



FATORES ASSOCIADOS À CONCLUSÃO DO ENSINO TÉCNICO: UMA ABORDAGEM VIA REGRESSÃO LOGÍSTICA MULTINÍVEL

Pedro Henrique Feitosa dos Santos¹;
Fernanda Esperidião²;
José Ricardo de Santana³;
Thiago Henrique Carneiro Rios Lopes⁴.

Resumo: Os estudos em Educação Profissional e Tecnológica no país acompanham as discussões das demais modalidades de ensino, especialmente no que diz respeito a taxas de evasão. O objetivo do artigo é investigar quais fatores individuais, institucionais e locais se associam às chances de conclusão de alunos dos cursos técnicos da rede federal de EPT a partir de regressões de probabilidade logística multinível entre os anos de 2017 e 2021, uma vez que variáveis em nível municipal foram utilizadas como *proxy* para dinamismo econômico. Os resultados mostram que características como curso técnico integrado, eixo tecnológico, raça, idade, sexo e renda estão significativamente relacionadas às probabilidades de conclusão/evasão, enquanto variáveis de dinamismo econômico não apresentam significância estatística no modelo final.

Palavras-chave: Educação profissional e Tecnológica; Regressão Logística Multinível; Conclusão escolar; Plataforma Nilo Peçanha.

Código JEL: I21; R11; C25

Área Temática: Economias Regionais e Urbanas

FACTORS ASSOCIATED WITH THE COMPLETION OF TECHNICAL EDUCATION: A MULTILEVEL LOGISTIC REGRESSION APPROACH

Abstract: Studies in Vocational and Technological Education in Brazil follow discussions on other educational modalities, especially regarding dropout rates. This article aims to investigate which

¹ Mestrando no Programa Acadêmico de Pós-Graduação em Economia da Universidade Federal de Sergipe. E-mail: pedrofeitosa@ufs.br

² Programa de Pós-Graduação em Economia da Universidade Federal de Sergipe. E-mail: feresper@academico.ufs.br

³ Programa de Pós-Graduação em Economia da Universidade Federal de Sergipe. E-mail: jrsantana@academico.ufs.br

⁴ Programa de Pós-Graduação em Economia da Universidade Federal de Sergipe. E-mail: thiagorios@academico.ufs.br

individual, institutional, and local factors are associated with the chances of graduation for students in technical courses within the federal vocational and technological education network using multilevel logistic probability regressions between 2017 and 2021, since municipal-level variables were used as a proxy for economic dynamism. The results show that characteristics such as integrated technical courses, technological axis, race, age, sex, and income are significantly related to graduation/dropout probabilities, while economic dynamism variables do not show statistical significance in the final model.

Keywords: Vocational and Technological Education; Multilevel Logistic Regression; School Completion; Nilo Peçanha Platform.

1. INTRODUÇÃO

As bases da Teoria do Capital Humano destacam o papel fundamental da educação enquanto investimento em capacitação para o mercado de trabalho, com vistas a maiores retornos futuros, como defendia Schultz (1961). Becker (1993) formalizou a teoria e dividiu a capacitação em dois grupos: treinamento geral e específico, onde o segundo treinamento, eleva a produtividade marginal mais na empresa que o treinou do que nas demais, e apesar dos desembolsos, os treinamentos específicos criam efeitos distribuídos na produtividade.

Krugman (1991), ao tratar de externalidades pecuniárias, explica que trabalhadores com habilidades específicas, como os industriais, procuram reduzir as probabilidades de desemprego ao se estabelecerem em locais onde há várias empresas do mesmo ramo, formando um mercado de mão de obra comum, tanto pela ótica da oferta como da demanda.

Consonante ao entendimento de que educação, treinamento e demanda por mão de obra específica se interligam, o presente artigo tratará da Educação Profissional e Técnica – EPT – no Brasil, cujas bases no início do século XIX remontam à preparação para o trabalho.

É o que relata Moura (2010), com a inauguração do Colégio das Fábricas e da Escola dos aprendizes artífices, cujo objetivo era assistencialista e preparatório à indústria. Ao longo da história, as práticas educativas se conectaram com as transformações estruturais do país, onde se pode citar a reforma da educação profissional dos anos 90, que trouxe uma escola funcional ao mercado (MOURA, 2010). Nas décadas seguintes, novos dispositivos normativos provocaram alterações na EPT brasileira, como a possibilidade do Ensino Profissional aliado ao Ensino Médio e a ampliação desse ensino integrado por meio da criação dos Institutos Federais em 2008 (BRASIL, 2004, 2008).

A importância da EPT no país esteve, sobretudo, na meta 11 do Plano Nacional da Educação (PNE), cuja proposta era triplicar as matrículas de EPT de nível médio até 2024 (BRASIL, 2014), entretanto, apesar do crescimento de matrículas em EPT, a meta não foi alcançada. Os painéis da Fundação Itaú (2026) mostram que em 2024 apenas 49,7% da Meta 11 foi atingida.

Nos últimos anos, os empregos de baixa qualificação diminuíram rapidamente e os de alta qualificação aumentaram. Comparativamente, mais de 30% de jovens e adultos brasileiros não possuem educação secundária, contra a média de 15% entre os países da OCDE. Entre os cursos superiores, o Brasil possui 21% de diplomação, a OCDE possui 45%. Já na faixa entre 20 e 24 anos, 30% dos indivíduos não estão nem estudando, nem trabalhando, média superior à OCDE e de vários países da América Latina (OECD, 2023).

Ante o exposto, o objetivo do artigo é investigar quais fatores individuais, institucionais e locais se associam às chances de conclusão de alunos dos cursos técnicos da rede federal de EPT. A opção pela rede federal é respaldada pelos dados escassos da EPT brasileira, onde os microdados fornecidos pela Plataforma Nilo Peçanha – PNP - formam a base mais abrangente para esta pesquisa, além da relevância dos IFs nesse cenário.

Adicionalmente, buscou-se incorporar variáveis de dinamismo econômico, como emprego, renda e atividade industrial dos locais onde se situam as instituições de ensino, através de dados da Relação Anual de Informações Sociais e do Observatório DATAVIVA, com vistas a investigar se locais de mais dinamismo econômico estão positivamente relacionadas às taxas de conclusão de alunos.

Na análise quantitativa utilizou-se o modelo de regressão logística multinível, já que os indicadores econômicos são agregados para municípios. Devido a elevada correlação entre as variáveis de dinamismo, também se utilizou Análise de Componentes Principais – PCA – como ferramenta estatística de redução da dimensionalidade e de multicolinearidade.

Além dessa introdução e das considerações finais, o artigo contará com mais três seções. A segunda seção trará a revisão da literatura e a terceira e abordará o papel dos Institutos Federais na EPT. A terceira seção cuidará da exposição da base de dados e da metodologia utilizada, e a quarta apresentará a análise dos resultados.

2. EDUCAÇÃO, TERRITÓRIO E DESIGUALDADE

Nesta seção, discutem-se as bases teóricas da educação brasileira e internacional que sustentam a análise do fluxo educacional – sobretudo a conclusão e a evasão escolar, apresentando contribuições da literatura acerca do ensino básico, profissional, técnico e superior. Aborda-se ainda o papel dos Institutos Federais e sua relevância para a educação brasileira.

2.1 DEBATES SOBRE A EDUCAÇÃO

As discussões que cercam a educação brasileira concentram-se especialmente nos elevados índices de evasão da educação básica e superior e nas desigualdades sociais. Silva (2017) e Soares et al. (2011), exploram os desafios de garantir a permanência dos alunos, não apenas até a conclusão do ano letivo, mas de todo o seu processo formativo, relacionando o problema às elevadas taxas de abandono escolar.

Medeiros e Oliveira (2014) propõem relacionar as condições de distribuição espacial, sobretudo das características populacionais, com as desigualdades educacionais intra e entre regiões, constatando a alta descentralização na educação primária e secundária brasileira. Já Almeida e Firme (2023) buscaram verificar o impacto do aumento da escolaridade no nível de renda per capita dos municípios do Sudeste brasileiro.

No âmbito da EPT, os artigos se concentram especialmente na revisão da literatura, a fim de verificar os impactos de programas para a educação de jovens e adultos e as causas da evasão, que englobam causas pessoais e institucionais, falta de identificação com o curso e dificuldades em conciliar estudos e trabalho (CARMO et al., 2020; FERNANDES; SILVA, 2024; CUNHA; FILHO, 2021).

A literatura educacional brasileira permeia, em linhas gerais, o atraso educacional do país, as razões para as elevadas taxas de evasão e abandono e as desigualdades sociais e econômicas entre os territórios, tanto entre como dentro das regiões. É válido pontuar que há relativamente mais trabalhos voltados à educação básica e superior, que à EPT.

A literatura internacional da EPT se difere da brasileira, especialmente pelas disparidades nos estágios do desenvolvimento da educação. Wolf (2002) critica a noção de que quanto mais alto o capital humano, maior a produtividade, trazendo luz à qualidade da educação antes do aspecto quantitativo. Wheelahan e Leesa (2007) reforçam esse posicionamento ao afirmarem que a EPT na Austrália exclui os grupos desfavorecidos do chamado “conhecimento poderoso”, reduzindo esse modelo de ensino ao treinamento para a indústria.

Este artigo dialoga com os seus pares da literatura, inclusive, tomando por base um método estatístico amplamente utilizado na literatura da educação: o modelo de regressão logística. Dentre os trabalhos que lançaram mão desse método, podem ser citados Silva (2017), David e Chaym (2019), Moraes e Souza (2024) e Soares et al. (2011). As variáveis utilizadas neste artigo são compatíveis com a literatura, tais como sexo, idade, raça, região territorial e características institucionais.

2.2 O PAPEL DOS INSTITUTOS FEDERAIS NO ENSINO TÉCNICO E PROFISSIONALIZANTE

O nascimento dos Institutos Federais data de 1909, quando o então presidente Nilo Peçanha sancionou a lei nº 7.566, que instaurava a Escola de aprendizes artífices. Dentre os objetivos dos Institutos Federais, cabe destacar não apenas a articulação do ensino tecnológico e profissional com o mundo do trabalho, mas a geração de emprego e renda, atuando na difusão do conhecimento e da tecnologia. (BRASIL, 2008)

A Lei 11.892 assegura que uma das finalidades dos Institutos Federais é formar e qualificar indivíduos, de modo que estes contribuam com o desenvolvimento socioeconômico local, regional e nacional. Pacheco et al. (2010) atribui aos IFs, a responsabilidade de resgatar a cidadania enquanto forma indivíduos conscientes no mundo do trabalho, capazes de humanizá-lo.

Em 2005, foi anunciado o Plano de Expansão da Rede Federal em três etapas, as quais priorizavam a implantação de Escolas Federais em estados desprovidos, regiões periféricas de grandes centros urbanos e municípios do interior, objetivando cobrir o maior número de mesorregiões brasileiras e contar, até 2020, com 1000 unidades (BRASIL, 2009; 2011).

Consonante às finalidades dos IFs, Silva e Terra (2013) defende que o seu papel de promoção

ao bem-estar social e combate às desigualdades sociais regionais é cumprido à medida que estes se relacionam com os territórios nos quais estão implantados.

3. BASE DE DADOS E METODOLOGIA

Nesta seção, será apresentada a metodologia aplicada na análise e as bases de dados educacionais e econômicas selecionadas. As variáveis, majoritariamente qualitativas, servem a este artigo como parâmetros para investigação de possíveis padrões nas decisões de concluir e evadir dos cursos técnicos da rede federal de EPT.

3.1 METODOLOGIA

O artigo contará com Análise Exploratória de Dados em termos relativos, com fulcro na razão entre concluintes e evadidos, por regiões brasileiras e de acordo com características individuais, a saber, raça/cor, gênero, renda, idade, combinadas com características institucionais, como o eixo tecnológico do curso ao qual o aluno está matriculado e o tipo de oferta – curso integrado ao Ensino Médio ou subsequente.

Além da estatística descritiva, os dados também serão submetidos ao método estatístico quantitativo de regressão logística multinível, complementada por uma Análise de Componentes Principais (PCA). As próximas subseções apresentarão e justificarão os métodos quantitativos, logo após, o artigo detalhará as bases de dados.

3.1.1 O modelo *logit*

Em estatística é comum, especialmente nas ciências sociais, o uso de regressandos de natureza qualitativa, isto é, cujos resultados se restringem a 1 ou 0. Nesses cenários, é necessário o uso de modelos de resposta qualitativas (GUJARATI, 2011). Aqui, será utilizado o modelo de regressão logística (Logit), em que a probabilidade de Y aumenta quando X aumenta, mas a taxa de aumento é menor à medida que se aproxima dos limites 0 e 1, assumindo uma curva sigmoideal retratada por uma Função de Distribuição Acumulada (FDA) logística.

$$P(y = 1|x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)}} \quad (1)$$

O uso da base logarítmica permite a não linearidade e mesmo com valores de x tendendo a $+\infty$ ou $-\infty$, a probabilidade P estará comportada entre 0 e 1 (GUJARATI, 2011).

Finalmente, a adequação do modelo logit também deve ser revista, pois há variáveis exclusivas a cada estudante, mas também há variáveis comuns a estudantes agrupados em um mesmo local, como os indicadores socioeconômicos de renda, emprego e remuneração municipais. Surge como alternativa ao logit tradicional o logit multinível, que se diferencia pela inclusão de *clusters* no modelo, sobretudo, devido ao fato de que o logaritmo de chances pode mudar entre os agrupamentos, que aqui serão territoriais (SOMMET, 2017). Segue multinível geral apresentado no manual do pacote estatístico Stata (2021)

$$\Pr(y_{ij} = 1 | x_{ij}, u_j) = H(x_{ij} \beta + z_{ij} u_j) \quad (2)$$

onde Pr é a probabilidade de sucesso, H é a função logística, $x_{ij} \beta$ são os efeitos fixos e $z_{ij} u_j$ são os efeitos aleatórios. É importante explicar que as equações estimadas neste artigo consideram apenas o intercepto aleatório, ou seja, $z_{ij}=1$ e u_j é a constante atribuída aos indivíduos, de acordo com o *cluster* em que estão inseridos.

A utilização do modelo logístico propõe verificar a razão de chances de conclusão do curso associada às variáveis explicativas apresentadas na subseção anterior. Por isso, os coeficientes das variáveis serão apresentados em forma de *odds ratio*, ou seja, a cada aumento de uma unidade da variável explicativa, o coeficiente reflete o aumento/redução de chances de sucesso do evento de interesse na variável dependente.

De acordo com Wooldridge (2010), um bom método para lidar com qualquer tipo de heterocedasticidade em modelos de resposta binária é a utilização de erros-padrão robustos, amplamente utilizados em modelos *cross section* com muitas amostras, assim como este. Nos modelos dessa pesquisa, foi adotado o estimador Huber-White nativo do pacote estatístico, com ajuste para *clusters* municipais, que visa garantir a validade dos testes de hipótese.

Já o *LR Test* é a base para a qualidade de predição do modelo, pois avalia se o estimador β é consistente e assintoticamente normal. No Logit multinível, o *LR test* mensura a diferença entre o resultado real de conclusão/evasão observado e a probabilidade predita no modelo, logo, quanto menor a diferença, quão mais próximo de 0, melhor (WOOLDRIDGE, 2010).

$$\ell_i(\beta) = y_i \log[G(x_i\beta)] + (1 - y_i) \log[1 - G(x_i\beta)] \quad (3)$$

Os critérios de Akaike (AIC) e Bayesiano (BIC) derivam do logaritmo de máxima verossimilhança, os quais traduzem o ajuste do modelo ponderado pela sua complexidade, mas no primeiro, o log de verossimilhança é somado ao dobro dos seus $k+1$ preditores, enquanto no segundo, é somado aos seus $k+1$ preditores, ponderados pelo logaritmo natural do número de observações da amostra (PENNSYLVANIA STATE UNIVERSITY, 2026). Esses critérios premiam os modelos que possuem as menores diferenças de verossimilhança, dada a quantidade de variáveis explicativas e o tamanho da amostra. Ambos são essenciais para a escolha do modelo a ser analisado, mas deve-se ressaltar que o BIC é mais sensível ao tamanho da amostra, resultando em penalidade maior a cada preditor incluído no modelo.

3.1.2 A Análise de componentes principais (PCA)

A PCA consiste na transformação de um grupo de variáveis iniciais em outro grupo, com a mesma dimensão, mas retendo o máximo de respostas, em detrimento à variação total contida nos dados originais (VARELLA, 2007).

O método parte de um conjunto de variáveis, cuja variabilidade seja de interesse e que não seja útil analisar suas correlações e covariâncias. A alternativa é derivar novas variáveis que consigam concentrar a maior parte das informações das variâncias (JOLLIFFE, 2002).

O método para essa derivação envolve, além da matriz $X(n \times k)$ dos dados, uma função linear ponderada por um vetor de pesos α_{ij} , responsável por transformar as coordenadas dos pontos das variáveis originais em um plano que maximize a distância entre eles, originando o primeiro componente principal, que em suma, explicará a maior parte da variação (JOLLIFFE, 2002; MELO, 2024).

$$\alpha'_1 X = \alpha_{\{11\}x_1} + \alpha_{\{12\}x_2} + \dots + \alpha_{\{1p\}x_p} = \sum_{\{j=1\}}^p \alpha_{\{1j\}x_j} \quad (4)$$

Depois, repete-se o processo de procurar uma função linear $\alpha'_2 X$, com variância máxima, mas sem correlação com a função linear anterior, até o k -ésimo componente, sendo que, a cada componente acrescido, reduz-se o percentual de variabilidade explicada (JOLLIFFE, 2002)

3.2 BASES DE DADOS

Essa pesquisa avaliou quais fatores se relacionam estatisticamente com a probabilidade de um aluno de curso técnico da rede federal ser concluinte ou evadido. Para isso, foi utilizada a Plataforma Nilo Peçanha, observatório de informações da Rede Federal de Educação Profissional, Científica e Tecnológica, cujo objetivo é fornecer dados atualizados das unidades federais de EPT do Brasil (BRASIL, 2026), complementadas por indicadores relativos a indústria, intensidade tecnológica e mercado de trabalho, obtidos via Relação Anual de Informações Sociais (RAIS) e DataViva (Cedeplar – UFMG).

3.2.1 Plataforma Nilo Peçanha

Apesar da limitada literatura referente à EPT, é possível citar autores que utilizaram a Plataforma Nilo Peçanha em suas análises, como Oliveira et. al (2022) e Silva et. al (2025), todos com trabalhos voltados às causas da evasão escolar.

Para este artigo, foram extraídos microdados da eficiência acadêmica de alunos da rede federal de ensino para os anos de 2017, 2019 e 2021, a fim de analisar a evolução do perfil dos concluintes e possíveis tendências na estrutura educacional. A base incorpora dados de todo o país, com todos os tipos de curso ofertados nos IFs, mas aqui, o alvo será o curso técnico, dado o seu aparato básico para ingresso no mercado de trabalho.

O tratamento da base ocorreu com a seleção das variáveis de interesse, divididas aqui em duas vertentes: individual e institucional. A primeira vertente abrange informações pessoais do aluno, tais como sexo, etnia, idade e faixa de renda familiar. Já a vertente institucional, trata das características do curso no qual o aluno está matriculado: local da unidade de ensino (município e região), eixo tecnológico e tipo de oferta do curso técnico.

As amostras que continham dados faltantes foram removidas, assim como as amostras de alunos em curso, pois o foco é analisar as disparidades entre o perfil dos concluintes e dos evadidos no ano de referência; Os alunos em curso podem se enquadrar futuramente em ambas as categorias. A análise dos estudantes pela ótica da conclusão, parte da conjectura de que ao formar indivíduos com capacitação profissional e técnica, estes estarão aptos a atuar no mercado de trabalho como capital humano de nível mais qualificado.

Após o tratamento, a amostra manteve 116.342 observações, sendo 21.872 para o ano de 2017, 55.773 de 2019 e 38.697 de 2021. Dos indivíduos que permaneceram na amostra, 64,4% são concluintes, 36,3% são brancos, 47,5% são mulheres, 60,4% fazem o curso técnico integrado ou concomitante ao Ensino Médio, 48% estão matriculados em cursos técnicos do eixo 1 – que na próxima subseção será associado ao conceito de STEM (*Science, Technology, Engineering and Mathematics*) – e 61,6% estão concentrados nas regiões Nordeste e Sudeste, distribuição semelhante à configuração demográfica entre as regiões do país. O quadro 1 traz as variáveis da Plataforma Nilo Peçanha (2026) selecionadas para a pesquisa.

Quadro 1 - Definições das variáveis individuais / institucionais analisadas

Nome	Descrição	Tipo de variável	Tratamento
<i>concluintes</i>	Variável que distribui os alunos em categorias, de acordo com a situação das matrículas. Concluintes são os que estão com matrículas concluídas ou integralizadas, evadidos são os que estão em abandono, reprovação, desligamento, cancelamento, transferência interna ou externa	Dummy	Essa é a variável dependente do modelo. Se o aluno está na categoria “concluintes”, recebe valor 1, se está na categoria “evadidos” recebe 0. Se o aluno estiver “em curso”, é removido da amostra, pois não é possível assegurar que ele concluirá ou abandonará.
<i>branco</i>	Classificação do aluno matriculado segundo cor/raça	Dummy	Se o aluno se identifica como pessoa branca, recebe valor 1, se não, recebe valor 0.
<i>eixo</i>	Classificação por eixo tecnológico do curso EPT ao qual o estudante está matriculado.	Dummy	O presente trabalho propôs uma divisão de cursos entre STEM e não STEM a partir da classificação ISCED F-05 da União Europeia.
<i>idade</i>	Anos completos do aluno matriculado na data de ocorrência da situação da matrícula	Discreta	A variável não foi submetida a tratamento.

Quadro 1 - Definições das variáveis individuais / institucionais analisadas

(continuação)

<i>regiao</i>	Define em qual das cinco grandes regiões está localizada a unidade do IF.	Dummy	Optou-se em utilizar dummies, pois o contexto socioeconômico da região em que a unidade se localiza pode influenciar na escolha de concluir ou evadir.
<i>ren_fam</i>	Renda per capita mensal da família do estudante em termos de salário mínimo, dividida em faixas de renda.	Dummy	A escolha em tornar essa variável dummy se deu, especialmente, para capturar a relação de não linearidade do impacto entre as diferentes faixas de renda.
<i>mulher</i>	Define se o indivíduo é mulher ou não mulher.	Dummy	Procedimento padrão de transformação de variável qualitativa em dicotômica.
<i>integrado</i>	Tipo de oferta do curso: integrado/concomitante ao Ensino Médio ou subsequente ao Ensino Médio.	Dummy	Se o curso é integrado ou concomitante ao ensino médio, recebe valor 1, se não, recebe valor 0.

Fonte: elaboração do autor (2026), a partir de dados da Plataforma Nilo Peçanha (2026)

A variável da renda familiar do aluno, “ren_fam”, é dividida na base da plataforma em seis faixas, aqui será dividida em classes, sendo a classe 1 a de menor renda e a classe 6 de maior renda. Para facilitar a compreensão das classes na seção de análise de resultados, aqui são listadas as faixas de renda abrangidas por cada classe:

- Classe 1: de 0 a 0,5 salário mínimo por pessoa;
- Classe 2: mais de 0,5 a 1 salário mínimo por pessoa;
- Classe 3: mais de 1 a 1,5 salário mínimo por pessoa;
- Classe 4: mais de 1,5 a 2,5 salário mínimo por pessoa;
- Classe 5: mais de 2,5 a 3,5 salário mínimo por pessoa;
- Classe 6: mais de 3,5 salário mínimo por pessoa.

3.2.1.1 Definição da variável “eixo”

Aqui, vale ressaltar a definição da variável “eixo” (tecnológico): considerando o caráter qualitativo da identificação do eixo tecnológico dos cursos presentes no Catálogo Nacional de Cursos Técnicos (CNCT), foi necessária uma distinção entre os eixos técnicos para a delimitação dos eixos que assumiriam valor 1 e 0. A Resolução Nº 02/2024 estabelece que as atualizações no Catálogo devem seguir a Classificação Internacional Normalizada da Educação - CINE BRASIL - já no seu primeiro nível, o Eixo Tecnológico. Apesar de a Resolução ter sido publicada após o período analisado, vale destacar que em edições anteriores, o catálogo possuía os mesmos eixos tecnológicos.

A CINE Brasil, por sua vez, é um par da ISCED – *Internacional Standard Classification Education* - formalizada pela Unesco para parametrizar a análise de estatísticas educacionais, permitindo que os países sistematizem os seus dados, tanto nacionais como internacionais.

O banco de dados de Mulheres com educação superior em áreas STEM da Eurostat (2026) explicita que os cursos STEM abrangem as áreas 05, 06 e 07 da ISCED-F, a saber, Ciências Naturais, Matemática e Estatística, Tecnologias da Informação e da Comunicação e Engenharias, Produção e Construção, respectivamente.

Sendo assim, a pesquisa buscou as áreas equivalentes à CINE-F, nos eixos tecnológicos do Catálogo Nacional de Cursos Técnicos - CNCT. Dentre os eixos, foram escolhidos para compor o STEM: Controle e Processos Industriais, Informação e Comunicação, Infraestrutura e Produção Industrial. Os eixos de Ambiente e Saúde e Recursos naturais não foram considerados STEM, pois há

diversos subeixos distintos, que poderiam se distanciar da definição base da ISCED-F.

3.2.2 Variáveis proxy para Dinamismo Econômico

A análise dos fatores educacionais que refletem as probabilidades de conclusão do curso técnico pode perceber heterogeneidade entre os territórios, dadas as suas diferenças socioeconômicas, assim como aponta Corbucci (2016).

Esse estudo propôs relacionar as características educacionais não apenas ao contexto de emprego e renda, variáveis usualmente atreladas ao capital humano, mas também com as atividades que envolvem maior nível de especialização no trabalho. Para isto, foram recolhidas variáveis de atividade industrial, incidência de P&D (Pesquisa e Desenvolvimento) e complexidade econômica dos empregos. O quadro 2, apresenta as variáveis utilizadas como *proxy* para o nível de dinamismo econômico dos municípios onde se situam unidades da rede federal de ensino técnico estudadas.

Quadro 2 - Definições das variáveis econômicas analisadas

Nome	Descrição	Tipo de variável	Tratamento	Fonte
<i>Lvab_ind</i>	log do valor adicionado bruto das indústrias por município, a preços correntes, em milhares de reais	Variável quantitativa contínua transformada.	Normalização via cálculo do log, para reduzir assimetria entre as observações.	Relação Anual de Informações Sociais (RAIS)
<i>Lvinc_tec</i>	log da frequência de vínculos ativos em atividades de media-alta e alta intensidade tecnológica, por município, de acordo com a nova classificação da OCDE.	Variável quantitativa contínua transformada.	Normalização via cálculo do log, para linearizar a relação. Para municípios em que não há vínculos formais de média-alta e alta intensidade tecnológica, foi mantido valor 0.	Relação Anual de Informações Sociais (RAIS); MORCEIRO (2019).
<i>Ice_emp</i>	índice de complexidade econômica dos empregos, por município, extraído do dataviva.	Variável quantitativa contínua.	Se o aluno se identifica como pessoa branca, recebe valor 1, se não, recebe valor 0.	DataViva
<i>L_emp</i>	quantidade de empregos formais, de todas as atividades econômicas, por município.	Variável quantitativa contínua transformada.	Normalização via cálculo do log, para reduzir assimetria entre as observações.	DataViva
<i>Ren_men_med</i>	Renda mensal média, por município, extraído do Dataviva.	Variável quantitativa contínua.	Não houve tratamento.	DataViva

Fonte: elaboração do autor (2026), a partir de dados brutos da RAIS (2026) e da plataforma DataViva (2026).

As variáveis selecionadas para representação de dinamismo econômico foram submetidas ao método de PCA (Análise de Componentes Principais), devido à elevada correlação entre as mesmas, buscando maximizar a variância das respostas fornecidas. Constatou-se que no primeiro componente concentra-se o fator que carrega a maior parte das informações referentes ao contexto de atividade econômica industrial e tecnológica, dinamismo de mercado de trabalho e nível médio de renda.

Outro ponto crucial para a análise dessas variáveis foi a distribuição territorial, pois, estão agregadas para municípios, já os dados educacionais aos quais elas serão associadas, estão em nível de unidade de ensino e, portanto, vários alunos da amostra podem pertencer a um mesmo município. Essa característica da pesquisa demandou o experimento de um modelo de regressão multinível, uma vez

que os dados estão hierarquizados territorialmente.

4. RESULTADOS E DISCUSSÕES

Aqui, apresentam-se e discutem-se os resultados obtidos na pesquisa. No primeiro momento, há a Análise Exploratória de dados, na qual se realiza o mapeamento das razões de conclusão e evasão entre as regiões, bem como a análise do comportamento das variáveis educacionais frente às características individuais dos estudantes ao longo do período selecionado. Na sequência, detalham-se os resultados do modelo de regressão logística, com a identificação do melhor ajuste e progredindo à análise estatística dos coeficientes determinantes das probabilidades de conclusão.

4,1 ANÁLISE EXPLORATÓRIA DE DADOS

O primeiro ponto de análise das amostras é a variação entre a razão de concluintes e evadidos por região entre os três anos: a região centro-oeste obteve maior proporção conclusão/evasão em 2017, mas, dois anos depois, apresentou o menor valor entre todas. O oposto ocorreu no Sudeste: segunda menor razão em 2017 e a melhor razão em 2021.

A significativa variação na proporção de concluintes e sua oscilação de médias gerais entre regiões ao longo do tempo, suscita a inclusão da variável binária de ano como controle no modelo, a fim de que os coeficientes não ignorem as diferenças entre os anos da amostra.

Outras variáveis dignas de atenção são “eixo” e “integrado”. Há algumas características individuais dos concluintes que se assemelham nos dois grupos: nos cursos classificados como STEM, em todos os anos, mulheres e brancos apresentaram maior percentual de conclusão, o mesmo vale para os cursos integrados nos anos de 2019 e 2021. Um contraponto está em 2017, onde não-brancos alcançaram maior percentual de conclusão tanto em cursos integrados, como em subsequentes.

Ao longo dos três anos, os perfis de renda familiar dos alunos também é semelhante: as faixas de renda mais elevadas estão entre os concluintes pertencentes aos cursos dos eixos STEM. Esses dados podem indicar que o acesso a cursos de maior viés técnico, bem como a sua conclusão é mais comum a indivíduos com renda familiar maior. Um fato interessante percebido em 2017 e 2019 é que a renda familiar dos evadidos em cursos subsequentes é maior que a dos concluintes, levantando uma suspeita de que indivíduos com maior segurança financeira podem abrir mão de maior qualificação técnica e profissional.

Por outro lado, é importante ressaltar que, à exceção de 2017, as taxas de conclusão nos cursos que não se enquadram como STEM são maiores que as taxas de conclusão em STEM. Ao relacionar as duas variáveis, obtém-se que a menor proporção de concluintes nos cursos STEM é encontrada na modalidade subsequente. Em 2019, do total da amostra, apenas 39,5% dos indivíduos matriculados em STEM eram concluintes, os outros 60,5% estavam em evasão. Esses resultados podem estar relacionados a fatores como o *gap* entre a conclusão do Ensino Médio e o início do curso técnico, menor disponibilidade de tempo para os estudos ou até o reflexo de bases educacionais defasadas.

Uma constatação que corrobora com o parágrafo anterior é que, a média de idade dos estudantes na modalidade subsequente é naturalmente maior que a dos estudantes na modalidade integrada, porém, a idade dos concluintes também é relativamente menor, estejam eles em cursos STEM ou não STEM, bem como na modalidade integrado. Em 2021, por exemplo, a média de idade dos concluintes frente aos evadidos é 4 anos menor em cursos STEM e 3, em cursos não STEM.

Aqui cabe um destaque para uma possível relação negativa entre idade e qualificação, especialmente se atendida a hipótese de que a cada um ano acrescido, a necessidade de acessar ao mercado de trabalho, independente do grau de qualificação, aumenta.

No mais, os dados comparativos entre concluintes e evadidos mostram que em 2019, houve mais evasão que conclusão nos cursos subsequentes, reforçando uma ideia de que os cursos técnicos integrados ou concomitantes ao Ensino Médio tendem a reter mais os estudantes.

4.2 RESULTADOS DO MODELO LOGIT

Conforme introduzido na seção 3, esse trabalho propôs a estimação de modelo econométrico de função logística - dada a composição dicotômica da variável dependente. O primeiro passo da análise foi definir qual o modelo mais adequado para o estudo. O quadro a seguir traz as diferentes abordagens tratadas e as informações de ajuste do modelo, como o Log de verossimilhança, critérios de Akaike e Bayesian e a Área sob a curva ROC:

Tabela 1 – Comparação de Modelos Estatísticos – segundo critérios de ajuste e especificação

Teste/variável	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6
Efeitos Aleatórios	Não	Não	Sim, por município.	Sim, por município.	Sim, por município.	Sim, por município.
Macrorregião	Sim	Sim	Sim	Não	Não	Não
Dinamismo econômico	Não	Sim	Sim	Sim	Sim, PCA.	Sim, PCA decomposta.
LR Test	-70.784,46	-70.762,21	-67.566,35	-67.569,94	-67.574,65	-68.350,04
AIC	141.602,90	141.568,40	135.178,70	135.177,90	135.179,30	135.169,30
BIC	141.767,20	141.781	135.401	135.361,50	135.324,30	135.324
AUC-ROC	0,6762	0,6759	0,7317	0,7317	0,7317	0,7317

Fonte: elaboração do autor (2026), com base nos dados da PNP/RAIS/DataViva (2017-2021)

Os dois primeiros modelos estimados são logit tradicionais, assim como o mencionado na equação 1. Frente aos demais modelos que incluem efeitos aleatórios, os logit tradicionais são inferiores tanto em ajuste, como em poder explicativo, por essa razão foram descartados.

Conforme apresentado na seção 3, inicialmente foi incluída uma variável que diferencia as macrorregiões brasileiras, contudo, as variáveis econômicas foram extraídas a nível municipal. Mediante o levantamento, se fez necessário definir qual a especificação da estrutura de variância e identificar a estrutura hierárquica que melhor captura a dependência espacial dos dados. Visando compreender qual a associação entre as unidades territoriais e as conclusões/evasões, foram estimados dois modelos logit multinível nulos, isto é, modelos restritos com apenas a variável dependente “concluintes” no primeiro nível e as respectivas variáveis geográficas no segundo nível.

Os testes demonstraram ampla superioridade da variável municipal frente a regional, com coeficiente de correlação intra-classe de 0,1595 para o modelo nulo 1, contra 0,0036 do modelo nulo 2. Um possível motivo para o melhor ajuste a nível municipal é a heterogeneidade dentro das regiões, o que dificulta sua eficácia como agrupamento. O município, por sua vez, tende a apresentar significativa correlação de vizinhança, que demanda efeitos aleatórios para correção de possíveis autocorrelações espaciais nos resíduos.

Os modelos de regressão multinível (modelos 3 a 6), também chamaram atenção por apresentar melhor valor de área sob a curva ROC, índice que de acordo com Lieli e Hsu (2018), testa a hipótese nula de que a área AUC é igual 0,5 e os preditores de resposta binária não possuem desempenho superior a um mero chute. Os resultados apontam que os preditores atingiram o limite de sensibilidade preditiva de conclusão com os dados disponíveis.

Vale destacar que o AIC dos modelos 3, 4 e 5 foram semelhantes, com o AIC do modelo 5 levemente superior, onde há chances de que a agregação de cinco variáveis explicativas em um único score, acarretou em perda de informações estatisticamente relevantes, que mesmo com a redução de parâmetros não foi compensada.

Por fim, o modelo 6 se diferenciou dos seus antecessores por seguir Mundlak (1978), dividindo o componente principal em dois índices, um que captura o dinamismo econômico médio dos municípios ao longo do período observado e outro que captura os desvios do score de dinamismo a cada ano, frente ao seu valor médio.

O ponto positivo dessa distinção é verificar quanto da razão de chances de conclusão é combinada com a estrutura econômica do município que sedia a unidade de ensino e quanto é combinado com as mudanças de dinamismo desse município entre os períodos. Esse mecanismo pode reduzir a endogeneidade entre as variáveis do nível 1 que se correlacionam com os erros da variável do nível 2.

Apesar de contar com uma variável a mais que seu imediato anterior, o modelo 6 não foi penalizado nos critérios de informação, pelo contrário, obteve os menores AIC e BIC, demonstrando melhora em verossimilhança. Também manteve poder discriminatório de 0,7317 no AUC-ROC, além de tratar os vieses de tempo e espaço dos modelos anteriores. Esses atributos indicam que a segmentação do componente principal não somente é aceitável, como recomendada, ainda que seus resultados não sustentem caráter explicativo.

Focando nas diferenças entre os modelos 5 e 6, é de suma importância relatar que o componente principal do modelo 5 rejeita a hipótese nula a 5%, já os dois componentes do modelo 6, não. Contudo, o melhor ajuste do modelo 6 revela que a significância da variável de dinamismo econômico do modelo anterior era fruto de um viés de especificação.

Os valores dos coeficientes também mudaram substancialmente, pois no quinto modelo, estar em um município mais dinâmico se relacionava a 5% mais chances de conclusão, resultado pouco expressivo. Após o cálculo de médias e desvio, o coeficiente de razões de chances de conclusão mediante o fator conjuntural do dinamismo econômico salta para 17,1%, número expressivo, porém, não significativo (p-valor 0,123), indicando que o dinamismo não seja parte fundamental na decisão de concluir ou não o curso técnico, ou ainda, a baixa variação temporal não foi capaz de sustentar o coeficiente com precisão.

Ante o exposto, serão discutidos os resultados do modelo 6, pois, na base em estudo há mais de 116 mil amostras e - dependendo da disposição do modelo - até 23 variáveis – evidenciando a importância de um bom critério BIC e um modelo parcimonioso e robusto, atributos em que o último modelo se destacou frente aos demais.

Cabe aqui o adendo de que a comparação entre os dois últimos modelos pode representar um panorama útil para essa pesquisa: a economia do local onde o indivíduo estuda tende a influenciar em sua retenção, evasão ou conclusão do curso técnico, porém, essa decisão também é envolta por muitas outras variáveis não observadas dos mais diversos contextos, fato que acaba reduzindo o poder de explicação, seja da estrutura, seja da conjuntura econômica local. A tabela 2 traz os resultados do modelo logístico hierárquico 6, cujas variáveis de proxy de dinamismo econômico foram convertidas em um componente principal, subdividido em dois índices.

Tabela 2 – Resultados de Modelo Logit multinível com efeitos aleatórios por município

Variáveis Independentes	Odds Ratio	Erro Padrão (Robust)	P-valor
Branco	1,136***	(0,023)	0,000
Mulher	1,202***	(0,023)	0,000
Idade	0,980***	(0,003)	0,000
Renda Familiar (Ref: Classe 1)			
Classe 2	1,059**	(0,029)	0,035
Classe 3	1,048	(0,035)	0,170
Classe 4	1,156***	(0,048)	0,001
Classe 5	1,181***	(0,055)	0,000
Classe 6	1,245***	(0,066)	0,000
Ano (Ref: 2017)			
2019	0,583***	(0,053)	0,000
2021	1,110	(0,118)	0,330
Eixo Tecnológico (STEM)	0,778***	(0,033)	0,000
Ensino Integrado	2,393***	(0,177)	0,000
Estrutura Econômica (PCA)	1,023	(0,018)	0,196
Conjuntura econômica (PCA)	1,171	(0,121)	0,123
Constante	2,454***	(0,250)	0,000
Efeitos Aleatórios			
Var(Cons) Município	0,439	(0,034)	—

Tabela 2 – Resultados de Modelo Logit multinível com efeitos aleatórios por município
(Continuação)

---	---	---	---
Estatísticas de Ajuste			
Observações	116.342		
Grupos (Municípios)	558		
Log-pseudolikelihood	-67.568,67		
AUC-ROC	0,7317		

Fonte: elaboração do autor (2026), com base nos dados da PNP/RAIS/DataViva (2017-2021)

Nota: *** significativo a 1%; ** significativo a 5%; * significativo a 10%

Em termos gerais, pode se afirmar que os coeficientes odds ratio calculados assumem valores consonantes às sinalizações da Análise Exploratória dos dados e trazem significância estatística ao modelo logístico de probabilidade.

Pessoas brancas e mulheres possuem aproximadamente 14% e 20% mais chances de concluir o curso técnico. Já no quesito idade, cada ano a mais de vida pode refletir menos 2% de chance de conclusão. Os resultados convergem com os encontrados em Soares et al. (2011) para a variável de sexo, mas divergem no que diz respeito à idade, assim como divergem de David e Chaym (2019), onde a chance de evasão é maior para o sexo feminino.

As variáveis de faixas de renda familiar apontam um salto nas razões de chances de conclusão entre a segunda faixa (6%) e a quarta faixa (16%). À medida que a renda cresce a porcentagem de probabilidade de conclusão também cresce, resultado que condiz com a literatura acadêmica atual (ALMEIDA; FIRME, 2023).

As variáveis referentes aos tipos de cursos em que os estudantes estão matriculados sugerem as seguintes tendências: alunos em cursos STEM possuem em média 22% menos chances de conclusão do Ensino Técnico, reforçando a ideia de que cursos inseridos nesse eixo tecnológico podem ser mais complexos, exigentes ou até mais sensíveis a demandas do mercado. Resultados semelhantes foram aferidos por David e Chaym (2019), onde a probabilidade de evasão em cursos de Engenharia de produção era elevada.

Já a informação do tipo do curso, carrega a maior razão de chance com a relação conclusão/evasão. Alunos em cursos integrados têm em média, 139% mais chances de concluir que alunos de cursos subsequentes. Os motivos para essa disparidade podem abranger desde a característica central do modelo do curso, que vincula a capacitação técnica à conclusão da educação básica, prevenindo um ingresso precoce no mercado de trabalho e auxiliando na fixação dos conhecimentos em ambas as esferas, até o perfil geral dos estudantes dessa modalidade, mais jovens e já incluídos no ecossistema da educação.

A relação entre os três períodos não é linear, pois em 2019 a razão de chances para conclusão era 42% menor que no ano base, 2017, enquanto em 2021 havia de 11% mais de chances de conclusão que em 2017. O elevado coeficiente para 2021, ainda que sem significância estatística, pode ser justificado pelas recomendações do Conselho Nacional de Educação (CNE), que sugeriu avaliações mais equilibradas, levando em consideração as diferentes situações dos alunos, especialmente após o período pandêmico (BRASIL, 2020).

Para além da escolha entre município e macrorregião como segundo nível do modelo, é válido observar a variância do efeito aleatório entre municípios. O valor de 0,439 demonstra que mesmo considerando as demais variáveis explicativas, o local da unidade de ensino pode ser associado com as probabilidades de conclusão ou evasão do estudante.

Ainda sobre a distinção entre os municípios, o modelo nulo 1 calculado para o nível municipal estava relacionado a praticamente 16% das razões de chance de conclusão do Ensino Técnico, já no modelo 6, a variável de efeitos relativos se relaciona a apenas 12% das razões de chance de conclusão, onde se intui que aproximadamente 25% da associação estatística entre a probabilidade de conclusão e o município de estudo é atrelada ao dinamismo econômico da cidade.

Os resultados permitem conjecturar sobre como as disparidades raciais e de renda, a idade e as bases

educacionais fragilizadas podem atuar como barreiras na qualificação de capital humano, especialmente em cursos de maior complexidade e intensidade tecnológica. Porém, não é possível afirmar que a estrutura econômica municipal se relaciona estatisticamente com o sucesso da formação do capital humano, fortalecendo a hipótese de que os fatores individuais e institucionais prevalecem nos resultados de fluxo educacional da EPT.

5. CONSIDERAÇÕES FINAIS

Mediante os baixos índices de formação educacional no Brasil e elevados índices de evasão, se comparados a outros países, essa pesquisa buscou investigar quais fatores individuais, institucionais e locais se associavam às chances de conclusão de alunos dos cursos técnicos da rede federal de EPT.

Através da Análise Exploratória de Dados, notou-se que, em linhas gerais, cursos técnicos integrados e de eixos não-STEM concentram a maior parcela dos concluintes. Para cursos agrupados no eixo STEM, as faixas de renda dos concluintes são superiores às dos evadidos e o percentual de mulheres e brancos em conclusão também supera o de homens e não-brancos.

Por outro lado, no STEM há menor percentual de conclusão, esse número reduz ainda mais em cursos do tipo subsequente, como evidenciado no ano de 2019, onde houve 60,5% de evasão. Outro destaque é que a média de idade dos concluintes é inferior, em ambos os grupos de eixo tecnológico e dentro dos cursos integrados.

Através do modelo de regressão logística multinível, obteve-se que classificações individuais, como de gênero, raça, renda e idade foram significativas para o modelo, onde indivíduos brancos possuem em média 13,6% mais chances de conclusão e indivíduos da maior faixa de renda familiar possuem 24,5% mais chances de conclusão que indivíduos da menor faixa.

Variáveis referentes ao eixo tecnológico do curso técnico e o tipo de oferta também foram estatisticamente significativas. Essa última variável apresentou probabilidade relativa de chance em cursos integrados muito superior às demais, com odds ratio de 2,139.

Já as variáveis de dinamismo econômico, não demonstraram significância estatística de relação com a variável dependente. A baixa variabilidade temporal e os diversos fatores externos não observados que envolvem a conclusão são candidatos a razões para esse resultado.

O artigo contribui com um modelo de probabilidade logística sob a ótica das conclusões de curso e as relaciona ao dinamismo econômico local, evidenciando que disparidades entre indivíduos – e, conseqüentemente, entre os cursos nos quais estão matriculados e a modalidade de oferta – são significativamente relacionados às probabilidades de conclusão de cursos EPT. A partir desses resultados, também é possível formular políticas educacionais que fortaleçam cursos integrados ao Ensino Médio, com atenção especial a cursos de ciências exatas e tecnologia, por exemplo, que demandem nivelamento pedagógico. Não esquecendo políticas voltadas a redução de desigualdades sociais e com incentivo à permanência de estudantes com mais idade, que já trabalham e se matriculam nos cursos técnicos subsequentes.

O artigo possui as seguintes limitações: não captura se os indivíduos concluintes de fato estarão inseridos no mercado de trabalho e se atuarão na área de formação, não consegue captar se a absorção do aluno concluinte se dará no mercado de trabalho da localidade onde ele estudou ou se há migração de capital humano, além de não considerar se o profissional formado em determinado eixo atuará, de fato, em sua área de formação.

Trabalhos futuros podem alterar a unidade geográfica de análise para regiões intermediárias, almejando capturar efeitos de transbordamento dos IFs, bem como estender o recorte temporal da amostra, com foco em estimativas mais precisas. Por fim, é útil incluir em trabalhos futuros, dados de EPT das redes estadual e privada de ensino, bem como experimentar outras variáveis explicativas de dinamismo econômico, de modo a verificar se há melhorias no poder explicativo do modelo.

A Educação Profissional e Tecnológica no país ainda não atingiu suas metas de adesão e retenção de alunos, mas, a redução das disparidades raciais e de renda, combinada a reforços na qualidade do ensino básico, juntamente à intensificação de oferta de cursos na modalidade integrada ao Ensino Médio, é um possível caminho de avanço da EPT.

REFERÊNCIAS

ALMEIDA, Leonardo Alexandrino de; FIRME, Vinícius de Azevedo Couto. **Educação e crescimento econômico: um estudo econométrico-espacial para os municípios do sudeste brasileiro**. 2023. Disponível em: <https://repositorio.ipea.gov.br/entities/publication/ed49e3cd-beef-490f-a8a0-952f81b7a817>. Acesso em 17 fev. 2026.

BECKER, G. S. **Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education**. 3rd ed. Chicago: University of Chicago Press, 1993.

BRASIL. **Lei nº 11.892, de 29 de dezembro de 2008**. Institui a Rede Federal de Educação Profissional, Científica e Tecnológica, cria os Institutos Federais de Educação, Ciência e Tecnologia, e dá outras providências. Brasília: Diário Oficial da União, 2008. Disponível: https://www.planalto.gov.br/ccivil_03/_ato2007-2010/2008/lei/l11892.htm. Acesso em 17 fev. 2026.

BRASIL. **Lei nº 13.005, de 25 de junho de 2014**. Aprova o Plano Nacional de Educação - PNE e dá outras providências. Brasília, DF: Presidência da República, [2014]. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l13005.htm . Acesso em: 14 fev. 2026.

BRASIL. Ministério da Educação. **CNE aprova diretrizes para escolas durante a pandemia**. Brasília: MEC, 28 abr. 2020. Disponível em: <https://portal.mec.gov.br/ultimas-noticias/12-aco-es-programas-e-projetos-637152388/89051-cne-aprova-diretrizes-para-escolas-durante-a-pandemia>. Acesso em: 16 fev. 2026.

BRASIL. Ministério da Educação. **Plataforma Nilo Peçanha**: indicadores da rede federal de educação profissional, científica e tecnológica. Brasília: MEC, [202-]. Disponível em: <https://www.gov.br/mec/pt-br/pnp>. Acesso em: 13 fev. 2026.

BRASIL. Ministério da Educação. Secretaria de Educação Profissional e Tecnológica. **Plano de expansão da Rede Federal de Educação Tecnológica – Fase II**. Brasília-DF, [200-]. Disponível em: <https://portal.mec.gov.br/setec-programas-e-aco-es/expansao-da-rede-federal>. Acesso em: 17 fev. 2026.

_____. Ministério do Trabalho e Emprego. **Relação Anual de Informações Sociais (RAIS)**: base de dados. Brasília: MTE, 2026. Disponível em: <https://bi.mte.gov.br/bgcaged/login.php>. Acesso em: 17 fev. 2026.

_____. **Prestação de Contas Ordinária Anual**: relatório de gestão 2010. Brasília-DF, 2011. Disponível em: https://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=8006-relatorio-gestao-2010-setec-versaofinal-cgu-pdf&category_slug=maio-2011-pdf&Itemid=30192. Acesso em: 17 fev. 2026.

CARMO, Ana Carolina Rigoni; DE AMORIM, Elizabeth de Jesus Moreira; DOS REMEDIOS, Sâmia Elene Lobato. O PROEJA como modalidade articulada à EPT: uma análise sobre evasão escolar. **Cadernos de Educação Básica**, v. 5, n. 2, p. 187-206, 2020. DOI: <https://doi.org/10.33025/ceb.v5i2.2785>

CORBUCCI, Paulo Roberto. **Desigualdades no acesso dos jovens brasileiros à educação superior**. Dimensões da experiência juvenil brasileira e novos desafios às políticas públicas. Brasília: Ipea, 2016. Disponível em: https://portalantigo.ipea.gov.br/agencia/images/stories/PDFs/livros/livros/livro_dimensoes_miolo_cap05.pdf. Acesso em: 17 fev. 2026.

CUNHA, Fabrício William da; LIMA FILHO, Adalberon Moreira de. **Revisão bibliográfica das pesquisas sobre evasão escolar na Educação Profissional e Tecnológica no Brasil**. 2021.

Disponível em: <http://repositorio.ufc.br/handle/riufc/65658>. Acesso em: 17 fev. 2026.

DATAVIVA. **Cedeplar, UFMG**, 2026. Disponível em: <https://legacy.dataviva.info/pt/rankings/>. Acesso em: 17 fev. 2026.

DAVID, Lamartine Moreira Lima; CHAYM, Carlos Dias. Evasão universitária: um modelo para diagnóstico e gerenciamento de instituições de ensino superior. **Revista de Administração IMED**, v. 9, n. 1, p. 167-186, 2019. DOI: <https://doi.org/10.18256/2237-7956.2019.v9i1.3198>.

DE MORAIS, Diogo Martins Gonçalves; DE SOUZA, Dallas Kelson Francisco. Modelos preditivos da evasão de estudantes: comparativo entre regressão logística, redes neurais e algoritmos genéticos. **South American Development Society Journal**, v. 10, n. 30, p. 182-182, 2024. DOI: <https://doi.org/10.24325/issn.2446-5763.v10i30p182-195>

EUROSTAT. **Female tertiary education graduates in STEM education fields**. [S. l.]: Eurostat, 2026. Disponível em: <https://ec.europa.eu/eurostat/databrowser/view/tps00217/default/table?lang=en>. Acesso em: 17 fev. 2026.

FERNANDES, Thiago; DA SILVA, Elizabete Maria. Desafios ocultos: o que leva estudantes a abandonarem a educação profissional e tecnológica no Brasil? **Revista Educação, Cultura e Sociedade**, v. 14, n. 2, p. 148-160, 2024. DOI: <https://doi.org/10.30681/ecs.v14i2.12931>.

FUNDAÇÃO ITAÚ. **Matrículas da EPTNM**: painel de dados. São Paulo: Observatório Fundação Itaú, [202-]. Disponível em: <https://fundacaoitau.org.br/observatorio/dados/paineis-dados/matriculas-da-eptnm>. Acesso em: 14 fev. 2026.

GUJARATI, Damodar N.; PORTER, Dawn C. **Econometria básica**. 5. ed. Porto Alegre: AMGH, 2011.

JOLLIFFE, I. T. **Principal Component Analysis**. 2. ed. New York: Springer, 2002.

KRUGMAN, Paul. Increasing returns and economic geography. **Journal of Political Economy**, v. 99, n. 3, p. 483-499, 1991.

LIELI, Robert P.; HSU, Yu-Chin. Using the area under an estimated ROC curve to test the adequacy of binary predictors. **Journal of Nonparametric Statistics**, v. 31, n. 1, p. 100-130, 2019. DOI: 10.1080/10485252.2018.1537440.

MEDEIROS, Marcelo; OLIVEIRA, Luís Felipe Batista de. Desigualdades regionais em educação:: potencial de convergência. **Sociedade e Estado**, [S. l.], v. 29, n. 2, p. 561–585, 2014. Disponível em: <https://periodicos.unb.br/index.php/sociedade/article/view/5901>. Acesso em: 17 fev. 2026.

MELO, Felipe Gondim. **Determinantes da economia informal**: uma análise de clusterização e componentes principais em países desenvolvidos e em desenvolvimento. 2024. Disponível em: <http://repositorio.ufc.br/handle/riufc/77313>. Acesso em: 17 fev. 2026.

MORCEIRO, Paulo César. Nova classificação de intensidade tecnológica da OCDE e a posição do Brasil. **Setor Externo**, v. 8, 2019.

MOURA, Dante Henrique. A Relação entre a educação profissional e a educação básica na CONAE 2010: possibilidades e limites para a construção do novo Plano Nacional de Educação. **Educação & Sociedade**, v. 31, p. 875-894, 2010. DOI:<https://doi.org/10.1590/S0101-73302010000300012>

MUNDLAK, Yair. On the Pooling of Time Series and Cross Section Data. **Econometrica**, v. 46, n. 1, p. 69-85, 1978.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **Engaging Employers in Vocational Education and Training in Brazil**: learning from international practices. Paris: OECD Publishing, 2023. Disponível em: https://www.oecd.org/en/publications/engaging-employers-in-vocational-education-and-training-in-brazil_d76a2fe6-en.html. Acesso em: 17 fev. 2026.

OLIVEIRA, Isleimar S.; MEDEIROS, Francisco Petrônio A.; ANDRADE, Fabio G.. Seleção de Atributos para Classificadores de Evasão Escolar com Dados da Plataforma Nilo Peçanha. In: WORKSHOP DE APLICAÇÕES PRÁTICAS DE LEARNING ANALYTICS E INTELIGÊNCIA ARTIFICIAL NO BRASIL (WAPLA), 1. , 2022, Manaus. **Anais [...]**. Porto Alegre: Sociedade Brasileira de Computação, 2022 . p. 30-39. DOI: <https://doi.org/10.5753/wapla.2022.226769>.

PENNSYLVANIA STATE UNIVERSITY. **Information Criteria and PRESS**. State College: PennState Eberly College of Science, [202-]. Disponível em: <https://online.stat.psu.edu/stat462/node/199/#:~:text=The%20BIC%20places%20a%20higher,to%20favor%20models%20that%20overfit>. Acesso em: 16 fev. 2026.

PACHECO, Eliezer Moreira; PEREIRA, Luiz Augusto Caldas; DOMINGOS SOBRINHO, Moisés. Institutos Federais de Educação, Ciência e Tecnologia: limites e possibilidades. **Linhas Críticas**, p. 71-88, 2010. Disponível em: http://educa.fcc.org.br/scielo.php?pid=S1981-04312010000100005&script=sci_abstract&tlng=es. Acesso em: 17 fev. 2026.

SCHULTZ, Theodore W. Investment in human capital. **The American Economic Review**, v. 51, n. 1, p. 1-17, 1961.

SILVA, Arthur Rezende da; TERRA, Denise Cunha Tavares. A expansão dos Institutos Federais de Educação, Ciência e Tecnologia e os desafios na contribuição para o desenvolvimento local e regional. In: **I Seminário Nacional de Planejamento e Desenvolvimento**, Curitiba, 2013. Disponível em: https://utfpr-ct-static-content.s3.amazonaws.com/utfpr.curitiba.br/wp-content/uploads/sites/20/2015/08/Arthur_Rezende.pdf. Acesso em: 17 fev. 2026.

SILVA, Fernanda Cristina da et al. **Gestão da evasão na EAD**: modelo estatístico preditivo para os cursos de graduação a distância da Universidade Federal de Santa Catarina. 2017. Disponível em: <https://repositorio.ufsc.br/xmlui/handle/123456789/179913>. Acesso em: 17 fev. 2026.

SILVA, Wicara Lima e et al. Análise da evasão escolar na Educação Profissional e Tecnológica a partir da plataforma Nilo Peçanha. **Revista Semiárido De Visu**, v. 13, n. 3, p. 388-406, 2025. DOI: <https://doi.org/10.31416/rsdv.v13i3.1307>.

SOARES, Tufi Machado; FERRÃO, Maria Eugénia; MARQUES, Cláudio de Albuquerque. Análise da evasão no ProJovem Urbano: uma abordagem através do Modelo de Regressão Logística Multinível. **Ensaio: Avaliação e Políticas Públicas em Educação**, v. 19, p. 841-860, 2011. Disponível em: <https://www.scielo.br/j/ensaio/a/SXwgJqx9m3K6tWfSh5GW5Tm/?format=pdf&lang=pt>. Acesso em: 17 fev. 2026.

STATACORP. **Stata: Release 17**. Statistical Software. College Station, TX: StataCorp LLC, 2021.

SOMMET, Nicolas; MORSELLI, Davide. Keep calm and learn multilevel logistic modeling: A simplified three-step procedure using Stata, R, Mplus, and SPSS. **International Review of Social Psychology**, v. 30, n. 1, p. 203-218, 2017. Disponível em: <https://rips-irsp.com/articles/irsp.90>. Acesso em: 17 fev. 2026.

TAMURA, Karin Ayumi. **Modelo logístico multinível**: um enfoque em métodos de estimação e predição. 2007. Tese (Doutorado em Estatística) – Universidade de São Paulo, São Paulo, 2007. DOI: <https://doi.org/10.11606/D.45.2007.tde-01072007-080446>.

VARELLA, Carlos Alberto Alves. **Análise de componentes principais**. Seropédica: Universidade Federal Rural do Rio de Janeiro, 2008.

WHEELAHAN, Leesa. How competency-based training locks the working class out of powerful knowledge: A modified Bernsteinian analysis. **British Journal of Sociology of Education**, v. 28, n. 5, p. 637-651, 2007. DOI:10.1080/01425690701505540.

WOLF, Alison. **Does education matter?:** Myths about education and economic growth. [S. l.]: Penguin UK, 2002. DOI: <https://doi.org/10.1590/S0034-75902003000400009>.

WOOLDRIDGE, Jeffrey M. **Econometric analysis of cross section and panel data**. [S. l.]: MIT Press, 2010.

_____. **Introdução à econometria**: uma abordagem moderna. 4. ed. São Paulo: Cengage Learning, 2010.