



**INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO PIAUÍ
CAMPUS URUÇUÍ
CURSO LICENCIATURA EM MATEMÁTICA**

ANA MARIA ARAUJO MOREIRA

**ANÁLISE DO PERFIL SOCIOECONÔMICO E DESEMPENHO EM MATEMÁTICA
DOS ESTUDANTES DO ENEM NO MUNICÍPIO DE URUÇUÍ - PI: UM ESTUDO COM
BASE NOS MICRODADOS DE 2023**

URUÇUÍ - PI

2024

ANÁLISE DO PERFIL SOCIOECONÔMICO E DESEMPENHO EM MATEMÁTICA DOS
ESTUDANTES DO ENEM NO MUNICÍPIO DE URUÇUI - PI: UM ESTUDO COM BASE
NOS MICRODADOS DE 2023

Trabalho de Conclusão de Curso apresentado como exigência para aprovação na disciplina TCC II do Curso de Licenciatura em Matemática do Instituto Federal de Educação, Ciência e Tecnologia do Piauí, Campus Uruçuí - PI.

Aprovada em: 24/02/2025.

BANCA EXAMINADORA

Prof. Dr. Ernandes Guedes Moura (Orientador)
Instituto Federal do Piauí (IFPI)

Prof. Dr. Bruno Ribeiro de Mesquita
Instituto Federal do Piauí (IFPI)

Prof. M.Sc Francisco Nordman Costa Santos
Instituto Federal do Piauí (IFPI)

RESUMO

O Exame Nacional do Ensino Médio (ENEM) desempenha um papel fundamental na educação brasileira, sendo não apenas uma ferramenta de avaliação, mas também um instrumento de transformação social, ao ampliar o acesso ao ensino superior. Este trabalho teve como objetivo mapear o perfil socioeconômico e desempenho em matemática dos participantes que realizam o ENEM, realizando uma análise preditiva com base nos microdados do ENEM 2023. Para esta pesquisa, foram analisadas as notas de Matemática e as questões do questionário socioeconômico, cujos dados foram obtidos do Instituto Nacional de Pesquisas Educacionais Anísio Teixeira (INEP). Os resultados sugerem que tanto características individuais, quanto fatores contextuais têm um papel relevante na nota de matemática. O coeficiente de determinação (R^2) foi de 0,32, indicando que o modelo explica cerca de 32% da variação nas notas.

***Palavras-chave:** ENEM; notas de matemática; perfil dos participantes.

ABSTRACT

The National High School Exam (ENEM) plays a fundamental role in Brazilian education, being not only an assessment tool, but also an instrument of social transformation, by expanding access to higher education. This study aimed to map the socioeconomic profile and mathematics performance of participants who take the ENEM, performing a predictive analysis based on microdata from ENEM 2023. For this research, the Mathematics scores and the questions from the socioeconomic questionnaire were analyzed, whose data were obtained from the National Institute of Educational Research Anísio Teixeira (INEP). The results suggest that both individual characteristics and contextual factors play a relevant role in the mathematics score. The coefficient of determination (R^2) was 0.32, indicating that the model explains approximately 32% of the variation in the scores.

Keywords: ENEM; math grades; participant profile.

LISTA DE ILUSTRAÇÕES

Gráfico 1	– Distribuição das Notas de Matemática.....	26
Gráfico 2	– Correlação Entre a Nota de Matemática e as Variáveis Explicativas dos Participantes do ENEM da cidade de Uruçuí – PI.....	27

LISTA DE TABELAS

Tabela 1	—	Modelo de Regressão Linear Múltipla.....	28
----------	---	--	----

SIGLAS

ENEM	Exame Nacional do Ensino Médio
FA	Frequência Absoluta
FR	Frequência Relativa
FIES	Fundo de Financiamento Estudantil
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
PROUNI	Programa Universidade para Todos
SISU	Sistema de Seleção Unificada

SUMÁRIO

1	INTRODUÇÃO	9
2	REFERENCIAL TEÓRICO	10
2.1	Cronologia do surgimento do Enem	11
3	MATERIAL E MÉTODOS.....	15
3.1	Processo de coleta e tratamento de dados.....	15
3.2	Correlação	16
3.3	Modelo de regressão linear múltipla	17
3.4	Implementação da análise	18
4	RESULTADOS	18
5	CONSIDERAÇÕES FINAIS.....	22
	REFERÊNCIAS.	27

1 INTRODUÇÃO

O Exame Nacional do Ensino Médio (ENEM) é um programa do governo federal que se iniciou no ano de 1998, cujo o principal objetivo é avaliar o desempenho escolar dos estudantes ao término da educação básica. (SILVEIRA, BARBOSA, SILVA, 2015).

Além de medir o nível de aprendizado, o ENEM também desempenha um papel fundamental na democratização do acesso ao ensino superior, servindo como porta de entrada para diversas universidades públicas e privadas no Brasil. Assim, o propósito do ENEM não se concentra na avaliação das instituições educacionais, mas sim na avaliação abrangente das competências e habilidades adquiridas pelos alunos ao fim da educação básica (DE SOUSA FERNANDES et.al,2023). A diversidade de perfil dos estudantes que realizam o ENEM influencia no estudo para saber as questões que os levam a realizar essa avaliação, se é com o intuito de ingressarem nas universidades ou simplesmente realizam ou por que o pai ou a mãe mandaram ou pelo fato de todos os seus colegas de sala de aula estão fazendo também. Os resultados do ENEM desempenham um papel importante ao fornecer informações relevantes aos gestores educacionais, permitindo-lhes tomar decisões fundamentadas sobre o progresso educacional dos alunos.

A partir do ano de 2009 medidas governamentais instigaram que o ENEM não fosse mais somente utilizado como um método de avaliação do Ensino Médio, porém como forma de ingresso ao ensino superior no Brasil. O Sistema de Seleção Unificada (Sisu) sobreveio a atuar em uma maior escala no processo de acomodação dos candidatos às vagas (SILVEIRA, BARBOSA, SILVA, 2015).

O processo de seleção realizado pelo ENEM e o Sisu, proporcionou a diversos alunos a oportunidade de ingressar em uma universidade e frequentar um curso de forma mais acessível, já que a nota obtida poderia ser usada em diversas universidades por todo o Brasil, permitindo também que sujeitos originários de regiões menos desenvolvidas desloquem-se para outras mais desenvolvidas, e com isso ter a oportunidade de mudar de vida. Esta forma de seleção é relevante para constituir um ambiente multicultural nas universidades.

¹ Ana Maria Araújo Moreira. Instituto Federal do Piauí - Campus Uruçuí. cauru.2020117lmat0098@aluno.ifpi.edu.br

² Ernandes Guedes Moura. Instituto Federal do Piauí - Campus Uruçuí. ernandes.guedes@ifpi.edu.br

2 REFERENCIAL TEÓRICO

A educação tem um papel fundamental na produção e reprodução cultural e social e é nesse processo de socialização, onde a educação tem duas dimensões: social – transmissão de herança cultural às novas gerações através do trabalho de varias instituições: e individual – formação de disposição e visões, aquisição de conhecimentos, habilidades e valores. A dimensão individual é subordinada à social no contexto de interesses objetivos e relação de poder, neste caso baseadas na categoria idade-geração, seja na família, seja na escola (SILVEIRA, BARBOSA, SILVA, 2015).

A educação é considerada como sendo reflexo das relações existentes em uma dada comunidade e, houve um momento da história a descoberta de que uns necessitavam ser os que comandam e outros necessitavam ser os comandados. Com isso, a sociedade humana, ao longo do tempo, passou de comunitária à corporativa, de corporativa à mercantilista, de mercantilista à industrial e de industrial à financeira. Até os dias atuais, se relacionamos como se configura. O sistema econômico. (CARVALHO, 2015).

Assim surgiu a necessidade do estudo sobre o perfil socioeconômico e desempenho em matemática dos estudantes do ENEM no município de Uruçuí – PI. A pesquisa foi realizada com base nos microdados do ENEM 2023 do município de Uruçuí-PI, em que os participantes revelam suas características quanto a participantes do Exame Nacional do Ensino Médio (ENEM).

O Exame Nacional do Ensino Médio (ENEM) foi instituído em 1998 e, desde este período ímpar na educação brasileira, ele vem sendo aplicado anualmente, e pode ser definido como política pública de avaliação do Ensino Médio pelo Governo Federal. Os alunos e todos os interessados podem, através desta ferramenta de avaliativa, realizar um julgamento de como estão seus conhecimentos, e ainda ter a oportunidade de ingressar no Ensino Superior e com isso obter benfeitorias de programas federais que utilizam dados estatísticos deste Exame como critérios seletivos (IGNÁCIO, 2010).

Diversos são os interesses e as demandas acordados pelo ENEM. Assim surgem diversos questionamos a respeito e um deles é a questão das ressignificações oferecidas pelos docentes na formulação de seus currículos e práticas escolares com base nesse exame, que é uma realidade no meio escolar, com isso fica evidente a importancia de adaptação dos

educadores para que seus alunos consigam obter um bom desempenho nessa importante avaliação educacional (IGNÁCIO, 2010).

O ENEM está contido dentro do quadro das políticas públicas de avaliação da qualidade da educação brasileira como parte de um sistema amplo que possui outras ferramentas de avaliação nos diferentes níveis de ensino, como o Sistema Nacional de Avaliação da Educação Básica (SAEB). No exercício docente, o exame age como determinantes orientadores e normativos da prática pedagógica (IGNÁCIO, 2010).

2.1 Cronologia do surgimento do Enem

No ano de 1998, surgiu o exame avaliativo brasileiro. A primeira edição do Enem registrou 157.221 inscrições e calculou uma quantidade de 115.575 participantes no dia 20 de agosto do referido ano. Dessa quantidade foi registrado que 83% tinha isenção da taxa de inscrição, que na época tinha o valor de R\$ 20. Entre os inscritos, 53% tinham 18 anos de idade ou menos, e 9% vinha de escolas públicas, o que foi um número muito baixo. Mesmo que o uso das notas do Enem fosse utilizado em apenas duas instituições de educação superior, as provas foram aplicadas em 184 municípios brasileiros (SOUZA, 2021).

No ano de 1999, o Enem começou a ser aderido por mais instituições de ensino e neste ano foi totalizado noventa e três estabelecimentos educacionais aderiram aos resultados provenientes do exame. Assim, o Enem expõe sua credibilidade. Em apenas um ano de efetivação, o número de instituições de educação superior que começaram a utilizar os resultados no Enem subiu de 2 para 93, expressando aumento significativo (SOUZA, 2021).

No ano de 2000, ocorreu uma grande conquista na realização do Enem, pois foi assegurado atendimento especializado para 376 pessoas com necessidades especiais, marcando o início da oferta de recursos de acessibilidade. A aplicação do Enem passa a ser acompanhado por observadores indicados pelas secretarias estaduais de educação e credenciados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), promovendo ainda mais credibilidade a esse sistema avaliativo (SOUZA, 2021).

No ano posterior, em 2001, o exame ganhou uma nova aquisição, as inscrições serem realizadas através da internet. Dessa forma, os concluintes do ensino médio passaram a ter direito à inscrição gratuita. A isenção da taxa também foi garantida aos inscritos que concluíram os estudos na modalidade Educação de Jovens e Adultos (EJA) 12 meses antes da realização

das inscrições, além dos concluintes e egressos do ensino médio que se declararam impossibilitados de pagar a taxa de inscrição (SOUZA, 2021).

Em 2003 ocorreu à sétima edição do Enem, nessa edição utilizou uma nova forma de mapeamento do perfil dos participantes. Desse modo, foi incluída no questionário socioeconômico uma questão referente ao ano de conclusão do ensino médio. Dessa forma, possibilitou aos estudantes dos outros anos realizarem a avaliação como teste. Já que anteriormente, todos os participantes se declaravam concluintes por falta de opção. Com a mudança, os “treineiros” passaram a ser identificados e, em 2003, representaram 19% do total de 1.882.393 inscritos. As provas foram aplicadas em 31 de agosto, em 605 municípios (SOUZA, 2021).

No ano de 2004 aconteceu uma nova oportunidade para os estudantes ingressarem nas universidades que foi a criação do Programa Universidade para Todos (ProUni) começou a utilizar a nota do Enem para o consentimento de bolsas de estudos integrais e parciais aos participantes. Outro marco importante foi a inserção do campo de Cadastro de Pessoa Física (CPF) na ficha de inscrição possibilitou o acompanhamento da direção dessa população, ao longo dos anos, mediante estudos realizados pelo Inep. (SILVEIRA, BARBOSA, SILVA, 2015).

Com o passar dos anos e aumento de universidades utilizando a nova do Enem para o ingresso no ensino superior à procura pela prova aumentou consideravelmente. No ano de 2008 o exame nacional alcançava uma década de execução no país e no seu aniversário ele chegou com novidades, o Inep e o Ministério da Educação (MEC) divulgaram que o Enem se tornaria o processo nacional de seleção para ingresso na educação superior e certificação do ensino médio. Neste ano, mais de 70% dos 4.018.050 inscritos afirmaram que fizeram o Enem para entrar na faculdade ou conseguir pontos para o vestibular. A aplicação em 1.437 municípios foi em 30 de agosto (SILVEIRA, BARBOSA, SILVA, 2015).

No ano de 2009, ocorreu à criação do Sistema de Seleção Unificada (Sisu), o Enem muda de formato. O exame passa a ter 180 questões objetivas, 45 para cada área do conhecimento, e a redação. A aplicação passa a ocorrer em dois dias e o exame começa a certificar a conclusão do ensino médio. Além disso, as matrizes de referência são reformuladas com base nas Matrizes de Referência do Exame Nacional para Certificação de Competências de Jovens e Adultos (Encceja). A edição de 2009 também foi marcada pelo vazamento da prova, que exigiu a preparação de um novo instrumento (SOUZA, 2021).

Em 2011, foi um marco importante, pois o Enem começou a atrelar em suas aplicações mais acessibilidade onde mais de 20 mil participantes com alguma deficiência tiveram direito a atendimento especializado. Neste mesmo ano, participantes que se declaram negros e pardos foram a maioria, totalizando 53% dos 5.366.949 inscritos. As provas foram aplicadas em 22 e 23 de outubro, em 1.603 cidades (IGNÁCIO, 2010).

Em 2013 o Enem torna-se porta de acesso para todas as instituições de educação superior públicas e pela primeira vez, quase todas as instituições federais adotam o Enem como critério de seleção. A nota do exame é utilizada na concessão de bolsas de estudos do programa Ciências sem Fronteiras e passa a ser divulgada por escola com estratificação nos níveis socioeconômicos (SILVEIRA, BARBOSA, SILVA, 2015).

No ano de 2014 ocorreu uma evolução internacional onde as Universidades de Coimbra e Algarve, em Portugal, passaram a aceitar o Enem, dando início as parcerias com instituições de ensino superior de Portugal, que foram autorizadas a valer-se das notas do Enem em seus processos seletivos. Neste mesmo ano, passou a ser permitido o uso do nome social do participante (IGNÁCIO, 2010).

No ano de 2016, para obterem mais segurança, foi iniciada a coleta de dado biométrico durante a aplicação da prova, para diminuir as fraudes que vinham acontecendo nos anos anteriores. As medidas de segurança do exame ficaram ainda mais rígidas, com o uso de detectores de metais nas entradas de todos os banheiros dos locais de prova (SOUZA, 2021).

Em 2017, desenvolveram e aplicaram uma consulta pública para com os resultados obtidos, implementarem melhorias no Enem. Assim após realização de consulta pública com a população para direcionar melhorias, o Enem passa a ser aplicado em dois domingos consecutivos. O exame ficou ainda mais acessível com a estreia da videoprova em Libras para surdos e deficientes auditivos. Outra novidade foi a estreia da prova personalizada com nome e número de inscrição do participante, e a adoção de novo recurso de segurança: identificador de receptor de ponto eletrônico (IGNÁCIO, 2010).

No ano de 2018 ocorreu a comemoração de duas décadas da aplicação do Enem, e com isso ele ganhou um logotipo comemorativo pelos seus 20 anos de existência, além de um documentário histórico e uma série de cinco minidocumentos sobre os bastidores do exame. A solicitação de isenção da taxa de inscrição passou a ser uma fase anterior à inscrição, e os isentos ausentes no ano anterior tiveram de justificar o motivo da falta para garantir a gratuidade novamente. A mudança trouxe bons resultados: o Enem 2018 teve o menor índice de faltosos

desde 2009, quando assumiu o formato em dois dias. O segundo domingo de aplicação ganhou 30 minutos a mais de duração e o quantitativo de detectores de ponto eletrônico aumentou cinco vezes. O número de instituições de educação superior portuguesas que usam as notas do Enem chegou a 35 (SILVEIRA, BARBOSA, SILVA, 2015).

3 MATERIAL E MÉTODOS

3.1 Processo de Coleta e Tratamento de Dados

Este trabalho teve como objetivo realizar uma análise preditiva utilizando os micros dados do ENEM 2023, fornecidos pelo INEP, para identificar os fatores socioeconômicos que mais influenciaram as notas de Matemática dos candidatos da cidade de Uruçuí-PI. Os micros dados incluem informações relacionadas às provas, gabaritos, questões, notas e respostas ao questionário socioeconômico dos participantes. Para esta pesquisa, foram analisadas as notas de Matemática e as questões do questionário socioeconômico, numeradas de Q001 a Q027, resultando em um conjunto de dados extenso com múltiplas opções de resposta.

Os dados foram obtidos do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), e o nosso foco foi apenas os candidatos que realizaram o ENEM de Uruçuí-PI. Inicialmente, o conjunto de dados contava com 1.330 candidatos e 76 variáveis. Após um processo rigoroso de limpeza e pré-processamento realizado no ambiente Jupyter Notebook com a linguagem Python, redundâncias, inconsistências e dados ausentes foram eliminados, reduzindo o conjunto para 935 candidatos e 16 variáveis selecionadas.

As variáveis escolhidas incluem dados sobre os candidatos, com a nota de Matemática (NU_NOTA_MT) como variável dependente. Entre as variáveis independentes estão: sexo (TP_SEXO), estado civil (TP_ESTADO_CIVIL), raça/cor (TP_COR_RACA), ano de conclusão do ensino médio (TP_ANO_CONCLUIU), tipo de escola (TP_ESCOLA), modalidade de ensino (TP_ENSINO) e faixa etária (TP_FAIXA_ETARIA). Também foram incluídas questões do questionário socioeconômico, como escolaridade do pai (Q001), escolaridade da mãe (Q002), ocupação do pai (Q003), ocupação da mãe (Q004), número de pessoas no domicílio (Q005), renda familiar (Q006), presença de empregada doméstica (Q007), acesso a computador (Q024) e uso de internet em casa (Q025). Mais detalhes sobre esses dados estão disponíveis no dicionário de dados no ANEXO.

Algumas variáveis precisaram ser transformadas para as análises de regressão. As variáveis ordinais, como os níveis de escolaridade (Q001) e categorias de renda (Q006), foram convertidas em valores numéricos usando a codificação ordinal (Ordinal Encoding), conforme

recomendado por Teixeira e Cavique (2023), que apontam essa técnica como ideal para preservar a ordem hierárquica das categorias.

Além disso, variáveis nominais, como sexo e estado civil, foram codificadas como variáveis binárias utilizando o método One-Hot Encoding. Segundo Lopez-Arevalo et al. (2020), essa técnica transforma dados categóricos em valores numéricos, aumentando a dimensionalidade e garantindo que os dados sejam adequados para os modelos de regressão.

Após essas transformações, o número de variáveis aumentou de 16 para 34, devido às novas categorias geradas. Essas etapas de pré-processamento foram essenciais para estruturar os dados de forma adequada, garantindo a qualidade das análises.

Por fim, foi gerado um histograma para visualizar a distribuição das notas de Matemática, permitindo observar a densidade das pontuações. Segundo Campello et al. (2019), o agrupamento baseado em densidade é eficiente para identificar padrões em dados, separando regiões de alta e baixa densidade.

3.2 Correlação

Após a fase inicial, foi realizada uma análise de correlação de Pearson para identificar quais variáveis socioeconômicas possuem uma relação linear com a nota de Matemática, a variável resposta. Schober, Boer e Schwarte (2018) explicam que a correlação de Pearson avalia a intensidade da associação entre variáveis, considerando que os dados apresentam uma distribuição normal. O coeficiente de correlação varia de -1 a +1, permitindo medir essa relação: quanto mais próximo de -1 ou +1, mais forte é a correlação.

A fórmula da correlação é dada por:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

Quando o valor de r se aproxima de 1 ou -1, a relação entre a variável explicativa e a resposta é mais forte, sendo positiva ou negativa, respectivamente. Já valores próximos de 0 indicam uma fraca correlação ou até mesmo ausência de associação.

3.3 Modelo de Regressão Linear Múltipla

Conforme Rencher e Schaalje (2008, p. 137), a regressão linear múltipla é empregada para prever uma variável dependente, y , que possui uma relação linear com diversas variáveis independentes, $x_1, x_2, \dots, x_n$. Segundo os autores, o modelo baseia-se na fórmula $y_i = f(x_i) + \varepsilon_i$, onde $f(x_i)$ é uma função de regressão desconhecida, e ε_i são erros independentes com variância constante σ^2 . A função $f(x)$ é obtida por meio da estimativa dos coeficientes $\beta_0, \beta_1, \dots, \beta_n$, conforme a seguinte equação:

$$y = (y_1 \ y_2 \ \dots \ y_m)^T \text{ e,} \quad (1)$$

$$X = \begin{pmatrix} 1 & X_{11} & X_{12} & \dots & X_{1n} \\ 1 & X_{21} & X_{22} & \dots & X_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{m1} & X_{m2} & \dots & X_{mn} \end{pmatrix} \quad (3)$$

$$y = X\beta + \varepsilon \quad (4)$$

Nesta pesquisa, o vetor y representa as notas de Matemática dos candidatos do Enem 2023 de Uruçuí- PI, enquanto a matriz X contém dados pessoais e socioeconômicos desses

participantes, como escolaridade e profissão dos pais. Cada coluna de X corresponde a uma variável coletada pelo INEP.

Rencher e Schaalje (2008, p. 141) destacam que os parâmetros de regressão são estimados pelo método dos mínimos quadrados ordinários (OLS), sendo β definido por:

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (5)$$

A função predita ($f(x)$) é então obtida por:

$$\hat{f}(x) = X\hat{\beta} \quad (6)$$

A qualidade do modelo foi avaliada com o coeficiente de determinação (R^2), conforme descrito por Rencher e Schaalje (2008, p. 161). Esse parâmetro mensura o quanto as variáveis independentes (x 's) explicam a variância da variável dependente (y), que, neste estudo, são as notas de Matemática. Em termos práticos, o R^2 reflete a proporção da variância da nota explicada pelas variáveis socioeconômicas.

3.4 Implementação da Análise

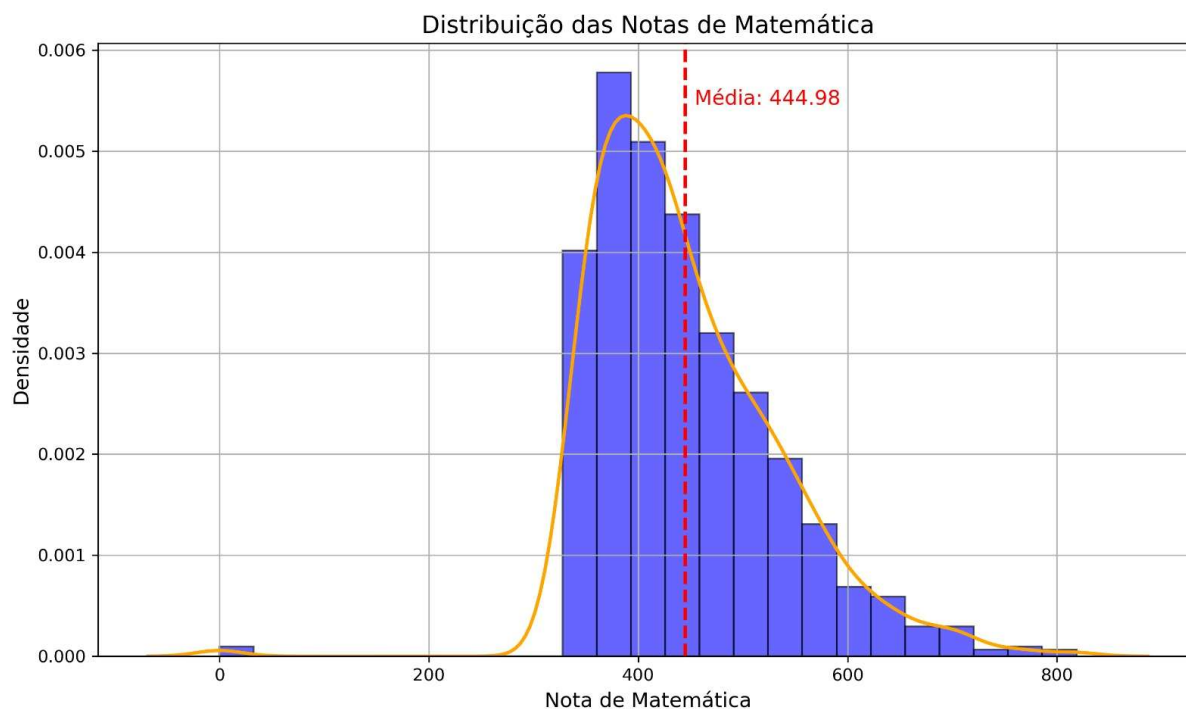
As análises de correlação e o tratamento de dados foram realizados utilizando Python (versão 3.9.13) e a biblioteca Scikit-learn (Python Software Foundation, 2024). O treinamento do modelo de regressão múltipla foi conduzido no software R, utilizando o método stepwise (R Core Team, 2018).

4 RESULTADOS

Para examinar como as notas de Matemática dos candidatos se distribuem, foi criada uma visualização em formato de histograma, destacando a densidade das pontuações. O objetivo dessa técnica foi mostrar a distribuição geral das notas. No Gráfico 01, percebe-se que

a maioria dos participantes do ENEM 2023, da cidade de Uruçuí – PI, obteve resultados próximos à média na prova de Matemática. Além disso, nota-se que poucos candidatos apresentaram notas muito distantes dessa média, tanto para baixo quanto para cima.

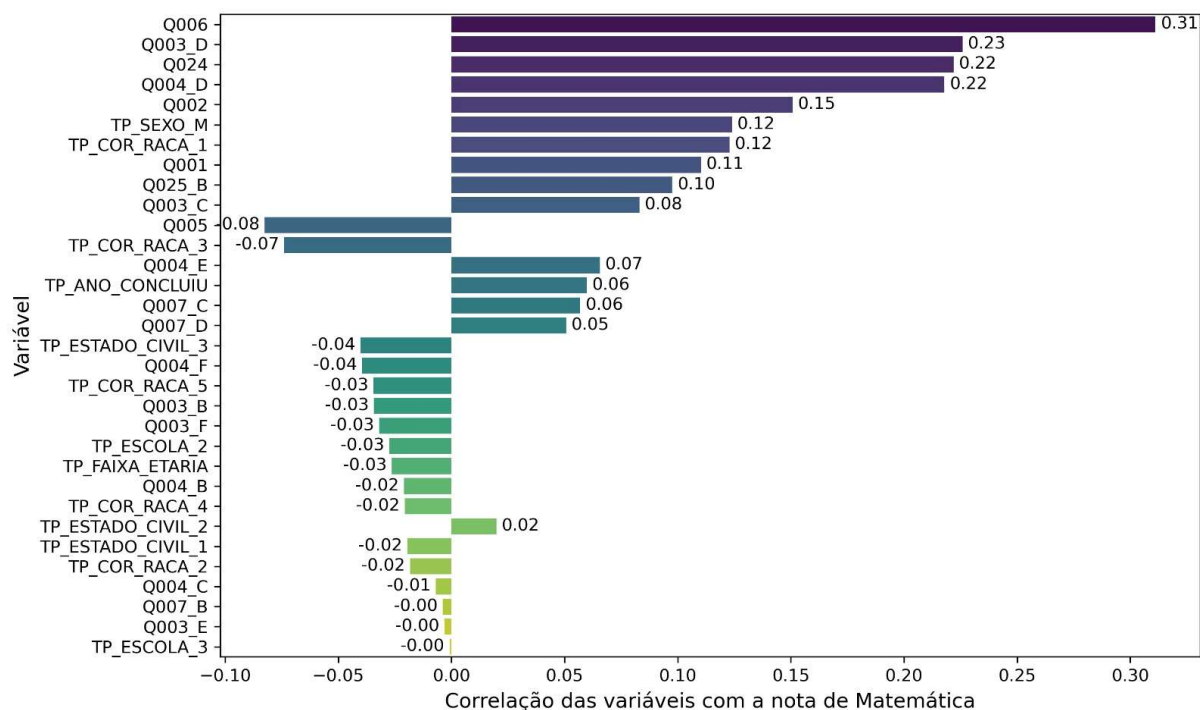
Gráfico 1 - Distribuição das Notas de Matemática



Fonte: Elaborada pelos autores, 2024.

O Gráfico 2 ilustra as correlações entre todas as variáveis explicativas escolhidas previamente e a variável dependente (NU_NOTA_MT). Percebe-se que a maioria das variáveis explicativas tem correlações positivas com a variável dependente.

Gráfico 2 - Correlação Entre a Nota de Matemática e as Variáveis Explicativas dos Participantes do ENEM da cidade de Uruçuí - PI



Fonte: Elaborada pelos autores, 2024.

O Gráfico 2 revela uma série de correlações entre a nota de Matemática e as variáveis explicativas analisadas. Observam-se correlações positivas e significativas com variáveis como Q001, Q002, Q003_C, Q003_D, Q004_D, Q006, Q0024, Q0025_B, TP_COR_RAÇA_1 e TP_SEXO_M, indicando uma relação favorável entre esses fatores e o desempenho em Matemática. Além disso, variáveis como Q004E, TP_ANO_CONCLUIU, Q007_C, Q007_D e TP_ESTADO_CIVIL_2 também mostram correlações positivas, embora menos intensas. Por outro lado, variáveis como TP_COR_RAÇA_3, Q005, TP_ESTADO_CIVIL_3, Q004_F, TP_COR_RAÇA_5, Q003_B, Q003_F, TP_ESCOLA_2, TP_FAIXA_ETARIA e Q004_B apresentam correlações negativas, sugerindo que esses fatores podem ter uma relação inversa com o desempenho dos alunos em Matemática. No geral, apenas um pouco menos da metade das variáveis analisadas mostrou correlação com a nota, destacando a complexidade dos fatores que influenciam o rendimento nessa disciplina.

Embora a correlação ofereça uma visão sobre a associação entre as variáveis, ela não implica causalidade. Por isso, foi realizada também uma análise de regressão. Todas as variáveis explicativas selecionadas anteriormente, conforme visto no Gráfico 1, foram inicialmente incluídas em um modelo de regressão linear múltipla (OLS). Após identificar que

algumas das variáveis não era estatisticamente significativa, utilizou-se a técnica de ajuste "stepwise", que realiza a seleção automática das variáveis preditoras. Como resultado, o número de variáveis foi reduzido de 34 para 9. Os resultados do modelo final estão apresentados na Tabela 1, onde todos os parâmetros indicam significância estatística, conforme demonstrado.

Tabela 01 - Modelo de Regressão Linear Múltipla

Predictors	Estimates	IC	p
(Intercept)	423.17	407.70 – 438.63	<0.001
Q006	8.34	5.15 – 11.54	<0.001
Q003 D	33.06	12.79 – 53.33	0.001
TP COR RAÇA 1	17.84	4.33 – 31.35	0.010
Q003 C	19.09	4.94 – 33.24	0.008
Q005	-5.00	-8.98 – -1.01	0.014
Q024	10.75	-0.29 – 21.79	0.056
TP SEXO M	13.72	2.36 – 25.07	0.018
Q003 E	-50.05	-104.44 – 4.34	0.071
Q004 D	12.29	-2.72 – 27.31	0.108

Observations: 935

R² / R² ajustado: 0.31

Fonte: Elaborado pelos autores, 2024.

A análise de regressão revelou que algumas variáveis, como Q006, Q003 D, TP COR RAÇA 1, Q003 C e TP SEXO M, exercem um impacto positivo e relevante, sugerindo que essas características estão associadas ao aumento nas notas de matemática. Em contrapartida, a variável Q005 apresentou um efeito negativo significativo, apontando para uma redução nas notas. Já variáveis como Q024, Q003 E e Q004 D não demonstraram significância estatística, mas podem ainda assim exercer alguma influência.

Esses resultados corroboram o estudo de Dutra, Firmino Júnior e Fernandes (2023), que identificou que 89,5% dos trabalhos analisados indicaram os fatores socioeconômicos como os principais influenciadores do desempenho acadêmico. Os elementos mais relevantes incluíram renda familiar, idade, sexo, raça, escolaridade dos pais, número de pessoas na residência e estado civil, todos com impacto considerável nas notas dos estudantes.

Da mesma forma, os achados estão em consonância com a pesquisa de De Sousa Fernandes, Pereira e Moura (2023), que constatou que alunos cujos pais possuem ensino

superior ou pós-graduação obtêm notas mais altas em comparação com aqueles cujos pais têm menor nível de escolaridade. Além disso, a média da renda familiar e o acesso à internet em casa foram identificados como fatores que contribuem positivamente para o desempenho acadêmico. Esses achados reforçam a importância das condições socioeconômicas no sucesso dos estudantes, especialmente em matemática.

De modo geral, os resultados sugerem que tanto características individuais, como sexo, raça e estado civil, quanto fatores contextuais, como o tipo de escola e o ano de conclusão do ensino médio, têm um papel relevante na nota de matemática. O coeficiente de determinação (R^2) foi de 0,32, indicando que o modelo explica cerca de 32% da variação nas notas.

5 CONSIDERAÇÕES FINAIS

Este estudo analisou o desempenho dos candidatos de Uruçuí - PI no Exame Nacional do Ensino Médio (ENEM), com foco na área de Matemática e suas Tecnologias. A pesquisa teve como objetivo identificar as variáveis que afetam esse desempenho. Para isso, foi empregado um modelo de regressão linear múltipla, que indicou que a maioria das variáveis analisadas exerceu influência significativa nas notas, revelando tanto efeitos positivos quanto negativos.

Os resultados apontaram que as notas dos participantes foram distribuídas de forma relativamente uniforme, com a maior parte dos candidatos obtendo pontuações próximas à média geral. As correlações encontradas sublinharam a importância de fatores socioeconômicos e demográficos, sugerindo que o contexto social tem um papel relevante no desempenho acadêmico.

As principais contribuições deste trabalho estão na identificação dos fatores socioeconômicos que influenciam as notas no ENEM, oferecendo informações úteis para educadores e tomadores de decisão. Contudo, o estudo apresentou limitações, como a ausência de variáveis que poderiam aprofundar a análise e a limitação geográfica ao município de Uruçuí.

Para pesquisas futuras, seria importante considerar outros contextos regionais e incorporar uma variedade maior de variáveis. Além disso, investigar possíveis relações causais entre as variáveis poderão proporcionar uma compreensão mais detalhada dos fatores que influenciam o desempenho no ENEM.

REFERÊNCIAS

ALVES, José Ronaldo. **Mapeamento das questões de Estatística no ENEM**. IX Encontro Paraibano de Educação Matemática. Anais do IX EPEM–Encontro Paraibano de Educação Matemática, 2016.

AREVALO-LÓPEZ, Ivan et al. **A Memory-Efficient Encoding Method for Processing Mixed-Type Data on Machine Learning**. *Entropy*. v.22, n.1391, 2020. Disponível em: <<https://www.mdpi.com/1099-4300/22/12/1391>>. Acesso em: 17 set. 2024.

BRASIL. Ministério da Educação. **Parâmetros de atualização do Exame Nacional do Ensino Médio-ENEM**. Brasília/DF, 2021. Disponível em: <https://www.gov.br/mec/pt-br/novo-ensino-medio-descontinuado/pdfs/Parametrosnovo_enem2022_compressed1.pdf>. Acesso em: 10 de set de 2024.

CAMPELLO, R. et al. Agrupamento baseado em densidade. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge. Discovery*. v. 10, 2019. Disponível em: <<https://doi.org/10.1002/widm.1343>>. Acesso em: 16 set. 2024.

Core Team R., 2018 **A language and environment for statistical computing**. R Foundation for Statistical Computing.

DUTRA, J. F.; FIRMINO JÚNIOR, J. B.; FERNANDES, D. Y. de S. **Fatores que podem interferir no desempenho de estudantes no ENEM: uma revisão sistemática da literatura**. *Revista Brasileira de Informática na Educação*. v. 31, p. 323–351, 2023. Disponível em: <<https://doi.org/10.5753/rbie.2023.3087>> Acesso em: 21 set. 2024

ENEM. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira/Inep. **Microdados do Enem 2023**. Disponível em: <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enem>>. Acesso em: 18 set. 2024.

FERNANDES, Luanda Carol de Sousa; PEREIRA, Lourivaldo Barreto; MOURA, Ernandes Guedes. **Variáveis que influenciaram o desempenho dos candidatos no ENEM em Matemática e suas tecnologias: um estudo com candidatos que realizaram a prova em Barra do Corda Maranhão em 2021**. *Revista Ibero-Americana de Humanidades, Ciências e Educação*. v. 9, n. 9, 2023. Disponível em: <<https://doi.org/10.51891/rease.v9i9.11454>>. Acesso em: 18 set. 2024.

GOULART, Amari et al. **Letramento Estatístico e o Exame Nacional de Ensino Médio**.

In: CONFERÊNCIA INTERAMERICANA DE EDUCAÇÃO MATEMÁTICA. 2015.

HASSANI, H.; SILVA, E. S. A. A Kolmogorov-Smirnov Based Test for Comparing the Predictive Accuracy of Two Sets of Forecasts. *Econometrics*. v. 3, n. 3, 2015. Disponível em: <<https://doi.org/10.3390/econometrics3030590>>. Acesso em: 16 set. 2024.

HOLLAS, Justiani; BERNARDI, Lucí; DOS SANTOS, T. M. **O Exame Nacional do Ensino Médio (Enem) e as competências para uma educação estatística crítica**. *Ensaio: avaliação e políticas públicas em educação*, v. 28, p. 110-134, 2019.

JALOTO, Alexandre; PRIMI, Ricardo. Fatores socioeconômicos associados ao desempenho no Enem. **Em Aberto**. v.34, n.112, 2021. Disponível em: <<https://doi.org/10.24109/2176-6673.emaberto.34i112.5002>>. Acesso em: 18 set. 2024.

LIMA, José Wenes Pereira et al. **Avaliação da nota de redação do Enem no estado do Ceará via modelo de regressão beta inflacionado**. João Pessoa: Universidade Federal da Paraíba/PB

2023. Disponível em: <https://repositorio.ufpb.br/jspui/bitstream/123456789/29828/1/Jos%C3%A9WenesPereiraLima_Dissert.pdf>. Acesso em: 17 set. 2024.

LUCENA, João Paulo Oliveira; SANTOS, Heric Nero Lisboa dos Santos. **A relação entre desempenho no Exame Nacional do Ensino Médio e o perfil socioeconômico: um estudo com os microdados de 2016**. Revista de Gestão e Secretariado. v.11. n.2, 2020. Disponível em: <<https://doi.org/10.7769/gesec.v11i2.994>>. Acesso em: 18 set. 2024.

MELO, Rafael Oliveira (2022). **Impacto das variáveis socioeconômicas no desempenho do Enem: Uma análise espacial e sociológica**. Revista de Administração Pública.v.55, n.6, 2021. Disponível:< <https://doi.org/10.1590/0034-761220200843>>. Acesso em: 18 set. 2024.

MIRANDA, Paula Roberta Miranda; AZEVEDO, Mário Luiz Neves de Azevedo. **Fies e Prouni na expansão da educação superior brasileira: políticas de democratização do acesso e/ou de promoção do setor privado-mercantil? Educação & Formação**. v.5, n.3, 2020. Disponível em: <<https://doi.org/10.25053/redufor.v5i15set/dez.1421>>. Acesso em: 16 set. 2024.

MISHRA, Srikanta.; DATTA-GUPTA. **Chapter 4 - Regression Modeling and Analysis**. 2018 Disponível em:<<https://doi.org/10.1016/B978-0-12-803279-4.00004-3>>. Acesso em: 16 set. 2024

NETO, Pedro Luiz de Oliveira Costa. **Estatística**. Editora Blucher, 2002.
Python Software Foundation. Python Language Reference, version 3.9.13. Available at: <https://www.python.org>. Acesso em: 28 agosto. 2024. .

PEDREGOSA, Fabian et al. **Scikit-learn: Machine learning in python journal of machine learning research**. Journal of machine learning research, v. 12, p. 2825-2830, 2011.

PORTO, Klark Gable Souza. Descoberta de conhecimento através da análise e mineração em dados do Enem. Brasília/DF: Universidade Federal de Brasília, 2019. Disponível em: <https://bdm.unb.br/bitstream/10483/28902/1/2019_KlarkGableSouzaPorto_tcc.pdf>. Acesso em: 17 set. 2024.

RENCHE, Alvin C.; SCHAALJE, G. Bruce. **Linear models in statistics**. 2. ed. John Wiley & Sons, 2008. Disponível em: <<https://www.utstat.toronto.edu/~brunner/books/LinearModelsInStatistics.pdf>>. Acesso em: Acesso em: 16 set. 2024.

RÊGO, Fabrício Emílio Dantas do. **Fatores que influenciam na nota da redação do ENEM no Rio Grande do Norte**. Natal/RN: Universidade Federal do Rio Grande do Norte, 2021. Disponível em: <https://repositorio.ufrn.br/bitstream/123456789/34324/1/FatoresQueInfluenciam_R%C3%A9go_2021.pdf>. Acesso em: 17 set. 2024.

SANTOS, Carla. Estatística descritiva. **Manual de auto-aprendizagem**, v. 2, 2007.
BUSSAB, Wilton de O.; MORETTIN, Pedro A. Estatística básica. In: **Estatística básica**. 2010. p. xvi, 540-xvi, 540.

SANTOS, Manoela Cassia dos Santos. **Desenvolvimento de modelos para estimar principais características socioeconômicas para as notas de redação e matemática para ENEM 2018**. Maceió/AL: Universidade Federal de Alagoas, 2020. Disponível em: <https://repositorio.ufpb.br/jspui/bitstream/123456789/29828/1/Jos%C3%A9WenesPereiraLima_Dissert.pdf>. Acesso em: 17 set. 2024.

SCHOBER, P.; BOER, C.; SCHWARTE, L. **Coefficientes de Correlação: Uso e Interpretação Apropriados. Anestesia e Analgesia**. v.126, 2018. Disponível em: <<https://doi.org/10.1213/ANE.0000000000002864>> Acesso em: 16 set. 2024.

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, v. 37, p. 1101, 2015.

SOUZA, Thiago Oliveira de. **Análise de dados: um estudo do perfil dos participantes do ENEM 2019.** 2021.

TEIXEIRA, Mariana Teixeira, CAVIQUE, Luís. **Feature engineering: techniques and applications.** Revista de Ciências da Computação, n.18, 2023. Disponível em: <https://www.oasisbr.ibict.br/vufind/Record/RCAP_43a8ec602f071f5561fd7164e1f7da3e>. Acesso em: 17 set. 2024

ANEXO

Anexo A - Descrição das variáveis utilizadas

NOME DA VARIÁVEL	DESCRIÇÃO
TP_FAIXA_ETARIA	Faixa etária
TP_SEXO M	Sexo masculino
TP_ESTADO_CIVIL 1	Estado Civil solteira
TP_ESTADO_CIVIL 2	Estado Civil Casado(a)/Mora com companheiro(a)
TP_ESTADO_CIVIL 4	Estado Civil Viúvo(a)
TP_COR_RACA 1	Raça/cor Branca
TP_COR_RACA 2	Raça/cor Preta
TP_COR_RACA 3	Raça/cor Parda
TP_COR_RACA 4	Raça/cor Amarela
TP_COR_RACA 5	Raça/cor indígena
TP_ANO_CONCLUIU	Ano de Conclusão do Ensino Médio
TP_ESCOLA 2	Pública
TP_ESCOLA 3	Privada.
Q001	Escolaridade do pai, ou homem responsável
Q002	Escolaridade da mãe, ou a mulher responsável
Q003 B ocupação da sua mãe ou da mulher responsável por você	Grupo 2: Diarista, empregado doméstico, cuidador de idosos, babá, cozinheiro (em casas particulares), motorista particular, jardineiro, faxineiro de empresas e prédios, vigilante, porteiro, carteiro, office-boy, vendedor, caixa, atendente de loja, auxiliar administrativo, recepcionista, servente de pedreiro, repositor de mercadoria.;
Q003 C	Grupo 3: Padeiro, cozinheiro industrial ou em restaurantes, sapateiro, costureiro, joalheiro, torneiro mecânico, operador de máquinas, soldador, operário de fábrica, trabalhador da mineração, pedreiro, pintor, eletricista, encanador, motorista, caminhoneiro, taxista.
Q003 D	Grupo 4: Professor (de ensino fundamental ou médio, idioma, música, artes etc.), técnico (de enfermagem, contabilidade, eletrônica etc.), policial, militar de baixa patente (soldado, cabo, sargento), corretor de imóveis, supervisor, gerente, mestre de obras, pastor, microempresário (proprietário de empresa com menos de 10 empregados), pequeno comerciante, pequeno proprietário de terras, trabalhador autônomo ou por conta própria.;
Q003 E	Grupo 5: Médico, engenheiro, dentista, psicólogo, economista, advogado, juiz, promotor, defensor, delegado, tenente, capitão, coronel, professor universitário, diretor em empresas públicas ou privadas, político, proprietário de empresas com mais de 10 empregados

Q003 F	Não sei.
Q004 B ocupação da sua mãe ou da mulher responsável por você.	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, auxiliar administrativa, recepcionista, servente de pedreiro, repositora de mercadoria.;
Q004 C	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, caminhoneira, taxista
Q004 D	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, pastora, microempresária (proprietária de empresa com menos de 10 empregados), pequena comerciante, pequena proprietária de terras, trabalhadora autônoma ou por conta própria.
Q004 E	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 empregados.
Q004 F	Não sei
Q005	Quantidade de pessoas que moram na residência
Q006	Renda mensal da família
Q007 B Em sua residência trabalha empregado(a) doméstico(a)?	Sim, um ou dois dias por semana.
Q007 C	Sim, três ou quatro dias por semana
Q007 D	Sim, pelo menos cinco dias por semana.
Q024	Uso de computador em casa
Q025	Acesso de internet em casa

Fonte: Elaborado pela autora (2024) com base no dicionário de dados disponibilizado pelo Enem. (2023)