

EVELINE JÚNIA BRANT MARIZ

APLICAÇÃO DO MODELO DE REGRESSÃO LOGÍSTICA
EM DADOS DE RENDIMENTO DO EXAME NACIONAL
DO ENSINO MÉDIO - ENEM

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação Mestrado Profissional em Matemática em Rede Nacional, para obter o título de *Magister Scientiae*.

Orientador: Fernando de Souza Bastos

Coorientador: Guaraci de Lima Requena

Ficha catalográfica elaborada pela Biblioteca da Universidade Federal de Viçosa - Campus Florestal

T

Mariz, Eveline Júnia Brant, 1987-
B821a Aplicação do modelo de regressão logística em dados de
2021 rendimento do exame nacional do ensino médio - Enem [recurso eletrônico] / Eveline Júnia Brant Mariz. – Florestal, MG, 2021.
65: il. (algumas color.).

Orientador: Fernando de Souza Bastos.
Dissertação (mestrado) - Universidade Federal de Viçosa.
Referências bibliográficas: f.63-65.

1. Exame Nacional do Ensino Médio-Brasil.
2. Autorregressão-estatística. 3. Rendimento escolar.
I. Universidade Federal de Viçosa. Instituto de Ciências Exatas e Tecnológicas. Mestrado Profissional em Matemática em Rede Nacional. II. Título.

519.53

Bibliotecário(a) responsável: Maria Aparecida Alves de Oliveira 1174

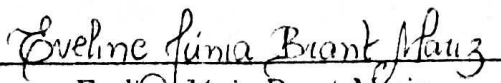
EVELINE JÚNIA BRANT MARIZ

APLICAÇÃO DO MODELO DE REGRESSÃO LOGÍSTICA
EM DADOS DE RENDIMENTO DO EXAME NACIONAL
DO ENSINO MÉDIO - ENEM

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação Mestrado Profissional em Matemática em Rede Nacional, para obter o título de *Magister Scientiae*.

APROVADA: 18 de junho de 2021

Assentimento:


Eveline Júnia Brant Mariz
Autora


Fernando de Souza Bastos
Orientador

Dedicatória

Dedico este trabalho ao meu filho Saulo, que me apresentou o amor mais genuíno do mundo e à minha mãe, pelas noites em claro, pela luta de vida em prol da minha formação e pelo amor incondicional.

Agradecimentos

Agradeço primeiramente a Deus, por me dar forças, discernimento e inspiração para insistir e persistir quando achei que poderia não ser capaz. À minha mãe pelas orações, incentivo e confiança. À minha irmã Lucelia pelo apoio e inspiração. Ao meu marido Anderson pelo apoio e amor incondicionais, me fazendo acreditar que eu era capaz. Ao meu cunhado Antônio Junior, pelo incentivo. Às minhas amadas sobrinhas Isabelle e Júlia e ao meu pequeno Saulo, pelo amor, carinho e sorrisos quando o peso do fardo insistia em me desanimar.

Agradeço aos professores pelo conhecimento a mim repassados ao longo dessa trajetória. Vocês foram fundamentais no meu crescimento pessoal e profissional.

E um agradecimento especial ao meu professor/orientador Fernando por todo o tempo dedicado a este trabalho. Sua orientação, motivação e cobrança foram fundamentais. Obrigada de coração!

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

“A educação tem raízes amargas, mas os seus frutos são doces.”
Aristóteles

Resumo

MARIZ, Eveline Júnia Brant, M.Sc., Universidade Federal de Viçosa, junho de 2021. **Aplicação do Modelo de Regressão Logística em Dados de Rendimento do Exame Nacional do Ensino Médio - Enem.** Orientador: Fernando de Souza Bastos. Coorientador: Guaraci de Lima Requena.

O Exame Nacional do Ensino Médio (Enem) foi aplicado pela primeira vez em 1998. Atualmente, o exame ocupa posição de destaque no sistema nacional de avaliação, composto também pelo Sistema de Avaliação da Educação Básica, conhecido como SAEB, e o Exame Nacional de Desempenho dos Estudantes (ENADE), além de possuir elevada aceitação por parte da sociedade e ser a principal ferramenta de seleção de alunos para ingresso no Ensino Superior. Diante desta importância, o Enem merece ser estudado, avaliado e debatido cientificamente através de diferentes metodologias e sob o prisma dos diversos ramos da ciência. Neste sentido, o objetivo geral deste trabalho foi avaliar a participação de variáveis de características pessoais, educacionais e socioeconômicas no rendimento dos candidatos na prova do Enem a partir de um modelo de regressão logística multinomial. O modelo estudado levou em consideração quinze variáveis identificadas no banco de dados do Enem e identificou uma relação sempre positiva entre o rendimento dos candidatos, a renda, a escolaridade da mãe e cursar o Ensino Médio em escola da rede privada. Por outro lado, foi identificada relação negativa entre o rendimento do candidato e características como idade e raça/cor preta. Além do trabalho inferencial, em que se identificou a participação das variáveis no rendimento dos candidatos, foi realizada uma análise descritiva de todo o banco de dados, determinando o Estado com maior número de concorrentes, o gênero dos candidatos e diversas outras características sociais, econômicas e fenotípicas.

Palavras-chave: Enem. Regressão. Multinomial. Rendimento.

Abstract

MARIZ, Eveline Júnia Brant, M.Sc., Universidade Federal de Viçosa, June, 2021. **Application of the Logistic Regression Model in Performance Data from the National High School Exam - Enem.** Adviser: Fernando de Souza Bastos. Co-adviser: Guaraci de Lima Requena.

The National High School Examination (Enem) was applied for the first time in 1998. Currently, the exam occupies a prominent position in the national assessment system, also composed of the Basic Education Assessment System, known as SAEB, and the National Examination of Basic Education. Student Performance (ENADE), in addition to being highly accepted by society and being the main tool for selecting students for entry into Higher Education. Given this importance, Enem deserves to be studied, evaluated and discussed scientifically through different methodologies and under the prism of the different branches of science. In this sense, the general objective of this work was to evaluate the participation of variables of personal, educational and socioeconomic characteristics in the candidates' performance in the Enem test using a multinomial logistic regression model. The model studied took into account fifteen variables identified in the Enem database and identified an always positive relationship between the candidates' income, income, mother's education and attending high school in a private school. On the other hand, a negative relationship was identified between the candidate's income and characteristics such as age and race / black color. In addition to the inferential work, in which the participation of the variables in the candidates' income was identified, a descriptive analysis of the entire database was also carried out, determining the state with the largest number of competitors, the gender of the candidates and several other social characteristics, economic and phenotypic.

Keywords: Enem. Regression. Multinomial. Performance.

Lista de Figuras

2.1 Logística do Enem. Fonte: INEP (2017)	16
4.1 Quantidade de Candidatos por Gênero	33
4.2 Média de idade por gênero	34
4.3 Média de idade por estado civil	35
4.4 Quantidade de candidatos por cor/raça	35
4.5 Quantidade de candidatos por unidade da federação	41
4.6 Número de candidatos a cada 1000 habitantes do Estado	42
4.7 Treemap da média das notas do Enem por Estado	48

Lista de Tabelas

3.1	Categorias da variável Nota	25
3.2	Variáveis independentes discriminadas de acordo suas respectivas características.	26
3.3	Descrição das variáveis independentes e seus respectivos níveis	27
3.4	Agrupamento de níveis de resposta	30
4.1	Candidatos por estado civil	34
4.2	Relação de candidatos por população	36
4.3	Escolaridade dos pais	37
4.4	Ocupação dos pais	38
4.5	Renda familiar dos candidatos	39
4.6	Renda média <i>per capita</i> dos candidatos	40
4.7	Tipo de escola do Ensino Médio	40
4.8	Candidatos por região do país	42
4.9	Quantidade e percentual de alunos por categoria de nota	43
4.10	Quantidade de alunos por categoria de nota por sexo	43
4.11	Quantidade de candidatos do estado civil por categoria de nota	44
4.12	Quantidade de candidatos da cor/raça por categoria de nota	44
4.13	Percentual de candidatos nas categorias por renda	45
4.14	Representatividade relativa da escola do Ensino Médio por categoria de nota .	46
4.15	Nota média dos candidatos em relação à escolaridade dos pais	47
4.16	Nota média dos candidatos em relação à ocupação dos pais	47
4.17	Categorias base das variáveis dependentes	49
4.18	Resultado de regressão - categoria 0	51
4.19	Resultado de regressão - categoria 1	54
4.20	Resultado de regressão - categoria 3	55
4.21	Resultado de regressão - categoria 4	57
4.22	Resultado de regressão - categoria 5	58
4.23	Impacto das características na nota	59

Sumário

1	Introdução	11
2	Revisão Bibliográfica	13
2.1	O Exame Nacional do Ensino Médio	13
2.2	Estudos sobre o Enem e determinantes do desempenho escolar	16
2.3	Modelos de regressão	18
2.3.1	Regressão logística	19
3	Materiais e Métodos	24
3.1	Base de dados e seleção de variáveis	24
3.1.1	Seleção de variáveis independentes	25
3.2	Estatística descritiva com <i>Shiny</i>	29
3.3	Modelo de regressão multinomial	30
4	Resultados e Discussão	33
4.1	Análise descritiva dos dados	33
4.1.1	Análise geral dos candidatos	33
4.1.2	Análise das notas dos candidatos	42
4.2	Análise do modelo de regressão	48
5	Conclusões	61

Introdução

Aplicado pela primeira vez em 1998, o Exame Nacional do Ensino Médio (Enem) passou por diversas modificações estruturais e, atualmente, ocupa posição de destaque no sistema nacional de avaliação educacional, sendo bem aceito pela sociedade e se tornando a principal ferramenta de seleção de alunos para ingresso no Ensino Superior.

Diante de suas características e magnitude, o exame é um sistema de avaliação em larga escala, na medida em que influencia no currículo das escolas do Ensino médio e contribui para a implantação de políticas públicas, visando a melhoria do ensino.

Seja como uma medida de avaliação da qualidade do Ensino Médio que possibilita melhorar a gestão desse nível de ensino ou como uma ferramenta de acesso ao Ensino Superior, a importância do Enem na vida dos estudantes brasileiros é inegável e pode ser atestada em números. Entre 2010 e 2019, mais de 65 milhões de estudantes confirmaram a inscrição no exame, sendo que em 2018, ano escolhido para avaliação neste trabalho, foram mais de 5,5 milhões de inscritos. Nesse sentido, estudar, avaliar e debater cientificamente tal Exame usando diferentes metodologias e sob o prisma dos diversos ramos da ciência é de extrema importância para a melhoria e consolidação deste Exame.

Nesta pesquisa foi realizada uma análise dos dados do Enem de 2018 divulgados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), visando identificar o quanto características pessoais, educacionais e socioeconômicas dos candidatos são capazes de determinar a nota desses participantes. Para esta análise foi utilizado o modelo de Regressão Logística Multinomial, que é adotado quando a variável dependente é do tipo categórica e possui três ou mais categorias.

O objetivo geral foi, por meio de um modelo logístico de regressão, avaliar a participação de variáveis de características pessoais, educacionais e socioeconômicas no rendimento dos candidatos na prova do Enem. Foi realizada ainda uma revisão de literatura acerca dos eixos temáticos e metodológicos, além de uma análise descritiva das informações apresentadas no banco de dados do exame. Assim, espera-se contribuir para o debate em torno da qualidade da educação e das necessidades de aprimoramento nos mecanismos que a avaliam.

O presente trabalho, além dessa breve introdução, é composto por outras quatro seções resumidamente descritas a seguir.

A segunda seção apresenta a revisão bibliográfica e foi subdividida em três subseções.

A primeira delas aborda o Enem, fazendo uma contextualização e um breve histórico do exame desde a sua criação até os dias atuais, destacando as principais mudanças e desafios pelos quais o exame passou até se tornar essencial para o acesso ao Ensino Superior no Brasil. A segunda subseção apresenta alguns estudos anteriores sobre o Enem e os fatores que possam determinar os resultados dos candidatos no exame. A terceira subseção aborda os aspectos conceituais e teóricos da regressão logística, sobretudo, em sua forma multinomial, que é a utilizada para ajuste do modelo do presente estudo.

A terceira seção aborda os aspectos metodológicos do presente estudo como aqueles relacionados a composição da base de dados, a categorização das notas, a definição das covariáveis, a avaliação dos modelos, bem como a escolha do melhor modelo ajustado.

Na quarta seção é realizada uma análise descritiva dos dados dos candidatos e são demonstrados os resultados obtidos a partir da aplicação do modelo de regressão na base de dados, além de ser realizada uma discussão acerca das principais considerações obtidas por meio das avaliações realizadas no modelo ajustado.

Por fim, a quinta seção apresenta brevemente as conclusões finais referentes a este estudo.

Revisão Bibliográfica

2.1 O Exame Nacional do Ensino Médio

Aplicado pela primeira vez em 1998, o Enem passou por diversas alterações ao longo dos anos, passando de um exame que visava exclusivamente a avaliação dos estudantes do Ensino Médio para uma importante ferramenta de seleção de alunos para ingresso no Ensino Superior. Para Travitzki (2013), a concepção inicial do Enem previa avaliar os alunos quanto ao raciocínio, as habilidades em resolver problemas e suas competências gerais para o exercício da cidadania e das atividades profissionais e acadêmicas. E esta opinião é corroborada por Silveira et al. (2015) ao afirmar que por mais de 10 anos o Enem não objetivou selecionar estudantes para o ingresso no Ensino Superior, visou única e exclusivamente avaliar as competências e habilidades dos alunos que concluíam o Ensino Médio.

Em 1999, o Enem passou a ser utilizado por Instituições de Ensino Superior como critério de acesso aos cursos de graduação. As primeiras instituições a utilizarem o exame foram a Pontifícia Universidade Católica do Rio de Janeiro (PUC-RJ) e a Universidade Federal de Ouro Preto (UFOP) (INEP, 2018).

Em 2004, com a criação do Programa Universidade para Todos (ProUni), o resultado do Enem passou a ser utilizado como critério para concessão de bolsas de estudo integral ou parcial em cursos de graduação de instituições privadas. Esta, além da avaliação do Ensino Médio, foi a principal função do Enem até o ano de 2009, quando foi criado o Sistema de Seleção Unificada (Sisu), que passou a utilizar as notas do Enem na seleção de alunos para os cursos de graduação das instituições federais de Ensino Superior.

Nesse período houve grandes transformações na prova do Enem. Aplicação de provas passou a ser em dois dias; as 180 questões passaram a ser classificadas dentro de quatro áreas de conhecimento: Ciência Humanas, Ciências da Natureza, Matemática e Linguagens e Códigos e foi implantada a Teoria de Resposta ao Item (TRI) para correção das provas, visando estimar a dificuldade dos itens e as habilidades de cada candidato. A partir desse ano, o exame também passou a ser utilizado para certificação de conclusão do Ensino Médio, dentro de critérios específicos (INEP, 2018).

A implantação do Sisu em 2009 foi um importante passo para o Enem chegar ao que é hoje, mas foi apenas em 2013 que, pela primeira vez, todas as instituições federais de ensino superior passaram a utilizar o Enem como critério de seleção de novos alunos. Nesse mesmo ano, a nota do exame também passou a ser utilizada na concessão de bolsas de estudos do Programa Ciência sem Fronteiras (INEP, 2018). A partir daí ainda houve muitos avanços e muitas mudanças em sua estrutura e logística., sendo que, nos dias atuais, o exame ocupa posição de destaque no sistema nacional de avaliação educacional (composto também por SAEB e ENADE), além de possuir elevada aceitação por parte da sociedade.

Para Travitzki (2013), dentre os fatores que impulsionaram o Enem a ocupar esta posição de destaque aponta-se como possibilidade o fato de a prova focar mais no raciocínio e não em ser um exame puramente conteudista. Outro ponto que proporcionou a elevação da importância do exame foi ter se consolidado como a principal alternativa ao modelo tradicional de vestibular. Apenas no início de 2019, através do Sisu, foram ofertadas 235.476 vagas para ingresso em 129 instituições públicas de ensino em todo o Brasil. Segundo o Ministério da Educação (MEC), em 2017 um total de 1.481 instituições de ensino (60% do total) utilizaram as notas do Enem para selecionar alunos ingressantes em cursos de graduação (MAIA, 2018).

Um aspecto positivo apontado pelos estudiosos é que o Exame possibilitou, principalmente a partir da implementação do Sisu, uma maior mobilidade entre os estudantes no território nacional. Esta mobilidade maximizou o intercâmbio cultural entre as diferentes regiões do país e aumentou as possibilidades de alunos egressos de regiões menos privilegiadas economicamente se deslocarem para regiões mais desenvolvidas e ingressarem nas universidades mais concorridas (SILVEIRA et al., 2015).

Há, contudo, que se balancear os aspectos positivos do intercâmbio de alunos das diferentes regiões do país com eventuais efeitos colaterais negativos, como, por exemplo, a migração de alunos dos estados mais ricos para estados considerados mais pobres. Além da concorrência de estudantes da mesma unidade da federação, os estudantes das regiões mais pobres acabam concorrendo com alunos que vieram de regiões mais ricas e, em tese, passaram por uma melhor preparação no decorrer dos estudos.

Assim como é possível destacar aspectos positivos a partir do Sisu, observa-se um ponto negativo também a partir de sua implementação: a ocorrência de um aumento da evasão de alunos do Ensino Superior (LI et al., 2017). Um dos motivos prováveis para esse aumento é que o Sisu possibilita o acesso ao curso a partir da nota. Assim, muitas vezes os estudantes acabam ingressando em cursos que não conhecem ou não refletem sua vocação e abandonam ao perceberem que não desejam aquela formação ou quando encontram as primeiras dificuldades. Por outro lado, há quem defenda o Sisu como uma ferramenta moderna que possibilita o acesso dos estudantes ao Ensino Superior de uma forma democrática e através do mérito.

Os 20 primeiros anos de elaboração e aplicação do Enem também foram marcados por polêmicas e problemas. Algumas dessas polêmicas repercutiram amplamente na mídia nacional e motivaram ajustes nas edições subsequentes como: o vazamento de dados dos candidatos das edições de 2007, 2008 e 2009; falhas nas correções das redações, como aprovação de alunos que zombaram do tema e da banca de correção (caso da receita do

miojo); o furto e vazamento de questões e cadernos de provas (talvez o problema mais grave da história do exame); e, mais recentemente, um episódio de erros de correção das provas que provocou alterações nas notas de aproximadamente 6.000 candidatos. Esses, e outros problemas detectados e combatidos possibilitaram o amadurecimento dos atores envolvidos e contribuíram para o aperfeiçoamento nas etapas de elaboração, gerenciamento e aplicação do exame.

Um conjunto de críticas ainda permeia sobre a realização do exame sendo talvez a mais pertinente aquela que aponta que o Enem, no decorrer dos anos, teve seu foco desvirtuado deixando de ser uma avaliação do Ensino Médio e se tornando um grande vestibular nacional. Neste sentido, a partir de 2009, as questões teriam passado a valorizar mais o conhecimento de determinados conteúdos pré-estabelecidos e menos a capacidade de raciocínio e resolução de problemas e de tomada de decisões.

Desta forma, o exame teria se fortalecido enquanto mecanismo de avaliação dos indivíduos (alunos) e estaria perdendo a capacidade de avaliar a qualidade das instituições (escolas) e do ensino como um todo, além de não empregar esforços em melhorar o rendimento dos alunos com desempenho regular ou baixo, podendo, inclusive, contribuir para aumentar a distância entre os alunos com melhores e piores desempenhos. Assim, apesar da manutenção do nome, a prova do Enem parece gradualmente deixar de se destinar a avaliar a educação no Ensino Médio e ter se transformado em um Exame de Acesso ao Ensino Superior, um grande vestibular nacional.

Em contraponto a isso, Machado et al. (2014) argumentam que um dos objetivos do Enem é, gradualmente, provocar transformações na educação brasileira na medida em que promove discussões sobre qualidade da educação dentro das próprias escolas, entre educadores e alunos e não apenas por meio de políticas centrais, repensando o currículo e a metodologia aplicada em relação àquela preconizada pela reforma do Ensino Médio.

Observa-se, contudo, que mesmo diante das críticas e divergências de opinião em relação ao Exame, é inegável a importância do mesmo como ferramenta de acesso às universidades e isso pode ser comprovado em números. Apenas nas 10 últimas edições (entre 2010 e 2019), mais de 65 milhões de estudantes confirmaram a inscrição no exame. No mesmo período, apenas a edição de 2010 não contou com mais de 5 milhões de participantes.

No que se refere à logística de aplicação do exame, necessita-se de uma estrutura robusta para garantir a segurança e lisura do certame. Este processo envolve aproximadamente 500 mil pessoas. Apenas em 2018, a impressão dos 11 milhões de cadernos de questão consumiu duas mil toneladas de papel e envolveu 600 funcionários em um período de 2 meses. Sob escolta policial, as provas armazenadas em malotes com lacres físicos e eletrônicos são distribuídas para batalhões do Exército em todas as unidades da federação, sendo que a abertura dos malotes ocorre apenas no dia e local de aplicação das provas, posteriormente ao fechamento dos portões e na presença de um representante do INEP. Após a realização das provas é necessário a aplicação de uma logística reversa visando possibilitar as etapas de correção das provas e divulgação dos resultados. O infográfico a seguir, elaborado pelo INEP, apresenta o desafio logístico no qual consiste a aplicação do Exame:



Figura 2.1: Logística do Enem. Fonte: INEP (2017)

2.2 Estudos sobre o Enem e determinantes do desempenho escolar

Alguns estudos já se ocuparam em analisar o Enem, seja enquanto ferramenta de avaliação do Ensino Médio ou enquanto mecanismo de ingresso no Ensino Superior. Outras pesquisas objetivaram avaliar as variáveis determinantes do desempenho escolar dos estudantes. Considerando que o objetivo geral do presente estudo consiste em, através de um modelo de regressão logística, identificar os principais fatores determinantes do desempenho dos estudantes na prova do Enem, os dois grupos de estudos mencionados anteriormente devem ser levados em consideração e nesta seção faz-se uma avaliação sobre uma amostra dos mesmos.

Viggiano et al. (2013a) utilizam os dados do Enem 2010 disponibilizados pelo INEP para analisar o desempenho de 847.621 estudantes distribuídos por todas as regiões do Brasil. Os autores concluíram por haver, quando comparadas as diferentes regiões do país, três agrupamentos de desempenho: superior (regiões Sul e Sudeste), médio (região Centro-Oeste) e inferior (regiões Norte e Nordeste).

Mont'Alvão et al. (2013) analisou as chances dos estudantes brasileiros ingressarem no Ensino Superior desde a década de 80 do século XX até o final da primeira década do século XXI. Como nos outros estudos, seus resultados indicaram influência relevante das origens socioeconômicas dos estudantes no desempenho do Enem de modo que são determinantes para a transição do indivíduo para o Ensino Superior. O estudo observou ainda que os fatores socioeconômicos são mais relevantes para a educação acadêmica,

quando comparada a tecnológica e que a variável renda familiar é determinante para o ingresso nos cursos com maior potencial de retorno financeiro. Ao longo do período, observou-se uma redução na influência do nível educacional dos pais, assim como uma queda do impacto das desigualdades raciais. No tocante às áreas de conhecimento, os resultados sugerem uma importância da variável sexo e apontam que existe uma concentração de mulheres nas áreas das ciências humanas e da saúde.

Figueiredo et al. (2014) procuraram analisar o grau de igualdade de oportunidades para o desempenho educacional dos estudantes que prestaram o Enem, considerando fatores sociais como origem familiar, renda, escolaridade dos pais, entre outros. Os autores observam que para estar entre os 5% das maiores notas os estudantes com baixo *background* social (filhos de mães analfabetas, com renda familiar de até 1 salário mínimo e que estudaram predominantemente em escola pública) precisam se esforçar em torno de 99,38% a mais do que aqueles com alto *background* familiar (filhos de mães com nível superior de escolaridade, renda familiar acima de 9 salários mínimos e que estudou predominantemente em escola particular). Os autores destacam ainda que quando observados os estudantes do Nordeste, onde a proporção de indivíduos com baixo *background* familiar é maior, os resultados se tornam ainda mais expressivos e indicam uma baixa mobilidade intergeracional educacional e propõem que uma parcela dos estudantes brasileiros é sumariamente excluída do estrato social superior.

Silveira et al. (2015), utilizando os dados do Enem de 2013 avaliaram o poder explicativo das variáveis socioeconômicas e geográficas para 14.715 escolas brasileiras. No estudo, os autores utilizam a média aritmética dos escores médios de proficiência nas provas de ciências da natureza, ciências humanas, linguagens e códigos, matemática e redação, visando apurar o Escore da Escola no Enem (EEE). O modelo elaborado pelos autores indica que a variável socioeconômica exclusivamente explica 66% da variação do EEE. Por sua vez, a variável geográfica explica 13% e um modelo conjugando as variáveis socioeconômica e geográfica explica em 71% o comportamento do Escore da Escola no Enem.

Haguette et al. (2016) avaliam as escolas do Ensino Médio da rede pública do Estado do Ceará com melhor desempenho e também aquelas com rendimento mais fraco no Enem de 2011, objetivando descobrir o que as escolas com bom desempenho possuíam que as diferenciavam daquelas com desempenho ruim. Os autores se valeram de visitações às escolas, entrevistas com a comunidade escolar (membros de núcleo gestor, alunos, professores e outros funcionários) e aplicação de questionário para construir um “retrato falado” de cada escola. Por fim, concluíram que as escolas com melhor desempenho se caracterizam predominantemente por: desejo de aprender dos alunos; autodisciplina coletiva; otimização do tempo letivo; ensino em tempo integral; estágio supervisionado; núcleo gestor dedicado ao planejamento e ao eixo pedagógico; maior parte dos professores trabalham em uma única escola e não exercem outra atividade; infraestrutura boa quando comparada às demais escolas da rede pública, contudo ainda insatisfatória; clima escolar positivo. Em contrapartida, identificaram que as escolas com desempenho fraco têm características predominantemente inversas as do grupo anterior, como: alunos não desejam aprender; indisciplina; erosão do tempo letivo; escola de turno único; núcleo gestor desperdiça parcela do tempo para tratar de questões disciplinares dos alunos; maior parte dos professores

temporários (contratados de forma precária); infraestrutura nitidamente insuficiente; clima escolar pobre, com acúmulo de aspectos negativos. Assim, o estudo verificou a coexistência de dois padrões de escolas mutuamente excludentes e irreconciliáveis.

Boneti et al. (2017) buscaram compreender como questões relacionadas a localização geográfica das escolas e a situação social e econômica das mesmas podem interferir nesse resultado. Avaliando os dados das edições do Enem de 2009 a 2013, observaram que escolas na região Nordeste ocupam, com maior frequência, posições consideradas de fragilidades no *ranking* das instituições. O estudo verificou ainda que escolas privadas, bem como escolas públicas federais, não figuraram entre as com pior desempenho no Enem em nenhuma das edições abrangidas pelo estudo. Diante dos dados, os autores concluíram que as desigualdades sociais e também as desigualdades regionais constituem uma explicação teórica plausível para o desempenho das escolas.

Pereira (2019), avaliando os dados dos estudantes das escolas das redes públicas e privadas de Ribeirão Preto que participaram do Enem em 2015, concluiu que alunos de grupo com nível socioeconômico mais alto possuem melhor desempenho no exame quando comparado aos candidatos com nível socioeconômico mais baixo, sendo o resultado do Enem um reflexo do nível socioeconômico dos candidatos. O trabalho destaca que 78% dos estudantes que tiraram as notas mais baixa (inferior a 450) declararam renda igual ou inferior a R\$ 2.364,00, enquanto 53% dos participantes que obtiveram notas mais elevadas (superior a 750) declararam renda familiar igual ou superior a R\$ 9.456,00. O estudo, contudo, questiona a efetividade do exame enquanto medição da qualidade do ensino e alega que o Enem possui caráter voluntário e existe uma diferença declarada entre a motivação dos alunos das escolas públicas e privadas quando da opção pela realização da prova. Nesta linha de raciocínio, o autor argumenta que apenas o *ranking* das notas do Enem é um descritor insuficiente da qualidade do ensino pois existem escolas que preparam o aluno para a realização do exame, mas não para a vida.

2.3 Modelos de regressão

O termo regressão foi encontrado pela primeira vez em 1885 na obra do antropologista Francis Galton, em que ele avaliou a relação entre a altura de pais e filhos. Galton verificou que pais altos tendiam a ter filhos altos e pais baixos tendiam a ter filhos baixos, entretanto, a altura dos filhos de pais mais altos ou mais baixos que o padrão tendia a mover-se em direção à altura média da população, o que ele chamou de *Regression toward the mean* (regressão para a média). (HOSMER JR et al., 2013)

Após anos de evolução do conceito e da ciência, é possível conceituar a análise de regressão como uma importante ferramenta estatístico-matemática que permite verificar o comportamento de determinadas variáveis, dadas as variações em outras. Ou seja, ela permite entender a influência que uma variável pode provocar nas demais. Há, contudo, que se observar que a relação entre duas ou mais variáveis pode ser de dependência funcional (quando o comportamento de uma variável depende do comportamento de outra) ou de mera associação (quando, apesar de oscilarem em conjunto, nenhuma variável pode ser considerada dependente de outra).

Hoje, modelos de regressão são utilizados em trabalhos dos mais variados ramos da ciência e são encontrados em praticamente todas as áreas do conhecimento, desde estudos meteorológicos visando a previsão da chuva em determinada região do globo terrestre até análises econômico-financeiras objetivando prever os resultados de companhias imobiliárias em determinado período. Em síntese, sempre que se tem por objetivo avaliar o comportamento de uma ou mais variáveis em detrimento do comportamento de outras, o modelo de regressão apresenta-se como uma alternativa possível. Como o presente trabalho se propõe a identificar como determinadas variáveis interferem nos candidatos do Enem, a regressão se torna uma importante aliada neste processo.

Existem diversas maneiras de se aplicar um modelo de regressão, dependendo do tipo de variável dependente (Regressão linear ou logística) e da quantidade de covariáveis – ou variáveis independentes – (Simples ou Múltipla). Apesar de o modelo mais comumente utilizado e conhecido ser a Regressão Linear, o recurso utilizado para análise dos dados deste trabalho foi o modelo de regressão logística.

2.3.1 Regressão logística

De acordo com Figueira (2006), diferente dos modelos de regressão linear, os modelos de regressão logística são caracterizados por possuírem a variável dependente do tipo categórica, podendo esta ser ordinal (quando existe uma relação de ordem entre as categorias da variável dependente) ou nominal (quando não existe essa relação de ordem entre as categorias ou a mesma é irrelevante).

Este modelo é interessante quando estão sendo trabalhadas variáveis que não possuem natureza contínua ou quando se deseja categorizar uma determinada variável para uma avaliação diferenciada, como é o caso deste estudo. Neste caso, a natureza das variáveis independentes (ou covariáveis) pode ser tanto qualitativa quanto quantitativa, tendo-se apenas a exigência de natureza categórica para a variável dependente.

Uma outra peculiaridade da regressão logística em relação à regressão linear (além da variável dependente dicotômica) é que a regressão logística não admite o método dos mínimos quadrados para a estimativa dos coeficientes de regressão. Segundo Maroco (2007), como os erros do modelo de regressão logística não seguem a distribuição normal, não é possível obter, por meio do método dos mínimos quadrados, estimativas ótimas dos coeficientes, como é feito na regressão linear. Na regressão logística o método usado para o ajustamento do modelo é chamado de Método da Máxima Verossimilhança (Maximum Likelihood) e estima os coeficientes de regressão para o modelo que maximizem a probabilidade de se encontrar as realizações da variável dependente. Neste método, não existe uma equação analítica que forneça os valores dos coeficientes, sendo o mesmo obtido a partir de estimativas iterativas realizadas por um algoritmo computacional (comumente é utilizado o método de Newton-Raphson). Além dos coeficientes, após atingido o ponto final da iteração, o software fornece as estatísticas necessárias para avaliar a significância dos coeficientes de regressão e a qualidade do modelo.

De acordo com quantidade de categorias da variável dependente ainda é possível dividir a regressão logística em dois modelos: o binário e o multinomial.

O modelo de regressão logística binário trabalha com variável dependente do tipo

binário (variáveis dummy). Assim, esse modelo se torna particularmente interessante para estudar questões que possam ser respondidas com sim ou não. Segundo Hair, Black, Babin e Anderson (2014), a regressão logística tem ampla aplicação em situações nas quais o principal objetivo é identificar o grupo ao qual pertence um objeto. As aplicações potenciais incluem prever situações em que o resultado é binário. Essas situações incluem o sucesso ou fracasso de um produto novo, decidir se uma pessoa deve ser concedida crédito ou prever se uma firma será bem sucedida. Em cada instância, os objetos caem em um dos dois grupos, e o objetivo é prever e explicar as bases para cada membro do grupo de objetos através de um conjunto de variáveis independentes selecionado pelo investigador.

Além da regressão logística binária, ainda existem a regressão logística multivariada, que é avaliada da mesma maneira que a binária, mas considerando a existência de mais de uma variável independente, e a regressão logística multinomial.

O modelo de regressão multinomial é uma generalização do modelo logístico binário, em que a variável dependente possui mais de duas categorias.

Considerando uma variável dependente aleatória Y com $q + 1$ categorias, tem-se que o modelo representando as chances de ocorrência de uma categoria em relação a categoria de referência é definido por q equações.

Segundo Hosmer Jr et al. (2013), o modelo logístico multinomial com n variáveis explicativas, definido pela função *logit* é:

$$\begin{aligned} g_i(x) &= \ln \left[\frac{\mathbb{P}(Y = i|X)}{\mathbb{P}(Y = 0|X)} \right] \\ &= \beta_{i0}X_{i0} + \beta_{i1}X_1 + \beta_{i2}X_2 + \dots + \beta_{in}X_n \\ &= x\beta_i, i = 0,1,2, \dots, q. \end{aligned} \tag{2.1}$$

onde $\beta_i = (\beta_{i0}, \dots, \beta_{in})$ e $X_{i0} = 1$.

A probabilidade de uma variável dependente Y tomar o valor de qualquer uma das $q + 1$ categorias existentes é calculada por:

$$\mathbb{P}(Y = i|X) = \frac{e^{g_i(x)}}{\sum_{i=0}^q e^{g_i(x)}} \tag{2.2}$$

onde $g_i(x)$ é dado pela expressão (2.1), para $i \in 1, 2, \dots, q$ e $g_0(x) = 0$.

A partir da expressão (2.1), com a utilização de propriedades logarítmicas, obtém-se que a chance de ocorrência de uma determinada categoria, dada a categoria de referência será de:

$$\frac{\mathbb{P}(Y = i|X)}{\mathbb{P}(Y = 0|X)} = e^{X\beta_i}, i = 1,2, \dots, q. \tag{2.3}$$

Uma vez calculadas as probabilidades de ocorrência é possível realizar a análise do *Relative Risk Ratios - RRR* (Razão de Risco Relativo), que refere-se a razão entre as

probabilidades de ocorrência de uma determinada categoria i em relação à categoria de referência.

Assumindo a variável dependente $Y = i$ em comparação a categoria de referência ($Y = 0$), o RRR considerando a covariável com valores $x = a$ e $x = b$ é:

$$RRR_i(a,b) = \frac{\mathbb{P}(Y = i|x = a)/\mathbb{P}(Y = 0|x = a)}{\mathbb{P}(Y = i|x = b)/\mathbb{P}(Y = 0|x = b)} \quad (2.4)$$

Quando RRR_i assumir valor maior do que 1, maior será a chance de um indivíduo com características $x = a$ em relação à indivíduos com características $x = b$ estar na categoria i , comparada à categoria de referencia 0. Desta maneira, a razão de risco relativo compara se a probabilidade de ocorrência de um dado evento é a mesma para dois grupos distintos de variável explicativa. Maiores informações ver (ANDRADE, 2006).

A estimação dos parâmetros do modelo logístico é realizada pelo Método da Máxima Verossimilhança (POWERS et al., 2008) que consiste em encontrar um estimador $\hat{\beta}$ que maximize a função de verossimilhança. Considerando a variável aleatória Y com Q valores possíveis, sendo que cada categoria q possui dois resultados possíveis: ($y_{iq} = 1$ se a categoria for a escolhida e $y_{iq} = 0$, caso contrário) e considerando $p_{iq} = \mathbb{P}(Y_i = q|X)$, a função de verossimilhança da variável Y é dada por:

$$L(\beta) = \prod_{i=1}^n \frac{e^{x_i\beta_q}}{Q-1 + \sum_{q=1}^{Q-1} e^{x_i\beta_q}} \quad (2.5)$$

Segundo Maroco (2007), o modelo de regressão logística deve obedecer a alguns pressupostos: Linearidade e Aditividade, Proporcionalidade, Constância de efeito, Erros independentes e com distribuição binomial e Não multicolinearidade dos preditores. Observados os pressupostos e uma vez ajustado mediante o método da máxima verossimilhança, deve-se avaliar a significância e a qualidade do modelo ajustado, assim como a significância dos coeficientes de regressão logística.

A significância do modelo ajustado pode ser verificada a partir do teste da Razão de Verossimilhanças, em que são testadas as hipóteses $H_0 : \beta_1 = \beta_2 = \dots \beta_q = 0$ (modelo não será estatisticamente significativo) e $H_1 : \exists i : \beta_i \neq 0$ ($i = 1, \dots, q$) (modelo será estatisticamente significativo, em que é possível rejeitar a hipótese nula).

Quando o modelo não for estatisticamente significativo, o mesmo não permite prever as probabilidades de sucesso ou fracasso a partir das variáveis independentes do modelo.

Por outro lado, a rejeição da hipótese nula implica a existência de pelo menos uma variável independente relacionada linearmente com o *Logit*. Assim, comumente utiliza-se o teste de Wald para identificar qual ou quais, dentre as variáveis independentes, influencia(m) o resultado do *Logit* de forma significativa. O teste consiste, basicamente, em verificar se um determinado coeficiente é nulo, considerando a variação estimulada dos demais coeficientes.

A estatística de teste é:

$$T_{Wald_i} = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}, \quad (2.6)$$

onde $\hat{\beta}_i$ é o estimador de β_i e $SE(\hat{\beta}_i)$ é o estimador do erro padrão de $\hat{\beta}_i$.

Na regressão logística, assim como na regressão linear, também é possível avaliar a qualidade do modelo de regressão logístico a partir da dimensão do efeito do modelo. Na regressão linear é utilizado o coeficiente de determinação, também chamado R^2 , para mensurar o efeito que as variáveis independentes provocam sob a variável dependente, contudo, na regressão logística o R^2 não se aplica da mesma maneira. Para os modelos generalizados, como o modelo logístico, foram desenvolvidos os chamados *pseudo- R^2* (*pseudo R-squared*) que devem ser interpretados de forma análoga ao R^2 . O R^2 de Cox & Snell, o de Nagelkerke e o de McFadden são alguns dos pseudo- R^2 mais conhecidos e bem desenvolvidos.

O R^2 de McFadden mede a qualidade do ajuste dos modelos estimados e é definido como:

$$R_{MF}^2 = 1 - \frac{\log(L_{UR})}{\log(L_R)}, \quad (2.7)$$

onde $\log(L_{UR})$ é máximo da função de verossimilhança da regressão não restrita e $\log(L_R)$ é o máximo da função de verossimilhança da regressão restrita (modelo nulo) quando $\beta_1 = \beta_2 = \beta_3 = \dots = \beta_q$.

Para a definição de variáveis que irão compor o modelo existem critérios matemáticos ou métodos que auxiliam a escolha. Os principais métodos são o forward, o backward e o Stepwise.

O método Forward consiste em no primeiro momento apenas a constante β_0 integrar o modelo. Acrescenta-se ao modelo a variável independente que apresenta maior potencial explicativo. Caso o ajuste do modelo aumente significativamente, a variável é mantida e repete-se o procedimento para a próxima variável independente até que não seja incluída mais nenhuma variável explicativa no modelo.

Já no método Backward, de modo oposto ao método forward, inicia-se o modelo com todas as variáveis independentes. A cada etapa, os previsores que não apresentam contribuição significativa para o modelo são retirados até que restem apenas variáveis com contribuição significativa.

O método Stepwise é uma modificação do Forward em que cada passo todas as variáveis. Inicia-se o modelo com uma variável, mas a cada passo do forward em que se inclui uma nova variável é realizado o backward para verificar a necessidade de descarte de alguma variável.

Existem diversas variações dos modelos forward e backward aplicáveis aos modelos de regressão logística (como os critérios baseados na razão de verossimilhança ou no teste de Wald). Além desses critérios matemáticos, a experiência do pesquisador, o conhecimento prévio sobre as variáveis do estudo e hipóteses teóricas relacionadas ao comportamento

das variáveis do estudo são levados em consideração quando da seleção de variáveis independentes para o modelo. Todos esses componentes é que irão determinar um modelo satisfatório e bem ajustado.

Materiais e Métodos

3.1 Base de dados e seleção de variáveis

Os dados para avaliação do desempenho dos candidatos do Enem foram colhidos da base de microdados do Enem 2018 disponibilizada pelo INEP. Nesta base consta a relação de todos os inscritos na referida edição do exame (por número de inscrição) com seus respectivos dados pessoais, respostas do questionário socioeconômico e as notas em cada uma das áreas de conhecimento.

Na edição de 2018 foram 5.513.747 inscritos no exame, sendo que apenas 70,62% desses candidatos compareceram para todas as etapas da prova.

As 27 perguntas do questionário socioeconômico e as informações pessoais, educacionais e gerais de cada aluno estão distribuídas nas colunas do banco de dados e são subdivididas de acordo com característica da informação, conforme abaixo:

- a. Dados do Participante;
- b. Dados da Escola;
- c. Dados dos Pedidos de Atendimento Especializado;
- d. Dados dos Pedidos de Atendimento Específico;
- e. Dados dos Pedidos de Recursos Especializados e Específicos para a realização da prova;
- f. Dados do Local de Aplicação da Prova;
- g. Dados da Prova Objetiva;
- h. Dados da Redação;
- i. Dados do Questionário Socioeconômico.

Inicialmente, diante de um banco de dados gigantesco, com mais de 5 milhões de linhas e 137 colunas, foi necessário realizar uma “limpeza”, retirando da base todos os candidatos que não compareceram em pelo menos uma das provas. Optou-se pela exclusão desses candidatos porque sua manutenção na base comprometeria a avaliação do modelo, uma vez que a inexistência de alguma nota impediria o cálculo da nota final e sua consequente categorização. Assim, foram mantidas 3.893.743 linhas, referente aos candidatos que compareceram em todas as etapas do exame.

O cálculo da nota final de cada aluno é realizado com base na média ponderada das notas de todas as áreas do conhecimento, cabendo a cada instituição definir o peso de cada nota, de acordo com o curso desejado pelo candidato ou pela área do conhecimento.

No presente estudo, o cálculo da nota final, em conformidade com o critério utilizado pela Universidade Federal de Viçosa (UFV) para ingresso no ano subsequente a aplicação do exame, considerou peso 1 para todas as áreas do conhecimento, independente do curso almejado pelo candidato. Assim, a nota final de cada candidato foi calculada conforme a média ponderada (ou aritmética, no caso) das notas obtidas na redação e nas quatro áreas do conhecimento, a saber, Ciências da Natureza, Ciências Humanas, Linguagens e Códigos e Matemática.

Após o cálculo da nota final, foi realizada a categorização dessas notas em seis grupos. Essa categorização é importante porque, em se tratando de Enem, não existe uma nota pré-definida em que o candidato está aprovado ou reprovado. O que define a sua entrada em uma universidade, além da nota, é o curso escolhido, a concorrência, o tipo de universidade, entre outros motivos. Como o foco deste trabalho foi avaliar o rendimento do aluno e não a sua aprovação ou reprovação no exame, a categorização das notas conforme descrita na Tabela 3.1 se apresentou como mais adequado à análise.

Tabela 3.1: Categorias da variável Nota

Categoria	Nota
0	0,00 † 406,98
1	406,98 † 497,22
2	497,22 † 587,46
3	587,46 † 677,70
4	677,70 † 767,94
5	767,94 † 858,18

Fonte: Elaborado pela autora.

O limite superior da categoria 0 (zero) foi determinado a partir da menor nota obtida por um ingressante na UFV em 2019 e o limite inferior foi a menor nota registrada no Enem no mesmo ano. A amplitude das categorias 1 a 5 foi definida pela diferença entre a maior nota registrada no Enem 2018 e o limite superior da categoria 0 (zero) dividida pela quantidade de categorias subsequentes.

3.1.1 Seleção de variáveis independentes

Definidas as categorias da variável resposta, foram avaliadas todas as variáveis existentes no banco de dados para identificação daquelas com maior relevância e que poderiam

apresentar relação com o rendimento dos candidatos.

Segundo Hair, Black, Babin, Anderson e Tatham (2009), a seleção de variáveis independentes podem ocorrer de duas maneiras, a primeira envolvendo a identificação de variáveis por meio de pesquisa prévia ou por meio de um modelo teórico base inerente da questão de pesquisa e a segunda que leva em consideração a intuição do pesquisador, que utilizará do seu conhecimento para identificar variáveis que ainda não foram analisadas conjuntamente, mas que poderiam se relacionar.

Assim, utilizando-se de estudos acerca do assunto já abordados neste trabalho e de informações obtidas ao longo da pesquisa, foram identificadas quais variáveis de caráter socioeconômico são determinantes para o resultado dos alunos no Enem. Além disso, a avaliação da participação de outras variáveis no rendimento dos candidatos foi considerada relevante na percepção do pesquisador. Desta maneira, as variáveis consideradas no modelo logístico multinomial deste estudo estão descritas na Tabela 3.2.

Tabela 3.2: Variáveis independentes discriminadas de acordo suas respectivas características.

Características	Variáveis
Pessoais	Cor/Raça
	Estado Civil
	Idade
	Sexo
Educacionais	Ano de Conclusão do Ensino Médio
	Conclusão do Ensino Médio
	Tipo de Escola em que Cursou o Ensino Médio
	Treineiro
Socioeconômicas	Acesso a Internet
	Computador em Casa
	Escolaridade Mãe
	Escolaridade Pai
	Ocupação Mãe
	Ocupação Pai
Renda	

Fonte: Elaborada pela autora.

A descrição dos níveis de resposta identificados no banco de dados para cada uma das variáveis está relacionada na Tabela 3.3:

Tabela 3.3: Descrição das variáveis independentes e seus respectivos níveis

Variável	Descrição
Cor/Raça	0 - Não Declarada
	1 - Branca
	2 - Preta
	3 - Parda
	4 - Amarela
	5 - Indígena
Estado Civil	0 - Solteiro(a)
	1 - Casado(a)/Mora com companheiro(a)
	2 - Divorciado(a)/Desquitado(a)/Separado(a)
	3 - Viúvo(a)
Idade	(Texto Livre)
Sexo	M - Masculino
	F - Feminino
Ano de Conclusão do Ensino Médio	0 - Não Informado
	1 - 2017
	2 - 2016
	3 - 2015
	4 - 2014
	5 - 2013
	6 - 2012
	7 - 2011
	8 - 2010
	9 - 2009
	10 - 2008
	11 - 2007
	12 - Antes de 2007
Conclusão do Ensino Médio	A - Já concluí o Ensino Médio.
	B - Estou cursando e concluirei o Ensino Médio em 2018.
	C - Estou cursando e concluirei o Ensino Médio após 2018.
	D - Não concluí e não estou cursando o Ensino Médio.
Tipo de Escola em que Cursou o Ensino Médio	A - Somente em escola pública
	B - Escola pública e privada SEM bolsa.
	C - Escola pública e privada COM bolsa.
	D - Em escola privada SEM bolsa de estudo integral.
	E - Em escola privada COM bolsa de estudo integral.
	F - Não frequentei a escola
Treineiro	0 - Não
	1 - Sim
Acesso a Internet	A - Não
	B - Sim
Computador em Casa	A - Não
	B - Sim, um.
	C - Sim, dois.
	D - Sim, três.
	E - Sim, quatro ou mais.
Escolaridade da Mãe	A - Nunca Estudou
	B - Não completou a 4ª série/5º ano.
	C - Completou a 4ª série/5º ano (não a 8ª série/9º).
	D - Completou a 8ª série/9º (não o Ensino Médio).

	E - Completou o Ensino Médio (não a Faculdade).
	F - Completou a Faculdade (não a Pós-graduação).
	G - Completou a Pós-graduação.
	H - Não sei.
Escolaridade do Pai	A - Nunca Estudou
	B - Não completou a 4ª série/5º ano.
	C - Completou a 4ª série/5º ano (não a 8ª série/9º).
	D - Completou a 8ª série/9º (não o Ensino Médio).
	E - Completou o Ensino Médio (não a Faculdade).
	F - Completou a Faculdade (não a Pós-graduação).
	G - Completou a Pós-graduação.
	H - Não sei.
Ocupação da Mãe	A - Grupo 1
	B - Grupo 2
	C - Grupo 3
	D - Grupo 4
	E - Grupo 5
	F - Não sei.
Ocupação do Pai	A - Grupo 1
	B - Grupo 2
	C - Grupo 3
	D - Grupo 4
	E - Grupo 5
	F - Não sei.
Renda	A - Nenhuma Renda
	B - Até R\$ 954,00
	C - De R\$ 954,01 até R\$ 1.431,00
	D - De R\$ 1.431,01 até R\$ 1.908,00
	E - De R\$ 1.908,01 até R\$ 2.385,00
	F - De R\$ 2.385,01 até R\$ 2.862,00
	G - De R\$ 2.862,01 até R\$ 3.816,00
	H - De R\$ 3.816,01 até R\$ 4.770,00
	I - De R\$ 4.770,01 até R\$ 5.724,00
	J - De R\$ 5.724,01 até R\$ 6.678,00
	K - De R\$ 6.678,01 até R\$ 7.632,00
	L - De R\$ 7.632,01 até R\$ 8.586,00
	M - De R\$ 8.586,01 até R\$ 9.540,00
	N - De R\$ 9.540,01 até R\$ 11.448,00
	O - De R\$ 11.448,01 até R\$ 14.310,00
P - De R\$ 14.310,01 até R\$ 19.080,00	
Q - Mais de R\$ 19.080,00	

Fonte: Elaborada pela autora.

- **Grupo 1:** Lavradora, agricultoras em empregados, bóia fria, criadorade animais (gado, porcos, galinhas, ovelhas, cavalos etc.), apicultora, pescadora, lenhadora, seringueira, extrativista.
- **Grupo 2:** Diarista, empregada doméstica, cuidadora de idosos, babá, cozinheira (em casas particulares), motorista particular, jardineira, faxineira de empresas e prédios, vigilante, porteira, carteira, office-boy, vendedora, caixa, atendente de loja.

- **Grupo 3:** Padeira, cozinheira industrial ou em restaurantes, sapateira, costureira, joalheira, torneira mecânica, operadora de máquinas, soldadora, operária de fábrica, trabalhadora da mineração, pedreira, pintora, eletricista, encanadora, motorista, caminhoneiro.
- **Grupo 4:** Professora (de Ensino Fundamental ou médio, idioma, música, artes etc.), técnica (de enfermagem, contabilidade, eletrônica etc.), policial, militar de baixa patente (soldado, cabo, sargento), corretora de imóveis, supervisora, gerente, mestre de obras.
- **Grupo 5:** Médica, engenheira, dentista, psicóloga, economista, advogada, juíza, promotora, defensora, delegada, tenente, capitã, coronel, professora universitária, diretora em empresas públicas ou privadas, política, proprietária de empresas com mais de 10 mil empregados. Não sei.

3.2 Estatística descritiva com *Shiny*

Todos os gráficos apresentados neste trabalho, bem como outros que nortearam toda a análise de dados aqui descrita foram elaborados dentro do próprio software R, com a interface do RStudio, por meio dos pacotes ggplot2 e ggplotly. Foram criadas interações com o pacote Shiny, que permite disponibilizar as visualizações numa página web.

Konrath et al. (2018) afirma que o Shiny é uma alternativa que contribui com inovações no ensino da Estatística, pois permite o desenvolvimento de aplicativos que podem ser manipulados pela Internet e as análises podem ser compartilhadas e discutidas entre estudantes e professores, ampliando as possibilidades de interação e de debates sobre os dados envolvidos.

Para este trabalho, o pacote Shiny permitiu a criação de um aplicativo com gráficos dinâmicos e interativos que garantem uma análise descritiva de informações presentes no questionário do Enem 2018. O aplicativo foi encapsulado em um pacote para o software R com o nome *statenem* e está disponível para instalação, juntamente com a base de dados e o código, no endereço <https://github.com/fsbmat-ufv/statenem>.

Para a elaboração dos gráficos dinâmicos utilizando o pacote Shiny foram criados dois scripts no diretório do R, um referente a interface do usuário (ui.R), onde é construído todo o design e layout do aplicativo e outro referente ao servidor (server.R), onde efetivamente são elaborados os gráficos.

As informações constantes no aplicativo foram divididas em três sessões: Dados do Participante, Desempenho e Questionário Socioeconômico. Na primeira sessão é apresentada a tabela completa com a base de dados e gráficos relacionados a informações pessoais, como quantidade de candidatos por gênero, raça/cor, média de idade, entre outros. Na segunda sessão são visualizadas informações acerca da nota média dos candidatos em cada uma das áreas de conhecimento e na redação e, por fim, a terceira sessão é reservada para análise da nota média final dos candidatos a partir das respostas ao questionário socioeconômico do exame. As informações mais relevantes para este trabalho são apresentadas na sessão de análise descritiva dos dados.

3.3 Modelo de regressão multinomial

Uma vez definidas as variáveis dependente e independentes do modelo, foi identificado que o modelo de regressão logística adequado para avaliação dos dados seria o de Regressão Multinomial, já que a variável dependente possui mais de duas categorias.

No decorrer do trabalho, observou-se que algumas variáveis independentes escolhidas para o modelo possuíam uma quantidade elevada de níveis de resposta e, visando possibilitar a obtenção de um modelo mais ajustado e maior qualidade no processamento deste, foi realizado o agrupamento nos níveis de resposta, de forma a deixá-los mais sucintos e coerentes, conforme detalhado na Tabela 3.4.

Tabela 3.4: Agrupamento de níveis de resposta

Variável	Níveis de Resposta	Níveis Agrupados
Estado Civil	0 - Solteiro(a)	0 - Solteiro
	1 - Casado(a)/Mora com companheiro(a)	1 - Não solteiro
	2 - Divorciado, etc	
	3 - Viúvo(a)	
Cor/raça	0 - Não declarado	Retirado do banco de dados ¹
	1 - Branca	1 - Branca
	4 - Amarela	
	2 - Preta	2 - Preta
	3 - Parda	
	5 - Indígena	
Escolaridade Pai	A - Nunca estudou.	A - Nunca estudou/Não sei
	H - Não sei.	
	B - Não completou a 4 ^a série do Fundamental.	B - Completou no máximo o Ensino Médio
	C - Ensino Fundamental Incompleto.	
	D - Ensino Médio Incompleto	
	E - Graduação Incompleta.	C - Completou a faculdade e/ou a pós graduação
	F - Pós-graduação Incompleta.	
G - Completou a Pós-graduação.		
Escolaridade Mãe	A - Nunca estudou.	A - Nunca estudou/Não sei
	H - Não sei.	
	B - Não completou a 4 ^a série do Fundamental.	B - Completou no máximo o Ensino Médio
	C - Ensino Fundamental Incompleto.	
	D - Ensino Médio Incompleto	
	E - Graduação Incompleta.	C - Completou a faculdade e/ou a pós graduação
	F - Pós-graduação Incompleta.	
G - Completou a Pós-graduação.		

¹Retirado apenas para o processamento do modelo de regressão. Esses dados são contemplados na análise descritiva.

Variável	Níveis de Resposta	Níveis Agrupados
Ocupação Pai	A - Grupo 1	A - Ocupação com menor requisito de formação
	F - Não sei.	
	B - Grupo 2	
	C - Grupo 3	B - Ocupação com maior requisito de formação
	D - Grupo 4	
	E - Grupo 5.	
Ocupação Mãe	A - Grupo 1	A - Ocupação com menor requisito de formação
	F - Não sei.	
	B - Grupo 2	
	C - Grupo 3	B - Ocupação com maior requisito de formação
	D - Grupo 4	
	E - Grupo 5.	
Renda	A - Nenhuma renda.	A - Até 1 salário mínimo
	B - Até R\$ 954,00.	B - De 1 a 2,5 salários mínimos
	C - De R\$ 954,01 até R\$ 1.431,00.	
	D - De R\$ 1.431,01 até R\$ 1.908,00.	
	E - De R\$ 1.908,01 até R\$ 2.385,00.	
	F - De R\$ 2.385,01 até R\$ 2.862,00.	C - De 2,5 a 7 salários mínimos
	G - De R\$ 2.862,01 até R\$ 3.816,00.	
	H - De R\$ 3.816,01 até R\$ 4.770,00.	
	I - De R\$ 4.770,01 até R\$ 5.724,00.	
	J - De R\$ 5.724,01 até R\$ 6.678,00.	D - De 7 a 12 salários mínimos
	K - De R\$ 6.678,01 até R\$ 7.632,00.	
	L - De R\$ 7.632,01 até R\$ 8.586,00.	
	M - De R\$ 8.586,01 até R\$ 9.540,00.	
	N - De R\$ 9.540,01 até R\$ 11.448,00.	E - Acima de 12 salários mínimos
O - De R\$ 11.448,01 até R\$ 14.310,00.		
P - De R\$ 14.310,01 até R\$ 19.080,00.		
	Q - Mais de R\$ 19.080,00.	
Computador em Casa	A - Não.	A - Não
	B - Sim, um.	B - Sim
	C - Sim, dois.	
	D - Sim, três.	
	E - Sim, quatro ou mais.	
Conclusão do Ensino Médio	A - Já concluí o Ensino Médio.	A - Já concluí.
	B - Concluirei o Ensino Médio em 2018.	B - Concluirei em 2018.
	C - Concluirei o Ensino Médio após 2018.	C - Não concluirei em 2018
	D - Não concluí e não cursando	
Tipo de Escola	A - Somente em escola pública.	A - Somente em escola pública.
	F - Não frequentei a escola	B - Parte em escola pública e parte em escola privada
	B - Escola pública e privada SEM bolsa	
	C - Escola pública e privada COM bolsa	C - Somente em escola privada
	D - Escola privada SEM bolsa	
E - Escola privada COM bolsa		

Fonte: Elaborada pela autora

Os níveis de resposta das variáveis Idade, Ano de conclusão do Ensino Médio, Sexo, Uso de internet e Treineiro não foram agrupados por se tratarem de variáveis contínua ou dicotômica.

Destaca-se que esse ajuste nos níveis de resposta ocorreu apenas para o processamento do modelo de regressão. A análise descritiva constante neste trabalho levou em consideração

todos os níveis de resposta existentes.

Para a manipulação e elaboração da análise estatística dos dados foi utilizado o software R, sendo o ambiente de desenvolvimento o RStudio, por se tratar de uma ferramenta gratuita.

Para confirmar a significância do modelo foram avaliados os resultados apresentados na estatística de Wald e o p-valor, garantindo assim a obtenção de um modelo ajustado. Além disso, o presente estudo avaliou os coeficientes apresentados após o processamento deste modelo, de forma a identificar o papel de cada variável no rendimento dos candidatos do Enem 2018.

Resultados e Discussão

4.1 Análise descritiva dos dados

4.1.1 Análise geral dos candidatos

O Exame Nacional do Ensino Médio 2018 contou com mais de 5 milhões de candidatos inscritos, sendo em sua maioria mulheres (59%). Entretanto, cerca de 30% desses inscritos não compareceram para a realização das provas e foram desconsiderados da base de dados deste estudo para avaliação do rendimento dos alunos. Assim, esta avaliação contou com 3.893.743 (três milhões oitocentos e noventa e três mil, setecentos e quarenta e três) candidatos que realizaram todas as etapas do Enem.

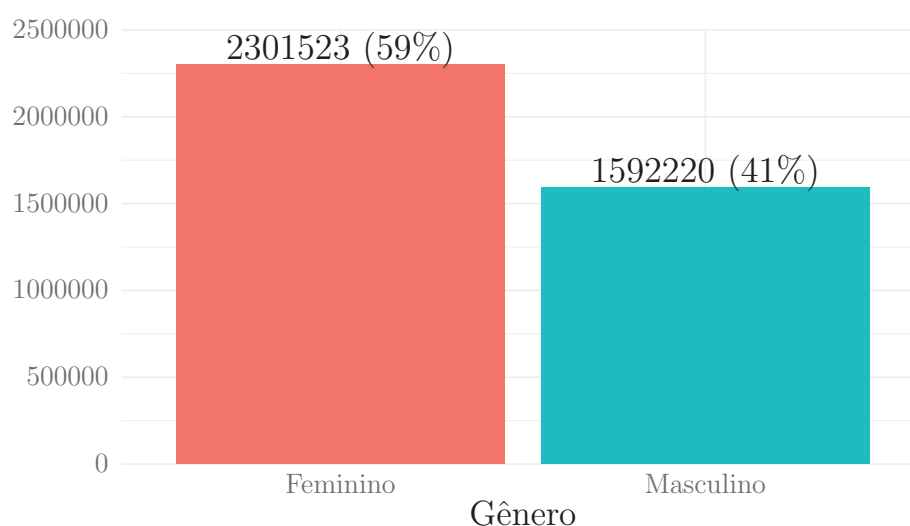


Figura 4.1: Quantidade de Candidatos por Gênero

Quando se leva em consideração apenas os candidatos que compareceram para a realização de todas as provas, o percentual de homens (41%) e mulheres (59%) se mantém

em relação ao número total de inscritos, indicando a desistência de candidatos em uma proporção aproximada para ambos os sexos.

Realizando uma comparação de ambos os sexos no tocante à idade média, é possível observar que a média etária dos candidatos do sexo masculino (21,5 anos) é ligeiramente superior à das candidatas do sexo feminino (21,34 anos).

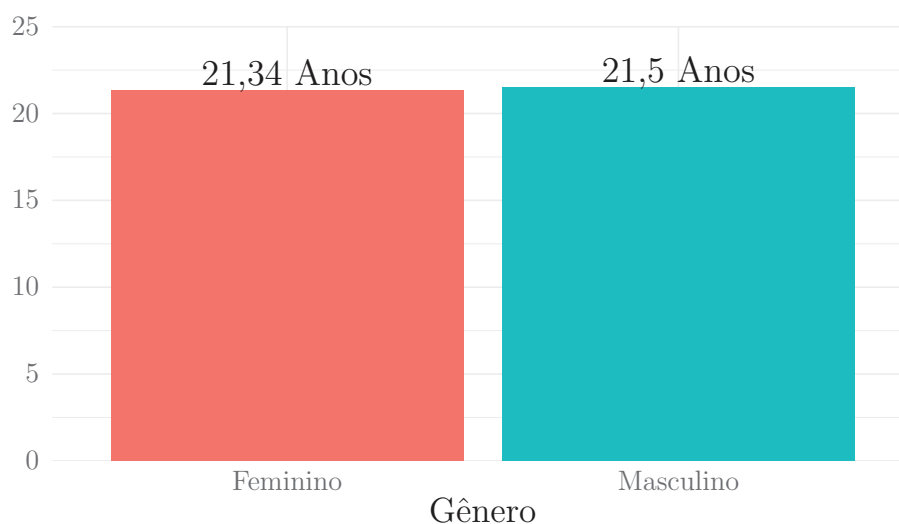


Figura 4.2: Média de idade por gênero

No que diz respeito ao estado civil dos candidatos, a maior parte (92,2%) se declarou solteiro. Os candidatos casados representam o segundo maior grupo com 6,7% do total de presentes. Os demais estados civis (que agrega divorciado/desquitado/separado e viúvo) representam apenas 1,1% do total.

Tabela 4.1: Candidatos por estado civil

Descrição	Quantidade	Percentual
Solteiro(a)	3.472.390	92,2%
Casado(a)/Mora com companheiro(a)	253.315	6,7%
Divorciado(a)/Desquitado(a)/Separado(a)	38.114	1,0%
Viúvo(a)	4.252	0,1%
Total	3.768.071	100,0%

Fonte: Elaborada pela autora

Considerando que a maior parte do público do exame é composta por estudantes que estão concluindo o Ensino Médio ou o fizeram em um passado recente, é possível afirmar que o fato de a maioria dos candidatos serem solteiros não chega a ser totalmente surpreendente. A idade média dos candidatos por sexo, apresentada no gráfico da Figura 4.2, também corrobora esse resultado (apenas 21,34 anos para as candidatas do sexo feminino e 21,5 anos para os candidatos do sexo masculino). Para se obter maiores conclusões ainda é interessante avaliar a idade média dos candidatos por estado civil.

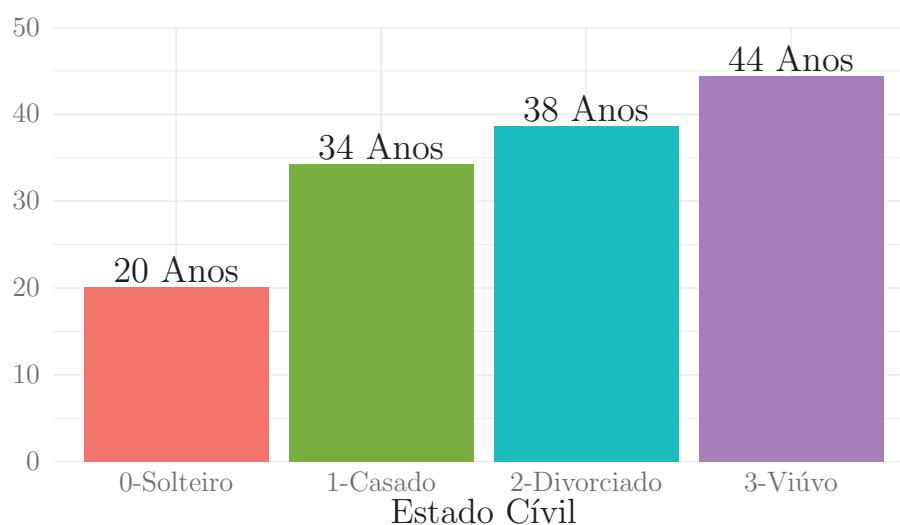


Figura 4.3: Média de idade por estado civil

Quando avaliada a faixa etária por estado civil, observa-se que os candidatos solteiros (média de 20 anos) são consideravelmente mais novos que os demais. Os candidatos casados possuem a idade média de 34 anos, enquanto os candidatos divorciados possuem 38 anos em média. Os candidatos viúvos apresentam, em média, mais que o dobro da idade dos candidatos solteiros.

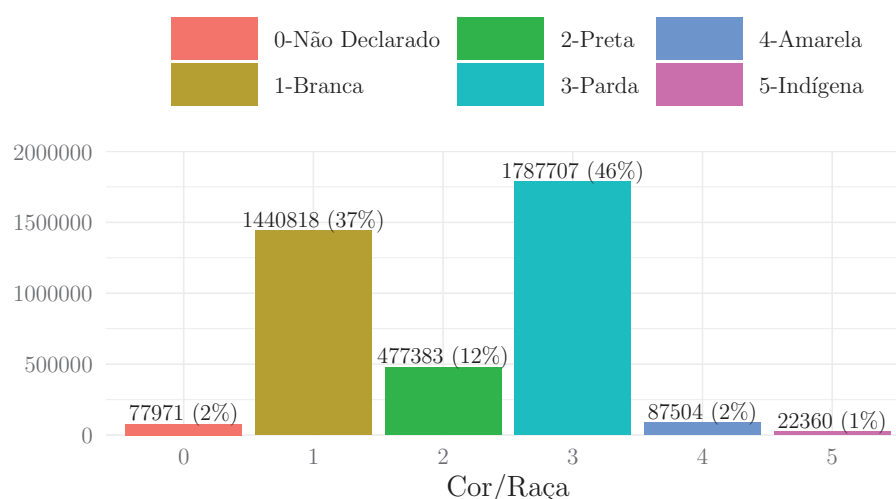


Figura 4.4: Quantidade de candidatos por cor/raça

Avaliando-se a cor ou raça, os grupos predominantes são os que se declararam da cor parda (46%), branca (37%) e preta (12%). Outras raças como amarela, indígena e não declarado, representam juntas 5% e dos candidatos. É interessante mencionar que comumente utiliza-se a denominação negro para abranger os indivíduos de cor/raça parda

ou preta. Assim, adotando esse critério, é possível afirmar que 58% dos candidatos do Enem 2018 se declararam negros.

Ainda no tocante a cor ou raça, é importante avaliar a representação de cada grupo em relação ao percentual da população brasileira. O Instituto Brasileiro de Geografia e Estatística (IBGE) pesquisa a cor ou raça da população com base na autodeclaração. A Tabela 4.2 agrega a informação da porcentagem da população por cor ou raça obtida na Pesquisa Nacional por Amostra de Domicílios (PNAD) 2015 aos dados do Enem e calcula a razão da porcentagem de candidatos do Enem sobre a porcentagem da população para cada cor ou raça:

Tabela 4.2: Relação de candidatos por população

Descrição	Candidatos		População	Candidatos/População
	Quantidade	%	%	
Branca	1.440.818	37%	45%	0,82
Preta	477.383	12%	9%	1,33
Parda	1.787.707	46%	45%	1,02

Fonte: Elaborada pela autora

A partir da relação Candidatos/População apresentada na Tabela 4.2 é possível observar que, percentualmente, a representatividade dos candidatos que se declaram da cor/raça branca no Enem é menor do que a representatividade desta raça na população brasileira. Por outro lado, a representatividade dos candidatos que declaram da cor/raça preta é maior no Enem do que na população brasileira como um todo. A razão entre a representatividade no Enem e a representatividade na população é de 1,33. Uma possível explicação para esse resultado é que existe uma maior parcela da população declarada preta que ainda não alcançou o Ensino Superior.

Ao se avaliar a escolaridade dos pais dos candidatos, chama a atenção o fato de as mães possuírem uma escolaridade maior em relação aos pais. Os grupos Nunca estudou, Não completou a 4^a série e Completou a 4^a série mas não completou o Ensino Fundamental foram mais indicados para os pais do que para as mães. Já os grupos que representam uma maior escolaridade como Completou pós-graduação, Completou a faculdade (mas não a pós graduação) e Completou o Ensino Médio (mas não a faculdade) foram mais indicados para as mães, quando comparado aos pais. Cabe ressaltar que quanto mais escolaridade o grupo indica, maior a proporção em que o mesmo foi utilizado para as mães, em comparação aos pais.

Uma outra afirmação possível sobre a escolaridade dos pais indicada pelos candidatos do Enem é que a mesma é, no geral, baixa: Cerca de 500 mil candidatos declararam que o pai completou faculdade e/ou pós-graduação e cerca de 740 mil candidatos declararam que a mãe alcançou esses níveis de instrução. No extremo oposto, quando considerados os pais que nunca estudaram ou não completaram a 4^a série, a declaração foi feita por mais de 1 milhão de candidatos para o pai e mais de 740 mil para mãe. Mais de 2 milhões de candidatos declararam que o pai não concluiu o Ensino Médio e 1,75 milhões de candidatos declararam o mesmo para mãe. Assim, considerando que os candidatos do

Enem já concluíram ou estão concluindo o Ensino Médio, é possível afirmar que a maior parte deles já superou a escolaridade dos respectivos pais.

Tabela 4.3: Escolaridade dos pais

Descrição	Pai		Mãe	
	Quantidade	Percentual	Quantidade	Percentual
A	200.902	5,2%	136.414	3,5%
B	813.670	20,9%	606.702	15,6%
C	549.969	14,1%	515.982	13,3%
D	460.454	11,8%	509.545	13,1%
E	1.043.472	26,8%	1.275.076	32,7%
F	311.918	8,0%	417.225	10,7%
G	187.705	4,8%	322.696	8,3%
H	325.653	8,4%	110.103	2,8%
TOTAL	3.893.743	100,0%	3.893.743	100,0%

Fonte: Elaborada pela autora

A - Nunca estudou. B - Não completou a 4^a série/5^o ano do Ensino Fundamental. C - Completou a 4^a série/5^o ano, mas não completou a 8^a série/9^o ano do Ensino Fundamental. D - Completou a 8^a série/9^o ano do Ensino Fundamental, mas não completou o Ensino Médio. E - Completou o Ensino Médio, mas não completou a Faculdade. F - Completou a Faculdade, mas não completou a Pós-graduação. G - Completou a Pós-graduação. H - Não sei.

A avaliação da ocupação dos pais dos candidatos trás informações relevantes quanto ao mercado de trabalho para os homens e as mulheres. Enquanto os pais apresentam representação menor em escolaridades de maior grau, quando comparados às mães, observa-se que em termos percentuais eles são maioria em atividades ocupacionais com níveis salariais mais elevados, como os descritos no Grupo 5 da Tabela 4.4.

Dentre as possíveis causas para haver mais pais no grupo 5 quando comparado às mães, em detrimento do maior grau de escolaridade dessas, podem estar barreiras ainda existentes para o mercado de trabalho das mulheres e barreiras de acesso aos cursos de graduação com maior notoriedade social que estão refletidos no grupo 5 como engenharia, medicina e direito.

Tabela 4.4: Ocupação dos pais

Descrição	Pai		Mãe	
	Quantidade	Percentual	Quantidade	Percentual
Grupo 1	785.584	20,2%	638.782	16,4%
Grupo 2	807.295	20,7%	1.637.076	42,0%
Grupo 3	921.025	23,7%	271.305	7,0%
Grupo 4	721.769	18,5%	841.174	21,6%
Grupo 5	229.267	5,9%	156.047	4,0%
Não sei.	428.803	11,0%	349.359	9,0%
TOTAL	3.893.743	100,0%	3.893.743	100,0%

Fonte: Elaborada pela autora

Grupo 1: Lavradora, agricultoras em empregados, bóia fria, criadorade animais (gado, porcos, galinhas, ovelhas, cavalos etc.), apicultora, pescadora, lenhadora, seringueira, extrativista. **Grupo 2:** Diarista, empregada doméstica, cuidadora de idosos, babá, cozinheira (em casas particulares), motorista particular, jardineira, faxineira de empresas e prédios, vigilante, porteira, carteira, office-boy, vendedora, caixa, atendente de loja. **Grupo 3:** Padeira, cozinheira industrial ou em restaurantes, sapateira, costureira, joalheira, torneira mecânica, operadora de máquinas, soldadora, operária de fábrica, trabalhadora da mineração, pedreira, pintora, eletricista, encanadora, motorista, caminhoneiro. **Grupo 4:** Professora (de Ensino Fundamental ou médio, idioma, música, artes etc.), técnica (de enfermagem, contabilidade, eletrônica etc.), policial, militar de baixa patente (soldado, cabo, sargento), corretora de imóveis, supervisora, gerente, mestre de obras. **Grupo 5:** Médica, engenheira, dentista, psicóloga, economista, advogada, juíza, promotora, defensora, delegada, tenente, capitã, coronel, professora universitária, diretora em empresas públicas ou privadas, política, proprietária de empresas com mais de 10 mil empregados. Não sei.

Quanto à renda familiar dos candidatos, a Tabela 4.5 demonstra que 52% dos mesmos possuíam uma renda média mensal inferior a 1,5 salários mínimos (salário mínimo em 2018 de R\$ 954,00). Por outro lado, cerca 4,4% dos candidatos possuíam renda superior a 10 salários mínimos.

Uma vez que a informação referente a renda familiar coletada no questionário socioeconômico do Enem é categórica, agrupada em diversas faixas de rendimento, não é possível obter os valores exatos da renda familiar *per capita*. Para estimar esses valores, o presente estudo dividiu o valor médio da respectiva faixa de remuneração de cada grupo no questionário socioeconômico pelo número de moradores na residência. Por fim, uma nova categorização foi criada para possibilitar uma distribuição dos candidatos de forma mais uniforme entre as mesmas. Os resultados desse cálculo e da categorização são apresentados na Tabela 4.6.

Tabela 4.5: Renda familiar dos candidatos

Descrição	Quantidade	% do grupo	% acumulado
Nenhuma renda.	153.283	3,9%	3,9%
Até R\$ 954,00.	891.153	22,9%	26,8%
De R\$ 954,01 até R\$ 1.431,00.	979.903	25,2%	52,0%
De R\$ 1.431,01 até R\$ 1.908,00.	364.313	9,4%	61,3%
De R\$ 1.908,01 até R\$ 2.385,00.	360.081	9,2%	70,6%
De R\$ 2.385,01 até R\$ 2.862,00.	175.316	4,5%	75,1%
De R\$ 2.862,01 até R\$ 3.816,00.	276.106	7,1%	82,2%
De R\$ 3.816,01 até R\$ 4.770,00.	167.665	4,3%	86,5%
De R\$ 4.770,01 até R\$ 5.724,00.	138.437	3,6%	90,0%
De R\$ 5.724,01 até R\$ 6.678,00.	82.322	2,1%	92,2%
De R\$ 6.678,01 até R\$ 7.632,00.	52.620	1,4%	93,5%
De R\$ 7.632,01 até R\$ 8.586,00.	42.056	1,1%	94,6%
De R\$ 8.586,01 até R\$ 9.540,00.	37.279	1,0%	95,6%
De R\$ 9.540,01 até R\$ 11.448,00.	53.790	1,4%	96,9%
De R\$ 11.448,01 até R\$ 14.310,00.	39.012	1,0%	97,9%
De R\$ 14.310,01 até R\$ 19.080,00.	35.383	0,9%	98,8%
Mais de R\$ 19.080,00.	45.024	1,2%	100,0%

Fonte: Elaborada pela autora

Mais de 50% dos candidatos ao Enem declararam renda *per capita* de até R\$ 400,00 enquanto apenas 8% dos candidatos declararam renda *per capita* superior a R\$ 2.000,00. Acredita-se que a renda familiar *per capita* seja positivamente associada ao desempenho dos candidatos do Enem pois oportuniza mais acesso aos recursos educacionais. Por outro lado, os candidatos com menor rendimento *per capita* familiar podem estar sujeitos a diversos tipos de barreiras e limitações como o acesso aos recursos tecnológicos, materiais didáticos voltados à preparação para o exame.

Tabela 4.6: Renda média *per capita* dos candidatos

Renda Média <i>per capita</i>	Quantidade	% do grupo	% acumulado
Nenhuma renda.	153.283	3,9%	3,9%
Até R\$ 200,00.	862.937	22,2%	26,1%
De R\$ 200,01 até R\$ 400,00.	966.387	24,8%	50,9%
De R\$ 400,01 até R\$ 600,00.	627.485	16,1%	67,0%
De R\$ 600,01 até R\$ 800,00.	218.051	5,6%	72,6%
De R\$ 800,01 até R\$ 1.000,00.	239.937	6,2%	78,8%
De R\$ 1.000,01 até R\$ 2.000,00.	517.502	13,3%	92,1%
De R\$ 2.000,01 até R\$ 3.000,00.	150.438	3,9%	95,9%
De R\$ 3.000,01 até R\$ 4.000,00.	68.860	1,8%	97,7%
De R\$ 4.000,01 até R\$ 5.000,00.	54.183	1,4%	99,1%
De R\$ 5.000,01 até R\$ 10.000,00.	32.605	0,8%	99,9%
Mais de R\$ 10.000,00.	2.075	0,1%	100,0%

Fonte: Elaborada pela autora

Uma outra variável analisada no presente estudo foi quanto ao tipo de escola que o candidato cursou o Ensino Médio. A maior parte dos candidatos, aproximadamente 73% do total, cursaram integralmente o Ensino Médio na rede pública de ensino. Levando em consideração também os candidatos que cursaram parcialmente o Ensino Médio em escola pública pode-se afirmar que 3,27 milhões de candidatos ao Enem 2018 passaram pela rede pública no decorrer do Ensino Médio em detrimento de 619,5 mil candidatos que estudaram exclusivamente em escola privada.

Tabela 4.7: Tipo de escola do Ensino Médio

Descrição	Quantidade	Percentual
A	2.841.735	72,98%
B	326.898	8,40%
C	103.853	2,67%
D	525.028	13,48%
E	94.567	2,43%
F	1.662	0,04%
TOTAL	3.893.743	

Fonte: Elaborada pela autora

A - Somente em escola pública. B - Parte em escola pública e parte em escola privada SEM bolsa de estudo integral. C - Parte em escola pública e parte em escola privada COM bolsa de estudo integral. D - Somente em escola privada SEM bolsa de estudo integral. E - Somente em escola privada COM bolsa de estudo integral. F - Não frequentou a escola.

No que diz respeito à origem dos candidatos, cerca de 70% vieram das regiões Sudeste (36%) e Nordeste (33%). Os Estados com maior número de participantes no exame são

respectivamente São Paulo (16%), Minas Gerais (11%), Bahia (7%), Rio de Janeiro (7%) e Ceará (6%). Juntos, estes Estados respondem por aproximadamente 48% do total de participantes.

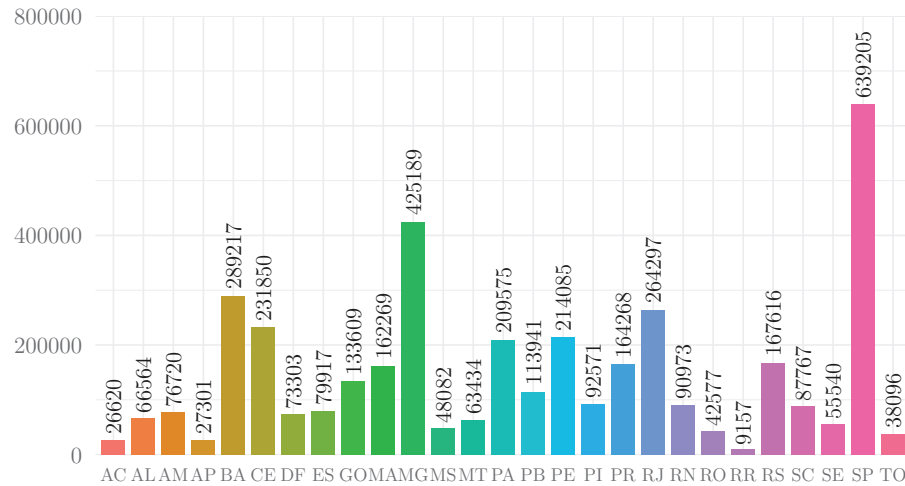


Figura 4.5: Quantidade de candidatos por unidade da federação

Quando considerado o número de participantes em relação à população dos respectivos Estados, observa-se que as regiões Norte e Nordeste registram maiores índices *per capita* de participantes. Dos sete Estados com maior proporção de candidatos por habitante, quatro são do Nordeste e três do Norte: os Estados do Amapá (32,28), Acre (30,18), Paraíba (28,36), Piauí (28,28), Rio Grande do Norte (25,94), Ceará (25,39) e Pará (24,36).

Ainda cabe destacar que entre as treze unidades da federação com maior índice de candidatos por 1.000 habitantes apenas uma (o Distrito Federal) não está situada nas regiões Norte ou Nordeste.

Os Estados de São Paulo, Minas Gerais e Rio de Janeiro (que estão entre os com maior quantidade absoluta de participantes) registraram respectivamente 13,92, 20,09 e 15,31 participantes por 1.000 habitantes. A região Sul apresenta a menor participação proporcional: enquanto o estado de Santa Catarina apresentou 12,25 candidatos por 1.000 habitantes (menor índice), os estados do Paraná e Rio Grande do Sul registraram respectivamente 14,37 e 14,73 candidatos por habitante.

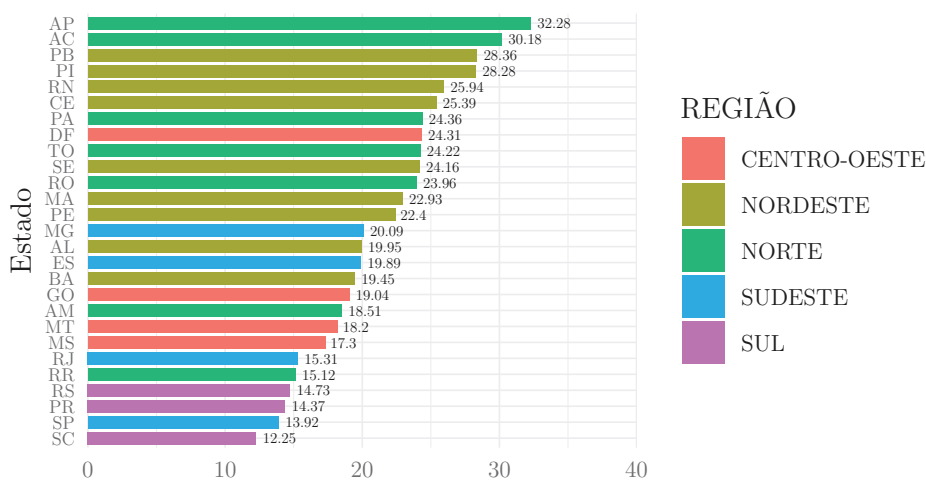


Figura 4.6: Número de candidatos a cada 1000 habitantes do Estado

As regiões Norte e Nordeste apresentam maior número de candidatos inscritos para grupo de 1.000 habitantes.

Tabela 4.8: Candidatos por região do país

Região	Candidatos	
	Total	Por 1.000 habitantes
Centro-Oeste	318.428	19,54
Nordeste	1.317.010	23,08
Norte	430.046	23,33
Sudeste	1.408.608	15,94
Sul	419.651	14,00

Fonte: Elaborada pela autora

4.1.2 Análise das notas dos candidatos

Esta sessão se dedica a avaliar as notas finais dos alunos no conjunto de provas em relação às variáveis já consideradas na primeira sessão. Para tanto, em algumas variáveis como sexo, estado civil, raça, renda e tipo de escola, é realizada uma análise bidimensional entre a respectiva variável e as categorias de nota. Para as variáveis representativas do grau de escolaridade de pai e mãe dos candidatos, assim como para as variáveis representativas da ocupação, optou-se por fazer a avaliação da nota média pois a avaliação conjuntamente dessas variáveis tanto do pai quanto da mãe dos candidatos pode permitir uma avaliação mais completa.

A análise quanto às respectivas categorias de nota é importante, pois é esta a variável

explicativa utilizada no modelo de regressão logístico multinomial que constitui a última parte do presente capítulo e o foco do estudo.

A avaliação geral das notas dos candidatos mostra que 94,5% obtiveram nota até 677,70 (categoria 3), sendo a categoria 2 (notas entre 497,22 e 587,46) a que possui maior quantidade de candidatos (37,86%). Os candidatos que obtiveram notas acima de 677,70 e ficaram nas categorias 4 e 5 representam, respectivamente, 5,2% e 0,3% do total de candidatos.

Tabela 4.9: Quantidade e percentual de alunos por categoria de nota

Categoria	Nota	Quantidade	Percentual
0	0,00 † 406,98	159.007	4,1%
1	406,98 † 497,22	1.363.963	35,0%
2	497,22 † 587,46	1.474.259	37,9%
3	587,46 † 677,70	683.219	17,5%
4	677,70 † 767,94	202.301	5,2%
5	767,94 † 858,18	10.994	0,3%
Total	-	3.893.743	100,0%

Fonte: Elaborada pela autora

Quando é avaliado o comportamento das notas em relação ao gênero dos candidatos, a Tabela 4.10 demonstra que, apesar de a maior parte dos candidatos ao Enem 2018 serem do sexo feminino, existe uma maior quantidade de candidatos do sexo masculino na categoria de nota mais alta. A quantidade absoluta de mulheres é superior em todas as demais categorias.

Ao avaliar isoladamente cada gênero, observa-se que eles apresentam comportamento semelhante ao da avaliação geral dos candidatos: 95% das mulheres estão distribuídas nas categorias de 0 a 3, contra 93,8% dos homens.

Segundo Morettin et al. (2017), como a relação percentual entre os gêneros se mantém similar a relação percentual total, pode ser que a dependência entre as duas variáveis (sexo e categoria de notas) não seja tão relevante.

Tabela 4.10: Quantidade de alunos por categoria de nota por sexo

Categorias	Feminino		Masculino	
	Quantidade	Percentual	Quantidade	Percentual
0	94.574	4,1%	64.433	4,0%
1	867.151	37,7%	496.812	31,2%
2	863.585	37,5%	610.674	38,4%
3	361.361	15,7%	321.858	20,2%
4	109.892	4,8%	92.409	5,8%
5	4.960	0,2%	6.034	0,4%
Total	2.301.523	100,0%	1.592.220	100,0%

Fonte: Elaborada pela autora

No tocante ao estado civil dos candidatos, observa-se na Tabela 4.11 que candidatos solteiros se diferenciam dos demais por serem mais representativos nas categorias de notas mais altas. Os solteiros representam cerca de 96,8% dos candidatos com nota nas categorias 4 e 5.

Tabela 4.11: Quantidade de candidatos do estado civil por categoria de nota

Estado Civil	Categorias						Total
	0	1	2	3	4	5	
Solteiro(a)	136.767	1.187.481	1.309.394	633.593	194.426	10.729	3.472.390
Casado(a)	13.936	110.891	99.126	26.090	3.178	94	253.315
Divorciado(a)	2.277	16.726	15.079	3.658	369	5	38.114
Viúvo(a)	542	2.261	1.185	239	24	1	4.252
Não declarado	5.485	46.604	49.475	19.639	4.304	165	125.672
Total	159.007	1.363.963	1.474.259	683.219	202.301	10.994	3.893.743

Fonte: Elaborada pela autora

Os números indicam uma maior dificuldade de acesso ao Ensino Superior através do Enem para os candidatos casados, divorciados e viúvos, que pode estar relacionada com a idade, bem como o tempo disponível aos estudos que os integrantes dessas classes possuem.

Quando o assunto é cor/raça, a distribuição dos candidatos considerando as categorias de nota na Tabela 4.12 demonstra que ainda existe uma disparidade racial ampla no Brasil. Candidatos das raças branca, amarela e aqueles que não declararam a raça representam, juntos, 72,2% dos candidatos que estão nas categorias de nota 4 e 5, sendo, portanto, maioria entre os candidatos que obtiveram maiores notas. Por outro lado, candidatos que declararam pertencer às raças preta, parda e indígena representam apenas 27,8% nessas categorias de notas mais elevadas, enquanto que, quando observadas as três categorias de notas mais baixas, candidatos das raças preta, parda ou indígena representam 66,7%.

Tabela 4.12: Quantidade de candidatos da cor/raça por categoria de nota

Cor/Raça	Categorias						Total
	0	1	2	3	4	5	
Não Declarado	3.410	23.990	26.683	17.176	6.306	406	77.971
Branca	35.598	375.720	552.037	344.852	124.937	7.674	1.440.818
Preta	24.131	197.421	186.638	59.188	9.744	261	477.383
Parda	90.390	724.083	669.301	245.761	55.911	2.261	1.787.707
Amarela	3.205	31.506	32.738	14.526	5.145	384	87.504
Indígena	2.273	11.243	6.862	1.716	258	8	22.360
Total	159.007	1.363.963	1.474.259	683.219	202.301	10.994	3.893.743

Fonte: Elaborada pela autora

Esse resultado sinaliza a importância de políticas afirmativas como as cotas raciais para reduzir a desigualdade histórica e a disparidade de oportunidades para as diferentes raças.

A Tabela 4.13 apresenta o percentual de alunos em cada categoria de acordo com o nível de renda familiar informado e demonstra o impacto da desigualdade de renda no rendimento dos candidatos.

Tabela 4.13: Percentual de candidatos nas categorias por renda

Renda	Categorias						Total
	0	1	2	3	4	5	
Nenhuma renda.	9,3%	53,9%	30,0%	6,1%	0,8%	0,0%	100,0%
Até R\$ 954,00.	7,5%	51,2%	33,6%	7,0%	0,7%	0,0%	100,0%
De R\$ 954,01 até R\$ 1.431,00.	4,5%	41,8%	40,7%	11,5%	1,5%	0,0%	100,0%
De R\$ 1.431,01 até R\$ 1.908,00.	3,1%	33,6%	43,7%	16,7%	2,8%	0,1%	100,0%
De R\$ 1.908,01 até R\$ 2.385,00.	2,5%	30,5%	44,5%	18,9%	3,5%	0,1%	100,0%
De R\$ 2.385,01 até R\$ 2.862,00.	1,9%	24,6%	43,2%	24,2%	5,8%	0,2%	100,0%
De R\$ 2.862,01 até R\$ 3.816,00.	1,6%	21,8%	42,4%	26,9%	7,0%	0,3%	100,0%
De R\$ 3.816,01 até R\$ 4.770,00.	1,2%	17,1%	39,1%	31,6%	10,6%	0,5%	100,0%
De R\$ 4.770,01 até R\$ 5.724,00.	1,0%	14,4%	36,1%	34,4%	13,4%	0,7%	100,0%
De R\$ 5.724,01 até R\$ 6.678,00.	0,8%	11,7%	32,9%	37,0%	16,5%	1,0%	100,0%
De R\$ 6.678,01 até R\$ 7.632,00.	0,7%	10,4%	30,8%	38,2%	18,7%	1,1%	100,0%
De R\$ 7.632,01 até R\$ 8.586,00.	0,6%	8,8%	28,3%	39,5%	21,4%	1,4%	100,0%
De R\$ 8.586,01 até R\$ 9.540,00.	0,5%	8,3%	27,2%	39,8%	22,6%	1,6%	100,0%
De R\$ 9.540,01 até R\$ 11.448,00.	0,5%	6,7%	25,1%	40,9%	24,9%	1,9%	100,0%
De R\$ 11.448,01 até R\$ 14.310,00.	0,4%	5,7%	22,2%	40,8%	28,5%	2,3%	100,0%
De R\$ 14.310,01 até R\$ 19.080,00.	0,3%	4,8%	20,1%	40,9%	31,1%	2,8%	100,0%
Mais de R\$ 19.080,00.	0,3%	3,9%	17,2%	39,6%	35,1%	3,9%	100,0%

Fonte: Elaborada pela autora

Considerando isoladamente cada nível de renda, observa-se que na medida em que a renda vai aumentando, o percentual de candidatos em categorias de notas maiores também se eleva, ao passo que o percentual dos candidatos com renda inferior é muito superior nas categorias de nota inferiores.

O resultado demonstra que candidatos que declararam uma renda familiar mais elevada apresentam melhores resultados e maior desempenho no Enem, possuindo mais chance de ingressarem em cursos com notas de corte mais elevadas. No extremo oposto, os candidatos das categorias que representam menor renda familiar possuem mais chances de obterem notas mais baixas e não conseguirem ingressar em nenhum curso¹.

No que se refere à escola em que o candidato estudou durante o Ensino Médio, observa-se que candidatos que estudaram em escola pública possuem uma maior representatividade entre aqueles que obtiveram notas nas duas categorias com notas mais baixas. Na Tabela 4.14 observa-se que 92% dos candidatos com nota na categoria zero estudaram o Ensino Médio integralmente em escola da rede pública. Os candidatos que estudaram parcialmente em escola pública e parcialmente em escola privada (com ou sem bolsa) são menos representativos tanto entre as categorias que representam as notas mais baixas quanto na categoria que representa as notas mais altas, indicando resultado intermediário. Por fim, candidatos que frequentaram unicamente a escola privada (com ou sem bolsa) possuem uma possibilidade bastante superior de alcançarem as categorias de notas mais

¹Considerando a nota mais baixa obtida por ingressantes na UFV pelo Enem 2018.

altas e uma menor possibilidade de estarem nos grupos de notas mais baixas: 84,2% dos candidatos com nota na categoria 5 estudaram o Ensino Médio integralmente em escola da rede particular de ensino.

Tabela 4.14: Representatividade relativa da escola do Ensino Médio por categoria de nota

Tipo de Escola	Categorias					
	0	1	2	3	4	5
A	92,0%	87,1%	75,2%	50,8%	24,9%	9,4%
B	4,5%	6,4%	9,3%	11,2%	8,7%	5,0%
C	1,1%	2,0%	3,1%	3,5%	2,7%	1,4%
D	1,8%	3,5%	10,1%	29,6%	56,4%	75,9%
E	0,4%	0,9%	2,2%	4,9%	7,2%	8,4%
F	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Total	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%

Fonte: Elaborada pela autora

A - Somente em escola pública. B - Parte em escola pública e parte em escola privada SEM bolsa de estudo integral. C - Parte em escola pública e parte em escola privada COM bolsa de estudo integral. D - Somente em escola privada SEM bolsa de estudo integral. E - Somente em escola privada COM bolsa de estudo integral. F - Não frequentou a escola.

Esses resultados demonstram uma nítida diferença entre o desempenho de alunos que cursaram o Ensino Médio em escolas públicas e o desempenho daqueles que estudaram em escola privada. Isso indica a baixa qualidade do ensino público e corrobora para o argumento da necessidade da política de cotas como forma de reduzir a disparidade de oportunidades entre os alunos egressos de escolas públicas e aqueles oriundos da escola privada.

A avaliação do rendimento dos candidatos em relação a escolaridade dos pais, diferente das avaliações anteriores, levou em consideração a nota média dos candidatos. Essa diferenciação se fez necessária para permitir uma avaliação conjunta de pai e mãe, demonstrando o quanto a escolaridade dos pais de uma maneira geral é capaz de interferir no rendimento dos candidatos.

A Tabela 4.15 mostra que os candidatos que possuem pais e mães com maior grau de escolaridade possuem uma nota média mais elevada em relação àqueles com pais com grau de escolaridade inferior. Desta maneira, os resultados demonstram que tanto o grau de escolaridade do pai quanto o da mãe está diretamente relacionado com o desempenho médio dos candidatos do Enem.

Tabela 4.15: Nota média dos candidatos em relação à escolaridade dos pais

Escolaridade do Pai	Escolaridade da Mãe						
	A	B	C	D	E	F	G
A	466,8	482,8	487,5	490,9	495,3	505,9	516,4
B	483,3	490,5	499,7	504,6	511,1	524,4	529,4
C	487,2	502,5	504,5	512,6	523,9	542,7	549,6
D	491,1	509,1	516,4	512,1	530,7	555,1	560,6
E	492,2	513,6	525,4	531,0	540,3	572,1	579,2
F	514,6	528,9	547,7	558,2	578,2	606,7	614,5
G	517,0	523,2	544,3	560,9	585,1	622,8	623,1

Fonte: Elaborada pela autora

A - Nunca estudou. B - Não completou a 4^a série/5^o ano do Ensino Fundamental. C - Completou a 4^a série/5^o ano, mas não completou a 8^a série/9^o ano do Ensino Fundamental. D - Completou a 8^a série/9^o ano do Ensino Fundamental, mas não completou o Ensino Médio. E - Completou o Ensino Médio, mas não completou a Faculdade. F - Completou a Faculdade, mas não completou a Pós-graduação. G - Completou a Pós-graduação.

Tabela 4.16: Nota média dos candidatos em relação à ocupação dos pais

Ocupação do Pai	Ocupação da Mãe				
	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5
Grupo 1	482,6	495,8	510,0	526,8	557,1
Grupo 2	498,8	510,6	520,0	545,0	570,3
Grupo 3	506,8	518,6	525,3	549,7	569,6
Grupo 4	526,5	553,7	558,4	583,7	616,8
Grupo 5	542,4	574,2	580,1	617,0	631,8

Fonte: Elaborada pela autora

Grupo 1: Lavradora, agricultoras em empregados, bóia fria, criadora de animais (gado, porcos, galinhas, ovelhas, cavalos etc.), apicultora, pescadora, lenhadora, seringueira, extrativista. **Grupo 2:** Diarista, empregada do doméstico, cuidadora de idosos, babá, cozinheira (em casas particulares), motorista particular, jardineira, faxineira de empresas e prédios, vigilante, porteira, carteira, office-boy, vendedora, caixa, atendente de loja. **Grupo 3:** Padeira, cozinheira industrial ou em restaurantes, sapateira, costureira, joalheira, torneira mecânica, operadora de máquinas, soldadora, operária de fábrica, trabalhadora da mineração, pedreira, pintora, eletricitista, encanadora, motorista, caminhoneiro. **Grupo 4:** Professora (de Ensino Fundamental ou médio, idioma, música, artes etc.), técnica (de enfermagem, contabilidade, eletrônica etc.), policial, militar de baixa patente (soldado, cabo, sargento), corretora de imóveis, supervisora, gerente, mestre de obras. **Grupo 5:** Médica, engenheira, dentista, psicóloga, economista, advogada, juíza, promotora, defensora, delegada, tenente, capitã, coronel, professora universitária, diretora empresas públicas ou privadas, política, proprietária de empresas com mais de 10 mil empregados.

Ao se avaliar a nota média dos candidatos em relação à ocupação do pai e da mãe,

observa-se uma relação direta entre esta e o desempenho dos candidatos no Enem 2018.

Os candidatos filhos de pai e mãe com ocupações relativas ao grupo 1 (que reúne profissões como lavrador, pescador, seringueiro e extrativista) possuem notas médias mais baixas. No extremo oposto, os candidatos que realizaram o Enem 2018 e obtiveram o maior desempenho são aqueles que pai e mãe estão no grupo 5 (que reúne profissões como médico, engenheiro, juiz, coronel e professor universitário).

Por fim, levando em consideração o Estado de origem de cada candidato é possível avaliar tanto a quantidade de inscritos quanto o rendimento desses candidatos por Estado. O treemap, Figura 4.7, resume os resultados do Enem por unidade da federação. O tamanho do Estado no gráfico representa a quantidade de candidatos, enquanto a cor representa a média das notas obtidas pelos candidatos sendo que a gradação em vermelho representa as melhores notas médias e a gradação em cinza representa as menores notas médias, conforme a escala abaixo do gráfico.

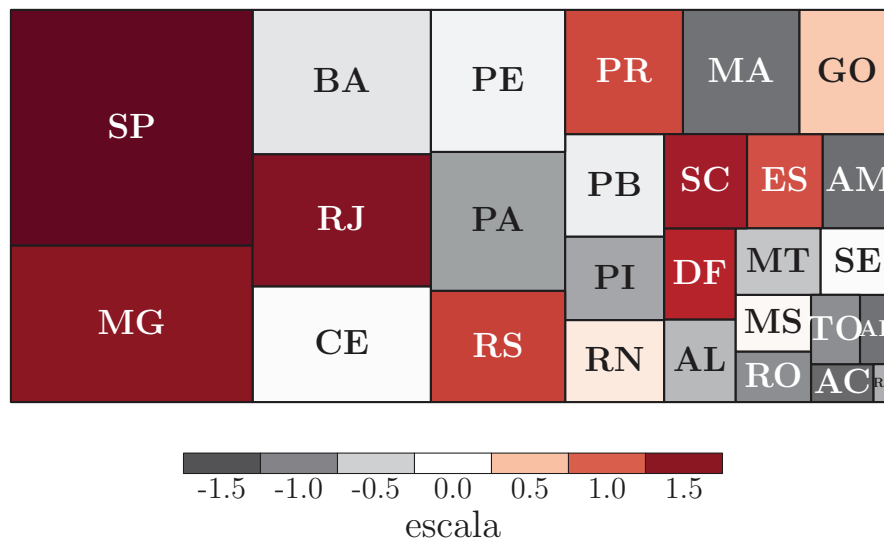


Figura 4.7: Treemap da média das notas do Enem por Estado

Quanto ao número de candidatos, é interessante verificar que somando-se os candidatos das 8 unidades da federação com mais participantes no Enem (SP, MG, BA, RJ, CE, PE, PA e RS) houve mais concorrentes do que na soma dos 19 Estados com menos candidatos. Isso ocorre principalmente por serem estados mais populosos. Observa-se ainda que as melhores notas médias foram obtidas por candidatos de Estados da região Sudeste (SP, RJ e MG) e de SC. No extremo oposto, existe uma concentração de Estados da região Norte com as piores notas médias (destacadamente AC, AM, RR e AP).

4.2 Análise do modelo de regressão

A seleção das variáveis submetidas ao modelo de regressão logística multinomial foi embasada na literatura pregressa sobre o desempenho dos alunos, a avaliação do pesquisador

e os resultados da análise da estatística descritiva. Dentre as variáveis selecionadas, a variável Idade assume valores contínuos, a variável Ano de Conclusão do Ensino Médio é ordinal, por existir uma gradação clara entre as possibilidades de resposta e as demais variáveis (Sexo, Estado Civil, Cor/Raça, Treineiro, Escolaridade do Pai, Escolaridade da Mãe, Renda, Uso de Computador, Acesso a Internet, Conclusão do Ensino Médio e Tipo de Escola) são variáveis categóricas nominais.

A interpretação dos coeficientes da regressão logística multinomial nas variáveis categóricas deve levar em consideração a categoria base para cada variável. No presente estudo, as categorias bases de cada variável categórica estão relacionadas na Tabela 4.17:

Tabela 4.17: Categorias base das variáveis dependentes

Variável	Categoria Base
TP_SEXO	F - Feminino
TP_ESTADO_CIVIL	0 - Solteiro
TP_COR_RAÇA	1 - Branca
IN_TREINEIRO	0 - Não
Escolaridade_Pai	A - Não estudou ou candidato não soube informar
Escolaridade_Mãe	A - Não estudou ou candidato não soube informar
Ocupação_Pai	A - Ocupação com menor requisito de formação
Ocupação_Mãe	A - Ocupação com menor requisito de formação
Renda	A - Renda de até 1 salário mínimo
Computador_em_Casa	A - Não possui computador em casa
Acesso_a_Internet	A - Não possui acesso à internet na residência
Conclusão_Ensino_Médio	A - Já ter concluído o Ensino Médio
Tipo_Escola	A - Ensino Médio somente em escola pública

No processamento do modelo de regressão foi escolhida como categoria de referência para a variável resposta, a categoria 2, que considera os candidatos com notas entre 497,22 e 587,46. A opção por adotar essa categoria como referência foi pelo fato de se utilizar, usualmente, a categoria com maior número de ocorrências (POWERS et al., 2008).

Assim, após o ajuste realizado nos níveis de cada resposta de variável independente, descrito na Sessão 3.3, e definida a categoria de referência da variável resposta, foi processada a regressão logística no software RStudio com o uso da linguagem R.

É importante destacar que os resultados apresentados pelo modelo de regressão não necessariamente irão refletir os resultados apresentados na análise descritiva. Isto porque, para permitir o processamento do modelo mais ajustado, alguns níveis de resposta das variáveis independentes foram adequados ou excluídos do banco de dados (conforme Sessão 3.3).

Os resultados do modelo de regressão logística serão apresentados separadamente, de acordo com as categorias da variável resposta em comparação com a categoria de referência. Inicialmente é realizada uma análise detalhada dos resultados da regressão logística da categoria 0 comparada à categoria 2 e, como as interpretações são realizadas de maneira análoga nas demais categorias, para estas será realizada uma análise geral dos dados. Nessa análise geral será usada a designação de candidato padrão para fazer menção ao candidato que pertence, para determinada variável independente, à categoria base. Destaca-se que

não está se designando por candidato padrão aquele que possui simultaneamente todas as características de base elencadas na 4.17, visto que a análise deve ser realizada para cada variável individualmente. Assim, por exemplo, quando mencionado o candidato padrão e se tratar da variável sexo abrange-se todas as candidatas do sexo feminino (ainda que sejam não solteiros). Por outro lado, quando tratar-se do estado civil, candidato padrão faz menção a todos os solteiros (ainda que do sexo masculino).

Antes da apresentação dos resultados ainda é importante destacar a diferença entre dois termos normalmente utilizados como sinônimos, mas que possuem diferença conceitual: probabilidade e chance. Probabilidade é a proporção de vezes que se espera que um evento ocorra em um processo realizado repetidas vezes. Chance é a razão entre a probabilidade de um evento ocorrer e a probabilidade dele não ocorrer. Neste trabalho esses termos são aplicados considerando tal diferença.

A Tabela 4.18 demonstra os resultados da regressão logística multinomial para a categoria 0 em relação a categoria de referência (2). Observa-se que, com exceção do Estado Civil, todas as demais variáveis apresentaram $p\text{-valor} < 0,05$ e, por isso, são estatisticamente significativas neste modelo. O intercepto representa uma estimativa logit em relação à categoria de referência quando as variáveis preditoras são avaliadas em zero. Entretanto, o objetivo deste trabalho é avaliar a participação das variáveis no rendimento dos candidatos, logo, o intercepto não trará uma interpretação prática ao estudo.

Os valores informados na coluna *Statistic* são referentes ao teste de Wald que, juntamente com o $p\text{-valor}$, define a significância do coeficiente.

A exponencial dos coeficientes refere-se a razão de risco relativo de cada variável estar na categoria analisada em relação à categoria de referência e é de grande relevância para a análise dos coeficientes. Quando os coeficientes são menores que zero, a exponencial do coeficiente é menor que 1 e indica uma redução na chance do candidato estar na categoria estudada quando comparada à categoria de referência, dadas as demais variáveis constantes; nos casos em que o coeficiente é maior que zero, a razão de risco relativo é maior que 1 e indica um aumento na chance do candidato estar na categoria estudada quando comparada à categoria de referência, dadas as demais variáveis constantes.

Analisando os resultados da Tabela 4.18, individualmente, observa-se que a cada aumento de um ano na idade, considerando as demais variáveis constantes, as probabilidades logísticas multinomiais de receber nota na categoria 0 em detrimento da categoria de nota 2 seriam aumentadas em 0,0812 unidades. Como as notas na categoria 0 são inferiores às notas da categoria 2, por meio da razão de risco relativo, pode-se concluir que candidatos mais velhos teriam a chance aumentada em 8,46% a cada ano a mais de ter nota inferior aos candidatos mais novos.

Quando se avalia o sexo, como se trata de uma variável preditora categórica dicotômica, ela compara homens e mulheres para a nota na categoria observada em relação a categoria de referência, mantendo as demais variáveis constantes. O logit multinomial para homens em relação a mulheres é 0,1287 unidades maior de estar na categoria 0 em relação à categoria 2, mantidas as demais variáveis do modelo constantes. A razão de risco relativo mostra que, os homens são mais propensos que as mulheres de estar na categoria 0 quando comparado com a categoria 2.

No processamento deste modelo, quando avaliada a categoria 0 em relação a categoria

2, a variável Estado Civil não se mostrou estatisticamente significativa e, por esse motivo, terá seus resultados desconsiderados neste caso específico.

Tabela 4.18: Resultado de regressão - categoria 0

Variável	Coef.	St.Err	Statistic	P.Value	ExpCoef
Categoria 0					
Intercept	-2,5797	0,0156	-165,1302	<0,001	0,0758
NU_IDADE	0,0812	0,0005	160,7383	<0,001	1,0846
TP_SEXOM	0,1287	0,0052	24,5821	<0,001	1,1374
TP_ESTADO_CIVIL1	0,0088	0,0109	0,8068	0,4198	1,0088
TP_COR_RACA2	0,3696	0,0059	62,1163	<0,001	1,4471
TP_ANO_CONCLUIU	-0,1299	0,0012	-108,9673	<0,001	0,8782
IN_TREINEIRO1	-0,6625	0,0164	-40,4982	<0,001	0,5156
Escolaridade_Pai_B	-0,1594	0,0072	-22,0535	<0,001	0,8526
Escolaridade_Pai_C	-0,0464	0,0150	-3,0907	0,0020	0,9547
Escolaridade_Mãe_B	-0,4826	0,0091	-52,8101	<0,001	0,6172
Escolaridade_Mãe_C	-0,3639	0,0138	-26,4568	<0,001	0,6949
Ocupação_Pai_B	-0,1671	0,0096	-17,3438	<0,001	0,8461
Ocupação_Mãe_B	0,0711	0,0095	7,5197	<0,001	1,0737
Renda_B	-0,4803	0,0059	-81,4077	<0,001	0,6186
Renda_C	-0,9315	0,0102	-91,1987	<0,001	0,3940
Renda_D	-1,5196	0,0288	-52,7895	<0,001	0,2188
Renda_E	-1,4037	0,0389	-36,0481	<0,001	0,2457
Computador_em_Casa_B	-0,5932	0,0063	-93,9425	<0,001	0,5526
Acesso_a_Internet_B	-0,2199	0,0062	-35,3996	<0,001	0,8025
Conclusão_Ensino_Médio_B	0,4239	0,0071	59,7379	<0,001	1,5280
Conclusão_Ensino_Médio_C	1,2072	0,0152	79,3039	<0,001	3,3441
Tipo_Escola_B	-0,2477	0,0095	-25,9818	<0,001	0,7806
Tipo_Escola_C	-0,2140	0,0121	-17,6331	<0,001	0,8073
Categoria de referência: 2					

Fonte: Elaborada pela autora

A razão de risco relativo da variável Cor/raça mostra que um candidato da raça preta tem 44,71% mais chance que o candidato branco de obter nota na categoria 0 se comparado com a categoria 2, sendo mantidas as demais variáveis constantes.

O ano de conclusão do Ensino Médio demonstra que os candidatos com mais tempo de conclusão do Ensino Médio tem possibilidade 0,1299 menor que os candidatos com conclusão de Ensino Médio mais recente de estar na categoria 0 comparada à categoria 2.

O candidato treineiro tem a possibilidade 0,6625 menor que o candidato não treineiro de estar na categoria 0 em relação à categoria 2, mantendo-se as demais variáveis constantes. A razão de risco relativo confirma esta perspectiva ao mostrar que o treineiro tem a chance reduzida em mais de 48% em comparação ao não treineiro de estar na categoria 0 em relação a categoria 2.

Ao se avaliar a escolaridade do pai e da mãe observa-se que os candidatos cujos pais estudaram, seja até no máximo o Ensino Médio, seja até a graduação e/ou pós-graduação, possuem menos chance que os candidatos cujos pais não estudaram de obter nota na categoria 0 comparada à categoria 2, mantidas as demais variáveis constantes. Comparados aos candidatos cujo pai nunca estudou, o candidato cujo pai concluiu no máximo o Ensino Médio tem a possibilidade 15% menor, aproximadamente, e o candidato cujo pai completou a faculdade e/ou a pós graduação, a redução da possibilidade de estar na categoria 0 é de apenas 4%, comparado à categoria de referência.

A avaliação da escolaridade da mãe pode ser realizada de maneira análoga. Os candidatos cujas mães estudaram, seja até no máximo o Ensino Médio, seja até a graduação e/ou pós-graduação, possuem menos chance que os candidatos cujas mães não estudaram de obter nota na categoria 0 comparada à categoria 2, mantidas as demais variáveis constantes. Comparados aos candidatos cuja mãe nunca estudou, o candidato cuja mãe concluiu no máximo o Ensino Médio tem chance 48% menor e o candidato cuja mãe concluiu a faculdade e/ou pós graduação, a chance é 30% menor, mantidas as demais variáveis constantes.

Os candidatos cujo pai possui ocupação com maior requisito de formação possui aproximadamente 16% menos chance que o candidato cujo pai possui ocupação com menor requisito de formação de estar na categoria de nota 0 em relação à categoria de referência. Entretanto, quando se avalia a ocupação da mãe, candidatos com mãe que possuem ocupação com maior requisito de formação possuem 7,37% mais chance que o candidato cuja mãe tem ocupação com menor requisito de formação de obter nota na categoria 0 em relação à categoria 2, mantidas as demais variáveis constantes.

Na variável renda observa-se que em todas as categorias de resposta há uma redução da chance em relação à categoria base (candidato com renda de até um salário mínimo) de se obter nota na categoria 0 em relação à categoria 2. Isso mostra que candidatos com renda superior possuem menos chance do que candidatos com renda de até um salário mínimo de estar com nota na categoria 0 em relação à categoria 2, mantidas as demais variáveis constantes. Além disso, observa-se que, até um certo ponto, quanto maior a renda, menor a chance de obter a nota nesta categoria avaliada em relação à categoria de referência.

Os candidatos que afirmaram ter computadores em casa têm, aproximadamente, 45% menos chance que os candidatos que não possuem o equipamento de estar na categoria de nota 0 em relação à categoria de referência. Desta maneira, observa-se que candidatos que utilizam o computador em casa são propensos a ter notas mais elevadas se comparado aos candidatos que não tem este acesso. Este resultado vai ao encontro do encontrado na variável Acesso a Internet: candidatos com acesso a internet tem 20% menos chance que os candidatos sem acesso a internet de obter nota na categoria em relação à categoria 2, mantidas as demais variáveis constantes.

Ao ser avaliado o ano de conclusão do Ensino Médio observa-se que candidatos que concluiriam o Ensino Médio no ano do exame (2018) possuíam 53% mais chance que os candidatos que já concluíram o Ensino Médio de obter nota na categoria 0 em relação à categoria de referência, já os candidatos que ainda não concluiriam o Ensino Médio em 2018 possuíam 234% mais chance que os candidatos que já concluíram o Ensino Médio de obter nota na categoria 0 em relação à categoria 2, mantidas as demais variáveis constantes.

Esse resultado pode estar associado ao fato de que o candidato que concluiu o Ensino Médio já foi apresentado a todo o conteúdo deste e, por este motivo, está mais propenso a obter notas maiores no exame.

Os resultados da variável Tipo de Escola mostram que candidatos que estudaram o Ensino Médio pelo menos parcialmente em escola da rede privada tem sua chance reduzida em relação ao candidato que estudou o Ensino Médio integralmente em escola da rede pública de estar na categoria 0 comparado à categoria de referência. Comparados ao candidato que estudou integralmente na rede pública, o candidato que estudou parcialmente em escola da rede privada tem 22% menos chance e o que estudou integralmente na rede privada tem, aproximadamente, 19% menos chance de estar na categoria 0 em relação à categoria 2.

Desta maneira, com a Tabela 4.18 é possível compreender como algumas características afetam positiva ou negativamente as chances de um candidato pertencer à categoria de nota 0 em relação à categoria de nota 2 (referência). Quando comparado ao candidato padrão (aquele que possui a característica base, quando avaliada uma variável independente específica), são características que, analisadas individualmente, reduzem as chances de o candidato pertencer à categoria de nota 0 em relação à categoria de nota 2: ser treineiro; ser filho de pai que completou no máximo o Ensino Médio ou que completou a faculdade ou a pós-graduação; ser filho de mãe que completou no máximo o Ensino Médio ou que completou a faculdade ou a pós-graduação; o pai possui ocupação com maior requisito de formação; pertencer a grupo familiar com renda superior a 1 salário mínimo (abrange todas as faixas superiores); possuir computador em casa; morar em residência com acesso à internet; ter estudado pelo menos parte do Ensino Médio em escola privada. Pertencer à categoria de nota 0 implica ter obtido uma nota menor do que aqueles que pertencem à categoria de nota 2. Por reduzirem as chances de obtenção de uma nota menor, as características citadas acima, analisadas individualmente, podem ser associadas a um melhor desempenho no exame quando comparadas às características do candidato padrão.

Por outro lado, são características que, analisadas individualmente, aumentam as chances de estar na categoria de nota 0 em relação a estar na categoria 2: ser homem; ser da cor/raça preta; a mãe possui ocupação com maior requisito de formação; ainda não ter concluído o Ensino Médio. Essas características podem ser associadas a retirar uma nota menor no exame, quando comparadas às características do candidato padrão. Embora a variável Estado Civil tenha apresentado coeficiente negativo, a mesma não se mostrou estatisticamente significativa.

A Tabela 4.19 demonstra os resultados da regressão logística multinomial para a categoria 1 em relação a categoria de referência. Neste caso, pela estatística de Wald e pelo p-valor, observa-se que todas as variáveis são estatisticamente significativas e, por meio da análise dos coeficientes e da razão de risco relativo é possível compreender como algumas características afetam positiva ou negativamente as chances de o candidato pertencer à categoria de nota 1 em relação à categoria de nota 2 (referência).

Quando comparado ao candidato padrão, são características que, analisadas individualmente, reduzem as chances de o candidato pertencer à categoria de nota 1 em relação à categoria de nota 2: ser homem; ser treineiro; ser filho de pai que completou no máximo o Ensino Médio; ser filho de mãe que completou no máximo o Ensino Médio ou que

completou a faculdade ou a pós-graduação; o pai possui ocupação com maior requisito de formação; pertencer a grupo familiar com renda superior a 1 salário mínimo (abrange todas as faixas superiores); possuir computador em casa; morar em residência com acesso à internet; ter estudado pelo menos parte do Ensino Médio em escola privada. Pertencer à categoria de nota 1 implica ao candidato ter obtido uma nota menor do que aqueles que pertencem à categoria de nota 2. Por reduzirem as chances de tirar uma nota menor, as características acima, quando analisadas individualmente dentro do modelo, podem ser associadas a um melhor desempenho no exame quando comparadas às características do candidato padrão.

Tabela 4.19: Resultado de regressão - categoria 1

Variável	Coef.	St.Err	Statistic	P.Value	ExpCoef
Categoria 1					
Intercept	-0,0717	0,0093	-7,7365	<0,001	0,9308
NU_IDADE	0,0401	0,0003	117,6122	<0,001	1,0409
TP_SEXOM	-0,0776	0,0026	-29,6419	<0,001	0,9253
TP_ESTADO_CIVIL1	0,2367	0,0055	42,9218	<0,001	1,2670
TP_COR_RACA2	0,1631	0,0028	59,3951	<0,001	1,1771
TP_ANO_CONCLUIU	-0,0658	0,0007	-99,5513	<0,001	0,9363
IN_TREINEIRO1	-0,2856	0,0109	-26,0964	<0,001	0,7516
Escolaridade_Pai_B	-0,0856	0,0039	-21,7258	<0,001	0,9179
Escolaridade_Pai_C	0,0520	0,0068	7,6812	<0,001	1,0534
Escolaridade_Mãe_B	-0,1712	0,0056	-30,8006	<0,001	0,8427
Escolaridade_Mãe_C	-0,1762	0,0072	-24,4514	<0,001	0,8385
Ocupação_Pai_B	-0,1211	0,0042	-28,8508	<0,001	0,8860
Ocupação_Mãe_B	0,0269	0,0043	6,3176	<0,001	1,0273
Renda_B	-0,3432	0,0031	-110,2043	<0,001	0,7095
Renda_C	-0,6092	0,0045	-135,7138	<0,001	0,5438
Renda_D	-0,7338	0,0095	-77,1802	<0,001	0,4801
Renda_E	-0,7682	0,0136	-56,3092	<0,001	0,4638
Computador_em_Casa_B	-0,3094	0,0031	-101,1165	<0,001	0,7339
Acesso_a_Internet_B	-0,1903	0,0033	-57,0082	<0,001	0,8267
Conclusão_Ensino_Médio_B	0,1496	0,0035	42,6104	<0,001	1,1613
Conclusão_Ensino_Médio_C	0,5117	0,0105	48,5602	<0,001	1,6681
Tipo_Escola_B	-0,2478	0,0044	-56,9239	<0,001	0,7805
Tipo_Escola_C	-0,4555	0,0053	-85,2699	<0,001	0,6341
Categoria de Referência: 2					

Fonte: Elaborada pela autora

Por outro lado, são características que, analisadas individualmente, aumentam as chances de o candidato estar na categoria de nota 1 em relação a estar na categoria 2: não ser solteiro; ser da cor/raça preta; ser filho de pai que completou a faculdade ou a pós-graduação; a mãe possui ocupação com maior requisito de formação; ainda não ter

concluído o Ensino Médio. Essas características, analisadas individualmente dentro do modelo, podem ser associadas a retirar uma nota menor no exame, quando comparadas às características do candidato padrão.

A Tabela 4.20 apresenta os resultados da regressão logística multinomial para a categoria 3 em relação a categoria 2 (referência). Neste caso, assim como na Tabela 4.19, todos os coeficientes apresentaram-se estatisticamente significativos e por meio deles e da razão de risco relativo é possível compreender como algumas características afetam positiva ou negativamente as chances de o candidato pertencer à categoria de nota 3 em relação à categoria de nota 2.

Tabela 4.20: Resultado de regressão - categoria 3

Variável	Coef.	St.Err	Statistic	P.Value	ExpCoef
Categoria 3					
Intercept	-0,8122	0,0174	-46,5959	<0,001	0,4439
NU_IDADE	-0,0633	0,0007	-87,8232	<0,001	0,9387
TP_SEXOM	0,2003	0,0032	62,4766	<0,001	1,2218
TP_ESTADO_CIVIL1	0,0558	0,0085	6,5858	<0,001	1,0574
TP_COR_RACA2	-0,1159	0,0033	-34,8702	<0,001	0,8906
TP_ANO_CONCLUIU	0,0464	0,0011	42,1017	<0,001	1,0475
IN_TREINEIRO1	0,0594	0,0172	3,4547	<0,001	1,0612
Escolaridade_Pai_B	-0,0285	0,0060	-4,7819	<0,001	0,9719
Escolaridade_Pai_C	0,2514	0,0077	32,7513	<0,001	1,2859
Escolaridade_Mãe_B	0,3811	0,0101	37,5640	<0,001	1,4638
Escolaridade_Mãe_C	0,5316	0,0110	48,1949	<0,001	1,7016
Ocupação_Pai_B	0,2266	0,0044	51,6458	<0,001	1,2543
Ocupação_Mãe_B	0,0569	0,0046	12,4422	<0,001	1,0586
Renda_B	0,1057	0,0050	21,1305	<0,001	1,1115
Renda_C	0,4480	0,0058	77,4992	<0,001	1,5652
Renda_D	0,5746	0,0085	67,3657	<0,001	1,7764
Renda_E	0,5813	0,0106	54,8852	<0,001	1,7883
Computador_em_Casa_B	0,3522	0,0044	81,0316	<0,001	1,4222
Acesso_a_Internet_B	0,1323	0,0054	24,4786	<0,001	1,1415
Conclusão_Ensino_Médio_B	-0,1303	0,0044	-29,4918	<0,001	0,8779
Conclusão_Ensino_Médio_C	-0,5680	0,0170	-33,4976	<0,001	0,5667
Tipo_Escola_B	0,2177	0,0049	44,4884	<0,001	1,2432
Tipo_Escola_C	0,6481	0,0048	136,5114	<0,001	0,1912
Categoria de Referência: 2					

Fonte: Elaborada pela autora

Quando comparado ao candidato padrão, as características ser da cor/raça preta; ser filho de pai que completou no máximo o Ensino Médio e ainda não ter concluído o Ensino Médio, quando analisadas de forma individual no modelo, reduzem as chances de o candidato pertencer à categoria de nota 3 em relação à categoria de nota 2:. Como a

categoria de nota 3 possui intervalo de notas maior que o intervalo de notas da categoria 2, pertencer à categoria de nota 3 implica ter obtido uma nota maior do que aqueles que pertencem à categoria de nota 2. Desta maneira, por reduzirem as chances de obtenção de uma nota maior, as características acima, analisadas individualmente, podem ser associadas a um desempenho inferior no exame quando comparadas às características do candidato padrão.

Em contrapartida, as características que, analisadas individualmente, aumentam as chances de estar na categoria de nota 3 em relação a estar na categoria 2 são: ser homem; não ser solteiro; ser treineiro; ser filho de pai que completou a faculdade ou a pós-graduação; ser filho de mãe que completou no máximo o Ensino Médio ou que completou a faculdade ou a pós-graduação; o pai possui ocupação com maior requisito de formação; a mãe possui ocupação com maior requisito de formação; pertencer à grupo familiar com renda superior a 1 salário mínimo (abrange todas as faixas superiores); possuir computador em casa; morar em residência com acesso à internet; ter estudado pelo menos parte do Ensino Médio em escola privada. Essas características, analisadas individualmente no modelo, podem ser associadas ao candidato obter uma nota maior no exame, quando comparadas às características do candidato padrão.

A Tabela 4.21 apresenta os resultados da regressão logística multinomial para a categoria 4 em relação a categoria 2 (referência). Neste caso, através da estatística de Wald e do p-valor todos os coeficientes apresentaram-se estatisticamente significativos e por meio deles e da razão de risco relativo é possível compreender como algumas características afetam positiva ou negativamente as chances de o candidato pertencer à categoria de nota 4 em relação à categoria de nota 2.

As características que, analisadas individualmente, reduzem as chances de um candidato, quando comparado ao candidato padrão, pertencer à categoria de nota 4 em relação à categoria de nota 2 são: não ser solteiro; ser da cor/raça preta; ainda não ter concluído o Ensino Médio. Pertencer à categoria de nota 4 implica ter obtido uma nota maior do que aqueles que pertencem à categoria de nota 2. Por reduzirem as chances de obtenção de uma nota maior, as características acima, analisadas individualmente, podem ser associadas a um desempenho inferior no exame quando comparadas às características do candidato padrão.

As características que, analisadas individualmente, aumentam as chances de o candidato estar na categoria de nota 4 comparado a estar na categoria 2 são: ser homem; ser treineiro; ser filho de pai que completou no máximo o Ensino Médio ou que completou a faculdade ou a pós-graduação; ser filho de mãe que completou no máximo o Ensino Médio ou que completou a faculdade ou a pós-graduação; o pai possui ocupação com maior requisito de formação; a mãe possui ocupação com maior requisito de formação; pertencer à grupo familiar com renda superior a 1 salário mínimo (abrange todas as faixas superiores); possuir computador em casa; morar em residência com acesso à internet; ter estudado pelo menos parte do Ensino Médio em escola privada. Essas características, analisadas individualmente, podem ser associadas a retirar uma nota maior no exame, quando comparadas às características do candidato padrão.

Tabela 4.21: Resultado de regressão - categoria 4

Variável	Coef.	St.Err	Statistic	P.Value	ExpCoef
Categoria 4					
Intercept	-2,1780	0,0346	-62,9876	<0,001	0,1133
NU_IDADE	-0,0648	0,0015	-43,9574	<0,001	0,9373
TP_SEXOM	0,0796	0,0051	15,6168	<0,001	1,0829
TP_ESTADO_CIVIL1	-0,3270	0,0182	-17,9316	<0,001	0,7211
TP_COR_RACA2	-0,1711	0,0054	-31,7873	<0,001	0,8427
TP_ANO_CONCLUIU	0,0061	0,0021	2,9473	0,0032	1,0061
IN_TREINEIRO1	0,0933	0,0371	2,5161	0,0119	1,0978
Escolaridade_Pai_B	0,0359	0,0115	3,1119	0,0019	1,0365
Escolaridade_Pai_C	0,5040	0,0131	38,5637	<0,001	1,6554
Escolaridade_Mãe_B	0,1917	0,0191	10,0552	<0,001	1,2114
Escolaridade_Mãe_C	0,4555	0,0199	22,9022	<0,001	1,5769
Ocupação_Pai_B	0,2655	0,0069	38,6327	<0,001	1,3040
Ocupação_Mãe_B	0,2063	0,0070	29,5061	<0,001	1,2291
Renda_B	0,1019	0,0102	9,9785	<0,001	1,1073
Renda_C	0,6227	0,0109	57,1171	<0,001	1,8639
Renda_D	0,9438	0,0133	70,7660	<0,001	2,5696
Renda_E	1,3439	0,0147	91,5828	<0,001	3,8339
Computador_em_Casa_B	0,5160	0,0083	62,3475	<0,001	1,6753
Acesso_a_Internet_B	0,1962	0,0111	17,7130	<0,001	1,2167
Conclusão_Ensino_Médio_B	-0,4843	0,0068	-71,7938	<0,001	0,6161
Conclusão_Ensino_Médio_C	-1,8087	0,0367	-49,3447	<0,001	0,1639
Tipo_Escola_B	0,2266	0,0087	26,0912	<0,001	1,2543
Tipo_Escola_C	1,3112	0,0069	190,2010	<0,001	3,7108
Categoria de Referência: 2					

Fonte: Elaborada pela autora

Por fim, a Tabela 4.22 apresenta os resultados da regressão logística multinomial para a categoria 5 em relação a categoria 2 (referência). Neste caso, com exceção da variável Sexo, todos os coeficientes apresentaram-se estatisticamente significativos e por meio deles e da razão de risco relativo é possível compreender como algumas características afetam as chances de o candidato pertencer à categoria de nota 5 em relação à categoria de nota 2.

As características que, analisadas individualmente, reduzem as chances de o candidato pertencer à categoria de nota 5 em relação à categoria de nota 2, quando comparado ao candidato padrão são: ser treineiro; ser filho de pai que completou no máximo o Ensino Médio; o pai possui ocupação com maior requisito de formação; possui renda de 1 até 7 salários mínimos; possuir computador em casa; morar em residência com acesso à internet. Pertencer à categoria de nota 5 implica ter obtido uma nota maior do que aqueles que pertencem à categoria de nota 2. Como essas características, analisadas individualmente, reduzem as chances de obtenção de uma nota maior, elas podem ser associadas a um desempenho inferior no exame quando comparadas às características do candidato

padrão. A variável sexo, embora o coeficiente seja negativo, foi desconsiderada da relação por não ser identificada significância estatística.

Tabela 4.22: Resultado de regressão - categoria 5

Variável	Coef.	St.Err	Statistic	P.Value	ExpCoef
Categoria 5					
Intercept	-5,8897	0,0674	-87,3873	<0,001	0,0028
NU_IDADE	0,0505	0,0012	41,4132	<0,001	1,0518
TP_SEXOM	-0,0057	0,0100	-0,5667	0,5709	0,9944
TP_ESTADO_CIVIL1	0,6438	0,0241	26,7300	<0,001	1,9038
TP_COR_RACA2	0,0500	0,0106	4,7224	<0,001	1,0512
TP_ANO_CONCLUIU	-0,1850	0,0027	-67,7839	<0,001	0,8311
IN_TREINEIRO1	-2,1540	0,0305	-70,6054	<0,001	0,1160
Escolaridade_Pai_B	-0,0801	0,0172	-4,6474	<0,001	0,9230
Escolaridade_Pai_C	0,5385	0,0227	23,7462	<0,001	1,7134
Escolaridade_Mãe_B	2,2160	0,0612	36,1849	<0,001	9,1703
Escolaridade_Mãe_C	2,2194	0,0628	35,3284	<0,001	9,2021
Ocupação_Pai_B	-0,2391	0,0150	-15,9194	<0,001	0,7873
Ocupação_Mãe_B	0,5200	0,0145	35,8099	<0,001	1,6820
Renda_B	-0,5540	0,0135	-40,9219	<0,001	0,5747
Renda_C	-0,3146	0,0174	-18,0938	<0,001	0,7301
Renda_D	0,2167	0,0237	9,1256	<0,001	1,2419
Renda_E	0,2495	0,0278	8,9765	<0,001	1,2833
Computador_em_Casa_B	-0,2786	0,0130	-21,4833	<0,001	0,7568
Acesso_a_Internet_B	-0,4858	0,0139	-34,8455	<0,001	0,6152
Conclusão_Ensino_Médio_B	0,0527	0,0128	4,1023	<0,001	1,0541
Conclusão_Ensino_Médio_C	0,8655	0,0260	33,3159	<0,001	2,3762
Tipo_Escola_B	0,2143	0,0167	12,8665	<0,001	1,2389
Tipo_Escola_C	1,1731	0,0144	81,2843	<0,001	3,2321
Categoria de Referência: 2					

Fonte: Elaborada pela autora

Por outro lado, as características que, analisadas individualmente, aumentam as chances de o candidato estar na categoria de nota 5 em relação a estar na categoria 2 são: não ser solteiro; ser da cor/raça preta; ser filho de pai que completou a faculdade ou a pós-graduação; ser filho de mãe que completou no máximo o Ensino Médio ou que completou a faculdade ou a pós-graduação; a mãe possui ocupação com maior requisito de formação; pertencer à grupo familiar com renda superior a 7 salários mínimos; ainda não ter concluído o Ensino Médio; ter estudado pelo menos parte do Ensino Médio em escola privada. Essas características podem ser associadas a retirar uma nota maior no exame, quando comparadas às características do candidato padrão.

A análise de todos os coeficientes do modelo permitiu identificar a maneira como cada característica do candidato interfere no modelo e, em consequência, na nota final

do candidato. A Tabela 4.23 resume esta análise, identificando quais características e em quais momentos houve impacto positivo e negativo na nota do candidato.

Tabela 4.23: Impacto das características na nota

Característica	Impacto sobre a nota				
	0	1	3	4	5
Idade	-	-	-	-	+
Ser homem	-	+	+	+	o
Não ser solteiro	o	-	+	-	+
Ser da cor/raça preta	-	-	-	-	+
Ter concluído Ensino Médio há mais tempo	+	+	+	+	-
Ser treineiro	+	+	+	+	-
Pai completou no máximo o Ensino Médio	+	+	-	+	-
Pai completou a faculdade ou a pós-graduação	+	-	+	+	+
Mãe completou no máximo o Ensino Médio	+	+	+	+	+
Mãe completou a faculdade ou a pós-graduação	+	+	+	+	+
Pai possui ocupação com maior requisito de formação	+	+	+	+	-
Mãe possui ocupação com maior requisito de formação	-	-	+	+	+
Renda de 1 a 2,5 salários mínimos	+	+	+	+	-
Renda de 2,5 até 7 salários mínimos	+	+	+	+	-
Renda de 7 até 12 salários mínimos	+	+	+	+	+
Renda acima de 12 salários mínimos	+	+	+	+	+
Possuir computador em casa	+	+	+	+	-
Morar em residência com acesso à internet	+	+	+	+	-
Concluir o Ensino Médio em 2018	-	-	-	-	+
Não concluir o Ensino Médio em 2018	-	-	-	-	+
Ensino Médio em escola pública/escola privada	+	+	+	+	+
Ensino Médio somente em escola privada	+	+	+	+	+

Fonte: Elaborada pela autora

o Resultado sem significância estatística

O resultado demonstra que a maior escolaridade da mãe, a renda familiar de 7 ou mais salários mínimos e ter estudado pelo menos parte do Ensino Médio em escola particular são características que estão sempre associadas positivamente ao desempenho do candidato. Além dessas características, destacam-se as seguintes características que geralmente estão associadas a um melhor desempenho: possuir computador em casa; morar em residência com acesso à internet; possuir renda de 1 até 7 salários mínimos; o pai possui ocupação com maior requisito de formação; ser candidato treineiro, e; ter concluído o Ensino Médio há mais tempo.

Por outro lado, a análise dos dados permite concluir que, geralmente, o fato de o candidato ainda não ter concluído o Ensino Médio, ser da raça/cor preta e ser mais velho está associado negativamente ao seu desempenho.

Esses resultados refletem as desigualdades de oportunidades decorrentes do processo de formação histórica da sociedade brasileira e corroboram com os estudos de Figueiredo et al. (2014) e Pereira (2019), que concluíram que para estudantes filhos de mães analfabetas, com renda familiar de até 1 salário mínimo e estudaram predominantemente em escola pública precisam se esforçar em torno de 99,38% a mais do que aqueles filhos de mães com nível superior de escolaridade, renda familiar acima de 9 salários mínimos e que estudou predominantemente em escola particular, sendo o Enem um reflexo do nível socioeconômico dos candidatos.

Conclusões

O presente estudo analisou o desempenho dos candidatos no Exame Nacional do Ensino Médio no ano de 2018. Para isso, valeu-se do banco de dados do Enem fornecido pelo INEP e utilizou os software RStudio para realizar uma avaliação por meio de uma análise descritiva e de modelos de regressão logística multinomial. O pacote *Shiny* do R mostrou-se útil por possibilitar uma visão ampla dos dados da pesquisa a partir da construção de um aplicativo dinâmico que apresenta, de maneira gráfica, as relações entre as variáveis e os resultados no exame. Por sua vez, o modelo de regressão logística multinomial foi capaz de confirmar as expectativas teóricas prévias e fornecer, em termos de chances, quais as probabilidades dos candidatos obterem diferentes padrões de notas em função das suas características pessoais, educacionais e socioeconômicas.

Através da análise descritiva foi possível identificar um perfil geral do candidato no Enem que, em sua maioria, é do sexo feminino (59%), solteiro (92,2%), possui renda familiar mensal inferior a 2 salários mínimos (61,3%) e cujo pai não completou o Ensino Médio. Constatou-se ainda que existe uma maior quantidade de candidatos por 1000 habitantes nas regiões Norte e Nordeste.

Quanto ao desempenho dos candidatos, a análise descritiva permitiu identificar que 94,5% dos candidatos obtiveram nota até 677,7 e, embora haja uma maior quantidade de candidatas do sexo feminino, entre os candidatos que obtiveram melhor desempenho no exame a maioria é do sexo masculino. Verificou-se ainda haver uma representatividade reduzida na categoria de notas mais elevadas dos candidatos declarados indígena, pretos e pardos. Os candidatos que estudaram durante o Ensino Médio somente em escola privada obtiveram desempenho melhor que os demais candidatos. Por fim, ainda na análise descritiva, foi possível identificar uma relação positiva entre a renda familiar e o desempenho no Enem, assim como entre a escolaridade da mãe e do pai e o resultado do exame.

A análise do modelo de regressão logística multinomial possibilitou identificar a participação de cada variável analisada no desempenho dos estudantes no Enem por meio dos coeficientes estimados e pela razão de risco relativo. Por meio da análise desses coeficientes foi possível identificar que características do candidato como renda familiar, escolaridade da mãe e estudo em escola privada impactam de forma positiva o seu rendimento no exame. Considerando individualmente cada variável, um candidato cuja mãe fez faculdade,

comparado a um candidato cuja mãe não estudou, tem mais de 8 vezes mais chances de obter nota na categoria mais elevada comparado à categoria 2, considerada de referência e que abrange notas entre 497,22 e 587,46. Quando é avaliado o tipo de escola que o candidato estudou o Ensino Médio, o candidato que estudou integralmente na rede privada, comparado ao que estudou integralmente na rede pública de ensino, tem 2 vezes mais chance de obter nota na categoria mais elevada comparada à categoria de referência.

A avaliação dos coeficientes da renda mostram que candidato com renda familiar superior a 12 salários mínimos, comparado a um candidato com renda familiar de até 1 salário mínimo possui, aproximadamente, 3 vezes mais chance de obter nota na categoria 4 (elevada) em relação à categoria 2. Esse resultado permite afirmar que existe uma desigualdade de oportunidades e que candidatos com maior renda possuem mais chances de ingressarem no Ensino Superior através do resultado do Enem. É possível afirmar ainda que quanto mais concorrido for o curso e mais elevada a nota de corte, maior se torna a referida desigualdade.

O modelo logístico de regressão confirmou a análise descritiva no que diz respeito à variável cor/raça. O candidato que se declarou da cor/raça preta (no modelo de regressão a cor/raça preta abrange pretos, pardos e indígenas), comparado ao candidato que se declarou branco (considerados no modelo os brancos e amarelos), possui 44% mais chance de estar na categoria de nota 0 (nota mais baixa) comparado à categoria de referência. Esses resultados refletem as desigualdades de oportunidades decorrentes do processo de formação histórica da sociedade brasileira.

Desta maneira, conjuntamente, os resultados da análise descritiva e dos modelos de regressão demonstram a relevância e o impacto de variáveis socioeconômicas e educacionais no desempenho dos candidatos ao Enem. Tais resultados apontam para a importância da manutenção/ampliação de políticas afirmativas que visam reduzir o impacto das desigualdades socioeconômicas históricas no resultado do exame. Contudo, para além disso, o resultado confirma a inviabilidade de obtenção do desenvolvimento educacional dissociado de desenvolvimento socioeconômico e cultural, demonstrando que apenas ações integradas são capazes de reverter o quadro educacional no Brasil.

Por fim, este estudo reconhece uma limitação de ordem prática, embora o estudo considere a premissa de que variáveis socioeconômicas influenciam significativamente o desempenho dos candidatos no Enem, o mesmo reconhece que outras variáveis como a capacidade cognitiva do candidato, o esforço individual, horas dedicadas ao estudo e as técnicas de estudo utilizadas na preparação podem ser ainda mais relevantes. Contudo, tal fato não invalida ou mesmo enfraquece os achados da presente pesquisa, que podem contribuir para a elaboração de políticas públicas que visem a redução da desigualdade no exame.

Referências

- ANDRADE, Adriana Oliveira. Aplicação do Modelo Logístico Multinomial no Estudo da Decisão do Voto. **Rio de Janeiro: ENCE**, 2006.
- BILDER, Christopher R; LOUGHIN, Thomas M. **Analysis of categorical data with R**. [S.l.]: Chapman e Hall, 2014.
- BONETI, Lindomar Wessler; OLIVEIRA, Geisa Melo de. Enem: análise do desempenho escolar nas edições de 2009 a 2013. **Revista Espaço Pedagógico**, v. 24, n. 2, 2017.
- CAMEY, Suzi Alves; TORMAN, Vanessa Bielefeldt Leotti; HIRAKATA, Vania Naomi; CORTES, Renan Xavier; VIGO, Alvaro. Viés da razão de chances estimada pela regressão logística multinomial para estimar o risco relativo ou a razão de prevalência e alternativas. **Cadernos de Saúde Pública**, SciELO Brasil, v. 30, n. 1, p. 21–29, 2014.
- CHANG, Winston; CHENG, Joe; ALLAIRE, J; XIE, Yihui; MCPHERSON, Jonathan et al. Shiny: web application framework for R. **R package version**, v. 1, n. 5, 2017.
- FÁVERO, Luiz Paulo; BELFIORE, Patrícia. **Manual de análise de dados: estatística e modelagem multivariada com Excel®[®], SPSS®[®] e Stata®[®]**. [S.l.]: Elsevier Brasil, 2017.
- FÁVERO, Luiz Paulo; BELFIORE, Patrícia; SILVA, FL da; CHAN, Betty Lilian. Análise de dados: modelagem multivariada para tomada de decisões. Rio de Janeiro: Elsevier, 2009.
- FIGUEIRA, Cleonis Viater. **Modelo de Regressão Logística. 2006. 149 f.** 2006. Tese (Doutorado) – Dissertação (Mestrado em Matemática)-Programa de Pós-Graduação em Matemática.
- FIGUEIREDO, Erik; NOGUEIRA, Lauro; SANTANA, Fernanda Leite. Igualdade de Oportunidades: Analisando o papel das circunstâncias no desempenho do ENEM. **Revista Brasileira de Economia**, SciELO Brasil, v. 68, n. 3, p. 373–392, 2014.
- FILHO, Dalson Figueiredo; ROCHA, Enivaldo; PARANHOS, Ranulfo; ALEXANDRE, José. Regressão Logística em Ciência Política. **Pernambuco, PE: UFPE**, 2015.
- GUJARATI, Damodar N; PORTER, Dawn C. **Econometria Básica**. [S.l.]: Amgh Editora, 2011.

HAGUETTE, André; PESSOA, Márcio Kleber Moraes; VIDAL, Eloísa Maia. Dez escolas, dois padrões de qualidade. Uma pesquisa em dez escolas públicas de Ensino Médio do Estado do Ceará. **Ensaio: Avaliação e Políticas Públicas em Educação**, SciELO Brasil, v. 24, n. 92, p. 609–636, 2016.

HAIR, Joseph F; BLACK, William C; BABIN, Barry J; ANDERSON, Rolph E. *Multivariate data analysis: Pearson new international edition*. **Essex: Pearson Education Limited**, 2014.

HAIR, Joseph F; BLACK, William C; BABIN, Barry J; ANDERSON, Rolph E; TATHAM, Ronald L. **Análise multivariada de dados**. [S.l.]: Bookman Editora, 2009.

HOSMER JR, David W; LEMESHOW, Stanley; STURDIVANT, Rodney X. **Applied logistic regression**. [S.l.]: John Wiley & Sons, 2013. v. 398.

INEP. **Operação Enem. Conheça todas as etapas do maior exame do Brasil**.

2017. Disponível em: <https://download.inep.gov.br/educacao_basica/enem/downloads/2017/operacao_enem.pdf>.

_____. **Primeira aplicação do Enem completa 20 anos nesta quinta-feira, 30 de agosto**. 2018. Disponível em: <http://portal.inep.gov.br/artigo/-/asset_publisher/B4AQV9zFY7Bv/content/primeira-aplicacao-do-enem-completa-20-anos-nesta-quinta-feira-30-de-agosto/21206>.

KLEINBAUM, David G; DIETZ, K; GAIL, M; KLEIN, Mitchel; KLEIN, Mitchell. **Logistic regression**. [S.l.]: Springer, 2002.

KONRATH, Andréa Cristina; SILVA, Silvio Aparecido da; HENNING, Elisa; SANTOS, Luciane Mulazani dos; MIRANDA, Rodrigo Gabriel de; SAMOBYL, Robert Wayne. Desenvolvimento de Aplicativos Web Com R e Shiny: inovações no ensino de Estatística. **Abakós**, v. 6, n. 2, p. 55–71, 2018.

LI, LD; CHAGAS, André Luis Squarize. Efeitos do SISU sobre a migração e a evasão estudantil. **XV ENCONTRO NACIONAL DA ASSOCIAÇÃO BRASILEIRA DE ESTUDOS REGIONAIS E URBANOS. Anais...** São Paulo: Aber, 2017.

LOPES, Letícia dos Santos. Definição de um modelo de cobrança (Collection Score) utilizando regressão logística multinomial, 2004.

MACHADO, Paulo Henrique Alves; LIMA, Elizeth. O ENEM no contexto das políticas para o Ensino Médio. **Perspectiva**, v. 32, n. 1, 2014.

MAIA, Dhiego. **Enem, 20, avança e democratiza acesso ao ensino superior no Brasil**. Out. 2018. Disponível em:

<<https://www1.folha.uol.com.br/educacao/2018/10/enem-20-avanca-e-democratiza-acesso-ao-ensino-superior-no-brasil.shtml>>.

MAROCO, João. **Análise estatística: com utilização do SPSS**. [S.l.: s.n.], 2007.

MONT'ALVÃO, Arnaldo; LÔPO, Arnaldo. Estratificação do acesso ao ensino superior no Brasil. **Belo Horizonte, MG: UFMT**, 2013.

MORETTIN, Pedro Alberto; BUSSAB, WILTON OLIVEIRA. **Estatística básica**. [S.l.]: Saraiva Educação SA, 2017.

PEREIRA, Fernanda Martins. **O peso do capital cultural nos resultados do Enem-2015 em Ribeirão Preto e o contrassenso de seu uso como indicador de qualidade das escolas**. 2019. Tese (Doutorado) – Universidade de São Paulo.

POWERS, Daniel; XIE, Yu. **Statistical methods for categorical data analysis**. [S.l.]: Emerald Group Publishing, 2008.

SALES, Leonardo Jorge. Fatores sociais como determinantes de rendimento do Enem, 2018.

SILVEIRA, Fernando Lang da; BARBOSA, Marcia Cristina Bernardes;

SILVA, Roberto da. Exame Nacional do Ensino Médio (ENEM): uma análise crítica.

Revista Brasileira de Ensino de Física, SciELO Brasil, v. 37, n. 1, p. 1101, 2015.

TRAVITZKI, Rodrigo. **ENEM: limites e possibilidades do Exame Nacional do Ensino Médio enquanto indicador de qualidade escolar**. 2013. Tese (Doutorado) – Universidade de São Paulo.

VIÇOSA, Universidade Federal de. **Simulador de Notas SISU**. Set. 2019. Disponível em: <<https://sisusimulator.com.br/ufv/universidade-federal-de-vicosa>>.

VIGGIANO, Esdras; MATTOS, Cristiano. O desempenho de estudantes no Enem 2010 em diferentes regiões brasileiras. **Revista Brasileira de Estudos Pedagógicos**, v. 94, n. 237, 2013.

_____. _____. **Revista Brasileira de Estudos Pedagógicos**, SciELO Brasil, v. 94, n. 237, p. 417–438, 2013.