



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
DEPARTAMENTO DE ESTATÍSTICA E INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA APLICADA

LHAÍSLLA EDUARDA CAVALCANTI RODRIGUES DA SILVA

TRAJETÓRIAS ACADÊMICAS EM RISCO: UMA
ABORDAGEM DE ANÁLISE DE SOBREVIVÊNCIA
PARA IDENTIFICAÇÃO DE DISPARIDADES
SOCIOECONÔMICAS NA EVASÃO

RECIFE – PE

2026

LHAÍSLLA EDUARDA CAVALCANTI RODRIGUES DA SILVA

**TRAJETÓRIAS ACADÊMICAS EM RISCO: UMA
ABORDAGEM DE ANÁLISE DE SOBREVIVÊNCIA
PARA IDENTIFICAÇÃO DE DISPARIDADES
SOCIOECONÔMICAS NA EVASÃO**

Dissertação submetida à Coordenação do
Programa de Pós-Graduação em Informática
Aplicada do Departamento de Estatística
e Informática - DEINFO - Universidade
Federal Rural de Pernambuco, como parte
dos requisitos necessários para obtenção do
grau de Mestre.

ORIENTADOR: Gabriel Alves de Albuquerque Junior

CO-ORIENTADORA: Maria da Conceição Moraes Batista

RECIFE – PE

2026

Dados Internacionais de Catalogação na Publicação
Sistema Integrado de Bibliotecas da UFRPE
Bibliotecário(a): Suely Manzi – CRB-4 809

S586t Silva, Lhaíslla Eduarda Cavalcanti Rodrigues da.
Trajetórias acadêmicas em risco: uma abordagem de
análise de sobrevivência para identificação de disparidades
socioeconômicas na evasão / Lhaíslla Eduarda Cavalcanti
Rodrigues da Silva. – Recife, 2026.
135 f.; il.

Orientador(a): Gabriel Alves de Albuquerque Junior.
Co-orientador(a): Maria da Conceição Moraes Batista.

Dissertação (Mestrado) – Universidade Federal Rural de
Pernambuco, Programa de Pós-Graduação em Informática
Aplicada, Recife, BR-PE, 2026.

Inclui referências.

1. Análise de sobrevivência. 2. Pessoal - Permanência. 3.
Evasão universitária. 4. Ensino Superior 5. Gestão baseada
em evidências. I. Albuquerque Junior, Gabriel Alves de,
orient. II. Batista, Maria da Conceição Moraes, coorient. III.
Título

CDD 004

LHAÍSLLA EDUARDA CAVALCANTI RODRIGUES DA SILVA

**TRAJETÓRIAS ACADÊMICAS EM RISCO: UMA
ABORDAGEM DE ANÁLISE DE SOBREVIVÊNCIA
PARA IDENTIFICAÇÃO DE DISPARIDADES
SOCIOECONÔMICAS NA EVASÃO**

Dissertação submetida à Coordenação do
Programa de Pós-Graduação em Informática
Aplicada do Departamento de Estatística
e Informática - DEINFO - Universidade
Federal Rural de Pernambuco, como parte
dos requisitos necessários para obtenção do
grau de Mestre.

Aprovada em: 27 de 02 de 2026.

BANCA EXAMINADORA

Gabriel Alves de Albuquerque Junior (Orientador)
Universidade Federal Rural de Pernambuco (UFRPE)
Departamento de Estatística e Informática (DEINFO)

Maria da Conceição Moraes Batista (Coorientadora)
Universidade Federal Rural de Pernambuco (UFRPE)
Departamento de Estatística e Informática (DEINFO)

Andrêza Leite de Alencar
Universidade Federal Rural de Pernambuco (UFRPE)
Departamento de Computação (DC)

Liliane Sheyla da Silva Fonseca
Universidade de Pernambuco (UPE)
Escola Politécnica de Pernambuco

Aos estudantes da UFRPE, que, apesar dos desafios, persistem no sonho da formação superior e contribuem para uma sociedade mais justa e fundamentada no conhecimento. E aos professores que atuam para que fatores institucionais não se tornem impeditivos à realização do objetivo de formação dos discentes.

Agradecimentos

A permanência e o desenvolvimento no ensino superior são influenciados por variáveis acadêmicas, além de fatores externos e internos que impactam a formação do estudante. A educação permite a superação desses obstáculos e a construção de caminhos. Nesse sentido, durante o mestrado, a pesquisa tornou-se um espaço de refúgio e amadurecimento, onde pude adquirir conhecimentos e fortalecer minha trajetória pelo apoio de pessoas em cada etapa.

Agradeço ao meu orientador, Professor Gabriel Alves, e à minha coorientadora, Professora Ceça, que me acompanham desde a graduação. Em *Ensinando a Transgredir*, bell hooks menciona, ao citar Thich Nhat Hanh sobre Paulo Freire, que a presença de certos mestres traz consigo uma "zona de luz" e uma atmosfera de paz. Essa descrição reflete o processo de orientação com o Professor Gabriel, onde o compartilhamento de conhecimento transmitiu segurança e clareza. Por tornar o processo acolhedor, onde por vezes pude ter a UFRPE como ambiente de refúgio, quando nem sempre minha casa o era, agradeço por ter disponibilizado o laboratório e o acesso aos inúmeros livros de sua biblioteca para construir e aprofundar meus conhecimentos.

Estendo esse reconhecimento às professoras Andréza Leite e Taciana Pontual por incentivarem e incluírem a participação de mais mulheres na área de tecnologia e na área acadêmica, assim como pelo impulso de coragem para que eu participasse de eventos da área de TI, compartilhando conhecimento. Aos professores do Programa de Pós-Graduação em Informática Aplicada (PPGIA), pelas contribuições ao aprendizado, e ao Professor Thiago Espinola, com quem cursei disciplinas desde a graduação, por mostrar a paixão pela pesquisa e demonstrar como a estrutura de um algoritmo impacta os resultados obtidos.

No âmbito institucional, manifesto gratidão à Universidade Federal Rural de Pernambuco (UFRPE) e ao Observatório de Dados da Graduação (ODG) pelo suporte institucional e pelos recursos disponibilizados, fundamentais para a viabilização desta pesquisa. De forma especial, ressalto que a concessão da bolsa de estudos foi o que viabilizou minha permanência e dedicação ao curso, reafirmando a relevância das políticas de assistência estudantil e do apoio institucional para a conclusão desta etapa acadêmica. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior — Brasil (CAPES).

Ninguém realiza trajetórias de forma isolada; nesse sentido, dedico um agradecimento especial aos meus avós Sheila e Paulo, que se fizeram presentes no apoio e na escuta. Às minhas tias Karina e Andressa, por sempre apontarem a direção da educação como um caminho de transformação. Ao meu tio Jerry, pelo cuidado e presença em minhas conquistas. E à minha irmã Thaíslla, por ser a maior incentivadora dos meus sonhos, que escuta e acompanha cada tentativa de início de sonho e que embarca comigo rumo ao desconhecido.

Quanto ao apoio pessoal, menciono meus amigos, que nesse processo me apoiaram não só academicamente, mas nas dificuldades da vida. À Evelyn, que espero ter sempre em minha vida, agradeço por estar ao meu lado e por tornar a vida mais leve, por me apoiar em sonhos distantes e por acreditar que cada etapa está me levando mais perto dos meus objetivos.

Somam-se a este grupo os meus amigos de laboratório, Arthur, Melissa, Nathan e Tarcísio pela divisão de aprendizados e pelos questionamentos que auxiliaram na evolução da pesquisa. À Bianka, meu agradecimento por ser parceria certa nos eventos de TI e pelas trocas e conversas que trouxeram leveza e brilho a este percurso; e ao Igor, pelas caminhadas e conversas sinceras na Transrural que enriqueceram essa trajetória, assim como a todos os outros amigos presentes.

Por fim, ressalto a companhia do meu gato Bartolomeu, que me permite desligar das preocupações da vida, oferecendo tranquilidade nos momentos difíceis. Encerro esta seção agradecendo ao meu amigo Abílio Nogueira, que me motivou a ingressar no mestrado, mesmo quando eu não acreditava que conseguiria, e que sempre compartilha comigo experiências acadêmicas e tecnologias em ascensão, assim como a motivação em estudar temas que tenham impacto social.

A academia não é o paraíso. Mas o
aprendizado é um lugar onde o paraíso pode
ser criado.

(bell hooks)

Resumo

A evasão no ensino superior caracteriza-se como um fenômeno multidimensional que gera repercussões econômicas, sociais e institucionais. Trata-se de um problema complexo que compromete a eficiência institucional, especialmente em áreas de Ciência, Tecnologia, Engenharia e Matemática (STEM), onde mais de 60% das desistências ocorrem nos primeiros dois anos. Diante disso, este trabalho objetiva consolidar um modelo computacional de análise de sobrevivência para identificar trajetórias acadêmicas em risco e quantificar o impacto de variáveis sociodemográficas (gênero e turno) e socioeconômicas (assistência estudantil) no fenômeno da evasão. A metodologia fundamenta-se no uso da análise de sobrevivência para observar o evento de abandono como uma função do tempo, integrando dados censurados de estudantes com vínculo ativo para evitar a subestimação do tempo de permanência. O modelo utiliza o estimador de Kaplan-Meier para examinar padrões de permanência baseados em gênero, turno, assistência estudantil e mudanças no Projeto Pedagógico de Curso (PPC). A validação estatística dos resultados é realizada por meio dos testes de Log-Rank (significância), Hazard Ratio (risco relativo) e Hedges'g (magnitude do efeito), permitindo mensurar a intensidade das disparidades encontradas. O modelo foi integrado à ferramenta de *Business Intelligence, System of Academic Business Intelligence and Analytics* - SABIA, utilizando uma base de dados dos 7.432 estudantes de cursos de STEM da UFRPE. Os resultados indicam que os quatro primeiros períodos letivos constituem a janela de maior risco de desligamento. No curso de Licenciatura em Computação (LC), a reforma curricular postergou o marco de 50% de evasão do 5º para o 9º período, enquanto no Bacharelado em Sistemas de Informação (BSI) a reestruturação reduziu o risco de desligamento em 46,43%, sustentando a sobrevivência discente em 64,80% ao final de dez períodos. No que tange às disparidades sociodemográficas, a estimativa de sobrevivência final para o turno noturno em STEM registrou 23,04%, valor inferior aos índices dos turnos diurnos, que apresentam 39,41% no matutino e 40,07% no vespertino. Essa diferença indica que a sobrecarga laboral atua como barreira estrutural à permanência nestas áreas. A assistência estudantil atua como fator de proteção institucional ao registrar impacto com g de Hedges de 0,8058 no turno noturno de STEM, o que demonstra a aplicação da intervenção em perfis de risco de desligamento. No cenário institucional, essa política mantém uma probabilidade de vínculo de 65,20% frente aos 38,70% observados entre discentes sem suporte, enquanto no recorte de STEM, a sobrevivência dos assistidos atinge

56,91% contra 31,95% dos não assistidos. A assistência estudantil atua como mecanismo de equidade ao neutralizar as disparidades de gênero, uma vez que a ausência de significância estatística entre as curvas de sobrevivência dos beneficiários equaliza as trajetórias de homens e mulheres. A vulnerabilidade ocorre no perfil de homens não assistidos em STEM, que apresentam estimativa de sobrevivência de 27,58% e *Hazard Ratio* de 1,1600. Esse indicador registra que o risco de desligamento para homens nessa condição é 16,00% superior ao grupo feminino sob a mesma condição socioeconômica, o que indica a necessidade de intervenções institucionais direcionadas para mitigar o risco de exclusão deste grupo.

Palavras-chave: Análise de Sobrevivência. Permanência Acadêmica. Evasão Estudantil. Ensino Superior. Gestão Baseada em Evidências.

Abstract

Dropout rates in higher education are characterised as a multidimensional phenomenon with economic, social and institutional repercussions. It is a complex problem that undermines institutional efficiency, particularly in Science, Technology, Engineering and Mathematics (STEM) fields, where over 60% of dropouts occur within the first two years. In view of this, this study aims to develop a computational survival analysis model to identify academic trajectories at risk and quantify the impact of sociodemographic (gender and study session) and socio-economic (student support) variables on the dropout phenomenon. The methodology is based on the use of survival analysis to observe the event of dropout as a function of time, integrating censored data from students with active enrolment to avoid underestimating the length of stay. The model uses the Kaplan-Meier estimator to examine retention patterns based on gender, shift, student support and changes to the Course Pedagogical Project (PPC). Statistical validation of the results is performed using the Log-Rank test (significance), Hazard Ratio (relative risk) and Hedges' g (effect size), enabling the intensity of the disparities found to be measured. The model was integrated into the *Business Intelligence* tool, *System of Academic Business Intelligence and Analytics* – SABIA, using a database of 7,432 students on STEM courses at UFRPE. The results indicate that the first four academic terms constitute the period of highest risk of dropout. In the Bachelor's degree in Computing (LC), the curriculum reform postponed the 50% dropout threshold from the 5th to the 9th term, whilst in the Bachelor's degree in Information Systems (BSI), the restructuring reduced the risk of dropout by 46.43%, maintaining student retention at 64.80% by the end of ten semesters. With regard to socio-demographic disparities, the estimated final survival rate for the evening shift in STEM stood at 23.04%, a figure lower than the rates for the day shifts, which were 39.41% for the morning shift and 40.07% for the afternoon shift. This difference indicates that work overload acts as a structural barrier to retention in these fields. Student support acts as an institutional protective factor, showing an impact with a Hedges' g of 0.8058 in the STEM evening shift, which demonstrates the effectiveness of the intervention in profiles at risk of dropping out. In the institutional context, this policy maintains a retention probability of 65.20% compared to the 38.70% observed among students without support, whilst in the STEM segment, the survival rate of those receiving support reaches 56.91% compared to 31.95% for those without support. Student support acts as a mechanism for equity

by neutralising gender disparities, since the lack of statistical significance between the survival curves of beneficiaries equalises the trajectories of men and women. Vulnerability is evident in the profile of unsupported men in STEM, who have an estimated survival rate of 27.58% and a hazard Ratio of 1.1600. This indicator shows that the risk of dropping out for men in this situation is 16.00% higher than for the female group under the same socioeconomic conditions, which points to the need for targeted institutional interventions to mitigate the risk of exclusion for this group.

Keywords: Survival Analysis. Academic Retention. Student Dropout. Higher Education. Evidence-Based Management.

Lista de Figuras

Figura 1 – Representação do ciclo de matrícula.	27
Figura 2 – Despesa governamental por aluno por nível de ensino (2022).	27
Figura 3 – Número de matrículas no ensino superior por grau acadêmico (2014-2024).	28
Figura 4 – Concluintes no ensino superior por gênero e categoria administrativa.	28
Figura 5 – Evolução média dos indicadores de trajetória dos ingressantes (2014-2023).	29
Figura 6 – Representação da censura sob a escala de tempo de estudo.	31
Figura 7 – Fluxo do modelo para análise de permanência discente.	52
Figura 8 – Fluxo de extração e integração de dados.	53
Figura 9 – Representação do modelo de sobrevivência.	58
Figura 10 – Interface inicial do ecossistema SABIA e acesso ao módulo de sobrevivência.	66
Figura 11 – Fluxograma da sequência das saídas analíticas na interface.	67
Figura 12 – Painel de filtros gerais para segmentação da amostra institucional.	68
Figura 13 – Seletor de evento para definição do desfecho da análise.	69
Figura 14 – Opções de visualização de intervalos e marcadores de censura.	69
Figura 15 – Controle de duração para delimitação do horizonte temporal.	69
Figura 16 – Interface de seleção da variável de agrupamento.	69
Figura 17 – Filtro secundário para seleção de grupos específicos.	69
Figura 18 – Seletor para exibição da curva de referência institucional.	70
Figura 19 – Tabela de análise descritiva gerada pela ferramenta.	71
Figura 20 – Visualização de sobrevivência, risco acumulado e probabilidades de desfecho.	72
Figura 21 – Indicadores de validação estatística e magnitude de efeito.	73
Figura 22 – Análise de sobrevivência e evasão do curso de LC ao longo do tempo.	80
Figura 23 – Análise de sobrevivência e evasão do curso de BSI ao longo do tempo.	83
Figura 24 – Análise de sobrevivência na instituição, por turno.	88
Figura 25 – Análise de sobrevivência na instituição, por gênero.	90
Figura 26 – Análise de sobrevivência na instituição, por assistência estudantil.	92
Figura 27 – Comparativo de sobrevivência por gênero segmentado por assistência estudantil.	93
Figura 28 – Análise de sobrevivência nos cursos de STEM, por turno.	102

Figura 29 – Análise de sobrevivência nos cursos de STEM, por gênero.	105
Figura 30 – Análise de sobrevivência nos cursos de STEM, por assistência estudantil.	106
Figura 31 – Comparativo de sobrevivência em STEM por gênero segmentado por assistência estudantil.	108

Lista de tabelas

Tabela 1 – Classificação de magnitude para diferentes métricas de tamanho de efeito	37
Tabela 2 – Revisão detalhada dos trabalhos relacionados.	39
Tabela 3 – Variáveis extraídas do arquivo <i>vinculo_periodo.parquet</i> .	54
Tabela 4 – Atributos selecionados para o modelo final.	55
Tabela 5 – Funções do módulo de processamento estatístico.	61
Tabela 6 – Funções do módulo de suporte e métricas estatísticas.	62
Tabela 7 – Detalhamento dos filtros gerais de entrada da ferramenta.	67
Tabela 8 – Exemplos de agrupamentos customizados via consultas Python.	70
Tabela 9 – Métricas de magnitude de efeito disponíveis na ferramenta.	74
Tabela 10 – Estatísticas descritivas e indicadores de sobrevivência por curso e vigência de PPC.	78
Tabela 11 – Probabilidade de sobrevivência e intervalo de Confiança (IC 95%) para o curso de LC por vigência de PPC.	81
Tabela 12 – Indicadores estatísticos para o curso de LC.	82
Tabela 13 – Probabilidade de sobrevivência e intervalo de Confiança (IC 95%) para o curso de BSI por vigência de PPC.	84
Tabela 14 – Indicadores estatísticos para o curso de BSI.	85
Tabela 15 – Caracterização sociodemográfica institucional (2014-2024)	87
Tabela 16 – Probabilidade de sobrevivência e intervalo de Confiança (IC 95%) para o cenário institucional por turno.	89
Tabela 17 – Probabilidade de sobrevivência e intervalo de confiança (IC 95%) por gênero no cenário institucional.	90
Tabela 18 – Probabilidade de sobrevivência e intervalo de Confiança (IC 95%) por Assistência no cenário institucional.	92
Tabela 19 – Análise longitudinal comparativa: probabilidades de sobrevivência por gênero e assistência (0-10)	94
Tabela 20 – Teste de hipótese <i>Log-Rank</i> para os Grupos Gerais e Combinações.	96
Tabela 21 – <i>Hazard Ratio</i> e coeficientes de risco para as combinações gerais.	98
Tabela 22 – Hedges'g para as comparações entre grupos gerais e combinações.	99
Tabela 23 – Indicadores de vínculo por curso em STEM.	101

Tabela 24 – Probabilidade de sobrevivência e intervalo de Confiança (IC 95%) para o cenário de STEM por turno.	103
Tabela 25 – Probabilidade de sobrevivência e intervalo de Confiança (IC 95%) para o cenário de STEM por gênero.	105
Tabela 26 – Probabilidade de sobrevivência e intervalo de Confiança (IC 95%) para o cenário de STEM por assistência estudantil.	107
Tabela 27 – Análise longitudinal comparativa em STEM: probabilidades de sobrevivência por gênero e assistência (0-10)	108
Tabela 28 – Teste de hipótese <i>Log-Rank</i> para cursos de STEM.	110
Tabela 29 – <i>Hazard Ratio</i> e coeficientes de risco para cursos de STEM.	111
Tabela 30 – Hedges’ g para comparações entre grupos de STEM.	113

Lista de Siglas

AA	<i>Academic Analytics</i>
AS	Análise de Sobrevivência
BCC	Bacharelado em Ciência da Computação
BI	<i>Business Intelligence</i>
BSI	Bacharelado em Sistemas de Informação
CLES	<i>Common Language Effect Size</i>
CRA	Coeficiente de Rendimento Acumulado
DEINFO	Departamento de Estatística e Informática
HR	<i>Hazard Ratio</i>
IES	Instituição de Ensino Superior
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
IRA	Índice de Rendimento Acadêmico
KM	<i>Kaplan-Meier</i>
LA	<i>Learning Analytics</i>
LC	Licenciatura em Computação
LF	Licenciatura em Física
MAT	Licenciatura em Matemática
OECD	<i>Organisation for Economic Co-operation and Development</i>
PPC	Projeto Pedagógico do Curso
SABIA	<i>System of Academic Business Intelligence and Analytics</i>
STEM	<i>Science, Technology, Engineering, and Mathematics</i>
UFRPE	Universidade Federal Rural de Pernambuco

Sumário

1	Introdução	20
1.1	Motivação e justificativa	22
1.2	Objetivos	23
1.3	Estrutura do trabalho	24
2	Referencial teórico	26
2.1	Panorama da evasão no ensino superior	26
2.2	Análise de sobrevivência	29
2.3	Métricas de análise e interpretação estatística	33
2.3.1	Intervalo de confiança	33
2.3.2	Teste de hipótese Log-Rank	34
2.3.3	Hazard Ratio	35
2.3.4	Effect-size	36
3	Trabalhos relacionados	39
3.1	Síntese da literatura e lacunas	50
4	Método	52
4.1	Extração e preparação de dados	53
4.2	Construção do modelo	56
4.2.1	Análise de sobrevivência	56
4.2.2	Métricas estatísticas	62
4.2.3	Limitações	64
5	Ferramenta	66
5.0.1	Parametrização e controles de execução	67
5.0.2	Indicadores de permanência	70
5.0.3	Interpretabilidade e suporte à decisão	74
6	Resultados	76
6.1	Impacto da mudança do projeto pedagógico de curso (PPC)	77
6.1.1	Análise descritiva da mudança do projeto pedagógico de curso (PPC)	77
6.1.2	Análise do curso: licenciatura em computação (LC)	79
6.1.3	Análise do curso: bacharelado em sistemas de informação (BSI)	83
6.2	Análise comparativa de permanência: cenário institucional e cursos de STEM	86

6.2.1	Análise descritiva do cenário institucional	86
6.2.2	Caracterização do cenário institucional	87
6.2.3	Análise descritiva do cenário em cursos de STEM	100
6.2.4	Caracterização do cenário em cursos de STEM	102
7	Discussão dos resultados	114
7.1	PP1: Quais os padrões temporais de sobrevivência e como os períodos de vulnerabilidade divergem entre as trajetórias analisadas?	114
7.1.1	Reformulação pedagógica curricular	114
7.1.2	Cenário institucional e áreas de STEM	116
7.2	PP2: De que maneira a reformulação curricular e as variáveis de gênero, turno e assistência estudantil impactam as disparidades no risco de evasão e na probabilidade de permanência discente?	118
7.2.1	Efeitos nas trajetórias dos cursos frente a reformulação pedagógica curricular	118
7.2.2	Efeitos nas trajetórias STEM e institucionais	119
7.3	PP3: De que forma a combinação de indicadores de significância (<i>Log – Rank</i>), risco (<i>Hazard Ratio</i>) e magnitude (<i>Effect Size</i>) permite quantificar a influência de fatores acadêmicos?	121
7.3.1	Métricas aplicadas à análise das mudanças de perfil pedagógico curriculares	121
7.3.2	Métricas aplicadas às trajetórias STEM e institucionais	122
7.4	Gestão baseada em evidências	124
8	Considerações finais	125
	Referências	128

1 Introdução

A evasão no ensino superior caracteriza-se como um fenômeno multidimensional com repercussões econômicas e institucionais, resultando no desperdício de recursos públicos e na interrupção de trajetórias sociais (SILVA; SAMPAIO, 2022; BRITO et al., 2020). No cenário brasileiro, a descontinuidade acadêmica compromete a sustentabilidade das políticas de democratização do acesso e a oferta de profissionais qualificados para o mercado de trabalho (GISI, 2006). Ao interromper o vínculo discente, o processo de exclusão impacta o retorno do capital investido e a função social das Instituições de Ensino Superior (IES) (FILIPAK; PACHECO, 2017).

Observa-se que a incidência deste fenômeno é superior nas áreas de Ciência, Tecnologia, Engenharia e Matemática (STEM). Nestes cursos, mais de 60% das desistências ocorrem nos primeiros dois anos, o que caracteriza a existência de janelas de risco nos períodos iniciais (CHEN et al., 2018). O insucesso em disciplinas de base, como Cálculo I e Programação, funciona como um gatilho para o abandono precoce, apresentando a recorrência de reprovações sobre a maioria dos ingressantes e aumento da probabilidade de descontinuidade do curso (VASCONCELOS; ANDRADE, 2018). Adicionalmente, a evasão tardia associa-se a barreiras estruturais e à necessidade de conciliação entre trabalho e estudo, afetando o tempo de permanência no sistema (GARCIA; GOMES, 2022).

No âmbito da administração das instituições de ensino superior, sistemas de dados são integrados para fundamentar os processos decisórios em métricas quantificáveis, permitindo que a gestão universitária abandone critérios subjetivos e adote uma gestão baseada em evidências (MARQUES et al., 2023; OKIMOTO et al., 2025). As políticas de assistência estudantil e os mecanismos de reserva de vagas constituem fatores de permanência acadêmica (SACCARO et al., 2019). O recebimento de auxílios financeiros e sociais correlaciona-se com a conclusão de cursos em áreas de STEM, especificamente no gênero feminino (CAVALCANTI et al., 2025). A identificação desses padrões permite que a gestão institucional direcione intervenções aos grupos em maior situação de vulnerabilidade (MENOLLI; NETO, 2021; SORGATTO; CÁCERES, 2023).

O fenômeno da evasão é estruturado como um problema de classificação binária em modelos preditivos (AJOODHA et al., 2020). O processamento computacional delimita o desfecho entre as categorias (VITELLI et al., 2010). Todavia, a análise por modelos estáticos

desconsidera a dimensão temporal e a distribuição de probabilidade dos eventos (MAO et al., 2024; COLOSIMO; GIOLO, 2021). A ausência de profundidade temporal impede a identificação de períodos em que fatores sociodemográficos tornam-se determinantes para o abandono, dificultando a segmentação do risco ao longo da trajetória (ORTIZ-MARTÍNEZ et al., 2023; SANTOS et al., 2024; HOED, 2016).

A integração da Análise de Sobrevivência ao monitoramento longitudinal viabiliza a observação do evento de evasão como uma função do tempo, o que permite quantificar o risco de interrupção em momentos específicos da trajetória acadêmica (LEE; WANG, 2003). Tal abordagem utiliza o estimador de *Kaplan-Meier* para a obtenção das curvas de sobrevivência, de modo que o teste *Log-Rank* atesta a significância da distinção entre os grupos, sendo complementado pelo *Hazard Ratio* e pelo *Effect-size* para quantificar, respectivamente, o risco relativo e a magnitude das disparidades. Dessa forma, a análise supera a limitação de testes de hipóteses isolados ao fornecer a intensidade e a proporção do efeito observado na permanência discente.

Em vista desse contexto, este trabalho utiliza a Análise de Sobrevivência integrada ao *System of Academic Business Intelligence and Analytics* (SABIA) para a identificação de trajetórias em risco e para mensurar disparidades sociodemográficas (gênero e turno) e socioeconômicas (assistência estudantil) na evasão. O escopo desta análise abrange os cursos de Bacharelado em Sistemas de Informação (BSI), Ciência da Computação (BCC), Licenciatura em Computação (LC), Licenciatura em Matemática e Licenciatura em Física da Unidade Acadêmica de Recife (Sede/UFRPE), permitindo comparar o cenário institucional com o panorama de áreas STEM. Contudo, adotam-se os cursos de BSI e LC para a avaliação do impacto das reformas curriculares, devido à reformulação de seus Projetos Pedagógicos de Curso (PPCs). Para tanto, busca-se responder às seguintes perguntas de pesquisa:

PP1: Quais os padrões temporais de sobrevivência e como os períodos de vulnerabilidade divergem entre as trajetórias analisadas?

PP2: De que maneira a reformulação curricular e as variáveis de gênero, turno e assistência estudantil impactam as disparidades no risco de evasão e na probabilidade de permanência discente?

PP3: De que forma a combinação de indicadores de significância (*Log – Rank*), risco (*Hazard Ratio*) e magnitude (*Effect Size*) permite quantificar a influência de

fatores acadêmicos?

1.1 Motivação e justificativa

A democratização do acesso ao ensino superior no Brasil representa um avanço nas políticas de inclusão, transformando a universidade em um espaço de mobilidade social. Entretanto, a expansão das vagas não assegura, por si só, a equidade nos resultados, uma vez que o fenômeno da evasão interrompe trajetórias e potencializa disparidades históricas (SILVA; SAMPAIO, 2022; BRITO et al., 2020). A magnitude deste desafio é evidenciada pelo contraste estatístico: enquanto a média de abandono no primeiro ano de curso na OCDE é de 13%, no Brasil este índice atinge 25% dos ingressantes em bacharelados (OECD, 2025). A persistência desse fenômeno resulta em uma baixa taxa de êxito acadêmico no Brasil, em que apenas 39% dos estudantes que iniciaram o ciclo em 2014 conseguiram concluir o curso até 2023 (INEP, 2023). Este cenário de vulnerabilidade é intensificado em áreas de STEM, onde os índices de descontinuidade são historicamente superiores à média nacional (OECD, 2024). Enquanto a taxa de conclusão em três anos extras para o conjunto dos bacharelados é de 70%, nos campos de STEM esse índice cai para 68%. Para efeito de comparação, a taxa de sucesso no mesmo período atinge 80% em saúde e bem-estar e 71% em artes e humanidades. Essa menor taxa de êxito é frequentemente atribuída à exigência acadêmica, uma vez que a incompatibilidade entre as expectativas dos alunos e o conteúdo rigoroso, ou uma preparação inadequada no ensino secundário, aumentam significativamente o risco de abandono (OECD, 2025).

Ademais, os estudantes de STEM demonstram maior probabilidade de trocar de área de estudo ao longo da trajetória universitária. Apenas 58% dos ingressantes em STEM concluem o bacharelado no campo original, enquanto 9% mudam para outro campo e 2% alteram o nível educacional antes de obterem o diploma. Outro fator que reduz o incentivo para a graduação formal é a alta demanda do mercado de trabalho: em áreas como engenharia e tecnologia da informação (TIC), o aquecimento do setor leva muitos alunos a aceitarem ofertas de emprego atraentes antes de concluírem o ciclo acadêmico. Esse quadro é agravado por uma desigualdade de gênero estrutural, na qual as mulheres representam apenas 30% dos novos ingressantes em cursos de STEM na média internacional e apenas 16% dos graduados de bacharelado no Brasil. Tal disparidade é alimentada por

fatores culturais e institucionais, como sentimentos de isolamento em ambientes dominados por homens, que levam as mulheres a mudarem de programa ou abandonarem o nível de ensino com maior frequência do que seus pares masculinos (OECD, 2025).

Diante desse contexto, o reconhecimento de disparidades e dificuldades enfrentadas por grupos em condição de vulnerabilidade orienta ações institucionais. Enquanto modelos tradicionais ignoram a cronologia dos eventos, a transição para abordagens temporais longitudinais permite identificar lacunas de formação e a ausência de suporte em marcos críticos (KLITZKE; CARVALHAES, 2023). O emprego da Análise de Sobrevivência justifica-se pela sua capacidade de lidar com dados censurados e observar o risco como uma função do tempo, permitindo identificar o impacto real das intervenções institucionais e detectar disparidades socioeconômicas ou demográficas nas curvas de permanência (LEE; WANG, 2003; COLOSIMO; GIOLO, 2021).

A identificação de padrões e tendências temporais permite visualizar quais grupos estão mais propensos à evasão e em quais períodos a vulnerabilidade ocorre. Grupos ou cursos que, apesar de apresentarem taxas de evasão similares, podem registrar a incidência de risco em períodos distintos. Assim, analisando essa trajetória pode-se elaborar políticas específicas ao longo do tempo. Ao integrar esta modelagem computacional, os registros administrativos são convertidos em indicadores interpretáveis de risco e magnitude de efeito, oferecendo suporte para o fortalecimento de práticas institucionais orientadas por evidências (MARQUES et al., 2023). A quantificação do impacto da assistência estudantil como fator de proteção é essencial para validar e orientar a aplicação de recursos institucionais (CAVALCANTI et al., 2025). Assim, este trabalho justifica-se pelo desenvolvimento de ferramentas que permitam acompanhar a trajetória discente de forma contínua, garantindo que o suporte institucional seja direcionado para neutralizar as disparidades identificadas e assegurar que o sistema de ensino superior cumpra seu papel de inclusão e sucesso acadêmico para todos os ingressantes.

1.2 Objetivos

Este trabalho tem como objetivo consolidar um modelo computacional de análise de sobrevivência para a permanência acadêmica no ensino superior, integrando métodos estatísticos e mecanismos de comparação temporal entre grupos de estudantes, para

identificar trajetórias acadêmicas em risco e quantificar disparidades sociodemográficas e socioeconômicas no fenômeno da evasão.

Diante do exposto, este trabalho tem os seguintes objetivos específicos:

- Implementar um modelo computacional de análise de sobrevivência para monitorar a permanência acadêmica;
- Desenvolver um painel de alto nível, integrado ao SABIA, para facilitar a utilização dos modelos de análise de sobrevivência;
- Identificar os padrões temporais e os marcos críticos de vulnerabilidade nas trajetórias discentes, comparando a mudança de Projetos Pedagógicos de Curso (PPCs) sobre a continuidade acadêmica;
- Analisar a influência de fatores socioeconômicos e sociodemográficos, como turno, gênero e assistência estudantil, sobre a probabilidade de sobrevivência e o tempo até o evento de evasão;
- Quantificar a magnitude do impacto de variáveis acadêmicas por meio da integração de métricas de significância (*Log-Rank*), risco (*Hazard Ratio*) e tamanho de efeito (*Hedges' g*).

1.3 Estrutura do trabalho

Este trabalho organiza-se em oito capítulos, estruturados de forma a fundamentar a modelagem proposta e apresentar os resultados obtidos. No Capítulo 1, foram delineados o objeto de estudo, a motivação baseada em indicadores educacionais e as perguntas de pesquisa. O Capítulo 2 detalha o referencial teórico, abordando as dimensões da permanência acadêmica e os fundamentos matemáticos da análise de sobrevivência. O Capítulo 3 estabelece o estado da arte por meio da revisão de trabalhos relacionados, contrastando abordagens preditivas estáticas com modelos longitudinais.

O Capítulo 4 descreve o método e a formalização do modelo computacional, detalhando as etapas de implementação e processamento dos dados. Complementarmente, o Capítulo 5 apresenta as funcionalidades da ferramenta desenvolvida, detalhando sua integração à plataforma SABIA para a visualização dos indicadores de risco e suporte à gestão acadêmica. Os resultados são discutidos nos Capítulos 6 e 7, que tratam, respectivamente, da análise de trajetórias e reformas curriculares e da identificação de

disparidades sociodemográficas (gênero e turno) e o impacto da assistência estudantil como fator socioeconômico. Por fim, o Capítulo 8 apresenta as considerações finais sobre os resultados obtidos com o modelo, a síntese das conclusões e as proposições de trabalhos futuros para a análise de diferentes perfis institucionais.

2 Referencial teórico

Este capítulo apresenta o embasamento teórico para construção do modelo. Inicialmente, realiza-se a contextualização da evasão no ensino superior, observando-se as trajetórias discentes com base em indicadores nacionais e internacionais. Em seguida, detalham-se os fundamentos da Análise de Sobrevivência, método estatístico para a modelagem do tempo até a ocorrência da evasão e identificação de fatores de risco. Por fim, descrevem-se as métricas estatísticas utilizadas na validação dos resultados e na comparação entre os grupos analisados.

2.1 Panorama da evasão no ensino superior

A evasão no ensino superior caracteriza-se como um fenômeno com implicações na gestão de recursos, no desenvolvimento social e no retorno do investimento público (HOED, 2016). Este evento impacta a eficiência das instituições e a formação de capital humano, gerando reflexos nos âmbitos econômico, social e institucional. A evasão ocorre em três níveis: a de curso, pelo desligamento da matriz curricular; a da instituição, pela saída da unidade de ensino; e a do sistema, que representa o abandono da educação superior (ANDIFES, 1996).

No cenário brasileiro, 25% dos ingressantes interrompem o vínculo após o primeiro ano, taxa superior à média internacional de 13% (OECD, 2025). A taxa de conclusão decorridos três anos do prazo regular é de 49%, enquanto a média internacional atinge 70%. Para analisar este fluxo é estabelecido o ciclo de matrícula (Ministério da Educação - Secretaria de Educação Profissional e Tecnológica, 2014), conforme ilustrado na Figura 1, que identifica o estado do discente por meio dos conceitos de permanência, desistência e conclusão.

A análise das trajetórias utiliza as situações registradas pelo Censo da Educação Superior anualmente (INEP, 2024). O estado *Concluído* compreende o discente que integralizou os créditos previstos, independentemente da participação em cerimônia de colação de grau ou no Exame Nacional de Desempenho de Estudantes (Enade). A categoria *Cursando* identifica o vínculo ativo sem a conclusão da carga horária no ano de referência. Interrupções temporárias constam como matrícula *Trancada*, enquanto a desvinculação

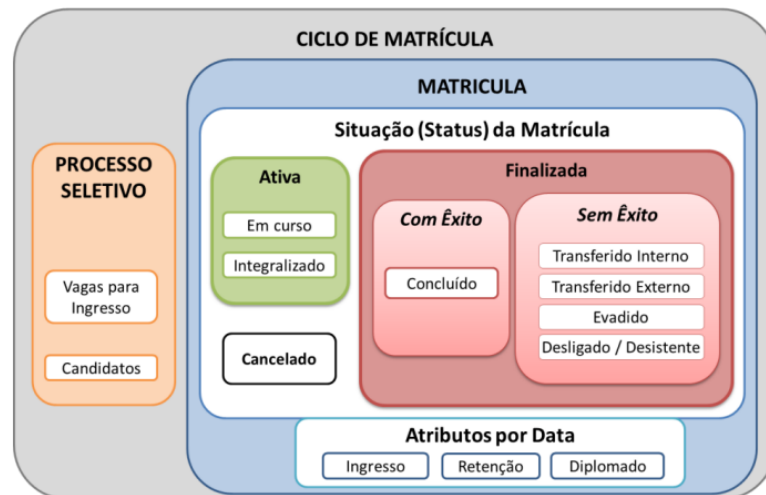


Figura 1 – Representação do ciclo de matrícula.

Fonte: [Ministério da Educação - Secretaria de Educação Profissional e Tecnológica \(2014\)](#).

definitiva ocorre por abandono, desligamento ou transferência. Estes indicadores permitem diferenciar a permanência da retenção, definida pela não progressão na velocidade prevista devido a reprovações ou trancamentos ([INEP, 2017](#)).

A gestão orçamentária e a alocação de recursos apresentam variações conforme os níveis de qualificação acadêmica. O investimento anual por estudante no Brasil é de US\$ 3.765, valor inferior à média da OCDE de US\$ 15.102 ([OECD, 2025](#)). Embora o financiamento público em educação corresponda a 4,3% do PIB, a distribuição de recursos por discente no ensino superior é equivalente à destinada à educação básica. A Figura 2 reflete a divergência deste padrão em relação à média internacional, voltada à priorização do nível terciário.

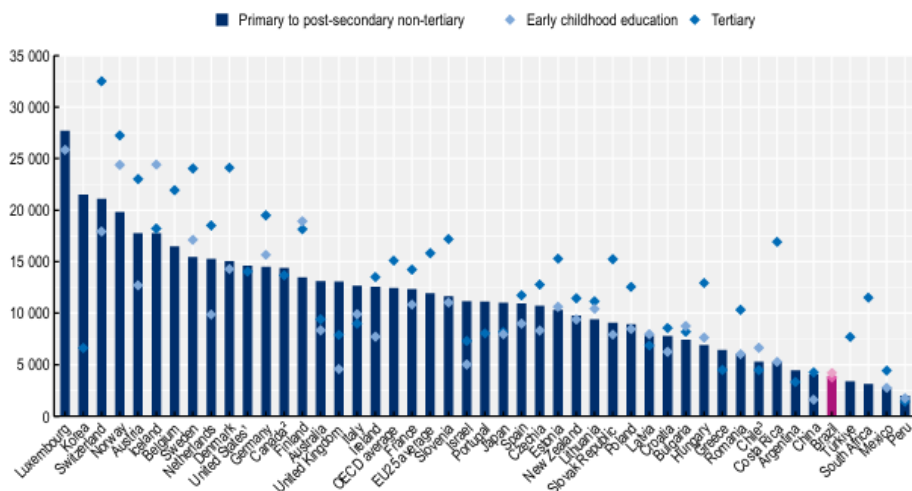


Figura 2 – Despesa governamental por aluno por nível de ensino (2022).

Fonte: [OECD \(2025\)](#).

Quanto ao perfil das matrículas, os bacharelados concentram 54,7% dos ingressantes, enquanto as licenciaturas respondem por 16,2%, apresentando 74,8% de participação feminina (INEP, 2024). A evolução desta distribuição por grau acadêmico é apresentada na Figura 3.

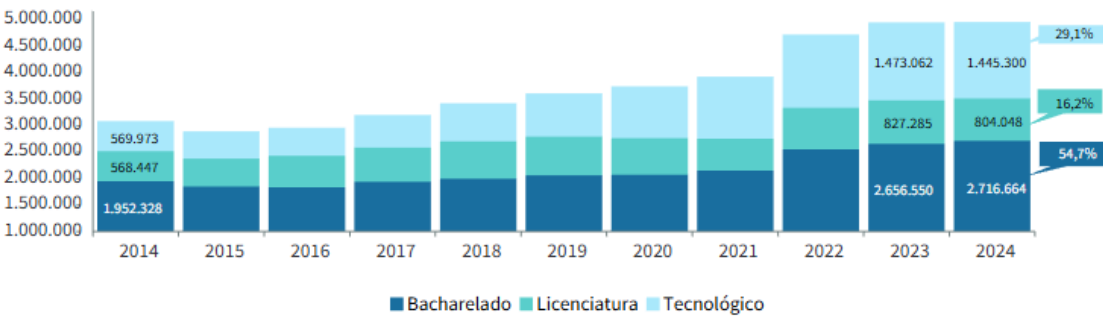


Figura 3 – Número de matrículas no ensino superior por grau acadêmico (2014-2024).
Fonte: INEP (2024).

No recorte de gênero, 53% das mulheres concluem a graduação no prazo previsto, contra 43% dos homens (OECD, 2025). A Figura 4 detalha a participação por sexo em diferentes categorias administrativas. Nas áreas STEM, o percentual de graduados é de 16%, valor inferior à média de 23% da organização.

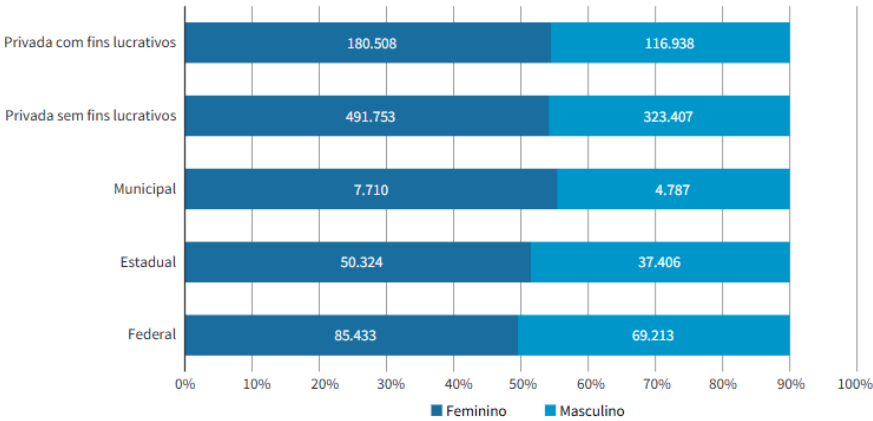


Figura 4 – Concluintes no ensino superior por gênero e categoria administrativa.
Fonte: INEP (2024).

O fluxo de longo prazo, entre 2014 a 2023 reforça este cenário, com a taxa de permanência declinando de 86% no primeiro ano para 1% ao final de dez anos, enquanto a desistência acumulada atinge 60% (INEP, 2023). Este comportamento da trajetória é ilustrado na Figura 5.

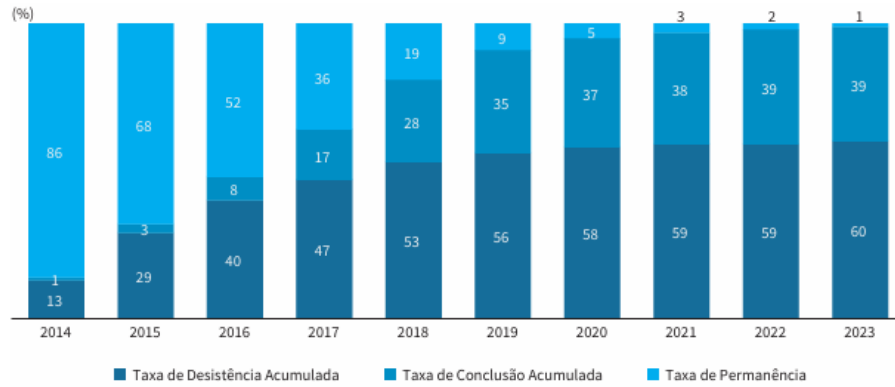


Figura 5 – Evolução média dos indicadores de trajetória dos ingressantes (2014-2023).
Fonte: INEP (2023).

Em síntese, a evasão no ensino superior manifesta-se como um fenômeno multidimensional. A configuração orçamentária, as assimetrias de gênero e as especificidades das áreas de conhecimento definem distintos perfis de permanência. Esses elementos contextuais demonstram que o fluxo acadêmico é suscetível a variáveis externas e institucionais, cuja complexidade demanda uma visão integrada das trajetórias discentes para além de indicadores transversais.

Nesse contexto, os conceitos de fluxo servem para mensurar o tempo que o estudante leva até a transição para um estado definitivo. No entanto, a gestão acadêmica demanda a identificação das variáveis associadas ao abandono e dos momentos de maior probabilidade de desistência. Portanto, a problemática deste trabalho reside na aplicação da Análise de Sobrevivência para detectar períodos críticos de risco, fornecendo evidências para o suporte institucional e base para a proposição de políticas fundamentadas.

2.2 Análise de sobrevivência

A análise de sobrevivência compreende um conjunto de procedimentos estatísticos voltados à modelagem e análise de dados em que a variável resposta é o tempo T até a ocorrência de um evento de interesse (COLOSIMO; GIOLO, 2021). A função de sobrevivência, denotada por $S(t)$, é definida como a probabilidade de um indivíduo sobreviver além de um tempo t específico, conforme a Equação 2.1.

$$S(t) = P(T > t) \quad (2.1)$$

Esta técnica baseia-se em cinco elementos distintos. O evento de interesse define o desfecho (falha). O tempo até o evento constitui a variável resposta T , que mensura o intervalo entre a origem e a falha. A censura representa os dados parciais onde o valor exato de T é desconhecido. A escala temporal define as unidades de medida para t . Por fim, a função $S(t)$ indica a probabilidade de o tempo de falha ser superior a um valor t determinado (LEE; WANG, 2003).

Destoando das técnicas de regressão convencional, esta abordagem permite lidar com dados incompletos e com a assimetria característica dos tempos de falha. As estratégias para a modelagem desses dados são classificadas de acordo com as suposições sobre a distribuição dos tempos e o efeito das covariáveis (HABERMAN, 1981). Os métodos não paramétricos não assumem uma distribuição de probabilidade teórica para os tempos de sobrevivência, sendo aplicados na estimação de curvas e na comparação exploratória entre grupos. Inversamente, os métodos paramétricos assumem que os tempos seguem distribuições específicas, como Weibull, Exponencial ou Log-normal, o que possibilita realizar estimativas baseadas em parâmetros fixos (LAWLESS, 2011).

A escolha das distribuições é realizada de acordo com a forma da função de risco e a natureza do fenômeno observado (RICH et al., 2010). Adicionalmente, os métodos semiparamétricos, representados pelo modelo de riscos proporcionais de Cox, não exigem o conhecimento da distribuição base do risco, mas estabelecem uma forma paramétrica para descrever o efeito das covariáveis no desfecho (COX, 2018).

A censura caracteriza-se como a situação em que se possui informação parcial sobre o tempo de sobrevivência, mas o instante exato da falha é desconhecido (COLOSIMO; GIOLO, 2021). A inclusão analítica desses dados é necessária para a validade das inferências, pois a exclusão das observações incompletas ou o seu tratamento como falhas introduziria viés, tendendo a subestimar o tempo médio de vida da população. O tratamento adequado da censura possibilita o aproveitamento total da informação disponível, utilizando o tempo de exposição de cada indivíduo até a perda de seguimento para gerar estimativas (KLEINBAUM et al., 2012). Conforme ilustrado na Figura 6, a análise exige a conversão da medida temporal para o tempo de estudo, em que a origem $t = 0$ marca o ingresso de cada indivíduo no acompanhamento.

Sob essa escala, a censura à direita ocorre quando o tempo real de sobrevivência (T) é conhecido apenas como sendo maior ou igual ao tempo de observação registrado

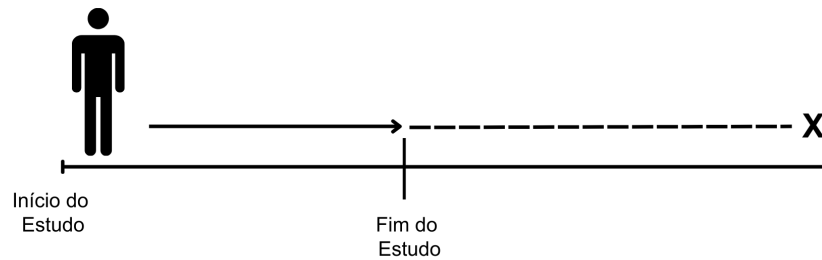


Figura 6 – Representação da censura sob a escala de tempo de estudo.

Fonte: Adaptado de [Kleinbaum et al. \(2012\)](#).

(t), ou seja, $T \geq t$. Este fenômeno manifesta-se pelo encerramento do estudo antes da ocorrência do evento, perda de acompanhamento ou saída do indivíduo por causas externas ([COLOSIMO; GIOLO, 2021](#)). Sendo subdividida em Tipo I, quando o tempo de encerramento do estudo é fixado previamente; Tipo II, se o monitoramento encerra após um número determinado de eventos; ou Aleatória, em casos de perda de seguimento ou retirada do sujeito por razões externas. Por outro lado, a censura à esquerda é registrada quando o evento ocorre em momento anterior ao início da coleta de dados, sendo conhecido apenas que $T \leq t$ ([LEE; WANG, 2003](#)). Adicionalmente, verifica-se a censura intervalar quando o desfecho é registrado apenas dentro de um intervalo de tempo específico, sem a precisão do instante exato da ocorrência ([LAWLESS, 2011](#)).

Diferente da censura, o truncamento é uma condição de seleção que exclui sistematicamente indivíduos da amostra por não atenderem a um critério de entrada. O truncamento à esquerda ocorre quando o sujeito apenas compõe a base de dados se sobreviver livre do evento até um tempo t_0 . Caso a falha ocorra antes de t_0 , o indivíduo torna-se invisível ao pesquisador e não é contabilizado. Já o truncamento à direita ocorre quando apenas indivíduos que já experimentaram o evento até um tempo fixado são incluídos na amostra, excluindo-se aqueles que ainda não falharam. ([MILLS, 2010](#)).

Neste trabalho, a opção pelo tratamento das observações como censura em vez de truncamento justifica-se pela necessidade de preservar a representatividade da amostra e evitar erros de subestimação. A exclusão de indivíduos censurados via truncamento

descartaria a informação de que esses discentes sobreviveram ao menos até o tempo t , resultando em estimativas de sobrevivência viciadas e na perda de dados sobre o tempo de exposição ao risco (MILLS, 2010).

Dentre os métodos não paramétricos, o Estimador de *Kaplan-Meier* (KM), ou estimador de limite-produto, é a técnica utilizada para estimar a função de sobrevivência $S(t)$ na presença de observações censuradas. A vantagem desta abordagem reside em sua natureza livre de distribuição, dispensando pressupostos sobre a forma funcional dos dados (KAPLAN; MEIER, 1958). O cálculo baseia-se no produto das probabilidades condicionais de sobreviver a intervalos de tempo cujos limites são definidos pelos instantes de falha observados, conforme a Equação 2.2.

$$\hat{S}(t) = \prod_{j:t_j < t} \left(1 - \frac{d_j}{n_j}\right) \quad (2.2)$$

Em que:

- d_j : representa o número de falhas observadas no tempo t_j ;
- n_j : indica o número de indivíduos sob risco imediatamente antes do instante t_j .

Matematicamente, a estimativa resulta em uma função escada com decréscimos nos tempos de ocorrência do evento (BLAND; ALTMAN, 1998). Para quantificar a variabilidade da curva e construir intervalos de confiança de 95%, utiliza-se a Fórmula de Greenwood presente na Equação 2.3, que estima a variância da função de sobrevivência (COLOSIMO; GIOLO, 2021).

$$\widehat{Var}[\hat{S}(t)] = [\hat{S}(t)]^2 \sum_{j:t_j < t} \frac{d_j}{n_j(n_j - d_j)} \quad (2.3)$$

Na área educacional, a função de sobrevivência é aplicada para analisar o tempo até a conclusão ou evasão do curso, considerando fatores como desistência ou transferência (HOED, 2016). Para os propósitos deste trabalho, adota-se a abordagem de censura à direita. Os estudantes são classificados como censurados quando o tempo de acompanhamento termina antes da falha, seja por permanência no curso ou encerramento do período de coleta. No contexto do desenvolvimento do modelo, o limite para o tempo de estudo é fixado entre dez e vinte semestres.

Este estudo utiliza o estimador não paramétrico de *Kaplan-Meier* como indicador para mensurar a dinâmica temporal da trajetória estudantil e identificar padrões de evasão. A análise permite avaliar o impacto de variáveis acadêmicas na probabilidade de permanência e determinar pontos de inflexão nos tempos de abandono. No contexto da evasão estudantil, o método estima a proporção de estudantes matriculados após intervalos específicos, incorporando as informações dos alunos que não atingiram o desfecho até o término da coleta de dados.

2.3 Métricas de análise e interpretação estatística

Esta seção fundamenta as métricas utilizadas para a interpretação dos resultados, estabelecendo os critérios de precisão, significância e magnitude dos efeitos no evento observado.

2.3.1 Intervalo de confiança

O Intervalo de Confiança (IC) define-se como um intervalo estimado para um parâmetro de interesse de uma determinada população. Em vez de representar o parâmetro por um único valor pontual, esta métrica estabelece um intervalo de estimativas prováveis para o valor real na população de origem ([CUMMING, 2013](#)).

O IC permite a inferência de significância estatística conforme o nível de confiança adotado, sendo o patamar de 95% ($\alpha = 0,05$) o padrão convencional na literatura científica. Embora esse nível seja amplamente aceito, o coeficiente de confiança pode ser definido em outros patamares, como 90% ou 99%, dependendo da precisão requerida e das potenciais consequências de um erro do tipo I no contexto do estudo. Sob o nível de confiança selecionado, a associação é validada quando o intervalo para medidas de diferença exclui o valor zero ([PATINO; FERREIRA, 2015](#)).

O cálculo fundamenta-se na estatística observada e na variabilidade amostral, conforme a Equação 2.4. Nesta equação, $\bar{X}$ representa a média amostral, 1,96 corresponde ao valor crítico para a distribuição normal com 95% de confiança, S é o desvio-padrão e n o tamanho da amostra ([OLIVE, 2005](#)).

$$IC = \bar{X} \pm \left(1,96 \times \frac{S}{\sqrt{n}} \right) \quad (2.4)$$

A precisão da estimativa correlaciona-se inversamente ao tamanho da amostra, dado que o incremento no número de observações resulta na redução do erro-padrão e da amplitude do intervalo de confiança (IC). Na análise de sobrevivência, o IC quantifica a incerteza da curva de *Kaplan-Meier* por meio da variância da probabilidade de sobrevivência obtida pela Fórmula de Greenwood (LEE; WANG, 2003). Este cálculo processa a relação entre o número de falhas e o total de indivíduos sob risco em cada intervalo de tempo. A aplicação do IC permite a inferência de significância estatística ao nível de 5%, com a validação da associação quando o intervalo para medidas de diferença exclui o valor zero ou quando o intervalo para razões, como o Hazard Ratio (HR), exclui o valor unitário (KLEINBAUM et al., 2012).

2.3.2 Teste de hipótese Log-Rank

O teste Log-Rank é um procedimento não paramétrico utilizado para a comparação de funções de sobrevivência entre dois ou mais grupos independentes (BLAND; ALTMAN, 1998). A hipótese nula (H_0) estabelece a igualdade entre as curvas de sobrevivência das populações, enquanto a hipótese alternativa (H_1) indica a existência de diferença estatística em ao menos um ponto do período de observação (WELLEK, 1993). O p -valor constitui a probabilidade de obter uma estatística de teste igual ou mais extrema que a observada, sob a premissa de que a hipótese nula é verdadeira. O método baseia-se na comparação entre o número de eventos observados e o número de eventos esperado sob a validade de H_0 em cada intervalo de tempo com ocorrência de falhas.

Para a comparação entre grupos, a estatística do teste (χ^2) é definida conforme a Equação 2.5.

$$\chi^2 = \frac{(O_g - E_g)^2}{Var(O_g - E_g)} \quad (2.5)$$

Em que:

- O_g : representa o total de eventos observados no grupo g ;
- E_g : representa o total de eventos esperados para o grupo g .

Esta estatística segue uma distribuição qui-quadrado com $G - 1$ graus de liberdade, sendo G o número de grupos comparados. A rejeição de H_0 ocorre quando o p-valor é inferior ao nível de significância de 5%, indicando disparidade entre as curvas de sobrevivência (CUZICK, 1982). O teste Log-Rank constitui a validação inferencial complementar às estimativas visuais obtidas pelo método de *Kaplan-Meier* (ELANDT-JOHNSON; JOHNSON, 1980).

No contexto deste trabalho, o teste é aplicado para comparar as curvas de sobrevivência entre as diferentes categorias analisadas, justificando-se por ser o procedimento não paramétrico. A escolha fundamenta-se na capacidade de utilizar todo o período de acompanhamento disponível, ao contrário de testes que comparam apenas pontos fixos no tempo, garantindo que as informações de toda a trajetória dos indivíduos e as observações censuradas sejam integralmente aproveitadas na análise. O teste não exige pressuposições sobre a forma da distribuição dos tempos de falha, fornecendo a evidência quantitativa para confirmar se as diferenças observadas são estatisticamente significantes.

2.3.3 Hazard Ratio

A *Hazard Ratio* (HR), ou razão de riscos, constitui a medida de efeito para a análise de sobrevivência (STARE; MAUCORT-BOULCH, 2016). Ele é definido pela razão entre as funções de taxa de falha de grupos independentes, em que a função de risco $h(t)$ representa o potencial instantâneo por unidade de tempo para a ocorrência do evento, sob a condição de sobrevivência até o tempo t .

A interpretação do HR baseia-se no valor unitário. O resultado $HR = 1$ indica equivalência de risco entre os grupos. Valores de $HR > 1$ indicam incremento na taxa de falha do grupo exposto em relação ao grupo de referência, enquanto $HR < 1$ indica menor taxa de ocorrência do evento (KLEINBAUM et al., 2012).

A fundamentação matemática do HR provém do Modelo de Riscos Proporcionais de Cox, conforme a Equação 2.6. Onde $h_0(t)$ representa a função de risco de base (*baseline hazard*) (COX, 2018). A estimativa do HR para a comparação entre preditores (X^* e X) é obtida por $HR = \exp[\sum \beta_i(X_i^* - X_i)]$. Para variáveis dicotômicas, o cálculo resulta em $HR = e^\beta$, sendo o coeficiente β estimado por máxima verossimilhança parcial (LEE; WANG, 2003).

$$h(t, X) = h_0(t) \exp \left(\sum_{i=1}^n \beta_i X_i \right) \quad (2.6)$$

A validade do HR pressupõe a proporcionalidade dos riscos, o que implica uma razão constante ao longo do período de observação. Caso as funções de risco se cruzem, a suposição de proporcionalidade é invalidada, demandando modelos com covariáveis dependentes do tempo (JR et al., 2008). Sob a premissa de riscos proporcionais, o HR sintetiza a força de associação entre as variáveis explicativas e o tempo até o desfecho; assim, enquanto a função de sobrevivência $S(t)$ foca na probabilidade de permanecer vivo, o hazard foca na taxa de falha e no potencial instantâneo do evento ocorrer naquele exato momento.

No âmbito deste trabalho, o Hazard Ratio é empregado para realizar análises multivariadas via modelo de Cox, possibilitando estimar o risco relativo de um grupo em relação a outro enquanto se controla simultaneamente a influência de potenciais fatores de confusão. Essa abordagem semiparamétrica permite obter inferências mais precisas sobre o impacto real de cada fator na sobrevivência, quantificando a força da associação entre as variáveis analisadas e o tempo até a ocorrência da evasão.

2.3.4 Effect-size

A magnitude das disparidades entre grupos ou a força de associação entre variáveis é quantificada pelo cálculo do tamanho do efeito (*Effect Size*). Diferente da significância estatística (p -valor), que mensura a probabilidade de os dados observados ocorrerem sob a hipótese nula e apresenta dependência do tamanho amostral, o tamanho do efeito constitui uma métrica direta da magnitude da diferença ou relação (SULLIVAN; FEINN, 2012). Esta métrica expressa o impacto observado em unidades de desvio-padrão, o que permite a comparação de resultados independentemente da escala original das variáveis.

O uso exclusivo de testes de significância de hipótese nula apresenta limitações, visto que em amostras de grande escala, efeitos de baixa relevância prática podem atingir significância estatística; em amostras reduzidas, efeitos práticos podem não ser detectados por insuficiência de poder estatístico (AARTS et al., 2014). O tamanho do efeito é utilizado em conjunto com intervalos de confiança para estabelecer a precisão da estimativa e a

relevância prática, sem restringir-se à lógica binária de aceitação ou rejeição da hipótese nula (CUMMING, 2013).

As medidas de tamanho do efeito são classificadas em não padronizadas e padronizadas. Na comparação entre médias, utiliza-se a diferença padronizada de médias, sendo o d de *Cohen* um dos estimadores aplicados. Este indicador apresenta viés positivo em estudos com amostras reduzidas (WANG et al., 2023). Para a correção deste viés, aplica-se o g de *Hedges*, que ajusta a diferença padronizada de médias para torná-la um estimador não enviesado (LAKENS, 2013), conforme a Equação 2.7.

$$g = d \times \left(1 - \frac{3}{4(n_1 + n_2) - 9} \right) \quad (2.7)$$

em que n_1 e n_2 representam os tamanhos amostrais dos grupos comparados. O fator de correção converge para 1 à medida que o tamanho da amostra aumenta, o que equaliza as métricas em amostras maiores.

A interpretação da magnitude baseia-se em limiares técnicos condicionados à natureza da métrica e ao delineamento experimental, conforme consolidado na Tabela 1.

Tabela 1 – Classificação de magnitude para diferentes métricas de tamanho de efeito

Métrica	Pequeno	Médio	Grande	Muito Grande
d de Cohen / g de Hedges	0,20	0,50	0,80	$\geq 1,20$
r de Pearson	0,10	0,30	0,50	$\geq 0,70$
Eta-quadrado (η^2)	0,01	0,06	0,14	–
Odds Ratio (OR)	1,50	3,00	10,00	–
AUC (ROC)	0,60	0,70	0,80	–
CLES	0,56	0,64	0,71	–

Nota: Valores adaptados de *Cohen* (2013) e *Lakens* (2013).

As métricas de variância (η^2) e associação (r) possuem escalas próprias, enquanto o *Common Language Effect Size* (CLES) fornece uma interpretação probabilística da superioridade entre indivíduos de grupos distintos (LAKENS, 2013). A seleção da métrica e dos limiares fundamenta a determinação da relevância dos dados observados e viabiliza a realização de metanálises (WANG et al., 2023).

Em termos operacionais, a aplicação do tamanho do efeito ocorre pela comparação entre uma categoria de referência, estabelecida como linha de base, e categorias alternativas. No contexto desta pesquisa, os grupos de referência são definidos pelos perfis de menor exposição teórica ao risco de evasão, como estudantes do sexo masculino o período anterior à alteração do PPC. Esta métrica padroniza a disparidade entre as categorias, permitindo mensurar o distanciamento das trajetórias dos grupos alternativos em relação à média da categoria de referência em unidades comparáveis.

A interpretação dos resultados via *Hedges' g* baseia-se em níveis de magnitude que classificam o efeito como pequeno (0,20), médio (0,50) ou grande (0,80). Esses limiares funcionam como guias para avaliar a intensidade da associação dos achados, distinguindo a significância estatística da magnitude do efeito observada.

3 Trabalhos relacionados

Esta seção apresenta uma revisão de estudos que investigam a evasão estudantil sob diferentes perspectivas metodológicas e temáticas, agrupando-os conforme as técnicas de modelagem e contextos acadêmicos. A literatura caracteriza a evasão como um fenômeno multifacetado, onde a permanência não é apenas a ausência do abandono, mas o resultado de um equilíbrio entre suporte institucional, desempenho acadêmico e integração social. Avanços metodológicos recentes propõem não apenas identificar “quem” evade, mas obter o momento da ocorrência para subsidiar políticas de permanência efetivas. A Tabela 2 [exibe](#) um resumo dos trabalhos relacionados citados.

Tabela 2 – Revisão detalhada dos trabalhos relacionados.

Trabalho	Ano	Objetivo	Método	Resultado
Hoed	2016	Analisar evasão em Computação.	Cox e Kaplan-Meier	Reprovações elevam o risco inicial.
Saccaro et al.	2019	Investigar fatores em STEM.	Kaplan-Meier	Auxílios reduzem a evasão.
Gutierrez-Pachas et al.	2022	Avaliar mudança curricular.	Sobrevivência e ML	Reestruturação elevou permanência.
Silva	2024	Impacto de nova matriz.	Sobrevivência	Reprovações reduzem a permanência.
Vitelli et al.	2010	Variáveis determinantes.	Regressão Logística	Acurácia 91,5%; auxílio é vital.
Dey e Astin	1993	Comparar modelos estatísticos.	Logit e Probit	Modelos similares em grandes bases.
Tsiakmaki et al.	2018	Prever notas do 2º semestre.	Regressão Linear	Notas do 1º sem. são preditoras.

Trabalho	Ano	Objetivo	Método	Resultado
Ajoodha et al.	2020	Classificar perfis de risco.	ML (6 algoritmos)	RF (82%) superou Logística.
Marques et al.	2023	Transformar dados em BI.	BI e Analytics	Identificação de tendências.
Nascimento et al.	2022	Revisão de PPC.	Análise Curricular	Redução de carga em exatas.
Vasconcelos e Andrade	2018	Base lógica/matemática.	Quanti-qualitativo	Defasagem gera retenção de 50%.
Falcão et al.	2018	Redução de choque acadêmico.	Reestruturação	Desmembramento de disciplinas.
Brito et al.	2020	Mapear atributos no Brasil.	Random Forest	Acurácia de 78% em Engenharias.
Barbosa et al.	2024	Explicabilidade (XAI).	RF e SHAP	gênero e raça para identificar vieses algorítmicos.
Duran et al.	2023	Permanência feminina.	Mann-Whitney	Ambiente gera autoexclusão.
Ortiz-Martínez et al.	2023	Empatia docente e STEM.	Qualitativo	Estereótipos fragilizam trajetória.
Pascoal et al.	2016	Monitoramento contínuo.	Naive Bayes	Instrução materna impacta evasão.
Pinheiro et al.	2018	Dados dinâmicos vs estáticos.	SVM e Tree	Dados dinâmicos elevam acurácia.
Carminati et al.	2020	Extração de regras.	CRISP-DM / J48	Mediana supera média como preditor.

Trabalho	Ano	Objetivo	Método	Resultado
Na et al.	2025	Aprendizado Profundo.	RNN (LSTM/GRU)	F1 Score de 0,9416; superior.
Lima e Zago	2023	Interpretabilidade em ML.	Revisão Crítica	Alerta para modelos “caixa preta”.
Mao et al.	2024	Explicabilidade no tempo.	SHAP e LIME	Necessidade de IA explicável.
Menolli et al.	2021	Grandes volumes (INEP).	BI e DW	Evasão maior em privadas (26%).
Sorgatto e Cáceres	2023	Perfil de ingressantes.	Learning Anal.	Crescimento de tecnólogos.
Gisi	2006	Desigualdade no acesso.	Análise Social	Expansão sem reduzir disparidades.
Oliveira et al.	2019	Capital cultural.	Análise Fatorial	Escolaridade materna é crucial.
Filipak e Pacheco	2017	Democratização vs Conclusão.	Teórico/Análise	Suporte institucional indispensável.
Silva e Sampaio	2022	Revisão de coorte (10 anos).	Revisão Sist.	Apenas 40% concluem graduação.
Silva et al.	2016	Reprovação em Cálculo.	Estatística	64% de reprovação no 1º ano.
Moura et al.	2020	Ciclo na educação básica.	Sistêmico	Evasão gera prejuízo social.
Almeida Schimiguel	2011	Ociosidade de vagas.	Avaliação	23,7% de ociosidade em Física.
Chen et al.	2018	Dinâmica temporal.	Cox / Survival	risco maior no 1º ano.

Trabalho	Ano	Objetivo	Método	Resultado
Zamith et al.	2025	Autonomia discente.	Modelo AFT	IRA alto eleva sobrevivência.
Cavalcanti et al.	2024	Mudança de PPC.	Kaplan-Meier	Aumento de 40% na sobrevivência.
Gutierrez-Pachas et al.	2023	Escala Longitudinal.	DeepSurv / CNN	Desempenho é 29% da relevância.
Garcia e Gomes	2022	Causas em Exatas.	Revisão	Dificuldade em 53% dos casos.
Sousa et al.	2010.	Perfis de cotistas.	Log-Normal	Necessidade de monitoria.
Junior et al.	2012	Diferença de gênero.	Sobrevivência	Mulheres resistem mais à retenção.
Okimoto et al.	2025	Suporte via Pesquisa.	Teste Hipótese	Projeto SUPER melhora rendimento.
Santos et al.	2024	Simulação de progresso.	Markov	Retenção altera conclusão final.
Pan et al.	2022	Dados voláteis (EAD).	CNN/LSTM	Combina DL com sobrevivência.
Costa et al.	2018	Riscos Proporcionais.	Modelo de Cox	Gênero e reprovações explicam.
Este trabalho	2026	Disparidade social.	Sobrevivência/ Métricas estatísticas	Modelo de análise de sobrevivência

O uso de modelos de regressão é usado para quantificar variáveis determinantes na trajetória acadêmica. [Vitelli et al. \(2010\)](#) utilizam a Regressão Logística via método

stepwise para demonstrar que a evasão vincula-se à indefinição na escolha profissional, à condição financeira e ao rendimento acadêmico. O modelo alcançou acurácia geral de 91,5% e capacidade de identificar 96,0% dos evadidos, apresentando um coeficiente de Nagelkerke R^2 de 0,815. Os resultados quantificam que o aumento de um grau na média de desempenho reduz o risco de evasão em 32,5%, enquanto a ausência de auxílio financeiro eleva a probabilidade de abandono em 259%.

Em análise de alternativas estatísticas, [Dey e Astin \(1993\)](#) comparam os modelos Logit, Probit e Regressão Linear em uma amostra nacional de 947 estudantes, registrando uma média de notas de 4,40 (em escala de 8 pontos) e uma taxa de retenção de 49,3% no período observado. Os autores observam que os métodos apresentam resultados similares na predição da retenção, com correlações entre as probabilidades previstas pelos modelos atingindo valores próximos a 1,0.

A influência do desempenho inicial como indicador de sucesso é descrita por [Tsiakmaki et al. \(2018\)](#), que aplicam técnicas de regressão em notas de uma amostra de 592 alunos, coletada ao longo de um período de quatro anos, para prever notas do segundo semestre. Ao comparar oito algoritmos de regressão e utilizar o Erro Médio Absoluto (MAE) como métrica de eficiência, os autores identificaram que a Regressão Linear (LR) apresentou o melhor desempenho na estimativa das notas em comparação a métodos como Random Forests e SVM e identificam que o desempenho acadêmico no primeiro semestre é o preditor para a continuidade da trajetória.

Na mesma linha, [Ajoodha et al. \(2020\)](#) comparam seis algoritmos de aprendizado de máquina, sendo eles Árvore de Decisão (C4.5), K-Star, Naïve Bayes, Support Vector Machines (SVMs), Florestas Aleatórias e Regressão Logística Linear, para classificar estudantes em quatro perfis de risco em universidades sul-africanas. Os resultados demonstram que, enquanto a Regressão Logística Linear alcançou 78% de acurácia, o algoritmo de Florestas Aleatórias obteve o melhor desempenho preditivo com 82% de acerto geral. Através do ranking de Ganho de Informação por entropia, o estudo identificou que o histórico familiar e as características pré-escolares são determinantes na estabilidade do vínculo acadêmico. Os preditores mais significativos foram o código do plano de carreira, com índice de 1,21, a descrição do curso, com 1,15, e a idade no primeiro ano, com 0,20, evidenciando que é possível definir o risco de desistência antes mesmo do ingresso do estudante na instituição

No contexto de gestão baseada em evidências, a plataforma SABIA utiliza *Business Intelligence* e *Analytics* para transformar microdados em indicadores de desempenho, permitindo que gestores identifiquem tendências de curso (MARQUES et al., 2023). Complementarmente, Nascimento et al. (2022) destaca a urgência na revisão do Projeto Pedagógico do Curso (PPC) para combater índices alarmantes de intenção de desistência. Para fortalecer a identidade acadêmica e mitigar a evasão, os autores propõem uma estrutura curricular que dedica 28% da carga horária à formação docente e 24% aos fundamentos da computação, reduzindo o peso de áreas de alta retenção, como a Matemática, para 14% do total. Argumentam ainda que a permanência deve ser estimulada pela melhoria da infraestrutura institucional e inclusão de atividades de extensão, elevando o protagonismo social do discente como mediador tecnológico.

No campo pedagógico, Vasconcelos e Andrade (2018) identificam que a falta de embasamento lógico-matemático desde o ensino médio é um gatilho crítico para a desistência, revelando que apenas 19,2% dos ingressantes avaliam seus conhecimentos prévios nessa área como “ótimos”. Por meio de um estudo quanti-qualitativo, os autores demonstram que essa defasagem resulta em altos índices de retenção precoce: 50% dos alunos precisaram cursar Matemática Discreta por duas vezes e 42,2% repetiram Introdução à Programação. Esse cenário impacta a persistência, levando 73,1% dos veteranos a considerarem o abandono da trajetória acadêmica. Como resposta, Falcão et al. (2018) relatam que o desmembramento de disciplinas densas e a introdução do Pensamento Computacional elevam a continuidade dos ingressantes ao reduzir o choque acadêmico inicial.

Buscando compreender os Dados Educacionais, Brito et al. (2020) mapeia atributos críticos no cenário brasileiro por meio do algoritmo Random Forest, analisando uma base de 51.175 registros de uma universidade federal integrados a mais de 200 indicadores demográficos e socioeconômicos do Censo Brasileiro. O estudo alcançou uma acurácia geral de aproximadamente 70%, com desempenho superior nas áreas de Engenharias (78,98%) e Ciências Exatas (76,19%), demonstrando ser possível identificar o risco de evasão utilizando apenas características disponíveis no momento do ingresso. Nessa mesma linha, Barbosa et al. (2024) propõem a utilização de inteligência artificial explicável e modelos fundamentados em algoritmos como Random Forest, SVM, Naive Bayes e KNN para a identificação de variáveis críticas no fenômeno da evasão. Utilizando a técnica

SHAP, os autores demonstraram que o número de cancelamentos de disciplinas é o fator de maior relevância na decisão do modelo, o qual apresentou uma acurácia de 97,85% no conjunto de treino e 90,46% nos dados de teste. O foco do estudo reside na garantia de justiça algorítmica ao monitorar métricas de desempenho estratificadas por gênero e raça, visando assegurar que as predições de risco não perpetuem desigualdades históricas.

No contexto de equidade, [Duran et al. \(2023\)](#) identificam que a permanência feminina em cursos de computação é fragilizada por ambientes competitivos e pela percepção de um campo predominantemente masculino. Utilizando uma metodologia quantitativa baseada em um levantamento com 3.193 estudantes e análise estatística via Mann-Whitney U e Estatística de Tipo Wald (WTS), os autores demonstraram que o fator campo predominantemente masculino é o principal preditor da intenção de abandono entre mulheres, com um tamanho de efeito médio de $r = -0,34$. Tais achados reforçam que a intenção de evasão está ligada tanto ao desempenho quanto à percepção do ambiente, evidenciando a percepção de que as disciplinas de programação são excessivamente difíceis ($r = -0,19$), o que resulta em uma lacuna de confiança, na qual mulheres realizam uma autoavaliação de competência inferior à dos homens, mesmo quando apresentam notas equivalentes, o que fragiliza o sentimento de pertencimento e a trajetória acadêmica em STEM. [Ortiz-Martínez et al. \(2023\)](#) demonstram que o desempenho acadêmico não é a causa principal da evasão feminina em STEM, visto que 75% das desistentes possuíam notas altas, concentrando-se o maior risco de desligamento nos primeiros quatro semestres.

A aplicação de técnicas de classificação demonstra a transição da análise de dados estáticos de matrícula para o monitoramento contínuo da trajetória acadêmica. [Pascoal et al. \(2016\)](#) utilizaram o classificador Naive Bayes em registros de Ciência da Computação, alcançando uma acurácia de 85,48% e uma AUC de 0,919, ao identificar que o grau de instrução materno é a variável social com maior impacto na probabilidade de evasão, enquanto o desempenho em disciplinas de exatas, como Cálculo e Física, atua como o principal preditor acadêmico. Reforçando essa perspectiva, [Pinheiro et al. \(2018\)](#) compararam os algoritmos SVM, Naive Bayes e Árvores de Decisão, constatando que o peso preditivo dos dados gerados durante o curso é significativamente superior aos dados de ingresso, uma vez que a acurácia do modelo elevou-se de 77,54% na base de matrícula para 98,76% com o uso do SVM na base unificada. Tais achados evidenciam que variáveis dinâmicas, como o coeficiente de rendimento (CR) e o tempo de curso, são determinantes

para a precisão do diagnóstico de risco, permitindo intervenções institucionais mais assertivas ao longo da vida acadêmica.

Complementarmente, [Carminati et al. \(2020\)](#) aplicam a metodologia *Cross-Industry Standard Process for Data Mining* (CRISP-DM) associada à Análise de Componentes Principais (ACP), utilizando o algoritmo J48 para a extração de regras de associação em uma base de 2.925 estudantes. Os autores alcançaram uma acurácia de 85,98% e um recall de 82,30%, identificando que a mediana das notas é um preditor mais robusto que a média simples e que alunos com rendimento inferior ao mínimo no semestre anterior possuem elevada propensão ao abandono antes do quinto período.

No campo do aprendizado profundo, [Na et al. \(2025\)](#) utilizam Redes Neurais Recorrentes (RNN) com arquiteturas LSTM e GRU e mecanismo de atenção multi-head para processar registros semestrais de 20.285 estudantes, alcançando um F1 score de 0,9416. Esse desempenho é 5% superior aos métodos tradicionais baseados em resumos estáticos, demonstrando que registros acadêmicos possuem padrões temporais inerentes que, quando explorados, elevam a precisão do diagnóstico de risco. Todavia, [Maia et al. \(2023\)](#) e [Mao et al. \(2024\)](#) observam que a complexidade desses modelos dificulta a interpretabilidade, denominando-os como modelos de “caixa preta”. Enquanto [Maia et al. \(2023\)](#) apontam que a maioria dos estudos prioriza métricas técnicas em detrimento do significado educacional, [Mao et al. \(2024\)](#) reforçam a necessidade de integrar técnicas de explicabilidade, como SHAP e LIME, para extrair a importância das variáveis e subsidiar decisões pedagógicas baseadas em evidências.

Para grandes volumes de dados, [Menolli e Neto \(2021\)](#) adotam o *Design Science Research (DSR)* para implantar um método de análise baseado em BI e *Data Warehouse (DW)*. Através da modelagem multidimensional e ferramentas OLAP, os autores integraram microdados do Censo da Educação Superior (INEP) para realizar análises em diferentes níveis de granularidade, identificando que a evasão apresenta índices superiores em instituições privadas, as quais registram taxas médias de 26% contra 12% na rede pública. De forma similar, [Sorgatto e Cáceres \(2023\)](#) utilizam *Learning Analytics* associado a um DW multidimensional para analisar a mudança no perfil dos ingressantes brasileiros entre 2009 e 2021, destacando que a procura por cursos tecnológicos superou as licenciaturas nos últimos três censos, evidenciando uma reconfiguração nas áreas de conhecimento e graus acadêmicos de escolha.

A dimensão social é explorada por [Gisi \(2006\)](#), que descreve o caráter de desigualdade no acesso e na permanência, argumentando que a expansão do ensino superior no Brasil não foi acompanhada pela redução das disparidades de condições entre os estudantes. Este cenário é corroborado por [Zamith et al. \(2025\)](#), que utilizou Análise Fatorial Exploratória e Confirmatória com 63 discentes evadidos para demonstrar que o *campus* de Crateús da UFC apresentou 46% de matrículas canceladas, contra 27% no *campus* de Russas. O estudo identifica que o grau de instrução da mãe é o atributo social com maior ganho de informação para a predição da evasão, validando o impacto do capital cultural na trajetória acadêmica. Adicionalmente, o estudo demonstra que o desempenho acadêmico configura-se como o fator de proteção mais expressivo. Observou-se que o incremento de uma unidade no Índice de Rendimento Acadêmico (IRA) amplia as chances de sobrevivência do discente em 279,4%, promovendo a estabilização da trajetória de permanência a partir do sétimo período.

Nesse contexto, [Filipak e Pacheco \(2017\)](#) reforçam que a democratização do acesso não garante a conclusão do curso sem mecanismos de suporte institucional. Dados nacionais analisados por [Silva e Sampaio \(2022\)](#), em revisão sistemática de 39 documentos, revelam que após dez anos de acompanhamento da coorte de 2010, apenas 40% dos estudantes concluíram a graduação, enquanto a taxa de desistência acumulada atingiu 59%.

O estudo de [Almeida e Schimiguel \(2011\)](#), realizado na Licenciatura em Física do IFMA, quantifica que de 320 vagas oferecidas, apenas 244 resultaram em matrículas, gerando uma ociosidade de 23,7% antes do início do período letivo. Os autores concluem que a decisão na escolha do curso (28%) e a desvalorização da profissão (16%) são os principais motivos para o abandono, transformando a desistência em um prejuízo social e econômico ao não suprir a carência de professores. Sob uma perspectiva sistêmica, [Moura et al. \(2020\)](#) reforçam que a evasão em licenciaturas gera um ciclo negativo para a educação básica.

A Análise de Sobrevivência (AS) é aplicada para o estudo da dinâmica temporal do abandono e o tratamento de dados censurados. [Chen et al. \(2018\)](#) realizam um estudo comparativo entre nove cursos da área de STEM com uma amostra de 12.293 estudantes, observando que 60% das desistências ocorrem nos primeiros dois anos e que a taxa de evasão acumulada atinge 23,62% ao final do sexto semestre. Os autores demonstram que modelos baseados em tempo apresentam desempenho superior aos de classificação

na identificação de risco; o modelo de Cox alcançou uma AUC de 0,948 na predição do terceiro semestre utilizando dados iniciais, além de uma acurácia de 61,4% na previsão do semestre exato da ocorrência da evasão.

O efeito de reformas na matriz acadêmica é descrito por [Cavalcanti et al. \(2024\)](#), que utiliza o estimador de Kaplan-Meier para demonstrar que atualizações pedagógicas e a adoção de metodologias ativas elevaram a probabilidade de sobrevivência entre 30% e 40% a partir do terceiro ano de vínculo. Sob a mesma perspectiva, [Gutierrez-Pachas et al. \(2022\)](#) analisam o impacto de mudanças no *design* curricular em programas de Ciência da Computação, verificando a conformidade com padrões internacionais (ACM/IEEE). Os autores compararam algoritmos de *Machine Learning* com métodos de AS, como o *Cox Proportional Hazards* (CPH) e o *DeepSurv*, identificando que, apesar do aumento na permanência geral pós-reestruturação, a representatividade feminina registrou um índice de apenas 16,8%. O estudo destaca que o modelo *DeepSurv* apresentou maior consistência na predição de risco, observando que a probabilidade de sobrevivência acadêmica declina após o primeiro ano de curso.

Em uma escala de análise longitudinal, [Gutierrez-Pachas et al. \(2023\)](#) utilizam a AS para identificar fatores que influenciam o tempo até o abandono em uma amostra de 13.696 estudantes. O trabalho demonstra a eficácia de modelos de *deep learning*, em que o algoritmo CNN alcançou acurácia de 98,7% em tarefas de classificação, enquanto o *DeepSurv* atingiu um índice de concordância (*C-index*) de 0,952 em Engenharia e 0,891 em Computação. Os resultados revelam que os fatores acadêmicos são os preditores dominantes na modelagem de risco: a proporção de aprovações obteve peso de 25,84% e o GPA atingiu 29,04% de relevância. Em contraste, variáveis como gênero (0,00% a 2,88%) e IDH de procedência (0,00% a 2,83%) apresentaram impacto marginal na predição do tempo até a evasão, indicando que o desempenho acumulado nos quatro semestres iniciais é o determinante da trajetória de risco e visando subsidiar decisões institucionais orientadas à equidade e ao suporte discente.

Em áreas de STEM, o insucesso em disciplinas de Cálculo e Algoritmos é identificado como um fator determinante para a interrupção da trajetória acadêmica. [Silva et al. \(2016\)](#) quantifica que as taxas de reprovação em disciplinas de primeiro ano, como Cálculo I e AVGA, atingem 64,47% e 62,02%, respectivamente, evidenciando alunos analisados com deficiências graves em conteúdos de matemática elementar do ensino médio. [Garcia e](#)

[Gomes \(2022\)](#) reforça essa criticidade ao apontar que a dificuldade de desempenho e a reprovação são as causas mais incidentes em 53% dos trabalhos revisados na área de exatas, onde a taxa média de evasão é quantificada em 39,21%, com picos de 53% em cursos de Física.

Sob a mesma perspectiva, [Silva \(2024\)](#) evidencia que a reforma curricular de 2018 na UFRPE elevou o Conceito ENADE de 4 para 5, dobrou o volume de matrículas e reduziu os trancamentos a um terço dos níveis anteriores. Esse êxito é atribuído à introdução de disciplinas como Pensamento Computacional e laboratórios práticos, que otimizaram a aprovação nas etapas iniciais e resultaram em 100% de concordância discente sobre o estímulo ao pensamento crítico.

Quanto aos mecanismos de suporte, [Saccaro et al. \(2019\)](#) e [Souza et al. \(2022\)](#) utilizam o estimador de *Kaplan-Meier* para avaliar a permanência. [Saccaro et al. \(2019\)](#) demonstram que o apoio financeiro e a reserva de vagas para cotistas resultam em probabilidades de evasão inferiores em comparação aos demais grupos. Por sua vez, [Souza et al. \(2022\)](#) utilizam a análise de sobrevivência associada a modelos preditivos, como o XGBoost, para analisar a evasão em Sistemas de Informação, identificando que 50% dos discentes evadem nos dois primeiros semestres, restando 23% de sobreviventes após o período regular de oito semestres. O modelo, com acurácia de 90,59%, aponta o desempenho em disciplinas de programação como o fator determinante para a continuidade.

De forma convergente, [Hoed \(2016\)](#) emprega a Regressão de Cox para demonstrar que a retenção nas etapas iniciais eleva o risco de evasão nos períodos subsequentes, enquanto [Souza et al. \(2024\)](#) utiliza o modelo Log-Normal na Licenciatura em Computação da UnB para evidenciar que alunos cotistas demandam políticas de monitoria e suporte financeiro específicas, ressaltando que o sentimento de pertencimento impacta a sobrevivência acadêmica de forma equivalente aos indicadores de rendimento.

No contexto da área de Física, [Junior et al. \(2012\)](#) observam, por meio de análise de sobrevivência, que as mulheres apresentam tempo de permanência no sistema superior ao dos homens, indicando que a decisão de evadir ocorre após períodos prolongados de retenção. O impacto do suporte institucional via pesquisa é quantificado por [Okimoto et al. \(2025\)](#), que analisam a participação de estudantes no projeto SUPER. Os resultados indicam que a inserção em atividades de capacitação tecnológica e o recebimento de suporte financeiro resultam em melhorias no rendimento acadêmico, com valores-p de 0,008 e

0,004 em cursos de computação, configurando a integração científica como método de manutenção do vínculo discente.

A assistência estudantil, conforme [Cavalcanti et al. \(2025\)](#), favorece a conclusão do curso em áreas de STEM, com impacto observado no público feminino que recebe auxílios financeiros e sociais. [Ribeiro et al. \(2021\)](#) descrevem o acolhimento institucional como um mecanismo de oposição à vulnerabilidade financeira, o que influencia a decisão de permanência. Em relação aos determinantes da persistência, [Nierotka et al. \(2023\)](#) e [Klitzke e Carvalhaes \(2023\)](#) observam que o coeficiente de rendimento acumulado (CRa) e a nota do Enem apresentam maior peso estatístico do que a origem sociodemográfica, sendo que o ingresso na primeira opção de curso atua como preditor de continuidade.

No âmbito das inovações metodológicas, a rede SAVSNet ([PAN et al., 2022](#)) combina redes neurais profundas (CNN e LSTM) com funções de perda baseadas em análise de sobrevivência para processar dados voláteis em cursos a distância. [Santos et al. \(2024\)](#) utilizam modelos de Cadeias de Markov Absorventes integrados à AS para simular o progresso acadêmico, identificando como a retenção em semestres específicos altera a probabilidade final de conclusão. Por fim, [Ameri et al. \(2016\)](#) aplicam modelos de Cox dependentes do tempo (TD-Cox) para integrar o GPA semestral, alcançando uma identificação de estudantes em risco com valor 9% superior aos métodos tradicionais. [Costa et al. \(2018\)](#) utilizam modelos de riscos proporcionais para indicar que o número de semestres cursados, o gênero e a ocorrência de reprovações explicam o risco de evasão.

3.1 Síntese da literatura e lacunas

A revisão da literatura demonstra que a evasão no ensino superior apresenta dependência temporal e acadêmica. Os estudos estabelecem a relação entre o insucesso em disciplinas iniciais e a redução do tempo de permanência, além de evidenciarem o impacto de suportes financeiros e institucionais na manutenção do vínculo discente. Observa-se que abordagens predominantes focam no desempenho acadêmico ou em reformas de matrizes curriculares, utilizando modelos de classificação estática que desconsideram a dimensão temporal da trajetória acadêmica.

Diferente dessas abordagens, este trabalho utiliza a análise de sobrevivência integrada a uma ferramenta de BI. Enquanto estudos como os de [Saccaro et al. \(2019\)](#)

e [Souza et al. \(2022\)](#) realizam análises de sobrevivência em contextos específicos e bases estáticas, a integração proposta com BI permite a filtragem dinâmica e o monitoramento segmentado em tempo real, suprimindo a lacuna de ferramentas longitudinais operacionais mencionada por [Vitelli et al. \(2010\)](#). A modelagem probabilística implementada permite observar o evento de evasão como uma função do tempo, identificando os períodos da trajetória em que as vulnerabilidades sociais atuam como determinantes para o abandono.

O diferencial metodológico reside no tratamento de dados censurados. Ao contrário de modelos de classificação aplicados por [Ajoodha et al. \(2020\)](#) e [Pascoal et al. \(2016\)](#), que descartam registros incompletos até o fim da coleta, a análise de sobrevivência incorpora essas observações ao cálculo das funções de risco. Enquanto as abordagens de ML mencionadas focam na predição binária (evadiu/não evadiu), este procedimento evita a subestimação do tempo de permanência e utiliza a totalidade dos registros para gerar estimativas de probabilidade sobre o comportamento da coorte.

Além do abandono, o modelo permite o monitoramento do tempo até a formação para avaliar o fluxo de graduação através de indicadores centralizados. Em contraste com o uso de Cadeias de Markov proposto por [Santos et al. \(2024\)](#) para simular o progresso semestral, este trabalho utiliza a função de risco para quantificar o impacto de variáveis.

Esta pesquisa utiliza o teste de *Log-Rank* e o cálculo de intervalos de confiança para validar as diferenças nas estimativas de *Kaplan-Meier*. O uso do *Hazard Ratio (HR)* e de medidas de *Effect Size* mensura o risco relativo entre grupos e a significância das trajetórias identificadas. Essa validação fornece evidências sobre o impacto das vulnerabilidades na permanência para subsidiar decisões direcionadas aos grupos em condição de risco.

Percebe-se que a abordagem proposta alinha-se aos estudos que descrevem a evasão como um fenômeno dependente do tempo e de fatores socioeconômicos. O método supre a lacuna de ferramentas para o monitoramento longitudinal integrado ao BI, permitindo a análise dinâmica do risco acadêmico. Adicionalmente, a aplicação de testes estatísticos para a validação das diferenças entre as curvas de sobrevivência fornece parâmetros de verificação de significância em modelos de acompanhamento discente. Ao incorporar o tratamento de dados censurados e o cálculo de risco relativo, o trabalho viabiliza a quantificação do impacto de variáveis sociais nas trajetórias de evasão e formação.

4 Método

Este capítulo detalha o método e as tecnologias utilizadas na elaboração do modelo de Análise de Sobrevivência. A estrutura divide-se em duas etapas: extração e preparação de dados e construção do modelo estatístico, conforme exhibe a Figura 7. A primeira parte, abordada na Seção 4.1, trata da extração e preparação de dados, detalhando a coleta de registros do sistema acadêmico e a definição das variáveis de tempo e evento.

Por sua vez, a segunda parte, discutida na Seção 4.2, trata da construção do modelo estatístico, fundamentado no método *Kaplan-Meier*. Esta etapa compreende a estimação das curvas de permanência, além da aplicação do teste de *Log-Rank* para a verificação de hipóteses entre grupos. O modelo utiliza o *Hazard ratio* (razão de risco) e métricas de *Effect-size* (tamanho do efeito) para mensurar as disparidades entre as categorias analisadas.

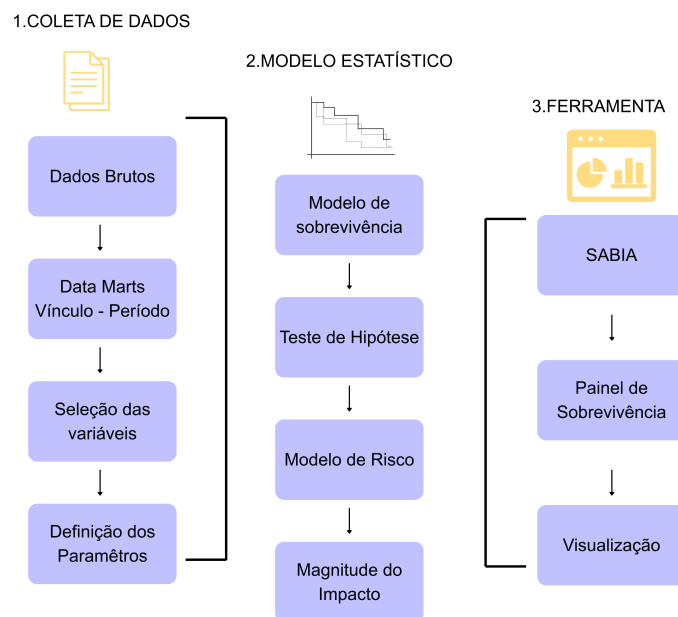


Figura 7 – Fluxo do modelo para análise de permanência discente.

Fonte: Elaborado pelo autor (2026).

4.1 Extração e preparação de dados

A coleta de dados deste trabalho fundamenta-se no fluxo de processamento estabelecido pela plataforma SABIA (MARQUES et al., 2023). A extração das informações ocorre semestralmente, conforme o cronograma de atualização das bases de dados do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Os dados obtidos resultam de uma combinação entre registros administrativos internos da Universidade Federal Rural de Pernambuco (UFRPE) e dados do INEP. O fluxo de integração dessas fontes é apresentado na Figura 8.

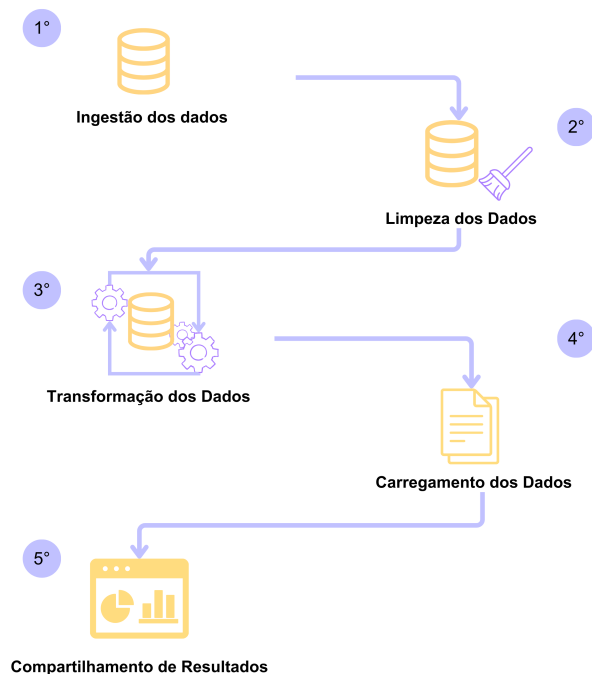


Figura 8 – Fluxo de extração e integração de dados.

Fonte: Elaborado pelo autor (2026).

O conjunto de dados compreende discentes dos cursos da área de STEM da Unidade Acadêmica de Recife (Sede/UFRPE) matriculados em Bacharelado em Sistemas de Informação (BSI), Bacharelado em Ciência da Computação (BCC), Licenciatura em Computação (LC), Licenciatura em Física e Licenciatura em Matemática. A escolha dos cursos de BSI e LC para a avaliação do impacto de reformas curriculares ocorreu em função da reformulação dos Projetos Pedagógicos de Curso (PPCs) no período de observação.

Os dados são anonimizados e segmentados por categoria em *Data Marts* no formato *.parquet* para armazenamento e leitura de registros (BARBOSA et al., 2024). O

arquivo *vinculo_periodo.parquet* centraliza as informações acadêmicas e sociodemográficas necessárias para a análise, compreendendo o histórico registrado na instituição. A Tabela 3 apresenta as colunas contidas no arquivo inicial.

Tabela 3 – Variáveis extraídas do arquivo *vinculo_periodo.parquet*.

Categoria	Colunas Extraídas
Perfil e Identificação	id_discente, data_nascimento, genero, estado_civil, funcionario, tipo_sanguineo
Localização	bairro, cidade, uf, cep
Ingresso Acadêmico	forma_ingresso, ano_ingresso, semestre_ingresso, periodo_letivo_ingresso, nota_vestibular, classificacao_vestibular, modalidade_ingresso, ingresso, ano_conclusao_medio, forma_ingresso_regular
Dados do Curso	curso, turno, cod_curriculo, cod_inep, campus, natureza, area_vestibular, area_cnpq, tipo_curso, is_stem
Desempenho	media_final, media_final_aprovado, ch_total, qtd_aprovacoes, qtd_reprovacoes, qtd_cancelamentos, qtd_outros, ch_aprovacoes, ch_reprovacoes, ch_cancelamentos, ch_outros, qtd_insucessos, ch_aprovacoes_acum, retido_reprovacoes, retido_cancelamentos, retido_progressao
Fluxo e Vínculo	situacao_periodo_letivo, tipo_periodo, duracao_vinculo, retido_duracao_vinculo, is_assistido, situacao_simplificada, situacao_final, ult_periodo_letivo, qtd_periodos_letivos, ult_situacao_periodo
Situação do Vínculo	ult_sit_formado, ult_sit_evadido, ult_sit_vinculado, ult_duracao_vinculo, ult_duracao_vinculo_evasao, ult_duracao_vinculo_formacao, ult_classe_evasao, duracao_final, duracao_curso_x, duracao_curso_y, conclusao_ideal

Fonte: Elaborado pelo autor (2026).

Entre as variáveis extraídas, selecionou-se um subconjunto para a composição do modelo de análise de sobrevivência. A Tabela 4 detalha os atributos mantidos, seus respectivos tipos de dados e descrições. A seleção baseia-se na disponibilidade de registros institucionais e na relevância para o estudo da permanência estudantil. Variáveis de localização geográfica e de desempenho acadêmico histórico, como o Índice de Rendimento Acadêmico (IRA) ou reprovações acumuladas, não foram incluídas no modelo final. O

recorte da pesquisa restringe-se aos dados cadastrais, de vínculo e de perfil de ingresso para permitir o uso de covariáveis fixas. Essa estrutura visa isolar o impacto das alterações curriculares e das disparidades entre os cursos de STEM e do cenário institucional, sem a interferência de métricas de desempenho dinâmicas durante a trajetória discente.

Tabela 4 – Atributos selecionados para o modelo final.

Variável	Tipo	Descrição
id_discente	String	Identificador único e anonimizado do estudante.
semestre_ingresso	Inteiro	Semestre letivo de entrada (1 ou 2).
periodo_letivo_ingresso	Float	Representação numérica do ano e semestre de ingresso.
duracao_vinculo	Inteiro	Tempo de vínculo transcorrido em semestres.
ult_duracao_vinculo	Inteiro	Tempo total acumulado até o desfecho ou última atualização.
forma_ingresso	String	Modalidade de entrada no curso superior.
curso	String	Nome do curso de graduação.
campus	String	Unidade acadêmica de vinculação.
genero	String	Categoria de identificação de gênero do discente.
turno	String	Período das atividades acadêmicas.
area_cnpq	String	Grande área do conhecimento conforme classificação.
is_assistido	Booleano	Indicador de recebimento de assistência estudantil.
perfil	String	Ocorrência da Mudança de Perfil Pedagógico Curricular (ANTES, DEPOIS)

Fonte: Elaborado pelo autor (2026).

Para a construção do modelo de sobrevivência, o parâmetro temporal é determinado pela variável *ult_duracao_vinculo*, que quantifica a permanência do discente na instituição em períodos semestrais. O status do evento é definido com base na variável *ult_situacao_periodo*, sendo caracterizado pelas ocorrências *EVADIDO* ou *FORMADO*. Registros com a situação *VINCULADO* no limite temporal da coleta de dados são classificados como dados censurados, pois o evento não ocorreu durante o período de

observação.

A definição do parâmetro de agrupamento utiliza os atributos categóricos como *curso*, *turno*, *campus*, *forma_ingresso* e *gênero*, além da variável booleana *is_assistido*, como variáveis independentes. Essa configuração permite a segmentação da análise conforme o perfil acadêmico e sociodemográfico.

4.2 Construção do modelo

A análise da probabilidade de permanência discente durante o vínculo institucional fundamenta-se no tratamento do evento de interesse e no tempo até sua ocorrência como variável dependente. A organização e o resumo dos dados, de natureza descritiva, embasam a interpretação dos fenômenos analisados (MONTGOMERY et al., 2000). A abordagem é quantitativa, pautada no processamento de registros numéricos para a identificação de padrões e a validação de hipóteses por meio de testes estatísticos. Embora o modelo permita a análise de diferentes desfechos acadêmicos (*EVADIDO* e *FORMADO*), este trabalho foca na evasão como o evento de interesse principal. Portanto, as análises subsequentes consideram a interrupção do vínculo como a falha a ser observada, tratando as demais situações como censura.

4.2.1 Análise de sobrevivência

O método proposto utiliza a análise de sobrevivência para descrever a dinâmica da permanência no ensino superior. O modelo estima a probabilidade de continuidade do vínculo discente ao longo do tempo, permitindo a comparação entre agrupamentos como gênero, turno, campus, semestre de ingresso, forma de ingresso e acesso à assistência estudantil. A variável gênero é processada conforme a classificação binária registrada nos sistemas administrativos e no Censo da Educação Superior (INEP), sendo limitada às categorias masculina e feminina e não abrange a pluralidade de identidades de gênero.

A aplicação do estimador de *Kaplan-Meier* justifica-se pela natureza do fenômeno da evasão, que resulta da interação de fatores individuais, socioeconômicos e institucionais, não aderindo a uma distribuição estatística padrão. Por ser um método não paramétrico baseado em estatísticas de ordem, a técnica permite identificar padrões nas trajetórias

acadêmicas sem a imposição de restrições ou suposições prévias sobre a distribuição dos dados (COLOSIMO; GIOLO, 2021).

O modelo utiliza a ordenação dos tempos observados, definidos pela duração de vínculo em períodos semestrais, para realizar inferências sobre a amostra, incorporando o tratamento de censura à direita. Este procedimento permite incluir registros de estudantes com trajetórias incompletas, preservando o tempo de contribuição dos discentes vinculados à instituição e garantindo a validade das estimativas de sobrevivência (KAPLAN; MEIER, 1958).

O conjunto de dados, resultante das etapas de extração, transformação e carregamento (ETL), é processado em linguagem *Python*¹. A manipulação e a análise dos registros fundamentam-se na biblioteca *Pandas*², enquanto o suporte às operações matemáticas ocorre via biblioteca *Numpy*³. As estimativas de sobrevivência e os testes estatísticos são executados com o auxílio da biblioteca *Lifelines*⁴, cujas rotinas de processamento estão distribuídas nos arquivos `Permanencia.py`, `survivability.py` e `survival_stats.py`.

Os módulos desenvolvidos estruturam responsabilidades distintas no processamento da análise. O arquivo `survivability_stats.py` é responsável pela execução dos cálculos estatísticos e testes de hipóteses. O módulo `survivability.py` realiza a orquestração, pré-processamento e agrupamento dos dados para as análises de sobrevivência. Por fim, o arquivo `Permanencia.py` é responsável pela interface visual, integrando os filtros e componentes interativos para a exibição dos resultados.

O ajuste da função de sobrevivência fundamenta-se na parametrização das variáveis temporais e de evento, conforme ilustrado na Figura 9. O marco inicial da definição dos parâmetros estabelece a origem da observação no ingresso do discente à instituição, representado pelo valor inicial da duração do vínculo. A escala de tempo é definida pela duração do vínculo em períodos semestrais.

O evento de interesse ou falha é configurado como a evasão por curso, caracterizada pelo desligamento do discente do curso de origem antes da integralização curricular. Esta definição abrange a interrupção do vínculo por abandono, jubramento, falecimento ou demais formas de desligamento registradas institucionalmente. No modelo, o valor 1

¹ <https://www.python.org/>

² <https://pandas.pydata.org/>

³ <https://numpy.org/pt/>

⁴ <https://lifelines.readthedocs.io/>

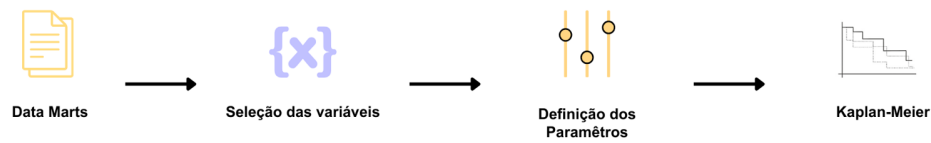


Figura 9 – Representação do modelo de sobrevivência.

Fonte: Elaborado pelo autor (2026).

corresponde à ocorrência da falha e o valor 0 identifica a permanência do estudante no sistema.

A censura ocorre quando o evento de interesse não é registrado durante o período de observação (KLEINBAUM et al., 2012). Neste trabalho, aplica-se a censura à direita, que ocorre quando o tempo até a evasão é superior ao período de acompanhamento ou quando o evento não é verificado até o término da observação. Esta condição considera discentes que mantêm o vínculo ativo ou que concluíram o curso dentro do intervalo temporal analisado (CAVALCANTI et al., 2024). Tal composição permite a inclusão de trajetórias incompletas no cálculo da função de sobrevivência, viabilizando identificar os padrões dos coortes sem comprometer a validade das estimativas.

A implementação dessas definições ocorre no arquivo `survivability.py`, que concentra a lógica de processamento e parametrização para a execução dos modelos estatísticos. O desenvolvimento do módulo utiliza a biblioteca *Lifelines* para a aplicação dos estimadores, por meio da classe `KaplanMeierFitter`.

O procedimento operacional compreende o ajuste do estimador aos vetores de duração temporal e aos indicadores binários de ocorrência do evento. A partir deste ajuste, extraem-se a função de sobrevivência pontual pelo atributo `survival_function_`, acompanhada dos intervalos de confiança calculados para cada ponto da linha do tempo das observações. Os resultados fornecem as probabilidades acumuladas de permanência e a variabilidade estatística das estimativas ao longo do período analisado.

Antes do processamento estatístico, os dados passam por um fluxo de filtragem no módulo para garantir a consistência das coortes. Registros sem informação de desfecho ou com duração de vínculo inconsistente são removidos. Para alunos com reingresso, considera-se apenas o vínculo mais recente, assegurando que o tempo de exposição ao evento reflita a realidade atual do currículo analisado.

As operações de cálculos iniciam-se na função `filter_prepare_dataframe`, que padroniza o tempo de início e a escala semestral das trajetórias acadêmicas. O parâmetro

de evento é processado pela função `__censor`, que identifica a ocorrência da falha (evasão ou formatura) e aplica a lógica de censura à direita. Nesta estrutura, a falha não censurada é registrada apenas quando o evento ocorre dentro do intervalo de observação e o tempo de vínculo é inferior ao limite máximo estabelecido. Caso contrário, o registro é classificado como censurado.

Neste trabalho, as estimativas fundamentam-se na ocorrência do evento *EVADIDO*, calculando-se a probabilidade de o estudante manter o vínculo institucional ao longo dos períodos semestrais. A configuração do modelo define a falha exclusivamente como o desligamento do curso, enquanto discentes que concluíram a graduação ou que permanecem matriculados ao final do período de observação são tratados via censura à direita. Esta parametrização assegura que a função de sobrevivência descreva especificamente o risco de abandono, permitindo a identificação dos períodos críticos para a permanência discente no sistema acadêmico.

Para a estimativa da probabilidade de continuidade, o módulo utiliza a função `__prepare_df_survival` como interface para os referidos estimadores, operacionalizando o cálculo das funções de sobrevivência e risco acumulado. Adicionalmente, o código executa análises descritivas sobre a estratificação dos dados por meio do parâmetro de agrupamento. As funções que compõem este arquivo e suas respectivas finalidades estão detalhadas na Tabela 5.

No contexto acadêmico, o modelo operacionaliza a equação de sobrevivência, para descreve a probabilidade de um discente permanecer com o vínculo ativo na instituição após decorridos t semestres. Nesta formulação, P indica a probabilidade e T é a variável aleatória correspondente ao tempo decorrido até a ocorrência da falha, como a evasão ou a diplomação.

Após os cálculos e definições dos parâmetros as estimativas são transformadas em um *DataFrame* de resultados estruturado para a visualização. Esse conjunto de dados final dispõe as informações em colunas que incluem o *time*, o *grupo*, a *prob*, bem como os limites inferior, *ic_lower* e superior, *ic_upper* do intervalo de confiança de 95% associados à probabilidade estimada em cada intervalo de acompanhamento. Adicionalmente, o *DataFrame* incorpora um indicador de censura, permitindo a identificação de observações com acompanhamento interrompido.

A estrutura de dados gerada pelo processamento estatístico serve de base para

as etapas subsequentes de validação e comunicação dos resultados. Enquanto o arquivo `survivability.py` consolida as curvas de probabilidade, as validações das estimativas entre as trajetórias dependem das métricas implementadas no módulo `survivability_stats.py`, conforme detalhado na Seção 4.2.2. Por fim, a integração desses componentes e a parametrização das consultas pelo usuário são gerenciadas pelo arquivo `permanencia.py`, cuja operação no ambiente SABIA é discutida no Capítulo 5.

Tabela 5 – Funções do módulo de processamento estatístico.

Função	Definição
<code>statistics_desc_group</code>	Agrega métricas descritivas por grupo, incluindo total de discentes, tempo médio, contagem de eventos, censuras e intervalos de confiança.
<code>compute_stats_por_grupo</code>	Calcula estatísticas individualizadas para um grupo e evento específico, tratando ausências de dados com marcadores nulos.
<code>__censor</code>	Aplica a lógica de censura à direita com base na situação acadêmica e no limite temporal de acompanhamento estabelecido.
<code>__select_columns</code>	Realiza a seleção de atributos relevantes e padroniza os tipos de dados para o processamento das funções de sobrevivência.
<code>__prepare_grupo</code>	Estrutura as coortes de análise conforme o critério de agrupamento (geral, perfil, customizado ou por variável específica).
<code>adicionar_grupo_geral</code>	Incorpora a amostra total aos dados estratificados para permitir a comparação entre grupos específicos e a média institucional.
<code>prepare_plot_dfs</code>	Coordena a extração dos conjuntos de dados necessários para as curvas de sobrevivência, risco acumulado, evasão e conclusão.
<code>filter_prepare_dataframe</code>	Executa o fluxo de limpeza e preparação inicial, garantindo a integridade dos vínculos e a aplicação das regras de censura.
<code>__prepare_df_survival</code>	Atua como interface para o cálculo de probabilidades via <i>Kaplan-Meier</i> .
<code>__prepare_df_logrank</code>	Processa o teste estatístico de <i>Log-Rank</i> para todas as combinações de grupos, validando a igualdade das funções de sobrevivência.
<code>prepare_effect_size</code>	Calcula a magnitude das diferenças estatísticas entre um grupo de referência e os demais agrupamentos da amostra.

4.2.2 Métricas estatísticas

As métricas de validação e inferência estatística são estruturadas no módulo `survivability_stats.py`, conforme detalhado na Tabela 6. A incerteza associada ao tempo médio de vínculo é calculada via método de *bootstrap* com 1000 reamostragens e técnica *basic*, operação executada pela função `__confidence_interval` e integrada à `statistics_desc_group`. Esta abordagem quantifica a variabilidade da média descritiva para cada agrupamento, considerando a distribuição dos dados e a presença de observações censuradas.

Tabela 6 – Funções do módulo de suporte e métricas estatísticas.

Função	Definição
<code>__put_compare_data</code>	Estrutura os dados de tempo e censura para análises comparativas.
<code>__func_fitter</code>	Interface para execução dos estimadores de Kaplan-Meier, retornando a linha do tempo, probabilidades e intervalos de confiança.
<code>__compare_groups</code>	Realiza o teste de Log-Rank entre pares de grupos para validar a hipótese nula de igualdade entre as distribuições de sobrevivência.
<code>__prepare_hazard_ratio</code>	Ajusta o modelo de riscos proporcionais de Cox para calcular o coeficiente e a razão de risco (<i>Hazard Ratio</i>) entre dois grupos específicos.
<code>__prepare_effect_size</code>	Mensura a magnitude das diferenças entre grupos utilizando a biblioteca <i>Pingouin</i> para o cálculo do tamanho do efeito.
<code>classificar_effect_size</code>	Categoriza o resultado numérico do tamanho do efeito em escalas qualitativas de acordo com o tipo de métrica aplicada.

Fonte: Elaborado pelo autor (2026).

Para as curvas de permanência, o intervalo de confiança de 95% é computado pela biblioteca *Lifelines* durante o ajuste do estimador de *Kaplan-Meier*. Os valores são extraídos do atributo `confidence_interval_` do objeto *fitter*. A função `__func_fitter` coordena a extração dessas estimativas, enquanto a função `__create_grupo_df` consolida as estimativas de sobrevivência em um *DataFrame* de resultados que incorpora os limites inferior e superior. Essa estruturação permite a visualização da precisão do modelo em cada ponto da trajetória acadêmica.

Além das estimativas intervalares, a comparação entre as curvas de permanência é realizada por meio do teste de *Log-Rank*, método não paramétrico que avalia a hipótese de inexistência de diferença estatística entre as funções de sobrevivência. A implementação deste procedimento ocorre no módulo `survivability_stats.py` através da função `__prepare_df_logrank`, que identifica os grupos únicos no conjunto de dados e gera combinações de pares para a execução de comparações bilaterais.

O processamento de cada par de grupos é gerenciado pela função `__compare_groups` que realiza a filtragem dos vetores de tempo e dos indicadores de censura. Esta função atua como interface para o método `logrank_test` da biblioteca *Lifelines*, que calcula a estatística do teste sob a hipótese nula definida na Equação 4.1

$$\begin{aligned} H_0 : h_1(t) &= h_2(t) \\ H_A : h_1(t) &= ch_2(t), \quad c \neq 1 \end{aligned} \tag{4.1}$$

onde $h(t)$ representa a função de risco no tempo t e c consiste no fator de proporcionalidade entre os grupos. O p-valor atua como indicador para a verificação da hipótese; caso este valor seja inferior ao nível de significância de 5%, ocorre a rejeição de H_0 , o que indica que as trajetórias acadêmicas dos grupos analisados apresentam divergência estatística (WELLEK, 1993). Em contrapartida, valores superiores a 0,05 sugerem a igualdade entre as distribuições (CUZICK, 1982). O teste de hipótese constata se as distinções observadas nos gráficos de *Kaplan-Meier* possuem fundamentação estatística.

Embora o estimador de *Kaplan-Meier* e o teste de *Log-Rank* identifiquem a existência de disparidades, tais métodos não paramétricos limitam-se à comparação de grupos isolados. Para mensurar o impacto simultâneo de múltiplas variáveis e quantificar o risco relativo entre os perfis discentes, implementou-se o Modelo de Riscos Proporcionais de Cox, caracterizado como uma abordagem semiparamétrica, por meio da função `__prepare_hazard_ratio`. O procedimento inicia-se na função `__prepare_pairwise_data`, que identifica os grupos únicos e gera todas as combinações de pares para comparação bilateral.

O processamento de cada par compreende a filtragem dos registros correspondentes e a codificação de uma variável indicadora binária (`grupo_bin`), onde o valor 0 define o grupo de referência e o valor 1 o grupo de comparação. O ajuste do modelo é executado pela classe `CoxPHFitter` da biblioteca *Lifelines*, utilizando como variáveis dependentes os

tempos de vínculo e os indicadores de evento.

A partir do ajuste, o algoritmo extrai o coeficiente de regressão (β) e calcula o HR por meio da exponencial deste coeficiente (e^β), o que permite mensurar a probabilidade relativa de ocorrência do evento entre os grupos. Um valor de $HR = 1$ indica riscos equivalentes, enquanto valores superiores ou inferiores à unidade apontam, respectivamente, maior ou menor propensão ao evento de interesse em relação ao grupo de referência (STARE; MAUCORT-BOULCH, 2016). Os resultados, incluindo o coeficiente, o HR e o p -valor, são consolidados em um *DataFrame* para a interpretação das diferenças de risco entre os agrupamentos.

Para quantificar a magnitude das diferenças entre agrupamentos, a implementação do cálculo do tamanho do efeito é realizada por meio da função `prepare_effect_size`. O módulo `survivability_stats.py` permite a seleção de um grupo de referência e a parametrização do tipo de métrica a ser aplicada, com suporte para indicadores como *Cohen's d*, *Hedges' g*, correlação de *Pearson*, *Eta-squared*, *Odds-ratio*, AUC e CLES.

O procedimento operacionaliza-se na função `__prepare_effect_size`, que realiza a filtragem dos tempos de permanência para cada par de grupos (referência *versus* comparação). A extração do indicador é executada via função `compute_effsize` da biblioteca *Pingouin*⁵. O valor absoluto resultante é processado pela função `classificar_effect_size`, que os categoriza em níveis de magnitude (pequeno, médio, grande ou muito grande).

As métricas são então consolidadas em *DataFrames* que discriminam o grupo de referência e os grupos comparativos. Essa estrutura de dados armazena os resultados das validações estatísticas para recuperação e exibição na interface do SABIA, cujas funcionalidades operacionais são detalhadas no Capítulo 5.

4.2.3 Limitações

O conjunto de técnicas estatísticas apresenta restrições relacionadas às premissas dos modelos e à natureza dos dados. O estimador de *Kaplan-Meier* possui limitações quanto ao viés da censura, exigindo a independência entre o mecanismo de censura e a ocorrência do evento de interesse para garantir estimativas não viciadas. O teste de

⁵ <https://pingouin-stats.org/build/html/index.html>

Log-Rank e o *Hazard Ratio* pressupõem riscos proporcionais, o que restringe a análise em casos de cruzamento de curvas de sobrevivência, onde a magnitude da diferença entre as coortes se altera conforme o avanço dos períodos letivos, limitando a precisão da razão de risco média calculada. Em cenários de amostras reduzidas ou com índices de censura, os intervalos de confiança apresentam variação na precisão. Adicionalmente, o cálculo do tamanho do efeito pelo g de *Hedges* não incorpora a censura, limitando a captura da dinâmica dos dados de sobrevivência.

No âmbito operacional da ferramenta, a precisão dos resultados está vinculada à integridade e à periodicidade de atualização dos registros administrativos nos *Data Marts* de origem. A dependência exclusiva de bases como as da UFRPE e do INEP constitui uma limitação, visto que tais registros não capturam variáveis subjetivas, como motivações psicológicas ou dinâmicas familiares, que influenciam a decisão de evadir. A ausência de dados provenientes de instrumentos qualitativos, como questionários aplicados no ato do desligamento, restringe o modelo ao ambiente institucional. A análise está sujeita a ameaças à validade, como o viés de seleção, visto que a amostra pode não representar a totalidade da diversidade dos estudantes. Fatores não controlados, como a qualidade do ensino e a motivação pessoal, influenciam os resultados de permanência. Essas restrições, somadas ao caráter observacional e longitudinal do estudo, limitam a inferência causal, indicando associações estatísticas entre as variáveis e o tempo até a ocorrência da evasão ou formação.

5 Ferramenta

A interface de usuário é implementada no formato de um *dashboard* interativo com a biblioteca *Streamlit*¹, utilizada na construção de painéis em linguagem *Python*, integrada ao ecossistema SABIA. A Figura 10 apresenta a tela de início do ecossistema, na qual o módulo Sobrevivência é disponibilizado. O arquivo `permanencia.py` é responsável pela integração entre as entradas fornecidas pelo usuário e os procedimentos estatísticos, permitindo a exploração dos dados de permanência em diferentes níveis de granularidade.

Figura 10 – Interface inicial do ecossistema SABIA e acesso ao módulo de sobrevivência.



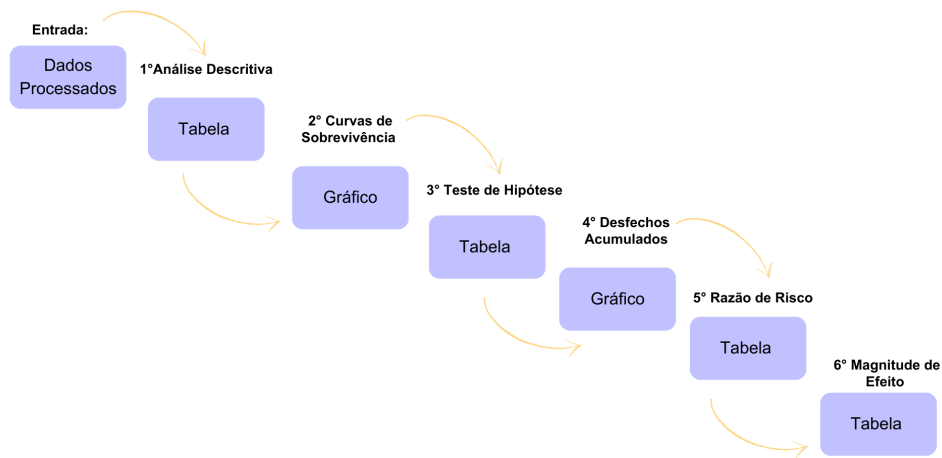
Fonte: Captura de tela do sistema SABIA (2026).

O fluxo operacional integra as entradas configuráveis à geração automática de indicadores. Essas entradas compreendem parâmetros institucionais e definições métricas, como a natureza do evento e o limite temporal. Uma vez processadas as seleções, o sistema gera saídas compostas por tabelas de caracterização e componentes gráficos interativos, desenvolvidos com a biblioteca *Plotly*², para inspeção visual das trajetórias. O encadeamento dos resultados constitui uma hierarquia que parte da estatística descritiva para os testes de hipótese e medidas de magnitude, conforme ilustrado na Figura 11.

¹ <https://streamlit.io/>

² <https://plotly.com/python/>

Figura 11 – Fluxograma da sequência das saídas analíticas na interface.



Fonte: Elaborado pelo autor (2026).

5.0.1 Parametrização e controles de execução

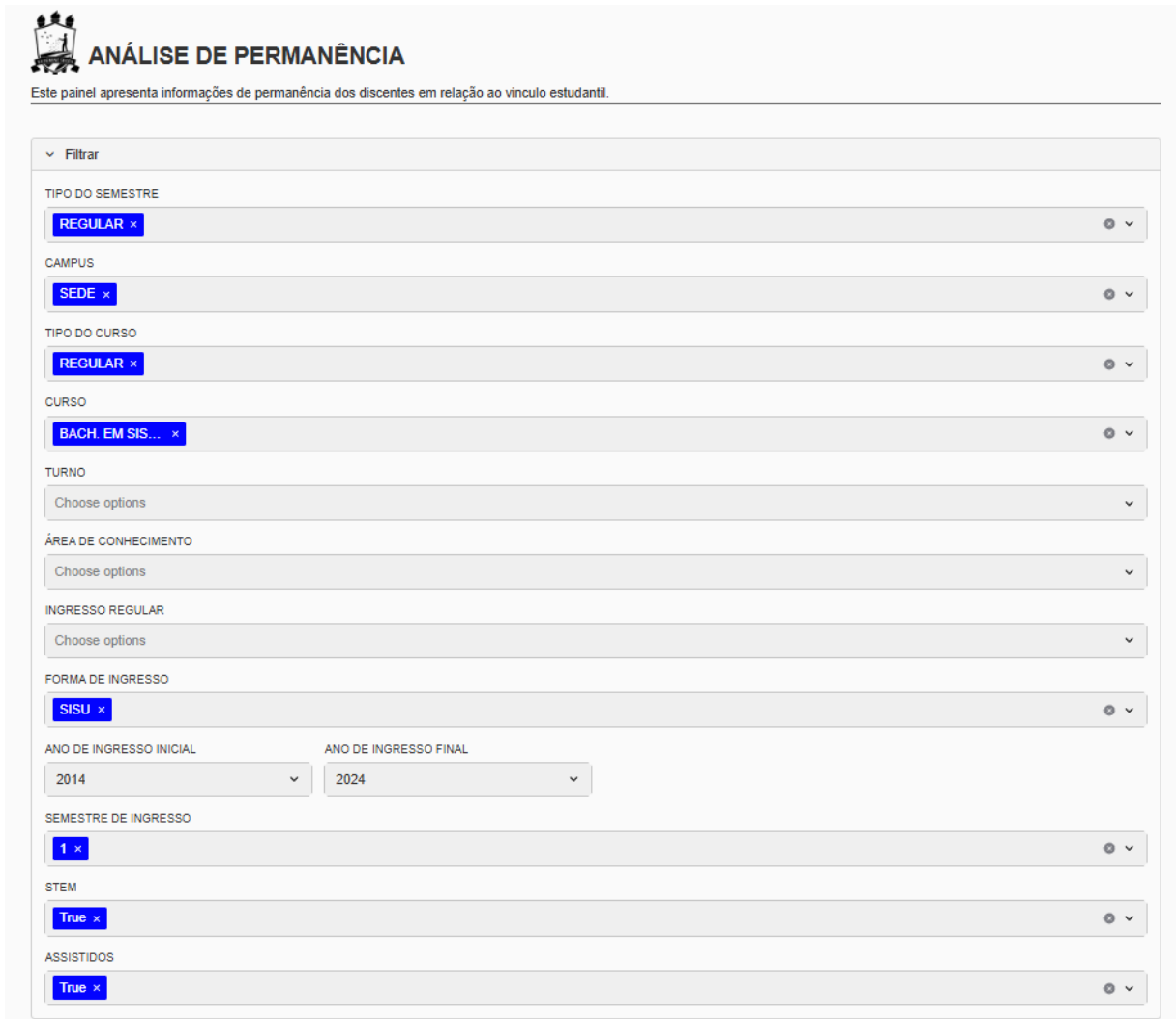
A configuração da análise fundamenta-se na seleção de filtros institucionais e na calibração do modelo estatístico. A função `create_controls` gera os componentes de interface necessários para a filtragem e a personalização dos dados em tempo real. Inicialmente, o usuário define o recorte da amostra por meio dos filtros gerais, conforme expresso na Figura 12, cujas variáveis e exemplos de seleção constam na Tabela 7. Na ausência de seleções, o sistema processa o cenário institucional total.

Tabela 7 – Detalhamento dos filtros gerais de entrada da ferramenta.

Variável	Tipo de Dado	Descrição	Exemplo de Seleção
Campus	Categórico	Localização física da unidade	SEDE
Curso	Categórico	Nome da graduação do discente	LIC. EM COMPUTAÇÃO
Turno	Categórico	Período de oferta das aulas	NOTURNO
Área	Categórico	Área de conhecimento CNPq	CIÊNCIAS EXATAS E DA TERRA
Forma de Ingresso	Categórico	Modalidade de entrada na IES	SISU
STEM	Booleano	Identificador de áreas tecnológicas	True
Assistidos	Booleano	Recebimento de auxílio estudantil	False
Ingresso Regular	Booleano	Filtro para Sisu e Vestibular	True
Ano Inicial	Inteiro	Intervalo de anos para recorte	2014
Ano Final	Inteiro	Intervalo de anos para recorte	2024
Semestre	Inteiro	Semestre letivo de entrada	1

Fonte: Elaborado pelo autor (2026).

Figura 12 – Painel de filtros gerais para segmentação da amostra institucional.



ANÁLISE DE PERMANÊNCIA

Este painel apresenta informações de permanência dos discentes em relação ao vínculo estudantil.

▼ Filtrar

TIPO DO SEMESTRE
REGULAR x

CAMPUS
SEDE x

TIPO DO CURSO
REGULAR x

CURSO
BACH. EM SIS... x

TURNOS
Choose options

ÁREA DE CONHECIMENTO
Choose options

INGRESSO REGULAR
Choose options

FORMA DE INGRESSO
SISU x

ANO DE INGRESSO INICIAL
2014

ANO DE INGRESSO FINAL
2024

SEMESTRE DE INGRESSO
1 x

STEM
True x

ASSISTIDOS
True x

Fonte: Captura de tela do sistema SABIA (2026).

Após a filtragem institucional, a interface disponibiliza os controles específicos para a execução do modelo. O seletor de evento representado na Figura 13 define o desfecho da análise entre as categorias *EVADIDO*, *FORMADO* ou a composição de ambos. Na seleção conjunta, o sistema registra o evento de falha para qualquer encerramento de vínculo, enquanto a censura é aplicada apenas aos discentes com matrícula ativa no período. Para o detalhamento gráfico, o seletor presente na Figura 14 habilita a exibição de intervalos de confiança e marcas de censura nos pontos de saída de dados.

O horizonte temporal é estabelecido pelo controle de duração detalhado na Figura 15, que limita o tempo de observação em períodos para o cálculo das estimativas. A segmentação da amostra ocorre pela escolha da variável de agrupamento ilustrada na Figura 16, que define os critérios de estratificação dos dados. Após essa definição, o controle de grupos exibido na Figura 17 permite selecionar os estratos específicos para a

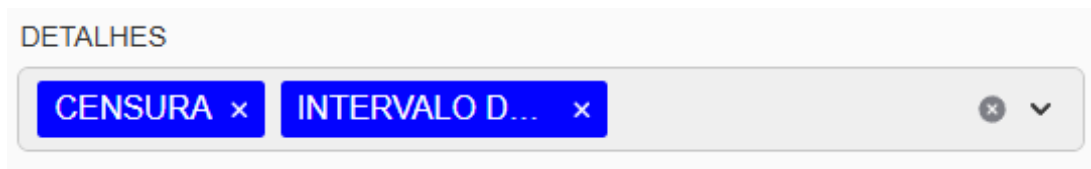
composição da análise comparativa. Por fim, o controle de grupo geral apresentado na Figura 18 permite a inclusão de uma curva de referência com a totalidade da amostra institucional para comparação com os grupos selecionados.

Figura 13 – Seletor de evento para definição do desfecho da análise.



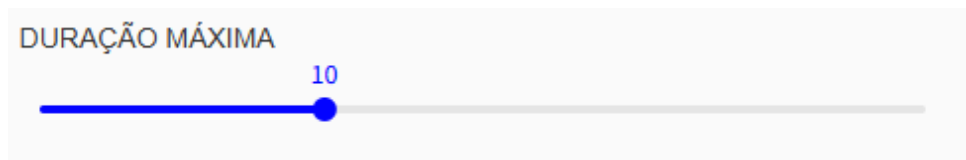
Fonte: Captura de tela do sistema SABIA (2026).

Figura 14 – Opções de visualização de intervalos e marcadores de censura.



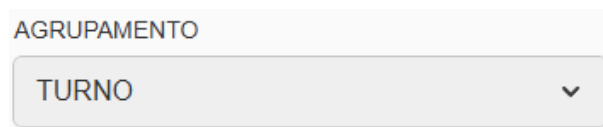
Fonte: Captura de tela do sistema SABIA (2026).

Figura 15 – Controle de duração para delimitação do horizonte temporal.



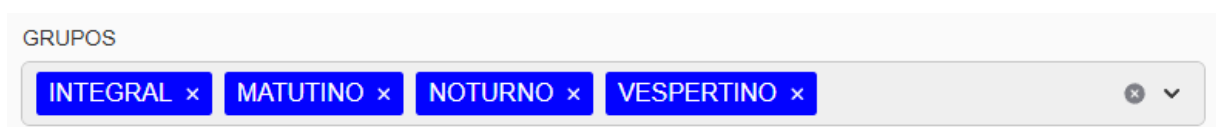
Fonte: Captura de tela do sistema SABIA (2026).

Figura 16 – Interface de seleção da variável de agrupamento.



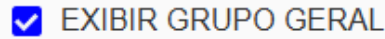
Fonte: Captura de tela do sistema SABIA (2026).

Figura 17 – Filtro secundário para seleção de grupos específicos.



Fonte: Captura de tela do sistema SABIA (2026).

Figura 18 – Seletor para exibição da curva de referência institucional.



Fonte: Captura de tela do sistema SABIA (2026).

Para segmentações de variáveis múltiplas, a ferramenta disponibiliza o agrupamento customizado via componente `data_editor`, que consiste em uma interface de tabela interativa para a inserção de nomes de grupos e lógicas de filtragem específicas por meio de *queries Python*, conforme exemplificado na Tabela 8. As instruções inseridas são capturadas pelo módulo `permanencia.py` e processadas pelo método `query()` da biblioteca *Pandas*, que realiza o recorte da base de dados antes do cálculo das estimativas de sobrevivência. A validação ocorre em tempo real e inconsistências na sintaxe interrompem o fluxo de cálculo.

Tabela 8 – Exemplos de agrupamentos customizados via consultas Python.

Nome do Grupo	Sintaxe da Query
Matutino & Assistido	<code>turno == 'MATUTINO' and is_assistido == True</code>
Vespertino & Assistido	<code>turno == 'VESPERTINO' and is_assistido == True</code>
Noturno & Assistido	<code>turno == 'NOTURNO' and is_assistido == True</code>
Integral & Assistido	<code>turno == 'INTEGRAL' and is_assistido == True</code>

Fonte: Elaborado pelo autor (2026).

5.0.2 Indicadores de permanência

A parametrização resulta no processamento dos registros para a geração de indicadores em três níveis. O primeiro consiste na tabela descritiva apresentada na Figura 19, que detalha o quantitativo de discentes, duração média de vínculo, desvio padrão e intervalo de confiança. A tabela informa a contagem de eventos, dados censurados e a mediana de sobrevivência estratificada pelos quartis $Q1$ e $Q3$, permitindo observar a distribuição temporal das saídas em diferentes pontos da amostra. O modelo trata dados censurados, o que mitiga o viés de exclusão de ativos e garante a integridade da análise. A mediana identifica o período em que 50% da amostra atingiu o desfecho, fornecendo um diagnóstico preliminar às hipóteses estatísticas.

Figura 19 – Tabela de análise descritiva gerada pela ferramenta.

▼ **Análise Descritiva Por Grupo**

A tabela abaixo apresenta informações descritivas a respeito dos dados agrupados por grupo. São exibidas as seguintes informações:

- **Mediana (Q1 e Q3):** Valores do primeiro (Q1) e terceiro quartil (Q3), indicando a dispersão dos tempos de formação.
- **Evento:** Número de eventos registrados no grupo.
- **Censurados:** Número de discentes cujo evento de interesse não ocorreu ou foi interrompido.
- **Intervalo de Confiança:** Intervalo de confiança (IC) de 95% para a média do tempo de formação.

ANÁLISE DESCRITIVA PARA EVADIDO									
	Grupo	Total de Discentes	Duração de Vínculo Média	Desvio Padrão	Intervalo de Confiança	Evento	Censurados	Mediana (Q1)	Mediana (Q3)
0	BACH. EM SISTEMAS DE INFORMAÇÃO (SEDE)	15	3.6667	2.0817	[1.3333; 5.3333]	3	12	2.5000	4.5000
1	GERAL	15	3.6667	2.0817	[1.3333; 5.3333]	3	12	2.5000	4.5000

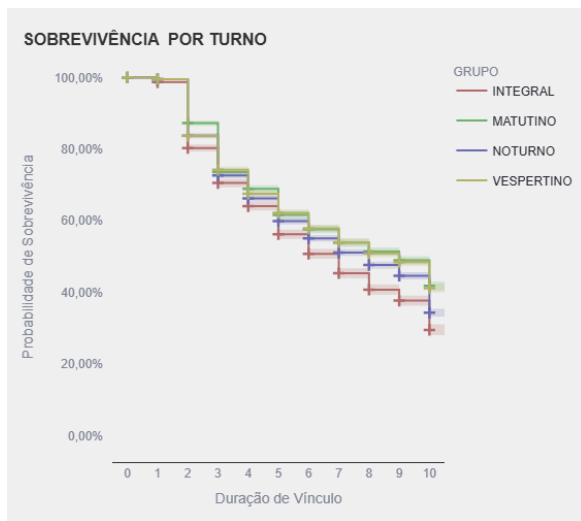
Fonte: Captura de tela do sistema SABIA (2026).

O segundo nível contempla a inspeção visual das trajetórias por meio dos painéis gráficos detalhados na Figura 20. Na curva de *Kaplan-Meier*, representada na Figura 20(a), o eixo vertical mensura a probabilidade de manutenção do vínculo institucional, com amplitude de 0% a 100%. O eixo horizontal indica a duração do vínculo em períodos semestrais, em que a inclinação da curva descreve a velocidade da evasão e o *plateau* representa a estabilidade da permanência residual ao final do horizonte observado. Marcadores em formato de cruz sinalizam a censura de dados, enquanto áreas hachuradas delimitam os intervalos de confiança das estimativas. A análise é complementada pelo risco acumulado de Nelson-Aalen exibido na Figura 20(b), além das curvas de conclusão detalhada na Figura 20(c) e de evasão na Figura 20(d), que identificam as janelas temporais de incidência para cada desfecho acadêmico.

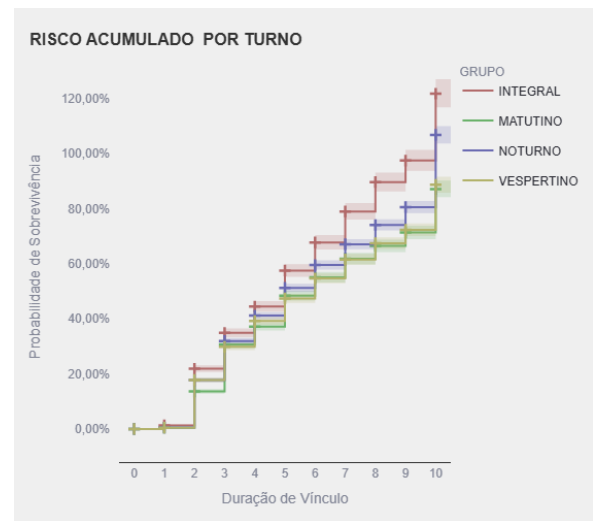
O terceiro nível compreende a validação estatística, consolidada pelas tabelas de resultados do teste *Log-Rank*, coeficientes de *Hazard Ratio* e magnitude de efeito, que são exibidas na Figura 21. O uso de métodos não paramétricos permite a análise das trajetórias educacionais sem a necessidade de assumir uma distribuição prévia para os dados, fundamentando a estatística dos resultados obtidos.

A significância da diferença entre as curvas é verificada pelos resultados do teste *Log-Rank* detalhados na Figura 21(a). A interpretação baseia-se no *p-value*: valores inferiores ao nível de significância adotado de 0,05 indicam que as curvas de sobrevivência dos grupos comparados não são equivalentes, permitindo a rejeição da hipótese nula de igualdade. A coluna *Iguais* consolida essa análise binária para facilitar a identificação de disparidades estatisticamente relevantes.

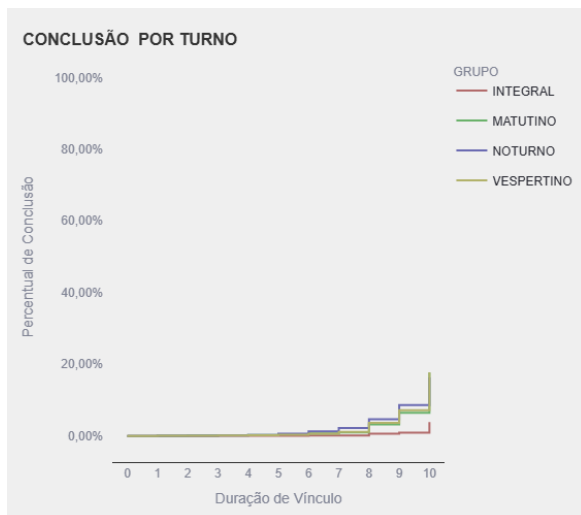
A razão de risco é quantificada pelo HR apresentado na Figura 21(b), fornece o coeficiente necessário para a comparação entre os pares de estratos institucionais. A



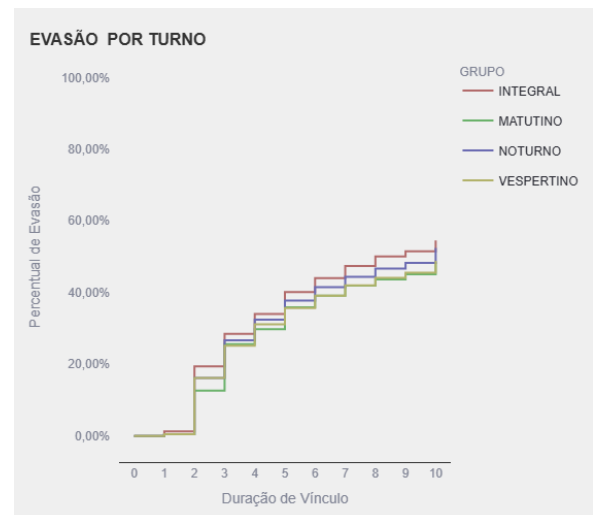
(a) Kaplan-Meier



(b) Nelson-Aalen



(c) Conclusão



(d) Evasão

Figura 20 – Visualização de sobrevivência, risco acumulado e probabilidades de desfecho.
 Fonte: Elaborado pelo autor (2026).

interpretação deste indicador fundamenta-se na relação com a unidade, em que um valor igual a 1 indica a ausência de diferença de risco entre os grupos comparados. Coeficientes superiores a 1 revelam que o grupo de teste apresenta uma probabilidade maior de sofrer o desfecho analisado, como a evasão, em relação ao grupo de referência. Por outro lado, valores inferiores a 1 sinalizam um efeito protetivo, indicando que o grupo de teste possui menor risco de ocorrência do evento quando comparado à base de referência institucional.

Para a análise da relevância prática, a ferramenta exhibe a magnitude de efeito, exemplificada pelo g de *Hedges* na Figura 21(c). Diferente do p -value, que aponta apenas a existência de diferença estatística, o g de *Hedges* quantifica a extensão dessa disparidade

em termos de desvios-padrão. A interpretação desta métrica fundamenta-se na coluna *Classificação*, que utiliza os limiares de Cohen para categorizar o impacto: valores próximos a 0,2 indicam um efeito pequeno, enquanto resultados em torno de 0,5 e acima de 0,8 representam efeitos médios e grandes, respectivamente. Esta abordagem permite determinar se a diferença encontrada possui relevância prática na trajetória acadêmica ou se o resultado estatístico é decorrente do volume da amostra observada.

A interface possibilita a alternância entre múltiplas métricas conforme os detalhes listados na Tabela 9, que incluem *d* de *Cohen*, correlação ponto-biserial, η^2 (*Eta-squared*), *Odds Ratio* e AUC, mediante a definição de um grupo de referência para o cálculo das magnitudes de impacto.

COMPARAÇÃO DA SOBREVIVÊNCIA (LOGRANK)				
	Referência	Grupo	Iguais	p-value
0	INTEGRAL	MATUTINO	NÃO	8.3849e-24
1	INTEGRAL	NOTURNO	NÃO	2.1175e-11
2	INTEGRAL	VESPERTINO	NÃO	2.1782e-19
3	MATUTINO	NOTURNO	NÃO	5.2874e-07
4	MATUTINO	VESPERTINO	SIM	0.0929
5	NOTURNO	VESPERTINO	NÃO	0.0006

(a) Resultados do teste Log-Rank.

ANÁLISE DE HAZARD RATIO					
	Grupo 1	Grupo 2	Coefficiente	Hazard Ratio	P-Valor
0	INTEGRAL	MATUTINO	-0.2780	0.7573	0.0000
1	INTEGRAL	NOTURNO	-0.1442	0.8657	0.0000
2	INTEGRAL	VESPERTINO	-0.2532	0.7763	0.0000
3	MATUTINO	NOTURNO	0.1394	1.1496	0.0000
4	MATUTINO	VESPERTINO	0.0236	1.0238	0.2716
5	NOTURNO	VESPERTINO	-0.1149	0.8914	0.0000

(b) Coeficientes de Hazard Ratio.

TAMANHO DO EFEITO

HEDGES' G

GRUPO DE REFERÊNCIA

NOTURNO

TAMANHO DO EFEITO

	Grupo 1	Grupo 2	Resultado	Classificação
0	NOTURNO	INTEGRAL	0.1677	Pequeno
1	NOTURNO	MATUTINO	0.0866	Pequeno
2	NOTURNO	VESPERTINO	0.0653	Pequeno

(c) Magnitude de efeito (*g* de Hedges).

Figura 21 – Indicadores de validação estatística e magnitude de efeito.

Fonte: Captura de tela do sistema SABIA (2026).

Tabela 9 – Métricas de magnitude de efeito disponíveis na ferramenta.

Métrica	Aplicação
d de Cohen	Diferença padronizada entre médias de sobrevivência
g de Hedges	Ajuste do d de Cohen para amostras pequenas ou desiguais
r_{pb} (Ponto-biserial)	Correlação entre variável binária e tempo de vínculo
η^2 (Eta-squared)	Proporção da variância explicada pelo agrupamento
Odds Ratio (OR)	Razão de chances de ocorrência do evento entre grupos
AUC (Área sob a curva)	Medida de separação entre as curvas de sobrevivência

Fonte: Elaborado pelo autor (2026).

5.0.3 Interpretabilidade e suporte à decisão

A interpretação dos resultados fundamenta-se na convergência das métricas geradas, funcionando como um mecanismo de explicabilidade para a identificação da intensidade e da precisão das evidências. A análise inicia-se pela inspeção visual das curvas de *Kaplan-Meier*, em que a inclinação e o distanciamento entre os segmentos indicam variações nas probabilidades de manutenção do vínculo institucional ao longo do tempo.

A precisão dessas estimativas é medida pela amplitude do intervalo de confiança (IC). Amplitudes estreitas indicam maior consistência dos dados observados, enquanto intervalos mais amplos sinalizam incerteza estatística, decorrente de amostras reduzidas ou alta variabilidade nas trajetórias. Na interface gráfica, a sobreposição das áreas de confiança entre diferentes estratos indica que os fatores de agrupamento podem não exercer influência distinta sobre a sobrevivência naquele ponto temporal específico.

A direção e a taxa do risco são quantificadas pela relação entre o coeficiente β do modelo de Cox e o *Hazard Ratio* ($HR = e^\beta$). O HR atua como o peso explicativo de cada variável: um coeficiente β positivo resulta em um $HR > 1,0$, o que indica que o atributo eleva a taxa de ocorrência do evento e atua como o fator que mais contribui para a elevação do risco na curva analisada. Inversamente, um β negativo ($HR < 1,0$) sinaliza um fator protetivo associado à maior permanência. A distância do HR em relação à unidade define a magnitude dessa influência: um HR de 1,25 representa um incremento de 25% no risco em relação ao grupo de referência institucional.

A integração final ocorre pela análise da magnitude de efeito (ex: g de Hedges), que

complementa a significância estatística ao mensurar a relevância prática da disparidade. Uma magnitude de efeito classificada como “grande” indica que a variável de agrupamento é o determinante principal na separação das trajetórias, independentemente do volume total de discentes. Assim, o conjunto de indicadores permite isolar a contribuição individual de cada fator e fundamentar a tomada de decisão institucional com base na relação entre risco, precisão e impacto real, utilizando a interpretabilidade intrínseca do modelo para justificar o comportamento das curvas observadas.

6 Resultados

Este capítulo apresenta os resultados obtidos por meio do modelo de análise de sobrevivência para mensurar padrões de permanência e evasão. A exposição dos dados estrutura-se em dois eixos. O primeiro eixo, intitulado Impacto da Mudança do PPC, analisa a transição entre matrizes curriculares nos cursos de Licenciatura em Computação (LC) e Bacharelado em Sistemas de Informação (BSI). O segundo eixo consiste na Avaliação Comparativa dos Indicadores de Permanência, estabelecendo o contraste entre o cenário institucional da UFRPE no campus SEDE e o recorte dos cursos de STEM.

Para as análises subsequentes detalhadas neste trabalho, o modelo foi parametrizado com o evento de falha definido como *Evasão* (desligamento do vínculo sem conclusão). O horizonte temporal foi fixado em 20 períodos semestrais para a análise do impacto da mudança do Projeto Pedagógico de Curso (PPC) e em 10 períodos para a comparação entre o cenário institucional e o recorte STEM. Os resultados oferecem suporte à gestão acadêmica e ao planejamento estratégico, identificando períodos associados ao abandono e padrões entre as trajetórias estudantis em diferentes contextos institucionais.

A forma de ingresso para que a comparação seja uniforme considerou discentes admitidos na modalidade regular, que inclui ingressos por Sisu ou Vestibular, desconsiderando categorias como transferência ou portadores de diploma para evitar distorções na amostra. O modelo utiliza a contagem sequencial de períodos letivos como unidade de tempo e, como a unidade de análise é o ciclo letivo independente do ano corrente, as interrupções de calendário referentes a 2020.1 não interferem nos resultados.

Os resultados fundamentam-se nas curvas de sobrevivência estimadas pelo método de Kaplan-Meier para determinar a probabilidade acumulada de permanência do discente ao longo do tempo. A validação das estimativas ocorre por meio de métricas estatísticas complementares com funções específicas. O teste de *Log-Rank* é aplicado para verificar a significância estatística, determinando se a separação entre as curvas resulta de disparidades reais entre os grupos ou se decorre de variações casuais na amostra. Complementarmente, os intervalos de confiança estabelecem a precisão dessas estimativas, delimitando a faixa de valores dentro da qual a verdadeira probabilidade de sobrevivência da população se encontra com um nível de confiança de 95%.

A intensidade da disparidade entre as categorias é mensurada pelo *Hazard Ratio*,

que quantifica a razão de risco para indicar a probabilidade de um grupo atingir o evento de evasão em relação a um estrato de referência. Adicionalmente, o tamanho do efeito *Hedges' g* avalia a relevância prática dos achados. Esta métrica permite distinguir diferenças estatisticamente significantes de impactos institucionais expressivos, quantificando a distância padronizada entre as trajetórias. Esta abordagem possibilita identificar as disparidades entre os grupos e os intervalos de desligamento associados a variáveis estruturais e socioeconômicas.

6.1 Impacto da mudança do projeto pedagógico de curso (PPC)

Esta seção analisa a transição entre as matrizes curriculares dos cursos de LC e BSI, cujas reformulações ocorreram em 2018.2 e 2020.1, respectivamente. O estudo utiliza uma janela de observação de 20 períodos letivos para o cálculo das funções de sobrevivência para o evento de EVASÃO.

A delimitação deste intervalo decorre da necessidade de estabelecer uma comparação entre os grupos de ingressantes antes e após as alterações nos projetos pedagógicos. O período total de 20 semestres é definido para viabilizar o acompanhamento equivalente ao tempo regular previsto para a conclusão (10 períodos) para os grupos de controle (PPC antigo). Para as coortes sob as novas matrizes, o acompanhamento não abrange a totalidade dos 20 períodos de observação devido à limitação cronológica imposta pela data de implementação dos novos PPCs em relação ao período final dos dados disponíveis.

O uso da metodologia de análise de sobrevivência nessa situação permite a comparação direta entre grupos com tempos de observação distintos. Através do tratamento de dados censurados à direita, o método integra as informações disponíveis das trajetórias recentes, viabilizando o contraste entre as probabilidades de permanência do novo currículo e o desempenho da matriz anterior em intervalos de tempo equivalentes, mesmo sem a integralização total do ciclo atual.

6.1.1 Análise descritiva da mudança do projeto pedagógico de curso (PPC)

A análise descritiva caracteriza o comportamento dos coortes de alunos sob a ótica da análise de sobrevivência, abrangendo o tempo de permanência no curso e a frequência

das censuras em cada currículo. Esta etapa estabelece o referencial para a interpretação das probabilidades de tempo de vínculo.

No curso de LC, a amostra compreende 550 discentes no grupo *Antes* e 442 no grupo *Depois*. Conforme os dados da Tabela 10, a média ($\bar{x}$) de permanência é de 5,53 períodos no grupo *Antes* e 4,21 no grupo *Depois*, com medianas ($\tilde{x}$) de 3,00 e 2,00. Quanto ao BSI, a amostra totaliza 796 discentes no grupo *Antes* e 317 no grupo *Depois*. Os tempos de permanência indicados na Tabela 10 revelam média de 5,55 períodos e mediana de 3,00 no grupo *Antes*, contra média de 3,46 e mediana de 2,00 no grupo *Depois*.

Tabela 10 – Estatísticas descritivas e indicadores de sobrevivência por curso e vigência de PPC.

Curso	n	Eventos	Censuras	$\bar{x}$ [IC 95%]	σ	Q_1	$\tilde{x}$	Q_3
LC	992	678	314					
Antes	550	471	79	5,5339 [5,2224; 5,8539]	3,3915	3,00	3,00	7,00
Depois	442	207	235	4,2174 [3,8791; 4,5652]	2,4919	2,00	2,00	6,00
BSI	1.113	641	472					
Antes	796	547	249	5,5503 [5,2431; 5,8447]	3,6823	3,00	3,00	7,00
Depois	317	94	223	3,4681 [3,1277; 3,7979]	1,6507	2,00	2,00	5,00

n : Total de discentes; $\bar{x}$: Média; σ : Desvio Padrão; $\tilde{x}$: Mediana; Q_n : Quartis.

Fonte: Elaborado pelo autor (2026). Tempos mensurados em semestres letivos.

Abreviações: LC: Licenciatura em Computação; BSI: Bacharelado em Sistemas de Informação; n : Total de discentes; $\bar{x}$: Média; σ : Desvio Padrão; $\tilde{x}$: Mediana; Q_n : Quartis; IC 95%: intervalo de Confiança de 95%.

Fonte: Elaborado pelo autor (2026). Tempos mensurados em semestres letivos.

A análise dos quartis permite observar a distribuição dos eventos de EVASÃO. No curso de LC, o 1º quartil (Q_1) reduziu de 3,00 para 2,00 períodos na transição para o grupo *Depois*, indicando que 25% dos desfechos ocorrem em intervalo iniciais. No curso de BSI, o Q_1 também se fixou em 2,00 períodos para o currículo reformulado. No que diz respeito à proporção de dados, as censuras em LC elevaram-se de 14,3% no grupo *Antes* para 53,1% no grupo *Depois*. Em BSI, a frequência de censuras supera a de eventos no grupo *Depois*, registrando 223 alunos vinculados contra 94 desfechos. Este incremento no volume de censuras indica a manutenção do vínculo discente (permanência) dentro da janela de observação disponível, caracterizando trajetórias acadêmicas que ainda não atingiram o desfecho de evasão ou conclusão.

Adicionalmente, a amplitude de permanência delimitada pelo 3º quartil (Q_3) indica que, no grupo *Antes* de LC, 75% da amostra permanecia vinculada por até 7 períodos, valor que se alterou para 6 no grupo *Depois*. No curso de BSI, o Q_3 passou de 7 períodos no grupo *Antes* para 5 períodos no grupo *Depois*. Esses indicadores, somados à redução do desvio padrão (σ), podem refletir a limitação do tempo de observação disponível para os currículos do grupo *Depois*.

6.1.2 Análise do curso: licenciatura em computação (LC)

As estimativas de sobrevivência para o curso de LC apresentam variações nas probabilidades de permanência conforme a vigência do PPC. De acordo com a Figura 22(a), os indicadores registram trajetórias de evasão com valores aproximados em ambos os grupos durante o 1º ano de vínculo acadêmico.

No intervalo inicial, até o 2º semestre, as probabilidades de sobrevivência apresentam valores próximos, registrando estimativas em torno de 84,00% para o grupo *Antes* e 83,67% para o grupo *Depois*. Os intervalo de confiança, representados visualmente pela área hachurada, correspondem a [80,66%; 86,81%] e [79,93%; 86,77%] respectivamente, conforme os dados da Tabela 11. A partir do 4º período, observa-se o distanciamento entre as curvas e o fim da sobreposição dos intervalos de confiança. Nesse ponto, o grupo *Depois* apresenta probabilidade de 69,22% e intervalo de confiança de [64,55%; 73,41%], enquanto o grupo *Antes* registra probabilidade de 60,18% e intervalo de confiança de [55,96%; 64,14%]. A ausência de intersecção entre os intervalo de confiança a partir deste marco indica uma

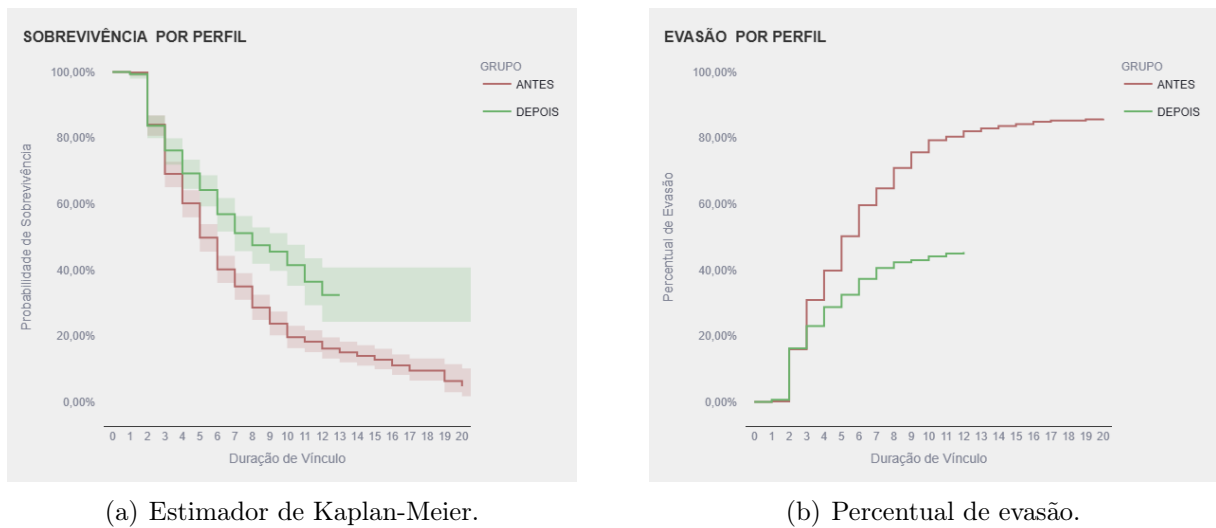


Figura 22 – Análise de sobrevivência e evasão do curso de LC ao longo do tempo.

distinção estatística entre as probabilidades de permanência, sinalizando que o coorte sob a nova matriz curricular retém um percentual superior de discentes vinculados em comparação ao currículo anterior.

Quanto a evasão, a Figura 22(b) demonstra que o grupo *Antes* atinge o patamar de 50% de desligamento no 5º período com aproximadamente 49,79%. Em contrapartida, no grupo *Depois*, essa estimativa de 50% de sobrevivência é deslocada para o intervalo entre o 7º e 11º período. No 10º período, a probabilidade de sobrevivência no grupo *Antes* é de aproximadamente 19,60%, ao passo que no grupo *Depois* o indicador situa-se próximo a 41,47%. No 13º período, a estimativa para o grupo *Depois* atinge cerca de 32,40%, valor que representa o dobro da probabilidade registrada no currículo do grupo *Antes* de 15,00%.

Ao final do horizonte de 20 períodos observado no grupo *Antes*, a estimativa de sobrevivência atinge o patamar de 4,75% e intervalo de confiança de [1,73%; 10,16%]. A comparação entre os indicadores demonstra que as alterações curriculares e de gestão acadêmica implementadas no projeto pedagógico do grupo *Depois* coincidem com a manutenção de probabilidades de sobrevivência superiores às registradas no grupo *Antes* em todos os períodos subsequentes ao 3º período de vínculo.

A Tabela 12 reforça as tendências observadas nas curvas de sobrevivência e evasão. O teste *Log-Rank* indica a rejeição da hipótese nula de igualdade entre as coortes, com p-valor de $1,1420e^{-75}$. O resultado valida a distinção estatística entre as funções de sobrevivência $S(t)$ das coortes *Antes* e *Depois*.

Tabela 11 – Probabilidade de sobrevivência e intervalo de Confiança (IC 95%) para o curso de LC por vigência de PPC.

Prd.	Antes		Depois	
	$S(t)$ (%)	IC 95%	$S(t)$ (%)	IC 95%
0	100,00	[100,00; 100,00]	100,00	[100,00; 100,00]
1	99,82	[98,72; 99,97]	99,34	[97,97; 99,79]
2	84,00	[80,66; 86,81]	83,67	[79,93; 86,77]
3	69,09	[65,04; 72,77]	76,17	[71,89; 79,89]
4	60,18	[55,96; 64,14]	69,22	[64,55; 73,41]
5	49,79	[45,54; 53,89]	64,19	[59,27; 68,69]
6	40,20	[36,08; 44,28]	56,91	[51,68; 61,80]
7	34,99	[31,00; 38,99]	51,15	[45,70; 56,33]
8	28,63	[24,88; 32,47]	47,49	[41,88; 52,88]
9	23,73	[20,23; 27,39]	45,54	[39,75; 51,14]
10	19,60	[16,33; 23,10]	41,47	[35,22; 47,60]
11	18,30	[15,11; 21,73]	36,45	[29,35; 43,56]
12	16,21	[13,15; 19,56]	32,40	[24,33; 40,71]
13	15,00	[12,02; 18,29]	32,40	[24,33; 40,71]
14	13,93	[11,02; 17,18]	–	–
15	12,80	[9,91; 16,07]	–	–
20	4,75	[1,73; 10,16]	–	–

A análise do impacto dessas mudanças é complementada pelos indicadores de risco. O modelo de riscos proporcionais revela um *Hazard Ratio* (HR) de 0,6332, indicando que, sob a vigência do novo projeto pedagógico, o risco de evasão dos discentes foi reduzido em 36,68% em relação ao currículo anterior. Esse indicador auxilia na compreensão da dinâmica de permanência, demonstrando que a reestruturação curricular incidiu sobre as frequências de desligamento dos estudantes.

A magnitude dessa variação é quantificada pelo coeficiente *Hedges'g* de 0,3359, valor que classifica o tamanho do efeito como médio. Esse indicador complementa a análise dos intervalos de confiança ao confirmar que a distância entre as probabilidades de permanência estabelece a redução da sobreposição entre as distribuições em favor do grupo *Depois*. A associação entre o tamanho do efeito e o distanciamento dos intervalos

de confiança a partir do quarto período ratifica o novo PPC como fator de influência na sobrevida estudantil em LC.

Tabela 12 – Indicadores estatísticos para o curso de LC.

Teste Log-Rank				
Grupo Ref.	Grupo Alt.	H_0	p-valor	
<i>Antes</i>	<i>Depois</i>	Rejeitado	1,1420e ⁻⁷⁵	

Modelo de Riscos Proporcionais de Cox				
Grupo Ref.	Grupo Alt.	β	HR	p-valor
<i>Antes</i>	<i>Depois</i>	-0,4569	0,6332	0,0000

Tamanho do Efeito (Hedges'g)			
Grupo Ref.	Grupo Alt.	Hedges'g	Classificação
<i>Antes</i>	<i>Depois</i>	0,3359	Médio

Notas: H_0 : Hipótese Nula de que as funções de sobrevivência são idênticas; **p-valor**: Probabilidade de significância (valores $< 0,05$ indicam diferença estatística); β : Coeficiente de regressão (valores negativos indicam redução do risco); **HR** (*Hazard Ratio*): Razão de riscos; indica que risco associado aos grupos; **Hedges'g**: Medida padronizada da magnitude da diferença entre os grupos.

Fonte: Elaborado pelo autor (2026).

Os indicadores em LC demonstram que o primeiro ano de vínculo constitui o momento mais crítico para o desligamento, com sobreposição dos intervalos de confiança até o segundo período. A distinção entre as matrizes ocorre a partir do terceiro período, quando o limite superior do grupo *Antes* [72,77%] torna-se inferior ao limite mínimo do grupo *Depois* [71,89%]. O *Hazard Ratio* de 0,6332 identifica o novo PPC como o fator de maior relevância na análise, reduzindo o risco de evasão em 36,68%. A magnitude do *Hedges'g* de 0,3359 confirma que o distanciamento entre as curvas reflete uma alteração estrutural nas trajetórias, com efeito médio. Conclui-se que o novo projeto pedagógico não altera o comportamento inicial de evasão, mas atua como suporte à permanência após o ciclo básico, resultando em probabilidades de sobrevivência superiores ao longo do fluxo curricular.

6.1.3 Análise do curso: bacharelado em sistemas de informação (BSI)

As estimativas de sobrevivência para o curso de BSI, conforme ilustrado na Figura 23(a), revelam variações nas probabilidades de permanência estudantil entre as vigências do PPC. As curvas de sobrevivência sugerem um comportamento inicial similar entre os grupos no 1º ano, seguido por um distanciamento em favor do projeto pedagógico atual.

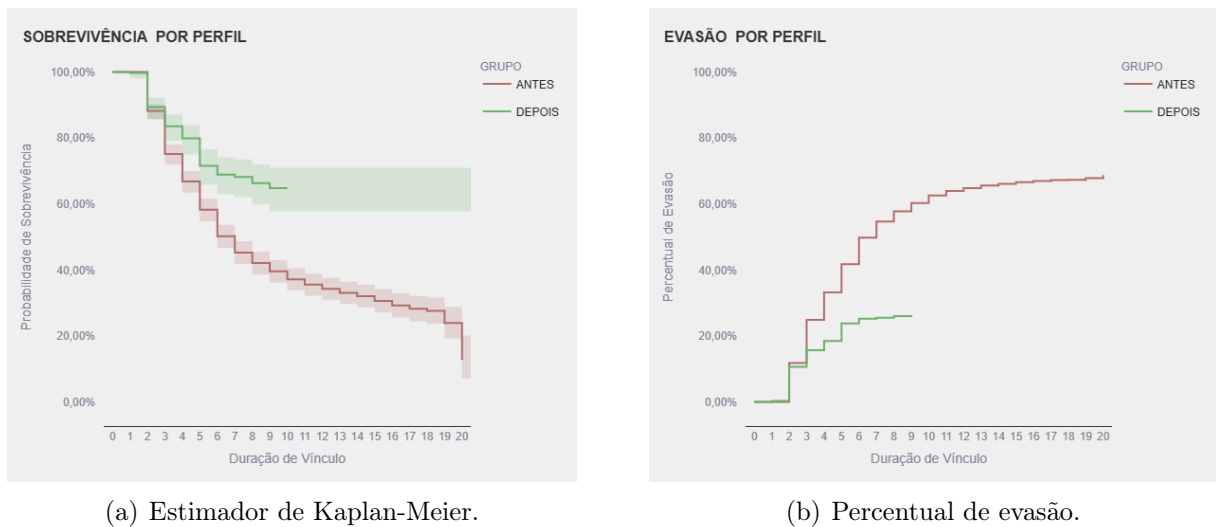


Figura 23 – Análise de sobrevivência e evasão do curso de BSI ao longo do tempo.

Até o 2º período, as probabilidades de sobrevivência apresentam valores próximos, registrando cerca de 88,19% de probabilidade de sobrevivência e intervalo de confiança de [85,74%; 90,24%] para o grupo *Antes* e probabilidade em torno de 89,33%, com intervalo de confiança de [85,63%; 92,12%] para o grupo *Depois*, conforme expresso na Tabela 13. A intersecção entre os intervalos de confiança evidencia a equivalência estatística entre as trajetórias iniciais de ambos os grupos.

A partir do 3º período, observa-se a interrupção da sobreposição entre as estimativas de permanência. Nesse estágio, o grupo *Depois* sustenta uma sobrevivência de 83,54%, enquanto o grupo *Antes* apresenta 75,13%, com intervalo de confiança de [79,13%; 87,10%] e [71,97%; 77,98%], respectivamente. No 4º período, essa disparidade se amplia, registrando-se 79,85% de discentes vinculados no grupo *Depois* contra 66,82% no corte anterior; os limites de confiança correspondentes situam-se em [75,02%; 83,84%] e [63,43%; 69,97%].

Quanto à evasão, a Figura 23(b) demonstra que o grupo *Antes* atinge o patamar

de 50% de desligamento entre o 6° e o 7° período, registrando aproximadamente 50,21% ao final do 6° semestre de vínculo acadêmico. Em contrapartida, no grupo *Depois*, a probabilidade de sobrevivência situa-se em 64,80% no 10° período, patamar superior ao registrado no grupo *Antes* para o mesmo intervalo. Ao atingir o horizonte de 20 períodos, a sobrevivência no grupo *Antes* alcança o valor de 12,75%.

Tabela 13 – Probabilidade de sobrevivência e intervalo de Confiança (IC 95%) para o curso de BSI por vigência de PPC.

Prd.	<i>Antes</i>		<i>Depois</i>	
	<i>S(t)</i> (%)	IC 95%	<i>S(t)</i> (%)	IC 95%
0	100,00	[100,00; 100,00]	100,00	[100,00; 100,00]
1	100,00	[100,00; 100,00]	99,72	[98,03; 99,96]
2	88,19	[85,74; 90,24]	89,33	[85,63; 92,12]
3	75,13	[71,97; 77,98]	83,54	[79,13; 87,10]
4	66,82	[63,43; 69,97]	79,85	[75,02; 83,84]
5	58,26	[54,75; 61,60]	71,56	[65,88; 76,46]
6	50,21	[46,68; 53,63]	68,87	[62,89; 74,08]
7	45,28	[41,79; 48,70]	68,17	[62,09; 73,49]
8	42,12	[38,66; 45,52]	66,30	[59,80; 72,00]
9	39,56	[36,14; 42,94]	64,80	[57,72; 70,99]
10	37,18	[33,81; 40,55]	64,80	[57,72; 70,99]
11	35,55	[32,20; 38,91]	–	–
12	34,30	[30,95; 37,67]	–	–
13	33,06	[29,70; 36,46]	–	–
14	32,06	[28,66; 35,50]	–	–
15	30,62	[27,11; 34,19]	–	–
20	12,75	[7,11; 20,10]	–	–

As tendências descritas são corroboradas estatisticamente pelos testes apresentados na Tabela 14. A rejeição da hipótese nula (H_0) resultante do teste *Log-Rank*, com um p-valor de $9,8017e^{-190}$, indica que a probabilidade das funções de sobrevivência serem idênticas é inferior ao nível de significância estabelecido. Esse resultado confirma a distinção estatística entre as curvas de permanência dos grupos *Antes* e *Depois*.

Adicionalmente, o modelo de riscos proporcionais de Cox quantifica a magnitude da disparidade entre os grupos. O coeficiente β de $-0,6241$ e o *Hazard Ratio* (HR) de $0,5357$ indicam que, sob a vigência do novo projeto pedagógico, o risco de evasão dos discentes foi reduzido em $46,43\%$ em relação ao currículo anterior.

A intensidade da variação observada entre as trajetórias é mensurada pelo coeficiente *Hedges'g*. O valor de $0,6818$ classifica o tamanho do efeito como grande, fornecendo a base estatística para o distanciamento das funções de sobrevivência evidenciado na Figura 23(a). Esta magnitude de efeito explica a configuração dos intervalos de confiança, uma vez que a ausência de sobreposição entre os limites de probabilidade a partir do 3º período decorre da amplitude da diferença entre as estimativas de permanência. Essa convergência entre a distinção das curvas, a redução de risco via HR e o tamanho de efeito calculado indica que a reestruturação do curso de BSI atua como fator de influência na permanência estudantil, retardando o tempo até a ocorrência do desligamento.

Tabela 14 – Indicadores estatísticos para o curso de BSI.

Teste Log-Rank				
Grupo Ref.	Grupo Alt.	H_0	p-valor	
<i>Antes</i>	<i>Depois</i>	Rejeitado	9,8017e ⁻¹⁹⁰	

Modelo de Riscos Proporcionais de Cox				
Grupo Ref.	Grupo Alt.	β	HR	p-valor
<i>Antes</i>	<i>Depois</i>	-0,6241	0,5357	0,0000

Tamanho do Efeito (Hedges'g)			
Grupo Ref.	Grupo Alt.	Hedges'g	Classificação
<i>Antes</i>	<i>Depois</i>	0,6818	Grande

Notas: H_0 : Hipótese Nula de que as funções de sobrevivência são idênticas; **p-valor**: Probabilidade de significância (valores $< 0,05$ indicam diferença estatística); β : Coeficiente de regressão (valores negativos indicam redução do risco); **HR** (*Hazard Ratio*): Razão de riscos entre os grupos (valores < 1 indicam redução de risco para o grupo de alteração); **Hedges'g**: Medida padronizada da magnitude da diferença entre os grupos.

Fonte: Elaborado pelo autor (2026).

A análise dos indicadores de BSI revela que, embora o primeiro ano apresente equivalência nas trajetórias, o novo projeto pedagógico estabelece uma divergência nas

probabilidades de sobrevivência. A interrupção na sobreposição dos intervalos de confiança ocorre no terceiro período, momento em que a distância entre o limite mínimo do grupo *Depois* [79,13%] e o limite máximo do grupo *Antes* [77,98%] confirma a distinção estatística entre as coortes. O *Hazard Ratio* de 0,5357 posiciona a reestruturação curricular como o determinante para a redução de 46,43% no risco de desligamento. Essa transformação é validada pelo *Hedges'g* de 0,6818, cujo efeito grande justifica o distanciamento entre as curvas e a manutenção de patamares de sobrevivência de 64,80% até o décimo período. Conclui-se que o novo PPC de BSI atua como um fator de proteção que retarda o momento do desfecho de evasão e altera a configuração do vínculo discente em comparação à matriz anterior.

6.2 Análise comparativa de permanência: cenário institucional e cursos de STEM

Esta seção investiga as trajetórias acadêmicas e o impacto do gênero, do turno e da assistência estudantil no evento de EVASÃO, contrastando os dados da UFRPE com o recorte da área STEM. A análise identifica disparidades socioeconômicas para fornecer subsídios à gestão acadêmica. O trabalho utiliza dados de discentes ingressantes entre 2014 e 2024, com uma janela de observação de 10 períodos letivos, intervalo que corresponde ao tempo regulamentar de integralização da maioria das graduações analisadas.

O recorte de STEM compreende os cursos de Licenciatura em Matemática (MAT), Licenciatura em Física (LF), Licenciatura em Computação (LC), Bacharelado em Sistemas de Informação (BSI) e Bacharelado em Ciência da Computação (BCC). A partir desse universo amostral, os dados foram segmentados nos agrupamentos de Turno (*MATUTINO*, *VESPERTINO* e *NOTURNO*), Gênero (*FEMININO* e *MASCULINO*) e Assistência Estudantil (*SIM* e *NÃO*).

6.2.1 Análise descritiva do cenário institucional

O panorama institucional da UFRPE, no período observado, compreende um universo de 36.895 matrículas ativas em cursos de graduação. Deste total, registraram-se 16.023 casos de evasão e 20.872 casos de censura. A duração média de permanência do

estudante na universidade é de 5,5 períodos letivos, com um intervalo de confiança de [5,4669; 5,5261]. Tais indicadores, detalhados na Tabela 15, estabelecem a base para a análise das trajetórias acadêmicas na instituição.

Tabela 15 – Caracterização sociodemográfica institucional (2014-2024)

Variável	Frequência (n)	Percentual (%)
Total Institucional	36.895	100,0%
<i>Gênero</i>		
Feminino	16.464	44,6%
Masculino	20.431	55,4%
<i>Assistência Estudantil</i>		
Assistido	858	2,3%
Não Assistido	36.037	97,7%

Fonte: Elaborado pelo autor (2026). Dados: UFRPE.

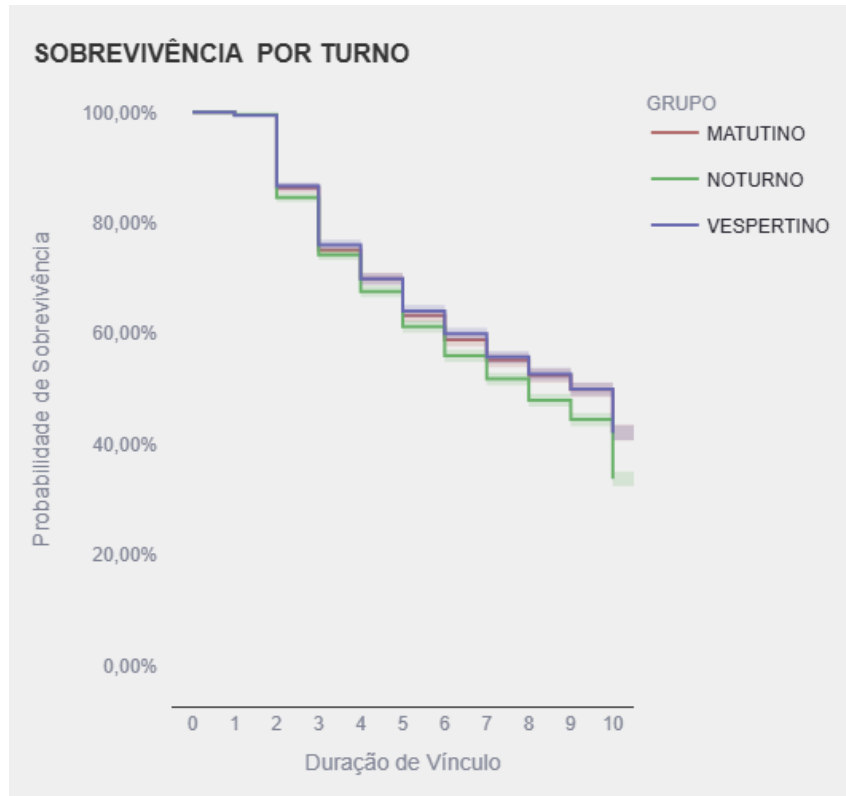
Fonte: Elaborado pelo autor (2025). Dados extraídos do sistema acadêmico da UFRPE.
Nota: O total institucional compreende os cursos de graduação presenciais das unidades Sede (Recife).

A composição por gênero indica uma distribuição de 44,6% de discentes do gênero *FEMININO* e 55,4% do gênero *MASCULINO*. Sobre a assistência estudantil, a instituição tem 2,3% (858) de discentes beneficiários de algum tipo de assistência estudantil. De forma geral, o desligamento por evasão atinge 43,4% do total de discentes, o que mostra que a saída de alunos acontece em toda a universidade, antes mesmo de olhar para os cursos separadamente.

6.2.2 Caracterização do cenário institucional

A Figura 24 e a Tabela 16 apresentam a distribuição das estimativas de sobrevivência na instituição, segmentadas pelos turnos *MATUTINO*, *VESPERTINO* e *NOTURNO*.

A trajetória longitudinal revela comportamentos distintos na ocorrência do evento de evasão entre as modalidades *MATUTINO* e *NOTURNO* ao longo dos dez períodos analisados.



(a) Estimador de Kaplan-Meier.

Figura 24 – Análise de sobrevivência na instituição, por turno.

Nos primeiros quatro períodos, as estimativas de permanência apresentam proximidade estatística entre os três grupos. No 4º período, os turnos *MATUTINO* com 69,95%, *VESPERTINO* com 69,92% e *NOTURNO* com 67,65% apresentam intervalos de confiança de [68,83%; 71,04%], [68,89%; 70,93%] e [66,61%; 68,66%], respectivamente, e registram sobreposição dos intervalos de confiança. A partir do 5º período, o turno *NOTURNO* distancia-se dos demais; enquanto os turnos *MATUTINO* e *VESPERTINO* mantêm probabilidades de 63,29% com intervalo de confiança de [62,10%; 64,46%] e 64,09% com intervalo de confiança de [63,00%; 65,16%], o grupo *NOTURNO* registra 61,31% com intervalo de confiança de [60,22%; 62,38%]. Esta divergência amplia-se no 8º período, no qual o turno *NOTURNO* registra 47,99% de sobrevivência, atingindo o limiar de 50% antes das modalidades diurnas.

O momento em que a probabilidade de permanência atinge 50% evidencia a

Tabela 16 – Probabilidade de sobrevivência e intervalo de Confiança (IC 95%) para o cenário institucional por turno.

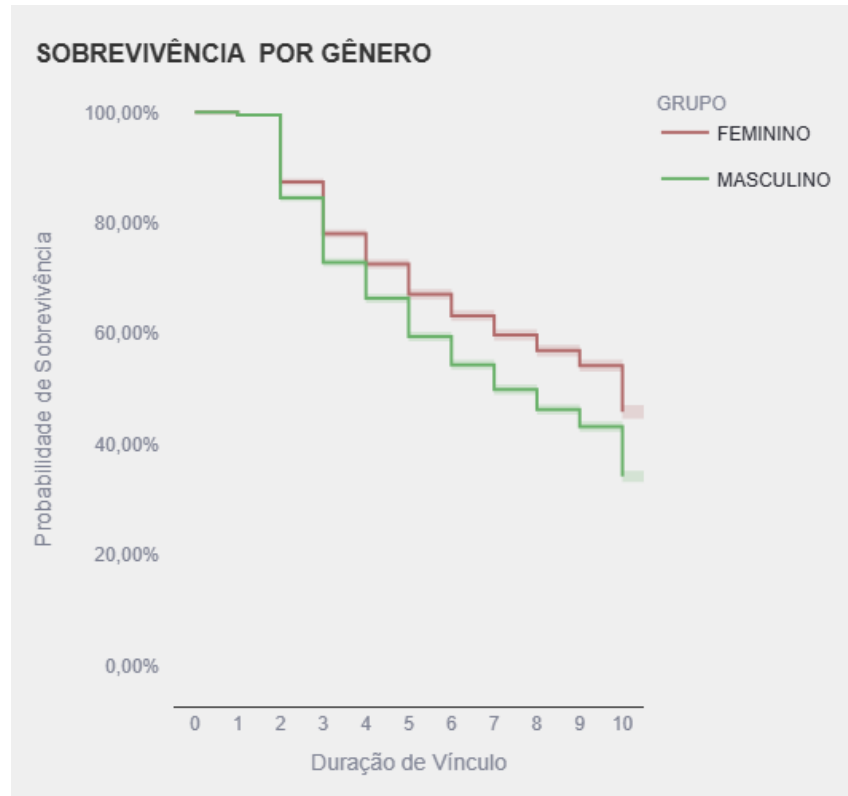
Prd.	Matutino		Vespertino		Noturno	
	$S(t)$ (%)	IC 95%	$S(t)$ (%)	IC 95%	$S(t)$ (%)	IC 95%
0	100,00	[100,00%; 100,00%]	100,00	[100,00%; 100,00%]	100,00	[100,00%; 100,00%]
1	99,56	[99,37%; 99,69%]	99,48	[99,29%; 99,61%]	99,58	[99,41%; 99,69%]
2	86,35	[85,52%; 87,14%]	86,68	[85,92%; 87,40%]	84,60	[83,81%; 85,35%]
3	75,15	[74,09%; 76,16%]	76,08	[75,13%; 77,00%]	74,29	[73,33%; 75,22%]
4	69,95	[68,83%; 71,04%]	69,92	[68,89%; 70,93%]	67,65	[66,61%; 68,66%]
5	63,29	[62,10%; 64,46%]	64,09	[63,00%; 65,16%]	61,31	[60,22%; 62,38%]
6	58,93	[57,70%; 60,14%]	59,99	[58,86%; 61,10%]	56,04	[54,91%; 57,16%]
7	55,25	[53,99%; 56,50%]	55,80	[54,64%; 56,95%]	51,84	[50,69%; 52,99%]
8	52,44	[51,16%; 53,71%]	52,74	[51,55%; 53,91%]	47,99	[46,81%; 49,15%]
9	49,90	[48,59%; 51,20%]	49,98	[48,76%; 51,19%]	44,51	[43,31%; 45,71%]
10	42,17	[40,75%; 43,58%]	42,11	[40,78%; 43,43%]	33,78	[32,48%; 35,09%]

disparidade temporal entre as modalidades. Para o turno *NOTURNO*, esse valor é alcançado entre o 7° com 51,84% das estimativas e o 8° período com 47,99%. Nos turnos *MATUTINO* e *VESPERTINO*, situam-se no limiar de 50% no 9° período, cruzando-o definitivamente no 10° período.

A distinção entre as curvas é verificada pela ausência de intersecção entre o limite superior do intervalo de confiança de 95% do turno *NOTURNO* e os limites inferiores dos turnos *MATUTINO* e *VESPERTINO* a partir do 8° período. Nesse estágio, o limite superior do turno *NOTURNO* é de 49,15%, valor inferior aos limites inferiores registrados pelos turnos *MATUTINO* de 51,16% e *VESPERTINO* de 51,55%. Ao final da janela de 10 períodos, a estimativa de sobrevivência para o *NOTURNO* é de 33,78% com intervalo de confiança de [32,48%; 35,09%], o que representa uma diferença de 8,39 pontos percentuais em relação aos turnos *MATUTINO* com 42,17% e *VESPERTINO* com 42,11%.

Ao expandir a análise para o agrupamento de gênero, a estratificação apresentada na Figura 25 e na Tabela 17 indica variações nas probabilidades de sobrevivência entre os grupos *FEMININO* e *MASCULINO*.

A trajetória das estimativas revela que as probabilidades de sobrevivência decaem



(a) Estimador de Kaplan-Meier por Gênero.

Figura 25 – Análise de sobrevivência na instituição, por gênero.

Tabela 17 – Probabilidade de sobrevivência e intervalo de confiança (IC 95%) por gênero no cenário institucional.

Prd.	<i>FEMININO</i>		<i>MASCULINO</i>	
	$S(t)$ (%)	IC 95%	$S(t)$ (%)	IC 95%
0	100,00	[100,00%; 100,00%]	100,00	[100,00%; 100,00%]
1	99,53	[99,38%; 99,64%]	99,54	[99,41%; 99,64%]
2	87,46	[86,80%; 88,08%]	84,57	[83,94%; 85,17%]
3	78,09	[77,27%; 78,89%]	72,89	[72,11%; 73,65%]
4	72,61	[71,72%; 73,48%]	66,40	[65,57%; 67,22%]
5	67,15	[66,20%; 68,09%]	59,53	[58,66%; 60,40%]
6	63,28	[62,29%; 64,25%]	54,40	[53,50%; 55,29%]
7	59,80	[58,78%; 60,81%]	49,93	[49,02%; 50,84%]
8	56,98	[55,93%; 58,01%]	46,35	[45,42%; 47,27%]
9	54,27	[53,19%; 55,34%]	43,24	[42,29%; 44,18%]
10	45,90	[44,68%; 47,12%]	34,24	[33,25%; 35,24%]

para ambos os grupos no intervalo entre o 1° e o 2° período. Para o grupo *FEMININO*, a probabilidade altera-se de 99,53% com intervalo de confiança de [99,38%; 99,64%] para 87,46% com intervalo de confiança de [86,80%; 88,08%], representando uma variação de 12,07 pontos percentuais nas estimativas de sobrevivência. No grupo *MASCULINO*, a variação no mesmo intervalo é de 14,97 pontos percentuais, passando de probabilidades de permanência de 99,54% com intervalo de confiança de [99,41%; 99,64%] para 84,57% com intervalo de confiança de [83,94%; 85,17%].

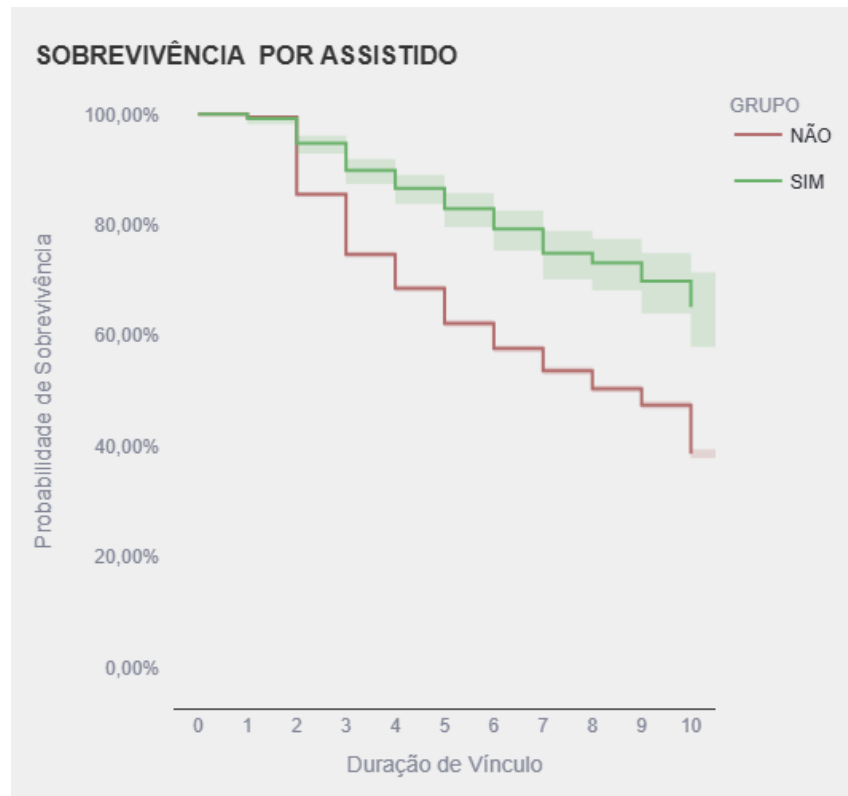
Quanto ao limiar de 50% de probabilidade de permanência, observa-se uma disparidade temporal entre as categorias. O grupo *MASCULINO* atinge este patamar no 7° período com 49,93% e intervalo de confiança de [49,02%; 50,84%]. Em contraste, o grupo *FEMININO* mantém-se acima do limite de 50% durante a observação do evento até o 9° período com 54,27% e intervalo de confiança de [53,19%; 55,34%], cruzando esta marca no 10° período, no qual registra 45,90% com intervalo de confiança de [44,68%; 47,12%].

A partir do 2° período, os limites superiores do intervalo de confiança para o grupo *MASCULINO* não interceptam os limites inferiores do grupo *FEMININO*, reforçando a distinção entre as trajetórias. Ao final de dez períodos, a diferença entre as estimativas de sobrevivência dos dois grupos é de 11,66 pontos percentuais.

Adicionalmente, ao considerar a assistência estudantil como fator de influência na permanência, a Figura 26 e a Tabela 18 evidenciam o comportamento das estimativas para discentes beneficiários de algum tipo de auxílio estudantil, grupo *SIM* e os não assistidos, grupo *NÃO*.

As estimativas de sobrevivência indicam variações distintas entre o 1° e o 2° período para ambos os grupos. No grupo *NÃO*, a probabilidade altera-se de 99,55% com intervalo de confiança de [99,45%; 99,63%] para 85,54% com intervalo de confiança de [85,07%; 85,99%], o que representa uma diferença de 14,01 pontos percentuais. Para os discentes assistidos *SIM*, a variação no mesmo intervalo é de 4,45 pontos percentuais, alterando-se de 99,21% com intervalo de confiança de [98,26%; 99,65%] para 94,76% com intervalo de confiança de [92,92%; 96,13%].

No que se refere ao alcance do patamar de 50% de probabilidade de sobrevivência, observa-se que o grupo *NÃO* atinge esse patamar no 8° período com 50,41% e intervalo de confiança de [49,70%; 51,11%]. Em contrapartida, o grupo *SIM* mantém-se acima do



(a) Estimador de Kaplan-Meier por Assistência.

Figura 26 – Análise de sobrevivência na instituição, por assistência estudantil.

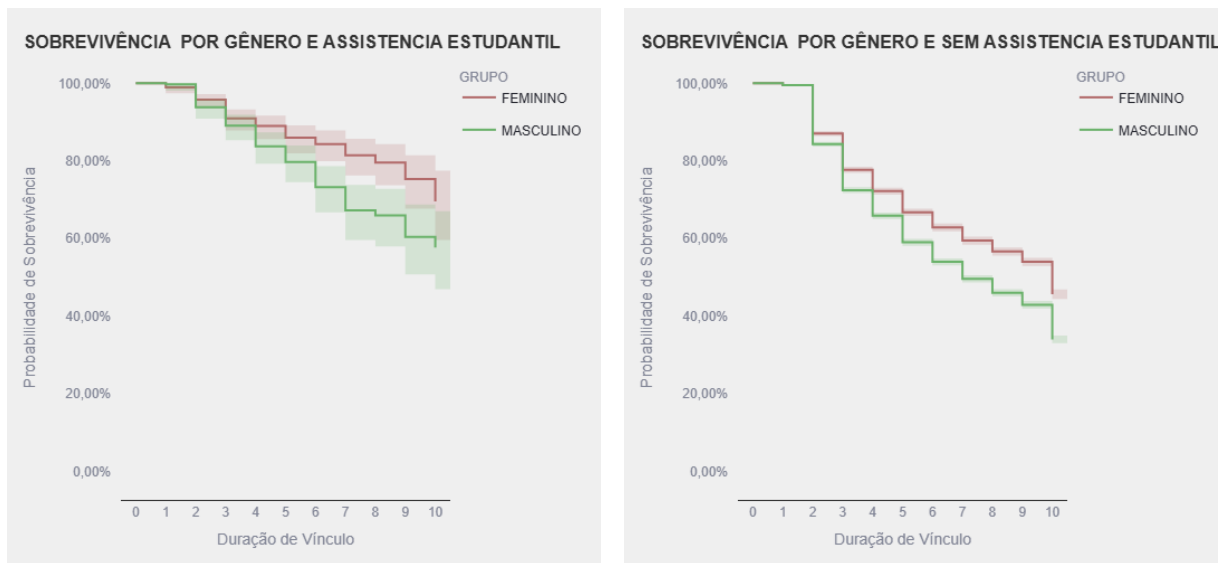
Tabela 18 – Probabilidade de sobrevivência e intervalo de Confiança (IC 95%) por Assistência no cenário institucional.

Prd.	Não		Sim	
	$S(t)$ (%)	IC 95%	$S(t)$ (%)	IC 95%
0	100,00	[100,00%; 100,00%]	100,00	[100,00%; 100,00%]
1	99,55	[99,45%; 99,63%]	99,21	[98,26%; 99,65%]
2	85,54	[85,07%; 85,99%]	94,76	[92,92%; 96,13%]
3	74,69	[74,11%; 75,25%]	89,90	[87,46%; 91,89%]
4	68,56	[67,94%; 69,18%]	86,64	[83,84%; 88,99%]
5	62,27	[61,61%; 62,92%]	82,99	[79,69%; 85,81%]
6	57,69	[57,01%; 58,36%]	79,30	[75,40%; 82,64%]
7	53,67	[52,98%; 54,36%]	74,89	[70,26%; 78,91%]
8	50,41	[49,70%; 51,11%]	73,19	[68,23%; 77,50%]
9	47,48	[46,76%; 48,20%]	69,86	[64,03%; 74,94%]
10	38,70	[37,92%; 39,48%]	65,20	[58,02%; 71,47%]

limite de 50% durante todos os dez períodos analisados, encerrando a janela de observação com uma probabilidade de sobrevivência de 65,20% e intervalo de confiança de [58,02%; 71,47%].

A diferenciação entre as curvas é corroborada pela ausência de intersecção entre os intervalo de confiança a partir do 2º período. Essa separação estatística reflete-se ao final do 10º período, com a diferença entre as estimativas de sobrevivência atingindo aproximadamente 26,50 pontos percentuais, indicando uma probabilidade de permanência superior para o grupo vinculado a programas de assistência.

Para analisar as interações entre gênero e assistência estudantil, as estimativas são detalhadas nas Figuras 27(a), 27(b) e na Tabela 19, revelando comportamentos distintos nas trajetórias de sobrevivência quando os grupos são segmentados pela vulnerabilidade socioeconômica.



(a) Gênero - Assistidos

(b) Gênero - Não Assistidos

Figura 27 – Comparativo de sobrevivência por gênero segmentado por assistência estudantil.

No grupo de discentes assistidos, as variações nas estimativas entre o 1º e o 2º período apresentam-se em 2,94% para o grupo *FEMININO*, passando de 98,78% com intervalo de confiança de [97,1%; 99,5%] para 95,84% com intervalo de confiança de [93,4%; 97,4%]. Para o grupo *MASCULINO*, a variação é de 6,22 pontos percentuais, alterando-se de 99,72% com intervalo de confiança de [98,0%; 100,0%] para 93,50% com intervalo de confiança de [90,4%; 95,6%]. No que se refere ao alcance do patamar de 50%

Tabela 19 – Análise longitudinal comparativa: probabilidades de sobrevivência por gênero e assistência (0-10)

Prd.	Assistidos				Não Assistidos			
	<i>FEMININO</i>		<i>MASCULINO</i>		<i>FEMININO</i>		<i>MASCULINO</i>	
	<i>S(t) (%)</i>	IC 95%	<i>S(t) (%)</i>	IC 95%	<i>S(t) (%)</i>	IC 95%	<i>S(t) (%)</i>	IC 95%
0	100,00	[100,0%; 100,0%]	100,00	[100,0%; 100,0%]	100,00	[100,0%; 100,0%]	100,00	[100,0%; 100,0%]
1	98,78	[97,1%; 99,5%]	99,72	[98,0%; 100,0%]	99,56	[99,4%; 99,7%]	99,54	[99,4%; 99,6%]
2	95,84	[93,4%; 97,4%]	93,50	[90,4%; 95,6%]	87,11	[86,4%; 87,8%]	84,32	[83,7%; 84,9%]
3	90,60	[87,2%; 93,1%]	89,10	[85,3%; 92,0%]	77,59	[76,7%; 78,4%]	72,46	[71,7%; 73,2%]
4	88,70	[85,0%; 91,5%]	84,21	[79,6%; 87,8%]	71,99	[71,1%; 72,9%]	65,94	[65,1%; 66,8%]
5	85,58	[81,2%; 89,0%]	79,92	[74,6%; 84,3%]	66,48	[65,5%; 67,4%]	59,04	[58,2%; 59,9%]
6	84,38	[79,7%; 88,1%]	73,00	[66,3%; 78,6%]	62,57	[61,6%; 63,6%]	53,96	[53,0%; 54,9%]
7	81,32	[75,7%; 85,7%]	66,74	[58,8%; 73,5%]	59,10	[58,1%; 60,1%]	49,53	[48,6%; 50,4%]
8	79,33	[73,1%; 84,3%]	65,41	[57,1%; 72,5%]	56,28	[55,2%; 57,3%]	45,94	[45,0%; 46,9%]
9	77,84	[70,9%; 83,3%]	59,80	[49,9%; 68,4%]	53,58	[52,5%; 54,7%]	42,86	[41,9%; 43,8%]
10	71,85	[62,1%; 79,5%]	56,96	[45,9%; 66,6%]	45,25	[44,0%; 46,5%]	33,89	[32,9%; 34,9%]

Fonte: Elaborado pelo autor (2026).

de probabilidade de sobrevivência, observa-se que ambos os gêneros assistidos mantêm-se acima deste limite durante todo o período de observação, encerrando o 10° período com 71,85% com intervalo de confiança de [62,1%; 79,5%] para o grupo *FEMININO* e 56,96% com intervalo de confiança de [45,9%; 66,6%] para o grupo *MASCULINO*.

Em contraste, no cenário de discentes não assistidos, as variações entre o 1° e o 2° período registram 12,45 pontos percentuais para o grupo *FEMININO* e 15,22 pontos percentuais para o *MASCULINO*. No limiar de 50%, o grupo *MASCULINO* não assistido cruza essa marca entre o 6° período, com estimativa de 53,96% e intervalo de confiança de [53,0%; 54,9%], e o 7° período, com 49,53% e intervalo de confiança de [48,6%; 50,4%]. O grupo *FEMININO* não assistido atinge patamar semelhante entre o 9° período, com 53,58% e intervalo de confiança de [52,5%; 54,7%], e o 10° período, com 45,25% e intervalo de confiança de [44,0%; 46,5%].

A diferenciação entre as trajetórias de gênero é corroborada pela ausência

de sobreposição dos intervalos de confiança em períodos mais avançados, sendo o distanciamento mais evidente no grupo não assistido. Ao final da janela de dez períodos, a diferença de sobrevivência entre gêneros no grupo assistido é de 14,89 pontos percentuais, enquanto no grupo não assistido essa diferença é de 11,36 pontos percentuais, evidenciando que a vantagem na probabilidade de permanência feminina persiste independentemente do suporte financeiro institucional.

A validação das diferenças observadas nas curvas de sobrevivência foi realizada por meio do teste de hipótese *Log-Rank*. Os resultados apresentados na Tabela 20 estão ordenados pelos p-valores para demonstrar as variáveis com maior impacto na trajetória discente. Os testes indicam que a assistência estudantil é o fator de diferenciação estatística com os menores p-valores. As comparações entre assistidos e não assistidos, segmentadas por gênero e turno, apresentam p-valores situados entre $1,5350e^{-106}$ para o grupo *MASCULINO* e $7,4072e^{-58}$ para o turno *NOTURNO*. Esses valores resultam na rejeição da hipótese nula (H_0) de igualdade entre as funções de sobrevivência e indicam que o suporte institucional altera a trajetória dos estudantes.

Os resultados indicam que a assistência estudantil associa-se à equalização das estimativas entre os grupos. No cenário geral, a diferença entre os gêneros apresenta um p-valor de $7,6669e^{-18}$, indicando maior evasão no grupo *MASCULINO*. No entanto, ao isolar os discentes assistidos, o teste *Log-Rank* não rejeita a hipótese de igualdade, com p-valor de 0,4750. Esse dado indica que o auxílio institucional reduz a disparidade de gênero, com probabilidades de permanência semelhantes entre os grupos *FEMININO* e *MASCULINO*. Raciocínio semelhante aplica-se aos turnos de oferta. Entre os estudantes que não recebem assistência, o turno *NOTURNO* apresenta trajetória de sobrevivência distinta do *MATUTINO*, com p-valor de $4,4300e^{-2}$. No grupo de assistidos, essa diferença não é verificada, apresentando $p = 0,4479$. Os dados sugerem que a assistência atua na redução das variações verificadas entre os turnos.

Tabela 20 – Teste de hipótese *Log-Rank* para os Grupos Gerais e Combinações.

Grupo de Ref.	Grupo de Alt.	H_0	p-valor
Assistido, Geral	Não Assistido, Geral	Rejeitado	$1,9262e^{-190}$
Assistido, Geral, Masculino	Não Assistido, Geral, Masculino	Rejeitado	$1,5350e^{-106}$
Assistido, Geral, Feminino	Não Assistido, Geral, Feminino	Rejeitado	$3,5420e^{-97}$
Assistido, Geral, Vespertino	Não Assistido, Geral, Vespertino	Rejeitado	$1,0285e^{-72}$
Assistido, Geral, Matutino	Não Assistido, Geral, Matutino	Rejeitado	$4,2035e^{-65}$
Assistido, Geral, Noturno	Não Assistido, Geral, Noturno	Rejeitado	$7,4072e^{-58}$
Feminino, Geral	Masculino, Geral	Rejeitado	$7,6669e^{-18}$
Masculino, Geral, Não Assistido	Feminino, Geral, Não Assistido	Rejeitado	$5,3315e^{-16}$
Masculino, Geral, Noturno	Feminino, Geral, Noturno	Rejeitado	$2,1592e^{-10}$
Masculino, Geral, Matutino	Feminino, Geral, Matutino	Rejeitado	$5,2225e^{-7}$
Masculino, Geral, Vespertino	Feminino, Geral, Vespertino	Rejeitado	$2,0000e^{-4}$
Matutino, Geral, Não Assistido	Noturno, Geral, Não Assistido	Rejeitado	$4,4300e^{-2}$
Matutino, Geral, Não Assistido	Vespertino, Geral, Não Assistido	Não Rejeitado	$8,1600e^{-2}$
Matutino, Geral	Vespertino, Geral	Não Rejeitado	$9,3700e^{-2}$
Matutino, Geral	Noturno, Geral	Não Rejeitado	$1,5730e^{-1}$
Noturno, Geral, Assistido	Vespertino, Geral, Assistido	Não Rejeitado	$3,7410e^{-1}$
Matutino, Geral, Assistido	Noturno, Geral, Assistido	Não Rejeitado	$4,4790e^{-1}$
Feminino, Geral, Assistido	Masculino, Geral, Assistido	Não Rejeitado	$4,7500e^{-1}$
Noturno, Geral, Não Assistido	Vespertino, Geral, Não Assistido	Não Rejeitado	$6,9680e^{-1}$
Vespertino, Geral	Noturno, Geral	Não Rejeitado	$8,6300e^{-1}$
Matutino, Geral, Assistido	Vespertino, Geral, Assistido	Não Rejeitado	$9,4960e^{-1}$

Nota: H_0 indica a hipótese nula de igualdade entre as funções de sobrevivência. Hipótese rejeitada para p-valor $< 0,05$.

Os resultados do modelo de riscos proporcionais presentes na Tabela 21 indicam que o gênero e o turno exercem influências distintas na probabilidade de evasão. No cenário geral, o gênero masculino apresenta um risco 34,40% superior ao *FEMININO* apresentando teste *Hazard Ratio* de 1,3440, com p-valor de 0,0000, padrão que se mantém nos recortes de turno e entre os discentes não assistidos. Em relação aos turnos, o período noturno registra um risco de evasão 16,25% maior quando comparado ao *MATUTINO* com HR de 1,1625 e $p = 0,0000$.

Por outro lado, a assistência estudantil atua como o principal fator de proteção contra o risco de interrupção do vínculo. Uma maior redução de risco é observada no turno matutino, onde o grupo assistido apresenta 36,33% do risco de evasão do grupo não assistido com $HR = 0,3633$ e p-valor de 0,0000. De maneira semelhante, o suporte institucional reduz o risco em 60,75% para as mulheres com HR de 0,3925 e 53,30% para os homens com HR de 0,4670, o que demonstra que a política de assistência consegue mitigar as vulnerabilidades associadas ao perfil do estudante, independentemente da categoria analisada.

Quando mensurada a magnitude desses efeitos utilizando o *Effect-Size* (Hedges' g), observa-se na Tabela 22 que a assistência estudantil apresenta os maiores índices de impacto prático no modelo. Todas as comparações diretas entre discentes assistidos e não assistidos resultam em tamanhos de efeito classificados como médios, com valores de $g = 0,3379$ para o turno *VESPERTINO* e $g = 0,3272$ para o gênero *FEMININO*. Esses valores confirmam que, para além da significância estatística, a assistência institucional exerce uma força de diferenciação sobre a permanência dos estudantes.

As variáveis de gênero e turno possuem tamanhos de efeito pequenos. A magnitude fora do eixo da assistência ocorre na comparação entre os turnos *VESPERTINO* e *NOTURNO*, com $g = 0,1191$, enquanto a diferença geral entre os gêneros *FEMININO* e *MASCULINO* registra $g = 0,1008$. Tais resultados indicam que, apesar de existirem disparidades demográficas e de horário, a intensidade dessas diferenças é reduzida quando comparada ao impacto exercido pelo recebimento de auxílios estudantis.

Tabela 21 – *Hazard Ratio* e coeficientes de risco para as combinações gerais.

Grupo Ref. (1)	Grupo Alt. (2)	(β)	HR	p-valor
Masculino, Geral, Matutino	Feminino, Geral, Matutino	0,2959	1,3443	0,0000
Feminino, Geral	Masculino, Geral	0,2956	1,3440	0,0000
Masculino, Geral, Não Assistido	Feminino, Geral, Não Assistido	0,2865	1,3317	0,0000
Masculino, Geral, Noturno	Feminino, Geral, Noturno	0,2855	1,3304	0,0000
Masculino, Geral, Vespertino	Feminino, Geral, Vespertino	0,2695	1,3093	0,0000
Matutino, Geral, Assistido	Vespertino, Geral, Assistido	0,2509	1,2852	0,2211
Matutino, Geral, Assistido	Noturno, Geral, Assistido	0,2495	1,2833	0,2712
Matutino, Geral	Noturno, Geral	0,1506	1,1625	0,0000
Matutino, Geral, Não Assistido	Noturno, Geral, Não Assistido	0,1431	1,1538	0,0000
Noturno, Geral, Assistido	Vespertino, Geral, Assistido	0,0121	1,0122	0,9529
Matutino, Geral	Vespertino, Geral	-0,0101	0,9900	0,6713
Matutino, Geral, Não Assistido	Vespertino, Geral, Não Assistido	-0,0130	0,9870	0,5848
Noturno, Geral, Não Assistido	Vespertino, Geral, Não Assistido	-0,1533	0,8562	0,0000
Vespertino, Geral	Noturno, Geral	-0,1597	0,8524	0,0000
Feminino, Geral, Assistido	Masculino, Geral, Assistido	-0,4656	0,6216	0,0048
Não Assistido, Geral, Vespertino	Assistido, Geral, Vespertino	-0,7476	0,4735	0,0000
Não Assistido, Geral, Masculino	Assistido, Geral, Masculino	-0,7614	0,4670	0,0000
Não Assistido, Geral	Assistido, Geral	-0,8777	0,4158	0,0000
Não Assistido, Geral, Noturno	Assistido, Geral, Noturno	-0,8842	0,4131	0,0000
Não Assistido, Geral, Feminino	Assistido, Geral, Feminino	-0,9352	0,3925	0,0000
Não Assistido, Geral, Matutino	Assistido, Geral, Matutino	-1,0125	0,3633	0,0000

Nota: HR > 1 indica risco aumentado para o Grupo Alt. em relação ao Ref.; HR < 1 indica efeito protetivo. β representa o coeficiente de regressão de Cox.

Tabela 22 – Hedges’g para as comparações entre grupos gerais e combinações.

Grupo de Ref.	Grupo de Alt.	Hedges’g	Tam. do efeito
Assistido, Geral, Vespertino	Não Assistido, Geral, Vespertino	0,3379	Médio
Assistido, Geral, Feminino	Não Assistido, Geral, Feminino	0,3272	Médio
Assistido, Geral, Noturno	Não Assistido, Geral, Noturno	0,3166	Médio
Assistido, Geral, Masculino	Não Assistido, Geral, Masculino	0,3144	Médio
Não Assistido, Geral	Assistido, Geral	0,3113	Médio
Assistido, Geral, Matutino	Não Assistido, Geral, Matutino	0,2941	Médio
Feminino, Geral, Assistido	Masculino, Geral, Assistido	0,1374	Pequeno
Vespertino, Geral	Noturno, Geral	0,1191	Pequeno
Matutino, Geral, Assistido	Noturno, Geral, Assistido	0,1140	Pequeno
Masculino, Geral, Vespertino	Feminino, Geral, Vespertino	0,1127	Pequeno
Masculino, Geral, Não Assistido	Feminino, Geral, Não Assistido	0,1043	Pequeno
Feminino, Geral	Masculino, Geral	0,1008	Pequeno
Masculino, Geral, Matutino	Feminino, Geral, Matutino	0,0901	Pequeno
Masculino, Geral, Noturno	Feminino, Geral, Noturno	0,0804	Pequeno
Noturno, Geral, Não Assistido	Vespertino, Geral, Não Assistido	0,0776	Pequeno
Matutino, Geral	Noturno, Geral	0,0701	Pequeno
Matutino, Geral, Assistido	Noturno, Geral, Assistido	0,0681	Pequeno
Matutino, Geral, Não Assistido	Vespertino, Geral, Não Assistido	0,0445	Pequeno
Matutino, Geral, Assistido	Vespertino, Geral, Assistido	0,0444	Pequeno
Matutino, Geral, Não Assistido	Noturno, Geral, Não Assistido	0,0072	Pequeno
Matutino, Geral	Vespertino, Geral	0,0024	Pequeno

Nota: Classificação do tamanho do efeito baseada nos limiares de Hedges’g: Pequeno ($\leq 0,2$), Médio ($\leq 0,5$) e Grande ($\leq 0,8$). No contexto educacional, valores próximos a 0,3 são considerados relevantes.

Os indicadores de permanência institucional indicam que a evasão é estruturada por condicionantes socioeconômicos, nos quais a assistência estudantil apresenta a maior magnitude de efeito com $g = 0,3113$ e atua como o principal fator de proteção com $HR = 0,4158$. A análise demonstra que o suporte institucional neutraliza disparidades de gênero e turno, visto que o teste Log-Rank não aponta diferença estatística entre assistidos com $p = 0,4750$. Em oposição, os grupos de maior risco concentram-se no gênero *MASCULINO* e no turno *NOTURNO*, com destaque para a combinação Masculino, Geral e Matutino que apresenta $HR = 1,3443$. A maior amplitude de sobrevivência ocorre no turno matutino, com 65,20% para assistidos contra 38,70% para não assistidos ao final de dez períodos.

Ao comparar o risco relativo com o tamanho do efeito (*Hedges'g*), observa-se que a assistência estudantil é o determinante de maior peso no modelo. Em contraste, as variáveis de gênero e turno apresentam baixa relevância prática, indicada por valores de $g \approx 0,10$, sugerindo que seu papel na permanência é menos expressivo. A disparidade entre a significância estatística do turno *NOTURNO* com $p = 0,0000$ e sua baixa força de impacto com $g = 0,0701$ indica que o horário de oferta não é a causa principal da evasão nesta etapa. Os resultados demonstram que o risco de desligamento não é um atributo fixo do discente, dado que o suporte financeiro apresenta maior capacidade de alteração da sobrevivência acadêmica com $HR = 0,4158$ do que as características como o gênero, que registra $HR = 1,3440$. Assim, a manutenção do vínculo institucional vincula-se à redução de riscos socioeconômicos por meio das políticas de assistência.

6.2.3 Análise descritiva do cenário em cursos de STEM

Segmentando a análise para o recorte dos cursos de STEM, detalhados na Tabela 23, observa-se um universo de 7.432 discentes, o que representa aproximadamente 20% do contingente total da universidade. O conjunto de dados abrange os agrupamentos de gênero, assistência estudantil e turno.

Em relação à composição demográfica, a amostra de STEM apresenta 5.893 registros do gênero masculino, o que corresponde a 79,7% do grupo, enquanto o gênero feminino conta com 1.539 registros, representando 20,3% do total. No que se refere à assistência estudantil, 158 discentes são beneficiários de auxílios, totalizando 2,1% da amostra, ao

passo que a maioria, composta por 7.274 estudantes, não possui vínculo com programas de assistência, perfazendo 97,9% do recorte. Tais proporções divergem do cenário institucional geral.

Tabela 23 – Indicadores de vínculo por curso em STEM.

Curso	Total Discentes	Duração Média [IC 95%]	Eventos	Censurados
BCC	1.323	6,2260 [6,0491; 6,3954]	552	771
BSI	1.360	6,0662 [5,8934; 6,2250]	676	684
LC	1.217	5,1980 [5,0353; 5,3698]	706	511
LF	1.271	4,9670 [4,8135; 5,1267]	803	468
MAT	2.261	4,9478 [4,8284; 5,0672]	1.394	867
INSTITUIÇÃO	36.895	5,4965 [5,4669; 5,5261]	16.023	20.872

Fonte: Adaptado de (CAVALCANTI et al., 2025).

Siglas: BCC: Bacharelado em Ciência da Computação; BSI: Bacharelado em Sistemas de Informação; LC: Licenciatura em Computação; LF: Licenciatura em Física; MAT: Licenciatura em Matemática.

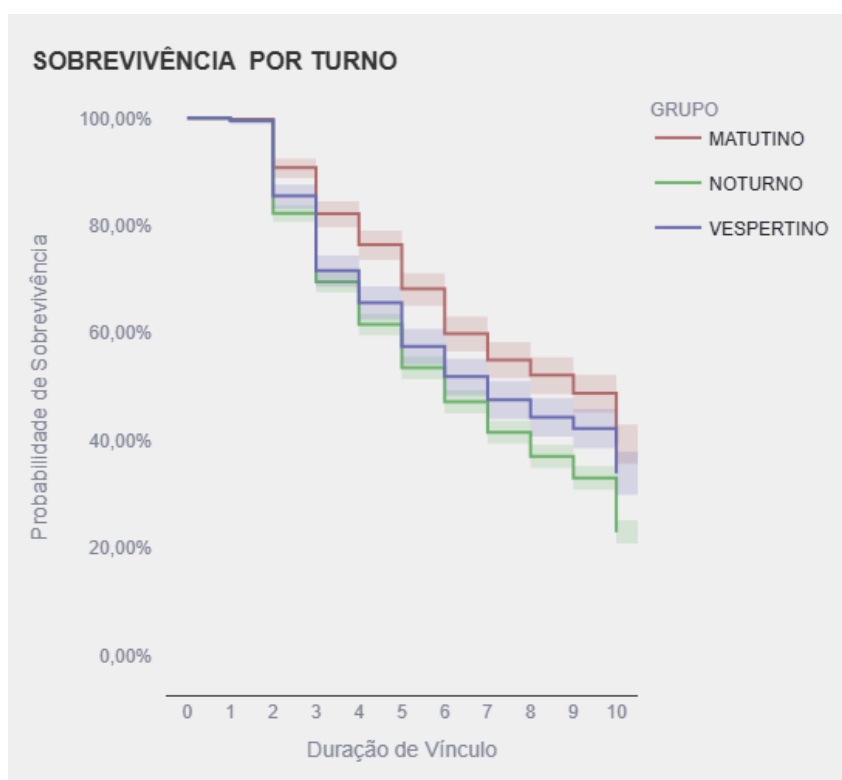
Nota: IC 95% representa o intervalo de Confiança de 95%; Eventos indicam o desligamento (evasão) e Censurados indicam a permanência ou conclusão.

Os dados indicam que os cursos de Bacharelado (BCC e BSI) apresentam maior duração média acadêmica. Nestes cursos, o número de registros censurados é superior ao número de eventos de evasão. No curso de BCC, registram-se 771 censuras e 552 eventos, com duração média de permanência de 6,22 períodos. Esse valor é superior à média institucional, que registra 5,49 períodos.

Em contrapartida, as Licenciaturas apresentam maior concentração de eventos de interrupção de vínculo, com o volume de eventos de evasão superando o de censuras em todas as modalidades do grupo. Na Licenciatura em Matemática, o total de 1.394 eventos é 60,8% superior aos 867 registros de censura. Na LF, essa diferença atinge 71,6%, com o registro de 803 eventos para 468 censuras, enquanto em LC a incidência de 706 evasões é 38,2% superior aos 511 casos de permanência ou conclusão. Observa-se que a duração média nessas licenciaturas é inferior a 5 períodos, o que indica uma interrupção de vínculo em estágio anterior ao verificado nos bacharelados analisados.

6.2.4 Caracterização do cenário em cursos de STEM

Inicialmente, a análise consiste na investigação das funções de sobrevivência por turno, utilizando o estimador de *Kaplan-Meier* para fundamentar as observações entre os grupos. A análise indica que o tempo de vínculo em *STEM* apresenta comportamentos distintos conforme o horário de oferta dos cursos. O estimador de *Kaplan-Meier* revela o turno *MATUTINO* com probabilidades de sobrevivência superiores, seguido pelo *VESPERTINO*, enquanto o turno *NOTURNO* registra os menores valores de probabilidade de sobrevivência ao longo de todo o intervalo de tempo acompanhado. Mediante a análise dos períodos de transição representado no Gráfico 28 e detalhados na



(a) Estimador de Kaplan-Meier.

Figura 28 – Análise de sobrevivência nos cursos de STEM, por turno.

Tabela 24, identifica-se que o turno *NOTURNO* atinge a probabilidade de sobrevivência de 50% entre o 5º período, no qual apresenta 53,61% com intervalo de confiança de [51,50%; 55,67%], e o 6º período, com 47,28% e limites situados entre [45,13%; 49,39%]. Os turnos *MATUTINO* e *VESPERTINO* atingem patamar semelhante apenas ao final do tempo de acompanhamento; o grupo *VESPERTINO* cruza o limiar de 50% entre o 9º período, com 50,21% e intervalo de confiança de [47,62%; 52,74%], e o 10º período. Já

o turno *MATUTINO* apresenta 48,81% no 9º período, porém com intervalo de confiança de [45,28%; 52,25%], indicando o cruzamento definitivo da marca de 50% apenas no 10º período. No 2º período de acompanhamento, a probabilidade de sobrevivência para o turno *NOTURNO* é de 82,33% com parâmetros de [80,74%; 83,80%], enquanto o turno *MATUTINO* apresenta 90,82% com intervalo de confiança de [88,84%; 92,46%]. A ausência de sobreposição entre os intervalo de confiança nestes pontos indica que a diferença nas funções de sobrevivência entre esses turnos é estatisticamente significativa neste intervalo de tempo.

Tabela 24 – Probabilidade de sobrevivência e intervalo de Confiança (IC 95%) para o cenário de STEM por turno.

Prd.	Matutino		Vespertino		Noturno	
	$S(t)$ (%)	IC 95%	$S(t)$ (%)	IC 95%	$S(t)$ (%)	IC 95%
0	100,00	[100,00; 100,00]	100,00	[100,00; 100,00]	100,00	[100,00; 100,00]
1	99,80	[99,20; 99,95]	99,63	[99,23; 99,82]	99,59	[99,23; 99,78]
2	90,82	[88,84; 92,46]	88,66	[87,14; 90,00]	82,33	[80,74; 83,80]
3	82,26	[79,71; 84,53]	78,22	[76,26; 80,04]	69,55	[67,64; 71,38]
4	76,51	[73,68; 79,09]	72,42	[70,28; 74,43]	61,68	[59,64; 63,64]
5	68,28	[65,14; 71,20]	65,89	[63,59; 68,07]	53,61	[51,50; 55,67]
6	59,93	[56,59; 63,11]	60,06	[57,66; 62,37]	47,28	[45,13; 49,39]
7	55,07	[51,65; 58,36]	56,15	[53,68; 58,55]	41,61	[39,45; 43,75]
8	52,20	[48,73; 55,55]	53,03	[50,51; 55,49]	37,11	[34,95; 39,27]
9	48,81	[45,28; 52,25]	50,21	[47,62; 52,74]	33,10	[30,93; 35,29]
10	39,41	[35,78; 43,01]	40,07	[37,25; 42,87]	23,04	[20,89; 25,25]

Ao atingir o décimo período, a probabilidade de sobrevivência no turno *NOTURNO* é de 23,04% com intervalo de confiança de [20,89%; 25,25%]. Os turnos *MATUTINO* e *VESPERTINO* apresentam probabilidades de 39,41% e 40,07%, respectivamente, com limites situados entre [37,45%; 41,37%] e [38,12%; 42,02%], os quais se interceptam ao final do período de integralização. Estes indicadores sugerem que as trajetórias de sobrevivência dos turnos diurnos apresentam comportamento similar a longo prazo, enquanto o turno *NOTURNO* mantém um perfil de sobrevivência inferior.

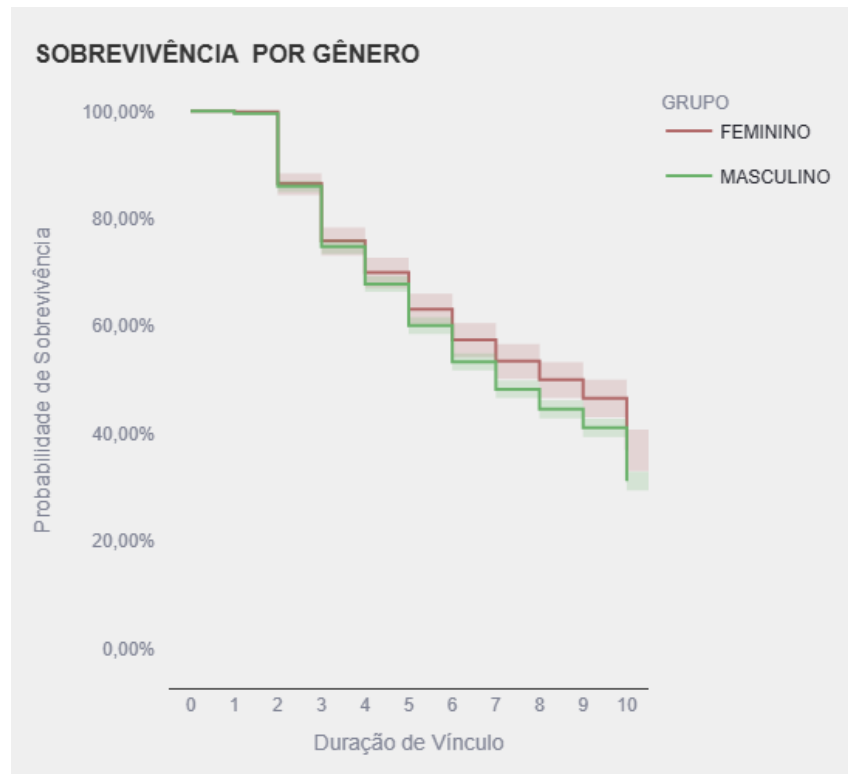
Complementarmente, a análise estende-se à influência do agrupamento de gênero nas funções de sobrevivência em STEM. O estimador de *Kaplan-Meier* indica que ambos os grupos apresentam trajetórias com sobreposição dos intervalos de confiança ao longo do intervalo observado. O grupo *FEMININO* registra probabilidades superiores ao grupo *MASCULINO* na maioria dos períodos, apresentando uma vantagem de 5,72 pontos percentuais ao final do acompanhamento.

A partir da identificação dos períodos de transição expressos na Figura 29 e detalhados na Tabela 25, verifica-se que as trajetórias de sobrevivência apresentam ritmos distintos. O grupo *MASCULINO* cruza o limiar de 50% de probabilidade de permanência no 7º período, ao registrar 48,30% com intervalo de confiança de [46,65%; 49,93%]. A consolidação estatística deste cruzamento ocorre neste ponto, dado que o limite superior do intervalo situa-se abaixo de 50%. Em contrapartida, o grupo *FEMININO* mantém-se acima do referido limiar até o 8º período, com 50,05% e intervalo de confiança de [46,67%; 53,34%].

O cruzamento para este grupo ocorre no 9º período, quando a estimativa reduz para 46,61% com limites situados entre [43,11%; 50,04%]. No 2º período, ambos os gêneros apresentam reduções na probabilidade de sobrevivência: o grupo *FEMININO* registra 86,55% com intervalo de confiança de [84,38%; 88,44%], enquanto o grupo *MASCULINO* apresenta 86,09% com parâmetros de [85,01%; 87,11%].

Ao atingir o 10º período, a probabilidade de sobrevivência é de 36,94% para o grupo *FEMININO* e 31,22% para o grupo *MASCULINO*. Embora o grupo *FEMININO* mantenha valores numericamente superiores, a proximidade dos intervalos de confiança sugere que o gênero apresenta uma homogeneidade maior em comparação à variável turno. Estes indicadores demonstram que, no cenário de *STEM*, as trajetórias de sobrevivência são aproximadas entre os grupos, com ambas as categorias apresentando uma permanência final próxima a um terço dos ingressantes.

Sob a ótica das políticas de permanência, a análise consiste no impacto da assistência estudantil sobre as funções de sobrevivência. O recebimento de auxílio evidencia o distanciamento mais expressivo entre as variáveis analisadas, demonstrando que o suporte institucional é um diferencial para a manutenção do vínculo. O estimador de *Kaplan-Meier* apresenta trajetórias em que o grupo assistido *SIM* mantém probabilidades de sobrevivência superiores durante todo o intervalo.



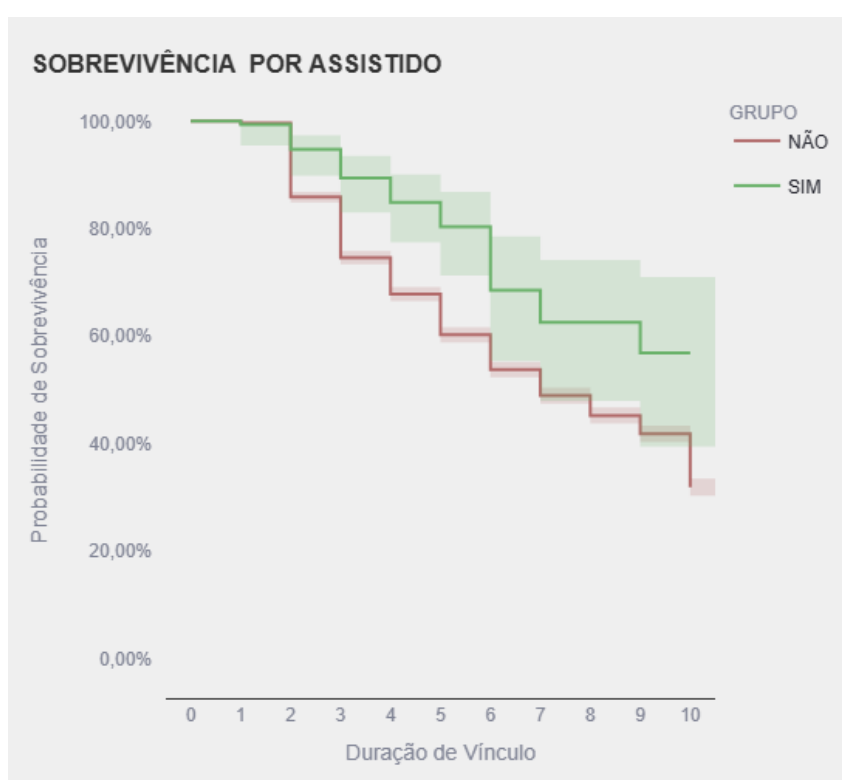
(a) Estimador de Kaplan-Meier por Gênero.

Figura 29 – Análise de sobrevivência nos cursos de STEM, por gênero.

Tabela 25 – Probabilidade de sobrevivência e intervalo de Confiança (IC 95%) para o cenário de STEM por gênero.

Prd.	FEMININO		MASCULINO	
	$S(t)$ (%)	IC 95%	$S(t)$ (%)	IC 95%
0	100,00	[100,00; 100,00]	100,00	[100,00; 100,00]
1	99,91	[99,36; 99,99]	99,57	[99,32; 99,73]
2	86,55	[84,38; 88,44]	86,09	[85,01; 87,11]
3	75,90	[73,20; 78,38]	74,84	[73,48; 76,14]
4	70,04	[67,14; 72,75]	67,88	[66,41; 69,31]
5	63,15	[60,05; 66,09]	60,15	[58,58; 61,68]
6	57,48	[54,25; 60,57]	53,41	[51,79; 55,00]
7	53,51	[50,19; 56,71]	48,30	[46,65; 49,93]
8	50,05	[46,67; 53,34]	44,57	[42,90; 46,22]
9	46,61	[43,11; 50,04]	41,12	[39,42; 42,80]
10	36,94	[33,10; 40,79]	31,22	[29,49; 32,96]

Conforme representado na Figura 30 e na Tabela 26, as estimativas de permanência apresentam distinção desde o 2º período, no qual o grupo *SIM* apresenta uma probabilidade de 94,81% com intervalo de confiança de [89,88%; 97,37%], enquanto o grupo *NÃO* registra 85,93% com limites situados entre [84,95%; 86,85%], o que consolida uma diferença de 8,88 pontos percentuais. Enquanto o grupo *NÃO* cruza o limiar de 50% de sobrevivência entre o 6º e o 7º período, o grupo *SIM* mantém-se acima desse valor em todo o intervalo observado, finalizando o 10º período com uma probabilidade de 56,91% e intervalo de confiança compreendido em [39,48%; 71,05%].



(a) Estimador de Kaplan-Meier por Assistência.

Figura 30 – Análise de sobrevivência nos cursos de STEM, por assistência estudantil.

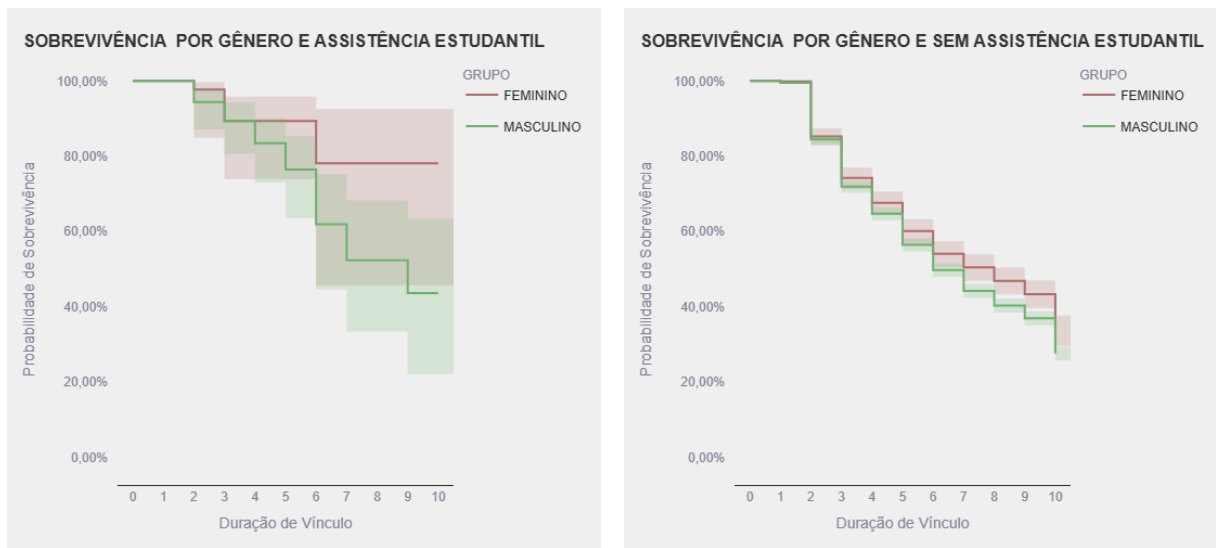
Ao final do 10º período, a diferença entre os grupos é de 24,96 pontos percentuais, visto que o segmento assistido mantém a probabilidade de sobrevivência em 56,91% com intervalo de confiança de [39,48%; 71,05%], enquanto o grupo *NÃO* apresenta 31,95% com limites situados entre [30,37%; 33,54%]. O afastamento entre os parâmetros dos intervalos de confiança ao término do acompanhamento confirma a distinção estatística entre as trajetórias de permanência. Estes indicadores demonstram que a assistência estudantil em STEM associa-se a índices de permanência superiores até o estágio final de observação, mantendo o grupo beneficiado acima do limiar de 50% de sobrevivência.

Tabela 26 – Probabilidade de sobrevivência e intervalo de Confiança (IC 95%) para o cenário de STEM por assistência estudantil.

Prd.	Assistido		Não Assistido	
	$S(t)$ (%)	IC 95%	$S(t)$ (%)	IC 95%
0	100,00	[100,00; 100,00]	100,00	[100,00; 100,00]
1	99,35	[95,48; 99,91]	99,65	[99,45; 99,78]
2	94,81	[89,88; 97,37]	85,93	[84,95; 86,85]
3	89,45	[83,07; 93,52]	74,65	[73,42; 75,83]
4	84,94	[77,50; 90,07]	67,86	[66,53; 69,15]
5	80,39	[71,28; 86,86]	60,27	[58,86; 61,65]
6	68,56	[55,38; 78,58]	53,83	[52,37; 55,26]
7	62,60	[47,95; 74,20]	48,97	[47,48; 50,44]
8	62,60	[47,95; 74,20]	45,28	[43,78; 46,77]
9	56,91	[39,48; 71,05]	41,85	[40,32; 43,37]
10	56,91	[39,48; 71,05]	31,95	[30,37; 33,54]

Ao segmentar a análise por gênero entre assistidos e não assistidos, conforme expresso na Figura 31 e na Tabela 27. No cenário dos assistidos, o grupo *FEMININO* mantém a probabilidade de permanência em 78,18% com intervalo de confiança de [45,8%; 92,5%] entre o 6º e o 10º período, o que indica a ausência de novos eventos de saída na metade final do acompanhamento. Em contrapartida, o grupo *MASCULINO* assistido apresenta reduções e cruza o limiar de 50% no 9º período, ao registrar 43,66% com limites de [22,2%; 63,4%]. A diferença de 34,52 pontos percentuais ao final do intervalo indica que o suporte institucional se associa à manutenção do vínculo para o grupo *FEMININO* em proporção superior à verificada para o grupo *MASCULINO*.

No cenário dos estudantes não assistidos, as trajetórias apresentam proximidade até o 5º período. A partir deste ponto, o grupo *MASCULINO* cruza o limiar de 50% de sobrevivência no 6º período, ao registrar 49,80% com intervalo de confiança de [48,0%; 51,6%], enquanto o grupo *FEMININO* atinge patamar similar no 8º período, com 46,90% e limites entre [43,3%; 50,4%]. No 10º período, o grupo *FEMININO* registra 33,66% de probabilidade de permanência, frente aos 27,58% verificados para o grupo *MASCULINO*.



(a) Gênero - Assistidos

(b) Gênero - Não Assistidos

Figura 31 – Comparativo de sobrevivência em STEM por gênero segmentado por assistência estudantil.

Tabela 27 – Análise longitudinal comparativa em STEM: probabilidades de sobrevivência por gênero e assistência (0-10)

Prd.	Assistidos				Não Assistidos			
	FEMININO		MASCULINO		FEMININO		MASCULINO	
	$S(t)$ (%)	IC	$S(t)$ (%)	IC	$S(t)$ (%)	IC	$S(t)$ (%)	IC
0	100,00	[100,0; 100,0]	100,00	[100,0; 100,0]	100,00	[100,0; 100,0]	100,00	[100,0; 100,0]
1	100,00	[100,0; 100,0]	100,00	[100,0; 100,0]	99,89	[99,2; 99,9]	99,55	[99,2; 99,7]
2	97,73	[84,9; 99,7]	94,44	[87,2; 97,6]	85,30	[82,8; 87,4]	84,46	[83,2; 85,7]
3	89,35	[74,0; 95,9]	89,41	[80,6; 94,4]	74,23	[71,2; 77,0]	71,93	[70,3; 73,4]
4	89,35	[74,0; 95,9]	83,45	[73,0; 90,1]	67,64	[64,4; 70,6]	64,76	[63,1; 66,4]
5	89,35	[74,0; 95,9]	76,49	[63,6; 85,3]	60,12	[56,7; 63,3]	56,45	[54,7; 58,2]
6	78,18	[45,8; 92,5]	61,92	[44,6; 75,3]	54,07	[50,6; 57,4]	49,80	[48,0; 51,6]
7	78,18	[45,8; 92,5]	52,40	[33,4; 68,3]	50,49	[46,9; 53,9]	44,26	[42,4; 46,1]
8	78,18	[45,8; 92,5]	52,40	[33,4; 68,3]	46,90	[43,3; 50,4]	40,36	[38,5; 42,2]
9	78,18	[45,8; 92,5]	43,66	[22,2; 63,4]	43,40	[39,7; 47,1]	36,96	[35,1; 38,8]
10	78,18	[45,8; 92,5]	43,66	[22,2; 63,4]	33,66	[29,7; 37,7]	27,58	[25,7; 29,5]

Fonte: Elaborado pelo autor (2026).

A validação das diferenças nas curvas de sobrevivência para os cursos de STEM, realizada via teste *Log-Rank* e detalhada na Tabela 28, confirma que a assistência estudantil é a variável de maior impacto estatístico com p-valor de $4,4667e^{-56}$. A rejeição da hipótese nula (H_0) em todos os recortes de assistência, por turno e gênero, demonstra que o suporte institucional altera a trajetória dos discentes nessas áreas com indicadores históricos de interrupção de vínculo.

Observa-se, contudo, um fenômeno de equalização proporcionado pelo auxílio. Enquanto no grupo de não assistidos as trajetórias entre gêneros e turnos são estatisticamente divergentes, como verificado na comparação entre *MASCULINO* e *FEMININO* não assistidos que apresenta um p-valor de $5,6640e^{-7}$, no agrupamento assistido essas disparidades perdem significância estatística. O p-valor de $1,0030e^0$ para gênero entre assistidos indica uma sobreposição das curvas de sobrevivência, sugerindo que a assistência atua como um nivelador, evidenciando que a política de assistência se associa à equalização das probabilidades de permanência entre subgrupos em STEM.

A análise de risco detalhada na Tabela 29 apresenta os coeficientes para as probabilidades de evasão nos cursos de STEM. A assistência estudantil atua como fator de proteção, conforme verificado no segmento *FEMININO* assistido, que apresenta 27,85% do risco de evasão em relação às estudantes de mesmo gênero não assistidas com HR de 0,2785 e p-valor de 0,0045.

Por outro lado, o turno *NOTURNO* em STEM apresenta riscos de interrupção do vínculo 59,70% superiores aos do turno *MATUTINO* entre os não assistidos que apresentam HR de 1,5970 e p-valor de 0,0010. Entre os estudantes que recebem suporte institucional, variações de risco, como a observada na comparação entre os grupos *MASCULINO* e *FEMININO* assistidos que refletem HR de 2,0780, não possuem significância estatística com p-valor igual a 0,1390. As estimativas sugerem que o auxílio institucional se associa à redução de disparidades de risco entre os subgrupos nessas áreas acadêmicas.

Tabela 28 – Teste de hipótese *Log-Rank* para cursos de STEM.

Grupo de Ref.	Grupo de Alt.	H_0	p-valor
Assistido, STEM	Não Assistido, STEM	Rejeitado	$4,4667e^{-56}$
Assistido, STEM, Matutino	Não Assistido, STEM, Matutino	Rejeitado	$7,1443e^{-21}$
Assistido, STEM, Noturno	Não Assistido, STEM, Noturno	Rejeitado	$1,5796e^{-20}$
Assistido, STEM, Feminino	Não Assistido, STEM, Feminino	Rejeitado	$7,5358e^{-18}$
Assistido, STEM, Vespertino	Não Assistido, STEM, Vespertino	Rejeitado	$1,6556e^{-10}$
Masculino, STEM	Feminino, STEM	Rejeitado	$1,0037e^{-9}$
Masculino, STEM, Não Assist.	Feminino, STEM, Não Assist.	Rejeitado	$5,6640e^{-7}$
Matutino, STEM	Vespertino, STEM	Rejeitado	$4,1753e^{-6}$
Matutino, STEM, Não Assist.	Vespertino, STEM, Não Assist.	Rejeitado	$4,6168e^{-6}$
Masculino, STEM, Noturno	Feminino, STEM, Noturno	Rejeitado	$< 0,0001$
Matutino, STEM, Não Assist.	Noturno, STEM, Não Assist.	Rejeitado	$1,2700e^{-2}$
Vespertino, STEM	Noturno, STEM	Rejeitado	$1,4900e^{-2}$
Matutino, STEM	Noturno, STEM	Rejeitado	$2,0000e^{-2}$
Noturno, STEM, Não Assist.	Vespertino, STEM, Não Assist.	Rejeitado	$2,8300e^{-2}$
Masculino, STEM, Vespertino	Feminino, STEM, Vespertino	Rejeitado	$3,1000e^{-2}$
Matutino, STEM, Assistido	Noturno, STEM, Assistido	Não Rejeitado	$7,0570e^{-2}$
Masculino, STEM, Matutino	Feminino, STEM, Matutino	Não Rejeitado	$1,0700e^{-1}$
Noturno, STEM, Assistido	Vespertino, STEM, Assistido	Não Rejeitado	$3,3410e^{-1}$
Matutino, STEM, Assistido	Vespertino, STEM, Assistido	Não Rejeitado	$6,4070e^{-1}$
Masculino, STEM, Assistido	Feminino, STEM, Assistido	Não Rejeitado	1,0030

Nota: H_0 indica a hipótese nula de igualdade entre as funções de sobrevivência. Hipótese rejeitada para p-valor $< 0,05$.

Tabela 29 – *Hazard Ratio* e coeficientes de risco para cursos de STEM.

Grupo Ref. (1)	Grupo Alt. (2)	(β)	HR	p-valor
Feminino, STEM, Assistido	Masculino, STEM, Assistido	0,7312	2,0780	0,1390
Matutino, STEM, Assistido	Noturno, STEM, Assistido	0,6935	2,0070	0,1340
Matutino, STEM, Não Assist.	Noturno, STEM, Não Assist.	0,4680	1,5970	0,0010
Matutino, STEM	Noturno, STEM	0,4664	1,5942	0,0000
Masculino, STEM, Noturno	Feminino, STEM, Noturno	1,1991	1,2200	0,0040
Masculino, STEM, Vespertino	Feminino, STEM, Vespertino	1,1584	1,1720	0,1030
Masculino, STEM, Não Assist.	Feminino, STEM, Não Assist.	0,1484	1,1600	0,0035
Feminino, STEM	Masculino, STEM	0,1301	1,1389	0,0700
Masculino, STEM, Matutino	Feminino, STEM, Matutino	0,0682	1,0710	0,8070
Matutino, STEM, Não Assist.	Vespertino, STEM, Não Assist.	0,0315	1,0320	0,5560
Matutino, STEM	Vespertino, STEM	0,0205	1,0207	0,7165
Não Assistido, STEM, Matutino	Assistido, STEM, Matutino	-0,3430	0,7100	0,4470
Não Assistido, STEM, Noturno	Assistido, STEM, Noturno	-0,4002	0,6700	0,3280
Vespertino, STEM	Noturno, STEM	-0,4416	0,6430	0,0000
Matutino, STEM, Assistido	Vespertino, STEM, Assistido	-0,5802	0,5598	0,3218
Noturno, STEM, Assistido	Vespertino, STEM, Assistido	-0,5962	0,5509	0,1738
Não Assistido, STEM, Masculino	Assistido, STEM, Masculino	-0,6443	0,5250	0,0021
Não Assistido, STEM	Assistido, STEM	-0,7405	0,4769	0,0000
Não Assistido, STEM, Vespert.	Assistido, STEM, Vespertino	-1,1469	0,3176	0,0013
Não Assistido, STEM, Feminino	Assistido, STEM, Feminino	-1,2783	0,2785	0,0045

Nota: HR > 1 indica risco aumentado para o Grupo Alt. em relação ao Ref.; HR < 1 indica efeito protetivo. β representa o coeficiente de regressão de Cox.

Para mensurar a magnitude desses efeitos, o uso do teste de tamanho de efeito *Effect-Size* (Hedges' g), cujos resultados estão detalhados na Tabela 30, indica que a assistência estudantil apresenta os valores de maior impacto prático nos cursos de STEM. No turno *NOTURNO*, a diferença entre assistidos e não assistidos registra um tamanho de efeito grande com $g = 0,8058$. As demais comparações ligadas ao suporte institucional, como nos grupos, *FEMININO* ($g = 0,4425$) e *MASCULINO* ($g = 0,3563$), apresentam magnitudes classificadas como médias.

Em contrapartida, as variáveis de gênero e turno registram tamanhos de efeito pequenos na maioria das combinações. A disparidade entre *MASCULINO* e *FEMININO* assistidos apresenta o menor índice da análise com $g = 0,0035$, o que indica a paridade nas condições de permanência entre os gêneros nesse grupo. Os resultados sugerem que, embora existam diferenças estatísticas em turnos e perfis demográficos, a intensidade dessas variações é reduzida quando comparada ao impacto exercido pelo recebimento de assistência estudantil.

Quando analisados os grupos de maior importância, os resultados em STEM demonstram que a assistência estudantil e a modalidade de curso são os principais determinantes da permanência, sobrepondo-se aos riscos associados ao turno e ao gênero. Embora o modelo de Cox identifique o turno *NOTURNO* como o fator de maior risco relativo entre os não assistidos ($HR = 1,5970$; $p = 0,0010$), a análise do tamanho do efeito pelo Hedges' g revela que a assistência estudantil possui uma força de impacto superior para neutralizar essa vulnerabilidade, atingindo magnitude grande ($g = 0,8058$) justamente nesse turno. Observa-se que o turno atua como um proxy para a natureza da formação acadêmica, uma vez que as Licenciaturas, concentradas no período noturno, registram volumes de evasão até 71,6% superiores aos de conclusão, o que desloca a relevância do risco do horário para a estrutura da modalidade de curso e o perfil socioeconômico dos ingressantes.

O significado técnico desses indicadores reside na capacidade de equalização proporcionada pelo suporte institucional. A convergência entre o risco (HR) e a intensidade do efeito (g) confirma que a assistência é a variável de maior peso hierárquico no modelo, sendo capaz de mitigar a janela crítica de evasão verificada entre o 2º e o 6º período. A evidência desse processo é a paridade de sobrevivência atingida entre homens e mulheres assistidos, que apresenta o menor tamanho de efeito da análise ($g = 0,0035$) e ausência

Tabela 30 – Hedges' g para comparações entre grupos de STEM.

Grupo de Ref.	Grupo de Alt.	Hedges' g	Tam. do efeito
Assistido, STEM, Noturno	Não Assistido, STEM, Noturno	0,8058	Grande
Assistido, STEM	Não Assistido, STEM	0,4442	Médio
Assistido, STEM, Feminino	Não Assistido, STEM, Feminino	0,4425	Médio
Assistido, STEM, Masculino	Não Assistido, STEM, Masculino	0,3563	Médio
Assistido, STEM, Matutino	Não Assistido, STEM, Matutino	0,3522	Médio
Matutino, STEM, Não Assist.	Noturno, STEM, Não Assist.	0,3486	Médio
Matutino, STEM	Noturno, STEM	0,3369	Médio
Noturno, STEM, Não Assist.	Vespertino, STEM, Não Assist.	0,3319	Médio
Noturno, STEM, Assistido	Vespertino, STEM, Assistido	0,2260	Médio
Vespertino, STEM	Noturno, STEM	0,2200	Médio
Assistido, STEM, Vespertino	Não Assistido, STEM, Vespertino	0,1839	Pequeno
Matutino, STEM, Assistido	Noturno, STEM, Assistido	0,1790	Pequeno
Masculino, STEM, Não Assist.	Feminino, STEM, Não Assist.	0,1191	Pequeno
Matutino, STEM, Não Assist.	Vespertino, STEM, Não Assist.	0,1181	Pequeno
Matutino, STEM	Vespertino, STEM	0,1127	Pequeno
Feminino, STEM, Matutino	Masculino, STEM, Matutino	0,1045	Pequeno
Matutino, STEM, Assistido	Vespertino, STEM, Assistido	0,0586	Pequeno
Masculino, STEM, Vespertino	Feminino, STEM, Vespertino	0,0529	Pequeno
Feminino, STEM	Masculino, STEM	0,0379	Pequeno
Masculino, STEM, Noturno	Feminino, STEM, Noturno	0,0071	Pequeno
Feminino, STEM, Assistido	Masculino, STEM, Assistido	0,0035	Pequeno

Nota: Classificação do tamanho do efeito baseada nos limiares de Hedges'g: Pequeno ($\leq 0,2$), Médio ($\leq 0,5$) e Grande ($\leq 0,8$). No contexto educacional, valores próximos a 0,3 são considerados relevantes.

de significância estatística de risco ($p = 0,1390$). Em suma, a interrupção de vínculo em STEM é condicionada por fatores financeiros e institucionais, visto que o suporte mantém a probabilidade de permanência do grupo beneficiado acima de 50% em todo tempo de observação, independentemente das pressões exercidas pelo turno ou pela modalidade do curso.

7 Discussão dos resultados

Este capítulo descreve a construção de um modelo de análise de sobrevivência para a obtenção de indicadores estatísticos de permanência acadêmica. Apresentam-se inferências sobre turno, gênero e assistência estudantil no cenário institucional e no recorte STEM. A abordagem utiliza dados para orientar decisões e a formulação de políticas de permanência. O objetivo consiste em identificar se os padrões temporais de sobrevivência e os períodos de vulnerabilidade divergem entre as trajetórias analisadas (PP1); de que forma fatores estruturais e socioeconômicos influenciam a probabilidade de permanência ao longo do tempo (PP2); e como a combinação de indicadores de significância (*Log – Rank*), risco (*Hazard Ratio*) e magnitude (*Effect Size*) quantifica a influência de fatores acadêmicos (PP3).

7.1 PP1: Quais os padrões temporais de sobrevivência e como os períodos de vulnerabilidade divergem entre as trajetórias analisadas?

7.1.1 Reformulação pedagógica curricular

A comparação entre os cursos de LC e BSI presentes na seção 6.1 indicam que a transição para o grupo *Depois* apresenta um aumento nos indicadores de permanência. Embora nos primeiros períodos, a sobreposição dos intervalos de confiança em ambos os cursos demonstra que essa mudança ainda sofre a ocorrência do evento de *EVASÃO*. No 2º período, LC apresenta $S(t)$ de 84,00% com intervalo de confiança de [80,66%; 86,81%] no grupo *Antes* e de 83,67% com intervalo de confiança de [79,93%; 86,77%] no grupo *Depois*. BSI registra 88,19% com intervalo de confiança de [85,74%; 90,24%] no grupo *Antes* e 89,33% com intervalo de confiança de [85,63%; 92,12%] no grupo *Depois*. Esse padrão sugere que as causas de desligamento no primeiro ano podem estar associadas a fatores externos à matriz curricular, como a reentrada em novos processos seletivos e variáveis externas ao controle institucional (JUNIOR et al., 2012; RIBEIRO et al., 2021). Esse cenário converge com as evidências que apontam a carência de embasamento lógico-matemático como um gatilho para a desistência precoce, ao descreverem o impacto do choque acadêmico inicial sobre a continuidade dos ingressantes (VASCONCELOS;

ANDRADE, 2018). Essa persistência da evasão no primeiro ano, a despeito das reformas, é corroborada por Souza et al. (2022), que identificou uma queda na probabilidade de sobrevivência para 56,48% já no semestre inicial, e por Pan et al. (2022), que associa essa volatilidade à inadaptação aos métodos de avaliação universitários.

As trajetórias apresentam distanciamento a partir da interrupção da sobreposição dos intervalos de confiança, o que ocorre no 3º período para BSI e no 4º período para LC. Em BSI, o grupo *Depois* sustenta uma sobrevivência de 83,54% com intervalo de confiança de [78,01%; 87,21%] no 3º período, valor superior aos 75,13% com intervalo de confiança de [71,97%; 77,98%] do grupo *Antes*. Essa elevação dos índices de permanência indica que a implementação dos Projetos Interdisciplinares (PIs) e a redução de pré-requisitos favoreceram a continuidade acadêmica (NASCIMENTO et al., 2022). A eficácia desta horizontalização de conteúdos via PIs corrobora os achados de Zamith et al. (2025), que associam a redução do risco de falha à integração curricular e observaram a estabilização da evasão a partir de matrizes mais flexíveis.

Entre as mudanças registradas na reformulação do BSI, ocorreu a substituição do modelo de dimensões (Gestão, Tecnologia e Humana) por uma organização em núcleos de formação suplementar e complementar. O PPC incluiu a introdução dos PIs, presentes do 1º ao 8º período, com a finalidade de relacionar os conteúdos das disciplinas de cada semestre. Essa estrutura foi acompanhada da flexibilização de disciplinas antes obrigatórias, como Economia e Física para Computação. Somam-se ao processo o deslocamento de componentes como Cálculo a Várias Variáveis para períodos posteriores e a definição de trilhas de formação (Graduação-Tecnologia ou Graduação-Ciência). Assim como relatado por Cavalcanti et al. (2024).

Em LC, a divergência estatística surge no 4º período, com o grupo *Depois* registrando 69,22% com intervalo de confiança de [64,85%; 74,44%] frente aos 55,96% com intervalo de confiança de [55,96%; 64,14%] do modelo anterior. O alcance do limiar de 50% de sobrevivência ocorria no 5º período com intervalo de confiança de [45,53%; 53,89%] no grupo *Antes*. Com a implementação do novo PPC, este indicador é deslocado para o 9º período no grupo *Depois* com intervalo de confiança de [39,75%; 51,14%], situando-se no intervalo entre o 7º e o 11º período, o que registra a extensão do vínculo discente sob a nova estrutura curricular. A fragmentação do fluxo de programação em componentes menores (Pensamento Computacional e Laboratórios) atua diretamente sobre o risco de

evasão precoce [Falcão et al. \(2018\)](#). Tal medida incide positivamente sobre o fenômeno da evasão ao substituir o choque acadêmico inicial por um fluxo de aprendizado mais distribuído [Zamith et al. \(2025\)](#).

Este cenário associa-se às alterações na estrutura curricular, que incluiu a substituição de uma disciplina única de Introdução à Programação com carga horária de 90h no 1º período por um fluxo composto por Pensamento Computacional, seguido pelas disciplinas de Programação I no 2º período e Programação II no 3º período, ambas com 60h cada. Estes componentes são acompanhados pelos Laboratórios de Programação I e II, com 30h cada e foco na prática laboratorial. Outro fator registrado foi a reformulação do Estágio Supervisionado Obrigatório, com a redução de três para dois pré-requisitos, representados por Metodologia do Ensino da Computação e Educação Brasileira, além da eliminação da dependência sequencial entre os estágios. A matriz passou a permitir a equiparação do estágio extracurricular como Vivência Profissional Complementar e o aumento de componentes optativos e em regime EAD ([SILVA, 2024](#)).

Contudo, o atraso no desfecho de evasão, retratado pelo deslocamento do marco de 50% nos cursos de LC e BSI, demanda uma análise sobre a eficiência dessa retenção. Embora a extensão do vínculo discente sob as novas matrizes indique redução do choque acadêmico inicial, é necessário verificar se este fenômeno resulta em incremento nas taxas de conclusão ou apenas nas estimativas de permanência do aluno em um fluxo que não culminará em diplomação. A manutenção do vínculo por períodos avançados para que a evasão ocorra próximo ao limite da integralização pode representar um custo de oportunidade superior ao desligamento precoce. Portanto, a horizontalização curricular deve ser avaliada sob a ótica da eficácia pedagógica, distinguindo o adiamento da falha da progressão acadêmica efetiva.

7.1.2 Cenário institucional e áreas de STEM

Os achados deste trabalho, expressos na Seção 6.2 demonstram que, embora o fenômeno da evasão seja transversal à universidade, sua dinâmica e intensidade variam conforme a área de conhecimento e o perfil socioeconômico ([JUNIOR et al., 2012](#)). No cenário institucional da UFRPE, a sobrevivência média registra 5,49 períodos, enquanto no recorte de STEM, as Licenciaturas em Matemática com média de 4,94 e Física com

4,96 períodos apresentam indicadores inferiores, o que indica uma interrupção de vínculo precoce nesta área. Esse comportamento associa-se à densidade das disciplinas de base, como Cálculo e Física, que atuam como barreiras iniciais ao fluxo acadêmico (MOURA et al., 2020; PASCOAL et al., 2016). Além disso, a distribuição dos dados revela a predominância de eventos de evasão sobre os casos de censura nas licenciaturas, com uma diferença de 71,6% no curso de Física. A interrupção precoce observada nessas licenciaturas converge com os achados de Chen et al. (2018), ao registrar que 60% das desistências em STEM ocorrem nos primeiros dois anos de curso.

A análise longitudinal identifica o intervalo entre o 1º e o 4º período como a fase de maior propensão ao desligamento em todos os cenários. No recorte de STEM, o decréscimo na probabilidade de sobrevivência encontra-se no intervalo onde o turno *NOTURNO* inicia o 2º período com 82,33% e intervalo de confiança de [80,60%; 83,93%] e atinge 61,68% com intervalo de confiança de [59,51%; 63,78%] no 4º período, representando uma perda de 20,65 pontos percentuais em apenas um ano letivo. Este declínio no turno noturno corrobora a tese de Garcia e Gomes (2022), que identifica a sobrecarga da jornada de trabalho como o principal impeditivo para a persistência discente, resultando em janelas de sobrevivência reduzidas quando comparadas aos turnos diurnos. No cenário institucional, metade da coorte permanece vinculada até o 6º período com intervalo de confiança de [54,91%; 57,16%], enquanto no turno *NOTURNO* de STEM, esse marco de 50% ocorre entre o 5º período, com registro de 53,61% e intervalo de confiança de [51,35%; 55,81%], e o 6º período, que apresenta 47,28% com intervalo de confiança de [45,03%; 49,50%]. Em contrapartida, os turnos *MATUTINO* e *VESPERTINO* em STEM mantêm a maioria vinculada até o 7º período. O risco relativo superior do turno noturno em STEM ($HR = 1,5970$) demonstra o argumento de Almeida e Schimiguel (2011) que discentes de exatas possuem probabilidade de abandono 1,32 vezes maior.

A análise por gênero e assistência estudantil revela disparidades nos marcos temporais de vulnerabilidade. O grupo *MASCULINO* atinge o limiar de 50% de sobrevivência no 7º período com registro de 49,93% e intervalo de confiança de [49,02%; 50,84%], enquanto o grupo *FEMININO* mantém-se acima deste limite até o 9º período, apresentando 54,27% com intervalo de confiança de [53,19%; 55,34%]. A assistência estudantil atua como o principal estabilizador: discentes não assistidos cruzam o marco de 50% no 8º período com 50,41% e intervalo de confiança de [49,70%; 51,11%], ao passo que

o grupo assistido não atinge este limiar durante os dez períodos analisados, encerrando a observação com 65,20% e intervalo de confiança de [58,02%; 71,47%] de permanência. Esses dados indicam que, embora os quatro primeiros períodos concentrem o maior risco, o suporte institucional e o perfil do discente alteram a velocidade da trajetória de evasão. A estabilidade desta trajetória assistida confirma a premissa de [Saccaro et al. \(2019\)](#) de que o apoio financeiro reduz as taxas de falha. Conforme argumentado por [Silva e Sampaio \(2022\)](#) e [Filipak e Pacheco \(2017\)](#), esses resultados demonstram que a democratização do acesso só se traduz em permanência quando acompanhada de suporte institucional contínuo, capaz de neutralizar a vulnerabilidade socioeconômica.

7.2 PP2: De que maneira a reformulação curricular e as variáveis de gênero, turno e assistência estudantil impactam as disparidades no risco de evasão e na probabilidade de permanência discente?

7.2.1 Efeitos nas trajetórias dos cursos frente a reformulação pedagógica curricular

A reformulação dos Projetos Pedagógicos de Curso (PPC) atuou como um fator estrutural de mitigação da evasão ao alterar a dinâmica de permanência nos períodos críticos. Na Subseção [6.1.3](#) a transição de BSI para o novo modelo resultou em uma redução de 33,20 pontos percentuais na evasão acumulada até o 7º período, com o índice caindo de 54,65% no grupo *Antes* para 21,45% no grupo *Depois*. Na Subseção [6.1.2](#), em LC, a redução foi ainda mais expressiva no 10º período, passando de 79,27% para 37,44%, o que representa uma diferença de 41,83 pontos percentuais entre as matrizes curriculares. Este aumento na sobrevivência converge com os resultados de [Zamith et al. \(2025\)](#), que ao analisar a mudança curricular na UFRRJ, observou um salto na sobrevida discente de 17,4% para 53%, além de uma redução de 21,3% no tempo para diplomação. Tais índices corroboram a premissa de [Cavalcanti et al. \(2024\)](#) de que a atualização dos Projetos Pedagógicos de Curso (PPC) é capaz de elevar a sobrevivência institucional entre 30% e 40%.

Esses resultados indicam que a flexibilização curricular, a remoção de pré-requisitos sequenciais e a introdução de Projetos Interdisciplinares (PIs) funcionam como mecanismos de proteção. Esta observação converge com o estudo de [Gutierrez-Pachas et al. \(2022\)](#),

que ao analisar a conformidade curricular com padrões internacionais, identificou que a reestruturação da matriz acadêmica resultou em uma elevação da permanência geral dos estudantes. Ao permitir que o discente adapte seu percurso, a estrutura reduz o impacto de reprovações em disciplinas de alta densidade, como as de programação e matemática, evitando o efeito cascata de trancamentos e desligamentos precoces (SOUSA et al., 2024).

7.2.2 Efeitos nas trajetórias STEM e institucionais

O agrupamento de turno constitui um eixo de temporalidade da evasão. No cenário institucional, os indicadores registrados na Subseção 6.2.2 estabelecem uma hierarquia de risco onde o turno *MATUTINO* apresenta a maior probabilidade de sobrevivência ao final de dez períodos com 42,17%, seguido pelo *VESPERTINO* com 42,11%, enquanto o *NOTURNO* registra 33,78%. No recorte STEM, essa disparidade é evidenciada no grupo *NOTURNO*, que cruza o limiar de 50% de permanência entre o 5º e o 6º período, onde registra 47,28%, e encerra o tempo de observação com 23,04%.

O impacto dessa condição é evidenciado na Subseção 6.2.4 pela combinação de maior risco entre homens não assistidos, os quais apresentam $HR = 1,2200$, configurando uma barreira estrutural para a permanência. Sob a ótica da sociologia da educação, a maior vulnerabilidade associada ao grupo masculino sugere uma intersecção entre a carência de suporte institucional e a construção social da masculinidade vinculada ao provimento financeiro imediato. Diferente do público feminino, que demonstra trajetórias de persistência mais tardias, o desligamento precoce masculino em STEM é tensionado pela expectativa de inserção no mercado de trabalho técnico.

Tal cenário converge com a análise de Gisi (2006), que argumenta que a democratização do acesso ao ensino superior no Brasil não foi acompanhada pela redução das disparidades de condições socioeconômicas. Nesse contexto, a vulnerabilidade no perfil dos ingressantes impõe novos desafios à permanência devido à necessidade de conciliação com atividades laborais, o que atua como um fator de risco adicional para o público masculino em cursos de alta densidade acadêmica (SORGATTO; CÁCERES, 2023; SANTOS et al., 2024).

No que tange às trajetórias de gênero em STEM, os dados revelam um contraste entre o volume de ingressos e a persistência acadêmica. Na UFRPE, as mulheres representam

20,3% dos ingressos em STEM, com sobrevivência final de 36,94%, superior à masculina de 31,22%. A análise longitudinal indica que o grupo *FEMININO* em STEM atinge o limiar de 50% de evasão no 9º período, enquanto o grupo *MASCULINO* atinge este limite entre o 6º e o 7º período. Esta disparidade temporal sugere que o desligamento masculino é um fenômeno de curto prazo, possivelmente associado ao desempenho nas disciplinas iniciais, enquanto a curva feminina apresenta um declínio mais distribuído, indicando que os desafios de permanência para as mulheres podem mudar de natureza ao longo da trajetória (SANTOS et al., 2024).

A manutenção do vínculo feminino acima de 50% até o 9º período em STEM encontra respaldo em Ortiz-Martínez et al. (2023), ao sugerir que as trajetórias femininas são caracterizadas por uma maior persistência temporal, com a decisão de evadir ocorrendo de forma mais tardia que a masculina. Contudo, a utilização de uma classificação binária baseada nos sistemas administrativos limita a identificação da pluralidade de identidades de gênero. No contexto da gestão baseada em evidências, a ausência dessas categorias nos registros de origem oculta disparidades que afetam a precisão das políticas de permanência para grupos como estudantes trans, não binários e outras identidades não catalogadas, cujas trajetórias de risco permanecem sem representação estatística.

A maior sobrevivência feminina observada nos dados corrobora as conclusões de Junior et al. (2012) na área de Física, onde se registra que as mulheres apresentam maior tempo de permanência no sistema e que a decisão de evadir ocorre após períodos prolongados de vínculo, contrastando com o perfil de desligamento precoce do grupo masculino. A distinção entre as curvas de permanência masculina e feminina reforça a necessidade de modelos explicáveis. Conforme Barbosa et al. (2024), a identificação de variáveis sensíveis em sistemas de suporte à decisão permite mitigar vieses, garantindo que as políticas institucionais não perpetuem desigualdades históricas no ambiente acadêmico. Essa persistência superior das mulheres em STEM pode estar alinhado aos resultados de Okimoto et al. (2025), que demonstra que a inserção em atividades de capacitação tecnológica e suporte institucional resulta em melhorias no rendimento e na manutenção do vínculo discente feminino.

No agrupamento de assistência estudantil, as estimativas comprovam o suporte institucional como fator de proteção. No recorte STEM, os beneficiários mantêm uma sobrevivência final de 56,91%, contra 31,95% dos não assistidos. As combinações revelam

que mulheres assistidas em STEM registram 78,18% de sobrevivência e 27,85% do risco de evasão do grupo não assistido com HR de 0,2785 e p-valor de 0,0045. Em contrapartida, homens não assistidos apresentam a menor estimativa de sobrevivência com 24,15%, reforçando que a assistência financeira mantém o vínculo de estudantes com vulnerabilidades externas ao desempenho acadêmico (SOUSA et al., 2024). A sobrevivência final de 56,91% do grupo assistido em STEM ratifica a revisão de Silva e Sampaio (2022), que aponta o auxílio financeiro como o principal preditor de permanência no cenário brasileiro, superando o impacto de variáveis demográficas isoladas."

7.3 PP3: De que forma a combinação de indicadores de significância (*Log – Rank*), risco (*Hazard Ratio*) e magnitude (*Effect Size*) permite quantificar a influência de fatores acadêmicos?

7.3.1 Métricas aplicadas à análise das mudanças de perfil pedagógico curriculares

A validade estatística da melhora curricular é corroborada pelos testes inferenciais. No curso de BSI detalhado na Subseção 6.1.3, o grupo *Depois* manteve a probabilidade de sobrevivência acima de 50,00% até o 8º período com intervalo de confiança de [58,94%; 73,44%], enquanto no grupo *Antes* esse patamar foi atingido no 6º período com intervalo de confiança de [46,68%; 53,63%]. Em LC, conforme os dados de transição de matriz descritos na Subseção 6.1.2, o desligamento de 50,00% no grupo *Antes* ocorria no 5º período com intervalo de confiança de [45,53%; 53,89%], ao passo que no grupo *Depois* este evento é postergado para o intervalo entre o 7º e o 11º período.

A significância dessas variações é confirmada pelo teste *Log-Rank* com p-valor inferior a 0,001 em ambos os cursos. O *Hazard Ratio* indica uma redução do risco de desligamento de 36,68% para LC, com HR de aproximadamente 0,6332, e de 46,43% para BSI, que apresenta HR de 0,5357. A elevação da sobrevida discente em BSI após a implementação dos PIs encontra respaldo em Duran et al. (2023), ao demonstrar que a exposição precoce a desafios práticos de computação fortalece a autoeficácia e atua como um preditor positivo para a permanência.

A magnitude do impacto, medida pelo *Hedges' g*, revela que a reforma em BSI teve um efeito de 0,6818, classificado como grande, enquanto em LC o efeito foi de 0,3359,

valor considerado médio. Essa distinção sugere que as alterações estruturais em BSI, como a inserção de Projetos Interdisciplinares em todos os períodos, surtiram maior impacto na trajetória discente quanto às mudanças implementadas em LC. Este diferencial de magnitude corrobora a tese de [Zamith et al. \(2025\)](#), que identifica a horizontalização de conteúdos e a redução do risco de falha inicial como consequências diretas de reformas que priorizam a integração curricular, como observado nos PIs do Bacharelado.

7.3.2 Métricas aplicadas às trajetórias STEM e institucionais

A validade das variações nas curvas de sobrevivência em STEM é confirmada pelo teste *Log-Rank* na Seção 6.2.4, que identifica a assistência estudantil como a variável de impacto estatístico com p-valor de $4,4667e^{-56}$. Enquanto no grupo de não assistidos as curvas entre gêneros são divergentes com p-valor de $5,6640e^{-7}$, no agrupamento assistido essa disparidade perde significância estatística com p-valor de $1,0030e^0$. Este indicador quantifica a capacidade do suporte institucional em neutralizar disparidades de gênero e promover a equidade na permanência, validando a premissa de [Silva e Sampaio \(2022\)](#) de que a assistência contínua é capaz de mitigar vulnerabilidades que interrompem a trajetória acadêmica.

As estimativas de risco resultantes do *Hazard Ratio* indicam a assistência como fator de proteção em STEM. O grupo *FEMININO* assistido apresenta 27,85% do risco de evasão em relação às estudantes não assistidas e p-valor de 0,0045, sendo o maior indicador de resiliência com 78,18% de sobrevivência no 10º período. Em contrapartida, o turno *NOTURNO* apresenta riscos de desligamento 59,70% superiores aos do turno *MATUTINO* entre os não assistidos, com HR de 1,5970 e p-valor de 0,0010. O intervalo entre o 2º e o 6º período concentra a maior probabilidade de desligamento; homens não assistidos atingem o limiar de 50,00% de sobrevivência no 6º período (49,80%), enquanto o grupo masculino assistido encerra a observação com 43,66% de vínculo, padrão que converge com os achados de [Souza et al. \(2022\)](#), que identificam a fase inicial do curso como o período de maior exposição ao insucesso em disciplinas de base.

A magnitude do impacto, medida pelo teste de *Effect-size* (Hedges' *g*), indica a predominância da assistência sobre as variáveis acadêmicas. No turno *MATUTINO*, a diferença entre assistidos e não assistidos registra um tamanho de efeito de $g = 0,3522$,

enquanto a disparidade entre *MASCULINO* e *FEMININO* assistidos apresenta $g = 0,0035$. A convergência entre o risco (HR) e a magnitude (g) indica que os intervalos entre o 1º e o 4º período constituem a fase de maior vulnerabilidade, sugerindo que o suporte financeiro e a tutoria pedagógica devem ser priorizados nesta janela para mitigar o desligamento nos perfis de maior risco.

Os resultados em STEM convergem com a tendência do cenário institucional, onde a assistência estudantil atua como o principal fator de proteção. A combinação de mulheres assistidas em STEM apresenta o maior indicador de permanência com 78,18% de sobrevivência no 10º período, apresentando HR de 0,2785 e p-valor de 0,0045. A tendência de risco do turno *NOTURNO* observada na instituição é corroborada em STEM, especialmente para homens não assistidos que atingem o limiar de 50,00% de sobrevivência no 6º período com registro de 49,80% e encerram a observação com 27,58%. O impacto da assistência neste agrupamento apresenta $g = 0,8058$, magnitude superior ao cenário geral da instituição. Essa proteção é superior à observada por [Oliveira et al. \(2019\)](#), sugerindo que o modelo assistencial da UFRPE atua de forma mais incisiva na manutenção do vínculo do que fatores estritamente pedagógicos.

Nesse contexto, o turno pode atuar como uma *proxy* para a condição de vulnerabilidade do aluno-trabalhador. Conforme os achados de [Sorgatto e Cáceres \(2023\)](#), a necessidade de conciliação entre a jornada laboral e as exigências acadêmicas, especialmente em cursos de alta densidade como os de STEM, é o principal vetor de fadiga e desmotivação, elevando o risco de trancamentos. Essa condição é agravada pela carência de suporte institucional específico para o período noturno, o que [Ribeiro et al. \(2021\)](#) identificam como uma falha na infraestrutura de apoio que impacta diretamente a satisfação e a persistência do discente.

A disparidade observada entre os turnos revela que a democratização do acesso, embora tenha ampliado o ingresso, apresenta limitações na garantia da equidade de permanência. [Filipak e Pacheco \(2017\)](#) argumentam que o acesso sem políticas de acompanhamento efetivas gera uma “exclusão interna”, onde o estudante do turno noturno é marginalizado por não conseguir acessar programas de monitoria e tutoria concentrados nos turnos diurnos. Esse cenário de desigualdade é corroborado por [Vitelli et al. \(2010\)](#), cujos resultados demonstram que a probabilidade de evasão é drasticamente reduzida apenas quando o auxílio financeiro é capaz de mitigar a dependência de rendas externas,

permitindo uma maior dedicação aos estudos.

Complementarmente, a análise de [Ajoodha et al. \(2020\)](#) em cursos de Ciência da Computação indica que variáveis como o tempo de deslocamento e a modalidade de ingresso são preditores críticos que interagem com o turno, aumentando o risco acadêmico. Por fim, [Alvim et al. \(2024\)](#) ressaltam que a rigidez das matrizes curriculares no período noturno ignora a heterogeneidade desse público, transformando o turno em um filtro seletivo que favorece o desligamento precoce, conforme observado na janela crítica entre o 2º e o 6º período.

7.4 Gestão baseada em evidências

O apoio acadêmico e a tomada de decisões institucional baseiam-se na capacidade de processar dados para subsidiar intervenções ([MARQUES et al., 2023](#)). Nesse sentido, o Painel de Permanência apresentado no Capítulo 5, ao consolidar informações, trata da integração de dados em conjunto com o modelo de análise de sobrevivência para transformar registros acadêmicos em indicadores. Dessa forma, esta integração atua entre a coleta de dados brutos e a geração de conhecimento estratégico, permitindo que a gestão identifique o volume de evasão, o tempo de ocorrência do evento e os fatores associados ao risco acadêmico ([CAVALCANTI et al., 2025](#)). A viabilidade desta análise em diferentes níveis de granularidade com curso, turno e perfil, corrobora o método de BI proposto por [Menolli e Neto \(2021\)](#). Ademais, a opção pela análise de sobrevivência evita a complexidade de modelos de “caixa preta”, conforme discutido por [Mao et al. \(2024\)](#), garantindo que os indicadores de risco sejam interpretáveis para a formulação de intervenções pedagógicas.

Indicadores interpretáveis em contextos sensíveis orientam o uso de métricas computacionais para decisões estratégicas ([BARBOSA et al., 2024](#)). A análise de sobrevivência supera as limitações de métricas transversais ao considerar o tempo de permanência e a censura dos dados [Klitzke e Carvalhaes \(2023\)](#), permitindo que a ferramenta identifique marcos de vulnerabilidade e quantifique o efeito de variáveis institucionais. Como resultado, observa-se que a permanência feminina em STEM atinge 62%, enquanto a masculina é de 56% no 6º semestre, fornecendo subsídios para o direcionamento de políticas de assistência.

8 Considerações finais

A evasão no ensino superior configura-se como um fenômeno multivariado e multifacetado, que envolve a interação de fatores internos e externos à instituição. Esta dissertação descreveu a construção e a aplicação de um modelo de análise de sobrevivência para a obtenção de indicadores estatísticos de permanência acadêmica. A investigação buscou compreender os padrões temporais e os períodos de vulnerabilidade discente, integrando a análise de sobrevivência aos fundamentos da gestão baseada em evidências para subsidiar decisões estratégicas.

Os resultados demonstram que a dinâmica de evasão varia conforme a estrutura curricular e o perfil socioeconômico, revelando que matrizes curriculares inflexíveis apresentam menor capacidade de absorção da pluralidade do corpo discente. A análise identificou que os quatro primeiros períodos letivos constituem a janela de maior risco de desligamento em todos os cenários. No entanto, a implementação dos novos Projetos Pedagógicos de Curso resultou no deslocamento temporal das curvas de sobrevivência, como observado em LC, onde o marco de 50% de evasão foi postergado do 5º para o intervalo entre o 7º e 11º período, ao promover a horizontalização de conteúdos e a redução de pré-requisitos sequenciais. Tais mudanças, que mantiveram a sobrevivência em BSI em 83,54% no 3º período, demonstram que a flexibilização da matriz atua como um mecanismo de justiça curricular ao mitigar o impacto do choque acadêmico inicial sobre os perfis que não dispõem de tempo integral para dedicação exclusiva aos estudos.

No que tange aos fatores estruturais e socioeconômicos, observa-se uma convergência de riscos onde a vulnerabilidade financeira potencializa a fragilidade logística, conforme evidenciado pela sobrevivência final de 23,04% no turno noturno de STEM. Este cenário caracteriza uma barreira de permanência para o aluno-trabalhador. Em contrapartida, os dados revelaram uma expressiva resiliência feminina no recorte tecnológico. Embora as mulheres representem a minoria com 20,3% dos ingressos, o grupo apresentou marcos de evasão significativamente mais tardios que o grupo masculino, com limiar de 50% de evasão apenas no 9º período, enquanto o grupo masculino atinge este marco entre o 6º e o 7º período. Este comportamento indica que, uma vez neutralizadas as barreiras materiais pela assistência estudantil, as trajetórias femininas tendem a uma estabilidade superior, posicionando o suporte financeiro como o principal regulador de equidade da instituição.

A eficácia desta assistência é confirmada pela magnitude do impacto ($g = 0,8058$) no turno noturno de STEM, que superou as variáveis puramente acadêmicas e manteve a maioria dos beneficiados vinculada durante o período observado. Diante dessas evidências, a gestão acadêmica deve considerar o escalonamento da assistência estudantil, priorizando o suporte financeiro e a tutoria pedagógica no primeiro biênio do curso para proteger os discentes na fase de maior exposição ao insucesso. Adicionalmente, a utilização de indicadores viabiliza fundamentar políticas de reingresso e reintegração voltadas ao acolhimento de discentes que interromperam o vínculo por vulnerabilidades materiais em períodos críticos identificados pelo modelo.

Complementarmente ao suporte financeiro, as políticas de permanência devem contemplar o fortalecimento de programas de mentorias e acolhimento voltados ao público feminino em áreas tecnológicas. A neutralização da disparidade de gênero pelo suporte institucional ($g = 0,0035$) justifica o investimento nessas ações para reduzir o isolamento em ambientes majoritariamente masculinos. No âmbito logístico, a identificação do turno noturno como um eixo de risco sinaliza a necessidade de oferta de monitorias e serviços de apoio acadêmico em horários alternativos, para que o aluno-trabalhador acesse recursos de suporte equivalentes aos disponíveis nos turnos diurnos.

Tais ações, integradas à revisão de cargas horárias e exigências de pré-requisitos em componentes de alta densidade matemática e computacional, formam um suporte para a neutralização das disparidades socioeconômicas. Ao alinhar a flexibilização curricular com a segurança material provida pela assistência, a instituição atua na qualificação do percurso acadêmico para que o suporte institucional seja proporcional ao risco identificado em cada trajetória. Essa abordagem permite que o investimento público seja aplicado de forma direcionada, focando nos períodos de exposição ao insucesso e nos perfis de vulnerabilidade.

O papel da universidade pública transcende a democratização do acesso; reside na garantia de condições que assegurem a permanência. Tratar a evasão não se limita a uma meta de eficiência administrativa, mas representa o compromisso institucional com a formação de qualidade de profissionais que, inseridos na sociedade, contribuirão para o desenvolvimento científico e social do país. Ao mitigar disparidades e fortalecer o suporte acadêmico através de políticas de assistência, a universidade cumpre seu imperativo ético de devolver cidadãos capacitados.

Dessa forma, o suporte à permanência discente fundamentado em dados atua como um mecanismo de mobilidade e transformação social. O uso de evidências estatísticas fornece a base para que o investimento institucional resulte em uma formação sólida e equânime, viabilizando que o sucesso acadêmico seja uma possibilidade real para todos os perfis discentes que ingressam na instituição.

Com base nos resultados, trabalhos futuros podem expandir a aplicação da modelagem para o monitoramento do tempo até a formação, permitindo estimar o período para a integralização curricular. No âmbito metodológico, a inclusão de dados com truncamento e censura à esquerda possibilita o tratamento de discentes que ingressaram em períodos anteriores ao início da observação, o que reduz o viés de seleção. A análise pode ser estendida ao nível das disciplinas para compreender os padrões temporais de aprovação e identificar os componentes curriculares que prolongam o fluxo dos cursos.

A aplicação do modelo em outros centros de ensino, campi ou áreas fora de STEM permite verificar a repetibilidade da capacidade de neutralização das disparidades de gênero pela assistência estudantil em diferentes contextos disciplinares. Além disso, a estratificação da análise por agrupamentos como etnia, renda e modalidade de ingresso possibilita identificar interseccionalidades no risco de desligamento. Adicionalmente, a realização de pesquisa qualitativa, por meio de entrevistas semiestruturadas e grupos focais com os perfis identificados como de risco, visa explorar fatores externos e dimensões subjetivas, como o clima organizacional, o sentimento de pertencimento e os conflitos de conciliação entre jornada laboral e vida acadêmica, que contribuem para a interrupção do vínculo.

Referências

- AARTS, S.; AKKER, M. V. D.; WINKENS, B. The importance of effect sizes. **The European Journal of General Practice**, Taylor & Francis, v. 20, n. 1, p. 61–64, 2014.
- AJODHA, R.; JADHAV, A.; DUKHAN, S. Forecasting learner attrition for student success at a south african university. In: **Conference of the South African Institute of Computer Scientists and Information Technologists 2020**. [S.l.: s.n.], 2020. p. 19–28.
- ALMEIDA, J. B.; SCHIMIGUEL, J. Avaliação sobre as causas da evasão escolar no ensino superior: estudo de caso no curso de licenciatura em física no instituto federal do maranhão. **Revista de Ensino de Ciências e Matemática**, v. 2, n. 2, p. 167–178, 2011.
- ALVIM, Í. V.; BITTENCOURT, R. A.; DURAN, R. S. Evasão nos cursos de graduação em computação no brasil. In: SBC. **Anais do IV Simpósio Brasileiro de Educação em Computação**. [S.l.], 2024. p. 1–11.
- AMERI, S.; FARD, M. J.; CHINNAM, R. B.; REDDY, C. K. Survival analysis based framework for early prediction of student dropouts. In: **Proceedings of the 25th ACM international on conference on information and knowledge management**. [S.l.: s.n.], 2016. p. 903–912.
- ANDIFES. **Diplomação, Retenção e Evasão nos Cursos de Graduação em Instituições de Ensino Superior Públicas**. 1996. Disponível em: <https://www.gov.br/inep/pt-br/centrais-de-conteudo/acervo-linha-editorial/publicacoes-institucionais/estatisticas-e-indicadores-educacionais/censo-da-educacao-superior-2022-notas-estatisticas>. Acesso em: 25 de janeiro 2024.
- BARBOSA, T.; FREITAS, N.; CAVALCANTI, L.; BATISTA, M. da C.; GOUVEIA, R.; ALVES, G. Construção dinâmica de modelos de learning analytics explicáveis e justos aplicados ao acompanhamento de estudantes de graduação. In: **Anais do XXXV Simpósio Brasileiro de Informática na Educação**. Porto Alegre, RS, Brasil: SBC, 2024. p. 1918–1930. ISSN 0000-0000. Disponível em: <https://sol.sbc.org.br/index.php/sbie/article/view/31367>.
- BLAND, J. M.; ALTMAN, D. G. Survival probabilities (the kaplan-meier method). **Bmj**, British Medical Journal Publishing Group, v. 317, n. 7172, p. 1572–1580, 1998.
- BRITO, B.; MELLO, R.; ALVES, G. Identificação de atributos relevantes na evasão no ensino superior público brasileiro. In: **Anais do XXXI Simpósio Brasileiro de Informática na Educação**. Porto Alegre, RS, Brasil: SBC, 2020. p. 1032–1041. ISSN 0000-0000. Disponível em: <https://sol.sbc.org.br/index.php/sbie/article/view/12859>.
- CARMINATI, G.; AUGUSTO, R.; DALLABRIDA, N.; TEIVE, R. Mineração de dados educacionais visando a identificação da evasão no ensino superior. **Anais do Computer on the Beach**, v. 11, n. 1, p. 461–468, 2020.
- CAVALCANTI, L.; BARROS, A.; FALCÃO, T.; BATISTA, M. da C.; CRISTINO, C.; ALVES, G. Avaliando o impacto da mudança do projeto pedagógico de cursos sobre a evasão através da análise de sobrevivência. In: **Anais do XXXV Simpósio Brasileiro**

de Informática na Educação. Porto Alegre, RS, Brasil: SBC, 2024. p. 2617–2626. ISSN 0000-0000. Disponível em: <<https://sol.sbc.org.br/index.php/sbie/article/view/31422>>.

CAVALCANTI, L.; FREITAS, N.; MARQUES, E.; LIMA, S.; MELLO, R. F.; CRISTINO, C. T.; ALVES, G. et al. Avaliação do impacto do gênero e da assistência estudantil sobre a sobrevivência dos estudantes em cursos de stem. In: SBC. **Simpósio Brasileiro de Informática na Educação (SBIE)**. [S.l.], 2025. p. 731–745.

CHEN, Y.; JOHRI, A.; RANGWALA, H. Running out of stem: a comparative study across stem majors of college students at-risk of dropping out early. In: **Proceedings of the 8th international conference on learning analytics and knowledge**. [S.l.: s.n.], 2018. p. 270–279.

COHEN, J. **Statistical power analysis for the behavioral sciences**. [S.l.]: routledge, 2013.

COLOSIMO, E. A.; GIOLO, S. R. **Análise de sobrevivência aplicada**. [S.l.]: Editora Blucher, 2021.

COSTA, F. J. d.; BISPO, M. d. S.; PEREIRA, R. d. C. d. F. Dropout and retention of undergraduate students in management: a study at a brazilian federal university. **RAUSP Management Journal**, SciELO Brasil, v. 53, p. 74–85, 2018.

COX, D. R. **Analysis of survival data**. [S.l.]: Chapman and Hall/CRC, 2018.

CUMMING, G. **Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis**. [S.l.]: Routledge, 2013.

CUZICK, J. The efficiency of the proportions test and the logrank test for censored survival data. **Biometrics**, JSTOR, p. 1033–1039, 1982.

DEY, E. L.; ASTIN, A. W. Statistical alternatives for studying college student retention: A comparative analysis of logit, probit, and linear regression. **Research in higher education**, Springer, v. 34, n. 5, p. 569–581, 1993.

DURAN, R.; BIM, S. A.; GIMENES, I.; RIBEIRO, L.; CORREIA, R. C. M. Potential factors for retention and intent to drop-out in brazilian computing programs. **ACM Transactions on Computing Education**, ACM New York, NY, v. 23, n. 3, p. 1–33, 2023.

ELANDT-JOHNSON, R. C.; JOHNSON, N. L. **Survival models and data analysis**. [S.l.]: John Wiley & Sons, 1980.

FALCÃO, T. P.; ARAÚJO, D.; FRANÇA, R.; ANDRADE, E.; FRANÇA, C. Currículo da licenciatura em computação: uma proposta alinhada às novas diretrizes e demandas contemporâneas. In: **Anais dos Workshops do Congresso Brasileiro de Informática na Educação**. [S.l.: s.n.], 2018. p. 1108.

FILIPAK, S. T.; PACHECO, E. F. H. A democratização do acesso à educação superior no brasil. **Revista Diálogo Educacional**, Pontifícia Universidade Católica do Paraná-PUCPR, v. 17, n. 54, p. 1241–1268, 2017.

GARCIA, L. M. L. d. S.; GOMES, R. S. Causas da evasão em cursos de ciências exatas: uma revisão da produção acadêmica. **Revista Educar Mais. Pelotas, RS. Vol. 6 (2022), p. 937-957.**, 2022.

GISI, M. L. A educação superior no brasil e o caráter de desigualdade do acesso e da permanência. **Revista Diálogo Educacional**, PUCPR, v. 6, n. 17, p. 97–112, 2006.

GUTIERREZ-PACHAS, D. A.; GARCIA-ZANABRIA, G.; CUADROS-VARGAS, A. J.; CAMARA-CHAVEZ, G.; POCO, J.; GOMEZ-NIETO, E. How do curricular design changes impact computer science programs?: A case study at san pablo catholic university in peru. **Education Sciences**, MDPI, v. 12, n. 4, p. 242, 2022.

GUTIERREZ-PACHAS, D. A.; GARCIA-ZANABRIA, G.; CUADROS-VARGAS, E.; CAMARA-CHAVEZ, G.; GOMEZ-NIETO, E. Supporting decision-making process on higher education dropout by analyzing academic, socioeconomic, and equity factors through machine learning and survival analysis methods in the latin american context. **Education Sciences**, MDPI, v. 13, n. 2, p. 154, 2023.

HABERMAN, S. Survival models and data analysis. by regina c. elandt-johnson and norman l. johnson. john wiley and sons, new york. 1980. xvi 457 pages. **Journal of the Institute of Actuaries**, v. 108, n. 2, p. 295–296, 1981.

HOED, R. M. Análise da evasão em cursos superiores: o caso da evasão em cursos superiores da área de computação. **Anais do XXXI Simpósio Brasileiro de Informática na Educação**, 2016.

INEP. Metodologia de Cálculo dos Indicadores de Fluxo da Educação Superior. 2017. Disponível em: <<https://portal.mec.gov.br/component/tags/tag/32044-censo-da-educacao-superior>>. Acesso em: 28 de janeiro de 2026.

INEP. RESUMO TÉCNICO DO CENSO DA EDUCAÇÃO SUPERIOR 2023 DIRETORIA DE ESTATÍSTICAS EDUCACIONAIS DE Educação. 2023. Disponível em: <<https://www.gov.br/inep/pt-br/centrais-de-conteudo/acervo-linha-editorial/publicacoes-institucionais/estatisticas-e-indicadores-educacionais/censo-da-educacao-superior-2022-notas-estatisticas>>. Acesso em: 27 de janeiro 2026.

_____. Censo Escolar da Educação Superior 2024: Notas Estatísticas. 2024. Disponível em: <https://download.inep.gov.br/publicacoes/institucionais/estatisticas_e_indicadores/notas_estatisticas_censo_escolar_da_educacao_superior_2024.pdf>. Acesso em: 27 de janeiro 2026.

JR, D. W. H.; LEMESHOW, S.; MAY, S. **Applied survival analysis: regression modeling of time-to-event data**. [S.l.]: John Wiley & Sons, 2008. v. 618.

JUNIOR, P. L.; SILVEIRA, F. L. d.; OSTERMANN, F. Análise de sobrevivência aplicada ao estudo do fluxo escolar nos cursos de graduação em física: um exemplo de uma universidade brasileira. **Revista brasileira de ensino de física**, SciELO Brasil, v. 34, p. 1403, 2012.

KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. **Journal of the American statistical association**, Taylor & Francis, v. 53, n. 282, p. 457–481, 1958.

KLEINBAUM, D. G.; KLEIN, M.; KLEINBAUM, D. G.; KLEIN, M. Parametric survival models. **Survival Analysis: A Self-Learning Text**, Springer, p. 289–361, 2012.

KLITZKE, M.; CARVALHAES, F. Fatores associados à evasão de curso na ufrj: Uma análise de sobrevivência. **Educação em Revista**, SciELO Brasil, v. 39, p. e37576, 2023.

LAKENS, D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and anovas. **Frontiers in psychology**, Frontiers Media SA, v. 4, p. 863, 2013.

LAWLESS, J. F. **Statistical models and methods for lifetime data**. [S.l.]: John Wiley and Sons, 2011. 362 p.

LEE, E. T.; WANG, J. **Statistical methods for survival data analysis**. [S.l.]: John Wiley & Sons, 2003. v. 476.

MAIA, J. de S. Z.; BUENO, A. P. A.; SATO, J. R. Applications of artificial intelligence models in educational analytics and decision making: A systematic review. **World**, MDPI, v. 4, n. 2, p. 288–313, 2023.

MAO, S.; ZHANG, C.; SONG, Y.; WANG, J.; ZENG, X.-J.; XU, Z.; WEN, Q. Time series analysis for education: Methods, applications, and future directions. **arXiv preprint arXiv:2408.13960**, 2024.

MARQUES, E.; CAVALCANTI, L.; BARBOSA, T.; ALVES, G. Sabia: Uma plataforma para auxiliar a gestão baseada em evidências nas instituições de ensino superior. In: SBC. **Anais do II Workshop de Aplicações Práticas de Learning Analytics em Instituições de Ensino no Brasil**. [S.l.], 2023. p. 71–80.

MENOLLI, A.; NETO, J. C. Uma análise do perfil dos cursos de licenciatura em computação no brasil. **Revista Brasileira de Informática na Educação**, v. 29, p. 01–24, 2021.

MILLS, M. **Introducing Survival and Event History Analysis**. London: SAGE Publications, 2010.

Ministério da Educação - Secretaria de Educação Profissional e Tecnológica. **Documento Orientador Para a Superação da Evasão e Retenção na Rede Federal de Educação Profissional, Científica e Tecnológica**. 2014. Disponível em: <http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=110401-documento-orientador-evasao-retencao-vfinal&category_slug=abril-2019-pdf&Itemid=30192>. Acesso em: 25 de janeiro 2024.

MONTGOMERY, D. C.; RUNGER, G. C.; CALADO, V. **Estatística aplicada e probabilidade para engenheiros**. 4. ed. [S.l.]: Grupo Gen-LTC, 2000. 491 p.

MOURA, F.; MANDARINO, P.; SILVA, S. da. Evasão escolar no ensino superior: Análise quantitativa no curso de licenciatura em física do ifpa campus bragança. **Revista Brasileira de Ensino de Física**, SciELO Brasil, v. 42, p. e20200044, 2020.

NA, J.; KIM, K. W.; KIM, H. G. A study on exploiting temporal patterns in semester records for efficient student dropout prediction. **Electronics**, MDPI, v. 14, n. 22, p. 4356, 2025.

NASCIMENTO, R. P.; SOUSA, G. D. de; NASCIMENTO, M. W.; REIS, J. R. Analisando os dados do percurso acadêmico do curso de sistemas de informação para auxiliar na revisão do ppc. In: SBC. **Anais do II Simpósio Brasileiro de Educação em Computação**. [S.l.], 2022. p. 189–196.

NIEROTKA, R. L.; SALATA, A.; MARTINS, M. K. Fatores associados à evasão no ensino superior: Um estudo longitudinal. **Cadernos de Pesquisa**, SciELO Brasil, v. 53, p. e09961, 2023.

OECD. **Education at a Glance 2024: OECD Indicators**. 2024. Disponível em: <https://www.oecd.org/en/publications/education-at-a-glance-2024_c00cad36-en.html>. Acesso em: 09 de fevereiro 2026.

_____. **Education at a Glance 2025: OECD Indicators**. 2025. Disponível em: <<https://doi.org/10.1787/1c0d9c79-en>>. Acesso em: 27 de janeiro 2026.

OKIMOTO, A. Y.; LOBATO, Y. de J.; SATO, S. d. C.; NOGUEIRA, J. C.; SANTOS, A. L. dos; CARVALHO, L. S. de; LAUSCHNER, T.; NAKAMURA, F. G.; NAKAMURA, E. F. et al. Quantificando o impacto do projeto super nos cursos de ciência da computação e engenharia de software da universidade federal do Amazonas. In: SBC. **Workshop sobre Educação em Computação (WEI)**. [S.l.], 2025. p. 1172–1182.

OLIVE, D. J. A simple confidence interval for the median. **Manuscript**, 2005.

OLIVEIRA, C. H. M. de; SANTOS, F. R. T.; LEITINHO, J. L.; FARIAS, L. G. A. T. Busca dos fatores associados à evasão: um estudo de caso no campus universitário da ufc em cratêus. **Revista Internacional de Educação Superior**, v. 5, p. e019006–e019006, 2019.

ORTIZ-MARTÍNEZ, G.; VÁZQUEZ-VILLEGAS, P.; RUIZ-CANTISANI, M. I.; DELGADO-FABIÁN, M.; CONEJO-MÁRQUEZ, D. A.; MEMBRILLO-HERNÁNDEZ, J. Analysis of the retention of women in higher education stem programs. **Humanities and Social Sciences Communications**, Palgrave, v. 10, n. 1, p. 1–14, 2023.

PAN, F.; HUANG, B.; ZHANG, C.; ZHU, X.; WU, Z.; ZHANG, M.; JI, Y.; MA, Z.; LI, Z. A survival analysis based volatility and sparsity modeling network for student dropout prediction. **PloS one**, Public Library of Science San Francisco, CA USA, v. 17, n. 5, p. e0267138, 2022.

PASCOAL, T.; BRITO, D. M. de; ANDRADE, L.; RÊGO, T. G. do. Evasao de estudantes universitários: diagnóstico a partir de dados acadêmicos e socioeconômicos. In: **Brazilian Symposium on Computers in Education (Simposio Brasileiro de Informática na Educação-SBIE)**. [S.l.: s.n.], 2016. v. 27, n. 1, p. 926.

PATINO, C. M.; FERREIRA, J. C. Intervalos de confiança: uma ferramenta útil para estimar o tamanho do efeito no mundo real. **Jornal Brasileiro de Pneumologia**, SciELO Brasil, v. 41, p. 565–566, 2015.

PINHEIRO, M. A. L.; SILVA, J. C. da; SOUZA, B. F. de. Aprendizado de máquina aplicado à análise de evasão no ensino superior. **Anais do Computer on the Beach**, v. 9, p. 512–521, 2018.

RIBEIRO, I. M.; CORREIA, W. F. M.; CAMPOS, F. Setores acadêmicos que interferem na satisfação do aluno no ensino superior. **Acta Scientiarum. Education**, Editora da Universidade Estadual de Maringá, v. 43, 2021.

RICH, J. T.; NEELY, J. G.; PANIELLO, R. C.; VOELKER, C. C.; NUSSENBAUM, B.; WANG, E. W. A practical guide to understanding kaplan-meier curves. **Otolaryngology—Head and Neck Surgery**, SAGE Publications Sage CA: Los Angeles, CA, v. 143, n. 3, p. 331–336, 2010.

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)**, SciELO Brasil, v. 49, p. 337–373, 2019.

SANTOS, M. T. S.; GASPARINI, I.; FRIGO, L. B. Análise da participação feminina nos cursos da área de computação: Um olhar sobre os projetos do programa meninas digitais. In: SBC. **Simpósio Brasileiro de Informática na Educação (SBIE)**. [S.l.], 2024. p. 2259–2272.

SILVA, A. C.; CORREA, C. S.; COELHO, D. de A.; SILVA, D. T. da; NETO, L. L. F.; XAVIER, M. M.; REIS, R. da S.; ADENILSON, F.; ROCHA, P. A. S. Análise quantitativa e qualitativa da evasão dos discentes do curso de engenharia ambiental do instituto federal da bahia analysis quantitative and qualitative of evasion of environmental engineering course students of the federal institute of bahia. **XIV International Conference on Engineering and Technology Education**, 2016.

SILVA, L. G. d. L. **O impacto da nova matriz curricular da Licenciatura em Computação no desempenho dos discentes**. Dissertação (B.S. thesis) — Brasil, 2024.

SILVA, P. T. d. F.; SAMPAIO, L. M. B. Políticas de permanência estudantil na educação superior: reflexões de uma revisão da literatura para o contexto brasileiro. **Revista de Administração Pública**, SciELO Brasil, v. 56, p. 603–631, 2022.

SORGATTO, D. W.; CÁCERES, E. N. Uma análise da mudança do perfil dos ingressantes do ensino superior brasileiro. In: SBC. **Anais do II Workshop de Aplicações Práticas de Learning Analytics em Instituições de Ensino no Brasil**. [S.l.], 2023. p. 32–41.

SOUSA, R.; FACHINI-GOMES, J.; HOLANDA, M.; LEÃO, M. Um estudo da evasão no curso de licenciatura em computação da universidade de Brasília. In: **Anais do XXXII Workshop sobre Educação em Computação**. Porto Alegre, RS, Brasil: SBC, 2024. p. 715–725. ISSN 2595-6175. Disponível em: <<https://sol.sbc.org.br/index.php/wei/article/view/29670>>.

SOUZA, J.; KOMATI, K.; ANDRADE, J. Análise de sobrevivência: um estudo de caso em um curso de sistemas de informação. In: **Anais do XXX Workshop sobre Educação em Computação**. Porto Alegre, RS, Brasil: SBC, 2022. p. 392–403. ISSN 2595-6175. Disponível em: <<https://sol.sbc.org.br/index.php/wei/article/view/20847>>.

STARE, J.; MAUCORT-BOULCH, D. Odds ratio, hazard ratio and relative risk. **Metodološki zvezki/Advances in Methodology and Statistics**, v. 13, n. 1, p. 59–67, 2016.

SULLIVAN, G. M.; FEINN, R. Using effect size—or why the p value is not enough. **Journal of graduate medical education**, The Accreditation Council for Graduate Medical Education Suite 2000, 515 . . . , v. 4, n. 3, p. 279–282, 2012.

TSIAKMAKI, M.; KOSTOPOULOS, G.; KOUTSONIKOS, G.; PIERRAKEAS, C.; KOTSIANTIS, S.; RAGOS, O. Predicting university students' grades based on previous academic achievements. In: IEEE. **2018 9th international conference on information, Intelligence, Systems and Applications (IISA)**. [S.l.], 2018. p. 1–6.

VASCONCELOS, V.; ANDRADE, E. Análise da evasão de alunos na licenciatura em computação. In: SBC. **Anais do XXVI Workshop sobre Educação em Computação**. [S.l.], 2018.

VITELLI, R. F.; ROCHA, C. S.; FRITSCH, R. Estudo sobre evasão nos cursos de graduação de uma instituição de ensino superior privada: aplicação de regressão logística. **Programa de Observatório de Educação INEP/CAPES, Núcleo em Rede, Projeto**, v. 44, 2010.

WANG, H.; SHENG, L.; CHEN, D. Perspective chapter: Using effect sizes to study the survival difference between two groups. In: **Recent Advances in Biostatistics**. [S.l.]: IntechOpen, 2023.

WELLEK, S. A log-rank test for equivalence of two survivor functions. **Biometrics**, JSTOR, p. 877–881, 1993.

ZAMITH, M. P. de M.; NASCENTE, J. M.; ZAMITH, S.; LESSA, N. C. Análise preliminar da influência da mudança curricular nos índices de evasão discente no curso de ciência da computação da ufrj. In: SBC. **Simpósio Brasileiro de Educação em Computação (EDUCOMP)**. [S.l.], 2025. p. 115–126.