

Análise Preliminar da Influência da Mudança Curricular nos Índices de Evasão Discente no Curso de Ciência da Computação da UFRRJ

Marcelo Panaro de Moraes Zamith¹, Juliana Mendes Nascente e Silva Zamith¹,
Natália Chaves Lessa¹

¹Departamento de Ciência da Computação (DCC)
Universidade Federal Rural do Rio de Janeiro (UFRRJ)
Nova Iguaçu/RJ – Brasil

{mzamith, julianascente, nataliach.lessa}@ufrrj.br

Abstract. *The Computing field faces a shortage of professionals in the job market due to various factors, including high dropout and retention rates in academic programs. This scenario is also observed in the Computer Science program at UFRRJ, where 4 out of 10 students drop out. In 2020, a new curriculum was implemented, allowing students to select specific areas of study aligned with their skills, aiming to enhance retention. This study aims to evaluate the effects of this change on student distribution using Survival Analysis. Initial results suggest a reduction in dropout rates during the early semesters.*

Resumo. *A área de Computação enfrenta falta de profissionais no mercado devido a diversos fatores, dentre eles a alta evasão e retenção nos cursos, cenário também observado no curso de Ciência da Computação da UFRRJ, onde 4 em cada 10 discentes abandonam o curso. Em 2020, foi implementada uma nova matriz curricular no curso, permitindo aos discentes escolher áreas específicas de estudo conforme suas aptidões, com o objetivo de aumentar sua permanência. Este trabalho tem como objetivo avaliar, através do uso da Análise de Sobrevivência, os efeitos dessa mudança na distribuição dos discentes. A análise inicial sugere uma redução do percentual de evasão nos primeiros períodos.*

1. Introdução

É notório que a Computação tem se tornado cada vez mais importante nos últimos anos em vários aspectos da sociedade moderna. Neste sentido, é imprescindível a formação de profissionais qualificados nesta área, aptos a atender às demandas tecnológicas e inovadoras da sociedade atual [BRASSCOM 2024].

Mesmo diante do cenário de alta empregabilidade na área de Tecnologia da Informação e Comunicação (TIC), os cursos na área de Computação apresentam uma alta taxa de evasão. Segundo [Duran et al. 2023], apenas 25,2% dos matriculados se formam em Computação no Brasil, considerando todos os cursos nessa área. O alto índice de evasão nos cursos de Computação implica na oferta de apenas 33% de profissionais para cobrir a demanda de 159.000 profissionais por ano [Alvim et al. 2024].

Segundo [Saccaro et al. 2019], a baixa quantidade de profissionais qualificados pode gerar um impacto negativo na economia, uma vez que a sociedade não poderá contar com esses profissionais para o crescimento da área e inovação tecnológica.

As altas taxas de evasão obtidas nos últimos anos têm preocupado as Instituições de Ensino Superior (IES) e sido tema de discussões que tentam entender e pontuar os motivos pelos quais os discentes abandonam seus cursos, além de buscar definir ações de permanência de modo a diminuir a taxa de evasão [Diogo et al. 2016]. O problema da evasão tem como consequência a perda de recursos, uma vez que o número de discentes matriculados contribui significativamente para compor a matriz orçamentária da instituição [Silva 2013].

Na Universidade Federal Rural do Rio de Janeiro (UFRRJ), os dados não são diferentes. O índice de evasão em 2018, considerando todos os cursos, alcançou mais de 50%, sendo 15% a mais comparado com o ano de 2008 [Carvalho 2018]. Esses dados são alarmantes, visto que a quantidade de concluintes é menor do que a quantidade de discentes que abandonam o curso. No curso de Ciência da Computação (CC) da UFRRJ, quatro de cada dez discentes que se matriculam não concluem o curso, considerando apenas os discentes que desistem. Esse número pode ser maior se forem incluídos os discentes que são cancelados por algum motivo relacionado ao regimento da instituição.

Desde sua criação em 2010, o curso de CC da UFRRJ teve duas atualizações consideráveis na matriz curricular. No processo de proposta dessas matrizes, foram revisados, inseridos e retirados conteúdos para que a mesma pudesse abranger os conteúdos específicos descritos na matriz de referência disponível pela Sociedade Brasileira de Computação (SBC) [SBC 2005], além de visar diminuir as taxas de retenção e de evasão observadas. Entretanto, nenhum estudo mais detalhado sobre evasão foi realizado desde então. De modo empírico, a evasão é conhecida, mas algumas questões – tais como: qual a probabilidade do discente evadir após o primeiro período? Os discentes evadem ao longo do curso? – poderiam auxiliar na proposta de ações concretas para evitar ou, pelo menos, diminuir a evasão.

Desta forma, este artigo apresenta uma análise preliminar que visa caracterizar o índice da evasão discente do curso de CC da UFRRJ ao longo do tempo, levando em consideração a mudança de matriz curricular. Esta análise foi realizada a partir da aplicação da técnica de Análise de Sobrevivência.

O restante do trabalho está dividido como segue. A Seção 2 apresenta estudos sobre a evasão em cursos de CC. A Seção 3 aborda a estrutura do curso de CC na UFRRJ, descrevendo a evolução das suas matrizes curriculares. Já a Seção 4 aborda alguns conceitos da Análise de Sobrevivência, método adotado no estudo. A Seção 5 discute os resultados das análises sobre a distribuição da evasão e os impactos da mudança curricular, além de outras variáveis influentes. Por fim, a Seção 6 traz as considerações finais, as limitações do estudo e sugestões para pesquisas futuras.

2. Evasão em Cursos de Ciência da Computação

Entre os trabalhos encontrados na literatura, muitos autores discutem a evasão no curso de CC no âmbito nacional e internacional. A principal motivação desses trabalhos é identificar e analisar variáveis que influenciam a evasão. Entre os trabalhos aqui

citados, algumas das causas da desistência do curso são: a falta de perspectiva na carreira, o baixo nível de comprometimento com o curso, a baixa participação em atividades acadêmicas, a falta de apoio familiar, as instalações precárias e o baixo desempenho escolar [Silva 2013, Saccaro et al. 2019, Hoed 2016, Schoeffel et al. 2018, Palmeira and Santos 2014, Santos 2016, Giannakos et al. 2016].

[Duran et al. 2023] apresentam uma pesquisa que visa identificar a intenção de abandono e quais fatores podem levar os discentes a permanecerem em um programa de Computação. A pesquisa envolveu o envio de um questionário que foi respondido por 3193 discentes de várias IES brasileiras. Os autores descrevem que, diante dos itens respondidos, os discentes consideram que a permanência na instituição ocorre, principalmente, devido a uma boa projeção da carreira no futuro. Eles consideram que bons salários os motivam a não evadir, enquanto a grande carga de conteúdo teórico, principalmente relacionada à matemática e à programação, é um grande desafio e motivador para deixarem o curso.

Alguns trabalhos consideram que a evasão ocorre principalmente nos primeiros períodos, devido à dificuldade do discente em algumas disciplinas da área, o que o leva a abandonar o curso, seja por não se identificar com a área, ou por ter muitas reprovações nas disciplinas iniciais, que se reflete no aumento do tempo do discente na universidade [Andriola et al. 2006, Chen et al. 2018, Hoed 2016].

O uso de técnicas de mineração e análise de dados para identificar as causas da evasão e retenção dos discentes tem sido adotado. Há vários trabalhos que analisam os dados dos cursos de CC em algumas universidades, tais como: na Universidade de Brasília (UnB) [Palmeira and Santos 2014], na Universidade Federal de Santa Catarina (UFSC) [Santos 2016] e na Universidade Federal da Bahia (UFBA) [Schoeffel et al. 2018]. Nesses trabalhos, fica evidente que a reprovação em disciplinas iniciais (dentre elas: Lógica, Matemática Discreta e Cálculo) é uma das principais causas da retenção e desistência dos discentes.

3. Caracterização do Curso de Ciência da Computação da UFRRJ

O curso de CC da UFRRJ teve sua primeira turma em 2010, com uma entrada anual de 60 discentes. Até o ano de 2013, esta entrada era dividida em 30 discentes no primeiro semestre e 30 no segundo semestre. Devido à adaptação do curso aos recursos da universidade, a entrada passou a ser anual no primeiro semestre. Ao mesmo tempo, foi implantada uma nova matriz curricular (*M_2014*) que teve como motivação a adequação àqueles conteúdos básicos descritos na matriz de referência da SBC.

Na matriz *M_2014*, as disciplinas da área de Matemática eram introduzidas no primeiro período e disciplinas mais específicas da área da Computação estavam dispostas apenas em períodos finais. Existiam apenas cinco disciplinas que abordavam tópicos especiais em áreas da Computação que poderiam ser utilizados para aprofundar o conhecimento em áreas específicas de forma mais dinâmica, sendo elas: Tópicos Especiais em Otimização, Tópicos Especiais em Banco de Dados e Engenharia de Software, Tópicos Especiais em Inteligência Artificial, Tópicos Especiais em Computação e Programação de Computadores. Essa matriz possuía 3.680 horas e as disciplinas eram distribuídas entre nove semestres.

Em 2016, o Núcleo Docente Estruturante (NDE) do curso iniciou um estudo sobre

a matriz curricular e identificou que uma reorganização poderia proporcionar uma melhor adaptação dos discentes ao curso. A proposta, inspirada em universidades como a Universidade Federal do ABC (UFABC), definiu que os conteúdos fossem apresentados de forma horizontal, ou seja, os fundamentos da área de Computação seriam abordados inicialmente em disciplinas obrigatórias, oferecendo uma visão geral da área. Posteriormente, os discentes poderiam aprofundar seus estudos em disciplinas optativas, escolhidas de acordo com suas aptidões e interesses.

Assim sendo, uma nova matriz curricular (M_{2020}) foi proposta com o intuito de dinamizar e acompanhar os avanços científicos e tecnológicos da Computação. Nessa matriz, foi dada uma autonomia maior aos discentes para cursar um conjunto de disciplinas escolhidas por eles de acordo com as áreas de conhecimento de seu interesse e aptidão. Além desta reestruturação de conteúdo, houve a redução de nove semestres para oito. A matriz M_{2020} totaliza 3.200 horas.

Foram criados quatro núcleos que nortearam a divisão dos conteúdos, são eles: Núcleo da Matemática, Núcleo da Programação, Núcleo de Formação Horizontal e Núcleo de Formação Específica. Cada núcleo tem como objetivo uma mobilização dos conhecimentos e esquemas para desenvolver respostas inéditas, criativas e eficazes para problemas da área da Computação.

No Núcleo de Formação Horizontal, os conteúdos básicos da Computação são apresentados em sua forma mais ampla e permitem que os discentes conheçam e desenvolvessem as habilidades necessárias básicas para um cientista da Computação. Esse primeiro contato permite que o discente aprofunde, posteriormente, o conhecimento no Núcleo de Formação Específica, que é formado por um conjunto de onze disciplinas optativas, divididas em quatro eixos de conhecimento, sendo eles: *i)* Algoritmos e Combinatória, *ii)* Engenharia de Sistemas e Informação, *iii)* Sistemas de Computação e *iv)* Sistemas Inteligentes.

Os discentes que cumprem um conjunto de disciplinas e horas mínimas de um eixo podem solicitar um Certificado de Estudos Especiais do eixo cumprido. Este certificado pode ser utilizado pelo discente como uma forma de certificação da área de aptidão e de desenvolvimento de aprendizado durante a graduação. A nova matriz entrou em vigor em 2020. Desde então, tem-se realizado um trabalho de divulgação dos eixos, bem como a criação de disciplinas optativas sob demanda.

Como uma forma de exemplificar a mudança realizada na matriz curricular, será apresentado a seguir como as disciplinas referentes à área de Arquitetura de Computadores são distribuídas nas duas matrizes em questão. Na matriz M_{2014} , eram ofertadas três disciplinas obrigatórias: *i)* Fundamentos de Sistemas, na qual se apresentava a Álgebra de Boole, fundamental para o bom desenvolvimento de circuitos digitais; *ii)* Arquitetura de Computadores I, que abrangia conceitos de organização de computadores, linguagem de máquina e desempenho; e *iii)* Arquitetura de Computadores II, que tinha foco no caminho de dados, hierarquia de memória (*cache*) entre outros. Já na matriz M_{2020} , são oferecidas as seguintes disciplinas obrigatórias referentes à área: 1) Circuitos Digitais, que contempla a Álgebra de Boole; e 2) Arquitetura de Computadores, na qual são vistos de forma horizontal os conceitos de organização dos computadores, hierarquia de memória, conjuntos de instruções (MIPS) e desempenho. Detalhes de cada um desses conteúdos

podem ser bem aprofundados pelos discentes interessados em disciplinas optativas como Tópicos Especiais em Arquitetura de Computadores e Arquitetura de Computadores II. Logo, todos os discentes têm a possibilidade de desenvolver habilidades na área, mas só os discentes que possuem mais aptidão e escolhem esta área irão se aprofundar nesses conhecimentos.

Essa mudança foi muito bem aceita pelos discentes que têm se apressado em cumprir determinado eixo de formação, já planejando sua empregabilidade.

No entanto, apesar de a matriz M_{2020} ter sido implantada há pouco tempo, surgiu a necessidade de avaliar se estas possibilidades dentro da matriz têm ajudado o discente a não desistir do curso, visto que a falta de afinidade com algumas áreas dentro da Computação se apresenta como uma das causas de evasão nos cursos [Klitzke and Carvalhaes 2023]. Logo, este estudo tem como objetivo responder se a alteração da matriz contribuiu para diminuir a evasão. Nas próximas seções serão apresentadas como este estudo preliminar foi conduzido.

4. Análise de Sobrevivência

Neste estudo, optou-se por utilizar a ferramenta estatística Análise de Sobrevivência [Mills 2010]. Apesar de ser muito utilizada na área da saúde, é uma ferramenta também utilizada em estudos sobre evasão no ensino superior. Em [Ishitani 2003] é avaliada a questão da evasão do curso de graduação pelos discentes que são os primeiros das suas famílias a ingressarem em cursos de nível superior. Os autores não especificam o curso, mas concluem que a maior probabilidade de abandono ocorre no primeiro período. Em [Radcliffe et al. 2006], os autores empregam a Análise de Sobrevivência para tentar prever se um discente irá abandonar o curso e quando isso ocorrerá. Por outro lado, [Murtaugh et al. 1999] buscam identificar fatores e propõem um modelo para prever a evasão na Universidade de Oregon. Em [Saccaro et al. 2019], os autores desenvolvem um estudo sobre a evasão dos cursos superiores na área de Ciências da Terra.

A abordagem da Análise de Sobrevivência observa a ocorrência de um evento ao longo de um tempo, onde o evento irá acontecer para uma parte de um grupo e para a outra não. Para isso, utiliza-se de uma função de sobrevivência definida por $S(t) = Pr(T > t)$, que é a probabilidade da ocorrência de um evento após um instante t para um indivíduo. A função de sobrevivência $S(t)$ pode ser interpretada como a fração de pessoas que sobreviveram a um evento. A partir disso, $F(t) = Pr(T \leq t)$ é, então, a fração de indivíduos atingidos pelo evento terminal em função do tempo. Logo, para qualquer instante de tempo t , tem-se que $F(t) + S(t) = 1$ e $F(t)$ é uma função crescente, enquanto $S(t)$ é decrescente.

Outra função relevante é a de densidade de um evento, dado por $f(t)$ e expressa a probabilidade de que o evento analisado ocorra para um indivíduo qualquer da população em um intervalo de tempo: $t \leq T < t + \Delta T$, conforme: $F(t + \Delta t) - F(t) \rightarrow f(t) = \frac{dF}{dt}$. A função de risco, $h(t)$, mede a taxa condicional de falha no tempo t , dado que o discente não evadiu até t , sendo expressa por: $h(t) = \frac{f(t)}{S(t)}$.

A Análise de Sobrevivência pode ser realizada adotando-se técnicas não paramétricas ou técnicas paramétricas. As técnicas não paramétricas são mais simples, flexíveis e não requerem nenhuma restrição sobre a distribuição dos tempos de vida dos

indivíduos. Por outro lado, são limitadas em relação ao número de variáveis trabalhadas ao longo do tempo. A técnica não paramétrica mais conhecida na literatura é chamada de estimador de Kaplan-Meier (EKM) [Kaplan and Meier 1958].

O EKM é compreendido como a probabilidade de ter sobrevivido a N eventos em um certo período de tempo analisado. Por exemplo, ao analisar um discente que cursa o oitavo período, pode-se dizer que esse discente sobreviveu aos oito períodos [Colosimo and Giolo 2021]. O estimador é construído a partir do produto das probabilidades de sobrevivência a cada intervalo de tempo.

Supondo que um certo estudo considere a sobrevivência em N medidas de tempo e que a amostra tenha k observações não censuradas. Considere ainda que essas medidas de tempo estejam ordenadas de forma cronológica $t_1 < t_2 < t_3 < \dots < t_{k-1} < t_k$. Dado que d_i seja o número de eventos terminais ocorridos no i -ésimo intervalo de tempo (t_i) e n_i o número de indivíduos que poderiam ser atingidos pelo evento terminal no tempo t_i .

Logo, o EKM pode ser definido como: $\hat{S}_{KM}(t) = \prod_{t_i=0}^t \left(1 - \frac{d_i}{n_i}\right)$.

A abordagem não paramétrica baseada em EKM permite realizar testes de significância estatística em amostras, destacando três principais métodos: Logrank, ou teste generalizado de Savage; Breslow, ou teste generalizado de Wilcoxon; e Tarone-Ware. São testes que avaliam a significância estatística entre amostras sob a hipótese nula, atribuindo pesos para melhor descrever as diferenças nas curvas ao longo do tempo. O teste Logrank dá maior ênfase às diferenças no final das distribuições de tempo de vida; por outro lado, o teste Breslow prioriza diferenças no início das distribuições; e, por fim, o teste Tarone-Ware foca nas diferenças em fases intermediárias da curva de distribuição de tempo de vida. Cada método adapta os pesos conforme o momento da evolução temporal, ajustando-se às características específicas das distribuições. A Seção 5 apresenta os resultados do EKM e os testes estatísticos dos três métodos adotados neste estudo.

Já as técnicas paramétricas, também conhecidas como modelos de vida acelerada, são caracterizadas por serem rígidas em relação à distribuição do tempo de vida do grupo de indivíduos. Embora permitam trabalhar com um maior número de variáveis, essas técnicas são menos utilizadas devido à sua falta de flexibilidade. Dentre os modelos paramétricos na literatura, adotou-se neste trabalho o modelo de Tempo de Falhas Acelerado (TFA) [Lima Junior et al. 2012, Saccaro et al. 2019], que emprega o logaritmo natural do tempo de sobrevivência, $\ln(T)$. Por sua vez, esse tempo T pode ser expresso como uma função linear das covariadas, conforme Eq.: $\ln(T) = X\hat{\beta} + \hat{\epsilon}$. Onde X é um vetor de covariadas, $\hat{\beta}$ é o vetor com os coeficientes obtidos com a regressão e o vetor de erros é dado por $\hat{\epsilon}$, dado pelos erros obtidos com a regressão. Os resultados são usualmente interpretados com o uso de Taxas de Tempo na estimação, onde cada variável vai afetar diretamente o tempo de sobrevivência. A Taxa de Tempo pode ser obtida pela Eq. 1.

$$e^{\beta_k} = \frac{T(X_k + 1)}{T(X_k)} \quad (1)$$

Desta forma, neste estudo, os dados são analisados em duas partes: na primeira, o EKM é empregado para analisar o conjunto de dados e o impacto da mudança para a matriz curricular M_{2020} na evasão. Na segunda parte, os dados são analisados através

do modelo paramétrico de TFA, o que permite verificar a influência de outras variáveis na taxa de evasão, além da variável matriz curricular.

Atualmente, há diversas ferramentas estatísticas para realizar a Análise de Sobrevida. Neste trabalho, foi utilizado o R e seus respectivos pacotes¹.

5. Análise dos Históricos e Resultados

Até 2024, ingressaram 934 discentes no curso de CC da UFRRJ. Desses, foram excluídos da análise 32 discentes, pois ingressaram no curso por meio de reintegração, o que os faria aparecer duas vezes na estatística: uma vez com a matrícula da entrada usual e outra com a matrícula da entrada por reintegração. Assim, a análise incluiu 902 discentes distribuídos ao longo dos períodos de ingresso e dos diferentes status, definidos como: **Ativos**, são discentes regularmente matriculados e cursando disciplinas; **Concluídos**, representam os discentes que terminaram o curso; **Evadidos**, são os discentes que ou abandonaram o curso ou foram desligados pela Universidade; e **Trancados**, são discentes com matrícula trancada no momento da análise dos históricos.

A Figura 1 mostra como os discentes estão distribuídos ao longo dos anos de existência do curso com seus status. Nota-se que os evadidos representam uma parcela grande dos discentes. Em média, há aproximadamente 12 discentes formados por ano, com um desvio padrão de 9,37. Ao passo que a média de evadidos é o dobro, com 26,6 por entrada e um desvio padrão de 12,01.

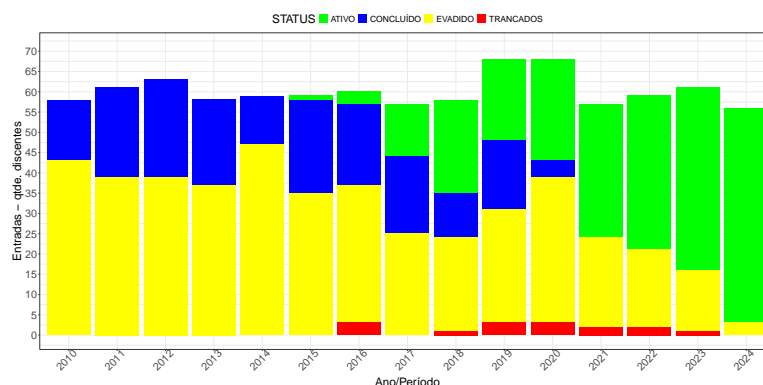


Figura 1. Distribuição dos discentes por entrada segmentado por status.

Na Análise de Sobrevida, foram considerados somente os ingressantes a partir de 2014, pois a matriz curricular utilizada antes de 2014 possuía um foco em Matemática Aplicada e, dessa forma, poderia enviesar o resultado.

Na Análise de Sobrevida não paramétrica, utilizou-se o EKM para verificar se a mudança da matriz curricular teve algum impacto na evasão. A variável considerada no modelo foi *matriz*, sendo atribuído o valor 0 à matriz M_{2014} , que inclui ingressantes de 2014 a 2019, e 1 à matriz M_{2020} , que inclui os discentes que entraram a partir de 2020.

As Figuras 2 e 3 mostram, respectivamente, a sobrevivência e o risco de evasão. A sobrevivência da matriz M_{2014} é ligeiramente menor que a M_{2020} ; apenas 17,4%

¹Dados e *scripts* em R utilizados neste trabalho estão disponíveis em: <https://github.com/mzamith-ufrrj/evacao-2024-ufrrj-im>

dos matriculados chegaram ao final do curso, com 50% de sobrevivência sendo alcançada no 14º período. Observando a curva de sobrevivência de M_{2020} , os resultados são mais otimistas, onde a sobrevivência se mantém em torno 53%. Entretanto, o padrão de alta evasão no início do curso é visível em ambas as curvas.

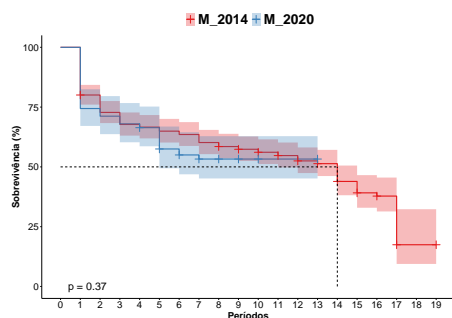


Figura 2. EKM: sobrevivência.

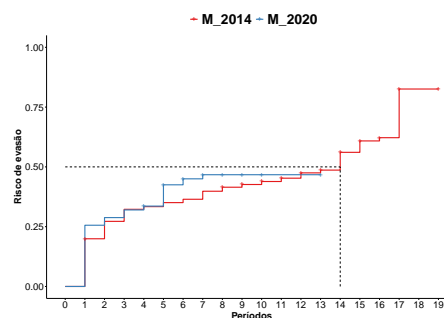


Figura 3. EKM: risco acumulado.

Ainda vale notar que a evasão na matriz M_{2014} ocorre regularmente ao longo dos períodos, ao passo que na matriz M_{2020} apresenta um platô a partir do 7º período. Acredita-se que, como a M_{2020} possui 8 períodos, o discente pode se sentir mais motivado a terminar o curso já que está próximo do cumprimento total da carga horária. Além disto, a partir do 6º período, os discentes iniciam o aprofundamento do conhecimento com as disciplinas optativas e começam, portanto, a idealizar sua carreira e área de trabalho de forma mais efetiva.

As Tabelas 1 e 2 mostram como a sobrevivência decai ao longo dos períodos de acordo com a matriz. A coluna censurado em ambas as tabelas mostra os discentes que terminaram o curso e, neste caso, eles figuram como “se estivessem desistido da pesquisa” [Hazra and Gogtay 2017]. Foram aplicados três testes de significância estatística para comparar as matrizes: Logrank, Breslow e Tarone-War que obtiveram, respectivamente, os p-valores de 41,24%, 30,76% e 34,31%. Embora as amostras sejam parecidas, a matriz M_{2020} apresenta uma taxa de sobrevivência maior quando comparada com a taxa da matriz M_{2014} . Nota-se ainda que a evasão dos discentes da M_{2020} torna-se constante a partir do 7º período com uma sobrevivência de $S(t) = 0,53$, ao passo que na M_{2014} a evasão é constante no curso, mesmo após os períodos iniciais.

O modelo não-paramétrico EKM foi utilizado para descrever a sobrevivência considerando apenas um único parâmetro, que neste caso foi a matriz curricular do ingressante. Por outro lado, os modelos paramétricos permitem a inclusão de múltiplas variáveis explicativas. Para avaliar outras variáveis, foi utilizado o modelo paramétrico TFA, estimado com as distribuições Weibull, Exponencial e Lognormal.

As variáveis selecionadas foram as seguintes: *i*) as **matrizes** M_{2014} e M_{2020} ; *ii*) o **Índice de Rendimento Acadêmico (IRA)**, que avalia o desempenho dos discentes, variando de 0 a 10, com média de aprovação em 5. Para análise, o IRA foi categorizado em faixas: valor 0 para $(IRA < 2,5)$, 1 para $(2,5 \leq IRA < 5,0)$, 2 para $(5,0 \leq IRA < 7,5)$ e 3 para $(IRA \geq 7,5)$; *iii*) as **Suspensões** que são variáveis inteiras que indicam o número de períodos em que os discentes trancaram o curso, interrompendo temporariamente o prazo de integralização; *iv*) as **Prorrogações** que são variáveis inteiras que indicam o número de períodos que estenderam o prazo após o período regular de integralização; e,

Tabela 1. Matriz Curricular M_{2014}

| Período | Discentes | Evento | Censurado | $S(t)$ | σ | IC superior | IC inferior |
|---------|-----------|--------|-----------|--------|----------|-------------|-------------|
| 1 | 361 | 72 | 3 | 0,801 | 0,026 | 0,843 | 0,760 |
| 2 | 286 | 26 | 0 | 0,728 | 0,032 | 0,775 | 0,683 |
| 3 | 260 | 18 | 0 | 0,677 | 0,036 | 0,728 | 0,631 |
| 4 | 242 | 4 | 0 | 0,666 | 0,037 | 0,717 | 0,619 |
| 5 | 238 | 6 | 0 | 0,649 | 0,039 | 0,701 | 0,602 |
| 6 | 232 | 5 | 0 | 0,635 | 0,040 | 0,687 | 0,588 |
| 7 | 227 | 12 | 0 | 0,602 | 0,043 | 0,655 | 0,553 |
| 8 | 215 | 6 | 3 | 0,585 | 0,044 | 0,638 | 0,536 |
| 9 | 206 | 4 | 19 | 0,574 | 0,046 | 0,627 | 0,525 |
| 10 | 183 | 4 | 17 | 0,561 | 0,047 | 0,615 | 0,512 |
| 11 | 162 | 4 | 35 | 0,547 | 0,049 | 0,602 | 0,498 |
| 12 | 123 | 5 | 27 | 0,525 | 0,052 | 0,581 | 0,474 |
| 13 | 91 | 2 | 13 | 0,513 | 0,054 | 0,571 | 0,462 |
| 14 | 76 | 11 | 19 | 0,439 | 0,072 | 0,506 | 0,381 |
| 15 | 46 | 5 | 12 | 0,391 | 0,088 | 0,466 | 0,329 |
| 16 | 29 | 1 | 15 | 0,378 | 0,095 | 0,455 | 0,314 |
| 17 | 13 | 7 | 3 | 0,174 | 0,314 | 0,323 | 0,094 |
| 19 | 3 | 0 | 3 | 0,174 | 0,314 | 0,323 | 0,094 |

Tabela 2. Matriz Curricular M_{2020}

| Período | Discentes | Evento | Censurado | $S(t)$ | σ | IC superior | IC inferior |
|---------|-----------|--------|-----------|--------|----------|-------------|-------------|
| 1 | 125 | 32 | 0 | 0,744 | 0,052 | 0,825 | 0,671 |
| 2 | 93 | 4 | 0 | 0,712 | 0,057 | 0,796 | 0,637 |
| 3 | 89 | 4 | 0 | 0,680 | 0,061 | 0,767 | 0,603 |
| 4 | 85 | 2 | 1 | 0,664 | 0,064 | 0,752 | 0,586 |
| 5 | 82 | 11 | 2 | 0,575 | 0,077 | 0,669 | 0,494 |
| 6 | 69 | 3 | 2 | 0,550 | 0,081 | 0,645 | 0,469 |
| 7 | 64 | 2 | 4 | 0,533 | 0,084 | 0,628 | 0,452 |
| 8 | 58 | 0 | 32 | 0,533 | 0,084 | 0,628 | 0,452 |
| 9 | 26 | 0 | 7 | 0,533 | 0,084 | 0,628 | 0,452 |
| 10 | 19 | 0 | 18 | 0,533 | 0,084 | 0,628 | 0,452 |
| 13 | 1 | 0 | 1 | 0,533 | 0,084 | 0,628 | 0,452 |

por fim, v) a **Idade**, que utiliza uma variável contínua.

Para determinar qual das três distribuições melhor descreve o conjunto de dados, foi aplicado o Critério de Informação de Akaike (CIA). Entre as distribuições avaliadas, a Lognormal apresentou o menor valor de CIA, indicando que é a mais adequada para descrever os dados, onde a distribuição Weibull obteve o valor de 1.352,324, a Exponencial obteve 1.363,205 e, com o menor valor, a Lognormal obteve 1.307,792.

A Tabela 3 apresenta os coeficientes da regressão por TFA com distribuição Lognormal, mostrando a influência de cada variável no tempo de sobrevivência acadêmica dos discentes.

Tabela 3. Coeficientes de regressão

| Variável | Coeficiente | σ_x | Z | p-valor |
|----------------------------|-------------|------------|-------|----------|
| β_0 | 0,4744 | 0,3080 | 1,54 | 0,123 |
| Matriz (β_1) | -0,2392 | 0,1695 | 0,13 | 0,897 |
| IRA (β_2) | 1,3333 | 0,0632 | 21,09 | 2,00E-16 |
| Suspensões (β_3) | 0,1135 | 0,0448 | 2,53 | 0,011 |
| Prorrogações (β_4) | 0,1316 | 0,1258 | 1,05 | 0,296 |
| Idade (β_5) | 0,0112 | 0,0153 | 0,73 | 0,464 |

Ao analisar esses resultados, observa-se que os discentes que cursam a matriz M_{2020} apresentam um tempo menor de conclusão, de 21,3%, o que traduz em um tempo menor de sobrevivência, quando comparado com a matriz M_{2014} . O IRA tem

um grande impacto na sobrevivência, pois cada faixa (1, 2 ou 3) aumenta as chances de sobrevivência em 279,4%. Logo, discentes com IRA entre 2,5 e 5 possuem quase três vezes mais chance de permanecer no curso quando comparados aos que possuem IRA abaixo de 2,5. Por sua vez, discentes com IRA acima de 5 têm até 10 vezes mais chance de sobreviver em relação aos da primeira faixa.

A variável suspensões aumenta o tempo de sobrevivência em 12,0% por período de suspensão. Vale ressaltar que suspensões não são contabilizadas no tempo de integralização, funcionando como uma ferramenta importante para reduzir a evasão. De forma similar, cada prorrogação aumenta a chance de sobrevivência em 14,1%. Ambas as variáveis oferecem suporte aos discentes, auxiliando-os a superar alguma dificuldade encontrada no decorrer do curso.

Já a variável idade contribui pouco para o tempo de sobrevivência, indicando um aumento de 1,1% na chance de sobrevivência para cada ano a mais de idade. Assim, discentes mais velhos, ao ingressar no curso, apresentam mais chances de permanecer, sugerindo que a maturidade e o foco nos objetivos contribuem para a conclusão do curso.

6. Considerações Finais e Trabalhos Futuros

Este trabalho teve como objetivo analisar os dados do curso de CC da UFRRJ em relação à evasão. Este foi o primeiro estudo mais detalhado realizado no curso, que passou por algumas alterações de matriz curricular, sendo duas mais significativas. A matriz proposta em 2020 foi inovadora, pois permite ao discente ser protagonista em sua formação.

A Análise de Sobrevivência revelou que a matriz M_{2020} apresenta um menor risco de evasão. Além disso, mostra que os discentes que alcançam o sétimo período têm uma alta probabilidade de concluir o curso. Outro aspecto relevante é que o tempo médio de conclusão do curso na matriz M_{2020} é menor em comparação à matriz M_{2014} .

O modelo paramétrico TFA revelou que um IRA mais baixo aumenta a probabilidade de evasão, enquanto discentes que utilizam suspensão temporária ou prorrogações têm maior chance de concluir o curso, embora com maior tempo de permanência.

Uma das limitações deste trabalho é o número reduzido de amostras disponíveis da matriz M_{2020} , o que torna os resultados obtidos preliminares e sujeitos a acompanhamento ao longo do tempo. Além disso, a implementação da nova matriz coincidiu com a pandemia da Covid-19, um dos períodos mais desafiadores vividos por todos. Dado que ambas as mudanças – a mudança curricular e os efeitos da pandemia – ocorreram simultaneamente, ainda não é possível avaliar isoladamente seus impactos no curso.

Como trabalho futuro, se faz necessário entender melhor os motivos pelos quais os discentes evadem nos períodos iniciais. Alguns trabalhos na literatura mostram que a dificuldade nas disciplinas está entre os motivos da evasão inicial; contudo, muitos discentes evadem mesmo antes das primeiras provas. Acredita-se que algumas variáveis (tais como: distância da universidade, dificuldade financeira, a falta de conhecimento da área da Computação e a falta de conhecimento do perfil do egresso do curso de CC) podem também impactar na decisão do discente de evadir. Esses aspectos não foram considerados neste trabalho e serão objetos de futuras investigações, seja por análise documental ou entrevistas com discentes. Aprofundar uma investigação neste sentido seria fundamental para apresentar propostas eficazes para a permanência dos discentes no curso.

Referências

- Alvim, I., Bittencourt, R., and Duran, R. (2024). Evasão nos Cursos de Graduação em Computação no Brasil. In *Anais do IV Simpósio Brasileiro de Educação em Computação*, pages 1–11, Porto Alegre, RS, Brasil. SBC.
- Andriola, W. B., Andriola, C. G., and Moura, C. P. (2006). Opiniões de Docentes e de Coordenadores acerca do Fenômeno da Evasão Discente dos Cursos de Graduação da Universidade Federal do Ceará (UFC). *Ensaio: Avaliação e Políticas Públicas em Educação*, 14:365 – 382.
- BRASSCOM, A. (2024). Relatório Setorial de TIC. Disponível em: <https://brasscom.org.br/inteligencia/relatorio-setorial/>, Acessado em: dezembro/2024.
- Carvalho, A. L. (2018). Construção de um Modelo de Acompanhamento do Desempenho Acadêmico (MADA) dos Cursos de Graduação da UFRRJ, por meio da Análise da Evasão Acadêmica. Master's thesis, Universidade Federal Rural do Rio de Janeiro, Mestrado Profissional em Gestão e Estratégia - MPGE.
- Chen, Y., Johri, A., Johri, A., Rangwala, H., and Rangwala, H. (2018). Running out of STEM: a Comparative Study across STEM Majors of College Students At-risk of Dropping Out Early. In *International Conference on Learning Analytics and Knowledge (LAK'18)*, page 270–279, Sydney, Australia. Association for Computing Machinery.
- Colosimo, E. A. and Giolo, S. R. (2021). *Análise de Sobrevivência Aplicada*. Editora Blucher.
- Diogo, M. F., Raymundo, L. S., Wilhelm, F. A., Andrade, S. P. C., Lorenzo, F. M., Rost, F. T., and Bardagi, M. P. (2016). Percepções de Coordenadores de Curso Superior sobre Evasão, Reprovações e Estratégias Preventivas. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, 21:125 – 151.
- Duran, R., Bim, S., Gimenes, I., Ribeiro, L., and Correia, R. (2023). Potential Factors for Retention and Intent to Drop-out in Brazilian Computing Programs. *ACM Transactions on Computing Education*, 23.
- Giannakos, M., Pappas, I., Jaccheri, L., and Sampson, D. (2016). Understanding Student Retention in Computer Science Education: The Role of Environment, Gains, Barriers and Usefulness. *Education and Information Technologies*, 22(5):2365–2382.
- Hazra, A. and Gogtay, N. (2017). Biostatistics series module 9: survival analysis. *Indian journal of dermatology*, 62(3):251–257.
- Hoed, R. M. (2016). Análise da Evasão em Cursos Superiores: o Caso da Evasão em Cursos Superiores da Area de Computação. Master's thesis, Universidade de Brasília, Instituto de Ciências Exatas Departamento de Ciência da Computação.
- Ishitani, T. T. (2003). A Longitudinal Approach to Assessing Attrition Behavior among First-generation Students: Time-varying Effects of Pre-college Characteristics. *Research in higher education*, 44(4):433–449.
- 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.
- 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.
- Mills, M. (2010). *Introducing Survival and Event History Analysis*. Sage.
- Murtaugh, P. A., Burns, L. D., and Schuster, J. (1999). Predicting the Retention of University Students. *Research in higher education*, 40(3):355–371.
- Palmeira, L. B. and Santos, M. P. (2014). Evasão no Bacharelado em Ciência da Computação da Universidade de Brasília: Análise e Mineração de Dados. Master's thesis, Universidade de Brasília, Instituto de Ciências Exatas Departamento de Ciência da Computação.
- Radcliffe, P. M., Huesman Jr, R. L., and Kellogg, J. P. (2006). Modeling the Incidence and Timing of Student Attrition: A Survival Analysis Approach to Retention Analysis. *Association for Institutional Research in the Upper Midwest (AIRUM)*.
- Saccaro, A., França, M. T. A., and Jacinto, P. 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(2):337–373.
- Santos, M. S. (2016). Uma Análise Probabilística da Retenção na Universidade Federal da Bahia: um Estudo de Caso no Curso de Ciência da Computação. Master's thesis, Universidade Federal da Bahia, Pós Graduação em Ciência da Computação.
- SBC, S. (2005). Currículo de Referência da SBC para Cursos de Graduação em Bacharelado em Ciência da Computação e Engenharia de Computação. Disponível em: <https://www.sbc.org.br/wp-content/uploads/2024/07/Curriculo-de-Referencia-CC-EC-versao2005.pdf>, Acessado em: dezembro/2024.
- Schoeffel, P., Wazlawick, R., Faria, V., and Ramos, C. (2018). Motivation and Engagement Factors of Undergraduate Students in Computing: A Systematic Mapping Study. In *2018 IEEE Frontiers in Education Conference (FIE)*, pages 1–5, San Jose, CA, USA. IEEE.
- Silva, G. P. (2013). Análise de Evasão no Ensino Superior: uma Proposta de Diagnóstico de seus Determinantes. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, 18:311 – 333.