

Avaliação do Impacto do Gênero e da Assistência Estudantil sobre a Sobrevivência dos Estudantes em Cursos de STEM

Lhaíslla Cavalcanti¹, Nathan Freitas¹, Ebony Marques¹, Maria da Conceição Moraes Batista¹, Socorro Lima¹, Rafael Ferreira Mello¹, Claudio Tadeu Cristino¹ e Gabriel Alves¹

¹Universidade Federal Rural de Pernambuco (UFRPE)
Recife – PE – Brasil

{lhaisslla.cavalcanti, nathan.freitas, ebony.marquesr, maria.cmbatista, socorro.lima, rafael.mello, claudio.cristino, gabriel.alves}@ufrpe.br

Abstract. *Courses in the fields of Science, Technology, Engineering, and Mathematics (STEM) often present challenges related to student retention, with disparities observed between different groups. This study seeks to analyse the academic retention of 7,432 students in STEM courses at UFRPE. To this end, it employs survival analysis, using the Log Rank, Effect Size, and Hazard Ratio tests to examine retention patterns based on gender, shift offered, and access to student assistance. The comparison between genders in the STEM area reveals a statistically significant difference from the 6th semester onwards, with women having an approximately 62% probability of retention and men around 56%, while students who do not receive assistance and are enrolled in evening classes have a significantly higher risk of dropping out. The results provide support for institutional policies that promote equity and retention.*

Resumo. *Cursos das áreas de Ciência, Tecnologia, Engenharia e Matemática (STEM) frequentemente apresentam desafios relacionados à permanência estudantil, com disparidades observadas entre diferentes grupos. Este estudo busca analisar a permanência acadêmica de 7.432 estudantes em cursos de STEM da UFRPE. Para isso, emprega análise de sobrevivência, utilizando os testes Log Rank, Effect Size e Hazard Ratio, a fim de examinar os padrões de permanência em função do gênero, turno de oferta e acesso à assistência estudantil. A comparação entre gêneros na área de STEM revela diferença estatisticamente significativa a partir do 6º período, com mulheres apresentando aproximadamente 62% de probabilidade de permanência e homens, cerca de 56%, enquanto estudantes não assistidos e matriculados no turno noturno têm um risco significativamente maior de evasão. Os resultados fornecem subsídios para políticas institucionais que promovam a equidade e permanência.*

1. Introdução

A evasão no ensino superior constitui um processo multifacetado que reflete desigualdades estruturais e fragilidades nas esferas institucional, individual e social [Santos et al. 2021, Alvim et al. 2024, Ribeiro et al. 2021]. Compreender os padrões de abandono e os fatores associados à permanência é fundamental para subsidiar políticas educacionais eficazes, bem como para aprimorar a gestão acadêmica e administrativa em múltiplos níveis [Vitelli et al. 2010].

Nos cursos de Ciência, Tecnologia, Engenharia e Matemática (STEM), esses desafios assumem contornos específicos, marcados pela persistente desigualdade de gênero, barreiras socioeconômicas, dificuldades acadêmicas [Santos et al. 2024, Silva 2024] e condição financeira do discente [Garcia and Gomes 2022, Almeida and Schimiguel 2011]. Tais especificidades evidenciam a necessidade de análises para esses cursos.

A Análise de Sobrevivência (AS) é um método estatístico que acompanha a ocorrência de eventos ao longo do tempo, permitindo identificar padrões e variáveis associadas à permanência e ao abandono no ensino superior [Souza et al. 2022, Nierotka et al. 2023, Saccaro et al. 2019, Chen et al. 2018]. Essa abordagem possibilita estimar a duração da trajetória acadêmica e responder a questões temporais, como o momento em que ela ocorre e os períodos de maior propensão ao abandono, subsidiando análises comparativas entre diferentes grupos [Klitzke and Carvalhaes 2023, Lima Junior et al. 2012, de Moraes Zamith et al. 2025, Sousa et al. 2024, Cavalcanti et al. 2024]. Embora útil para descrever padrões e orientar intervenções institucionais, a AS não permite inferências causais diretas, o que justifica o uso de abordagens estatísticas complementares.

Para aprofundar a compreensão dos padrões de permanência, este trabalho analisa os cursos de Licenciatura em Matemática, Física e Computação, além dos cursos de Bacharelado em Sistemas de Informação e Ciência da Computação na área de STEM da Universidade Federal Rural de Pernambuco (UFRPE), que possui particularidades curriculares e programas de acompanhamento acadêmico que justificam a escolha. A análise compara os estudantes com base em gênero, turno de oferta e acesso à assistência estudantil. A metodologia utilizada é o estimador de *Kaplan-Meier*, complementado pelas métricas estatísticas *Log-Rank*, *Hazard Ratio* e *Effect Size*, com o objetivo de responder às seguintes questões de pesquisa:

1. **PP 01:** Há diferenças nos padrões temporais de evasão entre estudantes dos cursos STEM e a permanência geral dos estudantes na instituição?
2. **PP 02:** Como gênero, assistência estudantil e turno de matrícula influenciam a probabilidade de permanência ao longo do tempo?
3. **PP 03:** De que forma as métricas estatísticas, como o teste *Log-Rank*, o *Effect Size* e o *Hazard Ratio*, permitem identificar e quantificar diferenças nos padrões de permanência estudantil na análise de sobrevivência?

2. Trabalhos Relacionados

A evasão no ensino superior é um fenômeno amplamente discutido na literatura, sendo analisado por meio de diferentes técnicas estatísticas e computacionais. No eixo acadêmico, diversos estudos utilizando a Análise de Sobrevivência identificaram que reprovações em disciplinas iniciais estão fortemente associadas à evasão, especialmente em cursos da área de STEM [Santos et al. 2021, Klitzke and Carvalhaes 2023]. O índice de rendimento acadêmico também se destaca como característica relevante [Santos et al. 2024], e a análise de sobrevivência tem sido eficaz na avaliação de reformas curriculares, indicando aumento na permanência discente após mudanças no perfil pedagógico dos cursos [Cavalcanti et al. 2024]. Outras abordagens, como a regressão logística, confirmam que o aumento no tempo de permanência tende a reduzir a probabilidade de evasão [Vitelli et al. 2010, Pachas et al. 2021].

Quanto às características sociodemográficas, estudantes do sexo feminino, com até vinte anos, e inseridos em cursos de bacharelado apresentam maior probabilidade de permanência [Nierotka et al. 2023, Almeida and Schimiguel 2011]. Em contraste, homens mais velhos tendem a evadir com maior frequência [Saccaro et al. 2019, Garcia and Gomes 2022]. O tempo de integralização do curso também aparece como variável explicativa relevante [Souza et al. 2022]. No campo do apoio institucional e motivacional, o recebimento de assistência social e financeira está positivamente associado à permanência. Análises que utilizam testes como o qui-quadrado de Pearson e medidas como a razão de chances e o V de Cramer confirmam essa relação [Alvim et al. 2024]. A satisfação acadêmica, a expectativa de carreira e a adaptação ao mercado de trabalho são fatores subjetivos igualmente relevantes, sendo investigados por meio de abordagens quantitativas [Ribeiro et al. 2021, Duran et al. 2023, Silva 2024]. Destacam-se ainda, estudos que associam a menor permanência em cursos STEM a barreiras institucionais e financeiras [de Lima and Zago 2018, Silva et al. 2016, Ortiz-Martínez et al. 2023].

Os achados relatados evidenciam a relevância de indicadores na compreensão da permanência estudantil sob diferentes perspectivas. Este trabalho se distingue por adotar uma abordagem estatística comparativa entre grupos, fundamentada na análise de sobrevivência e no teste de hipótese *Log-rank*, com a incorporação da medida de tamanho de efeito *Hedges' g* e da razão de risco *Hazard ratio*, sendo disponibilizados em uma plataforma interativa para análise de gestores. Essa combinação metodológica permite analisar como a probabilidade de permanência varia ao longo do tempo, bem como quantificar a magnitude e a direção dos efeitos observados. Isso amplia a capacidade explicativa dos resultados e facilita interpretações sobre padrões e limitações.

3. Método

O método apresentado adota a análise de sobrevivência como abordagem estatística para descrever a dinâmica da permanência no ensino superior ao longo do tempo, concentrando-se na comparação entre curvas gerais e aquelas correspondentes aos estudantes do gênero feminino e do gênero masculino, bem como daqueles que recebem assistência estudantil e daqueles que não a recebem, em cursos da área de STEM. Essas comparações têm como objetivo identificar quais fatores exercem maior impacto.

A análise foi conduzida a partir dos processos de aquisição e tratamento dos dados, seguida pela modelagem por Análise de Sobrevivência. As diferenças entre os grupos foram avaliadas utilizando o teste *Log-Rank* para significância estatística, o *Effect Size (ES)* para quantificar a magnitude prática das distinções e o *Hazard Ratio* para determinar os riscos associados. Esses testes foram escolhidos por combinarem robustez estatística e interpretabilidade, enquanto alternativas como *Wilcoxon* ou modelos paramétricos não foram adotadas por exigirem suposições sobre a distribuição dos dados ou serem menos intuitivas para gestores acadêmicos.

3.1. Coleta e Processamento dos dados

O presente trabalho foi realizado no contexto das ações do Observatório de Dados da graduação (ODG), órgão da UFRPE que assessora a Pró-Reitoria de Ensinos de Graduação (PREG) na gestão baseada em evidências. A plataforma *SABIA* segue as etapas do processo de *Business Intelligence (BI)*, abrangendo a ingestão, limpeza,

transformação e carregamento dos dados, os quais são estruturados em *Data Marts* ajustados às necessidades de cada painel do sistema [Marques et al. 2023], incluindo o painel que lida com a análise de sobrevivência, que foi utilizado nesse trabalho. Todos os dados utilizados nesse trabalho são então anonimizados e agregados, evitando a identificação dos estudantes.

A atualização dos dados ocorre semestralmente, em conformidade com o calendário acadêmico, garantindo que as informações processadas estejam atualizadas e reflitam as mudanças no ambiente acadêmico [Barbosa et al. 2024]. Esse ciclo de atualização favorece a integração de dados relacionados a matrícula, desempenho e outras variáveis relevantes para os cursos de graduação.

O processamento dos dados envolve etapas como a filtragem de registros inválidos, a remoção de vínculos acadêmicos inconsistentes, a aplicação de censura para discentes não concluintes e a normalização dos períodos de ingresso. Esse tratamento resulta em um conjunto de variáveis que inclui a forma de ingresso, curso, campus, semestre de ingresso, área, currículo, gênero e status de assistido, permitindo análises direcionadas a cursos ou unidades acadêmicas específicas. Essas variáveis são estruturadas para garantir a representatividade e a viabilidade estatística na análise da permanência estudantil.

3.2. Construção do Modelo

A construção do modelo empregado no Painel de Permanência, disponibilizado na plataforma System of Academic Business Intelligence and Analytics (SABIA), teve início com a definição dos *Data Marts*, organizados a partir das informações de vínculo institucional e dos períodos de matrícula dos discentes. Em seguida, foram consideradas as variáveis relacionadas ao vínculo acadêmico e às características dos estudantes, como identificador, duração do vínculo e situação acadêmica. A partir dessas informações, houve a definição dos parâmetros necessários ao cálculo das estimativas de sobrevivência, com o objetivo de estruturar uma análise temporal da ocorrência dos eventos.

O estimador de *Kaplan-Meier* é um método não paramétrico para análise de sobrevivência com dados censurados [Kaplan and Meier 1958]. Ao contrário dos modelos paramétricos, não exige suposições sobre a distribuição dos dados, o que o torna apropriado quando a distribuição é desconhecida [Bland and Altman 1998], como em estudos de evasão, onde múltiplos fatores interdependentes determinam o tempo até o evento, baseando-se na ordem dos eventos observados. No contexto do Painel de Permanência, o *Kaplan-Meier* foi utilizado para estimar a função de sobrevivência dos estudantes a partir da observação de sua situação de vínculo. As curvas resultantes permitem visualizar a evolução da permanência estudantil e realizar comparações entre diferentes grupos de discentes, sendo essa uma representação gráfica comum em estudos sociais e educacionais para sintetizar padrões entre coortes [Mills 2010].

Para este estudo, considera-se como evento de interesse a evasão por curso, definida como a perda do vínculo do discente com o curso de origem antes da conclusão da graduação, independentemente do motivo, como abandono, jubramento, falecimento ou qualquer outra forma de desligamento. A evasão é distinguida em tipos, como a evasão de curso (quando o discente se desliga do curso por diversas situações), a evasão da instituição (referente à sua saída da instituição) e a evasão do sistema (quando há um abandono definitivo do ensino superior) [sobre Evasão 1996]. A função de sobrevivência,

representada na Equação 1, expressa a probabilidade de um estudante permanecer vinculado ao curso após o tempo t , ou seja, sem ter se evadido até esse momento. É importante destacar que o valor de t é limitado pelo maior tempo de acompanhamento observado nos dados, determinado pelo tempo máximo de vínculo permitido ou pelo fim do período de observação.

$$S(t) = \Pr(T > t) \quad (1)$$

O tempo de acompanhamento corresponde ao intervalo entre a matrícula inicial do discente e a ocorrência do evento ou o encerramento do período de observação. Adotamos os períodos letivos como unidade temporal de análise, considerando censurados os casos em que o discente permanece vinculado ao curso ao final do tempo observado. A variância da função de sobrevivência é calculada pelo método de *Greenwood* [Colosimo and Giolo 2021]. Esse procedimento de cálculo da função de sobrevivência é replicável a diferentes agrupamentos de cursos.

Dessa forma, a operacionalização do modelo de análise de sobrevivência no Painel de Permanência demanda, como parâmetros de entrada, a definição do evento de interesse a ser analisado, o tempo máximo de observação considerado e a especificação do agrupamento correspondente ao grupo de estudo, bem como a indicação dos grupos de comparação, utilizados na geração das estimativas de sobrevivência [Cavalcanti et al. 2024]. Ressalta-se que os filtros disponíveis no *dashboard* contribuem para a personalização das análises, entre eles, o filtro de detalhes possibilita a inclusão dos intervalos de confiança e a identificação dos momentos em que ocorrem os casos censurados. Contudo, é importante destacar que a análise de gênero neste estudo está limitada à classificação binária, feminino e masculino do INEP, que não contempla a diversidade de identidades de gênero.

Os intervalos de confiança quantificam a incerteza das estimativas de sobrevivência, considerando a variabilidade amostral [Olive 2005]. Os limites superior e inferior incorporados ao modelo são obtidos automaticamente pelo método `confidence_interval` da biblioteca *Lifelines*, que calcula intervalos de 95% em cada ponto da curva gerada pelo estimador *Kaplan-Meier*, contemplando tanto eventos observados quanto censurados.

Complementarmente, o teste de *Log-Rank* avalia a hipótese nula de igualdade entre as distribuições do tempo até o evento nos grupos comparados [Wellek 1993, Bland and Altman 2004]. A hipótese nula considera a igualdade entre as curvas de sobrevivência dos grupos comparados, ou seja, assume que não há diferenças estatisticamente significativas nas distribuições de tempo até o evento, sendo rejeitada para valores de p inferiores a 0,05. A função `logrank_test`, da biblioteca *Lifelines*, foi usada para comparar os tempos de sobrevivência e os indicadores da ocorrência do evento extraídos para cada par de grupos. Essa métrica permite ao gestor identificar diferenças no tempo de permanência entre grupos de estudantes, fundamentando a priorização de ações e recursos institucionais.

Além da significância estatística, avaliou-se o *Effect Size (ES)* para mensurar a relevância prática das diferenças no tempo de permanência entre os grupos de estudantes. O ES quantifica a magnitude dessas variações, permitindo verificar se um resultado estatisticamente significativo também possui relevância no contexto institucional [Lakens 2013, Cumming 2013]. Para tal, utilizou-se a função `compute_essize` da

biblioteca *Pingouin*, adotando o *Hedges' g* por corrigir o viés de superestimação presente no *Cohen's d* em amostras pequenas. A incorporação dessa métrica permite priorizar intervenções com base não apenas na existência de diferenças estatísticas, mas também em sua relevância prática para a gestão acadêmica.

Para estimar o *Hazard Ratio* entre os grupos, utilizou-se o modelo de riscos proporcionais de Cox, ajustado por meio da função `CoxPHFitter` da biblioteca *Lifelines*. O modelo considera como variáveis dependentes o tempo até o evento e sua ocorrência, enquanto o grupo é a variável explicativa. A Equação 2 representa o modelo em que $h(t|X)$ denota o risco instantâneo condicional ao grupo X , $h_0(t)$ é a função de risco basal, e β é o coeficiente estimado. O (HR), calculado por $\exp(\beta)$, expressa a razão entre os riscos instantâneos de dois grupos ao longo do tempo, mesmo em presença de censura. Essa medida permite avaliar se há maior ou menor risco de ocorrência do evento entre os grupos, fornecendo subsídios para que gestores priorizem ações voltadas à permanência discente, com base no perfil de risco estimado.

$$h(t|X) = h_0(t) \cdot \exp(\beta X) \quad (2)$$

Apesar das vantagens, essas abordagens apresentam limitações. Em contextos com alta censura ou amostras pequenas, os intervalos de confiança perdem precisão. O teste de *Log-Rank* e o *Hazard Ratio* assumem riscos proporcionais, o que restringe a detecção de diferenças cujo efeito varia ao longo do tempo. O tamanho do efeito, estimado por *Hedges' g*, não considera a censura e pode não capturar toda a complexidade dos dados de sobrevivência. A definição dos grupos e o tamanho amostral também afetam diretamente os resultados. Adicionalmente, fatores não controlados, como qualidade do ensino ou motivação pessoal, podem influenciar os resultados, e a amostra pode apresentar viés de seleção, pois não representa completamente a diversidade dos estudantes de STEM. Apesar dessas restrições, a aplicação conjunta dessas técnicas amplia a análise comparativa entre grupos, contribuindo para interpretações mais informadas no contexto da permanência estudantil.

4. Resultados e Discussões

Para investigar diferenças na permanência estudantil em cursos STEM, foram analisados dados dos cursos presenciais e regulares de Licenciatura em Matemática (MAT), Física, Licenciatura em Computação (LC), Bacharelado em Sistemas de Informação (BSI) e Bacharelado em Ciência da Computação (BCC). A seleção considerou a disponibilidade dos dados e a existência de estudos prévios, favorecendo a continuidade da análise. Os dados contemplam discentes com ingresso entre 2014 e 2024 no campus sede da UFRPE. No mesmo período, a instituição registrou um total de 36.895 discentes matriculados em cursos de graduação. Considerou-se um recorte correspondente a dez períodos letivos, abrangendo o tempo previsto para integralização curricular. Os cursos analisados são ofertados em turnos distintos, incluindo matutino, vespertino e noturno.

A Tabela 1 sintetiza a situação acadêmica de 7.432 discentes nos cinco cursos analisados, além do panorama institucional abrangendo 36.895 matrículas. Observa-se maior incidência de desligamentos por evasão nos cursos de MAT e FÍSICA, enquanto os cursos MAT e BCC apresentam os maiores contingentes de matrículas censuradas. No âmbito institucional, o desligamento por evasão atinge 43,4% dos casos, com uma duração média de permanência próxima a 5,5 períodos letivos.

Tabela 1: Indicadores de Vínculo por Curso.

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
FÍSICA	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

Quanto à composição por gênero e assistência estudantil, institucionalmente, 44,5% dos discentes são do sexo *Feminino* (16.464) e 55,5% do sexo *Masculino* (20.431), com 2,3% (858) beneficiários de assistência estudantil. Na amostra dos cursos STEM, predomina o gênero masculino, com 5.893 discentes (79,7%) contra 1.539 do gênero feminino (20,3%), sendo 158 (2,1%) assistidos e 7.274 (97,9%) não assistidos.

As Figuras 1a e 1b apresentam o estimador de *Kaplan-Meier* em função da duração do vínculo (em períodos), segmentado por gênero para o conjunto geral de cursos e o subconjunto STEM, respectivamente. A área hachurada representa o intervalo de confiança, cuja amplitude é maior para o grupo *Feminino*, reflexo do menor tamanho amostral. Neste contexto, a notação $[l_i; l_s]$ representa os limites inferior (l_i) e superior (l_s) do intervalo de confiança, sendo a estimativa média obtida a partir da média aritmética simples de l_i e l_s .

No grupo Geral, as estimativas de sobrevivência para o grupo *Feminino* situam-se próximas de 50% entre o 8º e o 9º período, com valores de 54,46% e intervalos de confiança de [53,60%; 55,30%] e 52,04% [51,16%; 52,92%], indicando que atinge o *plateau* a partir desse intervalo. Para o grupo *Masculino*, o patamar de 50% é alcançado no intervalo entre 6º e o 7º períodos, cujos intervalos de confiança apresentados na Tabela 2 são [51,15%; 52,59%] e [46,68%; 48,13%]. Nos cursos STEM, observa-se padrão semelhante, com o grupo *Feminino* apresentando estimativa de 50,05% [46,67%; 53,34%] no 8º período, enquanto o grupo *Masculino* atinge esse nível entre o 7º e o 8º períodos, com 48,30% [46,65%; 49,93%] e 44,57% [42,90%; 46,22%], respectivamente. Os dados dos intervalos de confiança por período e gênero são apresentados na Tabela 2.

A análise se estende ao agrupamento por assistência estudantil, conforme expresso na Figura 2 e na Tabela 3. No conjunto geral de cursos, observa-se que as estimativas de sobrevivência do grupo *Não Assistido* reduzem-se de forma mais precoce ao longo dos períodos, atingindo aproximadamente 50% já no 7º período (49,81% [49,28%; 50,34%]), enquanto o grupo *Assistido* mantém-se acima desse patamar até o final da duração observada, com valor mínimo de 64,35% [60,45%; 67,96%] no 10º período. Padrão semelhante é verificado nos cursos da área STEM, em que o grupo *Não Assistido* atinge valores próximos de 50% entre o 6º e o 7º períodos (53,83% [52,37%; 55,26%] e 48,97% [47,48%; 50,44%], respectivamente), enquanto o grupo *Assistido* alcança esse nível apenas entre o 9º e o 10º períodos (56,91% [39,48%; 71,05%]), com maior amplitude nos intervalos de confiança. Esses resultados possivelmente refletem que o suporte financeiro e acadêmico favorece a conclusão do curso.

Os testes *Log-rank* da Tabela 4 correspondem à avaliação da hipótese nula de

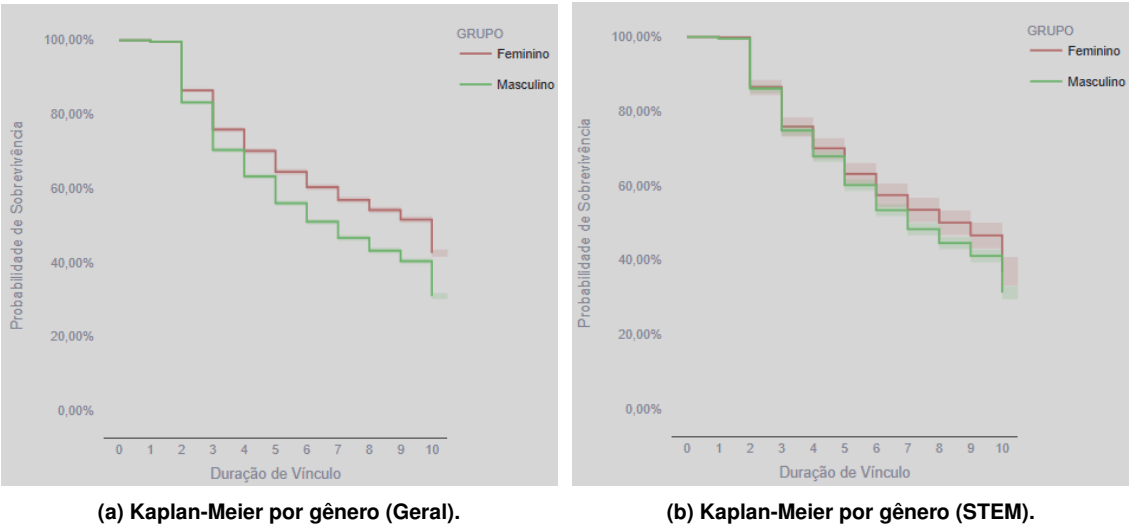


Figura 1: Análise de Sobrevivência por Gênero.

Tabela 2: Intervalo de Confiança (IC 95%) da probabilidade de sobrevivência por gênero para todos os cursos (Geral) e para cursos de STEM.

Prd.	Geral		STEM	
	Feminino	Masculino	Feminino	Masculino
0	[100,00;100,00]	[100,00;100]	[100,00;100,00]	[100,00;100,00]
1	[99,27;99,51]	[99,27;99,48]	[99,36;99,99]	[99,32;99,73]
2	[84,81;85,91]	[82,38;83,40]	[84,38;88,44]	[85,01;87,11]
3	[75,06;76,46]	[70,17;71,42]	[73,20;78,38]	[73,48;76,14]
4	[69,70;71,15]	[63,50;64,83]	[67,14;72,75]	[66,41;69,31]
5	[63,98;65,53]	[56,28;57,65]	[60,05;66,09]	[58,58;61,68]
6	[60,09;61,70]	[51,15;52,59]	[54,25;60,57]	[51,79;55,00]
7	[56,42;58,08]	[46,67;48,13]	[50,19;56,71]	[46,65;49,93]
8	[53,60;55,30]	[43,07;44,54]	[46,67;53,34]	[42,90;46,22]
9	[51,16;52,92]	[40,14;41,64]	[43,11;50,04]	[39,42;42,79]
10	[43,07;45,04]	[31,74;32,31]	[33,10;40,79]	[29,49;32,95]

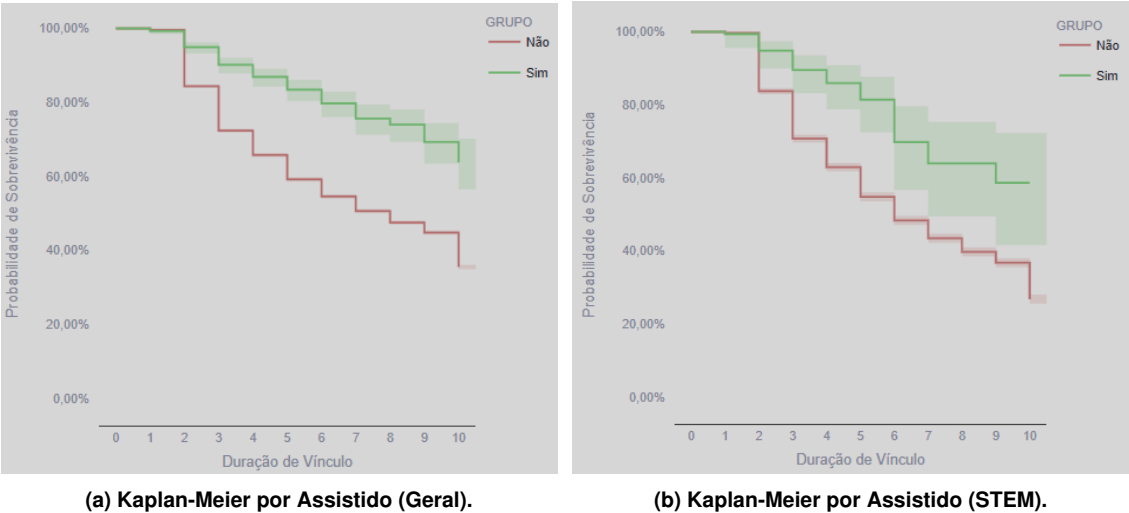


Figura 2: Análise de Sobrevivência por Assistido.

Tabela 3: Intervalo de Confiança (IC 95%) da probabilidade de sobrevivência por grupo Assistido para todos os cursos (Geral) e para cursos de STEM.

Prd.	Geral		STEM	
	Assistido	Não Assistido	Assistido	Não Assistido
0	[100,00;100,00]	[100,00;100,00]	[100,00;100,00]	[100,00;100,00]
1	[99,28;99,44]	[98,80;99,56]	[99,45;99,78]	[95,48;99,91]
2	[92,95;94,99]	[82,68;83,42]	[89,88;97,37]	[84,95;86,85]
3	[87,60;90,35]	[71,30;72,20]	[83,07;93,52]	[73,42;75,83]
4	[83,78;86,97]	[64,50;65,47]	[77,50;90,07]	[66,53;69,15]
5	[79,97;83,61]	[58,06;59,07]	[71,28;86,86]	[58,86;61,65]
6	[76,64;80,72]	[53,40;54,45]	[55,38;78,58]	[52,37;55,26]
7	[72,88;77,52]	[49,28;50,34]	[47,95;74,20]	[47,48;50,44]
8	[69,30;74,52]	[46,05;47,12]	[47,95;74,20]	[43,78;46,77]
9	[65,92;71,86]	[43,32;44,42]	[39,48;71,05]	[40,32;43,37]
10	[60,45;67,96]	[34,71;35,89]	[39,48;71,05]	[30,37;33,54]

igualdade entre as funções de sobrevivência no subconjunto STEM. A notação *G.Ref* indica o grupo de referência, enquanto o *G.Alt* corresponde ao grupo alternativo utilizado na comparação. A hipótese nula foi rejeitada para os grupos de gênero (Feminino e Masculino, $p = 1,0037 \times 10^{-9}$) e assistência estudantil (Assistido e Não Assistido, $p = 4,4667 \times 10^{-56}$), indicando diferenças nas distribuições de tempo de permanência. O agrupamento gênero-assistência mostrou igualdade para o grupo Feminino ($p = 1,0030$) e rejeição para o grupo Masculino ($p = 5,6640 \times 10^{-7}$). Entre estudantes não assistidos, os turnos Matutino versus Vespertino, Matutino versus Noturno e Noturno versus Vespertino apresentaram diferenças significativas ($p < 0,05$), enquanto para assistidos não houve diferenças significativas. Comparações por gênero nos turnos indicaram igualdade no grupo Matutino ($p = 0,107$) e diferenças nos turnos Vespertino e Noturno ($p = 3,1 \times 10^{-2}$ e $p < 10^{-4}$).

Tabela 4: Teste de Hipótese Log-Rank para cursos de STEM.

Grupo de Ref.	Grupo de Alt.	H_0	p-valor
Feminino, STEM, Noturno	Masculino, STEM, Noturno	Rejeitado	0,0000e ⁰
Assistido, STEM	Não Assistidos STEM	Rejeitado	4,4667e ⁻⁵⁶
Feminino, STEM	Masculino, STEM	Rejeitado	1,0037e ⁻⁹
Feminino, STEM, Não Assistido	Masculino, STEM, Não Assistido	Rejeitado	5,6640e ⁻⁷
Matutino, STEM, Não Assistido	Vespertino, STEM, Não Assistido	Rejeitado	4,6168e ⁻⁶
Noturno, STEM, Não Assistido	Vespertino, STEM, Não Assistido	Rejeitado	2,8300e ⁻²
Feminino, STEM, Vespertino	Masculino, STEM, Vespertino	Rejeitado	3,1000e ⁻²
Matutino, STEM, Não Assistido	Noturno, STEM, Não Assistido	Rejeitado	1,2700e ⁻²
Matutino, STEM, Assistido	Noturno, STEM, Assistido	Não Rejeitado	7,0570e ⁻²
Feminino, STEM, Matutino	Masculino, 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
Feminino, STEM, Assistido	Masculino, STEM, Assistido	Não Rejeitado	1,0030e ⁰

Uma vez estabelecida a significância entre os grupos, o tamanho do efeito indica a magnitude das diferenças, corroborando os resultados do teste *Log-rank*, conforme a Tabela 5. Valores de *Hedges' g* próximos a 0,2 indicam efeito pequeno, cerca de 0,5 efeito médio e acima de 0,8 efeito grande, permitindo quantificar a relevância prática das

diferenças observadas. No subconjunto STEM, a diferença entre gêneros foi pequena ($g = 0,0379$), enquanto a assistência estudantil apresentou efeito médio ($g = 0,4442$). Nas análises combinadas, as maiores magnitudes ocorreram nas comparações entre turnos, especialmente *Matutino* versus *Noturno* e *Noturno* versus *Vespertino*, com efeitos médios entre *Assistidos* e *Não assistidos*. O efeito permaneceu pequeno nas comparações de gênero por turno; no entanto, as diferenças entre turnos podem refletir a influência do regime de aulas (matutino, vespertino, noturno) sobre a permanência, possivelmente associada a perfis socioeconômicos distintos ou demandas externas.

Tabela 5: Hedges'g para comparações entre grupos para cursos de STEM.

Grupo de Ref.	Grupo de Alt.	Hedges'g	Tam. do efeito
Assistido, STEM	Não Assistido, STEM	0.4442	Médio
Matutino, STEM, Não Assistido	Noturno, STEM, Não Assistido	0.3486	Médio
Noturno, STEM, Não Assistido	Vespertino, STEM, Não Assistido	0.3319	Médio
Noturno, STEM, Assistido	Vespertino, STEM, Assistido	0.2260	Médio
Matutino, STEM, Não Assistido	Vespertino, STEM, Não Assistido	0.1181	Pequeno
Feminino, STEM, Não Assistido	Masculino, STEM, Não Assistido	0.1191	Pequeno
Matutino, STEM, Assistido	Vespertino, STEM, Assistido	0.0586	Pequeno
Feminino, STEM, Vespertino	Masculino, STEM, Vespertino	0.0529	Pequeno
Feminino, STEM, Matutino	Masculino, STEM, Matutino	0.1045	Pequeno
Feminino, STEM	Masculino, STEM	0.0379	Pequeno
Feminino, STEM, Noturno	Masculino, STEM, Noturno	0.0071	Pequeno
Feminino, STEM, Assistido	Masculino, STEM, Assistido	0.0035	Pequeno

O modelo de riscos proporcionais de Cox, apresentado na Tabela 6 para a área de STEM, revela um coeficiente positivo para o grupo *Masculino* ($\beta = 0,1311$), refletido no Hazard Ratio ($HR = 1,139$, $p = 0,0070$), indicando que o grupo *Masculino* apresenta uma razão de risco 13,9% superior à do grupo *Feminino*. De forma semelhante, estudantes *Não Assistidos* matriculados no turno *Noturno* apresentam coeficiente positivo ($\beta = 0,4680$) e HR igual a 1,597 ($p < 0,001$), o que indica risco instantâneo 59,7% maior em relação ao grupo *Assistido* no turno *Matutino*.

Tabela 6: Hazard Ratio para cursos de STEM.

Grupo Ref.	Grupo Alt.	(β)	HR	p-valor
Masculino, STEM, Assistido	Feminino, STEM, Assistido	0.7312	2.0780	0.1390
Matutino, STEM, Assistido	Noturno, STEM, Assistido	0.6935	2.0070	0.1340
Matutino, STEM, Não Assistido	Noturno, STEM, Não Assistido	0.4680	1.5970	0.0010
Masculino, STEM, Noturno	Feminino, STEM, Noturno	0.1991	1.2200	0.0040
Masculino, STEM, Não Assistido	Feminino, STEM, Não Assistido	0.1484	1.1600	0.0035
Masculino, STEM	Feminino, STEM	0.1311	1.1390	0.0070
Masculino, STEM, Vespertino	Feminino, STEM, Vespertino	0.1584	1.1720	0.1030
Masculino, STEM, Matutino	Feminino, STEM, Matutino	0.0682	1.0710	0.8070
Matutino, STEM	Vespertino, STEM	0.0466	1.0480	0.8850
Matutino, STEM, Não Assistido	Vespertino, STEM, Não Assistido	0.0315	1.0320	0.5560
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

4.1. Discussões

Os cursos da área de STEM são tradicionalmente associados a elevados índices de evasão estudantil, especialmente em contextos que combinam alta exigência acadêmica e limitações estruturais no apoio institucional [Ribeiro et al. 2021]. O maior engajamento feminino nesses cursos pode estar relacionado não apenas à motivação profissional das estudantes, mas também a variáveis socioeconômicas como a renda que afetam de modo distinto as populações em análise. Essa dimensão social juntamente com o perfil específico dos estudantes em cada curso, especialmente nos de caráter mais elitizado, pode influenciar os índices de evasão e demanda investigações que abordem essas questões [Silva et al. 2016].

PP 01: Há diferenças nos padrões temporais de evasão entre estudantes dos cursos STEM e a permanência geral dos estudantes na instituição?

Esta pergunta de pesquisa foi respondida por meio da análise das curvas de sobrevivência, com ênfase na observação da dinâmica da evasão ao longo dos períodos letivos. Os resultados obtidos corroboram a tendência identificada nos estudos de referência, evidenciando que a distribuição entre homens e mulheres ocorre de forma desigual nos cursos analisados, com predomínio masculino nas áreas STEM.

Observa-se que, a partir do 6º período, a diferença entre os intervalos de confiança para o gênero feminino entre STEM e todos os cursos passa a aumentar, deixando de haver interseção a partir do 8º período, o que indica diferença estatisticamente relevante entre esses grupos. Essa conclusão é embasada pelo teste *Log-Rank*, que rejeitou a hipótese de igualdade entre as curvas de sobrevivência a partir do 8º período, com $p < 0,05$. Já para os discentes do sexo masculino, essa diferença existe entre o 2º e o 5º períodos, mas deixa de ocorrer a partir do 6º período.

PP 02: Como gênero, assistência estudantil e turno de matrícula influenciam a probabilidade de permanência ao longo do tempo?

A comparação por grupos permitiu responder à pergunta de pesquisa ao evidenciar diferenças nos padrões de permanência ao longo do tempo, conforme as variáveis de gênero, assistência estudantil e turno de matrícula. Ao analisar o impacto desses grupos, verifica-se comportamento distinto entre homens e mulheres nos cursos da área STEM, com maior desligamento entre os homens nos períodos iniciais e entre as mulheres a partir da metade do curso. Os dados sugerem que programas de bolsas ou auxílios contribuem para reduzir a evasão, com efeito mais pronunciado entre estudantes do sexo feminino, enquanto padrões de desligamento variam entre gêneros ao longo do tempo. Essa configuração indica que os fatores associados ao abandono podem variar entre os grupos, o que justifica a realização de estudos qualitativos. Pesquisas como [Almeida and Schimiguel 2011] e [Garcia and Gomes 2022] vinculam o desligamento masculino a desempenho acadêmico, carga de trabalho, condições econômicas e elementos institucionais.

No que se refere às modalidades de curso, observa-se que os cursos de licenciatura tendem a apresentar padrões de evasão distintos em relação aos bacharelados, especialmente no turno noturno. A permanência nessa oferta pode estar associada à dificuldade de conciliar trabalho e estudo, mais recorrente entre estudantes que optam por esse turno [de Lima and Zago 2018]. Esse fator, quando combinado ao perfil socioeconômico po-

tencialmente mais vulnerável dos discentes de licenciatura, pode estar entre os motivos que influenciam a maior incidência de evasão.

PP 03: De que forma as métricas estatísticas, como o teste *Log-Rank*, o *Effect Size* e o *Hazard Ratio*, permitem identificar e quantificar diferenças nos padrões de permanência estudantil na análise de sobrevivência?

A análise das métricas estatísticas, como o teste *Log-Rank*, *Hedges' g* e *Hazard Ratio* permitiu identificar e quantificar diferenças significativas nos padrões de permanência estudantil, considerando gênero, assistência estudantil e turno de matrícula. O *Log-Rank* rejeitou a hipótese de igualdade nas comparações entre *Feminino* versus *Masculino*, *Assistidos* versus *Não assistidos*, e nos turnos *Matutino* versus *Vespertino* no grupo *Não assistido*, além de destacar particularidades no turno *Noturno* em algumas análises. Além disso, o tamanho do efeito indicou magnitude média nessas comparações, enquanto o *Hazard Ratio* apontou maior risco de desligamento para estudantes do gênero masculino, não assistidos e matriculados no turno noturno, em relação aos seus pares.

Ainda assim, é importante ressaltar que a adoção deste conjunto metodológico não deve ser interpretada como evidência exclusiva das causas da evasão estudantil, pois esta é um fenômeno complexo e multifatorial. A permanência acadêmica está condicionada a uma interação dinâmica entre aspectos acadêmicos, socioeconômicos e institucionais, os quais demandam abordagem integrada para melhor compreensão e enfrentamento.

Em decorrência dos achados, apesar do enfoque nas variáveis de gênero, assistência e turno, faz-se imprescindível o desenvolvimento de investigações futuras que ampliem a análise para outras dimensões contextuais, visando à elaboração de políticas institucionais direcionadas, os resultados reforçam a relevância de ações integradas. Tais políticas podem incluir programas de tutoria, concessão de bolsas, revisão dos projetos pedagógicos, flexibilização da oferta curricular, fortalecimento de redes de apoio e melhorias infraestruturais, visando mitigar a evasão nos grupos mais vulneráveis e promover equidade na permanência estudantil. A eficácia dessas ações dependerá da consideração integrada do contexto e dos diversos fatores que permeiam a permanência estudantil.

5. Considerações Finais

Este estudo, ao aplicar análise de sobrevivência em cursos STEM considerando gênero, turno e assistência estudantil, evidenciou maior permanência entre mulheres, beneficiários de assistência estudantil e alunos do turno diurno, enquanto homens e estudantes não assistidos apresentam risco significativamente maior de evasão, especialmente a partir do 6º período. Os achados, respaldados pelo *Log-Rank*, *Hazard Ratio* e *Effect Size*, indicam que a evasão reflete fatores estruturais e demográficos, reforçando a necessidade de intervenções institucionais fundamentadas em evidências, como acompanhamento acadêmico e políticas de apoio estudantil, apoiadas por ferramentas de gestão como o sistema SABIA da UFRPE [Marques et al. 2023].

Como trabalhos futuros, propõe-se a adoção de censura à esquerda, truncamento e inclusão do filtro de múltiplas covariáveis, permitindo maior precisão na análise da permanência e identificação do efeito simultâneo de fatores acadêmicos, socioeconômicos e institucionais. Comparações entre áreas do conhecimento, formas de ingresso, semestres e campi podem revelar padrões específicos de evasão, oferecendo subsídios mais robustos para políticas institucionais voltadas à promoção da permanência estudantil.

6. Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001. Os autores agradecem ao Observatório de Dados da Graduação (ODG) e à Gestão Institucional da UFRPE pelo essencial suporte e pelos recursos disponibilizados para a realização deste trabalho, desenvolvido no âmbito das iniciativas do ODG para fomentar a gestão universitária baseada em evidências. Ferramentas de Inteligência Artificial Generativa foram utilizadas como apoio na elaboração deste documento, incluindo o auxílio na redação, revisão textual e organização de conteúdo. Os autores revisaram e editaram criticamente todo o conteúdo gerado e assumem total responsabilidade pela obra final.

Referências

- Almeida, J. B. and Schimiguel, J. (2011). 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*, 2(2):167–178.
- Alvim, Í. V., Bittencourt, R. A., and Duran, R. S. (2024). Evasão nos cursos de graduação em computação no brasil. In *Anais do IV Simpósio Brasileiro de Educação em Computação*, pages 1–11. SBC.
- Barbosa, T., Freitas, N., Cavalcanti, L., da Conceicao Batista, M., Gouveia, R., and Alves, G. (2024). 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*, pages 1918–1930, Porto Alegre, RS, Brasil. SBC.
- Bland, J. M. and Altman, D. G. (1998). Survival probabilities (the kaplan-meier method). *Bmj*, 317(7172):1572–1580.
- Bland, J. M. and Altman, D. G. (2004). The logrank test. *Bmj*, 328(7447):1073.
- Cavalcanti, L., Barros, A., Falcão, T., da Conceicao Batista, M., Cristino, C., and Alves, G. (2024). 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*, pages 2617–2626, Porto Alegre, RS, Brasil. SBC.
- Chen, Y., Johri, A., and Rangwala, H. (2018). 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*, pages 270–279.
- Colosimo, E. A. and Giolo, S. R. (2021). *Análise de sobrevivência aplicada*. Editora Blucher.
- Cumming, G. (2013). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.
- de Lima, F. S. and Zago, N. (2018). Evasão no ensino superior: Tendências e resultados de pesquisa. *Reunião Científica Regional da ANPED*.
- de Moraes Zamith, M. P., Nascente, J. M., Zamith, S., and Lessa, N. C. (2025). Análise preliminar da influência da mudança curricular nos índices de evasão discente no curso de ciência da computação da ufrj. In *Simpósio Brasileiro de Educação em Computação (EDUCOMP)*, pages 115–126. SBC.

- Duran, R., Bim, S. A., Gimenes, I., Ribeiro, L., and Correia, R. C. M. (2023). Potential factors for retention and intent to drop-out in brazilian computing programs. *ACM Transactions on Computing Education*, 23(3):1–33.
- Garcia, L. M. L. d. S. and Gomes, R. S. (2022). Causas da evasão em cursos de ciências exatas: uma revisão da produção acadêmica. *Revista Educar Mais. Pelotas, RS. Vol. 6* (2022), p. 937-957.
- Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481.
- Klitzke, M. and Carvalhaes, F. (2023). Fatores associados à evasão de curso na ufrj: Uma análise de sobrevivência. *Educação em Revista*, 39:e37576.
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and anovas. *Frontiers in psychology*, 4:863.
- Lima Junior, P., Silveira, F. L. d., and Ostermann, F. (2012). 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*, 34:1403.
- Marques, E., Cavalcanti, L., Barbosa, T., and Alves, G. (2023). Sabia: Uma plataforma para auxiliar a gestão baseada em evidências nas instituições de ensino superior. In *Anais do II Workshop de Aplicações Práticas de Learning Analytics em Instituições de Ensino no Brasil*, pages 71–80. SBC.
- Mills, M. (2010). *Introducing Survival and Event History Analysis*. SAGE Publications, London.
- Nierotka, R. L., Salata, A., and Klitzke Martins, M. (2023). Fatores associados à evasão no ensino superior: Um estudo longitudinal. *Cadernos de Pesquisa*, 53:e09961.
- Olive, D. J. (2005). A simple confidence interval for the median. *Manuscript*.
- Ortiz-Martínez, G., Vázquez-Villegas, P., Ruiz-Cantisani, M. I., Delgado-Fabián, M., Conejo-Márquez, D. A., and Membrillo-Hernández, J. (2023). Analysis of the retention of women in higher education stem programs. *Humanities and Social Sciences Communications*, 10(1):1–14.
- Pachas, D. A. G., Garcia-Zanabria, G., Cuadros-Vargas, A. J., Camara-Chavez, G., Poco, J., and Gomez-Nieto, E. (2021). A comparative study of who and when prediction approaches for early identification of university students at dropout risk. In *2021 XLVII Latin American Computing Conference (CLEI)*, pages 1–10. IEEE.
- Ribeiro, I. M., Correia, W. F. M., and Campos, F. (2021). Setores acadêmicos que interferem na satisfação do aluno no ensino superior. *Acta Scientiarum. Education*, 43.
- Saccaro, A., França, M. T. A., and Jacinto, P. d. A. (2019). Fatores associados à evasão no ensino superior brasileiro: um estudo de análise de sobrevivência para os cursos das áreas de ciência, matemática e computação e de engenharia, produção e construção em instituições públicas e privadas. *Estudos Econômicos (São Paulo)*, 49:337–373.
- Santos, J., Sousa, J. D., Mello, R., Cristino, C., and Alves, G. (2021). Um modelo para análise do impacto da retenção e evasão no ensino superior utilizando cadeias de markov absorventes. In *Anais do XXXII Simpósio Brasileiro de Informática na Educação*, pages 813–823, Porto Alegre, RS, Brasil. SBC.

- Santos, M. T. S., Gasparini, I., and Frigo, L. B. (2024). Análise da participação feminina nos cursos da área de computação: Um olhar sobre os projetos do programa meninas digitais. In *Simpósio Brasileiro de Informática na Educação (SBIE)*, pages 2259–2272. SBC.
- Silva, A. C., Correa, C. S., de Assis Coelho, D., da Silva, D. T., Neto, L. L. F., Xavier, M. M., da Silva Reis, R., Adenilson, F., and Rocha, P. A. S. (2016). 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*.
- Silva, L. G. d. L. (2024). O impacto da nova matriz curricular da licenciatura em computação no desempenho dos discentes. B.S. thesis, Brasil.
- sobre Evasão, C. E. d. E. (1996). Diplomação, retenção e evasão nos cursos de graduação em ies públicas. *Avaliação, I* (2), pages 55–65.
- Sousa, R., Fachini-Gomes, J., Holanda, M., and Leão, M. (2024). Um estudo da evasão no curso de licenciatura em computação da universidade de Brasília. In *Anais do XXXII Workshop sobre Educação em Computação*, pages 715–725, Porto Alegre, RS, Brasil. SBC.
- Souza, J., Komati, K., and Andrade, J. (2022). Análise de sobrevivência: um estudo de caso em um curso de sistemas de informação. In *Anais do XXX Workshop sobre Educação em Computação*, pages 392–403, Porto Alegre, RS, Brasil. SBC.
- Vitelli, R. F., Rocha, C. S., and Fritsch, R. (2010). 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*, 44.
- Wellek, S. (1993). A log-rank test for equivalence of two survivor functions. *Biometrics*, pages 877–881.