





## Preditores do Desempenho em Matemática de Estudantes do Ensino Médio\*

Cristiano Mauro Assis Gomes<sup>1,\*\*</sup> , Denise de Souza Fleith<sup>2</sup> , Claisy Maria  
Marinho-Araujo<sup>2</sup> , & Mauro Luiz Rabelo<sup>2</sup> 

<sup>1</sup>Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil

<sup>2</sup>Universidade de Brasília, Brasília, DF, Brasil

**RESUMO** – Considerando a relevância da formação em matemática, assim como a evidência de preditores relacionados ao desempenho nesse domínio, realizou-se no presente estudo uma análise preditiva do desempenho matemático de inscritos no Exame Nacional do Ensino Médio de 2011, empregando a abordagem de Regressão em Árvore e um modelo com 53 preditores. Os resultados indicam que o modelo explicou 29,97% da variância do desempenho em matemática na amostra teste. Determinadas variáveis relacionam-se a um pior rendimento em matemática: renda familiar de até dois salários mínimos, sexo feminino, não ter cursado escolas particulares no ensino fundamental e no ensino médio, residir nas regiões Norte, Nordeste e Centro-Oeste, e estar altamente motivado para fazer o Exame para obter certificação ou bolsa de estudos. Os resultados obtidos salientam o papel de variáveis relacionadas ao indivíduo, à escola e à família como preditoras do desempenho em matemática.

**PALAVRAS-CHAVE:** matemática, ensino médio, predição, árvore de regressão, Exame Nacional do Ensino Médio

## Predictors of Students' Mathematics Achievement in Secondary Education

**ABSTRACT** – Acknowledging the relevance of mathematics education, as well the evidence about predictors related to achievement in this domain, the present study performed a predictive analysis of students' mathematics achievement in the National Exam for Secondary Education, employing the Regression Tree Method and a model with 53 predictors. Results indicated that the model explained 29.97% of the mathematics achievement variance. Certain variables are related to worse achievement in mathematics: Students' family monthly income equal or smaller than 2 minimum wages, be female, have not attended Primary and Secondary Education in private schools, live in North, North East and Center West regions of Brazil, be highly motivated to perform the exam to obtain Secondary Education certificate or scholarship. The results obtained highlight the role of variables related to the individual, school and family as predictors of mathematics achievement.

**KEYWORDS:** mathematics, secondary education, prediction, regression tree, National Exam for Secondary Education

O acesso à educação tem sido considerado requisito essencial para sobrevivência da e na sociedade moderna competitiva. Não fortuitamente, o crescimento econômico de uma nação está associado ao desenvolvimento das habilidades cognitivas de seus cidadãos (Hanushek & Woessmann, 2011). A competência em matemática, por exemplo, tem sido identificada em países europeus como de grande relevância para a realização pessoal, o exercício pleno da cidadania, a inclusão social e a empregabilidade no

mundo do século XXI. Contudo, o declínio no número de estudantes interessados em matemática, ciência e tecnologia e a predominância masculina nessas áreas têm preocupado gestores, educadores e pesquisadores europeus (*European Commission*, 2011).

No Brasil, o baixo desempenho de estudantes em matemática tem sido evidenciado em avaliações de larga escala, como o Exame Nacional de Ensino Médio (ENEM), o Sistema de Avaliação da Educação Básica (SAEB) e o

\* Apoio: Universidade Federal de Minas Gerais.

\*\* E-mail: [cristianogomes@ufmg.br](mailto:cristianogomes@ufmg.br)

■ Submetido: 08/03/2018; Revisado: 17/01/2019; Aceito: 09/04/2020.

PISA (*Programme for International Student Assessment*). De acordo com os resultados das provas do SAEB aplicadas em 2015, a proficiência média em matemática dos estudantes brasileiros concluintes do ensino médio foi de 267 pontos, o pior resultado desde 1995, início da construção da série histórica de desempenho nessa avaliação. Tal cenário demanda revisão das políticas públicas sociais e educacionais, assim como inovação das práticas de ensino e uma discussão aprofundada acerca da formação docente em matemática e das condições de ensino. Por sua vez, para fundamentar essa revisão, é preciso identificar que fatores (preditores) têm impactado o desempenho em matemática dos estudantes.

A literatura tem apontado fundamentalmente as características individuais, familiares e escolares como as três categorias de preditores mais relacionadas ao desempenho em matemática (Akben-Selcuk, 2017; Karakolidis et al., 2016; Lee & Stankov, 2013; Pangeni, 2014; Thien & Ong, 2014). No nível individual e familiar, salienta-se a contribuição de variáveis como sexo, status socioeconômico, frequência à educação infantil, bem como autoconceito, autoeficácia e ansiedade em relação à matemática. No que se refere à escola, sobressaem-se nível socioeconômico médio do grupo de estudantes, ambiente de gestão escolar, assim como disponibilidade de recursos materiais/equipamentos escolares adequados.

Vários estudos sustentam a afirmativa declarada de que as três classes de preditores apontadas - indivíduo, família e escola – destacam-se para explicar o desempenho discente em matemática (Akben-Selcuk, 2017; Karakolidis et al., 2016; Laros et al., 2010; Lee & Stankov, 2013; Martin & Lazendic, 2018; Pangeni, 2014; Pinto et al., 2016; Pipere & Mierina, 2017; Thien & Ong, 2015). Por exemplo, Karakolidis et al. (2016) analisaram dados do Programa Internacional de Avaliação de Estudantes (PISA) relativos à Grécia, e constataram que o sexo, a frequência à educação infantil, o autoconceito, a autoeficácia e o nível de ansiedade em matemática, bem como o status socioeconômico do aluno e da escola são preditores do desempenho em matemática. Da mesma forma, Thien e Ong (2015) analisaram dados do PISA sobre Cingapura e Malásia e observaram que o nível socioeconômico do estudante, a autoeficácia e o nível de ansiedade em matemática predizem o desempenho em matemática, em ambos os países, enquanto o nível socioeconômico da escola prediz o desempenho em matemática apenas no caso de estudantes malasianos.

Lee e Stankov (2013) investigaram a extensão em que autorenças acadêmicas, motivação, estratégias de aprendizagem e atitudes em relação à escola predizem desempenho em matemática, analisando dados do PISA de 41 países. Eles encontraram evidências de que a autoeficácia e o autoconceito são preditores relevantes do desempenho discente. Por sua vez, Akben-Selcuk (2017) analisou especificamente o desempenho de estudantes turcos no exame de matemática do PISA e verificou que sexo, idade, status socioeconômico, recursos da escola,

motivação intrínseca e variáveis de personalidade, tais como atribuição de causalidade externa ao fracasso e abertura às atividades de solução de problemas, apresentam papel preditivo. Já os achados de Pangeni (2014) chamam atenção para a contribuição dos fatores familiares e escolares para o desempenho em matemática. Os resultados revelaram que nível educacional dos pais, número de livros na residência, apoio dos pais na realização de tarefas escolares, formação dos professores, número de dias letivos e instalações físicas da escola predizem o desempenho matemático de uma amostra de 762 estudantes do ensino médio no Nepal.

No cenário nacional, as evidências também corroboram a afirmação de que as três classes de preditores apontadas - indivíduo, família e escola – são relevantes para explicar o desempenho discente em matemática. Laros et al. (2010), por exemplo, identificaram quais características do estudante e da escola estão associadas ao desempenho em matemática no ensino médio. Os resultados apontaram nível socioeconômico, recursos culturais, cobrança e incentivo dos pais, clima disciplinar na escola e trabalho colaborativo na escola. Em estudo de revisão integrativa, Pinto et al. (2016) identificaram os principais fatores que influenciam os resultados em matemática de estudantes brasileiros e portugueses no PISA. O contexto socioeconômico, o sistema educacional (por exemplo, altas taxas de retenção, desigualdade educacional, abandono escolar) e características da escola (pública ou particular, cultura da escola e desempenho do professor) foram considerados os fatores mais relacionados ao desempenho em matemática.

Em síntese, os estudos apresentam evidências de que um conjunto de preditores mostra-se relevante para explicar o desempenho em matemática. Identificar tais preditores não é uma tarefa científica trivial ou pouco relevante, pois o conhecimento a seu respeito permite a construção de informações bem embasadas sobre quais fatores se associam a um desempenho pior ou melhor. Dados acerca dos preditores permitem ainda a geração de políticas públicas embasadas empiricamente, capazes de atuar em fatores bem identificados. Schwartzman et al. (2017) afirmam que as políticas e práticas educacionais raramente têm sido embasadas em evidências empíricas. Nos países em desenvolvimento, como o Brasil, essa necessidade é urgente tendo em vista as condições de desigualdade social, econômica e educacional. Políticas e práticas baseadas em evidências, possivelmente, têm maior chance de se tornarem efetivas e alcançarem o objetivo de alavancar o desenvolvimento educacional do povo brasileiro.

Nesse sentido, foi objetivo deste estudo investigar um vasto número de variáveis preditoras do desempenho em matemática no ensino médio, assim como aplicar uma técnica de análise de dados não paramétrica, o método de regressão em árvore, com a finalidade de, especialmente, identificar relações não lineares existentes entre as variáveis do estudo e que não são facilmente identificáveis por métodos usuais de análise de dados, como, por exemplo, a análise de regressão múltipla ou a análise de regressão hierárquica. Este estudo

vai além da verificação do quanto as variáveis preditivas do estudo conseguem explicar a variância do desempenho em matemática. A análise preditiva aqui proposta tem como foco fornecer um “mapa” de interações entre as variáveis do estudo, com vistas a aprofundar o conhecimento teórico sobre o desempenho em matemática no ensino médio. Para tal intento, este estudo tem como variável de desfecho o desempenho dos participantes do Exame Nacional do Ensino Médio (ENEM) de 2011 no domínio de matemática e possui como variáveis predictoras um conjunto amplo de 53 variáveis dos microdados dessa edição do exame.

As razões pela escolha de analisar dados do ENEM são apresentadas, a seguir. Dado seu alcance atual, o ENEM pode ser encarado como uma política pública capaz de fornecer subsídios para o aprimoramento do sistema educacional brasileiro no nível da educação básica, particularmente no ensino médio. Instituído pela Portaria nº 438, de 28 de maio de 1998 (Ministério da Educação (MEC)/ Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), 1998), com o objetivo de avaliar as “competências e as habilidades desenvolvidas pelos examinandos ao longo do ensino fundamental e médio, imprescindíveis à vida acadêmica, ao mundo do trabalho e ao exercício da cidadania” (art. 2.º), o ENEM foi estruturado como uma avaliação individual de desenvolvimento de competências, tendo como eixos estruturantes a interdisciplinaridade e a contextualização de conhecimentos expressos na forma de situações-problema.

A criação do Programa Universidade para Todos (ProUni), com a consequente concessão de bolsas de estudos em instituições particulares de ensino superior com base no desempenho no ENEM, e a reformulação introduzida no exame em 2009, associada à implantação do Sistema de Seleção Unificada (SiSU), levaram o ENEM a assumir outra finalidade: a de processo seletivo de acesso ao ensino superior brasileiro (MEC/INEP, 2013). Isso fez com que o número de inscritos saltasse de 157,2 mil, quando de sua criação em 1998, para mais de 8,6 milhões em 2016. É o segundo exame de acesso ao ensino superior no mundo, em termos de quantidade de inscritos, ficando atrás apenas do Gaokao, exame criado em 1952 na China para selecionar os estudantes para as universidades daquele país. Desse modo, estudos que envolvem a análise do desempenho de estudantes em exames da magnitude do ENEM representam contribuições que extrapolam o contexto brasileiro, pois servem de base de comparação com resultados similares de outros países.

A partir das mudanças introduzidas no ENEM em 2009, além da tradicional redação, a prova do exame passou de 63 para 180 questões objetivas, distribuídas igualmente em quatro áreas de conhecimento: (a) Linguagens, Códigos e suas Tecnologias (incluindo redação), (b) Ciências Humanas e suas Tecnologias, (c) Ciências da Natureza e suas Tecnologias, e (d) Matemática e suas Tecnologias. No que se refere a Matemática e suas Tecnologias, a matriz de referência do exame, vigente a partir de 2009, relaciona

cinco eixos cognitivos – comuns a todas as quatro áreas avaliadas no exame –, com sete competências específicas de área, 30 habilidades e os objetos de conhecimentos específicos da Matemática. O enfrentamento de situações-problemas, a compreensão de fenômenos e a construção de argumentação estão evidentes na matriz de referência do exame. As sete competências da área de Matemática referem-se a conteúdos presentes na educação básica e estão organizadas por blocos temáticos: números, geometria, álgebra, grandezas e medidas, modelagem matemática, tratamento da informação e conhecimentos de estatística e probabilidade, enquanto os objetos de conhecimento estão subdivididos em cinco tópicos: conhecimentos numéricos; conhecimentos geométricos; conhecimentos de estatística e probabilidade; conhecimentos algébricos; conhecimentos algébricos/geométricos (Rabelo, 2013). De modo geral, as questões de Matemática são apresentadas sob a forma de situações-problema, as quais o participante deve resolver a partir da mobilização de saberes cognitivos e conceituais adquiridos ao longo da educação básica.

Hoje, quase todas as instituições de ensino superior (IES) brasileiras utilizam as notas obtidas pelos estudantes no ENEM como critério de seleção de ingresso em seus cursos de graduação. O exame transformou-se na avaliação de maior impacto e interesse da sociedade brasileira. Ao substituir o vestibular tradicional na maioria das IFES de todo o país, passou a ser visto como instrumento de democratização do acesso à universidade, uma vez que, com um único processo seletivo e pagamento de uma única taxa de inscrição, é possível concorrer a vagas em cursos superiores de graduação em todo o território nacional. O INEP, responsável pela operacionalização do ENEM, possui uma série de informações sobre inscritos e participantes no exame. Algumas dessas informações são coletadas diretamente no próprio ato de inscrição, enquanto aquelas que envolvem o desempenho na prova são, evidentemente, coletadas e registradas após sua realização e correção. Em suma, todo esse conjunto de informações compõe os chamados microdados do ENEM e tem a finalidade de armazenar dados importantes sobre o desempenho dos participantes, sobre suas características demográficas e socioeconômicas, além de informar se o inscrito apresenta alguma deficiência que requer a necessidade da realização da prova em condição especial, e com atendimento diferenciado (MEC/INEP, 2012). O INEP disponibiliza livremente o acesso a esses microdados, de modo que é possível explorá-los nos diferentes anos em que a prova do ENEM foi aplicada. Especificamente em relação à edição de 2011, Gomes et al. (2016, 2018) encontraram, por meio de análises fatoriais, evidências que corroboram a validade dos quatro domínios avaliados pelo Exame.

Refletindo sobre a relevância dos dados do ENEM para a investigação do desempenho em matemática, constata-se que os microdados possuem informações sobre diversas variáveis reconhecidas pela literatura como associadas ao desempenho em matemática, como, por exemplo, o sexo,

o nível socioeconômico (medido pela renda individual e familiar mensal, nível de escolaridade dos pais, local de residência, se exerce atividade remunerada etc), o tipo de escola na qual o inscrito cursou o ensino fundamental e o ensino médio, a localização e situação de funcionamento da escola do inscrito, entre outras. Considerando, pois, os

elementos expostos e argumentados, este estudo apresenta os resultados de uma análise preditiva que incorpora um modelo com 53 preditores, todos provenientes dos microdados do ENEM 2011, tomando como variável de desfecho o desempenho em matemática dos inscritos no ENEM 2011.

## MÉTODO

### Participantes

Foram utilizados para a análise os dados dos inscritos no ENEM 2011 que realizaram todas as provas, inclusive a de redação, e que responderam ao questionário socioeconômico. Essas condições geraram uma amostra de 3.670.089 participantes, a ser analisada neste estudo. Entre as características sociodemográficas mais marcantes dessa amostra, incluem-se: 59,51% eram do sexo feminino, 55,23% declararam já ter concluído o ensino médio, 91,24% informaram que concluíram ou estavam concluindo o ensino médio por meio de ensino regular, enquanto 86,23% sinalizaram ser solteiros, 43,51% de cor branca, 39,52% de cor parda e 11,57%, de cor preta. Por sua vez, 75,07% informaram cursar ou ter cursado escola estadual no ensino médio, enquanto 21,84% afirmaram cursar ou ter cursado escola privada durante o ensino médio, 97,58% declararam cursar ou ter cursado escola em região urbana, assim como 74,63% dos participantes indicaram ter renda familiar mensal de até dois salários mínimos, enquanto 56,54% declararam não possuir renda. Prosseguir os estudos no ensino superior e conseguir bolsa de estudos foram os dois motivos que mais fortemente impulsionaram os inscritos dessa amostra a participar do ENEM. Em uma escala de 0 a 5, indicando o valor 5 para a maior motivação, 90,60% selecionaram o valor máximo para a motivação em participar do ENEM como forma de prosseguir os estudos no ensino superior, assim como 82,81% selecionaram o valor 5 para a motivação em participar do ENEM como forma de conseguir uma bolsa de estudos.

### Variáveis dos Microdados do ENEM 2011 Utilizadas no Estudo

As variáveis preditivas utilizadas neste estudo provêm dos blocos de dados sobre o inscrito, a escola do inscrito e o questionário socioeconômico, referentes aos microdados do ENEM 2011 (MEC/INEP, 2012). Nem todas as variáveis desses blocos foram selecionadas, pois algumas focam informações para populações muito específicas. Em especial, a variável relacionada com a aplicação em sistema prisional não foi incluída, pois, conforme informação do INEP (2012), os inscritos que realizaram as provas nesta modalidade de opção não preencheram o questionário socioeconômico. Foram utilizadas 53 variáveis preditivas

neste estudo. Elas estão relacionadas na Tabela 1. A variável dependente envolve os escores padronizados dos estudantes no domínio de matemática, dispostos em uma escala criada pelo INEP e armazenada nos microdados de 2011. Essa escala padronizada possui o intervalo de 0 a 1000 pontos, com média de 500 pontos e desvio-padrão de 100 pontos (MEC/INEP, 2015).

### Procedimentos de Coleta de Dados

O INEP disponibiliza em seu site (<http://portal.inep.gov.br/web/guest/microdados>) arquivos que contêm os microdados do ENEM. Também fornece um código para transformar esses arquivos no formato .sav do pacote estatístico SPSS. Ao baixarmos os arquivos dos microdados do ENEM 2011, esses foram transformados para o formato .sav. Foram selecionados nesses arquivos apenas aqueles alunos que estiveram presentes nos dois dias de aplicação da prova do ENEM e que responderam ao questionário socioeconômico. Em seguida, salvamos os arquivos como objetos do software R (R Core Team, 2017), e realizamos todas as análises de dados com esse software. Vale destacar que os microdados do ENEM 2011 utilizados são de domínio público, tendo sido, contudo, assegurados o pleno sigilo e anonimato dos inscritos no exame.

### Análise de Dados

Utilizamos o método de regressão em árvore e o algoritmo CART (*Classification and Regression Trees*), criado por Breiman et al. (1984), com a finalidade de investigar o papel das variáveis preditivas na explicação da variável desfecho. O pacote rpart (Therneau & Atkinson, 2015) e o pacote caret (Kuhn, 2017), ambos do software R (R Core Team, 2017), foram empregados para realizar os procedimentos envolvidos na regressão em árvore. Considerando que o método de regressão em árvore, assim como o algoritmo CART, não fazem parte das técnicas mais conhecidas de análise de dados em psicologia, será explicado, muito brevemente, sua lógica. Detalhes maiores sobre essa abordagem podem ser encontrados em Gomes e Almeida (2017).

O algoritmo CART é uma técnica da área de *machine learning* que operacionaliza o método de regressão em árvore (James et al., 2013). Sobre o seu funcionamento, ele quebra

Tabela 1  
Variável Preditivas Utilizadas no Estudo

n	Variáveis Preditivas	n	Variáveis Preditivas
1	Idade	28	Indicador de Baixa Visão
2	Sexo	29	Indicador de cegueira
3	Unidade da Federação de Residência	30	Indicador de deficiência auditiva
4	Situação de Conclusão do Ensino Médio	31	Indicador de deficiência física
5	Tipo de instituição que concluiu ou concluirá o Ensino Médio	32	Indicador de deficiência mental
6	Solicitação de Certificação no Ensino Médio	33	Indicador de déficit de atenção
7	Solicitação de prova em Braille	34	Indicador de dislexia
8	Solicitação de prova ampliada	35	Indicador de gestante
9	Solicitação de auxílio de ledor	36	Indicador de lactante
10	Solicitação de sala de aula de fácil acesso	37	Indicador de leitura labial
11	Solicitação de transcrição	38	Solicitação de prova em horário diferenciado para os guardadores dos sábados
12	Solicitação de libras	39	Indicador de surdez
13	Estado Civil	40	Q2. Até quando seu pai estudou?
14	Cor/raça Declarada	41	Q3. Até quando sua mãe estudou?
15	Unidade da Federação da Escola	42	Q4. Renda Familiar Mensal
16	Dependência Administrativa da Escola	43	Q5. Qual a sua renda mensal, aproximadamente?
17	Localização da Escola	44	Q6. A casa onde você mora é?
18	Situação de Funcionamento da Escola	45	Q7. A sua casa está localizada em?
19	Q1. Quantas pessoas moram com você?	46	Q8. Você exerce ou já exerceu atividade remunerada?
20	Q15. Indique os cursos que você frequentou ou frequenta: curso profissionalizante	47	Q26. Indique o que levou você a participar do ENEM: obter a certificação do Ensino Médio ou acelerar meus estudos
21	Q16. Indique os cursos que você frequentou ou frequenta: preparatório para o vestibular	48	Q27. Indique o que levou você a participar do ENEM: conseguir uma bolsa de estudos
22	Q17. Indique os cursos que você frequentou ou frequenta: curso superior	49	Q28. Quantos anos você levou para concluir o Ensino Fundamental?
23	Q18. Indique os cursos que você frequentou ou frequenta: língua estrangeira	50	Q29. Você deixou de estudar durante o Ensino Fundamental?
24	Q19. Indique os cursos que você frequentou ou frequenta: curso de computação ou informática	51	Q30. Em que tipo de escola você cursou o Ensino Fundamental?
25	Q20. Indique os cursos que você frequentou ou frequenta: preparatório para concursos públicos	52	Q32. Você deixou de estudar durante o Ensino Médio?
26	Q24. Indique o que levou você a participar do ENEM: testar meus conhecimentos	53	Q33. Em que tipo de escola você cursou o Ensino Médio?
27	Q25. Indique o que levou você a participar do ENEM: prosseguir os estudos no Ensino Superior		

tantas vezes quanto possível os dados em pares de grupos distintos. Enquanto a amostra original, ainda não dividida pelo algoritmo, é denominada de nodo raiz ou nodo único, os grupos criados são chamados de nodos da árvore (Zhang & Singer, 2010). Cada nodo da árvore pode ser particionado, gerando novos pares de nodos. O nodo que dá origem a um novo par de nodos é denominado de nodo pai e os nodos que não geram nenhum par de nodos são denominados de folhas ou nodos terminais. Conforme pode ser constatado, toda a linguagem do método faz alusão à construção de uma árvore, que possui sua raiz de base, seus nodos e suas folhas (Lantz, 2015). Os modelos preditivos provenientes dessa

abordagem tendem a se tornar excessivamente ajustados ao dado analisado, trazendo problemas para a generalização da predição. A literatura em *machine learning* reconhece esse problema e o denomina de super aprendizagem, ou super ajuste. Essa literatura recomenda separar aleatoriamente os dados em duas partes, uma amostra treino e uma amostra teste, justamente como uma forma de tratar a super aprendizagem (James et al., 2013). De acordo com essa recomendação, foram separados de forma aleatória os dados dos microdados em duas amostras: treino (75% dos casos) e teste (25% dos casos). A proporção dos casos em 3/4 para a amostra treino e 1/4 para a amostra teste também seguiu

a recomendação da literatura em *machine learning* (James et al., 2013).

O algoritmo CART tem como estratégia inicial gerar uma árvore com o maior número possível de nodos. Em seguida, ele realiza uma “poda” dessa árvore, ou seja, elimina os nodos que pioram a predição do modelo. Buscando verificar os nodos que pioravam a predição do modelo e poderiam ser podados, realizou-se na amostra treino a técnica de validação cruzada 3-Fold, conforme recomendação da literatura em *machine learning* para amostras grandes (James et al., 2013; Lantz, 2015) e inspecionou-se o critério do custo de complexidade, que tem como função identificar o número de folhas da árvore em que há a melhor explicação

da variância da variável desfecho. Para a geração da árvore final, já “podada”, utilizamos o critério da parcimônia, ou seja, manteve-se apenas uma quantidade pequena de folhas da árvore original, de forma a permitir que esta fosse facilmente interpretável e gerasse informação substantiva sobre o desempenho em matemática (Rokach & Maimon, 2015). O modelo preditivo foi construído na amostra treino. Por sua vez, analisou-se sua capacidade preditiva, aplicando-o para prever a variável desfecho na amostra teste, conforme recomendação da literatura (James et al., 2013; Lantz, 2015). Para análise da capacidade preditiva, tomou-se como referência o  $R^2$ , ou seja, o percentual de variância explicada da variável desfecho.

## RESULTADOS E DISCUSSÃO

Optamos por apresentar os resultados e a discussão conjuntamente porque o método de regressão em árvore não é uma abordagem usual ou padrão na análise de dados em psicologia, demandando um esclarecimento mais alongado sobre a interpretação e a discussão de seus resultados. Antes de se adentrar especificamente nos resultados da árvore final gerada pelo modelo, é necessário descrever alguns achados referentes à construção da árvore. A árvore inicial criada, ainda sem o processo de “poda”, gerou número impressionante de 32.569 folhas, ou seja, nodos terminais. A técnica de 3-Fold permitiu identificar que a árvore com 2.830 folhas seria aquela que melhor explicaria a variância do desempenho em matemática, gerando um erro de predição de 61,58%. Não obstante, essa árvore possui um número enorme de folhas, impossibilitando a interpretação dos resultados da árvore, assim como a geração de conhecimento substantivo sobre o desempenho em matemática. Utilizando o critério da parcimônia, verificou-se um ponto de corte bastante interessante do ponto de vista interpretativo e que envolveria a seleção de apenas 20 folhas. Essa árvore geraria um erro relativo de cerca de 70,08% da variância da variável desfecho, ou seja, perderia cerca de 8,50% de explicação da variância, em relação ao melhor modelo preditivo possível de ser obtido. No entanto, ao invés de produzir milhares de folhas, esse modelo teria apenas 20 folhas, podendo ser facilmente interpretado, de forma a produzir resultados substantivos. Nesse sentido, os resultados deste estudo provêm dessa árvore. Em termos de acurácia, a árvore de 20 folhas explica 29,92% (100% menos o erro de predição do modelo) da variância do desempenho dos participantes no domínio de matemática do ENEM 2011 na amostra treino, assim como explica 29,97% da variância do desempenho em matemática na amostra teste, indicando um resultado bastante similar em ambas as amostras.

Das 53 variáveis preditivas utilizadas pelo algoritmo CART, apenas sete serviram para construir a árvore de 20 folhas. Em uma escala de 0 a 100 pontos fornecida pelo pacote rpart, a importância dessas variáveis para a predição da variável desfecho é a seguinte: [1] Q4. Renda Familiar Mensal (36 pontos), [2] Q30. Em que tipo de escola você

cursou o Ensino Fundamental? (15 pontos), [3] Q33. Em que tipo de escola você cursou o Ensino Médio? (14 pontos), [4] Sexo (nove pontos), [5] Q27. Indique o que levou você a participar do ENEM: conseguir uma bolsa de estudos, [6] Unidade da Federação de Residência (três pontos), [7] Q26. Indique o que levou você a participar do ENEM: obter a certificação do Ensino Médio ou acelerar meus estudos (dois pontos).

Sobre essas variáveis, é relevante constatar que parte considerável delas pertence ao rol de preditores do desempenho discente em matemática bem reconhecidos pela literatura da área (Akben-Selcuk, 2017; Karakolidis et al., 2016; Laros et al., 2010; Pangen, 2014; Pinto et al., 2016; Thien & Ong, 2015), como é o caso da renda familiar, do sexo, do tipo de escola e da região em que o estudante vive. Nesse sentido, os resultados da árvore de 20 folhas corroboram achados da literatura sobre os preditores do desempenho em matemática. A Figura 1 apresenta a árvore de 20 folhas. Devido à limitação de espaço, não é possível descrever todos os nodos nela apresentado. No entanto, a seguir, mostraremos como ler a árvore da figura, para que seja possível ao leitor acompanhar e compreender as informações contidas em todos os nodos e folhas geradas. Posteriormente, apresentar-se-á uma síntese dos resultados mais importantes.

No topo da Figura 1, há a informação de que o nodo raiz provém dos inscritos da amostra treino. O nodo raiz é a amostra treino ainda não quebrada pelo algoritmo CART. Relembrando, conforme argumentado previamente, a amostra treino foi a utilizada para a geração da árvore, enquanto a amostra teste foi utilizada para avaliar a capacidade preditiva do modelo gerado na amostra treino. Em relação ao nodo raiz, logo abaixo da frase “Nodo Raiz Amostra Treino” na Figura 1, encontra-se a frase “Q4. Renda Familiar”. Essa é uma das variáveis preditivas utilizadas neste estudo. Na medida em que ela está no topo da Figura 1, isso significa que ela foi a primeira variável utilizada para quebrar a amostra treino em dois nodos. Logo abaixo da frase “Q4. Renda Familiar”, encontram-se as frases “até 2 salários” e “<sim> <não>”. Essas informações indicam que

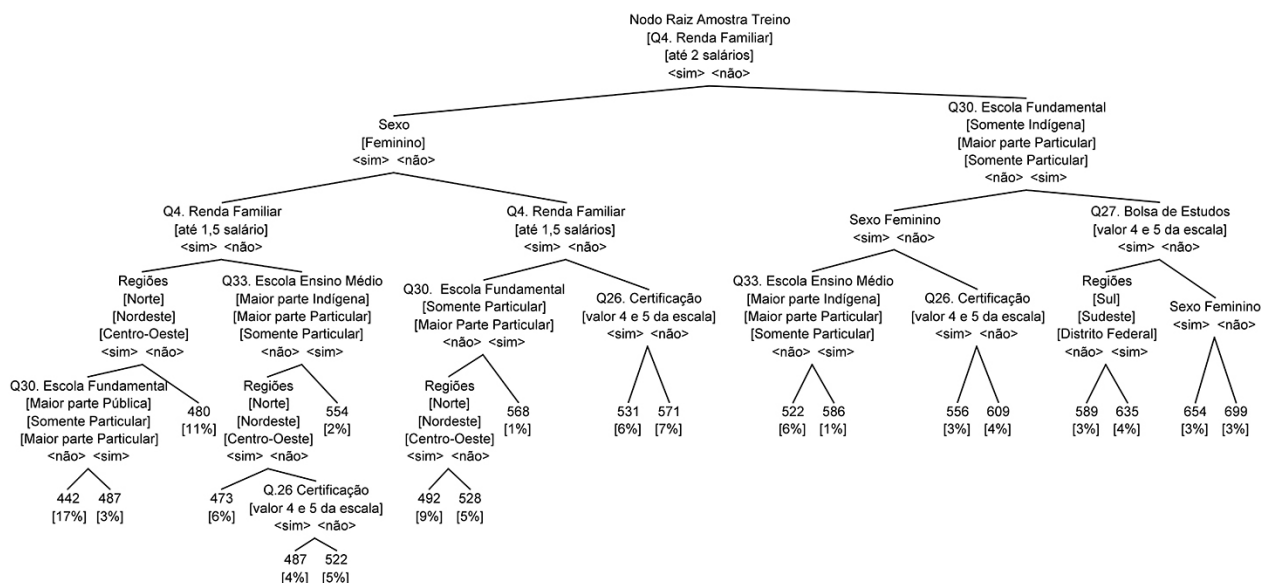


Figura 1. Árvore de 20 Folhas.

os inscritos do ENEM da amostra treino que informaram ter renda familiar de até dois salários mínimos formaram um nodo à esquerda do leitor, enquanto aqueles que informaram possuir renda familiar acima de dois salários mínimos, formaram um nodo à direita do leitor.

O nodo que compunha aqueles que informaram possuir renda familiar de até dois salários mínimos foi quebrado em dois novos nodos, por meio da variável sexo. Observe na Figura 1 que, na parte superior esquerda (tomando sempre o leitor como referência, para indicar os cantos superior, inferior, esquerda e direita), há as frases “Sexo” “Feminino” e “<sim> <não>”. Essas frases indicam que o nodo das pessoas que informaram possuir renda familiar de até dois salários mínimos foi quebrado em dois novos nodos, um de pessoas do sexo feminino e outro de pessoas do sexo masculino. As pessoas do sexo feminino foram alocadas em um nodo mais à esquerda, enquanto as do sexo masculino foram alocadas em um nodo mais à direita. O processo de quebra dos nodos continua até que seja informado na Figura 1 apenas um número, seguido de um percentual. Por exemplo, bem à esquerda, na parte inferior da figura, há o número 442 e o percentual 17%. Esse número indica um nodo terminal, no qual não há mais quebras. Conforme mencionado anteriormente, os nodos terminais são chamados de folhas. Esses números informam que as pessoas pertencentes a essa folha apresentam desempenho médio de 442 pontos no domínio de matemática do ENEM 2011 e que representam 17% da amostra treino.

Em um gráfico de árvore, dois tipos de informações são muito importantes. Ao longo da árvore, pode-se identificar quais variáveis preditivas geraram as diferentes quebras e os nodos subsequentes. Com essa identificação, é possível verificar o papel de cada uma das variáveis preditivas e sua importância para a composição dos nodos da árvore. No entanto, a informação mais importante está contida nas

folhas. Elas permitem identificar como o modelo prediz determinados resultados no domínio de matemática do ENEM 2011, articulando esta predição com as variáveis preditoras. A seguir, apresentamos um exemplo de como ler uma folha da Figura 1.

Bem no canto inferior esquerdo, há a folha com os números 442 (17%). Já foi indicado o que significam esses valores. Apenas reforçando, esse grupo de participantes do ENEM 2011 apresenta um desempenho médio de 442 pontos no domínio de matemática. Entre todas as folhas, esse é o grupo com pior desempenho em matemática. Para identificar o perfil de inscritos dessa folha, deve-se ler a Figura 1 da seguinte maneira. Iniciar pelo nodo raiz, no topo da figura, e verificar quais são as quebras que subsequentemente produziram o nodo das pessoas com o desempenho médio de 442 pontos. Partindo do nodo raiz, nota-se que as pessoas que relataram possuir renda familiar de até dois salários mínimos foram alocadas em um nodo à esquerda. Essas pessoas foram, em seguida, separadas novamente. As pessoas do sexo feminino com renda familiar de até dois salários mínimos foram alocadas mais à esquerda, em um novo nodo. Devemos continuar observando essa sequência, pois dela chegaremos à folha alvo. Em seguida, esse nodo de pessoas com renda familiar de até dois salários mínimos e do sexo feminino foi dividido em um par de nodos pela variável Renda Familiar. As pessoas com renda familiar de até 1,5 salário mínimo deste grupo foram alocadas em um novo nodo, mais à esquerda. Esse nodo foi quebrado em seguida por meio da variável Unidade da Federação de Residência. Pessoas do sexo feminino com renda familiar de até 1,5 salário mínimo e residentes nas regiões Centro-Oeste, Norte e Nordeste foram alocadas em um novo nodo, mais à esquerda. Finalmente, esse nodo foi quebrado por meio da variável Escola Fundamental cursada pelo inscrito. Aquelas pessoas desse nodo que não cursaram a maior parte

do ensino fundamental em escola pública ou em escola particular, ou não cursaram todo o ensino fundamental em escola particular, foram alocadas em um novo nodo. Esse nodo não foi novamente quebrado e é justamente a folha de pessoas com média de 442 pontos no domínio de matemática do ENEM 2011.

Em síntese, essa folha informa que pessoas do sexo feminino, com renda familiar de até 1,5 salário mínimo, residentes nas regiões Norte, Nordeste e Centro-Oeste do Brasil, e que não cursaram a maior parte do ensino fundamental em escola pública ou em escola particular, ou não cursaram todo o ensino fundamental em escola particular, apresentam o pior desempenho no domínio de matemática do ENEM 2011. Contrastando a essa folha, temos a folha na extrema direita da Figura 1, que apresenta o grupo com maior desempenho médio no domínio de matemática do ENEM 2011. Esse grupo possui desempenho médio de 699 pontos e representa 3% da amostra treino. Ele é composto por pessoas que declararam possuir renda familiar acima de dois salários mínimos, cursaram ensino fundamental somente em escola indígena ou particular, ou cursaram a maior parte em escola particular, não são do sexo feminino e não tiveram como forte motivação fazer o ENEM para obter bolsa de estudos (todos os itens relacionados à motivação em fazer o Exame possuíam uma escala de 0 a 5, sendo que maiores valores indicavam maior motivação; essas pessoas não marcaram os valores 4 e 5 para a variável Q.26. Indique o que levou você a participar do ENEM: obter a certificação do Ensino Médio ou acelerar meus estudos).

Sintetizando os resultados fundamentais contidos nas folhas da árvore da Figura 1, é possível afirmar que algumas variáveis estão relacionadas a um incremento ou um decréscimo no desempenho no domínio de matemática do ENEM 2011. Ser do sexo feminino está relacionado a um decréscimo importante. Por exemplo, as pessoas que possuem as mesmas características do grupo com maior desempenho, com exceção de serem do sexo feminino, apresentam um desempenho médio de 654 pontos, ao invés dos 699 pontos do grupo com maior desempenho, indicando uma queda de 45 pontos, o que corresponde a quase meio desvio-padrão da escala. Esse achado corrobora resultados de estudos anteriores, como os de Karakolidis et al. (2016) e Hampden-Thompson (2013), nos quais estudantes do sexo feminino tiveram desempenho inferior em matemática em comparação aos do sexo masculino. Apesar dos avanços e esforços no que tange ao alcance da igualdade de gênero, ainda persistem na sociedade brasileira estereótipos e práticas culturais e sociais que estabelecem limites ao que é permitido e até mesmo encorajado dependendo do gênero. Segundo Souza e Fonseca (2010), há uma intensa circulação de enunciados na sala de aula e em outros contextos que reforçam a superioridade masculina para a matemática, bem como características social e culturalmente atribuídas a mulheres (como docilidade, sensibilidade e gentileza) e aos homens (como racionalidade, coragem e ousadia). Como resultado, as práticas de numeramento são influenciadas por relações de gênero.

Além do sexo, a renda familiar relatada é uma variável importante. Pessoas que informaram possuir renda familiar de até dois salários mínimos apresentam desempenho pior, sendo intensificado o decréscimo no desempenho se a renda familiar relatada é de até 1,5 salário mínimo. No caso de estudantes com rendas mais elevadas, não se observou nenhum impacto relevante na diferenciação dos estudantes quanto ao desempenho matemático, divergindo da afirmação de Thien e Ong (2015) de que elevado nível socioeconômico teria um impacto positivo no desempenho acadêmico. Não obstante, é preciso cautela ao se comparar essas duas variáveis, pois o nível socioeconômico não se reduz à renda familiar. A variável socioeconômica é composta por vários outros fatores, como nível educacional e profissão dos pais, número de livros na residência, recursos culturais. Nesse sentido, comparar resultados referentes ao nível socioeconômico e à renda familiar é sempre um procedimento de aproximação, pois as informações geradas por ambas podem ser similares, mas não são idênticas.

Não ter cursado a maior parte do ensino fundamental em escola particular ou pública indica um decréscimo importante também. Por exemplo, o grupo com pior desempenho apresentou média de 442 pontos. Pessoas com as mesmas características desse grupo, mas que cursaram a maior parte do ensino fundamental em escola particular ou pública tiveram desempenho médio de 487 pontos, um incremento de 45 pontos. Por sua vez, os inscritos que relataram possuir alta motivação (valores 4 e 5 de uma escala de 0 a 5 pontos) em fazer o ENEM para obter certificação do Ensino Médio ou acelerar seus estudos apresentaram pior desempenho. As folhas com desempenho médio de 487 pontos e 522 pontos indicam isso. As pessoas dessas duas folhas apresentam as mesmas características, com exceção do grupo com pior desempenho, que relatou alta motivação para obter certificação do ensino médio ou acelerar seus estudos, com um decréscimo de 35 pontos em relação ao outro grupo. A motivação voltada para o alcance de metas externas e não para a aprendizagem em si pode ser um fator que interfere desfavoravelmente no desempenho acadêmico (Mello & Leme, 2016; Monteiro et al., 2012). É interessante constatar que a partir de 2017 o MEC (2017) tomou a decisão de retirar do ENEM a possibilidade de uso da nota no exame para obtenção do certificado de conclusão do ensino médio. O presente estudo apresenta evidências que poderiam subsidiar essa decisão.

Ser residente nas regiões Norte, Nordeste e Centro-Oeste do Brasil também está relacionado a ter pior desempenho no domínio de matemática do ENEM 2011. As folhas com desempenho médio de 492 pontos e 528 pontos indicam isso. Elas representam pessoas com as mesmas características, com exceção de que aquelas que residem nessas regiões apresentam pior desempenho, representando um decréscimo de 36 pontos. É possível que esse resultado esteja associado ao nível de desenvolvimento socioeconômico especialmente das regiões Norte e Nordeste, considerados inferiores quando comparados ao das regiões Sul e Sudeste (Instituto



Brasileiro de Geografia e Estatística (IBGE), 2019), indicando, possivelmente, menos oportunidades e condições aos seus cidadãos. Destaca-se que apenas os residentes do Distrito Federal que informaram renda familiar de até dois salários mínimos apresentaram desempenho em matemática similar aos residentes das regiões Norte e Nordeste. Residentes do Distrito Federal que relataram possuir renda familiar superior a dois salários mínimos apresentaram desempenho matemático similar aos residentes das regiões Sul e Sudeste. Essa particularidade do Distrito Federal merece ser investigada em novos estudos. Por fim, não ter

cursado a maior parte do ensino médio em escola indígena ou particular ou não ter cursado somente em escola particular também está relacionado a decréscimo no desempenho. As folhas com desempenho médio de 522 pontos e 586 pontos representam pessoas com as mesmas características, com a exceção de que o grupo com 586 pontos cursou a maior parte do ensino médio nessas escolas, indicando um incremento de 64 pontos. Esse último resultado leva-nos a refletir sobre as práticas de ensino adotadas nesses tipos de escolas, sendo necessário estudos mais aprofundados, inclusive de base qualitativa.

## CONCLUSÃO

O ENEM foi criado com o principal objetivo de avaliar o desempenho do estudante ao término da escolaridade básica, especificamente ao final do ensino médio. É uma proposta de avaliação diferenciada, distinguindo-se da maioria dos processos seletivos de acesso ao ensino superior, especialmente dos vestibulares aplicados pelas IES brasileiras. A partir de um forte investimento político nos anos de 2008 a 2012, esse processo avaliativo transformou-se em importante política pública brasileira. Além de induzir o ensino médio a reestruturações curriculares e adoção de propostas pedagógicas convergentes ao desenvolvimento de competências relevantes para o cidadão, o ENEM passou a ser uma forma de seleção para as IES, que utilizam o resultado do exame como critério para o ingresso em seus cursos de graduação, seja complementando ou substituindo (total ou parcialmente) o vestibular tradicional. Como desdobramentos, o ENEM configurou-se como seleção unificada nos processos seletivos das Instituições Federais do Ensino Superior (IFES), democratizando as oportunidades de acesso e possibilitando a mobilidade estudantil acadêmica.

Considerando o contexto dos avanços políticos e acadêmicos que a avaliação em larga escala do ensino médio provocou no país, oportunizados pelo ENEM, este estudo investigou um largo conjunto de variáveis dos microdados do ENEM 2011, buscando verificar se esse amplo espectro de informações poderia fornecer informações substantivas relevantes sobre o desempenho dos inscritos do ENEM no domínio de matemática. Nesse sentido, a utilização, no estudo aqui apresentado, de análises realizadas a partir dos microdados do INEP constitui efetiva contribuição às pesquisas de base metodológica quantitativa na área da educação e psicologia. O INEP é um dos maiores produtores de microdados relativos à educação: Censo Escolar, Censo da Educação Superior, Prova Brasil, SAEB, ANA, ENEM, ENADE, entre outros. Os microdados são os menores níveis de desagregação disponíveis de dados recolhidos por pesquisas, avaliações e exames realizados; eles não trazem a informação em si, não se encontram disponibilizados em outros produtos de levantamento estatístico e, por isso, precisam ser tratados para se extrair a informação desejada e os indicadores referentes aos objetivos das pesquisas. Ao

utilizar metodologias para manipulação dos microdados sobre o ENEM, evidenciam-se as contribuições deste artigo ao disseminar formas de acesso e compreensão das informações no processo de tratamento e cálculos estatísticos.

A análise preditiva realizada a partir dos resultados da presente investigação oportunizou a divulgação de um complexo “mapa” de interações entre as variáveis propostas nesta pesquisa, contribuindo para o avanço do conhecimento teórico sobre o desempenho em matemática no ensino secundário e consequentes subsídios para políticas públicas relacionadas tanto ao acesso ao ensino superior quanto às necessárias reformulações no currículo do ensino médio. Em termos de contribuições científicas, os resultados apresentados corroboraram achados de pesquisas anteriores nas quais variáveis como sexo, nível socioeconômico do estudante, região de residência, e tipo de escola explicam o desempenho escolar (Akben-Selcuk, 2017; Karakolidis et al., 2016; Laros et al., 2010; *Organization for Economic Co-operation and Development* (OECD), 2016; Pangen, 2014; Pinto et al., 2016; Thien & Ong, 2015). Por outro lado, há informações originais e inovadoras, a partir desse estudo, como a identificação de que renda superior a dois salários mínimos familiares é fator “protetivo” do desempenho em matemática, assim como a baixa motivação em fazer o Exame como meio de obter certificação, obter bolsas de estudo, ou acelerar os estudos. Além disso, o estudo mostra que das 53 variáveis preditivas utilizadas, apenas sete tiveram importância preditiva, de modo que um amplo conjunto de variáveis dos microdados usados neste estudo mostrou-se irrelevante para a compreensão do desempenho no domínio da matemática. O modelo preditivo deste estudo foi capaz de prever cerca de 30% do desempenho em matemática, o que é um resultado razoável apenas, afinal cerca de 70% do desempenho não foi explicado pelo modelo. Essa predição deixa uma parcela muita alta ainda sem explicação, o que pode ser apontado como possível limitação metodológica do estudo. É possível pressupor que a explicação de 30% reside no fato de que a árvore final selecionada foi criada pelo critério de parcimônia, diminuindo o poder preditivo do modelo. No entanto, mesmo se tivéssemos usado o critério do custo de complexidade

para gerar a árvore final, de maneira a produzir o modelo com melhor poder preditivo, mas com um número excessivo de folhas, esse modelo não explicaria mais do que cerca de 40% do desempenho em matemática. Tais resultados sugerem ser relevante o acréscimo de novas variáveis nos microdados do ENEM, com a finalidade de melhorar a predição do desempenho no domínio de matemática. É possível que a inclusão de novas variáveis educacionais, como, por exemplo, metodologia de ensino, formação de professores e organização curricular, assim como a inserção

de variáveis psicológicas nos microdados do ENEM, tais como criatividade, autoconceito e autoeficácia acadêmica, permitam melhor predição do desempenho no domínio de matemática. Este estudo revela a necessidade de se investir no potencial feminino para a matemática, encorajando as alunas desde cedo a se envolverem em atividades que exijam o uso do pensamento matemático, bem como na implementação de práticas docentes que despertem nos estudantes de distintos backgrounds socioeconômicos o interesse intrínseco pela área.

## REFERÊNCIAS

- Akben-Selcuk, E. (2017). Personality, Motivation, and Math Achievement among Turkish Students: Evidence from PISA Data. *Perceptual and Motor Skills*, *124*, 514-530. <http://doi.org/10.1177/0031512516686505>
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Chapman & Hall/CRC.
- European Commission. (2011). *Mathematics Education in Europe: Common Challenges and National Policies*. Education, Audiovisual and Cultural Executive Agency, Eurydice.
- Gomes, C. M. A., & Almeida, L. S. (2017). Advocating the Broad Use of the Decision Tree Method in Education. *Practical Assessment, Research & Evaluation*, *22*(10), 1-10. <https://doi.org/10.7275/2w3n-0f07>
- Gomes, C. M. A., Golino, H. F., & Peres, A. J. S. (2016). Investigando a Validade Estrutural das Competências do ENEM: Quatro Domínios Correlacionados ou Um Modelo Bifatorial? *Boletim na Medida*, *5* (10), 33-38.
- Gomes, C. M. A., Golino, H. F., & Peres, A. J. S. (2018). Análise da fidedignidade composta dos escores do ENEM por meio da análise fatorial de itens. *European Journal of Education Studies*, *5*, 331-344. <http://doi.org/10.5281/zenodo.2527904>
- Hampden-Thompson, G. (2013). Family Policy, Family Structure, and Children's Educational Achievement. *Social Science Research*, *42*, 804-817. <http://doi.org/10.1037/0022-0663.95.1.124>
- Hanushek, A. E., & Woessmann, L. (2011). *The Cost of Low Educational Achievement in the European Union* (Report No. 7). European Expert Network on Economics of Education (EENEE). [http://www.eenee.de/dms/EENEE/Policy\\_Briefs/PolicyBrief1-2011.pdf](http://www.eenee.de/dms/EENEE/Policy_Briefs/PolicyBrief1-2011.pdf)
- Instituto Brasileiro de Geografia e Estatística (IBGE) (2019). *Síntese de Indicadores Sociais: Uma Análise das Condições de Vida da População Brasileira*. <https://biblioteca.ibge.gov.br/visualizacao/livros/liv101678.pdf>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*. Springer. <https://doi.org/10.1007/978-1-4614-7138-7>
- Karakolidis, A., Pitsia, V., & Emvalotis, A. (2016). Examining Students' Achievement in Mathematics: A Multilevel Analysis of the Programme for International Student Assessment (PISA) 2012 Data for Greece. *International Journal of Educational Research*, *79*, 106-115. <http://doi.org/10.1016/j.ijer.2016.05.013>
- Kuhn, M. (2017). *caret: Classification and Regression Training*. <https://CRAN.Rproject.org/package=caret>
- Lantz, B. (2015). *Machine Learning with R*. Packt.
- Laros, J. A., Marciano, J. L. P., & Andrade, J. M. (2010). Fatores que Afetam o Desempenho na Prova de Matemática do SAEB: Um Estudo Multinível. *Avaliação Psicológica*, *9*, 173-186.
- Lee, J., & Stankov, L. (2013). Higher-Order Structure of Noncognitive Constructs and Prediction of PISA 2003 Mathematics Achievement. *Learning and Individual Differences*, *26*, 119-130. <http://doi.org/10.1016/j.lindif.2013.05.004>
- Martin, A. J., & Lazendic, G. (2018). Achievement in Large Scale National Numeracy Assessment: An Ecological Study of Motivation and Student, Home, and School Predictors. *Journal of Educational Psychology*, *110*, 565-482. <http://doi.org/10.1037/edu0000231>
- Mello, M. B. J. B., & Leme, M. I. S. (2016). Motivação de Alunos dos Cursos Superiores de Tecnologia. *Psicologia Escolar e Educacional*, *20*, 581-590. <http://doi.org/10.1590/2175-3539201502031053>
- Ministério da Educação (MEC)/Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). (1998). *Portaria Normativa nº 438, de 28 de maio de 1998*. DOU de 01 jun. 1998, nº 102-E, seção 1, p. 5. <http://www.normasbrasil.com.br/norma/?id=181137>
- Ministério da Educação (MEC)/Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). (2012). *Microdados do ENEM – 2011, Exame Nacional do Ensino Médio: Manual do usuário*. MEC/INEP. <http://portal.inep.gov.br/web/guest/microdados>
- Ministério da Educação (MEC)/Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). (2013). *Exame Nacional do Ensino Médio (Enem): Relatório pedagógico 2009-2010*. INEP/MEC. <http://portal.inep.gov.br/documents/186968/484421/Relatório+Pedagógico+ENEM+2009-2010/70890e24-a78a-44f8-a909-b235f02948f2?version=1.1>
- Ministério da Educação (MEC)/Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). (2015). *Relatório pedagógico: ENEM 2011-2012*. INEP. <http://www.publicacoes.inep.gov.br/portal/download/1401>
- Ministério da Educação (MEC) (2017). *Portaria nº 468 de 3 de abril de 2017*. [http://download.inep.gov.br/educacao\\_basica/enem/legislacao/2017/Portaria\\_mec\\_gm\\_n468\\_de\\_03042017\\_dispoe\\_sobre\\_a\\_realizacao\\_do\\_enem.pdf](http://download.inep.gov.br/educacao_basica/enem/legislacao/2017/Portaria_mec_gm_n468_de_03042017_dispoe_sobre_a_realizacao_do_enem.pdf)
- Monteiro, S. C., Almeida, L. S., & Vasconcelos, R. M. C. F. (2012). Abordagens à Aprendizagem, Autorregulação e Motivação: Convergência no Desempenho Acadêmico Excelente. *Revista Brasileira de Orientação Profissional*, *13*, 153-162.
- Organization for Economic Co-operation and Development (OECD) (2016). *Results from PISA 2015*. [http://download.inep.gov.br/acoes\\_internacionais/pisa/resultados/2015/pisa\\_2015\\_brazil.pdf](http://download.inep.gov.br/acoes_internacionais/pisa/resultados/2015/pisa_2015_brazil.pdf)
- Pangeni, K. P. (2014). Factors Determining Educational Quality: Student Mathematics Achievement in Nepal. *International Journal of Educational Research*, *34*(1), 30-41. <http://doi.org/10.1016/j.ijedudev.2013.03.001>
- Pinto, J., Carvalho e Silva, J., & Bixirão Neto, T. (2016). Fatores Influenciadores dos Resultados de Matemática de

- Estudantes Portugueses e Brasileiros no PISA: Revisão Integrativa. *Ciência e Educação*, 22(4), 837-853. <http://doi.org/10.1590/1516-731320160040002>
- Pipere, A., & Mierina, I. (2017). Exploring Non-Cognitive Predictors of Mathematics Achievement among 9<sup>th</sup> Grade Students. *Learning and Individual Differences*, 59, 65-77. <http://doi.org/10.1016/j.lindif.2017.09.005>
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <http://www.R-project.org>
- Rabelo, M. L. (2013). *Avaliação Educacional: Fundamentos, Metodologia e Aplicações no Contexto Brasileiro*. Rio de Janeiro, RJ: SBM.
- Rokach, L., & Maimon, O. (2015). *Data Mining with Decision Trees: Theory and Applications*. World Scientific Publishing.
- Schwartzman, S., Costin, C., & Coutinho, A. M. J. (2017). *Sociologia e Economia da Educação*. Rede Ciência para Educação.
- Souza, M. C. R. F., & Fonseca, M. C. R. F. (2010). *Relações de Gênero, Educação Matemática e Discurso: Enunciados sobre Mulheres, Homens e Matemática*. Autêntica.
- Therneau, T. M., & Atkinson, E. J. (2015). *An Introduction to Recursive Partitioning Using the rpart Routines*. <https://cran.rproject.org/web/packages/rpart/vignettes/longintro.pdf>
- Thien, L. M., & Ong, M. Y. (2015). Malaysian and Singaporean Students' Affective Characteristics and Mathematics Performance: Evidence from PISA 2012. *Springer Plus*, 4, 563. <http://doi.org/10.1186/s40064-015-1358-z>
- Zhang, H., & Singer, B. H. (2010). *Recursive Partitioning and Applications*. Springer. <https://doi.org/10.1007/978-1-4419-6824-1>