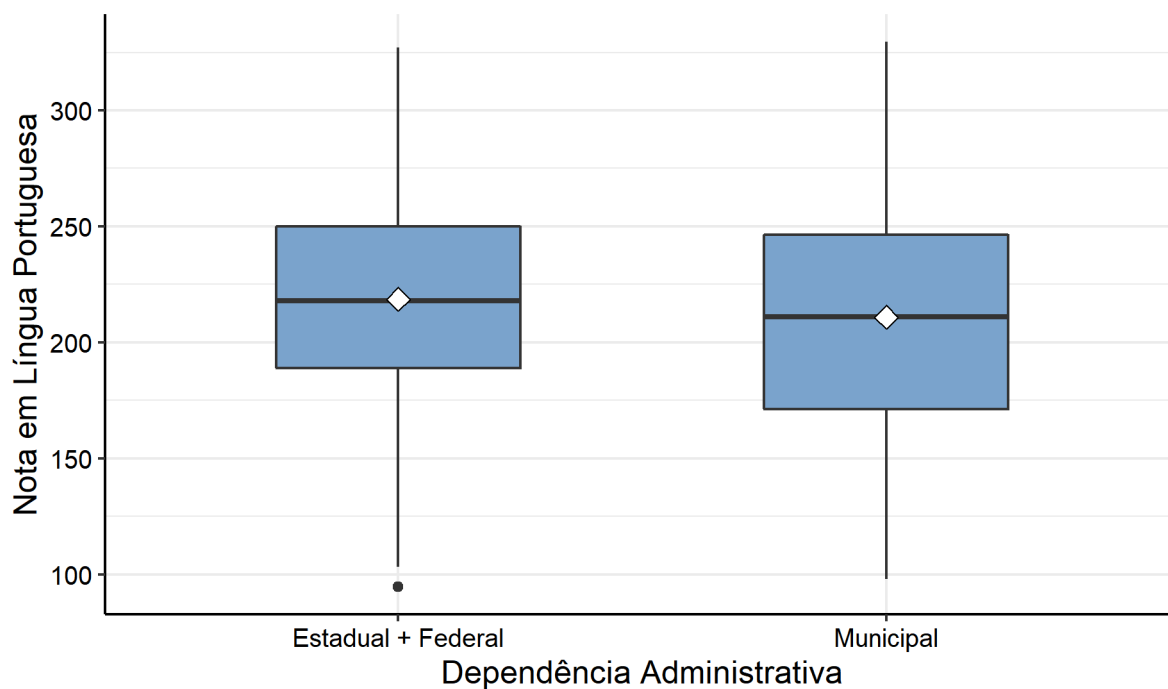
Variável	P-valor	Decisão do teste
T de Student	$3,312 \cdot 10^{-9}$	Rejeita-se H_0
Wilcoxon-Mann-Whitney	$7,143 \cdot 10^{-9}$	Rejeita-se H_0
Kolmogorov-Smirnov	$1,555 \cdot 10^{-8}$	Rejeita-se H_0

Fonte: INEP, SAEB, 2017.

Ainda que sejam consideravelmente maiores do que os observados para a amostra de tamanho 100, para todos os testes obteve-se o p-valores menores que o nível de significância (5%). Assim, há evidências para dizer que as notas em Matemática e a localização do estudantes não estão associadas.

2.3.2 Associação entre Notas em Língua Portuguesa e Dependência Administrativa

Figura 4 – Boxplot de Notas em Língua Portuguesa por Dependência Administrativa



Fonte: INEP, SAEB, 2017.

Quadro 2.4 – Quadro Resumo de Notas em Língua Portuguesa por Dependência Administrativa

Estatística	Estadual+Federal	Municipal
Média	218,59	210,81
Mínimo	94,61	97,85
1º Quartil	188,85	171,15
Mediana	217,95	211,09
3º Quartil	249,99	246,43
Máximo	327,11	329,61
Desvio padrão	46,14	48,85

Fonte: INEP, SAEB, 2017.

Observando o Gráfico 4, tem-se que ambas dependências administrativas apresentam estatísticas similares, diferenciando-se significativamente apenas no 1º Quartil, que leva a dizer que as notas dos estudantes de dependência Estadual+Federal estão mais concentradas. A análise descritiva não sugere associação entre as variáveis, mas para

testar se há ou não associação entre as notas em Língua Portuguesa e a dependência administrativa dos estudantes, tem-se as seguintes hipóteses:

$$\begin{cases} H_0 : \text{As variáveis Notas em Língua Portuguesa e Dependência Administrativa estão associadas.} \\ H_a : \text{Não há associação entre Notas em Língua Portuguesa e Dependência Administrativa.} \end{cases}$$

Para analisar as hipóteses serão feitos a seguir os testes T de Student, Wilcoxon-Mann-Whitney e Kolmogorov-Smirnov.

Tabela 4 – Teste para Notas em Língua Portuguesa por Dependência Administrativa

Variável	P-valor	Decisão do teste
T de Student	0,0106	Rejeita-se H_0
Wilcoxon-Mann-Whitney	0,01174	Rejeita-se H_0
Kolmogorov-Smirnov	0,002801	Rejeita-se H_0

Fonte: INEP, SAEB, 2017.

Os p-valores mostrados no Quadro 4 foram os maiores de todo o relatório, no entanto, ainda indicam que há evidências para dizer que não há associação entre as notas em Língua Portuguesa e a dependência administrativa.

3 Conclusão

A partir de todos os resultados conclui-se que apesar das análises descritivas mostrarem tendências à associação, os testes podem mostrar o contrário. Assim, depreende-se dos resultados que não há associação entre a nota em Matemática e a localização do estudante e que não há associação entre a nota em Língua Portuguesa e a dependência administrativa do estudante.

Referências

INEP. *SAEB*: 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 5.

ANEXO A – Código em R

Link para o diretório no GitHub:

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