



Universidade de Brasília
IE - Instituto de Exatas
Departamento de Estatística

Fatores associados ao desempenho escolar em Matemática do SAEB 2011 para alunos do 5^o ano no Distrito Federal

Um estudo multinível

Érica Lima Ambrosio

Brasília
2014

Érica Lima Ambrosio

Fatores associados ao desempenho
escolar em Matemática do SAEB 2011
para alunos do 5^o ano no Distrito Federal

Um estudo multinível

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

Orientador(a): M.^a Maria Teresa Leão
Costa

Brasília
2014

“To understand God’s thoughts we must study statistics,
for these are the measure of his purpose.”

Florence Nightingale

Agradecimentos

Agradeço primeiramente a Jeová Deus por ter me sustentando e pelas bênçãos que recebi durante toda a minha vida.

Aos meus pais Eliane e Marcos que sempre me ajudaram, aconselharam e me encorajaram a muito a cursar o bacharelado em Estatística. Também às minhas irmãs Daniela e Maísa que me apoiaram e me deram conselhos sábios durante momentos difíceis durante a graduação e principalmente durante a monografia. Aos meus tios Maurício e Catharine, e Everaldo e Elisabete que de forma direta e indireta sempre estiveram envolvidos na minha vida acadêmica. Aos meus primos Yuri e Emily que foram compreensivos nos momentos que eu não podia brincar pois precisava “fazer dever da escola”.

Aos meus amigos que se fosse citar todos por nome muitas páginas seriam utilizadas. Mas representando todos, cito os que participaram de forma direta no processo de confecção deste trabalho: Priscila Furtado dos Santos, Lucas Andrade dos Reis, Karina Caldas Pio Fernandes, Amanda Pimentel Carneiro, Mariana Rodrigues de Faria e Débora Alice Silva Fernandes.

Aos professores Donald Matthew Pianto e Maria Teresa Leão Costa que durante toda a graduação se mostraram acessíveis e com os quais eu pude contar em momentos de dificuldade. Agradeço também pelas orientações de Pibic e da monografia, respectivamente, que enriqueceram meu aprendizado e minha vida acadêmica.

E finalmente os professores Dalton Francisco de Andrade, Joaquim José Soares Neto e Ana Maria Nogales Vasconcelos que me disponibilizaram o banco de dados do nível de infraestrutura das escolas.

Resumo

Neste trabalho foram analisados os principais fatores associados ao resultado dos alunos de 5º ano do Distrito Federal para a prova do SAEB em Matemática no ano de 2011. A técnica utilizada foi a regressão multinível, sendo o primeiro nível o aluno e o segundo a escola. Os fatores que se mostraram mais significativos para um bom resultado no nível do aluno foram o fato do aluno ser do sexo masculino e o fato de ele nunca ter reprovado. No nível da escola a nota média da escola se mostrou o fator mais significativo. Estes resultados podem ser úteis para a motivação de projetos que aproximem as meninas das ciências exatas e para a realização de estudos que analisem o impacto da reprovação no rendimento acadêmico do aluno.

Palavras-chave: análise multinível, análise hierárquica, SAEB, Prova Brasil, Ensino Fundamental.

Lista de tabelas

Tabela 4.1: População de referência do SAEB do ano 2011	14
Tabela 5.1: Distribuição de frequência dos itens sobre o aluno por tipo de escola (em %)	19
Tabela 5.2: Distribuição de frequência da escolaridade da mãe do aluno por tipo de escola (em %)	20
Tabela 5.3: Distribuição de frequência para o histórico educacional do aluno por tipo de escola (em %)	21
Tabela 5.4: Distribuição de frequência do nível de estrutura das escolas (em %)	22
Tabela 5.5: Teste para a diferença das médias de notas para as variáveis candidatas à explicativas	23
Tabela 5.6: Modelo nulo	25
Tabela 5.7: Modelo com todas as variáveis explicativas do primeiro nível	26
Tabela 5.8: Modelo com o acréscimo das variáveis do segundo nível	28
Tabela 5.9: Modelo com o acréscimo dos efeitos aleatórios	29
Tabela A.1: Distribuição de frequência dos itens sobre bens e serviços na casa do aluno por tipo de escola (em %)	38
Tabela A.2: continuação – Distribuição de frequência dos itens sobre bens e serviços na casa do aluno por tipo de escola (em %)	39
Tabela A.3: Distribuição do número de pessoas que moram na casa do aluno por tipo de escola (em %)	40

Tabela A.4: Distribuição de frequência dos itens sobre os pais do aluno por tipo de escola (em %)	40
Tabela A.5: Distribuição de frequência para as respostas do envolvimento dos pais por tipo de escola (em %)	41
Tabela A.6: Distribuição de frequência para as respostas do incentivo dos pais por tipo de escola (em %)	41
Tabela A.7: Distribuição de frequência para os hábitos de leitura do aluno por tipo de escola (em %)	42
Tabela A.8: Distribuição de frequência para hábitos culturais do aluno por tipo de escola (em %)	43
Tabela A.9: Distribuição de frequência para o que o aluno faz quando não está na escola por tipo de escola (em %)	44
Tabela A.10: Distribuição de frequência para os hábitos de estudo e dos professores por tipo de escola (em %)	45
Tabela A.11: Distribuição de frequência da conservação de itens e equipamentos da escola (em %)	46
Tabela A.12: continuação – Distribuição de frequência da conservação de itens e equipamentos da escola (em %)	47
Tabela A.13: Distribuição de frequência dos itens sobre as salas de aula e a escola em si (em %)	48

Tabela A.14: continuação – Distribuição de frequência dos itens sobre as salas de aula e a escola em si (em %)	49
Tabela A.15: Distribuição de frequência da condição de recursos da escola (em %)	50
Tabela A.16: continuação – Distribuição de frequência da condição de recursos da escola (em %)	51
Tabela A.17: continuação – Distribuição de frequência da condição de recursos da escola (em %)	52
Tabela A.18: continuação – Distribuição de frequência da condição de recursos da escola (em %)	53
Tabela A.19: Distribuição de frequência dos itens sobre a biblioteca e a sala de leitura da escola (em %)	54
Tabela A.20: Distribuição de frequência dos itens sobre empréstimo de livros na escola (em %)	55

Sumário

1	Introdução	1
2	Objetivos	3
3	Referencial teórico	5
4	Metodologia	13
5	Resultados	17
6	Conclusão	33
	Referências Bibliográficas	35
A	Apêndice	38

Capítulo 1

Introdução

Frases como “O futuro é a educação”, “Educação é a base de um país forte” e outras do mesmo estilo são comumente utilizadas por candidatos à cargos políticos em campanhas eleitorais. Tais frases também são repetidas em conversas do cotidiano, mostrando a preocupação com o acesso a educação e a qualidade da mesma. Como grande parte da educação no Brasil é financiada com recursos públicos, existe a necessidade de saber se estes são bem empregados e se geram retorno, isto é, se os alunos recebem educação de qualidade. Assim, é natural que exista também a preocupação com o desempenho das crianças e jovens nas atividades escolares.

Neste sentido, o Sistema Nacional de Avaliação da Educação Básica - SAEB tem procurado avaliar a evolução da educação básica no Brasil a cada dois anos. Um dos pressupostos do SAEB é “nenhum fator determina, isoladamente, a qualidade do ensino” (INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS, 2013). Para analisar quais fatores determinam a qualidade do ensino, o SAEB é composto de uma prova de Português e Matemática, questionário sócio-econômico para os alunos e questionários para professores, diretores e um sobre as características da escola.

Identificar os fatores associados ao resultado dos alunos em uma avaliação como o SAEB pode mostrar para gestores de políticas públicas, diretores, coordenadores,

professores, aos pais e alunos onde a intervenção é necessária para que os resultados melhorem na próxima avaliação. E não apenas que os resultados da avaliação melhorem, mas que isto represente de fato que os alunos estão fixando os conhecimentos adquiridos e conseguindo utilizá-los quando necessário.

Já foram realizados estudos neste sentido por Laros *et al.* (2010) e Laros *et al.* (2012), que analisam os fatores associados aos resultados dos alunos do 3º ano do Ensino médio nas avaliações de Português e Matemática, respectivamente, no SAEB do ano 2001. Em ambos os estudos foi utilizada a análise multinível para a identificação dos fatores que podem estar associados aos resultados. A principal conclusão dos autores é que o nível socioeconômico é o fator que está mais associado ao resultados do alunos. Outros fatores associados são a escolaridade dos pais, recursos culturais na família do aluno, entre outros.

A regressão linear poderia ser a metodologia escolhida para identificar os fatores que afetam os resultados dos alunos, mas dada a estrutura da avaliação, esta metodologia não se mostra adequada. Alunos em uma mesma turma tendem a ter resultados parecidos, pois têm o mesmo professor e estão no mesmo ambiente. E isto pode comprometer os resultados apresentados pela regressão linear.

Um método apropriado para fazer a identificação dos fatores que determinam o resultados dos alunos é a análise multinível (HOX, 2010). Um estudo multinível se mostra adequado neste caso porque existem níveis claramente identificáveis na avaliação educacional, sendo estes: alunos, turma, escola, município e Unidade da Federação.

Sabe-se que nem sempre receber educação da melhor qualidade resultará em que o estudante obtenha o melhor resultado em uma avaliação. Tanto que é comum que alunos de uma mesma turma tenham resultados variados numa mesma avaliação. Fatores não necessariamente ligados à escola podem influenciar o resultado dos alunos. O objetivo principal deste estudo é identificar quais fatores — dentro ou fora da escola — influenciam o resultado dos alunos no SAEB do 5º ano no Distrito Federal.

Capítulo 2

Objetivos

Objetivo Principal

Identificar fatores associados ao desempenho acadêmico dos alunos de 5º ano do Distrito Federal através dos resultados do SAEB do ano 2011.

Objetivos Secundários

1. Avaliar a base de dados do SAEB do ano 2011 e verificar valores ausentes e extremos.
2. Estudar modelos hierárquicos de regressão.
3. Aplicar as metodologias citadas na base de dados do SAEB com o objetivo de identificar as características dos alunos e das escolas associadas ao desempenho em Matemática dos alunos de 5º ano do Distrito Federal no SAEB 2011.

Capítulo 3

Rereferencial teórico

Para a utilização da regressão linear, são exigidos pressupostos que não podem ser atendidos pelos dados do SAEB. A estrutura hierárquica dos dados — aluno, turma, escola, município, Unidade da Federação —, está de tal maneira que a suposição de independência não é atendida. Alunos de uma mesma escola tendem a ter vários fatores em comum, como meio de transporte à escola e índice sócio-econômico. Alunos de uma mesma turma têm o mesmo professor. Essa falta de independência das observações pode levar a vários resultados estatisticamente significantes, mas que na realidade têm uma associação espúria.

Uma das metodologias adequadas para a análise de dados com estrutura hierárquica é a análise multinível (HOX, 2010). O método multinível leva em consideração o nível em que cada variável é observada — por exemplo, o resultado de cada estudante no SAEB é observado no nível do aluno; a experiência de um professor é observada no nível da turma, pois em geral as escolas têm mais de um professor para a mesma disciplina; o fato de a instituição de ensino ser pública ou particular é observado no nível da escola.

Adaptando-se um exemplo encontrado em Hox (2010) para o projeto em desenvolvimento, um modelo multinível simples para a nota dos alunos de uma escola j no SAEB

é dado por

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + e_{ij}. \quad (3.1)$$

Neste caso:

Y_{ij} é a nota no SAEB do aluno i da escola j

β_{0j} é o intercepto

β_{1j} e β_{2j} são os coeficientes de regressão para cada uma das variáveis X_{1ij} e X_{2ij} , respectivamente

e_{ij} é o erro

Exemplos de variáveis X_{1ij} e X_{2ij} apropriadas para este projeto são a variável do nível sócio-econômico e se o aluno costuma fazer o dever de casa. Na análise multinível, as escolas — ou turmas, cidades, unidades da federação, conforme o agrupamento escolhido — podem ter interceptos e coeficientes de regressão diferentes entre si, por isso o índice j está presente nos parâmetros. Para o erro, assume-se que este tem média zero e uma variância a ser estimada. Também é assumido que este erro é o mesmo para todos os níveis.

O próximo passo é explicar a variação nos coeficientes de regressão com variáveis explicativas no nível da escola, a variável representada por Z . Para este projeto, a variável Z pode ser dada, por exemplo, pela experiência dos professores ou a proporção de alunos na escola que satisfazem a relação idade-série. Assim, segue-se que os coeficientes de regressão da equação 3.1 podem ser

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + \omega_{0j}, \quad (3.2)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + \omega_{1j}, \quad (3.3)$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}Z_j + \omega_{2j}. \quad (3.4)$$

Como β_{0j} é o intercepto, neste caso o coeficiente pode ser entendido como o valor estimado da nota média na escola j considerando a variável Z . As equações 3.3 e 3.4 mostram que a relação entre o resultado do aluno no SAEB e as variáveis X_{1ij} e X_{2ij} dependem da variável Z . Segundo Hox (2010), a variável Z pode ser chamada de variável moderadora, pois ela modera a relação entre a variável resposta e as variáveis explicativas no nível mais baixo. Os termos ω_{0j} , ω_{1j} e ω_{2j} são os termos para o erro no nível da escola, e estes incorporam a parte aleatória do modelo. É assumido que erros ω_{ij} têm média zero e variância com formato $\sigma_{\omega_i}^2$ e covariâncias $\sigma_{\omega_{01}}$, $\sigma_{\omega_{02}}$ e $\sigma_{\omega_{12}}$. Também assume-se que estes erros são independentes de e_{ij} .

Substituindo as equações 3.2 a 3.4 na equação 3.1, esta pode ser reescrita como:

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{1ij} + \gamma_{20}X_{2ij} + \gamma_{01}Z_j + \gamma_{11}X_{1ij}Z_j + \gamma_{21}X_{2ij}Z_j + \omega_{1j}X_{1ij} + \omega_{2j}X_{2ij} + \omega_{0j} + e_{ij}. \quad (3.5)$$

O segmento $[\gamma_{00} + \gamma_{10}X_{1ij} + \gamma_{20}X_{2ij} + \gamma_{01}Z_j + \gamma_{11}X_{1ij}Z_j + \gamma_{21}X_{2ij}Z_j]$ representa a parte fixa do modelo, enquanto $[\omega_{1j}X_{1ij} + \omega_{2j}X_{2ij} + \omega_{0j} + e_{ij}]$ contém os termos aleatórios que constituem a parte aleatória do modelo. Os parâmetros do modelo podem ser estimados pelo método da máxima verossimilhança, quadrados mínimos generalizados, estimação de equações generalizadas ou por Monte Carlo via Cadeias de Markov. Os métodos de máxima verossimilhança produzem estimativas que maximizam a probabilidade dos dados serem observados segundo o modelo, e são métodos robustos, com estimativas assintoticamente eficientes e consistentes (HOX, 2010). Neste projeto foi utilizado o método de estimação por máxima verossimilhança fixa — que inclui os efeitos aleatórios e fixos na função de verossimilhança. Apesar de em geral apresentar valores bem próximos aos obtidos por máxima verossimilhança restrita, o método escolhido tem a vantagem de poder ter seus modelos comparados por meio de testes de qui-quadrado, se a única diferença entre os modelos for os efeitos fixos. O cálculo das estimativas por meio desse método deve ser feito por métodos computacionais, pois

exige um procedimento iterativo. Segundo Hox (2010), o programa gera valores iniciais para os vários parâmetros, e a cada passo o procedimento tenta melhorar estes valores iniciais — este passo se repete inúmeras vezes. Cada vez que os valores iniciais são melhorados, avalia-se de quanto foi esta aparente melhora. Quando as mudanças são muito pequenas, o algoritmo para, pois entende que o modelo convergiu e os valores iniciais agora são as estimativas.

Da função de máxima verossimilhança pode-se calcular a estatística *deviance*, que mostra quão bem o modelo se ajusta aos dados. A *deviance* é definida como $-2 \times \ln(\text{verossimilhança})$, sendo $\ln(\text{verossimilhança})$ o logaritmo natural da função de verossimilhança no ponto de convergência. Em geral, quando dois modelos são comparados, o modelo com *deviance* menor é o que está melhor ajustado, este é o teste de qui-quadrado para diferença de *deviances*. Espera-se que ao se adicionar variáveis explicativas a um modelo nulo o valor da *deviance* caia. Do modelo nulo pode-se obter a correlação intraclassa, $\rho = \frac{\sigma_{\omega_0}^2}{\sigma_{\omega_0}^2 + \sigma_e^2}$, que é uma razão que mostra a proporção da variância que está entre grupos e também pode ser interpretada como a correlação esperada entre duas observações escolhidas aleatoriamente dentro do mesmo grupo. Para modelos encaixados (quando um modelo específico pode ser gerado de outro mais geral pela remoção de parâmetros do modelo mais geral) a diferença das *deviances* dos respectivos modelos tem distribuição qui-quadrado com graus de liberdade igual a diferença no número de parâmetros que são estimados nos dois modelos.

Para testar a significância dos coeficientes dos efeitos fixos pode-se realizar o teste de Wald, preferível para amostras com poucos grupos, ou teste de t de Student. Este último é mais conservativo e foi o utilizado neste projeto. As respectivas definições de cada teste podem ser encontradas em Agresti (2007) e Bussab e Morettin (2002). O teste de qui-quadrado para a diferença das *deviances* também é utilizado para se verificar a significância dos efeitos aleatórios. Isto é feito ao se comparar modelos com e sem efeitos aleatórios.

Segundo Hox (2010), a escolha de um modelo que se ajuste aos dados pode ser dada em cinco passos que serão analisados a seguir.

Primeiro passo

Sendo o primeiro deles analisar um modelo que contém apenas o intercepto e nenhuma variável explicativa. Assim, este modelo pode ser dado por

$$Y_{ij} = \beta_{0j} + e_{ij}. \quad (3.6)$$

O caso onde não existem variáveis explicativas, a equação 3.2 passa a ser $\beta_{0j} = \gamma_{00} + \omega_{0j}$. Substituindo, o a equação 3.6 é dada por

$$Y_{ij} = \gamma_{00} + \omega_{0j} + e_{ij}.$$

Este modelo é útil num primeiro momento pois com ele é possível estimar a correlação intraclasse e achar um valor de referência para o *deviance*.

Segundo passo

O segundo passo consiste em analisar um modelo com todas as variáveis explicativas do nível mais baixo — no caso do SAEB o nível mais baixo consiste nos alunos. Desta maneira é possível avaliar a contribuição de cada uma destas variáveis no modelo. O modelo a ser analisado é da forma

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + e_{ij} \quad (3.7)$$

onde X_{1ij} e X_{2ij} são as variáveis explicativas. O modelo deste passo pode ser comparado com o modelo sem variáveis explicativas por se comparar a *deviance* dos modelos.

Terceiro passo

No terceiro passo as variáveis explicativas de um nível mais alto são acrescentadas ao modelo. Para os dados do SAEB, um nível mais alto pode ser a escola. Assim, o modelo ser analisado neste passo é

$$Y_{ij} = \beta_{00} + \beta_{10}X_{1ij} + \beta_{20}X_{2ij} + \beta_{01}Z_{1j} + \omega_{0j} + e_{ij}, \quad (3.8)$$

onde Z_{1j} é uma variável de um nível mais alto. Com um modelo que consiste de variáveis no nível mais baixo e variáveis de um nível mais alto pode-se examinar se as variações entre grupos na variável dependente podem ser explicadas pelas variáveis de um nível mais alto.

Quarto passo

No próximo passo, é analisado se algum dos coeficientes das variáveis explicativas têm variância significativa entre os grupos. Isto é feito com base em um modelo de coeficientes aleatórios. Para este passo, o modelo é

$$Y_{ij} = \beta_{00} + \beta_{10}X_{1ij} + \beta_{20}X_{2ij} + \beta_{01}Z_{1j} + \omega_{1j}X_{1ij} + \omega_{0j} + e_{ij}. \quad (3.9)$$

O termo ω_{0j} é referente aos erros no nível da escola — no caso do SAEB — para os coeficientes das variáveis explicativas X_{1ij} . Os coeficientes que têm variância significativa são adicionados ao modelo final. É recomendável verificar se o modelo encontrado no quarto passo tem melhor adequabilidade aos dados do que o modelo do terceiro passo. O teste de χ^2 baseado nas *deviances* pode ser útil para esta comparação.

Quinto passo

O passo final consiste em adicionar ao modelo as interações entre os níveis de variáveis explicativas. Isso será feito para aquelas variáveis que tiveram sua variância detectada como significativa no quarto passo. O modelo completo para este passo é dado por

$$Y_{ij} = \beta_{00} + \beta_{10}X_{1ij} + \beta_{20}X_{2ij} + \beta_{01}Z_j + \beta_{11}X_{1ij}Z_j + \omega_{1j}X_{1ij} + \omega_{0j} + e_{ij}. \quad (3.10)$$

A ajustabilidade deste modelo final poderá ser testada por meio de teste de qualidade do ajustamento de χ^2 .

É digno de nota que ao acrescentar uma interação ao modelo, é necessário que as variáveis envolvidas também sejam incluídas — mesmo que individualmente elas não sejam significativas. E também existe uma mudança de interpretação dos coeficientes de regressão das variáveis que fazem parte da interação. Neste caso, o coeficiente de regressão representa o valor esperado deste coeficiente quando a outra variável da interação é igual a zero. Se as variáveis não podem assumir valor zero, não existe interpretação significativa destes coeficientes. Uma solução é a centralização, e assim a interpretação dos coeficientes de regressão não muda quando as interações passam a fazer parte do modelo.

Análise de resíduos

A análise de resíduos pode ser realizada por meio de análise gráfica. Ao contrário da regressão, na análise multinível existem resíduos diferentes para cada efeito aleatório do modelo. Assim é necessário que a análise seja feita para cada um dos resíduos para se verificar se os pressupostos de normalidade, não-linearidade e heterocedasticidade.

Para se analisar a normalidade dos resíduos pode ser feito um gráfico quantil-quantil dos resíduos padronizados e os quantis da distribuição normal. Se o pressuposto da normalidade for atendido com sucesso, uma linha diagonal à direita será formada no

gráfico.

Um gráfico de dispersão dos valores dos resíduos e os valores preditos pelo modelo para a variável resposta mostra se há falhas ao se atender os pressupostos de normalidade, não-linearidade e heterocedasticidade. Se os pressupostos forem atendidos, os pontos devem estar igualmente distribuídos acima e abaixo do valor 0 para os resíduos, e não devem existir padrões no gráfico.

Capítulo 4

Metodologia

Para este projeto são analisados três bancos de dados: dois provenientes do Sistema Nacional de Avaliação da Educação Básica (SAEB) referente ao ano de 2011, um com informações sobre o aluno e outro com os resultados do aluno na prova; e um banco de dados baseado nas informações do Censo Escolar 2011 com informações sobre a infraestrutura da escola. Um quarto banco de dados é utilizado somente durante a análise descritiva, que são os dados do SAEB com informações sobre a escola. Os microdados estão disponíveis gratuitamente no site do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira — INEP.

O SAEB é um sistema de avaliações aplicado a cada dois anos para alunos do 5º e do 9º ano do Ensino Fundamental e do 3º ano do Ensino Médio, de escolas públicas e particulares de todas as Unidades da Federação Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2013).

Os alunos participantes respondem a uma prova de Português e Matemática, e um questionário com perguntas referentes à escola, família, hábitos de leitura, trajetória escolar e motivação. Os professores das disciplinas avaliadas e os diretores respondem questionários referentes à sua trajetória profissional, estilo de liderança, práticas pedagógicas, entre outros tópicos. E existe também um questionário com perguntas sobre

características da escola.

O público-alvo da avaliação são todos os alunos de 5º e 9º ano do Ensino Fundamental de escolas públicas com 20 ou mais alunos, além de uma amostra composta por alunos de 5º e 9º ano do Ensino Fundamental de escolas particulares e do 3º ano do Ensino Médio de escolas públicas e particulares. Para fazer parte da amostra, as escolas devem ter pelo menos 10 alunos em suas turmas regulares.

Tabela 4.1: População de referência do SAEB do ano 2011

Série	Alunos	Turmas	Escolas
5o ano EF	3238506	128653	64812
9o ano EF	3037352	104591	48659
3o ano EM	2118160	66407	23233
Total	8394018	299651	87380

Fonte: Microdados do SAEB 2011, INEP (2013)

Com uma população de referência de tal magnitude, é possível entender o grande impacto que o SAEB tem.

Neste estudo, a população alvo é constituída pelos alunos do 5º ano do Ensino Fundamental do Distrito Federal.

Um desafio concernente ao SAEB é o grande número de observações sem resposta alguma, ou quase completamente composta de dados faltantes. O tamanho dos questionários aplicados aos alunos, professores, diretores e sobre a escola é muito grande e por este motivo muitos respondentes decidem por não preenchê-lo da maneira adequada. A estratégia aplicada para solucionar este problema foi considerar apenas os questionários que estavam pelo menos 70% respondidos, e aquelas variáveis que tinham pelo menos 90% de suas observações com respostas.

Com a população alvo e os requisitos mínimos de resposta escolhidos, a amostra a ser analisada neste projeto é composta por 32397 alunos e 371 escolas.

Para informações referentes à infraestrutura da escola foi utilizado um banco de

dados baseado nas informações do Censo Escolar 2011 feito por Soares Neto *et al.* (2013). A decisão de se utilizar este banco de dados e não o do SAEB com as respostas às perguntas do questionário para a escola foi feita para resumir a informação sobre as instituições de ensino. O banco de dados do SAEB para a escola tem muitas variáveis que isoladamente não contribuíram significativamente para o modelo, mas que têm importância no resultado do aluno mesmo que de forma indireta. O banco de dados desenvolvido por Soares Neto *et al.* (2013) traz um indicador da infraestrutura da escola, classificando-a como elementar, básica, adequada ou avançada.

A metodologia utilizada para a criação do indicador é a Teoria de Resposta ao Item, e algumas das variáveis consideradas são: a escola ter ou não copiadora, computador, biblioteca, acesso à internet e laboratório. No modelo final deste trabalho as variáveis consideradas no nível da escola são o indicador de infraestrutura e se a escola é pública ou particular.

Para separar os alunos em classes socioeconômicas de maneira adequada, foi utilizado o Critério de Classificação Econômica Brasil válido a partir de 01/01/2013, criado pela ABEP (ASSOCIAÇÃO BRASILEIRA DE EMPRESAS DE PESQUISA, 2012). Este critério tem como base a posse de televisão em cores, rádio, banheiro, automóvel, máquina de lavar, videocassete e/ou DVD, geladeira, freezer; a utilização de serviços de empregada doméstica; e a escolaridade do chefe da família. Todos os itens estão presentes no questionário socioeconômico respondido pelos alunos, exceto a escolaridade do chefe da família. No entanto, o questionário pergunta a escolaridade de ambos os pais. Como a variável escolaridade do pai não está sendo levada em consideração neste projeto, devido a sua taxa de não resposta, a escolaridade da mãe é considerada a escolaridade do chefe da família. Foi escolhido o Critério Brasil 2013 pois ele é baseado nas informações do Levantamento Socioeconômico do Ibope de 2011, o ano da edição do SAEB analisada nesse projeto.

O Critério Brasil separa as classes socioeconômicas em oito grupos, sendo estes A1,

A2, B1, B2, C1, C2, D e E. Em Associação Brasileira de Empresas de Pesquisa (2012) pode ser vista uma estimativa da renda média bruta familiar por mês para cada classe e devido à proximidade de valores entre algumas das classes, neste estudo as classes estão separadas por A (composta por A1 e A2), B1, B2, C (composta por C1 e C2) e DE (composta por D e E).

Por meio do pacote estatístico R (R Core Team, 2014), antes da análise descritiva foi aplicada a estratégia comentada na página 14 para que apenas os questionários com pelo menos 70% das perguntas respondidas e variáveis com no mínimo 90% de observações com resposta entrassem na análise final. Após este passo foi feita a análise descritiva dos dados para que se pudesse ter uma ideia das informações trazidas por estes a respeito da escola e dos alunos. Durante este passo foi construída a variável do Critério Brasil no banco de dados com informações sobre os alunos.

Em seguida, as variáveis que não apresentaram distribuição de frequências homogênea na análise descritiva foram comparadas com a nota obtida pelos alunos em Matemática. Com isso decidiu-se quais seriam candidatas à variáveis explicativas no modelo.

Após este passo foi iniciado o processo para a modelagem multinível, ou hierárquica, utilizando-se do pacote *lme4*, desenvolvido por Bates *et al.* (2014). Foram seguidos os passos para a escolha do modelo propostos em Hox (2010) e discutidos em mais detalhes no Referencial Teórico deste relatório. O último estágio foi realizar a análise do modelo final.

Capítulo 5

Resultados

Com a análise descritiva dos dados foi possível compreender quem são os alunos do 5º ano do Distrito Federal que prestaram o SAEB 2011 e como são as escolas que eles estudam. As tabelas referentes às distribuições de frequências para os questionários respondidos no SAEB estão no Apêndice. Para quase todas as variáveis que se tornaram candidatas à explicativas, as tabelas estão no corpo do texto.

O questionário respondido pelos estudantes conta com 54 perguntas. Os itens podem ser agrupadas da seguinte maneira:

- Sobre o aluno, como idade, gênero, cor/raça;
- Sobre bens na casa do aluno;
- Quantas pessoas moram com o aluno;
- Perguntas sobre a mãe;
- Perguntas sobre o pai — dado o índice de não resposta nesse grupo, apenas a pergunta se o aluno mora com o pai se manteve no banco de dados final;
- Envolvimento dos pais com o desenvolvimentos do aluno na escola;

- Sobre os hábitos de leitura — foram excluídas as perguntas referentes a livros e revistas, dado o índice de não respostas;
- Sobre os hábitos culturais — foi retirada a pergunta sobre museus;
- O que o aluno faz quando não está na escola;
- O que o aluno faz quando não está na escola;
- Histórico educacional do aluno, sobre o tipo de escola que ele frequentou, se já reprovou e assim por diante;
- Hábitos de estudo do aluno e se os professores corrigem as lições.

No primeiro grupo de variáveis, que são questões sobre o aluno, percebe-se na Tabela 5.1 uma diferença nas distribuições de cor/raça e distribuição da idade dos alunos entre as escolas públicas e particulares. O fato de por volta de de 30% dos alunos de cada tipo de escola ter 11 anos pode ser facilmente explicado pelo fato das provas serem aplicadas no segundo semestre, quando muitos aluno que não estão defasados já completaram a idade apropriada à série seguinte — neste caso onze anos. É interessante notar que a escola pública tem uma proporção bem maior de crianças mais velhas no 5^o ano quando comparada com a escola particular.

Tabela 5.1: Distribuição de frequência dos itens sobre o aluno por tipo de escola (em %)

Características	Pública n=31589	Particular n=808	Total n=32397
Sexo			
Masculino	50,22	49,50	50,20
Feminino	49,78	50,50	49,80
Cor			
Branca	22,59	41,57	23,07
Parda	50,69	37,26	50,36
Preta	8,70	3,55	8,57
Amarela	2,00	1,65	2,00
Indígena	3,12	3,68	3,13
Não sei	12,89	12,29	12,88
Idade			
8	0,03	0,12	0,03
9	1,31	2,12	1,33
10	43,59	65,17	44,13
11	39,00	30,34	38,78
12	10,52	1,50	10,30
13	3,64	0,62	3,57
14 ou mais	1,90	0,12	1,86

Fonte: Microdados do SAEB 2011, INEP (2013)

Para as variáveis referentes aos bens e serviços na casa do aluno — Tabelas A.1 e A.2 no apêndice — é interessante observar que 0,65% dos alunos da rede pública moram em um local que não tem banheiro. Os únicos itens que não apresentam grandes diferenças de distribuição entre cada tipo de escola são: DVD, geladeira e rádio. A diferença na distribuição heterogênea dos outros itens pode ser explicada pelo poder aquisitivo das famílias dos alunos de cada tipo de escola.

A Tabela A.3, presente no apêndice, mostra que os alunos da rede pública costumam morar em casas com mais pessoas do que seus colegas da rede privada. Pelo que pode ser observado na Tabela A.4, não existem diferenças entre as distribuições das respostas

às perguntas sobre aos pais e o tipo de escola do aluno.

Tabela 5.2: Distribuição de frequência da escolaridade da mãe do aluno por tipo de escola (em %)

Escolaridade da mãe	Pública n=31589	Particular n=808	Total n=32397
Ensino Fundamental 1 incompleto	43,35	25,22	42,89
Ensino Fundamental 2 incompleto	12,59	2,69	12,34
Ensino Médio Incompleto	13,20	2,94	12,94
Ensino Superior incompleto	18,07	17,41	18,05
Ensino superior completo	12,80	51,73	13,78

Fonte: Microdados do SAEB 2011, INEP (2013)

O Ensino Fundamental 1 é o período do 1º ao 5º ano, e o Ensino Fundamental 2 do 6º ao 9º ano. Existe uma diferença muito grande nas distribuições de escolaridade das mães e o tipo de escola, conforme a Tabela 5.2. A maioria dos alunos das escolas particulares têm mães com o ensino superior completo, enquanto que a maior proporção entre os alunos das escolas públicas é o de mães que completaram no máximo o Ensino Fundamental 1. Esta é a variável utilizada no Critério Brasil para a escolaridade do chefe da família neste projeto.

Não existem diferenças entre as distribuições das respostas para envolvimento e incentivo dos pais aos estudos para cada tipo de escola, como pode ser visto nas Tabelas A.5 e A.6, presentes no apêndice.

Na Tabela A.7, no apêndice percebe-se uma diferença significativa entre as distribuições para em jornais e sites da internet por tipo de escola. Os alunos das escolas públicas costumam ler jornais em uma frequência um pouco maior, enquanto os alunos da rede particular costumam acessar sites da internet com muito mais frequência. Também na Tabela A.8, no apêndice, pode-se notar que a distribuição de frequência para o cinema é bem maior para as escolas particulares. A maneira como os alunos

gastam seu tempo livre, Tabela A.9, é ligeiramente diferente para cada tipo de escola, mas aparentemente não é algo que possa contribuir para o desempenho do aluno no SAEB em Matemática.

Tabela 5.3: Distribuição de frequência para o histórico educacional do aluno por tipo de escola (em %)

Histórico Educacional	Pública n=31589	Particular n=808	Total n=32397
Entrou na escola			
Creche	32,47	61,12	33,19
Pré	46,78	29,73	46,35
1ª série	17,03	4,45	16,72
Depois da 1ª série	3,72	4,70	3,74
Tipo de escola			
Só pública	71,62	8,77	70,03
Só particular	5,68	72,43	7,36
As duas	22,70	18,8	22,6
Reprovação			
Não	72,48	93,00	73,00
Sim	27,52	7,00	27,00
Abandonou			
Não	95,41	97,62	95,47
Sim	4,59	2,38	4,53

Fonte: Microdados do SAEB 2011, INEP (2013)

A maior parte dos alunos que estudam na escola particular entrou na escola na creche — como pode ser visto na Tabela 5.3. A diferença da distribuição de reprovação em cada tipo de escola é muito grande e é algo a ser refletido, e que pode afetar o desempenho do aluno no SAEB.

E para o último grupo de variáveis referentes ao aluno, percebem-se na Tabela A.10, no apêndice, diferenças nas distribuições de fazer o dever de casa de cada matéria e o fato do professor corrigí-la.

As perguntas no questionário do SAEB referente à escola, que é respondido por um

funcionário do INEP que visita a escola, podem ser agrupadas da seguinte maneira:

- Sobre a conservação de itens e equipamentos da escola
- Sobre a sala de aula e a escola em si — as perguntas sobre se as sala são iluminadas ou arejadas não está mais no banco de dados, dado o índice de não resposta
- Condição de recursos como televisão, computadores e biblioteca, entre outros
- Sobre a biblioteca ou sala de leitura
- A respeito do empréstimo de livros — a pergunta sobre se os membros da comunidade podem pegar livros emprestados na biblioteca foi retirada banco de dados pela taxa de não resposta

Conforme as Tabelas A.11 e A.12, no apêndice, todas as escolas têm paredes e pisos. Dos outros itens, as escolas particulares apresentam todos — em estado bom ou regular —, exceto cozinha. Este fato pode ser explicado pela falta de obrigação dessas escolas de fornecer refeições para os alunos. E muitas vezes a escola não tem cozinha própria, mas sim de uma lanchonete ou algo do tipo. Tanto que no indicador da infraestrutura de Soares Neto *et al.*(2013), cozinhas não são levadas em consideração para as escolas particulares.

Tabela 5.4: Distribuição de frequência no nível de estrutura das escolas (em %)

Nível da estrutura	Escolas n=371
Avançada	24,53
Adequada	59,30
Básica	16,17
Elementar	0

Fonte: Soares Neto et al. (2013)

Em todos os quesitos do grupo sobre a sala de aula e a escola em si — Tabelas A.13 e A.14, no apêndice —, a distribuição da escola particular é maior do que a pública, exceto para vigilância durante a noite e os finais de semana. Quanto ao grupo dos recursos — Tabelas A.15 a A.18, apêndice —, se a escola particular tem o recurso, ele é bom ou regular, exceto no caso da quadra de esportes que algumas escolas têm quadras de esportes ruins. É de se tomar nota que algumas escolas pública não têm itens tão básicos como uma copiadora, e isso é refletido no número de escolas públicas que têm mimeógrafos. 75% das escolas particulares não têm mais mimeógrafos. É interessante ver também que 4% das escolas particulares não têm televisão, enquanto que apenas 0,58% das escolas públicas não possuem o mesmo recurso. Todas as escolas particulares têm linha telefônica, por outro lado, 11,57% das escolas públicas ainda não têm um recurso tão básico como telefone fixo.

Tabela 5.5: Teste para a diferença das médias de notas para as variáveis candidatas à explicativas

Características	Nota na prova de Matemática			Estatística do teste	p-valor
	n	média	desvio padrão		
Sexo					
Feminino	15865	-0,3888429	0,7623042	161249	<2,2e-16
Masculino	15731	-0,5250289	0,6385887		
Reprovação					
Sim	8574	-0,7841434	0,6775319	480178	<2,2e-16
Não	23143	-0,3421359	0,7458807		
Abandono					
Sim	1441	-0,6885264	0,7817806	116583	<2,2e-16
Não	30292	-0,4518210	0,7516360		
Tipo de escola					
Pública	31485	-0,47698945	0,7515128	-162061	<2,2e-16
Particular	808	-0,04286654	0,7652685		

Fonte: Microdados do SAEB 2011, INEP (2013)

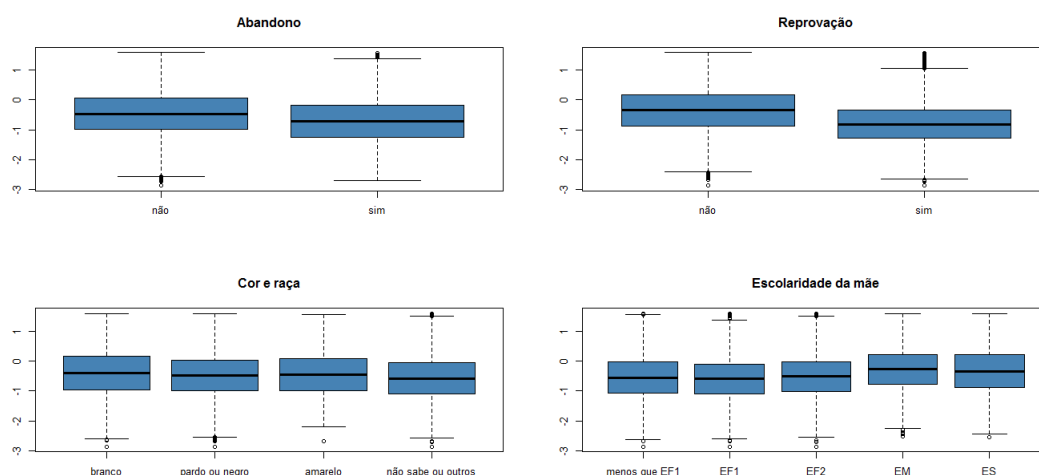
No itens das Tabelas A.19 e A.20, no apêndice, todas as distribuições de frequências

apresentam diferenças entre os tipos de escola. E é muito interessante notar que apenas nas escolas públicas o empréstimo de livros não é permitido para alunos e/ou professores.

Com as informações da Tabela 5.4 é possível ver que nenhuma das escolas analisadas neste trabalho está possui nível elementar de estrutura, que é o nível mais baixo.

O próximo passo se deu por avaliar de forma exploratória e bivariada a relação entre a nota obtida pelo aluno na prova de Matemática do SAEB e algumas das variáveis candidatas à variáveis explicativas do modelo final. Seguem algumas das relações mais notáveis representadas por boxplots, juntamente com alguns comentários:

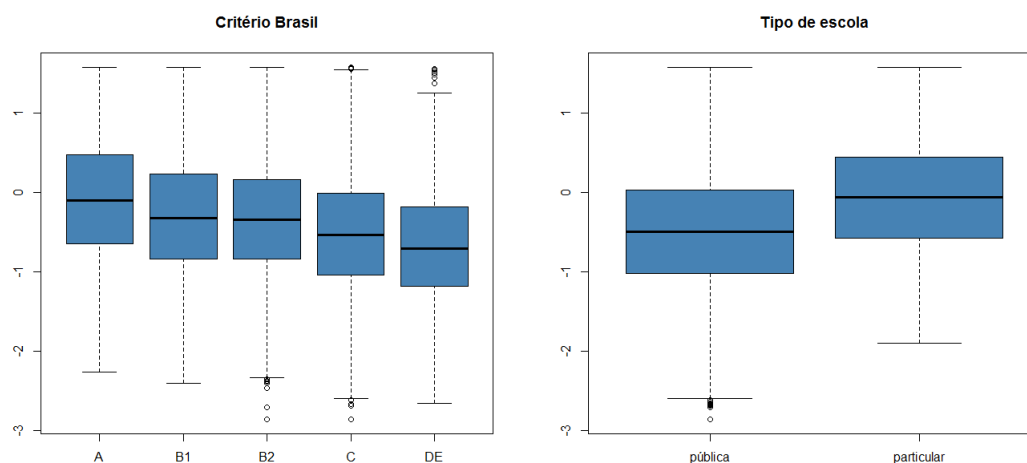
Figura 5.1: Distribuição das notas em Matemática pelo fato de o aluno já ter abandonado ou não a escola, por ele já ter reprovado ou não, pela cor e escolaridade da mãe



Percebe-se claramente pela Figura 5.1 que os alunos que já abandonaram a escola tem distribuição das notas inferior aos demais alunos. Se o aluno já reprovou, ele, em geral, tem notas menores que os demais colegas que nunca reprovaram. Também nota-se que os alunos pardos ou negros, e os que não declaram a raça, têm notas inferiores aos brancos e aos amarelos. E percebe-se que há diferença de distribuição das notas em comparação à escolaridade da mãe. A escolaridade da mãe não é analisada neste estudo

como um fator isolado, mas como um componente do Critério Brasil de Classificação Econômica.

Figura 5.2: Distribuição das notas em Matemática pelo Critério Brasil e tipo de escola do aluno



Na Figura 5.2 é possível ver que a nota varia conforme a classe do aluno no Critério Brasil. Há diferença na distribuição das notas segundo o tipo de escola.

Com isso, prosseguimos para a modelagem com base nessas e em outras variáveis explicativas. Para tal, seguimos os passos para a escolha de um modelo propostos por Hox (2010) com a variável resposta sendo a nota em Matemática no SAEB numa escala com média 0 e desvio-padrão 1. Tem-se o primeiro passo com um modelo nulo da forma:

Tabela 5.6: Modelo nulo

Parte fixa	Estimativa	Erro padrão	Estatística t
Intercepto	-0,44340	0,01506	-29,44
Parte aleatória	Estimativa	Desvio padrão	
Intercepto	0,07543	0,2746	
Resíduo	0,50715	0,7121	
<i>deviance</i>	70618,6		
Correlação intraclasse	0,1289		

No segundo passo são adicionadas as variáveis explicativas de um nível mais baixo. Para este projeto, as variáveis são as do nível do aluno. Foram escolhidas as variáveis:

sexo — a referência é o aluno ser do sexo masculino

cor — tendo por referência a cor branca

classe do Critério Brasil — com referência a classe B2, pois no agrupamento das classes em cinco grupos a classe B2 é a classe central.

reprovação

abandono, isto é, se o aluno alguma vez já abandonou a escola.

Segue o modelo para este passo:

Tabela 5.7: Modelo com todas as variáveis explicativas do primeiro nível

Parte fixa	Estimativa	Erro padrão	Estatística t
Intercepto	-0,117691	0,017814	-6,61
Sexo	-0,181972	0,008772	-20,74
Classe A	0,033254	0,026436	1,26
Classe B1	-0,027533	0,016592	-1,66
Classe C	-0,085832	0,010554	-8,13
Classe DE	-0,150827	0,019111	-7,89
Reprovação	-0,428918	0,010310	-41,60
Abandono	-0,039066	0,022061	-1,77
Pardo ou negro	-0,022548	0,010739	-2,10
Amarelo	0,0118268	0,031857	0,37
Outros	-0,097361	0,014261	-6,83
Parte aleatória	Estimativa	Desvio padrão	
Intercepto	0,05568	0,2360	
Resíduo	0,45885	0,6774	
<i>deviance</i>	51577,6		

Nota-se que até o momento o fator que mais tem influência na nota é o fato de o aluno já ter reprovado alguma série. Outro fator que tem grande contribuição é o sexo, pelos resultados do modelo, as meninas costumam ter notas mais baixas em Matemática

do que os meninos. Um resultado que está fora do esperado é a classe B1 ter a nota média menor do que a classe B2. Provavelmente o efeito dessa classe é confundido com o de alguma outra variável, mas seu efeito não é significativo, assim como o efeito da classe A. O efeito de a pessoa ser amarela também não é significante, porém o das outras cores é.

Como a categoria de cor/raça "Outros" inclui os indígenas e os que não souberam se classificar — no entanto, não inclui aqueles que deixaram a questão em branco — e seu efeito é mais significativo que os das outras cores, esse resultado é no mínimo fora do comum, dada que a população de indígenas no Distrito Federal não é grande. Foram feitos os passos de Hox (2010) para um modelo sem a categoria de cor Outro, mas este não mostrou muita diferença com o modelo que está sendo descrito no projeto. Assim, optou-se por continuar com o modelo que leva em consideração os indígenas e aqueles que não sabem declarar sua cor/raça.

Ao se comparar o modelo deste passo com o modelo nulo, é possível perceber uma grande diferença no valor da *deviance*, que de 70618,6 passou para 51577,6. O p-valor deu menor que $2,2 \times 10^{-16}$.

Na Tabela 5.8 está o modelo do terceiro passo, que é quando as variáveis do segundo nível são adicionadas. O segundo nível neste caso é a escola e as variáveis consideradas neste nível são o fato de a escola ser pública ou particular — a referência é a escola ser pública —, o nível da infraestrutura da escola e a média do rendimento da escola na prova de Matemática. Dos quatro níveis do indicador da infraestrutura, a referência é o nível básico. O nível elementar não entrou no modelo porque não existem escolas classificadas nesse nível na população dentre as escolas analisadas.

A Tabela 5.9 mostra que, das variáveis acrescentadas neste passo, o efeito da nota média da escola é o mais significativa para o modelo. Em um modelo com as mesmas variáveis, exceto a nota média, o efeito mais significativo é o tipo de escola. A estrutura e o tipo de escola terem coeficientes negativos neste modelo atual pode mostrar que há

um confundimento destes efeitos e o da nota média da escola, porque espera-se intuitivamente que estas variáveis contribuam com a melhora na nota do aluno, principalmente o nível de estrutura da escola. As variáveis no nível do aluno mantiveram a mesma significância e comportamento do passo anterior, exceto a classe A do Critério Brasil. Houve uma diminuição pouco expressiva no valor da *deviance* neste terceiro passo, que de 51577,6 passou a ser 50122,8, mas estatisticamente significativa, com p-valor $1,986 \times 10^{-5}$.

Tabela 5.8: Modelo com o acréscimo das variáveis do segundo nível

Parte fixa	Estimativa	Erro padrão	Estatística t
Intercepto	0,308489	0,020293	15,20
Sexo	-0,179721	0,008719	-20,61
Classe A	-0,016868	0,026592	-0,63
Classe B1	-0,045253	0,016471	-2,75
Classe C	-0,074291	0,010451	-7,11
Classe DE	-0,129737	0,018821	-6,89
Reprovação	-0,410561	0,010147	-40,46
Abandono	-0,033480	0,021938	-1,53
Pardo ou negro	-0,016646	0,010669	-1,56
Amarelo	0,022679	0,031641	0,72
Outros	-0,086464	0,014150	-6,11
Tipo de escola	-0,092022	0,028993	-3,17
Estrutura adequada	-0,037725	0,014569	-2,59
Estrutura avançada	-0,038693	0,015963	-2,42
Média da escola	0,916610	0,018067	50,73
Parte aleatória	Estimativa	Desvio padrão	
Intercepto	0,0000	0,0000	
Resíduo	0,4555	0,6749	
<i>deviance</i>	50122,8		

O quarto passo, na Tabela 5.9, consiste em verificar se algum dos coeficientes das variáveis explicativas têm variância significativa entre grupos. Para isto, utiliza-se um modelo de coeficientes aleatórios. Para testar o efeito aleatório foram escolhidas as variáveis cor e classe no Critério Brasil.

Ao se fazer o teste de χ^2 baseado nas *deviances* constata-se que os coeficientes das

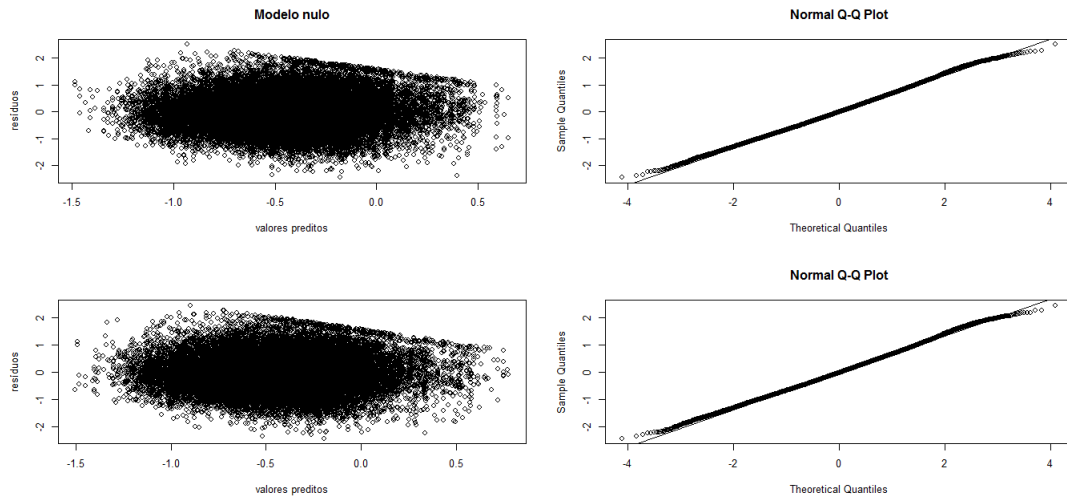
variáveis explicativas não têm variância significativa entre grupos. Foi feito o teste para modelos com apenas cor, e apenas classe do critério Brasil e ambos se mostraram não significativos neste respeito. Assim, não existem motivos para prosseguir para o quinto passo definido por Hox (2010) para a escolha de um modelo hierárquico.

Tabela 5.9: Modelo com o acréscimo dos efeitos aleatórios

Parte fixa	Estimativa	Erro padrão	Estatística t
Intercepto	0,308827	0,020341	15,18
Sexo	-0,179772	0,008714	-20,63
Classe A	-0,026880	0,029186	-0,92
Classe B1	-0,044670	0,016784	-2,66
Classe C	-0,073893	0,010747	-6,88
Classe DE	-0,132796	0,020044	-6,63
Reprovação	-0,410765	0,010141	-40,50
Abandono	-0,033736	0,021930	-1,54
Pardo ou negro	-0,017224	0,010873	-1,58
Amarelo	0,022571	0,032416	0,70
Outros	-0,085199	0,015061	-5,66
Tipo de escola	-0,101757	0,029685	-3,43
Estrutura adequada	-0,038330	0,014636	-2,62
Estrutura adequada	-0,038881	0,016028	-2,43
Média da escola	0,915962	0,018146	50,48
Parte aleatória	Estimativa	Desvio padrão	
Intercepto	0,00000	0,0000	
Pardo ou negro	0,001267	0,03559	
Amarelo	0,007664	0,08754	
Outros	0,007011	0,08373	
Classe A	0,020656	0,14372	
Classe B1	0,002645	0,05143	
Classe C	0,001765	0,04201	
Classe DE	0,010350	0,10173	
Resíduo	0,453042	0,67308	
<i>deviance</i>	51873		

Na análise de resíduos para o modelo nulo, Tabela 5.6, não foram constatadas violações aos pressupostos de normalidade, linearidade e heterocedasticidade, como pode ser observado nos gráficos abaixo.

Figura 5.3: Distribuição dos resíduos para os valores preditos e qq-plot para os modelos nulo e final



Ao se analisar os resíduos para o modelo com as variáveis do primeiro nível — Tabela 5.7, o nível do aluno —, é possível perceber que a distribuição dos resíduos para a nota prevista do aluno é parecida com a dos resíduos para o modelo nulo. O mesmo acontece para os resíduos no nível da escola.

Como não houveram efeitos aleatórios significativos no quarto passo na Tabela 5.9, o modelo final foi apresentado no terceiro passo da seguinte maneira:

$$\text{Nota em Matemática}_{ij} = 0,308489 - 0,179721 \times \text{ser menina}_{ij} - 0,016868 \times \text{Classe A}_{ij}$$

$$-0,045253 \times \text{Classe B}_{1ij} - 0,074291 \times \text{Classe C}_{ij} - 0,129737 \times \text{Classe DE}_{ij}$$

$$-0,410561 \times \text{reprovação}_{ij} - 0,033480 \times \text{abandono}_{ij} - 0,016646 \times \text{pardo ou negro}_{ij}$$

$$+ 0,022679 \times \text{amarelo}_{ij} - 0,086464 \times \text{outra cor ou não sabe}_{ij} - 0,092022 \times \text{escola particular}_j$$

$-0,037725 \times \text{estrutura adequada}_j - 0,038693 \times \text{estrutura avançada}_j + 0,916610 \times$
média da escola $_j$

Deste modelo, pode-se entender que, em média, um aluno que já reprovou tira - 0,410561 pontos do que um aluno que nunca reprovou. E, em média, as meninas tiram -0,179721 pontos a menos que os meninos na prova do SAEB de Matemática para alunos de 5º ano no Distrito Federal em 2011.

Capítulo 6

Conclusão

Dado o modelo final aquele que foi apresentado no terceiro passo, Tabela 5., os fatores que estão mais associados rendimento em Matemática de um aluno de 5^o ano no Distrito Federal que prestou o SAEB em 2011 são o fato de ele nunca ter reprovado e se ele é um menino. No nível da escola, o fator que está mais associado é a nota média da escola na prova — levando-nos a crer que se a escola tem em média uma nota boa, o aluno também terá.

O resultado deste trabalho de que meninos, em geral, obtiveram notas maiores na avaliação de Matemática do que a as meninas, pode ser motivador para o desenvolvimento de projetos que auxiliem as meninas a se interessar nas Ciências Exatas. A longo prazo isto pode contribuir para a melhora do resultado destas em Matemática e outras ciências puras.

Trabalhos futuros poderão analisar o impacto que a reprovação tem na vida acadêmica do aluno, e o porquê de a reprovação influenciar tanto o resultado do aluno no SAEB em Matemática. Algumas perguntas a serem respondidas são: Será que por ter reprovado uma das séries iniciais o aluno se acha inadequado para prosseguir os estudos? Será que os motivos que o levaram a reprovar são os mesmos que o impedem de ter um bom resultado no SAEB, como dificuldade de aprendizagem, falta de real apoio

SOARES NETO, J. J.; DE JESUS; G. R., KARINO; C. A., e DE ANDRADE, D. F. (2013). Uma escala para medir a infraestrutura escolar. *Estudos em Avaliação Educacional*, v. 24, n. 54, p. 78- 99, 2013.

Referências Bibliográficas

AGRESTI, A. *An introduction to categorical data analysis*. 2ª edição. John Wiley Sons, Inc. 2007.

ASSOCIAÇÃO BRASILEIRA DE EMPRESAS DE PESQUISA. *Critério de classificação econômica Brasil 2013*. São Paulo, Brasil. 2012.

BATES, D.; MAECHLER, M.; BOLKER, B.; e WALKER, S. lme4: *Linear mixed-effects models using Eigen and S4*. R package version 1.1-6. 2014

BUSSAB, W. O.; e MORETTIN, P. A. *Estatística Básica*. 5ª edição. Editora Saraiva. 2002.

HOX, J. *Multilevel Analysis: Techniques and Applications*. Taylor Francis. Quantitative Methodology Series. 2010.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS. *Microdados SAEB 2011: Manual do Usuário*. 2013.

LAROS, J. A.; MARCIANO, J. L.; e ANDRADE, J. M. Fatores associados ao desempenho escolar em português: um estudo multinível por regiões. *Ensaio: Avaliação e Políticas Públicas em Educação*, v. 20, p. 623-646, 2012,

LAROS, J. A.; MARCIANO, J. L. P., e ANDRADE, J. M. Fatores que afetam o desempenho na prova de Matemática do SAEB: Um estudo multinível. *Avaliação Psicológica*, v. 9 p. 173-186, 2010,

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. 2014.

SOARES NETO, J. J.; DE JESUS; G. R., KARINO; C. A., e DE ANDRADE, D. F. (2013). Uma escala para medir a infraestrutura escolar. *Estudos em Avaliação Educacional*, v. 24, n. 54, p. 78- 99, 2013.

Apêndice A

Apêndice

Tabela A.1: Distribuição de frequência dos itens sobre bens e serviços na casa do aluno por tipo de escola (em %)

Bens e Serviços	Pública n=31589	Particular n=808	Total n=32397
DVD			
sim	92,76	97,48	92,87
não	7,24	2,52	7,13
Geladeira			
sim	99,2	99,88	99,22
não	0,8	0,12	0,78
Freezer			
sim	79,33	94,32	79,71
não	20,67	5,68	20,29
Máquina			
sim	82,39	98,25	82,79
não	17,61	1,75	17,21
Computador			
sim	72,84	97,13	73,45
não	27,16	2,87	26,55
Empregada			
0	89,62	52,42	88,69
1	9,39	42,11	10,21
2 ou mais	0,99	5,47	1,11

Fonte: Microdados do SAEB 2011, INEP (2013)

Tabela A.2: continuação — Distribuição de frequência dos itens sobre bens e serviços na casa do aluno por tipo de escola (em %)

Bens e Serviços	Pública n=31589	Particular n=808	Total n=32397
TV			
0	2,81	0,51	2,75
1	34,03	13,65	33,52
2	38,53	31,51	38,35
3 ou mais	24,63	54,34	25,37
Rádio			
0	25,5	21,13	25,39
1	51,01	44,91	50,85
2	17,51	23,14	17,66
3 ou mais	5,98	10,82	6,1
Carro			
0	38,96	5,88	38,13
1	42,12	36,25	41,98
2	13,7	39,12	14,33
3 ou mais	5,22	18,75	5,55
Banheiro			
0	0,65	0,00	0,63
1	63,52	24,03	62,54
2	26,83	32,7	26,97
3	5,82	19,75	6,17
4 ou mais	3,19	23,52	3,69
Quarto			
0	0,66	0,25	0,65
1	11,57	2,26	11,33
2	39,91	23,74	39,51
3	33,31	35,05	33,35
4 ou mais	14,55	38,69	15,16

Fonte: Microdados do SAEB 2011, INEP (2013)

Tabela A.3: Distribuição do número de pessoas que moram na casa do aluno por tipo de escola (em %)

Quantas pessoas moram na casa	Pública n=31589	Particular n=808	Total n=32397
menos que 3	2,55	3,75	2,58
3	11,37	16,48	11,50
4	21,92	34,33	22,24
5	40,78	34,58	40,62
6	14,82	6,87	14,62
7 ou mais	8,56	4,00	8,44

Fonte: Microdados do SAEB 2011, INEP (2013)

Tabela A.4: Distribuição do frequência dos itens sobre os pais do aluno por tipo de escola (em %)

Relação com os pais	Pública n=31589	Particular n=808	Total n=32397
Mora com a mãe			
Sim	96,68	97,74	96,71
Não	3,32	2,26	3,29
Mora com o pai			
Sim	77,6	81,3	77,7
Não	22,4	18,7	22,3
Mãe alfabetizada			
Sim	96,88	99,5	97,5
Não	3,12	0,5	2,5
Vê a mãe lendo			
Sim	93,01	96,35	93,09
Não	6,99	3,65	6,91

Fonte: Microdados do SAEB 2011, INEP (2013)

Tabela A.5: Distribuição de frequência para as respostas do envolvimento dos pais por tipo de escola (em %)

Pais vão as reuniões	Pública n=31589	Particular n=808	Total n=32397
Sempre	71.70	68.33	71.61
De vez em quando	23.53	27.06	23.62
Nunca	4.77	4.61	4.76

Fonte: Microdados do SAEB 2011, INEP (2013)

Tabela A.6: Distribuição de frequência para as respostas do incentivo dos pais por tipo de escola (em %)

Incentivos	Pública n=31589	Particular n=808	Total n=32397
Incentivo aos estudos			
Sim	98,48	99,09	98,50
Não	1,52	0,91	1,50
Incentivo ao dever de casa			
Sim	97,11	98,25	97,14
Não	2,89	1,75	2,86
Incentivo à leitura			
Sim	95,05	94,94	95,05
Não	4,95	5,06	4,95
Incentivo para ir à escola			
Sim	97,3	97,62	97,30
Não	2,7	2,38	2,70
Conversa sobre a escola			
Sim	82,53	85,96	82,62
Não	17,47	14,04	17,38

Fonte: Microdados do SAEB 2011, INEP (2013)

Tabela A.7: Distribuição de frequência para hábitos de leitura do aluno por tipo de escola (em %)

Hábitos	Pública n=31589	Particular n=808	Total n=32397
Lê jornais			
Sempre	11,37	7,17	11,27
De vez em quando	48,38	39,97	48,16
Quase nunca	40,25	58,26	40,57
Lê livros em geral			
Sempre	43,19	45,93	43,25
De vez em quando	43,38	44,88	46,34
Quase nunca	10,44	9,19	10,41
Lê revistas em quadrinho			
Sempre	61,92	60,47	61,88
De vez em quando	30,31	31,99	30,35
Quase nunca	7,77	7,54	7,76
Lê sites da internet			
Sempre	51,57	68,77	52,01
De vez em quando	27,66	24,13	27,57
Quase nunca	20,77	7,10	20,42

Fonte: Microdados do SAEB 2011, INEP (2013)

Tabela A.8: Distribuição de frequência para hábitos culturais do aluno por tipo de escola (em %)

Hábitos	Pública n=31589	Particular n=808	Total n=32397
Frequenta biblioteca			
Sempre	19,3	19,17	19,30
De vez em quando	41,78	53,06	42,07
Quase nunca	38,92	27,77	38,63
Frequenta cinema			
Sempre	25,36	44,94	25,87
De vez em quando	52,42	49,94	52,35
Quase nunca	22,22	5,12	21,78
Frequenta eventos teatrais			
Sempre	18,48	13,84	18,36
De vez em quando	45,6	53,92	45,82
Quase nunca	35,92	32,25	35,82
Frequenta eventos musicais e de dança			
Sempre	29,24	22,96	29,08
De vez em quando	38,42	46,56	38,63
Quase nunca	32,35	30,48	32,30

Fonte: Microdados do SAEB 2011, INEP (2013)

Tabela A.9: Distribuição de frequência o que o aluno faz quando não está na escola por tipo de escola (em %)

Atividades	Pública n=31589	Particular n=808	Total n=32397
Internet e Tv - dia letivo			
1h	35,36	27,17	35,16
2h	18,98	22,89	19,08
3h	12,83	17,11	12,93
4h ou mais	32,83	32,83	32,83
Trabalho doméstico - dia letivo			
1h	48,19	45,05	48,11
2h	16,84	15,23	16,80
3h	7,47	4,57	7,40
4h ou mais	5,27	3,93	5,24
nada	22,23	31,22	22,46
Trabalha fora de casa			
Sim	6,03	5,01	6,01
Não	93,97	94,99	93,99

Fonte: Microdados do SAEB 2011, INEP (2013)

Tabela A.10: Distribuição de frequência para os hábitos de estudo e dos professores por tipo de escola (em %)

Atividades	Pública n=31589	Particular n=808	Total n=32397
Faz o dever de português			
Sempre	78,16	85,75	78,35
De vez em quando	20,13	12,5	19,93
Quase nunca	1,72	1,75	1,72
Professor corrige dever de português			
Sempre	84,52	92,35	84,72
De vez em quando	14,26	6,65	14,06
Quase nunca	1,22	1,00	1,22
Faz o dever de matemática			
Sempre	79,37	85,75	79,53
De vez em quando	19,36	12,75	19,19
Quase nunca	1,27	1,50	1,28
Professor corrige dever de matemática			
Sempre	87,08	93,61	87,25
De vez em quando	11,95	5,51	11,79
Quase nunca	0,97	0,88	0,96
Frequenta biblioteca da escola			
Sempre	35,16	35,11	35,16
De vez em quando	25,96	45,84	26,38
Quase nunca	38,89	22,05	38,46

Fonte: Microdados do SAEB 2011, INEP (2013)

Tabela A.11: Distribuição de frequência da conservação de itens e equipamentos da escola (em %)

Itens e Equipamentos	Pública n=346	Particular n=25	Total n=371
Telhado			
Bom	51,59	92,00	54,32
Regular	29,57	8,00	28,11
Ruim	18,55	0	17,30
Inexistente	0,29	0	0,27
Paredes			
Bom	58,6	100,00	61,41
Regular	31,78	0	29,62
Ruim	9,62	0	8,97
Inexistente			
Piso			
Bom	52,23	96,00	55,25
Regular	29,67	4,00	27,90
Ruim	18,1	0	16,85
Inexistente	0	0	0
Entrada do prédio			
Bom	58,06	96,00	60,66
Regular	29,91	4,00	28,14
Ruim	11,73	0	10,93
Inexistente	0,29	0	0,27
Pátio			
Bom	55,69	100,00	58,70
Regular	29,45	0	27,45
Ruim	12,24	0	11,41
Inexistente	2,62	0	2,45

Fonte: Microdados do SAEB 2011, INEP (2013)

Tabela A.12: continuação — Distribuição de frequência da conservação de itens e equipamentos da escola (em %)

Corredores			
Bom	58,06	95,83	60,55
Regular	25,51	4,17	24,11
Ruim	13,78	0	12,88
Inexistente	2,64	0	2,47
Salas de Aula			
Bom	51,16	100,00	54,35
Regular	37,79	0	35,33
Ruim	11,05	0	10,33
Inexistente	0	0	0
Portas			
Bom	48,97	96,00	52,19
Regular	31,67	4,00	29,78
Ruim	19,35	0	18,03
Inexistente	0	0	0
Janelas			
Bom	50,73	100,00	54,10
Regular	29,03	0	27,05
Ruim	19,65	0	18,31
Inexistente	0,59	0	0,55
Banheiros			
Bom	35,76	92,00	37,13
Regular	39,24	7,00	37,13
Ruim	24,71	0	23,04
Inexistente	0,29	0	0,27
Cozinha			
Bom	49,7	79,17	51,66
Regular	31,36	4,17	29,56
Ruim	18,34	0	17,13
Inexistente	0,59	16,67	1,66
Instalações Hidráulicas			
Bom	41,23	96,00	44,96
Regular	30,12	4,00	28,34
Ruim	28,36	0	26,43
Inexistente	0,29	0	0,27
Instalações Elétricas			
Bom	39,94	96,00	43,75
Regular	30,32	4,00	28,53
Ruim	29,15	0	27,17
Inexistente	0,58	0	0,54

Fonte: Microdados do SAEB 2011, INEP (2013)

Tabela A.13: Distribuição de frequência dos itens sobre as salas de aula e a escola em si (em %)

Itens sobre as salas de aula e a escola em si	Pública n=346	Particular n=25	Total n=371
Muros e Cercas			
sim	87,17	100,00	88,04
não	12,83	0	11,96
Controle de entrada e saída			
sim	94,22	100	94,61
não	5,78	0	5,39
Controle de entrada de pessoas estranhas			
sim	92,75	100,00	93,24
não	7,25	0	6,76
Portões trancados durante o funcionamento			
sim	85,42	100,00	86,41
não	14,58	0	13,59
Vigilância durante o dia			
sim	62,61	96,00	64,92
não	37,39	4,00	35,08
Vigilância durante a noite			
sim	97,38	79,17	96,20
não	2,62	20,83	3,80
Vigilância nos finais de semana			
sim	97,38	83,33	96,46
não	2,62	16,67	3,54
Policimento contra violência			
sim	21,70	50,00	23,56
não	78,30	50,00	76,44
Policimento contra tráfico dentro da escola			
sim	21,93	41,67	23,22
não	78,07	58,33	76,78

Fonte: Microdados do SAEB 2011, INEP

Tabela A.14: continuação — Distribuição de frequência dos itens sobre as salas de aula e a escola em si (em %)

Itens sobre as salas de aula e a escola em si	Pública n=346	Particular n=25	Total n=371
Policiamento contra o tráfico fora da escola			
sim	32,74	58,33	34,44
não	67,26	41,67	65,56
Proteção contra incêndio			
sim	77,06	100,00	78,63
não	22,94	0	21,37
Proteção na sala de equipamentos caros			
sim	90,56	91,67	90,63
não	9,44	8,33	9,37
Há depredação			
sim	26,65	30,43	26,89
não	73,35	69,57	73,11
Iluminação fora da escola			
sim	64,79	96,00	66,94
não	35,21	4,00	33,06
Proteção aos alunos nas imediações			
sim	39,47	70,83	41,53
não	60,53	29,17	58,47

Fonte: Microdados do SAEB 2011, INEP

Tabela A.15: Distribuição de frequência da condição dos recursos da escola (em %)

Recursos da Escola	Pública n=346	Particular n=25	Total n=371
Computadores para os alunos			
Bom	66,86	91,30	68,41
Regular	18,48	0	17,31
Ruim	4,11	0	3,85
Inexistente	10,56	8,70	10,44
Internet para os alunos			
Bom	42,23	83,33	44,93
Regular	19,65	8,33	18,90
Ruim	8,5	0	7,95
Inexistente	29,62	8,33	28,22
Computadores para os professores			
Bom	50,15	100,00	53,57
Regular	26,25	0	24,45
Ruim	10,91	0	10,16
Inexistente	12,68	0	11,81
Internet para os professores			
Bom	47,52	100,00	51,09
Regular	22,74	0	21,20
Ruim	7,58	0	7,07
Inexistente	22,16	0	20,65
Computadores exclusivos para uso adm.			
Bom	71,26	100,00	73,22
Regular	17,01	0	18,03
Ruim	3,52	0	7,38
Inexistente	2,64	0	1,37

Fonte: Microdados do SAEB 2011, INEP (2013)

Tabela A.16: continuação — Distribuição de frequência da condição dos recursos da escola (em %)

Recursos da Escola	Pública n=346	Particular n=25	Total n=371
Vídeos - educativos			
Bom	76,83	92,00	77,87
Regular	17,01	8,00	16,39
Ruim	3,52	0	3,28
Inexistente	2,64	0	2,46
Vídeos - lazer			
Bom	57,02	92,00	59,40
Regular	18,71	8,00	17,98
Ruim	6,73	0	6,27
Inexistente	17,54	0	16,35
Copiadora			
Bom	61,18	100,00	63,84
Regular	27,06	0	25,21
Ruim	8,24	0	7,67
Inexistente	3,53	0	3,29
Impressora			
Bom	55,52	96,00	58,27
Regular	31,69	4,00	29,81
Ruim	11,92	0	11,11
Inexistente	0,87	0	0,81
Retroprojektor			
Bom	72,06	79,17	72,53
Regular	10	8,33	9,89
Ruim	6,18	0	5,77
Inexistente	11,76	12,5	11,81
Projektor de slides			
Bom	48,66	79,17	50,69
Regular	9,5	4,17	9,14
Ruim	2,08	0	1,94
Inexistente	39,76	16,67	38,23
Videocassete ou dvd			
Bom	73,24	100,00	75,07
Regular	20,59	0	19,18
Ruim	4,41	0	4,11
Inexistente	1,76	0	1,64

Fonte: Microdados do SAEB 2011, INEP (2013)

Tabela A.17: continuação — Distribuição de frequência da condição dos recursos da escola (em %)

Recursos da Escola	Pública n=346	Particular n=25	Total n=371
Televisão			
Bom	74,78	96,00	76,22
Regular	22,03	0	20,54
Ruim	2,61	0	2,43
Inexistente	0,58	4,00	0,81
Mimeógrafo			
Bom	38,82	20,83	37,64
Regular	19,12	4,17	18,13
Ruim	16,18	0	15,11
Inexistente	25,88	75,00	29,12
Câmera Fotográfica			
Bom	52,21	100,00	55,49
Regular	11,21	0	10,44
Ruim	6,19	0	5,77
Inexistente	30,38	0	28,30
Televisão			
Bom	74,78	96,00	76,22
Regular	22,03	0	20,54
Ruim	2,61	0	2,43
Inexistente	0,58	4,00	0,81
Mimeógrafo			
Bom	38,82	20,83	37,64
Regular	19,12	4,17	18,13
Ruim	16,18	0	15,11
Inexistente	25,88	75,00	29,12
Câmera Fotográfica			
Bom	52,21	100,00	55,49
Regular	11,21	0	10,44
Ruim	6,19	0	5,77
Inexistente	30,38	0	28,30
Biblioteca			
Bom	32,54	88,00	36,39
Regular	22,39	4,00	21,11
Ruim	9,25	0	8,61
Inexistente	35,82	8,00	33,89

Fonte: Microdados do SAEB 2011, INEP (2013)

Tabela A.18: continuação — Distribuição de frequência da condição dos recursos da escola (em %)

Recursos da Escola	Pública n=346	Particular n=25	Total n=371
Quadra de Esportes			
Bom	19,47	87,50	23,97
Regular	19,17	4,17	18,18
Ruim	18,29	4,17	17,36
Inexistente	43,07	4,17	40,50
Laboratório			
Bom	20,83	76,00	24,65
Regular	5,65	0	5,26
Ruim	2,98	0	2,77
Inexistente	70,54	24,00	67,31
Auditório			
Bom	2,92	56,00	6,54
Regular	1,46	4,00	1,63
Ruim	1,46	0	1,36
Inexistente	94,15	40,00	90,46
Sala de Música			
Bom	3,22	58,33	6,83
Regular	1,46	4,17	1,64
Ruim	0,88	0	0,82
Inexistente	94,44	37,5	90,71
Sala de Artes			
Bom	7,04	70,83	11,23
Regular	2,93	4,17	3,01
Ruim	0,59	0	0,55
Inexistente	89,44	25,00	85,21
Sala de Leitura			
Bom	37,06	75,00	39,56
Regular	19,41	0	18,13
Ruim	10,29	0	9,62
Inexistente	33,24	25,00	32,69

Fonte: Microdados do SAEB 2011, INEP (2013)

Tabela A.19: Distribuição de frequência dos itens sobre a biblioteca e a sala de leitura da escola (em %)

Itens sobre a Biblioteca e a sala de Leitura	Pública n=346	Particular n=25	Total n=371
Acervo interessante aos alunos			
sim	72,75	96,00	74,37
não	27,25	4,00	25,63
Brinquedoteca			
sim	9,58	72,00	13,93
não	90,42	28,00	86,07
Espaço para estudo coletivo			
sim	40,67	92,00	44,32
não	59,33	8,00	55,68
Livros podem ser manuseados ou emprestados			
sim	84,98	96,00	85,75
não	15,02	4,00	14,25
Comunidade pode usar o espaço ou livros			
sim	33,53	54,17	34,92
não	66,47	45,83	65,08
Espaço arejado e bem iluminado			
sim	62,84	100,00	65,35
não	37,16	0	34,65
Existe um responsável pelo espaço			
sim	60,73	79,17	61,97
não	21,15	12,5	20,56
não existe biblioteca	18,13	8,33	17,46

Fonte: Microdados do SAEB 2011, INEP(2013)

Tabela A.20: Distribuição de frequência dos itens sobre empréstimo de livros na escola (em %)

Permissão	Pública n=346	Particular n=25	Total n=371
Alunos			
sim	85,4	100,00	86,43
não querem	2,54	0	2,36
a escola não permite	12,06	0	11,21
Professores			
sim	83,12	95,83	84,02
não querem	8,92	4,17	8,58
a escola não permite	7,96	0	7,40

Fonte: Microdados do SAEB 2011, INEP (2013)