

DESEMPENHO NO ENEM E CONDIÇÕES SOCIOECONÔMICAS: UMA ANÁLISE VIA REGRESSÃO LINEAR MÚLTIPLA

Felipe Yamamoto Tenedine (PIBIC/CNPq/FA/UEM), Vanderly Janeiro (Orientador).
E-mail: ra131790@uem.br.

Universidade Estadual de Maringá, Centro de Ciências Exatas, Maringá, PR.

Probabilidade e Estatística / Regressão e Correlação.

Palavras-chave: estatística, regressão, enem.

RESUMO

Este estudo analisa os fatores determinantes para o desempenho dos estudantes de Maringá-PR no Exame Nacional do Ensino Médio (ENEM) de 2023. O objetivo é identificar quais variáveis socioeconômicas, presentes no questionário do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), melhor explicam a nota média das quatro áreas de conhecimento. A análise baseou-se em uma amostra de 1.765 estudantes. A metodologia incluiu a construção de um modelo de regressão linear múltipla, com seleção de variáveis pelo método *stepwise* baseado no critério de informação de Akaike (AIC). A validação do modelo envolveu a verificação de multicolinearidade, por meio do Fator de Inflação de Variância (VIF), e a análise de resíduos com gráficos *half-normal* e envelopes simulados. Os resultados demonstram que maior renda familiar, frequência em escola privada, acesso à internet e posse de computadores no domicílio estão positivamente associados a um melhor desempenho. Em contrapartida, a defasagem idade-série e a autodeclaração como preto ou pardo apresentaram uma associação negativa com as notas. Adicionalmente, o estudo revelou achados contraintuitivos, como a associação negativa entre o desempenho e o número de quartos na residência, bem como a ausência de significância estatística da escolaridade parental no modelo final. Conclui-se que as desigualdades socioeconômicas e o acesso a recursos tecnológicos são fatores cruciais que explicam uma parcela significativa do desempenho dos estudantes em Maringá, reforçando a necessidade de políticas públicas estruturais para mitigar essas disparidades e promover a equidade educacional.

INTRODUÇÃO

O Exame Nacional do Ensino Médio (ENEM) é o principal mecanismo de acesso ao ensino superior no Brasil e uma ferramenta de avaliação do conhecimento adquirido pelos estudantes na educação básica. A literatura acadêmica tem consistentemente demonstrado que o desempenho dos estudantes está profundamente associado às condições socioeconômicas, familiares e escolares em que estão inseridos.

Estudos empíricos no Brasil mostram que os recursos familiares, incluindo o nível de instrução dos pais, são determinantes para o desempenho cognitivo dos estudantes, o que se reflete nos resultados de avaliações em larga escala como o ENEM (Soares e Collares, 2006).

Este trabalho surgiu da curiosidade em entender os determinantes do desempenho dos estudantes de Maringá no ENEM 2023. O objetivo é realizar análises descritivas e construir um modelo estatístico de regressão múltipla para identificar quais fatores socioeconômicos, dentre os coletados pelo questionário do INEP, melhor explicam o desempenho geral dos estudantes de Maringá-PR no ENEM 2023.

MATERIAIS E MÉTODOS

A pesquisa utilizou uma amostra de 1.765 observações dos Microdados do ENEM 2023, restrita a estudantes que concluíram o Ensino Médio e realizaram a prova em Maringá em 2023, com presença confirmada em todos os dias e nota superior a 100 em todas as áreas (Ciências Humanas, Ciências da Natureza, Linguagens e Códigos e Matemática). Essa nota mínima foi utilizada para excluir provas em branco, com pouquíssimas respostas ou com comportamento inconsistente nas respostas, visto que o ENEM utiliza a teoria de resposta ao item, que penaliza respostas inconsistentes, como acertar questões difíceis e errar as fáceis (INEP, 2021).

A variável de interesse foi a média aritmética das notas objetivas. A análise estatística foi conduzida no ambiente estatístico R em duas etapas: primeiro, calculou-se o V de Cramér para avaliar a associação entre as variáveis socioeconômicas e o Eta-quadrado (η^2), para calcular a variância nas notas explicada pelo fator socioeconômico. Em seguida, ajustou-se um modelo de regressão linear múltipla, selecionando um conjunto parcimonioso de preditores com o procedimento *stepwise* para minimizar o Critério de Informação de Akaike (AIC). A multicolinearidade foi avaliada pelo Fator de Inflação da Variância (VIF), e a qualidade do ajuste do modelo final foi diagnosticada visualmente com gráficos *half-normal* com envelopes de simulação.

RESULTADOS E DISCUSSÃO

A análise exploratória dos dados revelou que, embora as notas das quatro áreas do conhecimento apresentem distribuições variadas, com Matemática exibindo a maior variabilidade, a média geral das notas se aproxima de uma distribuição normal, validando sua utilização como variável resposta no modelo de regressão. A análise de associação (V de Cramér) mostrou correlação entre alguns indicadores socioeconômicos, especialmente entre a renda familiar e o número de banheiros. A análise do Eta-quadrado (η^2) identificou a escolaridade dos pais, a renda familiar e a quantidade de computadores como os fatores que mais explicam as notas.

O modelo final de regressão linear múltipla foi estatisticamente significativo ($p < 0,001$) e explicou 31,2% da variação nas notas dos estudantes (R^2 ajustado = 0,3124), um resultado considerado relevante para estudos em ciências sociais (Ozili,

2022). A ausência de multicolinearidade foi confirmada, com todos os preditores apresentando VIF inferior a 5, o que garante a confiabilidade dos coeficientes (Montgomery, Peck e Vining, 2021).

A interpretação dos resultados significativos apontou para desigualdades educacionais claras. Em relação à raça/etnia, em comparação com estudantes brancos, os autodeclarados pretos e pardos tiveram notas, em média, 17,53 e 10,16 pontos mais baixas, respectivamente. Em contrapartida, estudantes amarelos tiveram um desempenho 20,09 pontos superior. Quanto à trajetória escolar, a defasagem idade-série mostrou um impacto negativo, reduzindo a nota em 29,04 pontos. Além disso, cursar o ensino médio em escola particular esteve associado a um aumento de 30,12 pontos. Sobre os fatores socioeconômicos, a renda familiar foi um dos preditores mais fortes, com seu efeito positivo aumentando gradativamente até um ganho de 99,04 pontos na faixa mais alta. O acesso à tecnologia também se mostrou crucial, com a posse de 3 ou mais computadores e o acesso à internet associados a ganhos de 27,44 e 32,24 pontos, respectivamente.

De forma inesperada, a escolaridade dos pais não se mostrou significativa no modelo final, possivelmente porque seus efeitos foram capturados por variáveis mais diretas como renda e tipo de escola. Outro achado contraintuitivo foi a associação negativa do número de quartos com o desempenho, levantando a hipótese de que essa variável possa estar atuando como um indicador de alta densidade domiciliar, prejudicial aos estudos.

Finalmente, a análise de resíduos por meio de gráficos *half-normal* com envelopes simulados confirmou o bom ajuste do modelo, com quase 95% dos pontos dentro dos limites esperados, indicando que o modelo descreve adequadamente a variabilidade dos dados, sem violações severas de suas premissas (Moral, Hinde e Demétrio, 2017).

CONCLUSÕES

Os resultados deste estudo evidenciam que o desempenho dos estudantes de Maringá no ENEM 2023 é explicado por uma combinação complexa de fatores socioeconômicos. Mais do que a renda familiar, aspectos como o tipo de escola frequentada e a cor/raça declarada pelo estudante se mostram componentes centrais das desigualdades observadas. Além disso, o acesso a recursos tecnológicos, como computadores e internet, também mostrou impacto relevante nas notas.

Essas conclusões estão em consonância com achados de pesquisas anteriores que associam o desempenho no ENEM a desigualdades socioeconômicas persistentes (Melo *et al.*, 2021). Nesse contexto, os resultados reforçam a necessidade de políticas públicas que ultrapassem ações compensatórias baseadas apenas em renda. É fundamental promover intervenções estruturais voltadas à valorização e qualificação da rede pública de ensino, bem como ao enfrentamento das disparidades raciais e do acesso desigual à infraestrutura educacional.

AGRADECIMENTOS

Agradeço ao meu orientador, Vanderly Janeiro, pela orientação, apoio, paciência e contribuições ao longo do desenvolvimento deste trabalho. Seu apoio e incentivo foram fundamentais para a realização desta pesquisa. Expresso também minha gratidão ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro, que possibilitou a execução deste estudo.

REFERÊNCIAS

INEP. **Entenda a sua nota no Enem: Guia do participante**. Brasília, DF: Ministério da Educação, 2021.

MELO, R. O. *et al.* Impacto das variáveis socioeconômicas no desempenho do Enem: uma análise espacial e sociológica. **Revista de Administração Pública**, v. 55, n. 6, p. 1271–1294, nov. 2021.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to Linear Regression Analysis**. 6th. ed. Hoboken, NJ: John Wiley & Sons, 2021.

MORAL, R. A.; HINDE, J.; DEMÉTRIO, C. G. B. Half-Normal Plots and Overdispersed Models in R: The hnp Package. **Journal of Statistical Software**, v. 81, n. 10, p. 1–23, 2017.

OZILI, P. K. The Acceptable R-Square in Empirical Modelling for Social Science Research. **SSRN Electronic Journal**, jun. 2022..

SOARES, J. F.; COLLARES, A. C. M. Recursos familiares e o desempenho cognitivo dos alunos do ensino básico. **Dados**, v. 49, n. 3, p. 615–650, 2006.