



Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística

Luciano Moura Silva

**Desigualdade escolar na Região Integrada de
Desenvolvimento do Distrito Federal e Entorno**

Brasília - DF, 2020

Luciano Moura Silva

Desigualdade escolar na Região Integrada de Desenvolvimento do Distrito Federal e Entorno

Monografia apresentada ao Departamento de Estatística da Universidade de Brasília como requisito para a obtenção do grau de Bacharel em Estatística.

Orientador: Prof. **Dr. Luís Gustavo do Amaral Vinha**

Brasília - DF, 2020

Agradecimentos

Primeiramente quero agradecer à Deus, por ter me dado força pra chegar até aqui. A todas as pessoas que contribuíram de qualquer forma na minha formação, tanto pessoal quanto acadêmica. E ao meu amigo Vitor Macedo (RIPyGOD).

“Errar é humano, perdoar é divino; mas incluir erros em seu projeto é estatístico.”

-Leslie Kish

Conteúdo

| | | |
|----------|--------------------------------------------------------------------|-----------|
| 1 | Introdução | 9 |
| 2 | Referencial Teórico | 12 |
| 2.1 | Formulação do Modelo Multinível | 12 |
| 2.2 | Agregação e Desagregação | 13 |
| 2.3 | Construção do Modelo Multinível | 14 |
| 2.4 | Estimando parâmetros | 15 |
| 2.4.1 | Comparando os Modelos | 16 |
| 2.5 | Análise dos Resíduos | 16 |
| 3 | Metodologia | 17 |
| 4 | Resultados | 18 |
| 4.1 | Análise Descritiva | 18 |
| 4.1.1 | Desempenho em matemática | 18 |
| 4.1.2 | Desempenho em matemática com relação as demais variáveis | 19 |
| 4.1.3 | Análise Fatorial | 21 |
| 4.1.4 | Análise Multinível | 22 |
| 5 | Conclusão | 28 |
| | Referências | 29 |

Resumo

O presente estudo temo como objetivo apontar diferenças educacionais na Região Integrada de desenvolvimento do Distrito Federal e Entorno - (Ride-DF), com base no desempenho dos alunos nas provas de matemática, usando modelo multinível. Foram avaliados 57.000 alunos das redes pública e privada em 739 escolas, no qual após retirar os dados preenchidos de forma incorreta ou não preenchidos tivemos um total de 41.721 alunos e 711 escolas, ao qual fizeram parte do presente estudo. Na análise multinível a hierarquia adotada foi do nível 1 - aluno e nível 2 - Município / Região Administrativa - (Município/RA). No modelo final foram consideradas as variáveis Socioeconômico, que foi calculada com base na análise fatorial de alguns fatores de algumas respostas dos alunos, sexo, reprovação, Índice de Desenvolvimento Humano - Município / Região Administrativa - (IDH-Município/RA). Ao fim da análise multinível foi possível observar que algumas variáveis incidem fortemente na nota média do desempenho em matemática

Palavras-chave: Multinível, Distrito Federal, Desigualdade Escolar, Ride-DF, 5° ano SAEB-2017, Desempenho em Matemática.

1 Introdução

O termo desigualdade nos traz um conceito de diferença, não iguais. Quando falamos em desigualdade educacional queremos dizer que existem disparidades entre o desempenho dos estudantes em diferentes circunstâncias, isso pode ser por vários fatores, como por exemplo, investimento na educação, qualificação dos profissionais, método adotado de ensino, infraestrutura, dentre muitos outros fatores que contribuem direta ou indiretamente para essa desigualdade.

Quando citamos o Distrito Federal temos uma referência nacional em investimento na educação, mesmo apresentando muitas falhas. Por outro lado quando falamos de cidades do entorno do Distrito Federal-(DF), Minas Gerais e Goiás, temos resultados menos favoráveis. Por esse motivo o presente estudo visa analisar essa disparidades entre os municípios e as regiões administrativas que compõem a Região Integrada de Desenvolvimento do Distrito Federal e Entorno - (Ride-DF).

A Ride-DF, criada em 1998, que ao longo dos anos passou por diversas alterações, sendo que a última foi em 2018 ao qual incluiu mais 12 municípios, que mesmo não estando contemplando a área da Ride-DF na época da avaliação do Sistema de Avaliação da Educação Básica (Saeb), farão parte do estudo. Com isso temos que a composição da Ride-DF ficou da seguinte forma, DF, quatro municípios de Minas Gerais e 29 do estado de Goiás, com isso ela se torna mais ampla e acaba beneficiando mais municípios que vivem com parte da sua economia vinda do DF.



Figura 1: Região Integrada de desenvolvimento do Distrito Federal e Entorno .
(<http://www.sudeco.gov.br/municipios-ride-df>)

Nesse estudos serão utilizados dados do SAEB, disponível no site do INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira). Segundo (VINHA; KARINO; LAROS, 2016), "o principal objetivo do SAEB é avaliar a educação básica e contribuir para a

melhoria de sua qualidade e para a universalização do acesso às escolas, oferecendo subsídios concretos para a formulação, reformulação e controle de políticas públicas”. As avaliações do SAEB começaram em 1990 e passaram por inúmeras mudanças, ate chegarmos ao modelos atual, ao qual avalia os alunos do 5º e 9º anos do ensino fundamental e 3º ano do ensino médio, a cada dois anos, e por fim buscam avaliar todos os alunos das séries citadas.

Segundo (ARROYO, 2010), ao qual afirma que a desigualdade educacional e uma vergonha nacional, uma mancha e expressão de nosso atraso, além disso aponta como causa de nosso subdesenvolvimento nacional, regional, social, cultural, político e econômico. Infelizmente essa desigualdade faz parte da cultura do estado, que por sua vez e natural uma pessoa rica ter uma educação melhor que uma pessoa pobre, é aí que perguntamos, o que seria necessário para diminuir ou até mesmo acabar com essa desigualdade? Complicado responder à essa pergunta, pois a cultura de um povo está muito enraizada no seu dia-a-dia, ela vem sendo construída ao longo de muitos anos. Para que possamos ter de fato essa mudança, ela precisa ser implementada e aceita por todos, como sendo um processo a longo prazo e com dificuldades a serem superadas.

A escola tem um papel fundamental na vida de uma criança, ela poderia diminuir essa desigualdade, que existe entre os indivíduos. A desigualdade social no estado e refletida no desempenho acadêmico, o caso de ter uma casa boa, conviver com pessoas em boas condições financeiras e intelectuais, pode sim influenciar uma criança a querer ser mais, a querer ter mais que básico para sobrevivência, por outro lado crianças que convivem em um meio menos favorável tendem a não evoluir tanto quanto aquelas que viveram num meio mais favorável, pelos mesmos motivos que essas, pois conviveram com pessoas que não tiveram uma boa educação, não possuem uma condição financeira boa, dentre outros.

Temos que desigualdade educacional não está apenas na falta de um diploma de ensino fundamental ou na falta de um diploma de ensino médio, vai muito além de um indivíduo ter ou não um diploma de nível superior. Existe um grande hiato no ensino que cada pessoa recebe e o que realmente ela aprendeu, pois temos uma rede de ensino muito heterogênea, por inúmeros motivos, como, um espaço físico adequado, profissionais qualificados, falta de material adequado para ministra as aulas, dentre muitos outros. Podemos dizer que desigualdade educacional vai além de um simples conceito de diferença entre o ensino, abrange muitos fatores externos, no entanto o papel da escola é diminuir essas desigualdades, trazendo condições mínimas para seus alunos dentro do espaço escolar, onde a criança passa boa parte do seu dia, podendo assim dizer que ela é uma extensão de sua casa.

Esse estudo visa contribuir de forma literária, apontando quais pontos precisam de mais atenção para uma melhoria do ensino na Ride-DF, podendo até contribuir para políticas públicas que visam a melhoria do ensino nessa região de desenvolvimento.

O estudo tem como base (SANTOS, 2017), ao qual utiliza o mesmo método e a análise análoga de estudo, com poucas diferenças, como nível 2 (SANTOS, 2017) usa a escola e aqui

usaremos o município.

As regiões administrativas ou RA's do DF serão consideradas como municípios, terão o mesmo peso que as cidades de Minas Gerais e Goiás, e assim veremos quais fatores pesam mais no desempenho em matemática dos alunos dessas regiões.

2 Referencial Teórico

Na presente seção será aludido conceitos e definições com o objetivo de facilitar a compreensão deste trabalho. Primeiramente, será discutido a definição de modelo multinível, posteriormente as etapas que compõem a elaboração do modelo, estimadores de parâmetros, comparação de modelos e análise dos resíduos.

2.1 Formulação do Modelo Multinível

A principal finalidade dessa seção é trazer uma base teórica para uma melhor compreensão dos resultados. Segundo (HOX, 2010) o conceito geral é que os indivíduos interagem com o ambiente social contextos aos quais pertencem, que eles são influenciados pelo social, grupos ou contextos aos quais pertencem, e que esses grupos são, por sua vez, influenciados pelos indivíduos que compõem esse grupo.

A utilização de modelos multinível parte da premissa de que tenhamos dados dispostos como uma escada. Onde temos eles da seguinte forma, em avaliação educacional os alunos estão agrupados em escolas, essas escolas por sua vez estão agrupadas em cidades, as cidades em estados e assim por diante. Por outro lado temos o nível do aluno ao qual definimos o nível socioeconômico a partir de uma análise fatorial sugerida por (HAIR et al., 2009), e assim o modelo vai sendo construído de acordo com o que o pesquisador pretende analisar.

O nível mais baixo é indicado como sendo o indivíduo, ao qual é analisado pela variável socioeconômica, que é composta de várias respostas sobre a vida financeira da pessoa.

Uma situação comum é propor modelos com apenas dois níveis, onde por exemplo, temos que os alunos são agrupados e representados pelo índice i e as escolas pelo índice j e a variável resposta dada pelo desempenho escolar do aluno, com essa ilustração temos que o modelo multinível básico é apresentado da seguinte forma

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + e_{ij} \quad (1)$$

- Y_{ij} , desempenho escolar do estudante i na escola j ;
- β_{0j} , valor médio do desempenho dos alunos na escola j ;
- β_{1j} , coeficiente de regressão da escola j com relação a variável X_1 ;
- e_{ij} , erro aleatório no nível do estudante.

Na equação 1 o modelo pode ser interpretado da seguinte forma, existe uma variação de escola para escola no coeficiente de inclinação. Quando introduzimos uma variável explicativa Z associada a escola, teremos que os coeficientes de regressão pode ser apresentados como a seguir.

Os β_{0j} e β_{1j} , podem ser observados da seguinte forma,

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + \mu_{0j} \quad (2)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + \mu_{1j} \quad (3)$$

Temos que Z e a variável referente a escola, onde segundo (HOX, 2010) age como variável moderadora na relação entre a variável resposta e a variável do nível mais baixo, γ 's são fixos e μ 's são os erros.

Quando substituimos (2) e (3) em (1), temos a seguinte expressão:

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{1ij} + \gamma_{01}Z_j + \gamma_{11}X_{1ij}Z_j + \mu_{1j}X_{1ij} + \mu_{0j} + e_{ij} \quad (4)$$

Observe que um modelo com apenas dois níveis tivemos uma equação complexa, se acrescentarmos mais níveis essa equação fica cada vez mais complexa e mais difícil de se construir e interpretar.

As suposições do modelo em questão são:

- $e_{ij} \sim N_{0,1}$ - Erro do nível 1;
- u_{0j} e $u_{1j} \sim \text{Normal}$, com médias 0 e variâncias $\sigma_{\mu_0}^2$ e $\sigma_{\mu_1}^2$ - Erro nível 2;
- Covariância não nula.

No exemplo dado acima consideramos apenas duas variáveis explicativas, sendo que podemos inserir quantas forem necessárias para uma melhor explicação sobre os impactos estudados. Com isso a cada nova variável inserida no modelo teremos uma maior complexidade no mesmo, com isso a sua explicação se tornará mais extensa.

2.2 Agregação e Desagregação

A agregação acontece quando as variáveis de um nível inferior são movidas para um nível superior, por exemplo atribuir a renda dos alunos às escolas, ou seja estamos pegando a renda do aluno que está em um nível inferior e colocando como renda das escolas que está em um nível superior.

A Desagregação significa exatamente o contrário de agregação, aonde se pega uma atribuição dada a escola e passa aos alunos, por exemplo, atribuir a nota da escola em desempenho a todos os alunos de forma individual.

2.3 Construção do Modelo Multinível

Segundo (HOX, 2010), a construção do modelo segue cinco passos fundamentais, com isso iremos mostrar esses passos nessa seção.

- 1° passo:

Nesse primeiro momento vamos analisar apenas um modelo simples sem variáveis explicativas que pode ser dados pela seguinte equação.

$$Y_{ij} = \gamma_{00} + \mu_{0j} + e_{ij} \quad (5)$$

Temos da equação (5):

- γ_{00} , e o intercepto da regressão;
- μ_{0j} e e_{ij} , são os resíduos.

É útil passar por esse ponto pois ele nos fornece uma estimativa da correlação intraclasse ρ .

$$\rho = \frac{\sigma_{\mu 0}^2}{\sigma_{\mu 0}^2 + \sigma_e^2} \quad (6)$$

Temos da equação (6):

- $\sigma_{\mu 0}^2$, é a variância dos resíduos no nível do grupo μ_{0j} ;
- σ_e^2 , é a variância dos resíduos de nível individual e_{ij} .
- 2° passo:

Nesse passo iremos analisar todas as variáveis fixas do nível inferior, ou seja os componentes de variância são fixados em zero, onde o modelo pode ser descrito da seguinte forma

$$Y_{ij} = \gamma_{00} + \gamma_{\rho 0} X_{pij} + \mu_{0j} + e_{ij}. \quad (7)$$

Temos da equação (7):

- X_{pij} , são as variáveis p no nível individual.

Nesse momento analisaremos a contribuição de cada indivíduo, assim podemos analisar as mudanças que ocorrem no nível inferior, podemos também comparar esse resultado com o anterior e observar as mudanças.

- 3° passo:

Agora adicionaremos as variáveis explicativas do nível superior, e com isso iremos aumentar nosso modelo e conseqüentemente teremos um modelos mais complexo, assim teremos o seguinte modelo

$$Y_{ij} = \gamma_{00} + \gamma_{p0}X_{pij} + \gamma_{0q}Z_{qj} + \mu_{0j} + e_{ij}. \quad (8)$$

Temos da equação (8):

- Z_{qj} , são as variáveis explicativas q do nível do grupo.

Nesse modelo poderemos analisar se existe uma variação entre grupos. Esse modelos assim como o modelo da etapa 2 são conhecidos como componentes de variância.

- 4° passo:

Essa etapa investiga se alguma das variáveis explicativas do nível inferior, possuem um componente de variância significativo entre grupos

$$Y_{ij} = \gamma_{00} + \gamma_{p0}X_{pij} + \gamma_{0q}Z_{qj} + \mu_{pj}X_{pij} + \mu_{0j} + e_{ij}. \quad (9)$$

Temos da equação (9):

- μ_{pj} , são os resíduos de nível de grupo das inclinações de nível individual das variáveis X_{pij} .

Caso existe mais de dois níveis, essa etapa deve ser repetida nível a nível.

- 5° passo:

Agora será feito o cruzamento entre as variáveis do nível mais alto com aquelas do nível mais baixo, que nos dá o seguinte modelo

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}X_{ij}Z_j + \mu_{1j}X_{1ij} + \mu_{0j} + e_{ij}. \quad (10)$$

Em todas as etapas citadas acima foram procedimentos propostos por (HOX, 2010).

2.4 Estimando parâmetros

Segundo (HOX, 2010), o estimador mais utilizado é o de máxima verossimilhança pela vantagem de produzir estimativas que são assintoticamente eficientes e consistentes, sendo que podemos estimar também por métodos Bayesianos e mínimos quadrados.

Estimadores de máxima verossimilhança estimam os parâmetros de um modelo, fornecendo valores estimados para os parâmetros da população que maximizam a chamada 'função de verossimilhança', a função que descreve o problema da capacidade de observar os dados da amostra,

dados os valores específicos do parâmetro estimados, (HOX, 2010). Ou seja os estimadores de máxima verossimilhança são aqueles que maximizam a probabilidade de encontrarmos o que realmente está acontecendo com aquele conjunto dados, onde nem sempre são os resultados esperados, mas e o que ocorre de fato.

2.4.1 Comparando os Modelos

Para comparar os modelos usamos uma estatística chamada *deviance*, com ela podemos ver qual modelo proposto se ajusta melhor e conseqüentemente o mais adequado, pois esse apresentará uma *deviance* menor. A *deviance* é definida como

$$deviance = -2\ln(\text{Verossimilhança}) \quad (11)$$

Temos que ressaltar que a *deviance* não pode ser analisada sozinha, precisamos de mais de um modelo para podermos comparar, assim o modelo com melhor ajuste será aquele que apresentar uma *deviance* menor. Para a comparação deve-se utilizar o teste de χ^2 , com graus de liberdade igual a diferença no número de parâmetros estimados nos dois modelos.

2.5 Análise dos Resíduos

Após a construção do modelo e ajustes feitos na etapas anteriores, vamos analisar os resíduos. Nessa etapa temos que ter cuidado, pois devemos supor normalidade, linearidade e homocedasticidade destes.

Para o exemplo dado anteriormente, temos que os e_{ij} são os resíduos do nível mais baixo, para análise deles podemos fazer um simples boxplot que nos permitirá identificar outliers extremos. Segundo (HOX, 2010), uma suposição que geralmente é feita na análise de regressão multinível é que a variância dos erros residuais é a mesma em todos os grupos. Isso pode ser avaliado por computar uma análise unilateral de variância dos grupos sobre os valores absolutos do resíduos, que é o equivalente ao teste de Levene para igualdade de variâncias na análise de variância (STEVENS, 2012).

O μ_{0j} são os resíduos ao nível grupo, que também podem ser analisados da mesma forma que o e_{ij} , pelo boxplot. Por outro lado o μ_{1j} são os resíduos de inclinação entre grupos, com isso ao plotar as inclinações teremos uma visualização do quanto eles realmente diferem.

3 Metodologia

O banco de dados utilizado contém 41761 alunos que compõem os alunos que integram os municípios da Ride-DF, ao qual inicialmente eram 57.999, mas por conta de alguns alunos não possuírem respostas ou seja, por termos "NA's" tivemos que remove-los para uma melhor análise, pois caso contrário teríamos uma média puxada para baixo devido à esses "NA's". Esses alunos estão distribuídos em 717 escolas, com municípios no Estado de Goiás, Minas Gerais e as regiões administrativas do Distrito Federal.

O banco de dados possui informações relacionados aos alunos e as escolas, como suas características e atributos.

A primeira parte dos resultados consiste na análise descritiva dos dados, ao qual podemos ter uma análise exploratória e identificar qual a melhor forma de construir o modelo multinível, essa etapa é fundamental para que não se perca informações.

A segunda parte foi realizado a modelagem dos dados com a Análise de Regressão Multinível, a fim de identificar a relação da proficiência em matemática com as variáveis sexo, etnia, dentre outras. O primeiro nível, o aluno, e o segundo, o município, considerando as características do aluno e as do município. No município iremos considerar o IDH do ano de 2010, último disponível, para a faixa de cidade.

4 Resultados

Para as análises será utilizado o (R Core Team, 2020), juntamente com o pacote (BATES et al., 2015), ao qual se baseia na análise multinível.

4.1 Análise Descritiva

Primeiro iremos mostrar na Tabela 1 as variáveis e suas descrições, que compõem o estudo, ao qual foram obtidas através de um questionário aplicado aos alunos do 5° anos do ensino fundamental em 2017, e do teste de proficiência em matemática, essas serão as variáveis que vão compor o estudo.

Tabela 1: Variáveis que compõem o estudo.

| Variáveis | Descrição/Codificação |
|-------------------------------------|--------------------------------------|
| ID_UF | Código da Unidade da Federação |
| ID_MUNICÍPIO6 | Código do Município |
| ID_ALUNO | Código do aluno na ANEB/Prova Brasil |
| PROFICIENCIA_MT_SAEB | Proficiência do aluno em Matemática |
| Qual é o seu sexo? | 0:Masculino; 1: Feminino |
| Qual é a sua cor/raça/etnia? Branca | 0: Caso contrário; 1: Branca/Amarelo |
| Qual é a sua cor/raça/etnia? Parda | 0: Caso contrário; 1: Parda/Preto |
| Você já foi reprovado? | 0: Não; 1: Sim |
| IDH-Município/RA | Dados de 2010, extraídos do IBGE |

4.1.1 Desempenho em matemática

A Tabela 2 e a Figura 2 a seguir mostram o desempenho obtidos pelos alunos do 5° do ensino fundamental da Ride-DF.

Tabela 2: Medidas resumo do desempenho em matemática.

| n | Mínimo | 1°Q | Mediana | Média | 3°Q | Máximo | Desvio Padrão |
|--------|--------|-------|---------|-------|-------|--------|---------------|
| 42.888 | 109.5 | 195.3 | 224.8 | 224.9 | 253.6 | 357.0 | 41,38 |

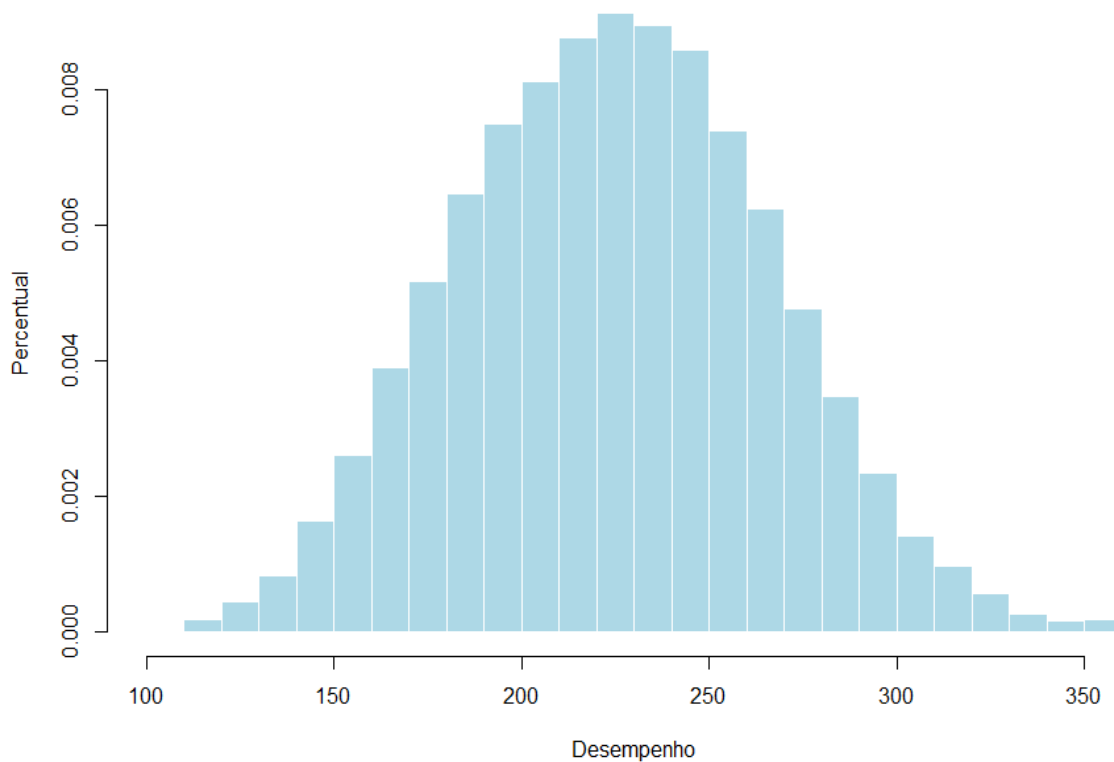


Figura 2: Distribuição do desempenho dos aluno em matemática.

Considerando a Ride-DF, temos uma nota média de 224,9 com uma variação em torno da média de 41,38 pontos. Observando o histograma podemos observar que o decréscimo à direita e levemente maior que o da esquerda. Podemos ainda observar no gráfico que as maiores notas ficam entre 200 e 250, quando passamos dos 300 pontos, podemos observar que poucos alunos atingiram essa nota, cerca de 3,57% dos alunos.

4.1.2 Desempenho em matemática com relação as demais variáveis

Para podermos fazer uma análise exploratória dos dados a fim de verificar possíveis variáveis que influenciam no desempenho em matemática, iremos realizar uma análise bivariada destes. Será utilizados gráficos e medidas resumos adequadas.

Sexo

Abaixo podemos observar a distribuição dos alunos segundo o sexo. Ao analisar a Tabela 3 podemos observar que temos mais mulheres em relação aos homens, considerando a média, notamos que os homens possuem uma média melhor comparado com mulheres.

Quando olhamos para a Figura 3 notamos que existem mais pontos discrepantes no sexo masculino, podendo assim explicar o fato de a média ser maior que a do sexo feminino, onde esses pontos puxam a média para cima.

Tabela 3: Medidas resumo do desempenho em matemática segundo sexo.

| Sexo | n | % | Média | Desvio Padrão |
|-----------|--------|--------|--------|---------------|
| Masculino | 20.776 | 49,75% | 227,02 | 42,49 |
| Feminino | 20.985 | 50,25% | 222,82 | 40,15 |

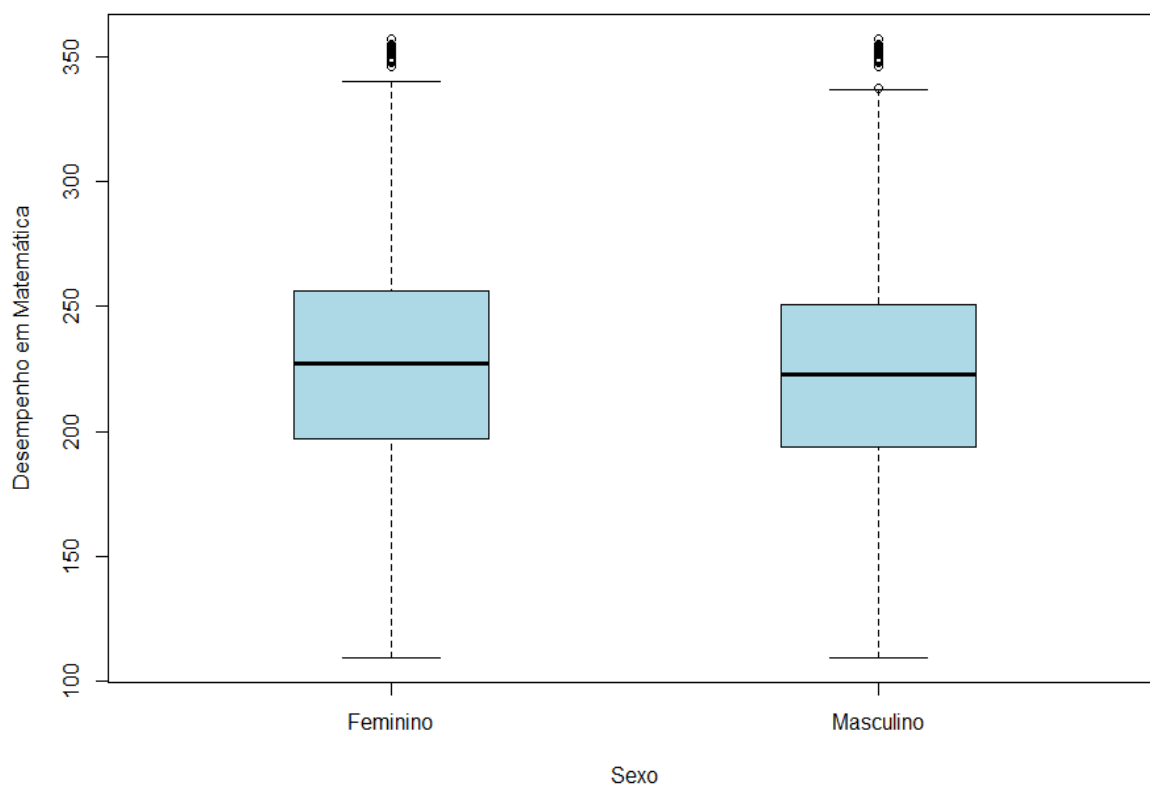


Figura 3: Desempenho em matemática por sexo.

Etnia autodeclarada

Ao analisarmos a Tabela e a Figura 4, podemos notar que a grande maioria, ou seja mais de 70% são as etnias autodeclaradas Branca e Parda, além de possuírem as maiores médias.

Tabela 4: Medidas resumo do desempenho em matemática segundo sexo.

| Cor/Raça/Etnia | n | % | Média | Desvio Padrão |
|-------------------|--------|--------|--------|---------------|
| Branca | 9.179 | 21,98% | 228,81 | 42,77 |
| Preta | 4.142 | 9,92% | 211,76 | 39,74 |
| Parda | 20.546 | 49,20% | 227,95 | 40,66 |
| Amarelo | 1.222 | 2,92% | 221,05 | 42,76 |
| Indígena | 1.190 | 2,85% | 225,56 | 39,95 |
| Não quis declarar | 5.482 | 13,13% | 217,67 | 39,94 |

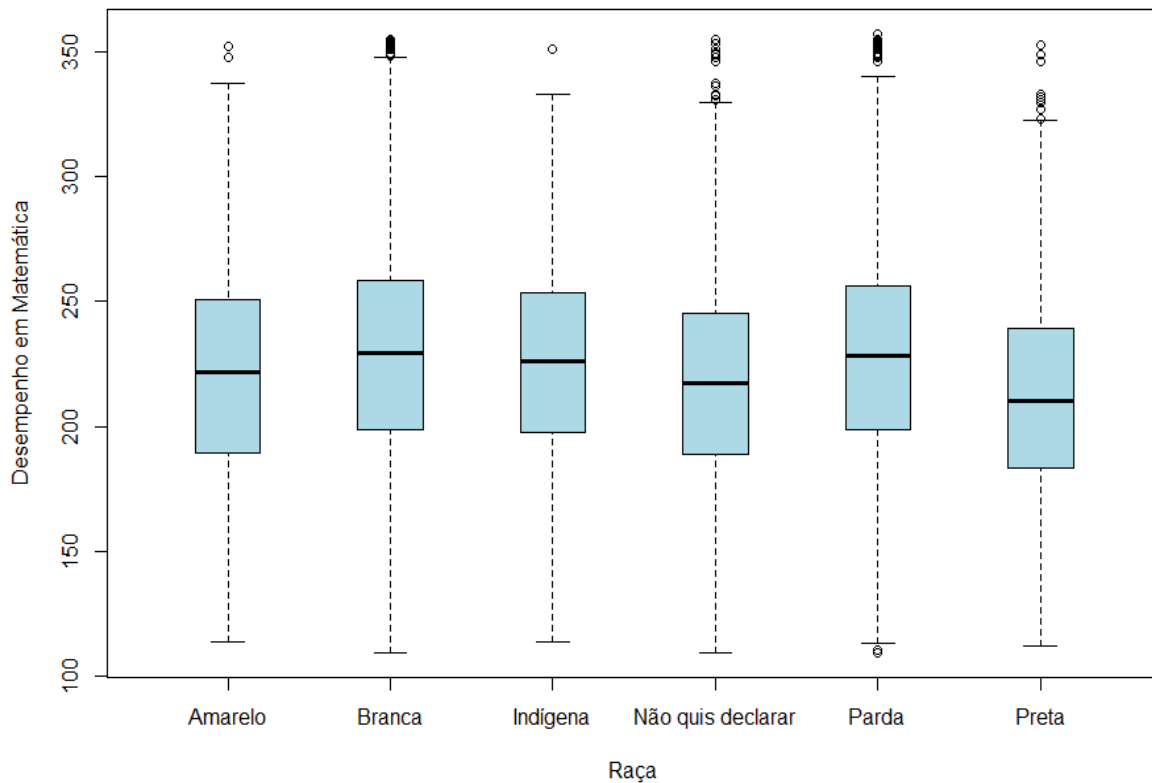


Figura 4: Desempenho em matemática por raça.

4.1.3 Análise Fatorial

Para a construção das condições socioeconômicas dos alunos, foi usado a análise fatorial, afim de verificar a relação de algumas respostas com o desempenho obtido por eles.

Segundo (HAIR et al., 2009), a análise fatorial deve ser feita com mais de cinco variáveis, sendo que foi feita com treze. Segue abaixo as respostas utilizadas para gerar a variável socioeconômico.

Tabela 5: Variáveis que compuseram a formulação da variável socioeconômico.

| Variáveis | Descrição/Codificação |
|---------------------------------------------------------------|-----------------------|
| Na sua casa tem televisão em cores? | 0: Não; 1: Sim |
| Na sua casa tem aparelho de rádio? | 0: Não; 1: Sim |
| Na sua casa tem videocassete e/ou DVD? | 0: Não; 1: Sim |
| Na sua casa tem geladeira? | 0: Não; 1: Sim |
| Na sua casa tem freezer (parte da geladeira duplex)? | 0: Não; 1: Sim |
| Na sua casa tem freezer separado da geladeira? | 0: Não; 1: Sim |
| Na sua casa tem máquina de lavar roupa? | 0: Não; 1: Sim |
| Na sua casa tem carro? | 0: Não; 1: Sim |
| Na sua casa tem computador? | 0: Não; 1: Sim |
| Na sua casa tem banheiro? | 0: Não; 1: Sim |
| Na sua casa tem quartos para dormir? | 0: Não; 1: Sim |
| Incluindo você, quantas pessoas vivem atualmente em sua casa? | 0: Não; 1: Sim |
| Em sua casa trabalha empregado(a) doméstico(a) 5 dias/semana? | 0: Não; 1: Sim |

4.1.4 Análise Multinível

Para a análise de regressão multinível, as variáveis foram recodificadas. A variável *Sexo*, (0=masculino, 1=feminino), *Cor/Raça/Etnia (Branca)*, (1=branca/amarelo, 0=caso contrário), *Cor/Raça/Etnia (Parda)*, (1=parda/preto, 0=caso contrário), *Você já foi reprovado?*, (0=não, 1=sim). Essa recodificação foi feita com análises feitas anteriormente, e com base nas literaturas citadas. Todas as variáveis que possuíam valores nulos foram retiradas da base, e as variáveis qualitativas e quantitativas foram padronizadas para facilitar a interpretação segundo os passos propostos por (HOX, 2010).

O primeiro passo proposto por (HOX, 2010), considera o modelo nulo, sem variáveis explicativas. Com base nesse modelo temos que a proficiência em matemática do 5º ano do Ensino Fundamental médio é representado pelo intercepto de 215,43 pontos. Podemos observar essa análise na Tabela 5.

Tabela 6: Passo1: Medidas resumo do desempenho em matemática segundo sexo.

| Efeito Fixo | Efeito | Erro Padrão | Estatística t |
|-------------------------------|-----------|---------------|---------------|
| Intercepto | 215,43 | 2,138 | 100,8 |
| Efeito randômico | Variância | Desvio Padrão | |
| Variância Nível 1 - aluno | 1.623,3 | 40,29 | - |
| Variância Nível 2 - município | 145,3 | 12,05 | - |
| ICC | | 8,22% | |
| Deviance | | 687.645,9 | |

Com os dados da Tabela 5 podemos calcular o coeficiente de correlação intraclasse,

$$ICC = \frac{\sigma_{\mu 0}^2}{\sigma_{\mu 0}^2 + \sigma_e^2} = \frac{145,3}{145,3 + 1623,3} = 8,22\% \quad (12)$$

O resultado mostra de 8,2% da variância do desempenho dos alunos está relacionada com o município, o que justifica o método utilizado para a análise, modelo de regressão multinível. Quando comparado o resultado com outros trabalhos (ANDRADE J. M.; LAROS, 2007), temos um ICC baixo, que pode ser facilmente explicado pelos municípios e regiões administrativas serem homogêneas.

Seguindo os passos propostos por Hox (2010), no segundo passo, foram inseridas as variáveis controle: *nível socioeconômico, sexo, etnia autodeclarada e Reprovação* (Tabela 6). O principal objetivo desse passo é verificar mais fatores que influenciam na variação do desempenho nos municípios/RA e dos alunos.

Tabela 7: Passo 2: modelo com variáveis de controle: Nível 1 - aluno.

| Efeito Fixo | Efeito | Erro Padrão | Estatística t |
|---------------------------------|--------|-------------|---------------|
| Intercepto | 218,34 | 1,92 | 113,18 |
| Variáveis controle | | | |
| Socioeconômico | 7,47 | 0,33 | 22,70 |
| Sexo (Feminino) | -7,32 | 0,30 | -24,58 |
| Reprovação | -27,81 | 0,35 | -78,65 |
| Cor/Raça/Etnia - Branca/Amarelo | 6,53 | 0,48 | 13,72 |
| Cor/Raça/Etnia - Parda/Preto | 4,90 | 0,42 | 11,68 |
| Efeito randômico | | Variância | Desvio Padrão |
| Variância Nível 1 - aluno | | 1.461,0 | 38,22 |
| Variância Nível 2 - município | | 112,1 | 10,59 |
| ICC | | 7,12% | |
| Deviance | | 680.558,4 | |
| N° de parâmetros | | 6 | |
| Teste de χ^2 p-valor | | <0,0001 | |
| Diferença Deviance | | 7.087,5 | |
| Diferença de parâmetros (g.l.) | | 3 | |
| Variância explicada | | | |
| Nível 1 - aluno | | 10,2% | |
| Nível 2 - município | | 25,81% | |

Usando o teste de de qui-quadrado indicado na Tabela 6 podemos observar que a diferença entre as deviances do Modelo nulo com o modelo com as variáveis de controle nível aluno é significativo, mostrando assim uma melhora no ajuste do modelo.

Observamos que a variável que teve um maior destaque foi a *Reprovação*, indicando que a diminuição de um desvio padrão, ou seja se o aluno não reprovar, aumenta em média 27,81 pontos no desempenho do aluno. Outro fator interessante é que as mulheres tem notas menores que as dos homens, se o aluno for do sexo feminino então a nota dela é 7,32 em média menor que se o aluno for do sexo masculino. O nível socioeconômico aumenta em média 7,48 pontos na nota dos alunos, na etnia autodeclarada foram feitos dois tipos, branca/amarelo e parda/preto, pois eram as que possuíam maior relevância na análise bivariada, quando falamos da branca podemos observar que se o aluno se declarou branca/amarelo, possui em média 6,53 pontos a

mais que os demais e quando ele se declarou pardo/preto ele possui em média 4,90 pontos a mais que os demais.

Podemos verificar que 25,81% da variância dos municípios ou regiões administrativas se explicam com as variáveis controle inseridas no modelo da Tabela 6. Por outro lado a inserção dessas variáveis explica cerca de 10,2% da variância dos alunos, sendo que a inserção dessas variáveis foi significativa nos dois níveis.

Lembrando que não podemos interpretar os efeitos erroneamente quando negativo, pois esse estudo é apenas observacional com corte transversal, onde não se pode falar em relações de causa e efeito.

A variável reprovação apresentou um efeito de -26,98, mostrando que em média o aluno que reprova tem uma nota 26,98 menor que os demais alunos.

Na Tabela 7 vamos ao passo 3, onde inserimos as variáveis de controle do nível mais alto, ou seja do nível 2 - município, a variável inserida será o *IDH-Município/RA* de 2010.

Com a inserção da variável *IDH-Município*, tivemos que ela explica para cada desvio padrão, 31,61 pontos em média no desempenho em matemática do aluno, tendo assim como a variável com a maior explicação, mostrando que o IDH é muito importante no desempenho dos alunos.

Tabela 8: Passo 3: modelo com variáveis de controle: Nível 2 - município.

| Efeito Fixo | Efeito | Erro Padrão | Estatística t |
|-------------------------------------|--------|-------------|---------------|
| Intercepto | 218,14 | 2,45 | 89,12 |
| Variáveis do nível aluno | | | |
| Socioeconômico | 7,47 | 0,33 | 22,70 |
| Sexo (Feminino) | -7,31 | 0,30 | -24,58 |
| Reprovação | -27,81 | 0,35 | -78,65 |
| Cor/Raça/Etnia - Branca/Amarelo | 6,53 | 0,48 | 13,72 |
| Cor/Raça/Etnia - Parda/Preto | 4,90 | 0,42 | 11,68 |
| Variáveis do nível da escola | | | |
| IDH Município/RA | 0,30 | 2,21 | 0,13 |
| Efeito randômico | | Variância | Desvio Padrão |
| Variância Nível 1 - aluno | | 1.461 | 38,22 |
| Variância Nível 2 - município | | 112 | 10,58 |
| ICC | | 7,12% | |
| Deviance | | 680.554,9 | |
| N° de parâmetros | | 6 | |
| Teste de qui-quadrado p-valor | | <0,001 | |
| Diferença Deviance | | 3,51 | |
| Diferença de parâmetros (g.l.) | | 1 | |
| Variância explicada | | | |
| Nível 1 - aluno | | 0,077% | |
| Nível 2 - município | | 40,05% | |

Ao analisar a inserção da variável *IDH-Município/RA*, podemos observar que a explicação da variabilidade do aluno foi quase nula, por outro lado tivemos uma explicação da variabilidade

do município de 40,05%, mostrando que o IDH influencia muito no desempenho do aluno, ou seja quanto melhor o IDH do município melhor será a nota do aluno.

No passo 4, vamos verificar o efeito randômico dos coeficientes de inclinação de todas as variáveis do nível aluno no município, ou seja se as variáveis do nível aluno se comportam de maneira diferente entre os municípios. Para testar o efeito iremos considerar todas as variáveis no efeito aleatório.

Tabela 9: Passo 4: Modelo com efeito randômico.

| Efeito Fixo | Efeito | Erro Padrão | Estatística t |
|----------------------------------------|--------|-------------|---------------|
| Intercepto | 216,83 | 18,75 | 11,57 |
| Variáveis do nível aluno | | | |
| Socioeconômico | 8,84 | 0,45 | 19,36 |
| Sexo (Feminino) | -7,95 | 4,71 | -1,69 |
| Reprovação | -26,17 | 25,80 | -1,01 |
| Cor/Raça/Etnia - Branca/Amarelo | 5,55 | 2,43 | 2,29 |
| Cor/Raça/Etnia - Parda/Preto | 4,37 | 0,66 | 6,14 |
| Variáveis do nível do município | | | |
| IDH Município/RA | 0,17 | 1,72 | 0,10 |
| Efeito randômico - Nível 1 | | Variância | Desvio Padrão |
| Variância Nível 1 - aluno | | 1.365,27 | 0,1863 |
| Efeito randômico - Nível 2 | | Variância | Desvio Padrão |
| Variância - município | | 98,38 | 9,92 |
| Variância - Socioeconômico | | 771,10 | 27,78 |
| Variância - Sexo (Feminino) | | 11,06 | 3,32 |
| Variância - Reprovação | | 332,72 | 18,24 |
| Variância - Branca/Amarelo | | 2,85 | 1,69 |
| Variância - Parda/Preto | | 0,13 | 0,36 |
| Deviance | | 671.446 | |
| N° de parâmetros | | 14 | |
| Diferença Deviance | | 9.108,9 | |
| Diferença de parâmetros (g.l.) | | 8 | |
| Teste de qui-quadrado p-valor | | 0,001 | |

Ao analisar os efeitos aleatórios do passo 4 podemos observar que a desigualdade escolar na Ride-DF fica clara quando observamos as variáveis *Socioeconômico*, que tem um efeito de 8,84 pontos no desempenho do aluno com um desvio padrão de 27,78, podendo em alguns municípios/RA chegar a anular esse efeito.

Na variável *Cor/Raça/Etnia - Branca/Amarelo*, temos um efeito de 5,55 pontos em média com um desvio padrão de 2,28 pontos, mostrando assim uma variabilidade grande entre os municípios/RA.

Quando falamos do *IDH Município/RA*, podemos observar um efeito de 0,18, mostrando que é a variável que teve o maior efeito médio, assim para cada 0,1 de IDH teremos cerca de 9,92 a mais na média do aluno do município.

Tivemos outras variáveis com peso grande, o *sexo*, teve um efeito médio de -7,94, indicando

que em média o sexo feminino possui uma nota média 7,94 pontos menor que o sexo masculino, com um desvio padrão de 3,35. Outra variável que pesou bastante foi a *Reprovação*, que teve um efeito de -26,17, indicando que se o aluno reprovou ele tem em média 26,17 pontos a menos em relação ao aluno que não reprovou, ainda tivemos um desvio padrão de 18,24.

O passo 5 não será feito nessa análise, que seria a parte da interação dos dados, pois não se mostrou significativo, onde o passo 4 já se mostrou suficiente para para o que foi proposto.

Nas Figuras 5 e 6 vamos vamos verificar as suposições de normalidade, homocedasticidade e linearidade para o passo 4. Foi feita a análise de resíduos dos níveis 1 e 2. Na Figura 5 podemos observar que não existe desvios dos pontos em relação a reta, indicando assim uma adequação da normalidade, indo para a Figura 6 temos que os pontos estão distribuídos de forma aleatória em torno da reta zero, mostrando assim a linearidade e a variância constante.

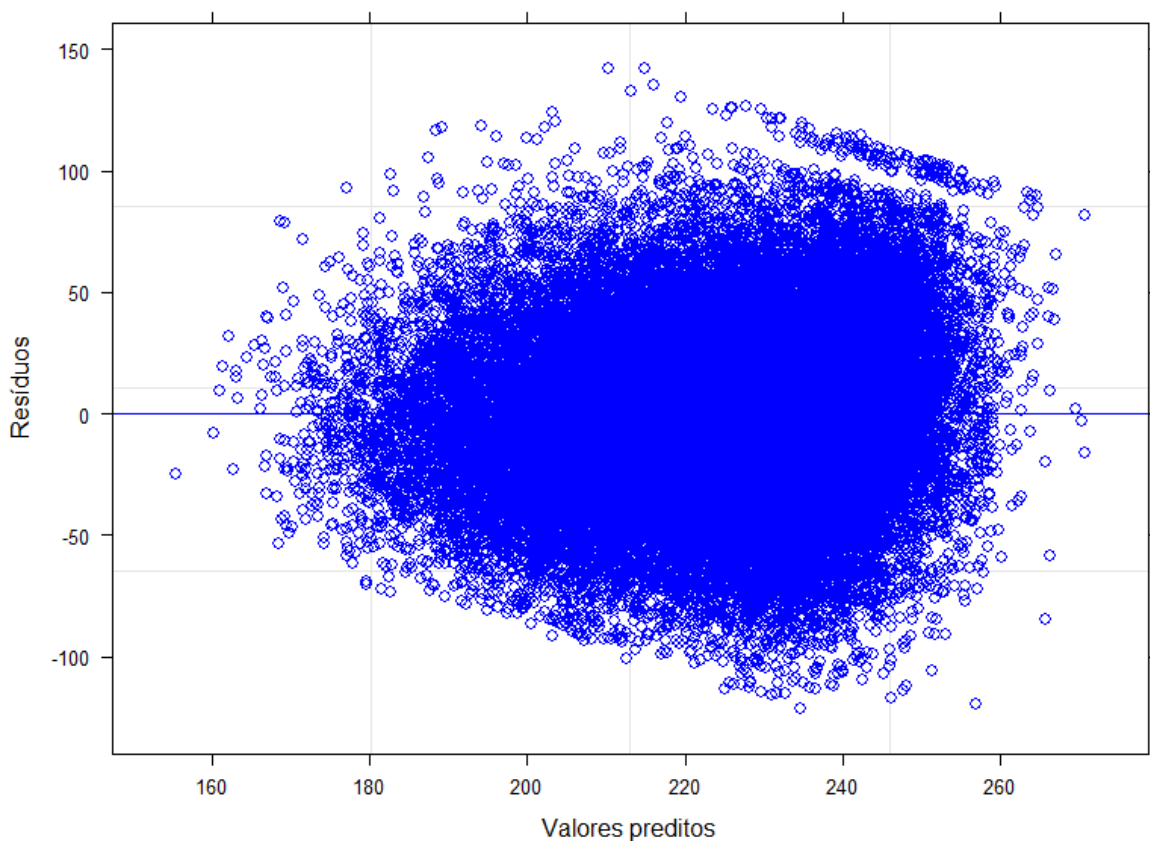


Figura 5: Gráfico dos resíduos versus valores preditos.

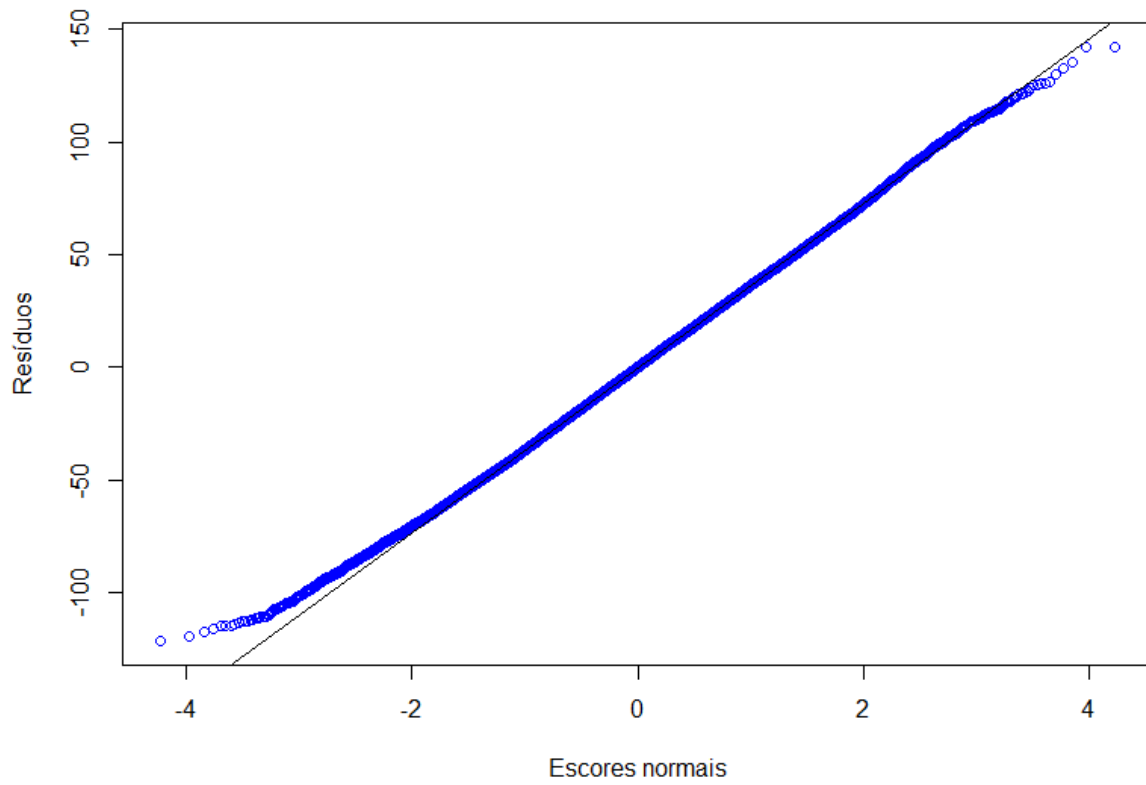


Figura 6: Gráfico dos resíduos padronizados versus escores normais.

5 Conclusão

No estudo será considerado o modelo 4, pois apesar de ele não apresentar uma deviance significativa em relação ao modelo 3, ele apresenta os desvios em torno das variáveis, assim sendo no estudo esse será o modelo final.

O estudo propõe analisar através do modelo de regressão multinível, alguns fatores que influenciam o desempenho de alunos da Ride-DF, considerando a Etnia autodeclarada, o nível socioeconômico, o *IDH - Município/RA* e a *reprovação do aluno*.

Considerando as variáveis Nível 1 - aluno, podemos observar que elas explicam 10,2% da variância do desempenho dos alunos, já no nível do Município/RA essa explicação chegou à 40,05%, mostrando o quão importante é o *IDH - Município/RA*.

Quando falamos da *etnia branca e parda*, observamos que é um fato que influencia bastante na nota média dos alunos, o nível socioeconômico ficou próximo da etnia.

No que diz respeito aos efeitos negativos, tivemos a variável reprovação, ao qual pesou muito na nota média do aluno, tirando em média 26,9 pontos a menos em média que os demais alunos. Em relação ao *sexo feminino* ao qual mostra que as mulheres possuem notas menores que os homens, em literaturas da área observamos o contrário, onde as mulheres possuem notas maiores que os homens na disciplina de português.

No passo 4 podemos observar os efeitos aleatórios de todas as variáveis e assim podemos ver de fato as diferenças e suas oscilações em cada variável.

Os estudos na área da educação ainda são poucos, precisando de uma maior atenção tanto de pesquisadores quanto de autoridades, para que aja uma aplicação dos estudos e assim um incentivo maior da parte do governo.

Referências

- ANDRADE J. M.; LAROS, J. A. *um estudo multinível com dados do SAEB/2001. Psicologia: Teoria e Pesquisa*. [S.l.: s.n.], 2007.
- ARROYO, M. G. Políticas educacionais e desigualdades: à procura de novos significados. *Educação & Sociedade*, SciELO Brasil, v. 31, n. 113, p. 1381–1416, 2010.
- BATES, D. et al. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, v. 67, n. 1, p. 1–48, 2015.
- HAIR, J. F. et al. *Análise multivariada de dados*. [S.l.]: Bookman editora, 2009.
- HOX, J. J. *Techniques and Applications. 2 ed. Grã - Bretanha: Routledge*. [S.l.: s.n.], 2010. v. 2.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2020. Disponível em: <http://www.R-project.org/>.
- SANTOS, T. *Avaliação da influência de fatores associados ao clima em sala de aula a partir da percepção de estudantes do 1º ano do Ensino Médio sobre o desempenho em matemática*. [S.l.: s.n.], 2017.
- STEVENS, J. P. *Applied multivariate statistics for the social sciences*. [S.l.]: Routledge, 2012.
- VINHA, L. G. d. A.; KARINO, C. A.; LAROS, J. A. Fatores associados ao desempenho em matemática no ensino fundamental no brasil. *Psico-USF*, SciELO Brasil, v. 21, n. 1, p. 87–100, 2016.