

Trabalho Final de Métodos Estatísticos 1

Consultores Responsáveis:

Felipe Bretas

Renan Marques

Tales Vaz

Victor Tavares

Requerente:

Maria Teresa

Sumário

	Página
1 Introdução	3
2 Objetivos	4
3 Metodologia	5
4 Resultados	6
4.1 Questao6:	6
4.2 Questao7:	6
4.3 Questao10:	7
5 Conclusões	8

1 Introdução

O presente estudo tem como objetivo analisar o desempenho dos estudantes do 5º ano e 9º ano do Ensino Fundamental e 3ª série do Ensino Médio na Prova Brasil de 2011. Para isso, foi utilizada uma amostra aleatória simples de 200 escolas, extraída do banco de dados dos resultados da avaliação. Além disso, foi selecionada uma subamostra de 50 escolas para comparação dos resultados obtidos. A pesquisa busca descrever as características das escolas participantes e o desempenho médio dos alunos. Entre os principais focos da análise estão a estimativa da nota média em Língua Portuguesa e Matemática, a verificação da normalidade das distribuições das notas e a comparação da proficiência em Matemática segundo o local da escola (urbana ou rural). Além disso, o estudo examina a evolução do desempenho entre 2009 e 2011 e investiga possíveis associações entre variáveis como região, categoria administrativa, tamanho da escola e tamanho do município. Para os testes de hipóteses foi utilizado um nível de significância de 5%. Os dados foram disponibilizados pelo Inep e se referem às escolas que participaram da Prova Brasil 2011, uma avaliação aplicada pelo Ministério da Educação para diagnosticar o desempenho dos estudantes em redes municipais, estaduais e federais. O software utilizado para análise estatística dos dados foi o R versão 4.4.2.. O R é um software de programação gratuito largamente usado na área de estatística e visualização de dados que permite não só o manuseio e análise de bancos de dados, como também a confecção de gráficos.

2 Objetivos

3 Metodologia

4 Resultados

4.1 Questao6:

A análise teve como objetivo investigar se existem diferenças significativas no desempenho em Matemática entre alunos de escolas urbanas e rurais. Para isso, foi utilizado o teste para comparação de duas médias, adequado para amostras independentes, uma vez que as variâncias entre os grupos podem ser diferentes. A hipótese nula (H_0) propõe que não há diferença no desempenho entre alunos de escolas urbanas e rurais, enquanto a hipótese alternativa (H_1) sugere que existe uma diferença significativa. A amostra total foi composta por 200 alunos, com $n = 200$. O teste gerou um valor t de 5.0852, com 56.626 graus de liberdade e um p -valor de $4.318e-06$, que é significativamente menor que o nível de significância de 5%. Isso indica que existe uma diferença significativa no desempenho entre os dois grupos. Os alunos de escolas urbanas apresentaram uma média de notas de 211.11, enquanto os alunos de escolas rurais obtiveram uma média de 187.70, com uma diferença de 23.41 pontos, o que reforça a superioridade dos alunos urbanos. O intervalo de confiança de 95% (14.19 a 32.63) também confirma essa diferença significativa, evidenciando que os alunos urbanos tendem a ter um desempenho melhor. O boxplot da amostra demonstra que as notas de Matemática são mais concentradas em torno da mediana para os alunos urbanos, com menor dispersão, enquanto os alunos rurais apresentam maior variação em suas pontuações.

4.2 Questao7:

O objetivo desta análise foi avaliar se existe uma diferença estatisticamente significativa entre as notas de Língua Portuguesa e Matemática dos alunos. Como se tratam das mesmas pessoas avaliadas nas duas disciplinas, foi escolhido o teste para comparação de duas médias pareadas, adequado para amostras dependentes. A hipótese nula (H_0) sugere que não há diferença nas médias das notas entre as duas disciplinas, enquanto a hipótese alternativa (H_1) indica que existe uma diferença significativa. O teste para a amostra de 200 alunos gerou um valor t de -28.28, com 199 graus de liberdade e um p -valor inferior a $2.2 \cdot 10^{-16}$, o que é altamente significativo. A diferença média entre as notas foi de -19.70, com intervalo de confiança de 95% variando entre -21.07 e -18.32. Isso confirma que as notas em Língua Portuguesa são significativamente menores que as de Matemática. O boxplot reforça essa diferença, mostrando uma clara disparidade entre as distribuições das duas disciplinas. Para a amostra de 50 alunos, os resultados foram semelhantes, com $t = -15.48$, $df = 49$ e p -valor inferior a $2.2 \cdot 10^{-16}$, evidenciando uma diferença média de -21.26, com intervalo de confiança

entre -24.02 e -18.50. Ambos os testes indicam que a diferença entre as notas de Matemática e Língua Portuguesa é estatisticamente significativa.

4.3 Questao10:

objetivo desta análise é investigar a relação entre o desempenho dos alunos nas disciplinas de Língua Portuguesa e Matemática. Para isso, foi utilizada a correlação de Pearson, que mede a força e a direção da associação linear entre as variáveis. A hipótese nula (H_0) propõe que não existe correlação entre as notas de Língua Portuguesa e Matemática, enquanto a hipótese alternativa (H_1) sugere que há uma correlação significativa entre elas. A correlação de Pearson para a amostra de 200 alunos apresentou um coeficiente de 0.9398, com um valor t de 38.69 e 198 graus de liberdade, com p -valor inferior a $2.2 \cdot 10^{-16}$. Isso indica uma forte associação positiva entre as notas das duas disciplinas. O gráfico de dispersão mostra um alinhamento claro entre as notas de Língua Portuguesa e Matemática, reforçando a forte correlação positiva. Para a amostra de 50 alunos, a correlação também foi alta, com coeficiente de 0.9067, $t = 20.26$ e p -valor inferior a $2.2 \cdot 10^{-16}$, confirmando que a relação entre as notas das duas disciplinas é estatisticamente significativa, independentemente do tamanho da amostra.

5 Conclusões

Portanto, a partir das análises descritivas realizadas foi possível ter uma visão geral dos estudantes e das escolas integrantes da amostra. Na primeira análises, pode-se observar as medidas resumos da proeficiência dos estudantes e gráficos que contribuíram para entender características pertinentes das escolas, não alterando de maneira significativa entre as amostras de tamanho 200 e de 50, sendo importante para uma boa visualização do panorama geral do projeto.

Já em relação a estimação da proporção de escolas com menos de 75% de participação, foi obtido para um intervalo de confiança de 95%, uma proporção de (0,04;0,11), ou 7,5% com erro 3,5% de erro, para as amostras de tamanho 200. Já para amostras de tamanho 50, teve a proporção de (0,00;0,16), ou 8% com margem de erro de 8%.

Ao se estimar a proeficiência média dos alunos nas matérias de português, foi obtido para um intervalo de confiança de 95%, uma média de (183,76;190,30) para amostras de tamanho 200, e para amostras de tamanho 50, o intervalo foi de (179,48;194,46). Foi estimado também a proeficiência média em matemática, onde foi obtido um intervalo de (202,90;210,65) para amostras de tamanho 200, e para amostras de tamanho 50, o intervalo foi de (199,82;216,64). Pode-se observar que as notas médias de português estimadas foram menor que as notas de matemática em ambas amostras e que, amostras de tamanho menor possuem um intervalo maior para um mesmo intervalo de confiança.