

UNIVERSIDADE FEDERAL DE ALFENAS

BRUNO PEREIRA COGO

**INFLUÊNCIA DAS CARACTERÍSTICAS SOCIOECONÔMICAS NO INGRESSO EM
CURSOS ALTAMENTE CONCORRIDOS NO SISU 2023.1: UM ESTUDO COM
ALUNOS DE MINAS GERAIS**

VARGINHA/MG

2023

BRUNO PEREIRA COGO

**INFLUÊNCIA DAS CARACTERÍSTICAS SOCIOECONÔMICAS NO INGRESSO EM
CURSOS ALTAMENTE CONCORRIDOS NO SISU 2023.1: UM ESTUDO COM
ALUNOS DE MINAS GERAIS**

Trabalho de conclusão de PIEPEX
apresentado ao Instituto de Ciências Sociais
Aplicadas da Universidade Federal de Alfenas
como requisito parcial à obtenção do título de
Bacharel em Ciência e Economia.

Orientador: Gislene Araujo Pereira

Varginha/MG

2023

RESUMO

Este estudo investiga o impacto de variáveis pessoais, educacionais e econômicas no ingresso em cursos altamente concorridos no Sistema de Seleção Unificada (SISU) na edição de 2023.1. Focalizando alunos do último ano do ensino médio em Minas Gerais que participaram do Exame Nacional do Ensino Médio (ENEM) em 2022, o objetivo é compreender como esses fatores contribuem para as desigualdades educacionais em um estado socioeconômico diversificado. A análise dos cursos mais concorridos, com destaque para Medicina e Direito, foi realizada utilizando o Relatório da edição Sisu 2023/1, obtido através da Chamada Regular (CR) do Ministério da Educação (MEC). Utilizando o modelo de regressão logística e microdados do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), variáveis como idade, sexo, raça, tipo de escola, escolaridade dos pais e renda familiar foram consideradas. Os resultados indicam que, para o curso de Medicina, candidatos com renda até R\$3.030,00 têm menor probabilidade de admissão, destacando a influência das condições financeiras nessa escolha acadêmica. Para o curso de Direito, também se observa um coeficiente negativo para a renda, evidenciando as persistentes desigualdades socioeconômicas no acesso, enquanto outras variáveis também exercem impacto na probabilidade de ingresso. A necessidade premente de políticas públicas voltadas para uma educação mais equitativa e acessível, visando reduzir as disparidades socioeconômicas e raciais, destaca-se como a principal conclusão a ser extraída desses resultados.

Palavras-chave: Enem; SiSU; desigualdades; socioeconômica.

ABSTRACT

This study investigates the impact of personal, educational, and economic variables on admission to highly competitive courses in the Unified Selection System (SISU) in the 2023.1 edition. Focusing on high school seniors in Minas Gerais who participated in the National High School Exam (ENEM) in 2022, the goal is to understand how these factors contribute to educational inequalities in a socioeconomically diverse state. The analysis of the most competitive courses, with a focus on Medicine and Law, was conducted using the Report from the SISU 2023/1 edition, obtained through the Regular Call (CR) of the Ministry of Education (MEC). Using logistic regression models and microdata from the National Institute for Educational Studies and Research Anísio Teixeira (INEP), variables such as age, gender, race, type of school, parents' education, and family income were considered. The results indicate that, for the Medicine course, candidates with income up to R\$3,030.00 have a lower probability of admission, highlighting the influence of financial conditions on this academic choice. For the Law course, a negative coefficient for income is also observed, highlighting the persistent socioeconomic inequalities in access, while other variables also impact the probability of admission. The pressing need for public policies focused on a more equitable and accessible education, aiming to reduce socioeconomic and racial disparities, stands out as the main conclusion to be drawn from these results.

Keywords: ENEM; SISU; inequalities; socioeconomic.

LISTA DE QUADROS E GRÁFICOS

Quadro 1 - Matriz de confusão	12
Quadro 2 - Descrição das Variáveis independentes	14
Gráfico 1 - Número de candidatos inscritos no Enem de 2022 por estado	16

LISTA DE TABELAS

Tabela 1 - Resumo das Características dos Participantes do ENEM 2022 em Minas Gerais	17
Tabela 2 - Notas de Corte dos Cursos no SiSU 2023.1 em Minas Gerais	18
Tabela 3 - Resultado de Regressão - Curso de Medicina	19
Tabela 4 - Resultado de Regressão - Curso de Direito	20
Tabela 5 - Impacto das característica nos cursos de Medicina e Direito	21
Tabela 6 - Matriz de Confusão para a Variável Dependente "Direito"	22
Tabela 7 - Matriz de Confusão para a Variável Dependente "Medicina"	22

SUMÁRIO

1 INTRODUÇÃO	7
2 FUNDAMENTAÇÃO TEÓRICA	9
2.1 Modelagem Estatística	9
2.2 Regressão Logística	9
3 METODOLOGIA	13
4 ANÁLISE DOS RESULTADOS	15
4.1 Análise Exploratória	15
4.2 Análise dos resultados da regressão logística	18
5 CONSIDERAÇÕES FINAIS	23
REFERÊNCIAS	25

1 INTRODUÇÃO

O Exame Nacional do Ensino Médio (Enem) foi criado em 1998 pelo Ministério da Educação com a intenção de avaliar o nível do ensino médio no Brasil (GOV, 2023). Inicialmente, apenas os alunos que estavam concluindo os estudos participavam do exame com o objetivo de identificar deficiências no sistema educacional e auxiliar na formulação de políticas públicas relacionadas à gestão escolar. Ao longo dos anos, o Enem passou por mudanças significativas que o transformaram em uma ferramenta importante para conectar os estudantes às instituições de ensino superior em todo o país. Essas transformações foram relevantes porque não apenas permitiram que uma gama mais ampla de alunos participasse do exame, mas também garantiram que ele se alinhasse mais de perto com as demandas das instituições de ensino superior, melhorando assim sua relevância e eficácia, como foi evidenciado na proposta apresentada à Associação Nacional dos Dirigentes das Instituições Federais de Ensino Superior (Andifes) para o Novo Enem em 2009:

Um exame nacional unificado, desenvolvido com base numa concepção de prova focada em habilidades e conteúdos mais relevantes, passaria a ser importante instrumento de política educacional, na medida em que sinalizaria concretamente para o ensino médio orientações curriculares expressas de modo claro, intencional e articulado para cada área de conhecimento (Brasil, 2009, p. 4).

A avaliação ocorre anualmente em todo o país, e os participantes são submetidos a testes em quatro áreas do conhecimento: Linguagens e Códigos e suas tecnologias, Matemática e suas tecnologias, Ciências da Natureza e suas tecnologias, e Ciências Humanas e suas tecnologias. São 180 questões objetivas e uma redação dissertativa argumentativa, que abordam conteúdos relacionados ao currículo do ensino médio. A pontuação das questões e da redação varia de 0 a 1.000 pontos, onde cada questão de múltipla escolha tem um valor específico de pontos, e a redação também recebe uma pontuação com base em critérios como coerência, organização, uso de evidências e poder persuasivo da escrita. A nota final é calculada como uma combinação das pontuações alcançadas em cada área do exame (Inep, 2023).

A partir da realização do Enem, os candidatos têm a oportunidade de concorrer ao Sistema de Seleção Unificada (SiSU), que é uma plataforma digital criada e gerenciada pelo Ministério da Educação que permite que os estudantes

utilizem suas pontuações como critério para ingressar em diversas instituições públicas de ensino superior em todo o país, incluindo universidades federais, institutos federais e algumas instituições estaduais, que oferecem vagas em cursos de graduação.

Uma característica proeminente do SiSU é sua intensa competição. Devido à grande procura de candidatos e ao número limitado de vagas disponíveis, muitos cursos apresentam notas de corte elevadas, enfatizando a necessidade de os alunos obterem notas excepcionais no Enem para maximizar suas chances de seleção. Portanto, é fundamental estudar a situação socioeconômica dos participantes, pois esses aspectos estão presentes no ambiente escolar e podem afetar o seu desempenho. Além disso, esta análise pode fornecer informações valiosas sobre os obstáculos que impedem o desenvolvimento e o progresso entre vários grupos sociais. Isto é particularmente relevante em regiões como Minas Gerais, com sua diversidade, onde prevalecem realidades distintas:

A grande extensão territorial, somada à sua posição geográfica central, conferiu ao estado um desenvolvimento histórico peculiar. Possui municípios no grupo seleta de melhores indicadores sociais e, também, no grupo de piores indicadores do país (Santos; Pales, 2014, p. 2).

Uma avaliação desses traços e sua associação com o desempenho acadêmico dos estudantes delimita áreas específicas que requerem atenção e melhoria por meio de políticas educacionais efetivas.

Nesse cenário, este estudo tem como objetivo analisar e compreender como as características socioeconômicas influenciam o ingresso em cursos altamente concorridos no SiSU 2023.1, com foco em alunos que estavam no último ano do ensino médio no estado de Minas Gerais e realizaram o ENEM de 2022. Para isso, será utilizada a metodologia de regressão logística nos microdados desse exame, disponibilizados em 02/05/2023 pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep).

A estrutura deste trabalho contempla cinco seções. A primeira seção trata-se da introdução, a segunda seção apresenta a fundamentação teórica que traz conceitos de modelagem estatística e regressão logística, a terceira seção apresenta a metodologia usada para realizar a análise dos dados. A quarta seção concentra-se na análise descritiva dos dados, seguida pela análise dos resultados. Por fim, têm-se as considerações finais, nas quais são apresentadas as conclusões e implicações derivadas dos resultados obtidos.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Modelagem Estatística

A modelagem estatística representa uma ferramenta valiosa para simplificar e compreender fenômenos complexos. Utilizando modelos matemáticos, seu objetivo principal é identificar padrões, relações causais e compreender variáveis interdependentes, desempenhando um papel crucial na formulação e teste de hipóteses. Apesar das limitações, como suposições e amostragens, a modelagem estatística é considerada um instrumento heurístico, buscando uma compreensão refinada dos fenômenos, sem necessariamente atingir a precisão absoluta. No contexto da análise de dados, os pesquisadores têm à disposição uma ampla gama de modelos estatísticos, incluindo os modelos de regressão (Neoway, 2022).

A análise estatística por meio de regressão desempenha um papel crucial na investigação da influência exercida por uma variável sobre outras (Hoffmann, 2016), sendo amplamente aplicada em diversas áreas da pesquisa científica, como saúde e finanças. Dentre os diversos modelos de regressão disponíveis, que incluem a regressão linear simples e múltipla, regressão logística, regressão polinomial, regressão de Poisson, regressão Ridge e Lasso, além da regressão não linear, a escolha do modelo adequado depende da natureza dos dados e dos objetivos da pesquisa.

Embora a regressão linear seja amplamente reconhecida e aplicada, este estudo específico opta pela utilização da regressão logística. Por meio desse modelo, busca-se analisar os dados e compreender de que maneira as variáveis independentes impactam a variável dependente. No contexto deste estudo, o foco recai sobre como as características socioeconômicas influenciam a admissão em cursos altamente concorridos no SiSU 2023.1.

2.2 Regressão Logística

De acordo com Figueira (2006), os modelos de regressão logística se destacam por possuírem a variável dependente categorizada como ordinal, quando há uma ordem natural entre as categorias, ou nominal, quando não há uma ordem definida ou esta é irrelevante. Esse enfoque torna-se particularmente útil ao lidar

com variáveis não contínuas, sendo aplicável quando a categorização diferenciada de uma variável é necessária.

Um ponto relevante a destacar na regressão logística em comparação com a regressão linear é a diferença no método de estimativa dos coeficientes de regressão. Na regressão logística, o método dos mínimos quadrados não é aplicável, em vez disso, utiliza-se o Método da Máxima Verossimilhança.

Fávero (2015) destaca que a regressão logística tem como objetivo analisar a probabilidade de ocorrência de um evento representado por Y (sendo $Y = 1$ para descrever a ocorrência do evento e $Y = 0$ para a ausência do evento). Esse método utiliza variáveis explicativas ($X_i, i=1, \dots, p$) para compreender de que maneira esses atributos impactam a probabilidade do evento em questão. No âmbito desta análise, o autor formula a expressão geral para a probabilidade estimada de ocorrência do evento como:

$$P(Y = 1) = \frac{1}{1 + e^{-g(x)}}, \quad (1)$$

onde $g(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$

Os coeficientes $\beta_0, \beta_1, \dots, \beta_p$ são determinados a partir do conjunto de dados por meio do método da máxima verossimilhança. Este método procura os valores dos parâmetros que apresentam a maior probabilidade de reproduzir o padrão de observações na amostra de dados (Gonzalez, 2018).

A interpretação dos coeficientes na regressão logística é crucial para compreender o efeito das variáveis independentes na probabilidade de um evento ocorrer (probabilidade de sucesso). Quando o coeficiente é positivo ($\beta_i > 0$), um aumento na variável independente está associado a um aumento na probabilidade de sucesso. Já quando o coeficiente é negativo ($\beta_i < 0$), um aumento na variável independente está associado a uma diminuição na probabilidade de sucesso. Além disso, o coeficiente β_0 representa o risco inicial e indica probabilidade inicial de sucesso. Em resumo, nota-se que o coeficiente β_i representa o quanto a exposição ao fator de risco (característica) contribuirá com o aumento/decréscimo da probabilidade de sucesso.

Outra interpretação útil na regressão logística envolve o impacto de cada coeficiente na razão de chances (*Odds Ratio-OR*), esse impacto expressa o grau de

associação entre y e X_i e é calculado ao elevar a constante de Euler à potência do coeficiente (Mesquita, 2014),

$$OR_i = e^{\beta_i}. \quad (2)$$

É essencial realizar uma avaliação abrangente do modelo ajustado, considerando tanto a sua significância quanto a qualidade. A avaliação da significância do modelo ajustado pode ser efetuada por meio de dois testes cruciais: o Teste da Razão de Verossimilhanças e o Teste de Wald.

O Teste da Razão de Verossimilhanças tem por objetivo verificar se é viável remover um subconjunto de variáveis X_i do modelo de regressão logístico, ou seja, se os coeficientes de regressão $\beta_i, i = 1, \dots, p$ são estatisticamente nulos. Por sua vez, o Teste de Wald é empregado para analisar a significância individual dos coeficientes no contexto do modelo de regressão logística. Ambos os testes são fundamentais para garantir uma análise robusta e abrangente do modelo ajustado (Mezzomo, 2009).

A qualidade de um modelo de regressão logística pode ser verificada por meio da matriz de confusão ou tabela de contingência. A matriz de confusão é uma tabela de contingência em que nas linhas estão os valores observados (y) e nas colunas estão valores previstos (\hat{y}), como ilustrado no Quadro 1.

Quadro 1 - Matriz de confusão

$y \setminus \hat{y}$	$\hat{y} = 0$	$\hat{y} = 1$
$y = 0$	VN	FP
$y = 1$	FN	VP

Fonte: Elaboração pelo autor (2023).

No Quadro 1 tem-se:

VN = Verdadeiro Negativo;

VP = Verdadeiro Positivo;

FN = Falso Negativo;

FP = Falso Positivo.

A partir desses valores são definidos três conceitos importantes: acurácia, sensibilidade e especificidade. A acurácia mensura a precisão geral do modelo ao prever corretamente tanto os verdadeiros positivos quanto os verdadeiros negativos. Por outro lado, a sensibilidade destaca a eficácia do modelo em identificar todos os casos positivos, minimizando a omissão de casos relevantes. Já a especificidade avalia a capacidade do modelo em classificar corretamente instâncias negativas (Mattos, 2009).

$$\text{Acurácia} = \frac{VP + VN}{n}$$

$$\text{Sensibilidade} = \frac{VP}{VP + FN}$$

$$\text{Especificidade} = \frac{VN}{VN + FP}$$

3 METODOLOGIA

Para este estudo, os dados foram obtidos a partir dos microdados do ENEM 2022, disponibilizados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). A análise foi direcionada aos indivíduos matriculados no último ano do ensino médio no estado de Minas Gerais e, para garantir a qualidade dos dados e focar na população-alvo, foram aplicados critérios de filtragem.

Inicialmente, a seleção concentrou-se nos candidatos que atendiam a esse critério específico. Em seguida, foram excluídos aqueles que não compareceram a todas as etapas do exame, bem como os que apresentavam dados faltantes. O resultado desse processo resultou em um banco de dados contendo um total de 67.159 observações.

A seleção das variáveis independentes para análise foi baseada em sua potencial relevância e na suposição de uma relação significativa com o desempenho dos participantes. Esse processo de seleção seguiu a abordagem preconizada por Hair et al. (2009), que orienta a identificação das variáveis por meio de pesquisa prévia e modelos teóricos relevantes para a questão de pesquisa. Adicionalmente, consideram-se estudos anteriores sobre o impacto de determinados fatores no desempenho dos participantes que realizaram o ENEM no município de Varginha-MG (Andrade e Pereira, 2022).

As variáveis independentes definidas para esta análise incluem faixa etária, sexo, raça, tipo de escola frequentada no ensino médio, escolaridade do homem e da mulher responsáveis e renda mensal familiar. Para representar essas variáveis categóricas numericamente, foram criadas variáveis dummies correspondentes a cada categoria, conforme apresentado no Quadro 2.

Quadro 2 - Descrição das variáveis independentes

Variável independentes	Descrição	Sigla
Faixa etária	0 – Até 18 anos	Faixa_etaria
	1 – Maior que 18 anos	

Sexo	0 – Masculino 1 – Feminino	Sexo
Raça	0 – Branca 1 – Preta, Parda, Amarela e Indígena	Raca
Tipo de escola frequentada	0 – Privada 1 - Pública	Tipo_escola
Escolaridade do pai ou homem responsável	0 – Ensino Médio Completo 1 – Ensino Médio Incompleto	E_pai
Escolaridade da mãe ou mulher responsável	0 – Ensino Médio Completo 1 – Ensino Médio Incompleto	E_mae
Renda mensal familiar	0 – Renda maior que R\$3.030,00 1 – Renda até R\$3.030,00	Renda

Fonte: Elaboração pelo autor.

Importante destacar que, no âmbito deste estudo, a variável dependente é representada pela admissão nos cursos, sendo atribuído o valor 1 para indicar a admissão e 0 para sinalizar a não admissão.

Para identificar os cursos mais concorridos no SiSU 2023.1, utilizou-se o relatório da edição Sisu 2023/1 com informações da Chamada Regular (CR) disponibilizado pelo Ministério da Educação (MEC, 2023). Foram escolhidos os cursos de ampla concorrência com o maior número de inscritos, e em seguida, calculou-se a média das notas de corte dessas opções. Destacam-se os cursos de Medicina e Direito como os mais concorridos.

A análise concentrou-se, portanto, nesses cursos, buscando avaliar o impacto das características socioeconômicas dos candidatos nas notas de corte. As notas dos participantes foram obtidas a partir da média das notas nas provas objetivas de Ciências da Natureza, Ciências Humanas, Linguagens e Códigos, Matemática e na Redação.

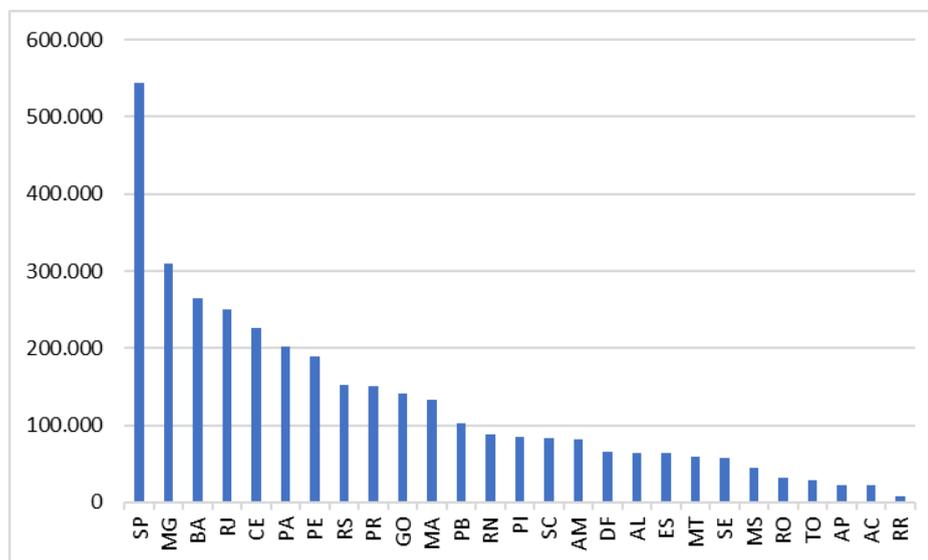
As variáveis de interesse passaram por um processo de adaptação e ajuste visando sua incorporação em um modelo de regressão logística, com o auxílio do software estatístico GRETL.

4 ANÁLISE DOS RESULTADOS

4.1 Análise Exploratória

De acordo com os microdados disponibilizados, Minas Gerais destaca-se como o segundo estado com o maior número de participantes no ENEM 2022, representando 9% do total, ficando apenas atrás do estado de São Paulo. A média das pontuações dos candidatos em Minas Gerais atingiu 533,72, variando de 140,34 a 849,86, com um desvio padrão de 96,40.

Gráfico 1 - Número de candidatos inscritos no Enem de 2022 por estado



Fonte: Microdados Enem (2022).

Ao explorarmos as características demográficas e socioeconômicas dos participantes no estado, são reveladas informações relevantes. A faixa etária dos candidatos é predominantemente composta por aqueles com até 18 anos, abrangendo 94% dos participantes, com uma média de pontuação de 535,73. Os 6% restantes têm mais de 18 anos, com uma média de pontuação de 504,06.

No que se refere ao gênero, observamos que 39% dos participantes são do sexo masculino, com uma média de pontuação de 540,26, enquanto 61% são do sexo feminino, com uma média de pontuação de 529,47.

Em relação à auto identificação racial, 52% dos participantes declararam-se brancos, alcançando uma média de pontuação de 552,51, enquanto 48% se

identificaram como pretos, pardos, amarelos ou indígenas, com uma média de pontuação de 513,43.

A escolha do tipo de escola frequentada pelos participantes também é diversificada, com 23% frequentando escolas privadas e 77% frequentando escolas públicas.

Sobre a escolaridade dos responsáveis pelos participantes, observam-se variações significativas, com 58% dos pais ou homens responsáveis possuindo ensino médio completo e 42% com ensino médio incompleto. No caso das mães ou mulheres responsáveis, destaca-se que 73% têm ensino médio completo, enquanto 27% apresentam ensino médio incompleto.

No que diz respeito à renda mensal familiar, 42% dos participantes têm uma renda mensal familiar superior a R\$3.030,00, com uma média de pontuação de 569,91, enquanto 58% possuem uma renda mensal familiar de até R\$3.030,00, com uma média de pontuação de 507,79.

Tabela 1 - Resumo das características dos participantes do ENEM 2022 em Minas Gerais

Variável	Descrição	Média	Máximo	Mínimo
Faixa etária	Até 18 anos	535,73	849,86	167,56
	Maior que 18 anos	504,06	791,56	140,34
Sexo	Masculino	540,26	849,86	140,34
	Feminino	529,47	814,36	165,86
Raça	Branca	552,51	849,86	167,56
	Preta, Parda, Amarela e Indígena	513,43	820,88	140,34
Tipo de escola frequentada	Privada	594,21	849,86	185,40
	Pública	515,93	828,42	140,34
Escolaridade do pai ou	Ensino Médio Completo	552,51	849,86	140,34

homem responsável	Ensino Médio Incompleto	507,36	803,38	165,86
Escolaridade da mãe mulher responsável	Ensino Médio Completo	546,23	849,86	140,34
	Ensino Médio Incompleto	499,82	828,42	165,86
Renda mensal familiar	Renda maior que R\$3.030,00	569,91	849,86	177,72
	Renda de até R\$3.030,00	507,79	814,68	140,34

Fonte: Microdados Enem (2022).

No que diz respeito aos cursos de ampla concorrência mais disputados no SiSU 2023.1, foram identificados os cinco principais, acompanhados pelo número de inscritos e pela média de notas de corte para cada curso (Tabela 2).

Tabela 2 - Notas de corte dos cursos no SiSU 2023.1 em Minas Gerais

Curso	Número de inscritos	Média de nota de corte
Medicina	111.482	811,43
Direito	53.385	732,61
Administração	39.798	639,46
Psicologia	35.746	729,74
Enfermagem	34.143	707,33

Fonte: Relatório da edição Sisu 2023/1.

Destaca-se que, no caso de Medicina, com seus 111.482 inscritos, a média de nota de corte foi de 811,43. Apenas 0,0208% dos participantes obtiveram a pontuação necessária para ingressar nessas vagas. Quanto a Direito, com 53.385 inscritos e média de nota de corte de 732,61, apenas 1,5545% dos candidatos alcançaram a pontuação adequada para concorrer a essas posições.

Esses dados estabelecem uma base sólida para análises subsequentes sobre como as características socioeconômicas dos participantes impactam o ingresso em cursos altamente concorridos no SiSU 2023.1.

4.2 Análise dos resultados da regressão logística

Na fase inicial da análise, o foco direcionou-se para o curso de Medicina ofertados no SiSU 2023.1, que se destacou como a opção mais procurada no SiSU 2023.1, registrando uma média de nota de corte de 811,43 e um total de 111.482 inscritos. Após o ajuste, as variáveis independentes faixa etária, sexo, raça, tipo de escola frequentada, escolaridade do pai ou responsável masculino, e escolaridade da mãe ou responsável feminina foram excluídas do modelo devido à falta de significância estatística. Desta forma, foi realizado um novo ajuste considerando apenas a variável independente renda, esse ajuste é apresentado na Tabela 3.

Tabela 3 - Resultado de Regressão - Curso de Medicina

	<i>Coefficiente</i>	<i>OR</i>	<i>Erro Padrão</i>	<i>z</i>	<i>p-valor</i>
Const	-7,6758	0,0005	0,2774	-27,67	1,65e-168
Renda	-2,8987	0,0550	1,0378	-2,79	0,0052

Fonte: Elaborada pelos autores com base nos dados da pesquisa (2023).

A análise da Tabela 3 destaca a relevância da renda familiar, representada pela variável "RENDA", na probabilidade de ingresso em Medicina. O coeficiente negativo sugere que candidatos com renda até R\$3.030,00 têm uma probabilidade menor (aproximadamente 0,06 vezes) de serem admitidos no curso, evidenciando a influência das condições financeiras nessa escolha acadêmica. Esta observação confirma as conclusões de Veras, Fernandez, Feitosa e Fernandes (2020), que identificou uma concentração significativa de renda alta entre os estudantes de Medicina na Universidade Federal da Bahia. Notavelmente, 38,8% dos participantes declararam que suas famílias possuíam uma renda superior a 10 salários-mínimos. Adicionalmente, a maioria dos estudantes afirmou que o ingresso na universidade não foi influenciado por políticas de afirmação, totalizando 58% nesse aspecto.

Na esfera da interação entre a pobreza e o processo de ensino-aprendizagem, nota-se que indivíduos com condições sociais desfavoráveis e renda familiar reduzida enfrentam desafios que prejudicam seu desempenho escolar, como destacado por Leonardeli, Ferreira e Ferreira (2020).

Ao analisar os dados relacionados ao curso de Direito, que registrou uma média de nota de corte de 732,61 com um total de 53.385 inscritos, os resultados apresentados na Tabela 4 destacam a influência de variáveis socioeconômicas no processo de admissão. Nota-se, mais uma vez, o coeficiente negativo da “REND A”, reforçando o padrão persistente das desigualdades socioeconômicas no acesso a cursos altamente competitivos.

Tabela 4 - Resultado de Regressão - Curso de Direito

	Coeficiente	OR	Erro Padrão	z	p-valor
Const	-2,4608	0,0854	0,0508	-48,39	0
Renda	-1,1979	0,3018	0,1027	-11,66	1,97E-31
Faixa_etaria	-0,4908	0,6122	0,1883	-2,61	0,0091
Sexo	-0,2153	0,8063	0,0637	-3,38	0,0007
Raca	-0,6971	0,4980	0,0828	-8,42	3,91E-17
Tipo_escola	-1,2644	0,2824	0,0765	-16,53	2,23E-61
E_pai	-0,7243	0,4847	0,1090	-6,64	3,05E-11
E_mae	-0,8183	0,4412	0,1635	-5,01	5,56E-07

Fonte: Elaborada pelos autores com base nos dados da pesquisa (2023).

Além da influência significativa da renda familiar, as demais variáveis analisadas apresentaram impactos semelhantes. Em outras palavras, ter idade superior a 18 anos, ser do sexo feminino, pertencer a uma das categorias raciais preta, parda, amarela e indígena, e possuir pais que não concluíram o ensino médio estão associados a uma probabilidade menor de admissão no curso.

O estudo realizado por Neto (2014, p. 21) reforça a ideia de que, apesar dos esforços para mitigar determinados problemas, as oportunidades de ingresso no ensino superior ainda são fortemente influenciadas por uma variedade de fatores. Dentre esses, destacam-se as origens socioeconômicas, a configuração familiar, a pertença a grupos raciais, o gênero, a experiência no mercado de trabalho e a localização regional. Essas considerações ressaltam a complexidade do cenário educacional e a importância de abordagens abrangentes para promover a equidade no acesso ao ensino superior.

A análise de todos os coeficientes nas razões de chances dos modelos proporcionou uma compreensão direta do impacto de cada atributo do candidato, conforme apresentado na Tabela 5.

Tabela 5 – Impactos negativos (OR) das características nos cursos de Medicina e Direito

Variável	Descrição	OR	
		Medicina	Direito
Faixa etária	Maior que 18 anos		0,6122
Sexo	Feminino		0,8063
Raça	Preta, Parda, Amarela e Indígena		0,4980
Tipo de escola frequentada	Pública		0,2824
Escolaridade do pai ou homem responsável	Ensino Médio Incompleto		0,4847
Escolaridade da mãe ou mulher responsável	Ensino Médio Incompleto		0,4412
Renda mensal familiar	Renda de até R\$3.030,00	0,0550	0,3018

Fonte: Elaborada pelos autores com base nos dados da pesquisa (2023).

Na avaliação dos resultados para o curso de Direito, destaca-se que a variável que exerceu o maior impacto negativo na probabilidade de admissão é a "Faixa etária" para indivíduos com mais de 18 anos, apresentando um Odds Ratio (OR) de 0,6122. Esses dados indicam que estudantes com idade superior a 18 anos têm uma probabilidade significativamente reduzida de serem admitidos em Direito.

Por outro lado, a variável com o menor impacto relativo foi o "Tipo de escola frequentada" (Pública), registrando um OR de 0,2824. Isso sugere que, embora estudantes provenientes de escolas públicas apresentem uma probabilidade inferior de admissão em Direito, o impacto é menos acentuado em comparação com outras variáveis consideradas.

A matriz de confusão apresentada nas Tabelas 6 e 7 revela informações cruciais sobre o desempenho do modelo para os cursos de Direito e Medicina.

Tabela 6 - Matriz de Confusão para a Variável Dependente "Direito"

$y \setminus \hat{y}$	$\hat{y} = 0$	$\hat{y} = 1$
$y = 0$	66115	0
$y = 1$	1044	0

Fonte: Elaborada pelos autores com base nos dados da pesquisa (2023).

$$\text{Acurácia} = \frac{0 + 66115}{67159} = 0,984 \quad (6)$$

$$\text{Sensibilidade} = \frac{0}{0 + 1044} = 0 \quad (7)$$

$$\text{Especificidade} = \frac{66115}{66115 + 0} = 1 \quad (8)$$

Notavelmente, no curso de Direito (conforme Tabela 6), apesar da alta acurácia de 98,4%, a sensibilidade de 0% indica uma falha significativa do modelo em identificar corretamente os casos de admissão. No entanto, é importante notar que o modelo demonstrou precisão na previsão dos casos de não admissão, conforme evidenciado pela especificidade de 100%.

Tabela 7 - Matriz de Confusão para a Variável Dependente "Medicina"

$y \setminus \hat{y}$	$\hat{y} = 0$	$\hat{y} = 1$
$y = 0$	67145	0

$y = 1$	14	0
---------	----	---

Fonte: Elaborada pelos autores com base nos dados da pesquisa (2023).

$$\text{Acurácia} = \frac{0 + 67145}{67159} = 0,999 \quad (3)$$

$$\text{Sensibilidade} = \frac{0}{0 + 14} = 0 \quad (4)$$

$$\text{Especificidade} = \frac{67145}{67145 + 0} = 1 \quad (5)$$

Essa mesma tendência é observada para o curso de Medicina (Tabela 7), onde a acurácia atinge impressionantes 99,9%, mas a sensibilidade permanece em 0%. A alta especificidade, novamente, sugere uma capacidade consistente de prever corretamente os casos de não admissão. O fenômeno da sensibilidade nula dos modelos pode ser atribuído à disparidade na distribuição das classes, um problema conhecido como desbalanceamento de dados. Neste contexto, a classe majoritária, representada pela 'não admissão', prevalece em quantidade significativamente maior do que a classe não majoritária, associada à 'admissão'. Esse desequilíbrio na representação de classes pode levar o modelo a favorecer a previsão da classe majoritária em detrimento da classe menos frequente (FIAMONCINI, 2023).

Os resultados destacando a importância das variáveis nos resultados. No entanto, é crucial interpretar esses resultados com cautela, considerando as complexidades inerentes aos modelos estatísticos e reconhecendo que fatores não contemplados nas variáveis podem influenciar o processo de admissão.

5 CONSIDERAÇÕES FINAIS

Neste estudo, realizou-se uma análise da influência das características socioeconômicas no ingresso em cursos altamente concorridos no SiSU 2023.1, com foco em alunos do último ano do Ensino Médio de Minas Gerais que realizaram o ENEM de 2022. Os dados, provenientes dos microdados do ENEM e do Relatório da edição Sisu 2023/1 do INEP, foram submetidos a uma análise estatística utilizando o software GRET, por meio de modelos de regressão logística.

Os resultados evidenciaram que a variável relacionada à renda familiar mensal exerceu um impacto significativo na probabilidade de admissão, tanto no curso de Medicina quanto em Direito. Candidatos com recursos financeiros mais limitados, representados por uma renda familiar de até R\$3.030,00, apresentaram uma probabilidade inferior de serem admitidos, destacando a influência das condições econômicas nessa decisão.

É relevante observar que, enquanto a renda familiar foi um fator determinante para Medicina e Direito, outras variáveis analisadas afetam predominantemente o curso de Direito. Faixa etária, sexo, tipo de escola frequentada e escolaridade dos pais ou responsável foram associados a uma probabilidade menor de admissão nesse curso. Essa diferenciação destaca a complexidade das variáveis que moldam o acesso a diferentes cursos, enfatizando a necessidade de abordagens específicas para compreender esses fatores em cada área de estudo.

Neste contexto, a urgência de políticas públicas que fomentem uma educação mais igualitária e acessível, reduzindo as disparidades socioeconômicas e raciais, torna-se evidente a partir desses resultados. Investimentos em programas de inclusão social, concessão de bolsas de estudo e aprimoramento na qualidade do ensino público são medidas cruciais para proporcionar oportunidades mais equitativas aos estudantes em Minas Gerais.

Contudo, é crucial destacar as limitações inerentes a este estudo. Os resultados fundamentam-se nos dados específicos do ENEM 2022, concentrando-se nos participantes que estavam cursando o último ano do ensino médio no estado de Minas Gerais. Essa especificidade pode restringir a generalização dos resultados para outras regiões ou para anos subsequentes. Além disso, é importante reconhecer que outras variáveis, que não foram incluídas no modelo, também podem exercer influência no contexto analisado.

É válido observar que o modelo pode ter uma capacidade limitada para prever corretamente os casos de não admissão. Essa consciência das possíveis limitações contribui para uma interpretação mais precisa dos resultados, ressaltando a importância de usar os achados com cautela e indicando áreas que podem ser exploradas em futuras pesquisas.

REFERÊNCIAS

ANDRADE T. P.; PEREIRA, G. A. Análise socioeconômica do desempenho de participantes do ENEM 2020 no município de Varginha-MG. **Cadernos de Estudos Interdisciplinares**, v. 5, n. 1, p. 37-51, 31 mar. 2023.

BRASIL. Ministério da Educação. **Exame Nacional do Ensino Médio (Enem)**. 2023 (On-line). Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enem>>. Acesso em: 15 set. 2023.

BRASIL. Ministério da Educação. **Proposta à Associação Nacional dos Dirigentes das Instituições Federais de Ensino Superior**. 2023 (On-line). Disponível em: <http://portal.mec.gov.br/component/docman/?task=doc_download&gid=768&Itemid=>. Acesso em: 15 set. 2023.

BRASIL. Portal Sisu. **Microdados**. 2023 (On-line). Disponível em: https://sisu.mec.gov.br/static/pdf/Portal%20Sisu_Sisu%202023-1_Inscri%C3%A7%C3%B5es%20e%20notas%20de%20corte.xlsx. Acesso em: 15 set. 2023.

FAVERO, L. P. **Análise de Dados**. Estatística e Modelagem Multivariada com Excel®, SPSS® e Stata®. Rio de Janeiro: Elsevier, 2017.

FIAMONCINI, Lucas. **Dados Desbalanceados: Como Equilibrar as Ocorrências das Classes**. Disponível em: <https://aquare.la/dados-desbalanceados-como-equilibrar-as-ocorrencias-das-classes/>. Acesso em: 15 set. 2023.

FIGUEIRA, C. **Modelo de Regressão Logística**. 2006. Dissertação (Mestrado em Matemática) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2006.

GONZALEZ, L. A. **Regressão logística e suas aplicações**. 2018. 45f. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Universidade Federal do Maranhão, São Luís, 2018. Disponível em: <<https://monografias.ufma.br/jspui/bitstream/123456789/3572/1/LEANDRO-GONZALEZ.pdf>>. Acesso em: 15 set. 2023.

HAIR, J. BLACK, W. BABIN, B. ANDERSON, R. TATHAM, R. **Análise multivariada de dados**. São Paulo: Bookman, 2009.

HOFFMANN, R. **Análise de regressão: uma introdução à econometria**. Piracicaba: Usp, 2016. Disponível em: <<https://repositorio.usp.br/bitstream/handle/BDPI/48616/REGRESS.pdf?sequence=>>. Acesso em: 15 set. 2023.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Edital nº 30, de 5 de maio de 2023**. Disponível em: <<https://www.in.gov.br/en/web/dou/-/edital-n-30-de-5-de-maio-de-2023-481632388>>. Acesso em: 15 set. 2023.

LEONARDELI, P. B.; FERREIRA, J. G.; FERREIRA, J. G. Interferência da renda na aprendizagem do educando em duas escolas públicas do município de Linhares/ES. **Revista Multidisciplinar do Vale do Jequitinhonha-Revivale**, v. 1, n. 1, 2021. Disponível em: <<https://revivale.ifnmg.edu.br/index.php/revivale/article/view/6/6>>. Acesso em: 15 set. 2023.

MATOS, P. F. **Projeto “Um ambiente para análise de dados da doença anemia falciforme”**. São Paulo: Universidade de São Paulo, Disponível em: <https://sites.icmc.usp.br/taspardo/TechReportUFSCar2009a-MatosEtAl.pdf> . Acesso em: 15 set. 2023.

MESQUITA, P S B. **Um modelo de regressão logística para avaliação dos programas de pós-graduação no Brasil**. 2014. 106. Dissertação (Dissertação em Ciências e Tecnologia) - Universidade Estadual do Norte Fluminense, Campos dos Goytacazes, 2014. Disponível em: <<https://uenf.br/posgraduacao/engenharia-de-producao/wp-content/uploads/sites/13/2013/04/Disserta%C3%A7%C3%A3o-Paulo-Mesquita.pdf> >. Acesso em: 15 set. 2023.

MEZZOMO, M. **Um estudo da mortalidade infantil: um modelo de regressão logística múltipla**. 2009. 59f. Trabalho de Conclusão de Curso (Especialização em estatística e modelagem educativa) - Universidade Federal de Santa Maria, Santa Maria, Disponível em: <https://repositorio.ufsm.br/bitstream/handle/1/2992/Mezzomo_Meire.pdf?sequence=1&isAllowed=y> . Acesso em: 15 set. 2023.

NEOWAY. **Modelos estatísticos**: o que são e como usá-los para tomar decisões. 2023 (On-line). Disponível em: < <https://blog.neoway.com.br/modelos-estatisticos/> >. Acesso em: 20 nov. 2023.

SANTOS, G. R.; PALES, R. C. Estratégias de desenvolvimento em Minas Gerais: uma análise comparada das macrorregiões de planejamento. **Cadernos do Desenvolvimento**, v. 9, n. 14, p. 163-185, 2018. Disponível em: < <http://www.cadernosdodesenvolvimento.org.br/ojs-2.4.8/index.php/cdes/article/view/129/132>> . Acesso em: 15 set. 2023.

VERAS, R. M. *et al.* Perfil socioeconômico e expectativa de carreira dos estudantes de Medicina da Universidade Federal da Bahia. **Revista Brasileira de Educação Médica**, v. 44, 2020. Disponível em: < <https://www.scielo.br/j/rbem/a/Kh6JgpZZSTZzSvhKrspRSGF/?format=pdf&lang=pt>> . Acesso em: 15 set. 2023.