



**Universidade de Brasília
Departamento de Estatística**

**Fatores Associados ao Desempenho de Alunos do 3^a Ano do Ensino Médio do
Distrito Federal:
Uma Análise Multinível Aplicada a Dados do SAEB 2013**

Guilherme Maia Batista

Projeto apresentado para obtenção do título
de Bacharel em Estatística.

**Brasília
2016**

Guilherme Maia Batista

**Fatores Associados ao Desempenho de Alunos do 3^a Ano do Ensino Médio do
Distrito Federal:
Uma Análise Multinível Aplicada a Dados do SAEB 2013**

Orientadora:
Profa. Maria Teresa Leão Costa

Projeto apresentado para obtenção do título de Bacharel em
Estatística.

**Brasília
2016**

AGRADECIMENTOS

Há alguns anos eu soube do resultado da minha aprovação na universidade, no começo eu quase não acreditei, mas quando caiu a ficha eu fiquei muito feliz de ter a oportunidade de estudar na Universidade de Brasília. O curso de Estatística me trouxe vários desafios gratificantes, sendo este trabalho um deles.

Estou muito satisfeito de estar nesta reta final, que eu não cheguei sozinho. Eu agradeço aos professores com quem eu pude ter aulas, em especial à professora Maria Teresa que me orientou na elaboração deste trabalho. Também agradeço a todos os meus amigos, inclusive os que ajudaram a tornar os estudos bem mais divertidos: Mariana, Geiziane, Victor, Agda, Andressa, Mateus, Márcia e Nahari. E por fim, eu agradeço à minha família, principalmente ao meu pai Marcial e à minha mãe Ianeli que me ensinaram tantas coisas.

RESUMO

Neste trabalho, foi estudado o desempenho dos alunos do 3º ano do Ensino Médio do Distrito Federal com base nos dados da Aneb (Avaliação Nacional da Educação Básica) de 2013, a mais recente. A Aneb foi criada pelo SAEB (Sistema de Avaliação da Educação Básica) e é composta por uma prova de matemática e outra de língua portuguesa. Tendo em vista a estrutura hierárquica dos dados, foi utilizada a técnica Análise Multinível. Sendo assim, foram construídos dois modelos: um para a proficiência em matemática e outro para a proficiência em língua portuguesa. A comparação entre os modelos mostrou algumas diferenças, como a variável sexo ser significativa para explicar apenas a disciplina matemática. Para o nível da escola, o único fator que se mostrou significativo foi o tipo de dependência administrativa para ambos os modelos. Aprimorar a educação deve ser um objetivo constante para uma sociedade, por isso, este trabalho pode servir de auxílio para conhecer os principais fatores associados ao desempenho dos alunos do 3º ano do Ensino Médio no Distrito Federal. **Palavras-chave:** regressão multinível, análise multinível, dados hierárquicos, dados aninhados, SAEB, Aneb, educação, Ensino Médio.

ABSTRACT

In this work, we studied the performance of students of the 3rd year of high school in Distrito Federal based on the data from Aneb (National Assessment of Basic Education) 2013, the latest. The Aneb was created by SAEB (System of Basic Education Evaluation) and consists of a math test and a portuguese test. Given the hierarchical structure of the data, we used the technique Multilevel Analysis. Hence, two models were built: one for proficiency in math and another for proficiency in portuguese. The comparison between the models showed some differences, such as the variable gender being significant only for mathematics. For the school level, the only significant factor was the type of administrative dependence for both models. Improving education must be a constant goal for a society, so this work can be of help to know the main factors associated with the performance of students of the 3rd year of high school in Distrito Federal.

Key-words: multilevel regression, multilevel analysis, hierarchic data, nested data, SAEB, Aneb, education, High School.

LISTA DE TABELAS

Tabela 4.1 - Pontuação recebida pela posse de cada item no CCEB 2014	18
Tabela 4.2 - Pontuação recebida pelo grau de instrução dos pais no CCEB 2014	18
Tabela 4.3 - Intervalos das categorias do CCEB 2014	18
Tabela 4.4 - Variáveis construídas para os professores	19
Tabela 4.5 - Variáveis construídas para os alunos	19
Tabela 5.1 - Distribuição das escolas conforme tipo de dependência administrativa e localização	21
Tabela 5.2 - Distribuição das escolas conforme existência e estado de conservação de estruturas	22
Tabela 5.3 - Distribuição das escolas conforme existência e estado de conservação de ambientes	22
Tabela 5.4 - Distribuição dos professores conforme sexo e idade	23
Tabela 5.5 - Distribuição dos diretores, professores e alunos conforme raça/cor	24
Tabela 5.6 - Distribuição dos professores e diretores conforme educação	24
Tabela 5.7 - Distribuição dos professores conforme características profissionais	25
Tabela 5.8 - Distribuição dos diretores conforme características	26
Tabela 5.9 - Distribuição dos alunos conforme características	27
Tabela 5.10 - Distribuição dos alunos conforme classe econômica para escolas públicas e privadas	28
Tabela 5.11 - Distribuição dos alunos conforme hábitos de leitura	28
Tabela 5.12 - Distribuição dos alunos conforme hábitos fora da escola	29
Tabela 5.13 - Distribuição dos alunos conforme hábitos de estudo	29
Tabela 5.14 - Modelos somente com o intercepto para matemática e língua portuguesa	35
Tabela 5.15 - Modelos com as variáveis do nível do aluno, com coeficientes de efeito fixo para as variáveis, para matemática e língua portuguesa	36
Tabela 5.16 - Modelos com as variáveis do nível do aluno e da escola, com coeficientes de efeito fixo para as variáveis, para matemática e língua portuguesa	37

LISTA DE FIGURAS

Figura 1.1 - Esquema da composição do SAEB	1
Figura 1.2 - Esquema de dados organizados em dois níveis	2
Figura 3.1 - Exemplo de gráfico de quantis	12
Figura 3.2 - Exemplo de gráfico de resíduos organizados em ordem temporal	13
Figura 3.3 - Exemplo de gráfico dos resíduos e valores esperados	13
Figura 5.1 - Diagramas de caixa da proficiência dos alunos em língua portuguesa e matemática	28
Figura 5.2 - Diagramas de caixa do desempenho do aluno conforme sexo	29
Figura 5.3 - Diagramas de caixa do desempenho do aluno conforme poder aquisitivo ...	29
Figura 5.4 - Diagramas de caixa do desempenho do aluno conforme número de reprovações ...	30
Figura 5.5 - Diagramas de caixa do desempenho do aluno conforme hábito de ler livros em geral	30
Figura 5.6 - Diagrama de caixa do desempenho do aluno que gosta ou não de estudar a matéria	31
Figura 5.7 - Diagramas de caixa do desempenho do aluno conforme raça/cor	31
Figura 5.8 - Diagramas de caixa do desempenho do aluno conforme tipo de administração da escola	32
Figura 5.9 - Gráfico de resíduos e valores esperados para os modelos finais	36
Figura 5.10 - Gráfico de quantis para os modelos finais	36

SUMÁRIO

1 INTRODUÇÃO	1
2 OBJETIVOS	5
3 REFERENCIAL TEÓRICO	7
3.1 Construção do Modelo de Regressão Multinível	8
3.2 Suposições do Modelo	9
3.3 Agregação e Desagregação	9
3.4 A Influência da Média	10
3.5 Estimação dos Parâmetros	10
3.6 Qualidade do Modelo	11
3.7 Inferência	11
3.8 Diagnóstico dos Resíduos	12
3.9 Etapas da Análise Multinível	14
4 METODOLOGIA	15
5 RESULTADOS	19
5.1 Apresentação dos Dados	19
5.2 Análise Bivariada	29
5.3 Aplicação do Modelo Multinível	33
6 CONCLUSÃO	39
REFERÊNCIAS	41
AANEXO	42

1 INTRODUÇÃO

Educação de qualidade é uma característica necessária para o desenvolvimento de um país. A estrutura educacional no Brasil é formada por dois grandes grupos: Educação Básica e Educação Superior, sendo o governo obrigado por lei a oferecer vagas para a Educação Básica. A estrutura da Educação Básica é formada por três etapas: Educação Infantil (3 a 5 anos de idade), Ensino Fundamental (6 a 14 anos de idade) e Ensino Médio (15 a 17 anos). Ao concluir a Educação Básica, o indivíduo pode escolher prosseguir os estudos ingressando na Educação Superior, a partir dos 18 anos (Portal Brasil, 2016).

O INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira), autarquia federal ligada ao Ministério da Educação, criou em 1990 o SAEB (Sistema de Avaliação da Educação Básica) com o seguinte objetivo: “produzir informações que subsidiem a formulação, reformulação e o monitoramento das políticas públicas, nas esferas municipal, estadual e federal visando a contribuir para a melhoria da qualidade, equidade e eficiência do ensino” (INEP).

O SAEB é composto por três tipos de avaliações: Aneb (Avaliação Nacional da Educação Básica), Anresc (Avaliação Nacional do Rendimento Escolar - também denominada Prova Brasil) e a ANA (Avaliação Nacional da Alfabetização). A Aneb é uma avaliação amostral feita a cada dois anos que busca colher dados representativos do país, de forma que a amostra abranja escolas públicas e privadas, em diferentes localizações (rural ou urbana) e áreas (Capital ou interior), nas unidades da Federação, para alunos da 5º ano, 9º ano do Ensino Fundamental e 3º ano do Ensino do Ensino Médio. A Anresc também é feita a cada dois anos, de forma censitária, para as escolas do país; ela é aplicada para os alunos da 5º e 9º ano do Ensino Fundamental. E, por último, o ANA (Avaliação Nacional da Alfabetização) que é uma avaliação censitária que envolve os alunos do 3º ano do Ensino Fundamental das escolas públicas do país.

Figura 1.1 – Esquema da composição do SAEB.



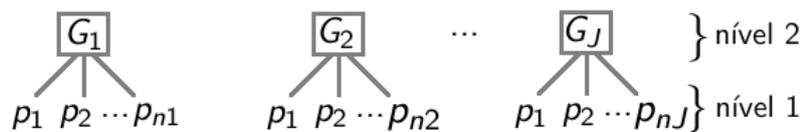
Fonte: INEP.

O 3º ano do Ensino Médio é o último ano da educação obrigatória, portanto conhecer o perfil e o desempenho dos alunos que estão concluindo esta etapa ajuda a mostrar os resultados desse processo de formação do indivíduo. Neste contexto, questões acerca da educação surgem, como qual o perfil dos estudantes prestes a concluírem a Educação Básica. Também, quais as características das suas escolas, professores e diretores. A partir dos dados da Aneb de 2013 serão estudados os fatores associados ao desempenho dos alunos. Tendo em vista a estrutura hierárquica dos dados, será utilizada a técnica Análise Multinível.

A análise multinível é um caso específico da regressão usado para estudar populações estruturadas de forma hierárquica, sendo um caso particular de modelos de efeitos aleatórios. Em estudos sociais, por exemplo, o indivíduo é influenciado pelo meio social em que ele vive, e em contrapartida, o meio é influenciado pelos indivíduos que o compõe. É de interesse do pesquisador poder medir essa influência, o que é possível através de uma análise multinível.

A partir dos dados se constrói um modelo de regressão multinível, tal que o nível mais baixo (nível 1) refere-se ao indivíduo, enquanto o nível 2 aos grupos de indivíduos, o nível 3 aos grupos de grupos, e assim por diante. Não há limite para o número de níveis no modelo, apenas a necessidade de amostras cada vez maiores à medida que se aumenta o número de níveis. Na figura abaixo é possível observar como estão organizados os dados quando há 2 níveis.

Figura 1.2 – Esquema de dados organizados em dois níveis.



G_j : j-ésimo grupo, $j=1,2,\dots,J$

p_i : i-ésimo indivíduo do j-ésimo grupo, $i=1,2,\dots,n_j$

Fonte: Adaptado do Hox.

A amostragem de dados estruturados hierarquicamente é realizada do nível mais alto para o nível mais baixo. Ou seja, no caso de dados escolares, primeiramente são selecionadas as escolas (nível 2) da amostra, e para cada escola é selecionado uma amostra de alunos (nível 1).

O indivíduo está submetido às particularidades do meio social em que ele vive, diferenciando-o daqueles que vivem em outro meio social. Por este motivo, características que se desejam estudar podem ser explicadas não pelo nível do indivíduo, mas sim pelo nível do grupo o qual ele pertence. Logo, um modelo que não discrimina qual o grupo do indivíduo pode trazer informações que não condizem com a realidade.

Por exemplo, supondo que queira se estudar questões ligadas à educação. Sendo a unidade de informação o estudante, espera-se que a escola em que ele estude tenha um papel importante na definição das suas características educacionais. Assim, o uso da análise multinível se mostra eficaz, uma vez que ela capta a variação entre os grupos (escolas, no caso).

Neste trabalho serão analisados os dados do SAEB (Sistema de Avaliação da Educação Brasileira) da Avaliação Nacional da Educação Básica para os alunos do 3º ano do Ensino Médio do Distrito Federal para o ano de 2013 em busca dos principais fatores associados ao desempenho dos alunos. Dada a estrutura dos dados do estudo, que são organizados de forma hierárquica, justifica-se a utilização da técnica Análise Multinível. Desta forma, será construído um modelo para explicar o desempenho do aluno na prova de língua portuguesa e outro para explicar o seu desempenho na prova de matemática.

2 OBJETIVOS

O objetivo principal deste trabalho é conhecer os principais fatores associados ao rendimento escolar dos alunos do 3º ano do Ensino Médio, utilizando os dados da Aneb 2013. Para isso, foi utilizada a técnica análise multinível. Abaixo está a lista dos objetivos específicos:

- Estudar o modelo de regressão multinível;
- Preparação e análise descritiva do banco de dados
- Estudar o método de uso do programa estatístico R para construir um modelo multinível;
- Identificar e interpretar os principais fatores associados ao rendimento escolar dos alunos através da construção de um modelo multinível.

3 REFERENCIAL TEÓRICO

O principal objetivo de uma análise de regressão é tirar conclusões de uma variável a partir de um conjunto de outras variáveis. Porém, esta relação não é perfeita, ou seja, uma parcela dessa relação é aleatória, não podendo ser prevista.

Em uma análise de regressão é necessário um conjunto de suposições probabilísticas bem definidas em relação à parte aleatória do modelo, conhecida como erro. Por meio dessas suposições é possível realizar inferências sobre o modelo (como intervalos de confiança, testes de hipótese para os parâmetros do modelo e para os valores esperados da variável resposta).

Entre os modelos, o mais utilizado é o linear, por sua fácil aplicação e interpretação. Por este motivo, ele será usado como um exemplo hipotético. Suponha que em uma escola deseja-se estudar o desempenho médio dos alunos de todas as turmas do último ano no vestibular, de acordo com o seu sexo e nota obtida em um teste de conhecimentos gerais. O modelo pode ser escrito como:

$$y_i = \beta_0 + \beta_1 \times x_{1i} + \beta_2 \times x_{2i} + e_i \quad (3.1)$$

Sendo

- y_i a nota no vestibular do i -ésimo aluno
- x_{1i} o sexo do i -ésimo aluno
- x_{2i} a nota no teste obtida pelo i -ésimo aluno
- e_i um erro aleatório atribuído ao i -ésimo aluno
- Todos os β 's são parâmetros fixos que podem ser estimados

O modelo linear acima assume que as turmas são homogêneas entre elas, implicando independência entre os erros, o que pode não ser verdade. Este é um caso específico onde pode-se aplicar o modelo de regressão multinível, que capta a informação de que existe variação de uma turma para outra. Uma regressão multinível pode ter vários níveis. Neste exemplo são dois níveis, onde o primeiro é o aluno e o segundo é a turma. Consequentemente o aluno está aninhado à turma que ele pertence. Todos os níveis podem ter variáveis explicativas.

Agora os β 's não são mais parâmetros fixos, mas sim parâmetros aleatórios que dependem da turma do aluno. O modelo pode ser escrito como:

$$y_{ij} = \beta_{0j} + \beta_{1j} \times x_{1ij} + \beta_{2j} \times x_{2ij} + e_{ij} \quad (3.2)$$

O índice j especifica a turma, enquanto o índice i especifica o i -ésimo aluno da respectiva turma. Suponha que o segundo nível (turma) tenha como variável explicativa os anos de experiência, medida em anos, do professor.

Sendo Z_j os anos de experiência do professor da j -ésima turma, pode-se escrever β_{0j} como:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} \times Z_j + u_{0j} \quad (3.3)$$

Tal que γ_{00} é um parâmetro fixo comum a todas as turmas; γ_{01} é um parâmetro fixo associado à experiência do professor; e u_{0j} é uma variável aleatória associada à turma j . Seguindo o mesmo caminho, obtemos β_{1j} e β_{2j} . Agora, ao usar esta forma expandida dos β 's no modelo, será obtida a seguinte equação:

$$y_{ij} = \gamma_{00} + \gamma_{01} \times Z_j + u_{0j} + \gamma_{10} \times x_{1ij} + \gamma_{11} \times Z_j \times x_{1ij} + u_{1j} \times x_{1ij} + \gamma_{20} \times x_{2ij} + \gamma_{21} \times Z_j \times x_{2ij} + u_{2j} \times x_{2ij} + e_{ij} \quad (3.4)$$

É fácil notar que à medida que o número de níveis aumenta o modelo fica cada vez mais complexo. Comparando o modelo com apenas um nível e o modelo com dois níveis, pode-se notar que enquanto o primeiro possui apenas um componente aleatório (e_{ij}), o segundo possui quatro (e_{ij} , u_{0j} , u_{1j} , u_{2j}).

3.1 Construção do Modelo de Regressão Multinível

Nesta seção será tratado, de forma mais geral, sobre o modelo com 2 níveis para variáveis respostas contínuas, por ser o mais comum, e também mais simples de se trabalhar e interpretar. Porém, modelos com mais níveis e/ou diferentes variáveis respostas possuem uma metodologia análoga, sendo as diferenças apenas adaptações (variável resposta qualitativa ou multivariada) ou inclusão de mais variáveis.

Suponha dados com p variáveis referentes ao nível 1, e k variáveis referentes ao nível 2. Temos a equação:

$$y_{ij} = \beta_{0j} + \beta_{1j} \times x_{1ij} + \beta_{2j} \times x_{2ij} + \dots + \beta_{pj} \times x_{pij} + e_{ij} \quad (3.5)$$

Sendo que o índice i indica o i -ésimo indivíduo do j -ésimo grupo ($i = 1, \dots, n_j$), enquanto o índice j indica em qual grupo pertence o indivíduo ($j = 1, \dots, J$). Os x 's são as variáveis referentes ao indivíduo (nível 1). O erro (e_{ij}) é o componente aleatório de cada indivíduo. Os β 's em uma regressão multinível não são parâmetros constantes, eles variam de acordo com a turma. Cada turma tem um conjunto de parâmetros diferente, que podem ser escritos na forma:

$$\begin{aligned}\beta_{0j} &= \gamma_{00} + \gamma_{01} \times Z_{1j} + \dots + \gamma_{0k} \times Z_{kj} + u_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11} \times Z_{1j} + \dots + \gamma_{1k} \times Z_{kj} + u_{1j} \\ &\vdots \\ \beta_{pj} &= \gamma_{p0} + \gamma_{p1} \times Z_{1j} + \dots + \gamma_{pk} \times Z_{kj} + u_{pj}\end{aligned}\tag{3.6}$$

Sendo que os γ 's são parâmetros fixos. Os Z 's são as variáveis referentes ao nível 2. Os u 's são os componentes aleatórios dos grupos.

3.2 Suposições do Modelo

Um modelo de regressão é estruturado sob suposições a respeito da parte aleatória do modelo. Neste caso são $p + 2$ componentes ($e_{ij}, u_{0j}, u_{1j}, \dots, u_{pj}$).

Supõe-se que $e_{ij}, u_{0j}, u_{1j}, \dots, u_{pj}$ seguem uma distribuição normal multivariada com médias iguais a zero e respectivas variâncias $\sigma_e^2, \sigma_{u_0}^2, \sigma_{u_1}^2, \dots, \sigma_{u_p}^2$. O erro do nível mais baixo não tem correlação com nenhum outro componente aleatório, por ele ser um distúrbio particular de cada indivíduo. Por outro lado, existem correlações entre os erros do nível 2, denotadas: $\sigma_{u_{01}}, \sigma_{u_{02}}, \dots, \sigma_{u_{(p-1)(p)}}$.

Essas suposições são necessárias para viabilizar inferências, como estimações dos parâmetros, intervalos de confiança, e previsões de novas observações.

3.3 Agregação e Desagregação

Em estudos, inicialmente, as variáveis estão definidas nos seus respectivos níveis, porém é possível realocá-las para diferentes níveis, isto é feito através de agregação ou desagregação de variáveis.

Agregação é quando se transfere uma variável do seu nível para o seu nível superior. Por exemplo, as rendas de cada domicílio (nível 1) transformadas na renda média

do bairro (nível 2). Enquanto que desagregação é quando se transfere uma variável do seu nível para o seu nível inferior. Por exemplo, utilizar uma variável indicadora de qual escola pertence cada aluno.

Hipoteticamente, seria possível agregar ou desagregar todas as variáveis, de forma que todas elas ficassem no mesmo nível. E, a partir desse ponto, faria-se uma análise de regressão padrão (com apenas 1 nível). Porém, esta abordagem está sujeita a erros que podem comprometer a análise.

Com a agregação perde-se informação e, conseqüentemente, o poder da análise. Já com a desagregação, vão surgir vários valores repetidos (já que os indivíduos do mesmo grupo têm o mesmo valor para as variáveis do nível superior), e isso causaria testes significativos quando, na realidade, não são.

3.4 A Influência da Média

Em alguns casos o uso da agregação/desagregação é benéfico para a análise. Um desses casos é para medir a influência da média do grupo no indivíduo. Por exemplo, em uma escola com média alta de desempenho, um aluno de médio desempenho pode se sentir desencorajado. Analogamente, em uma escola com média baixa de desempenho, um aluno de alto desempenho pode se sentir desmotivado. Portanto, nesta situação, pode-se usar além das notas de cada indivíduo (variável do nível 1), a nota média da escola (variável do nível 2).

3.5 Estimação dos Parâmetros

A estimação do modelo será feita pelo método de máxima verossimilhança (MV), que maximiza uma função de verossimilhança. Duas funções de verossimilhança são usadas na análise multinível: a função de máxima verossimilhança completa (MVC), e a função de máxima verossimilhança restrita (MVR). A primeira inclui tanto os coeficientes da regressão quanto os componentes da variância. Enquanto que a segunda função inclui apenas os componentes da variância. A MVC é enviesada, porém, esse viés costuma ser pequeno. Conseqüentemente, os dois métodos geram resultados bem próximos.

O método de máxima verossimilhança requer um processo iterativo, que pode

não convergir. Caso isso aconteça, uma alternativa é usar os mínimos quadrados generalizados (MQG), uma extensão dos mínimos quadrados ordinários que permite heterogeneidade e correlação entre as observações. Além disso, o MQG é computacionalmente bem mais rápido que a MV, por isso, pode ser usado como uma avaliação inicial do modelo.

3.6 Qualidade do Modelo

O objetivo do pesquisador é obter o melhor modelo possível, e para isso são necessárias maneiras de avaliar a qualidade de cada modelo, e comparação entre eles.

Através do método de MV, é obtida uma estatística denotada *deviance*, dada por

$$-2 \times \ln(MV) \quad (3.7)$$

A *deviance* mede o quão bem o modelo se ajusta aos dados. Valores baixos da *deviance* indicam um bom ajuste.

Dois modelos são aninhados quando o conjunto de parâmetros utilizado em um é subconjunto do outro. Para modelos aninhados, é possível realizar um teste qui-quadrado para definir se o ajuste do modelo mais completo difere significativamente do modelo mais simples. Este teste é obtido pela diferença das *deviances* que segue uma qui-quadrado com graus de liberdade igual à diferença de parâmetros entre os modelos.

Agora, para modelos não aninhados, uma medida geral de qualidade do ajuste do modelo é o AIC (*Akaike's Information Criterion*). Sendo p^* o número de parâmetros do modelo, o AIC é dado por:

$$AIC = deviance + 2 \times p^* \quad (3.8)$$

Além do *deviance*, o AIC utiliza na sua medida uma penalização do ajuste de acordo com o número de parâmetros a serem estimados.

3.7 Inferência

Os parâmetros fixos, de certa forma, quantificam a contribuição de uma variável no modelo. Por isso, são feitos testes de hipótese sobre os parâmetros para definir quais variáveis podem ser descartadas. Utiliza-se o teste de Wald, baseado na estatística

$$\frac{\hat{\theta}}{EP(\hat{\theta})} \quad (3.9)$$

que tem distribuição normal. Onde $\hat{\theta}$ é a estimativa do parâmetro, e EP o erro padrão dessa estimativa. A validade deste teste requer amostras grandes.

Para verificar a significância de um componente aleatório são comparadas as *deviances* do modelo com o componente e do modelo sem o componente. A partir desse ponto, utiliza-se o teste mostrado na seção anterior.

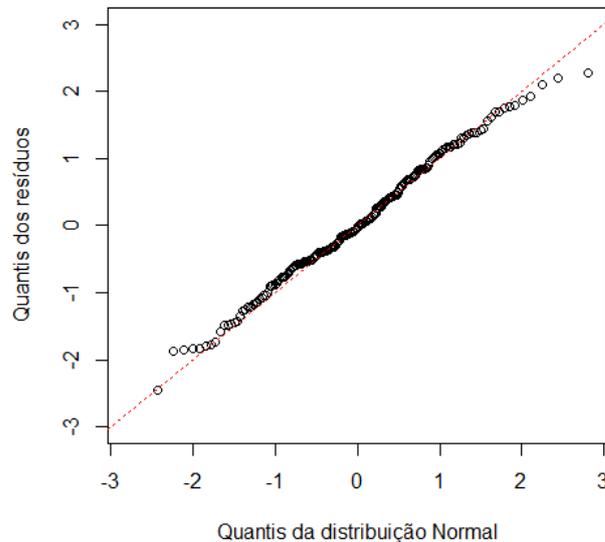
Com os parâmetros estimados, é possível realizar um dos principais objetivos da regressão: obter valores esperados da variável resposta para diferentes conjuntos de variáveis explicativas. Vale lembrar que a consistência de estimação do modelo se limita à vizinhança dos dados observados.

3.8 Diagnóstico dos Resíduos

Verificar se os resíduos atendem às suposições do modelo é essencial, uma vez que as suposições não sendo atendidas, a qualidade do modelo está comprometida. Na análise multinível, além do distúrbio aleatório do indivíduo, cada β tem um resíduo que deve ser estudado.

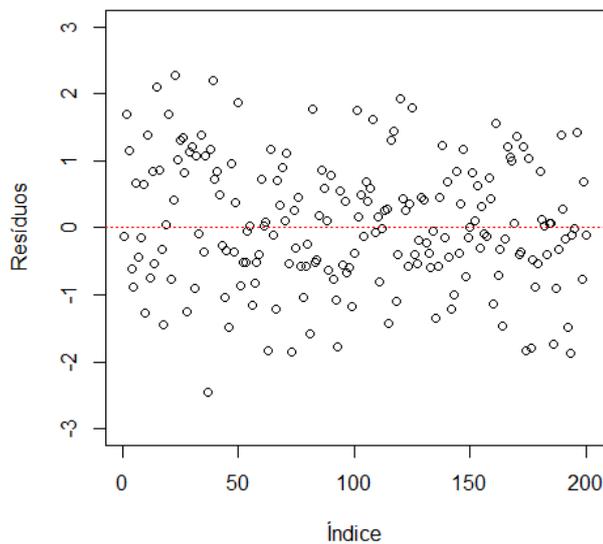
Através do gráfico dos quantis da distribuição normal contra os quantis dos resíduos padronizados, é possível verificar se existem violações das suposições. A situação desejada é quando este gráfico apresenta uma linha diagonal, ou seja, os quantis são praticamente iguais.

Figura 3.1 – Exemplo de gráfico de quantis.



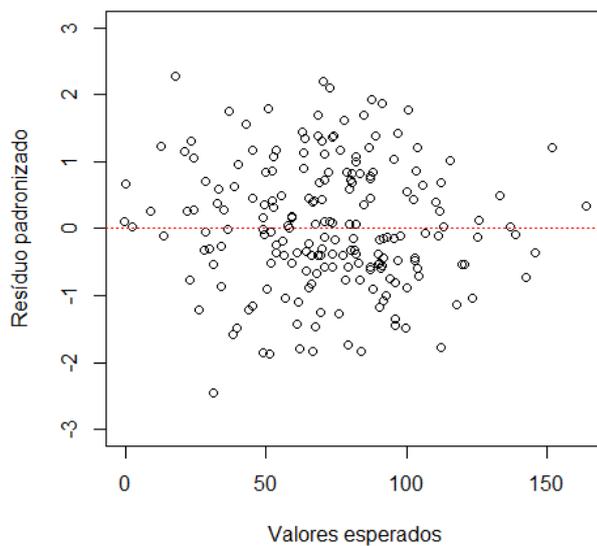
Para dados analisados em diferentes períodos, é interessante plotar os resíduos em ordem temporal, para verificar se houve alguma variação dos resíduos no tempo. Neste gráfico deseja-se observar resíduos alinhados horizontalmente em torno do zero (valor esperado dos resíduos).

Figura 3.2 – Exemplo de gráfico de resíduos organizados em ordem temporal.



Outro recurso é construir um gráfico de dispersão dos resíduos contra os valores esperados das variáveis repostas. Deseja-se observar uma nuvem sem nenhuma tendência.

Figura 3.3 – Exemplo de gráfico dos resíduos e valores esperados.



Ao detectar valores discrepantes, é necessário estudar o motivo dessa ocorrência. Essa discrepância pode ter sido consequência de um erro experimental. Caso contrário, a inclusão de novas variáveis pode expressar o que estava sendo equivocadamente captado pelo erro.

3.9 Etapas da Análise Multinível

Na construção de um modelo é sempre bom começá-lo da forma mais simples possível, e ir tornando-o mais complexo até se alcançar o modelo “ótimo”. Nesta seção será descrito um roteiro de como realizar uma análise multinível em cinco passos. À medida que se acrescentam mais informações, é comparado o modelo novo com o antigo para verificar se houve melhoria significativa.

Primeiro passo: Construir um modelo apenas com o intercepto. Este é um bom ponto inicial, já que com ele também é possível obter a correlação intraclasse, pela equação

$$\frac{\sigma_{u0}^2}{\sigma_e^2 + \sigma_{u0}^2} \quad (3.10)$$

Obter a correlação intraclasse é essencial para definir se é ou não necessária uma análise multinível dos dados.

Segundo passo: Agora são adicionadas as variáveis explicativas do primeiro nível. Nesta etapa é possível observar a contribuição de cada variável do nível individual.

Terceiro passo: São adicionadas as variáveis do segundo nível. É determinado se as variáveis do nível 2 explicam diferenças entre os grupos.

Quarto passo: É verificado se há algum componente de variância de alguma variável explicativa (u_{1j}, \dots, u_{pj}) com valor significativo.

Quinto passo: São adicionadas interações entre as variáveis do grupo e indivíduo. Vale notar que se a interação entre duas variáveis for incluída no modelo, deve-se incluir as variáveis separadamente, mesmo que elas não sejam significativas.

Seguidos estes passos, é obtido o modelo final. Porém, para concluir a análise multinível, ainda é necessário um diagnóstico dos resíduos e também a interpretação das variáveis e dos parâmetros do modelo, que é aonde se obtém mais informações a respeito dos dados.

4 METODOLOGIA

Neste estudo são usados os dados da Aneb. Para medir o desempenho do aluno são aplicadas duas provas: uma de língua portuguesa e outra de matemática. Além das provas, cada aluno responde um questionário que coleta informações referentes à sua condição financeira, social e escolar. Também são aplicados questionários que coletam informações sobre características dos professores, diretores e escolas.

Dados escolares têm, naturalmente, uma estrutura ideal para a aplicação de uma análise multinível. Neste caso serão dois níveis, de forma que o nível 1 refere-se ao aluno, enquanto o nível 2 refere-se à escola. A população de interesse são os alunos do 3º ano do Ensino Médio aninhados nas escolas do Distrito Federal. As informações dos diretores, professores e escolas fazem parte do nível 2. Foi escolhida a Aneb mais recente, realizada em 2013. Para o estudo é necessário obter informações específicas para cada indivíduo, para isso são utilizados os microdados, que estão disponibilizados gratuitamente no site do INEP (<http://portal.inep.gov.br>).

A Aneb é composta por duas provas, uma de língua portuguesa e outra de matemática. Sendo assim, serão construídos dois modelos: o primeiro terá como variável resposta o desempenho do aluno na prova de língua portuguesa, enquanto o segundo terá como variável resposta o desempenho em matemática. Com isso, deseja-se conhecer quais as principais características diferenciadoras do desempenho dos alunos para cada uma das disciplinas em questão.

Estão entre as perguntas do questionário de características sociais questões como a posse de diferentes bens materiais em quantidade e também a escolaridade dos pais, porém, analisar cada um desses fatores individualmente pode não trazer muitas informações. Por isso, nesse estudo será aplicado o CCEB 2014 (Critério de Classificação Econômica Brasil 2014) que atribui uma classe (A1, A2, B1, B2, C1, C2, D, E) para cada aluno de acordo com a quantidade de bens que o aluno possui e escolaridade dos pais, sendo A1 a classe com maior poder aquisitivo, enquanto E a classe com menor poder aquisitivo. O CCEB mais recente foi criado em 2015, mas as informações exigidas por ele não são disponibilizadas pelos dados do estudo. O CCEB é um índice criado pela Associação Brasileira de Estudos Populacionais, através de estudos demográficos, e é simples de ser aplicado: é gerada uma pontuação para

um indivíduo de acordo com os bens que ele possui e grau de instrução dos seus pais. Na tabela 4.1 pode ser vista a pontuação dada para cada item.

Tabela 4.1 – Pontuação recebida pela posse de cada item no CCEB 2014.

	Quantidade de itens				
	0	1	2	3	4 ou +
Televisão em cores	0	1	2	3	4
Rádio	0	1	2	3	4
Banheiro	0	4	5	6	7
Automóvel	0	4	7	9	9
Empregada mensalista	0	3	4	4	4
Máquina de lavar	0	2	2	2	2
Videocassete e/ou DVD	0	2	2	2	2
Geladeira	0	4	4	4	4
Freezer (aparelho independente ou parte da geladeira duplex)	0	2	2	2	2

Fonte: ABEP, CCEB 2014

Além da posse de bens, é também considerado o grau de instrução dos pais. A pontuação referente ao grau de instrução dos pais pode ser visto na tabela 4.2.

Tabela 4.2 – Pontuação recebida pelo grau de instrução dos pais no CCEB 2014.

Nomenclatura Antiga	Nomenclatura Atual	
Analfabeto/Primário incompleto	Analfabeto/Fundamental 1 Incompleto	0
Primário completo/Ginasial incompleto	Fundamental 1 Completo/Fundamental 2 Incompleto	1
Ginasial completo/Colegial incompleto	Fundamental 2 completo/Médio incompleto	2
Colegial completo/Superior incompleto	Médio Completo/Superior Incompleto	4
Superior completo	Superior Completo	8

Fonte: ABEP, CCEB 2014

A partir da pontuação obtida pelo aluno, uma categoria é atribuída a ele. Os intervalos de cada categoria pode ser visto na tabela 4.3.

Tabela 4.3 – Intervalos das categorias do CCEB 2014.

Classe	Pontos
A1	42-46
A2	35-41
B1	29-34
B2	23-28
C1	18-22
C2	14-17
D	8-13
E	0- 7

Fonte: ABEP, CCEB 2014

Todas as variáveis explicativas do banco de dados são categóricas. Uma variável com muitas classes pode trazer problemas de convergência para o modelo. Uma solução é agrupar classes similares, como foi feito para algumas variáveis deste estudo. Outro problema é o fato de haver vários professores em cada escola, enquanto no modelo só é possível incluir um único valor. Por essas razões, a partir das variáveis originais, foram criadas outras que pudessem ser incluídas no modelo. Na tabela 4.4 é possível ver as adaptações realizadas para as variáveis dos professores.

Tabela 4.4 – Variáveis construídas para os professores.

Variável	Descrição
Carga Horária	Proporção de professores que trabalham menos de 40h semanais
Dever de Casa	Proporção de professores que passa dever de casa semanalmente/diariamente
Educação	Proporção de professores com licenciatura na disciplina que eles atuam
Idade	Proporção de professores de 30 a 49 anos na escola
Raça/cor	Proporção de professores Negros (Pretos e Pardos) na escola
Renda	Proporção de professores que ganham 7 ou mais salários mínimos
Sexo	Proporção de professores do sexo masculino na escola
Situação Trabalhista	Proporção de professores Estatutários e CLT

Para as variáveis dos alunos, em relação à raça/cor, a categoria *Indígenas* não foi incluída no modelo por essa categoria representar menos de 2% da amostra, juntaram-se as categorias *Pretos* e *Pardos* criando a categoria *Negros*. Para o CCEB 2014 foi criado o grupo superior (A1, A2 e B1) e o grupo inferior (B2, C1, C2, D, E). E, referente à frequência de leitura, juntou-se a categoria *Sempre ou quase sempre* com a categoria *De vez em quando*.

Tabela 4.5 – Variáveis construídas para os alunos.

Variável	Descrição
CCEB 2014	Classe superior (A1, A2 e B1) e classe inferior (B2, C1, D, E)
Leitura de Livros em Geral	Frequentemente (<i>Sempre ou quase sempre</i> e <i>De vez em quando</i>) e Raramente (<i>Nunca ou quase nunca</i>)
Raça/cor*	Amarelos, Brancos e Negros(<i>Pretos e Pardos</i>)

* A categoria *Indígenas* não foi incluída no modelo

Para a escola, em algumas questões, além de ser perguntado se há uma estrutura na escola, pede-se para especificar a qualidade dessa estrutura. Por exemplo, em relação ao telhado são dadas as seguintes categorias: a) *Bom*; b) *Regular*; c) *Ruim* e d) *Inexistente*.

Porém, no modelo será desconsiderada a qualidade, de forma a ser avaliada apenas a existência ou não da estrutura de interesse, juntanto as categorias *Bom*, *Regular* e *Ruim* na categoria única *Existente*.

Na preparação e análise dos dados foi utilizado o programa estatístico R. O R é formado por uma estrutura base, e a partir dela é possível vincular pacotes que auxiliam em diversas aplicações. Para realizar a análise multinível foi usado o pacote *lme4*.

5 RESULTADOS

5.1 Apresentação dos Dados

Todas as séries em que um estudante passa são fundamentais, porém, o 3º ano do Ensino Médio tem um papel particularmente importante por ser o ano de conclusão da educação obrigatória, além de preceder o ingresso à Educação Superior para aqueles que desejam seguir a vida acadêmica. Diante disso, este trabalho busca conhecer as características que influenciam o desempenho de um estudante através da aplicação de um modelo multinível aos dados da Aneb do SAEB de 2013.

Primeiramente, deseja-se conhecer o perfil das escolas, diretores, professores e alunos em questão.

Tabela 5.1 – Distribuição das escolas conforme tipo de dependência administrativa e localização.

Características	Frequências	Porcentagens
Dependência Administrativa		
Estadual	47	58,0%
Privada	32	39,5%
Federal	2	2,5%
Localização		
Urbana	78	96,3%
Rural	3	3,7%
Total	81	100,0%

Fonte: INEP, SAEB, Aneb 2013

Foram observadas 81 escolas com turmas do 3º ano do Ensino Médio, sendo 49 públicas (47 estaduais e 2 públicas), representando 60,5% da amostra e 96,3% está localizadas em áreas urbanas (tabela 5.1).

Sobre o estado de conservação dos seguintes elementos das escolas: telhado, paredes, piso, entrada do prédio, pátio, corredores salas de aula, portas, janelas banheiros, cozinha, instalações hidráulicas e instalações elétricas; notou-se (tabela 5.2) que, para todos esses elementos, pelo menos 60% das escolas afirmaram *Bom* estado de conservação. Entretanto, 5 escolas afirmaram não ter cozinha e 1 afirmou não ter instalações hidráulicas.

Tabela 5.2 – Distribuição das escolas conforme existência e estado de conservação de estruturas.

Estrutura	Bom	Regular	Ruim	Inexistente	Sem Resposta	Total
Telhado	66 (81,5%)	9 (11,1%)	4 (4,9%)	-	2 (2,5%)	81 (100,0%)
Parede	67 (82,7%)	12 (14,8%)	-	-	2 (2,5%)	81 (100,0%)
Piso	62 (76,5%)	11 (13,6%)	6 (7,4%)	-	2 (2,5%)	81 (100,0%)
Entrada do Prédio	68 (83,9%)	10 (12,3%)	1 (1,2%)	-	2 (2,5%)	81 (100,0%)
Pátio	62 (76,5%)	14 (17,3%)	1 (1,2%)	-	4 (4,9%)	81 (100,0%)
Corredores	69 (85,2%)	9 (11,1%)	1 (1,2%)	-	2 (2,5%)	81 (100,0%)
Salas de Aula	63 (77,8%)	12 (14,8%)	4 (4,9%)	-	2 (2,5%)	81 (100,0%)
Portas	62 (76,5%)	17 (21,0%)	-	-	2 (2,5%)	81 (100,0%)
Janelas	57 (70,4%)	17 (21,0%)	5 (6,2%)	-	2 (2,5%)	81 (100,0%)
Banheiros	54 (66,7%)	23 (28,3%)	2 (2,5%)	-	2 (2,5%)	81 (100,0%)
Cozinha	54 (66,7%)	15 (18,5%)	2 (2,5%)	5 (6,2%)	5 (6,2%)	81 (100,0%)
Instalações Hidráulicas	52 (64,2%)	19 (23,5%)	7 (8,6%)	1 (1,2%)	2 (2,5%)	81 (100,0%)
Instalações Elétricas	50 (61,7%)	16 (19,8%)	8 (9,9%)	2 (2,5%)	5 (6,2%)	81 (100,0%)

Fonte: INEP, SAEB, Aneb 2013

- Dado numérico igual a zero não resultante de arredondamento

Em relação a alguns ambientes específicos, constatou-se (tabela 5.3) que apenas 5 escolas não têm biblioteca, enquanto 37% das escolas afirmaram não ter auditório, 58% afirmaram não ter sala de música e 43,2% afirmaram não ter sala de artes.

Tabela 5.3 – Distribuição das escolas conforme existência e estado de conservação de ambientes.

Ambientes	Bom	Regular	Ruim	Inexistente	Sem Resposta	Total
Biblioteca	49 (60,5%)	22 (27,2%)	3 (3,7%)	5 (6,2%)	2 (2,5%)	81 (100,0%)
Auditório	40 (49,4%)	6 (7,4%)	3 (3,7%)	30 (37,0%)	2 (2,5%)	81 (100,0%)
Sala de Música	25 (30,9%)	5 (6,2%)	1 (1,2%)	47 (58,0%)	3 (3,7%)	81 (100,0%)
Sala de Artes	33 (40,7%)	10 (12,3%)	-	35 (43,2%)	3 (3,7%)	81 (100,0%)

Fonte: INEP, SAEB, Aneb 2013

- Dado numérico igual a zero não resultante de arredondamento

A amostra em estudo continha 225 professores, que foram divididos em dois grupos de acordo com a sua disciplina (matemática ou língua portuguesa). Aqueles que afirmaram lecionar mais de uma disciplina fazem parte dos dois grupos, porém, apenas 1,3% dos professores responderam lecionar ambas as disciplinas.

Para a disciplina **língua portuguesa** há uma predominância de professores do sexo feminino (tabela 5.4), com 71% dos respondentes. O intervalo de idade mais frequente, com 65%, é de 30 a 49 anos. Ao ver a tabela 5.5, nota-se que 81% dos respondentes se

Tabela 5.4 – Distribuição dos professores conforme sexo e idade.

Características	Professores de MT	Professores de LP
Sexo		
Masculino	89 (69,0%)	35 (28,2%)
Feminino	39 (30,2%)	88 (71,0%)
Sem Resposta	1 (0,8%)	1 (0,8%)
Idade		
Até 24 anos	3 (2,3%)	3 (2,4%)
25 a 29	12 (9,3%)	19 (15,3%)
30 a 39	40 (31,0%)	45 (36,3%)
40 a 49	50 (38,8%)	36 (29,0%)
50 a 54	10 (7,8%)	12 (9,7%)
55 ou mais	11 (8,5%)	5 (4,0%)
Sem Resposta	3 (2,3%)	4 (3,2%)
Total	129 (100,0%)	124 (100,0%)

Fonte: INEP, SAEB, Aneb 2013

ou *Branco(a)* ou *Pardo(a)*, sendo as outras categorias *Preto(a)*, *Amarelo(a)* e *Indígena*, menos frequentes. Quanto à educação (tabela 5.6), 79% indicaram possuir Licenciatura em Letras. Referente às característica de trabalho (tabela 5.7), a carga horária mais recorrente, 50,8%, é de 40 horas semanais, enquanto 33,9% afirmaram ter carga horária semanal entre 20 e 39 horas. O intervalo de salário mais frequente, 36,3%, é entre R\$4.747 e R\$6.780, enquanto 29,8% apontaram receber R\$6.781 ou mais. As situações de trabalho que mais ocorrem são *Estatutário* e *CLT*, com 42,7% e 33,9% respectivamente. Contratos temporários foram indicados por 19,4% dos respondentes. A maioria, 47,6%, afirmou passar dever de casa semanalmente.

Tabela 5.5 – Distribuição dos diretores, professores e alunos conforme raça/cor.

Raça/Cor	Diretores	Professores de MT	Professores de LP	Estudantes
Branco(a)	37 (46,3%)	57 (44,2%)	55 (44,4%)	1076 (31,6%)
Pardo(a)	33 (41,3%)	57 (44,2%)	45 (36,3%)	1545 (45,3%)
Preto(a)	5 (6,3%)	4 (3,1%)	6 (4,8%)	422 (12,4%)
Amarelo(a)	2 (2,5%)	1 (0,8%)	6 (4,8%)	175 (5,1%)
Indígena	1 (1,3%)	-	1 (0,8%)	41 (1,2%)
Não quis declarar	-	2 (1,6%)	4 (3,2%)	-
Não sabe	1 (1,3%)	6 (4,7%)	4 (3,2%)	92 (2,7%)
Sem Resposta	1 (1,3%)	2 (1,6%)	3 (2,4%)	58 (1,7%)
Total	80 (100,0%)	129 (100,0%)	124 (100,0%)	3409 (100,0%)

Fonte: INEP, SAEB, Aneb 2013

- Dado igual a zero não resultante de arredondamento

Tabela 5.6 – Distribuição dos professores e diretores conforme educação.

Educação	Diretores	Professores de MT	Professores de LP
Ensino Médio	1 (5,0%)	-	1 (0,8%)
Pedagogia	19 (23,8%)	-	-
Curso Normal Superior	2 (2,5%)	1 (0,5%)	-
Licenciatura em Matemática	7 (8,8%)	97 (75,2%)	2 (1,6%)
Licenciatura em Letras	7 (8,8%)	2 (1,6%)	98 (76,0%)
Outras Licenciaturas	30 (37,5%)	23 (17,8%)	17 (13,2%)
Outras Áreas	10 (12,5%)	4 (3,1%)	1 (0,8%)
Sem Resposta	4 (5,0%)	2 (1,6%)	5 (3,9%)
Total	80 (100,0%)	129 (100,0%)	124 (100,0%)

Fonte: INEP, SAEB, Aneb 2013

- Dado numérico igual a zero não resultante de arredondamento

Agora, para os professores de **matemática** há uma inversão na predominância de sexo (tabela 5.4): 69% afirmaram ser do sexo masculino, enquanto 30,2% do sexo feminino. O intervalo de idade mais frequente, com 69,8%, também é de 30 a 49 anos. Ao observar a tabela 5.5, nota-se que 88,4% dos respondentes se considera ou *Branco(a)* ou *Pardo(a)*. Quanto à educação (tabela 5.6), 75,2% indicaram possuir Licenciatura em Matemática. Para as características de trabalho (tabela 5.7), observou-se que a carga horária mais frequente, 52,7%, é de 40 horas semanais, enquanto 27,9% afirmaram ter carga horária semanal entre 20 e 39 horas. 40,3% afirmaram receber um salário de R\$6.781 ou mais, enquanto 29,5% afirmaram receber entre R\$4.747 e R\$6.780. As situações de trabalho que mais ocorrem

são *Estatutário* e *CLT*, com 51,9% e 31,8% respectivamente. Contratos temporários foram indicados por 10% dos respondentes, pouco mais da metade do que foi observado em língua portuguesa. A maioria, 52,7%, afirmou passar dever de casa semanalmente.

Tabela 5.7 – Distribuição dos professores conforme características profissionais.

Características	Professores de MT	Professores de LP
Carga Horária		
Menos de 20 horas	6 (4,7 %)	5 (4,0%)
20 a 39 horas	36 (27,9%)	42 (33,9%)
40 horas	68 (52,7%)	63 (50,8%)
Mais de 40 horas	16 (12,4%)	9 (7,2%)
Sem Resposta	3 (2,3%)	5 (4,0%)
Salário		
Até R\$3390,00	12 (14,0%)	19 (15,3%)
R\$3391,00 a R\$4746,00	14 (14,0%)	20 (16,1%)
R\$4747,00 a R\$6780,00	38 (29,5%)	45 (36,3%)
R\$6781,00 ou mais	52 (40,3%)	37 (29,8%)
Sem Resposta	3 (2,3%)	3 (2,4%)
Situação Trabalhista		
Estatutário	67 (51,9 %)	53 (42,7%)
CLT	41 (31,8%)	42 (33,9%)
Contrato Temporário	13 (10,0%)	24 (19,4%)
Outros	4 (3,1%)	-
Sem Resposta	4 (3,1%)	5 (4,0%)
Costuma Passar Dever de Casa		
Diariamente	44 (34,1%)	36 (29,0%)
Semanalmente	68 (52,7%)	59 (47,6%)
Raramente	11 (8,5%)	18 (14,5%)
Nunca	1 (0,7%)	5 (4,0%)
Sem Resposta	5 (3,9%)	6 (4,8%)
Total	129 (100,0%)	124 (100,0%)

Fonte: INEP, SAEB, Aneb 2013

- Dado numérico igual a zero não resultante de arredondamento

A amostra era composta por 80 diretores e na tabela 5.8 observa-se que 60% dos respondentes afirmaram ser do gênero masculino, enquanto 40% do sexo feminino. O intervalo de idade mais recorrente, 37,5%, é entre 40 e 49 anos. Assim como os professores, a maioria dos diretores se considera ou *Branco(a)* ou *Pardo(a)*, assumindo 87,5% a soma das duas categorias (tabela 5.5). Referente à educação (tabela 5.6), 37,5% dos diretores indicaram possuir *Outras Licenciaturas* (diferente de matemática e língua portuguesa). A maioria, 61,3%, indicou receber um salário de R\$6.781 ou mais. Praticamente metade, 53,8%,

indicou uma carga horário de mais de 40 horas, enquanto 40% indicaram uma carga horária de 40 horas semanais. Em relação ao tempo de experiência, há uma concentração, 45%, no intervalo de 1 a 5 anos.

Tabela 5.8 – Distribuição dos diretores conforme características.

Características	Frequências	Porcentagens
Sexo		
Masculino	48	60,0%
Feminino	32	40,0%
Faixa Etária		
30 a 39 anos	19	23,8%
40 a 49 anos	30	37,5%
50 a 54 anos	20	25,0%
55 ou mais	10	12,5%
Sem Resposta	1	1,3%
Salário		
Até R\$3390,00	6	7,5%
R\$3391,00 a R\$4746,00	4	5,0%
R\$4747,00 a R\$6780,00	18	22,5%
Mais de R\$6781,00	49	61,3%
Carga Horária		
20 a 39 horas	4	5,0%
40 horas	32	40,0%
Mais de 40 horas	43	53,8%
Sem Resposta	1	1,3%
Tempo de Experiência		
Menos de 6 anos	24	52,5%
6 a 10 anos	15	18,8%
11 a 15 anos	9	11,3%
16 anos ou mais	12	15,0%
Sem Resposta	2	2,5%
Total	80	100,0%

Fonte: INEP, SAEB, Aneb 2013

O banco de dados dos alunos tem algumas variáveis que indicam informações sobre a realização ou não da prova e do questionário. Para definir a amostra, foram selecionados aqueles que indicaram ter respondido a prova, ter pelo menos três itens da prova respondidos e ter respondido o questionário. A amostra consiste de 3409 alunos. Pela tabela 5.9, vê-se que 54,9% dos respondentes afirmaram ser do sexo feminino, enquanto 43,8% do masculino. Referente à idade, 81,8% dos alunos afirmaram ter 17 ou 18 anos. Com relação

à condição social, aplicando o CCEB 2014, notou-se que as três classes mais frequentes, em ordem decrescente, são B2, B1 e C1, assumindo 68,2% a soma das três categorias. Reprovar uma série pode gerar um impacto negativo na vida do aluno, viu-se que 27% dos respondentes já reprovaram, sendo que 5,6% já reprovaram mais de uma vez.

Ao separar os alunos por tipo de escolas (públicas e privadas - Tabela 5.10), notou-se uma diferença substancial na distribuição das classes dos alunos por classe econômica. Os alunos das escolas públicas estão mais concentrados nas classes B2 e C1, 31,1% e 25,7% respectivamente. Enquanto os alunos das escolas privadas estão mais concentrados nas classes superiores A2 e B1, 33,1% e 34% respectivamente.

Tabela 5.9 – Distribuição dos alunos conforme características.

Características	Frequências	Porcentagens
Sexo		
Masculino	1493	43,8%
Feminino	1871	54,9%
Sem Resposta	45	1,3%
Idade		
Até 16 anos	117	3,4%
17 anos	1478	43,4%
18 anos	1311	38,5%
19 anos	348	10,2%
20 anos ou mais	123	3,6%
Sem Resposta	32	0,9%
CCEB 2014		
A1	64	1,9%
A2	471	13,8%
B1	742	21,8%
B2	933	27,4%
C1	651	19,1%
C2	390	11,4%
D	138	4,0%
Sem Resposta	20	0,5%
Já Reprovou		
Não	2391	70,1%
Sim, uma vez	731	21,4%
Sim, mais de uma vez	191	5,6%
Sem Resposta	96	2,8%
Total	3409	100,0%

Fonte: INEP, SAEB, Aneb 2013

Tabela 5.10 – Distribuição dos alunos conforme classe econômica para escolas públicas e privadas.

Classe	Pública	Privada
A1	4 (0,2%)	60 (5,5%)
A2	108 (4,6%)	363 (33,1%)
B1	369 (16,0%)	373 (34,0%)
B2	718 (31,1%)	215 (19,6%)
C1	595 (25,7%)	56 (5,1%)
C2	371 (16,1%)	19 (1,7%)
D	134 (5,8%)	4 (0,4%)
Sem Resposta	12 (0,5%)	8 (0,7%)
Total	2311 (100,0%)	1098 (100,0%)

Fonte: INEP, SAEB, Aneb 2013

A leitura tem um papel fundamental na formação de uma pessoa, por isso, foi perguntado a frequência de leitura de diferentes tipos de texto (tabela 5.11): revistas informativas em geral, gibis, livros em geral e notícias e/ou textos na internet. A maioria, 61%, afirmou ler revistas informativas *De vez em quando*. Para gibis, poucas pessoas indicaram o hábito de leitura: 49,5% afirmaram *Nunca ou quase nunca*. Apenas 12,9% indicaram ler *Nunca ou quase nunca* livros em geral, com o restante dos que responderam, 82,9%, divididos de forma equilibrada entre *Sempre ou quase sempre* e *De vez em quando*. Notícias e/ou textos na internet foi a categoria de leitura favorita, visto que 69% dos respondentes afirmaram ler *Sempre ou quase sempre*.

Tabela 5.11 – Distribuição dos alunos conforme hábitos de leitura.

Tipo de Leitura	Sempre ou Quase Sempre	De Vez em Quando	Raramente	Sem Resposta
Revistas Informativas	658 (19,3%)	2079 (61,0%)	516 (15,1%)	156 (4,6%)
Gibis	337 (9,9%)	1199 (35,2%)	1688 (49,5%)	185 (5,4%)
Livros em Geral	1302 (38,2%)	1525 (44,7%)	440 (12,9%)	142 (4,2%)
Notícia/Textos na Internet	2354 (69,0%)	847 (24,8%)	69 (2,0%)	139 (4,0%)

Fonte: INEP, SAEB, Aneb 2013

A forma como o aluno gasta o seu tempo pode influenciar o seu desempenho escolar. Com relação ao tempo gasto (tabela 5.12), em dia de aula, assistindo TV, navegando na internet ou jogando jogos eletrônicos foi visto que as categorias de tempo mais recorrentes são *Mais de três horas* e *Entre 1 e 2 horas*, com 31,9% e 26,3%, respectivamente. Somente 25,7% afirmaram trabalhar, recebendo ou não salário.

Tabela 5.12 – Distribuição dos alunos conforme hábitos fora da escola.

Característica	Frequência	Porcentagem
Horas diárias usando TV/Internet/Jogos Eletrônicos		
Mais de 3 horas	1088	31,9%
Mais 2, até 3 horas	800	23,5%
Entre 1 e 2 horas	895	26,3%
Menos de 1 hora	462	13,6%
Não usa TV/Internet/Jogos Eletrônicos	86	2,5%
Sem Resposta	78	2,3%
Trabalha		
Sim	875	25,7%
Não	2453	72,0%
Sem Resposta	81	2,4%
Total	3409	100,0%

Fonte: INEP, SAEB, Aneb 2013

É essencial ter afinidade com uma matéria para estudá-la. Pela tabela 5.13, vê-se que 62% dos respondentes afirmaram gostar de estudar língua portuguesa, enquanto 50,5% afirmaram gostar de estudar matemática. Em relação à frequência em que são feitos os deveres de casa, para língua portuguesa, as categorias mais frequentes foram *Sempre ou quase sempre* e *De vez em quando*, com 44,2% e 39,9%, respectivamente. Para matemática, ocorreu o mesmo padrão, com 41,9% e 41,6%, respectivamente.

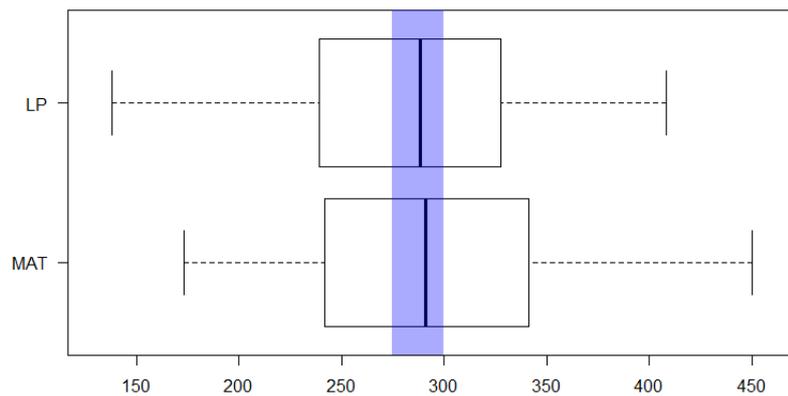
Tabela 5.13 – Distribuição dos alunos conforme hábitos de estudo.

Característica	Matemática	Língua Portuguesa
Gosta de Estudar		
Sim	1722 (50,5%)	2115 (62,0%)
Não	1565 (45,9%)	1204 (35,3%)
Sem Resposta	122 (3,6%)	90 (2,6%)
Faz Dever de Casa		
Sempre ou Quase Sempre	1427 (41,9%)	1506 (44,2%)
De Vez em Quando	1418 (41,6%)	1360 (39,9%)
Nunca	396 (11,6%)	344 (10,1%)
Não tem Dever de Casa	70 (2,1%)	114 (3,3%)
Sem Resposta	98 (2,9%)	85 (2,5%)
Total	3409 (100,0%)	3409 (100,0%)

Fonte: INEP, SAEB, Aneb 2013

Por fim, é importante conhecer o comportamento das proficiências que serão as variáveis respostas dos modelos. Para língua portuguesa, os valores estão concentrados em torno da média de 282,9 com o desvio padrão de 57,7, gerando um coeficiente de variação de 20,4%, indicando uma baixa homogeneidade dos dados. Enquanto para matemática, os valores estão mais espalhados, com uma média de 292,5 e um desvio padrão de 62,6, gerando um coeficiente de variação de 21,4%, também indicando uma baixa homogeneidade. A pontuação nas provas tem um significado que é explicado a partir de intervalos: pontuações dentro de um certo intervalo exigem habilidades, especificadas pelo INEP, na disciplina avaliada (ver Anexo). Essa relação de intervalos pode ser vista no anexo. Para língua portuguesa são 8 níveis, enquanto para matemática 10. Nota-se que na amostra ambas as matérias têm a média dentro do intervalo do nível 3 (275 a 300 pontos), destacado pela faixa azul na figura 5.1.

Figura 5.1 – Diagramas de caixa da proficiência dos alunos em língua portuguesa e matemática.

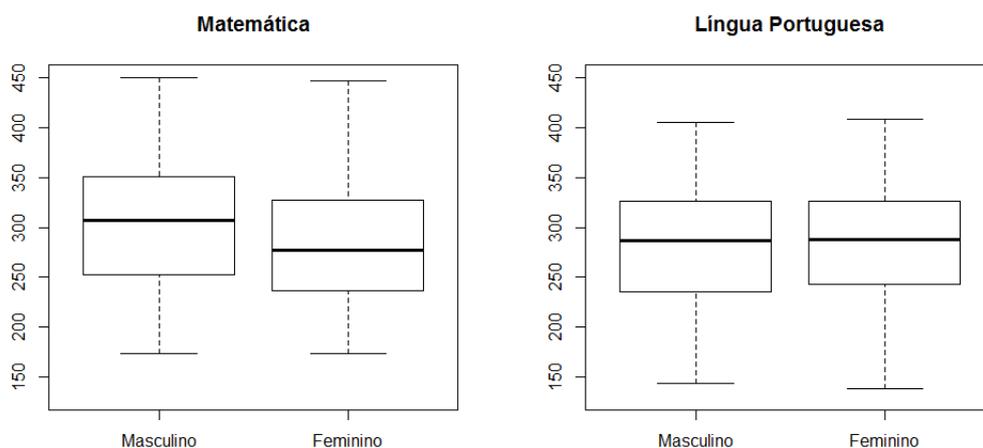


5.2 Análise Bivariada

Ao se realizar uma análise bivariada é possível identificar qual o tipo de relação das variáveis com o desempenho do aluno, obtendo-se assim variáveis candidatas a serem incluídas no modelo. Nesta seção algumas variáveis serão selecionadas para a análise.

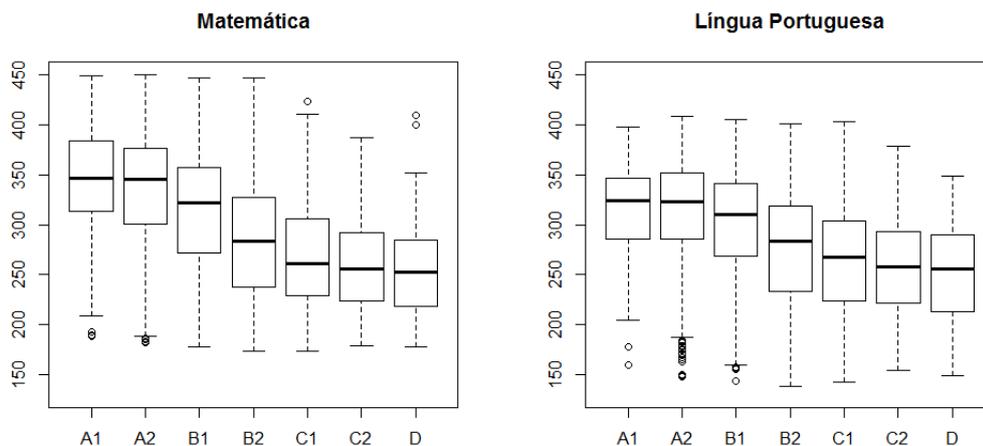
O desempenho em matemática parece ser diferente segundo o sexo, enquanto para língua portuguesa já não há uma diferença tão grande, o que pode ser visto na figura 5.2.

Figura 5.2 – Diagramas de caixa do desempenho do aluno conforme **sexo**.



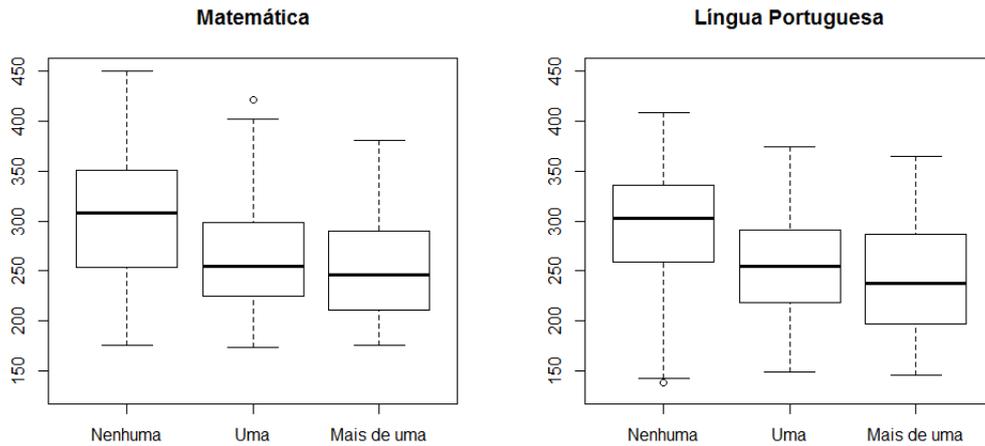
Infelizmente a equidade do ensino não é uma realidade no Brasil, alunos têm diferentes desempenhos dependendo do seu poder aquisitivo. Utilizando as classes do CCEB 2014, esse contraste pode ser observado na figura 5.3. Um fato interessante, é que essa diferença parece ser mais expressiva para matemática.

Figura 5.3 – Diagramas de caixa do desempenho do aluno conforme **poder aquisitivo**.



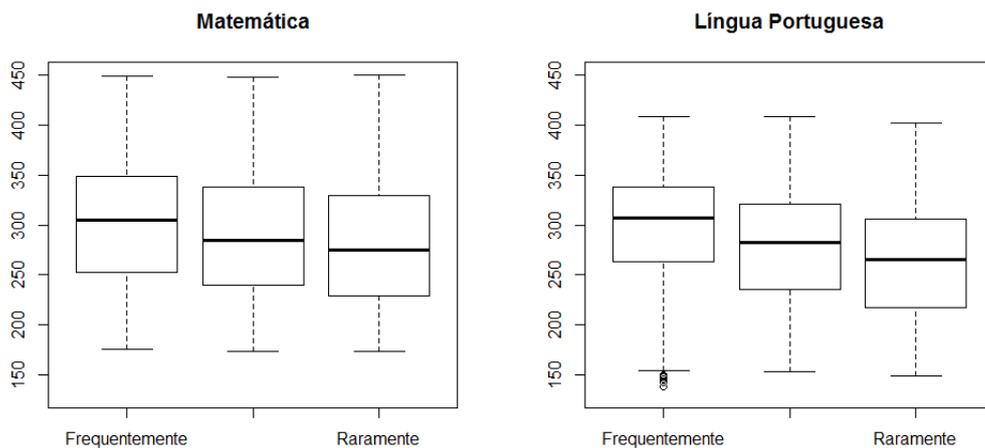
Duas variáveis observadas foram a idade e o número de reprovações. Há uma correlação entre a idade e o número de reprovações, fazendo com que usar as duas variáveis não acrescente muita informação em relação a usar apenas uma variável. Desta forma, foi escolhido o número de reprovações, que tem menos categorias, evitando possíveis problemas de convergência no modelo. Ao observar a figura 5.4, nota-se que aqueles que já reprovaram costumam ter um desempenho inferior aos que nunca reprovaram.

Figura 5.4 – Diagramas de caixa do desempenho do aluno conforme **número de reprovações**.



Para representar os hábitos de leitura, foi escolhido o tema *Livros em Geral*, que, como mostra a figura 5.5, sugere uma relação com o desempenho do aluno. Pode-se ver que o aluno mais habituado a ler livros tem, em média, um resultado melhor tanto em matemática quanto em língua portuguesa.

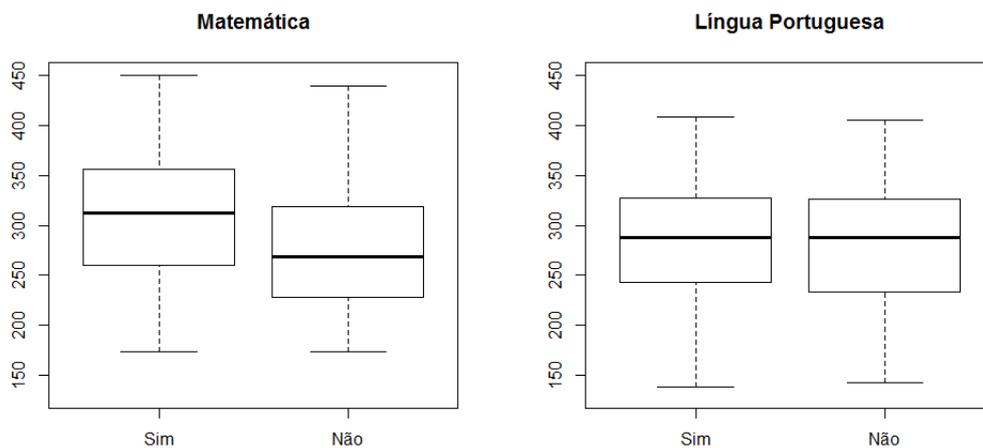
Figura 5.5 – Diagramas de caixa do desempenho do aluno conforme **hábito de ler livros em geral**.



Em relação à afinidade com a matéria, ao observar a figura 5.6, é possível

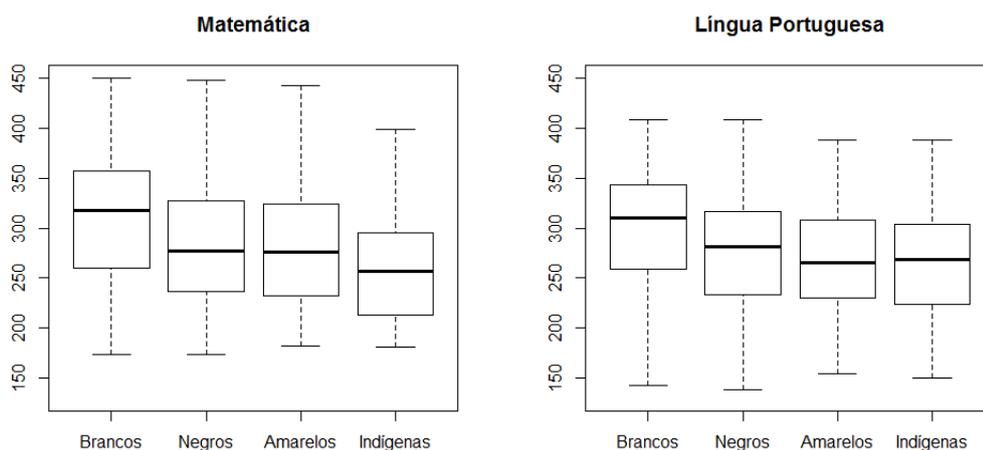
notar que os alunos que afirmaram gostar de estudar matemática tiveram um desempenho médio melhor. Porém, para língua portuguesa, não foram identificadas grandes diferenças no desempenho.

Figura 5.6 – Diagramas de caixa do desempenho do aluno que gosta ou não de **estudar a matéria**.



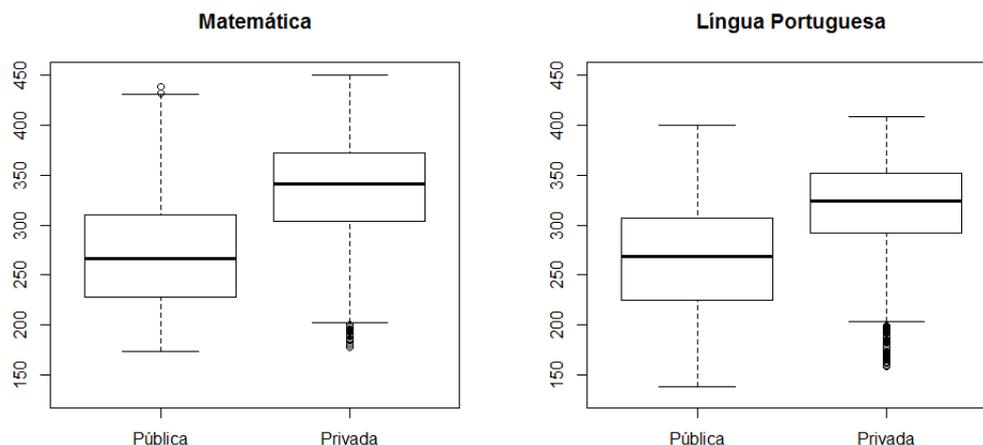
Foi perguntada a raça/cor que o aluno se considerava, e, ao observar a figura 5.7, notou-se que aqueles pertencentes à categoria *Branços* tiveram, em média, um desempenho maior que os pertencentes a outras categorias.

Figura 5.7 – Diagramas de caixa do desempenho do aluno conforme **raça/cor**.



Para o nível escola, uma característica estudada é se ela é pública ou privada, a sua importância para o modelo se torna evidente pela figura 5.8, que mostra uma grande diferença no desempenho médio de acordo com o tipo de administração. Esse comportamento ocorre tanto para matemática quanto para língua portuguesa.

Figura 5.8 – Diagramas de caixa do desempenho do aluno conforme **tipo de administração da escola**.



Além da análise gráfica, foram feitos testes para verificar se a diferença entre as categorias era significativa. Para as variáveis com duas categorias foi realizado o teste t de comparação de médias, enquanto para as variáveis com mais de duas categorias foi feita um análise de variância. Porém, todos os testes mostraram uma diferença significativa entre as categorias.

5.3 Aplicação do Modelo Multinível

Feita uma análise exploratória dos dados é possível iniciar a construção do modelo. Primeiramente é construído o modelo com apenas o intercepto, desta forma é obtida a correlação intraclasse, que varia de zero a um. Valores altos da correlação sugerem diferenças no comportamento da variável resposta conforme o grupo em que ela pertence.

Nas tabelas, os erros padrões dos parâmetros da parte fixa estão em parênteses, ao lado do coeficiente. A tabela 5.14 mostra os valores necessários para se calcular a correlação intraclasse utilizando a fórmula $\sigma_{u0}^2/(\sigma_e^2 + \sigma_{u0}^2)$. Para língua portuguesa a correlação intraclasse é de 0,33, ou seja, 33% da variação do desempenho acontece no nível da escola. Para matemática a correlação intraclasse é ainda maior: 0,45. Esses valores são fortes evidências de que há uma relação entre o aluno e a sua escola para ambas as matérias, tornando adequada a aplicação de um modelo multinível.

Tabela 5.14 – Modelos somente com o intercepto para matemática e língua portuguesa.

Coefficientes	Matemática	Língua Portuguesa
Parte fixa		
Intercepto	292,0 (4,6)	282,5 (3,7)
Parte aleatória		
σ_e^2	2130,0	2212,0
σ_{u0}^2	1748,0	1085,0
Deviance	36085,4	36172,7

Ao considerar uma variável categórica com p classes, o modelo gera $p-1$ coeficientes. O modelo sem coeficiente, para esta variável, representa a sua classe de referência. Por exemplo, para sexo é necessário apenas um coeficiente multiplicado por uma variável que indica se o indivíduo é mulher, de forma que quando o indivíduo for homem o coeficiente ligado às mulheres será nulo. Abaixo estão listadas as referências para as variáveis do nível do aluno:

- Sexo: Masculino
- CCEB 2014: Grupo superior
- Raça/cor: Branca
- Leitura: Sempre ou quase sempre
- Reprovação: Nenhuma vez
- Gosta de estudar: Sim

Os modelos com as variáveis do nível do aluno podem ser vistos na tabela 5.15. *Fazer o dever de casa e Trabalhar fora de casa* não foram significativos. Para língua portuguesa, o *Sexo*, ao incluir outras variáveis, se mostrou não significativo, ou seja, não foi rejeitada a hipótese de que o coeficiente associado ao gênero é igual a zero, portanto ele foi retirado do modelo. Ao fazer o teste de comparação de modelos, usando a deviance, conclui-se que houve melhora significativa (menor que 5%) ao incluir as variáveis do nível do aluno para os dois modelos.

Tabela 5.15 – Modelos com as variáveis do nível do aluno, com coeficientes de efeito fixo para as variáveis, para matemática e língua portuguesa.

Coeficientes	Matemática	Língua Portuguesa
Parte fixa		
Intercepto	330,4 (4,5)	307,1 (3,7)
Sexo: Feminino	-14,7 (1,7)	-
Reprovou: Uma vez	-15,5 (2,0)	-20,7 (2,1)
Reprovou: Mais de uma vez	-19,3 (3,8)	-29,1 (3,9)
CCEB 2014: Grupo inferior	-5,0 (2,0)	-5,4 (2,0)
Raça/cor: Negra	-4,7 (1,8)	-3,7 (1,9)
Raça/cor: Amarela	-5,8 (3,7)	-15,0 (3,9)
Leitura: Pouca frequência	-8,7 (1,7)	-13,2 (1,7)
Gosta de estudar: Não	-27,1 (1,6)	-7,8 (1,8)
Parte aleatória		
σ_e^2	1786,0	2007,9
σ_{u0}^2	1217,0	739,1
Deviance	31011,6	31805,2

Para matemática, o maior coeficiente, em módulo, foi para a categoria *Gosta de estudar: Não*. A sua interpretação é de que um aluno que não gosta de estudar matemática tem, em média, uma proficiência negativa em 27,1 pontos em relação a quem gosta. Enquanto para língua portuguesa, o maior coeficiente, em módulo, foi para *Reprovou: Mais de uma vez*. Sendo a sua interpretação: um aluno que já reprovou mais de uma vez tem, em média, um desempenho negativo em 29,1 pontos em relação a quem nunca reprovou.

A próxima etapa é incluir as variáveis significativas do nível da escola. As variáveis associadas aos professores foram todas não significativas. Outras variáveis foram consideradas também, como a existência de bibliotecas, de laboratórios de informática, laboratórios de ciências e salas de estudos, porém somente a *Dependência Administrativa* (pública ou privada) foi significativa.

Após incluir as variáveis do nível da escola no modelo, é verificado se há alguma componente aleatório significativo para os coeficientes das variáveis explicativas do nível do aluno. Não foi encontrado nenhum componente aleatório significativo nesta amostra. A significância de componentes aleatórios é verificada por meio do teste da diferença da *deviance* entre o modelo com e o modelo sem o componente. Para ilustrar a ocorrência de um componente aleatório, suponha que *Sexo: Feminino* tivesse um componente aleatório. A sua interpretação seria que uma aluna, mantida as outras variáveis constantes, teria diferentes valores médios de proficiência dependendo da sua escola.

Por fim, foi verificada a existência de interações entre as variáveis, porém nenhuma interação significativa foi encontrada. Portanto o modelo final é o ilustrado na tabela 5.16.

Tabela 5.16 – Modelos com as variáveis do nível do aluno e da escola, com os coeficientes de efeito fixo para as variáveis, para matemática e língua portuguesa.

Coefficientes	Matemática	Língua Portuguesa
Parte fixa		
Intercepto	308,8 (4,5)	290,3 (3,8)
Sexo: Feminino	-14,6 (1,7)	-
Reprovou: Uma vez	-15,1 (2,0)	-20,0 (2,1)
Reprovou: Mais de uma vez	-18,7 (3,8)	-28,2 (3,9)
CCEB 2014: Grupo inferior	-3,9 (2,0)	-3,8 (2,1)
Raça/cor: Negra	-4,4 (1,8)	-3,3 (1,9)
Raça/cor: Amarela	-5,8 (3,7)	-14,9 (3,9)
Leitura: Pouca frequência	-8,8 (1,7)	-13,3 (1,7)
Gosta de estudar: Não	-27,3 (1,6)	-7,8 (1,8)
Dependencia Administrativa: Privada	51,6 (5,9)	39,9 (4,9)
Parte aleatória		
σ_e^2	1785,1	2006,7
σ_{u0}^2	604,3	387,2
Deviance	30954,8	31754,4

Não foram detectadas fortes violações nos modelos em relação às suposições, como pode ser visto nas figuras 5.8 e 5.9.

Figura 5.9 – Gráfico de resíduos e valores esperados para os modelos finais

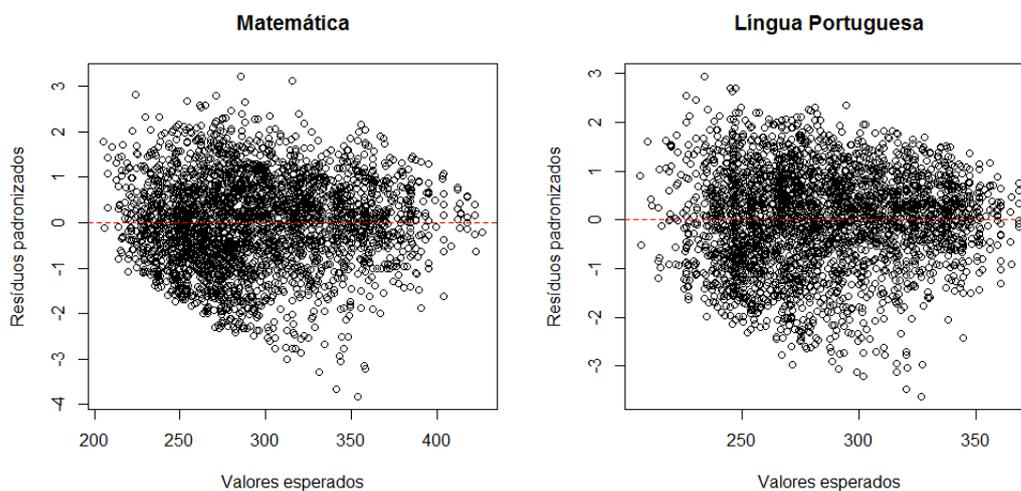
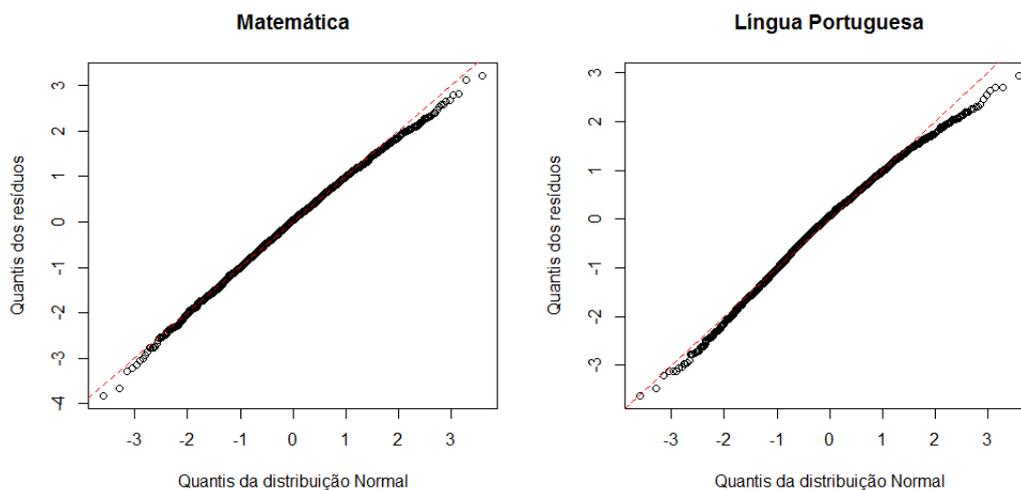


Figura 5.10 – Gráfico de quantis para os modelos finais



Desta forma, é possível escrever as equações finais do modelo. Para matemática a equação final é:

$$\begin{aligned}
 \text{proficiencia em matematica}_{ij} = & 308,8 - 14,6 \times \text{sexo} : \text{feminino}_{ij} \\
 & -15,1 \times \text{reprovou} : \text{uma vez}_{ij} - 18,7 \times \text{reprovou} : \text{mais de uma vez}_{ij} \\
 & -3,9 \times \text{CCEB2014} : \text{grupo inferior}_{ij} - 4,4 \times \text{raca/cor} : \text{negra}_{ij} \\
 & -5,8 \times \text{raca/cor} : \text{amarela}_{ij} - 8,8 \times \text{leitura} : \text{pouca frequencia}_{ij} \\
 & -27,3 \times \text{gosta de estudar} : \text{nao}_{ij} + 51,6 \times \text{dep. adm.} : \text{privada}_j + e_{ij} + u_{0j}
 \end{aligned} \tag{5.1}$$

Cada coeficiente tem uma interpretação, como para o gênero, que indica que uma aluna do gênero feminino tira, em média 14,6 pontos a menos que um aluno do gênero masculino. Aquele que reprovou uma vez tira, em média, 15,1 pontos a menos que quem nunca reprovou. Um resultado interessante foi a relevância da afinidade com a matéria: aqueles que responderam não gostar de estudar matemática tiram, em média, 27,3 pontos a menos que aqueles que gostam de estudar. O coeficiente para o tipo de dependência administrativa foi o maior: alunos de escolas privadas tiram em média 51,6 pontos a mais do que alunos de escolas públicas.

Para língua portuguesa, houve algumas diferenças que podem ser vistas na equação abaixo:

$$\begin{aligned}
 \text{proficiencia em lingua portuguesa}_{ij} = & 290,3 - 20 \times \text{reprovou} : \text{uma vez}_{ij} \\
 & -28,2 \times \text{reprovou} : \text{mais de uma vez}_{ij} - 3,8 \times \text{CCEB2014} : \text{grupo inferior}_{ij} \\
 & -3,3 \times \text{raca/cor} : \text{negra}_{ij} - 14,9 \times \text{raca/cor} : \text{amarela}_{ij} \\
 & -13,3 \times \text{leitura} : \text{pouca frequencia}_{ij} - 7,8 \times \text{gosta de estudar} : \text{nao} \\
 & +39,9 \times \text{dep. adm.} : \text{privada}_j + e_{ij} + u_{0j}
 \end{aligned} \tag{5.2}$$

A principal diferença entre os modelos é que o sexo não ajuda a explicar a variação da proficiência em língua portuguesa. A reprovação parece ter um impacto maior para língua portuguesa: um aluno que reprovou uma vez tira, em média, 20 pontos a menos que quem nunca reprovou, enquanto quem já reprovou mais de uma vez tira, em média, 28,2 pontos a menos que quem nunca reprovou. Diferente de matemática, para língua portuguesa

a afinidade com a matéria parece não ter tanta importância: quem respondeu não gostar de estudar língua portuguesa tirou, em média, 7,8 pontos a menos que quem respondeu gostar de estudar. O tipo de dependência administrativa continuou tendo o maior coeficiente: alunos de escolas privada tiram, em média, 39,9 pontos a mais que alunos de escolas públicas.

6 CONCLUSÃO

Modelos multiníveis têm o objetivo de analisar simultaneamente variáveis de diferentes níveis, e a importância dessa técnica é explicada pelo fato de que dados estruturados hierarquicamente acontecem frequentemente na prática.

A escala de proficiência formulada pelo INEP é segmentada em diferentes níveis para ambas as disciplinas. Em cada nível é listado um conjunto de habilidades que o aluno pode ter.

Para língua portuguesa, o desempenho médio foi de 282,5, dentro da faixa do nível 3 da escala de proficiência. Neste nível o estudante pode ser capaz de: a) localizar informação explícita em artigos de opinião; b) identificar a finalidade de relatórios científicos; c) reconhecer relações de sentido marcadas por conjunções, a relação de causa e consequência e a relação entre o pronome e seu referente em fragmentos de romances; d) reconhecer o tema de uma crônica; e) reconhecer variantes linguísticas em artigos; f) reconhecer o sentido e o efeito de sentido produzido pelo uso de recursos morfossintáticos em contos, artigos e crônicas; g) reconhecer opiniões divergentes sobre o mesmo tema em diferentes textos; h) inferir informação, o sentido e o efeito de sentido produzido por expressão em reportagens e tirinhas.

Enquanto para matemática, o desempenho médio foi de 292, também dentro da faixa do nível três da escala. Neste nível o estudante pode ser capaz de reconhecer: a) o valor máximo de uma função quadrática representada graficamente; b) em um gráfico o intervalo no qual a função assume o valor máximo. Também podem ser capazes de determinar: c) por meio de proporcionalidade o gráfico de setores que representa uma situação com dados fornecidos textualmente; d) o quarto valor em uma relação de proporcionalidade direta a partir de três valores fornecidos em uma situação do cotidiano; e) um valor reajustado de uma quantia a partir de seu valor inicial e do percentual de reajuste. Além disso, é provável que: f) resolvam problemas utilizando operações fundamentais com números naturais.

Para o Brasil as médias foram de 263,2 e 269,3 para língua portuguesa e matemática respectivamente, pertencendo, ambas as disciplinas, ao nível dois da escala de proficiência. Esses valores são indícios da desigualdade do ensino no país.

Neste estudo foi possível conhecer o perfil dos professores, diretores, alunos

e escolas do Distrito Federal. Foram construídos dois modelos multiníveis para explicar a variação da proficiência dos alunos em matemática e língua portuguesa.

Informações interessantes foram obtidas através dos modelos: o sexo ajuda a explicar a variação apenas para matemática, há um impacto negativo para alunos que já reprovaram, sendo ele maior para língua portuguesa. Também, o tipo de dependência administrativa mostrou-se a informação mais relevante para explicar as proficiências para ambas as matérias.

Foi verificada a existência de variação no nível da escola, porém somente uma variável foi incluída neste nível. Uma possível razão para isto é o fato das outras variáveis não acrescentarem muita informação sozinhas. Uma sugestão para estudos futuros é a utilização dos índices construídos pelo próprio INEP. Não foi possível o uso destes índices no estudo pois a identificação das escolas nos microdados é diferente da usada nos índices.

A educação de uma sociedade não é medida apenas pela qualidade, que é muito importante, mas também pela equidade, ou seja, todos os estudantes recebem o mesmo padrão de educação. Isso, infelizmente, não ocorre no país, inclusive no Distrito Federal. Possíveis soluções seriam incentivar os alunos a gostar mais da matéria; para matemática, criar projetos de inclusão para as mulheres; oferecer suporte a alunos que já reprovaram, e planejar metas para aprimorar o ensino da rede pública.

REFERÊNCIAS

ARRUDA, L. Desvendando Desigualdade de Oportunidades em Ciências e em Matemática Relacionadas ao Gênero do Aluno: Uma Aplicação de Modelagem Multinível ao SAEB. **Revista Brasileira de Pesquisa em Educação em Ciências**. v. 2, n. 3, p. 67-83, 2002

ASSOCIAÇÃO BRASILEIRA DE EMPRESAS DE PESQUISA. **Critério de Classificação Econômica Brasil 2014**. São Paulo, Brasil. 2013.

BARBOSA, M. E. F.; FERNANDES, C. Modelo Multinível: Uma Aplicação a Dados de Avaliação Educacional. **Estudos em Avaliação Educacional**. n. 22, p. 135-153, 2000.

BOLIN, J.E.; FINCH, W.H.; KELLEY, K. **Multilevel Analysis Using R**. 1st ed. CRC Press, 2014, 207 p.

DIEZ-ROUX, A.V. Bringing Context Back into Epidemiology: Variables and Fallacies in Multilevel Analysis. **American Journal of Public Health**. v. 88, p. 216-222, 1998.

HOX, J.J. **Multilevel Analysis: Techniques and Applications**. 2nd ed. Great Britain: Routledge, 2010. 393 p.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Nota Explicativa: Resultados Prova Brasil 2013**. Brasília, Brasil. 2015.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **SAEB 2013: Microdados da Aneb e Anresc (Prova Brasil)**. Brasília, Brasil. 2015.

LAROS, J.A.; MARCIANO, J.L.P. Análise Multinível Aplicada aos Dados do NELS:88. **Estudos em Avaliação Educacional**. v. 19, p. 263-278, 2008.

PORTAL BRASIL. **Saiba Como é a Divisão do Sistema de Educação Brasileiro**. Disponível em: <<http://www.brasil.gov.br/educacao/2014/05/saiba-como-e-a-divisao-do-sistema-de-educacao-brasileiro/view>> Acesso em: 5 de Julho de 2016.

A ANEXO

Nas próximas páginas estão anexadas as descrições dos níveis das escalas de proficiência para língua portuguesa e matemática.

A escala de Língua Portuguesa para o 3º ano do Ensino Médio

(continua)

LÍNGUA PORTUGUESA – 3º ANO DO ENSINO MÉDIO		
Nível		Descrição do nível
1	225-250	Nesse nível, o estudante pode ser capaz de identificar elementos da narrativa em história em quadrinhos; reconhecer a finalidade de recurso gráfico em artigos; reconhecer a relação de causa e consequência em lendas; inferir o sentido de palavra em letras de música e reportagens.
2	250-275	Nesse nível, o estudante pode ser capaz de reconhecer a ideia comum entre textos de gêneros diferentes e a ironia em tirinhas; reconhecer relações de sentido estabelecidas por conjunções ou locuções conjuntivas em letras de música e crônicas; reconhecer o uso de expressões características da linguagem (científica, profissional etc.) e a relação entre pronome e seu referente em artigos e reportagens; inferir o efeito de sentido da linguagem verbal e não verbal em notícias e charges.
3	275-300	Nesse nível, o estudante pode ser capaz de localizar informação explícita em artigos de opinião; identificar a finalidade de relatórios científicos; reconhecer relações de sentido marcadas por conjunções, a relação de causa e consequência e a relação entre o pronome e seu referente em fragmentos de romances; reconhecer o tema de uma crônica; reconhecer variantes linguísticas em artigos; reconhecer o sentido e o efeito de sentido produzido pelo uso de recursos morfossintáticos em contos, artigos e crônicas; reconhecer opiniões divergentes sobre o mesmo tema em diferentes textos; inferir informação, o sentido e o efeito de sentido produzido por expressão em reportagens e tirinhas.
4	300-325	Nesse nível, o estudante pode ser capaz de localizar informações explícitas em infográficos, reportagens, crônicas e artigos; identificar o argumento em contos; identificar a finalidade e a informação principal em notícias; reconhecer a relação entre os pronomes e seus referentes em contos; reconhecer elementos da narrativa em contos; reconhecer variantes linguísticas em contos, notícias e reportagens; reconhecer o efeito de sentido produzido pelo uso de recursos morfossintáticos em poemas; reconhecer ideia comum e opiniões divergentes sobre o mesmo tema na comparação entre diferentes textos; reconhecer ironia e efeito de humor em crônicas e entrevistas; reconhecer a relação de causa e consequência em piadas e fragmentos de romance; comparar poemas que abordem o mesmo tema; diferenciar fato de opinião em contos, artigos e reportagens; diferenciar tese de argumentos em artigos, entrevistas e crônicas; inferir informação, sentido de expressão e o efeito de sentido decorrente do uso de recursos morfossintáticos em crônicas; inferir o sentido decorrente do uso de recursos gráficos em poemas; inferir o efeito de sentido da linguagem verbal e não verbal e o efeito de humor em tirinhas.

(conclusão)

LÍNGUA PORTUGUESA – 3º ANO DO ENSINO MÉDIO		
Nível		Descrição do nível
5	325-350	Nesse nível, o estudante pode ser capaz de localizar informação explícita em resenhas; identificar a informação principal em reportagens; identificar elementos da narrativa e a relação entre argumento e ideia central em crônicas; reconhecer a finalidade de propagandas; reconhecer variantes linguísticas e o efeito de sentido de recursos gráficos em crônicas e artigos; reconhecer a relação de causa e consequência e relações de sentido marcadas por conjunções em reportagens, artigos e ensaios; reconhecer o tema em poemas; diferenciar fato de opinião em resenhas; inferir o sentido de palavras e expressões em piadas e letras de música; inferir informação em artigos; inferir o sentido de expressão em fragmentos de romances.
6	350-375	Nesse nível, o estudante pode ser capaz de reconhecer efeitos estilísticos em poemas; reconhecer ironia e efeitos de sentido decorrentes da repetição de palavras em sinopses; reconhecer opiniões distintas sobre o mesmo tema, na comparação entre diferentes textos; reconhecer finalidade e traços de humor em reportagens; reconhecer o efeito de sentido do humor em tirinhas; reconhecer o tema em contos e fragmentos de romances; reconhecer relação de sentido marcada por conjunção em crônicas; inferir informação e tema em reportagens, poemas, histórias em quadrinhos e tirinhas; inferir o sentido e o efeito de sentido de palavras ou de expressão em poemas, crônicas e fragmentos de romances.
7	375-400	Nesse nível, o estudante pode ser capaz de identificar a ideia central e o argumento em apresentações de livros, reportagens, editoriais e crônicas; identificar elementos da narrativa em crônicas, contos e fragmentos de romances; identificar ironia e tema em poemas e artigos; reconhecer relações de sentido marcadas por conjunção em artigos, reportagens e fragmentos de romances; reconhecer a relação de causa e consequência em reportagens e fragmentos de romances; reconhecer o efeito de sentido de recursos gráficos em artigos; reconhecer variantes linguísticas em letras de música e piadas; reconhecer a finalidade de reportagens, resenhas e artigos; inferir efeito de humor e ironia em tirinhas e charges.
8	400-425	Nesse nível, o estudante pode ser capaz de reconhecer o efeito de sentido resultante do uso de recursos morfosintáticos em artigos e letras de música.

A escala de Matemática para o 3º ano do Ensino Médio

(continua)

MATEMÁTICA – 3º ANO DO ENSINO MÉDIO		
Nível		Descrição do nível
1	225-250	<p>Espaço e forma Não existem itens âncora para esse nível.</p> <p>Grandezas e medidas Não existem itens âncora para esse nível.</p> <p>Números e operações; álgebra e funções Não existem itens âncora para esse nível.</p> <p>Tratamento de informações Nesse nível, o estudante pode ser capaz de associar uma tabela de até duas entradas a informações apresentadas textualmente ou em um gráfico de barras ou de linhas.</p>
2	250-275	<p>Espaço e forma Nesse nível, o estudante pode ser capaz de reconhecer as coordenadas de pontos representados em um plano cartesiano localizados no primeiro quadrante.</p> <p>Grandezas e medidas Não existem itens âncora para esse nível.</p> <p>Números e operações; álgebra e funções Nesse nível, o estudante pode ser capaz de reconhecer os zeros de uma função dada graficamente. Também é bem provável que os alunos determinem: o valor de uma função afim, dada sua lei de formação; um resultado utilizando o conceito de progressão aritmética.</p> <p>Tratamento de informações Nesse nível, o estudante pode ser capaz de associar um gráfico de setores a dados percentuais apresentados textualmente ou em uma tabela.</p>

(continuação)

MATEMÁTICA – 3º ANO DO ENSINO MÉDIO	
Nível	Descrição do nível
3 275-300	<p>Espaço e forma Não existem itens âncora para esse nível.</p> <p>Grandezas e medidas Não existem itens âncora para esse nível.</p> <p>Números e operações; álgebra e funções Nesse nível, o estudante pode ser capaz de reconhecer: o valor máximo de uma função quadrática representada graficamente; em um gráfico, o intervalo no qual a função assume valor máximo. Também podem ser capazes de determinar: por meio de proporcionalidade o gráfico de setores que representa uma situação com dados fornecidos textualmente; o quarto valor em uma relação de proporcionalidade direta a partir de três valores fornecidos em uma situação do cotidiano; um valor reajustado de uma quantia a partir de seu valor inicial e do percentual de reajuste. Além disso, é provável que resolvam problemas utilizando operações fundamentais com números naturais.</p> <p>Tratamento de informações Não existem itens âncora para esse nível.</p>
4 300-325	<p>Espaço e forma Não existem itens âncora para esse nível.</p> <p>Grandezas e medidas Nesse nível, o estudante pode ser capaz de resolver problemas envolvendo área de uma região composta por retângulos a partir de medidas fornecidas em texto e figura.</p> <p>Números e operações; álgebra e funções Nesse nível, o estudante pode ser capaz de reconhecer o gráfico de função a partir de valores fornecidos em um texto. Além disso, podem ser capazes de determinar: a lei de formação de uma função linear a partir de dados fornecidos em uma tabela; a solução de um sistema de duas equações lineares; um termo de progressão aritmética, dada sua forma geral; a probabilidade da ocorrência de um evento simples. Também é provável que resolvam: problemas utilizando proporcionalidade direta ou inversa, cujos valores devem ser obtidos a partir de operações simples; problemas de contagem usando princípio multiplicativo.</p> <p>Tratamento de informações Não existem itens âncora para esse nível.</p>

(continuação)

MATEMÁTICA – 3º ANO DO ENSINO MÉDIO	
Nível	Descrição do nível
5	325-350
	<p>Espaço e forma Não existem itens âncora para esse nível.</p> <p>Grandezas e medidas Nesse nível, o estudante pode ser capaz de determinar medidas de segmentos por meio da semelhança entre dois polígonos.</p> <p>Números e operações; álgebra e funções Nesse nível, o estudante pode ser capaz de determinar: o valor de variável dependente ou independente de uma função exponencial dada; o percentual que representa um valor em relação a outro; o valor de uma expressão algébrica; a solução de um sistema de três equações sendo uma com uma incógnita, outra com duas e a terceira com três incógnitas. Também é provável que sejam capazes de resolver problema envolvendo: divisão proporcional do lucro em relação a dois investimentos iniciais diferentes; operações, além das fundamentais, com números naturais; a relação linear entre duas variáveis para a determinação de uma delas; probabilidade de união de eventos. Além disso, é provável que os alunos sejam capazes de avaliar o comportamento de uma função representada graficamente, quanto ao seu crescimento.</p> <p>Tratamento de informações Não existem itens âncora para esse nível.</p>
6	350-375
	<p>Espaço e forma Nesse nível, o estudante pode ser capaz de reconhecer as coordenadas de pontos representados em um plano cartesiano e localizados em quadrantes diferentes do primeiro. É provável também que consigam associar um sólido geométrico simples a uma planificação usual dada. Além disso, há uma grande probabilidade de que resolvam problemas envolvendo Teorema de Pitágoras, para calcular a medida da hipotenusa de um triângulo pitagórico, a partir de informações apresentadas textualmente e em uma figura.</p> <p>Grandezas e medidas Nesse nível, o estudante pode ser capaz de determinar: a razão de semelhança entre as imagens de um mesmo objeto em escalas diferentes; o volume de um paralelepípedo retângulo, dada sua representação espacial.</p> <p>Números e operações; álgebra e funções Nesse nível, o estudante pode ser capaz de determinar os zeros de uma função quadrática, a partir de sua expressão algébrica. Além disso, é provável que resolvam problemas de porcentagem envolvendo números racionais não inteiros.</p> <p>Tratamento de informações Não existem itens âncora para esse nível.</p>

(continuação)

MATEMÁTICA – 3º ANO DO ENSINO MÉDIO	
Nível	Descrição do nível
7 375-400	<p>Espaço e forma Nesse nível, o estudante pode ser capaz de determinar: a medida de um dos lados de um triângulo retângulo, por meio de razões trigonométricas, fornecendo ou não as fórmulas; com o uso de do teorema de Pitágoras, a medida de um dos catetos de um triângulo retângulo não pitagórico.</p> <p>Grandezas e medidas Nesse nível, o estudante pode ser capaz de determinar a área de um polígono não convexo composto por retângulos e triângulos, a partir de informações fornecidas na figura. Além disso, é provável que consigam resolver problemas: por meio de semelhança de triângulos sem apoio de figura; envolvendo perímetros de triângulos equiláteros que compõem uma figura.</p> <p>Números e operações; álgebra e funções Nesse nível, o estudante pode ser capaz de reconhecer gráfico de função a partir de informações sobre sua variação descritas em um texto; os zeros de uma função quadrática em sua forma fatorada; gráfico de função afim a partir de sua representação algébrica; a equação de uma reta a partir de dois de seus pontos; as raízes de um polinômio apresentado na sua forma fatorada. Além disso, é provável também que os alunos sejam capazes de determinar os pontos de máximo ou de mínimo a partir do gráfico de uma função; o valor de uma expressão algébrica envolvendo módulo; o ponto de interseção de duas retas; a expressão algébrica que relaciona duas variáveis com valores dados em tabela ou gráfico; a maior raiz de um polinômio de 2º grau. Também é provável que os alunos sejam capazes de resolver problemas: para obter valor de variável dependente ou independente de uma função exponencial dada; que envolvam uma equação de 1º grau que requeira manipulação algébrica; envolvendo um sistema linear, dadas duas equações a duas incógnitas; usando permutação; utilizando probabilidade, envolvendo eventos independentes.</p> <p>Tratamento de informações Não existem itens âncora para esse nível.</p>

(continuação)

MATEMÁTICA – 3º ANO DO ENSINO MÉDIO	
Nível	Descrição do nível
8	400-425
	<p>Espaço e forma Nesse nível, o estudante pode ser capaz de reconhecer a proporcionalidade dos elementos lineares de figuras semelhantes. Também é provável que sejam capazes de determinar: uma das medidas de uma figura tridimensional, utilizando o Teorema de Pitágoras; a equação de uma circunferência, dados o centro e o raio; a quantidade de faces, vértices e arestas de um poliedro por meio da relação de Euler. É provável também que os alunos sejam capazes de resolver problema envolvendo razões trigonométricas no triângulo retângulo, com apoio de figura. Podem também ser capazes de associar um prisma a uma planificação usual dada.</p> <p>Grandezas e medidas Nesse nível, o estudante pode ser capaz de determinar a área da superfície de uma pirâmide regular; o volume de um paralelepípedo, dadas suas dimensões em unidades diferentes; o volume de cilindros.</p> <p>Números e operações; álgebra e funções Nesse nível, o estudante pode ser capaz de reconhecer: o gráfico de uma função trigonométrica da forma $y=\text{sen}(x)$; um sistema de equações associado a uma matriz. Também é provável que sejam capazes de determinar: a expressão algébrica associada a um dos trechos do gráfico de uma função definida por partes; o valor máximo de uma função quadrática a partir de sua expressão algébrica e das expressões que determinam as coordenadas do vértice; a distância entre dois pontos no plano cartesiano. É provável também que os alunos sejam capazes de resolver problema: usando arranjo; envolvendo a resolução de uma equação do 2º grau sendo dados seus coeficientes. Além disso, existe uma grande probabilidade de que sejam capazes de interpretar o significado dos coeficientes da equação de uma reta, a partir de sua forma reduzida.</p> <p>Tratamento de informações Não existem itens âncora para esse nível.</p>

(conclusão)

MATEMÁTICA – 3º ANO DO ENSINO MÉDIO	
Nível	Descrição do nível
9 425-450	<p>Espaço e forma Nesse nível, o estudante pode ser capaz de reconhecer a equação que representa uma circunferência, dentre diversas equações dadas. Também é provável que sejam capazes de determinar o centro e o raio de uma circunferência a partir de sua equação geral. É provável também que os alunos sejam capazes de resolver problemas envolvendo relações métricas em um triângulo retângulo que é parte de uma figura plana dada.</p> <p>Grandezas e medidas Nesse nível, o estudante pode ser capaz de determinar o volume de pirâmides regulares. É provável também que os alunos sejam capazes de resolver problema envolvendo: áreas de círculos e polígonos; semelhança de triângulos com apoio de figura na qual os dois triângulos apresentam ângulos opostos pelos vértices; envolvendo cálculo de volume de cilindro.</p> <p>Números e operações; álgebra e funções Nesse nível, o estudante pode ser capaz de reconhecer o gráfico de uma função exponencial do tipo $f(x)=10x+1$; o gráfico de uma função logarítmica dada a expressão algébrica da sua função inversa e seu gráfico. Também é provável que sejam capazes de determinar a expressão algébrica correspondente a uma função exponencial, a partir de dados fornecidos em texto ou gráfico; a inversa de uma função exponencial dada, representativa de uma situação do cotidiano; inclinação ou coeficiente angular de retas a partir de suas equações; um polinômio na forma fatorada, dadas as suas raízes.</p> <p>Tratamento de informações Não existem itens âncora para esse nível.</p>
10 450-475	<p>Espaço e forma Não existem itens âncora para esse nível.</p> <p>Grandezas e medidas Não existem itens âncora para esse nível.</p> <p>Números e operações; álgebra e funções Nesse nível, o estudante pode ser capaz de determinar a solução de um sistema de três equações lineares, a três incógnitas, apresentado na forma matricial escalonada.</p> <p>Tratamento de informações Não existem itens âncora para esse nível.</p>