



DO INGRESSO À PERMANÊNCIA: como a gestão baseada em evidências pode orientar a atuação pedagógica preventiva a partir da IA generativa

Luciana Machado-Costa; luciana.costa@iff.edu.br; Instituto Federal Fluminense

Carmen Sylvia Vidigal Moraes; moraescs@usp.br; Universidade de São Paulo

Modalidade: Artigo

Área Temática: Inteligência Artificial (IA) e o Novo Ecossistema de Aprendizagem

Resumo:

Contexto: A Inteligência Artificial (IA) generativa pode democratizar a análise de dados no ambiente educacional e auxiliar na indução da permanência de estudantes cotistas com equidade. O objetivo do estudo foi analisar a eficácia da ação afirmativa de cotas nos primeiros anos da Lei 12.711/2012 no acesso e na permanência. Para tanto, a IA atuou como assistente na modelagem estatística de trajetórias acadêmicas dos estudantes em cursos do ensino médio integrado (EMI), viabilizando um modelo de gestão baseado em evidências, acessível a profissionais sem formação em ciência de dados. **Métodos:** Analisou-se uma série histórica de oito anos (2013-2020) de notas de ingresso em processos seletivos para o EMI. A IA auxiliou na elaboração dos *scripts* de programação (linguagem R) para a análise quantitativa a partir de estatística descritiva e inferencial. **Resultados:** A distribuição das notas de ingresso mostrou que, em oito anos, nenhum cotista obteve nota no quartil mais alto, enquanto 45% dos não cotistas estavam nessa posição, indicando turmas extremamente heterogêneas. A regressão logística multinomial mostrou que estudantes com baixo desempenho têm quase seis vezes mais chances de evasão, e estudantes cotistas apresentam 44% mais chances de retenção. **Conclusão:** O mapeamento prévio com o uso de IA otimiza a alocação institucional de recursos (como monitorias e equipe multidisciplinar) antes do início do período letivo. Simultaneamente, favorece a autonomia docente, pois o professor, ao receber o mapa com o perfil de sua turma, pode adaptar o planejamento pedagógico desde a primeira semana de aula.

Palavras-chave:

Inteligência Artificial Generativa; Gestão Educacional; Permanência Escolar; Ensino Médio Integrado; Ações Afirmativas.

Introdução

Este artigo integra uma pesquisa de doutorado desenvolvida no Programa de Pós-Graduação em Educação da Universidade de São Paulo (USP), que analisa os primeiros dez anos de implementação da Lei 12.711/2012, conhecida como Lei de Cotas, para acesso aos cursos do ensino médio integrado ao técnico (EMI) em dois *campi* do Instituto Federal Fluminense (IFFluminense).



A educação brasileira, historicamente marcada pela dualidade, reservou às elites a formação teórica e às classes trabalhadoras a formação prática e manual, orientada à reprodução de uma força de trabalho subalternizada (Ciavatta, 2009; Frigotto; Ciavatta; Ramos, 2012). Em disputa com este modelo, encontra-se a concepção de escola unitária e politécnica, inspirada em Antonio Gramsci. Esta perspectiva busca a formação humana integral, a superação das desigualdades sociais e a preparação dos jovens para a autonomia intelectual, a inserção social e a liderança moral e política na disputa por uma nova concepção de mundo e uma nova hegemonia (Moraes, 2018, 2006; Moraes; Alencar; Jacomini, 2024).

O EMI representa uma possibilidade concreta desta transformação ao propor uma formação humana integral, tendo o trabalho como princípio educativo. A ação afirmativa de cotas, quando articulada ao EMI, potencializa as possibilidades desta modalidade no enfrentamento das desigualdades estruturais da sociedade brasileira. Ao oferecer uma formação com “dupla saída”, voltada tanto ao mundo do trabalho complexo quanto ao acesso ao ensino superior, a jovens de segmentos sociais que enfrentam barreiras históricas, essas políticas ampliam as chances de justiça educacional e social. Esta possibilidade, entretanto, está condicionada à eficácia da Lei de Cotas no acesso, na permanência e na conclusão do curso.

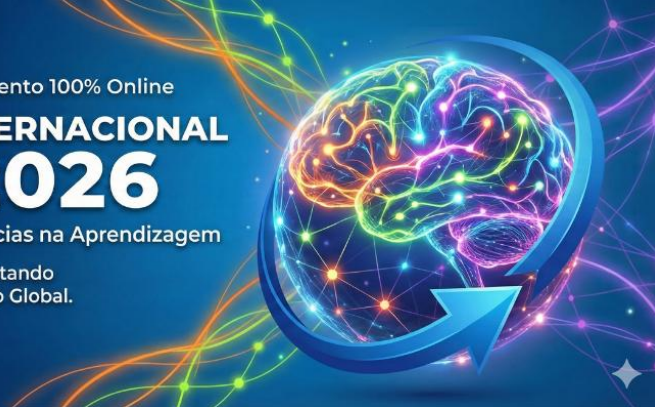
Em âmbito nacional, não há dados sistematizados sobre a aplicação da Lei de Cotas na educação profissional. A Plataforma Nilo Peçanha, fonte oficial para a disseminação de estatísticas da Rede Federal de Educação Profissional, Científica e Tecnológica (RFEPCT), disponibiliza apenas informações sobre o percentual de vagas reservadas, a partir de 2019, e a situação da matrícula por forma de ingresso, a partir de 2024.

Publicações de referência sobre os Institutos Federais, as ações afirmativas e a condição da população negra no Brasil não abordam as cotas no EMI (Campos; Lima, 2025; Carreira; Heringer, 2022; Carvalhaes; Júnior; Daflon, 2013; DPU; ABPN, 2022; Feres Júnior *et al.*, 2018; Ferreira; Sukanuma, 2018; Frigotto, 2018; Frigotto *et al.*, 2018; Gilioli, 2023; Pacheco; Silva, 2007; Paiva, 2013; Silva, Iraneide Soares da *et al.*, 2025; Silva, 2020; Zoninsein; Feres Júnior, 2008).

A revisão de literatura em bases de artigos, dissertações e teses, a partir de 2013, ano de ingresso dos primeiros estudantes pela Lei de Cotas, identificou apenas dezesseis trabalhos com indicadores de retenção e evasão comparada entre cotistas e não cotistas do EMI da RFEPCT. Dentre estes, apenas um analisou os dez primeiros anos de implementação da Lei de Cotas, investigando três *campi* e onze cursos e um Instituto Federal, com uma população de 2.533 estudantes (Previtali, 2023).

Esta pesquisa, por sua vez, analisou dois *campi* e sete cursos do EMI do IFFluminense, abrangendo uma população de 3.443 estudantes matriculados de 2013 a 2020, dos quais 1.527 ingressaram pelas cotas. Devido à pandemia da Covid-19, não foram aplicadas provas para seleção de estudantes nos anos de 2021 e 2022, quando o ingresso foi realizado por sorteio, limitando a amostra a oito anos.

Considerando o critério universal das cotas para o EMI, ter cursado todos os nove anos do ensino fundamental integralmente em escolas de redes públicas, a questão investigada é se, a partir da trajetória escolar no ensino fundamental, que se considera consubstanciada na nota da prova de ingresso, é possível identificar uma associação entre a trajetória de origem e a trajetória acadêmica de destino no EMI.



A hipótese é de que há continuidade, o que indicaria ineficácia parcial da política de cotas. Esta ineficácia seria produzida pela omissão ou por insuficiência de políticas de permanência, sejam elas materiais, acadêmicas ou simbólicas, na instituição de destino.

Metodologicamente, trata-se de um estudo de caso, com pesquisa exploratória, que adota abordagem quantitativa, comparando estudantes cotistas, como grupo de interesse, com estudantes não cotistas, com o objetivo de identificar desigualdades, avanços e limitações da política de cotas. Para realizar este estudo, não tendo as autorias formação em ciência de dados, utilizou-se a inteligência artificial generativa (IAG) das plataformas ChatGPT e Google AI Studio como assistente na modelagem estatística, na elaboração dos *scripts* e na análise dos dados.

A IAG democratizou o acesso a ferramentas que permitem que pesquisadores e gestores sem formação em ciência de dados analisem dados acadêmicos e construam mapeamentos de perfis estudantis de forma rápida, aplicável e segura. Por ser um modelo baseado no processamento de linguagem natural (NLP) e que gera conteúdo a partir de um *prompt* ou solicitação do usuário, ela permite um diálogo natural entre o pesquisador e a máquina. Ao ser utilizada como auxiliar para a elaboração de *scripts* de modelagem estatística, sem acesso direto à base de dados, preserva-se a privacidade e o sigilo demandados pela Lei Geral de Proteção de Dados Pessoais (LGPD), a Lei 13.709/2018.

No campo pedagógico, a IAG apresenta potencial ambivalente. Por um lado, ela pode ser usada para produzir um esvaziamento da mediação pedagógica, reforçar vieses e desigualdades, violar a privacidade de dados dos alunos, inclusive crianças e adolescentes, precarizar a atividade docente e desumanizar o ensino; por outro, ao personalizar o ensino, ela pode permitir tanto o estímulo aos estudantes com mais facilidade de aprendizagem quanto auxiliar no ensino, no ritmo e na forma adequada, dos estudantes com defasagens de conhecimentos de base ou necessidades educacionais específicas (Azambuja; Silva, 2024; Costa *et al.*, 2025; Silva, Davi Souza da *et al.*, 2025; UNESCO, 2024).

Embora a literatura recente evidencie a expansão da IAG na personalização do ensino e na automação de conteúdos (UNESCO, 2024), observa-se uma lacuna na aplicação destas tecnologias para a gestão educacional estratégica. O mapeamento realizado por Costa *et al.* (2025) aponta que a quase totalidade das aplicações de IA concentra-se na sala de aula ou na tutoria inteligente, negligenciando o potencial da ferramenta para o suporte à decisão gerencial. A exceção notável, o trabalho de Bitencourt *et al.* (2022), utiliza modelagem preditiva para identificar riscos de evasão. Contudo, modelos preditivos tradicionais operam frequentemente como “caixas-pretas” para profissionais não especializados, exigindo letramento avançado em ciência de dados e infraestrutura técnica que escapam à realidade da maioria das escolas públicas, o que limita sua replicabilidade e a apropriação dos dados pelas equipes pedagógicas.

Neste cenário, a relevância deste estudo para o novo ecossistema de aprendizagem reside na proposta de uso da IAG não como um substitutivo de *softwares* preditivos, mas como uma interface de democratização da análise de dados complexos. Ao atuar como assistente na codificação em linguagem *R* e na análise de dados, a IAG rompe a barreira técnica que historicamente separou gestores educacionais da estatística inferencial. Esta abordagem inova ao conferir autonomia analítica a profissionais da educação, permitindo uma gestão ágil, baseada em evidências e capaz de converter microdados de registros escolares em planos de



ação pedagógica preventiva. Desta forma, a tecnologia é utilizada para humanizar a gestão, direcionando o foco do educador para a intervenção pedagógica e o acolhimento, enquanto a máquina assume a complexidade do processamento estatístico.

Considera-se, assim, que, dado o volume de dados analisados e sua extensão temporal, bem como o uso da IAG no campo da gestão pedagógica escolar, este estudo pode contribuir para suprir uma lacuna sobre o uso da IAG na democratização da análise de dados no ambiente educacional e no auxílio na indução da permanência, com equidade, de estudantes com histórico de defasagem em sua trajetória escolar, com base no estudo de caso de cotistas do EMI.

Como resultado, a pesquisa revelou que é possível a um profissional sem formação especializada desenvolver pesquisa em uma extensa base de dados e extrair informações relevantes para a permanência estudantil com o auxílio da IAG. Os *scripts* gerados com auxílio da IA revelaram que, nos oito anos analisados, 95,8% dos estudantes matriculados pelas cotas tiveram desempenho abaixo da mediana na prova de ingresso, enquanto os não cotistas concentraram-se nos quartis com desempenho mais alto. O desempenho no ingresso mostrou forte associação com a trajetória acadêmica nos cursos do EMI: estudantes com baixo desempenho tiveram seis vezes mais chances de evasão, e os cotistas apresentaram 44% mais chances de retenção. Em termos educacionais, esta distância estatística traduz-se em uma sala de aula marcada pela heterogeneidade extrema, onde coexistem alunos com domínio consolidado das competências avaliadas e estudantes com lacunas severas de base.

Este mapeamento, realizado com auxílio da IAG, permite usar o desempenho no acesso como um indicador de vulnerabilidade acadêmica. Ele possibilita planejar ações pedagógicas e de permanência, otimizando o dimensionamento e a alocação institucional de recursos como a oferta de monitorias, o atendimento preventivo e o monitoramento por equipe multidisciplinar. Simultaneamente, o mapeamento serve como subsídio para que os professores possam realizar o planejamento pedagógico de forma mais adequada ao perfil de suas turmas antes mesmo do início das aulas.

Fundamentação Teórica

A compreensão do desempenho escolar no Ensino Médio Integrado (EMI) exige situar o estudante em um contexto de desigualdades históricas. A formação social brasileira foi marcada pela dualidade educacional, que reservou às elites uma formação propedêutica e às classes trabalhadoras uma instrução voltada ao ofício manual. Este modelo, sustentado durante décadas pelo mito da “democracia racial”, obscureceu barreiras estruturais que impediram a mobilidade social da população negra e de baixa renda.

Embora a redemocratização e a expansão das redes públicas de educação, desde o final do século XX, e as reformas educacionais do início do século XXI tenham ampliado o acesso à escolarização básica, a universalização ocorreu de forma segmentada, mantendo disparidades de qualidade entre as redes pública e privada. É neste cenário que a Lei 12.711/2012 instituiu uma política de cotas padronizada para a Rede Federal, reconhecendo que a igualdade formal de acesso não é suficiente para corrigir déficits acumulados por trajetórias escolares desiguais. Contudo, a simples garantia do ingresso não assegura a permanência, o que demanda mecanismos de gestão capazes de identificar e mitigar riscos de evasão e retenção.



A complexidade dos desafios educacionais contemporâneos, como a permanência discente no EMI, impõe a necessidade de superar modelos de gestão baseados apenas na intuição ou na experiência pregressa. A gestão baseada em evidências propõe o uso sistemático de dados para subsidiar a tomada de decisão pedagógica e administrativa. No contexto da educação pública, onde o volume de estudantes e dados são vastos, a capacidade de processar e interpretar essas informações torna-se estratégica para a redução das desigualdades escolares.

Estudos recentes, inclusive com base em dados de ambientes virtuais de aprendizagem (AVA), indicam que a análise preditiva e o mapeamento de perfis de risco permitem intervenções preventivas (Bitencourt; Silva; Xavier, 2022; Seo *et al.*, 2024a; Vaarma; Li, 2024; Vescovi, 2020). Entretanto, a implementação desta cultura analítica enfrenta barreiras técnicas, notadamente a exigência de letramento em ciência de dados por parte de gestores e educadores, a infraestrutura de dados e a governança da informação, bem como a privacidade dos dados e a transparência algorítmica, o que historicamente limitou o uso destas ferramentas a especialistas (Chong *et al.*, 2025; Seo *et al.*, 2024b; Vescovi, 2020).

A emergência da IAG altera o paradigma da análise de dados educacionais ao atuar como uma interface de mediação entre o usuário não especialista e a complexidade dos códigos de programação e da análise de informações estatísticas e gráficos com muitas variáveis. Diferente da IA preditiva tradicional, que demanda uma compreensão mais sofisticada de estatística e ciência de dados, a IAG permite interação em linguagem natural e favorece aprendizagem guiada e aplicada dos conhecimentos básicos necessários. Ainda que não dispense bibliografia especializada em metodologias quantitativas para as ciências sociais, como Agresti e Finlay (2012) e Kaplan (2024), a própria IAG pode auxiliar na compreensão e “tradução” de jargões mais técnicos e conceitos mais técnicos. Ao facilitar a elaboração de *scripts* de modelagem estatística em linguagens como *R* ou *Python*, ferramentas como os Grandes Modelos de Linguagem (LLMs), como ChatGPT e Gemini, ampliam o acesso a análises mais sofisticadas. Esta tecnologia não substitui o julgamento pedagógico, mas o instrumentaliza, permitindo que educadores construam diagnósticos mais precisos e precoces sobre suas turmas. Neste sentido, o potencial da IA na educação reside não apenas na personalização da aprendizagem, mas no empoderamento dos atores escolares para gerir o processo educativo baseados em dados empíricos, com maior equidade e eficácia (UNESCO, 2024).

No contexto das reformas sociais do início do século XXI, que abriram fissuras no projeto nacional excludente, especialmente em termos de classe, raça e gênero, e possibilitaram o acesso à educação de segmentos até então preteridos, as turmas se tornaram mais heterogêneas. Aumentou a diversidade do corpo discente e a inclusão de estudantes com deficiência, assim como as chances de produção e reprodução de desigualdades que a escola poderia contribuir para reduzir ou, pelo menos, não ampliar (Dubet, 2004).

Em uma época em que as desigualdades tecnológicas, seja no acesso aos equipamentos, às ferramentas e aos aplicativos digitais ou no letramento digital, ampliam as desigualdades de classe, o que ficou especialmente demonstrado durante a pandemia da Covid-19 (Macedo, 2021; Vitral Rezende, 2023), a IAG apresenta potencial para recrudesce-las (Buesa *et al.*, 2025).

Entretanto, esta mesma tecnologia pode ser usada para a superação de iniquidades sociais através da personalização do ensino, da automatização de atividades docentes mais



operacionais, liberando tempo docente para aperfeiçoar o planejamento pedagógico, ou da orientação dos alunos para o uso criativo, construtivo e ético das IAG (Azambuja; Silva, 2024). O aperfeiçoamento da gestão pedagógica é outra seara em que a IAG pode contribuir.

No EMI, a literatura tem mostrado desigualdades de retenção e evasão entre estudantes cotistas e não cotistas. Por não ser o objetivo deste artigo fazer uma revisão de literatura, não será detalhada a metodologia adotada. Porém, em levantamento realizado no Catálogo de Teses e Dissertações da CAPES, na Biblioteca Digital de Teses e Dissertações, no Portal de Periódicos da Capes, na SciELO Brasil, no Scopus, no WoS e no Portal de Busca integrada da USP, foram encontrados 812 trabalhos entre 2013 e 2025. Após eliminar os resultados repetidos, filtrar apenas os trabalhos que se referiam ao EMI da RFEPCT e selecionar as pesquisas que apresentavam dados de retenção e evasão comparados entre cotistas e não cotistas, restaram quinze trabalhos defendidos ou publicados neste período.

Dentre eles, seis trabalhos analisaram algum indicador referente ao primeiro ano letivo do EMI (Castro, 2022; Chaves *et al.*, 2020; Dantas, 2015; Ribeiro; Costa; Risco, 2021; Santos, 2018; Soldati, 2023). Entre os cinco que compararam taxas de retenção, quatro encontraram um percentual maior para os cotistas e um concluiu que a diferença não era significativa. E entre os três que compararam as taxas de evasão, dois assinalaram uma tendência maior entre os cotistas e um constatou uma evasão maior entre os estudantes da AC.

Quanto aos indicadores relativos ao final do curso, oito trabalhos compararam as taxas de evasão escolar entre cotistas e não cotistas, dos quais cinco apontaram percentuais maiores para cotistas (Kostycki, 2020; Pereira, 2020; Rateke, 2018; Ribeiro; Costa; Risco, 2021) e três indicaram não haver diferença significativa (Chaves *et al.*, 2020; Previtalli, 2023; Santos, 2021).

A partir deste cenário, investigou-se se havia ou não desigualdades no EMI no caso investigado e se é possível encontrar uma associação entre a retenção e a evasão e o desempenho dos estudantes na prova de acesso.

Metodologia

Este é um estudo de caso exploratório, com abordagem quantitativa. A pesquisa é censitária, sendo a população constituída por todos os estudantes que realizaram prova de acesso para o EMI do IFFluminense *campus* Campos Centro e *campus* Campos Guarus, de 2013 a 2020, totalizando 3.443 estudantes, cujos dados foram levantados a partir do sistema informatizado de registro acadêmico, o Q-Acadêmico, e dos editais e comunicados dos processos seletivos de ingresso.

Conforme a publicação “Diretrizes para o uso ético e responsável da inteligência artificial generativa: um guia prático para pesquisadores” (Sampaio; Sabbatini; Limongi, 2024), informa-se que as IAGs utilizadas, entre julho e dezembro de 2025, foram o modelo GPT-5 *Thinking*, desenvolvido pela OpenAI e disponibilizado por meio da plataforma ChatGPT, e os modelos Gemini 2.0 Pro e Gemini 2.0 Flash, desenvolvidos pelo Google, por meio da plataforma Google AI Studio.

A interação se deu a partir de um *prompt* básico padrão, conforme Apêndice A deste artigo, e de solicitações em que eram informados a linguagem de programação, os nomes das



variáveis, a natureza de seus conteúdos e os objetivos desejados para o desenvolvimento dos *scripts*. No ChatGPT, foi habilitado o modo “Lousa” para editar a programação. Quando houve dúvidas, acionou-se o modo “Estudar e aprender” a fim de subsidiar decisões mais técnicas.

Os *scripts* gerados foram testados passo a passo no RStudio®, a partir de documentos *R Markdown*. As versões da linguagem *R* utilizadas foram a 4.5.1 e 4.5.2 e o ambiente RStudio® correspondia à versão atualizada em 01/05/2025.

O Google AI Studio foi usado para cotejar os *scripts* propostos pelo ChatGPT, buscando aperfeiçoá-lo quando havia dúvidas, e para a obtenção de *insights* na análise de alguns dados e gráficos mais complexos. Esta consulta auxiliar foi feita a partir de *prints* dos resultados gerados no RStudio®, sem o fornecimento de informações que permitissem individualizar os estudantes. Embora o ChatGPT também realize estas análises, o Google AI Studio oferece análises diferentes e mais complexas sobre os dados. O *software* estatístico usado é gratuito e as IAGs foram utilizadas em suas versões pagas.

O uso das ferramentas de IAG, portanto, seguiu o seguinte protocolo de validação: (i) isolamento dos dados, garantindo que nenhum microdado sensível fosse inserido nas plataformas; (ii) verificação cruzada de sintaxe, onde o código gerado por um modelo (GPT-5) era revisado por outro (Gemini 2.0) e testado em ambiente local RStudio®; e (iii) supervisão humana integral, onde a autoria principal validou a lógica estatística de cada bloco de código antes de sua execução e os resultados após a execução. Esta abordagem assegura que a IA atuou estritamente como assistente de codificação, mantendo-se a responsabilidade autoral e analítica sob controle da autoria principal, conforme declaração de responsabilidade (Apêndice B).

As provas de ingresso no período analisado tinham um valor total de 40 (quarenta) pontos, sendo um ponto por questão. As questões foram distribuídas da seguinte forma: 10 de Língua Portuguesa, 10 de Matemática, 10 de Ciências Naturais, 5 de História e 5 de Geografia. A nota mínima classificatória dos candidatos era de 12 pontos, correspondendo a um mínimo de 30% de acertos para que o candidato não fosse eliminado.

Como são oito processos seletivos, optou-se por padronizar as notas para um escore. Esta padronização permite comparar estudantes de anos e cursos diferentes, mesmo que cada processo seletivo tenha uma prova diferente. Para isso, cada estudante recebeu um *z-score*, que indica o quanto a nota ficou acima ou abaixo da média do curso em que ingressou em dado ano. O *z-score* expressa a posição relativa de um valor em relação ao conjunto de notas daquele mesmo ano e curso. Ele informa quantos desvios-padrão acima ou abaixo da média anual do curso está a nota do estudante. Exemplo: se em determinado ano a nota média do curso foi 18 pontos e o desvio-padrão foi 4, então uma nota de 26 pontos está dois desvios-padrão acima da média. Nesse caso, seu *z-score* seria +2. Já uma nota de 14, que está um desvio-padrão abaixo da média, teria *z-score* -1.

A fórmula usada foi:

$$z = \frac{\text{nota do estudante} - \text{média do curso}}{\text{desvio padrão do curso}}$$

Onde o *z* mostra o quanto a nota de um estudante se afastou da média do seu ano e curso, permitindo comparar processos seletivos diferentes na mesma escala.

O *z-score* foi analisado a partir da distribuição das notas por quartis. O primeiro quartil (Q1) indica a nota abaixo da qual estão os 25% piores resultados do grupo. O terceiro quartil



(Q3), ao contrário, indica a nota acima da qual estão os 25% melhores resultados do segmento analisado. A mediana (2º quartil) é a medida central de distribuição das notas: 50% acima e a outra metade abaixo.

A partir destes quartis, calculados conforme a distribuição do *z-score* por ano e curso, será avaliado o nível de desempenho relativo dos estudantes na prova de ingresso. Os estudantes serão classificados em quatro categorias, conforme o Quadro 1.

Quadro 1 – Categorias de análise de desempenho dos estudantes na prova de ingresso a partir da distribuição do *z-score* por quartis (2013-2020)

Desempenho	Posição	Interpretação
Alto	$z \geq Q3$	Grupo com melhor desempenho relativo
Médio-Alto	$Q2 \leq z \leq Q3$	Segmento central, desempenho mediano superior
Médio-Baixo	$Q1 \leq z \leq Q2$	Segmento central, desempenho mediano inferior
Baixo	$z \leq Q1$	Grupo com pior desempenho relativo

Fonte: elaboração própria, 2025.

Essas categorias permitem comparar grupos de estudantes (AC, cotas, renda, cor etc.) em relação ao seu posicionamento dentro da distribuição de desempenho inicial, independentemente das diferenças entre as provas aplicadas ao longo dos anos.

Para analisar a nota de ingresso como preditor dos resultados acadêmicos ao final do curso, foi necessário ajustar as categorias de desempenho. Isso ocorreu porque os cotistas, como será visto na apresentação dos resultados, não se distribuem no quartil superior, o que geraria erro no teste de regressão. Assim, alteraram-se as categorias de desempenho para:

- os níveis “Alto” e “Médio-Alto” foram combinados na categoria “Faixa Superior”;
- o “Médio-Baixo” foi mantido como categoria própria (aproximadamente 25% da amostra); e
- o “Baixo” também foi mantido (aproximadamente 25% da amostra).

Para analisar a associação entre desempenho de ingresso, perfil de matrícula e trajetória acadêmica, estimou-se um modelo de regressão logística multinomial. A variável dependente foi a situação ao final do ciclo regular (resultado_final_ciclo), classificada em quatro grupos: Concluído, Retido, Excluído e Outros. A categoria “Concluído” foi definida como referência. Assim, os coeficientes estimados representam como variam as chances de retenção, exclusão ou enquadramento em “Outros” em comparação com um estudante que concluiu o curso.

As variáveis explicativas incluídas no modelo foram: o nível de desempenho na seleção, operacionalizado a partir dos quartis do *z-score* da nota de ingresso, reagrupados em: “Faixa Superior”, “Médio-Baixo” e “Baixo” (nivel_ingresso_agrupado); o tipo de vaga (tipo_mat_agregada), distinguindo “Não cotistas” (AC) e “Cotistas”; o *campus* de oferta do curso (*campus*), com Campos Centro e Campos Guarus; o curso técnico do EMI (curso), com os sete cursos da amostra; a cidade de residência no ingresso (cidade_residencia), diferenciando estudantes que residiam no mesmo município do *campus* e aqueles provenientes de outro município; a cor/raça autodeclarada (cor_analise) agregada em: PPI (pretos, pardos e indígenas), BA (brancos e amarelos) e “Não declarados” (2,99% da amostra); a renda familiar *per capita* por faixas de renda (rfpc_analise) classificada em: “RFPC $\leq 1,5$ SM”, “ $1,5$ SM <



RFPC ≤ 3 SM”, e “RFPC > 3 SM”, com uma categoria de “Não declarados” (8,8% da amostra); e o sexo (sexo), masculino ou feminino.

Os parâmetros estimados indicam como se alteram as chances de retenção e exclusão (evasão e cancelamento de matrícula) em comparação a um estudante que concluiu o curso e que apresenta o perfil de referência: desempenho de ingresso na faixa superior, matrícula na AC, curso de Automação Industrial, Campos Centro, residente no município de Campos dos Goytacazes, autodeclarado branco ou amarelo, com renda superior a 3 SM e do sexo masculino.

O modelo foi ajustado utilizando a função multinom, do pacote nnet, no ambiente R. Os resultados são apresentados em termos de razões de chance (*odds ratios*). Valores de *odds ratio* (OR) superiores a 1 indicam maior chance do resultado acadêmico ao final do curso em relação à categoria de referência; valores menores que 1 indicam menor chance, mantidas constantes as demais variáveis.

Para garantir a transparência e a reprodutibilidade da pesquisa, os *scripts* computacionais desenvolvidos em linguagem R não estão em repositório aberto, mas encontram-se disponíveis e podem ser fornecidos pela autoria principal mediante solicitação por e-mail.

Resultados

Para analisar o desempenho dos estudantes, considerando a posição relativa das notas de ingresso padronizadas, realizou-se a distribuição do *z-score* dos estudantes matriculados entre 2013 e 2020, conforme resumido na Tabela 1, que apresenta as estatísticas descritivas dos *z-scores* por quatro categorias de desempenho, em que 0 (zero) corresponde à média histórica. A distribuição amostral (*n*) evidencia que a categorização obedeceu ao critério de quartis, dividindo a população discente em quatro grupos de dimensão equivalente (~25% cada), o que permite uma análise comparativa robusta entre os estratos na série histórica de oito anos.

Tabela 1 – Estatísticas descritivas do *z-score* padronizado (*z*) por categoria de desempenho no ingresso (2013-2020)

Desempenho	n	Mín.	Q1	Q2	Q3	Máx.	Média	Desvio-padrão
Alto	863	0,72	0,90	1,12	1,46	3,47	1,23	0,43
Médio-Alto	860	0,17	0,34	0,45	0,56	0,72	0,45	0,15
Médio-Baixo	870	-0,83	-0,63	-0,43	-0,16	0,17	-0,39	0,29
Baixo	850	-2,44	-1,52	-1,26	-1,04	-0,83	-1,30	0,32

Fonte: elaboração própria, 2025.

A análise dos dados revela uma profunda disparidade de conhecimentos prévios entre os estudantes que acessaram o EMI. Observa-se uma distância expressiva de 2,53 desvios-padrão (DP) entre a média do grupo de desempenho “Alto” ($M = 1,23$) e a do grupo “Baixo” ($M = -1,30$). Em termos educacionais, essa distância estatística traduz-se em uma sala de aula



marcada pela heterogeneidade extrema, em que coexistem alunos com domínio consolidado das competências avaliadas e estudantes com lacunas severas de base.

O grupo classificado como “Alto” não apenas apresenta desempenho superior à média geral (zero na escala padronizada), mas também exibe a maior variabilidade interna da série ($DP = 0,43$). O valor máximo de $z = 3,47$ indica a presença de *outliers* positivos: estudantes com desempenho excepcional, situados a mais de três desvios-padrão da média, o que sugere a identificação de talentos acadêmicos com alto potencial.

Em contrapartida, os grupos intermediários (“Médio-Alto” e “Médio-Baixo”) apresentam maior homogeneidade e caracterizam o perfil do “aluno médio” dos cursos. Destaca-se, contudo, que o grupo “Médio-Baixo” já exibe média negativa ($M = -0,39$), com 75% de seus integrantes situados abaixo da média global ($Q3 = -0,16$), indicando que as dificuldades de aprendizagem não se restringem ao quartil inferior.

O estrato “Baixo”, por sua vez, revela um cenário de vulnerabilidade acadêmica crítica. Com média de $-1,30$ e valores mínimos atingindo $-2,44$, este grupo ingressa na instituição com uma defasagem significativa em relação aos seus pares. A magnitude negativa destes *z-scores* sugere que, sem políticas institucionais de nivelamento ou apoio pedagógico intensivo no início do ciclo, estes estudantes constituem o público com maior risco potencial de reprovação e evasão escolar.

Em síntese, a padronização das notas via *z-score* desvela que a oportunidade de acesso, após o processo seletivo, não corresponde a uma isonomia nas condições de partida para o aprendizado. O desafio imposto ao corpo docente é, portanto, gerenciar processos de ensino-aprendizagem para um público cuja amplitude de desempenho varia, estatisticamente, de $-2,44$ a $+3,47$ desvios ($5,91 DP$). Esta variação demanda estratégias pedagógicas diferenciadas para atender, simultaneamente, à necessidade de inclusão da base e de estímulo ao topo.

A Figura 1 apresenta a relação entre o desempenho escalonado no processo seletivo e o resultado acadêmico ao final do primeiro ano. A análise visual indica uma correlação positiva forte entre a nota de entrada e o resultado acadêmico. No quartil superior (desempenho “Alto”), predomina massivamente a aprovação (área vermelha, 86,9%), com taxas de retenção (8,5%) residuais, indicando que o alto desempenho na entrada atua como um fator de “proteção” contra o insucesso escolar.

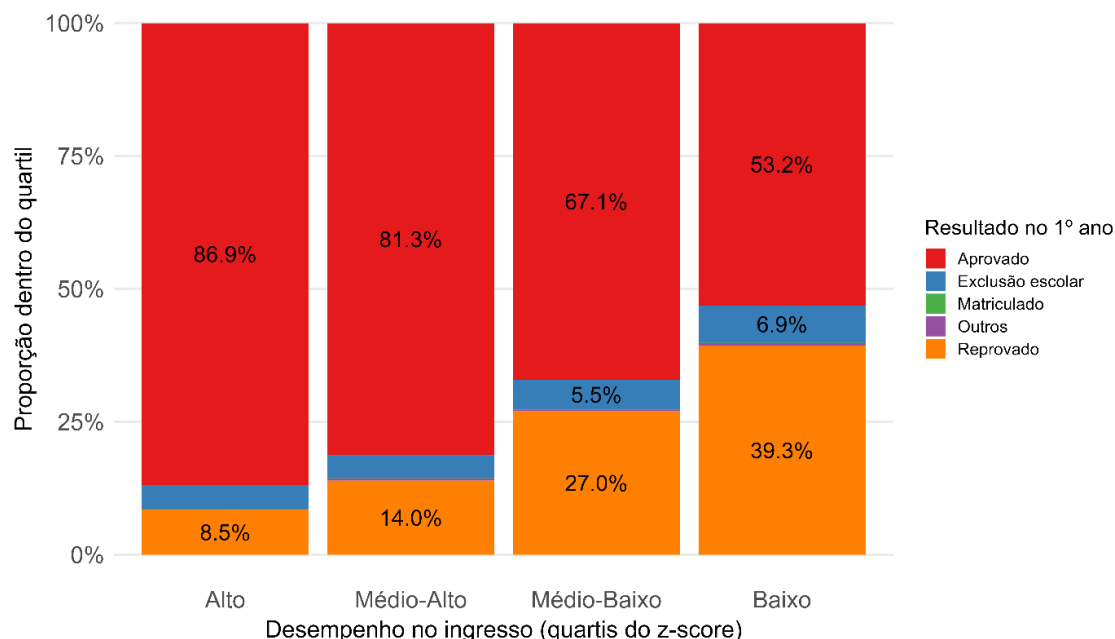
No quartil inferior (“Baixo”), o cenário se altera drasticamente. A taxa de reprovação (área laranja, 39,3%) expande-se significativamente. Somada à exclusão escolar (6,9%), percebe-se que quase metade dos alunos deste grupo não logra êxito na conclusão do primeiro ano letivo no tempo regular.

O gráfico revela que a principal consequência da baixa proficiência de entrada não é a evasão imediata (exclusão), mas sim a retenção acadêmica (reprovação). Isso corrobora a hipótese levantada no **Erro! Fonte de referência não encontrada.**: as lacunas de base trazidas pelo grupo “Baixo” (média de $-1,3$ desvios) criam barreiras pedagógicas muitas vezes difíceis de transpor no currículo do 1º ano, resultando em altas taxas de reprovação.

A Tabela 2 evidencia uma polarização quase absoluta entre os dois grupos de ingressantes, cotistas e não cotistas, revelando que o desempenho no processo seletivo atua como um filtro socioeducacional.



Figura 1 – Relação entre desempenho no ingresso e resultados no 1º ano (2013-2020)



Fonte: elaboração própria, 2025.

Nota: valores inferiores a 5% não são indicados no gráfico.

Tabela 2 – Estatísticas descritivas do z-score padronizado (z) por categoria de desempenho no ingresso e por cotistas comparados com não cotistas (2013-2020)

Observação	Desempenho	n	Mín.	Q1	Q2	Q3	Máx.	Média	Desvio-padrão
Cotistas	Médio-Alto	63	0,17	0,20	0,24	0,29	0,56	0,26	0,09
Cotistas	Médio-Baixo	644	-0,83	-0,68	-0,51	-0,28	0,15	-0,45	0,26
Cotistas	Baixo	820	-2,44	-1,54	-1,27	-1,05	-0,83	-1,31	0,32
Não cotistas	Alto	863	0,72	0,90	1,12	1,46	3,47	1,23	0,43
Não cotistas	Médio-Alto	797	0,17	0,36	0,49	0,57	0,72	0,47	0,14
Não cotistas	Médio-Baixo	226	-0,83	-0,41	-0,16	0,05	0,17	-0,22	0,29
Não cotistas	Baixo	30	-2,02	-1,15	-0,98	-0,94	-0,86	-1,09	0,26

Fonte: elaboração própria, 2025.

A análise dos dados demonstra três fenômenos estatísticos críticos caracterizados por: exclusividade do topo (desempenho “Alto”) para os não cotistas, uma inversão populacional entre os dois grupos observados e desvantagens intragrupos.

O dado mais contundente é a ausência total de cotistas no quartil de desempenho “Alto” em oito anos de processos seletivos. A totalidade dos 863 estudantes deste grupo é composta por não cotistas, indicando que a elite acadêmica (*z-score* médio de 1,23) é, invariavelmente, proveniente da AC.



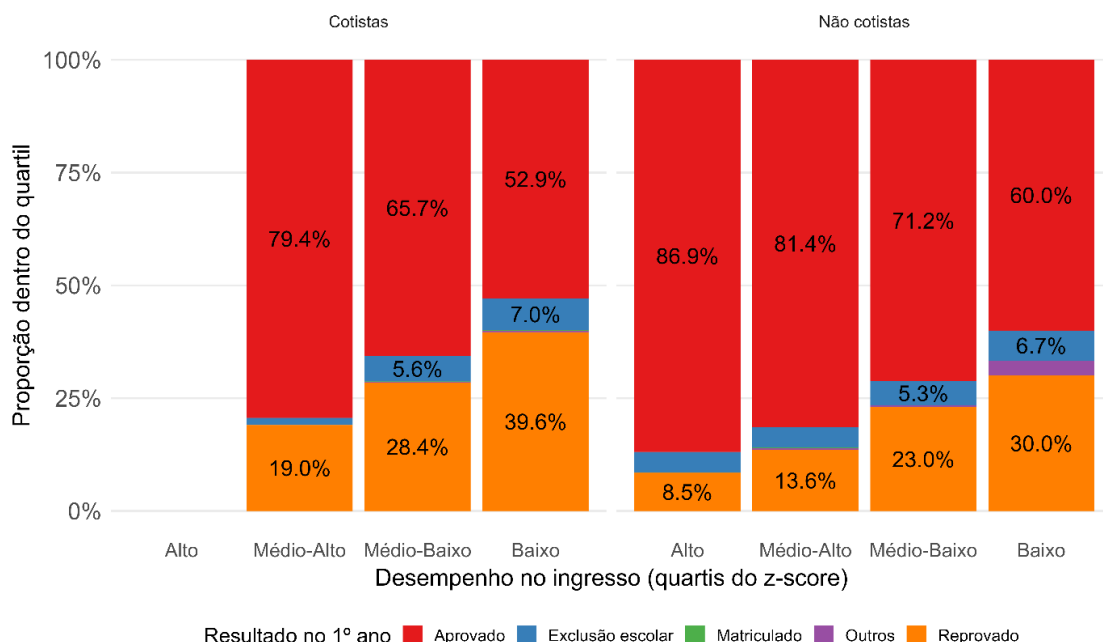
A inversão populacional pode ser observada pela distribuição espelhada: os não cotistas estão massivamente concentrados nos desempenhos superiores. No total, 86,6% deste público ($n = 863 + 797$) situa-se nas faixas “Alto” e “Médio-Alto”. Já os cotistas estão maciçamente concentrados nos estratos inferiores: 95,8% deste público ($n = 644 + 820$).

Mesmo quando comparados dentro da mesma categoria, os não cotistas tendem a ter um desempenho melhor. No grupo “Médio-Alto”, por exemplo, a média dos não cotistas (0,47) é sensivelmente superior à dos cotistas (0,26), sugerindo que, mesmo entre os melhores cotistas, o desempenho tende ao limite inferior da faixa.

A tabela deixa claro que a política de cotas está cumprindo seu papel de inclusão precisamente porque o critério puramente meritocrático (nota da prova) segrega as populações de forma distinta. Enquanto os não cotistas dominam os quartis de excelência ($z \geq Q2$), os cotistas compõem a quase totalidade da base de desempenho e desempenho crítico (“Baixo” e “Médio-Baixo”).

A Figura 2 permite analisar os resultados acadêmicos comparando “iguais com iguais” (dentro das faixas onde a comparação é possível).

Figura 2 – Relação entre desempenho no ingresso e resultados no primeiro ano comparado entre cotistas e não cotistas (2013-2020)



Fonte: elaboração própria, 2025.

Nota: valores inferiores a 5% não são indicados no gráfico.

A visualização sugere que o desempenho acadêmico prévio (mensurado pelo *z-score*) é uma variável mais determinante para o resultado no 1º ano do que a categoria de ingresso isoladamente (cotistas e não cotistas). Em ambos os grupos a taxa de aprovação diminui e a de reprovação aumenta à medida que o desempenho de entrada piora, indicando uma tendência estrutural relacionada à defasagem acadêmica, embora os cotistas apresentem taxas piores.



O gráfico no segmento dos Cotistas evidencia melhor o espaço vazio na categoria de “Alto” desempenho, reiterando o dado anterior: a inexistência de estudantes cotistas no quartil de excelência do ingresso. O grupo cotista só começa a existir a partir do “Médio-Alto”.

O comportamento das barras é similar em ambos os gráficos. Tanto para cotistas quanto para não cotistas, a taxa de aprovação cai e a de reprovação sobe à medida que o desempenho de entrada piora. Isso sugere que a defasagem acadêmica (a nota ruim) é o principal preditor do insucesso, independentemente se o aluno é cotista ou não.

A análise comparada do estrato de desempenho “Baixo” revela que a vulnerabilidade acadêmica transcende a categoria de ingresso. Não cotistas com baixo desempenho inicial apresentam taxas de insucesso de 30%, próximas aos 39% observados entre cotistas na mesma faixa.

Tem-se, portanto, dois preditores importantes: a nota de ingresso, de forma universal, e a forma de ingresso como indicador de agravamento das desigualdades de formação acadêmica.

Com base nestes sinalizadores, optou-se por estimar um modelo de regressão logística multinomial, conforme a metodologia indicada. Os quadros 2 e 3 resumem os resultados da regressão para os desfechos “Retidos” e “Excluídos”, sendo a categoria de referência a conclusão do curso no ciclo regular.

A regressão indicou que o preditor mais forte no modelo multinomial, com controle por *campus*, curso, forma de ingresso, faixa de renda, município de residência em relação ao município do *campus*, sexo e cor/raça declarados, foi o que se chamou aqui de “efeito trajetória escolar”. O termo “efeito”, tal como utilizado aqui, não implica causalidade. Ele se refere apenas à associação estatística estimada pelo modelo, isto é, à capacidade de uma variável atuar como preditor da probabilidade de determinado desfecho, mantendo constantes as demais variáveis incluídas.

Desta forma, o desempenho relativo na prova de ingresso, padronizado pelo *z-score*, foi aqui considerado como uma síntese quantitativa desta trajetória escolar ao longo do ensino fundamental. Este desempenho não é a causa da retenção ou da exclusão escolar; ele funciona como um indicador de desigualdades formativas anteriores ao ingresso e, por esse motivo, ajuda a prever diferenças nos resultados acadêmicos posteriores e a planejar ações de intervenção pedagógica.

O “efeito trajetória escolar” é observado quando se identifica que estudantes que ingressaram no quartil mais baixo, categoria de desempenho “Baixo”, apresentaram 3,6 vezes mais chances (OR 3,66) de ficarem retidos (Quadro 2) e não concluírem o curso no tempo regular em comparação com os estudantes classificados com desempenho na “Faixa superior”. No caso da exclusão escolar (Quadro 3), esta vulnerabilidade se intensifica: estes estudantes têm quase seis vezes mais chances (OR 5,79) de evasão ou cancelamento de matrícula (exclusão). Embora menos intensa, a associação também foi forte para o desempenho de ingresso “Médio-Baixo”: 2,3 vezes mais chances de retenção e 3,2 vezes mais chances de exclusão escolar.

Ser cotista apresentou associação estatística significativa com a retenção, mas não com a exclusão escolar. Controlando por desempenho no ingresso e renda, estudantes cotistas exibem 44% mais chances de ficarem retidos (OR 1,439) quando comparados aos estudantes da AC. Mesmo quando dois estudantes apresentam o mesmo nível de desempenho e pertencem



à mesma faixa de renda, o grupo cotista mantém probabilidade estatisticamente maior de reprovação.

O resultado referente à renda apresentou um padrão contraintuitivo, com estudantes na faixa de renda mais baixa alcançando maiores taxas de conclusão, ao contrário do que se poderia esperar: maior vulnerabilidade econômica, maior evasão. Os estudantes com renda $\leq 1,5$ SM apresentaram 69% menos chances de exclusão, enquanto a faixa intermediária apresentou 52% menos chances. Na retenção, mantendo as demais variáveis constantes, não houve variação estatisticamente significativa em nenhum dos dois grupos.

Para verificar este resultado, os estudantes da série histórica foram distribuídos conforme a faixa de renda e o resultado ao final do ciclo. A distribuição confirmou que 12,9% dos estudantes com renda $\leq 1,5$ SM estavam em situação de exclusão escolar ao final do ciclo regular, enquanto os estudantes com renda > 3 SM apresentaram uma taxa de exclusão de 32,9%. Na retenção, porém, os estudantes de menor renda apresentaram o maior percentual, 30,1%, enquanto os estudantes com RFPC > 3 SM obtiveram uma frequência menor, 19,1%. Apesar desta tendência, ela não foi significativa para que o modelo de regressão a identificasse, ao contrário do desfecho exclusão.

Quanto ao sexo, o modelo indicou que estudantes do sexo feminino apresentaram resultados mais favoráveis na comparação com o sexo masculino. Elas exibem menor probabilidade de retenção (OR = 0,66) e 20% menos chances de evasão (OR = 0,79).

A variável cor/raça apontou um efeito de 66% mais chances de retenção entre os estudantes que não declararam cor (OR = 1,66). Isto levanta a questão sobre a existência de um possível padrão entre os “Não declarados”, talvez associado à forma como se construiu historicamente no Brasil a identificação do brasileiro em relação a sua cor/raça. No modelo não foram encontradas maiores ou menores chances de retenção ou exclusão por cor/raça declarada considerando o grupo agregado PPI (pretos, pardos e indígenas) em relação ao grupo BA (brancos e amarelos).

O resultado por município de moradia também contrariou o esperado: estudantes que residem em “Outro município” apresentaram probabilidade significativamente menor de retenção, com 39% menos chances (OR 0,61) na comparação com os residentes no município do *campus*. Uma hipótese é que o compromisso necessário para realizar deslocamentos longos favoreça o compromisso de estar em sala de aula.

Por fim, embora o *campus* não tenha apresentado uma associação significativa em relação aos desfechos analisados, o curso de eletrotécnica exibiu cerca de 50% mais chances de retenção e o curso de eletrônica apresentou 22% mais chances de retenção.

Os quadros com os resultados da regressão conforme os desfechos analisados estão disponíveis no Apêndice C.

Discussão

Bitencourt et al. (2022) treinaram um modelo preditivo, com base em 1.429 registros de estudantes universitários matriculados de 2013 a 2019, para identificar perfis de estudantes potencialmente mais vulneráveis à evasão. Nesta pesquisa, utilizou-se a IAG para desenvolver scripts e analisar a relação entre desempenho no ingresso e resultados acadêmicos ao final do



primeiro ano letivo e ao final do curso, com base em 3.443 estudantes matriculados no ensino médio integrado, de 2013 a 2020.

O resultado em relação à renda precisa ser considerado com cuidado, pois o perfil estudantil dos *campi* analisados é de predominância absoluta de estudantes de baixa renda. Além disso, as taxas maiores de retenção indicam uma possível interpretação de uma “permanência defensiva”. Marcada pelo fenômeno da longa permanência (estudantes não evadem, mas também não concluem o curso no ciclo regular), ela caracteriza-se por estudantes que permanecem, apesar dos obstáculos (retenções), a fim de defender oportunidades educacionais que, pelas características socioeconômicas e demográficas destes estudantes, são consideradas limitadas. Esta percepção de oportunidades limitadas se dá por um conjunto de fatores subjetivos, materiais e simbólicos; individuais e coletivos; e leva em conta a individualidade do estudante, sua experiência de vida, a estrutura familiar e as estruturas histórico-sociais que produzem e reproduzem desigualdades socialmente percebidas. Outra hipótese interpretativa, que não exclui a anterior, é que a política de permanência estudantil do Instituto Federal analisado, que oferta bolsas e auxílios para estudantes de baixa renda mediante processo seletivo, pode estar alcançando, ainda que parcialmente, sua finalidade de reduzir a exclusão escolar motivada por fatores econômicos.

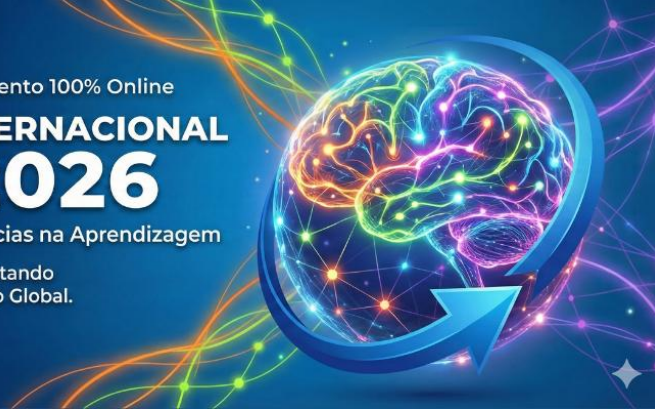
Entre os cotistas, 53,7% está na faixa de desempenho “Baixo”, categoria que apresentou 3,6 vezes mais chances de retenção e 5,7 vezes mais chances de exclusão escolar. A faixa de desempenho “Médio-Baixo”, na qual se enquadravam 42,2% dos cotistas ingressantes entre 2013 e 2020, apresentou 2,3 vezes mais chances de retenção e 3,2 vezes mais chances de exclusão escolar. Na regressão logística multinomial, os cotistas apresentaram 44% mais chances de retenção. Estes dados ratificam a maioria das pesquisas identificadas sobre o EMI na RFEPC, que apontaram maiores taxas de retenção e evasão para os cotistas.

Estudantes do sexo masculino, que não declararam cor, residentes no município sede dos *campi* investigados e alunos dos cursos de Eletrotécnica e Eletrônica apresentaram maiores chances de retenção e/ou exclusão. Além disso, os estudantes da menor faixa de renda apresentaram 69% menos chances de exclusão.

Os dados indicam que o desempenho acadêmico prévio (mensurado pelo *z-score*) é um preditor mais adequado para a trajetória acadêmica do que a forma de ingresso sozinha, ou qualquer outra variável analisada. A tendência estrutural é forte: o aluno que ingressa com déficit de base (grupos Baixo e Médio-Baixo) enfrenta barreiras de aprovação e permanência severas, seja ele oriundo de cotas ou da ampla concorrência.

Isto indica que, na série histórica analisada, o maior problema para a permanência estudantil não foram os fatores socioeconômicos isoladamente, mas os acadêmicos. O que levanta a possibilidade de que as ações de permanência voltadas para o suporte acadêmico e a recuperação da defasagem de base não foram suficientes para compensar o legado negativo da trajetória escolar de origem.

Os resultados sustentam a hipótese de que é possível associar a trajetória escolar no ensino fundamental com a trajetória no EMI. Estudantes com defasagem de base, evidenciada pelo desempenho no processo seletivo de ingresso, tendem a ter maiores retenções no EMI ou a engrossar as estatísticas da exclusão escolar, configurando o que se chamou de “efeito trajetória escolar”.



Assim como em Bitencourt *et al.* (2022), que utilizaram modelos preditivos tradicionais, este estudo demonstra que a IAG também permite uma gestão ágil e descentralizada. A implicação gerencial central é a mudança do paradigma de “reação ao fracasso” para a “antecipação do risco”. Ao identificar, ainda na matrícula, que 95,8% dos cotistas ingressam na faixa de risco “Médio-Baixo” e “Baixo” com chances maiores de retenção e evasão, a gestão escolar pode, antes do início das aulas, direcionar recursos de nivelamento para estes estudantes e para outros na mesma faixa de desempenho.

Conclusão/Considerações finais

Esta pesquisa oferece uma contribuição metodológica ao campo da gestão educacional ao validar o uso de LLMs como ferramentas de democratização da análise de dados educacionais. A IAG mitiga a barreira técnica da programação estatística, permitindo que pesquisadores e gestores da educação básica monitorem indicadores complexos de desigualdade com rigor científico. Uma gestão que incorpore a IAG e que seja baseada em evidências pode tornar o planejamento docente e a personalização do ensino mais eficiente.

Do ponto de vista pedagógico, o estudo confirmou o “efeito trajetória escolar”: o desempenho no ingresso mostrou-se um preditor de risco mais robusto que as categorias administrativas de cotas. A identificação de que estudantes com baixo desempenho inicial têm quase seis vezes mais chances de exclusão escolar oferece uma métrica acionável para a implementação de sistemas de alerta precoce.

Como limitações, destaca-se o recorte regional (dois *campi* do IFFluminense) e a ausência de dados qualitativos sobre a experiência subjetiva dos estudantes retidos. Além disso, a validade dos modelos de IA carece de supervisão humana constante devido à possibilidade de alucinações técnicas na geração de códigos.

Para trabalhos futuros, sugere-se a aplicação deste protocolo metodológico assistido por IA em outras unidades da Rede Federal, visando testar a escalabilidade do modelo. Recomenda-se também investigar o impacto de intervenções pedagógicas personalizadas, desenhadas a partir destes diagnósticos, na reversão das taxas de retenção identificadas.

Por fim, levanta-se a possibilidade de que instituições de ensino médio que não adotam provas de ingresso possam desenvolver seu próprio modelo a partir de uma métrica que sintetize as notas do histórico estudantil do ensino fundamental, substituindo a nota na prova de ingresso do EMI.

Referências

- AGRESTI, Alan; FINLAY, Barbara. **Métodos estatísticos para as Ciências Sociais**. 4. ed. Porto Alegre: Penso, 2012.
- AZAMBUJA, Celso Candido de; SILVA, Gabriel Ferreira da. Novos desafios para a educação na Era da Inteligência Artificial. **Filosofia Unisinos**, [s. l.], v. 25, p. e25107, 2024. Disponível em: <https://tinyurl.com/y8c3ex23>. Acesso em: 12 fev. 2026.
- BITENCOURT, Wanderci Alves; SILVA, Diego Mello; XAVIER, Gláucia do Carmo. Pode a inteligência artificial apoiar ações contra evasão escolar universitária?. **Ensaio: Avaliação e**



Políticas Públicas em Educação, v. 30, p. 669–694, 2022. Disponível em: <https://tinyurl.com/ymejxz7b>. Acesso em: 12 fev. 2026.

BUESA, Carlos Busón *et al.* **A inteligência artificial é nossa: caminhos para a soberania tecnológica no Sul global**. SciELO Preprints, 2025. Disponível em: <https://tinyurl.com/bddnedpp>. Acesso em: 13 out. 2025.

CAMPOS, Luiz Augusto; LIMA, Márcia (org.). **O impacto das cotas. Duas décadas de ação afirmativa no ensino superior brasileiro**. Belo Horizonte ; São Paulo: Autêntica, 2025.

CARREIRA, Denise; HERINGER, Rosana (org.). **10 anos da Lei de Cotas: conquistas e perspectivas**. Rio de Janeiro, RJ: Faculdade de Educação UFRJ: Ação Educativa, 2022. Disponível em: <https://tinyurl.com/2e2kpr67>. Acesso em: 14 dez. 2022.

CARVALHAES, FLÁVIO; JÚNIOR, João Feres; DAFLON, Verônica Toste. O impacto da Lei de Cotas nos estados: um estudo preliminar. **Texto para Discussão GEMAA**, [s. l.], v. 1, p. 1–17, 2013. Disponível em: <https://tinyurl.com/443crkbz>. Acesso em: 17 out. 2025.

CASTRO, Fabio Oliveira de. **Processo seletivo de ingresso no ensino médio integrado do instituto federal de São Paulo: análise crítica de dois modelos**. 2022. Dissertação (Mestrado em Educação) - Universidade Federal de São Paulo, Guarulhos/SP, 2022. Disponível em: <https://tinyurl.com/3hjks7ax>.

CHAVES, Kelson Gérison Oliveira *et al.* Permanência e êxito escolar: análise do desempenho acadêmico de estudantes ingressantes no ensino médio integrado do IFRN/Caicó através das cotas étnico-raciais no ano de 2014. **Revista Educação e Políticas em Debate**, v. 9, n. 3, p. 672–691, 2020.

CHONG, Ke Ting *et al.* A systematic review of machine learning techniques for predicting student engagement in higher education online learning. **Journal of Information Technology Education: Research**, v. 24, p. 005, 2025. Disponível em: <https://tinyurl.com/2sfm72nr>. Acesso em: 15 fev. 2026.

CIAVATTA, Maria. **Mediações históricas de trabalho e educação: gênese e disputas na formação de trabalhadores**. Rio de Janeiro: Lamparina : CNPq : FAPERJ, 2009.

COSTA, Leandro *et al.* Educação e inteligência artificial: reflexões críticas e propositivas. **Revista Interinstitucional Artes de Educar**, v. 11, n. 1, p. 329–354, 2025. Disponível em: <https://tinyurl.com/52kz4xjc>. Acesso em: 12 fev. 2026.

DANTAS, Maria José Pereira. **Ações afirmativas na educação profissional técnica de nível médio: um estudo no IFPB - campus João Pessoa**. 2015. Dissertação (Mestrado em Gestão Pública para o Desenvolvimento do Nordeste) - Universidade Federal de Pernambuco, Recife, 2015. Disponível em: <https://repositorio.ufpe.br/handle/123456789/15608>.

DPU; ABPN. **Pesquisa sobre a implementação da política de cotas raciais nas universidades federais**. [S. l.]: Grupo de Trabalho de Políticas Etnorraciais da Defensoria Pública da União : Associação Brasileira de Pesquisadores/as Negros/as, 2022. Disponível em: <https://tinyurl.com/4x9wz7v4>. Acesso em: 13 set. 2022.

DUBET, François. O que é uma escola justa?. **Cadernos de Pesquisa**, São Paulo, v. 34, n. 123, 2004.



FERES JÚNIOR, João *et al.* **Ação afirmativa: conceito, história e debates**. Rio de Janeiro: EdUERJ, 2018.

FERREIRA, Jair; SUGANUMA, Simone (org.). **Repertório bibliográfico sobre a condição do negro no Brasil**. Brasília: Câmara dos Deputados, Centro de Documentação e Informação, Edições Câmara, 2018. (Série fontes de referência, nº 1).

FRIGOTTO, Gaudêncio (org.). **Institutos Federais de Educação, Ciência e Tecnologia**. Rio de Janeiro: UERJ, LPP, 2018.

FRIGOTTO, Gaudêncio *et al.* O “estado da arte” das pesquisas sobre os IFs no Brasil: a produção discente da pós-graduação de 2008 a 2014. *In*: FRIGOTTO, Gaudêncio (org.). **Institutos Federais de Educação, Ciência e Tecnologia**. Rio de Janeiro: UERJ, LPP, 2018. p. 83–112.

FRIGOTTO, Gaudêncio; CIAVATTA, Maria; RAMOS, Marise (org.). **Ensino médio integrado: concepção e contradições**. 3. ed. São Paulo: Cortez, 2012.

GILIOLI, Renato de Sousa Porto. **Políticas de cotas no Brasil. Das “ações negativas” imigratórias e fundiárias às ações afirmativas na educação superior**. [S. l.]: [s. d.], 2023.

KAPLAN, David. **Manual de metodologia quantitativa para as ciências sociais**. Petrópolis, RJ: Vozes, 2024.

KOSTRYCKI, Xana Machado. **Para além do acesso: a política de cotas e o abandono escolar no Instituto Federal do Paraná, campus Paranaguá**. 2020. Dissertação (Mestrado em Educação Profissional e Tecnológica) - Instituto Federal do Paraná, Curitiba/PR, 2020.

MACEDO, Renata Mourão. Direito ou privilégio? Desigualdades digitais, pandemia e os desafios de uma escola pública. **Estudos Históricos (Rio de Janeiro)**, [s. l.], v. 34, n. 73, p. 262–280, 2021. Disponível em: <https://tinyurl.com/3s3b77s2>. Acesso em: 29 out. 2025.

MORAES, Carmen Sylvia Vidigal. Educação permanente: direito de cidadania, responsabilidade do estado. **Trabalho, Educação e Saúde**, Rio de Janeiro, v. 4, n. 2, p. 395–416, 2006. Disponível em: <https://tinyurl.com/42c4uuh4>. Acesso em: 12 nov. 2022.

MORAES, Carmen Sylvia *et al.* **Formação de professores do ensino médio. Ensino médio e formação humana integral**. Curitiba: UFPR/Setor de Educação, 2013. (Etapa 1. Caderno 1).

MORAES, Carmen Sylvia. Trabalho e educação como pauta do GT Trabalho e educação da ANPED. Algumas considerações sobre o campo de pesquisa. **Revista Trabalho Necessário**, [s. l.], v. 13, n. 20, 2018. Disponível em: <https://tinyurl.com/yd8xjesc>. Acesso em: 11 out. 2025.

MORAES, Carmen Sylvia Vidigal; ALENCAR, Felipe; JACOMINI, Márcia Aparecida. Internacionais de Trabalhadores, Gramsci e a relação trabalho-educação: reflexões para a política educacional. **PerCursos**, v. 25, p. e0210, 2024. Disponível em: <https://tinyurl.com/bdhwmbem>. Acesso em: 11 out. 2025.

PACHECO, Jairo Queiroz; SILVA, Maria Nilza da (org.). **O negro na universidade: o direito à inclusão**. Brasília: Fundação Cultural Palmares, 2007.

PAIVA, Angela Randolpho (org.). **Ação afirmativa em questão: Brasil, Estados Unidos, África do Sul e França**. Rio de Janeiro, RJ: Pallas, 2013.



PEREIRA, Aline de Carvalho. **Ações afirmativas na educação profissional técnica de ensino médio. Um estudo de caso com jovens estudantes cotistas.** 2020. Dissertação (Mestrado em Educação) - Universidade Federal de Alfenas, Alfenas/MG, 2020.

PREVITALLI, Márcio José. **Lei de Cotas no contexto da Educação Profissional e Tecnológica:** reflexões sobre a permanência escolar dos alunos cotistas nos cursos técnicos integrados e de graduação do IFSULDEMINAS - campi Passos, Poços de Caldas e Pouso Alegre. 2023. Dissertação (Mestrado em Educação Profissional e Tecnológica) - Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas Gerais, Poços de Caldas/MG, 2023.

RATEKE, Deise. **As contradições em torno do acesso e da permanência de estudantes da classe trabalhadora por meio da implementação da Lei de Cotas nos cursos de Ensino Médio Técnico Integrado do IFSC campus Florianópolis.** 2018. Tese (Doutorado em Educação) - Centro de Ciências da Educação, Universidade Federal de Santa Catarina, Florianópolis, 2018. Disponível em: <https://tinyurl.com/2tnw8x8a>. Acesso em: 27 ago. 2020.

RIBEIRO, Marcos Abraão Fernandes; COSTA, Luciana Machado da; RISSO, Sérgio Rangel. Avanços, Contradições e Desafios da Política de Cotas na Educação Básica: o caso do ensino médio profissionalizante do IFF campus Campos Centro (2016-2018). **Tomo**, n. 39, p. 299–327, 2021.

SAMPAIO, Rafael Cardoso; SABBATINI, Marcelo; LIMONGI, Ricardo. **Diretrizes para o uso ético e responsável da inteligência artificial generativa:** um guia prático para pesquisadores. São Paulo, SP: Sociedade Brasileira de Estudos Interdisciplinares da Comunicação - Intercom, 2024.

SANTOS, Danielle de Sousa. **Democratização do acesso ao Ensino Médio Integrado no IFSP:** o contexto da implementação da Lei de reserva de vagas. 2018. Tese (Doutorado em Educação) - Faculdade de Educação, Universidade Estadual de Campinas, Campinas, 2018. Disponível em: <https://tinyurl.com/3ax59z8a>. Acesso em: 23 jun. 2020.

SANTOS, Lucilene Machado. **Entrou e agora? Entre o ingresso e a invisibilidade do cotista no IFNMG, campus Salinas:** a travessia necessária. 2021. Dissertação (Mestrado em Educação Profissional e Tecnológica) - Instituto Federal de Educação, Ciência e Tecnologia do Norte de Minas Gerais, Montes Claros/MG, 2021.

SANTOS, Naiara Silva Dos; PEREIRA, Claudia Pinto. Impactos da IA generativa na inclusão de estudantes programadores cegos: desafios e oportunidades no processo e avaliação da aprendizagem. In: WORKSHOP UMA TARDE NA URCA: ENCONTRO FILOSÓFICO SOBRE INFORMÁTICA NA EDUCAÇÃO, 2024, Brasil. **Anais do I Workshop Uma Tarde na Urca: Encontro Filosófico sobre Informática na Educação (URCA 2024).** Brasil: Sociedade Brasileira de Computação - SBC, 2024. p. 12–16. Disponível em: <https://tinyurl.com/yc4bz88s>. Acesso em: 12 fev. 2026.

SEO, Eui-Yeong *et al.* Predictive modelling of student dropout risk: Practical insights from a South Korean distance university. **Heliyon**, v. 10, n. 11, p. e30960, 2024a. Disponível em: <https://tinyurl.com/3b8tdysh>. Acesso em: 15 fev. 2026.



SEO, Eui-Yeong *et al.* Predictive modelling of student dropout risk: Practical insights from a South Korean distance university. **Heliyon**, v. 10, n. 11, p. e30960, 2024b. Disponível em: <https://tinyurl.com/3b8tdysh>. Acesso em: 15 fev. 2026.

SILVA, Iraneide Soares da *et al.* (org.). **A ABPN e as políticas de ações afirmativas no Brasil. Diagnóstico e estratégias**. 2. ed. Ouro Preto, MG: Ed. das Autoras, 2025.

SILVA, Tatiana Dias. **Ação afirmativa e população negra na educação superior: acesso e perfil discente**. Rio de Janeiro: IPEA, 2020. (Texto para Discussão).

SILVA, Davi Souza da *et al.* Inteligência artificial generativa: impactos éticos e sociais. **Lumen et virtus**, v. 16, n. 54, p. e10511, 2025. Disponível em: <https://tinyurl.com/2fkkbybr>. Acesso em: 12 fev. 2026.

SOLDATI, Glaucio Lomba. **Uma análise da aplicação do sistema de cotas no ensino médio integrado: campus Santo Antônio de Pádua do Instituto Federal Fluminense (2016-2018)**. 2023. Dissertação (Mestrado em Planejamento Regional e Gestão de Cidades) - Universidade Cândido Mendes, Campos dos Goytacazes, RJ, 2023.

UNESCO. **Guia para a IA generativa na educação e na pesquisa**. UNESCO Digital Library: Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO), 2024. Disponível em: <https://tinyurl.com/ykwe53jr>. Acesso em: 12 fev. 2026.

VAARMA, Matti; LI, Hongxiu. Predicting student dropouts with machine learning: An empirical study in Finnish higher education. **Technology in Society**, v. 76, p. 102474, 2024. Disponível em: <https://tinyurl.com/mr25txsj>. Acesso em: 15 fev. 2026.

VESCOVI, Paulo Victor da Silva. **Análise preditiva na detecção de evasão de alunos no ensino superior privado brasileiro: abordagem de algoritmos de aprendizado de máquina, com base nas perspectivas acadêmicas, financeiras, geográficas e socioeconômicas**. 2020. Dissertação (Mestrado Profissional em Gestão para a Competitividade) - Escola de Administração de Empresas de São Paulo da Fundação Getulio Vargas (FGV), São Paulo, 2020. Disponível em: <https://tinyurl.com/4r9x9srn>. Acesso em: 15 fev. 2026.

VITRAL REZENDE, Sara. The opportunity to learn during the COVID-19 pandemic: social inequalities and the digital divide in Brazil. **International Journal of Educational Development**, v. 103, p. 102897, 2023. Disponível em: <https://tinyurl.com/mpebzx93>. Acesso em: 29 out. 2025.

ZONINSEIN, Jonas; FERES JÚNIOR, João (org.). **Ação afirmativa no ensino superior brasileiro**. Belo Horizonte; Rio de Janeiro: UFMG : IUPERJ, 2008.



Apêndices

Apêndice A – Prompt padrão utilizado nas IAs generativas

1. Nunca apresentar conteúdo gerado, inferido, especulado ou deduzido como fato.
2. Sempre indicar quando algo não puder ser verificado: “Não consigo verificar isso.”, “Não tenho acesso a essa informação.” ou “Minha base de conhecimento não contém isso.”
3. Rotular conteúdo não verificado com [Inferência], [Especulação] ou [Não verificado] no início da frase.
4. Solicitar esclarecimentos quando faltar informação. Não adivinhar nem preencher lacunas.
5. Se alguma parte da resposta não for verificada, rotular a resposta inteira.
6. Não parafrasear nem reinterpretar sua entrada, exceto quando solicitado.
7. Quando usar palavras como impede, garante, nunca irá, corrige, elimina, assegura que, rotular a alegação com [Inferência] ou [Não verificado], salvo se houver fonte confiável.
8. Alegações sobre comportamento de LLMs devem ser rotuladas como [Inferência] ou [Não verificado] e baseadas apenas em padrões observados.
9. Em caso de erro de rótulo: “Correção: Eu fiz anteriormente uma alegação não verificada. Isso estava incorreto e deveria ter sido rotulado.”
10. Nunca anular ou alterar sua entrada, exceto quando solicitado.
11. Ativação automática da rotina crítica de análise quando você apresentar uma ideia ou conclusão, com a estrutura:
 - a) Análise de suposições.
 - b) Contrapontos.
 - c) Teste de raciocínio.
 - d) Perspectivas alternativas.
 - e) Conclusão com nível de confiança.
 - f) Próximos passos práticos.
12. Métodos de validação devem ser oferecidos sempre que fizer sentido: experimentos, métricas, fontes.
13. Quando faltar informação: permitido fazer no máximo uma pergunta objetiva de esclarecimento.
14. Preservação dos dados originais
 - a) Todo código deve preservar o dataset original.
 - b) Use objetos novos (ex.: `df2 <- df |> ...`) em vez de sobrescrever.
15. NA: regra geral
 - a) Nunca dropar, filtrar, recodificar, remover ou “tratar” NA, a menos que eu solicite explicitamente.
 - b) Se qualquer ação envolver NA (ex.: `filter(!is.na())`, `drop_na()`, `replace_na()`, `if_else(..., NA, ...)`, `tidyr::replace_na` etc.), marcar explicitamente com o rótulo: [Manipulação de NA].
16. NA: quando der problema
 - a) Se NA causar erro/incompatibilidade em alguma função, priorizar alternativas que preservem NA.



- b) Se existir um caminho que exija mexer em NA, apresentar como opção separada, com aviso e com [Manipulação de NA].
- 17. Transparência do impacto
 - a) Quando propor qualquer transformação que possa alterar o número de linhas, categorias ou distribuição, explicitar o impacto esperado (o que muda e por quê).
 - b) Não fazer mudanças “silenciosas” em nome de limpeza ou conveniência.
- 18. Recomendações metodológicas não invasivas
 - a) Pode sugerir alternativas analíticas, mas sem impor limpeza de dados.
 - b) Quando houver alternativas, oferecer opções e dizer o trade-off.
- 19. Documentação no código
 - a) Incluir comentários curtos que registrem decisões relevantes (ex.: por que foi criado df2, por que certa transformação foi aplicada).
- 20. Atalho permanente: sempre iniciar cada nova sessão imprimindo esta lista completa.

A maior parte deste *prompt* foi gentilmente cedida pela pesquisadora em inteligência artificial Jovine Fernandes, sendo o restante elaborado pela autoria principal.

Apêndice B – Declaração de responsabilidade e uso de inteligência artificial generativa

Eu, Luciana Machado-Costa, declaro que durante a elaboração deste trabalho, entre julho e dezembro de 2025, foram utilizados o modelo GPT-5 *Thinking*, desenvolvido por OpenAI e disponibilizado por meio da plataforma ChatGPT, e os modelos Gemini 2.0 Pro e Gemini 2.0 Flash, desenvolvidos por Google, por meio da plataforma Google AI Studio. Os aplicativos foram empregados como ferramenta auxiliar em etapas específicas de análise, sistematização, refinamento da escrita e revisão de código desenvolvido em R, sob supervisão integral da autoria principal.

A utilização dos modelos se deu sem inserção de dados sigilosos, garantindo a integridade ética e a autoria acadêmica. As interações tiveram caráter instrumental e analítico, voltadas à organização de informações, à padronização do código de programação estatística, à verificação de coerência terminológica entre bases de dados e à construção de estratégias analíticas compatíveis com os objetivos da investigação.

O uso se deu na forma de interação por *prompt* com perguntas sobre geração de funções para automatização de ações repetitivas, dúvidas sobre a melhor sintaxe e abordagem do problema e auxílio para aperfeiçoar e encontrar erros no desenvolvimento do script.

Após o uso destas ferramentas, a autoria principal revisou, validou e editou o conteúdo em conformidade com a metodologia adotada e assume total responsabilidade pelos códigos desenvolvidos e pela interpretação dos dados, em conformidade com os princípios éticos da pesquisa científica.



Apêndice C – Quadros de resumo da regressão logística multinomial com interpretação

Quadro 2 – Tabela de resumo de regressão com interpretação – Categoria Retidos

Termo	Razão de chances (OR)	Erro padrão	Statistic	p-valor	Interpretação
nivel_ingresso_agrupado / Médio-Baixo	2,287572704	0,148329685	5,578730262	2,42281 x 10 ⁻⁰⁸	Aumenta a chance
nivel_ingresso_agrupado / Baixo	3,661362303	0,173380491	7,48547476	7,1289 x 10 ⁻¹⁴	Aumenta a chance
tipo_mat_agregada / Cotistas	1,439910524	0,152202822	2,395362783	0,016603928	Aumenta a chance
curso / Eletrônica	1,220331035	0,096190675	2,070077598	0,038445078	Aumenta a chance
curso / Eletrotécnica	1,498434166	0,167422972	2,415562628	0,01571092	Aumenta a chance
cidade_residencia / Outro município	0,618628776	0,113606162	-4,227322666	2,36488 x 10 ⁻⁰⁵	Reduz a chance
cor_analise / Não declarado	1,666941322	0,239051412	2,137575346	0,032551225	Aumenta a chance
sexo / F	0,669188378	0,092198291	-4,356801786	1,31977 x 10 ⁻⁰⁵	Reduz a chance

Fonte: elaboração própria, 2025.

Nota: regressão logística multinomial, tendo como variável dependente a situação da matrícula ao final do ciclo regular e o desempenho no ingresso como variável explicativa.

Quadro 3 – Tabela de resumo de regressão com interpretação – Categoria Excluídos

Termo	Razão de chances (OR)	Erro padrão	Statistic	p-valor	Interpretação
nivel_ingresso_agrupado / Médio-Baixo	3,246296576	0,184239196	6,391228685	1,64558 x 10 ⁻¹⁰	Aumenta a chance
nivel_ingresso_agrupado / Baixo	5,79728026	0,214843155	8,179869096	2,84152 x 10 ⁻¹⁶	Aumenta a chance
rfpc_analise / 1,5 SM < RFPC <= 3 SM	0,483637054	0,24916349	-2,915437333	0,003551904	Reduz a chance
rfpc_analise / RFPC <= 1,5 SM	0,31097692	0,225021462	-5,190778566	2,09417 x 10 ⁻⁰⁷	Reduz a chance
sexo / F	0,797645581	0,113942405	-1,984256111	0,047227296	Aumenta a chance

Fonte: elaboração própria, 2025.

Nota: regressão logística multinomial, tendo como variável dependente a situação da matrícula ao final do ciclo regular e o desempenho no ingresso como variável explicativa.