

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Curso de Especialização em Estatística

Júlia Drumond Campos e Silva

**PROFICIÊNCIA EM MATEMÁTICA DE ALUNOS NO PROEB:
análise exploratória por meio de modelos de regressão linear**

Belo Horizonte
2014

Júlia Drumond Campos e Silva

**PROFICIÊNCIA EM MATEMÁTICA DE ALUNOS NO PROEB:
análise exploratória por meio de modelos de regressão linear**

Monografia apresentada ao curso de Pós-Graduação em Estatística da Universidade Federal de Minas Gerais, como requisito parcial para a obtenção do título de Especialista em Estatística.

Orientadora: Ilka Afonso Reis

Belo Horizonte

2014

RESUMO

Este trabalho realizou uma análise exploratória dos dados relacionados à proficiência dos alunos em Matemática no Programa de Avaliação da Rede Pública de Educação Básica (Proeb). Para tanto, utilizou Modelos de Regressão Linear, elaborados de forma construtiva, com a análise da entrada de variáveis candidatas, uma a uma, até a composição do modelo final. Os dados selecionados para a análise são referentes aos alunos do nono ano do ensino fundamental que realizaram as provas do Proeb de Língua Portuguesa e Matemática no ano de 2013 e responderam a todas as perguntas elencadas no questionário contextual, aplicado na ocasião dos testes. Além dos resultados das proficiências e dos questionários, foi utilizado o Índice Socioeconômico para a construção dos modelos. Também foram consideradas para a análise as premissas da literatura, que embasaram os resultados observados. Verificou-se que os fatores mais relacionados às proficiências dos alunos em Matemática, de acordo com o estudo proposto, foram a Proficiência em Língua Portuguesa, a expectativa de continuidade dos estudos após a conclusão do ensino médio, a não procura por emprego, a expectativa de formação no ensino médio, a idade adequada para a etapa de escolaridade, a autoestima, a autoavaliação do aluno, a escolaridade da mãe e a variável relacionada ao motivo.

Palavras-chave: Regressão linear. Proficiência. Avaliação Educacional.

LISTA DE GRÁFICOS, TABELAS E QUADROS

GRÁFICO 1: Percentual de alunos, por rede de ensino e ano de escolaridade, no nível recomendado de desempenho nas avaliações de Língua Portuguesa e Matemática do ano de 2013.	15
GRÁFICO 2: Percentual de alunos, por rede de ensino, no nível recomendado de desempenho em Matemática, de 2006 a 2013.	16
GRÁFICO 3: Relação entre as proficiências de Língua Portuguesa e Matemática.....	23
GRÁFICO 4: Relação entre a proficiência de Matemática e o valor do Índice Socioeconômico.....	24
TABELA 1: Número de alunos, por dependência administrativa, que constam no Censo Escolar e na base selecionada, na rede estadual e municipal de ensino.....	20
TABELA 2: Percentual de alunos, por turno, no 9º ano em Belo Horizonte que constam na base selecionada e no Censo Escolar.....	21
TABELA 3: Estatísticas descritivas da proficiência em Língua Portuguesa e Matemática.....	21
TABELA 6: Comparação da diferença entre grupos e intra grupos (ANOVA) do Índice Socioeconômico das escolas.	24
TABELA 7: Maiores valores R^2 ajustados em modelos de regressão linear cuja resposta é o PROF_MT e a explicativa é a variável investigada.	27
QUADRO 1: Saída do MINITAB para o ajuste do Modelo 1.....	25
QUADRO 2: Saída do MINITAB para o ajuste do Modelo 2.....	26
QUADRO 3: Saída do MINITAB para o ajuste do Modelo 3.....	28
QUADRO 4: Saída do MINITAB para o ajuste do Modelo 4.....	29
QUADRO 5: Saída do MINITAB para o ajuste do Modelo 5.....	30
QUADRO 6: Saída do MINITAB para o ajuste do Modelo 6.....	31
QUADRO 7: Saída do MINITAB para o ajuste do Modelo 7.....	32
QUADRO 8: Saída do MINITAB para o ajuste do Modelo 8.....	33
QUADRO 9: Saída do MINITAB para o ajuste do Modelo 9.....	35
QUADRO 10: Saída do MINITAB para o ajuste do Modelo 10.....	36

LISTA DE ABREVIATURAS E SIGLAS

AUTO_AV: Variável que indica autoavaliação do estudante (escala de 0 a 10).

AUTO_EST: Variável que indica a autoestima do estudante (escala de 0 a 10).

CONTI: Variável que indica que o aluno tem a expectativa de dar continuidade aos estudos.

FOR_EM: Variável que indica que o aluno pretende se formar no Ensino Médio.

ISE: Variável que indica o valor do Índice Socioeconômico.

ID_ADEQ: Variável que indica se o aluno está na idade adequada para a série avaliada.

MAE_EM: Variável que indica que a mãe do aluno chegou ao Ensino Médio.

MAE_SUP: Variável que indica que a mãe do aluno chegou ao Ensino Superior.

MOTIVO: Variável que indica que o aluno discorda da frase “vou à escola somente porque sou obrigado”.

NE_CONTI: Variável que indica que o aluno não especificou a expectativa de continuidade dos estudos.

NP_EMPREGO: Variável que indica que o aluno não está procurando emprego.

PROEB: Programa de Avaliação da Rede Pública de Educação Básica.

PROF_LP: Variável que indica o valor da Proficiência de Língua Portuguesa.

PROF_MT: Variável que indica o valor da Proficiência de Matemática.

SUMÁRIO

1 INTRODUÇÃO	4
2 OBJETIVO	6
3 REVISÃO DA LITERATURA	7
4 MATERIAIS E MÉTODO	14
4.1 O banco de dados	14
4.2 Análise Estatística dos Dados	19
5 RESULTADOS	20
6 CONCLUSÃO	38
REFERÊNCIAS	41
APÊNDICE 1	43
APÊNDICE 2	54
ANEXO	55

1 INTRODUÇÃO

A educação é uma garantia constitucional que visa ao desenvolvimento pleno da pessoa. De acordo com o texto da Constituição Federal, em seu artigo 205, é direito de todos, assegurando o preparo para o exercício da cidadania e para o trabalho. Sua salvaguarda perpassa pela garantia material e não meramente formal. Não basta que a pessoa esteja inserida em um processo de escolarização, mas que esse processo seja capaz de garantir-lhe a aquisição de conhecimentos, competências e habilidades que atuarão como insumos, de modo direto ou indireto, para o gozo dos demais direitos.

O desempenho escolar é uma medida que visa a auferir se os alunos adquiriram os saberes e habilidades esperados para determinadas faixas etárias e níveis de escolaridade. Tal medida pode ser obtida por meio de avaliações externas, ou de larga escala, as quais consistem na aplicação de testes padronizados aos discentes e cujos resultados servem para gerar subsídios ao acompanhamento das políticas educacionais.

O estudo sobre os fatores potencialmente associados ao desempenho escolar faz-se necessário não só para ampliar o entendimento acerca do desenvolvimento cognitivo dos alunos, mas também para informar a maneira como as políticas educacionais podem contribuir mais efetivamente para esse progresso. Isso é especialmente importante em um contexto onde Administração Pública exerce seu dever e preza pelo bom resultado do investimento que, nesse caso, é a garantia do direito ao aprendizado.

Muitos autores se dedicaram ao tema do desempenho escolar e os fatores potencialmente relacionados a ele. Existe um extenso debate que, simplifiadamente, está dividido entre aqueles que focalizam as características do aluno, como o contexto socioeconômico em que vivem e outros que, controlado esse fator, se concentram no estudo do efeito que a escola pode ter para o desenvolvimento da aprendizagem.

As avaliações de larga escala são interessantes para esse debate, pois geram um elevado número de dados que podem gerar mais informação do que, unicamente, a proficiência obtida nos testes. Isso porque podem ser associados a outras medidas, com fontes externas aos testes (Índice de Desenvolvimento

Humano da região, investimentos em educação, etc.); ou internas, coletadas diretamente na aplicação deles, por meio de questionários contextuais.

Na busca por contribuir para esse debate, este trabalho intenta a realização de uma análise exploratória dos dados gerados pelo Programa de Avaliação da Rede Pública de Educação Básica – PROEB. Tais resultados são analisados conjuntamente com os questionários socioeconômicos, respondidos pelos alunos na hora dos exames. A pergunta de pesquisa está voltada para os fatores ali mensurados que podem ajudar na compreensão dos dados da proficiência.

Com essa finalidade, estruturou-se o trabalho da seguinte forma: Em primeiro lugar, é explicitada a proposição do trabalho, com o objetivo geral e os específicos. A fim de ampliar o entendimento sobre a questão proposta, recorre-se à literatura sobre o tema, na seção subsequente. A seguir, a metodologia é elucidada, atentando-se ali para as características da fonte geradora de dados e formas de seleção da base.

Uma vez expostos os objetivos e métodos, parte-se para o detalhamento dos resultados, os quais são incorporados passo a passo até a construção do modelo final. Segue-se a ele, a discussão à luz da literatura, que permite que sejam tecidas conclusões a respeito da análise exploratória de dados e seus resultados, essas elencadas em um capítulo à parte.

2 OBJETIVO

Realizar uma análise exploratória dos dados de alunos no Programa de Avaliação da Rede Pública de Educação Básica, por meio de modelos de regressão linear, a fim de identificar fatores possivelmente associados ao desempenho do aluno na prova de Matemática.

3 REVISÃO DA LITERATURA

A literatura acerca do desempenho escolar é ampla e historicamente debatida. Trabalhos dessa área de pesquisa apontam vários aspectos que influenciam o desempenho escolar. Entre os de destaque, têm-se fatores os associados aos alunos, como os socioeconômicos, e às escolas, como as práticas de gestão escolar e pedagógicas (BARROS *et al.*, 2001; LUZ, 2006; SOARES, 2004).

Segundo Andrade e Soares (2008), a proficiência dos alunos, mensurada em testes padronizados, estaria relacionada, por um lado, aos antecedentes sociodemográficos, opções pessoais, valores e estrutura da família e da sociedade em que os alunos vivem, e, por outro, também associadas à escola em que estudam. Sobre esse último aspecto, a literatura investiga o fato de algumas escolas parecerem mais capazes do que outras em promover o sucesso acadêmico dos seus educandos.

Vários estudos apontam que o desempenho escolar está associado a fatores relacionados aos alunos. (COLEMAN *et al.*, 1966; JENCKS, 1972; BARROS *et al.*, 2001; SOARES, 2008). O estudo considerado um divisor de águas do campo de pesquisas em educação foi o intitulado Relatório Coleman, de 1966, que fomentou a discussão sobre o tema à época, abrindo caminho para debates e estudos subsequentes. (MOSTELLER; MOYNIHAN, 1972, in BROOKE, 2008).

O trabalho de Coleman e colaboradores analisou associações entre o desempenho e diversos fatores como instalações, currículos, métodos de avaliação e atividades extracurriculares. Tal comparação também foi realizada em relação ao *background* dos responsáveis pelo aprendizado na escola, dos companheiros de turma e da família. Ademais, foram consideradas dimensões como a motivação dos alunos e confiança sobre o próprio futuro. (COLEMAN *et al.*, 1966, in BROOKE, 2008).

O Relatório Coleman apresentava, em seu norte, o conceito de igualdade de oportunidades educacionais, entendido como a capacidade das escolas em diminuir o *gap* de desempenho entre alunos de diferentes origens sociais. Entre as conclusões do estudo, foi verificado que a proficiência dos alunos cujos pais haviam cursado a faculdade era maior do que a daqueles com pais de menor escolaridade.

(COLEMAN *et al.*, 1966, in BROOKE, 2008). Além disso, mostrou que as escolas analisadas eram mais homogêneas internamente do que externamente. Ou seja, que docentes e discentes de uma mesma instituição de ensino tinham origens sociais muito parecidas. Segundo o estudo, isso fazia com que as diferenças de desempenho entre os grupos sociais tivessem maior propensão a se manter ao longo da vida acadêmica dos alunos.

Além de relacionado às características de *background* familiar, a análise indicava que o desempenho educacional estaria associado às aspirações dos alunos e dos colegas na escola. Verificou-se, por exemplo, que crianças de baixo *background* que conviviam com outras de nível social mais elevado melhoravam o desempenho mais do que se estivessem em escolas compostas por discentes de menor *background*.

Contudo, as conclusões do estudo geraram o cerne do debate que o seguiu, já que os resultados pareciam mostrar que as escolas teriam falhado em promover a isonomia entre discentes de diferentes *backgrounds*. Alguns enfatizavam a incapacidade desses locais em dirimir a desigualdade inicial. A respeito do Relatório Coleman, Mosteller e Moynihan (1972) apud Brooke (2001) exemplificavam que, analogamente a uma corrida, os alunos que “largavam na frente” apresentavam maior probabilidade de vencer, ou, em outras palavras, de alcançar um desempenho mais elevado. Não se poderia, equivocadamente, considerar que as escolas não faziam diferença no aprendizado dos alunos, já que todos os agregados sociais estudados apresentaram melhoras importantes no aprendizado. O que ocorria era que as diferenças entre os patamares iniciais não eram suprimidas.

Jenks (1973) apud Brooke (2008), em estudo posterior, analisava a correlação entre o desempenho e variáveis dos alunos e suas famílias. Segundo o autor, várias hipóteses poderiam ser tecidas a respeito das correlações verificadas. Entre elas, a de que filhos de pais do topo da pirâmide social tinham uma expectativa de continuidade dos estudos mais elevada do que aqueles de condição econômica inferior. Outra era a de que crianças de classe econômica mais alta tinham maiores probabilidades de conviver em um ambiente propício à aprendizagem, pois, por exemplo, não precisariam trabalhar durante a vida escolar.

Como resultados de sua pesquisa, Jenks (1973) afirmava três fatores importantes: igualmente a estudos anteriores como o de Coleman, a existência de

associação entre escolaridade e origem socioeconômica. Em seguida, que o sucesso acadêmico pode ser dividido entre condição financeira e aptidão acadêmica. Finalmente, que a motivação do corpo discente apresentou uma influência ainda maior que os demais fatores estudados no desempenho acadêmico.

Outro estudo que ganhou repercussão na mesma época foi o Relatório Plowden, que descrevia o estado da arte das pesquisas sobre a qualidade da educação básica. Dentre as conclusões, mostrava que a atitude dos pais, como o interesse pelo desempenho escolar dos filhos e o seu acompanhamento, estava muito relacionado a ele. Também constatava que, embora menores, as condições de domicílios dos alunos e instalações das escolas tinham efeitos importantes. Mas, em termos comparativos, os fatores relacionados às escolas não eram os principais preditores do desempenho.

Segundo Buchmann e Hannum (2001), o debate até então estava limitado a países industrializados até Heyneman divulgar os resultados de suas pesquisas, em 1976, no chamado "*Coleman Report for a developing country*". Heyneman verificou que os fatores escolares atuavam fortemente na determinação dos resultados em Uganda, na África. A partir disso, analisou a situação para países em desenvolvimento, em geral. Segundo Buchman e Hannum (2001), no estudo de Heyneman, conclui-se que, nesses países, o *background* familiar atuava menos fortemente do que a qualidade das escolas no desempenho dos alunos – comparativamente aos países desenvolvidos. Após a divulgação desses resultados, outros estudos começaram a surgir em países em desenvolvimento na tentativa de verificar a influência tanto de fatores de *background* familiar quanto de escolares na proficiência do corpo discente. (BUCHMANN e HANNUM, 2001).

No Brasil, em linhas de pesquisa mais recentes, Barros e colaboradores (2001) analisaram pessoas de onze a vinte e cinco anos residentes de zonas urbanas do nordeste e sudeste do país. O estudo partiu de variáveis como disponibilidade e qualidade de serviços educacionais, recursos familiares financeiros e não financeiros e recursos da comunidade. Intencionando verificar, entre esses, os principais determinantes do sucesso escolar, o estudo concluiu que a escolaridade da mãe foi o mais importante, o que não era um achado novo, pois pesquisas anteriores demonstravam tal influência. Além disso, a pesquisa indicou que, apesar da renda per capita ser amplamente estudada pela literatura, não se mostrou tão

significativa quanto a escolaridade em questão. Com a mesma linha de pesquisa, Luz (2006) mostra que o aprendizado cresce em mesma tendência que a escolaridade da mãe. Assim também, Louzano (2012) identifica que a probabilidade de fracasso escolar diminui com o incremento da escolaridade dos pais, bem como a chance de repetência e abandono.

Para Rodrigues, Rios-Neto e Pinto (2011), quanto maior o capital social, econômico e cultural dos pais, maiores as chances de que motivem, valorizem e incentivem os filhos aos estudos, o que pode ser influente para a agregação de novos saberes pelos alunos. Segundo Faria e Madalozzo (2012), crianças de menor condição econômica e com pais de menor escolaridade tenderiam a ter menos estímulos em casa, menos acesso a materiais propícios e exposição a um vocabulário menos abrangente. A esse respeito, Bordieu e Passeron, citados por Charlot (2007), apresentam o conceito de arbítrio cultural, no qual as famílias transmitem às crianças tanto bens e serviços, quanto relações mantidas pela família e o domínio da língua e relações com a cultura.

O arcabouço teórico acerca desses fatores é extenso. Sobre isso, Varani e Silva fazem a seguinte ressalva:

Embora a família seja fundamental no processo de desenvolvimento integral das crianças, ela não pode assumir sozinha a culpa pelo sucesso ou pelo fracasso escolar dos alunos, pois o bom ou o mau desempenho escolar não depende exclusivamente da participação/presença ou não da família na escola. Outros inúmeros fatores (sociais, políticos, econômicos e culturais) influem no desempenho. (VARANI; SILVA, 2010, p.524).

Estudos como o de Soares e Colhares (2006) afirmam que o sucesso escolar dos discentes parece estar associado às oportunidades que são oferecidas pela família e pela sociedade como um todo. Nesse sentido, para Arroyo (2000), o “fracasso escolar é uma expressão do fracasso social, dos complexos processos de reprodução da lógica e da política de exclusão que perpassa todas as instituições sociais e políticas”.

Tida como consensual na literatura, a influência das características dos alunos e suas famílias no sucesso acadêmico dos alunos é um dos fatores que o explica, mas não o único. Já à época de Coleman, pesquisas demonstravam que o efeito da escola, embora menor, não podia ser desconsiderado. Madaus e colaboradores (1980) *apud* Brooke (2008), por exemplo, apontavam que o fato de não conseguirem superar as diferenças iniciais entre o aprendizado dos alunos em

diferentes grupos não era o mesmo que dizer que escolas não faziam diferença. Outras variáveis, não medidas em testes padronizados, poderiam ser listadas entre a ampla gama de resultados desejáveis, como a redução da violência, a felicidade dos indivíduos e o convívio social.

Ademais, Madaus e colaboradores destacaram o assunto da qualidade do corpo docente. Segundo eles, vários estudos apontavam para a correlação, mesmo não muito forte, entre esse fator e aprendizagem dos alunos. Mais do que isso, afirmavam que variáveis para além da titulação acadêmica deveriam ser consideradas. Hanushek (1970) *apud* Brooke (2008) indicava a importância da fluência verbal e do tempo de formação dos professores nos resultados dos alunos. Sammons (1999) endossava que, controladas variáveis de desempenho inicial e socioeconômicas, era possível encontrar uma significativa relação entre a qualidade das escolas e a proficiência em leitura e matemática.

As críticas sobre a utilização de avaliações de larga escala para esse tipo de pesquisa ocorrem porque, muitas vezes, essas estão associadas a variáveis de status e de não processo. Talvez, por isso, não consigam identificar mais fortemente fatores relacionados às escolas. Madaus (1980) *apud* Brooke (2008) destacava que, quando isso é feito, a escola é tida como uma “caixa preta”, apenas sendo possível identificar o que está por fora dela. Analisar variáveis de processo, ou seja, o que está “dentro da caixa”, poderia representar mais fielmente a realidade.

Reforçando esse argumento, Rutter e colaboradores (1979) *apud* Brooke (2008) afirmavam que dados oriundos de testes padronizados diminuía a percepção sobre o efeito das escolas. Um dos motivos era que desconsideravam atitudes, valores e a vida interna dessas instituições. Outro, simplesmente porque a falta de correlação poderia significar que a variável utilizada não era tão relevante, mas a extrapolação da análise poderia implicar em um erro metodológico.

Diante das críticas a modelos que não captavam o efeito da escola, começaram a surgir vários estudos sobre a eficácia escolar e os fatores relacionados a ela. Sammons (1999) *apud* Brooke (2008) mostra que, apesar de não ter existido consenso sobre o que constituiria uma escola eficaz até o final dos anos 80, alguns fatores podiam ser listados como liderança profissional, objetivos e visões compartilhadas, incentivo positivo, altas expectativas, organização voltada para aprendizagem, entre outros.

Ainda na década de oitenta, Madaus e colaboradores citavam um estudo irlandês que avaliou variáveis de processo escolar, como pressões acadêmicas, administração de sala de aula e preocupação dos alunos com valores acadêmicos. As conclusões indicavam que as variáveis que mais se correlacionavam com o desempenho eram a estimativa dos professores sobre quanto a turma se adéqua à pressão acadêmica e, de outro lado, como os alunos percebem o que os professores esperam deles no que se refere a essa pressão. Ademais, destacava o nível de disciplina na escola e a quantidade de estudo, as duas positivamente correlacionadas.

Outros pesquisadores, ao invés de balizarem seus estudos em variáveis de processo para mensurar efeitos relacionados às escolas, se ativeram em utilizar outras metodologias, como os modelos hierárquicos. Assim fez Lee (2000), que, por meio desse tipo de análise, mensurou o efeito do tamanho das escolas no aprendizado dos alunos.

Sobre a consolidação dos estudos empíricos do efeito escola, Riddel (1995), citado por Sammons (1999), afirma que esses trabalhos têm sido notórios também em países em desenvolvimento. A constatação considera o fato de cada localidade possuir especificidades que moldam a sociedade de maneira suficientemente forte para atuar nos determinantes do resultado cognitivo. A exemplo disso, Albernaz, Ferreira e Franco (2002) informam que, no Brasil, a infraestrutura escolar e qualificação docente, diferentemente de outros países, podem afetar o desempenho escolar de forma significativa. Sant'Anna (2009) declara-se a favor da democratização de oportunidades dadas pela escola (e seus efeitos) para dirimir as desigualdades sociais do país.

Com amostras de escolas públicas no Brasil e por meio de entrevistas e testes padronizados aos alunos, Barbosa (2001) analisou um conjunto de itens determinantes para a avaliação da qualidade da escola, como a expectativa do professor e demais profissionais em relação ao desempenho dos discentes, objetivos disseminados no ambiente escolar, formas de acompanhamento da aprendizagem dos alunos, dentre outros. Concluiu que, nas escolas consideradas de boa qualidade, sob a ótica dos itens analisados, os alunos possuíam desempenho superior nos testes. Segundo Soares e colaboradores (2010), “a expectativa do professor provoca um impacto positivo na proficiência do aluno, mesmo

considerando-se o efeito de variáveis sociodemográficas tradicionalmente associadas ao desempenho”.

Em relação aos efeitos da entrada tardia na escola, reprovação ou abandono e posterior reingresso, Ferrão, Beltrão e Santos (2002) chamam a atenção para o acarretamento em defasagem idade-série. Jesus e Laros (2004), controlando a condição socioeconômica dos discentes, constataram que o atraso escolar era o fator que mais contribuía negativamente para o desempenho. No estudo de Riani e Rios-Neto (2008) sobre o tema, conclui-se que a qualidade da instituição de ensino é importante para aumentar a chance de o aluno frequentar a escola na idade correta.

Como visto, a literatura sobre os fatores relacionados ao desempenho abarca uma série de variáveis, relacionadas tanto ao aluno como ao processo escolar. Luz (2006), afirma que essa literatura:

reconhece o impacto da origem sócio-econômica na performance educacional do indivíduo, e enxerga nessa relação um caminho provável para a reprodução de desigualdades sociais. Entretanto, o sistema de ensino não é encarado como mero instrumento de endosso dessas determinações. Ao contrário, há um intenso debate sobre a atuação dos fatores escolares no desempenho do aluno, assim como o espaço reservado às políticas educacionais na democratização das oportunidades provenientes de uma formação educacional de qualidade. (LUZ, 2006, p.3)

Endossando esse argumento, Faria e Madalozzo (2012) afirmam que é possível superar condições adversas e atingir níveis elevados de desempenho acadêmico. Porém, para que isso ocorra, é necessário que condições adequadas à aprendizagem sejam asseguradas, possibilitando que os alunos “adquiram competências importantes, sejam capazes de desenvolver seu potencial e de ascender socialmente”. Nesse sentido, o estudo sobre o complexo rol de variáveis associadas ao desempenho, especialmente sobre os alunos, não é tido aqui como um fator de determinismo social. Antes, intenta contribuir para o debate, para que o processo de ensino aprendizagem seja melhor compreendido, e assim, o alcance de seus objetivos se torne mais factível.

4 MATERIAIS E MÉTODO

4.1 O banco de dados

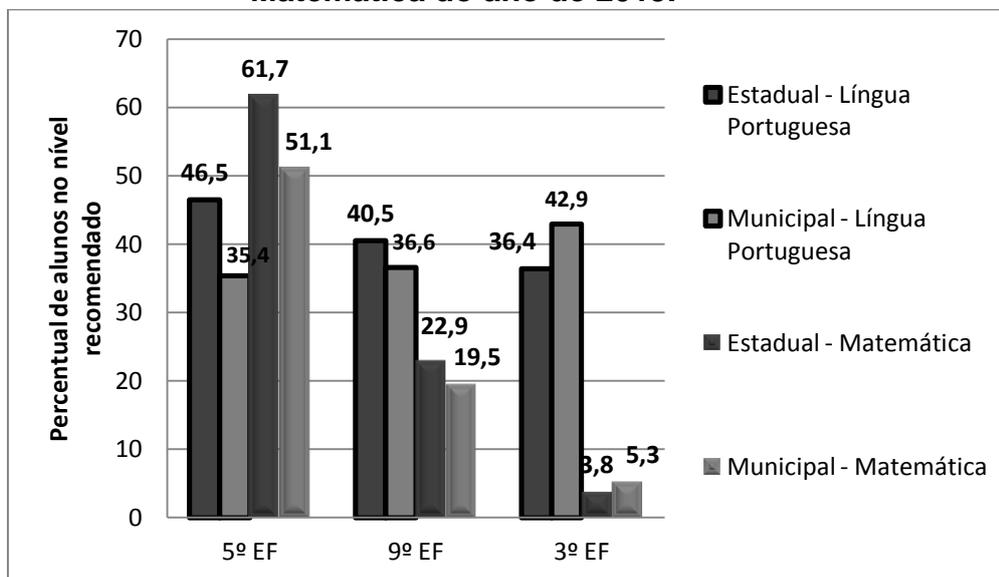
O Programa de Avaliação da Rede Pública de Educação Básica (Proeb) está inserido no contexto do Sistema Mineiro de Avaliação Educacional (Simave), que conta também com o Programa de Avaliação da Aprendizagem (Proalfa) e o Programa de Avaliação da Aprendizagem (Paae). O primeiro a ser criado dentro do referido Sistema foi o Proeb, constituindo-se como uma avaliação em larga escala da rede pública mineira desde os anos 2000. Esse Programa avalia estudantes do 5º e 9º ano do ensino fundamental e 3º ano do ensino médio em escolas municipais e estaduais, nas disciplinas de Língua Portuguesa e Matemática.

Os resultados obtidos por meio das avaliações são organizados em Padrões de Desempenho, os quais representam uma medida de alcance dos objetivos educacionais estabelecidos em uma Matriz de Referência. Essa é formada por uma gama de habilidades previstas nas diretrizes curriculares do estado, que dão suporte aos testes. Tais habilidades são selecionadas devido à importância curricular em determinada etapa avaliada e, também, por serem passíveis de medição em testes padronizados. Assim, os Padrões de Desempenho estabelecem metas para os alunos e, em cada um desses Padrões, são esperadas competências e habilidades cognitivas a serem alcançadas por eles.

Por competências, entende-se o grupo de habilidades que, em conjunto, contribui para o alcance de determinado resultado cognitivo, enquanto que habilidades são tidas como a capacidade de “saber fazer” algo. A proficiência dos alunos é, então, agrupada em três Padrões de Desempenho: Baixo, Intermediário e Recomendado. Dessa forma, é factível interpretar pedagogicamente quais são as habilidades desenvolvidas pelos educandos, uma vez que os localiza dentro de cada Padrão. Com isso, é possível não só calcular a proficiência média obtida, mas também o percentual de alunos por Padrão de Desempenho.

O gráfico a seguir apresenta o percentual de alunos no nível recomendado na avaliação de 2013, tanto em Língua Portuguesa quanto em Matemática, nas três etapas avaliadas.

GRÁFICO 1: Percentual de alunos, por rede de ensino e ano de escolaridade, no nível recomendado de desempenho nas avaliações de Língua Portuguesa e Matemática do ano de 2013.

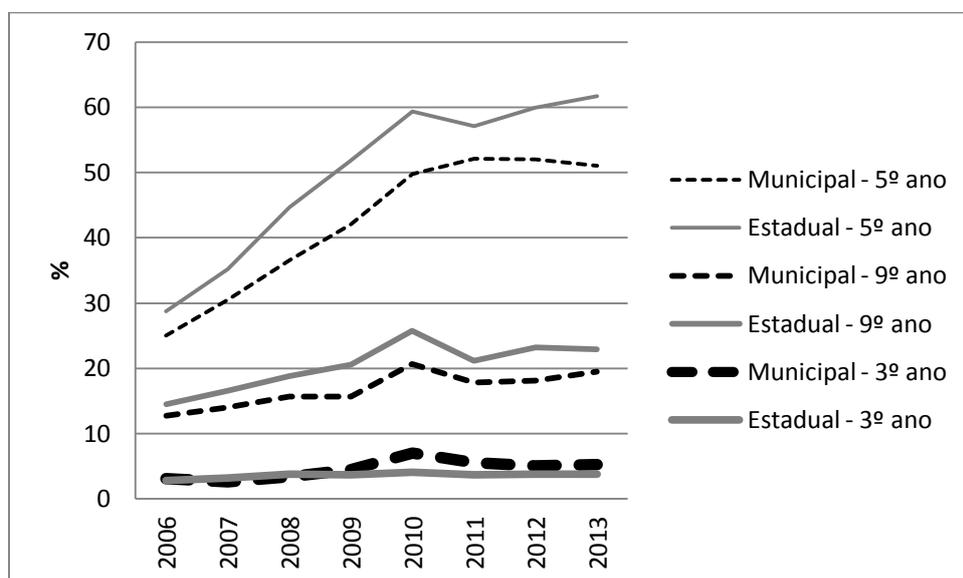


Fonte: Superintendência de Avaliação Educacional/SEEMG. Tabulação própria.

O Gráfico 1 mostra uma inversão entre os percentuais de alunos no nível recomendado de matemática nas etapas avaliadas, para ambas as redes de ensino: quanto maior a etapa de escolaridade, menor o percentual de alunos no nível desejável. No quinto ano do ensino fundamental, 61,7% dos alunos estão no nível recomendado de Matemática na rede estadual e 51,1 % alcançaram esse nível na rede municipal. No nono ano, o percentual decresce para 22,9% na rede estadual e para 19,5% na municipal. No terceiro ano do ensino médio, 5% ou menos dos alunos alcançaram o nível recomendado – ou seja, adquiriram as competências e habilidades esperadas para a última série da educação básica: 5,3% na municipal e 3,8% na estadual.

O cenário para as disciplinas de Língua Portuguesa e Matemática precisa ser interpretado, a fim de moldar ações que possam contribuir para a melhoria do desempenho. Entretanto, é para a disciplina de Matemática que se requer maior atenção, dados os percentuais alcançados. Por esse motivo, optou-se por utilizar a proficiência de matemática como variável resposta à investigação proposta. Como pode ser observado no gráfico a seguir, o patamar de desempenho em Matemática alcançado pelos alunos em 2013 não é atípico, mostrando certo padrão ao longo dos últimos anos, tanto na rede estadual quanto na municipal.

GRÁFICO 2: Percentual de alunos, por rede de ensino, no nível recomendado de desempenho em Matemática, de 2006 a 2013.



Fonte: Superintendência de Avaliação Educacional/SEEMG. Tabulação própria.

A análise do Gráfico 2 mostra que a evolução do desempenho em Matemática para as duas redes avaliadas foi similar entre 2006 e 2013. Por esse motivo, optou-se por utilizar dados de alunos das duas redes, indistintamente, na variável resposta. De 2006 a 2010, houve um crescimento mais acentuado no desempenho para o 5º ano, um decréscimo entre 2010 e 2011, tanto para o 5º quanto para o 9º, e uma estabilização dos resultados do 3º ano durante todo o período analisado. A comparação entre as etapas avaliadas evidencia, novamente, maiores percentuais para o 5º ano, intermediários para o 9º e menores para o 3º. Essa lógica é mantida em todo o período analisado.

Optou-se, então, por utilizar os resultados do 9º ano como recorte da variável resposta. Isso porque a etapa, intermediária entre as avaliadas, também apresenta, historicamente, resultados intermediários. Ademais, representa o final de um ciclo, o ensino fundamental, e transição para um novo, o ensino médio. Sendo assim, é uma etapa estratégica para a investigação sobre os fatores relacionados ao desempenho dos alunos, já que, antes dela, mais da metade dos estudantes dos anos iniciais do ensino fundamental alcançam o desempenho recomendado, enquanto que, ao final da educação básica, esse desempenho é alcançado por apenas cinco por cento ou menos dos discentes.

O estudo sobre os componentes da avaliação, bem como a análise dos resultados alcançados pelos alunos, delineou três estratégias iniciais para o desenho da exploração de dados. Em primeiro lugar, o uso do resultado de matemática como variável resposta, uma vez que o desempenho decresce ao longo dos anos de escolaridade. Em segundo, o uso indistinto de dados de alunos das redes municipais e estadual, observados certos padrões entre elas. Por fim, a utilização do recorte do 9º ano, como uma etapa de transição entre as avaliadas.

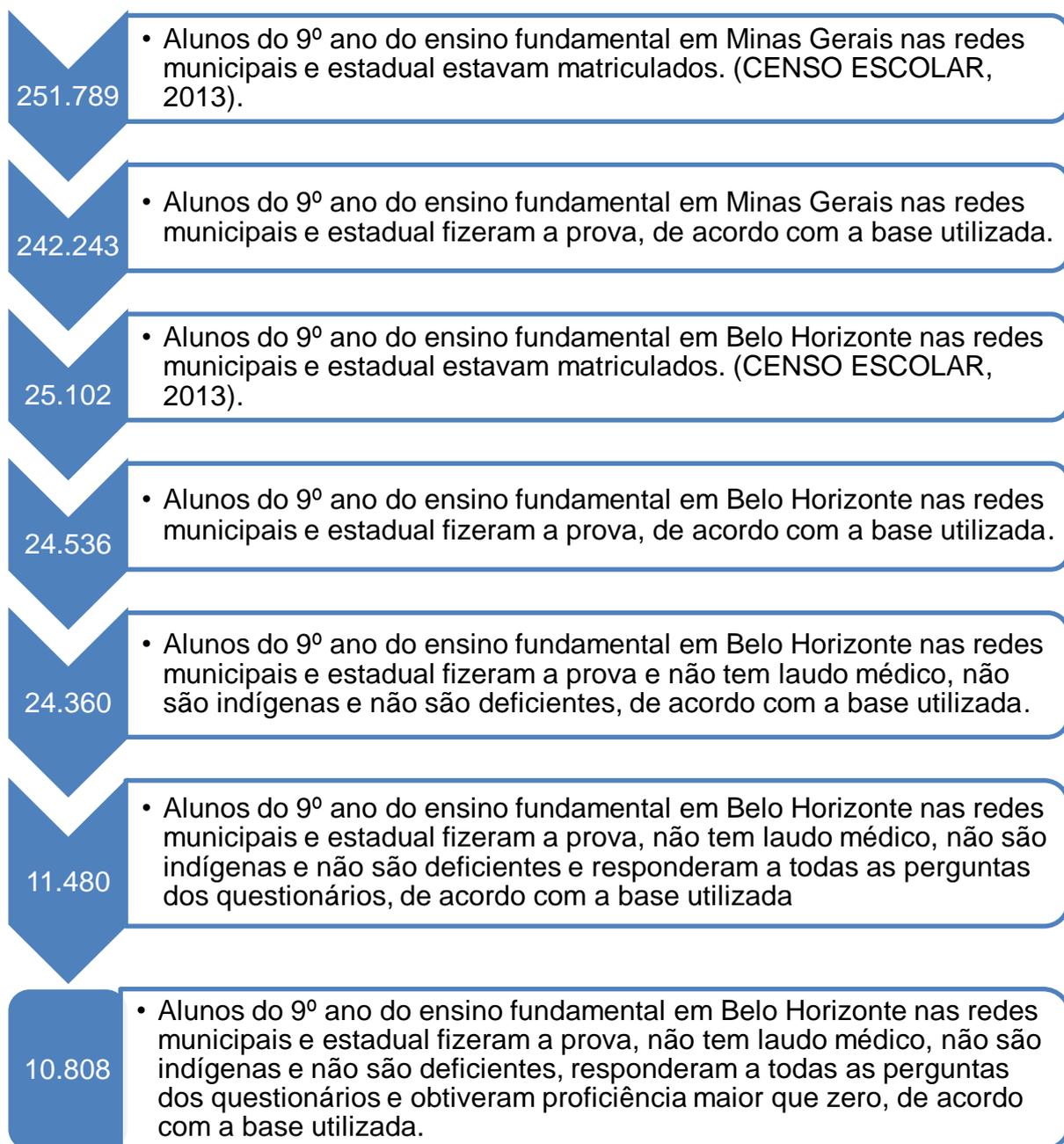
Uma vez definidas a disciplina, a série e as redes a serem utilizadas como variável resposta, partiu-se para a seleção da base de dados. Optou-se por utilizar a base de dados mais recente, referente às avaliações de 2013. As avaliações de 2014 do Proeb ainda não haviam sido realizadas. Outrossim, os dados referentes ao ano de aplicação só são divulgados no ano seguinte, pois passam por compilação e conferência em grande escala, após a realização dos testes, que normalmente são aplicados em outubro.

Os dados de 2013, disponibilizados em julho de 2014, estavam organizados com proficiências de Língua Portuguesa e Matemática, por aluno, em duas bases, uma com os resultados do terceiro ano do ensino médio e outra com os do nono e quinto do ensino fundamental. Em ambas as bases, para cada linha, que representava um aluno, havia, nas colunas, a proficiência obtida nas duas disciplinas avaliadas, bem como as respostas ao questionário contextual. O formulário utilizado para coletar tais respostas está disposto no ANEXO 1.

Adotou-se o software estatístico *IBM SPSS* para selecionar os casos de alunos do 9º ano. Em seguida, optou-se pelo enfoque nos alunos de Belo Horizonte, a fim de reduzir a variabilidade da resposta. No grupo selecionado, foram desconsiderados os resultados de alunos indígenas, com laudo médico e deficientes. Isso para se trabalhar apenas os perfis mais comuns de alunos da rede pública de ensino. Por fim, foram consideradas somente as proficiências de alunos que responderam a todas as perguntas e obtiveram proficiência maior que zero, a fim de simplificar o relacionamento entre as respostas aos questionários contextuais e as proficiências obtidas.

O afinamento da base de dados, até a seleção final, pode ser visualizado na Figura 1, que considera também números informados no Censo Escolar.

Figura 1: Seleção dos casos para a composição do banco de dados



4.2 Análise Estatística dos Dados

Selecionado o grupo de alunos a ser trabalhado, partiu-se para a construção dos modelos de regressão. Como primeiros passos, estudou-se a correlação entre as proficiências de Língua Portuguesa e Matemática, utilizando-se o coeficiente de correlação linear de Pearson e também os gráficos de dispersão. Além disso, considerou-se o índice socioeconômico como variável explicativa *a priori* no modelo, devido ao destaque dado pela literatura sobre o assunto e para que funcionasse como controle do efeito escola, visto que esse não foi captado diretamente pela metodologia utilizada¹.

Para a seleção das variáveis candidatas a entrar no modelo de regressão, utilizou-se os coeficientes de determinação (R^2) como critério de classificação, obtidos pelo ajuste de modelos regressão linear univariada cuja resposta, para todos eles, era a proficiência de Matemática e a única explicativa era a variável investigada. Como pode ser notado no questionário (ANEXO 1), a maioria das 58 perguntas gerou observações de variáveis do tipo categórica. Para essas variáveis, antes de gerar o R^2 no modelo univariado, foram geradas variáveis indicadoras (*dummies*), as quais, em alguns casos, agrupavam respostas possíveis em um rol menor de possibilidades, a fim de simplificar o modelo.

O software utilizado para a geração dos modelos e análises foi o MINITAB, versão 16. A abordagem utilizada para a construção do modelo de regressão linear final foi a construtiva, no qual, a cada passo, uma variável era adicionada ao modelo e tinha sua contribuição para o mesmo avaliada. A análise de resíduos para cada um dos modelos gerados consta no Apêndice. A próxima seção, dedicada aos resultados obtidos, detalha os passos até a composição do modelo de regressão final.

¹ O efeito é mais apropriadamente captado por meio de modelos de regressão hierárquicos, que não foram utilizados neste trabalho.

5 RESULTADOS

Esta seção é dedicada à apresentação da análise exploratória dos dados dos alunos do Proeb, por meio de modelos de regressão linear. Conforme explicitado nos objetivos propostos, pretende-se identificar fatores possivelmente associados ao desempenho dos alunos na disciplina de Matemática. Intenta-se assinalar, a partir dos questionários contextuais, quais os fatores mais relacionados à proficiência, selecionar as variáveis para a composição dos modelos, apontar, entre eles, os mais capazes de explicar a variabilidade da resposta e eleger um modelo final.

Para o cumprimento desses objetivos, foram utilizados os dados do Proeb 2013, cujo banco foi selecionado conforme exposto na metodologia. A partir da seleção, o modelo foi sendo moldado de forma construtiva, com a avaliação da entrada de variáveis uma a uma.

Antes do detalhamento dos passos da regressão, compara-se o banco utilizado com os dados informados no Censo Escolar quanto ao número de alunos por dependência administrativa e quanto ao turno. Ademais, são expostas as estatísticas descritivas das proficiências de Língua Portuguesa e Matemática, a utilização do Índice Socioeconômico com variável de controle e as correlações entre eles.

Observa-se, na Tabela 1, que os alunos estão bem distribuídos entre as redes na base de dados, com aproximadamente metade em cada uma delas. Essa distribuição é similar a das duas redes em Belo Horizonte, conforme demonstrado na referida tabela. Embora em Minas Gerais as matrículas do 9º ano sejam majoritariamente da rede estadual, em Belo Horizonte, as matrículas dessa etapa estão mais bem divididas entre as duas dependências administrativas.

TABELA 1: Número de alunos, por dependência administrativa, que constam no Censo Escolar e na base selecionada, na rede estadual e municipal de ensino.

Fonte	Ente Federado	Rede	Nº de alunos	Porcentagem
Censo Escolar	Minas Gerais	Estadual	178.349	70,8
		Municipal	73.440	29,2
	Belo Horizonte	Estadual	13.112	52,3
		Municipal	11.980	47,7
Base selecionada	-	Estadual	5.697	52,7
		Municipal	5.111	47,3

Fonte: Censo Escolar 2013 e Base selecionada.

Vale ressaltar que as demais dependências, federal e privada, respondem pela minoria das matrículas, tanto no Estado (10%) como Município de Belo Horizonte (24%), de acordo com dados do Censo Escolar 2013. Tais redes não participam das avaliações do Proeb.

A mesma similitude entre a base e o Censo Escolar é notada em relação ao número de alunos por turno. Os dados utilizados no trabalho apontam que a maioria dos alunos do 9º ano de Belo Horizonte estuda no turno matutino, por volta de 65%.

TABELA 2: Percentual de alunos, por turno, no 9º ano em Belo Horizonte que constam na base selecionada e no Censo Escolar.

Fonte	Turno	Nº de alunos	Percentual
Censo Escolar	manhã	16.221	64,8
	tarde	7.888	31,5
	noite	938	3,7
Base selecionada	manhã	7.048	65,2
	tarde	3.560	32,9
	noite	200	1,9

Fonte: Censo Escolar 2013 e Base de dados selecionada.

A característica é comum no Estado, especialmente para esta etapa de escolaridade. Os dados do Censo Escolar 2013 informam que aproximadamente 70% desses alunos estão matriculados no turno da manhã, ao passo que o noturno compõe uma parcela mínima desses alunos.

Visualizadas as características gerais do banco, parte-se para o detalhamento das proficiências. A esse respeito, a Tabela 5 é elucidativa. Sua análise, entretanto, deve levar em conta que, em Língua Portuguesa, o que caracteriza o padrão recomendado de desempenho é o alcance de 275 pontos e, para Matemática, o valor para a mesma faixa é de 300 pontos².

TABELA 3: Estatísticas descritivas da proficiência em Língua Portuguesa e Matemática.

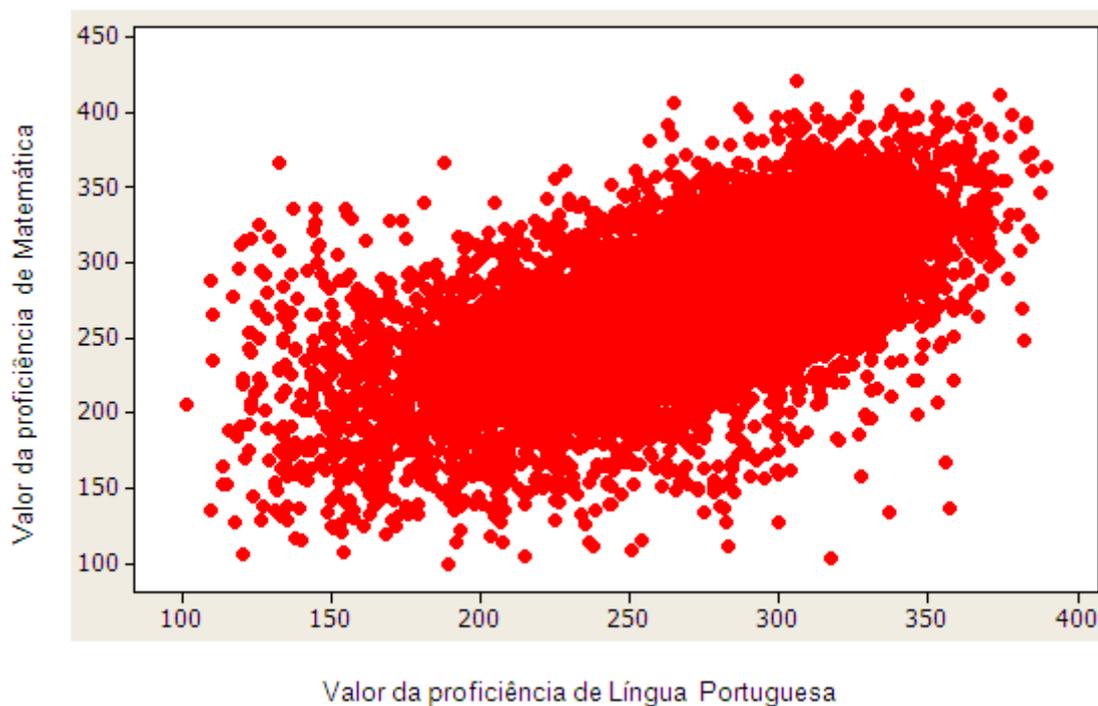
	Média	Mediana	Primeiro Quartil	Terceiro Quartil	Mínimo	Máximo	Desvio-padrão	Coef Variação	N	% Nível recomendado
Língua Portuguesa	266,64	270,66	238,21	299,48	101,42	389,84	46,45	17,04	10.808	24,5
Matemática	268,48	269,74	239,87	299,33	99,54	420,1	45,76	17,42	10.808	45,9

² Conforme descrito na metodologia, os padrões de desempenho são estipulados de acordo com as competências e habilidades cognitivas esperadas para cada disciplina na etapa avaliada.

Em Língua Portuguesa, portanto, a média e a mediana estão mais próximas do valor esperado para a etapa, comparativamente aos resultados de Matemática. Os valores do primeiro e terceiro quartis das duas disciplinas são próximos entre si, o que caracteriza uma distribuição similar dos dados. O valor observado do desvio-padrão também é similar nos dois casos. Em Matemática, a amplitude é maior do que em Língua Portuguesa. O coeficiente de variação identifica que a variação das notas é similar nas duas disciplinas. O valor de N é 10.808 em ambos os casos e o percentual de alunos no nível recomendado em Matemática é quase a metade do que se encontra no referido nível em Língua Portuguesa.

A análise a seguir pretende identificar a relação entre essas duas proficiências. Para tanto, foi construído o gráfico de dispersão dos pontos, a fim de possibilitar a verificação visual e calculado o Coeficiente de Correlação Linear de Pearson, para a quantificação da correlação linear entre as duas variáveis.

GRÁFICO 3: Relação entre as proficiências de Língua Portuguesa e Matemática.



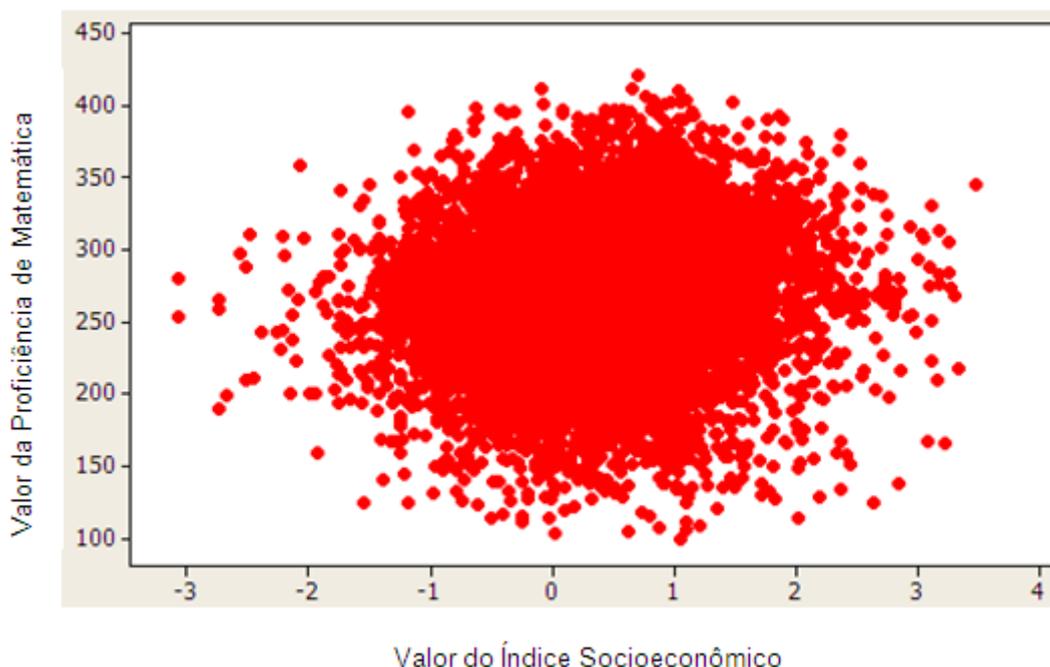
A visualização do Gráfico 5 indica uma correlação positiva entre a proficiência de Língua Portuguesa e Matemática³. Tal análise é confirmada pelo coeficiente de Pearson, que é de 0,614 (valor P igual a 0,000).

Considerado o destaque dado pela literatura para a relação do Índice Socioeconômico com o desempenho, decidiu-se investigar, também, sua relação com o desempenho de Matemática na base selecionada. Tal Índice leva em conta, sinteticamente a existência de itens de conforto na residência do aluno, cômodos, empregada doméstica e a escolaridade dos pais⁴. Essa relação está demonstrada no Gráfico 4.

³ Também se observa que há um grupo de alunos que apresentam valores mais baixos de proficiência em Língua Portuguesa e mais altos em Matemática, fora da tendência da curva. Isso é mais nítido para o grupo daqueles com proficiência em Língua Portuguesa abaixo de 150 e em Matemática acima de 250. Esse fato pode ser objeto de análises posteriores e possivelmente está impactando no leve decaimento observado nos gráficos de resíduos versus valores ajustados.

⁴ O Centro de Políticas Públicas de Avaliação Educacional da Universidade Federal de Viçosa foi o responsável pela elaboração dos testes em 2013 e elaborou, com base nas respostas às perguntas 4 a 14, 18 e 21 o Índice Socioeconômico, que constava na base utilizada com o score por aluno.

GRÁFICO 4: Relação entre a proficiência de Matemática e o valor do Índice Socioeconômico.



No caso do Índice Socioeconômico, a relação com o valor da Proficiência em Matemática não está nítida no gráfico de dispersão. Pode-se afirmar que há uma relação fraca e positiva entre as duas variáveis se considerado o Coeficiente de correlação linear de Pearson, que é de 0,103 (valor P igual a 0,000).

Ponderou-se que o referido Índice deveria ser considerado no modelo, devido a dois motivos. Em primeiro lugar, pelo papel que ocupa na literatura sobre o assunto e em segundo, como variável de controle do efeito-escola. Sobre esse segundo aspecto, verificou-se a composição do corpo discente em termos comparativos dentro de uma mesma escola e entre as escolas. Os resultados estão elencados na Tabela 6.

TABELA 4: Comparação da diferença entre grupos e intra grupos (ANOVA) do Índice Socioeconômico das escolas.

	Soma dos Quadrados	G.L.	Média dos Quadrados	F	Sig.
Entre Grupos	740,555	255	2,904	5,474	0
Nos grupos	5.598,148	10.552	0,531		
Total	6.338,703	10.807			

Fonte: Base de dados selecionada.

Verifica-se, com a Tabela 6, que existem evidências estatísticas de diferença no Índice Socioeconômico de alunos de escolas diferentes. Em outras palavras, que a composição do corpo discente quanto ao Índice Socioeconômico é mais similar dentro de uma mesma escola do que entre escolas diferentes. Tal fato supre, em parte, a limitação metodológica da não utilização de modelos multinível. Neste trabalho, o Índice Socioeconômico é utilizado como *proxy* da variável escola.

Na sequência, os modelos de regressão são desenhados com as referidas variáveis. Em primeiro lugar, construiu-se um modelo utilizando-se o resultado de Língua Portuguesa como preditora do resultado de Matemática. Os resultados estão dispostos no Quadro 1.

QUADRO 1: Saída do MINITAB para o ajuste do Modelo 1

PROF_MT versus PROF_LP					
<i>The regression equation is</i>					
PROF_MT = 107 + 0,605 PROF_LP					
Predictor	Coef	SE Coef	T	P	VIF
Constant	107,056	2,024	52,9	0,000	
PROF_LP	0,605422	0,007477	80,97	0,000	1,000
S = 36,1049 R-Sq = 37,8% R-Sq(adj) = 37,8%					
<i>Analysis of Variance</i>					
Source	DF	SS	MS	F	P
Regression	1	8545569	8545569	6555,53	0,000
Residual Error	10806	14086330	1304		
Lack of Fit	10709	13948454	1302	0,92	0,746
Pure Error	97	137876	1421		
Total	10807	22631899			

No Quadro 1, verifica-se que essa variável explica 37,8% da variabilidade da resposta e apresenta coeficientes estatisticamente significantes. A proficiência em Língua Portuguesa está captando o efeito de variáveis possivelmente ligadas à proficiência do aluno, mas não estão incluídas no modelo. Assim, a cada acréscimo

de um ponto no desempenho em Língua Portuguesa, o desempenho de Matemática aumenta, em média, 0,605 pontos. O teste da falta de ajuste, com Valor P maior que 0,05, indica que não há evidências estatísticas de falta de ajuste deste modelo.

Como será visto, o resultado da proficiência de Língua Portuguesa é o que mais acrescenta valor ao modelo entre as variáveis investigadas. Entretanto, o objetivo proposto não é de construção de um modelo inferencial sobre o resultado de Matemática, mas o estudo sobre as variáveis possivelmente relacionadas a ele. Por isso, as variáveis do questionário contextual são acrescentadas a ele, ainda que o R^2 não cresça de maneira acentuada.

O modelo regressão com base nos resultados de Língua Portuguesa e do Índice Socioeconômico está exposto no Quadro 2.

QUADRO 2: Saída do MINITAB para o ajuste do Modelo 2

PROF_MT versus PROF_LP; ISE

The regression equation is

$$\text{PROF_MT} = 107 + 0,601 \text{ PROF_LP} + 3,45 \text{ ISE}$$

<i>Predictor</i>	<i>Coef</i>	<i>SE Coef</i>	<i>T</i>	<i>P</i>	<i>VIF</i>
Constant	106,924	2,019	52,97	0,000	
PROF_LP	0,601196	0,007479	80,39	0,000	1,006
ISE	3,4503	0,4536	7,61	0,000	1,006

S = 36,0103 **R-Sq** = 38,1% **R-Sq(adj)** = 38,1%

Analysis of Variance

<i>Source</i>	<i>DF</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>P</i>
Regression	2	8620614	4310307	3323,95	0,000
Residual Error	10805	14011285	297		
Total	10807	22631899			

Nesse caso, o R^2 ajustado cresce para 38,1%, indicando uma leve melhora no poder explicativo do modelo. Ademais, o Quadro 2 mostra coeficientes estatisticamente significantes. Os valores de VIF indicam que não há problemas de multicolinearidade e o teste F da ANOVA indica a rejeição da hipótese nula de que todos os coeficientes de inclinação são iguais a zero. Para alunos com mesma

proficiência em Língua Portuguesa, o acréscimo de uma unidade no *score* do Índice Socioeconômico, aumenta o desempenho de Matemática, em média, em 3,45 pontos.

A fim de selecionar variáveis candidatas a entrar no modelo, optou-se por compará-las por meio do R^2 ajustado, utilizando-se com variável resposta a proficiência na prova de Matemática. A relação completa dos valores de R^2 e R^2 ajustados gerados para cada uma delas encontra-se no Apêndice 2, organizado por valores de R^2 , do maior para o menor.

Na Tabela 7, constam as variáveis que apresentaram valores de R^2 superiores 3% na investigação proposta. Todos os casos selecionados apresentaram pelo menos uma variável explicativa significativa⁵.

TABELA 5: Maiores valores R^2 ajustados em modelos de regressão linear cuja resposta é o PROF_MT e a explicativa é a variável investigada.

Item	Pergunta	Dummie	Classe de referência	R quadrado ajustado
54	O que você pretende fazer quando se formar no Ensino Médio?	Sim	Não vou me formar	4,60%
25	Atualmente, você está procurando por emprego?	Sim	Sim	4,50%
53	Você acredita que se formará no Ensino Médio?	Sim	Não	4,40%
28	De uma forma geral, de 0 (zero) a 10 (dez), qual nota você dá para você mesmo como estudante, considerando que "0" é muito ruim e "10" é muito bom?	Não	-	4,30%
3	Idade	Sim	Idade incorreta	3,70%
30	Pensando em sua ida como um todo, qual a sua opinião sobre a afirmativa "Sou capaz de fazer as coisas tão bem quanto a maioria das pessoas"?	Sim	Discordo muito	3,60%
13	Até que série/ano sua mãe ou responsável por você estudou?	Sim	Nunca estudou	3,20%
58	Pensando no que sua escola lhe oferece, qual a sua opinião sobre a afirmativa "Nada, vou a escola somente porque sou obrigado"?	Sim	Concordo muito	3,10%

Para a primeira variável elegível, questão 54, cujo enunciado é "O que você pretende fazer quando se formar no Ensino Médio?"⁶, gerou-se, a princípio, um modelo de regressão simples.

A variável da questão 54 foi então categorizada para discriminar aqueles que pretendem continuar os estudos e os que não pretendem fazê-lo. Assim, as categorias "Não vou me formar no ensino médio" e "Trabalhar", tornaram-se uma única categoria (não continuidade), que passou a ser a classe de referência. As

⁵ Para cada modelo, apenas uma questão investigada é a explicativa. Entretanto, nos casos de variáveis dummies, pôde existir mais de uma resposta, ainda que componentes do mesmo item do questionário.

⁶ Vide Anexo.

categorias “curso técnico” e “faculdade” foram agrupadas em “continuidade” e as categorias “outra opção” e “não sei”, como respostas “não especificadas”.

O Quadro 3 identifica o novo modelo gerado, considerando as variáveis já incorporadas anteriormente e somando-se as novas explicativas. O modelo compara aqueles que pretendem continuar os estudos (CONTI) ou não especificaram se pretendem continuar (NE_CONTI) com aqueles que afirmaram que não vão se formar no ensino médio ou vão trabalhar logo depois de se formarem (grupo de referência).

QUADRO 3: Saída do MINITAB para o ajuste do Modelo 3

PROF_MT versus PROF_LP; ISE; CONTINUIDADE.

The regression equation is

$$\text{PROF_MT} = 103 + 0,586 \text{ PROF_LP} + 3,10 \text{ ISE} + 9,68 \text{ CONTI} + 6,06 \text{ NE_CONTI}$$

<i>Predictor</i>	<i>Coef</i>	<i>SE Coef</i>	<i>T</i>	<i>P</i>	<i>VIF</i>
Constant	103,225	2,128	48,51	0,000	
PROF_LP	0,586108	0,007659	76,52	0,000	1,062
ISE	3,0979	0,4539	6,83	0,000	1,014
CONTI	9,676	1,147	8,44	0,000	2,398
NE_CONTI	6,063	1,288	4,71	0,000	2,297

S = 35,8861 **R-Sq** = 38,5% **R-Sq(adj)** = 38,5%

Analysis of Variance

<i>Source</i>	<i>DF</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>P</i>
Regression	4	8719698	2179925	1692,74	0,000
Residual Error	10803	13912201	1288		
Total	10807	22631899			

Controladas as demais variáveis, diz-se que os alunos que pretendem continuar a estudar após formarem-se no ensino médio alcançam, em média, 9,68 pontos a mais na proficiência de Matemática, se comparados aos que não pretendem. Analogamente, os que não especificaram se pretendem ou não dar continuidade aos estudos obtêm 6,06 pontos de acréscimo, em média, em relação à classe de referência.

Os coeficientes são estatisticamente significantes, o valor de R^2 ajustado é acrescido em 0,4 pontos percentuais comparativamente ao modelo anterior e a análise de resíduos, constante no Apêndice, é tida como aceitável no modelo. Ademais, o teste F da tabela ANOVA do Quadro 3 indica a rejeição da hipótese nula de que todos os coeficientes de inclinação são iguais a zero.

A próxima candidata é a variável do item 24, com o seguinte enunciado: “Atualmente, você está procurando por emprego?”. Para essa variável, as categorias são “Não” e “Sim”, sendo, essa última, a categoria de referência (Quadro 4).

QUADRO 4: Saída do MINITAB para o ajuste do Modelo 4

VL_PRF_ALN_1 versus VL_PRF_ALN_1; ISE; CONTINUIDADE; NÃO PROCURA EMPREGO.

The regression equation is

$$\text{PROF_MT} = 104 + 0,572 \text{ PROF_LP} + 2,45 \text{ ISE} + 8,53 \text{ CONTI} + 4,82 \text{ NE_CONTI} + 7,86 \text{ NP_EMPREGO}$$

<i>Predictor</i>	<i>Coef</i>	<i>SE Coef</i>	<i>T</i>	<i>P</i>	<i>VIF</i>
Constant	104,41	2,119	49,28	0,000	
PROF_LP	0,571846	0,007725	74,03	0,000	1,092
ISE	2,4519	0,4551	5,39	0,000	1,031
CONTI	8,525	1,145	7,45	0,000	2,418
NE_CONTI	4,816	1,286	3,75	0,000	2,315
NP_EMPREGO	7,859	0,7093	11,08	0,000	1,067

S = 35,6855 **R-Sq** = 39,2% **R-Sq(adj)** = 39,2%

Analysis of Variance

<i>Source</i>	<i>DF</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>P</i>
Regression	5	8876026	1775205	1394,01	0
Residual Error	10802	13755873	1273		
Total	10807	22631899			

Visualiza-se que a variável incorporada fez o valor de R^2 ajustado superar 39 pontos percentuais, incrementando o poder explicativo do modelo sobre a variabilidade da resposta. Os coeficientes obtidos são estatisticamente significantes e análise de resíduos é aceitável (Apêndice). Os valores de VIF não indicam

problemas de multicolinearidade. O Teste F permite rejeitar a hipótese nula de que os coeficientes de inclinação são iguais a zero.

Interpreta-se, portanto, que há um acréscimo médio de 7,86 pontos na proficiência de Matemática na comparação entre os alunos que não estão procurando emprego, em relação aos que estão, mantidas fixas todas as demais variáveis do modelo.

O próximo modelo, relacionado do Quadro 5, focaliza a expectativa de formação no ensino médio. Nesse caso, foi utilizada a resposta à pergunta 53, “Você acredita que se formará no Ensino Médio?”, sendo a resposta “Não” a classe de referência.

QUADRO 5: Saída do MINITAB para o ajuste do Modelo 5

PROF_MT versus PROF_LP; ISE; CONTINUIDADE; NÃO PROCURA EMPREGO; FORMARÁ NO ENSINO MÉDIO.

The regression equation is

$$\text{PROF_MT} = 102 + 0,562 \text{ PROF_LP} + 2,38 \text{ ISE} + 6,54 \text{ CONTI} + 3,99 \text{ NE_CONTI} + 7,67 \text{ NP_EMPREGO} + 8,27 \text{ FOR_EM}$$

<i>Predictor</i>	<i>Coef</i>	<i>SE Coef</i>	<i>T</i>	<i>P</i>	<i>VIF</i>
Constant	101,594	2,14	47,47	0,000	
PROF_LP	0,562245	0,00779	72,18	0,000	1,118
ISE	2,3826	0,4538	5,25	0,000	1,031
CONTI	6,545	1,167	5,61	0,000	2,33
NE_CONTI	3,989	1,286	3,1	0,002	2,33
NP_EMPREGO	7,6659	0,7075	10,83	0,000	1,069
FOR_EM	8,266	1,009	8,19	0,000	1,115

S = 35,5769 **R-Sq** = 39,6% **R-Sq(adj)** = 39,6%

Analysis of Variance

<i>Source</i>	<i>DF</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>P</i>
Regression	6	8960899	1493483	1179,95	0,000
Residual Error	10801	13671000	1266		
Total	10807	22631899			

Para o supracitado modelo, tem-se que os alunos que pretendem se formar no ensino médio têm, em média, um acréscimo de 8,27 pontos no desempenho de

Matemática confrontados com os que não pretendem (fixadas as demais variáveis). A análise de resíduos não mostra problemas com as suposições sobre os erros do modelo (Apêndice), valores de VIF não indicando multicolinearidade e Teste F rejeitando a hipótese nula de que os coeficientes de inclinação são iguais a zero. Visto isso, o modelo explica 39,6% da variabilidade dos dados (R^2 ajustado).

O modelo relacionado no Quadro 6 incorpora a autoavaliação do aluno, em uma escala de zero a dez. A pergunta é “De uma forma geral, de 0 (zero) a 10 (dez), qual nota você dá para VOCÊ MESMO COMO ESTUDANTE, considerando que “0” é muito ruim e “10” é muito bom?” Para essa pergunta não coube utilizar *dummies*, já que a variável é discreta ordinal.

QUADRO 6: Saída do MINITAB para o ajuste do Modelo 6

PROF_MT versus PROF_LP; ISE; CONTINUIDADE; NÃO PROCURA EMPREGO; FORMARÁ NO ENSINO MÉDIO; AUTOAVALIAÇÃO.

The regression equation is

$$\text{PROF_MT} = 89,0 + 0,554 \text{ PROF_LP} + 2,51 \text{ ISE} + 6,04 \text{ CONTI} + 4,18 \text{ NE_CONTI} \\ + 7,34 \text{ NP_EMPREGO} + 5,97 \text{ FOR_EM} + 2,04 \text{ AUTO_AV}$$

<i>Predictor</i>	<i>Coef</i>	<i>SE Coef</i>	<i>T</i>	<i>P</i>	<i>VIF</i>
Constant	89,044	2,484	35,85	0,000	
PROF_LP	0,554062	0,0078	71,03	0,000	1,131
ISE	2,5095	0,452	5,55	0,000	1,032
CONTI	6,035	1,163	5,19	0,000	2,532
NE_CONTI	4,184	1,28	3,27	0,001	2,33
NP_EMPREGO	7,3352	0,7052	10,4	0,000	1,071
FOR_EM	5,971	1,032	5,79	0,000	1,176
AUTO_AV	2,0393	0,2074	9,83	0,000	1,114

S = 35,4204 **R-Sq** = 40,1% **R-Sq(adj)** = 40,1%

Analysis of Variance

<i>Source</i>	<i>DF</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>P</i>
Regression	7	9082162	1297452	1034,15	0,000
Residual Error	10800	13549737	1255		
Total	10807	22631899			

No Modelo 6, verifica-se que o R^2 ajustado sobe para 40,1%, com coeficientes estatisticamente significantes, valores de VIF não indicando problemas de multicolinearidade, teste F da ANOVA indicando que se deve rejeitar a hipótese nula de que todos os coeficientes são iguais a zero. A análise de resíduos se mostra aceitável para a finalidade deste trabalho (Apêndice). Nesse modelo, para cada acréscimo de um ponto na escala de autoavaliação do aluno, o desempenho em Matemática cresce, em média, 2,04 pontos, mantidas fixas as demais variáveis.

O Quadro 7 dá continuidade aos passos dados no Quadro 6, com o acréscimo da variável sobre a idade do aluno avaliado. Para o item “Idade”, foram elencadas cinco opções de resposta, aqui categorizadas entre idade adequada (até 14 anos) e idade inadequada (todas as demais), sendo essa última a classe de referência.

QUADRO 7: Saída do MINITAB para o ajuste do Modelo 7 (continua)

PROF_MT versus PROF_LP; ISE; CONTINUIDADE; NÃO PROCURA EMPREGO; FORMARÁ NO ENSINO MÉDIO; AUTOAVALIAÇÃO; IDADE ADEQUADA.

The regression equation is

$$\text{PROF_MT} = 90,1 + 0,546 \text{ PROF_LP} + 2,46 \text{ ISE} + 5,57 \text{ CONTI} + 3,84 \text{ NE_CONTI} \\ + 6,48 \text{ NP_EMPREGO} + 5,88 \text{ FOR_EM} + 1,96 \text{ AUTO_AV} + 6,07 \text{ ID_ADEQ}$$

<i>Predictor</i>	<i>Coef</i>	<i>SE Coef</i>	<i>T</i>	<i>P</i>	<i>VIF</i>
Constant	90,116	2,479	36,36	0,000	
PROF_LP	0,545924	0,007831	69,71	0,000	1,147
ISE	2,4559	0,4505	5,45	0,000	1,032
CONTI	5,571	1,16	4,8	0,000	2,537
NE_CONTI	3,842	1,276	3,01	0,003	2,332
NP_EMPREGO	6,4762	0,7099	9,12	0,000	1,093
FOR_EM	5,875	1,028	5,71	0,000	1,176
AUTO_AV	1,9629	0,2069	9,49	0,000	1,116
ID_ADEQ	6,0708	0,7054	8,61	0,000	1,062

S = 35,3012 **R-Sq** = 40,5% **R-Sq(adj)** = 40,5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	8	9174451	1146806	920,26	0,000
Residual Error	10799	13457448	1246		
Total	10807	22631899			

Para esse modelo, atinge-se o patamar de 40,5% da variabilidade da resposta. A análise de resíduos é considerada aceitável (Apêndice) e os valores P dos testes T indicam que os coeficientes do modelo podem ser considerados estatisticamente significantes. Os valores de VIF, inferiores a 5, não mostram problemas de multicolinearidade das variáveis. O Teste F da ANOVA permite rejeitar a hipótese nula de que os coeficientes de inclinação são iguais a zero.

Interpreta-se que 6,07 pontos é o acréscimo no valor médio da proficiência em Matemática que os alunos na idade adequada têm em comparação com os que cursam a etapa com idade inadequada, mantidas fixas as demais variáveis.

A sequência da composição da análise refere-se à autoestima dos alunos, captada pela comparação da capacidade “fazer as coisas tão bem quanto à maioria das pessoas”. Nesse aspecto, os alunos foram inquiridos sobre a concordância ou não com a frase. As quatro possíveis respostas foram agrupadas em duas categorias: concordo e discordo, sendo essa de referência.

QUADRO 8: Saída do MINITAB para o ajuste do Modelo 8

(continua)

PROF_MT versus PROF_LP; ISE; CONTINUIDADE; NÃO PROCURA EMPREGO; FORMARÁ NO ENSINO MÉDIO; AUTOAVALIAÇÃO; IDADE ADEQUADA.; AUTOESTIMA.

The regression equation is

$$\begin{aligned} \text{PROF_MT} = & 87,1 + 0,543 \text{ PROF_LP} + 2,22 \text{ ISE} + 5,33 \text{ CONTI} + 3,80 \text{ NE_CONTI} \\ & + 6,25 \text{ NP_EMPREGO} + 5,24 \text{ FOR_EM} + 1,74 \text{ AUTO_AV} + 6,22 \text{ ID_ADEQ} \\ & + 8,09 \text{ AUTO_EST} \end{aligned}$$

<i>Predictor</i>	<i>Coef</i>	<i>SE Coef</i>	<i>T</i>	<i>P</i>	<i>VIF</i>
Constant	87,139	2,491	34,99	0,000	
PROF_LP	0,542594	0,00781	69,48	0,000	1,15
ISE	2,215	0,4496	4,93	0,000	1,036
CONTI	5,327	1,156	4,61	0,000	2,539
NE_CONTI	3,801	1,272	2,99	0,003	1,094
NP_EMPREGO	6,2533	0,7076	8,84	0,000	1,094
FOR_EM	5,236	1,027	5,10	0,000	1,181
AUTO_AV	1,7423	0,2075	8,40	0,000	1,132
ID_ADEQ	6,2167	0,7029	8,84	0,000	1,063
AUTO_EST	8,0929	0,88	9,16	0,000	1,045

S = 35,1665 **R-Sq** = 41,0% **R-Sq(adj)** = 40,9%

Analysis of Variance

<i>Source</i>	<i>DF</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>P</i>
Regression	9	9278174	1030908	833,61	0,000
Residual Error	10798	13353725	1237		
Total	10807	22631899			

Mantidas fixas todas as outras variáveis, verifica-se um acréscimo médio de 8,09 pontos na proficiência de Matemática no grupo de alunos que concordam com essa frase em contraste com os que não concordam. Conjuntamente, afere-se a significância estatística dos coeficientes de regressão, VIFs distantes de 5 e análise de resíduos aceitável (Apêndice). O Modelo 8 é capaz de explicar 40,9% da variabilidade da proficiência em Matemática.

Concernente ao Quadro 9, incorpora-se variável de escolaridade da mãe⁷ do discente ao Modelo anterior. Tal item foi assim categorizado: para os que responderam que a mãe nunca estudou, completou ou não a 4ª série/5º ano (antigo primário), mas não completou a 8ª série/9ºano (antigo ginásio), considerou-se que a mãe não completou o ensino fundamental (classe de referência); para as que completaram a 8ª série/9ºano, mas não completaram o Ensino Médio (antigo 2º grau), identificou-se que não completaram o ensino médio; as que chegaram à faculdade, as tendo completado ou não, assim foram classificadas. A categoria “Não sei” se mostrou não significativa estatisticamente e foi agrupada com a categoria de referência.

⁷ A pergunta é “Até que série/ano sua mãe ou o responsável por você estudou?”.

QUADRO 9: Saída do MINITAB para o ajuste do Modelo 9

PROF_MT versus PROF_LP; ISE; CONTINUIDADE; NÃO PROCURA EMPREGO; FORMARÁ NO ENSINO MÉDIO; AUTOAVALIAÇÃO; IDADE ADEQUADA; AUTOESTIMA; ESCOLARIDADE DA MÃE.

The regression equation is

$$\text{PROF_MT} = 86,9 + 0,539 \text{ PROF_LP} + 1,62 \text{ ISE} + 5,05 \text{ CONTI} + 3,94 \text{ NE_CONTI} \\ + 6,12 \text{ NP_EMPREGO} + 4,97 \text{ FOR_EM} + 1,76 \text{ AUTO_AV} + 6,08 \text{ ID_ADEQ} \\ + 7,98 \text{ AUTO_EST} + 2,19 \text{ MAE_EM} + 3,61 \text{ MAE_SUP}$$

<i>Predictor</i>	<i>Coef</i>	<i>SE Coef</i>	<i>T</i>	<i>P</i>	<i>VIF</i>
Constant	86,908	2,496	34,82	0,000	
PROF_LP	0,538759	0,007852	68,62	0,000	1,164
ISE	1,6208	0,4697	3,45	0,001	1,133
CONTI	5,047	1,157	4,36	0,000	2,546
NE_CONTI	3,943	1,271	3,10	0,002	2,334
NP_EMPREGO	6,1250	0,7078	8,65	0,000	1,096
FOR_EM	4,969	1,028	4,84	0,000	1,185
AUTO_AV	1,7648	0,2074	8,51	0,000	1,132
ID_ADEQ	6,0843	0,703	8,66	0,000	1,064
AUTO_EST	7,9835	0,8833	9,04	0,000	1,046
MAE_EM	2,1874	0,9466	2,31	0,021	1,220
MAE_SUP	3,6054	0,8140	4,43	0,000	1,397

S = 35,1375 **R-Sq** = 41,1% **R-Sq(adj)** = 41,0%

Analysis of Variance

<i>Source</i>	<i>DF</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>P</i>
Regression	11	9302693	845699	684,97	0,000
Residual Error	10796	13329206	1235		
Total	10807	22631899			

O Modelo 9 é capaz de explicar 41% da variabilidade da resposta. Além disso, os valores de VIF, abaixo de 5, não indicam problemas de multicolinearidade entre as variáveis. O valor P do teste F é menor do que o nível de significância (5%). Assim, deve-se rejeitar a hipótese nula de que os coeficientes de inclinação do modelo são iguais a zero. Os valores P do teste T dos coeficientes de regressão são estatisticamente significantes ao nível 5%. A análise de resíduos não mostra problemas com as suposições sobre os erros do modelo (Apêndice).

Fixadas as variáveis incorporadas anteriormente, nota-se que há um acréscimo médio de 2,19 pontos na proficiência de Matemática do aluno quando se passa do grupo de mães que não completaram o ensino fundamental para as que chegaram ao ensino médio. Analogamente, quando a escolaridade da mãe é maior, atingindo o ensino superior, há um ganho médio de 3,61 pontos na proficiência de Matemática alcançada pelo discente em relação ao aluno cujas mães não completaram o ensino médio, mantidas constantes as demais variáveis.

O Quadro 10 mostra o resultado do ajuste do Modelo 9 adicionado da variável “Motivo”, que é a última variável elencada na Tabela 7. A variável é relativa ao grau de concordância do aluno com a frase “Nada, vou à escola somente porque sou obrigado”, quando questionado sobre o que a escola lhe oferece.

QUADRO 10: Saída do MINITAB para o ajuste do Modelo 10

(continua)

PROF_MT versus PROF_LP; ISE; CONTINUIDADE; NÃO PROCURA EMPREGO; FORMARÁ NO ENSINO MÉDIO; AUTOAVALIAÇÃO; IDADE ADEQUADA; AUTOESTIMA; ESCOLARIDADE DA MÃE; MOTIVO.

The regression equation is

$$\text{PROF_MT} = 82,6 + 0,535 \text{ PROF_LP} + 1,75 \text{ ISE} + 4,59 \text{ CONTI} + 3,76 \text{ NE_CONTI} \\ + 6,11 \text{ NP_EMPREGO} + 4,40 \text{ FOR_EM} + 1,68 \text{ AUTO_AV} + 6,01 \text{ ID_ADEQ} \\ + 7,99 \text{ AUTO_EST} + 2,12 \text{ MAE_EM} + 3,57 \text{ MAE_SUP} + 7,47 \text{ MOTIVO}$$

<i>Predictor</i>	<i>Coef</i>	<i>SE Coef</i>	<i>T</i>	<i>P</i>	<i>VIF</i>
Constant	82,582	2,596	31,82	0,000	
PROF_LP	0,53514	0,007863	68,06	0,000	1,171
ISE	1,7526	0,4695	3,73	0,000	1,135
CONTI	4,589	1,158	3,96	0,000	2,558
NE_CONTI	3,757	1,269	2,96	0,003	2,336
NP_EMPREGO	6,1109	0,7066	8,65	0,000	1,096
FOR_EM	4,397	1,031	4,27	0,000	1,196
AUTO_AV	1,6797	0,2076	8,09	0,000	1,138
ID_ADEQ	6,0113	0,702	8,56	0,000	1,065
AUTO_EST	7,9941	0,8819	9,06	0,000	1,046
MAE_EM	2,1179	0,9451	2,24	0,025	1,220
MAE_SUP	3,5661	0,8127	4,39	0,000	1,397
MOTIVO	7,473	1,2560	5,95	0,000	1,052

S = 35,0816 R-Sq = 41,3% R-Sq(adj) = 41,2%

Analysis of Variance					
Source	DF	SS	MS	F	P
Regression	12	9346292	778858	632,85	0,000
Residual Error	10795	13285607	1231		
Total	10807				

Este último modelo da análise exploratória de dados proposta explica 41,2% da variabilidade da resposta, apresentando coeficientes estatisticamente significantes, teste F da ANOVA rejeitando a hipótese nula de que todos os coeficientes são iguais a zero, não tendo sido identificados problemas de multicolinearidade e tendo sido aceitas a análise de resíduos como não indicadora de problemas com as suposições de erros do modelo (Apêndice). Nesse caso, interpreta-se que há um acréscimo de 7,47 pontos, em média, na proficiência de Matemática no grupo dos que discordam que vão à escola somente porque são obrigados, em relação aos que concordam.

A análise exploratória dos dados dos alunos do Proeb, por meio de modelos de regressão linear, identificou fatores possivelmente associados ao desempenho dos alunos em Matemática, que estão elencados, conjuntamente, no Modelo 10, eleito como modelo final da análise exploratória.

6 CONCLUSÃO

Este trabalho visou identificar fatores possivelmente associados ao desempenho escolar na prova de matemática, por meio da utilização de métodos de regressão linear. Para tanto, foram utilizados dados da avaliação externa, Proeb, aplicada a alunos do nono ano matriculados em Belo Horizonte. Tais alunos foram selecionados entre aqueles que realizaram as provas de Língua Portuguesa e Matemática no ano de 2013 e responderam a todas as perguntas elencadas no questionário contextual.

Foi verificado que os fatores mais relacionados ao desempenho dos alunos na disciplina de Matemática são a proficiência em Língua Portuguesa, a expectativa de concluir o ensino médio e dar continuidade aos estudos para além dessa etapa, o fato de o aluno não estar procurando emprego, o de estar na idade adequada para a etapa de escolaridade, a autoavaliação como aluno e a autoestima relacionada à capacidade de fazer tão bem as coisas quanto a maioria das pessoas, a escolaridade da mãe e a variável relacionada motivo de ir à escola. Tais fatores e mais o Índice Socioeconômico compuseram análises de regressão linear, tendo sido eleito o Modelo 10, que explica 41,2%, da variabilidade da resposta, o modelo final.

A relação entre os resultados de Língua Portuguesa e Matemática foi a mais evidente. O gráfico de dispersão indicou uma clara correlação positiva entre as variáveis e, com a elaboração do modelo de regressão apenas com essa variável preditora, 37,8% da variabilidade da resposta já estava explicada. Contudo, o acréscimo das demais variáveis ao modelo representou ganhos para a finalidade deste trabalho de explorar fatores possivelmente relacionados ao desempenho.

Em relação ao Índice Socioeconômico, a literatura informa que, em geral, os resultados dos alunos estão relacionados, também, com antecedentes sociodemográficos. Optou-se por incluí-lo no modelo partindo-se dessa premissa e da observação de que os alunos componentes de uma mesma escola tinham composição de Índice Socioeconômico mais parecida do que alunos de escolas diferentes. Então, a variável foi utilizada como controle do efeito-escola.

Quanto às expectativas de continuidade aos estudos e ao fato de não estarem procurando emprego, identificadas no presente trabalho como fatores favoráveis ao desempenho, destaca-se o estudo de Jenks (1973) *apud* Brooke (2008), que afirmava que expectativa de dar continuidade aos estudos poderia

explicar diferenças no desempenho dos alunos. O mesmo indicava que essas diferenças poderiam se dar, também, pelo fato de alguns alunos terem que trabalhar durante a vida escolar.

Do mesmo modo, a variável relacionada à idade encontra respaldo na literatura, sendo válido retomar o estudo de Ferrão, Beltrão e Santos (2002) e de Jesus e Laros (2004), este último com a constatação de que o atraso escolar contribuía negativamente para o desempenho. Louzano (2012) considera que a chance de repetência e abandono diminui com o aumento da escolaridade dos pais. Sobre o tema da escolaridade dos responsáveis, recorre-se ao estudo de Barros e colaboradores (2001) que verifica que a escolaridade da mãe está entre os principais determinantes do sucesso escolar, o que corrobora com os achados deste trabalho.

As características dos pais podem relacionar-se, também, à autoavaliação e autoestima dos alunos e o fato de acreditarem que a escola lhes oferece algo e que não a frequentam somente porque são obrigados. Segundo Rodrigues, Rios-Neto e Pinto (2011), o capital cultural dos pais amplia as chances de que motivem, valorem e incentivem os filhos aos estudos. O levantamento de Sammos (1999) *apud* Brooke (2008) pode contribuir para a análise desta variável, pois elenca entre as características das escolas eficazes, fatores como incentivo positivo. Além disso, nessa investigação, podem ser considerados estudos afetos à expectativa do professor em relação aos alunos, como o de Soares e colaboradores (2010).

Os estudos e pesquisas sobre os fatores relacionados ao desempenho escolar são aqui retomados a título exemplificativo de como os fatores identificados nos modelos encontram eco na literatura. Essa análise pode ser feita sob duas óticas. A primeira, buscando identificar análises que corroboram com a hipótese de que as variáveis identificadas relacionam-se com a proficiência, a segunda com o intuito de verificar quais outros fatores estariam relacionados às variáveis identificadas. Nos dois casos, fica claro que qualquer análise será não exaustiva. Especialmente no campo de ciências sociais, diversos fatores não verificados influenciarão nos resultados obtidos. Por isso, usualmente investigações nesse campo consideram que valores baixos de correlação são suficientes para a proposta dos estudos.

Algumas das limitações deste estudo passam pela não utilização de teorias mais refinadas de análise, como métodos de regressão multinível, nos quais o efeito da escola poderia ser controlado de maneira apropriada. Além disso, retomando um destaque dado pelo referencial teórico, a limitação de ser uma análise baseada em avaliação externa, ou de larga escala, que tende a focalizar o que atua “fora da escola”, sendo o que está dentro da escola, como atitudes, valores e crenças, uma “caixa preta”.

Contudo, na análise proposta, pôde-se verificar e selecionar um rol de variáveis, por meio de modelos de regressão linear, e eleger um modelo de regressão capaz de explicar parte da variabilidade da resposta. Ou seja, por meio de análise estatística das respostas de um questionário contextual aplicado a um grupo de alunos, puderam-se vislumbrar fatores que contribuem para o debate acerca do desempenho escolar, pertinente ao desenvolvimento das políticas públicas educacionais.

REFERÊNCIAS

- ALBERNAZ, A. FERREIRA, F. e FRANCO, C. 2002. Qualidade e equidade no ensino fundamental brasileiro. *Pesquisa e Planejamento Econômico*, v.32, n.3. Rio de Janeiro, IPEA.
- ANDRADE, R. J.; SOARES, J.F. O efeito da escola básica brasileira. *Estudos em Avaliação Educacional*. • v. 19, n. 41, set./dez. 2008.
- BARROS, R. P. *et al.* Determinantes do Desempenho Educacional no Brasil. *Pesquisa e Planejamento Econômico*, v.31, n.1, p.1-42, abril 2001.
- BUCHMANN; HANNUN. Educação e estratificação nos países em desenvolvimento: uma revisão de teorias e pesquisas (2001).
- CHARLOT, Bernard. Sucesso e fracasso escolar: visões e proposições. In: UNESCO, MEC/INEP. *Repensando a escola: um estudo sobre os desafios de aprender, ler e escrever / coordenação de Vera Esther Ireland*. – Brasília: UNESCO, MEC/INEP, 2007.
- COLEMAN, J. S. et al. Igualdade de Oportunidades Educacionais (1966) In: Brooke, N., SOARES, J.F. *Pesquisa em eficácia escolar: origem e trajetórias*. Belo Horizonte: Editora UFMG, 2008.
- FARIA, E. M. ; MADALOZZO, R. C. . Excelência com Equidade: as lições das escolas brasileiras que oferecem educação de qualidade a alunos de baixo nível socioeconômico. 2012.
- FERRÃO; BELTRÃO e SANTOS. Políticas de não-repetência e a qualidade da educação: evidências obtidas a partir da modelagem dos dados da 4ª série do SAEB -99. *Estudos em Avaliação Educacional*, n. 26, jul-dez/2002.
- JESUS, G. R.; LAROS, J. A. Eficácia escolar: regressão multinível com dados de avaliação em larga escala. *Avaliação Psicológica*, v. 3, n. 2. Porto Alegre. 2004.
- JENCKS, Christopher. Desigualdade no Aproveitamento Educacional (1979) In: Brooke, N., SOARES, J.F. *Pesquisa em eficácia escolar: origem e trajetórias*. Belo Horizonte: Editora UFMG, 2008.
- LEE, V.E. Usando modelagem linear hierárquica para estudar contextos sociais: o caso do efeito das escola. *Educational Psychologist*, 2000
- LUZ, L. S. Os Determinantes do Desempenho Escolar: estratificação educacional e o efeito valor-adicionado. In: XV ENCONTRO NACIONAL DE ESTUDOS POPULACIONAIS, Caxambú, 2006.

MADAUS, F. G., *et al.* Estudos Empíricos (1980) In: Brooke, N., SOARES, J.F. Pesquisa em Eficácia Escolar: origem e trajetórias. Belo Horizonte: Editora UFMG, 2008.

MOSTELLER, F, MOYNIHAN, D. P. Um relatório inovador (1972) In: Brooke, N., SOARES, J. F., Pesquisa em Eficácia Escolar: origem e trajetórias. Belo Horizonte: Editora UFMG, 2008.

MONTGORMERY, D. C.; RUGER, G. C. Estatística Aplicada e Probabilidades para Engenheiros. Rio de Janeiro: LTC, 2003.

RIANI, Juliana de Lucena Ruas; RIOS NETO, E. L. G. Background familiar versus perfil escolar do município: qual possui maior impacto no resultado educacional dos alunos brasileiros?. Revista Brasileira de Estudos da População, v. 25, p. 251-269, 2008.

RODRIGUES, C. G.; RIOS-NETO, E. L. G.; PINTO, C. C. D. X. Diferenças intertemporais na média e distribuição do desempenho escolar no Brasil: o papel do nível socioeconômico, 1997-2005. Revista Brasileira de Estudos de População, v. 28, n. 1, p. 5-36, 2011.

RUTTER, Michael *et al.* Introdução: Estudos Anteriores (1979) In: Brooke, N., SOARES, J.F. Pesquisa em Eficácia Escolar: origem e trajetórias. Belo Horizonte: Editora UFMG, 2008.

SAMMONS, Pam. As características-chave da Escolas Eficazes (1999). In: Brooke, N., SOARES, J.F. Pesquisa em Eficácia Escolar: origem e trajetórias. Belo Horizonte: Editora UFMG, 2008.

SOARES, José Francisco. O efeito da escola no desempenho cognitivo de seus alunos. REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación. jul.-dez., 2004.

SOARES, José Francisco; COLLARES, Ana Cristina Murta. Recursos familiares e o desempenho cognitivo dos alunos do ensino básico brasileiro. Dados, Rio de Janeiro, v. 49, n. 3, 2006.

SOARES, SERGEI; SÁTYRO, NATÁLIA. O impacto da infra-estrutura escolar na taxa de distorção idade-série das escolas brasileiras de ensino fundamental – 1998 a 2005. Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2008.

SOARES, Tufi Machado *et al.* A expectativa do professor e o desempenho dos alunos. Psic.: Teor. e Pesq., Brasília, v. 26, n. 1, Mar. 2010.

VARANI, Adriana; SILVA, Daiana Cristina. A relação família-escola: implicações no desempenho escolar dos alunos dos anos iniciais do ensino fundamental. Revista Brasileira de Estudos Pedagógicos, v. 91, n. 229. 2010.

APÊNDICE 1
ANÁLISE DE RESÍDUOS DOS MODELOS AJUSTADOS

a) Análise de Resíduos do Modelo 1

Gráfico 1: Probabilidade normal dos resíduos do Modelo 1

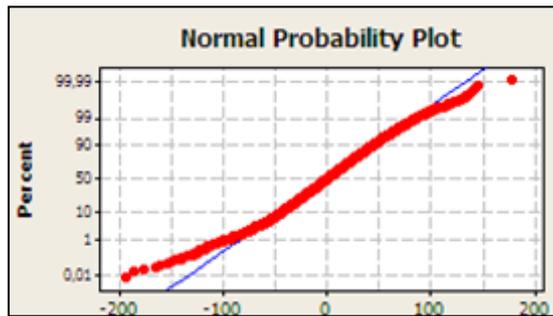


Gráfico 2: Resíduos versus valores ajustados do Modelo 1

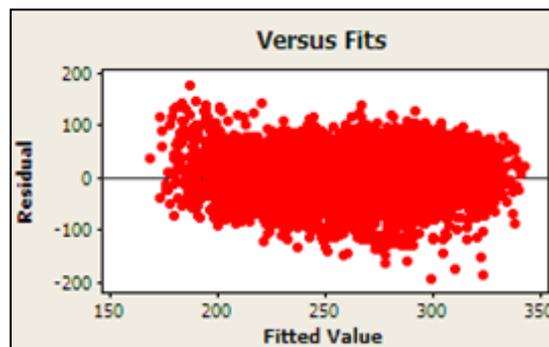
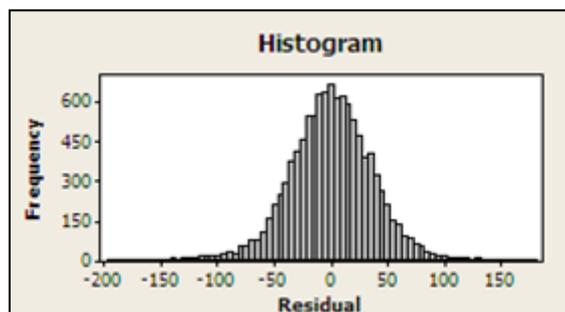
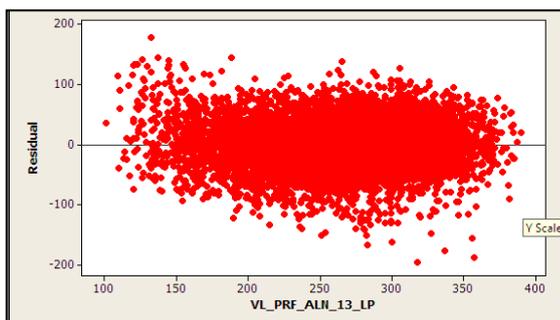


Gráfico 3: Histograma dos resíduos do Modelo 1



**Gráfico 4: Resíduos versus valores da proficiência em Língua Portuguesa -
Modelo 1**



b) Análise de Resíduos do Modelo 2

Gráfico 4: Probabilidade Normal dos resíduos do Modelo 2

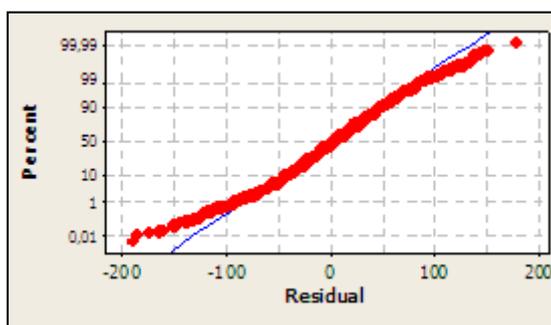


Gráfico 5: Gráfico de Resíduos versus valores ajustados do Modelo 2

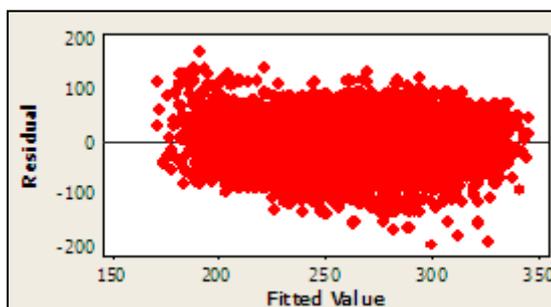


Gráfico 6: Histograma dos resíduos do Modelo 2

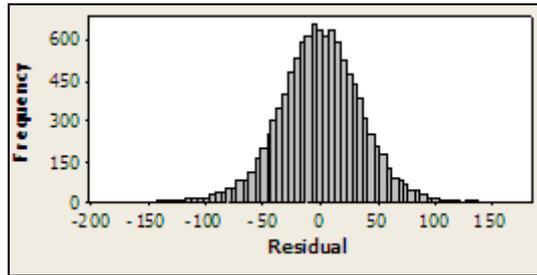
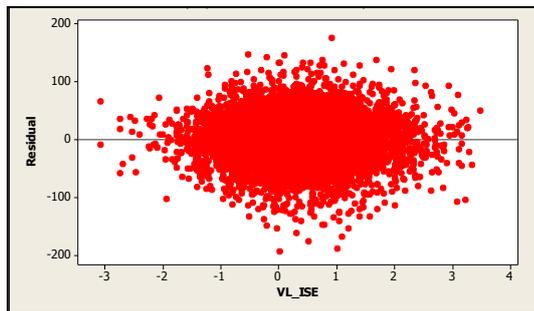


Gráfico 7: Gráfico de resíduos versus valor do Índice Socioeconômico – Modelo 2



c) Análise de Resíduos do Modelo 3

Gráfico 8: Probabilidade normal dos resíduos do Modelo 3

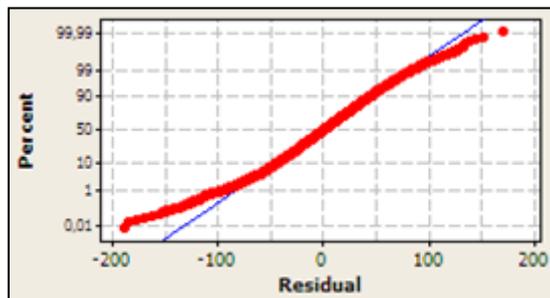


Gráfico 9: Resíduos versus valores ajustados do Modelo 3

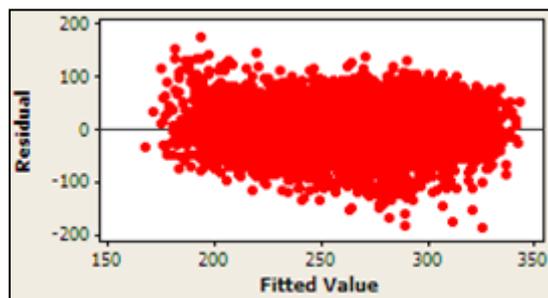


Gráfico 10: Histograma dos resíduos do Modelo 3

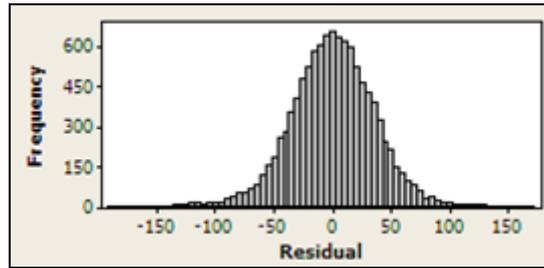


Gráfico 11: Resíduos versus variável CONTI – Modelo 3

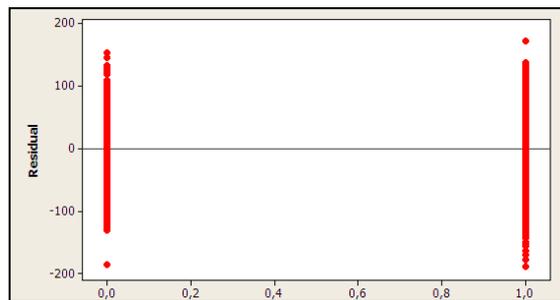
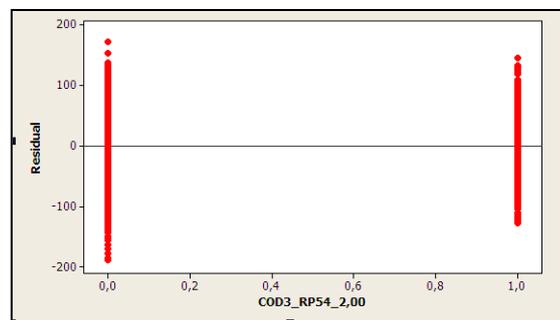


Gráfico 12: Resíduos versus variável NE_CONTI – Modelo 3



d) Análise de Resíduos do Modelo 4

Gráfico 13: Probabilidade normal dos resíduos do Modelo 4

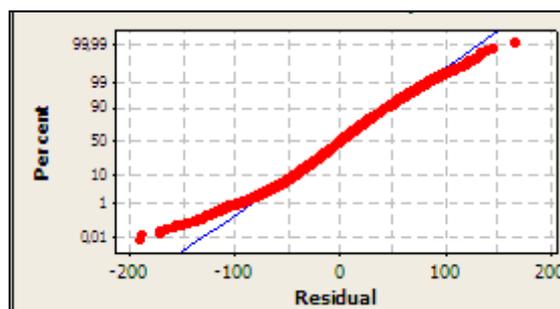


Gráfico 14: Resíduos versus valores ajustados do Modelo 4

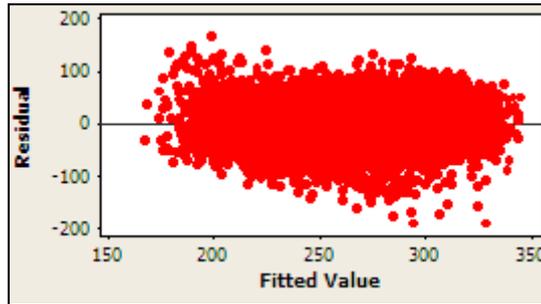


Gráfico 15: Histograma dos resíduos do Modelo 4

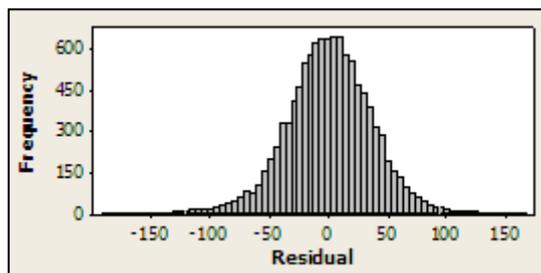
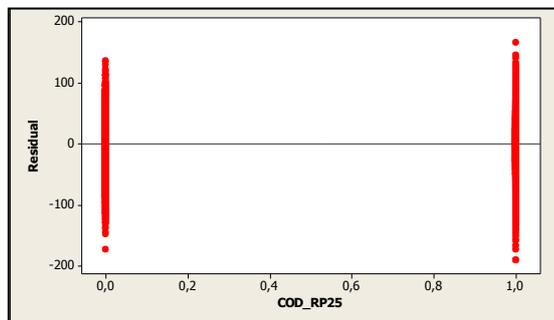


Gráfico 16: Resíduos versus variável NP_EMPREGO – Modelo 4



e) Análise de Resíduos do Modelo 5

Gráfico 17: Probabilidade normal dos resíduos do Modelo 5

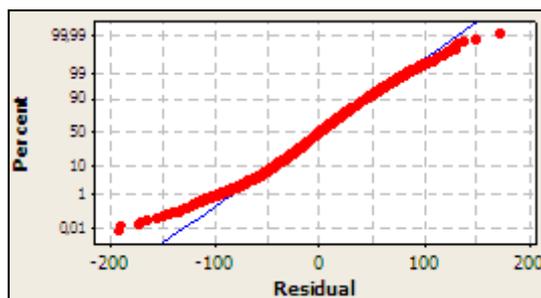


Gráfico 18: Resíduos versus valores ajustados do Modelo 5

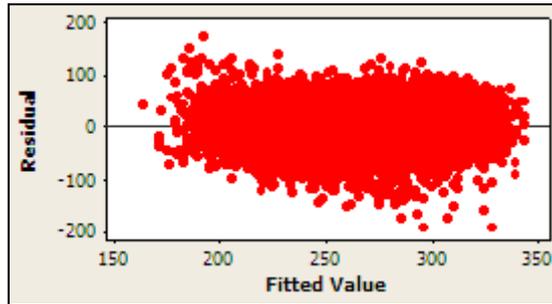


Gráfico 19: Histograma dos resíduos do Modelo 5

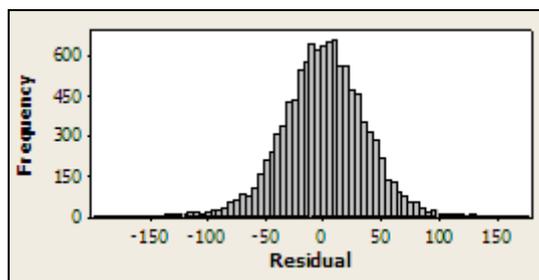
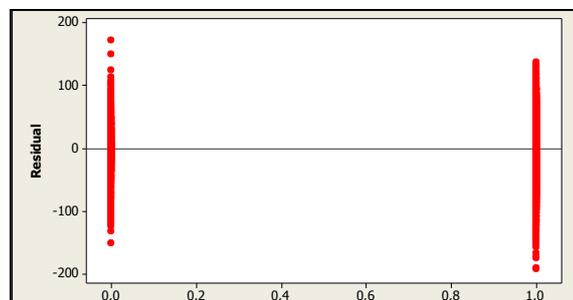


Gráfico 20: Resíduos versus variável FOR_EM – Modelo 5



f) Análise de Resíduos do Modelo 6

Gráfico 21: Probabilidade normal dos resíduos do Modelo 6

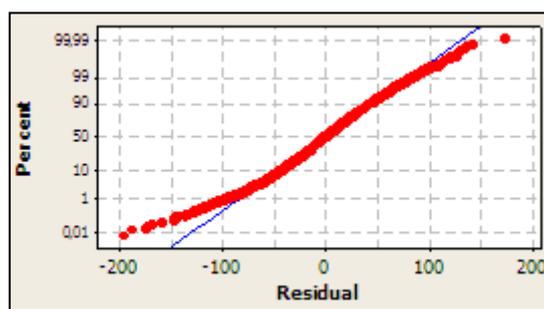


Gráfico 22: Resíduos versus valores ajustados do Modelo 6

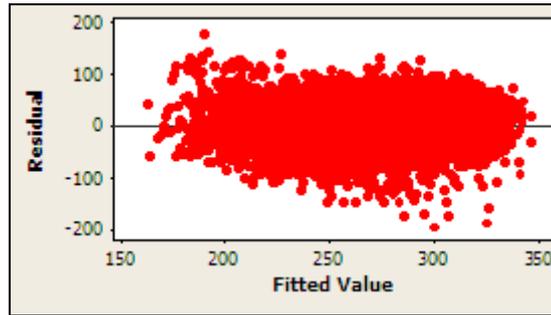


Gráfico 23: Histograma dos resíduos do Modelo 6

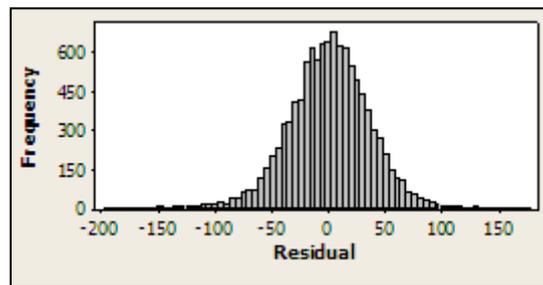
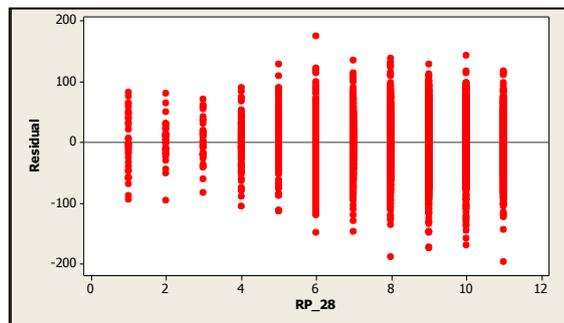


Gráfico 24: Resíduos versus variável AUTO_AV – Modelo 6



g) Análise de Resíduos do Modelo 7

Gráfico 25: Probabilidade normal dos resíduos do Modelo 7

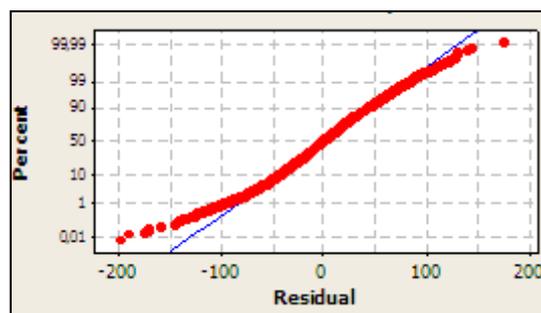


Gráfico 26: Resíduos versus valores ajustados do Modelo 7

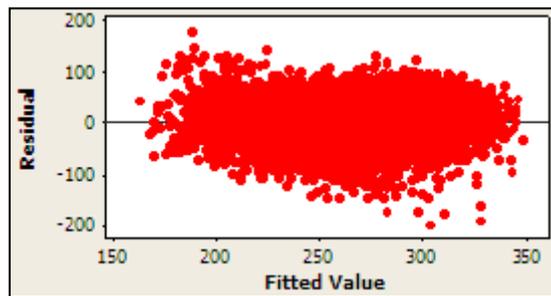


Gráfico 27: Histograma dos resíduos do Modelo 7

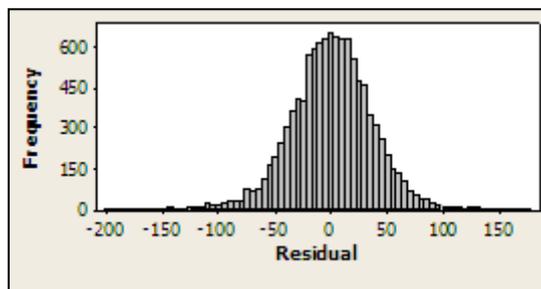
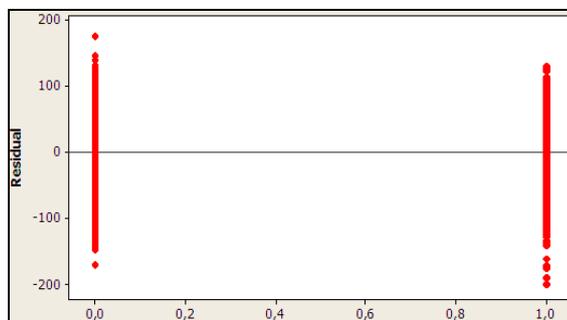


Gráfico 28: Resíduos versus variável ID_ADEQ – Modelo 7



h) Análise de Resíduos do Modelo 8

Gráfico 29: Probabilidade normal dos resíduos do Modelo 8

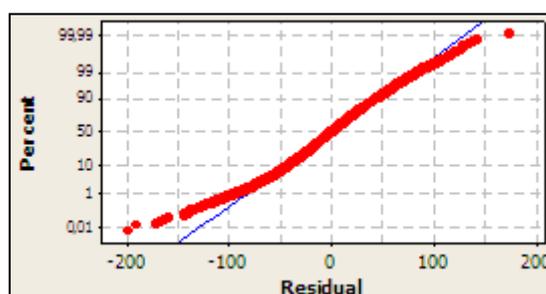


Gráfico 30: Resíduos versus valores ajustados do Modelo 8

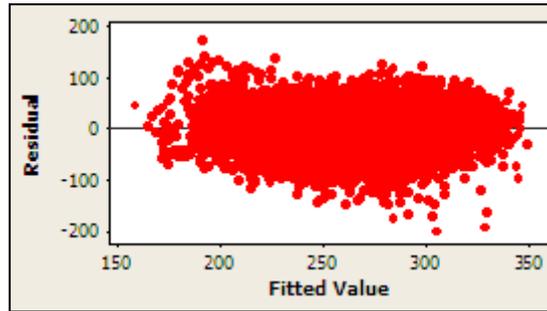


Gráfico 31: Histograma dos resíduos do Modelo 8

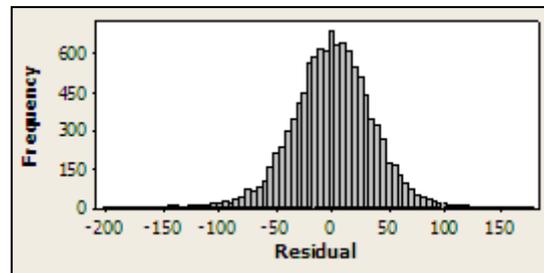
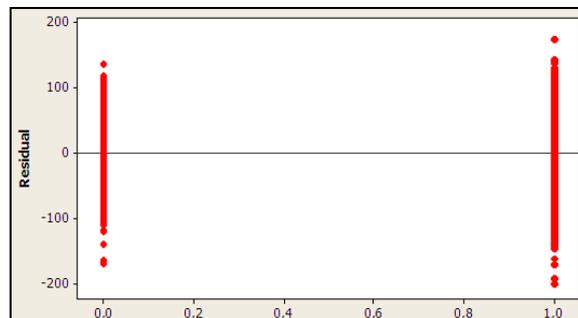


Gráfico 32: Resíduos versus variável AUTO_EST – Modelo 8



i) Análise de Resíduos do Modelo 9

Gráfico 33: Probabilidade normal dos resíduos do Modelo 9

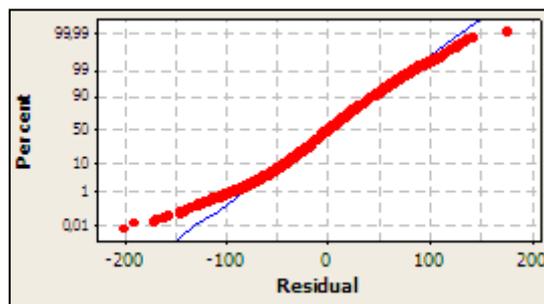


Gráfico 34: Resíduos versus valores ajustados do Modelo 9

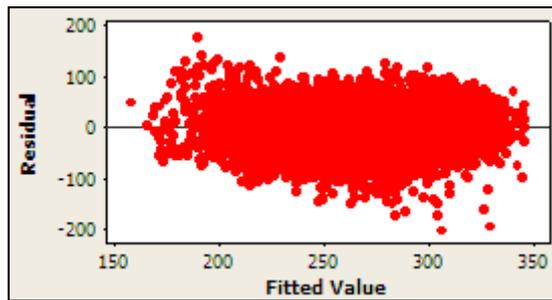


Gráfico 35: Histograma dos resíduos do Modelo 9

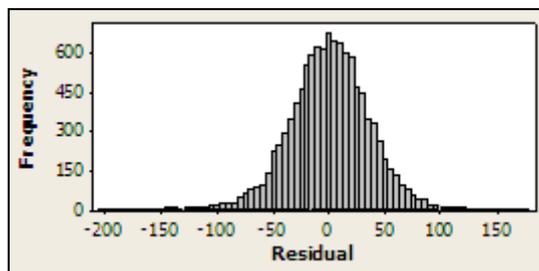


Gráfico 36: Resíduos versus variável MAE_EM – Modelo 9

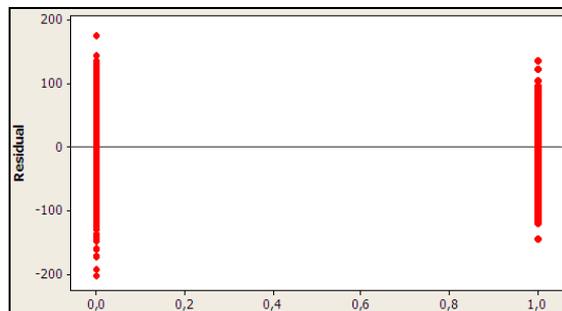
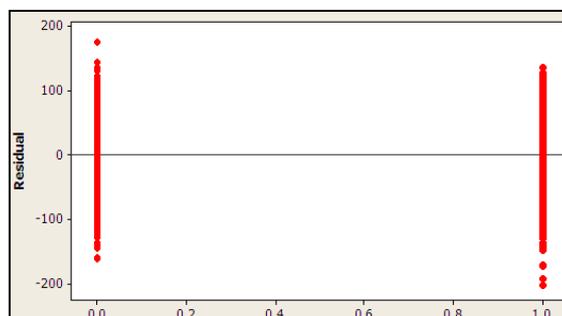


Gráfico 37: Resíduos versus variável MAE_SUP – Modelo 9



j) Análise de Resíduos do Modelo 10

Gráfico 38: Probabilidade normal dos resíduos do Modelo 10

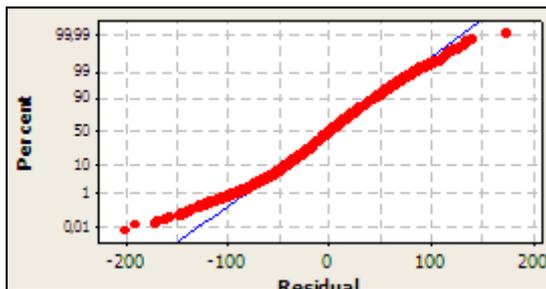


Gráfico 39: Resíduos versus valores ajustados do Modelo 10

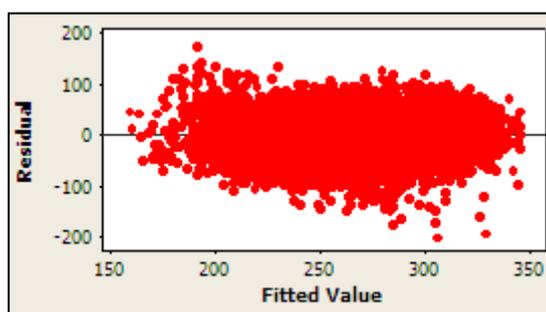


Gráfico 40: Histograma dos resíduos do Modelo 10

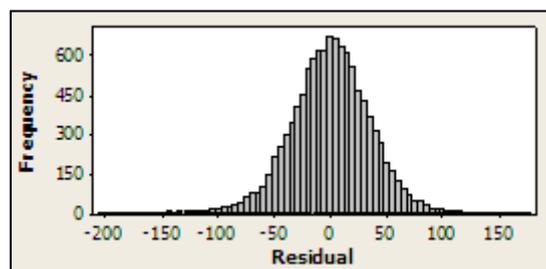
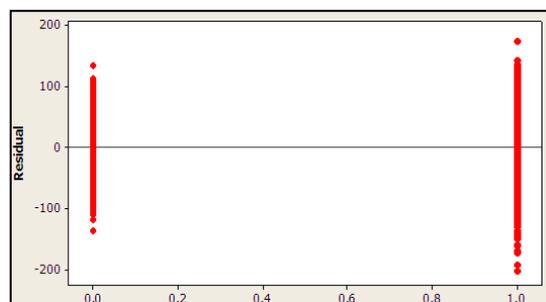


Gráfico 41: Resíduos versus variável MOTIVO – Modelo 10



APÊNDICE 2

VALORES DE R² PARA AS VARIÁVEIS DO QUESTIONÁRIO CONTEXTUAL

Questão	R ² ajustado	Classe de referência
54	4,60%	Não vou me formar
25	4,50%	Sim
53	4,40%	Não
28	4,30%	Não se aplica
3	3,70%	Idade incorreta
30	3,60%	Discordo muito
13	3,20%	Nunca estudou
58	3,10%	Concordo muito
32	2,90%	Discordo muito e discordo
14	2,60%	Nunca estudou
43	2,10%	Não se aplica
17	2,10%	Nunca
37	2,10%	Nunca
18	1,70%	Não
38	1,70%	Nunca
57	1,70%	Discordo
36	1,70%	Sempre
19	1,60%	3 ou mais
56	1,50%	Discordo muito
39	1,50%	Nunca
44	1,30%	Outra opção
35	1,30%	Nunca
34	1,20%	Nunca
40	1,00%	Nunca e a escola não tem
11	1,00%	Nenhum
22	0,80%	Com outras pessoas
33	0,80%	Nunca
24	0,70%	Sim (remunerado)
20	0,70%	7 ou mais
26	0,70%	Trabalho mais de 8
6	0,70%	Não tem banheiro
51	0,70%	Discordo muito
7	0,70%	Nenhum
31	0,60%	Discordo muito
1	0,60%	Mulher
21	0,60%	Nenhum
27	0,60%	Não se aplica
4	0,60%	Não tem TV
2	0,50%	Não Branco
29	0,50%	Discordo muito
52	0,50%	Discordo muito
42	0,50%	Nunca e minha escola não realiza
12	0,40%	Não
23	0,40%	Escola municipal
45	0,40%	Discordo muito
10	0,40%	Nenhum
41	0,30%	Não e a escola não tem
46	0,30%	Discordo muito
50	0,30%	Discordo muito
8	0,30%	Nenhuma
49	0,30%	Discordo muito
47	0,20%	Discordo muito e discordo
48	0,20%	Discordo
15	0,10%	Nunca
16	0,10%	Não
5	0,00%	Não tem rádio
9	0,00%	Nenhum
55	0,00%	Discordo muito

ANEXO
QUESTIONÁRIO CONTEXTUAL DO ALUNO DO NONO ANO