



UNIVERSIDADE DE BRASÍLIA
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO, CONTABILIDADE E
GESTÃO DE POLÍTICAS PÚBLICAS
DEPARTAMENTO DE ECONOMIA

Guilherme Viana Ferreira

**Dinâmicas da evasão definitiva do ensino superior: proposta
taxonômica, seus determinantes e a importância da
interoperabilidade dos dados educacionais**

Brasília

2025

GUILHERME VIANA FERREIRA

**Dinâmicas da evasão definitiva do ensino superior: proposta taxonômica,
seus determinantes e a importância da interoperabilidade dos dados
educacionais**

Tese apresentada como parte dos requisitos necessários à obtenção do Título de Doutor em Economia pelo Programa de Pós-Graduação em Economia Política da Universidade de Brasília.

Orientadora: Prof.^a Dr.^a Andrea Felipe Cabello

Brasília

2025

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Ficha catalográfica elaborada automaticamente, com os dados fornecidos pelo(a) autor(a)

FF383d

Ferreira, Guilherme Viana Dinâmicas da evasão definitiva do ensino superior: proposta taxonômica, seus determinantes e a importância da interoperabilidade dos dados educacionais / Guilherme Viana Ferreira; orientador Andrea Felipe Cabello. Brasília, 2025.
156 p.

Tese (Doutorado em Economia) – Faculdade de Economia, Administração, Contabilidade e Gestão De Políticas Públicas, Departamento de Economia, Universidade de Brasília, 2025.

1. Educação Superior. 2. Evasão estudantil. 3. Regressão logística multinível. 4. Integração de bases de dados. 5. Interoperabilidade de dados educacionais. I. Cabello, Andrea Felipe, orient. II. Título.

Nome: FERREIRA, Guilherme Viana

Título: Dinâmicas da evasão definitiva do ensino superior: proposta taxonômica, seus determinantes e a importância da interoperabilidade dos dados educacionais

Tese apresentada como parte dos requisitos necessários à obtenção do Título de Doutor em Economia pelo Programa de Pós-Graduação em Economia Política da Universidade de Brasília.

Aprovado em: ____ / ____ / 2025

Prof.^a Dr.^a Andrea Felipe Cabello

Universidade de Brasília

Prof.^a Dr.^a Ana Maria Nogales Vasconcelos

Universidade de Brasília

Dr. Carlos Eduardo Moreno Sampaio

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira

Dr. Adriano Souza Senkevics

Instituto de Pesquisa Econômica Aplicada

Brasília

2025

Agradecimentos

Os últimos quatro anos da minha vida, até o momento em que escrevo esta seção de agradecimentos, têm sido intensos. Não só pelo doutorado em si, mas pelo momento de amadurecimento pessoal e profissional que vem ocorrendo. No âmbito profissional, nunca fui tão reconhecido e, conseqüentemente, demandado, seja acadêmica ou profissionalmente. No entanto isso tem um preço, que é o meu tempo. E, ainda assim, nos últimos meses, aceitei o desafio de ser professor substituto no Departamento de Economia da UnB, lecionando duas disciplinas as quais eu nunca havia lecionado. Então meu tempo teve que ser eficientemente alocado entre diversas atividades, com variados graus de prioridade, para me permitir chegar a este momento, que eu, honestamente, tinha dificuldades de enxergar: o fim da escrita da tese. Soma-se a isso as constantes tentativas, nem sempre bem sucedidas, de ter equilíbrio entre vida pessoal e trabalho e estudo. Por isso, acredito que eu devo, antes de tudo, reconhecer a mim mesmo pelo esforço e dedicação ao longo desses anos. É uma vitória que será comemorada.

Concluir esse doutorado é uma vitória para minha família também. Serei o primeiro doutor da família, conquista que dedico à minha avó, Geralda, e à minha tia, Cristina, que foram minhas mães, e sempre fizeram além do que era possível para que eu pudesse me dedicar aos estudos e tivesse oportunidades melhores. Elas sempre acreditaram que a educação era a solução, e tinham razão.

À tia Rosana, ao tio Paulo, à Ângela, aos primos Ian, Zé Henrique, Michelle, às tias Márcia e Ruth, à irmã Rosana, ao pai José e à minha falecida mãe Lucimar, pela inspiração, pelo apoio, pelo incentivo e por sempre torcerem por mim.

Quero agradecer meu namorado e companheiro Mário Sérgio, que teve que lidar comigo nos pontos altos e baixos desse trajeto, sendo paciente e me encorajando a perseverar, mesmo sem entender nada do que eu falava, quando eu tentava explicar minhas frustrações acadêmicas ou profissionais.

À minha orientadora e amiga, Andrea, por ter me incentivado a entrar logo no doutorado, e por fazer eu me sentir acolhido na economia. Por me tranquilizar nos momentos mais difíceis e pela parceria, não só acadêmica, que temos já há muitos anos, e que continuaremos a ter, por muitos mais. Por ver meu potencial e por reconhece-lo, em todos os momentos possíveis.

À banca de defesa, de forma geral, que aceitou essa demanda extra de avaliar minha tese. De forma específica, ao Moreno, que fez uma análise minuciosa do trabalho, apontando para definições conceituais importantes, as quais foram incorporadas. Ao Adriano, pelas sugestões de agenda de pesquisa. À Ana, pelas reflexões sobre os conceitos utilizados.

À minha amiga Junia, que sempre viu em mim mais do que eu mesmo via, me incentivando sempre. Ao meu amigo João Paulo, pela companhia e momentos de lazer. À Rejane e Carol, por estarem lá quando eu precisei. Ao Vinícius, que mesmo distante continua torcendo por mim. Ao Roberto, pela amizade e apoio no cotidiano do trabalho na UnB. Ao Delano, pelo incentivo acadêmico e parceria. À Denise, por ter confiado na minha capacidade não só como estatístico, mas como gestor. À Mariana e Fabiana, que sempre torceram por mim. E a todos os demais amigos, familiares e colegas de trabalho que me acompanharam nessa jornada. Obrigado a todos.

Resumo

A expansão do acesso ao ensino superior no Brasil, evidenciada pelo aumento de 24% nas matrículas entre 2016 e 2023, contrasta com o crescimento modesto de 17,4% nos concluintes no mesmo período, revelando um descompasso estrutural agravado pelas altas taxas de evasão estudantil. Dados da OCDE sobre adultos com ensino superior completo indicam que o país permanece com taxas inferiores às dos vizinhos latino-americanos, como Peru, Chile e Colômbia, além de ter, também, um dos maiores prêmios salariais da graduação. Nesse contexto, a evasão definitiva do sistema educacional superior emerge como fenômeno crítico, comprometendo investimentos públicos e privados enquanto reflete desigualdades estruturais no acesso às oportunidades educacionais. A tese inicia estabelecendo uma estrutura taxonômica multidimensional para a evasão no ensino superior, integrando granularidade (curso, instituição e sistema) e temporalidade (imediata, temporária e definitiva) com os qualificadores (área do conhecimento, turno, modalidade), superando abordagens fragmentadas e oferecendo um instrumental analítico robusto para diagnósticos precisos. Em seguida, evidencia-se que a evasão definitiva do sistema atinge 45% dos estudantes que já tiveram seu vínculo finalizado nas coortes de 2005 a 2022, com taxas elevadas associadas a condições socioeconômicas desfavoráveis (renda familiar, ocupação parental), características institucionais (cursos noturnos, EaD, instituições privadas com fins lucrativos) e fatores individuais (sexo masculino, autodeclaração preta/parda/indígena, deficiência), além de revelar heterogeneidade regional e institucional, com valores maiores em instituições privadas. Utilizando amostragem estratificada em dois estágios, dos microdados identificados do Inep, e regressão logística multinível, o estudo demonstrou que a ocupação parental em grupos de maior complexidade reduz em 14,5% as chances de evasão, enquanto a renda familiar elevada diminui a propensão em 13,4%, corroborando o peso estrutural das desigualdades, representadas pelas dimensões socioeconômica e individual. A tese finaliza discutindo a interoperabilidade das bases educacionais, e reforça que a fragmentação de sistemas como o Censo da Educação Superior, Censo da Educação Básica e Enem limita o acompanhamento longitudinal preciso da evasão, ressaltando a necessidade de integração semântica e governança de dados para embasar políticas públicas eficazes.

Palavras-chave

Educação Superior. Evasão estudantil. Regressão logística multinível. Integração de bases de dados. Interoperabilidade de dados educacionais.

Abstract

The expansion of access to higher education in Brazil, evidenced by the 24% increase in enrollments between 2016 and 2023, contrasts with the modest 17.4% growth in graduates in the same period, revealing a structural imbalance aggravated by high student dropout rates. OECD data on adults with higher education indicate that the country remains with lower rates than its Latin American neighbors, such as Peru, Chile, and Colombia, in addition to also having one of the highest salary premiums for undergraduate studies. In this context, permanent dropout from the higher education system emerges as a critical phenomenon, compromising public and private investments while reflecting structural inequalities in access to educational opportunities. The thesis begins by establishing a multidimensional taxonomic framework for dropout in higher education, integrating granularity (course, institution, and system) and temporality (immediate, temporary, and permanent) with qualifiers (area of knowledge, shift, modality), overcoming fragmented approaches and offering a robust analytical tool for accurate diagnoses. Next, it is evident that permanent dropout from the system affects 45% of students who have already completed their studies in the 2005 to 2022 cohorts, with high rates associated with unfavorable socioeconomic conditions (family income, parental occupation), institutional characteristics (night courses, distance learning, private for-profit institutions) and individual factors (male gender, self-declared black/brown/indigenous, disability), in addition to revealing regional and institutional heterogeneity, with higher values in private institutions. Using stratified sampling in two stages, of the identified microdata from INEP, and multilevel logistic regression, the study demonstrated that parental occupation in groups of greater complexity reduces the chances of dropout by 14.5%, while high family income reduces the propensity by 13.4%, corroborating the structural weight of inequalities, represented by the socioeconomic and individual dimensions. The thesis concludes by discussing the interoperability of educational databases, and reinforces that the fragmentation of systems such as the Higher Education Census, Basic Education Census and Enem limits the accurate longitudinal monitoring of dropout, highlighting the need for semantic integration and data governance to support effective public policies.

Keywords:

Higher Education. Student dropout. Multilevel logistic regression. Database integration. Interoperability of educational data.

Lista de ilustrações

Figura 1 – Nível de instrução das pessoas de 25 a 64 anos. Brasil e países selecionados da América do Sul. 2022 e 2023.	25
Figura 2 – Série histórica do número, em milhões, de ingressantes, matriculados e concluintes no ensino superior. Brasil. 2016 a 2023.	26
Figura 3 – Série histórica do prêmio salarial do ensino superior (%). Categoria de referência: ensino superior completo. Brasil. 2012, 2015, 2020 e 2024.	27
Figura 4 – Prêmio salarial do ensino superior. Categoria de referência: ensino médio completo. Brasil e países selecionados na América do Sul. 2022.....	28
Figura 5 - Razão de chances dos determinantes socioeconômicos, com intervalo de confiança e nível de significância.	98
Figura 6 - Razão de chances dos determinantes individuais, com intervalo de confiança e nível de significância.	101
Figura 7 - Razão de chances dos determinantes pré-graduação, com intervalo de confiança e nível de significância.	103
Figura 8 - Razão de chances dos determinantes acadêmicos, com intervalo de confiança e nível de significância.	105
Figura 9 - Razão de chances dos determinantes institucionais, com intervalo de confiança e nível de significância.	109
Figura 10 - Razão de chances dos determinantes curriculares, com intervalo de confiança e nível de significância.	113
Figura 11 - Desenho esquemático simplificado da base de dados do CES.	134

Lista de tabelas

Tabela 1 - Comparação das estatísticas descritivas das bases utilizadas. Brasil. Coortes 2005 a 2022.	70
Tabela 2 – Frequência, percentual total e percentual válido das características individuais dos discentes. Brasil. Coortes 2005 a 2022.....	72
Tabela 3 – Frequência e percentual das características dos cursos dos discentes. Brasil. Coortes 2005 a 2022.	75
Tabela 4 – Frequência e percentual das características das instituições dos discentes. Brasil. Coortes 2005 a 2022.	76
Tabela 5 - Estatísticas descritivas das variáveis numéricas. Brasil. CES ano-base 2010 a 2022.	79
Tabela 6 - Média e erro-padrão da macroevasão por variáveis de nível institucional. Brasil. Coortes 2005 a 2022.	81
Tabela 7 - Média e erro-padrão da macroevasão por variáveis de nível curso. Brasil. Coortes 2005 a 2022.	84
Tabela 8 - Média e erro-padrão da macroevasão por variáveis de nível individual. Brasil. Coortes 2005 a 2022.	87
Tabela 9 - Resultados da regressão logística multinível da macroevasão do ensino superior, em razão de chances (odds ratio). Brasil (amostra 120n). Coortes 2005 a 2022.	90
Tabela 10 - Resultados da regressão logística multinível da macroevasão do ensino superior, em razão de chances (odds ratio). Brasil (amostra 240n). Coortes 2005 a 2022.....	94
Tabela 11 - Cinco estrelas dos dados abertos.	126
Tabela 12 - Pessoas de 25 anos ou mais de idade com superior completo. Brasil. 2016 a 2023.	167
Tabela 13 - Ingressantes, matriculados e concluintes do ensino superior. Brasil. 2016 a 2023.	167

Lista de quadros

Quadro 1 - Matriz de possibilidades de evasão do curso	39
Quadro 2 - Matriz de possibilidades de evasão da área do conhecimento	40
Quadro 3 - Matriz de possibilidades de evasão da IES	41
Quadro 4 – Proposta de estrutura taxonômica dos conceitos de evasão por granularidade, temporalidade e qualificadores	52
Quadro 5 – descrição da variável TP_SITUACAO.....	134
Quadro 6 - Cursos ABI desconsiderados.....	135
Quadro 7 – Exemplos da diferença entre as variáveis CO_ALUNO e CO_ALUNO_CURSO	136
Quadro 8 - Variáveis de apoio social, atividade extracurricular, financiamento e mobilidade acadêmica tratadas.....	137
Quadro 9 - Percentual de informação disponível em cada variável, ao longo do processo de junção das bases do CES, Enem e CEB	145

Lista de abreviaturas e siglas

ADH	Atlas do Desenvolvimento Humano
CEB	Censo da Educação Básica
CES	Censo da Educação Superior
CINE	Classificação Internacional Normalizada da Educação
CO	Condições Operacionais (campus, turno ou modalidade)
CPC	Conceito Preliminar de Curso
EJA	Educação de Jovens e Adultos
e-Mec	Sistema e-MEC (Sistema Eletrônico de Regulação do Ensino Superior)
ENEM	Exame Nacional do Ensino Médio
Fies	Fundo de Financiamento ao Estudante do Ensino Superior
FGV	Fundação Getúlio Vargas
IBGE	Instituto Brasileiro de Geografia e Estatística
IDH	Índice de Desenvolvimento Humano
IDHM	Índice de Desenvolvimento Humano Municipal
IES	Instituição de Ensino Superior
IESP	Instituições de Ensino Superior Públicas
IFES	Instituições Federais de Ensino Superior
IGC	Índice Geral de Cursos
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
INSE	Indicador de Nível Socioeconômico
LGPD	Lei Geral de Proteção de Dados
MEC	Ministério da Educação
OCDE	Organização para a Cooperação e Desenvolvimento Econômico
PAS	Programa de Avaliação Seriada
PcD	Pessoa com Deficiência
PNADC	Pesquisa Nacional por Amostra de Domicílios Contínua
Prouni	Programa Universidade para Todos
Reuni	Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais
SESU	Secretaria de Educação Superior
Sisu	Sistema de Seleção Unificada
STEM	Science, Technology, Engineering and Mathematics (Ciência, Tecnologia, Engenharia e Matemática)
TAE	Técnico-Administrativo em Educação
TIC	Tecnologias da Informação e Comunicação
UF	Unidade da Federação

Sumário

Resumo	9
Abstract	11
Lista de ilustrações	13
Lista de tabelas	15
Lista de quadros	17
Lista de abreviaturas e siglas.....	19
Introdução	25
1. Tipos de evasão no ensino superior: conceituação e proposta de estrutura taxonômica	33
1.1. Evolução teórica do conceito de evasão	34
1.2. Tipos de evasão e suas definições.....	35
1.2.1. Conceituação da evasão por granularidade e temporalidade	35
1.2.2. Qualificadores e matriz de possibilidades de evasão	37
1.2.3. Perspectivas da evasão e conceitos associados.....	42
1.2.4. Formas de cálculo da evasão	44
1.3. Categorização dos determinantes da evasão	48
1.4. Proposta de estrutura taxonômica da evasão.....	50
1.5. Considerações.....	53
2. Evasão definitiva do sistema de ensino superior	57
2.1. Fontes dos dados, variáveis e métodos	58
2.1.1. Fontes dos dados.....	60
2.1.2. Variáveis e hipóteses	62
2.1.3. Estratégia empírica	67
2.1.4. Amostragem.....	69
2.1.5. Perfil dos ingressantes de 2005 a 2022 com vínculo finalizado	71
2.2. Determinantes da evasão do sistema de ensino superior.....	80

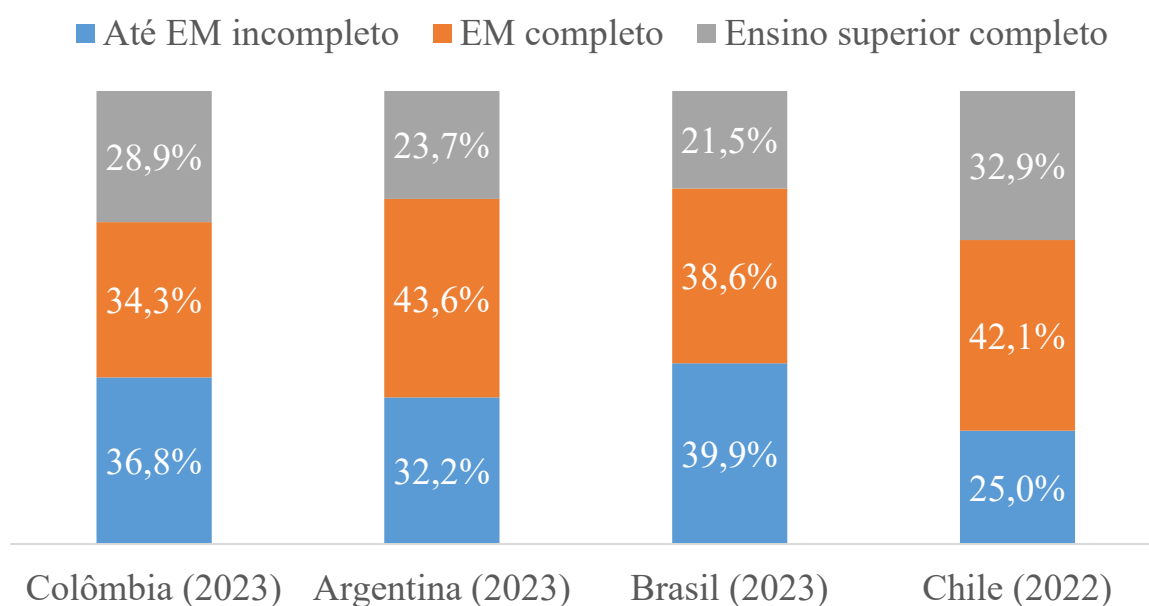
2.2.1. Evasão do sistema de ensino superior em números	80
2.2.2. Resultados das regressões logísticas multinível	89
2.2.2.1. Grupo de determinantes provenientes do CES.....	89
2.2.2.2. Grupo de determinantes provenientes do CES, CEB e Enem.....	93
2.2.3. Discussão conjunta dos determinantes.....	97
2.2.3.1. Determinantes socioeconômicos	98
2.2.3.2. Determinantes pessoais	100
2.2.3.3. Determinantes pré-graduação	102
2.2.3.4. Determinantes acadêmicos	105
2.2.3.5. Determinantes institucionais	109
2.2.3.6. Determinantes curriculares	112
2.3. Considerações.....	116
3. Interoperabilidade das bases de dados educacionais e o desafio do compartilhamento de dados	123
3.1. Interoperabilidade e reutilização de dados: conceitos e desafios	123
3.2. Dados abertos e estratégias de compartilhamento.....	124
3.3. Privacidade e soluções alternativas para compartilhamento	126
3.4. Compartilhamento de dados no Brasil	128
3.4.1. O Inep e as principais bases de dados educacionais brasileiras.....	129
3.4.2. O acesso aos dados identificados: a sala segura do Inep	131
3.5. Procedimento para limpeza e junção das bases de dados do CES, Enem e CEB	132
3.5.1. Censo da Educação Superior	133
3.5.1.1. Tratamentos específicos de variáveis do Censo da Educação Superior ..	137
3.5.2. Enem.....	139
3.5.2.1. Número de moradores.....	140
3.5.2.2. Escolaridade do pai e da mãe	140
3.5.2.3. Ocupação do pai e da mãe	141

3.5.2.4. Renda familiar	142
3.5.2.5. Tipos de escolas de ensino médio cursadas	143
3.5.2.6. Inconsistências entre versões do Enem	143
3.5.3. Censo da Educação Básica	145
3.5.4. Bases externas	146
3.5.5. A base consolidada	147
3.6. Considerações	147
Conclusão	151
Referências	153
Anexos	167

Introdução

Dados recentes sobre o nível de instrução no Brasil revelam alguns avanços significativos, bem como desafios contínuos no setor educacional – o percentual de adultos de 25 anos ou mais de idade que têm nível superior completo atingiu os 19,7%, conforme a última Pesquisa Nacional por Amostra de Domicílios Contínua – PNADC 2023 sobre educação¹, do Instituto Brasileiro de Geografia e Estatística (IBGE). Em comparação às edições anteriores da PNADC (Tabela 12), há uma clara tendência de aumento da escolaridade, pois em 2016 era 15,4 o percentual de brasileiros com nível superior completo. No entanto, a comparação com alguns países da América Latina, segundo a Organização para a Cooperação e Desenvolvimento Econômico (OCDE), como Peru (34,4%), Chile (32,9%) e Colômbia (28,9%) (OECD, 2024), indica que há bastante espaço para aprimoramento dessa estatística, segundo Figura 1.

Figura 1 – Nível de instrução das pessoas de 25 a 64 anos. Brasil e países selecionados da América do Sul. 2022 e 2023.



Fonte: OCDE, Education at a glance 2024.

Esse aumento no percentual de pessoas com nível superior foi acompanhado da expansão do acesso ao ensino superior, considerado o mesmo período. Conforme Figura 2, de 2016 a 2023, o número de matriculados ampliou quase 24% e o número de ingressantes cresceu mais de 67%,

¹ <https://www.ibge.gov.br/estatisticas/sociais/educacao/17270-pnad-continua.html?edicao=39295&t=resultados>. Acesso em março de 2024.

em contraste com o total de concluintes, que aumentou apenas 17,4%, além de ter tido um decréscimo de 3% entre 2021 e 2022 (Tabela 13). Essa evolução, especialmente dos ingressos e matrículas, é reflexo de uma democratização recente do acesso ao ensino superior, resultado da implementação de políticas que visam reduzir desigualdades e promover a inclusão de grupos historicamente sub-representados (Senkevics, 2021b), dentre outros fatores.

Figura 2 – Série histórica do número, em milhões, de ingressantes, matriculados e concluintes no ensino superior. Brasil. 2016 a 2023.



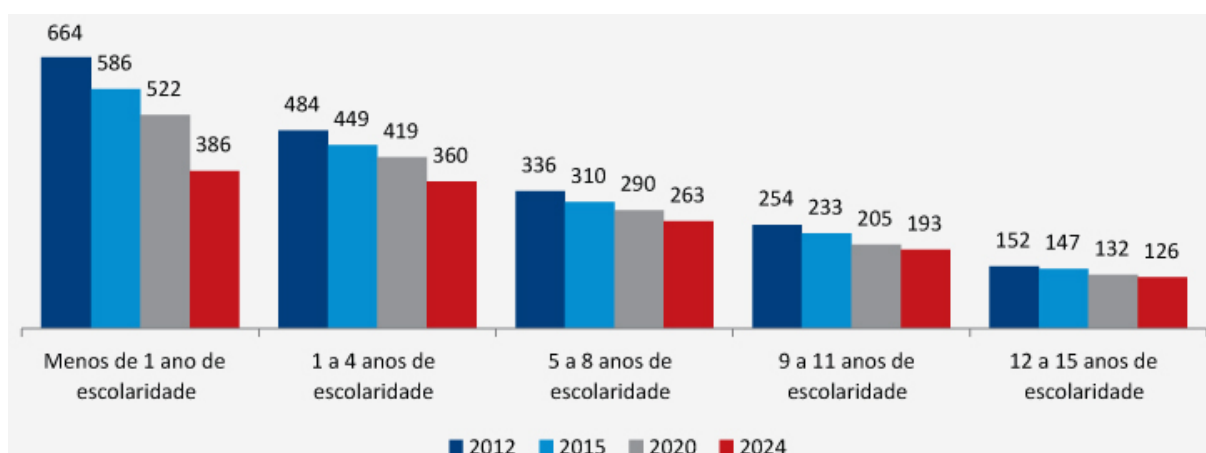
Fonte: Inep, Estatísticas do Censo da Educação Superior.

O desnível entre a alta quantidade de ingressantes e a baixa quantidade de concluintes pode ser atribuído à evasão dos alunos. A evasão estudantil é um fenômeno mundial que afeta os sistemas educacionais e gera custos não apenas ao aluno como também para as Instituições de Ensino Superior (IES) e para toda a sociedade. Além disso, é um tema que vem ganhando tração no Brasil, tanto pela sua importância quanto pelo cenário imposto pela pandemia de Covid-19, cujas consequências ainda são sentidas. O ensino remoto emergencial trouxe diversos novos fatores a serem considerados para a aprendizagem dos estudantes, e com isso novos desafios para garantir a permanência e diplomação deles.

No âmbito da discussão sobre evasão do ensino superior, há que se falar, também, do prêmio salarial da educação superior, que se refere à diferença de renda média entre trabalhadores que possuem diploma de ensino superior e aqueles que não o possuem. No contexto brasileiro, essa

diferença é expressiva e é acompanhada trimestralmente pelo IBGE, via PNADC². Estudo divulgado pela FGV³ indicou que trabalhadores com ensino superior completo têm, em média, renda 126% maior do que aqueles com ensino médio completo ou superior incompleto. E quando comparados a pessoas sem instrução formal, esse prêmio salarial chega a 386%, que pode ser visualizado na Figura 3.

Figura 3 – Série histórica do prêmio salarial do ensino superior (%). Categoria de referência: ensino superior completo. Brasil. 2012, 2015, 2020 e 2024.



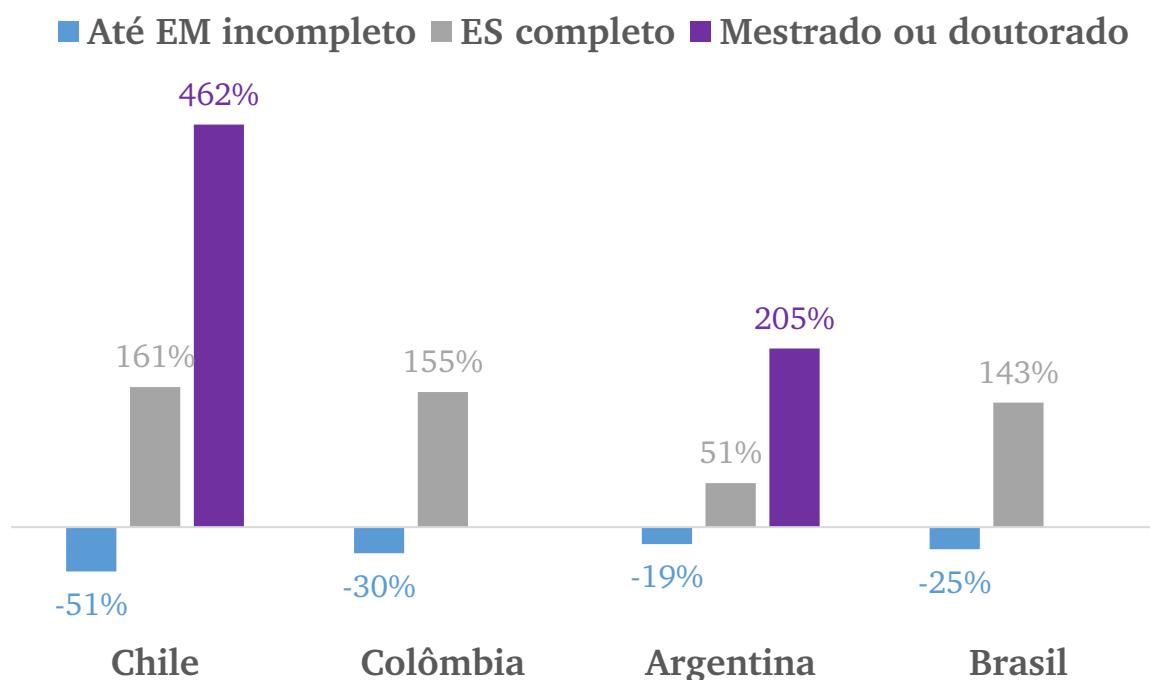
Fonte: FGV IBRE, com dados da PNAD Contínua/IBGE.

No contexto mundial, dados da OCDE demonstram que o Brasil tem um dos maiores prêmios salariais da graduação, de quase 150%, apesar de estar em linha com outros países da América do Sul (OECD, 2024), conforme Figura 4. Ou seja, o prêmio da graduação no Brasil é um dos mais altos do mundo, refletindo tanto a valorização do diploma no mercado de trabalho quanto as desigualdades estruturais do país, então ainda é muito importante discutir a evasão e estratégias para mitigá-la.

² <https://www.ibge.gov.br/estatisticas/sociais/educacao/9173-pesquisa-nacional-por-amostra-de-domicilios-continua-trimestral.html?edicao=42691>. Acesso em março de 2025.

³ <https://ibre.fgv.br/blog-da-conjuntura-economica/artigos/premio-salarial-do-ensino-superior-ainda-e-vantajoso>. Acesso em março de 2025.

Figura 4 – Prêmio salarial do ensino superior. Categoria de referência: ensino médio completo. Brasil e países selecionados na América do Sul. 2022.



Fonte: OCDE, Education at a glance 2024.

Enquanto a expansão do acesso à educação superior nas últimas décadas é um avanço notável, as altas taxas de evasão representam um problema persistente, com implicações significativas para o desenvolvimento socioeconômico do país. Este fenômeno não apenas compromete o investimento em educação, mas também reflete as dificuldades que os estudantes enfrentam no sistema educacional, considerados os diferentes perfis socioeconômicos e os fatores que levaram cada estudante a decidir ingressar no ensino superior e conseguir concluí-lo com sucesso.

No exterior, há vasta literatura sobre o fenômeno da evasão (Bradley & Migali, 2019). Já no Brasil, um marco para essa questão ocorreu em 1995, com a realização do Seminário sobre evasão nas universidades brasileiras, organizado pela Sesu/MEC e pela instituição da Comissão Especial de Estudos sobre a Evasão nas Universidades Públicas Brasileiras, no mesmo ano (Kipnis, 2000). Desde então, muitos estudos nessa temática foram publicados, mas em sua maioria focados em uma única IES ou mesmo um único curso de uma IES, o que dificulta a construção de um consenso sobre os fatores relacionados à evasão e caracteriza-se como uma lacuna a ser explorada.

Com essa lacuna em mente, o Capítulo 1 - Tipos de evasão no ensino superior: conceituação e proposta de estrutura taxonômica dedica-se a uma das principais dificuldades do estudo do fenômeno da evasão: a sua conceituação. Evidencia-se que a evasão no ensino superior é um fenômeno multifacetado, cuja compreensão exige a articulação de diferentes dimensões analíticas, tais como granularidade (curso, instituição e sistema) e temporalidade (imediate, temporária e definitiva), além da adição inovadora de qualificadores como área do conhecimento, campus, turno e modalidade. A literatura revisada demonstra que a ausência de consensos conceituais e a fragmentação das abordagens dificultam tanto a identificação dos determinantes do abandono quanto a formulação de estratégias de enfrentamento eficazes, especialmente em um contexto nacional marcado por profundas desigualdades sociais, econômicas e regionais.

Nesse sentido, o capítulo revisa criticamente os principais modelos teóricos — de Tinto, Bean, Bourdieu, Astin e Cabrera — e destaca a necessidade de abordagens integradas, que permitam captar a complexidade do fenômeno e suas múltiplas manifestações. Ao propor uma taxonomia multidimensional da evasão, que integra as dimensões de granularidade, temporalidade e qualificadores, o texto avança para além das definições tradicionais, oferecendo uma estrutura analítica capaz de diferenciar eventos de evasão em suas diversas formas e contextos. Essa proposta não apenas contribui para a padronização conceitual e metodológica do campo, mas também fundamenta a elaboração de políticas públicas e intervenções institucionais mais precisas, voltadas à mitigação da evasão e à promoção do sucesso acadêmico. Assim, o Capítulo 1 estabelece as bases teóricas e operacionais para o desenvolvimento das análises subsequentes, justificando a centralidade da discussão conceitual no enfrentamento do problema da evasão no ensino superior brasileiro.

O capítulo 2 - Evasão definitiva do sistema de ensino superior, utiliza dois dos conceitos trabalhados no Capítulo 1. O primeiro relativo às categorias dos determinantes da evasão, que podem ser classificadas em socioeconômicos, individuais, pré-graduação, acadêmicos, institucionais e curriculares, cada um representado por diversos indicadores quantificáveis. Sabe-se que estudantes de menor nível socioeconômico têm mais dificuldades de permanecer e concluir uma graduação (Coimbra, Silva, & Costa, 2021; Costa & Picanço, 2020; Farias, Gouveia, & Almeida, 2024; Martins, Fernandes, Mendes, & Magalhães, 2015; Bayma-Freire, Roazzi, & Roazzi, 2015), assim como estudantes pretos e do sexo masculino (Araújo, Mariano, & Oliveira, 2021; Costa & Picanço, 2020; Nangino & Paiva, 2022; Silva, 2013). Alunos que participam de atividades extracurriculares tendem a permanecer e concluir seus cursos

(Fagundes & Lima, 2021; Teodoro & Kappel, 2020), assim como aqueles que usufruíram de financiamento do Prouni (Carvalho & Lima, 2020; Silva & Santos, 2017). A qualidade da IES e de seus cursos tem, também, importância no controle da evasão (Silva & Santos, 2017), assim como aspectos curriculares, como carga horária, grau e turno das graduações (Assís, Bento, Pinho, Delani, & Tatiane, 2022).

E o segundo conceito, retirado do Capítulo 1, é o da evasão definitiva do sistema de ensino superior, tema central do segundo capítulo. Este conceito refere-se ao abandono completo, por parte do estudante, de qualquer modalidade de educação universitária, sem a intenção de retornar ou de se matricular em outra instituição ou curso de nível superior, caracterizando-se, assim, como uma desvinculação permanente do sistema educacional universitário. Tal fenômeno distingue-se das demais formas de evasão – de curso ou de instituição – nas quais o estudante pode migrar entre cursos, áreas ou instituições, mas permanece inserido no sistema de ensino superior. Na evasão definitiva, ocorre a interrupção integral da trajetória acadêmica, com impactos mais profundos tanto para o indivíduo quanto para a sociedade, pois implica a perda de capital humano e o comprometimento dos investimentos públicos e privados realizados na formação do estudante.

Aprofunda-se, então, a análise sobre a evasão definitiva, destacando os desafios metodológicos para sua mensuração, uma vez que a identificação desse fenômeno exige o acompanhamento longitudinal dos estudantes ao longo de extensos períodos e o acesso a bases de dados integradas de todo o sistema de ensino superior nacional. Além disso, o capítulo evidencia a escassez de estudos robustos sobre a evasão definitiva no Brasil, em parte devido à limitação de dados individualizados e à dificuldade de rastreamento das trajetórias acadêmicas após o desligamento do estudante. A pesquisa apresentada busca superar essas limitações ao utilizar microdados identificados do Inep, abrangendo as coortes de ingresso de 2005 a 2022, e adota uma abordagem metodológica inovadora, baseada em amostragem e regressão logística multinível, para captar as diferenças estruturais do fenômeno entre instituições, cursos e modalidades de ensino.

Nesse contexto, o capítulo não apenas delimita, de forma prática, a evasão definitiva do sistema, mas também investiga os determinantes socioeconômicos, individuais, pré-graduacionais, acadêmicos, institucionais e curriculares associados à decisão de abandono irreversível do ensino superior. A análise detalhada desses fatores permite compreender a complexidade do fenômeno e possibilita fundamentar, com dados, a proposição de políticas públicas e estratégias institucionais voltadas à redução da evasão do sistema, contribuindo para o aprimoramento do

sistema educacional brasileiro e para a promoção da equidade no acesso, permanência e sucesso no ensino superior, além de suprir uma lacuna significativa de pesquisa na literatura nacional (Coimbra, Silva & Costa, 2021).

O último capítulo, 3. Interoperabilidade das bases de dados educacionais e o desafio do compartilhamento de dados, dedica-se a explorar o problema de integração dos dados identificado no Capítulo 2. A análise parte do reconhecimento de que a compreensão aprofundada da evasão definitiva do sistema de ensino superior, bem como a produção de diagnósticos robustos e comparáveis sobre o fenômeno, depende fundamentalmente da disponibilidade e integração de bases de dados educacionais consistentes e interoperáveis.

No contexto brasileiro, a fragmentação das bases de dados, aliada à ausência de padrões consolidados de interoperabilidade, impõe severas limitações à pesquisa e à formulação de políticas públicas. O capítulo discute como diferentes sistemas – tais como o Censo da Educação Superior, o Censo da Educação Básica e os microdados do Enem – apresentam estruturas, identificadores e formatos de registro heterogêneos, dificultando o rastreamento longitudinal das trajetórias estudantis e a consolidação de informações essenciais para o estudo da evasão em múltiplos níveis de granularidade e temporalidade. Além disso, ressalta-se que, embora o Inep detenha a maioria dos dados considerados essenciais, a integração efetiva dessas bases ainda enfrenta desafios técnicos, institucionais e legais, especialmente no que se refere à compatibilização semântica dos dados e à observância da Lei Geral de Proteção de Dados (LGPD).

O capítulo aprofunda a discussão sobre o conceito de interoperabilidade, entendido como a capacidade de diferentes sistemas e aplicações trocarem informações de modo eficiente e compreensível, mesmo quando desenvolvidos de forma independente e com lógicas distintas. São detalhados os elevados custos e as ineficiências decorrentes da falta de interoperabilidade, que incluem não apenas a perda ou corrupção de dados durante transferências entre sistemas, mas também o alto custo operacional e o tempo despendido em processos de conversão e reformatamento de informações. Tais obstáculos são agravados pelo fato de que, a cada novo ciclo de coleta e divulgação de dados, alterações metodológicas e mudanças nos sistemas de registro acabam por comprometer a comparabilidade histórica e a consistência longitudinal dos dados educacionais.

Em resposta a esse cenário, o capítulo apresenta iniciativas recentes e perspectivas para a publicação e o compartilhamento de dados educacionais abertos, destacando a importância de modelos de infraestrutura baseados em padrões internacionais de qualidade e alinhamento

semântico (Penteado, Maldonado & Isotani, 2021). São discutidos exemplos de arcabouços tecnológicos que visam facilitar a publicação, o cruzamento e a reutilização de dados educacionais, com ênfase na necessidade de mecanismos robustos de controle de qualidade e de governança dos dados, capazes de garantir tanto a integridade quanto a segurança das informações sensíveis.

A presente tese compreende esta introdução, os capítulos 1 a 3 e a conclusão. A ordenação dos capítulos é apenas uma sugestão de leitura, que parte da conceituação geral da evasão, para a aplicação prática de um dos conceitos abordados, da evasão definitiva do sistema, finalizando com o relato das dificuldades superadas durante a pesquisa, no âmbito da integração e interoperabilidade. No entanto, os capítulos são, salvo algumas referências e detalhamentos indicados, autocontidos, então podem ser lidos em qualquer ordem, sem prejuízo para o leitor.

1. Tipos de evasão no ensino superior: conceituação e proposta de estrutura taxonômica

A evasão no ensino superior é um fenômeno amplamente reconhecido como um dos maiores desafios enfrentados pelas instituições de ensino e pelos sistemas educacionais em todo o mundo (Tinto, 1975). No Brasil, esse problema assume proporções ainda mais preocupantes, dada a complexidade das desigualdades sociais, econômicas e regionais que caracterizam o país. A literatura especializada converge para a ideia de que a evasão é um fenômeno multifacetado, envolvendo interações dinâmicas entre fatores individuais, institucionais e contextuais. Essa complexidade se reflete tanto na diversidade de tipos de evasão quanto nos determinantes que os influenciam, exigindo abordagens teóricas e metodológicas robustas para sua compreensão.

A análise dos tipos de evasão revela que diferentes formas de abandono ou migração no ensino superior estão associadas a causas específicas, que demandam intervenções igualmente específicas. A evasão de curso pode estar relacionada a questões vocacionais ou acadêmicas (Barroso et al., 2022; Santos, Pilatti & Bondarik, 2022), enquanto a evasão institucional pode refletir insatisfações com infraestrutura ou busca por melhores oportunidades em outras instituições (Filho & Nery, 2009). Já a evasão sistêmica, considerada a mais grave, indica uma desistência completa do sistema educacional superior e está frequentemente associada a barreiras socioeconômicas (Paula, 2021) ou falta de suporte adequado (Farias, Gouveia & Almeida, 2024).

Diante desse cenário, torna-se essencial compreender como o conceito de evasão tem evoluído ao longo do tempo, incorporando dimensões analíticas como granularidade (curso, instituição e sistema) e temporalidade (imediate, temporária e definitiva), juntamente a novos qualificadores para o fenômeno. Essas dimensões permitem uma caracterização mais precisa do fenômeno, superando as limitações das abordagens tradicionais que tratam a evasão como um evento uniforme. Além disso, discutir os determinantes da evasão no contexto brasileiro é fundamental para identificar os fatores estruturais e conjunturais que perpetuam as altas taxas de abandono no ensino superior.

Este capítulo pretende explorar as múltiplas facetas da evasão no ensino superior brasileiro, com foco na conceituação dos diferentes tipos de evasão e seus desdobramentos teóricos. Para isso, revisa-se a literatura teórica sobre o tema, analisam-se as principais contribuições metodológicas e propõe-se uma taxonomia abrangente que integre as dimensões de granularidade, temporalidade e fatores qualificadores. Ao final, busca-se oferecer subsídios

para pesquisadores do tema, gestores educacionais e formuladores de políticas públicas desenvolverem estratégias mais eficazes para mitigar os impactos da evasão e promover o sucesso acadêmico no Brasil, por meio de uma estrutura taxonômica que permite identificar detalhadamente cada tipo de evasão passível de ocorrer.

A narrativa será construída começando com a evolução teórica do conceito de evasão, conforme os autores mais proeminentes do tema. Em seguida serão explorados os tipos de evasão e suas definições, incluindo a conceituação de granularidade, temporalidade e fatores qualificadores, assim como discussões auxiliares do tema, finalizando a seção com a proposta de estrutura taxonômica para o estudo da evasão. A categorização dos determinantes da evasão é apresentada, frente à literatura analisada. São apresentadas, também, as diversas formas de cálculo da evasão, com sugestão de qual metodologia adotar. E o capítulo será concluído com as considerações finais pertinentes ao que foi apresentado.

1.1. Evolução teórica do conceito de evasão

O estudo precursor de Tinto (1975) foi o primeiro a estruturar teoricamente o conceito de evasão da educação superior. Seu modelo de integração acadêmica e social propõe que a permanência de um estudante em uma instituição depende de sua adaptação às demandas acadêmicas e à vida social universitária. Estudos brasileiros frequentemente utilizam esse modelo como base para investigações sobre evasão, embora algumas pesquisas apontem que seu uso pode ser limitado, ao não conseguir se adequar às desigualdades estruturais presentes em contextos como o Brasil (Castro & Teixeira, 2014). Além do modelo de Tinto, abordagens como o modelo de atrito de Bean, as teorias de capital cultural de Bourdieu e as contribuições de Astin e Cabrera são frequentemente adotadas para compreender o fenômeno de forma mais ampla.

O modelo de atrito de Bean (1980) propõe que a evasão no ensino superior é influenciada por fatores institucionais e não institucionais, incluindo satisfação acadêmica, apoio social, motivação pessoal e influências externas, como obrigações familiares e demandas financeiras. Diferente do modelo de Tinto, que enfatiza a integração acadêmica e social, Bean se baseia em modelos organizacionais do mundo corporativo, sugerindo que a decisão de permanecer na universidade é semelhante à decisão de um funcionário de continuar em um emprego. Estudos no Brasil utilizam esse modelo para analisar como fatores externos, como condições socioeconômicas e mercado de trabalho, podem afetar a permanência dos estudantes (Castro & Teixeira, 2014; Pinheiro, Ribeiro & Fernandes, 2023).

As teorias de Pierre Bourdieu (2014) aplicadas à evasão acadêmica argumentam que o sucesso ou abandono dos estudantes está ligado ao seu capital cultural, social e econômico. O conceito

de capital cultural refere-se aos recursos simbólicos e educacionais herdados de suas famílias, que influenciam sua adaptação ao ambiente acadêmico. Estudantes de origem socioeconômica desfavorecida muitas vezes enfrentam barreiras institucionais, pois podem ter menor familiaridade com a cultura acadêmica dominante, o que impacta sua participação e desempenho. No Brasil, essas teorias são frequentemente utilizadas para entender como desigualdades estruturais, incluindo raça, cor e nível socioeconômico, afetam a evasão e a eficácia das políticas afirmativas (Castro & Teixeira, 2014; Pinheiro, Ribeiro & Fernandes, 2023).

Astin (1998) introduziu a teoria do envolvimento estudantil, que propõe que quanto mais um aluno se engaje em atividades acadêmicas e interações institucionais, maior a probabilidade de sua permanência. Sua teoria tem sido aplicada em estudos sobre a importância do engajamento em ambientes universitários para reduzir a evasão. Já Cabrera *et al.* (1993) combinaram elementos dos modelos de Tinto e Bean, enfatizando que a evasão é fruto tanto da interação social e acadêmica dentro da instituição quanto da influência de fatores socioeconômicos e familiares externos. As pesquisas brasileiras utilizam essas abordagens para analisar como diferentes tipos de suporte institucional e envolvimento podem contribuir para a retenção dos estudantes (Castro & Teixeira, 2014; Pinheiro, Ribeiro & Fernandes, 2023).

Todas as teorias discutidas possuem aplicabilidade ao caso brasileiro, mas apenas se combinadas, pois, isoladamente, não são suficientes para compreensão do fenômeno por completo, dadas as complexas interações entre os inúmeros determinantes da evasão. Além disso, é necessário considerar a evasão sob a perspectiva dos cursos, das instituições e do sistema de ensino superior, fatores não abordados nesses estudos.

1.2. Tipos de evasão e suas definições

A literatura teórica sobre evasão carece, de forma geral, da utilização e integração dos conceitos de evasão por granularidade e temporalidade (Vitelli & Fritsch, 2016). A granularidade diz respeito ao contexto em que ocorreu a evasão: no curso, na instituição ou no sistema educacional, aspectos inicialmente definidos pela Comissão Especial de Estudos sobre a Evasão (MEC, 1996). E a temporalidade dimensiona a duração da evasão: imediata, se referente a até um ano; por período definido, se referente a mais de um ano; e definitiva, quando não há mais registro de matrícula, considerando todo o lapso temporal averiguável.

1.2.1. Conceituação da evasão por granularidade e temporalidade

Conceituando cada tipo de evasão, a granularidade de curso refere-se à saída definitiva do aluno do curso de origem sem o concluir. Segundo a Comissão Especial de Estudos sobre Evasão do

MEC, esta modalidade caracteriza-se em situações diversas como: abandono (deixa de matricular-se), desistência oficial, transferência ou reopção (mudança de curso), e exclusão por norma institucional (MEC, 1996). Uma característica distintiva deste tipo de evasão é que o estudante pode continuar na mesma instituição, apenas migrando para outro curso. Gilioli (2016) propôs uma nomenclatura alternativa denominando este fenômeno como microevasão – quando o estudante deixa o curso original, mas permanece na mesma IES ou no sistema. Complementarmente, Lima e Zago (2018) introduziram o conceito de nanoevasão, que ocorre quando o estudante permanece no mesmo curso e instituição, mas transfere-se para outro campus, turno, modalidade de ensino (presencial ou a distância) ou formação (licenciatura ou bacharelado).

A evasão da IES é definida como a situação em que o estudante se desliga da instituição na qual está matriculado. Este fenômeno ocorre quando há a saída do aluno da instituição de origem, que pode migrar para outra instituição de ensino superior, mas não abandona necessariamente o sistema educacional todo. Esta modalidade de evasão é classificada como mesoevasão (Gilioli, 2016), sendo caracterizada pela saída do curso e da IES, mas com a permanência do estudante no sistema de ensino superior, geralmente pelo ingresso em outra instituição. Representa uma perda direta para a instituição original, mesmo que o sistema educacional na totalidade mantenha o estudante.

A evasão do sistema caracteriza-se pelo abandono, de forma definitiva ou temporária, do ensino superior todo. Ocorre quando o estudante sai definitivamente do sistema de ensino, afastando-se do curso de origem sem ingressar em outro curso ou instituição de ensino superior. Este é considerado o tipo mais grave de evasão, pois implica na completa desistência do indivíduo em relação à educação superior. Na taxonomia proposta por Gilioli (2016), este fenômeno é denominado macroevasão, referindo-se à saída do curso e da IES não acompanhada pelo ingresso em outro curso/IES, representando a saída completa do sistema educacional superior. De maneira geral, no que concerne à granularidade, a literatura revisada indica que a maioria dos estudos sobre evasão no ensino superior tende a se concentrar isoladamente nos níveis de curso, institucional ou, raramente, sistêmico, sem uma abordagem que os integre de maneira abrangente. Essa fragmentação limita a compreensão plena dos fatores que contribuem para a evasão, dificultando tanto a formulação de estratégias eficazes de retenção quanto a avaliação do impacto de políticas públicas. Na maioria das pesquisas, a atenção está centrada em apenas um desses níveis, deixando de lado suas interconexões. Além disso, já há indícios de que o

comportamento da evasão e os determinantes que a influenciam dependem, também, da granularidade analisada (Paula, 2021).

Já a temporalidade nem é, geralmente, citada nos estudos consultados, mas pode-se afirmar que a maioria trata da evasão imediata, até por ser a de mais fácil mensuração, considerando a estruturação dos registros acadêmicos em cada IES. Deve-se destacar as dificuldades operacionais para implementação dos cálculos de evasão, no âmbito das IES, por granularidade e temporalidade, pois dependem de registros que estão além de sua jurisdição – cada IES consegue acompanhar seus estudantes que mudam de curso dentro da instituição, mas não sabem quando eles se matriculam em outra IES. Essa discussão será ampliada no capítulo 3.

A consequência da falta de uma teoria conceitual de evasão que considere as dimensões de granularidade e temporalidade está na impossibilidade de se afirmar quais determinantes tornam a evasão provável e em qual lapso temporal, informações essas que subsidiariam a melhoria de todo o ensino superior. A incapacidade de definir e analisar claramente a evasão por meio dessas dimensões limita o potencial de melhoria dos sistemas de ensino superior. Compreender as nuances da evasão é crucial para o desenvolvimento de estratégias direcionadas que aumentem a retenção e o sucesso dos alunos.

1.2.2. Qualificadores e matriz de possibilidades de evasão

Compreender a evasão por granularidade pressupõe o detalhamento do fenômeno, considerando diferentes perspectivas e fatores que influenciam o abandono ou a migração no ensino superior. Para tanto, propõem-se qualificadores que ajudam a identificar não apenas os tipos de evasão, mas também suas causas e implicações. Tais qualificadores incluem as granularidades de curso e IES, combinadas à área do conhecimento e às condições operacionais (CO) – campus, turno e modalidade – os quais complementam a análise ao refletirem aspectos logísticos que podem impactar a decisão do estudante.

A área do conhecimento, ou grande área⁴, segundo a classificação CINE, merece atenção especial por dois fatores principais. Primeiro, pela especificidade dos desafios acadêmicos associados a cada área do conhecimento: cursos de engenharia e ciências exatas demandam alta carga de matemática, o que pode gerar evasão por dificuldades acadêmicas, enquanto cursos de humanidades enfrentam problemas de desalinhamento vocacional (Santos, Pilatti & Bondarik, 2022). Segundo, conforme observam Miranda Júnior & Cabello (2019), a evasão de curso tende

⁴ De acordo com a classificação CINE Brasil, são dez as áreas gerais do conhecimento: 1) educação; 2) artes e humanidades; 3) ciências sociais, comunicação e informação; 4) negócios, administração e direito; 5) ciências naturais, matemática e estatística; 6) computação e Tecnologias da Informação e Comunicação (TIC); 7) engenharia, produção e construção; 8) agricultura, silvicultura, pesca e veterinária; 9) saúde e bem-estar; e 10) serviços.

a ocorrer dentro da mesma área do conhecimento, independente da área e do curso, o que pode estar relacionado com uma falta de conhecimento dos estudantes sobre nuances específicas dos cursos. Terceiro, pelo impacto futuro na força de trabalho: a evasão em áreas estratégicas, como saúde, educação e engenharias, compromete a formação de mão de obra qualificada, afetando setores essenciais da economia (Santos, Pilatti & Bondarik, 2022; Silva Filho et al., 2007) e o crescimento econômico da nação.

Ao integrar essas dimensões e qualificadores na análise da evasão, é possível compreender melhor as causas subjacentes e os padrões específicos de abandono ou mobilidade estudantil. Isso não apenas permite uma caracterização mais precisa do fenômeno como também fundamenta intervenções direcionadas para mitigar os impactos negativos da evasão no âmbito individual, institucional e sistêmico. Essa abordagem detalhada é essencial para desvendar a complexidade do fenômeno, especialmente porque eventos de evasão ocorrem raramente de forma isolada.

Iniciando esse detalhamento pela evasão do curso, ela pode ocorrer de forma isolada, porém a possibilidade de que outros qualificadores ajam de forma concomitante é alta, pois tal evento pode ser acompanhado por uma evasão da área de conhecimento, da IES, do campus ou do turno/modalidade, esses últimos considerados de forma conjunta porque cursos EaD não têm turno. Então um evento que, inicialmente, aparentava ser singular, torna-se múltiplo – são 16 as possibilidades de evasão na granularidade de curso, considerando as interações com as demais granularidades e com as chamadas condições operacionais, dispostas no Quadro 1.

A matriz de possibilidades de evasão do curso permite identificar e destacar padrões importantes em todas as dimensões analisadas. A mudança de área durante a evasão do curso frequentemente indica problemas de orientação vocacional ou expectativas não correspondidas com a realidade profissional (Barroso et al., 2022). Já a permanência na mesma área com troca de curso sugere uma afinidade com o campo de conhecimento, porém insatisfação com aspectos específicos do curso original ou mesmo um planejamento prévio, no sentido de ingressar, inicialmente, em um curso com menor nota de corte, para posteriormente solicitar a transferência de curso – também conhecido como efeito trampolim (Miranda Júnior & Cabello, 2019).

Quanto à mudança de IES, este qualificador está geralmente associado à busca por instituições com melhor infraestrutura ou reputação, enquanto a permanência na mesma instituição pode indicar satisfação com o ambiente acadêmico geral (Filho & Nery, 2009). A mudança de campus, por sua vez, relaciona-se frequentemente a questões de mobilidade urbana e

acessibilidade geográfica, sendo que estudantes que permanecem no mesmo campus tendem a valorizar a proximidade de sua residência ou local de trabalho. Por fim, migrações entre diferentes turnos (diurno para noturno) ou modalidades (presencial para EaD) refletem predominantemente dificuldades dos estudantes em conciliar estudos com trabalho ou responsabilidades familiares, fenômeno particularmente relevante em cursos noturnos e em períodos de crise econômica (Filho & Nery, 2009).

A análise conjunta desses qualificadores permite compreender a complexidade do fenômeno, revelando que eventos de evasão com múltiplas mudanças (curso, área, IES, campus e turno/modalidade simultaneamente) geralmente indicam problemas estruturais mais profundos no percurso acadêmico, enquanto mudanças isoladas tendem a refletir ajustes pontuais na trajetória educacional do estudante.

Quadro 1 - Matriz de possibilidades de evasão do curso

Mudança de área	Mudança de IES	Mudança de campus	Mudança de turno ou modalidade	Interpretação
Sim	Sim	Sim	Sim	Evasão com mudança completa (curso, área, instituição, campus e turno/modalidade)
Sim	Sim	Sim	Não	Mudança de curso, área, instituição e campus, mantendo turno/modalidade
Sim	Sim	Não	Sim	Mudança de curso, área, instituição e turno/modalidade, mantendo campus
Sim	Sim	Não	Não	Mudança de curso, área e instituição, mantendo condições operacionais
Sim	Não	Sim	Sim	Mudança de curso e área na mesma instituição com alterações operacionais completas
Sim	Não	Sim	Não	Mudança de curso, área e campus na mesma instituição, mesmo turno/modalidade
Sim	Não	Não	Sim	Mudança de curso, área e turno/modalidade na mesma IES e campus
Sim	Não	Não	Não	Mudança de curso e área, mantendo mesma IES, campus e turno/modalidade
Não	Sim	Sim	Sim	Mudança de curso mantendo área, com nova IES e alterações operacionais completas
Não	Sim	Sim	Não	Mudança de curso mantendo área, com nova IES e campus, mesmo turno/modalidade
Não	Sim	Não	Sim	Mudança de curso mantendo área, com nova IES, mesmo campus, novo turno/modalidade
Não	Sim	Não	Não	Mudança de curso mantendo área, com nova IES, mantendo condições operacionais
Não	Não	Sim	Sim	Mudança de curso na mesma área e IES, com novo campus e turno/modalidade
Não	Não	Sim	Não	Mudança de curso na mesma área e IES, com novo campus, mesmo turno/modalidade
Não	Não	Não	Sim	Mudança de curso na mesma área, IES e campus, com novo turno/modalidade

Não	Não	Não	Não	Mudança apenas de curso, mantendo área, IES, campus e turno/modalidade
-----	-----	-----	-----	--

Fonte: elaboração própria.

Analizadas as possibilidades de evasão do curso, pode-se discutir, também, a matriz de possibilidade de evasão da área do conhecimento, conforme Quadro 2. Quando um estudante evade da área do conhecimento a qual ele pertencia, automaticamente ocorre, também, a evasão do curso, pois não existem, na metodologia da Cine Brasil, cursos com mesmo código que façam parte de áreas diferentes. Devido a isso, a matriz em discussão tem apenas oito entradas, referentes às possíveis combinações da evasão da área do conhecimento, então os mesmos apontamentos feitos na apresentação do Quadro 1 aplicam-se aqui, tratando-se dos qualificadores IES, campus e turno/modalidade.

Quadro 2 - Matriz de possibilidades de evasão da área do conhecimento

Mudança de IES	Mudança de campus	Mudança de turno ou modalidade	Interpretação
Sim	Sim	Sim	Evasão com mudança completa (curso, área, instituição, campus e turno/modalidade)
Sim	Sim	Não	Mudança de curso, área, instituição e campus, mantendo turno/modalidade
Sim	Não	Sim	Mudança de curso, área, instituição e turno/modalidade, mantendo campus
Sim	Não	Não	Mudança de curso, área e instituição, mantendo condições operacionais
Não	Sim	Sim	Mudança de curso e área na mesma instituição com alterações operacionais completas
Não	Sim	Não	Mudança de curso, área e campus na mesma instituição, mesmo turno/modalidade
Não	Não	Sim	Mudança de curso, área e turno/modalidade na mesma IES e campus
Não	Não	Não	Mudança de curso e área, mantendo mesma IES, campus e turno/modalidade

Fonte: elaboração própria.

Para finalizar os cenários de evasão por granularidade, tem-se a matriz de possibilidades de evasão da IES, no Quadro 3, composta por 12 entradas distintas. Focando nos casos não discutidos, que dizem respeito à evasão da IES sem mudança de curso, e consequente manutenção da área do conhecimento, as possibilidades dessa evasão variam conforme a mudança de campus ou turno/modalidade.

Nestes cenários específicos, observa-se frequentemente que fatores geográficos e logísticos exercem papel determinante na decisão do estudante. A saída para outra instituição mantendo o mesmo curso e área, com mudança de campus, reflete geralmente a busca por proximidade à

residência ou local de trabalho, reduzindo tempo e custos de deslocamento (Ariovaldo, 2023). Quando ocorre mudança de turno/modalidade, a literatura sugere motivações relacionadas à flexibilidade temporal, especialmente entre estudantes que trabalham (Paula, 2021). A análise destes padrões revela que esta categoria de evasão, institucional, reflete menos insatisfação com aspectos acadêmicos e mais adaptação às necessidades práticas do estudante (Gama, 2018). Diferentemente da evasão com mudança de curso, acredita-se que este tipo de migração institucional tenha potencial positivo para a continuidade e conclusão dos estudos, pois mantém o alinhamento vocacional original e resulta frequentemente de decisões pragmáticas para viabilizar a permanência no sistema educacional, ainda que em outra IES (Ariovaldo, 2023).

Quadro 3 - Matriz de possibilidades de evasão da IES

Mudança de curso	Mudança de área	Mudança de campus	Mudança de turno ou modalidade	Interpretação
Sim	Sim	Sim	Sim	Evasão com mudança completa (curso, área, instituição, campus e turno/modalidade)
Sim	Sim	Sim	Não	Mudança de curso, área, instituição e campus, mantendo turno/modalidade
Sim	Sim	Não	Sim	Mudança de curso, área, instituição e turno/modalidade, mantendo campus
Sim	Sim	Não	Não	Mudança de curso, área e instituição, mantendo condições operacionais
Sim	Não	Sim	Sim	Mudança de curso mantendo área, com nova IES e alterações operacionais completas
Sim	Não	Sim	Não	Mudança de curso mantendo área, com nova IES e campus, mesmo turno/modalidade
Sim	Não	Não	Sim	Mudança de curso mantendo área, com nova IES, mesmo campus, novo turno/modalidade
Sim	Não	Não	Não	Mudança de curso mantendo área, com nova IES, mantendo condições operacionais
Não	Não	Sim	Sim	Mesmo curso e área, nova IES com novo campus e turno/modalidade
Não	Não	Sim	Não	Mesmo curso e área, nova IES e campus, mesmo turno/modalidade
Não	Não	Não	Sim	Mesmo curso e área, nova IES, mesmo campus, novo turno/modalidade
Não	Não	Não	Não	Transferência simples: mesmo curso, área e condições operacionais, apenas nova IES

Fonte: elaboração própria.

Entende-se que a matriz de possibilidades de evasão do sistema pode ser construída, porém, optou-se por não o fazer, devido à complexidade inerente ao tema e à dificuldade de caracterizar com precisão a evasão sistêmica como um evento definitivo. Diferentemente das outras granularidades (curso e IES), a evasão do sistema envolve incertezas significativas, pois o aluno

pode retornar ao ensino superior em qualquer momento futuro, com ou sem mudanças em quaisquer dos qualificadores. Outro ponto relevante é que os fatores associados à evasão sistêmica são amplos e multifacetados, extrapolando, possivelmente, o contexto educacional e envolvendo questões econômicas, sociais e individuais, o que dificulta a construção de cenários claros e objetivos.

1.2.3. Perspectivas da evasão e conceitos associados

Diante dessas limitações conceituais e metodológicas, o aprimoramento da compreensão da evasão requer priorizar análises em níveis onde os eventos são mensuráveis e as perspectivas dos atores são claramente delineadas. Enquanto a evasão do sistema permanece um fenômeno ambíguo – sujeito a retornos imprevisíveis e influenciado por fatores externos ao ambiente educacional –, a análise das granularidades de curso e IES permite focar em dinâmicas mais concretas. A interdependência entre granularidade e perspectiva torna-se evidente: para o aluno, a migração entre cursos, áreas ou instituições representa uma adaptação estratégica, enquanto para a IES e o sistema, configura-se como perda de recursos ou capital humano. Essa dualidade revela que, embora a macroevasão definitiva seja o evento mais crítico, sua complexidade operacional justifica concentrar esforços em níveis nos quais intervenções são viáveis – como políticas de permanência institucional ou redesenho curricular –, deixando a evasão sistêmica para estudos de grande porte que possam superar as atuais limitações de acompanhamento longitudinal dos egressos.

Para o aluno, quando ele muda de curso, área ou IES, ocorre apenas uma transferência ou migração para melhor se adaptar aos seus objetivos ou necessidades. A evasão só ocorreria, de fato, caso o indivíduo decida não mais retornar ao ensino superior, ou seja, no evento de uma macroevasão definitiva. Já na granularidade de curso ou área, duas possibilidades são aplicáveis, olhando pela perspectiva da IES. Na primeira, o aluno pode ter mudado de curso ou área, mas manteve-se na mesma IES. Na segunda, ocorreu a migração de curso ou área e também de instituição. Em ambos os casos, pode-se investigar, também, se houve mudança de campus, turno ou modalidade, para melhor qualificar o evento. Ou seja, os determinantes que levaram à evasão do curso ou da área podem ser diferentes, dependendo da perspectiva e dos fatores adicionais.

Quando o aluno muda de curso ou área, mas permanece na mesma instituição (transferência interna), configura-se uma microevasão ou nanoevasão, dependendo das características da migração. Nesse cenário, fatores como insatisfação com a grade curricular, dificuldades em disciplinas específicas ou descoberta de novas aptidões são possíveis determinantes (Pinheiro,

Ribeiro & Fernandes, 2023). Estudos destacam que mesmo mudanças para outro campus, turno ou modalidade geram impactos operacionais, como ociosidade de vagas em turmas específicas e desequilíbrio na alocação de recursos pedagógicos (Lima & Zago, 2018).

A mudança de campus, turno ou modalidade dentro da mesma IES (nanoevasão) pode mascarar evasões parciais. Um aluno que migra do turno diurno para o noturno pode estar ajustando-se a demandas laborais, mas essa adaptação nem sempre resolve problemas acadêmicos subjacentes (Santos, Pilatti & Bondarik, 2022). E a troca de modalidade, presencial para EaD, muitas vezes reflete dificuldades de conciliar estudos com trabalho (Pinheiro, Ribeiro & Fernandes, 2023).

Já na migração para outra IES (transferência externa), classificada como mesoevasão, os motivos frequentemente envolvem questões institucionais: falta de suporte acadêmico, infraestrutura inadequada ou busca por instituições com melhor reputação no curso desejado (Nierotka, Salata & Klitzke Martins, 2023). Estudos indicam que grande parte dos alunos que ingressaram via Sisu migraram de instituição, muitas vezes motivados por políticas de assistência estudantil ou oportunidades de mobilidade geográfica (Cabello et al., 2021). A análise desses casos requer diferenciar transferências estratégicas, que seriam a busca por cursos melhor avaliados, de evasões por insatisfação, sendo esta última associada a altos índices de reprovação ou desalinhamento entre expectativas e realidade do curso.

Pode-se dizer, então, que as transferências internas representam perda de capital humano específico (evasão em cursos de licenciatura agrava déficits docentes, por exemplo), enquanto as externas impactam diretamente a receita, se privada, e a competitividade institucional. Já na perspectiva do sistema, a migração entre cursos ou áreas pode indicar falhas na orientação vocacional pré-vestibular, enquanto a saída para outra IES expõe desigualdades na qualidade ou atratividade da oferta entre instituições (Li, 2016), dentre outros possíveis fatores.

Há ainda diversos termos utilizados pela literatura tratados como sinônimos de evasão, sem que isso seja necessariamente o caso. Palavras como abandono, desistência, fracasso, falha e saída têm conexão direta com o conceito de evasão, apesar de diversas dessas terem um julgamento negativo embutido. O abandono ou a desistência se referem à decisão do aluno de deixar seu programa acadêmico, geralmente sem notificação formal à instituição. Estão associados à falta de conexão ou engajamento com o curso, ou instituição (Mendes, 2023). A falha ou o fracasso normalmente estão associados à incapacidade do aluno de atender aos requisitos acadêmicos de um curso, levando à sua remoção do programa. Esses conceitos são frequentemente associados ao desempenho acadêmico, em vez de à saída voluntária (Branco, Conte & Habowski, 2020).

E a saída seria um termo geral para deixar um programa acadêmico, abrangendo saídas voluntárias e involuntárias. Pode ser devido a vários motivos, incluindo desafios pessoais, financeiros ou mesmo acadêmicos (Mendes, 2023; Oliveira & Barbosa, 2023). Importante notar que todos esses conceitos se direcionam à definição de evasão, a qual pode ser referente a qualquer granularidade ou temporalidade, dado a amplitude do conceito em discussão.

Já os termos migração e transferência têm uma conexão indireta com a evasão, pois sugerem que houve um evento de evasão seguido pela criação de um novo vínculo. Ambos os termos envolvem a transferência de um estudante de uma instituição ou programa para outro. É visto como uma mudança no caminho educacional e não como uma saída completa, muitas vezes em busca de um melhor alinhamento com os objetivos pessoais ou profissionais (Gama, 2018).

Além das dimensões e conceitos da evasão, deve-se considerar também a metodologia de cálculo utilizada. A literatura aponta que há muitas fórmulas diferentes para mensuração da evasão (Vitelli & Fritsch, 2016), o que é problemático, pois leva a resultados variados que não são comparáveis e têm aplicabilidade limitada. Isso demonstra que, para avançarmos efetivamente na quantificação do fenômeno e possibilitarmos comparações adequadas entre dados de diferentes estudos e períodos, é necessário estabelecer uma metodologia padronizada de diagnóstico, que deve incluir a classificação dos determinantes que levam à evasão e a forma de cálculo utilizada.

1.2.4. Formas de cálculo da evasão

No Brasil, o primeiro registro de fórmula de cálculo da evasão surgiu da Comissão Especial para o Estudo da Evasão, instituída em 1995, a qual fez um estudo abrangente sobre diplomação, retenção e evasão nos cursos de graduação das Instituições de Ensino Superior Públicas (IESP) brasileiras (MEC, 1996). A comissão buscou definir uma fórmula comum de cálculo de índices de evasão que pudesse ajudar a identificar causas e, possivelmente, propor soluções para o fenômeno na educação superior brasileira.

A metodologia adotada foi a de fluxo de acompanhamento de estudantes, chamada de técnica de painel ou coorte. O estudo acompanhou alunos ingressantes em determinados cursos até o prazo máximo de integralização curricular, considerando gerações completas cujo prazo máximo havia expirado entre o segundo semestre de 1992 e o segundo semestre de 1994. Para cada curso, foram analisadas três gerações completas (para cursos de ingresso anual) ou cinco gerações completas (para cursos de ingresso semestral). Uma geração completa foi definida como aquela em que o número de diplomados, evadidos e retidos é igual ao número de ingressantes no ano-base, considerando o tempo máximo de integralização do curso.

O cálculo da taxa de evasão foi expresso pela fórmula 1, onde Ni é o número de ingressantes, Nd o número de diplomados e Nr o número de retidos. Os dados foram coletados por cursos em cada universidade e posteriormente agrupados por áreas e subáreas de conhecimento, permitindo a comparação dos resultados. Tal metodologia permite o cálculo da evasão de curso e da instituição, no momento imediato, representando o percentual de discentes que não estão mais no curso ou instituição nem se diplomaram.

$$E^{CEEE} = \frac{Ni - Nd - Nr}{Ni} \times 100 \quad (1)$$

O Instituto Lobo é responsável por difundir algumas fórmulas de cálculo de evasão aqui no Brasil. A primeira delas teria granularidade por curso ou instituição e temporalidade imediata, sendo considerada como a evasão anual com dados agregados e de uso internacional (Silva Filho et al., 2007). A fórmula 2 descreve esse cálculo, o qual relaciona os matriculados (M_i) e ingressantes (I_i) do ano i aos matriculados e concluintes (C_{i-1}) do ano anterior.

$$E_i^{LOBO_2007} = 100 - \frac{M_i - I_i}{M_{i-1} - C_{i-1}} \times 100 \quad (2)$$

O resultado deste cálculo representa o percentual de alunos de um curso ou instituição, durante o período i (excluindo aqueles que ingressaram no mesmo período), que estavam matriculados no período anterior (sem considerar os formandos) e optaram por não realizar a matrícula no período i . Esse cálculo é utilizado para mensurar a evasão imediata, refletindo o percentual de não rematrículas. É importante observar que não são incluídos nesse cálculo os alunos que reingressam no curso ou na instituição e efetuam a matrícula em um determinado período, mas que não estavam matriculados no período anterior. Essa situação se aplica aos alunos que não se evadem de forma definitiva, mas que, na verdade, interrompem seus estudos por um tempo específico. Destaca-se, por fim, que apesar da definição feita pelos autores, não é possível garantir que esse cálculo represente as rematrículas, pois estão sendo usados dados consolidados, os quais, por definição, agregam a informação em um único número, impossibilitando garantir, de forma acurada, o conceito do indicador.

Em estudo posterior, o Instituto Lobo expandiu as fórmulas de evasão para considerar todas as granularidades, pelo menos de forma teórica, e mantendo a temporalidade imediata (Silva Filho; Lobo, 2012). A evasão dos cursos manteve-se a mesma da equação 2, mas a evasão das IES teve a adição do número de ingressantes que mudaram de curso, mas não de IES (ITC) e a evasão do sistema passou a considerar, além da variável ITC, também o número de estudantes que mudaram de IES via transferência (ITIES), conforme as equações 3 e 4.

$$E_i^{LOBO_2012_1} = 100 - \frac{M_i - I_i + ITC_i}{M_{i-1} - C_{i-1}} \times 100 \quad (3)$$

$$E_i^{LOBO_2012_2} = 100 - \frac{M_i - I_i + ITC_i + ITIES_i}{M_{i-1} - C_{i-1}} \times 100 \quad (4)$$

A evasão do sistema, conforme a fórmula 4, precisa do total de estudantes que se transferiram para outras IES (ITIES), dado esse que não era facilmente disponível na época (2012) e que continua de difícil acesso até hoje, especialmente no contexto dos registros gerenciais das instituições de ensino superior, as quais não têm acesso a dados de outras instituições. Essa informação só é possível de saber, com precisão, no âmbito da sala segura do Inep, ou apenas pelo Inep, que é o detentor de todos esses dados discutidos até o momento. Essa discussão será retomada no capítulo 3.

Serpa e Pinto (2000) propuseram alguns cálculos para obter a evasão do sistema, em temporalidade imediata. Partindo de três agregados (matrículas, concluintes e ingressos no vestibular), os autores chegaram à variação da matrícula de um ano para o outro, conforme equação 5, e à fórmula da evasão, conforme equação 6.

$$\Delta M_i^{S-P} = M_{i+1} - M_i = I_{i+1} - e_i - C_i \quad (5)$$

$$e_i^{S-P} = I_{i+1} - \Delta M_i - C_i \quad (6)$$

No entanto, alguns pontos devem ser destacados. Primeiramente, a fórmula 6 não trata de uma taxa de evasão, mas sim do número de evadidos, estimado pela diferença entre o total de ingressantes em um ano e a soma da variação de matriculados e concluintes do ano anterior. Os autores utilizam esse número como numerador frente às grandezas de matriculados, ingressantes e concluintes, mas sem afirmar qual seria, de fato, a evasão. Em segundo lugar, é aplicado um filtro para ingressos apenas por vestibular e não são considerados os trancamentos e flutuações anuais de outras movimentações, fato que limita a utilização e comparabilidade dos resultados. Por último, não se pode falar em evasão do sistema, pois não há detalhamento dessa informação ao se utilizar dados agregados, tratando-se, então, de uma suposição de que a evasão aferida é do sistema.

As diversas formas de mensuração da evasão indicam que não há consenso, na literatura, sobre como medir essa estatística (Cabello et al., 2018). No entanto, tanto Silva Filho et al. (2007) quanto Serpa e Pinto (2000) reconhecem que o acompanhamento das coortes seria a melhor metodologia para aferição das métricas de evasão, devido à própria natureza do método, que acompanha uma coorte de estudantes verificando seus desfechos. E é justamente isso feito no âmbito dos indicadores de fluxo da educação superior, do Inep, adaptados a partir dos conceitos discutidos na Comissão Especial de Estudos sobre a Evasão (1996).

Os indicadores de fluxo são compostos, principalmente, pelas taxas de permanência, conclusão e desistência, sendo essa última equivalente à taxa de evasão. A taxa de permanência refere-se

ao percentual de estudantes que continuam matriculados ao longo do tempo. A taxa de conclusão mede o percentual de estudantes que finalizam o curso, permitindo a análise da janela de tempo necessária para a formatura. E a taxa de desistência, ou evasão, é calculada como a porcentagem de estudantes que abandonam o curso sem o concluir. Os estudantes que faleceram ao longo do período acompanhado são desconsiderados, então, em qualquer momento do tempo, os alunos da coorte analisada só podem ter três desfechos: continuam matriculados (permanência), evadiram (desistência) ou concluíram o curso (conclusão). Essa dinâmica permite a fácil compreensão de todos os conceitos utilizados e garante a comparabilidade dos resultados obtidos.

Seu cálculo é feito por coorte de ingresso, ou seja, grupo de alunos que ingressou em determinado ano. A granularidade pode ser relativa ao curso, à IES ou ao sistema, dependendo do nível de acesso aos dados, e qualquer temporalidade pode ser calculada. Há, então, diversas vantagens para a utilização da metodologia do Inep. Sua principal desvantagem está na necessidade de acesso a dados detalhados de cada coorte que se pretende estudar, e no período necessário para acompanhar os principais desfechos, que seriam a conclusão ou evasão.

Ainda no âmbito dos indicadores de fluxo, uma discussão pertinente é de como operacionalizar o cálculo da evasão por diferentes granularidades, considerando que existem dois possíveis identificadores: o código e-Mec⁵ e o código CINE. O padrão atualmente utilizado pelo Inep é de usar o primeiro. Tal escolha é correta, mas em alguns casos pode viesar os resultados obtidos, pois um mesmo código e-Mec pode ter mais de um turno vinculado, e a literatura aponta que cursos noturnos têm evasão mais elevada, comparados aos cursos diurnos ou integrais (Lima & Fagundes, 2020; Paula, 2021). Para lidar com essa limitação, seria necessário incluir o turno do curso no agrupamento feito para calcular a evasão.

Já para calcular a evasão institucional, pelo menos isoladamente, não importa qual identificador será usado, pois será verificado se o código da IES mudou de um período para outro. No entanto, quando ocorre a evasão da IES, pode ou não ocorrer, também, a evasão do curso. Para capturar tal efeito o uso do código e-MEC não é adequado, pois cada IES tem um código diferente, mesmo que o nome do curso seja o mesmo. Ou seja, seria necessário utilizar o código CINE, juntamente com o código da IES, para conseguir capturar ambas as granularidades. No entanto, o uso do código CINE, apesar de permitir enquadrar corretamente a área dos cursos dentro do

⁵ O código e-MEC é um identificador único gerado pelo sistema e-MEC, utilizado para caracterizar a veracidade das informações prestadas sobre instituições e cursos de educação superior no Brasil. Esse código é emitido ao final de declarações de regularidade de cursos e serve como comprovação de que as informações são autênticas e válidas para todos os efeitos legais.

sistema educacional superior, desconsidera informações essenciais para a compreensão do fenômeno em estudo: o turno e a modalidade de ensino, este último já documentado na literatura como um importante fator associado a evasões mais altas no EaD, em comparação com o presencial (Bentes & Kato, 2014; Silva et al., 2025). Logo, na análise da evasão institucional, deve-se considerar o código CINE dos cursos em conjunto com os dados de turno e modalidade. Deve-se destacar que a mudança de grau incorreria em uma mudança de curso, pois os códigos CINE do bacharelado e licenciatura são diferentes.

Por fim, a evasão sistêmica tem a operacionalização mais direta dentre as granularidades abordadas, pois se trata de verificar, para cada indivíduo, se há um registro de conclusão em algum momento do tempo, em algum curso, de alguma IES, pois não há de se falar em evasão do sistema caso o estudante tenha se graduado no ensino superior, apenas em evasão do curso ou da IES. As dificuldades do cálculo estão na necessidade de registros extensivos dos alunos e na limitação do conceito “definitiva”, que, na verdade, refere-se ao período disponível dos dados.

1.3. Categorização dos determinantes da evasão

As teorias que definem o conceito de evasão no ensino superior abordam diferentes aspectos das possíveis categorias de determinantes que explicam a ocorrência do fenômeno. A combinação dos modelos teóricos de Tinto, Bean, Bourdieu, Astin e Cabrera com as variáveis disponíveis nos microdados identificados do Inep, especificamente do CES, CEB e Enem, permitiu a categorização dos determinantes da evasão em seis categorias, as quais serão detalhadas em seguida: socioeconômicos, individuais, pré-graduação, acadêmicos, institucionais e curriculares.

Os determinantes socioeconômicos têm origem em Bean e Bourdieu. O modelo de atrito de Bean (1980) considera o impacto de variáveis externas, como obrigações familiares e dificuldades financeiras, na decisão de evasão, enquanto Bourdieu (2014) explica como desigualdades de classe afetam o acesso e a permanência no ensino superior – estudantes com menor capital cultural e econômico têm dificuldades maiores para se integrar ao ambiente universitário. Variáveis como o índice socioeconômico (INSE), derivada do CEB, e escolaridade e profissão dos pais, número de moradores e renda familiar, provenientes do Enem, conseguem capturar a dimensão socioeconômica.

Cabrera e Tinto previram os determinantes individuais. O modelo de atração-ajuste de Cabrera *et al.* (1993) combina influências acadêmicas e externas, mostrando como discriminação e falta

de suporte institucional afetam a evasão. Já o modelo de integração de Tinto (1975) sugere que a adaptação acadêmica e social do estudante é fundamental para sua permanência, o que se torna especialmente relevante para grupos que historicamente enfrentam maiores desafios dentro do ambiente universitário. Esses conceitos, apesar de serem de difícil captura estruturada, podem ter como *proxy* variáveis como raça/cor, sexo, idade e PcD, provenientes do CES.

Os determinantes pré-graduação seriam fatores ligados à preparação acadêmica antes do ingresso, incluindo qualidade do ensino médio, habilidades adquiridas e expectativas formadas antes da universidade, conceitos abordados na teoria do capital cultural de Bourdieu (2014), a qual explica como diferenças na formação prévia impactam o desempenho e adaptação ao ensino superior, especialmente no Brasil, que padece de graves desigualdades estruturais. Para capturar essa dimensão, estão disponíveis as variáveis de nota no Enem, e informações de localização (urbana ou rural), tipo de ensino (regular, profissionalizante ou EJA) e dependência administrativa da escola de ensino médio, conforme o Enem e CEB.

Conceitos relacionados ao desempenho acadêmico, políticas de suporte, integração e gestão no âmbito dos cursos são caracterizados como determinantes acadêmicos. O modelo de envolvimento de Astin (1998) argumenta que maior participação ativa em atividades acadêmicas está associada a taxas menores de evasão, sugerindo que dificuldades acadêmicas podem ser mitigadas com estratégias de engajamento. Isso é corroborado pelo modelo de integração de Tinto (1975), o qual postula que ambientes que não favorecem o engajamento do estudante aumentam os riscos de evasão. Por fim, o modelo organizacional de Bean (1980) também deve ser considerado, pois sugere que a evasão é um fenômeno influenciado pela satisfação do estudante, incluindo serviços de apoio e integração social. O Inep dispõe de diversas variáveis para representar essa dimensão: forma de ingresso (vestibular, enem, sisu, etc.), reserva de vagas (cotas), atividades extracurriculares, financiamento estudantil, apoio social e existência de laboratório para o curso, todos provenientes do CES.

Os determinantes institucionais são fatores ligados à estrutura da universidade, características institucionais, infraestrutura, clima acadêmico e qualidade docente. O modelo de Tinto (1975) propõe que a permanência do estudante na universidade depende da sua integração acadêmica e social no ambiente institucional. Tal integração pode ser influenciada pelo tipo de IES – universidades públicas oferecem maiores oportunidades de suporte e envolvimento acadêmico, além de terem maior percentual de docentes com pós-graduação (Silva & Sampaio, 2022). Já o modelo de Bean (1980) sugere que a evasão é similar à decisão de um funcionário deixar um

emprego: os estudantes avaliam sua satisfação institucional e, se perceberem que os custos (financeiros, emocionais e acadêmicos) superam os benefícios, tendem a abandonar a universidade. Tais fatores podem ser identificados pelas variáveis do CES categoria administrativa, organização acadêmica, IES localizada na capital e número de servidores (docentes e técnicos), e por produtos derivados do CES, o Índice Geral de Cursos (IGC) e o percentual de docentes com pós-graduação *stricto sensu*.

Tem-se, por fim, os determinantes curriculares, relacionados à estrutura dos cursos, flexibilização curricular, carga horária e compatibilidade com as expectativas dos estudantes. O modelo de envolvimento de Astin (1998) sugere que currículos que incentivam maior participação estudantil favorecem a permanência. De forma complementar, Tinto (1975) indica que currículos excessivamente rígidos ou desalinhados com as expectativas dos alunos podem contribuir para desistências, especialmente em cursos com altas demandas técnicas. A captura desses conceitos pode ser operacionalizada por variáveis do CES: grau acadêmico, turno, modalidade, curso localizado na capital, carga horária do curso e área do conhecimento, assim como pelo Conceito Preliminar de Curso (CPC), um dos indicadores de qualidade da educação superior.

Definidas as variáveis do conjunto de dados do Inep que compõem cada determinante, pode-se proceder à proposta de estrutura taxonômica da evasão.

1.4. Proposta de estrutura taxonômica da evasão

Ao explorar detalhadamente a relação entre granularidade e temporalidade, e adicionar o efeito dos qualificadores na concepção dos diversos tipos de evasão, tem-se como resultado o Quadro 4. Essa matriz conceitual apresenta uma taxonomia abrangente que permite caracterizar o fenômeno da evasão em suas múltiplas facetas. A dimensão da granularidade estabelece três níveis de análise (curso, IES e sistema), representando os diferentes escopos onde a evasão pode ser observada e mensurada. Paralelamente, a temporalidade introduz uma perspectiva cronológica, diferenciando eventos de evasão imediata (menos de um ano e presença de nova vinculação), temporária (mais de um ano com novo vínculo) e definitiva (mais de um ano e sem registro de novo vínculo).

A interseção destas dimensões gera nove categorias analíticas distintas, cada uma com características próprias e implicações específicas. A evasão imediata do curso, por exemplo, pode representar um reajuste vocacional inicial, enquanto a evasão definitiva do sistema sugere um fenômeno mais grave de desistência completa da educação superior. Os qualificadores adicionais (mudança de curso, área do conhecimento, IES, campus ou turno/modalidade)

enriquecem a análise ao permitir identificar as nuances de cada evento de evasão, diferenciando uma simples troca de turno dentro do mesmo curso de uma completa mudança institucional, eventos que têm determinantes diferentes para explicar a evasão.

Essa estrutura taxonômica possibilita uma compreensão mais refinada do fenômeno, superando as abordagens tradicionais que frequentemente tratam a evasão como um evento uniforme. Ao decompor a evasão em suas diversas manifestações, o quadro pretende ser um robusto instrumental metodológico para pesquisadores e gestores educacionais, permitindo investigações mais precisas sobre os determinantes específicos de cada tipo de evasão e, consequentemente, o desenvolvimento de estratégias de intervenção mais eficazes e direcionadas, contribuindo com o avanço da literatura de evasão no ensino superior.

A proposição deste modelo bidimensional aperfeiçoado representa uma contribuição significativa à forma como o fenômeno da evasão tem sido tradicionalmente abordado. Esta proposta expande modelos anteriores (Vitelli & Fritsch, 2016) ao incorporar os qualificadores detalhados, permitindo uma categorização mais precisa dos eventos de mobilidade e evasão estudantil. A dimensão da temporalidade (imediate, temporária e definitiva) captura o aspecto dinâmico da trajetória estudantil, reconhecendo que mesmo períodos de afastamento podem não configurar desistência permanente.

Paralelamente, a granularidade (curso, IES e sistema) incorpora a perspectiva espacial-institucional do fenômeno, diferenciando eventos que, embora aparentemente similares (saída de um vínculo), têm implicações radicalmente distintas para os atores envolvidos. Esta distinção é particularmente relevante no contexto brasileiro contemporâneo, onde políticas como o SISU ampliaram significativamente a mobilidade estudantil entre instituições, gerando fluxos migratórios que, se analisados apenas da perspectiva da instituição de origem, poderiam ser incorretamente classificados como evasão definitiva (Cabello et al., 2021; Gilioli, 2016; Li, 2016).

Os qualificadores transversais (CO, curso, área, IES) funcionam como variáveis moderadoras que ampliam a capacidade explicativa do modelo, permitindo capturar a heterogeneidade de trajetórias dentro de cada categoria principal. Esta abordagem multifatorial reconhece que a evasão raramente tem causa única, sendo frequentemente resultado da interação entre fatores individuais, institucionais e contextuais. A mudança de turno/modalidade dentro do mesmo curso frequentemente reflete adaptações a necessidades socioeconômicas (Filho & Nery, 2009; Silva et al., 2025), enquanto a mudança simultânea de curso e área geralmente indica questões vocacionais ou acadêmicas mais profundas (Barroso et al., 2022).

Quadro 4 – Proposta de estrutura taxonômica dos conceitos de evasão por granularidade, temporalidade e qualificadores

TEMPORALIDADE	GRANULARIDADE						Sistema
	Curso			IES			
	Qualificadores						
	CO	Área	IES	CO	Curso	Área	
Imediata (até um ano)	Evasão do curso com retorno ao sistema em até um ano O aluno saiu do curso há um ano ou menos e tem novo vínculo de curso. Pode ter ocorrido uma mudança de área e/ou de IES, combinada com possível alteração de campus, turno ou modalidade.			Evasão da IES com retorno ao sistema em até um ano O aluno saiu da IES há um ano ou menos e tem novo vínculo com outra IES. Pode ter ocorrido uma mudança de curso e/ou área, combinada com possível alteração de campus, turno ou modalidade.			Evasão do sistema sem registro de retorno há até um ano O aluno não tem vínculo ativo com as IES há um ano ou menos. Todos os qualificadores são aplicáveis para caracterizar o fenômeno. A cada atualização dos dados disponíveis, este tipo específico de evasão será reclassificado em relação à granularidade e/ou temporalidade.
Temporária (maior que um ano)	Evasão do curso com retorno ao sistema após mais de um ano O aluno saiu do curso há mais de um ano e tem novo vínculo de curso. Pode ter ocorrido uma mudança de área e/ou de IES, combinada com possível alteração de campus, turno ou modalidade.			Evasão da IES com retorno ao sistema após mais de um ano O aluno saiu da IES há mais de um ano e tem novo vínculo com outra IES. Pode ter ocorrido uma mudança de curso e/ou área, combinada com possível alteração de campus, turno ou modalidade.			Evasão do sistema sem registro de retorno há mais de um ano O aluno não tem vínculo ativo com as IES há mais de um ano. Todos os qualificadores são aplicáveis. A cada atualização dos dados disponíveis, pode ocorrer a reclassificação da granularidade.
Definitiva (sem registro de retorno após a evasão)	Evasão do curso sem registro de retorno ao sistema O aluno saiu do curso há mais de um ano e não há registro de novo vínculo. Todos os qualificadores são aplicáveis para caracterizar o fenômeno. A cada atualização dos dados disponíveis, pode ocorrer a reclassificação da temporalidade.			Evasão da IES sem registro de retorno ao sistema O aluno saiu da IES há mais de um ano e não há registro de novo vínculo. Todos os qualificadores são aplicáveis para caracterizar o fenômeno. A cada atualização dos dados disponíveis, pode ocorrer a reclassificação da temporalidade.			Evasão do sistema sem registro de retorno O aluno não tem vínculo ativo com nenhuma IES há mais de um ano, nem registro de sucesso (formatura) em seu percurso na educação superior. Todos os qualificadores são aplicáveis. Supõe-se que o estudante não retornará ao ensino superior.

Fonte: elaboração própria.

Notas: CO são as Condições Operacionais que qualificam a evasão (campus, turno ou modalidade).

1.5. Considerações

A análise sobre os tipos de evasão no ensino superior revelou a complexidade do fenômeno, destacando a necessidade de abordagens integradas que considerem as dimensões de granularidade, temporalidade e qualificadores. A ausência de uma metodologia padronizada para calcular a evasão limita comparações entre estudos e dificulta a formulação de políticas públicas eficazes.

Com base na literatura e nas ideias desenvolvidas neste capítulo, pode-se afirmar que a evasão sistêmica não é apenas um indicador educacional; ela reflete falhas estruturais da sociedade em garantir oportunidades de desenvolvimento para jovens que abandonam o ensino superior definitivamente. Contudo, deve-se considerar que muitos casos inicialmente classificados como definitivos podem ser reclassificados ao longo do tempo como temporários ou imediatos, evidenciando a dinâmica fluida das trajetórias educacionais. Além disso, a ausência de um mecanismo padronizado para rastrear e confirmar a desistência definitiva do estudante no sistema educacional limita a análise. Por essas razões, optou-se por não incluir uma matriz específica para este nível de granularidade, priorizando análises mais concretas e mensuráveis nos outros níveis apresentados.

Em conjunto com a evasão do sistema, a evasão definitiva pode ser interpretada como um indicador das desigualdades sociais que afetam o acesso à educação superior no Brasil. Políticas públicas voltadas à inclusão socioeconômica devem ser priorizadas para reduzir as barreiras enfrentadas por estudantes vulneráveis.

A proposta de uma taxonomia bidimensional representa um avanço significativo para o estudo da evasão, ao permitir diagnósticos mais precisos e intervenções específicas. Para gestores educacionais, a taxonomia proporciona diagnósticos diferenciados que fundamentam intervenções específicas - desde redesenhos curriculares para mitigar evasões de curso até políticas assistenciais para reduzir evasões do sistema motivadas por fatores socioeconômicos. Para formuladores de políticas públicas, oferece indicadores com mais nuances, que podem orientar alocação de recursos e avaliação de programas segundo seus impactos em diferentes tipos de evasão.

Vislumbram-se, ainda, vantagens metodológicas substanciais no uso da taxonomia de evasão proposta. Primeiro, pode-se dizer da possibilidade de estudos comparativos mais precisos, minimizando distorções causadas por definições operacionais inconsistentes entre pesquisas. Segundo, possibilita o acompanhamento longitudinal sistemático e calcado em conceitos bem

definidos, essencial para avaliar impactos de políticas de permanência. Terceiro, fornece um modelo estruturado para sistemas de alerta precoce, ajudando instituições a identificar padrões específicos de evasão antes que se concretizem definitivamente.

Esta estrutura representa, portanto, não apenas um avanço teórico-conceitual, mas um instrumental analítico com potencial transformador para a gestão da permanência e sucesso no ensino superior brasileiro, possibilitando superar abordagens genéricas em favor de estratégias contextualizadas e preventivas baseadas em evidências. Os desafios relacionados à mensuração da evasão reforçam a importância de desenvolver metodologias padronizadas que permitam acompanhar longitudinalmente os estudantes em todas as granularidades do sistema educacional.

Há ainda fatores qualitativos que podem ser considerados na análise da evasão no ensino superior, como o alvitre, entendido como a capacidade de decisão individual influenciada por fatores subjetivos e contextuais (Mendes, 2023). Esses elementos, embora frequentemente negligenciados em estudos quantitativos, desempenham um significativo papel na compreensão das trajetórias acadêmicas, refletindo escolhas que não podem ser explicadas apenas por variáveis objetivas, e que estão profundamente enraizadas nas experiências individuais, nos valores pessoais e nas percepções subjetivas dos estudantes sobre o ambiente acadêmico e suas próprias condições de vida.

Decisões relacionadas à mudança de curso ou abandono definitivo podem estar associadas a percepções sobre a qualidade do ensino, identificação com a carreira escolhida ou até mesmo questões interpessoais e emocionais, como a sensação de pertencimento ou o impacto do adoecimento psíquico. Esses fatores qualitativos são difíceis de mensurar, mas têm implicações significativas para as políticas institucionais voltadas à retenção. Estudos apontam que a falta de identificação com o curso ou instituição, combinada com dificuldades na integração social e acadêmica, são determinantes importantes para a evasão (Lopes, 2023; Silva Filho et al., 2007).

Embora este trabalho se concentre em dimensões mais concretas e mensuráveis da evasão, como granularidade, temporalidade e seus qualificadores, é necessário reconhecer que a exclusão dos fatores qualitativos limita a compreensão integral do fenômeno. Pesquisas futuras poderiam explorar metodologias qualitativas, como histórias de vida ou entrevistas semiestruturadas, para captar as nuances e aspectos subjetivos que moldam as decisões de evadir dos estudantes.

Por fim, é essencial ampliar os esforços para integrar dados educacionais com informações provenientes de outros sistemas formativos, como o Sistema S, visando uma análise mais abrangente da evasão no contexto brasileiro. A implementação dessa abordagem pode transformar a gestão educacional ao fornecer ferramentas analíticas robustas para mitigar os impactos negativos da evasão nos níveis individual, institucional e sistêmico.

2. Evasão definitiva do sistema de ensino superior

A evasão definitiva do sistema de ensino superior caracteriza-se pelo abandono completo, por parte do estudante, de qualquer modalidade de educação universitária, sem a intenção de retornar, ou matricular-se em outra instituição ou curso de nível superior. Tal fenômeno implica que o discente não apenas desiste de um curso específico, área de estudo ou mesmo instituição, mas sim interrompe integralmente sua trajetória acadêmica no âmbito do ensino superior.

Essa definição distingue-se da evasão de curso, instituição ou área de conhecimento, na qual o estudante pode, ao longo do tempo, trocar de curso dentro da mesma instituição, transferir-se para outra universidade ou alterar sua área de estudo, permanecendo, contudo, inserido no sistema de ensino superior, apesar de terem ocorrido vários eventos distintos de evasão. Já na evasão definitiva, foco deste capítulo, há uma desvinculação permanente do estudante em relação ao sistema educacional universitário. Conceitualmente, pode-se classificar a evasão definitiva do sistema como macroevasão (Gilioli, 2016; Santos; Pilatti & Bondarik, 2022), termo que será utilizado como sinônimo neste capítulo.

No entanto, há um desafio metodológico para a correta mensuração da evasão do sistema, pois é necessário verificar um longo intervalo de tempo para ser possível classificar o último evento de evasão, de cada aluno, como evasão “definitiva”. E tal verificação só pode ser feita no âmbito de disponibilidade informacional de todo o ensino superior, abrangendo todos os alunos, algo possível apenas na sala segura do Inep. E, ainda assim, tal definição fica sujeita à restrição temporal do período analisado, pois é possível que, em momento futuro, a decisão de não cursar o ensino superior seja revertida, o que reverteria a evasão “definitiva” em outro tipo de evasão. Tal dificuldade metodológica fica evidente pela inexistência de estudos publicados que tratem do tema da macroevasão do ensino superior, no âmbito Brasil, abarcando um longo período de anos (2005 a 2022) e utilizando os microdados identificados do Inep, no contexto da sala segura. Desse forma, um diferencial adicional da contribuição deste estudo está na metodologia utilizada, de regressão logística multinível, para considerar as diferenças estruturais do fenômeno estudado entre IES, áreas de conhecimento e modalidade de ensino.

Ainda assim, foi encontrado um artigo (Marques, 2020) que tratou sobre a volta aos estudos dos alunos que evadiram do ensino superior, com base nos microdados do CES de 2009 a 2017, quando ainda eram publicamente disponíveis⁶. O autor conclui, baseando-se na análise descritiva, que uma significativa parcela dos alunos que se desligaram retornou ao ensino

⁶ Devido a restrições impostas pela Lei Geral de Proteção a Dados Pessoais, os dados foram temporariamente retirados do website do Inep e atualmente, parte deles, somente podem ser acessados pela sala segura do Inep, como será discutido adiante.

superior posteriormente. Embora não tenham optado por cursar as mesmas instituições de ensino, a maioria deles ingressou em cursos na mesma área do conhecimento do curso do qual se desligaram.

Estudo mais recente (Paula, 2021), também utilizando os microdados públicos do CES, mas atendo-se à coorte de ingresso de 2016, especificamente no contexto das universidades federais, conseguiu diferenciar a evasão do curso ou da instituição da evasão do sistema, ao verificar a situação dos alunos no ano seguinte, 2017.

Os demais estudos, mesmo quando reconhecem as diferentes medidas de evasão, não mensuram a evasão do sistema, pela já descrita dificuldade de trabalhar com esse conceito. Isso leva os pesquisadores a trabalharem com suas realidades mais próximas, como seu curso, área de conhecimento, instituição, modalidade de ensino, turno ou grau, cujos dados podem ser provenientes de sistemas acadêmicos ou pesquisas de percepção, segundo as revisões de literatura consultadas (Santos; Pereira & Pilatti, 2024; Silva et al, 2023).

Devido não só à escassez de estudos robustos sobre a macroevasão, mas também à relevância da discussão, o objetivo deste estudo é contribuir para a literatura do tema e analisar, dentre os fatores socioeconômicos, acadêmicos e institucionais, os mais associados com a saída definitiva dos discentes do sistema de ensino superior, considerado o período de ingresso de 2005 a 2022, com base nos microdados identificados do Inep. Utiliza-se aqui o conceito de evasão do sistema definitiva apresentado no Capítulo 2, conforme a proposta de estrutura taxonômica dos conceitos de evasão por granularidade, temporalidade e qualificadores (Quadro 4).

A primeira seção do capítulo apresenta a origem dos dados, o processo de tratamento, as variáveis disponíveis e as respectivas hipóteses, o método utilizado, as amostragens feitas e a análise descritiva do perfil dos estudantes. A segunda seção traz as estatísticas descritivas da macroevasão, os resultados das regressões logísticas multinível e a discussão de cada determinantes. E a última seção traz as considerações finais.

2.1. Fontes dos dados, variáveis e métodos

O desenvolvimento de políticas públicas eficazes, em educação ou quaisquer outras áreas, depende de acesso a dados detalhados e longitudinalmente consistentes. Os dados protegidos do Inep são essenciais para isso, pois oferecem uma visão abrangente e granular de diversos aspectos do sistema educacional, permitindo análises precisas de tendências e padrões ao longo do tempo. Estes dados incluem informações detalhadas sobre alunos, professores, cursos e instituições, dentre outros, cruciais para estudar e entender dinâmicas educacionais complexas.

Os microdados disponíveis publicamente no site do Inep, embora valiosos, apresentam um alto grau de agregação, limitando severamente a capacidade de realizar análises finas, como o acompanhamento longitudinal dos alunos para entender padrões de evasão e permanência no ensino superior. A indisponibilidade de detalhes individuais impede, por exemplo, a análise de trajetórias educacionais antes e após a admissão no ensino superior, um dos focos deste estudo. Além disso, os dados agregados publicados, tais como quantidades de ingressantes, matriculados e concluintes por curso, não permitem calcular taxas de evasão de maneira precisa, pois não é possível acompanhar os alunos individualmente através dos seus percursos acadêmicos específicos. Os indicadores de trajetória disponíveis publicamente limitam-se a métricas por curso e ano de ingresso, sem permitir a análise de outros determinantes críticos, como proposto neste projeto, em escala nacional.

Grande parte dos estudos sobre evasão no ensino superior no Brasil utiliza dados de sistemas internos e de cursos específicos, devido à limitada disponibilidade de dados identificados em larga escala e às restrições impostas pela Lei Geral de Proteção de Dados⁷ (LGPD). Essa lei garante a privacidade dos indivíduos, restringindo o uso de informações pessoais e, consequentemente, o escopo das pesquisas. Apesar dessas limitações serem necessárias para proteger dados pessoais, a necessidade de acesso a dados identificados permanece, desde que seguindo protocolos de segurança adequados, para enriquecer significativamente as análises educacionais e influenciar positivamente políticas públicas e práticas que melhorem, de forma geral ou específica, a educação no país.

Por esse motivo, foi necessária a utilização da sala segura do Inep, para ter acesso aos microdados identificados do CES, Enem e CEB, sendo toda a manipulação, tratamento e modelagem dos dados feitos nesse ambiente seguro.

Neste estudo, é utilizada a Classificação Internacional Normalizada da Educação (CINE) para os cursos de graduação. A CINE, desenvolvida pela UNESCO, oferece uma estrutura padronizada para categorizar os programas educacionais, facilitando comparações nacionais e internacionais e assegurando uma maior consistência na análise dos dados educacionais, evitando problemas de rastreamento das trajetórias dos alunos, que ocorreriam ao usar o código e-Mec do curso, por exemplo (Zalaf Caseiro & Ramos De Azevedo, 2019). Apesar do foco aqui não estar nos cursos ou áreas, é importante salientar tal escolha, considerando que não foram encontrados estudos que destacassem o uso da CINE, no contexto da evasão do ensino superior. A adoção dessa classificação só tem a contribuir para a literatura, favorecendo uma análise mais

⁷ https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm. Acesso em janeiro de 2025.

precisa dos padrões de evasão e de outros indicadores, além de contribuir para a harmonização dos resultados em estudos comparativos.

2.1.1. Fontes dos dados

O Censo da Educação Superior é a principal base deste estudo, por conter as informações de vínculo dos estudantes na graduação, sendo composta pelos conjuntos de dados de discentes, docentes, cursos e instituições. Estavam disponíveis os anos-base 2009 a 2022, totalizando 153.360.577 registros. No entanto, o CES 2009 não foi utilizado, por ser a primeira aplicação do instrumento e haver questionamentos sobre a consistência dos dados coletados (Silva, Cabral & Pacheco, 2016), restando 146.378.559 registros.

Devido ao escopo, o foco da pesquisa está na situação final do aluno em cada curso cursado, no sentido de saber se ele evadiu ou não. Dessa forma, foram filtrados apenas os discentes com situação de vínculo iguais a “Desvinculado do curso” ou “Formado”, correspondendo a 39.704.911 registros. Além disso, foram desconsiderados os discentes dos cursos sequenciais de formação específica, pela baixa frequência (apenas 88.109 registros). Devido às características dos cursos ABI⁸, nos quais não é possível formar, foram filtrados fora tais casos (61.879), ficando a base com 39.554.926 observações. Por fim, foram mantidos os ingressantes de 2005 em diante, por conta da reduzida quantidade de ingressos anteriores à 2005, finalizando a base com 39.195.072 registros de evasão ou formatura.

As informações de apoio social e atividade extracurricular foram tratadas separadamente, pois como foi aplicado um filtro para a situação de finalização do discente, em um determinado curso e ano-base do Censo, tal dado fica incompleto, devido ao fato que um determinado aluno pode ter tido apoio social ao longo de todo seu percurso até a formatura, informação essa que deve ser considerada na modelagem estatística. Dessa forma, foi feita a contagem das variáveis de apoio social e atividades extracurriculares para cada combinação de aluno-curso, para ter o número de períodos de participação nesses indicadores, os quais foram posteriormente agregados à base principal.

Às informações de alunos foram adicionados detalhes da base de cursos, com variáveis sobre a oferta do curso na capital, código CINE⁹, grau acadêmico, carga horária e totais de inscritos, vagas e ingressos. Da base de IES foram coletadas as variáveis de categoria administrativa, organização acadêmica, região, Unidade da Federação (UF) e total de servidores técnico-

⁸ Área Básica de Ingresso – representa um ingresso único para mais de um curso, comum para cursos de licenciatura e bacharelado que compartilham conteúdos ou ainda cursos de engenharia com disciplinas iniciais comum, por exemplo.

⁹ Classificação Internacional Normalizada da Educação é um sistema de classificação para cursos de graduação baseado em metodologia da Unesco.

administrativos (TAE). Por fim, da base de docentes, foi extraído o total de docentes efetivos em exercício em cada ano-base do Censo e em cada IES, para combinar com o total de TAEs e ter o total de servidores de cada IES em cada edição do CES.

Os microdados do Enem foram utilizados para melhor qualificar o perfil do estudante antes de ingressar no ensino superior, para ter dados socioeconômicos (escolaridade e ocupação dos pais, renda familiar e número de pessoas na residência) e informações da escola de ensino médio na qual se formou (escola pública ou privada, ensino regular, EJA ou profissionalizante, dependência administrativa e localização), provenientes do questionário socioeconômico e autodeclaradas pelos inscritos no momento da inscrição. A nota das quatro áreas de conhecimento da prova foi, também, utilizada. A junção das informações do Enem ao CES foi feita seguindo uma heurística de sucessivas tentativas, iniciando pelo ano imediatamente anterior de ingresso no ensino superior, de cada aluno, até esgotar as possibilidades de pareamento.

O Censo da Educação Básica serviu para melhorar a qualidade informacional de algumas variáveis de interesse que já existiam no CES. Considerando apenas dados do CES, a informação de cor/raça autodeclarada tinha 62,5% de preenchimento. Ao complementar tal variável com dados do Enem e do CEB, o preenchimento saltou para 77,7%, considerando o registro mais recente de disponibilidade de raça/cor. Importante ressaltar que há uma discussão sobre a qualidade informacional e estratégias para melhorá-la (Senkevics, Machado & Oliveira, 2016), que mostra como um mesmo dado tem diferentes perfis dependendo do instrumento analisado. As variáveis de aluno pessoa com deficiência (PcD) e tipo de escola de ensino médio também foram aprimoradas, porém com melhorias menos expressivas, passando de 0,46% a 0,56% e 83% a 86,2%, respectivamente.

Além disso, foi extraída do CEB a variável de código da escola de ensino médio, para permitir sua junção com o Indicador de Nível Socioeconômico (INSE), a qual permite ter informação adicional do nível socioeconômico das escolas dos estudantes, quando do ingresso no ensino superior, do período de 2011 a 2022.

Foram utilizados os microdados de dois dos indicadores de qualidade da educação superior: Índice Geral de Cursos (IGC) e Conceito Preliminar de Curso (CPC), para ter essa informação por curso e IES, de 2009 em diante. No caso do IGC, foram utilizados, também, as notas parciais de graduação, mestrado e doutorado, conforme metodologia de construção desse índice.

A base do percentual de docentes com pós-graduação *stricto sensu* foi também agregada à base principal, para ter esse dado por curso e IES, ao longo dos anos, de 2010 em diante. Para

finalizar a montagem da base principal, foram adicionadas duas informações: dados do Atlas do Desenvolvimento Humano (ADH), o qual traz o Índice de Desenvolvimento Humano (IDH) a nível municipal (IDHM), de 2010 a 2022; e dados sobre o Reuni, contendo as IFES que aderiram ao programa e seu ano de ingresso.

2.1.2. Variáveis e hipóteses

A principal variável do estudo é a evasão do sistema, ou macroevasão, que é a variável dependente. As variáveis independentes, representando os determinantes da macroevasão, aqui categorizados em socioeconômicos, individuais, pré-graduação, acadêmicos, institucionais e curriculares, serão descritas em seções específicas.

2.1.2.1. Variável dependente

A macroevasão (Santos, Pilatti & Bondarik, 2022), definida pelo evento de ingresso no ensino superior atrelado à evasão sem evento de reingresso, durante o período de acompanhamento do estudo, que compreende os anos-base do CES de 2010 a 2022 e coortes de ingresso de 2005 a 2022, é uma variável dicotômica de valor 1, caso todos os registros de um determinado aluno sejam de desvinculação, ou 0 caso contrário.

2.1.2.2. Variáveis independentes

Para atingir o objetivo de avaliar o impacto dos determinantes socioeconômicos, individuais, pré-graduação, acadêmicos, institucionais e curriculares na evasão do sistema, diversas variáveis foram estudadas.

- i. **Determinantes socioeconômicos** – variáveis como o índice socioeconômico, o nível educacional dos pais, o tamanho da família, a renda familiar e a ocupação dos pais podem influenciar a probabilidade de evasão dos alunos no ensino superior, por impactar sua estabilidade e capacidade de se manterem engajados nos estudos.
 - a. **INSE:** representa o índice socioeconômico da escola de ensino médio dos alunos, quando disponível, ou do município de origem. Alunos de menor índice socioeconômico, conforme a base de dados do INSE, têm maior probabilidade de evadir do ensino superior (Costa & Picanço, 2020; Farias, Gouveia & Almeida, 2024; Gama, 2018; Lima & Fagundes, 2020; Martins et al, 2015; Silva, 2013; Teodoro & Kappel, 2020);
 - b. **Escolaridade parental:** alunos cujos pais têm menos anos de estudo, conforme o questionário do Enem, têm maior probabilidade de evadir do ensino superior (Bayma-Freire, Roazzi & Roazzi, 2015; Costa & Picanço, 2020);

- c. **Ocupação parental:** alunos cujos pais têm ocupações que não exigem ensino superior, conforme o questionário do Enem, têm maior probabilidade de evadir do ensino superior (Bayma-Freire, Roazzi & Roazzi, 2015; Costa & Picanço, 2020);
 - d. **Número de moradores:** alunos provenientes de famílias maiores, conforme o questionário do Enem, têm maior probabilidade de evadir do ensino superior (Coimbra, Silva & Costa, 2021; Costa & Picanço, 2020);
 - e. **Renda familiar:** alunos provenientes de famílias de menor renda, conforme o questionário do Enem, têm maior probabilidade de evadir do ensino superior (Costa & Picanço, 2020; Farias, Gouveia & Almeida, 2024; Paula, 2021).
- ii. **Determinantes individuais** – a idade no ingresso, raça/cor, ser pessoa com deficiência (PcD) e sexo dos alunos são fatores individuais que podem afetar as taxas de evasão no ensino superior, refletindo como características individuais podem influenciar a permanência estudantil.
 - a. **Raça/cor autodeclarada:** alunos de raça/cor branca, conforme o CES, têm menor probabilidade de evadir do ensino superior (Araújo, Mariano & Oliveira, 2021; Arcanjo, 2022; Gomes & Hirata, 2024; Massini-Cagliari et al., 2021; Nierotka, Salata & Klitzke Martins, 2023);
 - b. **Sexo:** alunas do sexo feminino, conforme o CES, têm menor probabilidade de evadir do ensino superior (Costa; Bispo; Pereira, 2018; Li, 2016; Lopes Et Al., 2023; Nierotka, Salata & Klitzke Martins, 2023; Saccaro, França & Jacinto, 2019; Silva, 2018);
 - c. **Idade no ingresso:** alunos de menor idade no ingresso, conforme o CES, têm menor probabilidade de evadir do ensino superior (Lopes et al., 2023; Martins et al., 2015; Saccaro, França & Jacinto, 2019; Silva, 2013; 2018);
 - d. **PcD:** alunos com algum tipo de deficiência, conforme o CES, têm maior probabilidade de evadir do ensino superior (Pereira et al., 2020).
- iii. **Determinantes pré-graduação** – o desempenho no Enem, o contexto socioeconômico e educacional das escolas de nível médio, assim como a localização dos alunos, sugerem uma correlação significativa com a evasão no ensino superior, indicando que o preparo pré-universitário é crucial para a permanência estudantil.

- a. **Nota Enem:** estudantes que tiveram alto desempenho no Enem, conforme os dados do Enem, são menos suscetíveis à evasão no ensino superior (Nierotka, Salata & Klitzke Martins, 2023);
 - b. **Localização da escola do ensino médio:** alunos provenientes de escolas de nível médio rurais, conforme o questionário do Enem, têm maior probabilidade de evadir do ensino superior;
 - c. **Tipo de ensino da escola do ensino médio:** alunos provenientes de escolas de nível médio de ensino regular, conforme o questionário do Enem, têm menor probabilidade de evadir do ensino superior (Silva et al, 2020);
 - d. **Dependência administrativa de escola do ensino médio:** alunos provenientes de escolas de nível médio públicas, conforme o questionário do Enem, CEB ou CES, têm maior probabilidade de evadir do ensino superior (Mello et al, 2013).
- iv. **Determinantes acadêmicos** – a participação em atividades extracurriculares, a forma de ingresso, o uso de financiamento estudantil e o recebimento de apoio social podem influenciar as taxas de evasão no ensino superior, destacando a importância do engajamento e suporte acadêmico na retenção dos alunos.
- a. **Ingresso Enem ou Sisu:** alunos que ingressaram pelo Enem ou Sisu, conforme o CES, têm maior probabilidade de evadir do ensino superior (Cabello et al, 2021; Li, 2016; Teodoro & Kappel, 2020);
 - b. **Ingresso avaliação seriada:** alunos que ingressaram por processos de avaliação seriada, como o PAS, segundo o CES, têm menor probabilidade de evadir do ensino superior;
 - c. **Ingresso cotas:** alunos que ingressaram por reserva de vagas (cotas), conforme o CES, têm menor probabilidade de evadir do ensino superior (Lopes, 2023; Nangino & Paiva, 2022);
 - d. **Atividades extracurriculares:** a falta de envolvimento em atividades extracurriculares, conforme o CES, está significativamente relacionada com a evasão do ensino superior (Barroso et al., 2022; Farias, Gouveia & Almeida, 2022; Paula, 2021; Teodoro & Kappel, 2020; Wegner, 2022);
 - e. **Financiamento estudantil:** alunos que utilizaram algum tipo de financiamento estudantil, em especial do Prouni, conforme o CES, têm menor probabilidade de evadir do ensino superior (Gomes & Hirata, 2024; Lepine, 2019; Saccaro, França & Jacinto, 2019; Silva & Santos, 2017; Silva, 2013);

- f. **Apoio social:** alunos que receberam algum tipo de apoio social, conforme o Censo da Educação Superior, têm menor probabilidade de evadir do ensino superior (Araújo, Mariano & Oliveira, 2021; Colpo, Primo & Aguiar, 2024; Gama, 2018; Li, 2016; Lima & Fagundes, 2020, 2020; Nierotka, Salata & Klitzke Martins, 2023; Paula, 2021);
 - g. **Laboratório:** a presença de laboratórios está relacionada a uma maior permanência e menor evasão (Parente, 2014; Saccaro, França & Jacinto, 2019);
 - h. **Pandemia:** alunos que concluíram (formatura ou desligamento) nos anos da pandemia de Covid-19 (2020, 2021 e 2022) não têm padrão significativamente diferente na evasão total (Colpo, Primo & Aguiar, 2024).
- v. **Determinantes institucionais** – a categoria administrativa, a organização acadêmica, a localização geográfica, o tamanho, a qualificação docente e o desempenho institucional das IES podem influenciar a evasão no ensino superior, refletindo como a qualidade e acessibilidade do ensino oferecido são cruciais para o sucesso dos alunos.
- a. **Categoria administrativa:** alunos de IES de categoria administrativa privada, conforme o CES, têm maior probabilidade de evadir do ensino superior (Saccaro, França & Jacinto, 2019; Silva & Sauaia, 2014);
 - b. **Organização acadêmica:** alunos de IES de organização acadêmica universidade, conforme o CES, têm menor probabilidade de evadir do ensino superior (Cunha et al., 2015);
 - c. **IDH da UF:** alunos de IES sediadas em municípios de baixo IDH, conforme o IDHM, têm maior probabilidade de evadir do ensino superior;
 - d. **IES fora da capital:** alunos de IES sediadas fora da capital do Estado, conforme o Censo da Educação Superior, têm maior probabilidade de evadir do ensino superior;
 - e. **Número de servidores:** alunos de IES maiores, em relação ao número de servidores (docentes e técnico-administrativos), conforme o CES, têm menor probabilidade de evadir do ensino superior;
 - f. **Docentes pós-graduados:** alunos de IES com maior percentual de docentes com pós-graduação, conforme o CES, têm menor probabilidade de evadir do ensino superior;
 - g. **IGC:** alunos de IES com maior valor do Índice Geral de Cursos (IGC), conforme os Indicadores de Qualidade da Educação Superior, têm menor probabilidade de

evadir predo ensino superior (Araújo, Mariano & Oliveira, 2021; Silva & Santos, 2017; Silva & Sauaia, 2014);

- h. **Reuni:** alunos de Universidades Federais que aderiram ao Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais (Reuni), que ingressaram entre 2008 e 2012, conforme o CES, têm maior probabilidade de evadir do ensino superior (Prestes & Fialho, 2018).
- vi. **Determinantes curriculares** – a carga horária, turno, localização do curso, área, qualidade e natureza acadêmica dos cursos podem influenciar as taxas de evasão no ensino superior, destacando como a estrutura e o contexto do curso podem impactar a retenção de alunos.
- a. **Grau:** alunos de cursos de grau acadêmico licenciatura, conforme o CES, têm maior probabilidade de evadir do ensino superior (Assís et al., 2022; Paula, 2021);
 - b. **Turno:** alunos de cursos de turno noturno, conforme o CES, têm maior probabilidade de evadir do ensino superior (Lima & Fagundes, 2020; Paula, 2021);
 - c. **Modalidade:** alunos do EaD, segundo o CES, têm maior probabilidade de evadir do ensino superior (Bentes & Kato, 2014; Silva et al, 2025);
 - d. **Curso fora da capital:** alunos de cursos ofertados fora da capital, conforme o CES, têm maior probabilidade de evadir do ensino superior;
 - e. **Carga horária:** alunos de cursos com carga horária excessiva ou mal distribuída, conforme o CES, têm maior probabilidade de evadir do ensino superior (Teodoro & Kappel, 2020);
 - f. **Demanda:** alunos de cursos de alta demanda (relação entre o total de inscritos e vagas), conforme o CES, têm menor probabilidade de evadir do ensino superior (Hoed, Ladeira & Leite, 2018);
 - g. **Ocupação:** alunos de cursos de alta ocupação (relação entre o total de ingressantes e vagas), conforme o CES, têm menor probabilidade de evadir do ensino superior (Hoed, Ladeira & Leite, 2018);
 - h. **STEM:** alunos dos cursos das áreas de ciência, tecnologia, engenharia e matemática (*science, technology, engineering and mathematics* – STEM), conforme a classificação CINE, têm maior probabilidade de evadir do ensino superior (Mello et al, 2013);

- i. **CPC:** alunos de cursos com maior valor do Conceito Preliminar de Curso (CPC), conforme os Indicadores de Qualidade da Educação Superior, têm menor probabilidade de evadir do ensino superior.

2.1.3. Estratégia empírica

A análise da influência dos determinantes socioeconômicos, individuais, pré-graduação, acadêmicos, institucionais e curriculares na evasão do sistema foi operacionalizada por meio de regressões logísticas, estimadas pelo método de máxima verossimilhança (*maximum likelihood*), devido à característica dicotômica da evasão. Os dados educacionais, no entanto, dificilmente podem ser caracterizados por terem sua unidade de análise, alunos, independentes entre si, devido à inerente estrutura hierárquica de instituições de ensino, áreas de conhecimento, cursos e turmas (Theobald, 2018).

Por essa razão, para captar a influência dos fatores em diferentes níveis de agregação e considerar a estrutura hierárquica dos dados – estudantes aninhados em cursos e instituições – foi empregada a regressão logística multinível. Este método estende a regressão logística tradicional ao permitir a inclusão de variáveis em múltiplos níveis (individual, de curso e institucional), possibilitando a modelagem simultânea dos efeitos individuais e contextuais sobre a probabilidade de evasão. A abordagem multinível, também estimada pelo método de máxima verossimilhança, é especialmente adequada ao contexto deste estudo, quando há dependência entre as observações dentro dos agrupamentos, aprimorando a precisão das estimativas e proporcionando uma análise mais robusta dos determinantes da evasão no sistema de ensino superior.

A base de dados utilizada neste estudo, composta pelos registros de saída (evasão ou formatura) de todos os discentes registrados no CES 2010 a 2022, das coortes de ingresso de 2005 em diante, podem ser definidos como de classificação cruzada (*cross-classified models*), por não serem completamente aninhados (Browne, Goldstein & Rasbash, 2001). Um mesmo discente, ao longo do período considerado, pode ter cursado diversos cursos em diferentes IES, formando ou evadindo nesses vínculos. E ao contabilizar os registros, verifica-se que 33,9% dos alunos tiveram mais de um vínculo no CES. Portanto, ao analisar os dados, é fundamental levar em conta esses múltiplos vínculos para evitar vieses nas estimativas e obter uma compreensão mais precisa dos padrões de evasão, em particular da macroevasão.

A utilização dos modelos de regressão logística multinível ainda é incipiente no contexto brasileiro e no tema da evasão na educação superior. Devido a isso, a definição dos níveis da regressão logística multinível foi feita baseando-se em estudos internacionais (Oliveira &

Barbosa, 2023) e resultados empíricos das taxas de evasão do sistema (Seção 2.2.1.), o que resultou na especificação de três níveis: i) individual, representando as informações básicas dos alunos; ii) curso, representando os detalhes de vínculo de cada registro dos alunos; e iii) instituição e modalidade, a combinação de IES e modalidade (presencial ou EaD), representando as características institucionais, dado um curso e modalidade, de cada registro dos alunos. Além disso, controla-se pelo ano de ingresso, inserindo-a como efeito fixo e aleatório (no nível das instituições e modalidades), para considerar a mudança de perfil dos ingressantes do ensino superior brasileiro que vem sendo observada nas últimas décadas (Senkevics, 2021b). Tal estruturação metodológica reforça, mais uma vez, o caráter inovador do estudo.

O procedimento de estimação da regressão logística multinível consiste em iniciar com as estimativas do modelo nulo, que não inclui variáveis preditoras, mas apenas o intercepto e os efeitos aleatórios associados aos níveis hierárquicos do modelo. Neste modelo inicial, a variância total da variável dependente é decomposta em componentes atribuíveis a cada nível (individual, curso e instituição), permitindo avaliar a magnitude da variabilidade entre os grupos em comparação com a variabilidade dentro dos grupos (Snijders & Bosker, 2011).

A relevância da variância atribuída aos níveis hierárquicos é mensurada pelo coeficiente intraclass (ICC). O ICC indica a proporção da variância total devida às diferenças entre os grupos, em vez de diferenças individuais. No contexto de modelos logísticos multinível, devido à natureza dicotômica da variável dependente, o cálculo do ICC requer uma adaptação, já que a variância no nível individual é fixa e igual a $\frac{\pi^2}{3}$ (variância padrão da distribuição logística).

Um ICC maior que 0,10 (ou 10%) é considerado significativo no contexto educacional (Snijders & Bosker, 2011). Isso implica que pelo menos 10% da variação na probabilidade de evasão é atribuível às diferenças entre os grupos (cursos, instituições), justificando o uso de um modelo multinível. Caso o ICC seja inferior a esse valor, a variabilidade entre grupos pode ser considerada negligenciável, e um modelo de regressão logística tradicional poderia ser adequado.

Após confirmar que a variância atribuída aos níveis hierárquicos é relevante por meio do modelo nulo e do ICC, prossegue-se com a estimativa de modelos incrementais. São adicionadas, gradualmente, as variáveis explicativas referentes a cada nível – individual, curso e instituição – permitindo avaliar o impacto de cada conjunto de fatores na evasão do sistema. Este procedimento culmina na elaboração do modelo final, que incorpora os determinantes de

cada nível hierárquico e fornece uma compreensão abrangente dos fatores que influenciam a evasão do sistema no ensino superior.

A especificação simplificada do modelo base (modelo nulo com a inserção das variáveis ano de ingresso e modalidade) está na Equação 1:

$$\text{logit}(P(y_{ijk} = 1)) = \beta_{0j} + \beta_1 \text{AnoIngresso}_{ij} + u_{0j} + u_{1j} \text{AnoIngresso}_{ijk} + v_{0k} + \varepsilon_{ijk} \quad (7)$$

Onde $\text{logit}(P(y_{ijk} = 1))$ é o logaritmo da razão de chances de evasão para o indivíduo i , β_0 o intercepto fixo do modelo, β_1 o coeficiente fixo associado à variável ano de ingresso, u_{0j} o efeito aleatório do intercepto para o grupo j (instituição e modalidade de ensino), u_{1j} o efeito aleatório do coeficiente de ano de ingresso para o grupo j , v_{0k} o efeito aleatório do intercepto para o grupo k (curso) e ε_{ijk} o termo de erro. Assume-se que os efeitos aleatórios seguem distribuições Normais com média zero e variância σ^2 .

2.1.4. Amostragem

O tamanho da base final, 39.195.072 observações, juntamente com a complexidade computacional da técnica escolhida e as restrições impostas pela utilização da sala segura nas dependências do Inep, impossibilitou o uso de todo o conjunto de dados para a estimativa dos modelos de regressão logística multinível. Diante desse desafio, foi necessário planejar o procedimento amostral, para reduzir significativamente o número de registros mantendo a representatividade do novo conjunto de dados.

Tal dificuldade computacional é bem retratada na literatura – de acordo com estudos de simulação do tamanho de amostra suficiente para aplicação da regressão multinível, amostras maiores que 50 observações por grupo, considerando o maior nível, são necessárias para que as estimativas não sejam viesadas (Maas & Hox, 2005). Estudo mais recente, no qual ainda mais cenários foram simulados, concluiu que não só o tamanho da amostra por grupo é importante, mas também o número de grupos, postulando que os dados devem ter pelo menos 120 grupos, cada um com pelo menos 70 observações cada, para garantir estimativas acuradas dos efeitos fixos, aleatórios e seus respectivos erros-padrão (Ali et al., 2019). Corroborando com esses achados, também via simulação, estudo recente encontrou que, em distribuições moderadamente assimétricas (o que não é o caso aqui – ver Tabela 1), um conjunto de dados com pelo menos 30 grupos e mínimo de 100 registros cada obtém estimativas com mais de 90% de poder estatístico (Olvera Astivia, Gadermann & Guhn, 2019).

Seguindo as recomendações dos estudos supracitados, foi adotado um processo de amostragem estratificada em dois estágios. No primeiro estágio, a seleção foi realizada entre as IES que possuíam cursos com pelo menos 120 estudantes, abrangendo 98,6% do universo de IES consideradas. Dentro desse grupo majoritário, procedeu-se à seleção aleatória de 10% das IES, assegurando uma representatividade das diferentes características institucionais presentes no sistema educacional brasileiro. Essa abordagem permitiu captar a diversidade das IES em termos de localização geográfica, porte, natureza administrativa, oferta de cursos e demais características institucionais.

No segundo estágio, dentro de cada combinação de IES e curso, foram estabelecidos dois conjuntos de dados. O primeiro conjunto (amostra 120n ou enxuta) incluiu até 120 estudantes por combinação, utilizando apenas as variáveis com informações completas e oriundas do CES, totalizando 328.200 observações. O segundo conjunto (amostra 240n ou expandida), devido ao elevado número de valores ausentes nas variáveis provenientes do CEB e do Enem, ampliou o limite para até 240 estudantes por combinação de IES e curso, totalizando 397.231 observações. Essa ampliação do tamanho amostral visou compensar a perda de observações causada pelos dados faltantes, garantindo um número suficiente de casos para a modelagem estatística.

A análise descritiva dos três conjuntos de dados (base completa, Amostra 120n e Amostra 240n), presente na Tabela 1, mostra que as médias são, respectivamente, 44,96, 42,42 e 47,95, indicando taxas de evasão do sistema semelhantes entre eles. Os erros-padrão aumentam nas amostras em comparação com a base completa, refletindo o menor tamanho amostral. Os intervalos de confiança de 95% para as médias são estreitos, sugerindo estimativas precisas: [44,94 - 44,98] na base completa, [42,25 - 42,59] na Amostra 120n e [47,79 - 48,10] na Amostra 240n. As assimetrias próximas de zero indicam distribuições aproximadamente simétricas em todos os conjuntos. As curtoses negativas sugerem que as distribuições são platicúrticas, com caudas menos acentuadas do que a distribuição normal. Esses resultados indicam que as amostras mantêm as características estatísticas essenciais da base completa, reforçando a adequação do procedimento de amostragem adotado.

Tabela 1 - Comparação das estatísticas descritivas das bases utilizadas. Brasil. Coortes 2005 a 2022.

Estatística	Evasão do sistema		
	Base completa	Amostra 120n	Amostra 240n
Frequência	39.195.072	328.200	397.231
Evasão do sistema			
Assimetria	0,15	0,31	0,08
Curtose	-1,88	-1,91	-1,99

Média	44,96	42,42	47,95
Desvio-padrão	49,81	49,42	49,96
Erro-padrão	0,01	0,09	0,08
Intervalo de confiança 95%	[44,94 - 44,98]	[42,25 - 42,59]	[47,79 - 48,1]

Fonte: Elaboração própria a partir da base do CES complementada com dados do CEB e Enem (Inep).

Nota: Amostra 120n utiliza apenas variáveis do CES; Amostra 240n utiliza variáveis do CES, CEB e Enem.

Apesar dos cuidados metodológicos adotados no processo de amostragem, existe a possibilidade de viés de seleção nas amostras utilizadas no estudo. O primeiro estágio, ao selecionar apenas IES com cursos de pelo menos 120 estudantes, pode ter excluído sistematicamente instituições menores ou mais recentes, potencialmente introduzindo um viés em favor de IES mais estabelecidas ou de maior porte. Além disso, a seleção aleatória de 10% das IES nesse grupo, embora busque representatividade, pode não capturar completamente a heterogeneidade do sistema educacional brasileiro. No segundo estágio, o estabelecimento de limites de 120 e 240 estudantes por combinação de IES e curso nas amostras 120n e 240n, respectivamente, pode ter resultado em uma sub-representação de cursos menores ou menos populares. Adicionalmente, a ampliação da amostra 240n para compensar dados faltantes do CEB e Enem pode ter introduzido vieses relacionados às características dos estudantes com informações completas nessas bases. Embora a análise descritiva sugira que as amostras mantêm características estatísticas similares à base completa, é importante reconhecer que sutis diferenças nas distribuições podem impactar as inferências feitas a partir dos modelos de regressão logística multinível.

2.1.5. Perfil dos ingressantes de 2005 a 2022 com vínculo finalizado

Antes de discorrer sobre os resultados da modelagem, é importante verificar o perfil dos alunos presentes na base, com relação às características individuais, dos cursos e das instituições. A Tabela 2 descreve o perfil dos discentes, na qual é possível notar que algumas variáveis têm a categoria “sem informação”, que foi mantida para evidenciar uma das fragilidades dos dados educacionais, especialmente grave em variáveis-chave para as políticas públicas como raça/cor (Senkevics, Machado & Oliveira, 2016). Outras variáveis têm, também, alto percentual de falta de informação (localização, dependência administrativa e tipo de ensino da escola de ensino médio, escolaridade e ocupação dos pais e faixa do INSE), o que é justificado pelo fato dessas informações serem provenientes do CEB e Enem – apenas uma parte dos egressos do ensino superior, das coortes de 2005 a 2022, fizeram Enem ou têm registro no CEB.

Como já destacado, observa-se que a variável raça/cor autodeclarada possui um alto percentual de registros sem informação (21,8%), o que pode comprometer a análise dessa dimensão.

Utilizando os percentuais válidos, a maioria dos estudantes se autodeclara branca (52,4%) ou parda (35,6%). Em relação ao sexo, há predominância feminina (56,5%). De acordo com os dados do CES, a maioria estudou em escolas privadas no ensino médio (71,4%); no entanto, de acordo com o CEB, a dependência administrativa das escolas de ensino médio era privada em apenas 31,3% dos casos válidos, o que demonstra uma possível inconsistência. No que diz respeito ao tipo de ensino médio cursado, a modalidade regular é a mais frequente (77,6%), embora mais da metade dos registros não contenham essa informação. E apenas 2,1% dos estudantes eram provenientes de escolas secundárias áreas rurais. Há uma lacuna significativa de informações sobre a localização, dependência administrativa e tipo de ensino das escolas secundárias (51,3%, 58,4% e 56,5%, respectivamente).

Quanto à escolaridade dos parentes, verifica-se que cerca de 40% das informações estão ausentes, mas entre os registros disponíveis, o percentual válido de pais com ensino superior é substancialmente maior que o de mães (39,3% e 23,8%, respectivamente). A ocupação dos parentes também apresenta um alto percentual de dados faltantes (cerca de 59%), mas indicam o baixo percentual em ocupações de nível superior – 4,5% das mães e 6,1% dos pais. A faixa do INSE tem o maior percentual de falta de informação, 62,1%, e apenas 7% dos dados válidos são do nível VII¹⁰, corroborando com a escolaridade e ocupação dos pais. As faixas etárias no ingresso e saída indicam uma distribuição heterogênea. Os dados mostram o perfil de quem acessou o ensino superior e as fragilidades apontadas evidenciam limitações importantes para análises mais detalhadas e para a formulação de políticas públicas.

Tabela 2 – Frequência, percentual total e percentual válido das características individuais dos discentes. Brasil. Coortes 2005 a 2022.

Variável	Frequência	Percentual	Válidos
Raça/cor autodeclarada			
Parda	10.902.018	27,8%	35,6%
Branca	16.050.867	41,0%	52,4%
Preta	2.754.831	7,0%	9,0%
Indígena	227.430	0,6%	0,7%
Amarela	724.875	1,8%	2,4%
Sem informação	8.535.051	21,8%	
Sexo			

¹⁰ Neste nível, os estudantes estão de um a dois desvios-padrão acima da média nacional do Inse. A maioria dos estudantes respondeu ter em sua casa três ou mais quartos para dormir, dois ou mais computadores, garagem, mesa para estudar, *wi-fi*, máquina de lavar roupa, TV por internet, freezer, aspirador de pó, forno de micro-ondas e três ou mais celulares com internet. Algumas respostas não obtiveram maioria, mas indicam que parte dos estudantes afirmou possuir dois ou mais banheiros, um ou mais carros, duas ou mais televisões, uma ou mais geladeiras e escolaridade da mãe (ou responsável) e/ou do pai (ou responsável) variando entre ensino médio e ensino superior completo.

Variável	Frequência	Percentual	Válidos
Feminino	22.144.447	56,5%	56,5%
Masculino	17.050.625	43,5%	43,5%
Nacionalidade			
Brasileira	39.098.650	99,8%	99,8%
Estrangeira	96.422	0,2%	0,2%
Aluno PcD			
Sim	221.742	0,6%	0,6%
Não	38.239.898	97,6%	99,4%
Sem informação	733.432	1,9%	
Escola de ensino médio cursada			
Privada	24.281.294	61,9%	71,4%
Pública	9.727.354	24,8%	28,6%
Sem informação	5.186.424	13,2%	
Localização da escola de ensino médio cursada			
Urbana	18.691.898	47,7%	97,9%
Rural	392.816	1,0%	2,1%
Sem informação	20.110.358	51,3%	
Dependência administrativa da escola de ensino médio cursada			
Estadual	10.184.874	26,0%	62,5%
Privada	5.095.543	13,0%	31,3%
Federal	501.050	1,3%	3,1%
Municipal	508.930	1,3%	3,1%
Sem informação	22.904.675	58,4%	
Escolaridade da mãe			
Ensino Fundamental	8.422.925	21,5%	34,8%
Ensino Médio	10.021.437	25,6%	41,4%
Ensino Superior	5.752.179	14,7%	23,8%
Sem informação	14.998.531	38,3%	
Escolaridade do pai			
Ensino Fundamental	9.928.101	25,3%	42,2%
Ensino Médio	4.367.339	11,1%	18,6%
Ensino Superior	9.240.031	23,6%	39,3%
Sem informação	15.659.601	40,0%	
Ocupação da mãe			
Grupo 1	1.647.345	4,2%	10,1%
Grupo 2	7.319.178	18,7%	44,9%
Grupo 3	1.737.635	4,4%	10,6%
Grupo 4	4.872.581	12,4%	29,9%
Grupo 5	739.088	1,9%	4,5%
Sem informação	22.879.245	58,4%	
Ocupação do pai			
Grupo 1	2.488.771	6,3%	15,4%
Grupo 2	5.079.575	13,0%	31,5%
Grupo 3	2.867.830	7,3%	17,8%
Grupo 4	4.711.700	12,0%	29,2%
Grupo 5	976.563	2,5%	6,1%
Sem informação	23.070.633	58,9%	
Tipo de ensino médio cursado			
Regular	13.230.249	33,8%	77,6%
EJA	952.253	2,4%	5,6%

Variável	Frequência	Percentual	Válidos
Profissionalizante	2.864.681	7,3%	16,8%
Sem informação	22.147.889	56,5%	
Faixa do INSE			
Nível I	3.253	0,0%	0,0%
Nível II	425.368	1,1%	2,9%
Nível III	1.400.924	3,6%	9,4%
Nível IV	3.781.495	9,6%	25,5%
Nível V	5.564.504	14,2%	37,5%
Nível VI	2.626.635	6,7%	17,7%
Nível VII	1.045.158	2,7%	7,0%
Sem informação	24.347.735	62,1%	
Faixa etária no ingresso			
Menos de 21 anos	12.110.939	30,9%	30,9%
21 a 30 anos	16.387.270	41,8%	41,8%
31 a 40 anos	7.265.767	18,5%	18,5%
Mais de 40 anos	3.431.096	8,8%	8,8%
Faixa etária na saída			
Menos de 21 anos	3.522.339	9,0%	9,0%
21 a 30 anos	22.189.707	56,6%	56,6%
31 a 40 anos	8.933.755	22,8%	22,8%
Mais de 40 anos	4.549.271	11,6%	11,6%

Fonte: Elaboração própria a partir da base do CES complementada com dados do CEB e Enem (Inep).

A Tabela 3 apresenta informações sobre as características dos cursos a que os estudantes estão vinculados às instituições de ensino superior. Em relação ao grau acadêmico, a maioria dos estudantes está matriculada em cursos de bacharelado (59,4%), seguidos por tecnólogos (20,9%) e licenciaturas (19,7%). Sobre a modalidade de ensino, 68,2% dos estudantes cursam presencialmente, enquanto 31,8% estão em cursos EaD. Para os cursos presenciais, o turno noturno é o mais frequente (66,7%), seguido pelo integral (33,3%), dado que enfatiza o peso e importância social que os cursos noturnos têm para a formação universitária (Filho & Nery, 2009).

Quanto à situação de vínculo com o curso, 62,2% dos estudantes estão desvinculados (evasão do curso), enquanto 37,8% concluíram o curso. A forma de ingresso predominante é o vestibular (73,4%), com menor participação do Enem (15,9%) e da avaliação seriada (0,3%). De toda a base utilizada, apenas 2,5% dos estudantes ingressaram por cotas, sendo as cotas para ensino público (2,1%) e étnicas (1,1%) as mais comuns, e demonstrando que, apesar da ampliação do acesso ao longo dos anos, via políticas afirmativas, a participação efetiva ainda é muito baixa.

No que diz respeito às atividades acadêmicas e estrutura do curso, apenas 13,1% dos estudantes participaram de atividades extracurriculares, dos quais 27,8% receberam bolsa (3,6% do total).

A maior parte dos cursos está localizada fora das capitais das unidades federativas (67,2%), e apenas 15,1% dos estudantes frequentam cursos gratuitos, sendo que 99,8% são de IES públicas. Por fim, 89% dos cursos possuem laboratórios disponíveis para os alunos. Esses dados revelam um panorama diversificado do ensino superior no Brasil, mas também destacam tendências que devem ser consideradas ao se discutir a evasão.

Tabela 3 – Frequência e percentual das características dos cursos dos discentes. Brasil. Coortes 2005 a 2022.

Variável	Frequência	Percentual
Grau acadêmico		
Bacharelado	23.264.906	59,4%
Licenciatura	7.732.839	19,7%
Tecnólogo	8.196.003	20,9%
Sem informação	1.324	0,0%
Modalidade		
Presencial	26.725.574	68,2%
EaD	12.469.498	31,8%
Turno (apenas para modalidade presencial)		
Noturno	17.814.292	66,7%
Integral	8.911.282	33,3%
Situação de vínculo com o curso		
Desvinculado	24.375.139	62,2%
Formado	14.819.933	37,8%
Forma de ingresso		
Vestibular	28.759.600	73,4%
Enem	6.242.691	15,9%
Avaliação Seriada	117.880	0,3%
Ingresso por cota		
Sim	994.368	2,5%
Étnico	415.749	1,1%
Ensino público	817.110	2,1%
Renda familiar	295.321	0,8%
Atividade extracurricular		
Sim	5.118.864	13,1%
Recebeu bolsa	1.424.566	3,6%
Curso tem sede na capital da UF		
Sim	12.867.494	32,8%
Não	26.327.578	67,2%
Gratuidade		
Sim	5.906.144	15,1%
Não	33.288.928	84,9%
Laboratório		
Sim	34.880.112	89,0%
Não	4.314.960	11,0%

Fonte: Elaboração própria a partir da base do CES complementada com dados do CEB e Enem (Inep).

A Tabela 4 fornece uma visão detalhada de diferentes características das instituições de ensino superior. Em relação ao ano de referência do CES, a maior parte dos registros concentra-se em

2022 (10,9%), com um aumento gradual ao longo dos anos, demonstrando o aumento histórico do número de egressos. O ano de ingresso no CES mostra um aumento constante até 2014, com uma queda acentuada após 2020, refletindo a própria definição da variável, pois são os ingressantes que já evadiram ou formaram.

Sobre a organização acadêmica, mais da metade dos registros (53%) são de universidades, seguidas por faculdades (22,9%) e centros universitários (22,6%). Na categoria administrativa, as instituições privadas com fins lucrativos predominam (53%), juntamente com as privadas sem fins lucrativos (30,7%), enquanto as públicas federais representam 9,7% do total. Mesmo com a maioria de IES particulares, apenas 26,6% dos estudantes tiveram acesso a algum programa de financiamento estudantil, como Fies (6,1%) ou Prouni (4,9%). E do total de registros, menos de 10% recebeu algum tipo de apoio social, sendo o fornecimento de material didático o mais comum (6,4%).

A mobilidade acadêmica é praticamente inexistente, e pouco mais da metade dos registros são de IES que têm sede na capital da unidade federativa (53,7%). Apenas 7,7% dos alunos são de instituições que participaram do programa Reuni. Quanto à área de formação, negócios, administração e direito lideram com 36,5%, seguidos por educação (19,8%) e saúde e bem-estar (15%). Outras áreas apresentam menor representatividade, como ciências naturais e estatística (1,2%) e agricultura e veterinária (2%). Esses dados destacam a predominância de instituições privadas no sistema educacional brasileiro e a concentração de estudantes em áreas específicas de formação, além de evidenciar o espaço para ampliação disponível no acesso a apoio social e financiamento estudantil.

Tabela 4 – Frequência e percentual das características das instituições dos discentes. Brasil. Coortes 2005 a 2022.

Variável	Frequência	Percentual
Ano de referência do CES		
2010	1.956.814	5,0%
2011	2.252.872	5,7%
2012	2.433.375	6,2%
2013	2.421.851	6,2%
2014	2.633.030	6,7%
2015	2.867.860	7,3%
2016	3.182.688	8,1%
2017	3.088.649	7,9%
2018	3.441.038	8,8%
2019	3.544.266	9,0%
2020	3.419.618	8,7%
2021	3.667.543	9,4%
2022	4.285.468	10,9%
Ano de ingresso no CES		

Variável	Frequência	Percentual
2005	336.353	0,9%
2006	707.776	1,8%
2007	1.192.192	3,0%
2008	1.627.186	4,2%
2009	2.180.710	5,6%
2010	2.541.793	6,5%
2011	2.555.041	6,5%
2012	2.869.749	7,3%
2013	2.900.249	7,4%
2014	3.112.664	7,9%
2015	2.916.840	7,4%
2016	2.880.162	7,3%
2017	2.973.779	7,6%
2018	2.869.692	7,3%
2019	2.589.197	6,6%
2020	2.269.520	5,8%
2021	1.795.602	4,6%
2022	876.567	2,2%
Organização acadêmica		
Faculdade	8.974.050	22,9%
Centro Universitário	8.855.757	22,6%
Universidade	20.754.577	53,0%
Instituto Federal	579.342	1,5%
Centro Federal	31.346	0,1%
Categoria administrativa		
Privada com fins lucrativos	20.753.988	53,0%
Privada sem fins lucrativos	12.027.763	30,7%
Pública Federal	3.786.389	9,7%
Pública Estadual	2.130.951	5,4%
Pública Municipal	279.148	0,7%
Especial	216.833	0,6%
Apoio social		
Sim	3.769.134	9,6%
Alimentação	774.262	2,0%
Bolsa permanência	310.707	0,8%
Bolsa trabalho	423.947	1,1%
Material didático	2.504.192	6,4%
Moradia	138.642	0,4%
Transporte	466.415	1,2%
Financiamento estudantil		
Sim	10.414.145	26,6%
Fies	2.398.996	6,1%
Prouni	1.933.538	4,9%
Mobilidade acadêmica		
Sim	14.105	0,0%
Não	39.180.967	100,0%
Sede na capital da UF		
Sim	21.046.450	53,7%
Não	18.148.622	46,3%
Participou do Reuni		
Sim	3.004.718	7,7%

Variável	Frequência	Percentual
Não	36.190.354	92,3%
Área CINE		
Saúde e bem-estar	5.888.581	15,0%
Educação	7.762.010	19,8%
Negócios, administração e direito	14.289.752	36,5%
Computação e Tecnologias da Informação e Comunicação (TIC)	2.003.666	5,1%
Ciências sociais, comunicação e informação	1.679.003	4,3%
Serviços	1.027.527	2,6%
Engenharia, produção e construção	4.346.273	11,1%
Agricultura, silvicultura, pesca e veterinária	785.746	2,0%
Ciências naturais, matemática e estatística	474.333	1,2%
Artes e humanidades	938.181	2,4%

Fonte: Elaboração própria a partir da base do CES complementada com dados do CEB e Enem (Inep).

As estatísticas descritivas da base fornecem um panorama das variáveis numéricas utilizadas na modelagem, conforme a Tabela 5. Ao longo dos anos de referência 2010 a 2022 do CES, a média de servidores técnicos foi maior que a de docentes (1.208 e 546, respectivamente, arredondando para cima). Essa disparidade reflete a estrutura administrativa necessária para apoiar o corpo docente e as atividades acadêmicas, indicando que a maior parte das IES mantém uma proporção significativa de TAEs para garantir o funcionamento das atividades institucionais e administrativas. No entanto, chama a atenção os valores máximos dessas variáveis, indicando possíveis distorções históricas nos valores informados pelas IES.

As variáveis de carga horária, inscritos, ingressos e vagas aparentam, também, estar eivadas de distorções históricas registradas no CES: a carga horária de um dos cursos está informada ter mais de 606 mil horas; o curso com mais inscritos teve quase 400 mil interessados; o curso com mais ingressos teve quase 100 mil ingressantes; e o curso de maior número de vagas ofereceu quase 150 mil vagas. Tais valores afetaram, também, os máximos das estatísticas de concorrência e de ocupação.

O IGC contínuo médio no período foi igual a 2,73, o que é equivalente ao IGC faixa 3, conceito que, no contexto das IES, é o mínimo necessário para a manutenção do credenciamento institucional e para oferecer cursos de graduação e pós-graduação *stricto sensu* – em média, as IES atendem aos padrões de qualidade estabelecidos pelo Ministério da Educação. De forma semelhante, o CPC contínuo médio do período foi igual a 2,70, indicando que, em termos de qualidade dos cursos de graduação, as instituições também apresentam um desempenho

satisfatório, situando-se em um nível adequado, apesar do conceito máximo ser 5, tanto para o IGC quanto para o CPC.

Concluindo a leitura dos destaques da tabela, o número médio de moradores nas famílias dos egressos da base é de quatro pessoas, com renda familiar média de quase R\$ 4.500 (aproximadamente R\$ 1.800 *per capita*) e INSE médio 5,24, indicando nível socioeconômico médio a médio-alto (INEP, 2023a). A idade média ao ingressar no ensino superior foi de 26,73 anos, com mediana de 24 anos, indicando que muitos estudantes iniciam seus estudos universitários em uma fase mais madura da vida, possivelmente após experiências no mercado de trabalho ou em busca de desenvolvimento profissional (Gomes & Hirata, 2024; Paula, 2021). A idade média ao sair foi de 29,21 anos, com mediana de 26,58 anos, o que é consistente com a duração típica dos cursos de graduação. Esses valores sugerem que, embora haja uma variação na faixa etária dos estudantes, a maioria completa sua formação até os 30 anos, apesar de possíveis atrasos ou interrupções no percurso acadêmico.

Tabela 5 - Estatísticas descritivas das variáveis numéricas. Brasil. CES ano-base 2010 a 2022.

Variável	Média	Mediana	Mínimo	Máximo	Desvio-padrão	Sem informação
Nível Instituição						
Técnicos	1.207,49	545,86	0,00	17.661,00	1.927,85	-
Docentes	974,44	482,16	0,00	8.906,00	1.461,93	-
Servidores	2.181,93	1.070,47	0,00	24.021,00	3.283,85	-
Docentes com pós-graduação (%)	75,98	78,78	0,00	100,00	15,73	-
Conceito médio de graduação	2,66	2,67	0,22	4,86	0,39	14,6%
Conceito médio de mestrado	2,57	3,88	0,00	5,00	2,10	14,6%
Conceito médio de doutorado	2,09	0,43	0,00	5,00	2,21	14,6%
IGC contínuo	2,73	2,69	0,22	4,94	0,48	14,6%
Nível Curso						
CPC contínuo	2,70	2,67	0,09	4,97	0,50	27,6%
Carga horária	3.232,26	3.213,92	0,00	606.667,00	1.742,95	-
Inscritos	7.634,85	551,10	0,00	398.817,00	29.693,27	-
Ingressos	2.487,35	150,70	0,00	98.359,00	7.478,00	-
Vagas	4.860,75	289,13	0,00	149.153,00	12.808,77	-
Concorrência (inscritos por vaga)	3,29	1,36	0,00	6.629,06	8,55	-
Ocupação (%) (ingressos por vaga)	0,64	0,62	0,00	179,00	0,60	-
Nível Aluno						
Número de moradores	3,81	4,00	1,00	20,00	1,53	44,5%
Nota Enem	523,73	516,96	0,00	883,58	73,77	50,7%
Renda familiar	4.494,93	2.607,55	606,00	60.600,00	5.287,80	44,6%
INSE	5,24	5,31	0,96	7,72	0,47	0,0%
Idade no ingresso	26,73	24,00	6,00	110,00	8,76	-
Idade na saída	29,21	26,58	11,00	113,00	8,67	-

Fonte: Elaboração própria a partir da base do CES complementada com dados do CEB e Enem (Inep).

Feita a categorização e análise exploratória dos dados da pesquisa, a próxima seção apresentará as informações sobre evasão. Inicialmente, os resultados serão apresentados com foco nas estatísticas calculadas da macroevasão, e, em seguida, serão descritas as estimativas obtidas com os modelos de regressão logística multinível, levando em consideração as amostras descritas na seção 3.1.4.

2.2. Determinantes da evasão do sistema de ensino superior

A exploração dos determinantes da macroevasão será feita em três etapas. Na primeira (seção 2.2.1.), as estatísticas calculadas de evasão do ensino superior serão discutidas com o uso das médias e erros-padrão; na segunda etapa (seção 2.2.2.), as estimativas dos diversos modelos de regressão multinível aplicados às amostras 120n e 240n serão apresentadas com o uso das razões de chance, intervalos de confiança e significância estatística, usando o p-valor, além de estatísticas gerais de ajuste dos modelos; na última etapa (seção 2.2.3.) todos os resultados encontrados serão comparados com a literatura existente, buscando uma possível convergência dos achados, quando possível.

2.2.1. Evasão do sistema de ensino superior em números

Os números da macroevasão do ensino superior apresentam variações significativas em diferentes categorias, consideradas variáveis de nível individual, por curso e institucionais. Ponto em comum entre todas as estatísticas é o baixo erro-padrão, devido ao elevado tamanho da base utilizada. Isso dá precisão aos valores, porém a acurácia dos resultados depende também da qualidade e representatividade dos dados coletados. Mesmo com baixo erro-padrão, se houver vieses ou fatores não considerados na análise, a exatidão das conclusões pode ser afetada.

A análise da Tabela 6 revela tendências interessantes em relação ao ano-base do CES e ao ano de ingresso. A evasão do sistema, considerando ambas as variáveis, mostra uma tendência constante de crescimento até 2021. Por um lado, esse aumento pode ser explicado pela menor quantidade de tempo disponível para os estudantes concluírem o curso nos anos mais recentes, inclusive devido à pandemia (Colpo, Primo & Aguiar, 2024; Wegner, 2022). Por outro lado, pode indicar crescente desinteresse sistêmico pelo diploma de graduação, em contraste com a expansão que o ensino superior presenciou após os anos 2000 (Senkevics, 2021b), hipótese que demanda estudo específico para ser confirmada.

O tipo de organização acadêmica influencia as taxas de evasão estudantil. Centros Universitários têm a maior taxa média de evasão, enquanto Centros Federais têm a menor.

Diferenças estruturais, recursos disponíveis e políticas de apoio ao estudante variam entre essas organizações, impactando a permanência dos estudantes. A categoria administrativa também importa: instituições privadas com fins lucrativos apresentam taxa de evasão de 51,56%, em contraste com 31,6% nas públicas estaduais. Instituições públicas tendem a oferecer maior apoio socioeconômico e acadêmico (Carvalhaes, Senkevics & Ribeiro, 2023), reduzindo a evasão. Já as privadas com fins lucrativos podem ter custos mais elevados, por terem que constantemente investir na captação de “clientes” (Lopes, 2023), o que deixa pouca margem para iniciativas de suporte discente, aumentando a probabilidade de evasão.

O acesso a iniciativas de apoio social impacta significativamente a evasão. Estudantes com apoio social têm taxa de evasão do sistema menor que daqueles sem apoio. O apoio social fornece recursos emocionais e materiais que fortalecem os estudantes para superar os desafios individuais e acadêmicos (Paula, 2021). O mesmo pode ser dito para os egressos da rede privada: o uso das políticas de financiamento estudantil, como Prouni (Lepine, 2019) ou Fies (Oshiro, 2018), traduz-se em melhor desempenho discente e menores taxas de evasão.

Para finalizar a discussão da tabela, tem-se as taxas de evasão do sistema por área do conhecimento, de acordo com a CINE. As áreas de Serviços (51,13%) e Computação e TIC (50,86%) têm as maiores taxas de evasão, enquanto Ciências Naturais, Matemática e Estatística têm a menor (37,83%). Fatores como a dificuldade percebida dos cursos, expectativas do mercado de trabalho, satisfação com a escolha do curso e diversos outros (Assís et al, 2022) afetam a permanência dos estudantes, refletindo nos indicadores de evasão.

Tabela 6 - Média e erro-padrão da macroevasão por variáveis de nível institucional. Brasil. Coortes 2005 a 2022.

Variável	Evasão do sistema	
	Média	Erro-padrão
Ano de referência do CES		
2010	30,25	0,0328
2011	33,39	0,0314
2012	34,84	0,0305
2013	37,03	0,0310
2014	39,86	0,0302
2015	40,43	0,0290
2016	44,45	0,0279
2017	44,94	0,0283
2018	48,92	0,0269
2019	52,31	0,0265
2020	53,11	0,0270
2021	55,64	0,0259
2022	50,33	0,0242
Ano de ingresso no CES		
2005	23,46	0,0731

Variável	Evasão do sistema	
	Média	Erro-padrão
2006	19,19	0,0468
2007	19,41	0,0362
2008	24,27	0,0336
2009	31,32	0,0314
2010	35,17	0,0300
2011	36,40	0,0301
2012	38,22	0,0287
2013	38,83	0,0286
2014	42,09	0,0280
2015	43,82	0,0291
2016	47,14	0,0294
2017	50,25	0,0290
2018	56,85	0,0292
2019	65,18	0,0296
2020	72,30	0,0297
2021	79,62	0,0301
2022	28,54	0,0482
Organização acadêmica		
Faculdade	42,69	0,0165
Centro Universitário	48,33	0,0168
Universidade	44,61	0,0109
Instituto Federal	46,62	0,0655
Centro Federal	39,09	0,2756
Categoria administrativa		
Privada com fins lucrativos	51,56	0,0110
Privada sem fins lucrativos	40,07	0,0141
Pública Federal	34,08	0,0244
Pública Estadual	31,60	0,0318
Pública Municipal	33,49	0,0893
Especial	33,68	0,1015
Apoio social		
Sim	41,17	0,0253
Não	45,45	0,0084
Financiamento estudantil		
Sim	40,16	0,0152
Não	46,80	0,0093
Área CINE		
Saúde e bem-estar	43,69	0,0204
Educação	42,05	0,0177
Negócios, administração e direito	46,69	0,0132
Computação e Tecnologias da Informação e Comunicação (TIC)	50,86	0,0353
Ciências sociais, comunicação e informação	40,68	0,0379
Serviços	51,13	0,0493
Engenharia, produção e construção	46,28	0,0239
Agricultura, silvicultura, pesca e veterinária	41,52	0,0556

Variável	Evasão do sistema	
	Média	Erro-padrão
Ciências naturais, matemática e estatística	37,83	0,0704
Artes e humanidades	42,53	0,0510

Fonte: Elaboração própria a partir da base do CES complementada com dados do CEB e Enem (Inep).

Avançando para a discussão da evasão do sistema por variáveis referentes ao curso, disponíveis na Tabela 7, há diferenças já bem evidenciadas na literatura. A macroevasão do ensino a distância (54,91%) é bem superior à do ensino presencial (40,43%), reflexo da falta de orientações aos ingressantes do EaD e de formação didática adequada dos tutores e professores para atuação a distância (Kowalski et al., 2020), assim como da mercantilização dessa modalidade de ensino (Branco, Conte & Habowski, 2020). No quesito turno do curso, egressos do noturno apresentam, em média, evasão significativamente superior aos do turno integral, o que pode ser explicado, em parte, pelo diferente perfil dos alunos que procuram cursos noturnos ou a distância: pessoas que frequentemente conciliam os estudos com atividades profissionais e responsabilidades pessoais (Filho & Nery, 2009). Essa sobrecarga pode levar ao cansaço e ao estresse, dificultando a dedicação necessária aos estudos e aumentando a probabilidade de evasão. Além disso, a disponibilidade limitada de recursos institucionais e suporte acadêmico fora do horário comercial pode contribuir para desafios adicionais, impactando negativamente a permanência e sucesso desses alunos nos cursos.

O grau acadêmico tem, também, comportamento da evasão que depende da categoria analisada. Os cursos tecnólogos apresentam a maior taxa média de macroevasão, com 50,66%, enquanto os bacharelados têm 44,06% e as licenciaturas 42,00%, a menor taxa. Os cursos tecnólogos, por serem de curta duração e focados em habilidades específicas, de público predominantemente masculino, podem atrair estudantes que buscam inserção rápida no mercado de trabalho, mas que também podem abandonar o curso ao conseguir emprego na área (Vieira, Teló & Vieira, 2021). Apesar da evasão média das licenciaturas ser a menor entre os graus, a percepção na literatura é de que são cursos, em geral, menos atrativos, prestigiosos e de perspectivas profissionais limitadas (Gambirage et al., 2021; Rocha & Carvalhaes, 2023). Tais resultados demonstram a necessidade de aprofundamento no tema, para termos a percepção correta e informada sobre os determinantes da evasão sistêmica.

Com relação ao turno cursado, o menor valor da macroevasão está no integral (36,24%), com valor significativamente maior para o turno noturno (42,52%). Ao consolidar o turno noturno à modalidade EaD, a taxa de evasão aumentou ainda mais (47,62). Considerando que a maioria dos alunos do presencial estão no turno noturno (Tabela 3) e o avanço da popularidade do ensino

a distância (Senkevics, 2021b), tais resultados demonstram o grande desafio nesse tema. Uma estratégia de mitigação é o oferecimento de atividades extracurriculares – alunos que participaram, mesmo sem receber bolsa, têm macroevasão bem menor (28,67% geral e 25,14% com bolsa), especialmente em comparação a quem não participou (47,49%). A presença de laboratórios tem, também, um leve efeito na macroevasão, assim como o fato de o curso ser dado na capital das UFs.

Para concluir a análise da tabela, a forma de ingresso dos estudantes por vestibular ou enem não apresentou valores muito divergentes, tendo o primeiro macroevasão ligeiramente superior (46,97% e 45,78%, respectivamente). O ingresso por avaliação seriada teve a menor evasão, porém sua representatividade é muito limitada (apenas 0,3% do total). E o ingresso por reserva de vaga, no quesito macroevasão, representa valores menores (41,98%), em comparação ao ingresso por ampla concorrência (45,11%), indicando que, apesar das dificuldades adicionais dos alunos que ingressam por cotas (Pinheiro, Pereira & Xavier, 2021), sua evasão sistêmica é menor.

Tabela 7 - Média e erro-padrão da macroevasão por variáveis de nível curso. Brasil. Coortes 2005 a 2022.

Variável	Evasão do sistema	
	Média	Erro-padrão
Grau acadêmico		
Bacharelado	44,06	0,0103
Licenciatura	42,00	0,0177
Tecnólogo	50,66	0,0175
Modalidade		
Presencial	40,43	0,0095
EaD	54,91	0,0141
Turno		
Noturno / EaD	47,62	0,0091
Noturno	42,52	0,0096
Integral	36,24	0,0161
Forma de ingresso		
Vestibular	46,97	0,0093
Enem	45,78	0,0199
Avaliação Seriada	27,54	0,1301
Ingresso por cota		
Sim	41,98	0,0495
Não	45,11	0,0081
Atividade extracurricular		
Sim	28,67	0,0200
Recebeu bolsa	25,14	0,0363
Não	47,49	0,0086
Curso tem sede na capital da UF		
Sim	41,75	0,0137
Não	46,64	0,0097

Variável	Evasão do sistema	
	Média	Erro-padrão
Laboratório		
Sim	44,86	0,0084
Não	46,44	0,0240

Fonte: Elaboração própria a partir da base do CES complementada com dados do CEB e Enem (Inep).

A última tabela a ser discutida nesta seção, Tabela 8, tem os valores da evasão do sistema por características individuais dos egressos. A raça/cor autodeclarada, objeto de estudo de diversos autores em diferentes temas além da evasão, é significativamente maior para os alunos pretos, pardos ou indígenas (PPI), com média 50,1%, em comparação aos 43,4% dos discentes brancos ou amarelos. Essa diferença não pode ser explicada unicamente pela questão racial, pois está atrelada a um conjunto de fatores, como discriminação racial, falta de representação e apoio dentro das instituições de ensino, e a pressão econômica que esses estudantes frequentemente enfrentam, por fazerem parte de grupos raciais historicamente desfavorecidos (Silva, 2024). Portanto, compreender e abordar esses fatores é fundamental para desenvolver políticas e estratégias eficazes que promovam a inclusão e a permanência de estudantes PPI no ensino superior, contribuindo para a redução das desigualdades educacionais e sociais existentes.

Há também uma diferença notável das taxas de evasão do sistema por sexo. As mulheres, além de serem as que mais ingressam no ensino superior, são também as que mais concluem (Brasil, 2022) e menos evadem (42%), em média, em comparação aos homens (49%). Apesar das limitações de analisar uma característica isoladamente, estudos indicam que as mulheres tendem a evadir menos do que os homens, um padrão que se observa em diferentes contextos educacionais (Lopes et al., 2023; Saccaro, França & Jacinto, 2019) e que tem relação, também, com as motivações de aprendizagem e perspectivas futuras das mulheres (Cardoso & Nagai, 2019). Assim, considerar as diferenças de gênero é essencial para desenvolver políticas que promovam a equidade e reduzam a evasão no ensino superior.

Foi encontrada uma diferença significativa nas taxas de evasão do sistema entre alunos com deficiência (PcD) e aqueles sem deficiência. Os dados indicam que alunos PcD apresentam uma taxa média de evasão de 49,33%, enquanto alunos sem deficiência têm uma taxa de 45,11%. Essa discrepância pode ser atribuída a desafios adicionais enfrentados por estudantes com deficiência, como barreiras de acessibilidade física e tecnológica, falta de materiais adaptados, e insuficiente apoio pedagógico e institucional (Pereira et al., 2020). Esses obstáculos podem dificultar a integração e o desempenho acadêmico dos alunos PcD, aumentando a probabilidade de evasão.

A origem escolar do egresso afeta, também, as taxas de evasão do sistema. Aqueles que vieram de escolas públicas têm evasão menor (42%), em comparação aos de escolas de ensino médio privadas (48,55%). Tais valores são corroborados, em contextos mais específicos, pela literatura (Cunha et al, 2015), porém existe a hipótese de que alunos provenientes de escolas de ensino médio públicas sejam mais propensos a um menor rendimento acadêmico e, conseqüentemente, à evasão, devido à maior possibilidade de ter um contexto socioeconômico desfavorável (Bayma-Freire, Roazzi & Roazzi, 2015; Gama, 2018). Não há, no entanto, evidências de que os egressos de escolas de ensino médio públicas desistam mais (Sampaio et al., 2011). Alunos provenientes de escolas secundárias urbanas e federais têm, também, menor macroevasão (47,46% e 38,21%, respectivamente), porém há ressalvas devido à elevada falta de informação nessas variáveis (Tabela 2).

A escolaridade e ocupação dos pais ou responsáveis legais são fatores já conhecidos que afetam a macroevasão. Egressos cujos mãe ou pai tinham ensino superior têm evasão do sistema significativamente menor (40,36% e 38,94%, respectivamente). O mesmo ocorre quando se analisa a profissão das mães e pais – ocupações de nível superior (grupo 5) têm os menores valores de evasão, 34,91% e 39,86%, respectivamente. Esses resultados indicam que esses fatores estão fortemente associados à permanência no ensino superior, possivelmente devido ao maior suporte acadêmico, financeiro ou cultural oferecido aos estudantes (Paula, 2021). Já as altas taxas de evasão observadas entre estudantes cujos pais possuem menor escolaridade ou estão em ocupações de baixa qualificação reforçam a necessidade de políticas públicas que promovam maior inclusão e apoio a esses grupos.

As taxas de evasão no ensino superior variam significativamente de acordo com a faixa etária no ingresso. Os dados indicam que estudantes com menos de 21 anos apresentam taxas de evasão menores em comparação com os grupos etários mais avançados. Isso pode ser atribuído ao fato de que estudantes mais jovens geralmente têm menos responsabilidades familiares e profissionais, permitindo maior dedicação aos estudos (Lopes et al., 2023), assim como podem estar mais integrados ao ambiente acadêmico e social da instituição (Nóbrega et al., 2023). Portanto, é importante que as instituições de ensino superior ofereçam suporte específico para alunos ingressantes dependendo da faixa etária, visando reduzir a evasão de cada grupo.

Finalizando a análise da tabela, a faixa etária na saída dos estudantes também influencia as taxas de evasão do sistema. Nota-se que estudantes com menos de 21 anos têm taxas de evasão bem mais altas quando comparados aos de faixas etárias superiores. Isso pode indicar que alunos mais jovens podem enfrentar dificuldades de adaptação ao ensino superior ou indecisão quanto

à escolha do curso, levando ao abandono precoce (Nóbrega et al., 2023), motivado às vezes pela colocação no mercado de trabalho (Vieira, Teló & Vieira, 2021). Por outro lado, estudantes mais velhos tendem a ser mais resilientes e determinados a concluir seus estudos (Rosa & Oliveira, 2022). Dessa forma, oferecer orientação vocacional, apoio acadêmico e integração institucional aos estudantes mais jovens pode ser fundamental para reduzir a evasão nesse segmento.

Tabela 8 - Média e erro-padrão da macroevasão por variáveis de nível individual. Brasil. Coortes 2005 a 2022.

Variável	Evasão do sistema	
	Média	Erro-padrão
Raça/cor autodeclarada		
PPI	50,10	0,0134
Branca / amarela	43,38	0,0121
Sexo		
Feminino	41,94	0,0105
Masculino	49,05	0,0121
Aluno PcD		
Sim	49,33	0,1062
Não	45,11	0,0080
Escola de ensino médio cursada		
Privada	48,55	0,0101
Pública	41,92	0,0158
Sem informação	34,43	0,0209
Localização da escola de ensino médio cursada		
Urbana	47,76	0,0116
Rural	51,13	0,0798
Dependência administrativa da escola de ensino médio cursada		
Estadual	50,89	0,0157
Privada	40,27	0,0217
Federal	38,21	0,0686
Municipal	60,91	0,0684
Escolaridade da mãe		
Ensino Fundamental	47,90	0,0172
Ensino Médio	50,48	0,0158
Ensino Superior	40,36	0,0205
Escolaridade do pai		
Ensino Fundamental	47,70	0,0159
Ensino Médio	50,14	0,0239
Ensino Superior	38,94	0,0160
Ocupação da mãe		
Grupo 1	50,48	0,0390
Grupo 2	54,50	0,0184
Grupo 3	56,75	0,0376
Grupo 4	49,95	0,0227
Grupo 5	34,91	0,0554
Ocupação do pai		
Grupo 1	49,62	0,0317
Grupo 2	55,62	0,0220

Variável	Evasão do sistema	
	Média	Erro-padrão
Grupo 3	55,36	0,0294
Grupo 4	49,09	0,0230
Grupo 5	39,86	0,0495
Tipo de ensino médio cursado		
Regular	46,00	0,0137
EJA	69,10	0,0474
Profissionalizante	46,58	0,0295
Faixa etária no ingresso		
Menos de 21 anos	38,18	0,0140
21 a 30 anos	47,16	0,0123
31 a 40 anos	48,93	0,0185
Mais de 40 anos	50,82	0,0270
Faixa etária na saída		
Menos de 21 anos	61,78	0,0259
21 a 30 anos	41,12	0,0104
31 a 40 anos	46,67	0,0167
Mais de 40 anos	47,95	0,0234

Fonte: Elaboração própria a partir da base do CES complementada com dados do CEB e Enem (Inep).

Os resultados analisados, da macroevasão do ensino superior, apesar de metodologicamente simples, já indicam diversos desafios a serem superados, por meio das variações significativas em diversas categorias, incluindo características individuais dos estudantes, especificidades dos cursos e aspectos institucionais. Observou-se que fatores como raça/cor, sexo, presença de deficiência, faixa etária, tipo de instituição, modalidade de ensino e área de conhecimento, dentre outras, influenciam diretamente as taxas de evasão. Essas disparidades ressaltam a complexidade do fenômeno e a necessidade de abordagens multifacetadas para enfrentá-lo.

Apesar da precisão estatística proporcionada pelo grande volume de dados e baixo erro-padrão, é fundamental considerar a qualidade e a representatividade das informações coletadas. Vieses ou fatores não contemplados na análise podem afetar a acurácia dos resultados e, conseqüentemente, das conclusões. Portanto, aprofundar a compreensão desses fatores, utilizando metodologias estatísticas robustas, é essencial para embasar políticas e estratégias eficazes que promovam a inclusão, a permanência e o sucesso dos estudantes no ensino superior.

Nas próximas seções, serão apresentados os resultados das modelagens estatísticas realizadas por meio da regressão logística multinível. Tal abordagem permite identificar e quantificar o impacto dos diferentes fatores na evasão, considerando a hierarquia dos dados e as interações entre variáveis individuais, de curso e institucionais. Espera-se que os resultados contribuam para o desenvolvimento de intervenções direcionadas e políticas públicas que reduzam a evasão e promovam a equidade educacional.

2.2.2. Resultados das regressões logísticas multinível

Concluída a discussão das taxas médias de evasão do ensino superior por cada determinante, que apontou diversas conclusões – embora pontuais por estarem isoladas do contexto complexo da evasão –, o próximo passo é entender de que maneira e intensidade cada determinante influencia a macroevasão, utilizando a regressão logística multinível. As próximas duas subseções irão descrever o procedimento de modelagem, as estatísticas de ajuste e a tabela de resultados, com estimativa e intervalo de confiança, de cada amostra utilizada.

2.2.2.1. Grupo de determinantes provenientes do CES

Esta seção utilizará a amostra 120n, de tamanho 328.200, composta apenas por variáveis provenientes do CES, porém considerando as melhorias de qualidade informacional implementadas pelo uso do CEB e Enem.

Conforme descrito na seção 3.1.4, o procedimento de aplicação da regressão logística multinível deve iniciar pelo modelo nulo, de forma a justificar o uso dessa técnica de modelagem. O ICC foi igual a 0,164, indicando que 16,4% de toda a variância da métrica em estudo é devido às diferenças entre os grupos, IES e curso, o que é suficiente para dar prosseguimento às modelagens.

A próxima iteração, denominada modelo inicial ou Modelo I, incluiu a variável ano de ingresso, que foi inserida tanto na parte fixa quanto na parte aleatória do modelo, juntamente com o grupo das IES. Além disso, o agrupamento das IES foi modificado: cada IES foi combinada com sua modalidade de ensino, formando uma combinação IES-modalidade. Ambas as modificações foram feitas para considerar as diferentes estruturas de evasão que existem ao se levar em conta a coorte de ingresso e a modalidade de ensino, conforme verificado na Tabela 6 e Tabela 8, respectivamente.

O modelo individual, ou Modelo II, agregou ao modelo inicial as variáveis de nível individual, mantendo a estrutura anterior dos grupos e efeitos. O próximo modelo, Modelo III, expandiu pela utilização das variáveis de nível de curso. E a última iteração, Modelo IV, considerou todos os níveis dos determinantes, adicionando as variáveis de nível institucional. As estatísticas de ajuste dos modelos estão dispostas ao fim da Tabela 9, e indicam que houve uma melhoria progressiva no ajuste e na capacidade explicativa das variáveis incluídas na modelagem da macroevasão. Observa-se que o R^2 marginal aumenta de 0,077 no modelo I para 0,207 no modelo IV, indicando que a proporção da variância explicada apenas pelos efeitos fixos quase triplica ao longo dos modelos. Da mesma forma, o R^2 condicional cresce de 0,225 para 0,339, mostrando que a variância total explicada pelos efeitos fixos e aleatórios também aumenta.

Além disso, os valores do AIC e BIC diminuem consistentemente de um modelo para o outro, passando de 403.824 e 403.888 no modelo I para 381.139 e 381.620 no modelo IV, respectivamente. Isso sugere que os modelos posteriores têm melhor qualidade de ajuste e que a adição de variáveis explicativas melhorou o modelo de forma significativa, superando a penalização pela complexidade adicional. O ICC mantém-se relativamente estável, variando ligeiramente entre 0,16 e 0,167, indicando que a proporção da variância atribuída às diferenças entre os grupos permanece consistente. Por fim, o RMSE (*Root Mean Square Error*) diminui de 0,46 no modelo I para 0,44 no modelo IV, reforçando a melhoria na precisão dos modelos ao longo das iterações.

Durante o processo de ajuste incremental dos modelos, a análise da estatística VIF (*Variance Inflation Factor*) revelou que algumas variáveis apresentavam multicolinearidade significativa, com valores superiores a 1,5. As variáveis afetadas foram idade no fim do vínculo, ingresso via vestibular, total de inscritos, total de ingressos, Reuni e área CINE. Devido a esse problema, optou-se por não incluir essas variáveis nos modelos finais, pois a multicolinearidade pode distorcer as estimativas dos coeficientes e comprometer a confiabilidade dos resultados. Além disso, algumas variáveis não foram utilizadas devido à falta de informações completas nos dados disponíveis. Essas variáveis incluem o IDHM, IGC, CPC, concorrência e ocupação.

Procedendo à análise dos resultados obtidos pela aplicação incremental da regressão logística multinível aos dados da amostra 120n, cujos valores das estimativas, com intervalo de confiança a 95%, estão dispostos na Tabela 9, a discussão será feita por grupo de determinantes, na seção

2.2.3.

Tabela 9 - Resultados da regressão logística multinível da macroevasão do ensino superior, em razão de chances (*odds ratio*). Brasil (amostra 120n). Coortes 2005 a 2022.

Variável	Modelos 120n			
	I	II	III	IV
	Inicial	Individual	Individual + Curso	Individual + Curso + Instituição
Intercepto	0,826*** (0,776, 0,879)	0,619*** (0,582, 0,658)	0,772*** (0,704, 0,847)	0,450*** (0,357, 0,566)
Ano de ingresso	1,774*** (1,728, 1,821)	1,886*** (1,836, 1,938)	2,469*** (2,397, 2,544)	2,588*** (2,510, 2,668)
Cor/raça: PPI (branca/amarela)		1,137*** (1,116, 1,158)	1,119*** (1,098, 1,140)	1,140*** (1,119, 1,163)
Sexo: Masculino (feminino)		1,406*** (1,382, 1,430)	1,395*** (1,371, 1,419)	1,386*** (1,362, 1,411)
Idade no ingresso		1,052***	1,027***	1,022***

Variável	Modelos 120n			
	I	II	III	IV
	Inicial	Individual	Individual + Curso	Individual + Curso + Instituição
		(1,043, 1,061)	(1,017, 1,036)	(1,013, 1,032)
PcD		1,127* (1,022, 1,243)	1,153** (1,044, 1,273)	1,149** (1,039, 1,270)
Escola EM: Pública (privada)		0,925*** (0,907, 0,944)	0,954*** (0,934, 0,974)	0,911*** (0,892, 0,931)
INSE		0,921*** (0,911, 0,931)	0,896*** (0,886, 0,906)	0,877*** (0,868, 0,888)
Grau: Licenciatura (bacharelado)			0,761*** (0,668, 0,867)	0,750*** (0,658, 0,855)
Grau: Tecnólogo (bacharelado)			0,774*** (0,718, 0,833)	0,744*** (0,690, 0,802)
Turno: Noturno (integral)			1,092*** (1,064, 1,120)	1,095*** (1,066, 1,124)
Modalidade: EaD (presencial)			1,238*** (1,103, 1,389)	1,305*** (1,152, 1,477)
Ingresso Enem			0,932*** (0,909, 0,956)	1,138*** (1,108, 1,168)
Ingresso Avaliação Seriada			1,169* (1,029, 1,328)	1,268*** (1,116, 1,442)
Ingresso Cotas			1,273*** (1,216, 1,334)	1,271*** (1,213, 1,332)
Atividade extracurricular			0,666*** (0,657, 0,675)	0,701*** (0,692, 0,711)
Atividade extracurricular com bolsa			0,935*** (0,921, 0,949)	0,930*** (0,916, 0,944)
Curso fica na Capital			0,963 (0,918, 1,009)	0,919** (0,869, 0,972)
Carga horária do curso			1,022 (0,994, 1,050)	1,025+ (0,997, 1,054)
Curso possui laboratório			0,987 (0,953, 1,021)	1,003 (0,969, 1,039)
Total de vagas do curso			0,992 (0,896, 1,099)	1,004 (0,905, 1,115)
Egresso durante a pandemia do Covid 19			0,430***	0,384***

Variável	Modelos 120n			
	I	II	III	IV
	Inicial	Individual	Individual + Curso	Individual + Curso + Instituição
			(0,419, 0,441)	(0,374, 0,395)
Organização acadêmica: Centro Universitário (universidade)				0,861** (0,782, 0,949)
Organização acadêmica: Faculdade (universidade)				1,017 (0,922, 1,121)
Organização acadêmica: Instituto Federal (universidade)				1,083 (0,855, 1,372)
Organização acadêmica: Centro Federal (universidade)				0,485 (0,171, 1,374)
Apoio Social				0,882*** (0,871, 0,894)
Financiamento estudantil				0,736*** (0,725, 0,747)
Financiamento estudantil Fies				0,920*** (0,908, 0,932)
Financiamento estudantil Prouni				0,851*** (0,841, 0,862)
Categoria administrativa: Pública Estadual (pública federal)				1,087 (0,892, 1,325)
Categoria administrativa: Pública Municipal (pública federal)				1,907*** (1,487, 2,447)
Categoria administrativa: Privada com fins lucrativos (pública federal)				1,810*** (1,438, 2,279)
Categoria administrativa: Privada sem fins lucrativos (pública federal)				1,806*** (1,439, 2,268)
Categoria administrativa: Especial (pública federal)				1,787*** (1,382, 2,309)
IES fica na Capital				1,184*** (1,090, 1,286)
Total de servidores (TAEs e docentes)				1,029 (0,967, 1,094)
Percentual de docentes com pós-graduação				0,909***

Variável	Modelos 120n			
	I	II	III	IV
	Inicial	Individual	Individual + Curso	Individual + Curso + Instituição
				(0,894, 0,925)
Num.Obs.	328.200	328.200	328.200	328.200
R ² Marg.	0,077	0,093	0,156	0,207
R ² Cond.	0,225	0,23	0,292	0,339
AIC	403.824	400.595	390.074	381.139
BIC	403.888	400.745	390.374	381.620
ICC	0,16	0,151	0,161	0,167
RMSE	0,46	0,45	0,45	0,44

Fonte: Elaboração própria a partir da base do CES complementada com dados do CEB e Enem (Inep).

Nota: + $p < 0,1$, * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$; EM = Ensino Médio; categoria de referência e intervalo de confiança a 95% entre parênteses.

2.2.2.2. Grupo de determinantes provenientes do CES, CEB e Enem

Esta seção abordará a amostra 240n, de tamanho 397.231, composta por variáveis provenientes do CES, CEB e Enem, consideradas as melhorias de qualidade informacional implementadas durante o processo de tratamento dos dados.

Os indicadores de ajuste dos modelos, apresentados ao final da Tabela 10, evidenciam uma melhoria progressiva tanto no ajuste quanto na capacidade explicativa das variáveis incluídas na modelagem da macroevasão. Observa-se que o R² marginal aumenta de 0,121 no Modelo I para 0,348 no Modelo IV, indicando que a proporção da variância explicada apenas pelos efeitos fixos mais que dobrou ao avançar dos modelos. De forma similar, o R² condicional cresce de 0,261 para 0,462, mostrando que a variância total explicada pelos efeitos fixos e aleatórios também aumentou significativamente.

Além disso, os valores do AIC e BIC diminuem consistentemente de um modelo para o outro, passando de 483.824 e 483.889 no Modelo I para 431.179 e 431.745 no Modelo IV, respectivamente. Essa redução sugere que os modelos subsequentes apresentam melhor qualidade de ajuste, e que a inclusão de novas variáveis explicativas aprimorou o modelo de forma significativa, superando a penalização pela complexidade adicional inerente à adição de mais parâmetros. O ICC mantém-se relativamente estável, variando ligeiramente entre 0,159 e 0,174, o que indica que a proporção da variância atribuída às diferenças entre os grupos permanece consistente ao longo dos modelos. Por fim, o RMSE diminui de 0,46 no Modelo I para 0,42 no Modelo IV, reforçando a melhoria na precisão e na capacidade explicativa dos modelos ao longo das iterações.

Durante o processo de ajuste incremental dos modelos, a análise da estatística VIF revelou que algumas variáveis apresentavam multicolinearidade significativa, com valores superiores a 1,5. As variáveis afetadas foram idade no fim do vínculo, total de inscritos, total de ingressos, tipo de ensino médio regular, financiamento estudantil, curso na capital, IDHM, Reuni e área CINE. Devido a esse problema, optou-se por não incluir essas variáveis nos modelos finais, pois a multicolinearidade pode distorcer as estimativas dos coeficientes e comprometer a confiabilidade dos resultados. E a variável de categoria administrativa da IES teve de ser transformada em variável dicotômica, se IES privada ou não, para poder ser mantida nos modelos.

Tabela 10 - Resultados da regressão logística multinível da macroevasão do ensino superior, em razão de chances (*odds ratio*). Brasil (amostra 240n). Coortes 2005 a 2022.

Variável	Modelos 240n			
	I	II	III	IV
	Inicial	Individual	Individual + Curso	Individual + Curso + Instituição
Intercepto	0,740*** (0,673, 0,813)	1,192** (1,047, 1,357)	1,223** (1,064, 1,404)	1,678*** (1,404, 2,005)
Ano de ingresso	2,922*** (2,808, 3,041)	3,145*** (3,019, 3,276)	4,931*** (4,713, 5,158)	5,056*** (4,835, 5,287)
Cor/raça: PPI (branca/amarela)		1,073*** (1,057, 1,090)	1,052*** (1,035, 1,069)	1,068*** (1,050, 1,085)
Sexo: Masculino (feminino)		1,569*** (1,545, 1,594)	1,543*** (1,519, 1,568)	1,539*** (1,514, 1,565)
Idade no ingresso		0,875*** (0,859, 0,891)	0,825*** (0,810, 0,841)	0,841*** (0,825, 0,858)
PcD		1,038 (0,951, 1,133)	1,076 (0,984, 1,178)	1,087+ (0,992, 1,191)
Escola EM: Pública (privada)		1,001 (0,980, 1,021)	0,995 (0,974, 1,016)	0,957*** (0,936, 0,978)
INSE		1,058*** (1,050, 1,067)	1,045*** (1,036, 1,054)	1,024*** (1,016, 1,033)
Localização escola EM: Rural (urbana)		0,937** (0,893, 0,983)	0,922** (0,877, 0,969)	0,924** (0,878, 0,972)
Mãe e/ou pai com ensino superior		1,020* (1,004, 1,038)	1,017+ (0,999, 1,034)	0,999 (0,982, 1,017)
Mãe e/ou pai em ocupações que exigem ensino superior		0,867*** (0,844, 0,890)	0,883*** (0,859, 0,908)	0,855*** (0,831, 0,879)

Variável	Modelos 240n			
	I	II	III	IV
	Inicial	Individual	Individual + Curso	Individual + Curso + Instituição
Nota Enem		0,629*** (0,623, 0,635)	0,628*** (0,622, 0,635)	0,673*** (0,666, 0,680)
Tamanho da família		1,060*** (1,052, 1,068)	1,059*** (1,051, 1,067)	1,059*** (1,051, 1,067)
Renda familiar		0,891*** (0,882, 0,900)	0,900*** (0,891, 0,910)	0,866*** (0,857, 0,876)
Fez EM EJA		1,892*** (1,816, 1,971)	1,875*** (1,797, 1,956)	1,840*** (1,762, 1,921)
Fez EM Profissionalizante		0,994 (0,972, 1,016)	0,988 (0,966, 1,011)	1,023+ (1,000, 1,048)
Dependência da escola EM: Estadual (federal)		1,064** (1,020, 1,111)	1,039+ (0,994, 1,086)	1,075** (1,028, 1,125)
Dependência da escola EM: Municipal (federal)		1,273*** (1,200, 1,350)	1,232*** (1,159, 1,309)	1,279*** (1,202, 1,360)
Dependência da escola EM: Privada (federal)		1,096*** (1,051, 1,143)	1,091*** (1,044, 1,139)	1,067** (1,021, 1,115)
Grau: Licenciatura (bacharelado)			0,747** (0,623, 0,895)	0,714*** (0,595, 0,857)
Grau: Tecnólogo (bacharelado)			0,579*** (0,525, 0,638)	0,553*** (0,501, 0,611)
Turno: Noturno (integral)			1,047*** (1,026, 1,069)	1,070*** (1,048, 1,093)
Modalidade: EaD (presencial)			1,186*** (1,134, 1,240)	1,145*** (1,094, 1,199)
Ingresso Enem			1,074*** (1,052, 1,097)	1,249*** (1,222, 1,276)
Ingresso Avaliação Seriada			0,867** (0,781, 0,962)	0,93 (0,838, 1,032)
Ingresso Cotas			0,992 (0,953, 1,033)	0,969 (0,930, 1,009)
Atividade extracurricular			0,642*** (0,634, 0,649)	0,680*** (0,672, 0,688)
Atividade extracurricular com bolsa			0,952*** (0,941, 0,963)	0,944*** (0,933, 0,955)
Carga horária do curso			1,060***	1,047**

Variável	Modelos 240n			
	I	II	III	IV
	Inicial	Individual	Individual + Curso	Individual + Curso + Instituição
			(1,030, 1,090)	(1,017, 1,078)
Curso possui laboratório			0,914*** (0,881, 0,948)	0,901*** (0,869, 0,935)
Total de vagas do curso			1,152*** (1,120, 1,184)	1,158*** (1,126, 1,191)
Concorrência do curso			1,007 (0,994, 1,020)	1,004 (0,990, 1,017)
Ocupação do curso			1,605*** (1,550, 1,662)	1,381*** (1,332, 1,432)
CPC contínuo			1,017** (1,006, 1,028)	1,037*** (1,025, 1,049)
Egresso durante a pandemia do Covid 19			0,299*** (0,292, 0,306)	0,296*** (0,289, 0,303)
Organização acadêmica: Centro Universitário (universidade)				1,05 (0,957, 1,152)
Organização acadêmica: Faculdade (universidade)				1,335*** (1,207, 1,476)
Organização acadêmica: Instituto Federal (universidade)				0,907 (0,686, 1,200)
Organização acadêmica: Centro Federal (universidade)				1,219 (0,331, 4,486)
Apoio Social				0,859*** (0,850, 0,869)
Financiamento estudantil Fies				0,768*** (0,762, 0,773)
Financiamento estudantil Prouni				0,787*** (0,781, 0,793)
IES Privada				0,863* (0,760, 0,980)
IES fica na Capital				1,250*** (1,146, 1,363)
Total de servidores (TAEs e docentes)				1,183*** (1,133, 1,235)
Percentual de docentes com pós-graduação				0,775***

Variável	Modelos 240n			
	I	II	III	IV
	Inicial	Individual	Individual + Curso	Individual + Curso + Instituição
				(0,760, 0,790)
IGC contínuo				0,919*** (0,896, 0,941)
Num.Obs.	397.231	397.231	397.231	397.231
R2 Marg.	0,121	0,199	0,292	0,348
R2 Cond.	0,261	0,316	0,396	0,462
AIC	483.824	467.749	444.462	431.179
BIC	483.889	468.010	444.897	431.745
ICC	0,159	0,146	0,147	0,174
RMSE	0,46	0,45	0,43	0,42

Fonte: Elaboração própria a partir da base do CES complementada com dados do CEB e Enem (Inep).

Nota: + $p < 0,1$, * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$; EM = Ensino Médio; categoria de referência e intervalo de confiança a 95% entre parênteses.

2.2.3. Discussão conjunta dos determinantes

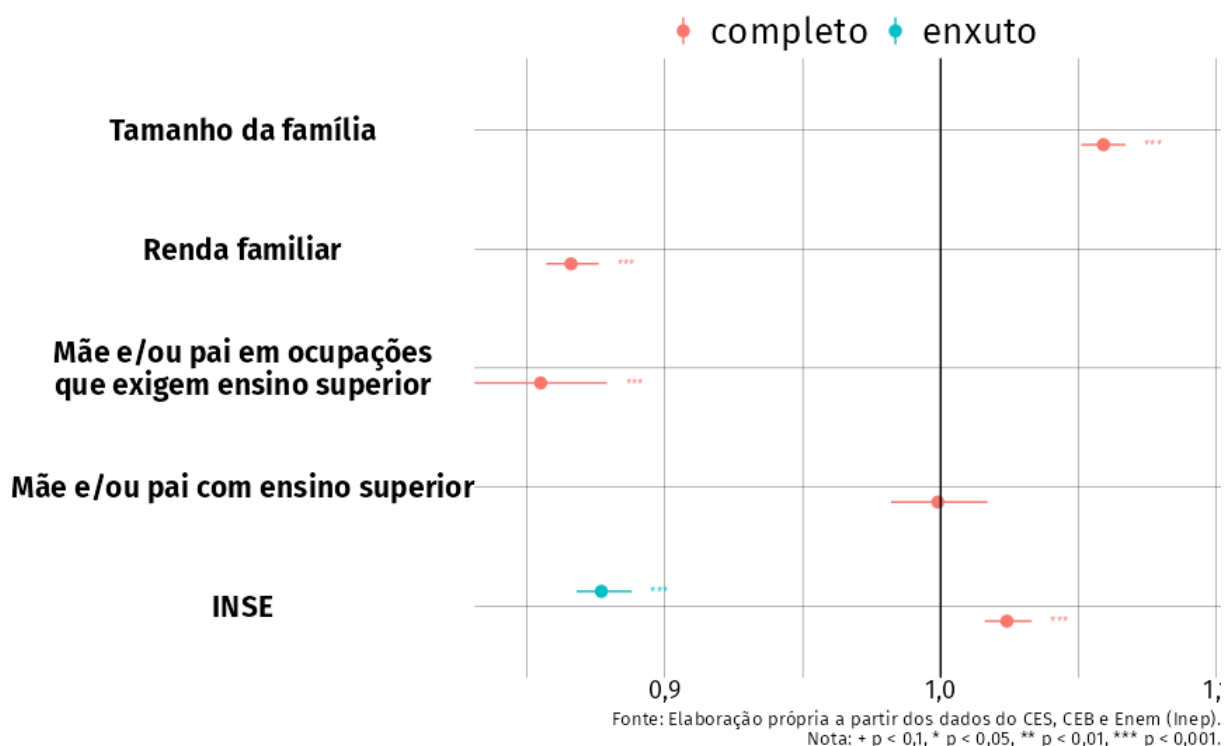
Esta seção irá fazer a análise dos resultados obtidos pela aplicação incremental da regressão logística multinível aos dados das amostras 120n e 240n, com base nos valores das estimativas, com intervalo de confiança a 95%. A abordagem de cada determinante será feita utilizando os resultados de ambas as amostras (disponíveis na Tabela 9 e Tabela 10), destacando quando houver discordância entre os modelos finais de cada grupo, e situando a análise no contexto dos estudos de evasão no ensino superior.

O ano de ingresso não está sendo tratado como um determinante da evasão, mas sim como uma variável de controle, no sentido de considerar a influência, tanto fixa quanto aleatória, de cada coorte nas curvas de evasão do sistema. Foi visto que há uma relação positiva e crescente entre essas duas variáveis (Tabela 6), fato confirmado na modelagem, pelo alto valor da razão de chances (2,588 no Modelo 120n IV e 5,056 no Modelo 240n IV). A literatura aponta a importância de se considerar as coortes de ingresso dos alunos na análise da evasão (Costa & Picanço, 2020; Gama, 2018; Nangino & Paiva, 2022), apesar de não haver resultados para se comparar, pois os estudos usam, geralmente, apenas uma coorte, e o foco fica, principalmente, na variável de idade no ingresso dos discentes, a qual será analisada juntamente aos demais determinantes individuais. Portanto, incluir o ano de ingresso como variável de controle é fundamental para captar adequadamente as influências temporais na evasão e fortalecer a robustez da modelagem.

2.2.3.1. Determinantes socioeconômicos

A Figura 5 mostra, visualmente, os resultados encontrados na modelagem. Será feita a discussão de cada determinante.

Figura 5 - Razão de chances dos determinantes socioeconômicos, com intervalo de confiança e nível de significância.



INSE

Na amostra 120n, o INSE é a única variável entre os determinantes socioeconômicos. Em todos os modelos sua razão de chances foi menor que um, sendo 0,877 no Modelo IV, indicando que quanto maior, ou seja, melhor, o indicador socioeconômico da escola de ensino médio (ou município) de origem, menor a propensão a evadir do sistema de ensino superior. Essa conclusão está fortemente alinhada aos achados da literatura (Costa & Picanço, 2020; Farias, Gouveia & Almeida, 2024; Gama, 2018; Lima & Fagundes, 2020; Martins et al., 2015; Silva, 2013; Teodoro & Kappel, 2020), apesar de não ter sido encontrado estudo que relacione evasão com o INSE do Inep, mas sim com outras variáveis socioeconômicas.

No entanto, esse resultado foi o oposto na amostra 240n. Apesar da estimativa estar bem próxima de 1, com valor 1,024 no modelo final, isso indica um pequeno aumento de 2,4% na razão de chances da evasão do sistema com o aumento de uma unidade do INSE. Alguns poucos estudos chegaram ao mesmo resultado, porém em contextos bem particulares: acompanhando

a coorte de 2013 que ingressou via Enem, na Universidade Federal da Fronteira do Sul (UFFS), até 2018, Nierotka, Salata & Klitzke Martins (2023) encontraram que o indicador de nível socioeconômico, construído pelos autores, não foi um fator relevante para explicar a evasão; em outro estudo, tese de doutorado, utilizando a coorte de 2016, dos ingressantes das universidades federais e da Universidade Federal de Minas Gerais (UFMG), Paula (2021) verificou que, em ambos os cenários, as variáveis relacionadas à situação socioeconômica dos alunos tiveram influência menor na evasão dos cursos e do sistema de ensino superior. Então é importante avaliar qual o contexto de cada resultado, e a interação com as demais variáveis, para assim tecer conclusões, que serão feitas na seção 3.3.

Escolaridade parental

Disponível apenas na amostra 240n, a escolaridade parental foi trabalhada de forma a identificar se a mãe ou o pai do egresso tinham ensino superior. No Modelo II, que incluiu apenas as variáveis de nível individual, havia uma leve influência desse fator na evasão do sistema (1,02), porém no modelo final a estimativa não foi estatisticamente significativa. No contexto dos ingressantes do ano de 2016 das universidades federais (Paula, 2021), foi encontrado resultado semelhante, e o autor sugere que essa variável está, na verdade, entrelaçada a diversos outros fatores. Além disso, os estudos que encontraram relação entre evasão e escolaridade dos pais estão inseridos em contextos e épocas específicos [Clique ou toque aqui para inserir o texto.](#)(Bayma-Freire, Roazzi & Roazzi, 2015; Costa & Picanço, 2020), limitando a abrangência dos achados e as possibilidades de discussão desse resultado em específico.

Ocupação parental

Variável disponível, também, apenas na amostra 240n, representa os egressos cujos pais estão em ocupações do grupo 5¹¹, que seriam as ocupações de maior complexidade e, consequentemente, melhor nível socioeconômico. Os resultados da modelagem indicam que há influência dessa variável na evasão do sistema, no sentido de alunos que têm o pai ou a mãe nesse grupo de ocupações têm 14,5% menos chances de evadir (razão de chances de 0,855 no Modelo IV), em linha com a literatura (Bayma-Freire, Roazzi & Roazzi, 2015; Costa & Picanço, 2020). Isso indica que a ocupação parental, em comparação à escolaridade parental, é um melhor preditor do efeito da influência familiar na macroevasão.

Número de moradores

¹¹ Grupo 5: Médico, engenheiro, dentista, psicólogo, economista, advogado, juiz, promotor, defensor, delegado, tenente, capitão, coronel, professor universitário, diretor em empresas públicas e privadas, político, proprietário de empresas com mais de 10 empregados.

O número de moradores, disponível apenas na amostra 240n, representa o tamanho da família do egresso. Todos os modelos apontam para a leve influência dessa variável na evasão (1,059 no Modelo IV), apontando que estudantes provenientes de famílias maiores podem enfrentar desafios adicionais, como responsabilidades domésticas ou limitações financeiras, que podem afetar negativamente sua permanência no ensino superior. Os estudos abordando esse determinante são limitados, mas corroboram com os achados (Coimbra, Silva & Costa, 2021), apesar do uso de variáveis como ter filhos ou cônjuge como *proxy* do tamanho da família (Costa & Picanço, 2020), variáveis essas não disponíveis nos microdados utilizados.

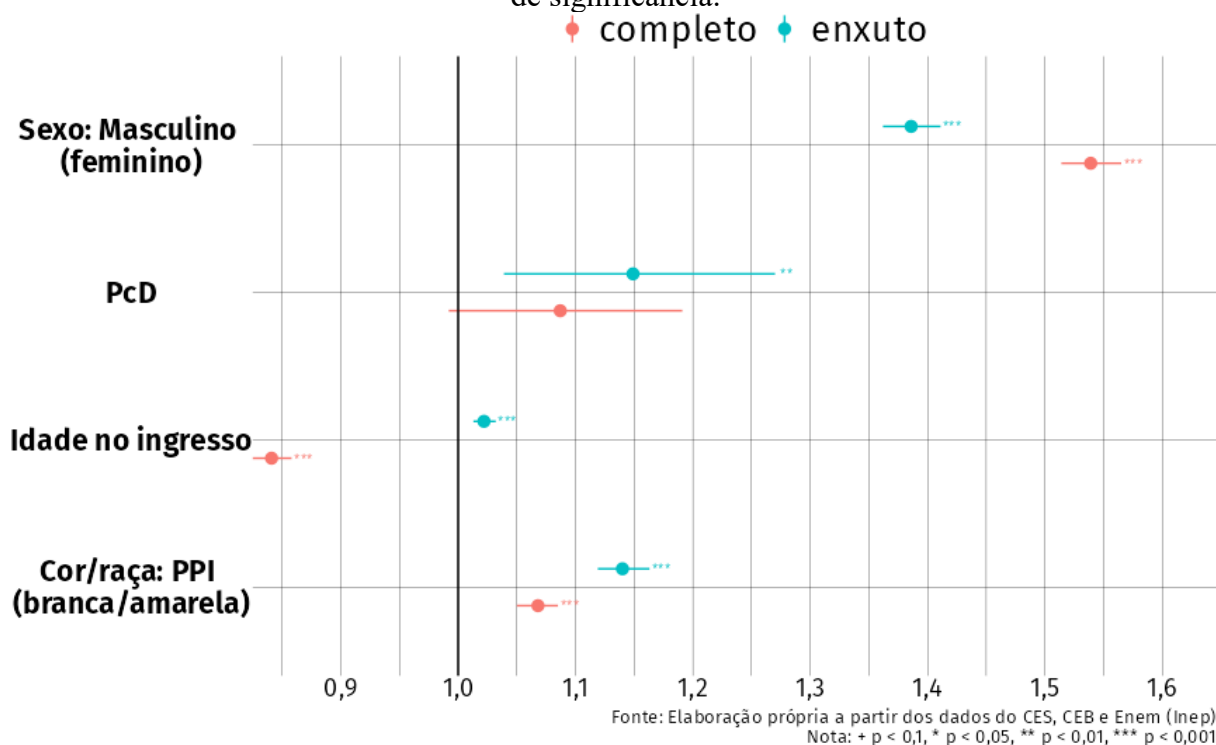
Renda familiar

A renda familiar, última variável dos determinantes socioeconômicos, também disponível apenas na amostra 240n, traz a renda familiar conforme declarada pelos estudantes, no momento do preenchimento do questionário Enem. Os resultados sugerem que quanto maior a renda familiar, menor a propensão do aluno a evadir do ensino superior (0,866 no Modelo IV), reforçando o peso dos fatores socioeconômicos quando se trata da evasão. Diversos estudos, baseados em diferentes contextos, conseguiram chegar à mesma conclusão (Costa & Picanço, 2020; Farias, Gouveia & Almeida, 2024; Paula, 2021), indicando que a falta de recursos financeiros mínimos é um entrave significativo ao sucesso educacional.

2.2.3.2. Determinantes individuais

A *Figura 6* mostra, visualmente, os resultados encontrados na modelagem. Será feita a discussão de cada determinante.

Figura 6 - Razão de chances dos determinantes individuais, com intervalo de confiança e nível de significância.



Cor ou raça autodeclarada

A variável cor/raça mostra uma associação significativa com a maior probabilidade de evasão entre estudantes autodeclarados pretos, pardos ou indígenas (PPI) em comparação aos brancos ou amarelos. Em todos os modelos, as razões de chances para estudantes PPI são positivas e significantes, indicando que, mesmo controlando por variáveis individuais, de curso e institucionais, estudantes PPI têm aproximadamente 13% a 14% mais chances de evadir em relação aos brancos ou amarelos. Tal resultado está de acordo com a literatura que estudou a relação entre raça e evasão (Araújo, Mariano & Oliveira, 2021; Arcanjo, 2022; Gomes & Hirata, 2024; Massini-Cagliari et al, 2021; Nierotka, Salata & Klitzke Martins, 2023). No entanto, apesar da raça ou cor ser um aspecto relevante contextualmente, sua interpretação isolada, por vezes, não aponta para diferenças significativas (Costa & Picanço, 2020; Klitzke & Carvalhaes, 2023; Li, 2016; Paula, 2021). Portanto, a cor/raça é um determinante significativo da evasão, especialmente quando combinada com outras variáveis de cunho socioeconômico.

Sexo

Em ambas as amostras, os homens apresentaram maior propensão a evadir, em comparação às mulheres. O maior efeito ocorreu na amostra 240n, com razão de chances igual a 1,539 no modelo final. Há uma forte concordância, na literatura, no sentido de os homens evadirem mais

que as mulheres (Costa; Bispo; Pereira, 2018; Gomes & Hirata, 2024; Li, 2016; Lopes et al., 2023; Nierotka, Salata & Klitzke Martins, 2023; Saccaro, França & Jacinto, 2019; Silva, 2018), mesmo considerando os diversos cenários em que tais resultados surgiram. Essa tendência pode ser explicada por diversos fatores, como menor engajamento acadêmico dos homens, pressões socioeconômicas que os levam a priorizar o trabalho em detrimento dos estudos e diferenças nas motivações educacionais entre os gêneros.

Idade no ingresso

Conforme visto na Tabela 8, a evasão do sistema aumenta à medida que avançamos nas categorias da faixa etária no ingresso. Os resultados da modelagem com a amostra 120n confirmam as médias calculadas, indicando que quanto maior a idade na data do ingresso, maior a propensão ao abandono do ensino superior (razão de chances de 1,022), o que está de acordo com boa parte da literatura (Lopes et al, 2023; Martins et al., 2015; Saccaro, França & Jacinto, 2019; Silva, 2013; 2018). Já na amostra 240n os resultados foram no sentido contrário: com uma razão de chances de 0,841, a probabilidade de evadir diminui com o avançar da idade no ingresso. Não foram encontrados estudos que confirmem esse achado, mas sabe-se que estudantes mais velhos e de melhores condições financeiras têm maiores taxas de sucesso na graduação (Gomes & Hirata, 2024; Paula, 2021), então considerando que a amostra 120n conta com apenas uma variável dos determinantes socioeconômicos (INSE), é possível que a interação dos indicadores tenha feito surgir tal efeito, refletindo a complexidade do fenômeno em análise.

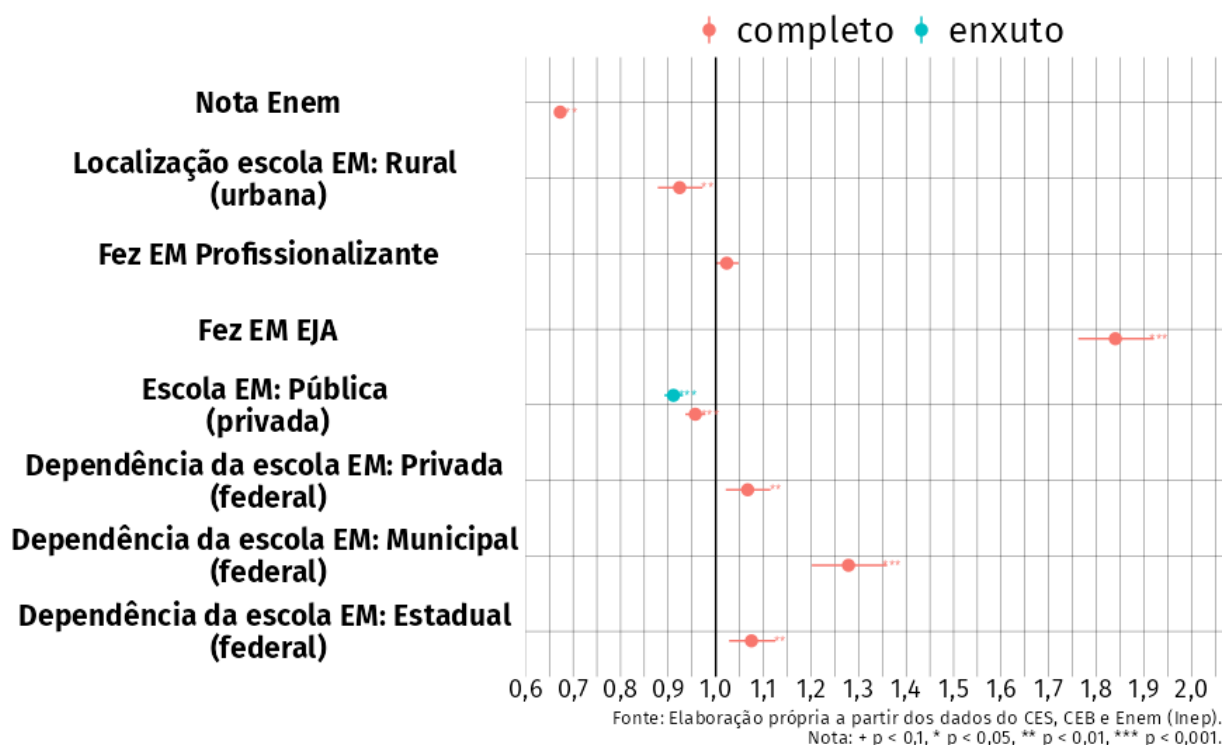
PcD

Os dados empíricos do CES apontam na direção de que a evasão do sistema dos alunos com deficiência é significativamente maior, em relação aos não PcD (Tabela 8). E, de acordo com as modelagens empreendidas, em ambas as amostras, esse dado é confirmado – o efeito é mais forte na amostra 120n, a qual não possui a maioria dos determinantes socioeconômicos, indicando que os egressos PcD têm propensão 15% maior a evadir. Tal achado está em consonância com a literatura (Pereira et al, 2020), demonstrando que alunos com deficiência enfrentam desafios adicionais no ensino superior, aumentando a probabilidade de evasão.

2.2.3.3. Determinantes pré-graduação

A Figura 7 mostra, visualmente, os resultados encontrados na modelagem. Será feita a discussão de cada determinante.

Figura 7 - Razão de chances dos determinantes pré-graduação, com intervalo de confiança e nível de significância.



Nota Enem

A nota Enem é uma das variáveis disponíveis apenas na amostra 240n. A modelagem indicou que há uma forte relação negativa e estatisticamente significativa entre as notas no Enem e a evasão do ensino superior – a razão de chances de 0,673 aponta que a propensão a evadir é 32,7% menor, quanto maior forem as notas no Enem do aluno. Resultado semelhante foi encontrado na literatura, no sentido de que estudantes com notas mais altas no Enem eram mais resilientes e tinham menor risco de abandono (Nierotka Salata & Klitzke Martins, 2023). Notas altas no Enem servem também como sinalizador significativo do desempenho dos estudantes, aumentando a probabilidade de ingresso no ensino superior (Carvalhaes, Senkevics & Ribeiro, 2023). Ainda assim, deve-se pontuar que estudantes que tiveram alto desempenho no Enem são mais suscetíveis à evasão nos primeiros anos, caso não estejam cursando sua primeira opção de curso desejada (Li, 2016).

Localização da escola do ensino médio

A especificação se a escola de ensino médio é rural ou urbana está presente apenas na amostra 240n. Com base nos resultados da modelagem, pode-se afirmar que estudantes provenientes de escolas secundárias localizadas em áreas rurais têm menor propensão a evadir, comparativamente aos alunos de escolas privadas. Não foram encontrados estudos que

utilizaram essa variável, e deve-se lembrar do alto percentual sem informação, 51,3% (Tabela 2). A hipótese inicial era de que os alunos de escolas secundárias rurais teriam maior propensão a evadir, a qual não foi confirmada empiricamente, possibilitando relacionar esse achado com a resiliência e determinação para concluir o ensino superior, frente às adversidades suportadas por quem mora no campo.

Tipo de ensino da escola do ensino médio

O detalhamento do tipo de ensino no ensino médio, se regular, educação de jovens e adultos (EJA) ou profissionalizante, está disponível apenas na amostra expandida. No entanto, mais uma vez tem-se um alto percentual sem informação, 56,5% (Tabela 2). Os resultados da modelagem apontaram que estudantes que fizeram ensino médio EJA têm alta propensão a evadir do ensino superior, com chances 84% maiores de abandono do que estudantes que não fizeram EJA. Há indícios de que o mesmo seja verdade para estudantes que vieram do ensino médio profissionalizante, em menor escala, mas a razão de chances de 1,023 é significativa apenas no intervalo de confiança a 90%. No que diz respeito aos cursos de engenharia da UNESP, foi encontrado o mesmo resultado relativo ao ensino médio profissionalizante (Silva et al, 2020), então pode-se inferir que o ensino médio cursado no ensino regular aumenta as chances de sucesso na graduação.

Dependência administrativa da escola de conclusão do ensino médio

A informação da escola de ensino médio em que os alunos concluíram seus estudos, se pública ou privada, foi utilizada em ambos os modelos, porém há ressalvas na qualidade da informação, devido ao alto percentual sem informação, 58,4% (Tabela 2). A regressão multinível de ambas as amostras indicou que a propensão a evadir dos estudantes que concluíram o ensino médio em escolas públicas é menor, comparativamente aos que concluíram em escolas privadas, o que não era esperado, de acordo com a hipótese inicial. São poucos os estudos que utilizaram esse determinante, e em contextos muito específicos, o que dificulta a comparabilidade. Nos cursos de administração e ciências contábeis verificou-se que a maioria dos evadidos eram de escolas secundárias particulares (Cunha et al, 2015), enquanto na UFPEL quase 82% eram oriundos de escolas secundárias públicas (Mello et al, 2013). Apesar das literaturas divergentes, infere-se que, ao controlar por diversos determinantes, os estudantes provenientes de escolas públicas têm maior sucesso na graduação.

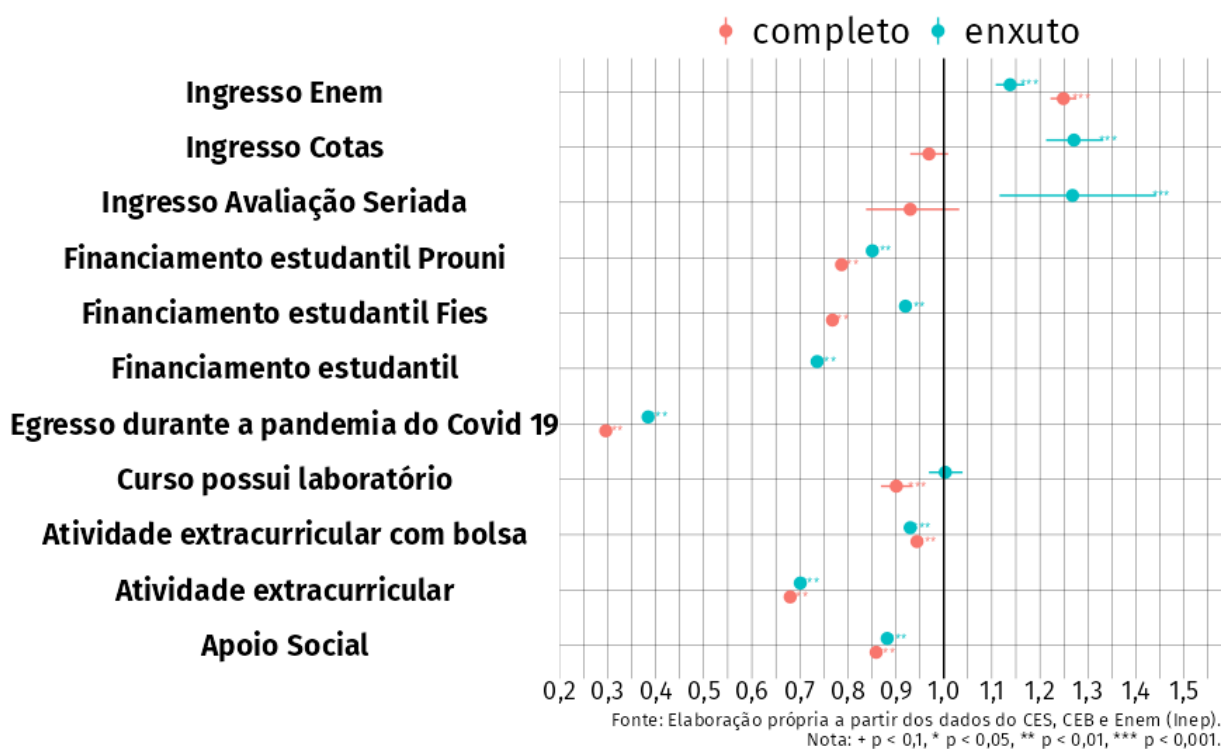
A modelagem da amostra expandida foi capaz de utilizar e testar, separadamente, a influência da escola cursada no ensino médio ser estadual, municipal ou privada, em comparação às escolas secundárias federais. Todas as estimativas indicaram que a propensão a evadir dos

alunos provenientes das escolas secundárias estaduais, municipais ou privadas são maiores que a dos alunos federais, reforçando o achado empírico de que os alunos de escolas de ensino médio públicas têm menores chances de evadir, em comparação com as escolas particulares, além de destacar o efeito positivo das federais no sucesso acadêmico dos estudantes.

2.2.3.4. Determinantes acadêmicos

A Figura 8 mostra, visualmente, os resultados encontrados na modelagem. Será feita a discussão de cada determinante.

Figura 8 - Razão de chances dos determinantes acadêmicos, com intervalo de confiança e nível de significância.



Ingresso Enem ou Sisu

Os modelos finais de ambas as amostras convergem ao achado de que os estudantes ingressantes via Enem ou Sisu têm maior propensão a evadir, com razão de chances de 1,138 no modelo enxuto e 1,249 no modelo completo. Esse comportamento já está documentado na literatura (Cabello et al, 2021; Teodoro & Kappel, 2020) e está relacionado ao sistema de seleção do Sisu, que permite aos candidatos ter uma segunda escolha de curso, o qual é abandonado assim que possível, levando à evasão (Li, 2016), além da preferência regional dos candidatos, que o fará abandonar o curso assim que conseguir uma vaga próxima à família (Campos et al, 2017). Pode-se afirmar que o Sisu facilita a mobilidade estudantil (Rangel et al, 2019), ao custo da evasão

nos cursos. Ainda assim deve-se ressaltar que o ingresso via Enem ou Sisu ampliou o acesso ao ensino superior, especialmente para a população de baixa renda, mesmo que exigindo alto desempenho (Carvalhaes, Senkevics & Ribeiro, 2023).

Ingresso avaliação seriada

O efeito do ingresso via avaliação seriada, como o Programa de Avaliação Seriada (PAS), foi mensurado em ambas as amostras, sendo significativo apenas na amostra 120n. A razão de chances de 1,268 indica que estudantes que ingressaram por avaliação seriada têm 26,8% mais chances de evadir do ensino superior. Devido à particularidade do tipo de ingresso, não foram encontrados estudos relacionados à evasão. No entanto, era esperado o comportamento oposto, devido à baixa taxa de evasão do sistema desse ingresso, 27,5% (Tabela 7), o que confirma a complexidade da interação entre os determinantes individuais, dos cursos e das instituições e o resultado final, a macroevasão.

Ingresso cotas

As diversas cotas de ingresso foram agrupadas e utilizadas em ambas as modelagens. O modelo com variáveis do CES teve significância estatística para afirmar que os estudantes ingressantes via reserva de vaga têm maior propensão a evadir, com 27,1% mais chance de não terminar a graduação. Porém ao considerar todos os determinantes, provenientes do CEB e Enem, a razão de chances foi estimada em 0,969, mas não foi estatisticamente significativa. Não há, na literatura, estudo de larga escala que permita comparação direta, então não é possível falar em consenso nesse item. São todos estudos em contextos específicos e medidas diferentes de evasão, com conclusões variadas, discutidas em seguida.

Na Unesp, os alunos que ingressaram entre 2014 e 2018, por cotas para pretos, pardos e indígenas (PPI) ou escola pública, tiveram taxas de evasão do curso maiores do que a ampla concorrência. Resultado semelhante ocorreu nos cursos de exatas e tecnologia da Universidade Federal do Recôncavo da Bahia (UFRB) (Lopes, 2023). Já no Campus Avançado Bom Sucesso, pertencente ao IF Sudeste MG, a evasão dos ingressantes por cotas foi significativamente menor (Nangino & Paiva, 2022); o mesmo ocorreu na UFRB, nos cursos não ligados às áreas de exatas ou tecnologia (Lopes, 2023). E na UFPEL não foi identificada diferença nos níveis de evasão entre os cotistas ou ampla concorrência (Campos et al, 2017), nem na coorte de ingressantes das universidades federais de 2016 (Paula, 2021).

Ou seja, apesar da falta de consenso sobre a relação entre reserva de vagas e evasão, assim como as estimativas opostas encontradas na modelagem, fatos que refletem a complexidade do tema, pode-se afirmar que a expansão do acesso à universidade por grupos sub-representados é

consequência positiva das políticas de cotas (Senkevics, 2021b), mas há que se pensar no apoio que esse novo público necessita para garantir sua permanência e conclusão.

Atividade extracurricular (com ou sem bolsa)

A participação dos alunos em atividades extracurriculares, conforme coletado pelo Inep (pesquisa, extensão, monitoria ou estágio não obrigatório), foi mensurada pelo total de semestres em que cada aluno participou de alguma atividade extracurricular em seu curso. Ambas as modelagens convergiram à conclusão de que quanto maior o engajamento do estudante nessas atividades, recebendo ou não bolsa, menor é a propensão a evadir – o modelo enxuto estimou redução de 29,9% e o modelo completo, 32% na chance de evasão do ensino superior, ambas estatisticamente significantes. Esse resultado já está bem consolidado na literatura e está relacionado, também, a melhoria do clima institucional (Barroso et al, 2022; Farias, Gouveia & Almeida, 2022; Paula, 2021; Teodoro & Kappel, 2020; Wegner, 2022). Há, então, um consenso sobre a importância das atividades extracurriculares não só para o sucesso acadêmico, mas também para o desenvolvimento interpessoal dos alunos. Deve-se atentar, no entanto, para que grupos marginalizados tenham acesso a essas oportunidades (Costa & Picanço, 2020).

Financiamento estudantil (geral, Fies e Prouni)

A utilização de financiamento estudantil para cursar as IES privadas foi mensurada de forma geral (apenas no modelo enxuto, devido à presença de multicolinearidade no modelo completo), considerando qualquer tipo de financiamento (reembolsável ou não, de qualquer esfera administrativa) e de forma específica, considerando apenas Fies ou Prouni. Os resultados de ambas as modelagens apontam para a reduzida e significativa propensão a evadir quando os alunos foram contemplados com algum financiamento estudantil – chance 26,4% menor de evasão, no modelo enxuto. Fies e Prouni têm resultados semelhantes, especialmente no modelo completo. Tais resultados estão, em sua maioria, de acordo com a literatura (Gomes & Hirata, 2024; Oshiro, 2018; Saccaro, França & Jacinto, 2019; Silva, 2013), indicando inclusive que bolsistas integrais têm melhor desempenho acadêmico (Lepine, 2019). Apenas um dos estudos da literatura (Silva & Santos, 2017) encontrou uma relação positiva entre evasão e o Fies, mas o cálculo da evasão foi feito de forma genérica, usando os estoques de matriculados, ingressantes e concluintes, gerando ressalvas quanto à pertinência do resultado encontrado. Além do efeito na evasão, deve-se destacar também o efeito na ampliação e viabilização do acesso ao ensino superior proporcionado por essas políticas, especialmente de estudantes de

primeira geração e estudantes trabalhadores (Farias, Gouveia & Almeida, 2022; Senkevics, 2021b).

Apoio social

O recebimento de alguma das categorias de apoio social, conforme instrumento do Inep (alimentação, moradia, transporte, material didático, bolsa trabalho ou bolsa permanência), foi mensurado pelo total de semestres nos quais os alunos tiveram acesso a algum dos apoios sociais, durante seu curso. Em ambas as modelagens, os resultados indicaram a significativa menor propensão a evadir dos alunos que foram amparados por alguma política de apoio social, com razão de chances 0,882 no modelo enxuto e 0,859 no modelo completo. A literatura chega à mesma conclusão (Araújo, Mariano & Oliveira, 2021; Colpo, Primo & Aguiar, 2024; Gama, 2018; Lima & Fagundes, 2020; Nierotka, Salata & Klitzke Martins, 2023; Paula, 2021), destacando que o efeito pode ser amplificado ao receber o apoio durante o primeiro ano da graduação (Li, 2016) ou quando recebido por alunos em situação de vulnerabilidade socioeconômica (Colpo, Primo & Aguiar, 2024). Pode-se afirmar, então, que o efeito e a importância das políticas de apoio social estão amplamente documentados na literatura, e confirmados empiricamente pelos modelos aplicados.

Laboratório

A presença de estruturas identificadas como laboratório, de acordo com o CES, foi caracterizada de forma dicotômica (sim ou não) e testada no contexto da macroevasão, com a ressalva de que não é possível capturar a qualidade da infraestrutura dos laboratórios. Enquanto que o modelo enxuto não identificou efeito significativo, o modelo completo, que considera as variáveis do CES, CEB e Enem, estimou a razão de chances de 0,901, ou seja, alunos que passaram por cursos com laboratório têm 9,9% menos chances de evadir do ensino superior. Isso está de acordo com a literatura encontrada (Saccaro, França & Jacinto, 2019), a qual identificou, de forma qualitativa, que a falta de acesso a práticas supervisionadas era um fator recorrente em entrevistas com evadidos (Parente, 2014). Ou seja, mesmo sem considerar fatores qualitativos sobre os laboratórios, ou categorizar por áreas do conhecimento que mais dependem desse tipo de estrutura, como saúde ou engenharias, infere-se que a prática e experiência oferecidas pelos laboratórios é relevante para o sucesso na graduação.

Pandemia

O impacto da pandemia de Covid-19 foi mensurado considerando especificamente os alunos que formaram ou evadiram em 2020, 2021 e 2022. Os resultados sugerem que os egressos de 2020 a 2022 têm uma significativa menor propensão a evadir do sistema, em ambos os

Categoria administrativa / IES privada

O detalhamento completo das categorias administrativas, de acordo com o Inep, possui as opções pública estadual, pública municipal, privada com fins lucrativos, privada sem fins lucrativos, especial e pública federal, esta última sendo usada como referência para as análises. Foi possível ter todo esse detalhamento no modelo enxuto, porém no modelo completo foi necessário consolidar em apenas uma variável, indicando se a IES era privada ou não, para evitar a multicolinearidade. As estimativas no modelo enxuto indicam que os alunos de todas as categorias, com exceção da pública estadual, têm significativamente maior propensão a evadir, quando comparados a alunos das públicas federais. Já ao incluir as variáveis do CEB e Enem e consolidar a variável de interesse, a razão de chances fica em 0,863, indicando que estudantes que passaram por IES particulares têm 13,7% menos chance de evadir do ensino superior, o que não está de acordo com a limitada literatura encontrada no tema (Saccaro, França & Jacinto, 2019; Silva & Sauaia, 2014) nem com a hipótese inicial. Ainda assim, deve-se salientar que os fatores econômicos que afetam a evasão são particularmente significativos para alunos de instituições privadas (Lopes, 2023), então há que se investigar possíveis interações não exploradas da relação entre macroevasão e IES privadas.

Organização acadêmica

Buscou-se aferir se o tipo de organização acadêmica da IES (centro universitário, faculdade, instituto federal ou centro federal) afeta a macroevasão, usando as universidades como categoria de referência da análise. No contexto da amostra com apenas as variáveis do CES, os alunos de centros universitários têm 13,9% menos chance de evadir do ensino superior, em comparação aos alunos das universidades, não havendo estimativas significativas para as demais categorias. E na amostra expandida, os estudantes provenientes das faculdades têm 33,5% mais chance de evadir da graduação, em comparação aos estudantes das universidades. Não foram encontrados estudos diretamente comparáveis, mas no contexto dos cursos de administração e ciências contábeis, entre 2001 e 2010, as taxas de evasão das faculdades, institutos federais e centros federais eram significativamente maiores (Cunha et al, 2015), corroborando com o achado da amostra expandida. Conclui-se que a variável de organização acadêmica foi pouco explorada na literatura, mas tem potencial para ajudar a entender o fenômeno da evasão.

IES na capital

A presença física do campus das IES pode ser um fator que influencie a macroevasão. Os resultados indicam que houve concordância entre os ambos os modelos, no sentido de que,

quando a IES cursada é na capital do estado, os alunos têm uma maior propensão a evadir do ensino superior. Não foram localizados estudos comparáveis que tenham utilizado essa variável, e o resultado empírico não está de acordo com a hipótese inicial. No entanto, pode-se inferir que estudantes nas capitais dos estados enfrentam custos de vida mais altos, o que aumenta a pressão financeira e contribui para o aumento das taxas de evasão. Além disso, a presença de oportunidades de emprego mais imediatas nas capitais pode incentivar os estudantes a abandonarem os estudos em favor da entrada no mercado de trabalho. Ou seja, é necessário explorar a interação dessa variável com outros determinantes para compreender seu efeito na evasão.

Número de servidores (docentes e TAEs)

Pretendeu-se testar o efeito do tamanho do corpo docente e administrativo na macroevasão dos alunos, especialmente por não terem sido encontradas evidências na literatura dessa relação. Ambos os modelos indicaram que quanto maior a estrutura de recursos humanos da IES, maior a propensão do aluno a evadir do ensino superior, apesar de ter sido estatisticamente significativa apenas no modelo que considera as variáveis do CES, CEB e Enem. Tal resultado, divergente do esperado nas hipóteses iniciais, pode indicar que o tamanho da instituição aumenta, também, a complexidade de todo o sistema, prejudicando um possível olhar mais atento aos alunos que mais necessitam, dificultando seu acompanhamento e diminuindo suas chances de sucesso na graduação.

Docentes com pós-graduação

Foi utilizada a métrica do percentual de docentes com pós-graduação *stricto sensu* (mestrado ou doutorado) em cada curso e IES. As modelagens resultaram na concordância de que quanto maior esse percentual, menor é a propensão dos alunos a evadir do ensino superior – diminuição das chances de evasão em 9,1% e 22,5% nos modelos enxuto e completo, respectivamente, sendo ambas estatisticamente significativas. Não foi encontrada literatura que relacionasse alguma métrica de evasão à escolaridade dos docentes. Ainda assim, habilidades pedagógicas deficientes dos educadores podem levar ao aumento das taxas de abandono escolar entre os alunos, particularmente nos primeiros anos de sua educação (Machado et al, 2021). E a relevância do currículo dos cursos, influenciada pela formação de professores, é fundamental para o envolvimento dos alunos (Barroso et al, 2022). Então depreende-se que graus maiores de formação acadêmica sinalizam positivamente os alunos no sentido de buscarem e conseguirem o sucesso na graduação.

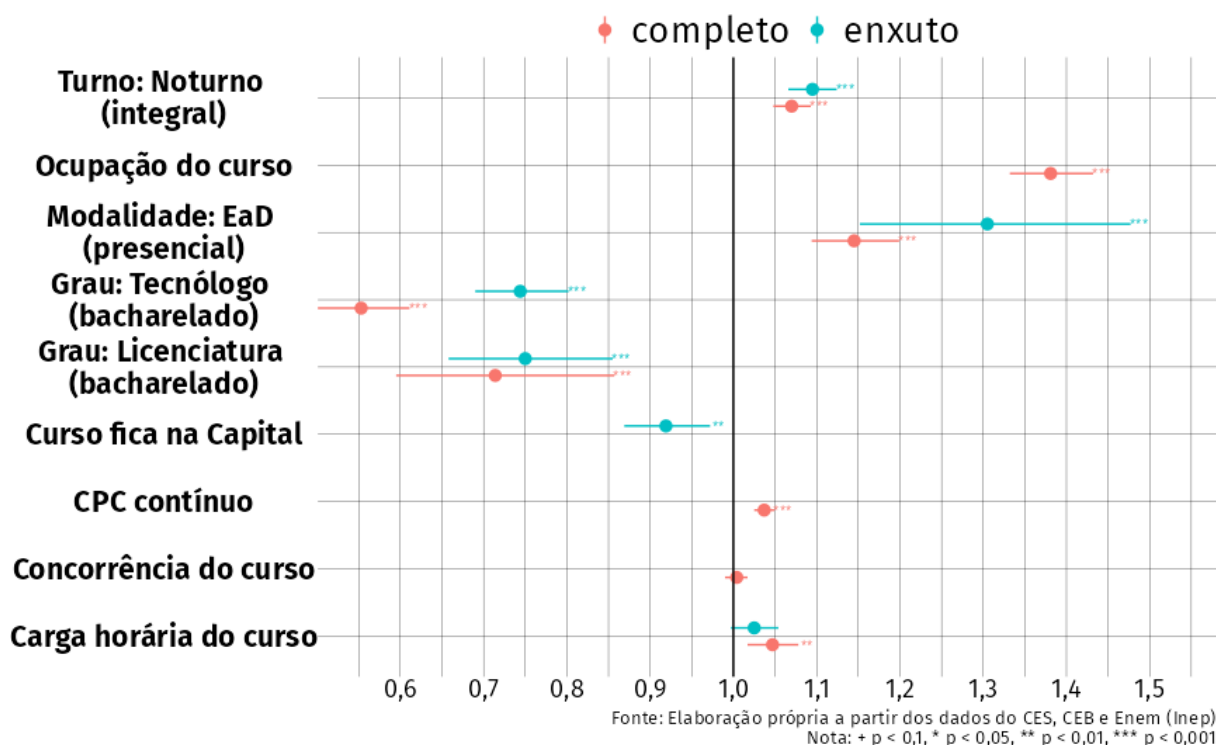
IGC

O Índice Geral de Cursos (IGC) é um dos indicadores de qualidade da educação superior que avalia as IES. Sua metodologia considera a média do Conceito Preliminar de Curso (CPC) ponderada pelos matriculados e a média dos conceitos de avaliação da pós-graduação ponderada, também, pelos matriculados. Disponível apenas no modelo completo, a razão de chances foi estimada, significativamente, em 0,919, indicando que estudantes de IES com maiores valores de IGC contínuo têm menor propensão a evadir do ensino superior. Isso está de acordo com a literatura, mesmo quando utilizada outra forma de cálculo da evasão (Silva & Sauaia, 2014). Mas olhando para a retenção, há evidências de que o IGC tem um efeito positivo marginal, sugerindo que instituições de maior qualidade podem levar a menores taxas de abandono (Araújo, Mariano & Oliveira, 2021; Silva & Santos, 2017; Silva & Sauaia, 2014). Então considerando o objetivo do IGC como métrica e a percepção de sua função como indicador que avalia a qualidade das IES (Silva & Santos, 2017), apoiado no achado empírico, pode-se concluir que um maior valor de IGC contribui para a redução das taxas de evasão no ensino superior, alinhando-se ao seu propósito de avaliar a qualidade das instituições de ensino superior.

3.2.3.6. Determinantes curriculares

A Figura 10 mostra, visualmente, os resultados encontrados na modelagem. Será feita a discussão de cada determinante.

Figura 10 - Razão de chances dos determinantes curriculares, com intervalo de confiança e nível de significância.



Grau

A macroevasão foi avaliada também de acordo com o grau acadêmico dos estudantes: bacharelado, licenciatura e tecnólogo, sendo o primeiro utilizado como referência na análise. Os resultados de ambas as modelagens convergiram na conclusão de que tanto os alunos da licenciatura quanto os alunos do tecnólogo têm menor propensão a evadir, quando comparados aos alunos do bacharelado, com todas as estimativas estatisticamente significativas. Apesar de não haver estudos diretamente comparáveis na literatura, alguns deles, em seus contextos específicos, encontraram resultados discordantes com as modelagens empreendidas e de acordo com a hipótese inicial postulada para esta variável. No contexto das IFES e tendo como referência cursos de bacharelado de maior prestígio (Medicina, Direito, Odontologia e Engenharias), a propensão a evadir dos estudantes do bacharelado e licenciatura era maior (Assís et al, 2022; Paula, 2021), semelhante ao encontrado nos cursos de educação física de Porto Velho, Rondônia, nos quais a licenciatura tinha evasão significativamente maior que o bacharelado (Assís et al, 2022). Conclui-se, então, pela necessidade de aprofundar a análise da variável de grau acadêmico, incluindo interações com outros determinantes que ajudem a explicar o resultado encontrado, como turno ou modalidade, por exemplo.

Turno

O efeito na macroevasão de se cursar a graduação no turno noturno foi aferido frente ao turno integral. Ambos os modelos convergiram à conclusão de que estudantes de cursos noturnos têm maiores chances de evadir do ensino superior: 9,5% no modelo enxuto e 7% no completo. Tal resultado está de acordo com a literatura que tratou da macroevasão (Lima & Fagundes, 2020; Paula, 2021) e da evasão dos cursos em Portugal (Realinho et al., 2022). É importante destacar que muitos estudantes do noturno precisam trabalhar para se manter, então não há dedicação exclusiva à graduação, o que pode influenciar negativamente no sucesso acadêmico (Lima & Fagundes, 2020). E apesar da maior propensão a evadir, os cursos noturnos prestam um papel social de maior democratização do acesso à educação superior (Zalaf Caseiro & Ramos De Azevedo, 2019), sendo um componente essencial da educação de adultos no Brasil (Filho; Nery, 2009).

Modalidade

A modalidade dos cursos, se presencial ou EaD, é um tema que vem ganhando espaço não só na mídia e nas discussões acadêmicas, mas principalmente quando se fala na expansão do EaD (Senkevics, 2021b). Dessa forma, foi testada a influência de se cursar uma graduação a distância, tendo como referência cursos presenciais. Os resultados indicaram que alunos do EaD têm maior propensão a evadir, quando comparados aos alunos presenciais, sendo o efeito mais forte no modelo enxuto (razão de chances 1,305), que considera apenas as variáveis do CES. Não foram encontrados estudos diretamente comparáveis, mas a literatura fala, de forma geral, sobre as altas taxas de evasão do EaD, especialmente das instituições privadas (Bentes & Kato, 2014) e entre estudantes de baixa renda e aqueles com responsabilidades familiares (Silva et al, 2025). O perfil dos alunos que fazem presencial é diferente dos que fazem EaD, sendo os últimos, em média, mais velhos (Senkevics, 2021a), o que pode corroborar a compreensão das maiores taxas de evasão dessa modalidade. Por fim, quando se fala em ensino a distância, há toda uma discussão sobre a mercantilização dessa modalidade (Branco, Conte & Habowski, 2020), que afeta não só a evasão e o sucesso dos alunos, mas também seu futuro no mercado de trabalho.

Curso na capital

A localização do curso na capital do estado foi avaliada como um possível determinante da macroevasão. O resultado na amostra enxuta, que tem apenas as variáveis do CES, teve razão de chances igual a 0,919, indicando que os estudantes que cursaram sua graduação nas capitais tinham menor propensão a evadir do ensino superior. Essa estimativa é o oposto da encontrada entre macroevasão e IES na capital, apontando para uma possível interação entre IES e curso

na capital, que poderia ser explorada para elucidar seus efeitos na evasão. Foi encontrado apenas um estudo nesse tema, mas sobre retenção, o qual afirma que os estudantes que frequentam instituições na capital experimentam uma probabilidade reduzida de períodos de retenção prolongados em comparação com aqueles em outros locais (Araújo, Mariano & Oliveira, 2021). Infere-se que há outros fatores afetando o sucesso na graduação desses alunos das capitais, como nível socioeconômico mais alto, disponibilidade de serviços de apoio e oportunidades acadêmicas e econômicas mais diversificadas, além de acesso à educação de qualidade.

Carga horária

Foi explorada a relação entre a carga horária do curso e seu efeito na macroevasão. Os resultados, em ambos os modelos, apontam para o achado de que quanto maior a carga horária, maior a propensão a evadir do ensino superior. O efeito em si não é forte, mas é significativo apenas no modelo completo: 4,7% mais chances de evasão, ou seja, é necessário incluir as variáveis do CEB e Enem para que esse efeito surja. Na literatura abordada, a carga horária total é vista como uma das características mais determinantes da evasão (Teodoro & Kappel, 2020). No âmbito da educação a distância, os alunos relataram que as altas cargas horárias dos cursos, em conjunto com as demais demandas de trabalho e pessoais, contribuíram na decisão de evadir (Bentes & Kato, 2014). Importante ressaltar que cursos de maior prestígio, que têm também elevadas cargas horárias, não padecem do problema de maior evasão (Paula, 2021), então acredita-se que a relação entre carga horária e macroevasão poderia ser mais bem explicada com a interação de outros determinantes.

Total de vagas

Buscou-se identificar se o total de vagas oferecidas tinha efeito na macroevasão. A estimativa no modelo enxuto não foi estatisticamente significativa, mas ao adicionar as variáveis do CEB e Enem, o modelo completo apontou que a propensão a evadir foi 15,8% maior, dependendo do número de vagas oferecidas pelo curso. Não foram encontrados estudos comparáveis na literatura, apesar de um dos estudos ter achado uma leve correlação negativa (-1,41%) entre evasão e total de vagas (Teodoro & Kappel, 2020). Como cursos com mais vagas atraem, geralmente, mais candidatos, infere-se que isso crie pressão sobre o desempenho dos alunos e leve-os a evadir, após o ingresso. Outra explicação alternativa seria a de que, quanto mais vagas, maior é a diluição dos diversos recursos institucionais, pode levar a uma diminuição percebida ou real na qualidade da educação, levando os alunos a abandonar o curso (Li, 2016). Seria necessário um melhor detalhamento dessa variável, preferencialmente com análise conjunta da demanda e ocupação, para compreender esse resultado.

Demanda

A demanda ou concorrência, definida como a relação entre o total de candidatos e o total de vagas, foi analisada no modelo completo. A estimativa obtida não foi estatisticamente significativa, impossibilitando análises empíricas posteriores.

Ocupação

A ocupação, definida como a razão entre o total de ingressantes e o total de vagas, foi testada no âmbito da macroevasão, no modelo completo. O resultado indica que estudantes que ingressaram em cursos de alta ocupação têm maior propensão a evadir do ensino superior, com razão de chances de 1,381. Isso não está de acordo com a literatura encontrada sobre esse tema específico (Hoed, Ladeira & Leite, 2018) nem com a hipótese inicialmente definida. No entanto, a alocação de vagas nem sempre está alinhada às preferências ou capacidades dos alunos, levando a possíveis ineficiências. Os alunos podem se matricular em cursos que não são sua primeira escolha, resultando em insatisfação e potencial desistência (Zalaf Caseiro & Ramos De Azevedo, 2019), o que sugere a necessidade de melhor explorar empiricamente essa variável em conjunto com outros determinantes.

CPC

A relação entre o Conceito Preliminar de Curso (CPC) contínuo, um dos indicadores de qualidade da educação superior, e a macroevasão, foi testada no modelo completo. A estimativa da razão de chances de 1,037 aponta que, quanto maior o CPC contínuo do curso, maior a propensão dos alunos a evadir do ensino superior. Não foram encontrados estudos na literatura que utilizassem diretamente esse indicador de qualidade. Considerando que a hipótese inicial era de que alunos provenientes de cursos com melhor CPC teriam mais chances de sucesso na graduação, pode-se concluir pela necessidade de interações adicionais dessa variável com os outros determinantes, especialmente os socioeconômicos, presentes apenas na amostra expandida, para melhor compreender quais relações pesam nessa estimativa.

2.3. Considerações Finais

Este capítulo avançou na compreensão da macroevasão do ensino superior brasileiro entre 2005 e 2022, utilizando microdados identificados longitudinais do Inep e modelos multiníveis. Devido ao tamanho e complexidade do problema, foram utilizadas duas amostras diferentes, uma considerando apenas variáveis do CES e outra utilizando todas as variáveis disponíveis do CES, CEB e Enem. A amostragem foi utilizada, inclusive, para lidar com os problemas de qualidade dos dados destacados ao longo do texto, considerando que os registros do Inep são imutáveis. Os resultados confirmam que a evasão definitiva é um fenômeno multifatorial,

marcado por desigualdades estruturais e lacunas institucionais. A seguir, sintetizam-se as principais descobertas e implicações.

Pode-se dizer que os determinantes socioeconômicos e individuais são o eixo central dos achados sobre o fenômeno. Estudantes de maior renda familiar, de famílias menores, cujos pais estão em ocupações de nível superior, apresentaram até 14,5% menos chances de abandonar o sistema. Apesar da escolaridade dos pais não ter sido estatisticamente significativa na modelagem, a macroevasão calculada confirmou, também, essa desigualdade, evidenciando que a importância que os pais dão à educação tem mais peso do que o nível educacional deles (Foley, Gallipoli & Green, 2014). Paradoxalmente, o Indicador de Nível Socioeconômico (INSE) da escola de origem revelou efeitos contraditórios: reduziu a probabilidade de macroevasão na amostra restrita (CES), mas teve impacto residual na expandida (com Enem e CEB), sugerindo a fragilidade do indicador, que é uma medida agregada de diversas métricas. Há também invisibilidades demográficas. Alunos pretos, pardos e indígenas (PPI) tiveram risco até 14% maior de evasão, mesmo após controle por todos os demais determinantes. Homens tiveram até 54% mais chances de evadir que mulheres, possivelmente por questões culturais e pressões para ingressar precocemente no mercado de trabalho. Pessoas com deficiência (PcD) enfrentaram risco até 15% maior de evasão, destacando falhas em políticas de acessibilidade física e pedagógica. Estudantes de escolas secundárias rurais, embora minoritários, mostraram resiliência, com chance de evasão 8% menor em comparação às urbanas, sugerindo potencial subestimado nesse grupo.

O desempenho no Enem é um indicador de sucesso para se concluir a graduação, mas já a forma de ingresso via Enem ou Sisu aumenta consideravelmente a chance de evadir do ensino superior. Esse paradoxo reflete dinâmicas complexas do sistema: estudantes com notas altas no Enem têm 32,7% menos propensão a evadir, evidenciando que o desempenho prévio é um preditor robusto de resiliência acadêmica. No entanto, ingressantes pelo Sisu apresentam até 24,9% mais chances de abandonar o sistema, possivelmente devido a três fatores-chave: a) mobilidade estudantil induzida, pois o sistema de seleção permite que alunos migrem para cursos ou instituições mais alinhados às suas preferências após o ingresso inicial, gerando rotatividade (Li, 2016); b) desencontros de expectativas, pois parte dos estudantes opta por cursos fora de sua primeira escolha durante o processo seletivo, levando a insatisfação e desistência precoce (Nierotka, Salata & Klitzke Martins, 2023; Silva, 2018; Cabello et. al, 2021); c) desafios de adaptação regional, pois alunos que se deslocam para outras regiões

enfrentam custos adicionais de moradia, distância familiar e dificuldades de integração, fatores que ampliam o risco de evasão (Li, 2016).

O tipo de ensino médio cursado emergiu como um determinante pré-graduação crítico da macroevasão, com resultados que expõem desigualdades profundas no sistema educacional brasileiro. Estudantes que cursaram o Ensino de Jovens e Adultos (EJA) apresentaram 84% mais chances de evadir do ensino superior em comparação aos egressos do ensino regular, enquanto aqueles provenientes de escolas profissionalizantes tiveram risco 2,3% maior. Esses achados destacam fatores como o desalinhamento curricular e lacunas formativas desses tipos de ensino, que focam em formação técnica acelerada. Além disso, há uma sobrecarga de responsabilidades por parte desse público, que deve conciliar trabalho e estudo (Silva et al., 2020), dificultando ainda mais o sucesso acadêmico.

Não foi possível chegar a uma conclusão empírica e convergente sobre o ingresso por reserva de vagas. Apesar da estatística de macroevasão calculada ser menor para os cotistas e o modelo expandido não ter encontrado relação estatisticamente significativa entre cotas e evasão, o modelo enxuto apontou que os ingressantes por cota têm maior propensão a evadir do ensino superior. Serão necessárias novas modelagens, considerando a interação desse tipo de ingresso com determinantes socioeconômicos e individuais, para chegar a um consenso empírico, sem desconsiderar a importância dessa política para o aumento da diversidade de acesso à educação superior.

Os determinantes institucionais e de qualidade acadêmica são mais úteis para ter um panorama da macroevasão do que para apontar possíveis políticas, dada a característica dessas variáveis. Ao contrário da hipótese inicial, os estudantes de universidades não têm, no geral, menor propensão a evadir, confirmado pela estatística de macroevasão calculada. Já os egressos das IES públicas federais, no modelo enxuto, apresentam menores chances de evadir, em comparação com as demais categorias administrativas, com exceção das públicas estaduais. No entanto, esse resultado foi contrariado pelo modelo expandido, no qual os alunos de IES privadas tiveram menor propensão a evadir. Os indicadores de qualidade utilizados, IGC e CPC, também apontaram em direções opostas. Ou seja, será necessária nova rodada de modelagens, incluindo novas interações, para chegar a um consenso nesses determinantes.

Ainda nos determinantes institucionais, alunos de IES com alto percentual de docentes com mestrado ou doutorado têm menor propensão a evadir, sinalizando a importância da formação docente de qualidade e seu impacto na formação discente. Essa métrica, inclusive, faz parte do cálculo do CPC, o qual faz parte do cálculo do IGC, reforçando ainda mais sua relevância

(Machado, Terra & Tannuri-Pianto, 2024), mostrando que políticas que incentivem o aumento do percentual de professor com pós-graduação *stricto sensu* têm efeito em cascata não só nos indicadores de qualidade da educação superior, mas também em métricas relacionadas à permanência e sucesso acadêmico.

A análise integrada dos determinantes acadêmicos financiamento estudantil, atividades extracurriculares, apoio social e laboratório mostra que são pilares interdependentes para o sucesso discente. Programas como Fies e ProUni, apesar de cobrirem apenas 26,6% dos alunos analisados, reduzem a probabilidade de evasão, comprovando seu papel crítico na democratização do acesso e na permanência de estudantes de baixa renda (Lepine, 2019). Essas políticas não apenas mitigam pressões financeiras imediatas, mas também funcionam como mecanismos de inclusão ao permitir que alunos concentrem esforços nos estudos, por não precisar conciliá-los a jornadas extenuantes de trabalho.

As atividades extracurriculares emergiram como um fator transformador: alunos que participaram de projetos de extensão, pesquisa ou monitoria tiveram até 32% menos chances de evadir do sistema. Essa redução está associada ao desenvolvimento de habilidades socioemocionais, como liderança, gestão do tempo e trabalho em equipe, que fortalecem o vínculo institucional e ampliam perspectivas profissionais (Barroso et al, 2022). Além disso, a interação com pares e docentes nessas atividades cria redes de apoio informais, essenciais para a adaptação ao ambiente acadêmico, especialmente entre estudantes de primeira geração (Wegner, 2022).

O apoio social integrado – incluindo bolsas-permanência, auxílio alimentação e moradia – mostrou-se igualmente decisivo, reduzindo a propensão à evasão em até 14%. Alunos beneficiados por essas políticas relatam maior engajamento nas disciplinas e menor sensação de isolamento, fatores críticos para a conclusão da graduação (Colpo, Primo & Aguiar, 2024; Gama, 2018; Paula, 2021). Por fim, a presença de laboratórios também se mostrou positiva para diminuir as chances de evadir, apesar da fragilidade informacional de não ser possível saber a qualidade das instalações.

A convergência dessas iniciativas demonstra que a permanência e consequente sucesso no ensino superior depende de uma estratégia multidimensional. Financiamento assegura acesso inicial; atividades extracurriculares e laboratórios fortalecem o engajamento; e políticas de apoio social garantem condições materiais para a continuidade. Juntas, essas ações não apenas reduzem a evasão, mas também nivelam oportunidades, transformando a educação superior em

um efetivo instrumento de equidade social, pavimentando o caminho para cumprimento da meta 13.8 do Plano Nacional de Educação¹².

Para finalizar as considerações deste capítulo, é importante trazer as limitações da pesquisa. O principal desafio foi a disponibilidade e qualidade dos dados. A falta de interoperabilidade dos sistemas governamentais, no âmbito dos dados educacionais, dificulta sobremaneira o desenvolvimento de estudos complexos e robustos – não é possível calcular a macroevasão, por exemplo, sem ser no ambiente seguro do Inep. E mesmo nesse ambiente, falta a integração das bases de dados, onerando os pesquisadores interessados e diminuindo a qualidade informacional – foram realizadas diversas etapas de melhoria informacional que poderiam ser mais facilmente efetuadas pelos detentores dos dados. Deve-se atentar, também, para a possível presença de vieses de seleção, não só por conta das amostras realizadas, mas também pela complexidade do fenômeno estudado.

Outra limitação reside na limitação das variáveis disponíveis. O Inep padronizou, ao longo de muitos anos, o instrumento de coleta de dados de suas pesquisas, como o CES, CEB e Enem, e ao mesmo tempo que isso gera vantagens, surgem também desvantagens, pela perda de informações mais específicas que auxiliariam a melhor qualificar não só a evasão, mas qualquer fenômeno relacionado da área. Ainda assim, entende-se que o *trade off* é necessário para a uniformização dos dados educacionais. Isso poderia ser resolvido pelas informações disponíveis nos sistemas acadêmicos de cada IES, porém a falta de interoperabilidade é, mais uma vez, um obstáculo relevante.

Por fim, deve-se destacar uma possível subestimação da evasão, por viés de seleção temporal. Isso porque a análise se restringiu a estudantes com vínculos finalizados, de evasão ou conclusão, decisão que pode ter consequências no cálculo das taxas das extremidades do período analisado. Pode-se falar, também, em diferentes tempos de exposição das coortes estudadas, decisão que tem potencial de viesar as estimativas obtidas.

Feitas essas ressalvas, os resultados aqui apresentados constituem contribuições à literatura da evasão do sistema de ensino superior, estimadas com rigor metodológico e de forma inovadora por regressões logísticas multinível. O modelo hierárquico permitiu capturar a variabilidade não observada entre instituições e cursos, e a inclusão de interceptos e coeficientes aleatórios

¹² Elevar gradualmente a taxa de conclusão média dos cursos de graduação presenciais nas universidades públicas, de modo a atingir 90% (noventa por cento) e, nas instituições privadas, 75% (setenta e cinco por cento), em 2020, e fomentar a melhoria dos resultados de aprendizagem, de modo que, em 5 (cinco) anos, pelo menos 60% (sessenta por cento) dos estudantes apresentem desempenho positivo igual ou superior a 60% (sessenta por cento) no Exame Nacional de Desempenho de Estudantes - ENADE e, no último ano de vigência, pelo menos 75% (setenta e cinco por cento) dos estudantes obtenham desempenho positivo igual ou superior a 75% (setenta e cinco por cento) nesse exame, em cada área de formação profissional.

para ano de ingresso nas IES reconheceu trajetórias distintas de macroevasão entre coortes, enquanto o efeito aleatório por curso Cine controlou diferenças estruturais. A abordagem multinível superou limitações de modelos tradicionais ao incorporar a natureza hierárquica dos dados educacionais, mesmo frente aos desafios citados. Essa estratégia metodológica reforçou a robustez das estimativas, oferecendo resultados inéditos sobre a macroevasão como fenômeno sistêmico, influenciado por diversos determinantes.

3. Interoperabilidade das bases de dados educacionais e o desafio do compartilhamento de dados

3.1. Interoperabilidade e reutilização de dados: conceitos e desafios

Os dados são vistos, atualmente, como uma mercadoria valiosa. Muitos estão sendo produzidos por governos, empresas e instituições, já que a tecnologia permitiu que fossem melhor analisados e que decisões fossem aprimoradas por meio de técnicas de estatística, *big data*, aprendizado de máquina e inteligência artificial, entre outras. Uma revolução parece estar acontecendo.

Como uma mercadoria valiosa, os dados têm custos: são caros para produzir, armazenar e analisar, entre outros aspectos. Portanto, é eficiente reutilizar dados sempre que possível. Governos e organizações internacionais, como as Nações Unidas (ONU) ou a Organização para a Cooperação e Desenvolvimento Econômico (OCDE), têm promovido três princípios: de reutilização de dados, do “uma vez só” e da “fonte única da verdade”. O primeiro trata de não criar novas bases de dados quando uma informação já existe em algum lugar. O segundo refere-se a não solicitar mais de uma vez a mesma informação ao fornecer um serviço público. O terceiro diz respeito à unificação das fontes de informação para evitar instruções e fatos conflitantes. Parent & Spaccapietra (2000) destacam, por exemplo, a importância da reutilização de dados e como as bases de dados devem ser construídas a partir da integração, e não do zero. Pode-se afirmar que os dados tendem a aumentar seu valor à medida que são reutilizados repetidamente (Boz & Simsek, 2022).

Embora louváveis, esses princípios não são tão fáceis de colocar em prática. A digitalização é um processo relativamente recente e, em muitas instituições, os processos de coleta de dados evoluíram a partir de práticas e culturas com décadas de existência que precedem os computadores ou a internet. Isso significa que enfrentamos um desafio de integração e interoperabilidade, senão também cultural/institucional. Como Ribes (2017) coloca: “A interoperabilidade dos dados é histórica, infraestrutural, relativamente irreversível, negociada, epistêmica, com falhas e sem falhas aparentes, e está se aproximando do status de um valor geral em vez de um meio específico para um fim”.

Em termos simples, integração tem a ver com combinar dois conjuntos de dados. Interoperabilidade refere-se a garantir que duas variáveis ou dois conjuntos de dados sejam compatíveis. Por exemplo, se o número de escolas em um conjunto de dados e o número de instituições educacionais em outro tiverem a mesma definição e considerarem um conjunto igual de critérios em suas regras (e desconsiderarem também os mesmos critérios), há

compatibilidade aqui. Nas palavras de Nasfi *et al.* (2023): “A avaliação da interoperabilidade baseia-se em uma avaliação quantitativa do grau em que um banco de dados gerido localmente e um banco externo podem ser integrados. O sucesso dessa integração depende da quantidade de identificadores correspondentes entre ambos os bancos”. Ou ainda: permite o uso dos dados de um contexto em outro, de forma automatizada e com baixo custo (Boz & Simsek, 2022).

No entanto, interoperabilidade vai além da questão técnica. A parte técnica é certamente um pré-requisito para a interoperabilidade, mas há também uma camada não digital. Ou seja, interoperabilidade é composta por interoperabilidade legal (organizações sob diferentes estruturas legais encontram formas de trabalhar juntas); organizacional e cultural (processos alinhados); semântica (sistemas diferentes podem interpretar informações de maneiras compatíveis); e técnica (inclui especificações de interface, questões de integração dos dados, apresentação dos dados e outros protocolos) (World Bank, 2022).

Lewis *et al* (2008) alertam que padrões não são suficientes para garantir interoperabilidade porque os dados dependem da interpretação dos usuários. Madnick & Zhu (2006) seguem na mesma linha, argumentando que problemas relacionados à qualidade dos dados são problemas de má interpretação e, portanto, questões semânticas dos dados. Além disso, interoperabilidade e integração dependem tanto das imperfeições dos dados, como sua completude (Nutt; Razniewski & Vegliach, 2012). Muitos países em desenvolvimento possuem infraestrutura precária para registro de dados administrativos, resultando em conjuntos com baixa qualidade. Isso às vezes é remediado por pesquisas ou censos; no entanto, restrições como falta de capital humano qualificado, financiamento inadequado ou interferência política podem levar à baixa qualidade dos dados (Chege & Wanjohi, 2023).

3.2. Dados abertos e estratégias de compartilhamento

Interoperabilidade pode parecer algo distante do cotidiano das pessoas. No entanto, é ela que viabiliza grande parte dos serviços públicos digitais. Na educação, por exemplo, é importante que os dados fornecidos por uma escola ao órgão local sejam legíveis e compatíveis com os recebidos do ministério da educação ou outras fontes relevantes. Dados incompatíveis significam perda de informações, mais trabalho, mais custo e mais tempo desperdiçado. O compartilhamento eficaz requer interoperabilidade (Boz & Simsek, 2022).

A maior parte da literatura foca na interoperabilidade dentro das organizações, mais especificamente entre agências governamentais. Contudo, o compartilhamento entre essas

agências ainda é um desafio globalmente. Os principais obstáculos listados¹³ incluem falta de recursos financeiros; ausência de catálogos ou padrões comuns; falta de protocolos claros; treinamento inadequado; desconfiança no processo; níveis variados de maturidade digital; engajamento institucional insuficiente; ou desconhecimento sobre bases existentes, o que leva à subutilização dessas bases (World Bank, 2018). Esses desafios são ainda maiores em países federativos devido à necessidade adicional de compartilhar informações entre diferentes níveis administrativos, o que leva a mais camadas de burocracia.

Esses problemas não ocorrem apenas nos governos: até mesmo na ONU enfrentam-se dificuldades semelhantes relacionadas à governança inadequada (Shamoug et al., 2020). Gasco (2012) propõe quatro níveis para interoperabilidade: i) intra-administrativa (dentro da mesma organização); ii) horizontal (entre organizações no mesmo nível); iii) vertical (entre diferentes níveis governamentais); iv) transfronteiriça (entre países). Muitas vezes isso representa níveis crescentes na dificuldade do compartilhamento. Dentro de uma instituição, o compartilhamento de dados é usualmente feito por meio de um sistema ou integração de dados. Entre instituições, acordos de cooperação são assinados para permitir esse compartilhamento.

A interoperabilidade de dados enfrenta desafios significativos devido à separação entre estruturas legais que regem a coleta de dados estatísticos e regulatórios, cada uma com regras rigorosas de confidencialidade. Esses dados têm propósitos distintos, o que complica sua integração e reutilização. Países com diferentes níveis de maturidade na coleta de dados enfrentam desafios variados, sendo que muitos desses precedem a digitalização (BIS, 2015).

Além disso, conjuntos de dados possuem utilidades distintas. Dados demográficos, indicadores socioeconômicos, informações de saúde e dados geográficos são cruciais para a prestação de serviços públicos. Por outro lado, dados parlamentares, sobre despesas públicas, orçamentos e estatísticas oficiais são mais relevantes para o engajamento cidadão (UN, 2013). A falta de interoperabilidade está relacionada a esses diferentes propósitos e à ausência de reutilização eficiente dos dados coletados, o que exige mudanças culturais para superar tais barreiras.

Uma estratégia eficiente de dados abertos poderia resolver muitos desses problemas. Dados abertos devem ser acessíveis por padrão, abrangentes, utilizáveis, comparáveis e interoperáveis¹⁴. Eles devem ser legíveis por máquinas¹⁵¹⁶, disponibilizados em formatos não proprietários e acompanhados por metadados detalhados e licenças abertas.

¹³ Para uma listagem completa, consulte CDO Council (2022), Texas Department of Information Resources (2023), Busby *et al.* (2020), DLUHC (2023) e Freeguard & Shepley (2023).

¹⁴ <https://opendatacharter.org/>. acesso em maio 2024.

¹⁵ <https://opendefinition.org/od/2.1/en/>. acesso em maio 2024.

¹⁶ <https://opendatawatch.com/>. Acesso em maio 2024.

O modelo “Cinco Estrelas dos Dados Abertos”¹⁷, proposto por Tim Berners-Lee, classifica iniciativas de dados abertos com base na qualidade técnica e na interoperabilidade dos dados. Ele inclui níveis como uso de identificadores únicos (URIs) e conexão com outros conjuntos de dados para permitir interpretações contextualizadas e respostas a questões complexas. Esse modelo destaca que a interoperabilidade é essencial para maximizar o valor dos dados abertos e promover sua integração em sistemas maiores.

Tabela 11 - Cinco estrelas dos dados abertos.

Classificação	Tipo de arquivo
*	PDF
**	Excel
***	CSV
****	RDF
*****	RDFs com URIs para reuso de dados integrados.

Fonte: elaboração própria com base em <https://5stardata.info/en/>.

A disponibilidade de dados abertos varia entre diferentes áreas e países. De acordo com a ONU (2020) e a OCDE (2023b), regiões como a Europa se destacam em todas as áreas devido à maior maturidade na coleta de dados. Entretanto, mesmo em países com menor disponibilidade geral de dados, conjuntos relacionados ao meio ambiente e geoespaciais tendem a ser mais acessíveis. Isso reflete a heterogeneidade no processo de abertura de dados, influenciado por fatores econômicos e prioridades temáticas.

3.3. Privacidade e soluções alternativas para compartilhamento

O problema do compartilhamento de dados e o desafio dos dados abertos podem ser vistos como interligados. A melhor solução para compartilhar informações seria por meio de estratégias de dados abertos, que promovem transparência e responsabilização. Contudo, isso exige tecnologias que protejam a privacidade, conhecidas como *Privacy-Enhancing Technologies* (PETs). Segundo a OCDE (2023a), PETs são ferramentas digitais que permitem o processamento e compartilhamento de dados enquanto protegem sua confidencialidade. Elas se dividem em quatro categorias principais:

¹⁷ <https://5stardata.info/en/>. Acesso em maio 2024.

- Obfuscação de dados: Alteração ou adição de ruído aos dados para prevenir identificação;
- Processamento de dados criptografados: Permite o uso de dados sem necessidade de descriptografá-los, garantindo maior segurança;
- Análises federadas e distribuídas: Executam tarefas analíticas sem acesso direto aos dados;
- Ferramentas de responsabilização de dados: Embora não sejam PETs propriamente ditas, aumentam o controle dos indivíduos sobre seus próprios dados.

Essas tecnologias são essenciais para viabilizar estratégias eficientes de dados abertos, especialmente em contextos onde privacidade e segurança são preocupações primordiais. Além disso, PETs promovem conformidade regulatória e confiança, permitindo que organizações compartilhem dados para fins benéficos sem comprometer informações sensíveis.

Embora as tecnologias de aprimoramento de privacidade ofereçam soluções robustas para proteger dados sensíveis, sua adoção ainda é limitada devido à falta de conhecimento sobre seus benefícios, custos e riscos associados à reidentificação. Isso leva à necessidade de uma solução alternativa: o compartilhamento de dados por meio de acordos de cooperação técnica baseados em bancos de dados interoperáveis e na disponibilização de dados agregados não identificáveis ou bases restritas anonimizadas em portais de dados abertos. Esses portais devem seguir princípios como legibilidade por máquinas, uso de formatos não proprietários, múltiplas opções de download, metadados detalhados e licenças abertas.

Essa abordagem alternativa enfrenta desafios técnicos e financeiros significativos:

- Análise caso a caso: Cada conjunto de dados exige avaliação específica para garantir anonimização e compatibilidade com os padrões de interoperabilidade;
- Ambientes seguros para acesso: Em alguns casos, é necessário criar salas seguras para acesso aos dados;
- Monitoramento constante dos riscos: É imprescindível acompanhar continuamente os riscos de reidentificação, especialmente em bases anonimizadas;
- Custos operacionais e técnicos: Incluem a manutenção da infraestrutura, limpeza dos dados e conformidade com regulamentações.

Além disso, o compartilhamento eficiente requer infraestrutura robusta e governança clara. A falta dessas condições pode levar a problemas como baixa interoperabilidade, perda de controle sobre os dados e desafios relacionados à segurança e privacidade¹⁸. Portanto, embora essa

¹⁸ <https://www.striped-giraffe.com/en/blog/data-sharing-challenges-and-opportunities/>. Acesso em abril 2025.

solução seja viável para promover transparência e reutilização dos dados, sua implementação demanda investimentos significativos em tecnologia, treinamento e monitoramento contínuo.

3.4. Compartilhamento de dados no Brasil

A interoperabilidade de dados e os desafios associados à integração de sistemas ganharam relevância no contexto brasileiro, especialmente após a promulgação de marcos legais como a Lei de Acesso à Informação (Lei nº 12.527/2011) e o Decreto nº 8.777/2016, que instituiu a Política de Dados Abertos (PDA). Esses instrumentos refletem esforços legislativos para modernizar a gestão pública, embora apresentem lacunas específicas na regulamentação do compartilhamento intersetorial. Como destacado por Boz & Simsek (2022), a reutilização de dados depende criticamente de estratégias de interoperabilidade, aspecto que assume contornos particulares no cenário educacional nacional.

A legislação brasileira concentra-se predominantemente em i) interoperabilidade intrasetorial, com ênfase na integração de sistemas governamentais, mediante iniciativas como a Estratégia de Governança Digital (Decreto nº 10.332/2020); e ii) transparência ativa, através de portais de dados abertos que seguem princípios de formatos não proprietários e licenciamento livre (World Bank, 2022). Entretanto, a ausência de normativas específicas para compartilhamento público-privado cria assimetrias informacionais, particularmente sensíveis no setor educacional. Conforme alertado por Chege & Wanjohi (2023), países com infraestrutura precária de registros administrativos tendem a supervalorizar instrumentos censitários, fenômeno observado no Brasil através da centralidade do Censo Escolar (CES) e Superior (CES).

A discussão sobre privacidade e soluções alternativas para o compartilhamento de dados se conecta diretamente aos desafios do compartilhamento e da interoperabilidade no contexto brasileiro, descritas na próxima seção, pois, embora tecnologias como as PETs ofereçam caminhos promissores para viabilizar dados abertos e proteger informações sensíveis, sua adoção efetiva depende de um ambiente regulatório, institucional e técnico que ainda está em consolidação no país. Assim, enquanto as soluções técnicas avançam, a implementação prática dessas estratégias esbarra em limitações legais, operacionais e culturais, tornando fundamental que as políticas nacionais de dados abertos e de interoperabilidade evoluam em paralelo ao desenvolvimento e à difusão dessas tecnologias, a fim de garantir tanto a segurança quanto a utilidade social dos dados educacionais.

3.4.1. O Inep e as principais bases de dados educacionais brasileiras

O esforço para coletar informações sobre a educação no Brasil remonta ao início do século XX. O primeiro *Anuário Estatístico Brasileiro*, de 1916, já apresentava dados sobre o ensino superior da época. Contudo, foi apenas 20 anos depois que essas informações começaram a ser coletadas de forma sistemática e anual (Inep, 2023b). A partir da década de 1990, o Inep tornou-se o órgão responsável por “organizar e manter o sistema de informações e estatísticas educacionais”. Esse processo é realizado de forma descentralizada e autodeclarada, abrangendo todas as instituições de ensino superior públicas e privadas, e com diversos propósitos – estatístico, regulatório, avaliação de qualidade e classificatório para a ingresso no ensino superior.

Inicialmente, os dados eram enviados por disquete, mas a partir dos anos 2000 o processo migrou para plataformas digitais com acesso restrito às instituições. Em 2010, o sistema censitário começou a importar informações de instituições de ensino e cursos do cadastro do Ministério da Educação (sistema e-Mec), além de tornar obrigatório o uso do CPF para docentes e estudantes. Essa mudança representou um passo inicial para a interoperabilidade com outras bases de dados (Inep, 2023b).

O Inep coleta anualmente dois censos educacionais – o Censo Escolar (CEB), referente à educação básica, e o Censo da Educação Superior (CES). O próprio Inep considera o Censo Escolar “o principal instrumento de coleta de informações da educação básica e a mais importante pesquisa estatística educacional brasileira”¹⁹. A previsão legal do Censo Escolar vem da Lei nº 9.394, de 20 de dezembro de 1996, que estabelece as diretrizes e bases da educação nacional. Em dois artigos – o artigo 5º, parágrafo 1º, inciso I e o artigo 9º, inciso V – é colocada a obrigação de se “recensear a população em idade escolar para o ensino fundamental, e os jovens e adultos que a ele não tiveram acesso” e de “coletar, analisar e disseminar informações sobre a educação”. Entretanto, foi somente a Portaria nº 316, de 4 de abril de 2007 que regulamentou o Censo Escolar e atribuiu ao Inep a competência de conduzir o processo.

Atualmente, sua execução é coordenada pelo Inep em parceria com as secretarias estaduais e municipais de educação, e a participação é obrigatória para todas as instituições de ensino da educação básica, incluindo a educação profissional e tecnológica e educação de jovens e adultos. Sua coleta é realizada de forma descentralizada com dados declaratórios pelas

¹⁹ <https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-escolar>. Acesso em novembro de 2024.

instituições de ensino e periodicamente o Inep realiza pesquisas de controle de qualidade, com a contratação de uma empresa externa. Nessas pesquisas, é feita a comparação entre os dados declarados os registros administrativos e acadêmicos verificados *in loco*, buscando minimizar possíveis discrepâncias nas informações coletadas, melhorar o processo de coleta e aumentar a qualidade dos dados²⁰.

O segundo censo é o Censo da Educação Superior, que abrange todos os estabelecimentos de ensino superior públicos e privados e adota como unidades de informação alunos, professores e instituições. Sua coleta também é descentralizada e declaratória. As unidades de informação incluem estudantes, professores e instituições, que têm seus dados preenchidos nominalmente em sistema próprio de coleta desses dados, os quais permitem, após sua consolidação, obter informações sobre qualificação docente, despesas institucionais e informações acadêmicas, como matrículas, vagas oferecidas, taxas de evasão e tempo médio de conclusão, dentre diversas outras informações.

Para a coleta do censo, cada instituição deve nomear um recenseador institucional, responsável pelo processo de coleta de dados na instituição. A metodologia de coleta tem sido atualizada a cada edição e a memória é catalogada nos documentos publicados no site do Inep²¹.

Além desses dois censos, foram utilizados também dados coletados no Exame Nacional do Ensino Médio (Enem). O Enem data de 1998, e o aumento de estudantes que fazem a prova ocorreu de forma gradativa ao longo dos anos, motivado principalmente pelo uso da prova em programas educacionais oficiais²². Atualmente tem três propósitos: avaliação, instrumento de ingresso no ensino superior e acesso a financiamento estudantil. Ele também já foi utilizado em processos de seleção de mobilidade internacional, como o Programa Ciências sem Fronteiras. Hoje em dia, os estudantes que realizam a prova podem utilizar a nota no Sistema de Seleção Unificada (Sisu), no Programa Universidade para Todos (ProUni) e no Fundo de Financiamento Estudantil (Fies) e em instituições estrangeiras com convênio com o Inep.

Atualmente, as provas contemplam quatro áreas de conhecimento: linguagens, códigos e suas tecnologias; ciências humanas e suas tecnologias; ciências da natureza e suas tecnologias; e

²⁰ <https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-escolar/outros-documentos/pesquisa-de-controle-de-qualidade> Acesso em novembro de 2024.

²¹ Disponíveis em: <https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-da-educacao-superior/outros-documentos/metodologia-de-coleta> Acesso em novembro de 2024.

²² <https://ces.ibge.gov.br/base-de-dados/metadados/inep/exame-nacional-do-ensino-medio-enem.html> Acesso em novembro de 2024.

matemática e suas tecnologias, totalizando 180 questões objetivas, além de um texto dissertativo-argumentativo²³.

Desde 2009, o cálculo da nota é feito utilizando a metodologia de teoria de resposta ao item (TRI), em que se busca mensurar a proficiência do participante em cada um dos conteúdos considerados. Isso significa que, pelo menos no caso da prova objetiva, notas de provas de anos diferentes seriam, em tese, comparáveis, como o próprio Inep coloca (“Entenda a sua nota no Enem: guia do participante”, 2021): “Uma das vantagens de se utilizar a TRI é que sua nota nos possibilita fazer comparações entre notas de diferentes provas da mesma área, mas os acertos de provas diferentes não são diretamente comparáveis.” Além das provas e das notas, o Inep disponibiliza também informações relacionadas ao questionário respondido pelos inscritos ao realizarem a prova²⁴.

3.4.2. O acesso aos dados identificados: a sala segura do Inep

Há duas versões dessas bases de dados – uma não identificada, disponível na página do Inep, e outra identificada, disponível apenas na sala segura da instituição, localizada em sua sede na cidade de Brasília, Distrito Federal²⁵.

Importante mencionar a intervenção da Autoridade Nacional de Proteção de Dados Pessoais (ANPD) que, por meio da Nota Técnica nº 46/2022/CGF/ANPD, solicitou que o Inep elaborasse um Relatório de Impacto à Proteção de Dados (Inep, 2022) para fins de adequação da divulgação dos microdados do Censo Escolar e do Enem à LGPD. A conformidade à lei levou à retirada temporária dos dados da página da instituição para adequação das bases às exigências legais. Deve-se enfatizar que o desafio é enorme, pois as bases são muito grandes, o que significa que o risco de reidentificação, mesmo após a anonimização, tende a ser elevado, já que são muitas variáveis²⁶.

A sala segura é oficialmente conhecida como Sedap, ou serviço de acesso a dados protegidos do Inep²⁷, e trata-se de uma política institucional consolidada, que tem como objetivo garantir

²³ <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enem> Acesso em novembro de 2024.

²⁴ <https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enem> Acesso em novembro de 2024.

²⁵ Por meio de convênios, o Inep recentemente inaugurou outros acessos à sala segura fora de sua sede.

²⁶ Segundo a nota de esclarecimento do Inep sobre o caso, “os resultados das análises promovidas pela UFMG apontam que, somente no Censo Escolar da Educação Básica 2019, o uso de três identificadores (mês, ano de nascimento e código da escola em que estuda) permite a identificação com probabilidade de acerto de até 29,64%. Se usados quatro identificadores, a chance de sucesso aumenta para 49,86% e, com o uso de todos os dez identificadores, o risco é elevado para 75,51%.” Disponível em: <https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enem> Acesso em novembro de 2024.

²⁷ Para mais informações, ver <https://www.gov.br/pt-br/servicos/usar-o-servico-de-acesso-a-dados-protegidos-do-inep> Acesso em novembro de 2024.

o acesso a dados estatísticos identificados, possibilitando importantes estudos e pesquisas. É necessário solicitar acesso previamente, apresentando, além de documentos pessoais, projeto de pesquisa a ser desenvolvido e documento de aceite emitido pela instituição de vínculo do pesquisador. Não é possível retirar os dados de lá, sendo possível a extração de resultados estatísticos previamente inspecionados pelos técnicos da instituição para garantir o não vazamento de informações.

Além dos dados, o Inep disponibiliza *softwares* estatísticos para uso dos pesquisadores e bases de dados auxiliares (sob requisição) para pesquisas que necessitem delas. Os dicionários de dados disponíveis publicamente na página do Inep referem-se somente aos microdados disponíveis também publicamente. Sugere-se a inclusão dos dicionários de dados disponíveis na sala segura para que os pesquisadores possam preparar melhor suas pesquisas de forma antecipada, reduzindo assim o tempo de utilização desse recurso e otimizando o processo como um todo. A atenção às tecnologias disponíveis é algo também muito importante, algo que deve ser motivo de atenção por parte dos órgãos que disponibilizam tal serviço, para diminuir o ônus aos usuários.

Todo o processo de tratamento para utilização das bases será descrito na seção seguinte.

3.5. Procedimento para limpeza e junção das bases de dados do CES, Enem e CEB

As três principais bases de dados consideradas neste estudo (CES, CEB e Enem) permitem o retrato longitudinal da educação brasileira desde a educação infantil até o final da graduação. Ainda que elas já sejam bastante úteis de forma separada, para algumas pesquisas, principalmente aquelas focadas no ingresso no ensino superior ou em fenômenos ocorridos no ensino superior influenciados por trajetórias anteriores, é importante que esses dados longitudinais sejam considerados de forma ampla. Logo, chegamos ao objetivo principal deste artigo – descrever o desafio de realizar essa junção e tratamento de dados para estudos de sucesso e evasão no ensino superior, utilizando indicadores da trajetória anterior do estudante como variáveis de controle.

A descrição feita aqui considera os dados disponíveis na sala segura do Inep, uma vez que a junção de bases, tomando indivíduos como unidade de observação, só é possível por meio de bases identificadas. Em outros níveis de análise menos desagregados, como, por exemplo, unidades federativas, é possível fazer a junção com dados não identificados disponíveis na página do Inep. Entretanto, chamamos a atenção que, dependendo do propósito da pesquisa, perde-se muito detalhamento de informação e enfatiza-se o ganho que a base identificada traz

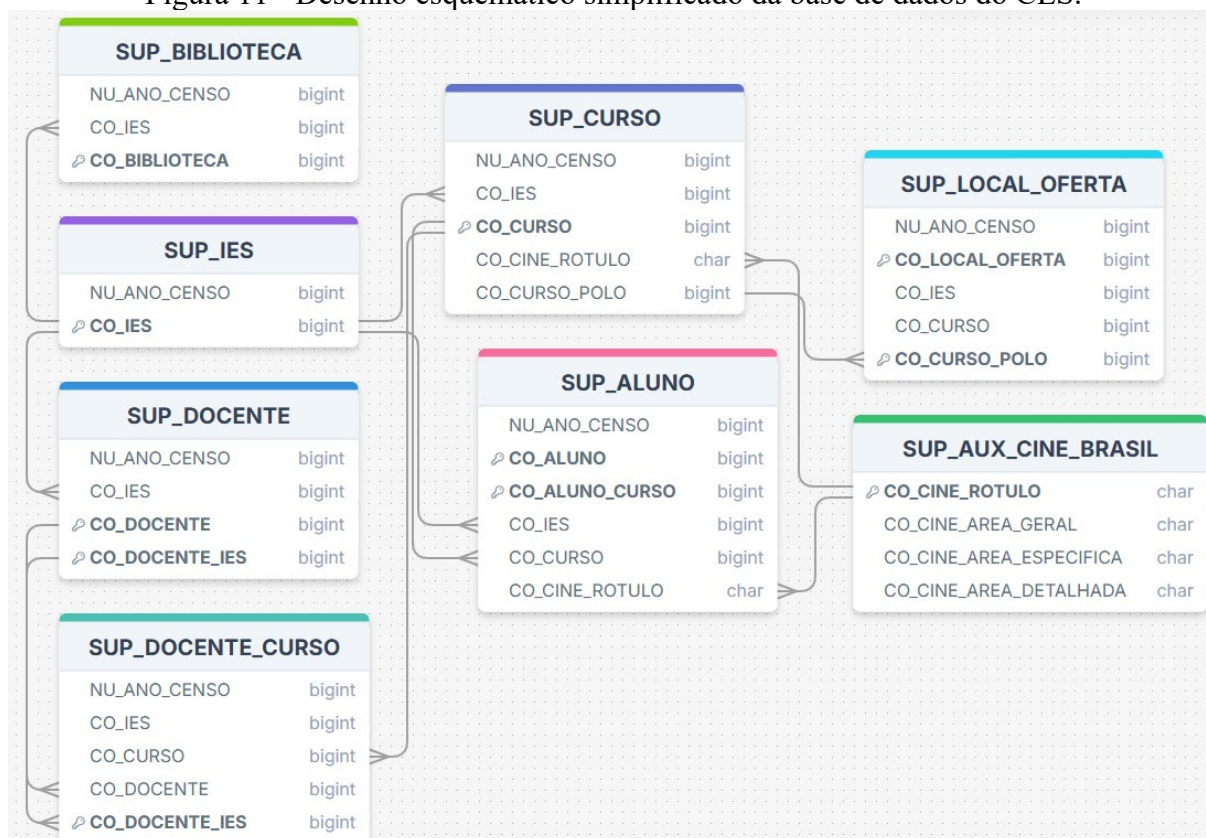
à análise. Inicialmente discutiremos as três bases de forma separada para depois falarmos de sua junção.

3.5.1. Censo da Educação Superior

O ponto de partida para o empilhamento do CES é considerar que a coleta do CPF só está disponível a partir do ano de 2009. Considerando todos os registros de 2009 a 2022, há um total de 153.360.577 observações. Entretanto, é importante observar que, no ano de 2009, a qualidade de dados ainda não era satisfatória, já que se tratava da primeira coleta por CPF (Silva; Cabral & Pacheco, 2016). Assim, em nosso esforço, preferimos não utilizar esse ano e somente considerar a série histórica a partir de 2010. Logo, retiram-se 6.982.018 observações, restando um total de 146.378.559 registros, referentes aos dados empilhados de 2010 a 2022.

A base de dados identificada do CES é estruturada em módulos, sendo quatro principais e quatro secundários, com diversas conexões entre si, conforme Figura 1. Os módulos principais têm os dados nominais dos alunos (SUP_IES) e dos docentes (SUP_DOCENTE), as características das instituições (SUP_IES) e dos cursos (SUP_CURSO). Os módulos secundários expandem detalhes mais específicos de alguns dos módulos principais. A base de cursos tem informações adicionais sobre os locais de oferta (SUP_LOCAL_OFERTA); a base dos docentes está conectada com os cursos a que cada docente tem vínculo (SUP_DOCENTE_CURSO); os dados das IES têm o detalhamento da sua estrutura de bibliotecas (SUP_BIBLIOTECA); e as bases de alunos e cursos podem ser ligadas à tabela auxiliar da classificação Cine Brasil (SUP_AUX_CINE_BRASIL).

Figura 11 - Desenho esquemático simplificado da base de dados do CES.



Fonte: elaboração própria usando [drawSQL](#).

Entendida a estrutura da base, o próximo passo é considerar a variável TP_SITUACAO, disponível na SUP_ALUNO, que indica a situação de vínculo do aluno ao seu curso, na instituição de ensino, em cada ano-base e período. Suas possíveis categorias estão dispostas no Quadro 5. Como o nosso foco é direcionar o tratamento para estudos de sucesso e evasão no ensino superior, utilizando as variáveis da trajetória anterior do estudante como controle, consideramos apenas alunos que finalizaram seu vínculo com a instituição de ensino superior, ou seja, aqueles que desvincularam do curso ou formaram (TP_SITUACAO = 4 ou 6). Tem-se 39.704.911 observações nessa situação.

Quadro 5 – descrição da variável TP_SITUACAO

Rótulo	Descrição
2	Cursando
3	Matrícula trancada
4	Desvinculado do curso
5	Transferência interna
6	Formado

7	Falecido
---	----------

Fonte: elaboração própria com base em Inep.

Em relação ao nível acadêmico, desconsideramos os cursos sequenciais de formação específica, pela irrisória representatividade frente ao total de registros (0,22%), mantendo apenas cursos em que $TP_NIVEL_ACADEMICO = 1$, ou seja, graduação, totalizando 39.616.802 registros até então.

Segundo Inep (2015, p. 2), ABI, ou Área Básica de Ingresso, “refere-se à situação em que uma única ‘entrada’ no curso possibilitará ao estudante, após a conclusão de um conjunto básico de disciplinas (denominado de ‘ciclo básico’ por algumas instituições de educação superior) a escolha de uma entre duas ou mais formações acadêmicas”. Esse é um tipo de arranjo comum em cursos de engenharia, licenciatura ou cursos que têm mais de uma especialização. Nós excluimos os cursos do tipo ABI pois a saída de um curso ABI no sistema brasileiro não significa a finalização do curso superior do estudante e sim o ingresso sequencial na segunda parte de seu curso de graduação. Para excluir os registros referentes às ABIs, consideramos a variável CO_CINE_ROTULO , que traz os códigos de acordo com a CINE Brasil, e excluimos as observações de código de curso iguais aos dispostos no Quadro 6 (61.876 observações). A base resultante ficou com 39.554.926 observações.

Quadro 6 - Cursos ABI desconsiderados

Código	Rótulo Cine Brasil
0011A01	ABI Educação
0011A02	ABI Artes e humanidades
0011A03	ABI Ciências sociais, comunicação e informação
0011A04	ABI Negócios, administração e direito
0011A05	ABI Ciências naturais, matemática e estatística
0011A06	ABI Computação e Tecnologias da Informação e Comunicação (TIC)
0011A07	ABI Engenharia, produção e construção
0011A08	ABI Agricultura, silvicultura, pesca e veterinária
0011A09	ABI Saúde e bem-estar
0011A10	ABI Serviços

Fonte: elaboração própria com base na classificação de cursos CINE Brasil.

Nossa série histórica abrange os anos-base de 2010 a 2022. Entretanto, pela forma como os dados são registrados no CES e o fato de termos considerado somente aqueles estudantes com vínculos finalizados, os estudantes do ano-base 2010, por exemplo, que satisfazem as condições de $TP_SITUACAO$ 4 ou 6 (desvinculados ou formados), são estudantes que ingressaram em anos anteriores. Aqui, tomamos uma decisão metodológica e consideramos que, na média, os

curso de graduação brasileiros têm duração aproximada de 5 anos. Logo, excluímos todas as observações em que o ingresso no ensino superior ocorreu antes do ano de 2005 (359.854 observações). Dessa forma, ficamos com 39.195.072 observações.

Dessas 39.195.072 observações, identificamos 29.488.269 registros únicos, ou seja, diversos alunos têm mais de um registro de desvínculo ou conclusão no período analisado. É importante ressaltar que, apesar do CPF ser coletado a partir de 2009, ele não é prontamente disponibilizado mesmo na sala segura do Inep, pois a variável que identifica os indivíduos na base do CES é a CO_ALUNO, um identificador único usado no âmbito do Inep. Para identificar as movimentações dos alunos, é necessário utilizar outro identificador, o CO_ALUNO_CURSO, variável que registra cada vínculo dos estudantes em cada curso a qual eles tenham sido vinculados. Então se o aluno tiver, por exemplo, evadido do curso A e iniciado o curso B durante o período que estamos considerando, é esperado que ele apareça duas vezes em nossa base de dados. O Quadro 7 ilustra alguns exemplos adicionais da diferença das duas variáveis.

Quadro 7 – Exemplos da diferença entre as variáveis CO_ALUNO e CO_ALUNO_CURSO

	Situação	CO_ALUNO	CO_ALUNO_CURSO
Aluno 1	Ingressou e finalizou o curso A apenas	Aparece uma vez	Aparece uma vez
Aluno 2	Ingressou e evadiu do curso A apenas	Aparece uma vez	Aparece uma vez
Aluno 3	Ingressou e finalizou o curso A. Logo depois, iniciou e finalizou o curso B	Aparece uma vez	Aparece duas vezes
Aluno 4	Ingressou e evadiu o curso A. Logo depois, iniciou e finalizou o curso B	Aparece uma vez	Aparece duas vezes
Aluno 5	Ingressou e evadiu o curso A. Logo depois, iniciou e ainda não finalizou o curso B	Aparece uma vez	Aparece uma vez
Aluno 6	Ingressou e evadiu o curso A. Logo depois, iniciou e evadiu o curso B	Aparece uma vez	Aparece duas vezes

Fonte: elaboração própria.

Devido à falta do CPF como variável na base do CES, foi preciso solicitar ao Inep uma base adicional para fazer a junção do CES com as demais bases de interesse, o CEB e o Enem. A variável para ligação seria a CO_ALUNO com a NU_CPF_MASC dessa nova base. Entretanto, ao fazer a junção dessas duas bases, observamos que 569.943 registros (1,5% do total) não tinham CPF registrado, o que impediu a junção posterior com o censo escolar e o Enem. No entanto, considerando que a base do CES dispõe da maioria das informações de interesse,

caracterizando-se como a principal base da análise, os registros sem CPF foram mantidos, e a base final do CES continuou em 39.195.072 registros.

Por fim, ao contar a quantidade de registros com CPF únicos, chegamos ao valor de 25.362.498, dos quais 16.775.227 (66,1%) aparecem uma única vez em toda a base, considerando apenas as observações com CPF preenchido. Isso quer dizer que a maioria dos alunos que passou pelo ensino superior, nas edições de 2010 a 2022 do CES, teve apenas um registro em um curso de alguma IES, com o desfecho de evasão ou conclusão.

3.5.1.1. Tratamentos específicos de variáveis do Censo da Educação Superior

Diversas variáveis do CES exigiram um tratamento específico e/ou acesso a bases auxiliares, para agregar mais informações e enriquecer a base final.

O primeiro tratamento feito diz respeito às variáveis de apoio social, atividade extracurricular, financiamento e mobilidade acadêmica, conforme Quadro 8. Devido à decisão metodológica de filtrar os registros de conclusão e desvinculação, essas informações se perderiam caso fossem consideradas apenas no momento final do vínculo do aluno. Para evitar essa limitação, optou-se por realizar a contagem acumulada das ocorrências dessas variáveis ao longo do tempo. Essa abordagem permitiu capturar o efeito longitudinal dos apoios recebidos pelos estudantes durante sua trajetória acadêmica, utilizando o identificador único CO_ALUNO, semelhante ao CPF, como chave para o acompanhamento individual.

Essa estratégia possibilitou uma análise mais detalhada e robusta, permitindo observar não apenas a presença ou ausência de apoios no momento final do vínculo, mas sim como esses apoios se acumularam ao longo dos anos, durante o percurso acadêmico de cada aluno, oferecendo uma visão mais ampla sobre o impacto cumulativo dos apoios no sucesso acadêmico dos estudantes, conhecimento já consolidado na literatura do tema (Araújo; Mariano & Oliveira, 2021; Colpo; Primo & Aguiar, 2024; Gama, 2018; Lima & Fagundes, 2020; Nierotka; Salata & Klitzke Martins, 2023; Paula, 2021).

Quadro 8 - Variáveis de apoio social, atividade extracurricular, financiamento e mobilidade acadêmica tratadas

Categoria	Variável	Descrição
Apoio social	IN_APOIO_SOCIAL	Informa se o aluno recebe algum tipo de apoio social na forma de moradia, transporte, alimentação, material didático e bolsas (trabalho ou permanência)
	IN_APOIO_ALIMENTACAO	Informa se o aluno recebe apoio alimentação
	IN_APOIO_BOLSA_PERMANENCIA	Informa se o aluno recebe auxílio financeiro destinado a alunos em situação de vulnerabilidade socioeconômica ou pertencente a grupos étnicos específicos

	IN_APOIO_BOLSA_TRABALHO	Informa se o aluno recebe remuneração referente a trabalhos prestados nas dependências da IES ou unidades vinculadas
	IN_APOIO_MATERIAL_DIDATICO	Informa se o aluno recebe apoio para aquisição de material didático
	IN_APOIO_MORADIA	Informa se o aluno recebe apoio moradia
	IN_APOIO_TRANSPORTE	Informa se o aluno recebe apoio para transporte até a IES
Atividade extracurricular	IN_ATIVIDADE_EXTRACURRICULAR	Informa se o aluno participa de algum tipo de atividade extracurricular (estágio não obrigatório, extensão, monitoria e pesquisa)
	IN_BOLSA*	Informa se o aluno recebe bolsa/remuneração por participar de alguma atividade extracurricular
Financiamento	IN_FINANCIAMENTO_ESTUDANTIL	Informa se o aluno utiliza financiamento estudantil
	IN_FIN_REEMB_FIES	Informa se o aluno utiliza o Fundo de Financiamento Estudantil (Fies) como forma de financiamento estudantil reembolsável
	IN_FIN_PROUNI*	Informa se o aluno é bolsista integral ou parcial do Programa Universidade para Todos (Prouni), tipo de financiamento estudantil não reembolsável
Mobilidade acadêmica	IN_MOBILIDADE_ACADEMICA	Informa se o aluno está regularmente matriculado em curso de graduação, que se vincula temporariamente a outra instituição, sendo ela nacional ou internacional

Fonte: elaboração própria com base nas variáveis do CES.

Nota: * variáveis construídas para o estudo.

A informação de turno do curso (TP_TURN0) dos alunos foi, também, tratada, para dicotomizar a variável em apenas duas categorias: integral, que incluiu os turnos integral, matutino e vespertino; e noturno. O mesmo foi feito com a nacionalidade (TP_NACIONALIDADE), agrupando nas categorias brasileira ou estrangeira. A raça/cor autodeclarada (TP_COR_RACA) foi consolidada em PPI (pretos, pardos e indígenas) e não PPI (brancos e amarelos), em uma nova variável. Por fim, todas as variáveis que tinham a categoria “sem informação” ou similar tiveram esses registros específicos tratados como se realmente não existisse a informação (NA ou *not available*), finalizando os ajustes do módulo de alunos.

No módulo de cursos, foi necessário ajustar a variável de carga horária mínima do curso (NU_CARGA_HORARIA), pois existiam diversos registros de carga horária igual a zero e alguns registros de carga horária igual a 606.667, indicando falhas de preenchimento do CES.

De acordo com a Resolução CNE/CES nº 2/2007²⁸, os cursos de graduação devem ter entre 2.400 e 7.200 horas de carga horária mínima, enquanto que os tecnólogos variam de 1.600 a 2.400 horas, conforme disposto no Catálogo Nacional de Cursos Superiores de Tecnologia²⁹. Verificou-se que 80.912 alunos estavam em cursos com registro de carga horária menor que 1.600 horas, e 329.175 estudantes tinham vínculo com cursos com carga horária mínima maior que 7.200 horas. Esses casos foram tratados substituindo-os pela carga horária mínima média de cada curso, de acordo com a classificação Cine (CO_CINE_ROTULO), considerando o intervalo usual de carga horária (1.600 a 7.200 horas mínimas).

E do módulo docente extraiu-se apenas o total de docentes por IES e ano-base do CES, por tratar-se de um determinante da evasão. Já o número total de técnico-administrativos em educação (TAEs) está facilmente disponível no módulo IES, logo, a princípio, ambas variáveis poderiam estar prontamente disponíveis, facilitando o trabalho dos pesquisadores. Os demais módulos não precisaram de tratamentos de consistência adicionais.

3.5.2. Enem

A base de dados do Enem, mesmo no ambiente da sala segura, veio com o CPF mascarado, impossibilitando a integração com as demais bases de dados. Foi necessário fazer um requerimento especial ao Inep, para fornecimento de uma chave que permitisse a integração. Questiona-se a necessidade disso, visto o ambiente seguro em que estávamos. Além disso, os anos-base 2015 a 2018 não estavam disponíveis no ambiente seguro, logo foi necessário, mais uma vez, fazer um requerimento especial para conseguir os anos faltantes. E quando a disponibilização foi feita, verificou-se que a variável com o identificar único da escola de educação básica cursada (CO_ESCOLA) não estava presente, causando ainda mais atrasos para a continuação da pesquisa. No entanto, após a inserção dessa variável, verificou-se que a falta de preenchimento desse indicador é tão grande que limita seu uso para muitas pesquisas – apenas 13.044.321 registros, ou 33,3% do total considerado.

Além disso, conforme já é documentado na literatura, o questionário socioeconômico do Enem sofreu muitas modificações ao longo do ano, o que gerou falta de consistência entre as variáveis socioeconômicas ao longo dos vários anos da base de dados. O processo de tratamento e harmonização de cada variável será descrito a seguir. Deve-se destacar que se utilizou o ano-base 2009 do Enem, apesar do primeiro ano-base do CES ser o 2010 – isso foi feito pois os alunos que fizeram Enem em 2009 podem ter entrado na edição de 2010 em diante do CES.

²⁸ http://portal.mec.gov.br/cne/arquivos/pdf/2007/rces002_07.pdf. Acesso em março de 2025.

²⁹ <https://cncst.mec.gov.br/>. Acesso em março de 2025.

3.5.2.1. Número de moradores

A variável NU_MORADORES, que representa o número de pessoas residentes no domicílio do participante, apresentou significativas alterações em sua nomenclatura e codificação ao longo das edições do ENEM entre 2009 e 2022, exigindo um cuidadoso processo de harmonização para viabilizar análises em série histórica. Nas edições iniciais, observa-se maior inconsistência na forma de coleta deste dado. Em 2009, a informação era obtida através da questão Q15 (“Quantidade de pessoas que moram em sua casa”), enquanto em 2010 passou a ser coletada via Q01 “Quantas pessoas moram com você? (incluindo filhos, irmãos, parentes e amigos)”. A diferença na formulação das perguntas é sutil, mas importante, pois a primeira possivelmente incluía o respondente na contagem, enquanto a segunda especificamente perguntava sobre “outras pessoas” além do respondente.

A partir de 2011, houve uma tentativa de padronização com a utilização da questão Q01, mas observa-se que em 2012 ocorreu nova alteração para Q04 “Quantas pessoas moram em sua casa (incluindo você)?”, explicitando claramente a inclusão do respondente na contagem total. Nos anos de 2013 e 2014, a codificação mudou para Q004, mantendo a mesma formulação da pergunta, mas alterando apenas o identificador da variável. Entre 2015 e 2018, o INEP estabeleceu um padrão diferente com o identificador QUESTAO_05, que posteriormente foi substituído por Q005 a partir de 2019, mantendo-se estável até 2022. Esta sequência de mudanças reflete a evolução do questionário socioeconômico do ENEM e os desafios de manter consistência ao longo do tempo.

Para harmonizar esta variável e permitir análises em série histórica, foi necessário criar um mapeamento sistemático que considerasse não apenas as diferentes codificações (Q15, Q01, Q04, Q004, QUESTAO_05, Q005), mas também as potenciais diferenças conceituais na formulação das perguntas. As edições iniciais (2009-2010) exigiram um tratamento mais complexo para adequar os valores, enquanto nas edições posteriores foi possível realizar um mapeamento direto.

3.5.2.2. Escolaridade do pai e da mãe

As variáveis referentes à escolaridade dos pais ou responsáveis (ESCOLARIDADE_PAI e ESCOLARIDADE_MAE) constituem importantes indicadores socioeconômicos nos microdados do ENEM, apresentando transformações significativas em sua nomenclatura, codificação e categorização ao longo das edições entre 2009 e 2022.

No período analisado, observa-se uma evolução nas designações dos questionários socioeconômicos. Em 2009, as informações sobre escolaridade parental eram coletadas através

das questões Q17 ("Até quando seu pai estudou") e Q18 ("Até quando sua mãe estudou"). Em 2010, houve uma modificação para Q02 ("Qual é o nível de escolaridade do seu pai?") e Q03 ("Qual é o nível de escolaridade da sua mãe?"), com sutis diferenças na formulação que podem influenciar as respostas dos participantes. Em 2011, retornou-se a uma formulação similar à de 2009 com Q02 e Q03, mantendo-se estável em 2012 com Q01 e Q02, respectivamente, evidenciando uma tentativa de padronização. Entre 2013 e 2014, a codificação permaneceu como Q001 e Q002, enquanto no período de 2015 a 2018, o INEP adotou a nomenclatura QUESTAO_01 e QUESTAO_02. A partir de 2019 até 2022, estabeleceu-se o padrão Q001 e Q002, demonstrando uma possível estabilização na forma de categorização destas variáveis. Para harmonização dessas variáveis em uma série histórica, foi necessário implementar um mapeamento consistente que respeitasse as diferenças conceituais entre os anos. A transformação envolveu:

- Padronização da nomenclatura das variáveis (ESCOLARIDADE_PAIS e ESCOLARIDADE_MAE);
- Uniformização das categorias de resposta nos três níveis educacionais: até o ensino fundamental completo (EF), até o ensino médio completo (EM) e ensino superior em diante (ES);
- Tratamento adequado para respostas "Não sei", "Sem resposta" ou semelhantes, convertendo-as sistematicamente para valores ausentes (NA).

3.5.2.3. Ocupação do pai e da mãe

Em 2009, as ocupações dos pais ou responsáveis eram registradas nas questões Q19 ("Área que o pai trabalha ou trabalhou, na maior parte da vida") e Q20 ("Área que sua mãe trabalha ou trabalhou, na maior parte da vida"). As respostas eram categorizadas em dez grupos que indicavam setores de atividade econômica ou ocupacional. As edições de 2010 a 2014 não registraram esses dados, e em 2015 a questão voltou com as nomenclaturas QUESTAO_03 e QUESTAO_04, mantendo-se assim até 2018, com novas categorias baseadas em cinco grupos de ocupação:

- **Grupo 1:** Lavradora, agricultora sem empregados, boia-fria, criadora de animais (gado, porcos, galinhas, ovelhas, cavalos etc.), apicultora, pescadora, lenhadora, seringueira, extrativista;
- **Grupo 2:** Diarista, empregada doméstica, cuidadora de idosos, babá, cozinheira (em casas particulares), motorista particular, jardineira, faxineira de empresas e prédios,

vigilante, porteira, carteira, office-boy, vendedora, caixa, atendente de loja, auxiliar administrativa, recepcionista, servente de pedreiro, repositora de mercadoria;

- **Grupo 3:** Padeira, cozinheira industrial ou em restaurantes, sapateira, costureira, joalheira, torneira mecânica, operadora de máquinas, soldadora, operária de fábrica, trabalhadora da mineração, pedreira, pintora, eletricista, encanadora, motorista, caminhoneira, taxista;
- **Grupo 4:** Professora (de ensino fundamental ou médio, idioma, música, artes etc.), técnica (de enfermagem, contabilidade, eletrônica etc.), policial, militar de baixa patente (soldado, cabo, sargento), corretora de imóveis, supervisora, gerente, mestre de obras, pastora, microempresária (proprietária de empresa com menos de 10 empregados), pequena comerciante, pequena proprietária de terras, trabalhadora autônoma ou por conta própria;
- **Grupo 5:** Médica, engenheira, dentista, psicóloga, economista, advogada, juíza, promotora, defensora, delegada, tenente, capitã, coronel, professora universitária, diretora em empresas públicas ou privadas, política, proprietária de empresas com mais de 10 empregados.

Em 2019, as variáveis mudaram novamente de nomenclatura (Q003 e Q004), mas mantiveram-se estáveis com os mesmos identificadores até 2022, indicando uma consolidação na forma de coleta. Para padronizar as variáveis ao longo dos anos, as diferentes nomenclaturas foram unificadas, com o mapeamento para os cinco grupos padronizados de resposta, e as respostas “Não sei” e semelhantes foram sistematicamente convertidas para valores ausentes (NA) para evitar interpretações equivocadas nas análises.

3.5.2.4. Renda familiar

A harmonização da variável RENDA_FAM nos microdados do ENEM (2009-2022) envolveu um processo técnico-metodológico complexo para garantir comparabilidade temporal, considerando mudanças nos intervalos de renda e a necessidade de ajuste monetário. Inicialmente, identificou-se que as categorias de renda familiar sofreram alterações significativas ao longo dos anos, tanto na formulação das questões (Q21, Q04, Q006) quanto nos valores dos intervalos monetários. Para unificar esses dados, adotou-se o ponto médio dos intervalos declarados em cada ano.

Posteriormente, implementou-se um procedimento de deflação para padronizar os valores em relação ao último ano disponível (base 2022). Utilizou-se o fator $\frac{1212}{\text{valor base original}}$, onde 1212, ou R\$ 1.212, representa o salário mínimo vigente no último ano da série (2022). Esse ajuste

permitiu converter todos os valores históricos para o poder de compra equivalente ao ano-base, eliminando distorções causadas pela inflação e mudanças reais nos patamares de renda.

Além disso, observou-se que o número de categorias de renda expandiu ao longo dos anos – de 7 opções em 2009 para 17 em 2012 –, exigindo mapeamentos específicos para manter a coerência dos intervalos. O tratamento de valores ausentes ou não respondidos foi padronizado para NA, preservando a integridade estatística. Este duplo processo (cálculo de pontos médios e deflação) permitiu criar uma métrica unificada de renda familiar real, essencial para análises longitudinais sobre desigualdade socioeconômica e desempenho educacional no Brasil.

3.5.2.5. Tipos de escolas de ensino médio cursadas

As variáveis ESCOLA_EM_QENEM e ESCOLA_EM_QENEM_BOLSA, presentes apenas nas edições do ENEM de 2015 e 2016, fornecem informações sobre o tipo de escola de ensino médio frequentada pelos participantes e se houve ou não a concessão de bolsa de estudos integral para aqueles que cursaram escolas particulares.

A variável ESCOLA_EM_QENEM classifica as escolas em oito categorias, as quais foram consolidadas em três categorias principais: “Somente ou principalmente pública”, “Somente ou principalmente particular” e “Somente ou principalmente indígena ou quilombola”. Já a outra variável, ESCOLA_EM_QENEM_BOLSA, detalha a situação dos estudantes que cursaram escolas particulares, diferenciando entre aqueles que receberam bolsa integral e os que não receberam. O tratamento dessas variáveis envolveu a padronização das respostas categóricas, convertendo-as em rótulos consistentes para facilitar a análise. A harmonização também incluiu o tratamento de valores ausentes (NA) para casos em que os participantes não responderam ou forneceram informações inconsistentes.

A curta duração dessas variáveis na base do ENEM reflete mudanças nas prioridades do questionário socioeconômico ao longo do tempo. Embora úteis para estudos específicos sobre desigualdades no acesso à educação e impacto da rede escolar no desempenho acadêmico, sua ausência em edições posteriores limita análises comparativas mais amplas. Ainda assim, durante os anos de 2015 e 2016, essas variáveis oferecem um panorama valioso sobre as condições educacionais dos participantes, destacando as diferenças entre redes públicas e privadas e o possível papel das bolsas integrais na inclusão de estudantes em escolas particulares.

3.5.2.6. Inconsistências entre versões do Enem

Como um mesmo aluno poderia ter participado do ENEM em diferentes anos, informando dados distintos a cada edição, foi necessário criar uma estratégia de harmonização que

garantissem a consistência e a representatividade dos dados ao longo da série histórica. O objetivo era consolidar as variáveis de cada aluno, priorizando os registros mais recentes e preenchendo lacunas nos anos em que as informações estavam ausentes.

Para alcançar esse objetivo, o processo começou com a desagregação das variáveis em arquivos separados, reduzindo a complexidade computacional e permitindo o tratamento individualizado de cada variável. Por exemplo, variáveis como NU_MORADORES, ESCOLARIDADE_PAI, ESCOLARIDADE_MAE, OCUPACAO_PAI, OCUPACAO_MAE e RENDA_FAM foram extraídas de cada ano e organizadas em arquivos específicos. Em seguida, foi implementado um sistema de hierarquia temporal para selecionar o registro mais recente de cada aluno (identificado pelo CPF), assumindo que as informações mais recentes seriam mais precisas ou mais representativas.

Uma etapa crucial foi o preenchimento retroativo das lacunas. Quando um aluno não fornecia determinada informação em um ano específico, mas havia informado em anos anteriores ou posteriores, o valor foi propagado para os anos faltantes. Por exemplo, se um participante informou NU_MORADORES = 5 em 2015 e deixou essa variável em branco em 2016, o valor de 2015 foi replicado para preencher a lacuna de 2016, garantindo que todos os anos tivessem valores consistentes para as variáveis analisadas.

O mesmo foi feito com a RENDA_FAM, atentando para a necessidade de deflação. Variáveis categóricas, como ESCOLARIDADE_PAI e ESCOLARIDADE_MAE, já estavam padronizadas em três níveis principais, então foi feita a priorização do maior nível educacional reportado pelo aluno ao longo dos anos. Da mesma forma, as ocupações dos pais, previamente agrupadas em cinco grandes categorias (Grupos 1 a 5), e os registros inconsistentes ou ausentes foram tratados para garantir coerência.

Além dos ganhos de consistência, foi possível melhorar a quantidade de informação disponível. Na variável categórica de raça/cor autodeclarada (TP_COR_RACA), implementou-se uma hierarquia de fontes: dados do Enem substituíram categorias sem informação (0 ou 9) do CES, aumentando a taxa de preenchimento de 62,5%, considerando apenas CES, para 77,7%, considerando CES e Enem conjuntamente.

A variável CO_ESCOLA, que é o número único gerado pelo Inep como identificador da escola de ensino médio, foi também adicionada à base do CES, para ser utilizada posteriormente na modelagem estatística – nesse primeiro momento, apenas 13.044.321 (33,3%) dos registros tinham essa informação preenchida, isto é, dos estudantes com vínculo finalizada na educação superior, apenas 1/3 possuía informação de qual escola secundária finalizaram o ensino médio.

A integração final das variáveis socioeconômicas do Enem na base do CES seguiu essa abordagem metodológica rigorosa para garantir consistência e confiabilidade nos cruzamentos de dados. Utilizando o CPF mascarado (NU_CPF_MASC) como chave primária, o processo precisou, também, ser otimizado no desempenho computacional, para o tratamento dos quase 40 milhões de registros. Para cada variável realizou-se um *merge* específico, preservando todos os registros do CES e atualizando valores ausentes com informações do Enem, quando disponíveis. A próxima seção discute a integração das informações do CEB.

3.5.3. Censo da Educação Básica

Os dados provenientes do Censo da Educação Básica são relativos aos anos de referência 2007 a 2022, totalizando quase 800 milhões de registros, e foram utilizados para melhorar a qualidade informacional de algumas variáveis. Para a integração com o CES, foram necessárias algumas etapas de processamento. A primeira foi filtrar apenas os concluintes, utilizando a variável IN_CONCLUINTE, disponível apenas no âmbito da sala segura do Inep, resultando em 84.149.324 ocorrências.

Devido ao fato do CES também não possuir, entre suas variáveis, o CPF do aluno, foi preciso solicitar, mais uma vez, outra base que relacionasse o identificador disponível (CO_PESSOA_FISICA) ao CPF – tal junção encontrou 67.989.759 (80,8%) correspondências, causando possível perda de informação. Essa base foi então cruzada com os CPFs dos alunos do CES, obtendo 22.069.703 observações, ou seja, 32,5% dos quase 68 milhões. Por fim, foram mantidos apenas os registros mais recentes, finalizando em 12.910.740 registros, correspondentes a 32,9% do total da base final.

Com relação ao ganho informacional, os resultados estão dispostos no Quadro 9. A variável CO_ESCOLA, que apresentava 33,3% de dados preenchidos, após a combinação com o CEB elevou a cobertura para 48,9%, representando um ganho substancial, apesar de indicar, ainda assim, que mais da metade dos registros não têm a informação da escola de ensino médio dos estudantes. A variável de raça/cor teve também um ganho expressivo de informação, passando de 63,3% a 78,7% de preenchimento. Os indicadores de pessoa com deficiência (IN_DEFICIENCIA) e tipo de escola de conclusão do ensino médio (TP_ESCOLA_CONCLUSAO_ENS_MEDIO), se pública ou privada, tiveram melhorias menos expressivas nesse quesito. A próxima seção trata da integração das outras bases utilizadas no estudo.

Quadro 9 - Percentual de informação disponível em cada variável, ao longo do processo de junção das bases do CES, Enem e CEB

Variável	CES	CES + Enem	CES + Enem + CEB
Código identificador único da escola (CO_ESCOLA)	Indisponível	33,3%	48,9%
Cor/raça autodeclarada (TP_COR_RACA)	63,3%	76,8%	78,7%
Pessoa com deficiência - PcD (IN_DEFICIENCIA)	0,5%	Indisponível	0,6%
Escola de conclusão do ensino médio, pública ou privada (TP_ESCOLA_CONCLUSAO_ENS_MEDIO)	84,1%	Indisponível	87,3%

Fonte: elaboração própria com base na integração do CES, Enem e CEB.

3.5.4. Bases externas

A integração de variáveis externas seguiu o mesmo processo sistemático para garantir consistência e relevância analítica. O Indicador de Nível Socioeconômico (INSE) foi incorporado e consolidou dados históricos disponíveis (2011-2022) em dois níveis: escolar e municipal. Para escolas, os microdados originais do Inep foram padronizados em sete níveis (I a VII) e expandidos temporalmente para cobrir anos sem dados específicos. No nível municipal, calculou-se a média ponderada pelo número de alunos. Esses dados foram posteriormente vinculados à base principal usando CO_ESCOLA e CO_MUNICIPIO, com atualização retroativa para garantir a completude.

O indicador de Percentual de Docentes com Pós-Graduação stricto sensu foi uma das bases externas utilizadas. A variável PERC_DOCENTES_IES, o percentual de docentes com pós-graduação, foi tratada e agregada por IES e município, considerando a categoria administrativa, organização acadêmica e área CINE. Para as IES, utilizou-se o valor direto do arquivo de docentes, enquanto que para municípios, aplicou-se uma média ponderada pela quantidade de alunos em cada curso. Essa abordagem permitiu capturar tanto a qualidade docente institucional quanto regional.

Indicadores de qualidade como o IGC (Índice Geral de Cursos) e o CPC (Conceito Preliminar de Curso) foram integrados a partir de bases públicas do Inep. O IGC, contínuo e categorizado em faixas (1-5), foi vinculado via CO_IES e NU_ANO_CENSO, enquanto o CPC foi associado a cursos específicos usando CO_IES e CO_CURSO. Para o Programa REUNI, criou-se uma variável binária (REUNI = 1) identificando IES participantes entre 2008 e 2022, com base em relatório oficial³⁰ divulgado.

³⁰ [Reuni 2008 - Relatório de Primeiro Ano](#). Acesso em janeiro de 2024.

O Índice de Desenvolvimento Humano (IDH) foi incorporado em nível estadual usando dados anuais de 2010 a 2021, enquanto que a inclusão municipal, ou seja, do Índice de Desenvolvimento Humano Municipal (IDHM), limitou-se ao ano-base de 2010, devido à indisponibilidade de dados mais recentes. Dessa forma, essa hierarquia (UF > município) teve que priorizar a cobertura temporal em detrimento da granularidade espacial. A seção seguinte conclui os procedimentos de tratamento e manipulação das bases.

3.5.5. A base consolidada

Após os procedimentos descritos, de junção das bases do CES, Enem, CEB e demais informações disponíveis em bases externas, a base final continuou com 39.195.072 registros. Entretanto, observamos que apenas 4.701.156, ou 12% do total, apresentam informações completas em todos os indicadores, revelando um problema estrutural de qualidade das bases de dados de educação brasileiras. Essa fragmentação reflete problemas sistêmicos documentados na literatura, como a ausência de protocolos unificados de coleta e a heterogeneidade temporal nas variáveis socioeconômicas e institucionais (BARROS et al., 2023).

A interoperabilidade limitada entre sistemas – evidenciada pela necessidade de tratamentos complexos para harmonizar códigos de escolas e categorias de renda familiar, por exemplo – reproduz padrões identificados em estudos sobre governança de dados educacionais, onde grande parte das inconsistências derivam da falta de padrões semânticos entre as bases (Barros et al., 2023). Para contornar essas limitações na análise da evasão definitiva do sistema, adotou-se uma abordagem de amostragem estratificada em dois estágios, procedimento descrito no capítulo 2 (2.1.4. Amostragem). Essa estratégia permitiu isolar vieses decorrentes da falta de informação, ainda que reforce a urgência de políticas públicas para padronização de metadados e implementação de frameworks de interoperabilidade, como proposto no Plano de Dados Abertos³¹.

3.6. Considerações

A experiência de integração das bases do Enem, Censo da Educação Básica e Censo da Educação Superior revelou desafios estruturais que transcendem questões técnicas, refletindo lacunas institucionais e culturais na gestão de dados educacionais. A harmonização de variáveis aparentemente simples, como número de moradores ou cor/raça autodeclarada, exigiu esforços metodológicos complexos devido à instabilidade temporal dos questionários e à ausência de padrões semânticos entre sistemas. Estudos como o de Senkevics *et al.* (2016) já alertavam para

³¹ [Plano de Dados Abertos](#). Acesso em março de 2025.

a volatilidade na declaração de raça/cor entre edições do Enem, fenômeno que se repete em outras variáveis socioeconômicas, comprometendo análises longitudinais caso os pesquisadores não se atentem a essas inconsistências e façam os tratamentos específicos necessários.

A implementação da Lei Geral de Proteção de Dados em 2018 introduziu novos obstáculos no âmbito da integração. A remoção temporária de microdados públicos pelo Inep, motivada por riscos de reidentificação (como demonstrado em estudos da UFMG com o Censo Escolar de 2019), reduziu drasticamente a transparência. Atualmente, o acesso a bases completas restringe-se ao Serviço de Acesso a Dados Protegidos, aonde pesquisadores enfrentam limitações operacionais: impossibilidade de exportar dados brutos, dependência de autorizações burocráticas e ausência de metadados completos para preparação prévia de análises. Esse modelo, embora essencial para privacidade, cria um *trade-off* entre proteção individual e utilidade social dos dados, especialmente crítico em um contexto de infraestrutura digital ainda em maturação.

A fragmentação institucional agrava esses desafios. Sistemas municipais, estaduais e federais operam com protocolos desconexos, dificultando a interoperabilidade. Essa desarmonia reproduz padrões observados em países com infraestrutura administrativa limitada, onde censos assumem papel central por falta de alternativas (Shamoug et al., 2020). No Brasil, a centralidade dos censos é exacerbada pela baixa qualidade de metadados em bases locais e pela resistência cultural à padronização, fatores que impedem a construção de um ecossistema integrado de dados educacionais.

Ainda assim, é necessário diferenciar a esfera da interoperabilidade, que abrange os dados abertos, da esfera de microdados sensíveis. A interoperabilidade refere-se à capacidade de sistemas compartilharem dados não identificáveis por meio de padrões técnicos e semânticos. Tem como objetivo eliminar duplicações de coleta, garantir transparência na gestão e viabilizar políticas baseadas em evidências. Já os microdados sensíveis são uma exceção regulada, que devem estar sujeitos a anonimização robusta, ambientes controlados e ter finalidade específica, com relevância social comprovada, critérios esses todos atendidos e mantidos durante o uso do ambiente da sala segura.

Para avançar, é urgente equilibrar proteção de dados e abertura. Tecnologias como *Privacy-Enhancing Technologies* – incluindo técnicas de difusão de dados e síntese estatística – poderiam viabilizar o acesso a microdados anonimizados sem comprometer identidades. Paralelamente, a adoção de *frameworks* de interoperabilidade permitiria conciliar a

heterogeneidade histórica das variáveis. Do ponto de vista institucional, a governança de dados precisa ser priorizada, com a definição de padrões comuns e adoção de identificadores únicos (como o CPF mascarado) em todas as bases educacionais.

Para superar os desafios estruturais identificados, três eixos estratégicos emergem como prioritários. O primeiro envolve o desenvolvimento de infraestruturas técnicas unificadas, com a criação de plataformas integradas para coleta e armazenamento de dados, alinhadas a padrões internacionais como o DDI³² (*Data Documentation Initiative*). Essa padronização de metadados permitiria não apenas a harmonização de bases históricas, mas também a adoção de protocolos de interoperabilidade que facilitem o compartilhamento seguro entre sistemas municipais, estaduais e federais.

Em paralelo, é essencial investir na capacitação institucional, com programas de treinamento contínuo para gestores públicos em governança de dados e práticas FAIR³³ (*Findable, Accessible, Interoperable, Reusable* ou Encontrável, Acessível, Interoperável, Reutilizável). Isso incluiria a formação de equipes especializadas em metodologias de harmonização temporal, tratamento de *missing data* e uso de PETs, assegurando que a qualidade e a utilidade analítica dos dados sejam preservadas, sem prejudicar a anonimidade.

Por fim, a atualização do marco regulatório é urgente. A revisão da LGPD deve incorporar cláusulas específicas para dados educacionais, estabelecendo critérios claros para o uso de informações anonimizadas em pesquisas e políticas públicas, sem sacrificar a privacidade individual. Isso exigiria a criação de comitês interdisciplinares com participação de juristas, técnicos do Inep e representantes da sociedade civil, dentre outros, garantindo que o acesso a microdados seja simultaneamente ético e funcional. A sinergia entre avanços técnicos, capacitação humana e ajustes normativos representa o caminho mais viável para transformar os sistemas educacionais brasileiros em exemplos de transparência e eficiência baseada em evidências.

Enquanto o Brasil não superar essas barreiras, seguirá dependente das salas seguras – um modelo consolidado e necessário, mas insuficiente para democratizar, de forma ampla, o conhecimento gerado a partir de seus sistemas educacionais. A interoperabilidade não é apenas uma questão técnica: é um imperativo ético para garantir que políticas públicas sejam construídas com evidências robustas e inclusivas. Mas apesar de a interoperabilidade ser um direito coletivo para melhorar políticas públicas, o acesso a microdados é um privilégio

³² <https://ddialliance.org/>. Acesso em janeiro de 2025.

³³ <https://www.go-fair.org/fair-principles/>. Acesso em janeiro de 2025.

condicional, que demanda contrapartidas éticas irrenunciáveis. A verdadeira transparência não é expor indivíduos, mas garantir que sistemas educacionais sejam aprimorados sem sacrificar direitos fundamentais.

Conclusão

A conclusão deste trabalho está organizada em duas seções: a primeira sintetiza os principais achados dos capítulos, e a segunda discute as lacunas e limitações da pesquisa, apontando possíveis agendas para investigações futuras.

A importância dos determinantes socioeconômicos e individuais

Ao estudar a evasão definitiva do sistema de educação superior, identificou-se que grande parte dos estudos não usavam uma definição estruturada do fenômeno geral da evasão. O primeiro capítulo evidenciou a complexidade do fenômeno da evasão no ensino superior, e propôs uma taxonomia que integra as dimensões de granularidade (curso, instituição e sistema) e temporalidade (imediata, temporária e definitiva) com alguns qualificadores (área do conhecimento, campus, turno e modalidade). Essa estrutura conceitual tem como objetivo superar a fragmentação das abordagens tradicionais, oferecendo um instrumental robusto para a análise diferenciada dos tipos de evasão e de seus determinantes.

Com a clareza teórica do conceito de evasão definitiva do sistema, foi possível avançar para a sua aplicação e mensuração do fenômeno, no segundo capítulo, utilizando dados identificados do Inep, das coortes de ingresso no ensino superior de 2005 a 2022, combinando dados do CES com registros do Enem e CEB.

Devido ao volume dos dados (quase 40 milhões de registros) e à técnica empregada (regressão logística multinível), foi necessário usar um processo de amostragem estratificada em dois estágios. Essa etapa, combinada com os desafios informacionais identificados ao longo do tratamento, gerou duas amostras distintas: uma que considerou apenas as variáveis do CES, e outra que considerou também as variáveis provenientes do Enem e CEB. A essas amostras foram aplicadas regressões logísticas multinível, que consideraram os níveis de curso, conforme classificação Cine, e IES e modalidade (conjuntamente).

Os resultados obtidos destacaram a importância dos indicadores relativos aos determinantes socioeconômicos e individuais, por não só confirmarem as desigualdades estruturais como para reiterar a necessidade de políticas públicas voltadas à população mais vulnerável. Por um lado, temos variáveis como ocupação parental (razão de chances de 0,855 para famílias em ocupações de maior complexidade) e renda familiar (0,866 de razão de chances), que reduzem significativamente a probabilidade de evasão definitiva do sistema. Por outro, temos fatores individuais, como raça/cor autodeclarada mostrando-se crítica, com estudantes pretos, pardos e indígenas (PPI) apresentando até 14% mais chances de evasão em relação a brancos e amarelos. Sexo masculino emergiu como forte preditor de evasão (razão de chances de 1,539), enquanto

idade no ingresso revelou efeito paradoxal: na amostra enxuta associou-se a maior risco (1,022), mas na expandida a menor (0,841), sugerindo interações complexas com variáveis socioeconômicas. Estudantes PcD também apresentaram risco 15% maior de evasão, corroborando desafios estruturais de inclusão.

As dificuldades relatadas durante todo o processo de manipulação e tratamento dos dados foram tema do terceiro e último capítulo. Foram abordados os desafios relativos à integração e interoperabilidade das bases de dados educacionais, ressaltando que a fragmentação e a ausência de padrões consolidados dificultam o acompanhamento longitudinal das trajetórias estudantis e, consequentemente, a mensuração precisa da evasão definitiva.

Argumenta-se que o avanço na integração e interoperabilidade das bases de dados educacionais é condição indispensável para a superação das limitações metodológicas identificadas. A consolidação de uma infraestrutura nacional de dados educacionais interoperáveis permitiria não apenas o acompanhamento longitudinal das trajetórias estudantis e a mensuração precisa de todos os tipos de evasão, mas também a avaliação do impacto de políticas públicas e a formulação de estratégias mais eficazes para a redução das desigualdades, não só educacionais, no Brasil.

Limitações e agenda de pesquisa

Em termos metodológicos, a pesquisa enfrentou restrições decorrentes da incompletude de variáveis-chave, como raça/cor e origem escolar, e da necessidade de amostragem estratificada, que pode ter sub-representado realidades de instituições menores ou interiorizadas. Além disso, a classificação da evasão como “definitiva” está condicionada ao período de observação dos dados, não sendo possível descartar retornos futuros ao sistema.

No âmbito conceitual, permanece o desafio de incorporar de modo mais sistemático fatores qualitativos e subjetivos, como motivações pessoais, saúde mental e processos de decisão vocacional, que escapam às análises quantitativas baseadas em registros administrativos. Com base nessas limitações, propõe-se uma agenda de pesquisa que inclua: aplicação e validação da taxonomia proposta; realização de estudos qualitativos sobre as trajetórias e motivações dos estudantes evadidos; desenvolvimento de análises longitudinais que considerem os efeitos de eventos recentes, como a pandemia de Covid-19, sobre os padrões de evasão; e a inclusão de efeitos moderadores nas modelagens, de forma a verificar como interações de determinadas variáveis, como raça/cor e sexo, por exemplo, influenciam os resultados identificados.

Referências

Ali, A. et al. Sample size issues in multilevel logistic regression models. **PLoS ONE**, v. 14, n. 11, 1 nov. 2019.

Araújo, A.; Mariano, F.; Oliveira, C. Academic determinants of retention in Higher Education. **Ensaio: aval. pol. públ. Educ., Rio de Janeiro**, v. 29, n. 113, p. 1045–1066, out. 2021.

Arcanjo, J. **Fatores associados à permanência de estudantes no ensino superior no Distrito Federal**. [s.l.] Universidade de Brasília, 2022.

Ariovaldo, T. **Evasão nos cursos de licenciatura: análise a partir do Censo do Ensino Superior e de dados da Universidade Federal de Minas Gerais**. Belo Horizonte: UNIVERSIDADE FEDERAL DE MINAS GERAIS, 2023.

Assís, A et al. Graduação em Educação Física: motivos de ingresso, interrupção e permanência em cursos de Licenciatura e Bacharelado. **Revista Brasileira de Ciências do Esporte**, v. 44, 2022.

Astin, A.. Student Involvement: A Developmental Theory for Higher Education. Em: **College Student Development and Academic Life**. 1. ed. [s.l.] Routledge, 1998.

Barros, A. et al. **Arquitetura de Dados Educacionais como Plataforma para Governo Inteligente - Utilizando dados abertos para apoio à gestão educacional baseada em evidências**. Anais do XI Workshop de Computação Aplicada em Governo Eletrônico (WCGE 2023). **Anais...Sociedade Brasileira da Computação**, 6 ago. 2023.

Barroso, P. et al. DROPOUT FACTORS IN HIGHER EDUCATION: A LITERATURE REVIEW. **Psicologia Escolar e Educacional**, v. 26, 2022.

Bayma-Freire, H.; Roazzi, A.; Roazzi, M. O nível de escolaridade dos pais interfere na permanência dos filhos na escola? **Revista de Estudios e Investigación en Psicología y Educación**, v. 2, n. 1, p. 35–40, 31 jul. 2015.

Bean, J. Dropouts and turnover: The synthesis and test of a causal model of student attrition. **Research in Higher Education**, v. 12, n. 2, p. 155–187, 1980.

Bentes, M.; Kato, O. Fatores Que Afetam A Evasão Na Educação A Distância: Curso De Administração. **Psicol. educ., São Paulo**, n. 39, p. 31–45, dez. 2014.

BIS. **Data-sharing: issues and good practices**: IFC Report. [s.l: s.n.].

Bourdieu, P.; Passeron, J.. **A reprodução: Elementos para uma teoria do sistema de ensino**. 7. ed. [s.l.] Editora Vozes, 2014.

Boz, M.; Simsek, I. Analysis of Education Management Information Systems of the Ministry of National Education in Terms of Interoperability. **Journal of Qualitative Research in Education**, v. 22, n. 32, 17 out. 2022.

Bradley, S.; Migali, G. The effects of the 2006 tuition fee reform and the Great Recession on university student dropout behaviour in the UK. **Journal of Economic Behavior & Organization**, v. 164, p. 331–356, ago. 2019.

Branco, L. Conte, E.; Habowski, A. Evasão na educação a distância: pontos e contrapontos à problemática. **Avaliação: Revista da Avaliação da Educação Superior (Campinas)**, v. 25, n. 1, p. 132–154, abr. 2020.

Browne, W. Goldstein, H.; Rasbash, J. Multiple membership multiple classification (MMMC) models. **Statistical Modelling**, n. 1, p. 103–124, 2001.

Cabello, A. et al. Formas de ingresso em perspectiva comparada: por que o SISU aumenta a evasão? O caso da UNB. **Avaliação: Revista da Avaliação da Educação Superior (Campinas)**, v. 26, n. 2, p. 446–460, maio 2021.

Cabrera, A. Nora, A.; Castañeda, M. College Persistence: Structural Equations Modeling Test of an Integrated Model of Student Retention. **The Journal of Higher Education**, v. 64, n. 2, p. 123–139, mar. 1993.

Campos, L. et al. Cotas sociais, ações afirmativas e evasão na área de Negócios: análise empírica em uma universidade federal brasileira. **Revista Contabilidade e Finanças**, v. 28, n. 73, p. 27–42, 1 jan. 2017.

Cardoso, A. Nagai, N. Diversidade De Gênero E A Evasão Universitária Em Cursos De Graduação Em Administração, Ciências Contábeis, Ciências Econômicas E Sistemas De Informação Da Universidade Federal De Mato Grosso (Ufmg- Campus Rondonópolis). **Revista de Estudos Sociais**, v. 20, n. 41, p. 61, 7 fev. 2019.

Carvalhaes, F.; Senkevics, A. S.; Ribeiro, C. The intersection of family income, race, and academic performance in access to higher education in Brazil. **Higher Education**, v. 86, n. 3, p. 591–616, 1 set. 2023.

Castro, A. Teixeira, M. Evasão universitária: modelos teóricos internacionais e o panorama das pesquisas no Brasil. **Psicol. Argum**, v. 32, n. 79, p. 9–17, 2014.

Chege, Z. Wanjohi, P. Value Chain Approach to Data Production, Use, and Governance for Sound Policymaking in Africa. Em: **Data Governance and Policy in Africa**. Cham: Springer International Publishing, 2023. p. 31–70.

Coimbra, C.; Silva, L.; Costa, N. Evasion in higher education: definitions and trajectories. **Educacao e Pesquisa**, v. 47, p. 1–18, 2021.

Colpo, M.; Primo, T.; Aguiar, M. Lessons learned from the student dropout patterns on COVID-19 pandemic: An analysis supported by machine learning. **British Journal of Educational Technology**, v. 55, n. 2, p. 560–585, 1 mar. 2024.

Costa, A.; Picanço, F. Para Além Do Acesso E Da Inclusão: Impactos da raça sobre a evasão e a conclusão no ensino superior. **Novos Estudos CEBRAP**, v. 39, n. 2, p. 281–306, 1 ago. 2020.

Costa, F. Bispo, M. Pereira, R.. Dropout and retention of undergraduate students in management: a study at a Brazilian Federal University. **RAUSP Management Journal**, v. 53, n. 1, p. 74–85, jan. 2018.

Cunha, J. et al. Quem está ficando para trás? Uma década de evasão nos cursos brasileiros de graduação em Administração de Empresas e Ciências Contábeis. **Revista de Educação e Pesquisa em Contabilidade (REPeC)**, v. 9, n. 2, 28 jun. 2015.

Farias, R. Gouveia, V.; Almeida, L. Indicadores Do Sucesso Acadêmico Na Educação Superior: Análise Segundo Natureza Dos Cursos. **Psicologia Escolar e Educacional**, v. 28, 2024.

Farias, R.; Gouveia, V.; Almeida, L. Adaptação e sucesso acadêmico em estudantes brasileiros do primeiro ano da educação superior. **Revista de Estudios e Investigacion en Psicologia y Educacion**, v. 9, n. 1, p. 58–75, 1 jul. 2022.

Filho, A.; Nery, A. Ensino superior noturno no Brasil: história, atores e políticas. **Revista Brasileira de Política e Administração da Educação**, v. 25, n. 1, 2009.

Foley, K.; Gallipoli, G.; Green, D. Ability, Parental Valuation of Education, and the High School Dropout Decision. **Journal of Human Resources**, v. 49, n. 4, p. 906–944, 2014.

Gama, B.. **Determinantes da evasão universitária e impacto no gasto público**. Uberlândia: Universidade Federal de Uberlândia, 24 ago. 2018.

Gambirage, C. et al. Entre razões e emoções da evasão universitária, o contexto importa? Uma análise das instituições comunitárias catarinenses. **Interações (Campo Grande)**, p. 685–700, 3 nov. 2021.

Gascó, M. Approaching E-Government Interoperability. **Social Science Computer Review**, v. 30, n. 1, p. 3–6, 12 fev. 2012.

Gilioli, R.. **EVASÃO EM INSTITUIÇÕES FEDERAIS DE ENSINO SUPERIOR NO BRASIL: EXPANSÃO DA REDE, SISU E DESAFIOS**. Brasília: [s.n.].

Gomes, M.; Hirata, G. Determinantes da evasão no ensino superior: uma abordagem de riscos competitivos. **Pesquisa e Planejamento Econômico (PPE)** : v. 52, n. 03, dez. 2022., v. 52, n. 3, p. 9–37, 6 mar. 2024.

HOED, R.; LADEIRA, M.; LEITE, L.. Influence of algorithmic abstraction and mathematical knowledge on rates of dropout from Computing degree courses. **Journal of the Brazilian Computer Society**, v. 24, n. 1, 1 dez. 2018.

Inep. **Entenda a sua nota no Enem: guia do participante**. Brasília, DF: INEP: [s.n.]. Disponível em: <https://download.inep.gov.br/publicacoes/institucionais/avaliacoes_e_exames_da_educacao_basica/entenda_a_sua_nota_no_enem_guia_do_participante.pdf>. Acesso em: 16 abr. 2025.

Inep. **Tabelas de Divulgação do Censo da Educação Superior 2022**. Brasília: [s.n.]. Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-da-educacao-superior/resultados>>. Acesso em: 3 out. 2024.

Inep. **Saeb 2021: Indicador de Nível Socioeconômico do Saeb 2021: nota técnica**. Brasília: [s.n.]. Disponível em: <https://download.inep.gov.br/areas_de_atuacao/Indicadores_de_nivel_Nota_tenica_2021.pdf#page=22.20>. Acesso em: 4 mar. 2025a.

Inep. **Histórico**. Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-da-educacao-superior/historico>>. Acesso em: 16 abr. 2025b.

Kipnis, B. A pesquisa institucional e a educação superior brasileira: um estudo de caso longitudinal da evasão. **Linhas Críticas**, v. 6, n. 11, p. 109–130, 31 dez. 2000.

Klitzke, M.; Carvalhaes, F. Student Dropout In A Brazilian Public University: A Survival Analysis. **Educação em Revista**, v. 39, 2023.

Kowalski, A. et al. Evasão no Ensino Superior a Distância: Revisão da Literatura em Língua Portuguesa. **EaD em Foco**, v. 10, n. 2, 7 jul. 2020.

Lepine, A. Financial Aid and Student Performance in College: Evidence from Brazil. **Brazilian Review of Econometrics**, v. 38, n. 2, p. 221–261, 4 jan. 2019.

Lewis, G. et al. **Why Standards Are Not Enough to Guarantee End-to-End Interoperability**. Seventh International Conference on Composition-Based Software Systems (ICCBSS 2008). **Anais...IEEE**, fev. 2008.

Li, D. **O novo Enem e a plataforma Sisu: efeitos sobre a migração e a evasão estudantil**. São Paulo: UNIVERSIDADE DE SÃO PAULO, 2016.

Lima, F. Zago, N. Desafios conceituais e tendências da evasão no ensino superior. **Revista Internacional de Educação Superior**, v. 4, n. 2, p. 366–386, 2 abr. 2018.

Lima, M.; Fagundes, R. Educational Data Mining: A Study of the Factors That Cause School Dropout in Higher Education Institutions in Brazil. **Revista Novas Tecnologias na Educação**, v. 18, n. 1, jul. 2020.

Lopes, R.; Ribeiro, G.; Lisboa, L. S.; Silva, J. L. P.; Taconeli, C. A.. Fatores associados à evasão de calouros no ensino superior: um estudo com dados da Universidade Federal do Recôncavo da Bahia. **Revista Brasileira de Educação**, v. 28, 2023.

Lopes, T.. Estudo Dos Motivos Da Evasão Em Instituições De Ensino Superior. **Revista H-Tec Humanidades e Tecnologia**, v. 7, n. 2, p. 6–183, 2023.

Maas, C. ; Hox, J. Sufficient sample sizes for multilevel modeling. **Methodology**, v. 1, n. 3, p. 86–92, 2005.

Machado, A.; Terra, R.; Tannuri-Pianto, M. Higher education responses to accountability. **Economics of Education Review**, v. 98, 1 fev. 2024.

Machado, C.. et al. Atribuição de causalidade à evasão dos graduandos de ciências contábeis de uma instituição de ensino superior pública. **Pensar Contábil, Rio de Janeiro**, v. 23, n. 81, p. 25–35, maio 2021.

Madnick, S.; Zhu, H. Improving data quality through effective use of data semantics. **Data & Knowledge Engineering**, v. 59, n. 2, p. 460–475, nov. 2006.

MARQUES, F. T. A VOLTA AOS ESTUDOS DOS ALUNOS EVADIDOS DO ENSINO SUPERIOR BRASILEIRO. **Cadernos de Pesquisa**, v. 50, n. 178, p. 1061–1077, out. 2020.

MARTINS, E. DA C. et al. Um estudo dos preditores de sucesso acadêmico no ensino superior. **Revista de Psicología y Educación**, v. 10, n. 1, p. 77–100, 2015.

MASSINI-CAGLIARI, G. et al. Impacto da Política de Reserva de Vagas nas taxas de Evasão na Unesp. **Avaliação: Revista da Avaliação da Educação Superior (Campinas)**, v. 26, n. 1, p. 197–217, abr. 2021.

MEC. **Diplomação, Retenção e Evasão nos Cursos de Graduação em Instituições de Ensino Superior Públicas**. [s.l: s.n.].

MELLO, S. P. T. DE et al. **O FENÔMENO EVASÃO NOS CURSOS SUPERIORES DE TECNOLOGIA: UM ESTUDO DE CASO EM UMA UNIVERSIDADE PÚBLICA NO SUL DO BRASIL**. XIII Coloquio de Gestión Universitaria en Américas. *Anais...*Buenos Aires: nov. 2013.

MENDES, H. C. **Granularidade, Temporalidade e Alvitre: novas dimensões da evasão**. [s.l.] UNIVERSIDADE FEDERAL DE GOIÁS (UFG), 2023.

MIRANDA JÚNIOR, N. DA S.; CABELLO, A. F. Atratividade de cursos de graduação e a política institucional de mudança de curso: efeito trampolim? *Cadernos Gestão Pública e Cidadania*, v. 24, n. 77, 1 mar. 2019.

NANGINO, M. V. M.; PAIVA, J. A. DE. **AÇÕES AFIRMATIVAS: PERFIL DOS INGRESSANTES PELAS COTAS E ÍNDICE DE EVASÃO NOS CURSOS SUPERIORES DO CAMPUS AVANÇADO BOM SUCESSO**. IX ENCONTRO BRASILEIRO DE ADMINISTRAÇÃO PÚBLICA. *Anais...*São Paulo: SBAP, 5 out. 2022.

NASFI, R.; BRONSELAER, A.; DE TRÉ, G. A novel approach to assess and improve syntactic interoperability in data integration. *Information Processing & Management*, v. 60, n. 6, p. 103522, nov. 2023.

NIEROTKA, R. L.; SALATA, A.; KLITZKE MARTINS, M. FATORES ASSOCIADOS À EVASÃO NO ENSINO SUPERIOR: UM ESTUDO LONGITUDINAL. *Cadernos de Pesquisa*, v. 53, 2023.

NÓBREGA, B. S. DA et al. Evasion of students in higher education: An analysis of academic and socio-economic variables. *Cuadernos de Educación y Desarrollo*, v. 15, n. 1, p. 611–619, 19 abr. 2023.

NUTT, W.; RAZNIEWSKI, S.; VEGLIACH, G. Incomplete Databases: Missing Records and Missing Values. Em: [s.l: s.n.]. p. 298–310.

OECD. **Emerging privacy-enhancing technologies**. [s.l: s.n.].

OECD. **Open Government for Stronger Democracies**. [s.l: s.n.].

OECD. **Education at a Glance 2024**. [s.l.] OECD, 2024.

OLIVEIRA, M. M. DE; BARBOSA, E. F. **Multilevel modeling for the analysis and prediction of school dropout: A systematic review**. 47th International Computer Software and Applications Conference. **Anais...**IEEE Computer Society, 2023.

OLVERA ASTIVIA, O. L.; GADERMANN, A.; GUHN, M. The relationship between statistical power and predictor distribution in multilevel logistic regression: A simulation-based approach. **BMC Medical Research Methodology**, v. 19, n. 1, 9 maio 2019.

OSHIRO, C. H. **FINANCIAL AID AND DROPOUTS FROM PRIVATE HIGHER EDUCATION INSTITUTIONS IN BRAZIL**. 12th International Technology, Education and Development Conference. **Anais...**Valencia, Spain: mar. 2018. Disponível em: <<http://library.iated.org/view/OSHIRO2018FIN>>

PARENT, C.; SPACCAPIETRA, S. Database Integration: The Key to Data Interoperability. Em: **Advances in Object-Oriented Data Modeling**. [s.l.] The MIT Press, 2000. p. 221–254.

PARENTE, N. N. **AS CONDIÇÕES DE ACESSO E PERMANÊNCIA DOS ESTUDANTES DO CURSO DE LICENCIATURA EM FÍSICA DO IFCE, CAMPUS DE SOBRAL**. FORTALEZA: UNIVERSIDADE FEDERAL DO CEARÁ, 2014.

PAULA, G. B. DE. **DESIGUALDADES SOCIAIS E EVASÃO NO ENSINO SUPERIOR: UMA ANÁLISE EM DIFERENTES NÍVEIS DO SETOR FEDERAL BRASILEIRO**. [s.l.] UNIVERSIDADE FEDERAL DE MINAS GERAIS, 2021.

PENTEADO, B. E.; MALDONADO, J. C.; ISOTANI, S. **Modelo de infraestrutura para publicação de dados abertos educacionais conectados de qualidade**. Anais Estendidos do X Congresso Brasileiro de Informática na Educação (CBIE 2021). **Anais...**Sociedade Brasileira de Computação, 22 nov. 2021.

PEREIRA, R. R. et al. Alunos com Deficiência na Universidade Federal do Pará: Dificuldades e Sugestões de Melhoramento. **Revista Brasileira de Educação Especial**, v. 26, n. 3, p. 387–402, jul. 2020.

PINHEIRO, C. B.; RIBEIRO, J. L. L. DE S.; FERNANDES, S. A. F. Modelos teóricos da evasão no ensino superior e notas sobre o contexto nacional. **Avaliação: Revista da Avaliação da Educação Superior (Campinas)**, v. 28, 2023.

PINHEIRO, D. C.; PEREIRA, R. D.; XAVIER, W. S. Impactos das cotas no ensino superior: um balanço do desempenho dos cotistas nas universidades estaduais. **Revista Brasileira de Educação**, v. 26, 2021.

PRESTES, E. M. DA T.; FIALHO, M. G. D. Dropout rates in higher education and institutional management: The case of the Federal University of Paraíba. **Ensaio: aval. pol. públ. Educ., Rio de Janeiro**, v. 26, n. 100, p. 869–889, 1 jul. 2018.

RANGEL, F. DE O. et al. Evasão ou mobilidade: conceito e realidade em uma licenciatura. **Ciência & Educação (Bauru)**, v. 25, n. 1, p. 25–42, jan. 2019.

REALINHO, V. et al. Predicting Student Dropout and Academic Success. **Data**, v. 7, n. 11, 1 nov. 2022.

RIBES, D. **Notes on the Concept of Data Interoperability**. Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. **Anais...**New York, NY, USA: ACM, 25 fev. 2017.

ROCHA, D. N. DA; CARVALHAES, F. QUEM SÃO OS FUTUROS PROFESSORES DO BRASIL? O PERFIL SOCIOECONÔMICO DOS CURSOS DE LICENCIATURA DO ENSINO SUPERIOR. **Sociologia & Antropologia**, v. 13, n. 2, 2023.

ROSA, F. DE; OLIVEIRA, T. L. DE. Student Retention at a Higher Educational Institution: A Comparison Between Evaded and Nonevaded Groups Using Partial Least Square. **Journal of College Student Retention: Research, Theory & Practice**, v. 24, n. 2, p. 499–525, 19 ago. 2022.

SACALĂ, M. D. et al. Econometric research of the mix of factors influencing first-year students' dropout decision at the faculty of agri-food and environmental economics. **Economic Computation and Economic Cybernetics Studies and Research**, v. 55, n. 3, p. 203–220, 2021.

SACCARO, A.; FRANÇA, M. T. A.; JACINTO, P. DE A. Fatores Associados à Evasão no Ensino Superior Brasileiro: um estudo de análise de sobrevivência para os cursos das áreas de Ciência,

Matemática e Computação e de Engenharia, Produção e Construção em instituições públicas e privadas. **Estudos Econômicos (São Paulo)**, v. 49, n. 2, p. 337–373, abr. 2019.

SAMPAIO, B. et al. Desempenho no vestibular, background familiar e evasão: evidências da UFPE. **Economia Aplicada**, v. 15, n. 2, p. 287–309, jun. 2011.

SANTOS, C. A. DOS; PEREIRA, G. DE Q.; PILATTI, L. A. Higher Education Dropout: A Scoping Review. **Revista de Gestão Social e Ambiental**, v. 18, n. 8, p. e07156, 13 jun. 2024.

SANTOS, C. O. DOS; PILATTI, L. A.; BONDARIK, R. Evasão no ensino superior brasileiro: conceito, mensuração, causas e consequências. **Debates em Educação**, v. 14, n. 35, p. 294–314, 31 ago. 2022.

SENKEVICS, A. S. **O acesso, ao inverso: desigualdades à sombra da expansão do ensino superior brasileiro, 1991-2020**. São Paulo: Universidade de São Paulo, 2021a.

SENKEVICS, A. S. A expansão recente do ensino superior. **Cadernos de Estudos e Pesquisas em Políticas Educacionais**, v. 3, n. 4, p. 48, 22 abr. 2021b.

SENKEVICS, A. S.; MACHADO, T. DE S.; OLIVEIRA, A. S. DE. **A COR OU RAÇA NAS ESTATÍSTICAS EDUCACIONAIS: UMA ANÁLISE DOS INSTRUMENTOS DE PESQUISA DO INEP**. [s.l.: s.n.].

SERPA, L. F. P.; PINTO, N. M. A. C. A Evasão no ensino superior no Brasil. **Estudos em Avaliação Educacional**, n. 21, p. 109, 30 jun. 2000.

SHAMOUG, A. et al. **Conflict exit data management and integration at the United Nations**. New York: [s.n.].

SILVA, A. M. DA; SANTOS, B. C. S. Eficácia de políticas de acesso ao ensino superior privado na contenção da evasão. **Avaliação: Revista da Avaliação da Educação Superior (Campinas)**, v. 22, n. 3, p. 741–757, dez. 2017.

SILVA, A. M. DA; SAUAIA, A. C. A. EVASÃO E QUALIDADE EM INSTITUIÇÕES DE ENSINO SUPERIOR PRIVADAS: UMA ANÁLISE DA ECONOMIA DOS CUSTOS DE TRANSAÇÃO. **Administração: Ensino e Pesquisa**, v. 15, n. 4, p. 805, 31 dez. 2014.

SILVA, A. R. DA et al. Uma Análise dos Determinantes da Evasão no Ensino Superior EaD. **EaD em Foco**, v. 15, n. 1, p. e2322, 31 jan. 2025.

SILVA, E. C. DA. **Previsão da evasão acadêmica no ensino superior: O caso dos cursos de graduação presencial da Universidade de Brasília**. Brasília: Universidade de Brasília, 2023.

SILVA, F. C. DA; CABRAL, T. L. DE O.; PACHECO, A. S. V. **EVASÃO EM CURSOS DE GRADUAÇÃO: UMA ANÁLISE A PARTIR DO CENSO DA EDUCAÇÃO SUPERIOR BRASILEIRA**. XVI Coloquio Internacional de Gestion Universitaria. **Anais...** Arequipa, Perú: nov. 2016.

SILVA, G. P. DA. ANÁLISE DE EVASÃO NO ENSINO SUPERIOR: UMA PROPOSTA DE DIAGNÓSTICO DE SEUS DETERMINANTES. **Avaliação, Campinas**, v. 18, n. 2, p. 311–333, jul. 2013.

SILVA, M. E. N. F. DA. **Contextos e trajetórias do aluno na explicação do sucesso no Ensino Superior: A contribuição da modelação multinível**. [s.l.] Universidade do Minho, out. 2018.

SILVA, M. L. DA et al. Uma análise da evasão discente em cursos de Engenharia de uma Universidade Pública Brasileira. **Research, Society and Development**, v. 9, n. 8, 24 jun. 2020.

SILVA, E. C. et al. **A systematic review of the factors that impact the prediction of retention and dropout in higher education**. 56th Hawaii International Conference on System Sciences. **Anais...** Hawaii: HICSS, 3 jan. 2023. Disponível em: <<https://hdl.handle.net/10125/102787>>. Acesso em: 18 set. 2024

SILVA FILHO, R. L. L. E et al. A evasão no ensino superior brasileiro. **Cadernos de Pesquisa**, v. 37, n. 132, p. 641–659, dez. 2007.

SILVA FILHO, R. L. L. E; LOBO, M. B. DE C. M. **ESCLARECIMENTOS METODOLÓGICOS SOBRE OS CÁLCULOS DE EVASÃO**. [s.l.: s.n.].

SILVA, M. O. S. Evasão escolar: Desafios e Perspectivas da Educação no Brasil. Em: **Rumo ao futuro da Educação: tendências e desafios**. [s.l.] Editora Licuri, 2024. p. 239–251.

SILVA, P. T. DE F. E.; SAMPAIO, L. M. B. Student retention policies in higher education: reflections from a literature review for the Brazilian context. **Revista de Administração Pública**, v. 56, n. 5, p. 603–631, set. 2022.

SNIJDERS, T. A. B.; BOSKER, R. **Multilevel analysis: an introduction to basic and advanced multilevel modeling**. 2. ed. [s.l.] Sage, 2011.

TEODORO, L. DE A.; KAPPEL, M. A. A. Aplicação de Técnicas de Aprendizado de Máquina para Predição de Risco de Evasão Escolar em Instituições Públicas de Ensino Superior no Brasil. **Revista Brasileira de Informática na Educação**, v. 28, p. 838–863, 20 nov. 2020.

THEOBALD, E. Students are rarely independent: When, why, and how to use random effects in discipline-based education research. **CBE Life Sciences Education**, v. 17, n. 3, 1 set. 2018.

TINTO, V. Dropout from Higher Education: A Theoretical Synthesis of Recent Research. **Review of Educational Research**, v. 45, n. 1, p. 89–125, 1 mar. 1975.

UN. **Guidelines on open government data for citizen engagement**. New York: [s.n.].

UN. **United Nations E-Government Survey 2020: Digital Government In The Decade Of Action For Sustainable Development**. [s.l.: s.n.].

VIEIRA, J. DE A.; TELÓ, E. DE S.; VIEIRA, M. M. M. Fatores determinantes da evasão de estudantes de cursos superiores de tecnologia. **Debates em Educação**, v. 13, n. 32, p. 297–322, 31 ago. 2021.

VITELLI, R. F.; FRITSCH, R. Evasão escolar na educação superior: de que indicador estamos falando? **Estudos em Avaliação Educacional**, v. 27, n. 66, p. 908, 22 dez. 2016.

WEGNER, R. C. EVASÃO NO ENSINO SUPERIOR: DIGRESSÕES MOTIVADAS A PARTIR DA PANDEMIA DO NOVO CORONAVÍRUS. **Revista Docência e Ciberultura**, v. 6, n. 1, p. 01–22, 13 set. 2022.

WORLD BANK. **Information and Communication for Development 2018 : Data-Driven Development**. Washington, D.C.: [s.n.]. Disponível em:

<<http://documents.worldbank.org/curated/en/987471542742554246>>. Acesso em: 16 abr. 2025.

WORLD BANK. **Interoperability - Towards a Data-Driven Public Sector**. Washington, D.C.: [s.n.]. Disponível em: <<http://documents.worldbank.org/curated/en/099550101092318102>>. Acesso em: 16 abr. 2025.

ZALAF CASEIRO, L. C.; RAMOS DE AZEVEDO, A. Eficiência e potencial de expansão da educação superior pública. **Cadernos de Estudos e Pesquisas em Políticas Educacionais**, v. 2, p. 107–158, 22 nov. 2019.

Anexos

Tabela 12 - Pessoas de 25 anos ou mais de idade com superior completo. Brasil. 2016 a 2023.

Variável	2016	2017	2018	2019	2022	2023
Superior completo	15,4%	15,8%	16,6%	17,5%	19,2%	19,7%

Fonte: IBGE. Pesquisa Nacional por Amostra de Domicílios Contínua.

Tabela 13 - Ingressantes, matriculados e concluintes do ensino superior. Brasil. 2016 a 2023.

Ano-base	Ingressantes	Matriculados	Concluintes
2016	2.985.644	8.048.701	1.170.960
2017	3.226.249	8.286.663	1.201.454
2018	3.445.935	8.450.755	1.264.778
2019	3.633.320	8.603.824	1.250.239
2020	3.765.475	8.680.354	1.278.755
2021	3.944.897	8.986.554	1.327.188
2022	4.756.957	9.444.116	1.287.635
2023	4.994.192	9.977.217	1.374.789

Fonte: Inep. Estatísticas do Censo da Educação Superior.