

UNIVERSIDADE FEDERAL DE SÃO JOÃO DEL-REI
MESTRADO PROFISSIONAL EM ADMINISTRAÇÃO PÚBLICA - PROFIAP

JULIANA MARA PEREIRA DA CUNHA

EVASÃO ACADÊMICA DAS UNIVERSIDADES FEDERAIS: UMA ANÁLISE COM DADOS EM PAINEL DAS VARIÁVEIS DETERMINANTES NOS PERÍODOS PRÉ, DURANTE E PÓS-PANDEMIA DE COVID-19

São João del-Rei

2025

EVASÃO ACADÊMICA DAS UNIVERSIDADES FEDERAIS: UMA ANÁLISE COM DADOS EM PAINEL DAS VARIÁVEIS DETERMINANTES NOS PERÍODOS PRÉ, DURANTE E PÓS- PANDEMIA DE COVID-19

Dissertação de mestrado apresentada ao Programa de Mestrado Profissional em Administração Pública - PROFIAP, da Universidade Federal de São João del- Rei como parte dos requisitos para a obtenção do título de Mestre em Administração Pública.

Área de concentração: Administração Pública.

ORIENTADOR: Saulo Cardoso Maia

São João del-Rei

2025

JULIANA MARA PEREIRA DA CUNHA

EVASÃO ACADÊMICA DAS UNIVERSIDADES FEDERAIS: UMA ANÁLISE COM DADOS EM PAINEL DAS VARIÁVEIS DETERMINANTES NOS PERÍODOS PRÉ, DURANTE E PÓS-PANDEMIA DE COVID-19

Esta dissertação foi julgada adequada para a obtenção do título de Mestre em Administração Pública e aprovada em sua forma final pelo Orientador e pela Banca Examinadora.

Orientador:

Prof. Dr. Saulo Cardoso Maia, UFSJ. Doutor pela Universidade Federal de Lavras, Brasil, com doutorado sanduíche na University of Queensland, Austrália.

Banca Examinadora:

Prof. Dr. Cledinaldo Aparecido Dias, UFMG. Doutor pela Universidade de Brasília, Brasil.

Prof. Dr. Douglas Marcos Ferreira, UFSJ. Doutor pela Universidade Federal de Viçosa, Brasil.

Coordenador do PROFIAP:

_____

Profª. Drª. Caroline Miriã Fontes Martins São João del-Rei, agosto de 2025.

AGRADECIMENTOS

Primeiramente, agradeço a Deus, pela oportunidade de seguir essa experiência, por me proporcionar condições físicas e intelectuais, além de me guiar e sustentar ao longo de toda esta jornada acadêmica.

Agradeço à minha família, pelo apoio incondicional, incentivo constante e compreensão diante dos desafios enfrentados durante a realização deste trabalho.

Ao meu orientador, expresso minha sincera gratidão pela orientação, paciência e valiosas contribuições, que foram essenciais para o desenvolvimento desta dissertação.

Por fim, agradeço à UFSJ, que proporcionou os recursos, o ambiente acadêmico e as oportunidades necessárias para a realização desta pesquisa.

RESUMO

A evasão acadêmica é um problema que atinge universidades do Brasil e do mundo todo. Estudos apontam que esse fenômeno é complexo e pode ser decorrente de fatores sociais, pessoais, políticos, econômicos e acadêmicos. No Brasil, apesar da expansão do acesso ao ensino superior, impulsionada por políticas públicas do governo nas últimas décadas, verifica-se que pouca atenção é dada a políticas de permanência, o que resulta em índices elevados de deserção universitária. O objetivo deste trabalho é realizar um estudo sobre evasão nos cursos de graduação das universidades públicas federais brasileiras no período de 2018 a 2023 e, analisar separadamente os períodos anterior à pandemia de COVID-19, nos anos de 2018 e 2019; durante a pandemia, 2020 e 2021 e; posterior à pandemia, 2022 e 2023, para verificar possíveis interferências que a emergência de saúde pública vivenciada no país provocou no ensino superior. Para avaliar as principais variáveis relacionadas à evasão e a influência da pandemia de COVID-19 nesses índices acadêmicos, a metodologia utilizada foi a análise de dados em painel, tendo em vista identificar relações entre grupos específicos de discentes, a crise sanitária e a evasão. Foram utilizados os microdados sobre a educação superior disponibilizados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – INEP. Os resultados indicam que discentes do sexo feminino e de faixa etária abaixo dos 34 anos têm menor propensão à evasão. Com exceção dos alunos com menos de 18 anos que apontam para grandes chances de evasão. As análises indicam também que a pandemia está associada a uma tendência de aumento na evasão nos grupos mais velhos, acima dos 50 anos, nos discentes matriculados em cursos noturnos e matriculados por meio de reserva de vagas por renda familiar. Em contrapartida, alunos matriculados por meio de políticas de ações afirmativas em outras categorias como escola pública e deficiência apresentaram associação com a redução da evasão acadêmica, com diminuição ainda maior no período de 2020 a 2021. Discentes matriculados por reserva de vagas por etnia apresentaram comportamento oscilante ao longo do período analisado. O estudo revela ainda, que a concorrência nos cursos de graduação influencia inversamente a evasão, ao modo que existe grande diferença entre cursos de diferentes áreas, com tendência à menor evasão nos cursos da área de Saúde e bem-estar e maiores tendências nos cursos da área de Ciências naturais, matemática e estatística, principalmente no período pandêmico. Além disso, a pesquisa mostrou que os cursos presentes nas regiões Centro-Oeste, Nordeste e Norte apresentaram maiores efeitos na evasão acadêmica em comparação com a região Sudeste, principalmente durante a pandemia, o que corrobora desigualdades estruturais no sistema educacional brasileiro. A região Sul, por sua vez, revela maiores problemas no período de 2022 a 2023, indicando possivelmente reflexos de efeitos pós-pandemia, como impactos psicológicos e aumento das desigualdades sociais, que podem dificultar a integração dos estudantes ao ambiente universitário. De modo geral, os resultados apontam que a evasão acadêmica sofreu maiores impactos durante e após o período de pandemia.

Palavras-chave: Evasão Acadêmica, Censo da Educação Superior, Pandemia de COVID-19.

ABSTRACT

Academic dropout is a problem that affects universities in Brazil and around the world. Studies show that this phenomenon is complex and may be due to social, personal, political, economic, and academic factors. In Brazil, despite the expansion of access to higher education, driven by government public policies in recent years, little attention is given to retention policies, resulting in high rates of university dropouts. The objective of this study is to conduct a study on dropout rates in undergraduate courses at Brazilian federal public universities from 2018 to 2023 and, to analyze separately the periods prior to the COVID-19 pandemic, from 2018 to 2019; during the pandemic, from 2020 to 2021; and after the pandemic, from 2022 to 2023, to verify possible interferences that the public health emergency experienced in the country caused in higher education. To assess the main variables related to dropout rates and the influence of the COVID-19 pandemic on these academic indices, the methodology used was panel data analysis, with a view to identifying relationships between specific groups of students, the health crisis, and dropout rates. Microdata on higher education provided by the Anísio Teixeira National Institute for Educational Studies and Research (INEP) were used. The results indicate that female students under the age of 34 are less likely to drop out. The exception is students under the age of 18, who are more likely to drop out. The analyses also indicate that the pandemic is associated with an upward trend in dropout rates among older groups, those over 50 years of age, students enrolled in night courses, and those enrolled through family income-based seat reservations. In contrast, students enrolled through affirmative action policies in other categories, such as public school and disability, showed an association with a reduction in academic dropout rates, with an even greater decrease in the period from 2020 to 2021. Students enrolled through reserved spots based on ethnicity showed fluctuating behavior throughout the period analyzed. The study also reveals that competition in undergraduate courses has an inverse influence on dropout rates, as there is a large difference between courses in different areas, with a tendency toward lower dropout rates in courses in the area of health and well-being and higher tendencies in courses in the area of natural sciences, mathematics, and statistics, especially during the pandemic period. In addition, the research showed that courses in the Midwest, Northeast, and North regions had a greater effect on academic dropout rates compared to the Southeast region, especially during the pandemic, which corroborates structural inequalities in the Brazilian educational system. The southern region, in turn, reveals greater problems in the period from 2022 to 2023, possibly indicating the effects of the post-pandemic period, such as psychological impacts and increased social inequalities, which may hinder students' integration into the university environment. Overall, the results indicate that academic dropout rates were more severely impacted during and after the pandemic period.

Keywords: Academic Dropout, Higher Education Census, COVID-19 Pandemic**.**

LISTA DE GRÁFICOS

Gráfico 1- Dispersão da evasão em relação à concorrência 53
Gráfico 2- Histograma do índice geral de evasão no período de 2018 a 2023 54
Gráfico 3- Evasão média ao longo do período analisado (2018-2023) 55
Gráfico 4- Dispersão da evasão em relação à concorrência e ao percentual de matriculados por gênero, por idade, por turno e por reserva de vagas em ações afirmativas 56
Gráfico 5- Histograma da evasão em relação às áreas gerais, às regiões geográficas e aos períodos antes, durante e após a pandemia de COVID-19 57
Gráfico 6- Visualização dos resíduos de Pearson em relação às probabilidades previstas – dispersão, quantil normal e histograma 60
Gráfico 7- Distribuição dos Valores Preditivos 61
Gráfico 8- Distribuição das observações sobre a evasão, por área geral, por meio de diagrama de caixas 64
Gráfico 9- Evasão média por área geral ao longo tempo (2018-2023) 65
Gráfico 10- Distribuição das observações sobre a evasão, por região geográfica, por meio de diagrama de caixas 69
Gráfico 11- Distribuição das observações sobre a evasão pela proporção de matriculadas do gênero feminino, por meio de diagrama de caixas, divididas em tercis 78

LISTA DE TABELAS

Tabela 1- Fator de Inflação da Variância (VIF) considerando as variáveis relacionadas aos matriculados em valores percentuais no modelo de regressão 58
Tabela 2- Resultados de Wald Chi2 para os modelos com link probit ou logit, corr(ar1), corr(ar2) ou corr(unstructured) 59
Tabela 3- Estatísticas descritivas da evasão por área geral 62
Tabela 4- Estatísticas descritivas da evasão por região geográfica dos cursos 68
Tabela 5- Estatísticas descritivas da concorrência por área geral 71
Tabela 6- Estatísticas descritivas da concorrência por região geográfica dos cursos 72
Tabela 7- Evasão média pela proporção de matriculados no turno noturno dividida em tercis 73
Tabela 8- Estatísticas descritivas da evasão pela proporção de matriculados por reserva de vagas (RV) 75
Tabela 9- Evasão média pela proporção de matriculadas do gênero feminino dividida em tercis 77
Tabela 10- Estatísticas descritivas da evasão pela proporção de matriculados por faixa etária 79
Tabela 11- Coeficientes resultantes da regressão utilizando modelo de equações de estimação generalizadas (GEE) com distribuição binomial, função de ligação “logit”, estrutura de correlação “ar1” e erros robustos, nos 4 períodos analisados 82
Tabela 12- Número, incidência de casos e óbitos por COVID-19 segundo as Grandes Regiões do Brasil - Dados acumulados até 29/01/2024. 88

LISTA DE QUADROS

Quadro 1- Classificação das áreas gerais, áreas específicas e áreas detalhadas, conforme Classificação Internacional Normalizada da Educação CINE/UNESCO 31

Quadro 2- Variáveis comuns ao banco de dados sobre os cursos e ao banco de dados sobre as IES.....34

Quadro 3- Variáveis excluídas por não haver relação com a proposta do estudo.....36

Quadro 4- Variáveis selecionadas para análise em relação à evasão 50

LISTA DE ABREVIATURAS E SIGLAS

ABI - Área Básica de Ingresso

CINE - Classificação Internacional Normalizada da Educação

COE-MEC - Comitê Operativo de Emergência do Ministério da Educação

COVID - Doença por Coronavírus 2019 (Coronavirus Disease 2019)

CV - Coeficientes de Variação

FIES - Programa de Financiamento Estudantil

GEE - Equações de Estimação Generalizadas

IES - Instituições de Ensino Superior

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira

ISCED-F 2013 - Fields of Education and Training 2013

LDB - Lei de Diretrizes e Bases da Educação

LGPD - Lei Geral de Proteção de Dados

LM - Multiplicador de Lagrange

MEC - Ministério da Educação

MWNLS - Modelo dos Mínimos Quadrados não Lineares Ponderados Multivariados

N - Número de Observações

PARFOR - Plano Nacional de Formação de Professores da Educação Básica

PROUNI - Programa Universidade para Todos

RE - Efeitos Aleatórios

REUNI - Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais

RV - Reserva de Vagas

SD - Desvio Padrão (Standard Deviation)

TIC - Tecnologias da Informação e Comunicação

UFC - Universidade Federal do Ceará

UERJ - Universidade do Estado do Rio de Janeiro

UFMG - Universidade Federal de Minas Gerais

UFPI - Universidade Federal do Piauí

UNESCO - Organização das Nações Unidas para a Educação, a Ciência e a Cultura

VIF - Fator de Inflação da Variância

SUMÁRIO

1. INTRODUÇÃO 13
1.1 O PROBLEMA DE PESQUISA E SUA IMPORTÂNCIA 15
1.2 OBJETIVO GERAL 17
1.3 OBJETIVOS ESPECÍFICOS 18
1.4 HIPÓTESES 18
1.5 ORGANIZAÇÃO DO TEXTO 19
2. REVISÃO DE LITERATURA 20
2.1 A EVASÃO NO ENSINO SUPERIOR 20
2.2 MÉTODOS DE CÁLCULO DA EVASÃO 25
2.3 A PANDEMIA DE COVID-19 27
3. METODOLOGIA 29
3.1 MICRODADOS DO INEP SOBRE A EDUCAÇÃO SUPERIOR NO BRASIL 30
3.2 IMPORTAÇÃO DOS DADOS 34
3.3 ORGANIZAÇÃO DAS VARIÁVEIS 35
3.4 CÁLCULO DA EVASÃO E DA CONCORRÊNCIA 39
3.4.1 Cálculo da Evasão 39
3.4.2 Cálculo da Concorrência 41
3.5 REGRESSÃO COM DADOS EM PAINEL 42
3.5.1 Estimação do modelo 44
3.5.2 Testes de validação do modelo 48
4. DISCUSSÃO E ANÁLISE DOS RESULTADOS 52
4.1 TESTES DE CORRELAÇÃO E MULTICOLINEARIDADE 55
4.1.1 Teste de Correlação de Spearman 55
4.1.2 Fator de Inflação da Variância (VIF) 57
4.2 TESTES DO MODELO DE REGRESSÃO PARA RESPOSTA FRACIONÁRIA 58
4.3 ABORDAGEM DESCRITIVA DOS DADOS 61
4.3.1 Áreas gerais do conhecimento 61
4.3.2 Regiões geográficas do Brasil 67
4.3.3 Concorrência 70
4.3.4 Turno 73
4.3.5 Política de reserva de vagas 74
4.3.6 Gênero 77
4.3.7 Faixa etária 79
4.4 RESULTADOS PARA OS DADOS EM PAINEL 82

4.4.1 Efeitos das áreas gerais do conhecimento 84
4.4.2 Efeitos das regiões geográficas 87
4.4.3 Efeitos da concorrência 90
4.4.4 Efeitos do turno 91
4.4.5 Efeitos da política de reservas de vagas 92
4.4.6 Efeitos do gênero 95
4.4.7 Efeitos da faixa etária 96
4.4 DISCUSSÃO PARA AS HIPÓTESES 98
5. CONSIDERAÇÕES FINAIS 100
REFERÊNCIAS 104
APÊNDICES E ANEXOS 112
APÊNDICE A – ANÁLISES DAS VARIÁVEIS 112
APÊNDICE A.1 – REGRESSÃO ENTRE AS VARIÁVEIS EVASÃO E CONCORRÊNCIA APÓS TRANSFORMAÇÕES DE LINEARIZAÇÃO (LN, 1/X E LOG-LOG) PARA VERIFICAR A TRANFORMADA MAIS ADEQUADA 112
APÊNDICE A.2 – MATRIZ DE CORRELAÇÃO ENTRE AS VARIÁVEIS UTILIZADAS NA PESQUISA – DADOS DO INEP 114
APÊNDICE B – RESULTADOS DOS PROCEDIMENTOS OPERACIONAIS DOS DADOS DO CENSO DA EDUCAÇÃO SUPERIOR INEP 118
APÊNDICE B.1 – APLICAÇÃO DE TESTES PARA DEFINIÇÃO DO MODELO DE REGRESSÃO COM RESPOSTA FRACIONÁRIA 118
APÊNDICE B.2 – MODELO DE REGRESSÃO PARA ANÁLISES QUANTO À INFLUÊNCIA DOS PARÂMETROS (VARIÁVEIS INDEPENDENTES) EM RELAÇÃO À EVASÃO (VARIÁVEL DEPENDENTE) NO PERÍODO PRÉ-PANDEMIA (2018-2019) 126
APÊNDICE B.3 – MODELO DE REGRESSÃO PARA ANÁLISES QUANTO À INFLUÊNCIA DOS PARÂMETROS (VARIÁVEIS INDEPENDENTES) EM RELAÇÃO À EVASÃO (VARIÁVEL DEPENDENTE) NO PERÍODO DURANTE A PANDEMIA (2020-2021) 128
APÊNDICE B.4 – MODELO DE REGRESSÃO PARA ANÁLISES QUANTO À INFLUÊNCIA DOS PARÂMETROS (VARIÁVEIS INDEPENDENTES) EM RELAÇÃO À EVASÃO (VARIÁVEL DEPENDENTE) NO PERÍODO PÓS-PANDEMIA (2022-2023) 129
APÊNDICE B.5 – MODELO DE REGRESSÃO PARA ANÁLISES QUANTO À INFLUÊNCIA DOS PARÂMETROS (VARIÁVEIS INDEPENDENTES) EM RELAÇÃO À EVASÃO (VARIÁVEL DEPENDENTE) EM TODO O PERÍODO (2018 A 2023) 130

1. INTRODUÇÃO

A expansão do acesso ao ensino superior público no Brasil foi impulsionada pelo governo brasileiro por meio do Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais (REUNI), criado em 2007 (REUNI, 2010). Na rede privada, os investimentos foram por meio do Programa de Financiamento Estudantil (FIES) e do Programa Universidade para Todos (PROUNI) criados em 1999 e 2005, respectivamente (Santos, Amaral e Luz, 2023). Essa expansão foi fomentada pela ideia de promover mais intensamente o desenvolvimento e reduzir as desigualdades sociais do país. De acordo com Rodrigues *et al*. (2021), isso se justifica porque os avanços e as inovações tecnológicas, e também o crescimento econômico de um país, demandam recursos humanos qualificados e investimento em educação em diversas áreas do conhecimento, criando um círculo virtuoso para o desenvolvimento e a queda das desigualdades sociais e econômicas.

De modo geral, as ações que foram adotadas pelo Governo Federal, a partir de 2003, no sentido de universalizar o acesso à educação superior, devem ser entendidas tendo como base uma visão de longo prazo. Essa democratização do acesso teve como intuito a tentativa de reversão de um quadro histórico, onde o acesso à universidade era marcadamente elitizado (BRASIL, 2015).

Todas essas iniciativas do Governo, o aumento da oferta de cursos superiores a distância e as políticas de cotas vêm exercendo papel importante, porém limitado na redistribuição de oportunidades (Barros, 2015). Sendo assim, o processo de universalização da Educação Superior no Brasil apresenta inúmeras fragilidades (Lima, 2014). Embora haja esforços no sentido de expandir o acesso, Rodrigues et al. (2021) esclarecem que o êxito desse propósito ganha novos desafios quando se leva em consideração os índices de evasão. Por mais pressuroso que sejam as políticas de acesso, mais importante ainda é a concepção e a implementação de políticas públicas de permanência do aluno na universidade, tendo em vista que o ingresso por si só não traz os resultados educacionais esperados, bem como não promove a evolução da universalidade do ensino superior (Silva, 2022).

Por essa razão a evasão acadêmica tem sido objeto de trabalhos e pesquisas no mundo todo (Silva Filho *et al.*, 2007). Esse fenômeno tem atraído a atenção de pesquisadores e formuladores de políticas, com vistas à melhoria dos índices institucionais, melhoria pedagógica, economia nos gastos públicos, no ensino público, e diminuição de prejuízos, no

ensino privado. De acordo com Addison e Williams (2023), a retenção de alunos é uma grande preocupação para os administradores e o governo devido aos custos associados à evasão estudantil. Além de outros fatores político-sociais.

Lopes *et al.* (2023) explicam que a evasão na graduação pode estar atrelada a fatores demográficos, como idade e sexo, por exemplo; a fatores institucionais, como o ambiente, as estruturas das instituições e a qualidade do ensino; a fatores psicológicos e individuais, relacionados à motivação pessoal e ao contexto social no qual o aluno está inserido, além de fatores políticos, relacionados aos investimentos e outras políticas públicas na educação. Nesse sentido, a evasão acadêmica não pode ser caracterizada como um "fracasso individual", pois, conforme afirmam Shirasu e Arraes (2016), ela é resultado de uma das principais falhas estruturais no sistema educacional brasileiro.

Vitelli e Fritsch (2016), em consonância com outros autores, complementam, ainda, que os fatores que interferem no fenômeno da evasão possuem natureza diversificada, incluindo fatores sociais (Spady, 1970; Teixeira, Mentges e Kampff, 2018; Tinto, 1975; Tinto, 1988; Thomas, 2002), econômicos (Araujo *et al.*, 2023; Gomes e Hirata, 2022), de desempenho (Tinto, 1975) e de escolha (Teixeira, Mentges e Kampff, 2018). Entretanto, existe uma escassez de informações detalhadas sobre as metodologias e ferramentas usadas para calcular a evasão nos estudos existentes, o que dificulta o entendimento preciso dos fatores envolvidos.

Ainda, é importante frisar que a maior parte da oferta de cursos acadêmicos é feita pelas universidades privadas. Em 2019, 75% alunos de graduação estavam matriculados no setor privado (Almeida e Silva, 2022). De fato, existe no ensino superior brasileiro uma predominância quantitativa de cursos em instituições privadas, o que permite inferir que os dados referentes aos alunos das instituições particulares afetam de forma decisiva os indicadores globais do ensino superior brasileiro, inclusive a evasão (Silva Filho *et al.*, 2007). Nesse sentido, essa pesquisa se concentra em analisar a evasão no ensino superior federal para subsidiar medidas mais precisas quanto às políticas de gestão pública propriamente, embora a evasão acadêmica implique também em prejuízos pessoais, sociais e atraso no desenvolvimento econômico do país (Araujo *et al.*, 2023; Gomes e Hirata, 2022), o que também constituem, em essência, problemas públicos.

Além disso, grande parte dos trabalhos acadêmicos brasileiros sobre evasão são estudos de caso, que se concentram em estados, instituições, cursos ou categorias específicas de

estudantes (Adachi, 2009). Essa abordagem localizada dificulta a obtenção de um panorama nacional e totalizante (Coimbra, Silva e Costa, 2021). Um estudo específico aponta que os trabalhos encontrados no Brasil, até certo momento, analisam o problema da evasão em instituições específicas (Saccaro, França e Jacinto, 2019). Esse trabalho, por outro lado, se atentar em trazer uma abordagem a nível nacional.

1.1 O PROBLEMA DE PESQUISA E SUA IMPORTÂNCIA

O abandono de cursos superiores não é uma realidade local, é um assunto universal. Sendo assim, a evasão nas universidades tem se mostrado um problema recorrente em todo o mundo (Corrêa, Noronha e Miura, 2004; Santos, 2013; Silva Filho *et al.*, 2007). Segundo Zhang e Rangwala (2018) quase a metade dos estudantes universitários nos Estados Unidos não concluem o curso de graduação em seis anos e 60% deles desistem nos primeiros dois anos de estudo.

Essa análise traz certa cumplicidade de comportamentos em algumas áreas de estudo, sem, no entanto, deixar de reconhecer que existem diferenças entre instituições de ensino e peculiaridades socioeconômica e cultural em cada país (Santos, 2013).

No Brasil, de acordo com os dados apresentados pelo Censo da Educação Superior (INEP, 2021), a taxa de desistência do ensino superior, acumulada desde o ingresso em 2012, foi de 59% no ano de 2021. Quando analisadas instituições públicas e privadas separadamente, este índice cai para 48% nas estaduais e 55% na federais, ao passo que se eleva para 61% em instituições privadas (INEP, 2021). No ano seguinte, o Censo da Educação Superior (INEP, 2022) mostrou que essa taxa permanecia praticamente a mesma: 58% de desistência acumulada desde 2013 e, que quando separadas por rede de ensino, nas instituições privadas o índice de abandono é maior: 59% de desistência, contra 52% nas instituições públicas. Essa taxa de desistência se trata da taxa acumulada no período de 10 anos em cada curso de ingresso, ou seja, é o acompanhamento da trajetória dos estudantes nos cursos de graduação que ingressaram há10 anos (INEP, 2022).

O fenômeno da evasão pode ser decorrente de inúmeras razões. Os fatores acadêmicos e sociais podem ser considerados as principais causas para o abandono estudantil (Spady, 1970; Teixeira, Mentges e Kampff, 2018; Tinto, 1975; Tinto, 1982; Tinto, 1988; Tinto, 1993; Thomas,

2002). Sendo assim, além de questões de ordem pessoal, como expectativas com o curso e condição financeira do aluno, aspectos institucionais relacionados à infraestrutura da universidade e os métodos pedagógicos também constituem motivos para a desistência da graduação (Mello *et al.*, 2013).

Esse problema vem sendo discutido no Brasil e no mundo há algumas décadas. Isso reforça a importância de se atentar para essa realidade complexa e de difícil solução, uma vez que ela traz impactos significativos para a sociedade, para as instituições de ensino e para o país (Cunha e Maia, 2025). Portanto, pode-se categorizar a evasão acadêmica como um problema público, os quais necessitam de medidas de política pública. Clemente (2018) ressalta a metáfora de Secchi (2016) que diz, que o problema público é como uma doença, e a política pública é um tratamento.

Isso reforça o quão relevante é o assunto porque, conforme afirmam Coimbra *et al.* (2021), as políticas de combate à evasão, assim como quaisquer outras políticas públicas, devem partir de um diagnóstico ou de uma apreciação, onde se consideram os dados e as evidências acerca do problema social.

O impacto da evasão atinge diversas esferas. Do ponto de vista pessoal, o abandono da universidade pode gerar frustração, interrupção de sonhos e metas pessoais, perda de tempo e recursos investidos - ainda que tenha ingressado em curso de graduação público - e, menores oportunidades profissionais (Pinheiro, Ribeiro e Fernandes, 2023), podendo perder o acesso a empregos qualificados e mais bem remunerados. Ademais, Sachuck, Takahashi e Augusto (2008) afirmam que a qualificação capacita o trabalhador a operar a partir de novas tecnologias.

Por outro lado, com a evasão dos alunos, as instituições públicas de ensino superior perdem recursos financeiros, afetando seu planejamento orçamentário. Além disso, conforme dizem Pinheiro, Ribeiro e Fernandes (2023) e Silva Filho *et al.* (2007), a ineficiência no uso de recursos públicos se dá em consequência do gasto que cada estudante gerou pelo tempo que esteve matriculado, mas não concluiu, não gerou retorno, comprometendo os indicadores de gestão.

Para além dessas análises, há os prejuízos sociais e econômicos que muitas vezes não são percebidos pelos cidadãos. A redução de capital humano qualificado no mercado de trabalho e a baixa produtividade e inovação tecnológica provocam atraso no desenvolvimento social e econômico do país, gerando desigualdade (Araujo *et al.*, 2023; Gomes e Hirata, 2022).

Além do que, há um baixo retorno sobre o investimento público, reduzindo os efeitos positivos das políticas governamentais de acesso ao ensino superior, como explicam Araújo *et al*. (2023).

De modo geral, pode-se afirmar que o Brasil avançou em relação ao acesso à educação superior, mas que pouca atenção foi dada à questão da permanência. Por isso é de grande relevância o desenvolvimento de estudos que permitam identificar com maior precisão os fatores que levam os estudantes ao abandono do ensino superior (Nierotka, Salata e Martins, 2023).

Assim sendo, o estudo sobre a evasão acadêmica e os fatores que levam a esse comportamento se mostra relevante, uma vez que se trata de um assunto de interesse da sociedade e do governo, já que repercute em questões sociais e em gasto público.

É observável, pelo contexto acima apresentado, que a pesquisa pode ser pertinente, que pode contribuir para a formulação de políticas públicas acerca da evasão, trazendo à luz uma visão mais clara para a gestão pública, de como se comporta a evasão acadêmica no Brasil.

Diante dessa realidade, este trabalho buscar discutir o seguinte questionamento relacionado ao abandono acadêmico no Brasil: **quais as principais variáveis determinantes da evasão a nível de cursos de graduação nas universidades federais no período de 2018 a 2023?**

1.2 OBJETIVO GERAL

O objetivo deste trabalho é avaliar as principais variáveis determinantes na evasão acadêmica no Brasil no período de 2018 a 2023, em relação aos cursos presenciais de graduação das universidades federais.

Para isso, propõe-se realizar um estudo sobre evasão no ensino superior baseado nos dados do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), abordando análises baseadas nas áreas gerais do conhecimento e na concorrência, além de outros parâmetros sociodemográficos, como idade, gênero, política de reserva de vagas, turno e regiões geográficas, verificando as diferenças entre os períodos antes (2018-2019), durante (2020-2021) e após a pandemia de COVID-19 (2022-2023).

A pesquisa adota uma abordagem quantitativa, utilizando análise de regressão múltipla com dados em painel.

1.3 OBJETIVOS ESPECÍFICOS

Os objetivos específicos do trabalho são:

- Analisar se existe correlação entre taxas de evasão e a relação candidatos/vaga dos cursos de graduação.
- Analisar as áreas dos cursos de graduação que são mais suscetíveis à evasão.
- Analisar se o turno do curso interfere nos índices de evasão.
- Analisar se a evasão se apresenta de forma diversa de acordo com a região geográfica do país.
- Analisar se as características sociais como gênero e idade são fatores que podem influenciar os índices de evasão acadêmica.
- Analisar se discentes que ingressam por meio de reserva de vagas (conhecidas como cotas) têm maior propensão à evasão.
- Analisar a possível influência da pandemia da doença por coronavírus 2019 (COVID-19) em relação ao abandono do ensino superior, trazendo uma perspectiva dos anos de 2018 a 2023.

1.4 HIPÓTESES

A evasão no ensino superior é um fenômeno multifatorial associado a variáveis acadêmicas, institucionais, socioeconômicas e contextuais, que afetam diretamente a permanência e conclusão nos cursos (Gomes e Hirata, 2022; Barros, 2022; Feitosa, 2016). Estudos indicam que fatores como seletividade no ingresso, área do conhecimento, turno de oferta, região geográfica e perfil sociodemográfico influenciam significativamente os índices de evasão. Além disso, eventos conjunturais como a pandemia da COVID-19 alteraram substancialmente a dinâmica de permanência estudantil (Souza et al., 2022).

Tendo em vista os objetivos específicos delineados e a literatura que trata dos determinantes da evasão no ensino superior, as hipóteses de pesquisa a seguir são investigadas

ao longo desse trabalho. Tais hipóteses buscam orientar a investigação empírica, permitindo testar, por meio de análise estatística, possíveis associações e diferenças entre parâmetros acadêmicos e sociodemográficos dentro do contexto pandêmico.

H_1: Cursos de graduação com maior relação candidatos/vaga apresentam menores taxas de evasão, em razão de um perfil de ingresso mais competitivo e motivado.

H_2: Cursos das áreas de Licenciaturas e de Ciências Exatas, como Matemática e Estatística, apresentam maiores índices de evasão em comparação a cursos da área de Saúde.

H_3: Cursos ofertados no turno noturno apresentam maiores taxas de evasão em relação aos cursos diurnos, devido à maior sobreposição com atividades laborais.

H_4: A taxa de evasão apresenta variação significativa entre as regiões geográficas do Brasil, sendo mais elevada nas regiões Norte e Nordeste.

H_5: Estudantes do sexo masculino apresentam maior probabilidade de evasão em cursos de graduação.

H_6: Estudantes mais velhos apresentam maior propensão à evasão no período pandêmico.

H_7: Estudantes de baixa renda sofreram mais impactos durante a pandemia impactando a evasão.

H_8: As taxas de evasão no ensino superior aumentaram durante os anos de 2020 e 2021 em comparação ao período pré-pandemia (2018–2019).

1.5 ORGANIZAÇÃO DO TEXTO

O trabalho foi organizado em cinco capítulos: o primeiro discorre sobre a introdução do assunto estudado trazendo uma contextualização, seguida de problema de pesquisa, justificativa, e objetivos propostos. O segundo capítulo apresenta a revisão de literatura. O terceiro descreve a metodologia utilizada e o modelo econométrico adotado. No quarto capítulo

são apresentadas as análises obtidas no estudo. E por fim, o quinto capítulo traz as considerações finais.

2. REVISÃO DE LITERATURA

Nesse capítulo são abordadas revisões sobre a evasão acadêmica sob as perspectivas conceituais encontradas na literatura, atravessando constatações sobre fatores determinantes de acordo com abordagens de autores nacionais e internacionais que subsidiarão as análises futuras. Apresentando também alguns métodos de cálculo de evasão. O capítulo faz, ainda, um recorte sobre a evasão acadêmica e a pandemia de Covid-19 trazendo algumas normativas publicadas à época para propor medidas emergenciais de enfrentamento no que tange à educação.

2.1 A EVASÃO NO ENSINO SUPERIOR

Inicialmente, o termo “evasão acadêmica” requer uma definição sobre seu conceito, visto que pode ser interpretado de formas diferentes na literatura.

Nesse sentido, Tinto (1975) se atenta à dificuldade de pesquisadores diferenciarem o que se chama “fracassos acadêmicos” (saída involuntária) dos abandonos voluntários. Ou também as evasões definitivas daquelas temporárias. Ou ainda, a dificuldade em separar o conceito de abandono acadêmico das transferências entre instituições.

As autoras Abbad, Carvalho e Zerbini (2006), sob uma ótica mais genérica, apontam para a evasão como a desistência definitiva do aluno em qualquer etapa do curso. Ainda nesse sentido, Cardoso (2008) entende que a evasão acadêmica se refere aos alunos que saíram da universidade sem a conclusão do curso e sem formalizar transferência para outra universidade.

Vitelli e Fritsch (2016) explicam que a evasão no ensino superior pode ser analisada em diferentes níveis de granularidade: na disciplina, quando há reprovação por falta ou trancamento; no curso, quando o estudante se desliga dele por abandono, desistência, transferência, mudança interna ou exclusão por normas; na instituição, quando perde totalmente

o vínculo com a universidade ou faculdade em que estava matriculado; e no sistema, quando abandona, temporária ou definitivamente, o ensino superior como um todo.

Nos estudos de Adachi (2009) a autora considera como evadidos os casos de estudantes desligados dos cursos tanto por solicitação da instituição quanto por solicitação do estudante. Dessa forma, entende a evasão como desistência voluntária ou involuntária. Para as mudanças de vínculo por transferência interna (por meio dos processos de reopção de curso, de turno ou de grau acadêmico) todos são interpretados como evadidos. A autora explica, ainda, que a reopção depende da existência de outras vagas, desta maneira a ocupação de determinado lugar em um curso foi precedida por um desligamento de estudante (Adachi, 2009).

Para Silva Filho *et al.* (2007) a evasão é vista como o número de estudantes que ingressou em um curso, mas não obteve o diploma ao final de um tempo. De modo semelhante, este estudo adota a abordagem de Adachi (2009) e Silva Filho *et al.* (2007), e reconhece que qualquer vaga ocupada que não levou à diplomação é entendida como evasão, seja por jubilamento, desistência voluntária, transferência externa, reopção etc.

Partindo da abordagem conceitual, a busca pela definição de fatores que determinam a evasão no ensino superior tem despertado a atenção de pesquisadores e administradores públicos de todo o mundo. Há algumas décadas o abandono dos cursos de graduação tem sido alvo de estudos para tentar entender este fenômeno. A compreensão de suas causas é objeto de muitos trabalhos e pesquisas educacionais (Samasonok, Kamienas e Juškevičienė, 2023; Silva *et al.*, 2022).

No contexto internacional, Behr *et al.* (2020) levantam três dimensões de fatores associados à evasão: nível nacional, relacionados aos tipos de cursos e às desigualdades socioeconômicas; nível institucional, no que tange às condições físicas e administrativas da instituição e à relação com os docentes; e nível individual, determinados por fatores demográficos e por motivação pessoal.

Sendo assim, a desmotivação dos estudantes universitários para concluir os cursos de graduação podem ser diversos. Nesse sentido, pode haver desistência em função da necessidade de trabalhar, motivados por razões financeiras, problemas institucionais, problemas pessoais ou circunstâncias relacionadas à família.

Segundo Mello *et al.* (2013), os fatores para evasão são divididos em internos e externos. Desse modo, tanto os métodos didáticos pedagógicos e a infraestrutura da universidade, quanto as questões de ordem pessoal, como a realidade financeira ou a expectativa com o curso escolhido, são agentes capazes de influenciar na desistência.

Tinto (1975) faz um paralelo da evasão e a relação com fatores externos, fatores pessoais e o nível de integração social e acadêmica. Segundo o autor, a decisão de permanecer ou abandonar o ensino superior resulta de um processo dinâmico de interação entre o estudante e a instituição, no qual se entrelaçam fatores pessoais, como expectativas e motivações próprias, e fatores externos, como pressões econômicas, obrigações profissionais e domésticas e eventos da vida cotidiana. Esses elementos influenciam diretamente a forma como o aluno se integra ao ambiente acadêmico e social: no plano acadêmico, a qualidade das interações com docentes, o desempenho nas disciplinas e o envolvimento intelectual; no plano social, as relações com colegas, a participação em grupos e o sentimento de pertencimento (Tinto, 1975).

Assim, pode-se dizer o abandono estudantil é proveniente de fatores acadêmicos e sociais (Spady, 1970; Teixeira, Mentges e Kampff, 2018; Tinto, 1975; Tinto, 1982; Tinto, 1988; Tinto, 1993; Thomas, 2002). De acordo com Tinto (1975), quanto maior o grau de integração nessas duas dimensões, maior a probabilidade de o estudante persistir. Nesse sentido, os sistemas acadêmicos e sociais das universidades constituem estruturas importantes nos processos de evasão de alunos e devem ser estudados (Spady, 1970).

Em uma pesquisa feita na Universidade Federal do Ceará (UFC) em relação à evasão nos anos de 1999 e 2000, Andriola *et al.* (2006) constataram que dentre os principais motivos para a evasão, encontram-se o baixo conhecimento sobre o curso no momento de ingresso, a desmotivação com os estudos e a precariedade das condições físicas das instituições.

Por outro lado, a deficiência na qualidade do ensino básico e médio permitem que alunos cheguem ao ensino superior com preparo insuficiente para conseguir manter o curso. Dessa forma, o baixo desempenho escolar pode estar intimamente relacionado ao desligamento do aluno por abandono involuntário, mas não à forma de desistência voluntária (Tinto, 1975). Para Tinto (1975) e Pietras (2010), a saída voluntária parece estar mais relacionada à falta de congruência entre o indivíduo e o clima intelectual da instituição, assim como a relação com os pares. Nesse sentido, as instituições de ensino maiores, ao fornecerem uma maior diversidade de culturas estudantis e docentes e, portanto, levarem a uma maior probabilidade de

correspondência social e intelectual, parecem reduzir a taxa de evasão de seus alunos (Tinto, 1975).

Kehm *et al.* (2019) constataram que o interesse pelo assunto estudado, o esforço para aprendizagem autodirigida e a possibilidade de conseguir um emprego no futuro constituem elementos essenciais para motivar a continuidade no ensino superior. Contudo, a falta de conexão entre as expectativas dos jovens com relação ao mercado de trabalho pode tornar a trajetória acadêmica desmotivante. Os jovens são delimitados pelo imediatismo, tornando a permanência em um curso superior com duração de alguns anos um tanto quanto árdua.

Em outra perspectiva, foi feito um estudo na Universidade do Estado do Rio de Janeiro (UERJ) sobre o ingresso por meio de políticas de ações afirmativas, analisando a progressão de estudantes cotistas e não cotistas ingressantes em 2005 até o ano de 2011. O autor Mendes Junior (2014) constatou que, neste grupo, os alunos cotistas apresentaram maiores taxas de graduação e menores taxas de evasão. Ele acredita que esses alunos tendem a ter um maior senso de valor ao curso, apesar de apresentarem coeficientes de rendimento inferiores.

Já por uma abordagem econômica, o abandono acadêmico, de acordo com Silva Filho *et al.* (2007), afeta todos os sistemas educacionais, causando desperdícios financeiros. A evasão no ensino superior está presente em instituições públicas e privadas. No setor privado, resulta em perdas de receitas e, no setor público, caracteriza-se por recursos públicos sem o devido retorno (Silva Filho *et al.*, 2007).

Isso é identificado como ineficiência no gasto dos recursos públicos, visto que as universidades mantêm a mesma estrutura projetada para atender a um número de estudantes maior do que aquele efetivamente atendido, ou seja, a universidade falha em cumprir a missão de formar pessoas que contribuam com a sociedade, o progresso e o desenvolvimento da nação (Cardoso e Nagai, 2019).

Gomes e Hirata (2022) explicam que o investimento realizado pela sociedade na educação superior é desperdiçado a cada aluno que evade, uma vez que grande parte dos estudantes de nível superior é financiada pelo setor público de alguma forma, seja pela universidade pública, seja por meio de financiamento das mensalidades no setor privado.

Ainda no âmbito público, é relevante relembrar que a ampliação do acesso à educação superior no Brasil buscou alcançar as metas definidas pelo governo. Essa tentativa foi válida

no sentido de que a participação dos jovens na educação superior é fundamental para estimular a inovação e o desenvolvimento econômico e tecnológico, garantir os níveis adequados de produção de bens e serviços especializados e favorecer a promoção da justiça social (Junior *et al.*, 2019; Gomes e Hirata, 2022). Todavia, promover a entrada na universidade não garante esses resultados se não houver uma política que forneça condições de permanência dos estudantes até a sua conclusão (Silva, 2022; Junior *et al.*, 2019). Nesse sentido, a compreensão aprofundada da evasão é fundamental para entender o fenômeno que afeta a educação superior.

De forma geral, os principais motivos que fomentam a necessidade de um aluno abandonar a graduação são falta de motivação dos alunos e professores, problemas pessoais e socioeconômicos, insatisfação com o curso/instituição, problemas de aprendizagem associados com metodologias de ensino, restrições do mercado de trabalho, nível de conhecimento prévio à graduação, desconhecimento sobre o curso, entre outros (Kantorski *et al.*, 2016).

De acordo com Silva e Roman (2021) é preciso identificar as características apresentadas por grupos de alunos que estão em risco de deserção com o intuito de auxiliar as universidades a adotarem medidas mais assertivas, uma vez que os estudos de evasão constituem um suporte importante para os processos de avaliação institucional (Corrêa, Noronha e Miura, 2004).

Rodrigues (2012) explica que, por mais que a evasão possa existir devido a problemas de ordem pessoal, o importante é que o aluno não abandone o curso por motivo que poderia ser evitado por parte da instituição, sobretudo a pública.

Cardoso e Nagai (2019) salientam que, com os altos índices de evasão, as universidades públicas falham em cumprir o papel de formar pessoas que contribuam com progresso da sociedade, tornando-se ineficientes no processo ensino-aprendizagem. Além disso, a sociedade de uma forma geral também acaba perdendo em cidadania e formação de capital social. Dessa forma, identificar características de ordem acadêmica, social e econômica por trás da evasão pode auxiliar os gestores públicos na formulação de políticas públicas para reverter esse quadro.

2.2 MÉTODOS DE CÁLCULO DA EVASÃO

O cálculo da evasão pode ser encontrado na literatura científica por meio de abordagens diversas. Os pesquisadores tendem a adotar um método que mais se adequa ao seu trabalho.

No ano de 1995 foi instituída a Comissão Especial de Estudos sobre Evasão atrelada ao Ministério da Educação (MEC) que contou com o apoio das universidades para a formulação de um relatório que abordasse as causas gerais e específicas do fenômeno evasão, e apresentasse sugestões para minimizar tais índices (ANDIFES, 1996). Nesse trabalho, a evasão é considerada a partir de uma série histórica de dados sobre uma geração de discentes ingressantes e o tempo máximo de integralização. No relatório fica identificado como evadidos os alunos que não se diplomaram no período do curso e que não estão mais vinculados (ANDIFES, 1996), de acordo com o cálculo (1) a seguir, definido pela Comissão:

$$E = \left[\frac{N_i - N_d - N_r}{N_i}\right] x100 \quad (1)$$

Onde:

E = Percentual de evadidos;

N_i = Número de ingressantes no ano-base;

N_d = Número de diplomados;

N_r = Número de retidos;

Por outro lado, Adachi (2009) adota a fórmula (2) utilizada para o cálculo da evasão nos cursos de graduação da Universidade Federal de Minas Gerais (UFMG), onde considera-se o número de ingressantes menos o número de diplomados, dividido pelo número de ingressantes, em valores percentuais:

$$E = \left[\frac{N_i - N_d}{N_i}\right] x100 \quad (2)$$

Onde:

E = Percentual de evadidos;

N_i = Número de ingressantes no ano-base;

N_d = Número de diplomados;

Um dos métodos de cálculo de evasão muito difundido é o de Silva Filho *et al.* (2007) *et al*. (2007). Em nota de esclarecimentos metodológicos sobre os cálculos de evasão, Silva Filho e Lobo (2012) definem que é possível se fazer os cálculos baseados na evasão por instituições de ensino superior (IES), retirando-se do cálculo os ingressantes oriundos de transferências entre IES, ou entre cursos na mesma instituição. Por outro lado, também é possível calcular a evasão utilizando informações gerais de todo o sistema, como se apresentam os dados no INEP.

A equação (3) de Silva Filho *et al*. 2007 foi o método de cálculo adotado neste trabalho. Primeiramente, pelo formato de dados disponíveis, que são as informações gerais de todo o sistema educacional no Brasil apresentados pelo INEP. Além disso, é considerado um cálculo adequado devido à capacidade de medir a perda de alunos de um ano para o outro em relação ao total de alunos que, potencialmente, deveriam se matricular, uma vez que não concluíram o curso (Silva Filho e Lobo, 2012).

Assim sendo, Silva Filho *et al*. (2007) utilizam como cálculo básico da evasão a comparação entre o número de alunos que estavam matriculados num determinado ano, subtraídos os concluintes, com a quantidade de alunos matriculados no ano seguinte, subtraindo-se deste último os ingressantes desse ano. Dessa forma, é possível medir a perda de alunos de um ano para outro, o que se entende por evasão anual (Silva Filho *et al.*, 2007), conforme a equação (3):

$$E_n = 1 - \left[\frac{M_n - I_n}{M_{n-1} - C_{n-1}}\right] x100 \quad (3)$$

Onde:

E = Percentual de evadidos;

M = Número de matriculados;

I = Número de ingressantes;

C = Número de concluintes;

n = Ano do estudo;

n-1 = Ano anterior.

Hoed (2016) explica que a maioria dos estudos de Silva Filho *et al.* (2007) sobre evasão acadêmica lida com dados agregados, não sendo possível a análise por uma coorte em que se acompanha individualmente os discentes, uma vez que os dados do INEP disponibilizam informações totalizadas de matriculados, concluintes e ingressantes.

Essa metodologia de cálculo foi amplamente adotada por pesquisadores sobre evasão acadêmica. Dentre os principais autores que utilizaram o cálculo de evasão de Silva Filho *et al*. (2007) destacam-se os trabalhos publicados de Pinheiro *et al*. (2023), Faustino-Ferber e Martins (2022) Coimbra et al. (2021), Hoed (2016), Santos (2016), Gerab (2016) e Fialho (2014).

Ao todo, o trabalho entitulado como "A Evasão no Ensino Superior Brasileiro" (Silva Filho *et al.*, 2007), onde foi lançada a metodologia de cálculo utilizada nesta pesquisa, obteve até o momento 1025 citações no Google Scholar e 75 na Scopus. Esses números mostram a relevância e aceitação do método proposto pelos autores para o cálculo da evasão em outros trabalhos acadêmicos.

2.3 A PANDEMIA DE COVID-19

A pandemia de COVID-19 foi um marco histórico para o Brasil e o mundo em diversos aspectos. Costa e Silva (2024) afirmam que a crise ocasionada pela pandemia teve repercussões não apenas no sistema de saúde, mas também nos âmbitos social, econômico, cultural, educacional e político.

No ano de 2020, aproximadamente um mês após ser declarada a situação emergencial de saúde pública em decorrência da COVID-19 no Brasil (Brasil, 2020a), foi instituído o Comitê Operativo de Emergência do Ministério da Educação (COE-MEC) (Brasil, 2020b) para propor medidas de enfrentamento no que tange à educação. A partir dele foram publicadas a Portaria nº 343/2020, que foi alterada pelas Portarias nº 345/2020 e nº 395/2020, e a Medida Provisória nº 934/2020, onde dispunham sobre a substituição de aulas presenciais por aulas remotas, utilizando meios digitais, a fim de evitar o contato social (Brasil, 2020c; Brasil, 2020d). Além disso, essa normativas flexibilizavam os dias letivos, as horas-aula e o calendário de férias das instituições (Brasil, 2020c; Brasil, 2020d).

A pandemia de COVID-19 afetou mais de 1,6 bilhão de estudantes de todas as idades. Isso representa cerca de 94% da população estudantil mundial (UNESCO, 2022). Em nota, o chefe da seção de Saúde e Educação da UNESCO, Christopher Castle, a especialista de Programas, IIEP- UNESCO, Leonora MacEwen e a especialista associada de Programas, IIEP-UNESCO, Thalia Séguin, afirmam que a pandemia de COVID-19 possivelmente não será a última crise enfrentada pelo mundo, e que os sistemas educacionais necessitam se preparar para adaptar-se e recuperar-se, tornando-se mais fortes e equitativos (UNESCO, 2022).

Diante do contexto pandêmico, manter-se frequente e ativo nas escolas e universidades se revelou um desafio para muitos estudantes, seja pelas limitações econômicas, pela dificuldade de adaptação ao formato on-line das aulas, ou ainda pelas dificuldades psicológicas e emocionais associadas ao isolamento social (Aristovnik *et al.*, 2021).

Gusso *et al.* (2020) afirmam que, embora não seja possível evidenciar os efeitos do ensino emergencial remoto (à época), já era possível perceber consequências já com o fim do primeiro semestre acadêmico afetado pela pandemia, como o aumento da probabilidade de evasão do ensino superior.

Após esse período, outros estudos apontaram para efeitos mais evidentes. Em relação à saúde mental, Araujo *et al.* (2023) afirmam que esse problema foi indicado como o maior motivo para evasão durante a pandemia. Além disso, de acordo com esses autores, as responsabilidades relacionadas ao trabalho e ao contexto familiar, especialmente para as mulheres como o cuidado com filhos ou parentes mais velhos, indicaram fortes correlações com o abandono acadêmico (Araujo *et al.*, 2023).

Já os autores Martínez-Líbano e Yeomans-Cabrera (2023) explicam que o desempenho acadêmico prejudicado durante a pandemia e o perfil financeiro fragilizado pela instabilidade econômica foram identificados como principais preditores do risco de evasão. Chagas (2019) afirma que o baixo desempenho acadêmico, especialmente nos anos iniciais do curso, é um fator significativo de evasão, sendo muitas vezes um sintoma de distanciamento do aluno em relação ao curso.

Além disso, a implementação do ensino remoto emergencial por meio de plataformas digitais não favoreceu o engajamento dos alunos, que foram pegos de surpresa por esse modelo (Faria, 2022). Faria (2022) afirma, ainda, que a falta de condições mínimas impostas para o

ensino remoto, como dificuldades no uso de meios digitais, causou desmotivação com o andamento dos cursos.

De modo geral, a pandemia evidenciou vulnerabilidades existentes na sociedade em suas várias dimensões, com repercussões psicológicas e sociais que perpassaram as questões educacionais (Barros, 2022). As literaturas indicam que a pandemia de COVID-19 teve um impacto significativo na evasão no ensino superior, tanto ao exacerbar vulnerabilidades existentes quanto ao introduzir novos desafios para os estudantes e instituições.

3. METODOLOGIA

A metodologia de estudo empregada explica como os dados foram obtidos, tratados e analisados. Trata-se de uma pesquisa com abordagem predominantemente quantitativa, com aplicação de técnicas estatísticas. Os estudos apresentados no capítulo de discussão e análise dos resultados foram realizados a partir do tratamento dos microdados sobre a educação superior no Brasil dos anos de 2017 a 2023, fornecidos pelo INEP.

Para analisar as possíveis relações de correlação entre a evasão acadêmica e fatores sociodemográficos foi utilizado um conjunto de variáveis determinadas com base na revisão de literatura, adaptando-se os modelos e cálculos já utilizados, seguindo a metodologia de Regressão com Dados em Painel, técnica adequada para tratamento de dados longitudinais (Gujarati e Poter, 2008).

Para cada censo, de 2017 a 2023, é disponibilizado pelo INEP um arquivo com os microdados em formato .csv (*Comma-Separated Values*) com informações sobre os cursos superiores e outro sobre as instituições de ensino superior (IES). É fornecido também um arquivo em formato .pdf (*Portable Document Format*) denominado “leia-me” trazendo instruções fundamentais para a correta extração e manipulação dos dados. Além disso, são disponibilizados 2 anexos: um arquivo em formato .pdf (*Portable Document Format*) contendo os questionários utilizados no censo da educação superior para obtenção dos dados, o qual não foi utilizado por não se aplicar a esse estudo, e outro anexo em .xlsx (*Excel*) contendo o dicionário dos dados, onde é possível visualizar com clareza todas as variáveis disponíveis e uma descrição breve sobre cada uma delas.

Os microdados sobre a educação superior em 2024 não foram disponibilizados, pois até a data deste estudo, ainda estava em período de coleta de dados pelo Censo da Educação Superior (INEP, 2025).

Para os estudos estatísticos foi utilizado o software Stata®, versão 16.1 SE.

3.1 MICRODADOS DO INEP SOBRE A EDUCAÇÃO SUPERIOR NO BRASIL

O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) é uma organização criada em 1937 sob a denominação de "Instituto Nacional de Pedagogia", e transformado em autarquia federal vinculada ao Ministério da Educação (MEC) em 1997, é o órgão federal responsável pelas evidências educacionais e atua em três esferas: avaliações e exames educacionais; pesquisas estatísticas e indicadores educacionais; e gestão do conhecimento e estudos (INEP, 2025).

Os microdados fornecidos pelo INEP reúnem um conjunto de informações detalhadas relacionadas às pesquisas, aos exames e avaliações do Instituto (INEP, 2025). Como ferramenta de transparência, fornecem um amplo panorama da educação brasileira, fomentando um grande acervo de dados sobre a educação superior no país. São uma fonte segura e eficaz de obtenção de dados, acessíveis aos pesquisadores, estudantes, gestores e sociedade em geral (INEP, 2023).

Tendo em vista o atendimento às normas previstas na Lei Geral de Proteção de Dados (LGPD) – Lei nº 13.709, de 14 de agosto de 2018, os arquivos contendo os microdados do INEP estão passando por uma reestruturação para suprimir a possibilidade de qualquer tipo de identificação de pessoas (INEP, 2025), sendo assim, os dados disponibilizados se resumem às Instituições de Ensino Superior (IES) e aos cursos.

O INEP classifica os cursos de graduação e sequenciais de acordo com a Classificação Internacional Normalizada da Educação (CINE) da Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO) apresentada no anexo I do documento Dicionário de Dados (INEP, 2023). De acordo com o site do INEP (2025), a CINE tem por referência a metodologia *Fields of Education and Training* 2013 (ISCED-F 2013), com adequações à realidade educacional brasileira. A ISCED é produzida pelo Instituto de Estatísticas da

UNESCO com a finalidade de reunir, compilar e analisar estatísticas educacionais comparáveis com as de outros países, regiões e estados, e que possibilita ordenar cursos por níveis de ensino e áreas de formação (INEP, 2025).

De acordo com o CINE (2018), o uso de uma metodologia reconhecida internacionalmente permite que as estatísticas geradas sejam mais fidedignas, pois são formuladas a partir de uma metodologia de classificação comum, baseada em terminologia padronizada, conceitos, definições, métodos de tabulação e classificação. Para Hoed (2016) a aplicação de padrões de classificação como o da ISCED é relevante porque permite comparar dados estatísticos nacionais com os internacionais.

Dessa forma, o INEP (2017; INEP, 2018; 2019; 2020; INEP, 2021) distribui os 5075 cursos superiores em 11 áreas gerais, 35 áreas específicas e 84 áreas detalhadas, conforme a taxonomia apresentada no *Error! Reference source not found.*. Contudo, existem cursos de mesmo nome registrados com códigos diferentes no banco de dados. Sendo assim, o número efetivo de cursos é de 1462. Neste estudo, serão abordadas análises sobre evasão baseadas nas áreas gerais do conhecimento, além de outros parâmetros sociodemográficos e acadêmicos, como gênero, idade, política de reserva de vagas, região geográfica, turno e concorrência.

Quadro 1- Classificação das áreas gerais, áreas específicas e áreas detalhadas, conforme Classificação Internacional Normalizada da Educação CINE/UNESCO

Áreas gerais, específicas e detalhadas, conforme Classificação Internacional Normalizada da Educação CINE/UNESCO	**Número de cursos**
Engenharia, produção e construção	**788**
Arquitetura e construção	**151**
Arquitetura e planejamento urbano	57
Engenharia civil e construção	94
Engenharia e profissões correlatas	**431**
Eletricidade e energia	84
Eletrônica e automação	105
Engenharia mecânica e metalurgia	83
Engenharia química e de processos	61
Tecnologia de proteção ambiental	82
Veículos a motor, construção naval, aeronáutica, ferroviária e metroviária	15
Engenharia e profissões correlatas sem definição precisa	1
Produção e processamento	**202**
Materiais	29
Mineração e extração	30
Processamento de alimentos	61
Produção e processos de fabricação	79
Têxteis (vestuário, calçados e couro)	3
Programas interdisciplinares abrangendo engenharia, produção e construção	**4**
Programas interdisciplinares abrangendo engenharia, produção e construção	4
Negócios, administração e direito	**358**
Direito	**73**
Direito	73

Negócios e administração	**283**
Contabilidade e tributação	70
Gestão e administração	159
Gestão comercial	3
Secretariado e trabalhos de escritório	11
Marketing e propaganda	37
Finanças, bancos e seguros	3
Programas interdisciplinares abrangendo negócios, administração e direito	**2**
Programas interdisciplinares abrangendo negócios, administração e direito	2
Ciências sociais, comunicação e informação	**406**
Ciências sociais e comportamentais	**270**
Economia	64
Ciências sociais e políticas	150
Psicologia	56
Comunicação e informação	**122**
Ciência da informação e museologia	58
Comunicação e reportagem	64
Programas interdisciplinares abrangendo ciências sociais, comunicação e informação	**14**
Programas interdisciplinares abrangendo ciências sociais, comunicação e informação	14
Educação	**1463**
Educação	**1441**
Formação de professores sem áreas específicas	159
Formação de professores em áreas específicas (exceto Letras)	966
Formação de professores de letras	314
Formação de professores de educação infantil	1
Ciência da educação	1
Programas interdisciplinares abrangendo educação	**22**
Programas interdisciplinares abrangendo educação	22
Saúde e bem-estar	**490**
Bem-estar	**43**
Serviço social	42
Assistência a idosos e a deficientes	1
Saúde	**438**
Medicina	81
Enfermagem	74
Promoção, prevenção, terapia e reabilitação	159
Odontologia	32
Farmácia	50
Tecnologia de diagnóstico e tratamento médico	26
Medicina e terapia tradicional e complementar	2
Saúde pública e saúde coletiva	14
Programas interdisciplinares abrangendo saúde e bem-estar	**9**
Programas interdisciplinares abrangendo saúde e bem-estar	9
Agricultura, silvicultura, pesca e veterinária	**299**
Agricultura	**176**
Produção agrícola, agropecuária e zootecnia	176
Silvicultura	**43**
Silvicultura	43
Veterinária	**47**
Veterinária	47
Pesca	**28**
Pesca	28
Programas interdisciplinares abrangendo agricultura, silvicultura, pesca e veterinária	**5**
Programas interdisciplinares abrangendo agricultura, silvicultura, pesca e veterinária	5
Ciências naturais, matemática e estatística	**443**
Ciências físicas	**180**
Ciências da terra	55

Química	76
Física	49
Matemática e estatística	**89**
Estatística	36
Matemática	53
Ciências biológicas e correlatas	**110**
Biologia	81
Bioquímica e biotecnologia	29
Meio ambiente	**15**
Ciências ambientais	15
Programas interdisciplinares abrangendo ciências naturais, matemática e estatística	**49**
Programas interdisciplinares abrangendo ciências naturais, matemática e estatística	49
Computação e Tecnologias da Informação e Comunicação (TIC)	**177**
Computação e Tecnologias da Informação e Comunicação (TIC)	**168**
Ciência da computação	69
Desenvolvimento de sistemas que integram software e hardware	13
Gestão e desenvolvimento de sistemas de informação	69
Produção de software	11
Infraestrutura e gestão de TIC	3
Soluções computacionais para domínios específicos	3
Programas interdisciplinares abrangendo computação e Tecnologias da Informação e Comunicação (TIC)	**9**
Programas interdisciplinares abrangendo computação e Tecnologias da Informação e Comunicação (TIC)	9
Artes e humanidades	**403**
Artes	**232**
Produção audiovisual, de mídia e cultural	40
Moda, design de interiores e desenho industrial	40
Belas artes	31
Música e artes cênicas	120
Artesanato	1
Humanidades (exceto línguas)	**87**
História e arqueologia	54
Filosofia e ética	31
Religião e teologia	2
Programas interdisciplinares abrangendo artes e humanidades	**15**
Programas interdisciplinares abrangendo artes e humanidades	15
Línguas	**69**
Letras	69
Serviços	**53**
Serviços pessoais	**49**
Serviços domésticos	4
Turismo e hotelaria	37
Serviços de alimentação	7
Esportes e lazer	1
Serviços de segurança	**2**
Proteção de pessoas e de propriedades	2
Serviços de transporte	**2**
Serviços de transporte	2
Programas básicos	**195**
Programas básicos	**195**
Programas básicos	195
Total Geral	**5075**

Fonte: elaborado pela autora (2025).

3.2 IMPORTAÇÃO DOS DADOS

Para realizar o carregamento dos dados no Stata®, as pastas foram salvas no Disco Local (C:) do computador, dentro de uma pasta criada com o nome “INEP” e cada arquivo foi renomeado, em letras minúsculas, com o nome dos dados e o ano correspondente, exemplo: “cursos2017”, “cursos2018”, “ies2017”, “ies2018”, e assim sucessivamente. Esse processo é importante porque os comandos iniciais do Do-file que foi construído para o processamento dos dados, e que será disponibilizado para trabalhos futuros, utilizam esse caminho e essa nomenclatura para importar os dados. Dessa forma, qualquer pesquisador que se interesse em verificar os algoritmos ou utilizá-los em pesquisas semelhantes, possam fazer o download dos dados no site do INEP e executar o programa de forma automática.

Foi utilizado o comando *import* do Stata® para carregar os dados salvos conforme especificado anteriormente e o comando *append* para anexar, em um mesmo arquivo, os dados ano a ano. Criando assim, 2 arquivos novos salvos automaticamente na pasta “INEP”: *dadoscursos.dta* e *dadosies.dta*. O comando *append* no Stata® serve para anexar conjuntos de dados no formato Stata® (.dta) armazenados no disco, ao final do conjunto de dados que está atualmente na memória. Caso seja especificado algum nome de arquivo sem uma extensão, o Stata® assume automaticamente a extensão .dta.

Para juntar os 2 arquivos anteriores em um único arquivo que correlaciona os dados de uma planilha com os de outra, sem perder a correspondência correta entre as variáveis, foi utilizado o comando *merge*. O comando *merge* do Stata® serve para juntar observações correspondentes do conjunto de dados atualmente na memória, chamado de *master dataset*, com aquelas de um arquivo externo, chamado de *using dataset*, realizando a correspondência com base em uma ou mais variáveis-chave (StataCorp, 2025). O propósito do *merge* é adicionar novas variáveis de um segundo conjunto de dados às observações existentes no conjunto de dados que está na memória.

Foi verificado nos bancos de dados que havia 5 variáveis comuns aos 2 arquivos (cursos e IES), conforme

Quadro ***2***:

Quadro 2- Variáveis comuns ao banco de dados sobre os cursos e ao banco de dados sobre as IES

Variável	Descrição
nu_ano_censo	Ano de referência do Censo da Educação Superior
tp_organizacao_academica	Tipo de Organização Acadêmica da IES (universidade, faculdade, IF, CEFET)
tp_rede	Rede de Ensino (pública ou privada)
tp_categoria_administrativa	Tipo de Categoria Administrativa da IES (federal, estadual, municipal, privada)
co_ies	Código único de identificação da IES

Fonte: elaborado pela autora (2025).

As variáveis-chave não podem ser do tipo strL (strings longas) (StataCorp, 2025). Por essa razão, as variáveis "co_ies" e "nu_ano_censo" (que posteriormente vêm a ser renomeadas para "codies" e "ano", respectivamente) foram as variáveis utilizadas como variáveis-chave para o programa identificar e correlacionar todos os dados de ambas as planilhas a partir dos códigos de identificação das instituições e os respectivos anos do censo.

O comando *merge* pode realizar junções de diferentes tipos, como *one-to-one* (um para um), quando há apenas uma variável-chave no banco *master* para uma variável-chave no banco *using*; *one-to-many* (um para muitos), quando há apenas uma variável-chave no banco *master* para muitas no banco *using*; *many-to-one* (muitos para um), quando há muitas variáveis-chave no banco *master* para uma no banco *using* e; *many-to-many* (muitos para muitos), quando há muitas variáveis-chave no banco *master* para muitas no banco *using* (StataCorp, 2025). Neste estudo foi utilizada a modalidade *many-to-one* (muitos para um). Dessa forma a mescla reconhecerá 7 linhas de uma mesma universidade ("codies") para cada curso, ou seja, uma para cada ano, de 2017 a 2023 ("ano").

Por padrão, o comando *merge* cria uma nova variável chamada "merge", que contém códigos numéricos que indicam a origem e o conteúdo de cada observação no conjunto de dados combinado. Esses códigos são explicados em uma tabela de resultados de correspondência (*matched*). Nesse estudo houve 100% de correspondência, pois os 2.589.935 códigos das IES (não corresponde ao número de IES, pois os códigos de uma IES aparecem várias vezes) que existem no banco de dados dos cursos também existem igualmente no banco de dados das instituições. Portanto, pode-se afirmar que aconteceu a correspondência correta.

3.3 ORGANIZAÇÃO DAS VARIÁVEIS

As variáveis utilizadas neste trabalho foram renomeadas em um Do-file à parte. Esse passo é importante para facilitar a identificação da variável reduzindo seu tamanho e simplificando sua nomenclatura. Além disso foi atribuído um *label* para melhor identificação de cada uma delas.

O arquivo é carregado no arquivo principal a partir do comando *do*. A proposta é diminuir o tamanho do arquivo do Do-file principal, a fim de melhorar a organização e a visualização dos comandos de cada procedimento.

O estudo tem como propósito avaliar a evasão nos cursos presenciais de graduação das universidades federais brasileiras. Portanto, foi necessário selecionar apenas as observações relacionadas a esse público aplicando filtro por meio do comando *keep if* no Stata®, excluindo os cursos à distância, os cursos tecnólogos, as instituições privadas, as instituições estaduais e municipais e as instituições do tipo Centro Universitário, Faculdade, Instituto Federal de Educação, Ciência e Tecnologia e Centro Federal de Educação Tecnológica.

Foi necessário preparar os dados eliminando variáveis impertinentes ao estudo proposto a fim de facilitar a leitura, interpretação e processamento. A priori havia 200 variáveis sobre os cursos superiores e 81 variáveis sobre as IES. Contudo, algumas variáveis não foram utilizadas porque o estudo proposto não avalia a evasão acadêmica sobre sua ótica. Com o uso do comando *drop* no Stata® as variáveis do Quadro 3 foram excluídas:

Quadro 3- Variáveis excluídas por não haver relação com a proposta do estudo

Variável	**Justificativa**
tp_dimensao	Não se aplica porque no passo anterior já foi feito o filtro para se utilizar apenas cursos presenciais no Brasil.
co_local_oferta	
no_local_oferta	
in_comunitaria	O trabalho não engloba instituições privadas comunitárias ou confessionais porque trata-se de estudo apenas com IES públicas.
in_confessional	
no_cine_rotulo	Para os cursos de graduação, foi utilizada a nomenclatura adotada pelas IES[i].
co_cine_rotulo	
co_cine_rotulo2	
in_gratuito	O estudo trabalha apenas com IES públicas (gratuitas).
qt_vg_total_ead	O estudo trabalha apenas com cursos presenciais.
qt_inscrito_total_ead	
qt_ing_nacbras	O estudo não se propõe a avaliar a evasão acadêmica separadamente para discentes estrangeiros ou brasileiros, sejam eles natos ou naturalizados.
qt_ing_nacestrang	

qt_mat_nacbras	
qt_mat_nacestrang	
qt_conc_nacbras	
qt_conc_nacestrang	
qt_ing_financ	O trabalho não analisa a evasão acadêmica em universidades particulares.
qt_ing_financ_reemb	
qt_ing_fies	
qt_ing_rpfies	
qt_ing_financ_reemb_outros	
qt_ing_financ_nreemb	
qt_ing_prounii	
qt_ing_prounip	
qt_ing_nrpfies	
qt_ing_financ_nreemb_outros	
qt_mat_financ	
qt_mat_financ_reemb	
qt_mat_fies	
qt_mat_rpfies	
qt_mat_financ_reemb_outros	
qt_mat_financ_nreemb	
qt_mat_prounii	
qt_mat_prounip	
qt_mat_nrpfies	
qt_mat_financ_nreemb_outros	
qt_conc_financ	
qt_conc_financ_reemb	
qt_conc_fies	
qt_conc_rpfies	
qt_conc_financ_reemb_outros	
qt_conc_financ_nreemb	
qt_conc_prounii	
qt_conc_prounip	
qt_conc_nrpfies	
qt_conc_financ_nreemb_outros	
co_projeto	
qt_parfor	O trabalho não analisa separadamente discentes que participam do programa especial para a formação de professores em exercício na rede pública de educação básica por meio do Plano Nacional de Formação de Professores da Educação Básica (PARFOR).
qt_ing_parfor	
qt_mat_parfor	
qt_conc_parfor	

no_mesorregiao_ies	O estudo analisa as IES, do ponto de vista geográfico, apenas pela região[ii].
co_mesorregiao_ies	
no_microrregiao_ies	
co_microrregiao_ies	
no_mantenedora	Dados das mantenedoras e de contato das IES não foram utilizados nos estudos de evasão acadêmica[iii].
co_mantenedora	
ds_endereco_ies	
ds_numero_endereco_ies	
ds_complemento_endereco_ies	
no_bairro_ies	
nu_cep_ies	
qt_tec_fundamental_incomp_fem	O estudo não faz relação da evasão com quadro de servidores técnicos das IES quanto às suas características de gênero e formação.
qt_tec_fundamental_incomp_masc	
qt_tec_fundamental_comp_fem	
qt_tec_fundamental_comp_masc	
qt_tec_medio_fem	
qt_tec_medio_masc	
qt_tec_superior_fem	
qt_tec_superior_masc	
qt_tec_especializacao_fem	
qt_tec_especializacao_masc	
qt_tec_mestrado_fem	
qt_tec_mestrado_masc	
qt_tec_doutorado_fem	
qt_tec_doutorado_masc	
in_acesso_portal_capes	O estudo não utiliza essas variáveis porque entende-se que o acesso a essas bases de dados e periódicos se aplicam mais aos discentes de pós-graduação.
in_acesso_outras_bases	
in_assina_outra_base	
in_repositorio_institucional	
in_busca_integrada	
qt_periodico_eletronico	
qt_doc_ex_preta	O estudo não avalia a relação da evasão com a raça ou cor do quadro de docentes.
qt_doc_ex_parda	
qt_doc_ex_amarela	
qt_doc_ex_indigena	
qt_doc_ex_cor_nd	
qt_doc_ex_bra	O estudo não avalia a relação da evasão com a nacionalidade do quadro de docentes.
qt_doc_ex_est	
qt_doc_ex_com_deficiencia	O estudo não avalia a relação da evasão com docentes com deficiência.

Fonte: elaborado pela autora (2025).

Além disso, para as variáveis relacionadas aos matriculados foram criadas *proxies* como indicadores da proporção em relação ao todo. Exemplo: em vez de se analisar a variável da quantidade de matriculadas do sexo feminino, foi-se utilizada a taxa de matriculadas do sexo feminino em relação ao total de discentes matriculados (propfem = qtmatfem / qtmat). Isso porque a variável dependente "eva" é apresentada em percentual. Portanto, para facilitar a sua interação e interpretação em relação às variáveis independentes, estas últimas foram convertidas de valores absolutos para valores também percentuais. Dessa forma, para cada ponto percentual aumentado ou diminuído nas taxas calculadas, há uma associação de "X" pontos percentuais alterados na evasão, de acordo com o resultado da regressão.

3.4 CÁLCULO DA EVASÃO E DA CONCORRÊNCIA

3.4.1 Cálculo da Evasão

Utilizando os dados do INEP e suas terminologias, algumas definições se fazem pertinentes: candidatos inscritos se referem ao número de classificados para determinado curso, considerando que possam constar em uma lista de espera. Por outro lado, os ingressantes se referem àqueles efetivamente cadastrados para o curso naquele ano, dentro do número de vagas disponibilizadas. Nesse ponto o candidato passa a ser discente. Os discentes matriculados são aqueles cadastrados acumulados de anos anteriores, comumente entendido por alunos "ativos".

Serão comparados índices de evasão em 6 anos: 2018, 2019, 2020, 2021, 2022 e 2023. O ano de 2017 serve apenas para subsidiar o cálculo da evasão em 2018, não sendo possível calcular o índice nesse ano porque isso dependeria dos dados do ano de 2016, que não pertence ao escopo. Por essa razão, após o cálculo da evasão, os resultados para o ano de 2017 foram excluídos.

Silva Filho *et al* (2007) apresentam uma situação específica em que o curso de Formação de Professor da Educação Básica/Normal Superior no ano de 2001 atingiu taxa de evasão correspondente a - 46%. Explicam que o valor atípico (negativo) se deve ao aumento da busca por esse curso em função da expansão da oferta entre 2000 e 2001, por conta de uma exigência da Lei de Diretrizes e Bases da Educação - LDB nº 9.394/1996, e que pode ter atraído discentes

já formados em outros cursos ou que mudaram de curso, e que não foram considerados como ingressantes, mas apenas como matriculados Silva Filho (2007). Portanto, esse valor negativo não condiz com a realidade.

Hoed (2016) explica ainda que em se tratando de dados agregados, fica difícil, em certas circunstâncias, obter uma precisão no cálculo da evasão. Taxas mais coerentes de evasão poderiam ser obtidas se os dados disponibilizados pelo INEP fossem individualizados, com a condição de matrícula informada para cada discente em cada curso. Ou seja, se está cursando, se é concluinte, se foi desvinculado, se foi transferido etc.

No caso deste trabalho, o índice de evasão calculado para alguns cursos em alguns anos resultou em valores abaixo de zero ou acima de 1. Acredita-se que essa discrepância possa ter ocorrido em função de algum problema no fornecimento de dados pelas instituições ou por alguma situação semelhante ao caso anterior.

Em uma série de dados de índices de evasão acadêmica, valores de percentual de evasão maiores que 100% (1 em formato decimal) ou menores que 0% (valores negativos) são inválidos ou impossíveis por razões lógicas e estatísticas. O percentual de evasão representa a proporção de alunos que abandonaram a graduação em relação ao total de alunos matriculados. No caso desse estudo, é feita uma comparação do total líquido de alunos de um ano (os matriculados excluídos os ingressantes), com o total líquido de alunos do ano anterior (os matriculados excluídos os concluintes). A diferença resulta exatamente no número de evadidos. Um percentual de evasão superior a 100% significaria que o número de alunos que evadiram é maior do que o total de alunos matriculados. Isso não faz sentido, pois é impossível ter mais alunos evadidos do que o número total de alunos na universidade.

Por outro lado, o percentual de evasão também não pode ser negativo, pois isso sugeriria que houve algum tipo de "evasão negativa", ou seja, mais alunos retornaram do que saíram. Isso também é logicamente impossível.

O melhor cenário ou tido como o ideal é aquele em que esse índice tem valor zero, pois logicamente explicaria que o quantitativo total líquido (excluindo os ingressantes) de um ano é exatamente igual ao quantitativo total líquido (excluindo os concluintes) do ano anterior. Ou seja, exceto os discentes que concluíram, ninguém mais saiu.

Para tratar esses valores, é comum excluir ou corrigir as observações com valores fora do intervalo esperado (0 a 1 ou 0% a 100%). Dessa forma, os valores negativos ou maiores que 1 foram excluídos.

Todavia, para se realizar a comparação dos índices de evasão proposta, ano a ano, é necessário que se tenha dados em todos os anos. Sendo assim, aqueles cursos que apresentaram situação como descrito anteriormente precisaram ser totalmente excluídos da amostra, e não apenas excluído aquele ano em que a evasão foi maior que 1 ou menor que zero. Para isso foram utilizados os comandos *egen tag=count(ano), by(codcurso), keep if tag==6*, de modo que o algoritmo fizesse o rastreio de cada ano para cada curso, se o número encontrado fosse diferente de 6, ou seja, não tivesse sido encontrado índice de evasão para os 6 anos propostos (2018 a 2023), esse curso seria inteiramente retirado da amostra.

3.4.2 Cálculo da Concorrência

Foi criada variável *proxy* para a concorrência. Como explicitado nos conceitos, o número de inscritos difere do número de ingressantes. Observa-se que o número de ingressantes, por via de regra, é o mesmo número de vagas, exceto quando não há preenchimento total delas. Nesse caso, há sobra de vagas. Logo, o número de ingressantes sempre será menor ou igual ao número de vagas.

O número de inscritos, por sua vez, se refere ao número de candidatos que se inscreveram no processo seletivo. Ou seja, o número de inscritos é, geralmente, superior ao número de ingressantes, visto que pode haver candidatos para além do número de vagas. Por essa razão, o cálculo de concorrência (4) se baseia na divisão de inscritos por vagas, pois representa uma relação de procura e oferta.

$$C = \left[\frac{I}{V}\right] \quad (4)$$

Onde:

C = Concorrência;

I = Número de inscritos;

V = Número de vagas;

Após o cálculo da concorrência foi feita a exclusão das observações maiores que 500 (relação candidato/vaga>500), pois foram consideradas valores atípicos (*outliers*).

3.5 REGRESSÃO COM DADOS EM PAINEL

De acordo com Gujarati (2008), os dados utilizados em uma análise estatística podem ser classificados em 3 tipos: a) séries temporais, cujos valores são observáveis ao longo de um período; b) cortes transversais, onde as variáveis em um mesmo lapso temporal são apresentadas para diferentes entidades amostrais e; c) painéis, onde as variáveis são apresentadas para diferentes entidades ao longo do tempo, como se mostra neste trabalho, onde serão observados dados de diversos cursos de graduação, de diversas instituições de ensino superior, em diferentes momentos do tempo.

Serão analisados os dados sobre evasão, concorrência, percentual de discentes matriculados por gênero, por idade, por turno e por reserva de vagas em ações afirmativas; além de dados sobre as áreas gerais do conhecimento às quais estão contidos os cursos de graduação e as 5 regiões geográficas do Brasil, variando nos períodos pré, pós e durante a pandemia de COVID-19. De acordo com Gujarati (2008), essa variação de cortes transversais no tempo constitui-se, essencialmente, os dados em painel.

Os dados em painel também podem ser classificados como dados empilhados (*pooled*), combinação de séries temporais e dados de corte transversal, painel de microdados, dados longitudinais ou análise histórica de eventos (Gujarati, 2008). Todas essas nomenclaturas podem ser encontradas nesse trabalho.

A análise de dados em painel justifica-se devido a algumas vantagens em relação aos demais modelos, conforme explica Gujarati (2008):

- as técnicas de estimação dos dados em painel podem levar em consideração a heterogeneidade, devido à tendência da sua presença entre os dados, por se tratar de indivíduos que podem sofrer influências diversas;
- os dados em painel possuem maior variabilidade, menos colinearidade entre variáveis e mais graus de liberdade;

- os dados em painel são mais adequados para examinar a dinâmica da mudança das observações de corte transversal,
- os dados em painel podem detectar e medir melhor os efeitos não observáveis em um corte transversal puro ou em uma série temporal pura.

Em relação à heterogeneidade, esclarece Hsiao (2014), que se os dados disponíveis fossem gerados por experimentos controlados simples, poderiam ser aplicados métodos estatísticos padrão. Isso significa que para especificar a forma completa de uma distribuição de probabilidade bastaria indicar um vetor de poucos parâmetros. Uma vez que esses parâmetros estão fixados, toda a distribuição ficaria determinada. Ou seja, seria modelada por um número finito de parâmetros (Wasserman, 2004). Entretanto, a maioria dos dados em painel provém de processos complexos do cotidiano. Via de regra, os indivíduos podem estar sujeitos à influência de vários fatores distintos àqueles analisados. Dessa forma, quando fatores importantes peculiares a um dado indivíduo são omitidos, a suposição de que a variável gerada é dada por uma função de distribuição de probabilidade paramétrica pode não ser realista (Hsiao, 2014).

De acordo com Hsiao (2014) e Baltagi (2005), os estudos com séries temporais ou com dados de corte transversal são frequentemente afetados por multicolinearidade e por problemas devido à escassez de graus de liberdade. Isso significa que a informação fornecida pela amostra não é suficientemente rica para atender às exigências do modelo especificado. Geralmente, é necessário impor restrições à pesquisa ou buscar aumentar a informação da amostra (Hsiao, 2014). Para isso os dados em painel se mostram mais eficientes. Além disso, eles contêm informações tanto sobre diferenças entre unidades de corte transversal quanto sobre dinâmicas dentro delas, ao longo do tempo, o que pode aumentar substancialmente a informação da amostra (Hsiao, 2014).

Baltagi (2005) e Gujarati (2008) afirmam, ainda, que os dados em painel podem levar em consideração os efeitos não diretamente observáveis. Ou seja, no modelo, mesmo que uma variável não seja observada (ou conhecida), é possível mensurá-la por um efeito, chamado efeito não observado. Uma das formas de examinar a confiabilidade de uma medida é mensurá-la várias vezes ao longo do tempo. Logo, os dados em painel ajudam a superar eventuais erros de mensuração (Mesquita, Fernandes e Filho, 2020).

3.5.1 Estimação do modelo

Para estimar o modelo de regressão com dados em painel mais adequado à pesquisa, são realizados testes e análises de acordo com outros estudos sobre o modelo para respostas fracionárias. A princípio, considerando a possibilidade de se utilizar modelos convencionais, são aplicados os testes: teste do modelo *Pooled*, teste do modelo dos Efeitos Aleatórios, teste de Breusch-Pagan para comparar os dois modelos anteriores por meio da detecção da presença de efeitos aleatórios, teste de Wooldrigde para verificar se há correlação entre os erros e o teste do modelo Robusto Clusterizado usando Mínimos Quadrados Generalizados Factíveis (FGLS) para heterocedasticidade e a autocorrelação. O teste de Efeitos Fixos não é aplicado porque esse modelo não permite analisar variáveis que não variam no tempo, como é o caso deste estudo (Gujarati e Poter, 2008).

No entanto, Wooldridge (2013) explica que modelos de efeitos aleatórios convencionais, formulados no contexto de regressões lineares para dados em painel, não são adequados para variáveis resposta fracionárias porque a estrutura linear desses modelos não impõe restrições para que as previsões permaneçam dentro do intervalo admissível, podendo gerar valores previstos negativos ou superiores a 1, o que é incoerente do ponto de vista substantivo.

Além disso, o uso direto de modelos lineares também ignora a necessidade de funções de ligação não lineares, como o “logit” ou o “probit”, que garantem que a resposta estimada se mantenha nos limites da escala. Em dados de painel, embora os efeitos aleatórios tradicionais tentem capturar a correlação intra-indivíduo por meio de um termo de erro específico de cada unidade, essa abordagem não garante estimadores consistentes quando a variável dependente é fracionária e a relação com as covariáveis é modelada de forma não linear. Por essas razões, a literatura recomenda abordagens específicas.

Dessa forma, estudos empíricos tentando explicar respostas fracionárias se multiplicaram nos últimos anos (Papke e Wooldridge, 2008). Papke e Wooldridge (1996) propuseram modelos diretos para a média condicional da resposta fracionária que mantêm os valores previstos dentro do intervalo unitário, como é o caso da variável dependente “evasão”. Os autores forneceram um código para o Stata® e ficou denominado como método de estimação por quase-verossimilhança com função média logística como “fractional logit” (Papke e Wooldridge, 2008).

Entretanto, posteriormente surge o questionamento sobre como lidar com a heterogeneidade não observada que pode estar correlacionada com as variáveis explicativas. Para isso Papke e Wooldridge (2008) propuseram o modelo dos mínimos quadrados não lineares ponderados multivariados (MWNLS) para estimar médias condicionais em dados em painel com regressores estritamente exógenos na presença de correlação serial e heterocedasticidade, que é semelhante à abordagem das equações de estimação generalizadas (GEE) de Liang e Zeger (1986), que permitem modelar diretamente a natureza restrita e a estrutura de variância da variável resposta a partir do uso de função de ligação apropriada. Papke e Wooldridge (2008) entendem que GEE e MWNLS são assintoticamente equivalentes. A equação de estimação de GEE (5) é dada por meio de:

$$\sum_{i=1}^{K} D_i^T V_i^{-1} S_i = 0 \quad (5)$$

Onde:

S_i é o vetor de resíduos, dado em função do vetor de respostas Y_i e o vetor de médias $\mu_i(\beta)$, modelado pela função de ligação $g(.)$;

D_i é a matriz das derivadas da média em relação aos parâmetros;

T significa que se utiliza a transposta da matriz D;

K é o número de grupos;

V_i é a matriz de covariância "trabalhada" formada pela matriz de variâncias A_i e pela matriz de correlação $R(\alpha)$ (Liang e Zeger, 1986).

A solução dessa equação fornece os coeficientes estimados para cada variável explicativa (β) levando em conta a correlação entre observações dentro de cada grupo para obter estimativas consistentes e erros padrão robustos.

Para dados em painel, Wooldridge (2011) afirma que é possível utilizar o comando xtgee no Stata®. Papke e Wooldridge (2008) consideram, ainda, que é possível utilizar abordagem com efeitos aleatórios correlacionados (CRE) para estimar modelos de resposta fracionária em dados em painel para painéis balanceados, como é o caso deste estudo.

Bendig *et al*. (2025) utilizam em seu estudo com variável dependente fracionária, o modelo GEE com distribuição binomial e função de ligação "probit". Usam estrutura de correlação "exchangeable" e erros robustos de Huber–White para reduzir heterocedasticidade.

Para mitigar problemas de endogeneidade, usam o método de função de controle (control function approach), adequado para modelos não lineares com interações.

Gujarati (2008) afirma que a aplicação do modelo “logit” ou “probit” é muito semelhante e define mais uma preferência do pesquisador. Em regra, os resultados tendem a ser semelhantes, mas o “logit” é melhor interpretável devido à função logística. Além disso, ele também apresenta caudas mais espessas, ou seja, se aproxima de 0 ou 1 em um ritmo mais lento, o que pode ter um ajustamento melhor.

Em relação à correlação, a opção “exchangeable” adotada por Bendig *et al.* (2025) assumiria que, no caso deste estudo, as medidas anuais seriam parecidas entre si. Mas já se pôde perceber que a pandemia interferiu de alguma forma na evasão. Então há evidência de uma estrutura temporal definida. É possível que haja uma autocorrelação ano a ano “ar1”, ou de dois em dois anos “ar2” devido ao agrupamento de 2 anos antes da pandemia, 2 anos durante e 2 anos após. A opção “unstructured” estima todas as correlações possíveis, mas apresenta uma opção mais genérica.

Outro problema atrelado ao uso de modelos com dados em painel é a heterocedasticidade (Gujarati e Poter, 2008). De acordo com Hsiao (2014) e Wooldridge (2013), a heterocedasticidade ocorre quando a variância dos erros em um modelo de regressão não é constante ao longo das observações (pode ocorrer entre indivíduos ou ao longo do tempo). Quer dizer que os resíduos têm variâncias diferentes para diferentes valores de x, o que viola uma das suposições fundamentais da regressão linear clássica (homocedasticidade). A homocedasticidade falha sempre que a variância dos fatores não observados varia entre diferentes segmentos da população (Wooldridge, 2013). Isso afeta a eficiência, pois, segundo Wooldridge (2013) e Mesquita *et al*. (2020) os erros-padrão estarão incorretos e os testes t e F podem fazer inferências enganosas.

Neste trabalho considera-se também os erros de Huber-White aplicados por Bendig *et al.* (2025), pois, conforme o Manual do Stata® (StataCorp, 2025), para o comando xtgee, o modelo permanece válido (erros padrão válidos) mesmo se os pressupostos em relação à estrutura de correlação aplicadas anteriormente estiverem errados. Mas se estiverem certos, o modelo fica mais bem ajustado. Ou seja, os coeficientes serão mais precisos.

É relevante, ainda, considerar a possibilidade de endogeneidade de algumas variáveis explicativas neste estudo, ou seja, considerar que elas podem estar correlacionadas com

variáveis não observadas que variam ao longo do tempo, conforme considerou Bendig *et al*. (2025). A endogeneidade decorre da correlação entre o erro que pode estar correlacionado tanto com a heterogeneidade não observada quanto com o fator omitido que varia no tempo (Papke e Wooldridge, 2008).

Virgillito (2020) explica que em um modelo de regressão simples supõe-se que o valor médio esperado (esperança matemática) do termo erro deve ser sempre zero, ou seja, o termo de erro u_i não mantém correlação com os regressores (os regressores são exógenos). Contudo, se uma variável independente for uma variável correlacionada com o termo erro será dito que este é um regressor endógeno.

A regressão em dois estágios, também chamada de Two-Stage Least Squares (2SLS), é uma técnica econométrica usada principalmente quando há endogeneidade entre as variáveis explicativas. Neste trabalho foi considerada a possibilidade de endogeneidade nas seguintes variáveis: lnconc (concorrência), propescopub (RV escola pública), propetn (RV etnia), proppdef (RV deficiência), proprenda (RV renda familiar) e propoutros (RV outras opções). Pressupondo que podem estar correlacionadas com o erro da regressão por algum fator, por exemplo, por haver variável omitida que afeta tanto a explicativa quanto a dependente, ou, por haver causalidade reversa (nesse caso, é evidente na concorrência) etc. Pode-se suspeitar também que políticas de reserva de vagas por ações afirmativas podem influenciar a evasão por pressão de desempenho, por exemplo.

Conforme Gujarati (2008) e Lin e Wooldridge (2019), esta técnica foi aplicada seguindo os seguintes estágios:

1º: Foi feita a regressão (xtreg) de cada variável supostamente endógena em relação às demais exógenas e aos instrumentos (nesse caso foram utilizadas as médias das próprias endógenas como instrumento para prever uma nova variável endógena - resíduo);

2º: Foram incluídos os valores previstos das variáveis endógenas obtidos no estágio 1 (resíduos) na versão da regressão principal da evasão (xtgee) para obter estimativas consistentes.

Considerando a possibilidade de endogeneidade, a aplicação da regressão em 2 estágios foi realizada, a priori, utilizando como instrumentos as defasagens das variáveis supostamente endógenas (L.var). Porém, quando o instrumento adotado é a defasagem da própria variável, o

painel perde a primeira observação de cada unidade porque não se pode estimar a defasagem sem o dado anterior. Neste contexto, tem-se dados de 2018 a 2023. Quando se assume a defasagem como variável instrumental, o ano de 2018 não apresenta resultados porque não tem o ano de 2017 para estimá-lo. Como alternativa, foram utilizadas as médias das variáveis supostamente endógenas. Assim o instrumento pode capturar a parte exógena da variável endógena, assumindo que o efeito individual médio afeta a variável endógena, conforme aplicaram Lin e Wooldridge (2019).

3.5.2 Testes de validação do modelo

Para verificação da consistência do modelo foram levantados os seguintes diagnósticos:

I. Análise da significância das variáveis incluídas (resíduos);

II. Validação do modelo analisando os instrumentos. Essa parte consiste em um teste simulando o modelo de regressão (xtreg) das variáveis supostamente endógenas, seus instrumentos, seus resíduos e as demais variáveis explicativas; e em seguida, a aplicação do teste nos instrumentos;

III. Análise do pseudo R2;

IV. Cálculo dos resíduos Pearson. Essa ferramenta serve para visualizar padrões sistemáticos e possíveis desvios;

V. Análise da correlação entre valores preditos e reais.;

VI. Avaliação da Distribuição dos Valores Preditivos.

Gujarati (2008) afirma que, com base na teoria ou em trabalhos aplicados, é possível desenvolver um modelo capaz de captar a essência do assunto estudado. Primeiro, submete-se o modelo a testes e depois de obter os resultados, iniciam-se as análises tendo em mente os critérios de um bom modelo discutido anteriormente, examinando aspectos como o grau de explicação da variância da variável dependente pelas variáveis independentes, a significância global do modelo, os sinais dos coeficientes estimados em relação às expectativas anteriores, técnicas para detectar autocorrelação nos resíduos, entre outros. Se esses diagnósticos forem razoáveis, pode-se afirmar que o modelo escolhido é uma representação adequada da realidade (Gujarati e Poter, 2008).

Dessa forma, o roteiro descrito consiste, principalmente, na modelagem de acordo com a literatura que versa sobre sua utilização, baseando-se na semelhança das características entre os estudos e as adequações que podem ser realizadas de acordo com as particularidades desta pesquisa, além dos resultados dos diagnósticos.

Além dos testes aplicados, a estimação do Fator de Inflação da Variância (VIF – *Variance Inflation Factor*) é importante para medir o nível de multicolinearidade entre uma variável explicativa e as demais do modelo. Kim (2019) elucida que o VIF pode indicar o quanto a variância de um coeficiente está inflada devido à correlação com outras variáveis independentes no modelo, o que pode dificultar a estimativa precisa dos coeficientes de regressão e comprometer a interpretação dos resultados. O VIF fornece uma indicação razoável e intuitiva dos efeitos da multicolinearidade sobre a variância do i-ésimo coeficiente de regressão (O'Brien, 2007).

Embora não haja um valor universalmente definido como limite para a estatística VIF, usualmente quando seu valor é igual a 1 significa que não há nenhuma correlação entre a variável explicativa e as demais, ou seja, indica ausência de multicolinearidade. Quando o VIF está entre 1 e 5 indica que há correlação moderada, geralmente aceitável e, se o VIF for maior que 5 significa forte indício de multicolinearidade e merece atenção, pois pode comprometer a confiabilidade das estimativas dos coeficientes (O'Brien, 2007; Kim, 2019). Portanto, em consonância com os autores, nesse trabalho foi adotado o valor de VIF menor ou igual a 5 como uns dos critérios para seleção de variáveis com baixa multicolinearidade.

Outra ferramenta utilizada para modelagem e que sugere a seleção das variáveis é o teste de correlação de Spearman. Zar (2005) explica que o coeficiente de correlação mede quantitativamente a força da correlação entre duas variáveis, que expressa o quão perto uma mudança na magnitude de uma das variáveis é acompanhada por uma mudança na magnitude da outra variável. A correlação de Spearman é uma medida não paramétrica (exige menos suposições) que avalia a associação monotônica entre duas variáveis, isto é, se uma variável tende a aumentar (ou diminuir) à medida que a outra também o faz. É particularmente útil quando os dados não seguem uma distribuição normal ou quando a relação entre as variáveis não é linear. Além disso, é apropriado quando há presença de *outliers* (Siegel, 1957; Zar, 2005; Miot, 2018).

A seleção das variáveis utilizadas também foi baseada nos estudos teóricos e em evidências empíricas anteriores, conforme Quadro 4.

Quadro 4- Variáveis selecionadas para análise em relação à evasão

Variável	Descrição	Característica	Referências
propfem	Percentual de matriculadas do sexo feminino	Gênero	(Nierotka, Salata e Martins, 2023) (Faria, 2022) (Gomes e Hirata, 2022) (Chagas, 2019) (Cardoso e Nagai, 2019) (Hoed, 2016)
propmasc	Percentual de matriculados do sexo masculino*		
prop0a17	Percentual de matriculados com menos de 17 anos	Idade	(Lopes *et al.*, 2023) (Nierotka, Salata e Martins, 2023) (Saccaro, França e Jacinto, 2019) (Guerra, Ferraz e Medeiros, 2019)
prop18a24	Percentual de matriculados de 18 a 24 anos*		
prop25a29	Percentual de matriculados de 25 a 29 anos		
prop30a34	Percentual de matriculados de 30 a 34 anos		
prop35a39	Percentual de matriculados de 35 a 39 anos		
prop40a49	Percentual de matriculados de 40 a 49 anos		
prop50a59	Percentual de matriculados de 50 a 59 anos		
prop60mais	Percentual de matriculados com mais de 60 anos		
propdiu	Percentual de matriculados no turno diurno*	Turno	(Guerra, Ferraz e Medeiros, 2019) (Adachi, 2009) (Dias, Theóphilo e Lopes, 2010)
propnot	Percentual de matriculados no turno noturno		
propescopub	Percentual de matriculados que participam do programa de reserva de vagas - escola pública	Política afirmativa	(Lopes *et al.*, 2023) (Faria, 2022) (Barros, 2022) (Chagas, 2019) (Mendes Junior, 2014) (Dias, Theóphilo e Lopes, 2010)
propetn	Percentual de matriculados que participam do programa de reserva de vagas - cunho étnico		
proppdef	Percentual de matriculados que participam do programa de reserva de vagas - pessoas com deficiência		
proprenda	Percentual de matriculados que participam do programa de reserva de vagas - renda familiar		
propoutros	Percentual de matriculados que participam do programa de reserva de vagas - outros tipos		
lnconc	Concorrência por curso (aplicado o método de linearização)	Demanda	(Saccaro, França e Jacinto, 2019) (Hoed, 2016) (Dias, Theóphilo e Lopes, 2010)
dg1	Área geral: Agricultura, silvicultura, pesca e veterinária	Área do conhecimento	(Lopes *et al.*, 2023) (Barros, 2022) (Saccaro, França e Jacinto, 2019) (Chagas, 2019) (Hoed, 2016) (Silva Filho *et al.*, 2007)
dg2	Área geral: Artes e humanidades		
dg3	Área geral: Ciências naturais, matemática e estatística		
dg4	Área geral: Ciências sociais, comunicação e informação		
dg5	Área geral: Computação e Tecnologias da Informação e Comunicação (TIC)		
dg6	Área geral: Educação		
dg7	Área geral: Engenharia, produção e construção		
dg8	Área geral: Negócios, administração e direito		

dg9	Área geral: Programas básicos**		
dg10	Área geral: Saúde e bem-estar*		
dg11	Área geral: Serviços		
dreg1	Região: Norte	Região Geográfica	(Gomes e Hirata, 2022) (Saccaro, França e Jacinto, 2019) (Adachi, 2009) (Silva Filho *et al.*, 2007)
dreg2	Região: Nordeste		
dreg3	Região: Centro-Oeste		
dreg4	Região: Sudeste*		
dreg5	Região: Sul		
covid	Período da pandemia de COVID-19 no Brasil (2020-2021)	Período	(Barros, 2022) (Faria, 2022) (Rosa, Santos e Gonçalves, 2021)
poscovid	Período após a pandemia de COVID-19 no Brasil (2022-2023)		
antescovid	Período antes da pandemia de COVID-19 no Brasil (2018-2019)*		
***Excluída do modelo por efeito de multicolinearidade.**			
****Excluída do modelo por razões apontadas no item 4.3.1.**			

Fonte: elaborado pela autora (2025).

O quadro apresenta alguns autores que também estudaram a evasão acadêmica sob a perspectiva das características definidas para este trabalho, o que confere relevância ao que se propõe.

[i] Embora tenha sido utilizada a classificação da CINE para a nomenclatura das áreas, a nomenclatura para os cursos adotada pelas IES são mais objetivas e classificam mais claramente os cursos como Biologia, Química, Matemática e Física, por exemplo. Ex. nome do curso, segundo a nomenclatura das IES: “Ciências - Biologia e Química”; nomenclatura adotada pela CINE: “Ciências naturais formação de professor”. A notação da CINE utiliza o termo "formação de professor", para os cursos de licenciatura. Essa nomenclatura fica grande e atrapalha a visualização da variável, uma vez que no estudo proposto não há distinção de cursos de bacharelado e licenciatura. Sem contar que a nomenclatura “Ciências naturais formação de professor” também é utilizada para o curso “Ciências – Matemática e Física”, por exemplo. Seria um equívoco englobar todos esses cursos como sendo um único curso: “Ciências naturais formação de professor”.

[ii] A mesorregião divide as regiões do país em regiões menores, por exemplo, a região sudeste se divide em metropolitana de São Paulo, Campinas, Zona da Mata, Araraquara, metropolitana de Belo Horizonte etc. Ou ainda, a região nordeste pode ser dividida em Oeste Potiguar, Mata Pernambucana, Sertão Pernambucano etc. A microrregião subdivide essas últimas em regiões menores ainda, por exemplo, Oeste Potiguar em Mossoró, Mata Pernambucana em Mata Meridional Pernambucana ou até mesmo Sertão Pernambucano em Sertão do Moxotó.

[iii] As variáveis “no_mantenedora” e “co_mantenedora” foram excluídas pela possibilidade de gerar conflitos com relação à identificação das instituições. A mantenedora responde pela gestão global do patrimônio (dinheiro, estrutura, aquisições) e a mantida responde pela gestão acadêmica, portanto utilizaremos essa última. Por exemplo a Universidade Federal do Piauí (UFPI) tem como mantenedora a Fundação Universidade Federal do Piauí, sob código 14054 e como mantida a Universidade Federal do Piauí, sob código 5.

4. DISCUSSÃO E ANÁLISE DOS RESULTADOS

As variáveis analisadas podem ser classificadas em 2 grupos:

- Acadêmicas:
  - Variável relacionada à concorrência dos cursos;
  - Variáveis relacionadas às áreas gerais dos cursos;
  - Variáveis relacionadas ao turno dos cursos.

- Sociodemográficas:
  - Variáveis relacionadas ao gênero dos discentes;
  - Variáveis relacionadas à idade dos discentes;
  - Variáveis relacionadas à região geográfica do Brasil onde se encontram os cursos;
  - Variáveis relacionadas ao ingresso por meio de política de reserva de vagas.

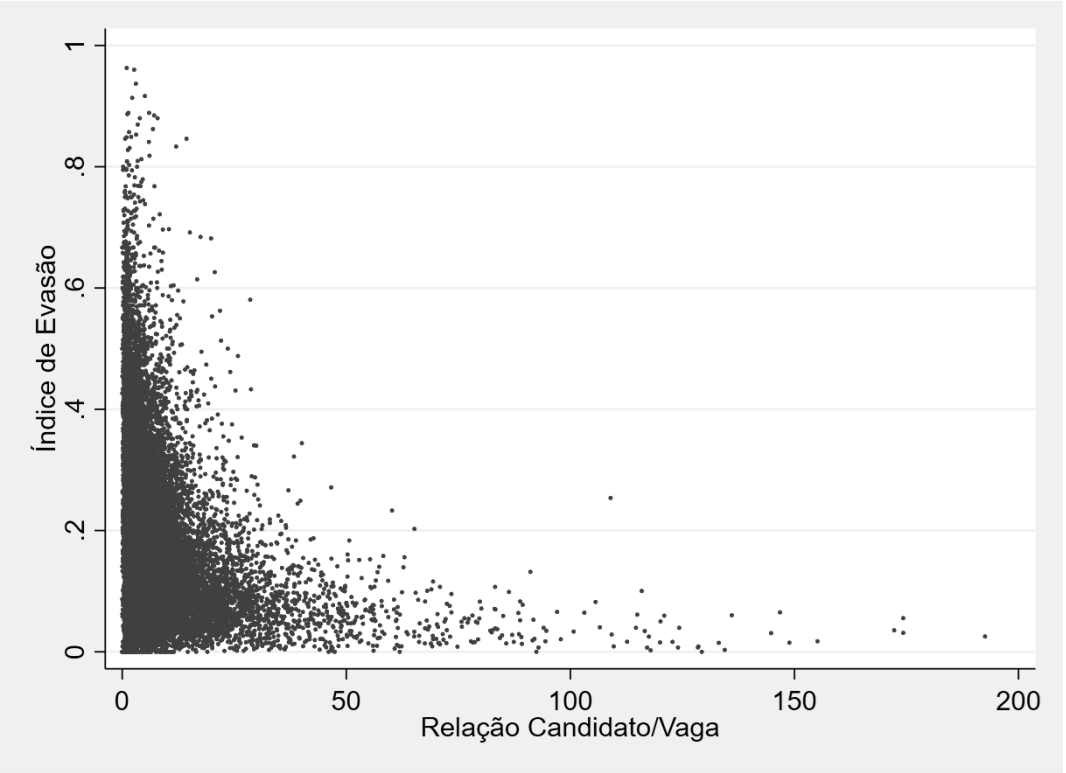
Neste capítulo são apresentadas análises de resultados estatísticos dessas variáveis trazendo suas relações com os períodos mencionados na pesquisa.

O

Gráfico *1* mostra a dispersão da evasão em relação à concorrência para os dados de 2018 a 2023. Ele indica uma tendência a uma relação inversa, onde a evasão diminui à medida que a concorrência aumenta.

Para linearizar essa relação foram aplicadas as transformações: logarítmica (ln(conc)), inversa (1/conc) e log-log (log_eva e log_conc) para verificar qual delas apresenta o método de linearização mais adequado. Os resultados após a aplicação da linearização podem ser verificados no Apêndice A.1. O coeficiente de determinação R^2 indica quanto da variável dependente (evasão) pode ser explicada pelas variáveis independentes (nesse caso, a concorrência). Ou seja, quanto maior o valor de R^2 (mais próximo de 1) maior o percentual de explicação. Por essa razão a transformada ln foi utilizada (R^2= 0.1480).

Gráfico 1- Dispersão da evasão em relação à concorrência

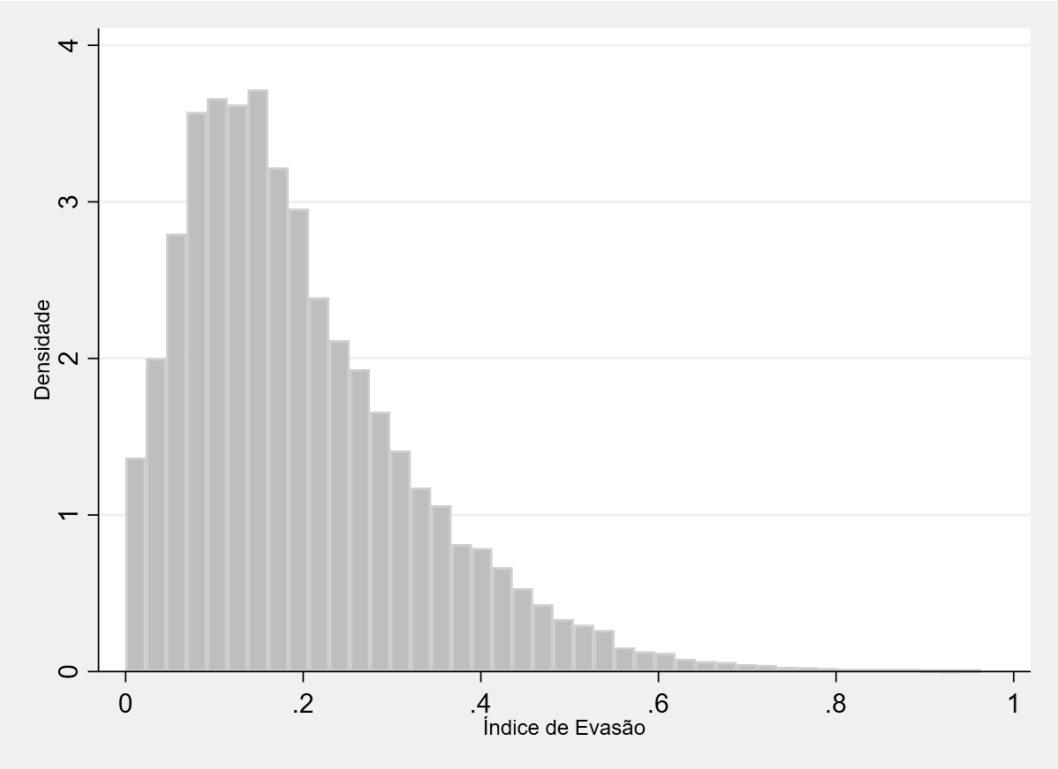


Fonte: resultados da pesquisa.

O

Gráfico ***2***, por sua vez, mostra uma distribuição assimétrica à direita da densidade de cursos em suas taxas de evasão nos anos de 2018 a 2023. A maior parte dos cursos está concentrada nos índices de evasão entre 0 e 30%. Poucas observações apresentam índices de evasão acima de 50%. O pico do histograma ocorre próximo de 15% a 20%, indicando que esse intervalo é onde está a maior densidade de evasão. Existem cursos com índice de evasão acima de 60%, mas são mais escassos.

Gráfico 2- Histograma do índice geral de evasão no período de 2018 a 2023

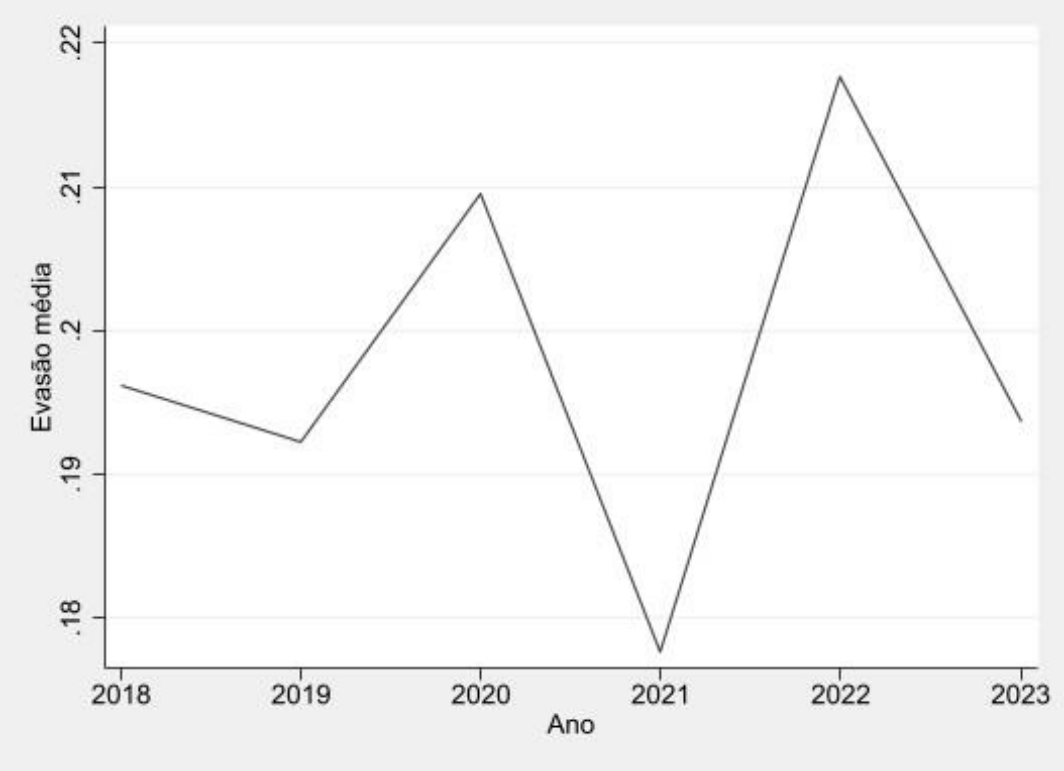


Fonte: resultados da pesquisa.

O

Gráfico *3* aponta um aumento perceptível na média da evasão acadêmica nacional no primeiro ano da pandemia, 2020, em relação ao ano anterior, 2019. O cálculo da evasão média segue a metodologia de Silva Filho *et al.* (2007) por meio da fórmula (3), utilizando os dados do Censo da Educação Superior disponibilizados pelo INEP.

Gráfico 3- Evasão média ao longo do período analisado (2018-2023)



Fonte: resultados da pesquisa.

É possível verificar ainda, que a média da evasão decresce no segundo ano de pandemia, 2021, e torna a aumentar após esse ano, evidenciando uma instabilidade de comportamento ao longo do período.

4.1 TESTES DE CORRELAÇÃO E MULTICOLINEARIDADE

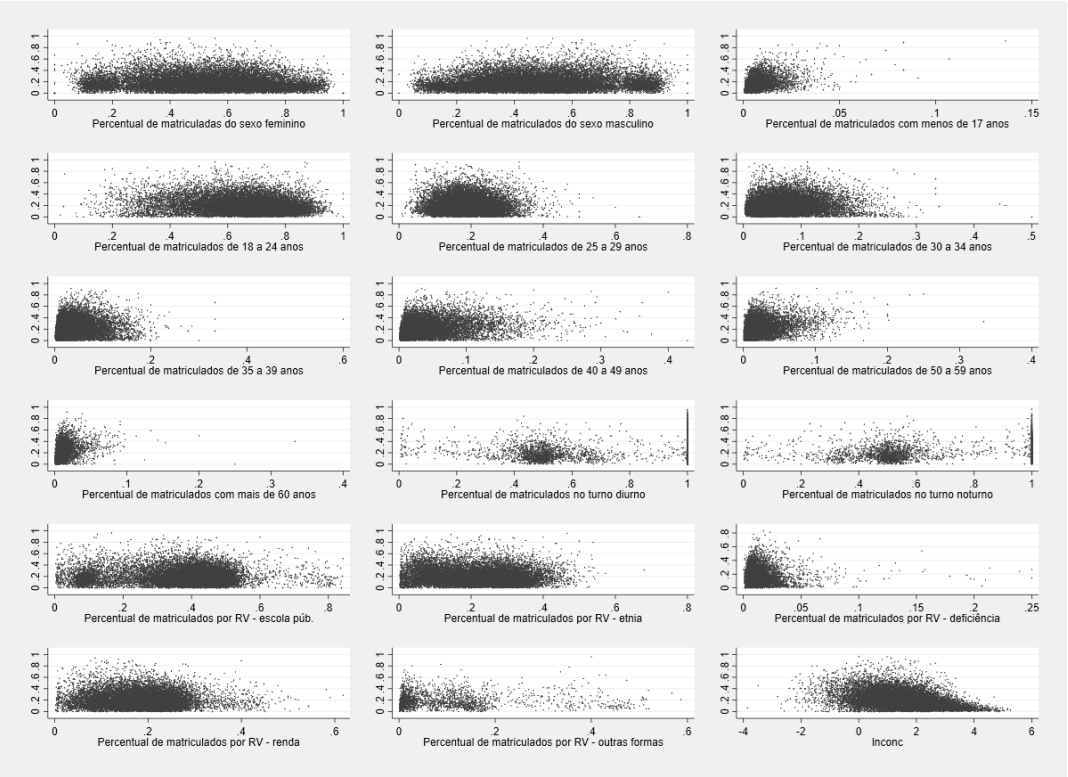
4.1.1 Teste de Correlação de Spearman

Para analisar a correlação entre duas variáveis, primeiramente, é importante a visualização do diagrama de dispersão e do histograma. Nesses gráficos é possível verificar se há relação monotônica ascendente ou descente, se a relação possui uma tendência linear e se a distribuição dos dados é normal (Zou, Tuncali e Silverman, 2003).

O Gráfico 4 apresenta a dispersão da evasão em relação aos parâmetros relacionados aos matriculados por gênero, idade, turno e reserva de vagas em ações afirmativas, e em relação à concorrência. O gráfico evidencia claramente uma relação não-linear das variáveis

explicativas com a variável explicada. Miot (2018) afirma que, ao contrário do teste de Pearson, o uso do teste não paramétrico de Spearman é indicado quando a relação entre as variáveis é não-linear.

Gráfico 4- Dispersão da evasão em relação à concorrência e ao percentual de matriculados por gênero, por idade, por turno e por reserva de vagas em ações afirmativas

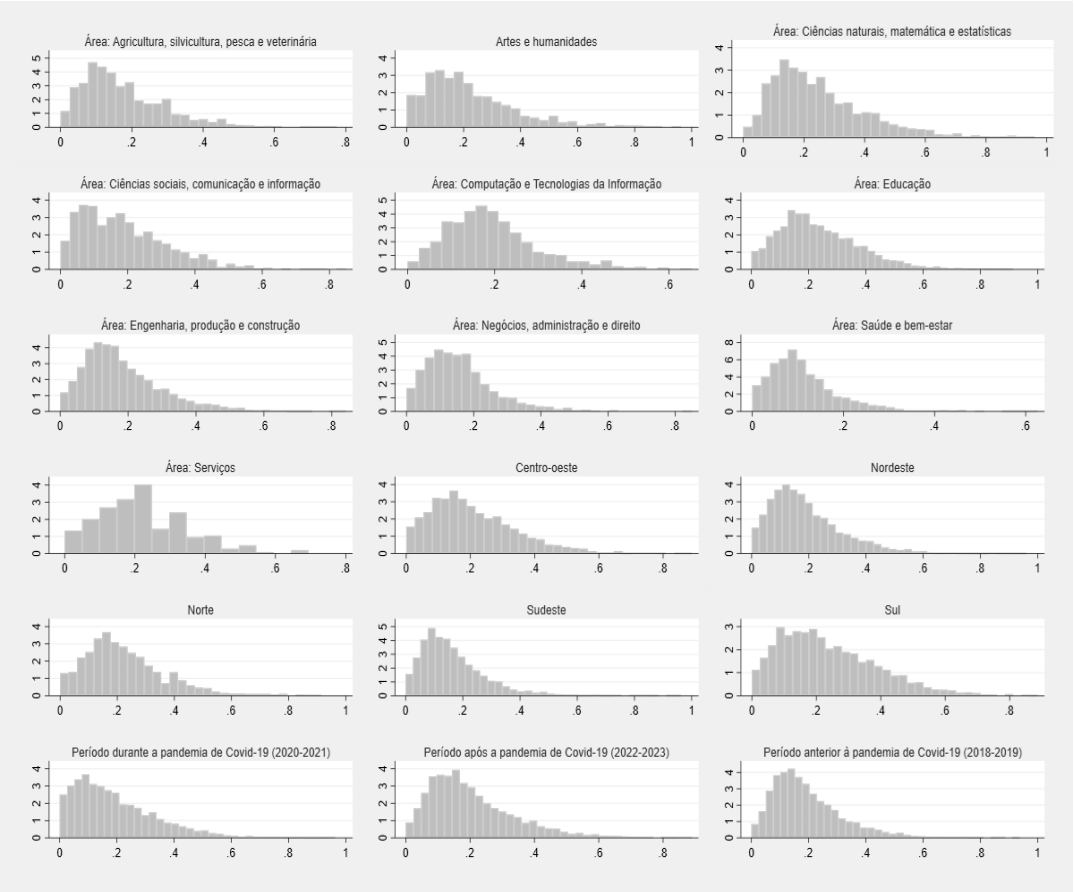


Fonte: resultados da pesquisa.

O

Gráfico ***5*** apresenta as distribuições da taxa de evasão para as áreas gerais, as regiões geográficas e os períodos analisados. É possível identificar uma distribuição assimétrica positiva com um padrão de assimetria tendendo mais à esquerda. Miot (2018) salienta, ainda, que a abordagem normalmente utilizada na investigação da correlação entre variáveis que não apresentam distribuição normal é a correlação da ordem de postos de Spearman. Esse método também se aplica em casos que ao menos uma das variáveis apresenta características ordinais (Zou, Tuncali e Silverman, 2003; Miot, 2018).

Gráfico 5- Histograma da evasão em relação às áreas gerais, às regiões geográficas e aos períodos antes, durante e após a pandemia de COVID-19



Fonte: resultados da pesquisa.

O resultado do teste de correlação de Spearman é apresentado no apêndice A.2. Ele mostra que os parâmetros analisados, em sua maioria, demonstraram haver alguma correlação com a variável explicada com um nível de significância adequado ($p<0.05$).

4.1.2 Fator de Inflação da Variância (VIF)

Os Fatores de Inflação da Variância (VIF) resultantes do modelo apresentaram os coeficientes dispostos na Tabela 1.

Usualmente o valor do VIF igual a 1 indica que não há nenhuma correlação entre as variáveis. Quando o coeficiente encontrado está entre 1 e 5 indica que há correlação moderada, e, se o VIF for maior que 5 significa forte indício de multicolinearidade.

Tabela 1- Fator de Inflação da Variância (VIF) considerando as variáveis relacionadas aos matriculados em valores percentuais no modelo de regressão

Variável	VIF	Variável	VIF
dg6	3.39	dg1	1.75
dg7	3.09	poscovid	1.69
prop40a49	2.93	lnconc	1.67
propescopub	2.91	dg2	1.67
proprenda	2.87	propfem	1.67
propetn	2.74	propnot	1.57
prop35a39	2.59	covid	1.43
prop30a34	2.32	dreg3	1.43
dreg5	2.20	prop60mais	1.42
dg3	2.14	dreg1	1.41
prop50a59	2.09	prop25a29	1.28
dreg2	1.88	proppdef	1.20
dg8	1.85	prop0a17	1.17
dg5	1.83	dg11	1.15
dg4	1.76	propoutros	1.09
Média 1.92			

Fonte: resultados da pesquisa.

Os valores do VIF das variáveis utilizadas estão abaixo de 5, mostrando baixo indício de multicolinearidade entre as variáveis, e, em alguns casos, os valores estão próximos de 1, indicando ausência de multicolinearidade entre as variáveis explicativas utilizadas no modelo.

4.2 TESTES DO MODELO DE REGRESSÃO PARA RESPOSTA FRACIONÁRIA

Os resultados dos testes desse modelo podem ser verificados no Apêndice B.1.

Seguindo Bendig *et al*. (2025) para estimar erros robustos de Huber–White, partindo da premissa que os erros variam em magnitude ao longo dos valores da variável independente, portanto existe heterocedasticidade; o modelo utilizado adota a opção de erros robustos “vce(robust)”.

O Teste de Wooldridge para autocorrelação em painel é tradicionalmente aplicado a modelos xtreg. Contudo, para modelos GEE (xtgee), a abordagem precisa ser adaptada. Dessa forma, foram testadas as correlações “unstructured”, “ar1” e “ar2” separadamente e analisados os seus resultados.

Considerando as opções de link “probit” ou “logit” combinadas com as opções de correlações “corr(unstructured)”, “corr(ar1)” e “corr(ar2)”, os resultados estão resumidos na

Tabela ***2***:

Tabela 2- Resultados de Wald Chi² para os modelos com link probit ou logit, corr(ar1), corr(ar2) ou corr(unstructured)

link	(corr)	Wald Chi²	Prob > Chi²	Nº de variáveis - coeficiente não significativo p>0.1
probit	ar1	2341.09	0	9
probit	ar2	2551,01	0	7
probit	unstructured	2160.92	0	4
logit	ar1	2490.83	0	6
logit	ar2	2303.58	0	5
logit	unstructured	2139.81	0	4

Fonte: resultados da pesquisa.

O Wald Chi² testa a significância global do modelo. Os modelos link(probit) corr(ar2) e link(logit) corr(ar1) apresentaram os resultados de Wald Chi² (Prob > Chi² = 0.0000) maiores que os demais, 2551,01 e 2490.83, respectivamente. Contudo, embora o valor de Wald Chi² no modelo link(logit) corr(ar1) tenha sido menor, ou seja, tem um poder de explicação global do modelo ligeiramente inferior, ele apresenta 6 variáveis sem significância estatística, 1 a menos que no modelo link(probit) corr(ar2).

Considerando esses resultados e os argumentos de Gujarati (2008), onde explica que o “logit” é melhor interpretável e tem um ajuste melhor devido ao ritmo de resposta um pouco mais lento (o modelo link(probit) corr(ar2) gastou 9 iterações para convergir, enquanto o link(logit) corr(ar1) gastou 11 iterações), neste trabalho o modelo adotado foi o link(logit) corr(ar1).

Para definição do uso de “logit”, “ar1” e “vcs(robust)” os testes foram aplicados ao período todo, de 2018 a 2023, e depois de definido, foi estendido ao demais modelos (2018-2019, 2020-2021 e 2022-2023).

A seguir seguem os resultados dos testes de diagnóstico para verificação da consistência do modelo:

I. Para todos os 4 casos (períodos pré-pandemia, durante a pandemia, pós-pandemia e período total), foram analisadas as significâncias das variáveis incluídas (resíduos “_hat”). Os

resultados mostram que nem todas as variáveis "_hat" foram significativas, especialmente no período pós-pandemia, a maioria não foi significativa (os resultados se encontram nos Apêndices B2, B3, B4 e B5).

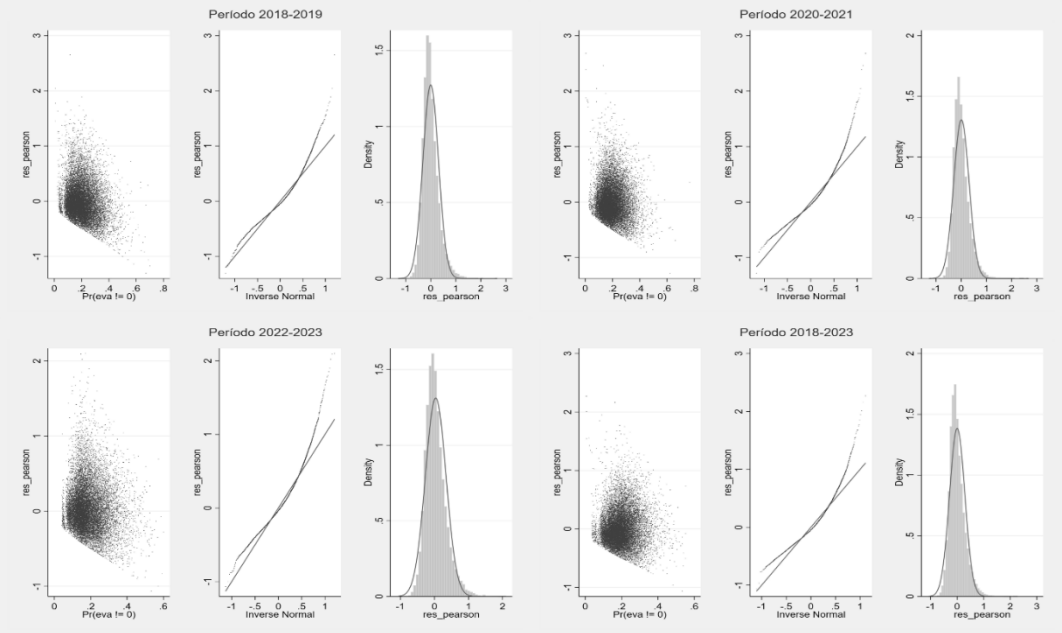
II. Após a aplicação do teste de validade dos instrumentos, os resultados de Chi^2 foram significativos para todos os 4 casos, conforme resultados apresentados no Apêndice B.1.

III. A análise do pseudo R2 não pôde ser concretizada porque não está previsto para o xtgee no Stata®.

IV. O

Gráfico ***6*** ilustra os resíduos de Pearson, onde "p_hat" contém as probabilidades previstas, em todos os 4 períodos. Em todos os períodos, a dispersão não apresentou um padrão como uma tendência crescente ou decrescente, que poderia indicar problemas de especificação do modelo. Os resíduos estão distribuídos aleatoriamente em torno de zero. No gráfico quantil normal, os pontos estão aproximadamente alinhados com a diagonal do gráfico, ou seja, os resíduos seguem uma distribuição normal. E o histograma mostra uma distribuição simétrica com centro próximo de zero. Essas características indicam uma escolha apropriada do modelo.

Gráfico 6- Visualização dos resíduos de Pearson em relação às probabilidades previstas – dispersão, quantil normal e histograma



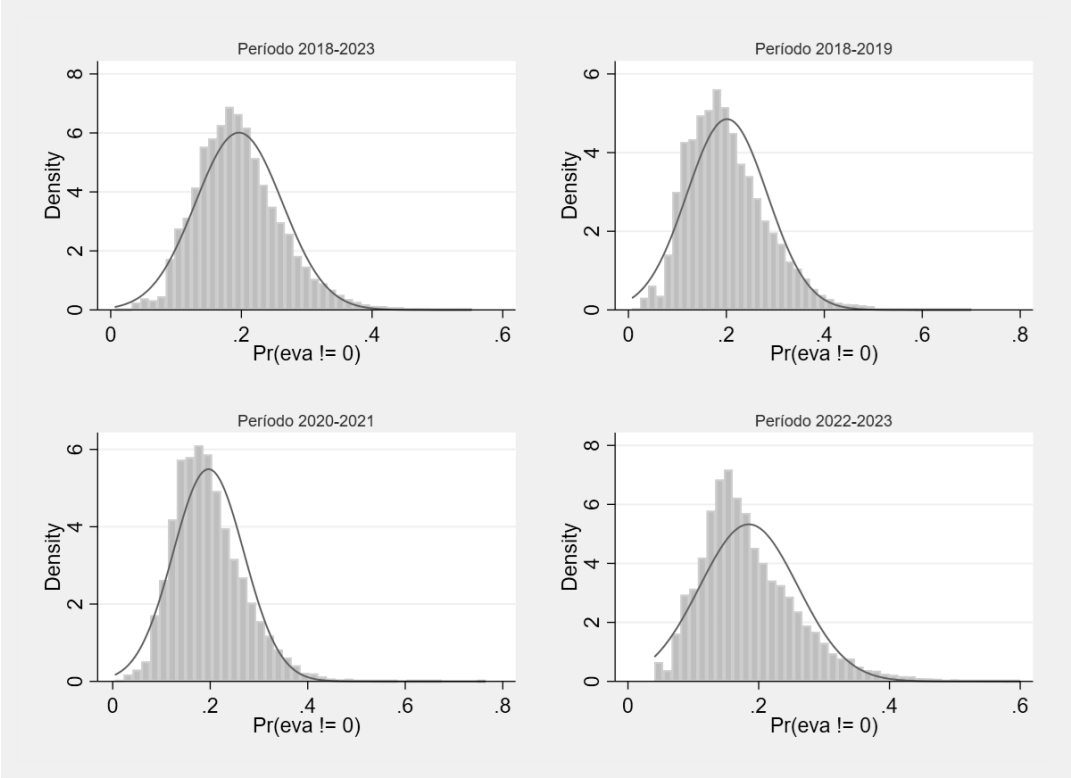
Fonte: resultados da pesquisa.

V. Os resultados da correlação entre “eva” e “p_hat” fornecem uma ideia da adequação preditiva. Para todo o período a correlação é 0.5128. Para os períodos antes, durante e após a pandemia as correlações foram 0.4140, 0.4510 e 0.4725, respectivamente. Pode-se considerar uma correlação moderada. Os resultados podem ser verificados no Apêndice B.1.

VI. O

Gráfico 7 traz o resultado da avaliação da distribuição dos valores preditivos por meio do histograma.

Gráfico 7- Distribuição dos Valores Preditivos



Fonte: resultados da pesquisa.

É possível verificar que as probabilidades estimadas “p_hat” apresentaram uma distribuição razoável. Ou seja, é um resultado positivo, visto que não há uma concentração em 0 ou 1 em nenhum dos 4 períodos.

4.3 ABORDAGEM DESCRITIVA DOS DADOS

4.3.1 Áreas gerais do conhecimento

A pesquisa analisa a evasão, entre outras variáveis, nas 11 áreas gerais do conhecimento, de acordo com a classificação utilizada pelo INEP, apresentadas na Tabela 3. Todavia, dentre elas, a área intitulada como "Programas Básicos" chama atenção devido aos índices estatísticos em relação às demais áreas. Nela estão contidos os cursos de graduação cuja forma de ingresso se dá por meio da chamada "entrada única", ou por Área Básica de Ingresso (ABI), onde é possível escolher o curso posteriormente. Explica o INEP (2015) que a ABI se refere à situação em que uma única entrada no curso possibilitará ao estudante, após a conclusão de um conjunto básico de disciplinas (ou ciclo básico), a escolha efetiva da sua formação acadêmica.

Foi verificado que o número de observações dentro desse grupo é baixo em relação às outras áreas e que os cursos são muito diversos (ABI - Música; ABI - Teatro; ABI - Matemática, ABI - História; ABI - Letras Espanhol; ABI - Letras Francês, ABI - Educação Intercultural etc), podendo comprometer a essência do estudo, uma vez que cada curso apresenta características particulares e impacta de forma diferente nos índices de evasão. Além disso, o estudo mostra que a evasão dentro dessa área tem alta dispersão.

Para compreender a distribuição dos dados e suas características principais, como tendência central, dispersão e amplitude, foram levantadas as estatísticas descritivas da evasão em relação às áreas gerais, conforme disposto na Tabela 3.

Tabela 3- Estatísticas descritivas da evasão por área geral

Área Geral	N	Mean	SD	CV	P25	P50	P75	Min	Max
Agricultura, silvicultura, pesca e veterinária	1194	0.1762878	0.1172442	0.6650727	0.0892857	0.1498032	0.2374101	0	0.775
Artes e humanidades	950	0.2222146	0.1609642	0.7243636	0.1071429	0.1865537	0.3023256	0	0.962963
Ciências naturais, matemática e estatísticas	1661	0.250739	0.1577727	0.6292309	0.1333333	0.2173913	0.3349754	0	0.96
Ciências sociais, comunicação e informação	1372	0.1895821	0.1309301	0.6906248	0.0852283	0.1673559	0.2676495	0	0.8484849
Computação, TI e Comunicação	665	0.1941839	0.1063074	0.547457	0.1190476	0.1783784	0.2478261	0	0.6521739
Educação	4440	0.2366619	0. 1416432	0. 5985045	0.1333333	0.2115784	0.3205618	0	0.9135802

Engenharia, produção e construção	2952	0.184518	0.1178284	0.6385739	0.1005223	0.1572758	0.2437237	0	0.8409091
Negócios, administração e direito	1374	0.1603883	0.10522	0.6560325	0.0851064	0.1424045	0.2087912	0	0.849315
Programas básicos	258	0.4333456	0.2122859	0.4898767	0.2627986	0.3831827	0.625	0.0353698	0.9790576
Saúde e bem-estar	1904	0.1137092	0.0802228	0.7055082	0.0593907	0.0972837	0.1495614	0	0.6261683
Serviços	210	0.2224307	0.1295334	0.5823539	0.1302326	0.2102355	0.3076923	0	0.6940299
Total global	**16980**	**0.2013058**	**0.1391166**	**0.6910708**	**0.0996092**	**0.1699188**	**0.2708333**	**0**	**0.9790576**

N:	Número de observações encontradas para evasão
Mean:	Média da taxa de evasão
SD:	Desvio-padrão
CV:	Coeficiente de variação
P25:	Percentil 25
P50:	Percentil 50
P75:	Percentil 75
Min:	Valor mínimo
Max:	Valor máximo

Fonte: resultados da pesquisa.

O número de observações (N) varia bastante entre as áreas. O maior é da área de Educação (4440) e o menor em Programas básicos (258). Em relação à média (Mean) da taxa de evasão, observa-se que é mais alta em Programas básicos (0,43), uma taxa consideravelmente maior do que a da próxima área mais alta: Ciências naturais, matemática e estatísticas. Além disso, a média da evasão em Programas básicos é mais que o dobro da média global.

O desvio padrão (SD) é a medida mais comum para avaliar a dispersão dos dados. Ele indica o quanto os valores individuais de um conjunto de dados se distanciam da média (Wooldridge, 2013). Na amostra utilizada os desvios-padrão variam muito, sendo que a área de Programas básicos também apresenta o maior valor (0,21), o que sugere uma maior dispersão dos dados nesta área. No entanto, foi calculado também o coeficiente de variação (CV), que ao contrário do desvio-padrão, indica uma dispersão relativa e, nesse caso a área de Programas básicos apresentou o menor índice. O CV expressa o desvio padrão como uma porcentagem da média, tornando possível comparar a variabilidade entre conjuntos de dados que têm unidades diferentes ou escalas diferentes (Fialho, 2014). O coeficiente de variação é maior em Artes e humanidades (0,72), indicando que os dados nessa área são mais dispersos em relação à média.

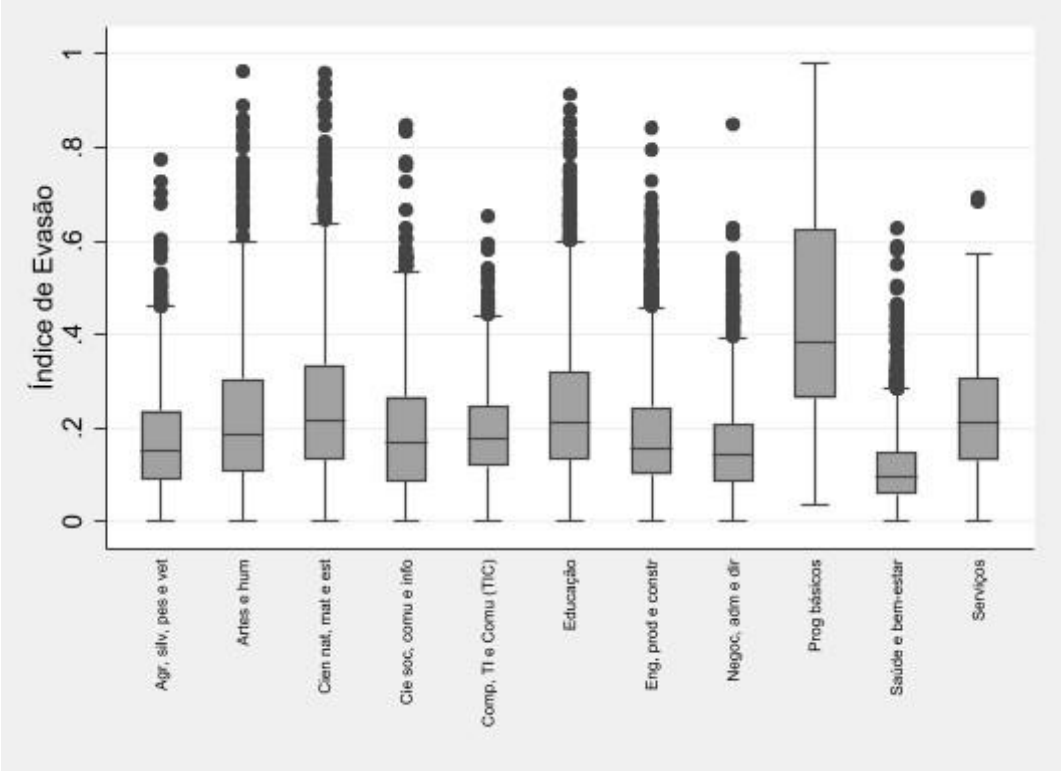
O percentil P25 conhecido como o primeiro quartil, é o marco abaixo do qual estão os 25% menores valores. Ele representa o ponto de corte para os 25% valores inferiores. Do mesmo modo o percentil P50 (mediana) separa os valores 50% inferiores e os 50% superiores, assim como o percentil P75, também conhecido como o terceiro quartil, é o valor abaixo do qual compreendem 75% dos menores valores (Hoed, 2016). Em outros termos, os percentis indicam a porcentagem de dados que estão abaixo de um valor específico e ajudam a entender a dispersão dos dados de maneira mais robusta, sem a influência de valores extremos (*outliers*). O

Gráfico 8 revela que a área Programas básicos possui a maior amplitude interquartílica, isso implica que, quanto maior a diferença entre P25 e P75, maior é a dispersão dos dados na parte central da distribuição.

Ainda nesse sentido, o

Gráfico 8 demonstra uma maior assimetria para a área de Programas básicos em função da não centralização na mediana no *box*. Além disso, os "bigodes", que indicam a variabilidade dos dados fora da caixa, apresentam um intervalo de valores muito maior que a amplitude interquartílica a partir de P25 e P75. Portanto, os bigodes, quando longos, indicam que os dados se espalham bastante (maior dispersão). Quando curtos, indicam que os dados estão mais concentrados em torno da mediana (menor dispersão) (Hoed, 2016).

Gráfico 8- Distribuição das observações sobre a evasão, por área geral, por meio de diagrama de caixas



Fonte: resultados da pesquisa.

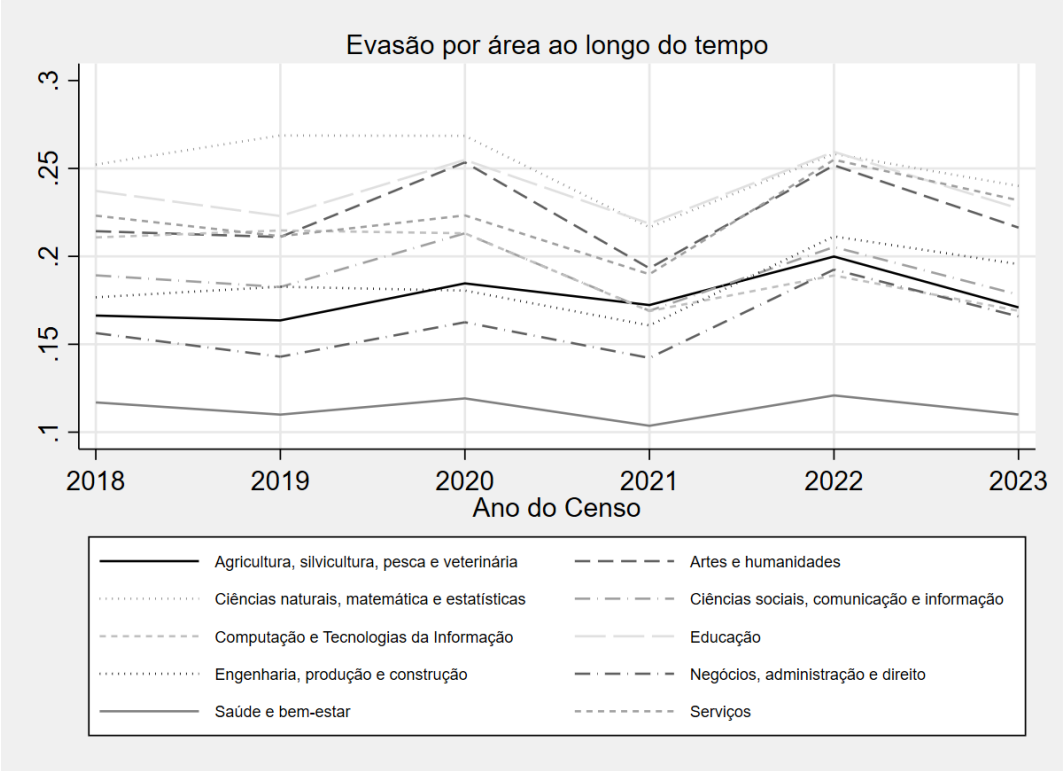
Também é possível verificar a dispersão por meio da largura da caixa, ou seja, a amplitude entre os quartis. Quanto maior a caixa, maior a variabilidade entre o 25º e o 75º percentil. A área de Programas básicos é, evidentemente a área com maior amplitude.

Portanto, no que tange às áreas analisadas, foi necessário excluir a área Programas básicos pela possibilidade de influenciar a resposta do comportamento da evasão além de comprometer os resultados das outras áreas. As demais análises do trabalho já não consideram essa área.

O

Gráfico ***9*** mostra a evasão média por área geral (excluída a área de Programas Básicos) ao longo de todo o período analisado.

Gráfico 9- Evasão média por área geral ao longo tempo (2018-2023)



Fonte: resultados da pesquisa.

O Gráfico 9 revela ainda que, em qualquer ano, a evasão na área de Ciências naturais, matemática e estatísticas é maior que todas as outras, com exceção do período de 2021 a 2022 que se assemelha à área de Educação. Essas duas áreas apresentam média de evasão maiores que os valores médios da evasão total apresentados na *Error! Reference source not found.*3. Esse achado se contrapõe à pesquisa de Gomes e Hirata (2022). Nos resultados da pesquisa, os autores afirmam que os cursos na área de Educação geralmente apresentam menor risco de evasão.

Por outro lado, a área de Saúde e bem-estar se mostra a de menor evasão ao longo do período, bem abaixo da média. Silva Filho *et al* (2007) também encontraram taxas médias anuais de evasão menores que a média nacional. Pode-se inferir que um dos fatores que podem influenciar esse índice é a alta concorrência, principalmente porque os cursos de medicina, que estão contidos nesse grupo, são reconhecidos no Brasil por alta procura. Adachi (Adachi, 2009) afirma que cursos como Medicina e Odontologia apresentam menores taxas de evasão acadêmica.

As demais áreas do conhecimento têm evasão mais próxima da média global, numa escala de classificação da maior à menor seguindo as áreas: Serviços, Artes e Humanidades, Computação e Tecnologias da Informação e Comunicação (TIC), Ciências sociais,

comunicação e informação, Engenharia, produção e construção, Agricultura, silvicultura, pesca e veterinária e Negócios, administração e direito, podendo intercambiar um pouco entre elas em algum ano específico.

A área de Serviços, especificamente, com índice médio de evasão de 22%, aproximadamente, se mostra como a terceira maior taxa dentre as outras áreas. Verifica-se que essa área apresentou índices mais elevados em pesquisa anterior, com uma média de 29% entre 2001 e 2005, sendo a área com a maior taxa de evasão à época (Silva Filho *et al.*, 2007).

Essas análises revelam também que os cursos das áreas de Ciências sociais, comunicação e informação e Negócios, administração e direito, particularmente, aparentam redução dos índices médios de evasão em relação a períodos anteriores. Silva Filho *et al* (2007) afirmam que a taxa média anual de evasão para essas áreas era de aproximadamente 25% no período de 2001 a 2005[1]. No período de análise da pesquisa essa média não alcança 20%.

De 2018 a 2019 houve um leve declínio da evasão em todas as áreas, excetuando-se Ciências naturais, matemática e estatísticas, Computação e Tecnologias da Informação e Comunicação (TIC) e Engenharia, produção e construção que apresentaram ligeiro aumento, podendo inferir, nesse período, algum tipo de problema dos discentes em relação aos cursos de maior abstração. Esses resultados vão ao encontro com os achados de Hoed (2016), o autor concluiu que os cursos com maiores taxas de evasão são cursos com grande exigência matemática.

Em 2020 observa-se um aumento da evasão em todas as áreas, salvo Engenharia, produção e construção que se manteve praticamente estável, seguida de uma queda em todas as áreas no ano de 2021. O ano de 2020 foi marcado pelo início da pandemia de COVID-19 no Brasil. Muitas instituições de ensino adotaram as instruções de suspensão das aulas presenciais de acordo com as recomendações das organizações nacionais e internacionais de saúde, à época. Algumas encontraram dificuldades para remodelar sua dinâmica de ensino em um novo ambiente, o virtual. Ainda nesse sentido, o período pandêmico trouxe consigo sentimentos de medo e incertezas quanto à saúde e segurança da população, seguidos de problemas psicológicos advindos do isolamento social. Uma pesquisa realizada com estudantes de graduação no Brasil identificou aumento significativo na ansiedade, instabilidade emocional e

[1] Nesse período, a classificação e subdivisão das áreas de conhecimento adotadas pelo Inep era um pouco diferente, sendo que essas duas áreas eram compreendidas dentro de uma única: Ciências Sociais, Negócios e Direito.

piora na saúde mental em geral (Valesan *et al.*, 2024). Esse fenômeno pode ter influenciado os discentes em relação à decisão de se manterem matriculados, ou não, nos cursos de graduação.

Em contrapartida, a queda expressiva dos índices de evasão no ano seguinte pode se justificar em função da adoção, por algumas instituições de ensino, da não obrigatoriedade de inscrição em componentes curriculares, entendendo-se que os discentes optantes de cursos presenciais não poderiam ser obrigados a cursar a graduação em formato à distância, além da flexibilização do calendário acadêmico das instituições (Brasil, 2020c; Brasil, 2020d). Essas medidas podem ter causado impacto na queda da evasão.

O ano de 2022 teve um aumento expressivo de evasão em todas as áreas com uma leve queda em 2023, mas ainda assim, mantendo-se maior do que no ano de 2021. Isso pode ser justificado pela dificuldade de retorno dos discentes à universidade após um período de aproximadamente 2 anos em isolamento social.

É observável ainda, que a evasão obedece a um comportamento padrão ao longo do período analisado, independentemente da área de conhecimento. Ou seja, as razões que motivam a tendência de aumento ou diminuição da evasão pelos discentes, nesse período, podem transcender às questões acadêmicas e estarem também relacionadas às questões sociais.

4.3.2 Regiões geográficas do Brasil

A mesma análise das estatísticas descritivas foi feita para a evasão em relação às 5 regiões geográficas brasileiras: Centro-Oeste, Nordeste, Norte, Sudeste e Sul, conforme Tabela 4 (excluída a área de Programas Básicos). Essas estatísticas ajudam a entender como a evasão varia entre as diferentes regiões e a identificar possíveis padrões ou inconsistências.

De acordo com os valores apresentados na Tabela 4 e o com o

Gráfico ***10***, pode-se verificar que o desvio padrão (SD) e o coeficiente de variação (CV) indicam que, embora a média da evasão seja relativamente baixa, a dispersão dos dados nas regiões pode ser significativa. Em algumas regiões, como Sudeste e o Nordeste, essa dispersão é particularmente alta: 0,69 e 0,67, respectivamente.

Tabela 4- Estatísticas descritivas da evasão por região geográfica dos cursos

Região	N	Mean	SD	CV	P25	P50	P75	Min	Max
Centro-Oeste	2.076	.2089176	.1359996	.6509725	.1077769	.1836735	.2884064	0	.8888889
Nordeste	4.914	.1909979	.1290107	.6754561	.0982503	.1632226	.255814	0	.96
Norte	1.302	.2290145	.1453971	.6348817	.1282051	.2	.3052632	0	.9135802
Sudeste	5.046	.1646845	.1136982	.6903997	.0845987	.141065	.2171717	0	.962963
Sul	3.384	.23786	.1511863	.6356101	.1193182	.2097338	.3333333	0	.8867925
Total global	**16.722**	**.1977258**	**.134595**	**.6807154**	**.0986842**	**.1680272**	**.2673267**	**0**	**.962963**

N:	Número de observações encontradas para evasão
Mean:	Média da taxa de evasão
SD:	Desvio-padrão
CV:	Coeficiente de variação
P25:	Percentil 25
P50:	Percentil 50
P75:	Percentil 75
Min:	Valor mínimo
Max:	Valor máximo

Fonte: resultados da pesquisa.

A região Sul apresentou a maior média de evasão (0,2379), seguida pela região Norte (0,2290) e pelo Centro-Oeste (0,2089). Em contrapartida, a menor média foi observada no Sudeste (0,1647) para esse recorte temporal. Apesar das diferenças nas médias, o coeficiente de variação (CV) se manteve próximo em todas as regiões (variando de 0,6349 a 0,6904), o que indica que a dispersão relativa das taxas foi semelhante, sugerindo que a variação na evasão é consistente entre as regiões.

No sentido do que os resultados apresentam, uma pesquisa com dados de 2011 a 2016 indicou que a evasão ocorre mais frequentemente na região Sul e em universidades localizadas no interior. Instituições interiorizadas, com um perfil de estudantes mais popular, enfrentam índices de evasão superiores a 50% (Nierotka, Salata e Martins, 2023).

O percentil 75 (P75) para as regiões Sul (0,3333) e Norte (0,3053) apresentam valores bem mais elevados que o Sudeste (0,2171) e o Nordeste (0,2558), reforçando que, mesmo entre as instituições com menores taxas, ainda existem picos mais altos nessas duas regiões.

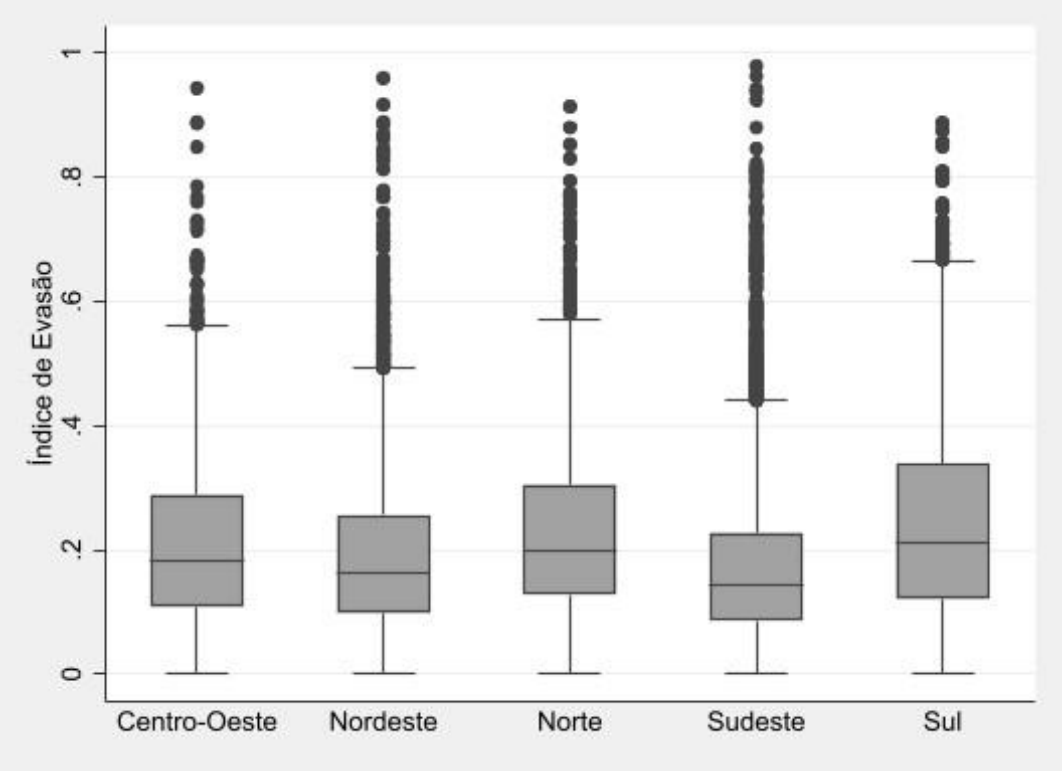
De forma agregada, a média nacional de evasão foi de 0,1977, com desvio-padrão de 0,1346, indicando que, no geral, cerca de 20% dos estudantes evadiram, mas com variação considerável entre cursos e contextos regionais. Esses resultados sugerem que políticas de permanência estudantil podem demandar estratégias diferenciadas por região, levando em conta

que, apesar da dispersão relativa semelhante, os níveis médios de evasão não são uniformes no território nacional.

O

Gráfico ***10*** revela ainda que todas as regiões possuem *outliers*, o que indica vários cursos com evasão muito alta, mesmo em regiões com menor mediana. O valor mínimo de 0 em todas as regiões mostra que, em alguns casos, a evasão é nula, enquanto os máximos são bem elevados, com destaque para o Sudeste (0,96), indicando que, embora a evasão média seja inferior às demais, ainda existem picos de evasão significativos, chegando a quase 100%.

Gráfico 10- Distribuição das observações sobre a evasão, por região geográfica, por meio de diagrama de caixas



Fonte: resultados da pesquisa.

Embora a mediana do índice de evasão seja relativamente baixa em todas as regiões, há variações importantes. A dispersão dos dados, indicada pelo comprimento das caixas e pelos limites dos "bigodes", mostra que existe variabilidade considerável dentro de cada região, sendo o Nordeste e o Sudeste os casos com maiores extremos positivos.

A distribuição das evasões tende a ser assimétrica em todas as regiões, com uma maior concentração de valores no lado inferior, ou seja, valores mais baixos de evasão, mas com uma cauda que pode alcançar valores mais altos de evasão.

Nesse sentido, a análise permite inferir que a maioria das regiões apresenta índices de evasão abaixo de 0,4, mas estratégias de mitigação deveriam focar nos casos extremos, especialmente nas regiões Sudeste e Nordeste, onde há maior concentração de valores atípicos. Esses achados fornecem uma base visual para investigar fatores socioeconômicos ou estruturais que influenciam a evasão escolar.

4.3.3 Concorrência

O

Gráfico *1* mostra a dispersão entre a evasão e a concorrência. Ele apresenta uma relação inversa entre evasão e concorrência, concentrando-se a maior parte da evasão nos menores índices de candidato/vaga.

A Tabela 5 apresenta as estatísticas descritivas da concorrência em relação às áreas do conhecimento. Ela revela que a maior média entre as áreas está na área Saúde e bem-estar, aproximadamente 21,6 candidatos por vaga. Essa área também apresenta o maior valor absoluto: 192,52 candidatos por vaga. Muito possivelmente, esses índices estão atrelados à procura elevada pelos cursos de Medicina nos últimos anos no Brasil (Andrade, 2025). Esses cursos, que geralmente têm níveis elevados de concorrência, apresentam baixa evasão. Isso é justificado pela satisfação e motivação do estudante ao ser aprovado em um curso tão concorrido, além dos anseios de realização profissional e financeira (Hoed, 2016). Isso sugere, de antemão, uma comparação da elevada concorrência nessa área com as menores médias de evasão. Silva Filho *et al.* (2007) e Adachi (2009) admitem encontrar correlação negativa da evasão com a demanda, explicando que áreas com alta relação candidato por vaga tendem à menor desistência.

Ainda nesse sentido, a menor procura pelos cursos da área de Ciências naturais, matemática e estatísticas, média de 4,44 candidatos por vagas, pode estar intimamente relacionado com a média de evasão, visto que a área apresentou a maior saída.

Tabela 5- Estatísticas descritivas da concorrência por área geral

Área Geral	N	Mean	SD	CV	P25	P50	P75	Min	Max
Agricultura, silvicultura, pesca e veterinária	1194	7.831867	8.242876	1.052479	2.491525	5.165151	9.56701	.0752688	56.97143
Artes e humanidades	950	6.484781	6.168914	.9512909	2.009091	4.555	9.1	.0694444	42.91667
Ciências naturais, matemática e estatísticas	1661	4.43844	3.729237	.8402135	1.804878	3.5	5.925	.1025641	30.68182
Ciências sociais, comunicação e informação	1372	10.30462	12.26399	1.190145	3.189167	6.227778	11.81057	.2093023	94.6383
Computação e Tecnologias da Informação	665	7.036306	5.421202	.7704614	3.454545	5.690909	9.333333	.4129032	47.12
Educação	4440	5.256261	5.350405	1.017911	1.724447	3.628719	6.995833	.0111111	62.4375

Engenharia, produção e construção	2952	5.053695	4.702123	.9304326	2.094885	3.803913	6.333333	.144	51.92
Negócios, administração e direito	1374	11.89458	10.37639	.8723629	4.270492	8.948822	16.43518	.3658537	73.46
Saúde e bem-estar	1904	21.59696	23.08879	1.069076	8.017424	14.53137	26.01082	.3902439	192.5238
Serviços	210	8.882685	10.00258	1.126076	2.215909	5.020759	11.4	.094697	56.59259
Total global	16722	8.329539	11.47813	1.378003	2.386364	4.887301	9.717391	.0111111	192.5238

N:	Número de observações encontradas para evasão
Mean:	Média da taxa de evasão
SD:	Desvio-padrão
CV:	Coeficiente de variação
P25:	Percentil 25
P50:	Percentil 50
P75:	Percentil 75
Min:	Valor mínimo
Max:	Valor máximo

Fonte: resultados da pesquisa.

A mesma observação pode ser feita para a concorrência nas regiões geográficas, conforme Tabela 6. Ela mostra que as regiões Sudeste e Nordeste, que apresentam as maiores concorrências, também vão ao encontro da análise dos dados que indicam essas regiões com as menores médias de evasão, o que pode estar associado ao número de oferta de cursos e ao quantitativo populacional. Ao passo que a região Sul, aquela com a maior média de evasão, apresenta a menor procura por cursos: 4,35 candidatos por vaga, em média, apesar de ser a segunda região com maior densidade demográfica.

Tabela 6- Estatísticas descritivas da concorrência por região geográfica dos cursos

Região	**N**	**Mean**	**SD**	**CV**	**P25**	**P50**	**P75**	**Min**	**Max**
Centro-Oeste	2076	6.898608	9.165097	1.328543	2.073052	4.103333	8.127307	.0111111	117.1239
Nordeste	4914	9.697247	11.32559	1.167918	3.376923	6.350824	11.575	.1035857	155.1389
Norte	1302	8.199226	11.8553	1.445904	2.454545	5.29686	9.606061	.1025641	146.775
Sudeste	5046	10.29008	13.15794	1.278701	3.4375	6.268075	11.87097	.0574713	192.5238
Sul	3384	4.347997	8.693209	1.99936	1.3	2.359775	4.358289	.0655738	174.275
Total global	16722	8.329539	11.47813	1.378003	2.386364	4.887301	9.717391	.0111111	192.5238

N:	Número de observações encontradas para evasão

Mean:	Média da taxa de evasão
SD:	Desvio-padrão
CV:	Coeficiente de variação
P25:	Percentil 25
P50:	Percentil 50
P75:	Percentil 75
Min:	Valor mínimo
Max:	Valor máximo

Fonte: resultados da pesquisa.

Os resultados apresentam também valores mínimos e máximos, em todas as regiões, bastante semelhantes, revelando que em qualquer região, existem cursos com procura muito alta e outros com procura muito baixa. Nesse sentido, os coeficientes de variação significativamente elevados, denotam um comportamento heterogêneo na concorrência entre os cursos presentes em cada região.

4.3.4 Turno

No que diz respeito à distribuição dos estudantes por turno, de acordo com os dados analisados, a maioria está matriculada no turno diurno, representando 72,0% do total. O turno noturno concentra os 27,9% restantes dos alunos. Essa predominância do diurno pode refletir tanto a maior oferta de vagas nesse período quanto a menor disponibilidade de cursos noturnos em algumas regiões ou áreas de conhecimento.

Na Tabela 7 é apresentada a evasão média de acordo com a proporção de alunos matriculados no turno noturno, dividindo as observações em tercis. Cada tercil representa as proporções de discentes do menor para o maior, de modo que 1 representa menor proporção, 2 representa média proporção e 3 representa alta proporção. A análise foi feita sob a perspectiva apenas do turno noturno, pois esse parâmetro tem caráter complementar em relação ao turno diurno, dispensando a análise desse último.

Tabela 7- Evasão média pela proporção de matriculados no turno noturno dividida em tercis

	Tercil	Mean	SD
Noturno	1 (menor prop. no noturno)	.1835757	.1290315
	2 (média prop. no noturno)	.2079584	.1109474
	3 (alta prop. no noturno)	.2257678	.1411199

Tercil:	Divisão da proporção de discentes em 3 grupos
Mean:	Média da taxa de evasão
SD:	Desvio-padrão

Fonte: resultados da pesquisa.

A análise da evasão no turno noturno mostra diferenças relevantes entre os grupos avaliados. No grupo 1, a média da evasão é de aproximadamente 18,4%, indicando a menor média dentre os tercis. O desvio padrão de 0,129 mostra uma variação significativa no grupo, inferindo que a evasão pode ser muito diferente dependendo da unidade analisada. No segundo tercil, a média da evasão aumenta para cerca de 20,8%, com um desvio padrão um pouco menor, 0,11, o que indica uma tendência de maior evasão no turno noturno para esse grupo, mas com variações menos acentuadas em relação ao grupo anterior, ou seja, com menor dispersão entre os dados. O grupo 3 apresenta a maior média de evasão, cerca de 22,6%, acompanhada do maior desvio padrão, 0,14, mostrando que, apesar de ser o grupo com maior taxa média de evasão, também possui maior dispersão nos índices observados. Esse tipo de análise permite verificar que, quanto mais se eleva a proporção de alunos matriculados no turno noturno de determinado

curso, mais se eleva as taxas de evasão, em consonância com Gomes e Hirata (2022) e Adachi (2009), que afirmam que estudos indicam risco maior de evasão em cursos noturnos.

Essa maior evasão nos cursos noturnos é frequentemente associada à dificuldade em conciliar trabalho e estudo (Gomes e Hirata, 2022). Desse modo, o aumento gradual das médias de evasão do grupo 1 ao grupo 3 sugere a necessidade de investigar os fatores que contribuem para essa diferença, como condições socioeconômicas, já que muitas vezes o discente estuda neste turno devido à necessidade de trabalhar, além de se verificar características dos cursos e suporte institucional, para que possam ser adotadas estratégias mais eficazes de retenção dos estudantes no turno noturno.

Sob a perspectiva do turno diurno, no contexto do 1º tercil (menor proporção no noturno e, portanto, maior no diurno), a evasão média é de 0,1836, a menor entre os grupos. Essa evidência, somada ao peso de 72% das matrículas serem no turno diurno, mostra que este turno pode ter um papel protetivo estrutural contra a evasão, influenciando positivamente o desempenho do sistema acadêmico como um todo. A predominância no diurno pode estar associada a fatores como maior alinhamento com a idade universitária regular, rotinas mais estáveis de estudos e menor sobreposição com atividades laborais. Esse contexto pode contribuir para menores chances de reprovação, o que, segundo Chagas (2019) pode favorecer a permanência.

Além disso, de acordo com Adachi (2009), alunos de cursos diurnos podem ter um perfil socioeconômico diferente dos alunos do noturno, uma vez que ingressam na universidade com algum tipo de apoio financeiro familiar, não necessitando se dividir entre trabalho e estudo.

O desvio-padrão mais baixo no 1º tercil (0,1290) e no 2º tercil (0,1109) sugere que, nos cenários onde a participação do diurno é alta, a variação da evasão entre cursos é menor. Em outras palavras, a permanência na universidade tende a ser mais uniforme, reforçando a ideia de que o diurno proporciona um ambiente mais previsível para o acompanhamento e controle da evasão.

4.3.5 Política de reserva de vagas

A Tabela 8 apresenta as estatísticas descritivas das proporções de estudantes matriculados por tipo de reserva de vagas. As variáveis analisadas incluem: escola pública, etnia, deficiência, renda e outros.

Tabela 8- Estatísticas descritivas da evasão pela proporção de matriculados por reserva de vagas (RV)

RV	N	Mean	SD	CV	P25	P50	P75	Min	Max
Escola Pública	8317	.1718276	.1034311	.6019469	.0968553	.1564208	.2283465	0	.8029197
Etnia	8080	.1674267	.0984141	.5878039	.0943831	.153558	.2244561	0	.8
Deficiência	2203	.1307328	.0750013	.5736994	.0757576	.1254436	.1740394	0	.5393258
Renda	6878	.1669976	.0960433	.5751179	.095427	.1556626	.2227488	0	.6980198
Outros	1448	.1901609	.1149855	.6046749	.1025937	.1705607	.254951	0	.7288136

N:	Número de observações encontradas para evasão
Mean:	Média da taxa de evasão
SD:	Desvio-padrão
CV:	Coeficiente de variação
P25:	Percentil 25
P50:	Percentil 50
P75:	Percentil 75
Min:	Valor mínimo
Max:	Valor máximo

Fonte: resultados da pesquisa.

A análise das taxas de evasão entre os grupos de ingresso por reserva de vagas revela diferenças importantes na permanência estudantil conforme o tipo de política afirmativa utilizada. Os dados mostram que estudantes que ingressaram pela categoria “outros” apresentam a maior média de evasão, com 19,0%, seguidos pelos que entraram por escola pública (17,18%), etnia (16,74%) e renda (16,70%). A menor média de evasão é observada entre estudantes que ingressaram por meio da reserva para pessoas com deficiência, com 13,07%, valor notavelmente inferior aos demais.

Ainda que as médias variem relativamente pouco (entre 13% e 19%) os desvios padrão apontam para diferenças significativas. A evasão por meio da categoria “outros” apresenta a maior dispersão (SD = 0,115), indicando que a experiência de permanência dos estudantes dessa categoria varia amplamente conforme o contexto. Essa heterogeneidade pode refletir a própria natureza indefinida dessa categoria, que possivelmente envolve processos institucionais de cotas ou a combinação delas. Em contraste, a evasão entre estudantes com deficiência tem o menor desvio padrão (0,075), sugerindo uma distribuição mais concentrada, e com a média relativamente baixa. Isso pode indicar que, onde há políticas específicas de suporte, a permanência desses estudantes é mais estável, embora o número absoluto de observações (N =

2203) seja bem menor que nas demais categorias, porque, na prática, o número de alunos matriculados com essa condição é muito menor do que os demais.

Os valores mínimos e máximos reforçam a variação na evasão conforme o tipo de reserva de vaga. Todas as categorias apresentam cursos com evasão nula (mínimo = 0), indicando que há cursos nos quais a permanência foi total para aquele grupo. No entanto, os valores máximos são elevados: em alguns cursos, cerca de 80% dos estudantes que ingressaram por escola pública ou etnia evadiram, e cerca de 70% entre os que entraram pelas categorias "renda" ou "outros" também deixam a universidade antes de completar a graduação. A evasão máxima entre pessoas com deficiência também é alta (53,9%), mas ainda inferior às demais. Contudo, como dito anteriormente, a entrada desse grupo de alunos é baixa, tornando essas taxas, na prática, pouco expressivas.

Os alunos provenientes de escola pública, por vezes, enfrentam problemas de aprendizagem. Dias, Theóphilo e Lopes (2010) justificam que parte dos estudantes evadidos que alegaram dificuldades acadêmicas tinham estudado integralmente em escolas públicas, o que pode refletir a precária formação oferecida pelo ensino público.

Em relação ao grupo étnico, Gomes e Hirata (2022) afirmam que estudantes não brancos (negros e pardos, por exemplo) têm um risco de evasão 1,08 vezes maior em comparação com estudantes brancos. Especificamente, estudantes autodeclarados negros têm um risco de evasão quase 30% maior do que os brancos, mesmo quando outros fatores são controlados, indicando que a etnia continua a criar desigualdades mesmo após o ingresso no ensino superior (Nierotka, Salata e Martins, 2023).

Para o grupo de ingressantes por reserva de vagas por renda, Nierotka, Salata e Martins (2023) sugerem que o apoio financeiro por meio de programas de assistência estudantil, como transporte, moradia e alimentação (PNAES), podem reduz significativamente as chances de evasão. Guerra, Ferraz e Medeiros (2019) complementam que, em pesquisa, a minoria dos alunos evadidos utilizou a política de assistência social, sugerindo que os estudantes assistidos por essa política tendem a não evadir.

Esses resultados sugerem que, embora as políticas de reserva de vagas contribuam para o acesso ao ensino superior, elas não garantem, por si só, a permanência. A evasão expressiva entre alunos de políticas de reserva de vagas nas categorias como escola pública, renda e etnia pode refletir fragilidades socioeconômicas persistentes, falta de suporte institucional ou outras

barreiras enfrentadas após o ingresso. Por outro lado, a menor taxa de evasão entre pessoas com deficiência, embora promissora, deve ser interpretada com cautela, dada a menor representatividade desse grupo e a possível seletividade na entrada.

A elevada variabilidade entre cursos, evidenciada pelos desvios padrão e pelos máximos, aponta para a importância de políticas complementares ao ingresso, voltadas à permanência e ao acompanhamento individualizado dos estudantes beneficiados pelas ações afirmativas que estão experimentando dificuldades acadêmicas.

4.3.6 Gênero

A proporção de mulheres matriculadas nas instituições analisadas neste trabalho é de 52,8%, enquanto a proporção de homens é de 47,2%. Esses valores revelam uma leve predominância feminina no meio acadêmico. Contudo, a relação entre a composição de gênero e os índices de evasão apresenta nuances importantes.

A Tabela 9 revela a evasão média de acordo com a proporção de alunas matriculadas, dividindo as observações em tercis. Cada tercil representa as proporções de discentes do menor para o maior, de modo que 1 representa menor proporção, 2 representa média proporção e 3 representa alta proporção. A análise foi feita sob a perspectiva apenas do gênero feminino, pois esse parâmetro tem caráter complementar em relação ao gênero masculino (só existem essas duas possibilidades no banco de dados), dispensando a análise desse último.

Tabela 9- Evasão média pela proporção de matriculadas do gênero feminino dividida em tercis

	Tercil	Mean	SD
Fem	1 (menor prop. de mulheres)	.2108537	.1408038
	2 (média prop. de mulheres)	.2055622	.1356509
	3 (alta prop. de mulheres)	.1767614	.1243127

Tercil:	Divisão da proporção de discentes em 3 grupos
Mean:	Média da taxa de evasão
SD:	Desvio-padrão

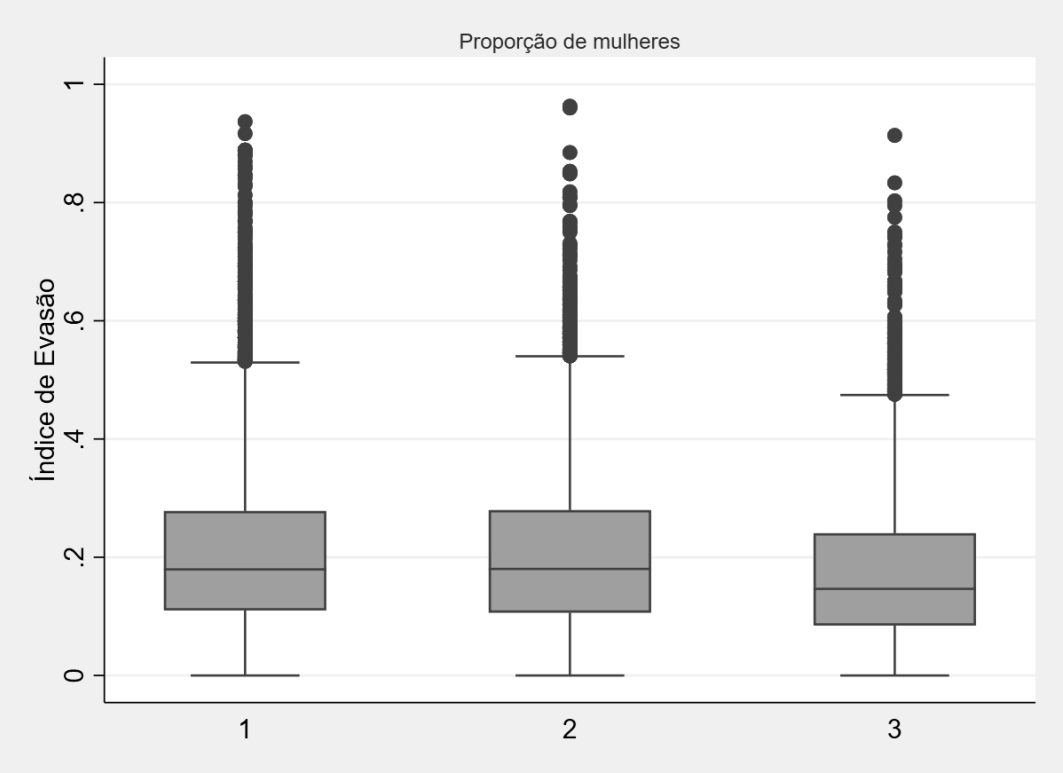
Fonte: resultados da pesquisa.

Ao dividir as observações da proporção de mulheres em tercis nota-se que a evasão média diminui conforme aumenta a presença feminina. Nos cursos com menor proporção de mulheres, como visto no 1º tercil, a evasão média foi de 21,1%. No grupo intermediário, 2º

tercil, essa média caiu para 20,6%, aproximadamente, e, nos cursos com maior proporção feminina, 3º tercil, a evasão média foi a menor, cerca de 17,7%. Esse comportamento sugere que a maior presença de mulheres pode estar associada a taxas menores de evasão.

Essas informações são visualmente corroboradas pelo Gráfico 11, que apresenta a distribuição das observações sobre a evasão pela proporção de matriculadas do gênero feminino, por meio de diagrama de caixas, divididas em tercis. A análise dos boxplots revela uma relação inversa entre a presença feminina e o índice de evasão. Ao observar a distribuição da evasão por tercis, verifica-se que os cursos com menor proporção de alunas (primeiro tercil) apresentam uma mediana de evasão mais elevada em comparação aos cursos com proporção intermediária (segundo tercil) e, especialmente, em relação aos cursos com maior presença feminina (terceiro tercil). Essa análise é importante porque a mediana, que é o ponto médio da amostra, sugere claramente uma tendência de redução nos níveis de evasão no box em que a participação feminina é mais concentrada.

Gráfico 11- Distribuição das observações sobre a evasão pela proporção de matriculadas do gênero feminino, por meio de diagrama de caixas, divididas em tercis



Fonte: resultados da pesquisa.

Além da mediana, nota-se que a dispersão dos dados, expressa pela amplitude interquartil e pela presença de valores extremos, é relativamente semelhante entre os três grupos. Contudo, o terceiro tercil, que representa os cursos com maior proporção de mulheres, aparenta concentrar um número ligeiramente menor de *outliers* no extremo superior, o que reforça a hipótese de que é mais raro encontrar índices de evasão extremamente altos em cursos com maior concentração de mulheres.

Essa menor probabilidade de evasão para as mulheres foi observada também em um estudo longitudinal com ingressantes de 2010 a 2017 (Gomes e Hirata, 2022). Além disso, em uma pesquisa com estudantes de alguns cursos específicos, como Ciências Contábeis, por exemplo, mostrou chance de evadir para estudantes do sexo masculino de 2,19 vezes maior do que para estudantes do sexo feminino (Lopes *et al.*, 2023).

Esses resultados indicam que, nos contextos analisados, há uma correlação negativa entre proporção de mulheres e evasão média, ou seja, quanto maior a participação feminina, menor tende a ser a evasão. O padrão denotado pelas estatísticas e pelo gráfico pode indicar que a presença feminina desempenha um papel moderador na permanência estudantil, sendo um elemento relevante a ser considerado em análises institucionais e políticas educacionais voltadas à redução da evasão.

4.3.7 Faixa etária

Tabela 10 apresenta as estatísticas descritivas da evasão em relação às proporções de estudantes matriculados por faixa etária.

Tabela 10- Estatísticas descritivas da evasão pela proporção de matriculados por faixa etária

Faixa Etária	N	Mean	SD	CV	P25	P50	P75	Min	Max
17 anos ou menos	1490	.1586879	.0971444	.6121731	.0936304	.1417329	.2025	0	.9166667
18 a 24 anos	9150	.1732746	.1055199	.6089753	.0959504	.1561243	.2338143	0	.8275862
25 a 29 anos	6818	.1598478	.092539	.5789194	.091498	.1485798	.2145036	0	.7524753
30 a 34 anos	5596	.1581769	.0911079	.5759873	.0913347	.1469989	.2121465	0	.7692308
35 a 39 anos	4498	.1588966	.0927526	.5837295	.0916115	.1457285	.2118831	0	.7524753
40 a 49 anos	4316	.170301	.1042099	.6119156	.0950255	.1526775	.2282159	0	.8461539
50 a 59 anos	2.778	.1751225	.1100419	.6283711	.0953846	.1553863	.2318008	0	.8461539

60 anos ou mais	1.435	.1827384	.1096573	.6000778	.1063218	.1625485	.2451225	0	.6551724

N:	Número de observações encontradas para evasão
Mean:	Média da taxa de evasão
SD:	Desvio-padrão
CV:	Coeficiente de variação
P25:	Percentil 25
P50:	Percentil 50
P75:	Percentil 75
Min:	Valor mínimo
Max:	Valor máximo

Fonte: resultados da pesquisa.

A análise das taxas médias de evasão em relação à proporção de matriculados por faixa etária revela padrões relevantes sobre a permanência no ensino superior em diferentes fases da vida. As faixas etárias mais elevadas concentram as maiores médias de evasão, com destaque para os grupos acima de 60 anos (18,27%) e de 50 a 59 anos (17,51%). Essa tendência decrescente de evasão conforme diminui a idade também é visível nos dados: estudantes entre 40 e 49 anos apresentam evasão média de 17,03%, enquanto os de 30 a 39 anos apresentam índices em torno de 15,8%. A menor evasão média está associada ao grupo de 0 a 17 anos, com 15,86%, embora represente um número menor de observações (N = 1490), o que requer uma análise mais cuidadosa desse grupo, já que a presença de alunos com menos de 18 anos na universidade é reduzida.

Os desvios padrão também acompanham essa tendência, revelando maior variabilidade nas taxas de evasão entre cursos para as faixas etárias mais velhas, chegando a 0,11 entre os cursos com estudantes de 50 a 59 anos, e a 0,109 nos com estudantes acima de 60 anos, indicando que, nesses grupos, a evasão varia bastante conforme o curso ou instituição. Em contraste, as faixas entre 25 e 39 anos, que concentram o maior número de observações, apresentam desvios padrão mais baixos, entre 0,091 e 0,093, o que sugere um comportamento mais homogêneo da evasão entre essas idades.

As medidas de posição central corroboram essa leitura. A mediana da evasão gira em torno de 0,14 a 0,15 em quase todos os grupos, o que indica uma distribuição relativamente simétrica. No entanto, ao observar os valores do terceiro quartil (Q3), especialmente nas faixas mais avançadas, nota-se que 25% dos cursos com estudantes acima de 60 anos apresentam evasão superior a 0,24, o que é expressivo. De modo semelhante, Q3 supera 0,23 nas faixas de

50 a 59 e 40 a 49 anos, evidenciando que, embora a média se mantenha razoável, há uma proporção relevante de cursos com evasão bastante elevada nesses grupos.

Todos os grupos apresentam valores mínimos de evasão igual a zero, o que indica que há cursos nos quais não se observou nenhuma evasão no período analisado. Por outro lado, os valores máximos variam de forma importante, com destaque para o grupo de 0 a 17 anos, que chega à evasão máxima de 91,6%, embora isso provavelmente reflita um número pequeno de casos extremos. Já entre os grupos de 18 a 49 anos, os máximos oscilam entre 75% e 84%, indicando que há cursos, mesmo entre faixas etárias com menor evasão média, com taxas de abandono consideravelmente elevadas.

Em concordância com essa pesquisa, uma análise sobre evasão e idade indicou um aumento no risco de evasão para estudantes de 21 a 24 anos em comparação com alunos mais jovens (Nierotka, Salata e Martins, 2023), tornando-se um ponto de reflexão, uma vez que essa é a faixa etária mais representativa dentro da universidade.

Gomes e Hirata (2022) apontam um estudo que analisou dados longitudinais de universitários ingressantes de 2010 a 2017. Os resultados indicam que ingressantes mais velhos possuem menor risco de evasão. Todavia, esse grupo apresenta resultado inverso no período dessa pesquisa, 2018 a 2023, o que indica que o período pandêmico pode ter influenciado esses índices.

Esses resultados indicam que a faixa etária exerce influência sobre a evasão, especialmente nos extremos. Estudantes mais velhos, ainda que representem menor volume, apresentam maior tendência à evasão e maior variabilidade entre os cursos, possivelmente em função de fatores como conciliação entre estudo e trabalho, responsabilidades familiares, desafios com tecnologias ou menor integração institucional. Por outro lado, os cursos com predominância de estudantes jovens, embora com algumas exceções, tendem a apresentar menor evasão média e menor variação entre casos, o que pode refletir tanto maior engajamento inicial quanto políticas institucionais mais bem direcionadas a esse público.

Em termos de formulação de políticas públicas, os dados apontam para a importância de ações diferenciadas de permanência, que levem em conta as especificidades etárias dos estudantes e as condições estruturais dos cursos em que estão inseridos. Mas sem deixar de se considerar a forte influência de características pessoais que podem estar por trás desses índices,

já que as pessoas mais velhas, via de regra, possuem menos argumentos para concluir um curso de graduação, uma vez que estão em outro estágio da vida.

4.4 RESULTADOS PARA OS DADOS EM PAINEL

Os coeficientes resultantes da regressão utilizando o modelo de equações de estimação generalizadas (GEE) com distribuição binomial, função de ligação "logit", estrutura de correlação "ar1" e erros robustos estão apresentados na Tabela 11. Foram assumidas variáveis *dummies* para representar variáveis categóricas. Os coeficientes indicam o quanto a evasão média em cada área difere da área de referência em pontos percentuais (p.p.), controlado pelos demais fatores do modelo

Foram retiradas da regressão para evitar multicolinearidade perfeita as variáveis propmasc, prop18a24, propdiu, dg10 (área Saúde e bem-estar) e dreg4 (região Sudeste). As excluídas representam variáveis de referência para as análises das regressões. A constante apresenta o valor médio de todas as categorias que foram omitidas na estimação do modelo. No grupo de proporções de matriculados por reserva de vagas não foi excluída nenhuma variável. Isso porque o grupo não abarca a totalidade de alunos matriculados, podendo haver também discentes matriculados por outras formas de ingresso que não a reserva de vagas, por exemplo, ampla concorrência ou transferência externa. Dessa forma, não ocorre o problema da multicolinearidade perfeita dentro desse grupo.

Tabela 11- Coeficientes resultantes da regressão utilizando modelo de equações de estimação generalizadas (GEE) com distribuição binomial, função de ligação "logit", estrutura de correlação "ar1" e erros robustos, nos 4 períodos analisados

	C1: todo o período	C2: pré-covid	C3: covid	C4: pós-covid
Média evasão	.1977258	.1940976	.1936298	.2054499
CV	.68071537	.63267118	.74993908	.65480767
Wald chi^2 (X)	2490.83	639.30	1156.02	2210.40
(Prob > chi^2 = 0.0000)	(34)	(28)	(34)	(32)
propfem	-.1288954**	-.3483443**	-.0690701	-.3116563*
prop0a17	9.380611*	7.900508*	15.10884*	12.07854*
prop25a29	-.9062395*	-1.054288*	-1.456314*	-.608329*
prop30a34	-.8869397*	-1.167272	-1.470132*	-.0210755
prop35a39	-.3968079	1.743683	-.2579279	-.3174327
prop40a49	.6790236	1.09523	.3516967	.9303209
prop50a59	3.526481*	5.245386*	4.623109*	3.184801*
prop60mais	2.021287**	-.1122027	4.356468*	1.445697

propnot	.2815054*	(omitted)	.3267528*	.1703273*
lnconc	-.0341619*	-.0741469**	-.0390381**	-.1004267*
propescopub	-.2394057*	-.2026603	-.7050226*	.0347496
propetn	-.2843431**	.4780288**	.6416547*	-.3952725**
proppdef	-2.057186***	-.8009767	-2.531535	-.7144636
proprenda	.5358608*	-.0999524	1.192972*	-.1395539
propoutros	.2160265	-1.812366*	-.1381871	.5597618*
lnconc_hat	-.2191185*	-.5117176*	-.0955733*	-.3532211*
propescopub_hat	-1.24258*	-1.843612**	-.8106708***	.0378022
propetn_hat	.3394512	4.77747*	-1.740437*	2.288081*
proppdef_hat	-19.77077*	-18.03111***	-22.66643*	-3.748199
proprenda_hat	.6355176	1.164652	.8579891	-1.08395
propoutros_hat	-3.179631*	-2.171366	-3.432187*	-.3614428
dg1	.0464245	-.0590086	.1772333**	(omitted)
dg2	.1085972***	-.3644149*	.301875*	.0198971
dg3	.3535997*	.1579893***	.4636107*	.1546413**
dg4	.1441368*	(omitted)	.2459799*	.0425673
dg5	.1197685***	.048408	.2065458**	-.0501618
dg6	.1807266*	-.0816206	.3274877*	.0135472
dg7	.0669168	-.1334198	.1769367**	(omitted)
dg8	.0583338	(omitted)	.0611569	.0905295***
dg11	.1189976***	(omitted)	.1662571	.0670391
dreg1	.2041089*	-.1513647	.3507591*	.007575
dreg2	.2968689*	-.155657	.5242665*	-.0920136
dreg3	.4425602*	(omitted)	.9345163*	-.0003345
dreg5	.2822047*	(omitted)	.0767401	.4791003*
covid	.0345825***	-	-	-
poscovid	.1282054*	-	-	-

* significativo a 1% de significância
** significativo a 5% de significância
*** significativo a 10% de significância

Fonte: resultados da pesquisa.

Algumas variáveis foram omitidas em certos períodos, o que pode sugerir falta de variação nos dados ou colinearidade com outras variáveis. O significado dos termos na saída dos modelos estimados estão explicados em nota[2].

A média da evasão em cada um dos períodos analisados, pré-pandemia de COVID (2018-2019), durante a pandemia de COVID (2020-2021) e pós-pandemia de COVID (2022-

[2] O Wald Chi² e respectivo p-valor indicam que o conjunto das variáveis explicativas é estatisticamente significativo em todos os períodos (Prob > chi² = 0.0000). O coeficiente de variação (CV) mede a dispersão relativa da evasão em cada período, ou seja, a variação relativa da evasão ao longo do tempo e entre as unidades. A unidade é representada por curso. O número entre parênteses (34, 28, 34, 32) representa o número de variáveis testadas. Cada linha é uma variável explicativa (propfem, prop0a17, propnot, lnconc etc.) com o coeficiente estimado para cada período. O sinal indica a direção da relação com a evasão (positivo = aumenta evasão; negativo = reduz evasão). Em alguns casos aparece (*omitted*), que indica que a variável foi omitida por colinearidade perfeita ou por não variar no período específico.

2023), não destoam muito da média de todo o intervalo (~19,77%), observando-se um ligeiro aumento no período após a pandemia (~20,54%).

Os coeficientes de variação (CV) se mostraram altos em todos os períodos, mostrando dispersão relativa em relação à média bastante significativa. Ou seja, os desvios-padrão em relação às médias dos 4 intervalos foram de 0.6807, 0.6327, 0.7499 e 0.6548, respectivamente. Eles apresentam alta variabilidade na taxa de evasão entre as unidades (cursos).

Os resultados de Wald Chi^2 (Prob>Chi^2 =0.0000) indicam que as variáveis independentes em conjunto têm efeito explicativo sobre a variável dependente. Em outros termos, o Wald Chi^2 testa as hipóteses:

- H_0: todos os coeficientes (exceto o intercepto) são iguais a zero. Indica que nenhuma variável independente explica a evasão;
- H_1: pelo menos um coeficiente é diferente de zero, ou seja, alguma variável tem efeito sobre a evasão.

Quanto maior o valor de Wald Chi^2 mais forte é a evidência para rejeitar H_0. Mas esse valor está atrelado à quantidade de variáveis da amostra e, portanto, é mais relevante observar sua significância. Nessas regressões o Wald Chi^2 foi significativo em todos os 4 períodos, isso significa que a hipótese nula deve ser rejeitada e assume-se que o modelo tem poder explicativo global.

4.4.1 Efeitos das áreas gerais do conhecimento

Para os efeitos das áreas gerais na evasão, em todo o período (2018 a 2023), as variáveis apresentaram coeficientes positivos significativos, com exceção das áreas de Agricultura, silvicultura, pesca e veterinária, Engenharia, produção e construção e Negócios, administração e direito. Isso indica que as demais áreas têm evasão sistematicamente maior que a área comparada (Saúde e bem-estar), com destaque para a área de Ciências naturais, matemática e estatística (dg3) e Educação (dg6) que apresentam os maiores coeficientes. Os cursos da área de Ciências naturais, matemática e estatística têm, em média, uma evasão 0,3536 p.p. maior que os cursos da saúde, mantendo-se todas as outras áreas constantes. Uma pesquisa realizada com dados de 2009 a 2014 mostrou que cursos de Matemática e Engenharias Metalúrgica e

Mecânica apresentaram taxas de evasão próximas a 90% e 60% respectivamente, preocupantes devido à relação com a geração de inovações tecnológicas e aumento da produtividade (Saccaro, França e Jacinto, 2019). Isso pode se justificar em razão das fragilidades da formação de base (ensinos fundamental e médio). Além disso, Saccaro, França e Jacinto (2019) justificam que os currículos dos cursos de graduação de Engenharia, com muitos componentes curriculares das áreas de Ciências Exatas (matemática e física), geram altas taxas de reprovação e podem explicar resultados de evasão.

Já os cursos de Educação apresentaram uma tendência de 0,1807 p.p. a mais de evasão que a referência. Esse efeito é estatisticamente robusto e não se deve ao acaso. As formações em licenciatura, em geral, sofrem com baixa remuneração no mercado de trabalho. As demais áreas também apresentam, em média, evasão maior que a área base, mas com valores percentuais mais moderados: de aproximadamente 0,1086 a 0,1441 p.p.

Para o período anterior à pandemia, o modelo não pode explicar a evasão em muitas das áreas por não apresentar significância estatística ($p>0,1$). Contudo, analisando aquelas que apresentaram significância estatística, verifica-se que a área Ciências naturais, matemática e estatística apresenta novamente uma tendência a uma maior evasão em relação à área de Saúde e bem-estar, 0,158 p.p. Por outro lado, a área de Artes e humanidades aparenta uma tendência de queda no índice de evasão de 0,3644 p.p. em relação à área de Saúde e bem-estar no período anterior à pandemia.

Para o período de pandemia, todas as variáveis relacionadas às áreas se mostraram significativas, menos as áreas de Negócios, administração e direito e Serviços. As áreas que apresentam maior tendência de evasão em relação à Saúde continuam sendo Ciências naturais, matemática e estatística e Educação, com a inclusão agora da área de Artes e humanidades que, ao contrário do período anterior marcado por forte queda na evasão, nesse período apresenta evasão aproximadamente 0,3019 p.p. maior que a referência.

Os alunos dos cursos de Agricultura, silvicultura, pesca e veterinária; Ciências sociais, comunicação e informação; Computação e Tecnologias da Informação e Comunicação (TIC) e Engenharia, produção e construção têm, em média, uma tendência a evadir maior do que os alunos de Saúde, 0,1772, 0,2460, 0,2065 e 0,1769 p.p., respectivamente.

Em relação ao período pós-pandemia, apenas as áreas Ciências naturais, matemática e estatística e Negócios, administração e direito apresentaram significância estatística e, portanto,

a evasão só pode ser explicada por essas áreas nesse período. A primeira permanece a área mais afetada pela evasão com um coeficiente que indica 0,1546 p.p. a mais de evasão do que a área base. Para os cursos pertencentes à segunda área, a regressão mostra uma evasão em torno de 0,09 p.p. maior que a área da Saúde no pós-pandemia.

Em resumo, é possível verificar que a área de Ciências naturais, matemática e estatística, em qualquer tempo, se apresenta com a maior probabilidade de evasão em relação à referência. Contudo, é perceptível uma oscilação marcada por um brusco aumento no período pandêmico, chegando a uma tendência 3 vezes maior que nos períodos anterior e posterior à pandemia. Essa taxa de evasão significativamente maior do que os da área de Saúde pode indicar desafios específicos, como dificuldades acadêmicas, menor alinhamento vocacional, problemas com o nível de abstração dos cursos ou problemas estruturais que impactam a permanência dos alunos (Barros, 2022; Hoed, 2016; Dias, Theóphilo e Lopes, 2010).

Esse dado indica a necessidade de ações de apoio acadêmico, como tutorias e reforço em conteúdos de maior dificuldade, além da promoção de metodologias ativas e de aprendizagem colaborativa.

A área de Artes e humanidades, embora não apresente significância no último período, revela um comportamento curioso nos períodos anteriores. No período de 2018 a 2019 apresentava uma tendência de evasão muito menor do que a área de Saúde (coeficiente negativo). Entretanto, nos 2 anos seguintes o resultado mostra exatamente o inverso. Demonstra um comportamento de instabilidade dos discentes dos cursos dessa área.

Os cursos da área de Educação também aparentam um ponto de atenção no que tange à evasão. Os índices podem estar relacionados à desvalorização da profissão com baixa remuneração e perspectivas de carreira (Feitosa, 2016), além até mesmo ao perfil dos estudantes, já que os cursos de educação no Brasil costumam atrair alunos que já trabalham ou que tenham família, o que pode dificultar a continuidade dos estudos. Nesse sentido, muitos estudantes percebem que a realidade da profissão docente envolve desafios como dificuldades no ambiente escolar e baixo prestígio social e valorização financeira (Adachi, 2009), e acabam desistindo do curso. Para isso é importante desenvolver políticas de valorização profissional, com oferta de melhores postos no mercado de trabalho, o que pode, inclusive, melhorar o reconhecimento social da carreira.

As áreas Ciências sociais, comunicação e informação e Computação e Tecnologias da Informação e Comunicação (TIC), embora não apresentem algum nível de significância nos períodos pré e pós-pandemia separadamente, revelam que a evasão no período da pandemia é significativamente superior se comparado com todo o período de 2018 a 2023, chegando a quase o dobro. Isso pode decorrer de fatores como a desmotivação de discentes de cursos que dependem de interação presencial, como Comunicação e Ciências Sociais. Além disso, alunos daqueles cursos que envolvem discussões, debates e trabalho em grupo, como Ciências Sociais e Comunicação, podem ter perdido parte de sua essência na modalidade de ensino remoto. Ainda nesse sentido, pode-se pressupor que nos cursos com disciplinas mais técnicas, como Ciência da Computação, por exemplo, a falta de suporte adequado para aprendizado prático e projetos colaborativos pode ter impactado negativamente a experiência acadêmica.

4.4.2 Efeitos das regiões geográficas

A Tabela 11 revela, ainda, o comportamento dos cursos de graduação presentes nas regiões geográficas do Brasil tomando-se por base a região Sudeste omitida na regressão. Os coeficientes do período anterior à pandemia não apresentaram significância estatística e, por essa razão, não foram conclusivos.

A análise da evasão acadêmica com base nesses dados de diferentes regiões geográficas é crucial por vários motivos, principalmente para entender as disparidades regionais e melhorar a gestão educacional. Os coeficientes da regressão medem o efeito relativo em comparação com a referência (Sudeste). Eles indicam o impacto de cada região na evasão em relação ao Sudeste, e não a taxa absoluta de evasão.

Para o período de 2018 a 2023, a regressão mostra que a região Centro-Oeste teve uma tendência de 0,4426 p.p. a mais de evasão do que a região Sudeste. Enquanto que nas regiões Nordeste, Sul e Norte os coeficientes indicam que estar matriculado em um curso de graduação nessas regiões aumenta a evasão em 0,2969 p.p., 0,2822 p.p. e 0,2041 p.p., respectivamente, comparado à região Sudeste.

A regressão revela, ainda, que no período de 2020 a 2021 esse comportamento se intensificou, indicando os efeitos na evasão que cada região apresentou em relação ao Sudeste durante a pandemia. Os coeficientes mostram que a região Centro-Oeste impactou 0,9345 p.p. mais a evasão do que a região Sudeste nesse período.

De modo geral, durante a pandemia as tendências a evasão se mantiveram positivas, porém com coeficientes bem mais elevados, indicando um aumento muito significativo dos efeitos da evasão em todas as regiões comparadas ao Sudeste, com exceção da região Sul que não apresentou significância estatística e, portanto, não se pode fazer inferências.

A Tabela 12 apresenta a incidência de casos de COVID-19 e óbitos pela doença, a cada 100.000 habitantes, em todas as grandes regiões geográficas do Brasil, acumulados até janeiro de 2024. Por meio dela é possível se fazer correlações entre a evasão nas regiões do Brasil, durante e após a pandemia, com a incidência de casos e óbitos.

Tabela 12- Número, incidência de casos e óbitos por COVID-19 segundo as Grandes Regiões do Brasil - Dados acumulados até 29/01/2024.

Grandes Regiões do Brasil	Casos		Óbitos	
	Acumulados	Incidência de casos/ 100 mil habitantes	Total	Óbitos/ 100 mil habitantes
Brasil	38 303 320	18 226,90	709 195	337,50
Norte	2 941 127	15 957,50	51 886	281,50
Nordeste	7 518 402	13 173,60	136 557	239,30
Sudeste	15 273 600	17 283,40	341 492	386,40
Sul	8 137 841	27 147,90	112 488	375,30
Centro-Oeste	4 432 350	27 197,20	66 772	409,70

Fonte: Anuário Estatístico do Brasil – IBGE: Painel Covid-19. Ministério da Saúde (IBGE, 2024).

Os dados mostram que a região Centro-Oeste se mostrou com maior incidência de casos e óbitos a cada 100.000 habitantes. Essa situação pode também ter influenciado o comportamento da evasão acadêmica.

Apesar de os coeficientes da região Centro-Oeste serem consideravelmente maiores, não significa, necessariamente, que sua taxa de evasão absoluta seja a maior. Significa apenas que, quando comparado ao Sudeste, essa região tem um efeito maior de aumento na evasão. Isso porque a regressão isola o impacto da variável regional enquanto controla para outros fatores, ou seja, mostra o efeito regional comparativo, não apenas a taxa observada, o que pode fazer os coeficientes divergirem da média. Nesse sentido, apesar de não ter a maior média de evasão, a região Centro-Oeste pode apresentar um efeito mais forte, principalmente durante a pandemia, porque sua evasão, quando comparada ao Sudeste, pode ser particularmente mais sensível a fatores estruturais, como a menor oferta de instituições (8 Universidades Federais no Centro-Oeste, em um total de 69 no Brasil), disparidade econômica e desafios no ensino remoto, por exemplo (Costa e Silva, 2024).

Após a pandemia, a região Sul apresentou um aumento expressivo nos efeitos da evasão em comparação à base, possivelmente indicando efeitos pós-pandemia, como impactos psicológicos e emocionais além do aumento das desigualdades sociais, que podem favorecer a desconexão dos estudantes com a universidade. Nos anos de 2022 e 2023, a região apresentou 0,4794 p.p a mais de efeitos sobre a evasão se comparado ao Sudeste. Isso pode ser decorrente dos elevados números de óbitos e incidência de casos, proporcionais à população na região Sul, conforme Tabela 12. Esses índices ultrapassaram as médias nacionais no período acumulado até janeiro de 2024.

A região Norte, por sua vez, apresenta-se como aquela com menor efeito em relação à região Sudeste no que tange à evasão acadêmica. Todavia, esses resultados podem ir de encontro com os pressupostos dos efeitos que a pandemia pode ter deixado na educação superior. Mendonça *et al*. (2020) explicam que a região Norte registrou as proporções mais baixas de médicos, leitos de UTI e ventiladores, enfrentando dificuldades com a doença. Conforme a Tabela 12, os números de incidência e mortes por COVID-19 se mostraram expressivos, apesar da baixa densidade populacional, o que em tese, favoreceria a baixa contaminação. Orellana *et al*. (2020) afirmam que a região foi duramente atingida pela COVID-19 como resultado de vários fatores, incluindo a resistência do governo local e federal em impor medidas de mitigação não farmacológicas, como distanciamento social e uso de máscaras, impulsionando o surgimento de uma nova variante altamente transmissível.

Na região Nordeste, embora tenha apresentado as taxas proporcionais de incidência e de morte por COVID-19 bem abaixo da média nacional; durante o período de 2020 a 2021, o resultado da regressão mostra que houve um aumento expressivo da influência dessa região na evasão acadêmica em relação à referência, comprando-se ao período todo.

Os fatores e desfechos da pandemia COVID-19 são influenciados por desigualdades sociodemográficas, localização geográfica, ideologia política e religiosa (Brito *et al.*, 2023). Sendo assim, cada região do país apresentaram suas respostas à evasão acadêmica. Nesse sentido, o perfil socioeconômico dos estudantes, a infraestrutura educacional e as políticas regionais são alguns dos fatores que podem ter influenciado a evasão de forma diferente em cada região do Brasil.

Esses resultados reafirmam a existência de desigualdades estruturais entre regiões e reforçam a necessidade de ações governamentais de financiamento diferenciado, com foco em

equidade. Ações essas que têm a finalidade de promover o fortalecimento da infraestrutura das instituições de ensino superior nessas regiões, aumentar o investimento na inclusão digital e o aumento da cobertura da assistência estudantil.

4.4.3 Efeitos da concorrência

Os valores apresentados na Tabela 11 mostram o resultado das regressões da variável concorrência linearizada. Para o período de 2018 a 2023, o coeficiente -0.0341619 indica que um aumento de 1 unidade no logaritmo da concorrência está associado a uma redução média de aproximadamente 0,0342 unidades na evasão, ou seja, 3,42 pontos percentuais, devido ao sinal negativo, mantendo todas as outras variáveis constantes, apresentando-se estatisticamente significativo a 1%. Por ser uma transformação logarítmica, um aumento de 1 unidade em "lnconc" significa que a concorrência foi multiplicada por aproximadamente 2,718 vezes (porque o log é na base *e*, ou seja, e^1 é aproximadamente 2,718). Ou seja, quando o número de candidatos por vaga aumenta a ponto de o ln da concorrência subir 1 unidade, que significa que essa concorrência quase triplicou (2,7 vezes). Assim, a evasão média cai cerca de 3,42 pontos percentuais. Em outros termos, se um curso que apresenta uma concorrência de 10 candidatos por vaga, por exemplo, aumentar sua procura ao nível de 27,18 candidatos por vaga, as taxas de evasão tendem a reduzir cerca de 3,42 p.p.

O resultado para o período da pandemia, 2020 a 2021, foi relativamente semelhante ao período todo: O aumento de 1 unidade no logaritmo da concorrência está associado a uma redução média de aproximadamente 3,9 pontos percentuais na evasão.

Observa-se que este efeito foi mais intenso ainda nos períodos 2018 a 2019 e 2022 a 2023. O aumento de 1 unidade no logaritmo da concorrência está associado a uma redução média de aproximadamente 7,42 pontos percentuais na evasão para o primeiro período citado e 10 p.p. para o segundo.

O período pós-pandemia apresenta um coeficiente negativo quase 3 vezes maior que o do período inteiro. Isso sugere que o efeito pós-pandemia foi significativo no que se refere à relação de redução da evasão quanto se aumenta a concorrência nos cursos. Esse fenômeno pode ocorrer devido à fatores como maior motivação dos alunos quando aprovados em cursos mais concorridos ou de maior prestígio e, com a retomada das aulas presenciais a pressão

acadêmica por se manter no curso de maior concorrência exigem maior esforço por parte dos alunos (Bohnenberger, 2025).

De modo geral, as regressões indicam que a concorrência nos cursos pode estar intimamente ligada à redução da evasão acadêmica (Adachi, 2009). Isso pode explicar o porquê de a área de saúde apresentar menor evasão, visto que os cursos de medicina têm alta procura no Brasil.

Isso reforça a importância de fortalecer o suporte em cursos com menor concorrência, com iniciativas que promovam maior valorização do curso e reestruturação pedagógica. Contudo, é relevante analisar também as políticas de criação de novas vagas, com análises mais aprofundadas sobre a real necessidade da oferta, uma vez que um dos parâmetros que influenciam a concorrência é a disponibilidade de vagas (relação inversa).

4.4.4 Efeitos do turno

No período anterior à pandemia (2018-2019) os dados foram omitidos, possivelmente por falta de variação suficiente ou colinearidade com outras variáveis no modelo pré-pandemia (C2). Dessa maneira, não é possível extrair conclusões para esses anos.

Nos resultados referentes ao período de 2018 a 2023, a cada aumento de 1 unidade na proporção de alunos no turno noturno (ou seja, passando de 0% para 100% dos alunos no noturno) está associado a um aumento de aproximadamente 28,15 pontos percentuais na taxa de evasão, em relação ao turno diurno. Ou seja, a cada aumento de 10 pontos percentuais na proporção de alunos no turno noturno (ex: de 20% para 30% de alunos no noturno), a evasão aumenta, em média, cerca de 2,8 pontos percentuais.

Ainda nessa tendência, o período da pandemia apresentou coeficiente ainda maior, inferindo em um aumento de 3,3 p.p., aproximadamente, à medida que a taxa de alunos em cursos noturnos aumenta 10 p.p., quando comparados ao turno diurno. Essa tendência pode estar relacionada a características específicas dos alunos do turno noturno como necessidade de trabalhar concomitantemente aos estudos, maior idade média, responsabilidades familiares e menor tempo disponível para os estudos. Fatores esses que podem ter gerado ainda mais dificuldades considerando as instabilidades econômicas provocadas pela pandemia de Covid-19 (Barros, 2022).

No período pós-pandemia o coeficiente reduziu quase à metade, mas ainda indica relação positiva da evasão com o aumento de alunos que estudam à noite: 1,7 p.p. de aumento da evasão a cada 10 p.p. de aumento na proporção de matriculados em cursos noturnos.

Os resultados apontam de forma robusta que a proporção de discentes no turno noturno estão mais propensos à evasão acadêmica, mesmo após o retorno às aulas presenciais. A diminuição da intensidade pode denotar uma readaptação às rotinas acadêmicas, porém ainda revela a necessidade de atenção a esse grupo.

Em geral, eles são mais vulneráveis social e economicamente. Normalmente são mais propensos a trabalhar durante o dia e com isso têm mais chance de abandono por conciliação estudo-trabalho (Trópia e Souza, 2023; Pires, 2021). A pandemia teve maior impacto sobre esse público no que se refere à evasão acadêmica, isso pode decorrer de dificuldades de acesso a ensino remoto ou problemas como desemprego, por exemplo.

O estudo reforça a necessidade de políticas específicas voltadas ao grupo de estudantes do turno noturno, que historicamente enfrenta maiores dificuldades de permanência e conclusão do curso, normalmente pela dificuldade em conciliar trabalho e estudo (Feitosa, 2016). Este efeito pode estar associado a sobrecarga gerada pela conciliação entre trabalho e estudo, somada à possibilidade de menor oferta de serviços institucionais no período da noite. Nesse sentido, seria importante ampliar o atendimento acadêmico, psicológico e social no turno noturno, tanto em questões administrativas, quanto em disponibilidade docente, além de considerar estratégias como o ensino híbrido parcial e programas de apoio financeiro específicos para esse público, fornecendo a possibilidade de redução da jornada de trabalho.

4.4.5 Efeitos da política de reservas de vagas

No período de 2018 a 2023 a evasão em cursos de graduação por alunos matriculados por meio de reserva de vagas (RV) em ações afirmativas se mostrou negativa para os grupos: RV para alunos de escola pública, RV por etnia e RV para alunos com deficiência. Para cada ponto percentual de aumento na proporção de alunos de RV para escola pública espera-se uma redução de 0,24 p.p. na evasão. Essa tendência segue na redução 0,28 p.p. e 2,06 p.p. no grupo de RV por etnia e deficiência, respectivamente. Isso indica que, mesmo controlando a evasão pelas demais variáveis (idade, região, turno, cotas, pandemia etc.), o efeito de ter mais alunos por esses grupos de reserva de vagas está associado a menor evasão.

Esses dados podem indicar que os discentes com deficiência talvez recebam maior atenção institucional, adaptação curricular ou têm um perfil altamente resiliente. Todavia, geralmente o número de alunos com deficiência é baixo e a interpretação quantitativa precisa ser feita com cuidado, pois pequenas mudanças nas matrículas podem gerar grandes efeitos no coeficiente.

Ao contrário disso, a regressão indica que cada ponto percentual de aumento na proporção de alunos matriculados por reserva de vagas por renda familiar sugere um aumento de 0,54 p.p. na evasão acadêmica, no período de 2018 a 2023.

Durante o período da pandemia, o grupo de RV por escola pública indica uma redução na evasão ainda mais vultuosa, ao passo que o grupo de RV por renda familiar também apresenta uma tendência ao aumento da evasão ainda mais intensa.

Isso pode insinuar que alunos de escola pública que acessam a universidade por cota podem ter maior motivação ou apoio institucional. Ao passo que alunos de RV de baixa renda podem ter mais dificuldades materiais para permanecer (mesmo com a cota de ingresso), devido à necessidade de trabalhar, devido à outras carências como falta de transporte, moradia, alimentação etc, reforçando a hipótese de que a vulnerabilidade econômica agravou a evasão durante a pandemia.

O grupo de discentes matriculados por RV por etnia que, ao período todo mostrou relação inversa com a evasão, quando analisados os períodos separadamente foi possível notar instabilidade. Nos períodos anterior e durante a pandemia, a cada ponto percentual de aumento na proporção de alunos nesse grupo foi associado um aumento de 0,48 e 0,64 p.p. na evasão, respectivamente. Contudo, após a pandemia a regressão mostrou índice negativo, ou seja, associação inversa entre a proporção do grupo e a evasão, o que pode refletir desigualdade no acesso ao ensino remoto durante a pandemia, mas maior suporte social, psicológico ou financeiro, que favorecem a permanência, após a pandemia.

Os alunos matriculados por meio de “outros” tipos de ações afirmativas apresentaram relação inversa com a evasão antes da pandemia de COVID-19, mas relação direta com ela após a pandemia (2022-2023). Essa oscilação sugere que a heterogeneidade dessa categoria dificulta conclusões consistentes. Esse grupo de reserva de vagas pode abarcar diversas modalidades, e seus efeitos em relação à evasão podem sofrer interferências diferentes de acordo com contextos institucionais ou regionais específicos.

De modo geral, os resultados sugerem que a política de reserva de vagas por ações afirmativas tem efeitos diferenciados sobre a evasão, dependendo do critério de inclusão e do período analisado. A existência de discentes que ingressaram por meio de políticas de cotas por escola pública, etnia e deficiência está associada à redução da evasão, sugerindo a eficácia das políticas para permanência. Apesar disso, o grupo associado à etnia mostrou volatilidade marcada pela influência positiva na evasão no período pré e durante a pandemia, e negativa após a crise sanitária.

Por outro lado, a reserva de vagas por renda apresenta relação positiva com a evasão, indicando que a condição socioeconômica dos alunos ainda representa uma barreira significativa à permanência, especialmente durante a pandemia. Os alunos matriculados nessa condição se mostram fragilizados em relação à permanência devido à vulnerabilidade econômica. Valesan *et al* (2024) afirmam que grande parte dos estudantes tiveram sua renda reduzida durante o isolamento, o que agravou as condições de permanência e desempenho acadêmico. Nesse sentido, Martínez-Líbano e Yeomans-Cabrera (2023) complementam que o status socioeconômico prejudicado durante a pandemia foram identificados como um dos principais preditores do risco de evasão. Esses resultados reforçam a importância de combinar ações de acesso com políticas de permanência estudantil, especialmente para grupos economicamente vulneráveis (Silva, 2022).

Em relação aos alunos que ingressaram por meio de reserva de vagas por escola pública, os índices de evasão podem ser um reflexo da precária formação escolar, devido à deficiência do sistema de ensino básico do país, que é um fator determinante das dificuldades enfrentadas por eles (Dias, Theóphilo e Lopes, 2010). De acordo com Jardim e Almeida (2016), a escola pública muitas vezes não consegue cumprir a grade curricular do ensino médio exigida na maioria dos exames de seleção das universidades, o que sempre a colocou em desvantagem competitiva frente às escolas de elite da rede privada. Dias, Theóphilo e Lopes (2010) corroboram com essa afirmação por meio um estudo que mostra que mais de 60% dos alunos com dificuldades acadêmicas haviam estudado integralmente em escolas públicas.

No que tange o grupo étnico, Mendes Junior (2014) afirma em um estudo feito na Universidade Estadual do Rio de Janeiro que, apesar de os cotistas apresentarem um desempenho médio inferior aos não cotistas, e aqui se incluem os ingressantes por reserva de vagas por etnia, eles estão se formando em taxas maiores. A hipótese é que a utilidade da faculdade é maior para os cotistas, o que se traduz em maior persistência. No entanto, em outro

estudo de mesmo caráter, Nierotka, Salata e Martins (2023) contataram que os discentes homens e autodeclarados negros evadem mais frequentemente. Esses contrastes trazem o reflexo do comportamento instável em relação à evasão que esse grupo apresentou durante os períodos analisados.

Esses dados indicam a importância de políticas étnico-raciais de permanência, especialmente em contextos de crise. Para estudantes de baixa renda, por sua vez, reforça-se a necessidade de fortalecimento da assistência estudantil, com ampliação de auxílios financeiros, apoio à inclusão digital, moradia, alimentação e transporte, visando à redução da vulnerabilidade econômica e o risco de evasão.

Barros (2022) afirma que as ações afirmativas são mecanismos político-sociais que visam proporcionar igualdade de oportunidades a grupos socialmente marginalizados. Barros (2015) explica que essas políticas têm o intuito de anular os efeitos culturais e comportamentais que a discriminação de determinados grupos gerou ao longo dos anos. Sendo assim, os resultados da pesquisa propõem aos gestores públicos uma maior análise da complexidade da efetividade dessas políticas, que dependem da combinação de medidas de acesso e permanência, reforçando que o ingresso à universidade, por si só, não garante o sucesso das políticas de educação (Silva, 2022; Junior *et al.*, 2019).

4.4.6 Efeitos do gênero

A variável referente à proporção de discentes matriculados do gênero masculino foi omitida nos modelos devido à multicolinearidade perfeita com a variável de proporção de alunas matriculadas, dado que juntas somam 1, e, portanto, não é necessário incluí-la simultaneamente.

Em todos os períodos os coeficientes da regressão indicam que uma maior proporção de discentes do sexo feminino está associada à menor evasão em relação a discentes do sexo masculino, com exceção do período durante a pandemia que não apresentou significância estatística e, portanto, não se pode fazer inferências.

Para o período todo, tem-se que o aumento de 1 ponto percentual na proporção de alunas mulheres está associado a uma redução de 0,13 p.p. na taxa de evasão, comparado à proporção de alunos do sexo masculino. Esse resultado é estatisticamente significativo ao nível de 5%, o que reforça a robustez da associação.

Nos períodos anterior e posterior à pandemia o coeficiente foi ainda mais elevado, apontando que o aumento de 1 ponto percentual na proporção de alunas diminui a taxa de evasão em 0,35 e 0,31 p.p., respectivamente, em relação à referência, mas sem denotar uma mudança muito abrupta entre os dois períodos. Os resultados são estatisticamente significativos ao nível de 5% e 1%, respectivamente. Esses valores indicam que, antes da pandemia, o efeito da presença feminina sobre a evasão era mais intenso, podendo estar relacionado a características como maior comprometimento acadêmico ou maior facilidade de adesão às rotinas institucionais. No período pós-pandemia (2022–2023), embora o nível de significância seja inferior ao observado nos períodos anteriores, 10%, o efeito é relevante e sugere uma constância da tendência de menor evasão em cursos com maior presença feminina. Esse padrão pode estar vinculado à recuperação gradual da estabilidade acadêmica após um período de crise sanitária, além da reorganização das condições socioeconômicas e institucionais.

De forma geral, os resultados indicam que uma maior proporção de discentes do sexo feminino nos cursos está negativamente associada à evasão em todos os dos períodos analisados, com exceção do intervalo pandêmico, em que o efeito se torna estatisticamente nulo. Isso reforça a importância de considerar as questões de gênero na formulação de políticas de permanência estudantil, uma vez que, historicamente, às mulheres são atribuídas maiores responsabilidades domésticas e com filhos (Lopes *et al.*, 2023), além de piores colocações em posições no mercado de trabalho. Contudo, as regressões mostram que, apesar disso, as alunas são mais persistentes (Stinebrickner e Stinebrickner, 2014; Nierotka, Salata e Martins, 2023; Adachi, 2009). Isso sugere que mais iniciativas voltadas ao suporte de alunas podem contribuir indiretamente para a redução das taxas de evasão, como garantir o acesso dos filhos à escola ou em creches, e aumento da licença à maternidade para estudantes puérperas, por exemplo.

Também é importante criar políticas de permanência com foco específico no público masculino. Criar estratégias de engajamento, fortalecimento de vínculos institucionais e ações de acolhimento poderiam contribuir para mitigar os fatores que levam ao abandono por parte desses estudantes.

4.4.7 Efeitos da faixa etária

Para o grupo de discentes com 17 anos ou menos, os resultados das regressões foram significativos a 1% em todos os períodos analisados. Em todo o período, para cada ponto

percentual de aumento na proporção de alunos dessa faixa etária há uma tendência de aumento de 9,38 p.p. na evasão acadêmica em relação ao grupo de 18 a 24 anos. Analisando separadamente, o período que antecedeu a pandemia apresentou tendência de aumento de 7,9 p.p. de aumento na evasão à medida que a proporção de alunos com menos de 18 anos aumenta 1 ponto percentual. No período de pandemia esse resultado quase dobrou, 15,1 p.p., e o período após a pandemia houve uma leve redução, 12 p.p., mas ainda sim valores elevados.

Esses resultados evidenciam uma tendência à evasão muito maior para o grupo de estudantes abaixo de 17 anos do que a faixa de referência, de 18 a 24 anos, em todos os períodos analisados. O que pode configurar falta de maturidade acadêmica em um grupo muito jovem, ingresso precoce na universidade com abandono logo em seguida ou incerteza em relação às perspectivas com o curso escolhido.

Os grupos de discentes de 25 a 29 anos e de 30 a 34 anos são aqueles que apresentaram relação negativa com a evasão em relação ao grupo de 18 a 24 anos. Para todo o período, os dois grupos apresentaram tendências semelhantes de redução da evasão: a cada 1 p.p. de aumento na proporção desses alunos pode-se inferir em diminuição da evasão de 0,9 p.p., aproximadamente, com significância a 1%. No período anterior à pandemia, os números indicam uma redução de 1,05 p.p. na evasão à medida que aumenta 1 p.p. na proporção de alunos de 25 a 29 anos. Nesse mesmo período, o grupo de alunos de 30 a 34 anos não apresentou significância estatística e, portanto, não se pode fazer inferências. Durante o período de pandemia de Covid-19, de 2020 a 2021, curiosamente, os resultados indicam redução ainda maior na evasão acadêmica para esses dois grupos. À medida que aumenta 1 ponto percentual na proporção de alunos matriculados de 25 a 29 anos e de 30 a 34 anos, o painel sugere uma redução na evasão de 1,45 p.p. e 1,47 p.p., respectivamente. No período pós-pandemia foi apresentado um coeficiente menor para o primeiro grupo, 0,6, mas ainda apresentando tendência de redução. O segundo grupo não foi estatisticamente significativo. Infere-se que alunos nessa faixa etária tendem a ser mais decididos, não são tão jovens, já trabalham e buscam o curso por objetivos claros (Gomes e Hirata, 2022). Além disso podem apresentar maior resiliência e motivação com o curso do que os alunos mais jovens devido a um certo grau de maturidade.

As análises para os estudantes de 35 a 49 anos não apresentaram significância estatística e, portanto, não se pode fazer inferências.

Estudos apontam que a evasão acadêmica tende a ser maior em estudantes mais velhos (Lopes *et al.*, 2023; Gomes e Hirata, 2022). Nessa pesquisa, os discentes de graduação de 50 a 59 anos apresentaram relação positiva com as taxas de evasão, quando comparados aos alunos de 18 a 24 anos, com significância estatística a 1%. Em todo o período, a regressão indica que um aumento de 1 p.p. na proporção de alunos na faixa etária pode ocasionar um aumento de 3,53 p.p. na evasão. Quando analisado o período pré-pandemia o coeficiente é ainda maior, 5,24 p.p. Nos anos seguintes a relação de alunos de 50 a 59 anos com a evasão veio reduzindo, mas ainda se mantendo positiva, 4,62 p.p. e 3,18 p.p. para os períodos de pandemia e pós-pandemia, respectivamente. Essa condição pode decorrer do fato de pessoas mais velhas enfrentam barreiras maiores para manter os estudos, como saúde, adaptação à tecnologia e o distanciamento com os conteúdos escolares, por exemplo (Lopes *et al.*, 2023).

Nos anos de 2018 a 2023, o aumento de 1 p.p. na proporção de alunos com mais de 60 anos implicou em um aumento de 2 p.p. na evasão na universidade, em relação ao grupo de referência, com significância estatística de 5%. Durante a pandemia de COVID-19, de 2020 a 2021, esse índice mais que dobrou, 4,35 p.p. de aumento na evasão a cada aumento de 1 ponto percentual na proporção de discentes dessa faixa etária, com significância de 1%. Nos períodos anterior e posterior à pandemia não ocorreu significância estatística. O efeito desses resultados pode ser justificado tendo em vista que esse grupo foi o mais vulnerável às dificuldades de saúde, sendo o mais afetado pela COVID-19, e também por haver maior dificuldade de adaptação ao ensino remoto e ao uso de tecnologias (Andrade, Silva e Scortegagna, 2023).

Esse resultado aponta para a importância de políticas diferenciadas conforme a faixa etária. Para os estudantes mais jovens, com menos de 17 anos, recomenda-se ações de ambientação acadêmica, tutoria e apoio psicopedagógico. Para os estudantes mais velhos pode ser interessante criar um canal de suporte institucional para conciliar estudo, trabalho e responsabilidades familiares, além de flexibilização curricular e de horários.

4.4 DISCUSSÃO PARA AS HIPÓTESES

H1: Confirmada. Os resultados indicam que cursos com maior relação candidatos/vaga apresentam menores taxas de evasão. Esse padrão pode ser explicado pelo perfil de ingresso

mais competitivo e motivado desses estudantes, que demonstram maior planejamento acadêmico e engajamento com os estudos, reduzindo a probabilidade de abandono.

H2: Confirmada. As áreas de Licenciaturas e Ciências Exatas mostraram índices de evasão mais elevados em comparação a cursos da área de Saúde. Fatores como percepção limitada de aplicabilidade imediata no mercado de trabalho e valorização profissional contribuem para esse resultado. Por outro lado, cursos da área de Ciências Naturais, matemática e estatística apresentaram maior evasão, possivelmente em função do grau de dificuldade e nível de abstração.

H3: Confirmada. O turno mostrou influência nos índices de evasão. Cursos noturnos apresentaram taxas mais elevadas, em grande parte devido à sobreposição de compromissos profissionais e familiares.

H4: Confirmada parcialmente. Houve variação regional na evasão acadêmica. A região Norte apresentou menor efeito da evasão em comparação ao Sudeste, enquanto a região Nordeste demonstrou aumento expressivo da evasão durante 2020 e 2021, apesar de ter sido proporcionalmente menos afetada pela COVID-19.

H5: Confirmada. Os dados indicam que homens apresentam maior probabilidade de evasão em comparação às mulheres.

H6: Confirmada. Estudantes com idade superior a 50 anos mostraram maior propensão à evasão durante a pandemia., principalmente no grupo acima de 60 anos.

H7: Confirmada. Alunos de baixa renda foram mais severamente afetados durante a pandemia, enfrentando dificuldades de permanência advindos da instabilidade econômica provocada pela crise sanitária.

H8: Confirmada. Em 2020, observou-se aumento expressivo da evasão em todas as áreas. Já em 2021, houve tendência de queda, sugerindo adaptação gradual das instituições e dos estudantes às novas modalidades de ensino, embora os impactos da pandemia ainda fossem perceptíveis.

5. CONSIDERAÇÕES FINAIS

A análise estatística com dados em painel revelou determinantes robustos e contextuais da evasão no ensino superior brasileiro, abrangendo o período de 2018 a 2023, com análises comparativas entre os períodos pré-pandemia, durante a pandemia de COVID-19 e o cenário pós-pandêmico. A evasão mostrou-se sensível tanto a características individuais dos estudantes quanto a fatores institucionais e regionais.

A proporção de discentes do sexo feminino apresentou relação negativa com a evasão, especialmente nos períodos pré e pós-pandemia, e também no período todo, indicando maior permanência entre as mulheres, mesmo em situações de dificuldades como a pandemia. Esse achado vai ao encontro com estudos anteriores que identificam maior resiliência e desempenho acadêmico feminino no ensino superior (Nierotka, Salata e Martins, 2023; Lopes *et al.*, 2023; Gomes e Hirata, 2022; Hoed, 2016; Adachi, 2009). O que se pode inferir desses estudos é que as mulheres tendem a apresentar menor desistência dos cursos de graduação possivelmente por diferenças culturais, sociais ou motivacionais, além de maior responsabilidade econômica e familiar (Lopes *et al.*, 2023; Faria, 2022; Saccaro, França e Jacinto, 2019).

O estudo mostra ainda, que estudantes entre 25 e 29 anos apresentaram menor propensão à evasão em comparação ao grupo de referência (18 a 24 anos). O aumento da proporção de discentes de 30 a 34 anos também mostrou estar relacionada à maior probabilidade de permanência estudantil, sobretudo durante a pandemia. Isso pode refletir maior estabilidade e motivação. Pode-se inferir que os alunos dessa faixa etária já atingiram certo grau de maturidade, apresentam maior resiliência e buscam fazer um curso com objetivos mais bem definidos (Gomes e Hirata, 2022).

Em contrapartida, os grupos de idade mais avançada (mais de 50 anos) e mais jovens (0 a 17 anos) foram mais suscetíveis à evasão, especialmente durante a pandemia, em consonância com estudos que apontam desafios de adaptação em um novo ambiente de estudos e múltiplas responsabilidades no caso dos mais jovens (Jardim e Almeida, 2016). No caso dos discentes mais velhos, essa condição pode ser decorrente das dificuldades de adaptação ao uso de novas tecnologias e da vulnerabilidade na saúde durante e após a crise sanitária, principalmente nos grupos mais idosos (Andrade, Silva e Scortegagna, 2023).

Estudantes do turno noturno tiveram maiores taxas de evasão, o que pode ser atribuído à sobreposição de atividades profissionais e acadêmicas (Faria, 2022). Os índices se mostraram ainda mais evidentes durante a pandemia, levando a inferir que esses estudantes podem ter sofrido maior impacto por situações como desemprego ou insegurança de saúde. Esses resultados destacam a importância de políticas de apoio específicas para estudantes trabalhadores.

A concorrência do curso teve coeficientes negativos consistentes, sugerindo que cursos mais disputados atraem estudantes com maior motivação ou capital intelectual, reduzindo a evasão Silva Filho *et al.* (2007). Isso se alinha ao argumento de que cursos com maior seletividade tendem a reter estudantes com melhor preparação ou maior comprometimento (Adachi, 2009; Hoed, 2016). Para além dessa análise, é importante destacar também, que a oferta dos cursos também é um fator que afeta a concorrência, já que ela é definida pela relação candidato por vaga. Ou seja, para cursos com baixa procura em que ocorre sobra de vagas é pertinente analisar, do ponto de vista da administração pública, se é interessante manter tantas vagas abertas.

As políticas de reserva de vagas produziram efeitos heterogêneos. A reserva por deficiência teve forte efeito redutor da evasão, indicando eficácia no suporte à permanência dessa população. O estudo mostrou também que o aumento da proporção de alunos em reversa de vagas para escola pública reduziram a evasão, especialmente no período pandêmico, propondo uma ideia de que essas políticas afirmativas foram eficientes.

Já os efeitos das ações afirmativas por etnia foram instáveis, revelando que antes e durante a pandemia o aumento da proporção de alunos matriculados nos cursos de graduação por meio de política de reserva de vagas na modalidade etnia estavam diretamente atrelados a uma maior evasão acadêmica. Contudo, no período após a pandemia esse quadro se reverteu. Esse cenário pode denotar que esse grupo sofreu maiores dificuldades de acesso ao ensino remoto durante a pandemia, mas maior suporte social, psicológico ou financeiro após a crise sanitária.

Ainda nesse sentido, o grupo de alunos matriculados por reserva de vagas por renda familiar se mostrou relacionado à evasão acadêmica em todo o período, mas principalmente durante a pandemia, revelando que o grupo pode ser fortemente influenciado por outros tipos

de necessidades materiais, como dificuldades de transporte, acesso à moradia, à alimentação etc.

As áreas gerais dos cursos também influenciam a evasão. Quando comparadas à área de Saúde e bem-estar, as áreas em geral aparentam sofrer problemas com a evasão, principalmente no período da pandemia de COVID-19. A área de Ciências naturais, matemática e estatística se mostra com a mais problemática, sugerindo problemas de permanência de alunos em cursos de maior abstração e que exigem maior raciocínio lógico-matemático. Esses cursos merecem atenção em relação à reformulação de métodos de ensino e estratégias para fomentar o mercado de trabalho que não é tão atrativo para os graduados em cursos dessa área.

Regionalmente, identificou-se que os cursos presentes nas regiões Centro-oeste, Nordeste e Norte apresentaram maiores efeitos na evasão acadêmica em relação à região Sudeste, principalmente durante a pandemia, o que corrobora desigualdades estruturais no sistema educacional brasileiro.

Por fim, nota-se que a pandemia de COVID-19 teve impacto significativo sobre a evasão, com aumento durante e após o período crítico. Esse efeito pode ser atribuído a fatores como desmotivação, dificuldades no ensino remoto e agravamento das desigualdades sociais e digitais, já que a crise sanitária agravou problemas econômicos, sociais e psicológicos preexistentes (Barros, 2022).

Com base nos resultados empíricos desta pesquisa, é possível propor medidas institucionais e políticas públicas que visam combater a evasão acadêmica de forma mais direcionada.

Em síntese, os resultados indicam que a evasão acadêmica é um fenômeno multifatorial que atravessa dimensões sociais, econômicas, institucionais e regionais. As propostas aqui apresentadas partem do princípio de que políticas de permanência devem ser desenhadas com base em evidências, considerando as especificidades dos diferentes grupos sociais e contextos regionais. Para além de garantir o acesso ao ensino superior, é imprescindível assegurar condições reais de permanência a fim de promover o sucesso acadêmico.

Apesar dos avanços obtidos neste estudo, reconhece-se que a análise esteve limitada ao período de 2018 a 2023. Embora esse período permita identificar tendências antes, durante e após a pandemia, fenômenos como evasão acadêmica podem ter causas estruturais de longo

prazo não captadas nesse intervalo. Em função disso, pesquisas futuras poderiam ampliar o recorte para incluir dados anteriores, observando comportamentos de longo prazo e avaliando o efeito de políticas educacionais implementadas ao longo do tempo. Isso permitiria identificar se a pandemia realmente foi um ponto de inflexão ou apenas acelerou tendências já existentes.

Outra proposta seria avaliar a eficácia de programas de permanência estudantil no combate ao abandono acadêmico, verificando se universidades que adotam esses programas têm menor evasão. Essas análises podem ajudar a desenhar políticas públicas mais eficazes.

REFERÊNCIAS

ABBAD, G.; CARVALHO, R. S.; ZERBINI, T. Evasão em curso via internet: explorando variáveis explicativas. **Rae - Revista de Administracao de Empresas**, São Paulo, 5, jul-dez 2006. Acesso em: 09 ago. 2025.

ADACHI, A. A. C. T. **Evasão e Evadidos nos Cursos de Graduação da Universidade Federal de Minas Gerais**. UFMG. Belo Horizonte. 2009.

ADDISON, L.; WILLIAMS, D. Predicting student retention in higher education institutions (HEIs). **Higher Education, Skills and Work-Based Learning**, 2023. Disponivel em: <https://doi.org/10.1108/HESWBL-12-2022-0257>.

ALMEIDA, D. D. C. L.; SILVA, S. M. C. D. Expansão e Interiorização do Ensino Superior no Brasil – um estudo de caso em Minas Gerais. **SciELO Preprints**, 2022. Acesso em: 01 ago. 2024.

ANDIFES. **Ministério da Educação (MEC). Secretaria de Educação Superior.** Comissão especial de estudos sobre a evasão nas universidades públicas brasileiras. [S.l.]: [s.n.]. 1996.

ANDRADE, M. G. D. Ainda vale a pena fazer medicina? **g1**, 2025. Disponivel em: <https://g1.globo.com/educacao/noticia/2025/04/07/medicina-alunos-particulares-cursos-negocio-no-brasil.ghtml>. Acesso em: 23 jun. 2025.

ANDRADE, S. J. V. D.; SILVA, E. J. L. D.; SCORTEGAGNA, P. A. O Desafio do Ensino Remoto com Idosos durante a Pandemia de Covid-19. **Educação e Realidade**, 48, 2023. Acesso em: 17 jun. 2025.

ANDRIOLA, W. B.; ANDRIOLA, C. G.; MOURA, C. P. Opiniões de docentes e de coordenadores acerca do fenômeno da evasão discente dos cursos de graduação da Universidade Federal do Ceará (UFC). **Ensaio: Avaliação e Políticas Públicas em Educação**, Rio de Janeiro, 14, n. 52, set 2006. 365-382. Acesso em: 28 maio 2025.

ANDRIOLA, W. B.; ANDRIOLA, C. G.; MOURA, C. P. Opiniões de docentes e de coordenadores acerca do fenômeno da evasão discente dos cursos de graduação da Universidade Federal do Ceará (UFC). **Ensaio: Avaliação e Políticas Públicas em Educação**, Rio de Janeiro, 14, n. 52, 2006. 365-382. Acesso em: 28 maio 2025.

ARAUJO, A. C. D. C. et al. Motivos da evasão universitária e serviços de apoio ao estudante. **Revista Brasileira de Orientação Profissional**, Campinas, 24, Jan/Jun 2023. 3-16. Acesso em: 15 abril 2025.

ARISTOVNIK, A. et al. Impacts of the Covid-19 Pandemic on Life of Higher Education Students: Global Survey Dataset from the First Wave. **Data in Brief**, 39, 2021. Acesso em: 26 maio 2025.

BALTAGI, B. H. **Econometric Analysis of Panel Data**. 3ª. ed. Chichester : John Wiley & Sons, 2005.

BARROS, A. D. S. X. Expansão da educação superior no Brasil: limites e possibilidades. **Educação & Sociedade**, Campinas, junho 2015. 361-390.

BARROS, C. S. D. O. **Influência do Plano Nacional de Assistência Estudantil sobre Indicadores de Desempenho Acadêmico, Evasão e Retenção dos Discentes Cotistas Assistidos no Período da Pandemia de Covid-19 na UFPB**. UFPB. João Pessoa. 2022.

BEHR, A. et al. Dropping out of university: a literaturereview. **Review of Education**, 8, n. 2, jun 2020. 614-652. Acesso em: 28 maio 2025.

BENDIG, D. et al. Fear the loss or welcome the gains? How stock options influence CEO risk-taking in corporate cleantech investments. **Journal of Business Venturing**, 40, mar 2025. Acesso em: 2025 jun. 03.

BOHNENBERGER, P. Concorrência acirrada e cursos lotados: os desafios dos graduados nas áreas mais populares do Brasil. **CNN**, 01 mar 2025. Disponivel em: <https://cbn.globo.com/brasil/noticia/2025/03/01/concorrencia-acirrada-e-cursos-lotados-os-desafios-dos-graduados-nas-areas-mais-populares-do-brasil.ghtml>. Acesso em: 23 jun. 2025.

BRASIL. Ministério da Educação. **A democratização e expansão da educação superior no país 2003-2014**, Brasília, 2015. Disponivel em: <https://bit.ly/46H5CoZ>. Acesso em: 02 julho 2024.

BRASIL. **Ministério da Saúde. Gabinete do Ministro. Portaria n. 188, de 03 de fevereiro de 2020**. Brasília, p. 1. 2020a. (ed. 24-A, seção 1).

BRASIL. **Ministério da Educação. Gabinete do Ministro. Portaria n. 329, de 11 de março de 2020.** Brasília, p. 165. 2020b. (ed. 49, seção 1).

BRASIL. **Ministério da Educação. Gabinete do Ministro. Portaria n. 343, de 17 de março de 2020**. Brasília, p. 39. 2020c. (ed. 53, seção 1).

BRASIL. **Atos do Poder Executivo. Medida Provisória nº 934, de 01 de abril de 2020**. Brasília, p. 1. 2020d. (ed. 63-A, seção 1).

BRESSAN, V. G. F. **Seguro depósito e Moral Hazard nas cooperativas de crédito brasileiras**. UFV. Viçosa. 2009.

BRITO, S. D. A. et al. Três anos de pandemia da COVID-19: análise comparativa da incidência, letalidade e mortalidade entre os Estados da região sul do Brasil. **Journal of Human Growth and Development**, 33, 2023. Acesso em: 18 jun. 2025.

CARDOSO, A. L. J.; NAGAI, N. P. 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 (UFMT- Campus Rondonópolis). **Revista de Estudos Sociais - UFTM**, Cuiabá, 20, fev 2019. Acesso em: 15 maio 2025.

CARDOSO, C. B. **Efeitos da política de cotas na Universidade de Brasília: uma análise do rendimento e da evasão**. Universidade de Brasília - UNB. Brasília. 2008
.
CHAGAS, T. M. **Análise da Evasão dos Alunos dos Cursos da Unb: Um Estudo no Âmbito da Graduação**. Unb. Brasília. 2019.

CINE. **Classificação Internacional Normalizada da Educação**. Brasil, p. 166. 2018.

CLEMENTE, A. J. Leonardo Secchi. Análise de Políticas Públicas: Diagnóstico de Problemas, Recomendação de Soluções. **Revista Brasileira de Ciência Política**, São Paulo, Mai/Ago 2018. Acesso em: 15 abril 2025.

COIMBRA, C. L.; SILVA, L. B. E.; COSTA, N. C. D. A evasão na educação superior: definições e trajetórias. **Educação e Pesquisa**, São Paulo, Abril 2021. Acesso em: 15 abril 2025.

CORRÊA, A. C. C.; NORONHA, A. B.; MIURA, I. K. Avaliação da evasão e permanência prolongada em um curso de graduação em administração de uma universidade pública, São Paulo, 2004. Disponivel em: <https://repositorio.usp.br/item/001516817>.

COSTA, N. B. D.; SILVA, G. R. D. A pandemia de covid-19 nos estados das regiões norte e centro-oeste do Brasil. **Revista Hygeia**, Uberlândia, 20, 2024. Acesso em: 16 jun. 2025.

CUNHA, J. M. P. D.; MAIA, S. C. Evasão no Ensino Superior: Um Mapeamento da Produção Científica Mundial com Base na Web of Science. **Revista Lumen et Virtus**, 49, 2025. 7020–7044. Acesso em: 23 jun. 2025.

DIAS, E. C. M.; THEÓPHILO, C. R.; LOPES, M. A. S. **Evasão no Ensino Superior:** Estudo dos Fatores Causadores da Evasão no Curso de Ciências Contábeis da Universidade Estadual de Montes Claros – Unimontes – MG. Congresso USP de Controladoria e Contabilidade. São Paulo: [s.n.]. 2010.

FARIA, A. A. G. D. B. T. **Estudantes do Ensino Superior: variáveis pessoais e contextuais relacionadas ao insucesso acadêmico e evasão escolar**. Tese de Doutoramento em Ciências da Educação - Universidade do Minho. Portugal. 2022.

FAUSTINO-FERBER, A. P.; MARTINS, A. M. Formação inicial de professores: análise da evasão em cursos de Pedagogia no Instituto Federal de São Paulo. **Revista Principia**, 59, n. 4, ago 2022. 1137-1157. Acesso em: 21 fev. 2025.

FEITOSA, J. M. **Análise de evasão no ensino superior: uma proposta de diagnóstico para o campus de Laranjeiras**. Universidade Federal do Sergipe. São Cristóvão. 2016.

FIALHO, M. G. D. **A evasão escolar e a gestão universitária: O caso da Universidade Federal da Paraíba**. Universidade Federal da Paraíba. João Pessoa. 2014.

GERAB, F. **Cálculo da evasão e do índice de titulação em cursos usando uma abordagem probabilística**. XLIV Congresso Brasileiro de Educação em Engenharia. Natal: [s.n.]. 2016.

GOMES, M.; HIRATA, G. Determinantes da Evasão no Ensino Superior: Uma Abordagem de Riscos Competitivos. **Pesquisa e Planejamento Econômico**, 52, dez 2022. Acesso em: 15 maio 2025.

GUERRA, L. C. B.; FERRAZ, R. M. C.; MEDEIROS, J. P. D. Evasão na educação superior de um Instituto Federal do nordeste brasileiro. **Revista Eletrônica de Educação**, São Carlos, 13, mai-ago 2019. 533-553. Acesso em: 25 abr. 2025.

GUJARATI, D. N.; POTER, D. C. **Econometria Básica**. 5ª. ed. Porto Alegre: AMGH Editora Ltda, 2008.

GUSSO, H. L. et al. Ensino Superior em Tempos de Pandemia: Diretrizes à Gestão Universitária. **Educação & Sociedade**, Campinas, 41, 2020. Acesso em: 25 maio 2025.

HOED, R. M. **Análise da evasão em cursos superiores: o caso da evasão em cursos superiores da área de Computação**. Unb. Brasília. 2016.

HSIAO, C. **Analysis of Panel Data**. 3ª. ed. New York: Cambridge University Press, 2014.

IBGE. **Anuário Estatístico do Brasil: Painel Covid-19**. Ministério da Saúde. Brasília - DF. 2024.

INEP. **Censo da Educação Superior 2014 (Glossário)**. Ministério da Educação (MEC). Brasília. 2015.

INEP. Microdados do Censo da Educação Superior. **Ministério da Educação - BRASIL**, 2017. Disponivel em: <https://download.inep.gov.br/microdados/microdados_censo_da_educacao_superior_2017.zip>. Acesso em: 17 set. 2024.

INEP. Microdados do Censo da Educação Superior. **Ministério da Educação - BRASIL**, 2018. Disponivel em: <https://download.inep.gov.br/microdados/microdados_censo_da_educacao_superior_2018.zip>. Acesso em: 17 set. 2024.

INEP. Microdados do Censo da Educação Superior. **Ministério da Educação - BRASIL**, 2019. Disponivel em: <https://download.inep.gov.br/microdados/microdados_censo_da_educacao_superior_2019.zip>. Acesso em: 17 set. 2024.

INEP. Microdados do Censo da Educação Superior. **Ministério da Educação - BRASIL**, 2020. Disponivel em: <https://download.inep.gov.br/microdados/microdados_censo_da_educacao_superior_2020.zip>. Acesso em: 17 set. 2024.

INEP. Censo da Educação Superior. **Ministério da Educação - BRASIL**, 2021. Disponivel em: <https://download.inep.gov.br/educacao_superior/censo_superior/documentos/2021/apresentacao_censo_da_educacao_superior_2021.pdf>. Acesso em: 08 novembro 2023.

INEP. Microdados do Censo da Educação Superior. **Ministério da Educação - BRASIL**, 2021. Disponivel em: <https://download.inep.gov.br/microdados/microdados_censo_da_educacao_superior_2021.zip>. Acesso em: 17 set. 2024.

INEP. Censo da Educação Superior. **Ministério da Educação - BRASIL**, 2022. Disponivel em: <https://download.inep.gov.br/educacao_superior/censo_superior/documentos/2022/apresentacao_censo_da_educacao_superior_2022.pdf>. Acesso em: 08 novembro 2023.

INEP. Microdados do Censo da Educação Superior. **Ministério da Educação - BRASIL**, 2022. Disponivel em: <https://download.inep.gov.br/microdados/microdados_censo_da_educacao_superior_2022.zip>. Acesso em: 17 set. 2024.

INEP. Microdados do Censo da Educação Superior. **Ministério da Educação - BRASIL**, 2023. Disponivel em: <https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/censo-da-educacao-superior#:~:text=Microdados%20do%20Censo%20da%20Educa%C3%A7%C3%A3o%20Superior%202023>. Acesso em: 17 set. 2024.

INEP. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Ministério da Educação - BRASIL**, 2025. Disponivel em: <https://www.gov.br/inep/pt-br>. Acesso em: 05 março 2025.

JARDIM, F. A. A.; ALMEIDA, W. M. D. Expansão recente do ensino superior brasileiro: (novos) elos entre educação, juventudes, trabalho? **Juventude, Educação e Trabalho** , 22, n. 47, 2016. 63-85. Acesso em: 01 ago. 2024.

JUNIOR, P. L. et al. Taxas longitudinais de retenção e evasão: uma metodologia para estudo da trajetória dos estudantes na educação superior. **Ensaio: Avaliação e Políticas Públicas em Educação**, Rio de Janeiro, 27, jan/mar 2019. 157-178. Acesso em: 15 maio 2025.

KANTORSKI, G. Z. et al. **Predição da Evasão em Cursos de Graduação em Instituições Públicas**. V Congresso Brasileiro de Informática na Educação. Santa Maria: [s.n.]. 2016.

KEHM, B. M.; LARSEN, M. R.; SOMMERSEL, H. B. Student dropout from universities in Europe: A review of empirical literature. **Hungarian Educational Research Journal**, 2019. Disponivel em: <https://doi.org/10.1556/063.9.2019.1.18>.

KIM, J. H. Multicollinearity and misleading statistical results. **Korean Journal of**, Daegu, jul 2019. Acesso em: 16 maio 2025.

LIANG, K.-Y.; ZEGER, S. L. Longitudinal data analysis using generalized linear models. **Biometrika**, 73, abril 1986. 13-22. Acesso em: 2025 fev. 02.

LIMA, P. G. Universalização da Educação Superior no Brasil: contrapontos e possibilidades. **Educar em Revista**, Curitiba, março 2014. 243-264.

LIN, W.; WOOLDRIDGE, J. M. Testing and Correcting for Endogeneity in Nonlinear Unobserved Effects Models. In: TSIONAS, M. **Panel Data Econometrics Teory**. [S.l.]: Elsevier, 2019. p. 21-43. Acesso em: 05 fev. 2025.

LOPES, R. et al. Fatores associados à evasão de calouros no ensino superior: um estudo com dados da Universidade Federal do Recôncavo da Bahia. **Revista Brasileira de Educação**, Rio de Janeiro, 28, 2023. Acesso em: 09 maio 2025.

MAIA, S. C. **Gerenciamento de Resultados em Cooperativas de Crédito no Brasil**. UFMG. Belo Horizonte. 2012.

MARTÍNEZ-LÍBANO, J.; YEOMANS-CABRERA, M.-M. Post-pandemic psychosocial variables affecting academic dropout in a sample of Chilean higher-education students. **Frontiers in Education**, 8, dez 2023. Acesso em: 03 fev. 2025.

MELLO, S. P. T. D. 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. [S.l.]: [s.n.]. 2013.

MENDES JUNIOR, A. A. F. Uma análise da progressão dos alunos cotistas sob a primeira ação afirmativa brasileira no ensino superior: o caso da Universidade do Estado do Rio de Janeiro. **Ensaio: Avaliação e Políticas Públicas em Educação**, Rio de Janeiro, 22, n. 82, jan-mar 2014. 31-56. Acesso em: 15 maio 2025.

MENDONÇA, F. D. et al. Região Norte do Brasil e a pandemia de COVID-19: análise socioeconômica e epidemiológica. **Journal Health Npeps**, 5, jan-jun 2020. 20-37. Acesso em: 17 jun. 2025.

MESQUITA, R.; FERNANDES, A. A. T.; FILHO, D. B. F. Uma introdução à regressão com dados de painel. **Política hoje**, 29, 2020. 434-507. Acesso em: 20 maio 2025.

MIOT, H. A. Análise de correlação em estudos clínicos e experimentais. **Jornal Vascular Brasileiro**, 17, out-dez 2018. 275-279. Acesso em: 21 maio 2025.

NIEROTKA, R. L.; SALATA, A.; MARTINS, M. K. Fatores Associados à Evasão no Ensino Superior: Um Estudo Longitudinal. **Cadernos de Pesquisa**, São Paulo, 53, 2023. Acesso em: 15 maio 2025.

O'BRIEN, R. M. A Caution Regarding Rules of Thumb for Variance Inflation Factors. **Quality & Quantity**, 41, out 2007. 673–690. Acesso em: 06 abr. 2025.

ORELLANA, J. D. Y. et al. Explosão da mortalidade no epicentro amazônico da epidemia de COVID-19. **Cadernos de Saúde Pública**, 37, 2020. Acesso em: 17 jun. 2025.

PAPKE, L. E.; WOOLDRIDGE, J. M. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. **Journal of Applied Econometrics**, 11, 1996. 619-632. Acesso em: 2025 fev. 02.

PAPKE, L. E.; WOOLDRIDGE, J. M. Panel data methods for fractional response variables with an application to test pass rates. **Journal of Econometrics**, 145, 2008. 121-133. Acesso em: 2025 fev. 02.
PIETRAS, S. A. **The impact of academic advising on gpa and retention at the community college level**. Indiana: [s.n.], 2010. Disponivel em: <https://eric.ed.gov/?id=ED521991>.

PINHEIRO, C. B.; RIBEIRO, J. L. L. D. S.; FERNANDES, S. A. F. Modelos teóricos da evasão no ensino superior e notas sobre o contexto nacional. **Revista da Avaliação da Educação Superior**, Campinas, Outubro 2023. Acesso em: 15 abril 2025.

PIRES, L. M. **A Universidade Multifuncional nos Cursos Noturnos: Uma Análise da Formação Acadêmica do Estudante-Trabalhador e do Trabalhador-Estudante**. Universidade Federal do Pampa. [S.l.]. 2021.

REUNI. Reestruturação e Expansão das Universidades Federais. **Ministério da Educação - BRASIL**, 2010. Disponivel em: <https://reuni.mec.gov.br/o-que-e-o-reuni>. Acesso em: 11 junho 2024.

RODRIGUES, L. B. et al. Evasão e retenção no Ensino Superior: abordagem baseada em taxas quantitativas. **Revista Contemporânea de Educação**, 16, 21 outubro 2021. 4-21.

RODRIGUES, S. M. Y. O. **Investigando a evasão acadêmica para subsidiar propostas de políticas públicas de acesso e permanência na UNESPAR / FECILCAM**. Universidade Estadual de Maringá. Maringá. 2012.

ROSA, C. D. M.; SANTOS, F. F. T. D.; GONÇALVES, A. M. Os efeitos da pandemia da COVID-19 na permanência na educação superior. O cenário de uma universidade federal brasileira. **Revista Iberoamericana de Educación**, Madri, 86, 2021. 61-76. Acesso em: 03 fev. 2025.

SACCARO, A.; FRANÇA, M. T. A.; JACINTO, P. D. 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, 49, abr-jun 2019. 337-373. Acesso em: 05 maio 2025.

SACHUK, M. I.; TAKAHASHI, L. Y.; AUGUSTO, C. A. Impactos da inovação tecnológica na competitividade e nas relações de trabalho. **Caderno de Administração**, 16, Jul/Dez 2008. 57-66. Acesso em: 15 abril 2025.

SAMASONOK, K.; KAMIENAS, E.; JUšKEVIčIENė, A. Factors determining dropouts from higher education institutions. **Entrepreneurship and sustainability issues**, 2023. Disponivel em: <https://doi.org/10.9770/jesi.2023.10.3(11)>.

SANTOS, F. P.; AMARAL, P.; LUZ, L. Expansão do ensino superior e a distribuição regional das universidades brasileiras. **Revista Brasileira de Estudos Uurbanos e Regionais**, 25, 03 novembro 2023.

SANTOS, G. M. D. O. **Estatísticas sobre os cursos de licenciatura em Física**. X Colóquio Internacional "Educação e Contemporaneidade". Aracajú: [s.n.]. 2016. p. 1-10.

SANTOS, P. V. S. **Adaptação à Universidade dos Estudantes Cotistas e Não Cotistas:** Relação entre Vivência Acadêmica e Intenção de Evasão. Salvador: [s.n.], 2013. Disponivel em: <https://repositorio.ufba.br/handle/ri/14513>.

SECCHI, L. **Análise de Políticas Públicas. Diagnóstico de Problemas, Recomendação de Soluções**. São Paulo: Cengage Learning, 2016. Acesso em: 15 abril 2025.

SHIRASU, M. R.; ARRAES, R. D. A. E. **Determinantes da evasão e repetência escolar**. XLIII Encontro Nacional de Economia. Florianópolis: [s.n.]. 2016. p. 1-19.

SIEGEL, S. Nonparametric Statistics. **American Statistical Association**, Pennsylvania, 11, jun 1957. 13-19. Disponivel em: <http://www.jstor.org/stable/2685679>. Acesso em: 21 maio 2025.

SILVA FILHO, R. L. L. et al. A evasão no ensino superior brasileiro. **Cadernos de Pesquisa**, 2007. Disponivel em: <https://publicacoes.fcc.org.br/cp/article/view/346>.

SILVA FILHO, R. L. L.; LOBO, M. B. D. C. M. **Esclarecimentos Metodológicos Sobre os Cálculos de Evasão**. Instituto Lobo. [S.l.]. 2012.

SILVA, D. B. D. et al. Evasão no ensino superior público do Brasil: estudo de caso da Universidade de São Paulo. **Avaliação: Revista da Avaliação da Educação Superior**, 2022. Disponivel em: <https://doi.org/10.1590/S1414-40772022000200003>.

SILVA, J. J. D.; ROMAN, N. T. **Predicting Dropout in Higher Education:** a Systematic Review. X Congresso Brasileiro de Informática na Educação. [S.l.]: [s.n.]. 2021.

SILVA, R. D. L. M. A universalidade do acesso e permanência no ensino superior. **Revista Concilium**, 22, setembro 2022. 441-451.

SPADY, W. G. Dropouts from higher education: An interdisciplinary review and synthesis. **Interchange**, 1970. Disponivel em: <https://doi.org/10.1007/BF02214313>.

STATACORP. **Stata 19 Base Reference Manual**. College Station, TX: Stata Press. [S.l.]. 2025.

STINEBRICKNER, R.; STINEBRICKNER, T. R. A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout. **The Review of Economic Studies**, 81, jan 2014. 426-472. Acesso em: 07 fev. 2025.

TEIXEIRA, R. P.; MENTGES, M. J.; KAMPFF, A. J. C. **Evasão no ensino superior:** Um estudo sistemático. Congresso Ibero-americano de Docência Universitária. Porto Alegre: [s.n.]. 2018.

THOMAS, L. Student retention in higher education: the role of institutional habitus. **Journal of Education Policy**, 2002. Disponivel em: <https://doi.org/10.1080/02680930210140257>.

TINTO, V. Dropout from Higher Education: A Theoretical Synthesis of Recent Research. **Review of Educational Research**, 1975. Disponivel em: <https://doi.org/10.2307/1170024>.

TINTO, V. Defining dropout: A matter of perspective. **New directions for institutional research**, 1982. 3-15. Disponivel em: <https://doi.org/10.1002/ir.37019823603>.

TINTO, V. Stages of student departure: reflections on the longitudinal character of student leaving. **The Journal of Higher Education**, 1988. Disponivel em: <https://doi.org/10.2307/1981920>.

TINTO, V. Leaving college: Rethinking the causes and cures of student attrition. **University of Chicago Press**, 1993. Disponivel em: <https://doi.org/10.2307/40250027 >.

TRÓPIA, P. V.; SOUZA, D. C. C. D. As portas permanecem semiabertas: estudantes trabalhadores nas universidades federais. **Dossiê "Jovens, Trabalho e Educação"**, 34, 2023. Acesso em: 17 jun. 2025.

UNESCO. Organização das Nações Unidas para a Educação, a Ciência e a Cultura. **Interrupção educacional e resposta à COVID-19**, 08 fev 2022. Disponivel em: <https://www.iiep.unesco.org/en/articles/covid-19-educational-disruption-and-response>. Acesso em: 19 maio 2025.

VALESAN, L. F. et al. Effects of the COVID-19 Pandemic on College Students in Brazil. **Health Sciences**, 96, jul 2024. Acesso em: 09 ago. 2025.

VIRGILLITO, S. B. Regressão a dois estágios (2SLS): o Método das Variáveis Instrumentais em Microeconometria. **LinkedIn**, jul 2020. Disponivel em: <https://pt.linkedin.com/pulse/regress%C3%A3o-dois-est%C3%A1gios-2sls-o-m%C3%A9todo-das-vari%C3%A1veis-em-virgillito>. Acesso em: 10 jun. 2025.

VITELLI, R. F.; FRITSCH, R. Evasão escolar na educação superior: de que indicador estamos falando? **Estudos em Avaliação Educacional**, São Paulo, 27, dez 2016. 908-937. Acesso em: 09 ago. 2025.

WASSERMAN, L. **All of Statistics:** A Concise Course in Statistical Inference. Pittsburgh: Springer Science Business Media, 2004.

WOOLDRIDGE, J. M. **Fractional Response Models with Endogeneous Explanatory Variables and Heterogeneity**. CHI11 Stata Conference 12, Stata Users Group. [S.l.]: [s.n.]. 2011.

WOOLDRIDGE, J. M. **Introductory Econometrics:** A Modern Approach. 5ª. ed. Mason: South-Western, Cengage Learning, 2013.

ZAR, J. H. Spearman Rank Correlation. **John Wiley & Sons**, New Jersey, jul 2005. Acesso em: 21 maio 2025.

ZHANG, L.; RANGWALA, H. Early Identification of At-Risk Students Using Iterative Logistic Regression. **Artificial Intelligence in Education AIED 2018**, London, 20 jun 2018. 613-626.

ZOU, K. H.; TUNCALI, K.; SILVERMAN, S. G. Correlation and Simple Linear Regression. **Radiology**, 227, jun 2003. Acesso em: 21 maio 2025.

APÊNDICES E ANEXOS

APÊNDICE A – ANÁLISES DAS VARIÁVEIS

APÊNDICE A.1 – REGRESSÃO ENTRE AS VARIÁVEIS EVASÃO E CONCORRÊNCIA APÓS TRANSFORMAÇÕES DE LINEARIZAÇÃO (LN, 1/X E LOG-LOG) PARA VERIFICAR A TRANFORMADA MAIS ADEQUADA

```
. reg eva lnconc
```

Source	SS	df	MS	Number of obs	=	16,758
				F(1, 16756)	=	2909.62
Model	45.0332145	1	45.0332145	Prob > F	=	0.0000
Residual	259.338528	16,756	.015477353	R-squared	=	0.1480
				Adj R-squared	=	0.1479
Total	304.371743	16,757	.018163856	Root MSE	=	.12441

eva	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lnconc	-.0473659	.0008781	-53.94	0.000	-.0490871	-.0456447
_cons	.2713114	.0016676	162.69	0.000	.2680427	.2745802

Fonte: Resultados da pesquisa (saída do software Stata®)

```
. reg eva inv_conc
```

Source	SS	df	MS	Number of obs	=	16,750
				F(1, 16748)	=	398.11
Model	7.05554003	1	7.05554003	Prob > F	=	0.0000
Residual	296.819894	16,748	.017722707	R-squared	=	0.0232
				Adj R-squared	=	0.0232
Total	303.875434	16,749	.0181429	Root MSE	=	.13313

eva	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
inv_conc	.018666	.0009355	19.95	0.000	.0168323	.0204997
_cons	.1900774	.0010987	173.00	0.000	.1879238	.1922311

Fonte: Resultados da pesquisa (saída do software Stata®)

```
. reg log_eva log_conc
```

Source	SS	df	MS	Number of obs	=	16,662
				F(1, 16660)	=	2866.25
Model	1580.73471	1	1580.73471	Prob > F	=	0.0000
Residual	9187.96868	16,660	.551498721	R-squared	=	0.1468
				Adj R-squared	=	0.1467
Total	10768.7034	16,661	.64634196	Root MSE	=	.74263

log_eva	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
log_conc	-.2829392	.0052849	-53.54	0.000	-.2932981	-.2725803
_cons	-1.435758	.0100261	-143.20	0.000	-1.455411	-1.416106

Fonte: Resultados da pesquisa (saída do software Stata®)

APÊNDICE A.2 – MATRIZ DE CORRELAÇÃO ENTRE AS VARIÁVEIS UTILIZADAS NA PESQUISA – DADOS DO INEP

```
. spearman eva propfem propmasc prop0a17 prop18a24 prop25a29 prop30a34 prop35a39 prop40a49 prop50a59 prop60mais pr
> pdiu propnot propescopub propeti proppdef proprenda propoutros lnconc dg1 dg2 dg3 dg4 dg5 dg6 dg7 dg8 dg10 dg11
> reg1 dreg2 dreg3 dreg4 dreg5 covid poscovid antescovid, stats (rho p)
(obs=16722)
```

Key
rho
Sig. level

	eva	propfem	propmasc	prop0a17	prop1~24	prop2~29	prop3~34	prop3~39	prop4~49	prop5~59	prop60~s
eva	1.0000										
propfem	-0.1155 0.0000	1.0000									
propmasc	0.1155 0.0000	-1.0000 1.0000	1.0000								
prop0a17	-0.0355 0.0000	-0.0754 0.0000	0.0754 0.0000	1.0000							
prop18a24	-0.2136 0.0000	-0.0339 0.0000	0.0339 0.0000	0.1415 0.0000	1.0000						
prop25a29	-0.0135 0.0814	-0.1115 0.0000	0.1115 0.0000	-0.1649 0.0000	-0.6240 0.0000	1.0000					
prop30a34	0.1920 0.0000	0.0532 0.0000	-0.0532 0.0000	-0.0924 0.0000	-0.8353 0.0000	0.3266 0.0000	1.0000				
prop35a39	0.1887 0.0000	0.1219 0.0000	-0.1219 0.0000	-0.0791 0.0000	-0.7651 0.0000	0.2063 0.0000	0.7061 0.0000	1.0000			
prop40a49	0.2434 0.0000	0.1650 0.0000	-0.1650 0.0000	-0.0912 0.0000	-0.7299 0.0000	0.1279 0.0000	0.6661 0.0000	0.7207 0.0000	1.0000		
prop50a59	0.1970 0.0000	0.1655 0.0000	-0.1655 0.0000	-0.0860 0.0000	-0.5815 0.0000	0.0709 0.0000	0.5122 0.0000	0.5538 0.0000	0.6288 0.0000	1.0000	
prop60mais	0.1233 0.0000	0.0804 0.0000	-0.0804 0.0000	-0.0707 0.0000	-0.3684 0.0000	0.0332 0.0000	0.2923 0.0000	0.3403 0.0000	0.4057 0.0000	0.4323 0.0000	1.0000
propdiu	-0.1736 0.0000	0.0079 0.3074	-0.0079 0.3074	0.0829 0.0000	0.4452 0.0000	-0.1590 0.0000	-0.4535 0.0000	-0.4463 0.0000	-0.4391 0.0000	-0.3326 0.0000	-0.1495 0.0000
propnot	0.1736 0.0000	-0.0079 0.3074	0.0079 0.3074	-0.0829 0.0000	-0.4452 0.0000	0.1590 0.0000	0.4535 0.0000	0.4463 0.0000	0.4391 0.0000	0.3326 0.0000	0.1495 0.0000
propescopub	-0.0722 0.0000	0.0940 0.0000	-0.0940 0.0000	0.0399 0.0000	0.1069 0.0000	-0.0793 0.0000	-0.0594 0.0000	-0.0382 0.0000	-0.0526 0.0000	-0.0620 0.0000	-0.0582 0.0000
propeti	-0.0921 0.0000	0.1111 0.0000	-0.1111 0.0000	0.1337 0.0000	0.0308 0.0001	-0.0676 0.0000	0.0173 0.0251	0.0428 0.0000	0.0215 0.0055	0.0109 0.1580	-0.0132 0.0890
proppdef	-0.1876 0.0000	0.0642 0.0000	-0.0642 0.0000	0.1031 0.0000	-0.0028 0.7163	-0.0234 0.0024	0.0331 0.0000	0.0551 0.0000	0.0414 0.0000	0.0095 0.2181	0.0085 0.9526
proprenda	-0.0440 0.0000	0.1296 0.0000	-0.1296 0.0000	0.0450 0.0000	0.0401 0.0000	-0.0586 0.0000	0.0079 0.3083	0.0290 0.0002	0.0094 0.2246	-0.0194 0.0122	-0.0397 0.0000
propoutros	0.0174 0.0243	0.0291 0.0002	-0.0291 0.0002	-0.0075 0.3300	-0.0106 0.1720	0.0285 0.0002	-0.0002 0.9765	0.0059 0.4453	0.0089 0.2583	-0.0007 0.9272	0.0152 0.0495
lnconc	-0.3936 0.0000	0.1641 0.0000	-0.1641 0.0000	0.1673 0.0000	0.1848 0.0000	-0.1016 0.0000	-0.1193 0.0000	-0.1058 0.0000	-0.1493 0.0000	-0.1024 0.0000	-0.0984 0.0000
dg1	-0.0409 0.0000	-0.0172 0.0259	0.0172 0.0259	0.0207 0.0073	0.0504 0.0000	0.0607 0.0000	-0.0561 0.0000	-0.0825 0.0000	-0.1033 0.0000	-0.0551 0.0000	-0.0082 0.2865

	0.0000	0.0000	0.0000								
dg8	-0.0895	-0.0609	-0.1799	-0.1385	1.0000						
	0.0000	0.0000	0.0000	0.0000							
dg10	-0.1072	-0.0729	-0.2155	-0.1660	-0.1073	1.0000					
	0.0000	0.0000	0.0000	0.0000	0.0000						
dg11	-0.0337	-0.0230	-0.0678	-0.0522	-0.0337	-0.0404	1.0000				
	0.0000	0.0030	0.0000	0.0000	0.0000	0.0000					
dreg1	0.0024	0.0069	0.0447	-0.0516	0.0062	0.0158	-0.0327	1.0000			
	0.7539	0.3718	0.0000	0.0000	0.4211	0.0414	0.0000				
dreg2	0.0224	-0.0097	0.0545	-0.0921	-0.0037	0.0167	0.0051	-0.2429	1.0000		
	0.0038	0.2103	0.0000	0.0000	0.6310	0.0305	0.5133	0.0000			
dreg3	-0.0234	-0.0112	0.0739	-0.0573	0.0252	-0.0578	-0.0087	-0.1094	-0.1875	1.0000	
	0.0024	0.1487	0.0000	0.0000	0.0011	0.0000	0.2595	0.0000	0.0000		
dreg4	0.0190	-0.0031	-0.0855	0.0277	-0.0003	0.0580	0.0031	-0.2475	-0.4241	-0.1910	1.0000
	0.0141	0.6873	0.0000	0.0003	0.9699	0.0000	0.6906	0.0000	0.0000	0.0000	
dreg5	-0.0334	0.0163	-0.0501	0.1533	-0.0174	-0.0597	0.0234	-0.1896	-0.3249	-0.1464	-0.3311
	0.0000	0.0348	0.0000	0.0000	0.0247	0.0000	0.0025	0.0000	0.0000	0.0000	0.0000
covid	0.0003	0.0002	0.0000	0.0000	0.0000	-0.0003	0.0000	0.0000	0.0000	0.0000	0.0000
	0.9682	0.9777	1.0000	1.0000	1.0000	0.9725	1.0000	1.0000	1.0000	1.0000	1.0000
poscovid	0.0003	0.0009	0.0000	0.0000	0.0000	-0.0003	0.0000	0.0000	0.0000	0.0000	0.0000
	0.9682	0.9109	1.0000	1.0000	1.0000	0.9725	1.0000	1.0000	1.0000	1.0000	1.0000
antescovid	-0.0006	-0.0011	0.0000	0.0000	0.0000	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000
	0.9365	0.8887	1.0000	1.0000	1.0000	0.9451	1.0000	1.0000	1.0000	1.0000	1.0000

	dreg5	covid	poscovid	antesc~d
dreg5	1.0000			
covid	0.0000	1.0000		
	1.0000			
poscovid	0.0000	-0.5000	1.0000	
	1.0000	0.0000		
antescovid	0.0000	-0.5000	-0.5000	1.0000
	1.0000	0.0000	0.0000	

propoutros	-0.0127	0.0127	-0.1421	-0.0571	-0.0337	-0.1068	1.0000				
	0.1005	0.1005	0.0000	0.0000	0.0000	0.0000					
lnconc	0.0775	-0.0775	0.1010	0.2679	0.2639	0.1263	-0.0409	1.0000			
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000				
dg1	0.1927	-0.1927	0.0138	0.0132	-0.0128	0.0267	0.0186	0.0063	1.0000		
	0.0000	0.0000	0.0746	0.0889	0.0971	0.0005	0.0163	0.4139			
dg2	-0.0063	0.0063	-0.0709	-0.0255	-0.0643	-0.0682	-0.0067	-0.0255	-0.0681	1.0000	
	0.4122	0.4122	0.0000	0.0010	0.0000	0.0000	0.3873	0.0010	0.0000		
dg3	0.1462	-0.1462	-0.0627	-0.0285	-0.1249	-0.0429	0.0129	-0.1282	-0.0921	-0.0815	1.0000
	0.0000	0.0000	0.0000	0.0002	0.0000	0.0000	0.0955	0.0000	0.0000	0.0000	
dg4	-0.0409	0.0409	0.0139	0.0430	0.0684	0.0214	-0.0323	0.0712	-0.0829	-0.0734	-0.0993
	0.0000	0.0000	0.0714	0.0000	0.0000	0.0057	0.0000	0.0000	0.0000	0.0000	0.0000
dg5	0.0059	-0.0059	-0.0083	-0.0475	0.0332	-0.0199	-0.0045	0.0252	-0.0564	-0.0499	-0.0676
	0.4447	0.4447	0.9650	0.0000	0.0000	0.0102	0.5576	0.0011	0.0000	0.0000	0.0000
dg6	-0.3463	0.3463	0.0083	0.0493	-0.0552	0.0363	0.0176	-0.1849	-0.1667	-0.1476	-0.1997
	0.0000	0.0000	0.9674	0.0000	0.0000	0.0000	0.0227	0.0000	0.0000	0.0000	0.0000
dg7	0.2221	-0.2221	-0.0370	-0.1596	-0.0615	-0.1061	-0.0124	-0.1291	-0.1284	-0.1136	-0.1538
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1095	0.0000	0.0000	0.0000	0.0000
dg8	-0.2322	0.2322	0.0417	0.0471	0.1225	0.0471	-0.0064	0.1545	-0.0830	-0.0734	-0.0994
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.4110	0.0000	0.0000	0.0000	0.0000
dg10	0.1683	-0.1683	0.0994	0.1094	0.1432	0.0995	0.0122	0.3343	-0.0994	-0.0880	-0.1190
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1154	0.0000	0.0000	0.0000	0.0000
dg11	-0.0292	0.0292	-0.0107	-0.0007	-0.0092	-0.0032	-0.0227	0.0038	-0.0313	-0.0277	-0.0375
	0.0002	0.0002	0.1662	0.9263	0.2352	0.6747	0.0034	0.6209	0.0001	0.0003	0.0000
dreg1	-0.0399	0.0399	0.0574	0.0216	-0.0200	0.0213	-0.1347	-0.0618	0.0012	0.0110	-0.0268
	0.0000	0.0000	0.0000	0.0053	0.0096	0.0060	0.0000	0.0000	0.8721	0.1544	0.0005
dreg2	0.0179	-0.0179	0.1618	0.4690	0.2236	0.2346	0.0148	0.1512	-0.0137	0.0016	0.0171
	0.0206	0.0206	0.0000	0.0000	0.0000	0.0000	0.0549	0.0000	0.0764	0.8357	0.0273
dreg3	0.0066	-0.0066	-0.0774	0.1055	0.0824	0.0253	-0.0275	0.0056	0.0858	-0.0482	-0.0025
	0.3933	0.3933	0.0000	0.0000	0.0000	0.0011	0.0004	0.4672	0.0000	0.0000	0.7481
dreg4	-0.0161	0.0161	-0.1448	-0.0610	-0.1498	-0.1112	0.0809	0.1728	-0.0396	0.0120	0.0378
	0.0369	0.0369	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1206	0.0000
dreg5	0.0265	-0.0265	-0.0135	-0.5502	-0.1209	-0.1732	0.0196	-0.3219	0.0025	0.0076	-0.0389
	0.0006	0.0006	0.0804	0.0000	0.0000	0.0000	0.0112	0.0000	0.7438	0.3286	0.0000
covid	-0.0001	0.0001	0.0797	0.0537	0.0199	0.0793	-0.0346	0.0148	0.0000	-0.0004	0.0001
	0.9872	0.9872	0.0000	0.0000	0.0100	0.0000	0.0000	0.0558	1.0000	0.9623	0.9854
poscovid	-0.0027	0.0027	0.0942	0.0419	0.0988	0.0599	0.0775	-0.2402	0.0000	-0.0004	-0.0003
	0.7231	0.7231	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.9623	0.9708
antescovid	0.0029	-0.0029	-0.1738	-0.0955	-0.1188	-0.1392	-0.0429	0.2254	0.0000	0.0007	0.0001
	0.7110	0.7110	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.9247	0.9854

	dg4	dg5	dg6	dg7	dg8	dg10	dg11	dreg1	dreg2	dreg3	dreg4
dg4	1.0000										
dg5	-0.0608	1.0000									
	0.0000										
dg6	-0.1798	-0.1224	1.0000								
	0.0000	0.0000									
dg7	-0.1384	-0.0942	-0.2784	1.0000							

dg2	0.0312	0.0416	-0.0416	-0.0592	-0.1300	0.0578	0.0901	0.0617	0.0653	0.1014	0.1161
	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
dg3	0.1188	-0.0804	0.0804	0.0070	0.1514	-0.1044	-0.1250	-0.1230	-0.1298	-0.0967	-0.0714
	0.0000	0.0000	0.0000	0.3652	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
dg4	-0.0187	0.0592	-0.0592	0.0014	-0.0161	-0.0577	0.0205	0.0520	0.0733	0.1188	0.1311
	0.0155	0.0000	0.0000	0.8521	0.0368	0.0000	0.0080	0.0000	0.0000	0.0000	0.0000
dg5	0.0131	-0.3155	0.3155	0.0431	0.1209	-0.0480	-0.0843	-0.1111	-0.1431	-0.1257	-0.1068
	0.0903	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
dg6	0.1858	0.1430	-0.1430	-0.0577	-0.3438	0.0361	0.3594	0.3644	0.3974	0.2946	0.1416
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
dg7	-0.0322	-0.3063	0.3063	0.0533	0.1855	0.1071	-0.2692	-0.2920	-0.2982	-0.2737	-0.1557
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
dg8	-0.0786	-0.0156	0.0156	0.0101	-0.0386	-0.0033	0.0869	0.0931	0.0729	0.0233	-0.0298
	0.0000	0.0436	0.0436	0.1932	0.0000	0.6664	0.0000	0.0000	0.0000	0.0026	0.0001
dg10	-0.2546	0.3584	-0.3584	0.0042	0.1613	-0.0919	-0.1332	-0.0930	-0.0969	-0.1028	-0.0826
	0.0000	0.0000	0.0000	0.5858	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
dg11	0.0274	0.0743	-0.0743	-0.0134	-0.0621	0.0065	0.0377	0.0499	0.0641	0.1086	0.1038
	0.0004	0.0000	0.0000	0.0841	0.0000	0.3997	0.0000	0.0000	0.0000	0.0000	0.0000
dreg1	0.0369	0.0325	-0.0325	-0.0533	0.0681	-0.1363	-0.0540	-0.0158	0.0307	0.0068	-0.0155
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0405	0.0001	0.3795	0.0450
dreg2	-0.0267	-0.0502	0.0502	0.2844	-0.0424	-0.0510	0.1077	0.1192	0.0749	0.0411	0.0218
	0.0006	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0049
dreg3	0.0703	0.0395	-0.0395	0.0455	-0.1614	0.0373	0.1656	0.1895	0.1705	0.0528	-0.0474
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
dreg4	-0.1625	0.0253	-0.0253	-0.2067	0.1570	0.0699	-0.2036	-0.2488	-0.2448	-0.1095	-0.0282
	0.0000	0.0011	0.0011	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0003
dreg5	0.1387	-0.0250	0.0250	-0.0729	-0.0796	0.0649	0.0443	0.0358	0.0558	0.0377	0.0519
	0.0000	0.0012	0.0012	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
covid	-0.0510	0.0018	-0.0018	-0.0424	0.0226	-0.0387	0.0054	0.0090	-0.0048	-0.0003	0.0040
	0.0000	0.8173	0.8173	0.0000	0.0035	0.0000	0.4819	0.2449	0.5350	0.9701	0.6041
poscovid	0.0472	0.0271	-0.0271	-0.2118	-0.0724	0.1434	-0.0133	-0.0006	0.0368	0.0235	0.0535
	0.0000	0.0005	0.0005	0.0000	0.0000	0.0000	0.0855	0.9391	0.0000	0.0024	0.0000
antescovid	0.0037	-0.0288	0.0288	0.2542	0.0498	-0.1048	0.0079	-0.0084	-0.0320	-0.0232	-0.0575
	0.6287	0.0002	0.0002	0.0000	0.0000	0.0000	0.3095	0.2773	0.0000	0.0027	0.0000

	propdiu	propnot	propes~b	propeti	proppdef	propre~a	propou~s	lnconc	dg1	dg2	dg3
propdiu	1.0000										
propnot	-1.0000	1.0000									
	1.0000										
propescopub	-0.0232	0.0232	1.0000								
	0.0027	0.0027									
propeti	-0.0080	0.0080	0.5068	1.0000							
	0.2989	0.2989	0.0000								
proppdef	0.0134	-0.0134	0.2544	0.3344	1.0000						
	0.0839	0.0839	0.0000	0.0000							
proprenda	-0.0231	0.0231	0.7524	0.5967	0.2872	1.0000					
	0.0028	0.0028	0.0000	0.0000	0.0000						

Fonte: Resultados da pesquisa (saída do software Stata®)

APÊNDICE B – RESULTADOS DOS PROCEDIMENTOS OPERACIONAIS DOS DADOS DO CENSO DA EDUCAÇÃO SUPERIOR INEP

APÊNDICE B.1 – APLICAÇÃO DE TESTES PARA DEFINIÇÃO DO MODELO DE REGRESSÃO COM RESPOSTA FRACIONÁRIA

1. Teste de modelos link(probit), link(logit), corr(unstructured), corr(ar1) e corr(ar2)

```
. xtgee eva propfem prop0a17 prop25a29 prop30a34 prop35a39 prop40a49 prop50a59 prop60mais propnot lnconc propescopu
> b propeti proppdef proprenda propoutros lnconc_hat propescopub_hat propeti_hat proppdef_hat proprenda_hat propout
> ros_hat dg1 dg2 dg3 dg4 dg5 dg6 dg7 dg8 dg11 dreg1 dreg2 dreg3 dreg5 covid poscovid, family(binomial) link(probit
> ) corr(unstructured) vce(robust)

Iteration 1: tolerance = 1.0909853
Iteration 2: tolerance = .20456244
Iteration 3: tolerance = .03405963
Iteration 4: tolerance = .0073664
Iteration 5: tolerance = .00157254
Iteration 6: tolerance = .00032777
Iteration 7: tolerance = .00006603
Iteration 8: tolerance = .00001344
Iteration 9: tolerance = 2.600e-06
Iteration 10: tolerance = 4.956e-07

GEE population-averaged model                   Number of obs     =     16,722
Group and time vars:          codcurso ano      Number of groups  =      2,787
Link:                              probit       Obs per group:
Family:                          binomial                    min =          6
Correlation:                 unstructured                    avg =        6.0
                                                             max =          6
                                                Wald chi2(36)     =    2160.92
Scale parameter:                         1      Prob > chi2       =     0.0000

                                  (Std. Err. adjusted for clustering on codcurso)
```

eva	Coef.	Semirobust Std. Err.	z	P>|z|	[95% Conf.	Interval]
propfem	-.009237	.0367656	-0.25	0.802	-.0812963	.0628223
prop0a17	4.123957	.6750364	6.11	0.000	2.80091	5.447004
prop25a29	-.4665976	.0771564	-6.05	0.000	-.6178214	-.3153737
prop30a34	-.580522	.1537337	-3.78	0.000	-.8818344	-.2792095
prop35a39	-.315138	.2133459	-1.48	0.140	-.7332883	.1030122
prop40a49	.1366945	.230639	0.59	0.553	-.3153495	.5887386
prop50a59	1.672402	.2987592	5.60	0.000	1.086845	2.25796
prop60mais	.9108457	.4246289	2.15	0.032	.0785883	1.743103
propnot	.1804502	.0180137	10.02	0.000	.1451439	.2157565
lnconc	-.0113742	.00601	-1.89	0.058	-.0231537	.0004052
propescopub	-.1361914	.0415572	-3.28	0.001	-.2176421	-.0547408
propeti	-.0273083	.0579989	-0.47	0.638	-.140984	.0863673
proppdef	-1.642425	.5440967	-3.02	0.003	-2.708835	-.5760156
proprenda	.3121703	.0753	4.15	0.000	.164585	.4597556
propoutros	.2232961	.0798546	2.80	0.005	.0667838	.3798083
lnconc_hat	-.07239	.0123805	-5.85	0.000	-.0966554	-.0481246
propescopub_hat	-.8072032	.1901899	-4.24	0.000	-1.179968	-.4344378
propeti_hat	-.0065804	.2460762	-0.03	0.979	-.4888809	.4757201
proppdef_hat	-7.823844	1.978694	-3.95	0.000	-11.70201	-3.945675
proprenda_hat	.1914648	.3036444	0.63	0.528	-.4036673	.7865969
propoutros_hat	-1.504947	.3277062	-4.59	0.000	-2.147239	-.8626546
dg1	.1028104	.0299592	3.43	0.001	.0440913	.1615294
dg2	.1408216	.0314743	4.47	0.000	.0791331	.20251
dg3	.2771258	.0301543	9.19	0.000	.2180244	.3362272
dg4	.1248886	.0270594	4.62	0.000	.0718531	.1779241
dg5	.1014849	.0344412	2.95	0.003	.0339814	.1689883
dg6	.1878801	.0250428	7.50	0.000	.1387972	.236963
dg7	.1059496	.0283893	3.73	0.000	.0503075	.1615916
dg8	.0639133	.0270076	2.37	0.018	.0109795	.1168472
dg11	.1331423	.0449395	2.96	0.003	.0450625	.221222
dreg1	.1641376	.0248762	6.60	0.000	.1153811	.2128941
dreg2	.1619226	.0230644	7.02	0.000	.1167173	.2071279
dreg3	.2555707	.0294533	8.68	0.000	.1978433	.3132982
dreg5	.2574114	.0363058	7.09	0.000	.1862532	.3285695
covid	.0618612	.009184	6.74	0.000	.0438608	.0798615
poscovid	.0818959	.0120912	6.77	0.000	.0581976	.1055942
_cons	-.7022284	.0579413	-12.12	0.000	-.8157913	-.5886655

Fonte: Resultados da pesquisa (saída do software Stata®)

```
. xtgee eva propfem prop0a17 prop25a29 prop30a34 prop35a39 prop40a49 prop50a59 prop60mais propnot lnconc propescopu
> b propeti proppdef proprenda propoutros lnconc_hat propescopub_hat propeti_hat proppdef_hat proprenda_hat propout
> ros_hat dg1 dg2 dg3 dg4 dg5 dg6 dg7 dg8 dg11 dreg1 dreg2 dreg3 dreg5 covid poscovid, family(binomial) link(probit
> ) corr(ar1) vce(robust)

Iteration 1: tolerance = .7893121
Iteration 2: tolerance = .12868695
Iteration 3: tolerance = .02429297
Iteration 4: tolerance = .00581438
Iteration 5: tolerance = .00142979
Iteration 6: tolerance = .00035445
Iteration 7: tolerance = .00008803
Iteration 8: tolerance = .00002187
Iteration 9: tolerance = 5.436e-06
Iteration 10: tolerance = 1.351e-06
Iteration 11: tolerance = 3.357e-07

GEE population-averaged model                   Number of obs     =     16,722
Group and time vars:          codcurso ano      Number of groups  =      2,787
Link:                               probit      Obs per group:
Family:                           binomial                    min =          6
Correlation:                         AR(1)                    avg =        6.0
                                                              max =          6
                                                Wald chi2(36)     =    2551.01
Scale parameter:                         1      Prob > chi2       =     0.0000

                                  (Std. Err. adjusted for clustering on codcurso)
```

eva	Coef.	Semirobust Std. Err.	z	P>|z|	[95% Conf.	Interval]
propfem	-.0742934	.0370932	-2.00	0.045	-.1469947	-.0015922
prop0a17	5.352585	.7398499	7.23	0.000	3.902506	6.802665
prop25a29	-.5206278	.0849514	-6.13	0.000	-.6871295	-.3541261
prop30a34	-.5302671	.1663154	-3.19	0.001	-.8562394	-.2042948
prop35a39	-.2361436	.2437267	-0.97	0.333	-.7138393	.241552
prop40a49	.4022669	.2648891	1.52	0.129	-.1169061	.92144
prop50a59	2.108859	.3560769	5.92	0.000	1.410961	2.806757
prop60mais	1.268142	.5007996	2.53	0.011	.286593	2.249691
propnot	.1616716	.0189078	8.55	0.000	.124613	.1987302
lnconc	-.0208846	.0058054	-3.60	0.000	-.0322629	-.0095063
propescopub	-.1380286	.0460157	-3.00	0.003	-.2282178	-.0478395
propeti	-.1557462	.0640624	-2.43	0.015	-.2813062	-.0301861
proppdef	-1.055581	.6032259	-1.75	0.080	-2.237882	.1267205
proprenda	.3070819	.0826512	3.72	0.000	.1450885	.4690753
propoutros	.1230451	.0905732	1.36	0.174	-.0544751	.3005653
lnconc_hat	-.128497	.0132537	-9.70	0.000	-.1544739	-.1025202
propescopub_hat	-.7116473	.2043433	-3.48	0.000	-1.112153	-.3111419
propeti_hat	.1480857	.2773024	0.53	0.593	-.395417	.6915884
proppdef_hat	-9.703529	3.098076	-3.13	0.002	-15.77565	-3.631412
proprenda_hat	.3104616	.3428043	0.91	0.365	-.3614224	.9823456
propoutros_hat	-1.708399	.435748	-3.92	0.000	-2.562449	-.8543484
dg1	.0145122	.02997	0.48	0.628	-.044228	.0732523
dg2	.0486365	.0311366	1.56	0.118	-.0123901	.1096631
dg3	.188463	.030084	6.26	0.000	.1294994	.2474266
dg4	.0638893	.0252482	2.53	0.011	.0144037	.1133748
dg5	.0537805	.0352858	1.52	0.127	-.0153784	.1229394
dg6	.0908853	.0243278	3.74	0.000	.0432036	.1385669
dg7	.0210736	.0281645	0.75	0.454	-.0341278	.0762751
dg8	.0177698	.025459	0.70	0.485	-.032129	.0676685
dg11	.0616098	.0408325	1.51	0.131	-.0184206	.1416401
dreg1	.1151925	.02543	4.53	0.000	.0653506	.1650343
dreg2	.1659609	.024352	6.82	0.000	.1182319	.2136898
dreg3	.2447813	.0337319	7.26	0.000	.1786681	.3108946
dreg5	.1502681	.0403599	3.72	0.000	.0711642	.2293721
covid	.0197337	.0102487	1.93	0.054	-.0003533	.0398207
poscovid	.069497	.0135131	5.14	0.000	.0430118	.0959821
_cons	-.5025565	.0629828	-7.98	0.000	-.6260005	-.3791125

Fonte: Resultados da pesquisa (saída do software Stata®)

```
. xtgee eva propfem prop0a17 prop25a29 prop30a34 prop35a39 prop40a49 prop50a59 prop60mais propnot lnconc propescopu
> b propeti proppdef proprenda propoutros lnconc_hat propescopub_hat propeti_hat proppdef_hat proprenda_hat propout
> ros_hat dg1 dg2 dg3 dg4 dg5 dg6 dg7 dg8 dg11 dreg1 dreg2 dreg3 dreg5 covid poscovid, family(binomial) link(probit
> ) corr(ar2) vce(robust)

Iteration 1: tolerance = .79641672
Iteration 2: tolerance = .07729489
Iteration 3: tolerance = .01241585
Iteration 4: tolerance = .00250201
Iteration 5: tolerance = .00050653
Iteration 6: tolerance = .00010108
Iteration 7: tolerance = .00002034
Iteration 8: tolerance = 4.059e-06
Iteration 9: tolerance = 8.166e-07

GEE population-averaged model                   Number of obs     =     16,722
Group and time vars:          codcurso ano      Number of groups  =      2,787
Link:                              probit       Obs per group:
Family:                          binomial                     min =          6
Correlation:                        AR(2)                     avg =        6.0
                                                              max =          6
                                                Wald chi2(36)     =    2341.09
Scale parameter:                         1      Prob > chi2       =     0.0000

                                (Std. Err. adjusted for clustering on codcurso)
```

eva	Coef.	Semirobust Std. Err.	z	P>|z|	[95% Conf.	Interval]
propfem	-.0550042	.0365873	-1.50	0.133	-.1267139	.0167055
prop0a17	4.972502	.6623728	7.51	0.000	3.674275	6.270728
prop25a29	-.4547744	.0829885	-5.48	0.000	-.617429	-.2921199
prop30a34	-.5681877	.1657152	-3.43	0.001	-.8929837	-.2433918
prop35a39	-.3958586	.2394286	-1.65	0.098	-.86513	.0734128
prop40a49	.2647634	.2625873	1.01	0.313	-.2498983	.7794251
prop50a59	1.885393	.3301419	5.71	0.000	1.238326	2.532459
prop60mais	1.124488	.4709129	2.39	0.017	.2015161	2.047461
propnot	.1734581	.0181161	9.57	0.000	.1379512	.2089649
lnconc	-.0101897	.0057084	-1.79	0.074	-.021378	.0009986
propescopub	-.1890477	.0419717	-4.50	0.000	-.2713108	-.1067847
propeti	-.0947911	.0605322	-1.57	0.117	-.2134321	.0238498
proppdef	-.9362503	.5566612	-1.68	0.093	-2.027286	.1547856
proprenda	.3749333	.0763406	4.91	0.000	.2253085	.5245581
propoutros	.2469924	.086768	2.85	0.004	.0769302	.4170546
lnconc_hat	-.1284933	.012593	-10.20	0.000	-.153175	-.1038115
propescopub_hat	-.691399	.1942593	-3.56	0.000	-1.07214	-.3106577
propeti_hat	.17427	.2622391	0.66	0.506	-.3397091	.6882491
proppdef_hat	-8.572461	2.501231	-3.43	0.001	-13.47478	-3.670137
proprenda_hat	.1366948	.3060972	0.45	0.655	-.4632446	.7366342
propoutros_hat	-1.702435	.3645524	-4.67	0.000	-2.416944	-.987925
dg1	.0347942	.0293708	1.18	0.236	-.0227715	.09236
dg2	.0758266	.0310784	2.44	0.015	.0149141	.1367392
dg3	.2177619	.0297524	7.32	0.000	.1594482	.2760756
dg4	.0836103	.0256158	3.26	0.001	.0334043	.1338163
dg5	.0740429	.0346051	2.14	0.032	.006218	.1418677
dg6	.1209348	.0240562	5.03	0.000	.0737855	.1680841
dg7	.0421089	.0278405	1.51	0.130	-.0124574	.0966751
dg8	.0295583	.0256885	1.15	0.250	-.0207903	.0799069
dg11	.0905928	.0406139	2.23	0.026	.0109909	.1701946
dreg1	.1241603	.0247838	5.01	0.000	.075585	.1727357
dreg2	.1582566	.0235212	6.73	0.000	.1121559	.2043574
dreg3	.2240858	.031315	7.16	0.000	.1627096	.2854621
dreg5	.1940518	.038681	5.02	0.000	.1182385	.2698651
covid	.0145045	.0097478	1.49	0.137	-.0046009	.0336099
poscovid	.047472	.0126325	3.76	0.000	.0227127	.0722313
_cons	-.5422551	.0591702	-9.16	0.000	-.6582266	-.4262836

Fonte: Resultados da pesquisa (saída do software Stata®)

```
. xtgee eva propfem prop0a17 prop25a29 prop30a34 prop35a39 prop40a49 prop50a59 prop60mais propnot lnconc propescopu
> b propeti proppdef proprenda propoutros lnconc_hat propescopub_hat propeti_hat proppdef_hat proprenda_hat propout
> ros_hat dg1 dg2 dg3 dg4 dg5 dg6 dg7 dg8 dg11 dreg1 dreg2 dreg3 dreg5 covid poscovid, family(binomial) link(logit)
>  corr(unstructured) vce(robust)

Iteration 1: tolerance = 1.8870038
Iteration 2: tolerance = .27546981
Iteration 3: tolerance = .03975957
Iteration 4: tolerance = .00765759
Iteration 5: tolerance = .00109963
Iteration 6: tolerance = .00029358
Iteration 7: tolerance = .00012092
Iteration 8: tolerance = .0000433
Iteration 9: tolerance = .00001344
Iteration 10: tolerance = 3.755e-06
Iteration 11: tolerance = 9.680e-07

GEE population-averaged model                   Number of obs     =     16,722
Group and time vars:          codcurso ano      Number of groups  =      2,787
Link:                                logit      Obs per group:
Family:                          binomial                    min =          6
Correlation:                  unstructured                   avg =        6.0
                                                             max =          6
                                                Wald chi2(36)     =    2139.81
Scale parameter:                         1      Prob > chi2       =     0.0000

                                  (Std. Err. adjusted for clustering on codcurso)
```

eva	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
propfem	-.0134333	.0644963	-0.21	0.835	-.1398436	.112977
prop0a17	7.330615	1.161444	6.31	0.000	5.054226	9.607004
prop25a29	-.802168	.1353773	-5.93	0.000	-1.067503	-.5368333
prop30a34	-.9864023	.2652964	-3.72	0.000	-1.506374	-.4664309
prop35a39	-.540264	.3677512	-1.47	0.142	-1.261043	.1805151
prop40a49	.2133865	.3964503	0.54	0.590	-.5636419	.9904148
prop50a59	2.810423	.5036143	5.58	0.000	1.823357	3.797489
prop60mais	1.480656	.7018887	2.11	0.035	.1049797	2.856333
propnot	.3147691	.0310102	10.15	0.000	.2539903	.3755479
lnconc	-.0180279	.0105878	-1.70	0.089	-.0387797	.0027239
propescopub	-.2343523	.0744282	-3.15	0.002	-.3802289	-.0884756
propeti	-.0564881	.1029279	-0.55	0.583	-.2582231	.145247
proppdef	-3.061522	.992212	-3.09	0.002	-5.006222	-1.116822
proprenda	.5437954	.1345936	4.04	0.000	.2799969	.807594
propoutros	.400637	.1430129	2.80	0.005	.1203369	.6809371
lnconc_hat	-.121148	.0217407	-5.57	0.000	-.1637591	-.078537
propescopub_hat	-1.41542	.3327416	-4.25	0.000	-2.067582	-.7632585
propeti_hat	.0225263	.4288053	0.05	0.958	-.8179167	.8629693
proppdef_hat	-15.32849	3.392629	-4.52	0.000	-21.97792	-8.679058
proprenda_hat	.4119161	.5374369	0.77	0.443	-.6414408	1.465273
propoutros_hat	-2.750287	.5711275	-4.82	0.000	-3.869676	-1.630897
dg1	.2068895	.0539521	3.83	0.000	.1011454	.3126336
dg2	.2787737	.0567348	4.91	0.000	.1675755	.3899719
dg3	.5137912	.053205	9.66	0.000	.4095113	.6180711
dg4	.2532243	.0492082	5.15	0.000	.1567781	.3496705
dg5	.2088948	.0619674	3.37	0.001	.087441	.3303487
dg6	.3552554	.0450554	7.88	0.000	.2669485	.4435624
dg7	.2192165	.0509367	4.30	0.000	.1193825	.3190505
dg8	.1407892	.0495352	2.84	0.004	.0437019	.2378765
dg11	.2483679	.0794463	3.13	0.002	.0926559	.4040799
dreg1	.2928104	.0436715	6.70	0.000	.2072157	.378405
dreg2	.2947508	.0407797	7.23	0.000	.2148241	.3746776
dreg3	.4598572	.0516135	8.91	0.000	.3586966	.5610179
dreg5	.462778	.0630183	7.34	0.000	.3392644	.5862917
covid	.1092243	.0162461	6.72	0.000	.0773825	.1410661
poscovid	.1504817	.0212253	7.09	0.000	.1088809	.1920824
_cons	-1.209037	.1011945	-11.95	0.000	-1.407375	-1.0107

Fonte: Resultados da pesquisa (saída do software Stata®)

```
. xtgee eva propfem prop0a17 prop25a29 prop30a34 prop35a39 prop40a49 prop50a59 prop60mais propnot lnconc propescopu
> b propeti proppdef proprenda propoutros lnconc_hat propescopub_hat propeti_hat proppdef_hat proprenda_hat propout
> ros_hat dg1 dg2 dg3 dg4 dg5 dg6 dg7 dg8 dg11 dreg1 dreg2 dreg3 dreg5 covid poscovid, family(binomial) link(logit)
>  corr(ar1) vce(robust)

Iteration 1: tolerance = 1.3782338
Iteration 2: tolerance = .19387396
Iteration 3: tolerance = .03562538
Iteration 4: tolerance = .0082055
Iteration 5: tolerance = .00199252
Iteration 6: tolerance = .00048919
Iteration 7: tolerance = .00012036
Iteration 8: tolerance = .00002963
Iteration 9: tolerance = 7.296e-06
Iteration 10: tolerance = 1.797e-06
Iteration 11: tolerance = 4.424e-07

GEE population-averaged model                   Number of obs     =     16,722
Group and time vars:          codcurso ano      Number of groups  =      2,787
Link:                                logit      Obs per group:
Family:                           binomial                   min =          6
Correlation:                        AR(1)                    avg =        6.0
                                                               max =          6
                                               Wald chi2(36)     =    2490.83
Scale parameter:                         1     Prob > chi2       =     0.0000

                                (Std. Err. adjusted for clustering on codcurso)
```

eva	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
propfem	-.1288954	.0649702	-1.98	0.047	-.2562347	-.0015562
prop0a17	9.380611	1.272192	7.37	0.000	6.88716	11.87406
prop25a29	-.9062395	.1488178	-6.09	0.000	-1.197917	-.6145619
prop30a34	-.8869397	.2851245	-3.11	0.002	-1.445774	-.3281059
prop35a39	-.3968079	.4153892	-0.96	0.339	-1.210956	.4173401
prop40a49	.6790236	.4518048	1.50	0.133	-.2064975	1.564545
prop50a59	3.526481	.6023643	5.85	0.000	2.345869	4.707093
prop60mais	2.021287	.8364624	2.42	0.016	.3818512	3.660723
propnot	.2815054	.0316326	8.90	0.000	.2195067	.3435041
lnconc	-.0341619	.0102487	-3.33	0.001	-.054249	-.0140748
propescopub	-.2394057	.0823622	-2.91	0.004	-.4008326	-.0779789
propeti	-.2843431	.1140143	-2.49	0.013	-.5078071	-.0608791
proppdef	-2.057186	1.140237	-1.80	0.071	-4.292009	.1776374
proprenda	.5358608	.1479871	3.62	0.000	.2458114	.8259101
propoutros	.2160265	.1635592	1.32	0.187	-.1045437	.5365967
lnconc_hat	-.2191185	.0227512	-9.63	0.000	-.2637101	-.1745269
propescopub_hat	-1.24258	.3561479	-3.49	0.000	-1.940617	-.5445427
propeti_hat	.3394512	.4775954	0.71	0.477	-.5966185	1.275521
proppdef_hat	-19.77077	4.819056	-4.10	0.000	-29.21595	-10.32559
proprenda_hat	.6355176	.5954468	1.07	0.286	-.5315367	1.802572
propoutros_hat	-3.179631	.7590697	-4.19	0.000	-4.66738	-1.691882
dg1	.0464245	.0529845	0.88	0.381	-.0574233	.1502723
dg2	.1085972	.0557938	1.95	0.052	-.0007567	.217951
dg3	.3535997	.0524454	6.74	0.000	.2508086	.4563909
dg4	.1441368	.0460289	3.13	0.002	.0539218	.2343518
dg5	.1197685	.0625739	1.91	0.056	-.0028741	.2424111
dg6	.1807266	.043497	4.15	0.000	.095474	.2659792
dg7	.0669168	.0502397	1.33	0.183	-.0315512	.1653848
dg8	.0583338	.0466784	1.25	0.211	-.0331542	.1498217
dg11	.1189976	.0720831	1.65	0.099	-.0222826	.2602779
dreg1	.2041089	.0446855	4.57	0.000	.116527	.2916908
dreg2	.2968689	.0424816	6.99	0.000	.2136065	.3801313
dreg3	.4425602	.0572098	7.74	0.000	.330431	.5546894
dreg5	.2822047	.0694797	4.06	0.000	.146027	.4183823
covid	.0345825	.0180511	1.92	0.055	-.000797	.0699619
poscovid	.1282054	.0230138	5.57	0.000	.0830993	.1733116
_cons	-.8608079	.1084344	-7.94	0.000	-1.073335	-.6482804

Fonte: Resultados da pesquisa (saída do software Stata®)

```
. xtgee eva propfem prop0a17 prop25a29 prop30a34 prop35a39 prop40a49 prop50a59 prop60mais propnot lnconc propescopu
> b propeti proppdef proprenda propoutros lnconc_hat propescopub_hat propeti_hat proppdef_hat proprenda_hat propout
> ros_hat dg1 dg2 dg3 dg4 dg5 dg6 dg7 dg8 dg11 dreg1 dreg2 dreg3 dreg5 covid poscovid, family(binomial) link(logit)
>  corr(ar2) vce(robust)

Iteration 1: tolerance = 1.3839255
Iteration 2: tolerance = .12161162
Iteration 3: tolerance = .01878854
Iteration 4: tolerance = .00384562
Iteration 5: tolerance = .00075983
Iteration 6: tolerance = .00014985
Iteration 7: tolerance = .00002949
Iteration 8: tolerance = 5.809e-06
Iteration 9: tolerance = 1.143e-06
Iteration 10: tolerance = 2.252e-07

GEE population-averaged model                 Number of obs     =     16,722
Group and time vars:          codcurso ano    Number of groups  =      2,787
Link:                                logit    Obs per group:
Family:                          binomial                   min =          6
Correlation:                       AR(2)                    avg =        6.0
                                                            max =          6
                                            Wald chi2(36)     =    2303.58
Scale parameter:                         1  Prob > chi2       =     0.0000

                                (Std. Err. adjusted for clustering on codcurso)
```

eva	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
propfem	-.0936358	.0643056	-1.46	0.145	-.2196725	.0324009
prop0a17	8.713635	1.135301	7.68	0.000	6.488485	10.93878
prop25a29	-.7817174	.1459597	-5.36	0.000	-1.067793	-.4956417
prop30a34	-.9611711	.2858976	-3.36	0.001	-1.52152	-.400822
prop35a39	-.6832065	.4124859	-1.66	0.098	-1.491664	.125251
prop40a49	.4370249	.4520132	0.97	0.334	-.4489046	1.322954
prop50a59	3.158598	.556654	5.67	0.000	2.067576	4.24962
prop60mais	1.789015	.779478	2.30	0.022	.261266	3.316764
propnot	.3026895	.0308262	9.82	0.000	.2422711	.3631078
lnconc	-.0153698	.0101068	-1.52	0.128	-.0351788	.0044392
propescopub	-.3297689	.0751493	-4.39	0.000	-.4770587	-.182479
propeti	-.1784716	.1076734	-1.66	0.097	-.3895077	.0325645
proppdef	-1.808957	1.037147	-1.74	0.081	-3.841728	.2238144
proprenda	.6549573	.1367347	4.79	0.000	.3869621	.9229524
propoutros	.4405774	.1552241	2.84	0.005	.1363438	.7448111
lnconc_hat	-.2189471	.0220614	-9.92	0.000	-.2621867	-.1757075
propescopub_hat	-1.213712	.338118	-3.59	0.000	-1.876411	-.5510129
propeti_hat	.3746645	.4534243	0.83	0.409	-.5140307	1.26336
proppdef_hat	-17.05825	4.087537	-4.17	0.000	-25.06967	-9.046821
proprenda_hat	.3122549	.5356215	0.58	0.560	-.7375439	1.362054
propoutros_hat	-3.120261	.640293	-4.87	0.000	-4.375213	-1.86531
dg1	.0860112	.0526005	1.64	0.102	-.0170838	.1891062
dg2	.1596531	.055878	2.86	0.004	.0501342	.2691719
dg3	.4085457	.0521403	7.84	0.000	.3063526	.5107388
dg4	.1799605	.046643	3.86	0.000	.0885419	.2713791
dg5	.1587465	.0617845	2.57	0.010	.0376511	.2798418
dg6	.2372779	.043181	5.49	0.000	.1526447	.3219111
dg7	.1067421	.0498377	2.14	0.032	.0090621	.2044222
dg8	.0799855	.0471069	1.70	0.090	-.0123423	.1723132
dg11	.1734448	.0721211	2.40	0.016	.03209	.3147995
dreg1	.2221351	.0434607	5.11	0.000	.1369537	.3073165
dreg2	.2820572	.0413336	6.82	0.000	.2010448	.3630695
dreg3	.4005145	.0542482	7.38	0.000	.2941899	.506839
dreg5	.3583938	.0668298	5.36	0.000	.2274099	.4893778
covid	.0256767	.0172286	1.49	0.136	-.0080907	.0594441
poscovid	.0897528	.0218131	4.11	0.000	.0469998	.1325057
_cons	-.9315001	.1026315	-9.08	0.000	-1.132654	-.730346

Fonte: Resultados da pesquisa (saída do software Stata®)

2. Teste dos instrumentos para os períodos pré-pandemia (2018-2019), durante a pandemia (2020-2021), pós-pandemia (2022-2023) e todo o período (2018-2023), respectivamente

```
. test mean_propescopub mean_propeti mean_proppdef mean_proprenda mean_propoutros mean_lnconc

 ( 1)  mean_propescopub = 0
 ( 2)  mean_propeti = 0
 ( 3)  mean_proppdef = 0
 ( 4)  mean_proprenda = 0
 ( 5)  mean_propoutros = 0
 ( 6)  mean_lnconc = 0

           chi2(  6) =  188.84
         Prob > chi2 =    0.0000
```

Fonte: Resultados da pesquisa (saída do software Stata®)

```
. test mean_propescopub mean_propeti mean_proppdef mean_proprenda mean_propoutros mean_lnconc

 ( 1)  mean_propescopub = 0
 ( 2)  mean_propeti = 0
 ( 3)  mean_proppdef = 0
 ( 4)  mean_proprenda = 0
 ( 5)  mean_propoutros = 0
 ( 6)  mean_lnconc = 0

           chi2(  6) =  229.61
         Prob > chi2 =    0.0000
```

Fonte: Resultados da pesquisa (saída do software Stata®)

```
. test mean_propescopub mean_propeti mean_proppdef mean_proprenda mean_propoutros mean_lnconc

 ( 1)  mean_propescopub = 0
 ( 2)  mean_propeti = 0
 ( 3)  mean_proppdef = 0
 ( 4)  mean_proprenda = 0
 ( 5)  mean_propoutros = 0
 ( 6)  mean_lnconc = 0

           chi2(  6) =  473.33
         Prob > chi2 =    0.0000
```

Fonte: Resultados da pesquisa (saída do software Stata®)

```
. test mean_propescopub mean_propeti mean_proppdef mean_proprenda mean_propoutros mean_lnconc

 ( 1)  mean_propescopub = 0
 ( 2)  mean_propeti = 0
 ( 3)  mean_proppdef = 0
 ( 4)  mean_proprenda = 0
 ( 5)  mean_propoutros = 0
 ( 6)  mean_lnconc = 0

           chi2(  6) =   19.32
         Prob > chi2 =    0.0037
```

Fonte: Resultados da pesquisa (saída do software Stata®)

3. Teste de correlação entre valores preditos e reais para os períodos pré-pandemia (2018-2019), durante a pandemia (2020-2021), pós-pandemia (2022-2023) e todo o período (2018-2023), respectivamente

```
. correlate eva p_hat
(obs=16,722)
```

	eva	p_hat
eva	**1.0000**	
p_hat	**0.4140**	**1.0000**

Fonte: Resultados da pesquisa (saída do software Stata®)

```
. correlate eva p_hat
(obs=16,722)
```

	eva	p_hat
eva	**1.0000**	
p_hat	**0.4510**	**1.0000**

Fonte: Resultados da pesquisa (saída do software Stata®)

```
. correlate eva p_hat
(obs=16,722)
```

	eva	p_hat
eva	**1.0000**	
p_hat	**0.4725**	**1.0000**

Fonte: Resultados da pesquisa (saída do software Stata®)

```
. correlate eva p_hat
(obs=16,722)
```

	eva	p_hat
eva	**1.0000**	
p_hat	**0.5128**	**1.0000**

Fonte: Resultados da pesquisa (saída do software Stata®)

APÊNDICE B.2 – MODELO DE REGRESSÃO PARA ANÁLISES QUANTO À INFLUÊNCIA DOS PARÂMETROS (VARIÁVEIS INDEPENDENTES) EM RELAÇÃO À EVASÃO (VARIÁVEL DEPENDENTE) NO PERÍODO PRÉ-PANDEMIA (2018-2019)

```
. xtgee eva propfem prop0a17 prop25a29 prop30a34 prop35a39 prop40a49 prop50a59 prop60mais propnot lnconc propescopu
> b propeti proppdef proprenda propoutros lnconc hat propescopub hat propeti hat proppdef hat proprenda hat propout
> ros_hat dg1 dg2 dg3 dg4 dg5 dg6 dg7 dg8 dg11 dreg1 dreg2 dreg3 dreg5 if ano<2020, family(binomial) link(logit) co
> rr(ar1) vce(robust)
note: propnot != 0 predicts success perfectly
      propnot dropped and 1867 obs not used

note: dg4 != 0 predicts success perfectly
      dg4 dropped and 273 obs not used

note: dg8 != 0 predicts success perfectly
      dg8 dropped and 132 obs not used

note: dg11 != 0 predicts success perfectly
      dg11 dropped and 40 obs not used

note: dreg3 != 0 predicts success perfectly
      dreg3 dropped and 266 obs not used

note: dreg5 != 0 predicts success perfectly
      dreg5 dropped and 734 obs not used

note:  some groups have fewer than 2 observations
       not possible to estimate correlations for those groups
       8 groups omitted from estimation


Iteration 1: tolerance = .37436924
Iteration 2: tolerance = .02694723
Iteration 3: tolerance = .00153749
Iteration 4: tolerance = .00010371
Iteration 5: tolerance = 9.198e-06
Iteration 6: tolerance = 3.084e-07

GEE population-averaged model                   Number of obs     =      2,254
Group and time vars:          codcurso ano      Number of groups  =      1,127
Link:                               logit       Obs per group:
Family:                          binomial                    min =          2
Correlation:                        AR(1)                    avg =        2.0
                                                             max =          2
                                                Wald chi2(28)     =     639.30
Scale parameter:                         1      Prob > chi2       =     0.0000

                                (Std. Err. adjusted for clustering on codcurso)
------------------------------------------------------------------------------
             |               Robust
         eva |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     propfem |   -.3483443   .1408498    -2.47   0.013    -.6244047   -.0722838
    prop0a17 |    7.900508   1.930302     4.09   0.000     4.117185    11.68383
   prop25a29 |   -1.054288   .3733527    -2.82   0.005    -1.786045   -.3225298
   prop30a34 |   -1.167272   .8415873    -1.39   0.165    -2.816753    .4822085
   prop35a39 |    1.743683   1.227889     1.42   0.156    -.6629353    4.150302
   prop40a49 |     1.09523    .923778     1.19   0.236    -.7153421    2.905801
   prop50a59 |    5.245386   1.235218     4.25   0.000     2.824403     7.66637
  prop60mais |   -.1122027   3.050676    -0.04   0.971    -6.091417    5.867012
     propnot |           0  (omitted)
      lnconc |   -.0741469   .0296735    -2.50   0.012    -.1323059   -.0159878
  propescopub |   -.2026603   .2004021    -1.01   0.312    -.5954412    .1901206
     propeti |    .4780288   .2427863     1.97   0.049     .0021764    .9538813
    proppdef |   -.8009767   2.296899    -0.35   0.727    -5.302815    3.700862
   proprenda |   -.0999524   .3509717    -0.28   0.776    -.7878443    .5879396
  propoutros |   -1.812366   .4736682    -3.83   0.000    -2.740738   -.8839929
  lnconc hat |   -.5117176   .0765538    -6.68   0.000    -.6617603    -.361675
propescopub_hat |   -1.843612   .7969041    -2.31   0.021    -3.405515   -.2817087
  propeti_hat |    4.77747   1.068518     4.47   0.000     2.683212    6.871727
 proppdef hat |   -18.03111   10.07356    -1.79   0.073    -37.77492    1.712709
proprenda_hat |    1.164652   1.178623     0.99   0.323    -1.145407    3.474711
propoutros_hat |   -2.171366   2.465857    -0.88   0.379    -7.004358    2.661625
         dg1 |   -.0590086    .08551    -0.69   0.490    -.2266052     .108588
         dg2 |   -.3644149   .1033256    -3.53   0.000    -.5669293   -.1619004
         dg3 |    .1579893   .0921429     1.71   0.086    -.0226076    .3385861
         dg4 |           0  (omitted)
```

dg5	.048408	.1186745	0.41	0.683	-.1841898	.2810057
dg6	-.0816206	.1032992	-0.79	0.429	-.2840832	.1208421
dg7	-.1334198	.093996	-1.42	0.156	-.3176485	.0508089
dg8	0	(omitted)				
dg11	0	(omitted)				
dreg1	-.1513647	.0939895	-1.61	0.107	-.3355808	.0328515
dreg2	-.155657	.1068983	-1.46	0.145	-.3651739	.0538599
dreg3	0	(omitted)				
dreg5	0	(omitted)				
_cons	-.5361995	.2592333	-2.07	0.039	-1.044287	-.0281116

Fonte: Resultados da pesquisa (saída do software Stata®)

APÊNDICE B.3 – MODELO DE REGRESSÃO PARA ANÁLISES QUANTO À INFLUÊNCIA DOS PARÂMETROS (VARIÁVEIS INDEPENDENTES) EM RELAÇÃO À EVASÃO (VARIÁVEL DEPENDENTE) NO PERÍODO DURANTE A PANDEMIA (2020-2021)

```
. xtgee eva propfem prop0a17 prop25a29 prop30a34 prop35a39 prop40a49 prop50a59 prop60mais propnot lnconc propescopu
> b propeti proppdef proprenda propoutros lnconc_hat propescopub_hat propeti_hat proppdef_hat proprenda_hat propout
> ros_hat dg1 dg2 dg3 dg4 dg5 dg6 dg7 dg8 dg11 dreg1 dreg2 dreg3 dreg5 if ano==2020|ano==2021, family(binomial) lin
> k(logit) corr(ar1) vce(robust)

Iteration 1: tolerance = 1.7254513
Iteration 2: tolerance = .1605599
Iteration 3: tolerance = .01769262
Iteration 4: tolerance = .00231458
Iteration 5: tolerance = .00031672
Iteration 6: tolerance = .00004264
Iteration 7: tolerance = 5.828e-06
Iteration 8: tolerance = 7.848e-07

GEE population-averaged model                   Number of obs      =      5,574
Group and time vars:          codcurso ano      Number of groups   =      2,787
Link:                               logit       Obs per group:
Family:                          binomial                    min =          2
Correlation:                       AR(1)                     avg =        2.0
                                                             max =          2
                                               Wald chi2(34)      =    1156.02
Scale parameter:                        1      Prob > chi2        =     0.0000

                                (Std. Err. adjusted for clustering on codcurso)
```

eva	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
propfem	-.0690701	.0910324	-0.76	0.448	-.2474903	.1093501
prop0a17	15.10884	2.198442	6.87	0.000	10.79998	19.41771
prop25a29	-1.456314	.2784905	-5.23	0.000	-2.002145	-.9104822
prop30a34	-1.470132	.5278422	-2.79	0.005	-2.504683	-.4355799
prop35a39	-.2579279	.6986595	-0.37	0.712	-1.627275	1.11142
prop40a49	.3516967	.7399259	0.48	0.635	-1.098531	1.801925
prop50a59	4.623109	1.011421	4.57	0.000	2.640759	6.605459
prop60mais	4.356468	1.642446	2.65	0.008	1.137333	7.575604
propnot	.3267528	.0444199	7.36	0.000	.2396915	.4138142
lnconc	-.0390381	.0186591	-2.09	0.036	-.0756092	-.002467
propescopub	-.7050226	.130635	-5.40	0.000	-.9610625	-.4489827
propeti	.6416547	.1880658	3.41	0.001	.2730525	1.010257
proppdef	-2.531535	1.746772	-1.45	0.147	-5.955146	.8920762
proprenda	1.192972	.2212306	5.39	0.000	.7593682	1.626576
propoutros	-.1381871	.2118983	-0.65	0.514	-.5535001	.2771259
lnconc_hat	-.0955733	.0310387	-3.08	0.002	-.1564081	-.0347385
propescopub_hat	-.8106708	.4886607	-1.66	0.097	-1.768428	.1470866
propeti_hat	-1.740437	.6521437	-2.67	0.008	-3.018615	-.4622588
proppdef_hat	-22.66643	5.442419	-4.16	0.000	-33.33338	-11.99949
proprenda_hat	.8579891	.8762953	0.98	0.328	-.8595181	2.575496
propoutros_hat	-3.432187	.838509	-4.09	0.000	-5.075635	-1.78874
dg1	.1772333	.0739235	2.40	0.017	.0323458	.3221208
dg2	.301875	.0789079	3.83	0.000	.1472184	.4565315
dg3	.4636107	.0731505	6.34	0.000	.3202383	.606983
dg4	.2459799	.066066	3.72	0.000	.1164928	.3754669
dg5	.2065458	.0924907	2.23	0.026	.0252674	.3878242
dg6	.3274877	.0625366	5.24	0.000	.2049182	.4500572
dg7	.1769367	.0706641	2.50	0.012	.0384377	.3154358
dg8	.0611569	.0660839	0.93	0.355	-.0683652	.190679
dg11	.1662571	.1235566	1.35	0.178	-.0759095	.4084236
dreg1	.3507591	.0637868	5.50	0.000	.2257392	.475779
dreg2	.5242665	.0558753	9.38	0.000	.4147529	.6337801
dreg3	.9345163	.0713667	13.09	0.000	.7946401	1.074393
dreg5	.0767401	.0926337	0.83	0.407	-.1048187	.2582989
_cons	-.9958074	.1547047	-6.44	0.000	-1.299023	-.6925918

Fonte: Resultados da pesquisa (saída do software Stata®)

APÊNDICE B.4 – MODELO DE REGRESSÃO PARA ANÁLISES QUANTO À INFLUÊNCIA DOS PARÂMETROS (VARIÁVEIS INDEPENDENTES) EM RELAÇÃO À EVASÃO (VARIÁVEL DEPENDENTE) NO PERÍODO PÓS-PANDEMIA (2022-2023)

```
. xtgee eva propfem prop0a17 prop25a29 prop30a34 prop35a39 prop40a49 prop50a59 prop60mais propnot lnconc propescopu
> b propeti proppdef proprenda propoutros lnconc_hat propescopub_hat propeti_hat proppdef_hat proprenda_hat propout
> ros_hat dg1 dg2 dg3 dg4 dg5 dg6 dg7 dg8 dg11 dreg1 dreg2 dreg3 dreg5 if ano>2021, family(binomial) link(logit) co
> rr(ar1) vce(robust)
note: dg1 != 0 predicts success perfectly
      dg1 dropped and 398 obs not used

note: dg7 != 0 predicts success perfectly
      dg7 dropped and 984 obs not used


Iteration 1: tolerance = .20795164
Iteration 2: tolerance = .01106209
Iteration 3: tolerance = .00048262
Iteration 4: tolerance = .00002414
Iteration 5: tolerance = 1.220e-06
Iteration 6: tolerance = 6.168e-08

GEE population-averaged model                   Number of obs     =      4,192
Group and time vars:          codcurso ano      Number of groups  =      2,096
Link:                                logit      Obs per group:
Family:                          binomial                    min =          2
Correlation:                       AR(1)                     avg =        2.0
                                                             max =          2
                                                Wald chi2(32)     =    2210.40
Scale parameter:                        1       Prob > chi2       =     0.0000

                                  (Std. Err. adjusted for clustering on codcurso)
```

eva	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
propfem	-.3116563	.0867167	-3.59	0.000	-.4816179	-.1416948
prop0a17	12.07854	3.431793	3.52	0.000	5.35235	18.80473
prop25a29	-.608329	.2287393	-2.66	0.008	-1.05665	-.1600083
prop30a34	-.0210755	.4070137	-0.05	0.959	-.8188077	.7766568
prop35a39	-.3174327	.5873636	-0.54	0.589	-1.468644	.8337787
prop40a49	.9303209	.5848752	1.59	0.112	-.2160134	2.076655
prop50a59	3.184801	.8921791	3.57	0.000	1.436162	4.93344
prop60mais	1.445697	1.139333	1.27	0.204	-.787355	3.67875
propnot	.1703273	.033995	5.01	0.000	.1036983	.2369562
lnconc	-.1004267	.0149653	-6.71	0.000	-.1297583	-.0710952
propescopub	.0347496	.1618859	0.21	0.830	-.2825408	.3520401
propeti	-.3952725	.1943986	-2.03	0.042	-.7762868	-.0142582
proppdef	-.7144636	1.313057	-0.54	0.586	-3.288007	1.85908
proprenda	-.1395539	.3353017	-0.42	0.677	-.7967333	.5176254
propoutros	.5597618	.191073	2.93	0.003	.1852655	.934258
lnconc_hat	-.3532211	.0441724	-8.00	0.000	-.4397974	-.2666448
propescopub_hat	.0378022	.5578134	0.07	0.946	-1.055492	1.131096
propeti_hat	2.288081	.6293169	3.64	0.000	1.054642	3.521519
proppdef_hat	-3.748199	3.251667	-1.15	0.249	-10.12135	2.624952
proprenda_hat	-1.08395	1.070253	-1.01	0.311	-3.181606	1.013707
propoutros_hat	-.3614428	.439654	-0.82	0.411	-1.223149	.5002632
dg1	0	(omitted)				
dg2	.0198971	.0627186	0.32	0.751	-.1030291	.1428233
dg3	.1546413	.0604492	2.56	0.011	.0361631	.2731196
dg4	.0425673	.0503293	0.85	0.398	-.0560763	.141211
dg5	-.0501618	.0726421	-0.69	0.490	-.1925378	.0922141
dg6	.0135472	.0541244	0.25	0.802	-.0925347	.1196291
dg7	0	(omitted)				
dg8	.0905295	.0528806	1.71	0.087	-.0131146	.1941736
dg11	.0670391	.0931065	0.72	0.472	-.1154463	.2495244
dreg1	.007575	.0615896	0.12	0.902	-.1131384	.1282883
dreg2	-.0920136	.058124	-1.58	0.113	-.2059345	.0219073
dreg3	-.0003345	.0650231	-0.01	0.996	-.1277775	.1271085
dreg5	.4791003	.0870922	5.50	0.000	.3084028	.6497978
_cons	-.9962355	.1383384	-7.20	0.000	-1.267374	-.7250972

Fonte: Resultados da pesquisa (saída do software Stata®)

APÊNDICE B.5 – MODELO DE REGRESSÃO PARA ANÁLISES QUANTO À INFLUÊNCIA DOS PARÂMETROS (VARIÁVEIS INDEPENDENTES) EM RELAÇÃO À EVASÃO (VARIÁVEL DEPENDENTE) EM TODO O PERÍODO (2018 A 2023)

```
. xtgee eva propfem prop0a17 prop25a29 prop30a34 prop35a39 prop40a49 prop50a59 prop60mais propnot lnconc propescopu
> b propeti proppdef proprenda propoutros lnconc_hat propescopub_hat propeti_hat proppdef_hat proprenda_hat propout
> ros_hat dg1 dg2 dg3 dg4 dg5 dg6 dg7 dg8 dg11 dreg1 dreg2 dreg3 dreg5 covid poscovid, family(binomial) link(logit)
>  corr(ar1) vce(robust)

Iteration 1: tolerance = 1.3782338
Iteration 2: tolerance = .19387396
Iteration 3: tolerance = .03562538
Iteration 4: tolerance = .0082055
Iteration 5: tolerance = .00199252
Iteration 6: tolerance = .00048919
Iteration 7: tolerance = .00012036
Iteration 8: tolerance = .00002963
Iteration 9: tolerance = 7.296e-06
Iteration 10: tolerance = 1.797e-06
Iteration 11: tolerance = 4.424e-07

GEE population-averaged model                   Number of obs     =     16,722
Group and time vars:          codcurso ano      Number of groups  =      2,787
Link:                                logit      Obs per group:
Family:                          binomial                    min =          6
Correlation:                       AR(1)                     avg =        6.0
                                                             max =          6
                                                Wald chi2(36)     =    2490.83
Scale parameter:                        1       Prob > chi2       =     0.0000

                                (Std. Err. adjusted for clustering on codcurso)
```

eva	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
propfem	-.1288954	.0649702	-1.98	0.047	-.2562347	-.0015562
prop0a17	9.380611	1.272192	7.37	0.000	6.88716	11.87406
prop25a29	-.9062395	.1488178	-6.09	0.000	-1.197917	-.6145619
prop30a34	-.8869397	.2851245	-3.11	0.002	-1.445774	-.3281059
prop35a39	-.3968079	.4153892	-0.96	0.339	-1.210956	.4173401
prop40a49	.6790236	.4518048	1.50	0.133	-.2064975	1.564545
prop50a59	3.526481	.6023643	5.85	0.000	2.345869	4.707093
prop60mais	2.021287	.8364624	2.42	0.016	.3818512	3.660723
propnot	.2815054	.0316326	8.90	0.000	.2195067	.3435041
lnconc	-.0341619	.0102487	-3.33	0.001	-.054249	-.0140748
propescopub	-.2394057	.0823622	-2.91	0.004	-.4008326	-.0779789
propeti	-.2843431	.1140143	-2.49	0.013	-.5078071	-.0608791
proppdef	-2.057186	1.140237	-1.80	0.071	-4.292009	.1776374
proprenda	.5358608	.1479871	3.62	0.000	.2458114	.8259101
propoutros	.2160265	.1635592	1.32	0.187	-.1045437	.5365967
lnconc_hat	-.2191185	.0227512	-9.63	0.000	-.2637101	-.1745269
propescopub_hat	-1.24258	.3561479	-3.49	0.000	-1.940617	-.5445427
propeti_hat	.3394512	.4775954	0.71	0.477	-.5966185	1.275521
proppdef_hat	-19.77077	4.819056	-4.10	0.000	-29.21595	-10.32559
proprenda_hat	.6355176	.5954468	1.07	0.286	-.5315367	1.802572
propoutros_hat	-3.179631	.7590697	-4.19	0.000	-4.66738	-1.691882
dg1	.0464245	.0529845	0.88	0.381	-.0574233	.1502723
dg2	.1085972	.0557938	1.95	0.052	-.0007567	.217951
dg3	.3535997	.0524454	6.74	0.000	.2508086	.4563909
dg4	.1441368	.0460289	3.13	0.002	.0539218	.2343518
dg5	.1197685	.0625739	1.91	0.056	-.0028741	.2424111
dg6	.1807266	.043497	4.15	0.000	.095474	.2659792
dg7	.0669168	.0502397	1.33	0.183	-.0315512	.1653848
dg8	.0583338	.0466784	1.25	0.211	-.0331542	.1498217
dg11	.1189976	.0720831	1.65	0.099	-.0222826	.2602779
dreg1	.2041089	.0446855	4.57	0.000	.116527	.2916908
dreg2	.2968689	.0424816	6.99	0.000	.2136065	.3801313
dreg3	.4425602	.0572098	7.74	0.000	.330431	.5546894
dreg5	.2822047	.0694797	4.06	0.000	.146027	.4183823
covid	.0345825	.0180511	1.92	0.055	-.000797	.0699619
poscovid	.1282054	.0230138	5.57	0.000	.0830993	.1733116
_cons	-.8608079	.1084344	-7.94	0.000	-1.073335	-.6482804

Fonte: Resultados da pesquisa (saída do software Stata®)