**RELATÓRIO TÉCNICO**

**DISSERTAÇÃO DE DOUTORADO EM LINGUÍSTICA**

**DOUTORANDA MARIA LÚCIA**

**Análise estatística dos erros ortográficos encontrados em produções escritas dos alunos do Ensino Fundamental II de três escolas públicas de Belém-PA.**

**1. Introdução**

O presente trabalho tem por objetivo apresentar a análise estatística para os dados de natureza ortográfica da doutoranda Maria Lúcia.

Neste trabalho foram estudadas as características de escrita de alunos do 5° ano de três diferentes escolas públicas das dependências administrativas federal, estadual e municipal na cidade de Belém do Pará. Para avaliar possíveis diferenças entre palavras grafadas de acordo com a dependência administrativa da escola em que estuda, será realizada uma análise estatística com base em análise exploratória de dados, testes paramétricos e não-paramétricos e análise de variância.

**2. Materiais e Métodos**

Os dados foram coletados pela doutoranda, tratados e apresentados em uma planilha do programa Excel, contendo informações de 132 estudantes submetidos ao estudo oriundos de três escolas de dependências administrativas federal, estadual e municipal. Em cada escola foram pesquisados alunos de duas turmas do 5º ano do ensino fundamental.

Os dados coletados são de natureza qualitativa e quantitativa e compreendem as informações do aluno: código, escola, turma, turno, sexo e idade. Além das 25 palavras informadas para serem grafadas: porcos, borboletas, formigas, carneiro, garças, golfinho, falcão, alce, polvo, pulga, avestruz, mosca, ostra, cisne, esquilo, onça, ganso, caranguejo, elefante, chimpanzé, águia, boi, égua, papagaio e gaivotas.

A análise estatística será inicialmente realizada por meio de análise exploratória de dados, objetivando fornecer uma visão geral do estudo e delinear possíveis características estruturais que possam ser avaliadas em maiores detalhes com o auxílio de testes de hipóteses.

Inicialmente procedeu-se a elaboração do banco de dados para a análise estatística com a junção das informações fornecidas trocando as palavras escritas corretamente por zero (0) e as escritas incorretamente por um (1). Em seguida foi criada a variável “Erros” que corresponde à soma dos valores 0 e 1 para cada aluno, o que resultará no total de palavras escritas incorretamente por cada aluno dentre as 25 apresentadas para ele escrever.

A análise exploratória de dados será conduzida por meio de tabelas e gráficos. Além do gráfico de setores e de colunas, que serão utilizados para traçar perfil da amostra, serão apresentados também gráficos de caixa (box-plot), os quais são apropriados para visualizar a distribuição de uma variável quantitativa relacionando-a a fatores ou subgrupos. Nos gráficos em formas de caixas são representados a média, o primeiro e o terceiro quartis dos dados. Destacam-se também os valores mínimo e máximo e valores discrepantes dos demais.

Como o objetivo maior do trabalho é comparar os erros gramaticais, que é uma variável quantitativa, por características como gênero e dependência administrativa, que são variáveis categóricas, esse gráfico nos proporcionará uma indicação visual da possibilidade de diferença entre os grupos.

Para comparar estatisticamente os erros gramaticais quanto à dependência administrativa das escolas, turmas e gêneros, serão realizados testes de hipóteses. Os testes de hipóteses são técnicas científicas para a comprovação de hipóteses teóricas a partir de um conjunto de dados que sejam representativos da população. Os testes de hipóteses podem ser paramétricos ou não paramétricos.

Os testes paramétricos são mais fidedignos que os não paramétricos, mas foram desenvolvidos para certo tipo de variável (variável quantitativa mensurada a um nível intervalar) e exige o cumprimento de pressupostos pelas subamostras que estarão sendo comparadas, dentre os quais, a normalidade dos conjuntos de dados e a homogeneidade das variâncias das subamostras. A violação dos pressupostos invalida a aplicação do teste paramétrico, mas para esses casos existe a contrapartida não paramétrica.

Os testes não paramétricos foram desenvolvidos para dados ordinais e nominais e são muito uteis para grandes amostras em que as suposições de uso do teste paramétrico não se verifiquem ou para pequenas amostras, para as quais não é possível a comprovação precisa dos pressupostos. Esses testes são menos exigentes quanto às condições para sua utilização porém menos potentes para detectar diferenças do que os testes paramétricos, mas pode ser a única opção no caso de amostras muito pequenas.

Para a comparação de duas subamostras o teste paramétrico utilizado é o teste t de Student, desenvolvido para testar a hipótese de que as médias entre as duas amostras não diferem. Sua contrapartida não paramétrica utilizada neste trabalho foi o teste de U de Mann-Whitney, o qual transforma as mensurações originais da variável sob teste em postos e compara as amostras por meio de uma estatística que leva em consideração a soma dos postos em cada amostra.

Para o caso de mais de três subamostras, a comparação paramétrica é realizada por meio do teste F na Análise de Variância (Anova) e o teste não paramétrico é feito por meio da estatística de Kruskal-Wallis que é um teste robusto no caso de não adequação dos dados aos pressupostos da Anova.

**3. Resultados**

**3.1 Perfil da amostra**

Participaram do estudo 132 estudantes do 5º ano com idade entre 9 e 12 anos e média de 11 anos dos quais 45% são do sexo feminino (Figura 1).

**Figura 1**. Características dos indivíduos sob estudo.

Fonte: A aut**o**ra

Na distribuição por escola observou-se predominância de alunos das escolas federal e municipal em relação à escola estadual na amostra (Figura 2).

**Figura 2:** Distribuição das Escolas sob estudo.

Fonte: A autora

Quanto ao número de alunos das diferentes dependências administrativas, verificou-se que não há grande diferença na quantidade entres as turmas, no máximo 3 alunos de diferença na dependência administrativa federal.

**Figura 3:** Quantitativo de alunos por turma e dependência administrativa.

Fonte: A autora

Na distribuição por gênero nas escolas verificou-se predominância do gênero masculino nos três tipos de dependência administrativa conforme a Figura 4.

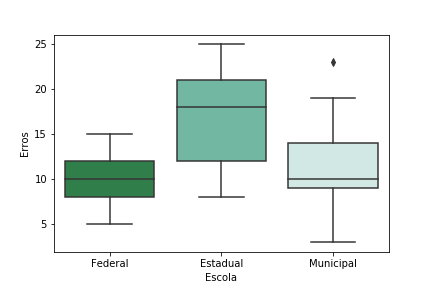
**Figura 4:** Quantitativo de alunos por gênero e dependência administrativa.

Fonte: A autora

**3.2 Características Ortográficas**

Quanto ao erro gramatical das palavras entre as escolas estudadas verifica-se na Figura 5 que a escola estadual destaca-se negativamente com a maior média de erros da amostra, enquanto que não se visualiza grande diferença entra as outras duas escolas.

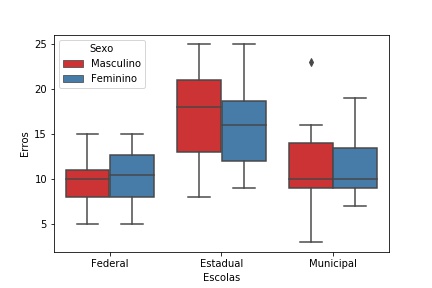
**Figura 5:** Erros gramaticais por dependência administrativa.

****

Fonte: A autora

Na comparação por gênero, considerando a amostra toda, a média de erros dos estudantes do sexo feminino foi de 12,12 e do sexo masculino 12,33. Levando em conta a dependência administrativa, a Figura 6 não evidencia muita diferença nas médias de erros dos estudantes de escolas de dependência administrativa federal e municipal, somente na escola estadual observa-se a média de erros dos alunos do sexo masculino superior às do sexo feminino.

**Figura 6:** Erros gramaticais por gênero e dependência administrativa.



Fonte: A autora

**3.3 Testes estatísticos**

O estudo para avaliar se os estudantes diferem no que se refere ao total de erros cometidos em razão das categorias de gênero desconsiderando a dependência escolar, constatou que a diferença entre as médias de erros não é significativa (Teste U de Mann-Whitney, p = 0,958), ou seja, não há diferença significativa entre os erros dos sexos masculino e feminino.

Em relação à dependência administrativa, a Tabela 1 descreve os resultados exploratórios, onde se evidencia que a média de erros na escola estadual é superior à média das demais dependências administrativas, possuindo também maior variabilidade.

**Tabela 1:** Estatísticas de erros de grafia por dependência administrativa da escola.

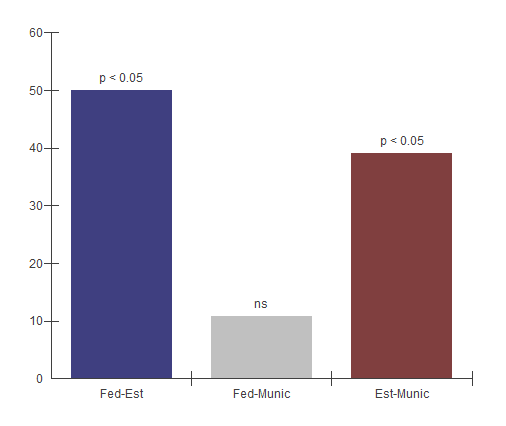
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dependência | Número de estudantes | Média de erros | Desvio padrão | Mínimo | Máximo |
| Federal | 49 | 9,92 | 2,691 | 5 | 15 |
| Estadual | 35 | 17,00 | 5,275 | 8 | 25 |
| Municipal | 48 | 11,13 | 3,955 | 3 | 23 |
| Total | 132 | 12,23 | 4,900 | 3 | 25 |

Fonte: A autora

A comparação dos erros cometidos em relação às categorias de dependência administrativa das escolas, baseada na Análise de Variância não-paramétrica de Kruskal-Wallis, rejeitou a hipótese de que as médias dos erros são iguais entre as escolas mostrando que a diferença entre as médias de erros cometidos não pode ser atribuída ao acaso (H =37,237; p < 0,001). Desse modo, aceitamos a hipótese de que pelo menos uma das médias difere das demais e vai ser necessário realizar um teste a posteriori para aprofundar a análise.

A Figura 7 mostra em maiores detalhes os resultados do teste de comparação de Dunn, que é realizado após a Análise de variância para verificar quais médias diferem entre si por meio de testes pareados. Observa-se que os alunos de escolas Federal e Municipal não diferem significativamente entre si no que diz respeito aos erros de escrita cometidos, mas essas duas categorias diferem da Estadual, sendo a maior diferença estatisticamente significativa a observada entre a Federal e Estadual a maior diferença.

**Figura 7:** Testes de comparação entre médias



Fonte: A autora