

Um Estudo da Evasão no Curso de Licenciatura em Computação da Universidade de Brasília

Richard Wallan Paulino de Sousa¹, Juliana Betini Fachini-Gomes¹,
Maristela Holanda², Maria Teresa Costa Leão¹

¹Departamento de Estatística - Universidade de Brasília (UnB)

²Departamento de Ciências da Computação - Universidade de Brasília (UnB)

{richard, jfachini, mholanda, teresa}@unb.br

Abstract. *The Computing Licenciature major at the University of Brasília (UnB), similar to other Computing courses in Brazil, is impacted by dropouts. Therefore, the objective of this work is to answer the following research question “What academic and social factors impact the probability of dropping out of the Computing Degree course at the University of Brasília?”. The data used for this study corresponds to the academic trajectory of undergraduate students in Computing from 2012/2 to 2019/2. The survival analysis methodology was the technique used, more specifically the Log-Normal regression model. The model proved to be robust in the analysis of residues and in presenting results consistent with the evasion literature. The results of this article can help in the development of educational policies and strategies to reduce the number of dropouts in the LC course at UnB.*

Resumo. *O curso de Licenciatura em Computação (LC) da Universidade de Brasília (UnB), semelhante a outros cursos de Computação no Brasil, é impactado com a evasão. Por isso, o objetivo deste trabalho é responder a seguinte pergunta de pesquisa “Quais fatores acadêmicos e sociais impactam na probabilidade de evadir do curso de Licenciatura em Computação da Universidade de Brasília?”. Os dados utilizados para este estudo corresponde a trajetória acadêmica dos alunos de LC no período de 2012/2 a 2019/2. A metodologia de análise de sobrevivência foi a técnica utilizada, mais especificamente o modelo de regressão Log-Normal. O modelo mostrou-se robusto na análise de resíduos e em apresentar resultados consistentes com a literatura de evasão. Por exemplo, através do modelo é possível entender que os alunos com IRA alto têm maior probabilidade de sobreviver à evasão. Como, cursar disciplinas estratégicas durante o verão, diminui a evasão. Esse e os demais resultados deste artigo evidenciam a necessidade de desenvolvimento de políticas e estratégias educacionais para diminuir a evasão no curso de LC na UnB.*

1. Introdução

A evasão escolar está presente desde os níveis básicos de educação até o ensino superior. E está relacionada a fatores de diferentes naturezas, tais como, a satisfação com a instituição de ensino [Ribeiro et al. 2021] e outras variáveis relacionadas a fatores sociais e individuais do aluno como: desempenho acadêmico, escolha profissional e indicadores

econômicos [Fritsch et al. 2015]. Especificamente em computação destaca-se a dificuldade nas disciplinas de programação e cálculo [Duran et al. 2023].

Existe diferentes definições para evasão, neste artigo é usada a presente em [Fritsch et al. 2015]; “a evasão escolar está relacionada à perda de estudantes que iniciam, mas não concluem seus cursos” ou seja, “é a desistência do aluno do seu curso de origem por qualquer motivo, exceto conclusão ou diplomação”.

Neste contexto, este artigo tem a seguinte questão de pesquisa: “Quais fatores acadêmicos e sociais impactam na probabilidade de evadir do curso de Licenciatura em Computação da Universidade de Brasília?”. E a metodologia utilizada para responder essa questão foi análise de sobrevivência. Alguns fatores como, por exemplo: idade, sistema de cotas, tipo de escola, entre outros, irão compor o modelo de regressão. A organização do banco de dados e todas as análises foram realizadas no *software* R.

O presente artigo está organizado nas seguintes seções: Seção 2 introduz os trabalhos relacionados; Seção 3 descreve a metodologia aplicada nesta pesquisa; Seção 4 apresenta os resultados e discussões do estudo; Seção 5 aborda as limitações do trabalho. Por fim, as conclusões estão na Seção 6.

2. Trabalhos Relacionados

Alguns trabalhos na literatura envolvem o estudo de evasão em cursos de Computação no Brasil como [Duran 2024, Duran et al. 2023]. Porém, especificamente em Licenciatura, são apresentados os seguintes trabalhos.

Em [Souza et al. 2015] é apresentando um estudo sobre a evasão no curso de Licenciatura em Informática do IFRN – Campus Natal – em que aplicou um questionário com o objetivo de conhecer o perfil dos alunos desistentes e as razões que os levaram a abandonar o curso. Em [Vasconcelos and Andrade 2018] também utilizou questionários aplicados a professores e alunos do curso, na Universidade Federal Rural de Pernambuco, e foram identificados problemas que podem levar os alunos a evadir no curso.

Em [Calixto 2015] um estudo sobre as causas de evasão discente no curso de Licenciatura em Computação na Universidade Federal da Paraíba (UFPB) virtual no formato UAB foi apresentado. Os dados foram de entrevista semi-estruturada. As causas identificadas para evasão discente foram: falta de tempo para a realização das atividades propostas, crença de que os cursos à distância requerem menos esforços e dificuldades em participar das atividades presenciais, dentre outros.

Em [Prietch and Pazeto 2010] apresentou possíveis fatores da evasão no curso de Licenciatura Plena em Informática do Campus Universitário de Rondonópolis, Universidade Federal de Mato Grosso (UFMT/CUR), por meio de informações obtidas no sistema acadêmico da Instituição, bem como mediante depoimento dos alunos.

Diferentemente dos artigos apresentados anteriormente, o presente trabalho traz um estudo usando análise de sobrevivência em dados acadêmicos e sociais dos alunos de licenciatura da Universidade de Brasília.

3. Metodologia

A metodologia desta pesquisa é descrita nesta seção, sendo composta de: Seção 3.1 onde os dados da análise são apresentados e Seção 3.3 com o método de análise de sobre-

vivência aplicado.

3.1. Dados

Os dados utilizados para este estudo foram disponibilizados pela UnB, e corresponde a trajetória acadêmica dos alunos de licenciatura em Computação (LC) no período de 2012/2 a 2019/2. Esse período foi considerado por conter informações de dois currículos. O que torna possível estudar o efeito do currículo na evasão do aluno, na qual a hipótese inicial é de que o currículo novo aumenta a probabilidade do aluno não evadir. O período final considerado foi o momento anterior a pandemia de Covid-19 (2019).

O banco de dados original para o curso de LC tinha 348.679 observações e 22 variáveis. As variáveis utilizadas neste estudo são: Índice de rendimento acadêmico (IRA) que representa numericamente o rendimento acadêmico geral do estudante - valores entre 0,0 e 5,0; Data de nascimento; Sistema de cotas (sim ou não); Escola (pública ou privada); Período de ingresso na UnB (ano/semestre); Período de ingresso no curso (ano/semestre); Forma de ingresso na UnB (PAS, Vestibular,...); Período de saída do curso (ano/semestre); Forma de saída do curso (informação se o aluno se formou, se está ativo ou se mudou de curso); Período que cursou a disciplina (ano/semestre); Mínimo de créditos para se formar; Créditos total no período; Total de créditos cursados pelo aluno; Créditos aprovados no período; Código da disciplina; Nome da disciplina; Créditos da disciplina; Menção do aluno na disciplina - SR, II, MI, MM, MS e SS.

3.2. Transformação do dados

A variável resposta do estudo, tempo até a evasão, foi construída utilizando as variáveis período de entrada no curso de LC e o período de saída do curso de LC. A variável tempo, medida em unidades semestrais, é criada ao subtrair o período de saída pelo período de entrada em LC. Essa variável informa o tempo que o aluno passou no curso e esses tempos são tanto de alunos que evadiram como de alunos que não evadiram. Então, é necessário criar uma variável indicadora de falha e censura, denominada *status*. Como pode ser observado na Tabela 1, caso o estudante esteja matriculado no curso de LC até 2019/2 ou se o estudante concluiu o curso de LC, a variável *status* assume que o tempo até a evasão é um tempo censurado. Caso contrário, a variável *status* assume que o tempo até a evasão é um tempo de falha. Para construir esta variável foi utilizada a variável original forma de saída do curso.

Durante o verão, a universidade oferece algumas disciplinas com os mesmos conteúdos das disciplinas que acontecem durante os tradicionais semestres acadêmicos. Porém, as disciplinas de verão acontecem de forma concentrada, com mais horas de aula durante a semana. Para a construção da variável Verão foi utilizado o período que o aluno cursou a disciplina. Essa variável assume resposta "Sim" se o aluno cursou alguma disciplina no verão e "Não", caso contrário.

O período de estudo (2012/2 até 2019/2) considerado neste artigo contempla dois diferentes currículos do curso de LC. Desse modo, este trabalho também teve o interesse em investigar se os diferentes currículos contribuem na probabilidade de um estudante evadir. Então, a variável Currículo foi criada e recebe resposta "Velho" se o estudante ingressou em LC no período de 2012/2 a 2015/1. E "Novo" se o estudante ingressou em LC no período de 2015/2 a 2019/2.

Tabela 1. Formas de saída do curso e a construção da variável *status*

Forma de saída	<i>Status</i>
Ativo	Censura
Formatura	Censura
Deslig - não cumpriu condição	Falha
Deslig - abandono	Falha
Desligamento voluntário	Falha
Deslig - decisão judicial	Falha
Mudança de curso	Falha
Novo vestibular	Falha
Reprovou três vezes na mesma disciplina	Falha
Transferência	Falha

Para avaliar se algumas informações relacionadas as disciplinas impactam na probabilidade do aluno evadir do curso, também foram construídas as variáveis Média de créditos p/ semestre, Soma de créditos reprovados e Soma de créditos trancados. A variável Média de créditos p/ semestre foi construída para medir quantos créditos, em média, o aluno i cursa por semestre até o fim do seu curso. Foi considerado disciplinas cursadas aquelas com menções: SR, II, MI, MM, MS ou SS. Pois, também tem a menção CC quando o estudante cursa com aprovação alguma disciplina em outra instituição e valida a mesma em seu currículo na UnB.

Para construir a variável Média de créditos p/ semestre, o banco de dados foi agrupado por aluno e pelo período que o aluno cursou a disciplina. O cálculo é definido por: $\bar{Y} = \frac{\sum_i^n \sum_j^m x_{ij}}{n}$, em que \bar{Y} é a média de créditos cursados por semestre e x é a quantidade de créditos da disciplina j do semestre i .

Para a construção da variável Soma de créditos reprovados somou a quantidade de créditos de cada disciplina que o aluno obteve menções: SR, II, MI, MM, MS ou SS. A variável Soma de créditos trancados foi construída ao somar a quantidade de créditos de cada disciplina que o aluno obteve menções: TJ (trancamento justificado) ou TR (trancamento). Segundo Chagas (2019), reprovações obtidas durante os dois primeiros anos exercem uma grande influência sobre a probabilidade de evasão para os recém ingressados. Com o intuito de avaliar essa reflexão no curso de LC, foi construída a variável Reprovou nos 2 primeiros anos. Caso o aluno tenha reprovado, a variável recebe resposta "Sim", caso contrário, recebe "Não".

Originalmente, a variável Forma de ingresso do aluno no curso de LC possui sete categorias, sendo que quatro delas não representavam 10% da amostra de alunos. Como mostra a Tabela 2. Portanto, decidiu-se agrupar as categorias: "Enem UnB", "portador de diploma de curso superior" e "transferência facultativa" e "obrigatória em uma única categoria chamada "outros". Logo, as categorias ficaram divididas em: "Vestibular", "Programa de avaliação seriada (PAS)", "Sistema de seleção unificada (SISU)" e "Outros".

3.3. Análise de Sobrevivência

A análise de sobrevivência consiste em um conjunto de técnicas estatísticas com o propósito de estudar dados em que a variável resposta é o tempo até a ocorrência de

Tabela 2. Distribuição de frequência da forma de ingresso dos alunos

Forma de ingresso	Percentual
Vestibular	43%
Programa de Avaliação Seriada	25%
Sisu-Sistema de Seleção Unificada	18%
Portador Diplom Curso Superior	8%
Enem UnB	4%
Transferência Obrigatória	1%
Transferência Facultativa	1%
Total	100%

um determinado fenômeno de interesse, denominado tempo de falha. No entanto, em algumas situações, o estudo terminou antes da ocorrência do evento de interesse. Nesses casos, a resposta é chamada de incompleta ou parcial, mas ainda assim importante para o estudo. Então, para essas situações, a variável resposta tempo é chamada de tempo censurado. Neste artigo, o fenômeno de estudo é a evasão do curso de LC da UnB. Logo, o objetivo de estudo é avaliar o tempo até a ocorrência da evasão no cursos de LC, e esse tempo será o tempo de falha. Para os estudantes que não evadiram (formaram ou estão cursando LC), os tempos associados serão tempos de censura. O que caracteriza dados de análise de sobrevivência. Por isso, neste estudo será utilizada essa técnica estatística [Lawless 2011] [Kalbfleisch and Prentice 2011].

O primeiro passo é saber qual distribuição de probabilidade pode ser usada para modelar o tempo até a evasão. Então, faz-se os ajustes das distribuições de probabilidade aos dados, estimando a probabilidade de sobreviver de forma paramétrica, e compara os ajustes com a probabilidade de sobreviver estimada pelo método empírico de Kaplan-Meier. A curva de sobrevivência estimada pelo modelo paramétrico que mais se aproximar da curva de sobrevivência estimada por Kaplan-Meier é a distribuição candidata [Kaplan and Meier 1958][Bland and Altman 1998]. Outra metodologia que auxilia na escolha da distribuição de probabilidade é calcular o critério de Akaike (AIC). A distribuição candidata é a distribuição que obtiver menor valor de AIC.

Para responder a pergunta de pesquisa "Quais fatores acadêmicos e sociais impactam na probabilidade de evadir do curso de LC da UnB?", é necessário estender a distribuição de probabilidade candidata para modelar os dados ao introduzir as variáveis explicativas, dessa forma tem-se um modelo de regressão em análise de sobrevivência. O vetor de variáveis explicativas, $x^T = (1, x_1, x_2, \dots, x_p)$ deve ser incluído ao reparametrizar um dos parâmetros da distribuição de probabilidade candidata. Então, um parâmetro θ pode ser reescrito pela relação $\theta = x^T \beta$, em que $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_p)^T$ são os parâmetros desconhecidos associados as variáveis explicativas. Originando um modelo de regressão paramétrico para os dados de evasão.

Para estimar os parâmetros desconhecidos do modelo, deve-se utilizar o método de máxima verossimilhança [Fonseca et al. 2005], pois esse método é capaz de incorporar os tempos de falhas e censuras do estudo. Um importante passo é a seleção das variáveis explicativas associadas ao modelo de regressão para os dados de evasão. Para realizar essa etapa o método de seleção de variáveis *Backward* é utilizado. Após o ajuste do modelo

de regressão é necessário verificar a qualidade global do ajuste por meio do resíduo de Cox-Snell [Cox and Snell 1968].

4. Resultados e Discussões

Para responder a questão de pesquisa é necessário primeiro deixar claro quais variáveis explicativas serão consideradas e na sequência, realizar um estudo descritivo dessas variáveis. Sendo assim, neste artigo, as variáveis explicativas consideradas são: idade, tipo de escola, modalidade de admissão, sistema de cotas, IRA, verão, currículo, média de créditos p/ semestre, soma de créditos reprovados, soma de créditos trancados e reprovou nos 2 primeiros anos. A variável resposta é composta pelo tempo até a ocorrência da evasão (tempo) e a variável indicadora (*status*) que descreve se o tempo é um falha ou censura.

Tabela 3. Estatísticas descritivas do tempo até a evasão

Variável	Estatísticas				
	Min	Quartil 1	Quartil 2	Quartil 3	Max
Tempo (todos)	1,00	3,00	5,00	7,00	16,00
Tempo falha	1,00	3,00	4,00	6,00	11,00
Tempo censura	2,00	3,00	6,00	8,00	16,00

Da Tabela 3 tem-se que 75% dos alunos que evadiram estavam matriculados no curso de LC até o sexto semestre e a evasão ocorre até o décimo primeiro semestre. Ou seja, a grande maioria da evasão ocorre até o sexto semestre. Ao considerar a estrutura curricular do curso de LC da UnB, o prazo mínimo para concluir o curso é de seis semestres, o prazo médio é de nove semestres e o prazo máximo é de dezesseis semestres. Da Tabela 3 é possível verificar que apenas 50% dos estudantes conseguem concluir o curso em até seis semestres.

Tabela 4. Estatísticas descritivas de Status, Tipo de Escola e Currículo

Variáveis	Estatísticas	
	Falha	Censura
Status	52%	48%
Tipo de Escola	Pública	Privada
	52%	48%
Currículo	Velho	Novo
	46%	54%

A Tabela 4 mostra que 52% dos estudantes evadiram do curso de LC, um pouco mais da metade. Esse resultado torna ainda mais importante entender quais fatores aumentam a probabilidade de evasão. Ainda da Tabela 4, observa-se que o percentual de estudantes que vieram de escola pública (52%) é próximo do percentual de estudantes que vieram de escola particular (48%). Também é possível observar que o percentual de estudantes que cursaram o currículo velho (46%) é próximo do percentual de estudantes que cursaram o currículo novo (54%).

A Tabela 5 informa que apenas 25% dos estudantes entraram via sistema de cotas, quantidade que é a metade do percentual estabelecido por lei. Também é baixo o percentual de estudantes que cursam disciplina no verão (24%). Em contrapartida, quase todos os alunos reprovaram (90%) alguma disciplina nos dois primeiros anos do curso.

Tabela 5. Sistema de Cotas, Verão e Reprovou nos 2 Primeiros Anos

Variável	Estatística	
	Sim	Não
Sistema de Cotas	25%	75%
Verão	24%	76%
Reprovou nos 2 Primeiros Anos	90%	10%

Tabela 6. Distribuição de frequência da forma de ingresso

Forma de ingresso	Porcentagem
Vestibular	43%
Programa de Avaliação Seriada	25%
Sisu-Sistema de Seleção Unificada	18%
Outros	14%
Total	100%

Da Tabela 6 nota-se que a maioria dos alunos ingressaram por meio do vestibular (43%), seguido por aqueles que entraram pelo Programa de Avaliação Seriada (PAS) com 25% dos alunos, 18% sendo do Sisu e 14% ingressaram de outra forma. A Tabela 7 apresenta as estatísticas descritivas das variáveis quantitativas. Ao analisar a variável idade, verifica-se que 50% dos alunos têm idade entre 16 e 21 anos. E que 50% dos alunos têm idade entre 21 e 60 anos, tendo a segunda metade maior variação do que a primeira metade. 50% dos estudantes possuem valor de IRA entre 2,58 e 5,00. Indicando um bom resultado, já que o valor 3,0 é um limite para os alunos competirem por diferentes tipos de bolsas na UnB.

Tabela 7. Estatísticas descritivas de Idade, IRA, Média de Créditos p/ Semestre, Soma de Créditos Reprovados, Soma de Créditos Trancados

Variável	Estatísticas				
	Min	Quartil 1	Quartil 2	Quartil 3	Max
Idade	16,00	19,00	21,00	27,00	60,00
IRA	0,00	1,45	2,58	3,29	5,00
Média de Créditos p/ Semestre	0,00	12,00	14,00	16,25	30,00
Soma de créditos reprovados	0,00	8,00	18,00	28,00	110,00
Soma de créditos trancados	0,00	0,00	4,00	10,00	60,00

Ao analisar a média de créditos por semestre, observa-se que 50% dos estudantes não cursam o mínimo de créditos indicados na estrutura curricular do curso de LC da UnB (mínimo = 14 créditos e máximo 27 créditos). Também é possível observar da Tabela 7 que 25% dos estudantes reprovaram até 8 créditos e que 75% dos estudantes reprovaram entre 8 e 110 créditos. 25% dos estudantes não utilizaram a opção de trancar alguma disciplina. Após a análise descritiva apresentada acima, a última etapa do estudo é a modelagem dos dados. Nessa etapa, o primeiro passo é saber qual distribuição de probabilidade melhor se ajusta aos dados: as distribuições log-logística, log-logística discreta e log-normal foram ajustadas aos dados. Com as estimativas dos parâmetros, estimou-se a função de sobrevivência de cada distribuição e comparou com o ajuste da função de sobrevivência estimada pelo método de Kaplan-Meier [Kaplan and Meier 1958]. Esse resultado, bem como os valores do critério de Akaike (AIC) estão na Figura 1.

Da Figura 1, é possível verificar que a distribuição log-normal, ao longo de todo tempo, acompanha mais o ajuste da função de sobrevivência estimada pelo método de Kaplan-Meier. Também verifica-se que o menor valor de AIC é da distribuição log-normal. Então, a distribuição log-normal é a distribuição que melhor se ajusta aos dados, dentre as distribuições analisadas. Conforme descrito na Seção 3.3, o próximo passo é estender a distribuição log-normal por meio da reparametrização do parâmetro $\mu = x^T \beta$ para construir um modelo de regressão log-normal com o objetivo de verificar quais fatores acadêmicos e sociais impactam na evasão do curso de LC. Nessa etapa da modela-

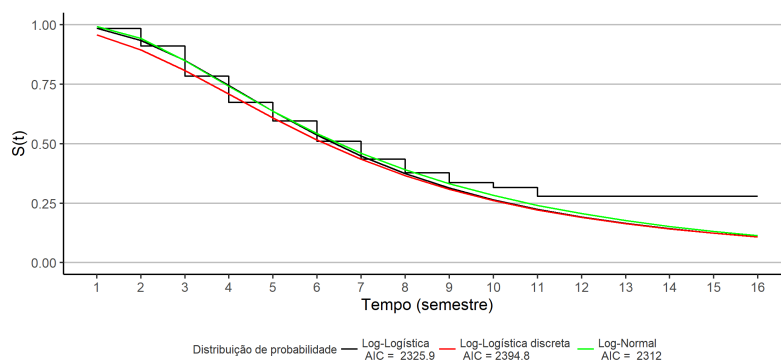


Figura 1. Estimação da função de sobrevivência e valor de AIC das distribuições log-logística, log-logística discreta e log-normal

gem, primeiro foi ajustado o modelo de regressão log-normal com apenas uma variável explicativa. O resultado foi analisado e então, algumas variáveis foram selecionadas para compor um modelo de regressão log-normal. Na sequência foi utilizado o método de seleção de variáveis *Backward* para selecionar as várias que realmente impactam significativamente na evasão. O modelo de regressão log-normal final para os dados de evasão do curso de LC estão na Tabela 8. Os resultados do modelo de regressão log-normal mostram que com o aumento do IRA, aumenta a probabilidade do aluno sobreviver a evasão, mantidas as demais variáveis constantes. Ou seja, alunos com maior IRA têm menor probabilidade de evadir do curso. Em relação a variável sistema de cotas, verifica-se pela

Tabela 8. Coeficientes estimados, erro padrão, estatística do teste e p-valor para o modelo de regressão log-normal final

Variável	Estimativa	Erro padrão	P-valor
β_0	-0,0675	0,1695	0,6905
β_{IRA}	0,3784	0,0180	;0,0001
$\beta_{Sistema\ de\ Cotas\ Sim}$	-0,0827	0,0443	0,0617
$\beta_{Verao\ Sim}$	0,1327	0,0562	0,0182
$\beta_{Curriculo\ Velho}$	-0,1436	0,0400	0,0003
$\beta_{Media\ de\ Creditos\ Semestre}$	0,0269	0,0126	0,0328
$\beta_{Soma\ de\ Creditos\ Reprovados}$	0,1193	0,0345	0,0005
$\beta_{Soma\ de\ Creditos\ Trancados}$	0,0167	0,0029	;0,0001
$\beta_{Reprovou\ nos\ 2\ Primeiros\ Anos\ Sim}$	0,8106	0,1769	;0,0001
$\beta_{Media\ de\ Creditos\ Semestre:Reprovou\ nos\ 2\ Primeiros\ Anos\ Sim}$	-0,0467	0,0137	0,0006
$\beta_{Soma\ de\ Creditos\ Reprovados:Reprovou\ nos\ 2\ Primeiros\ Anos\ Sim}$	-0,1025	0,0344	0,0029
log(scale)	-0,8720	0,0371	;0,0001

Tabela 8 que alunos cotistas têm maior probabilidade de evadir do que alunos não co-

tistas, mantidas as demais variáveis constantes. Também é possível observar que alunos que cursaram o verão têm menor probabilidade de evadir do curso do que alunos que não cursaram o verão, mantidas as demais variáveis constantes. E que alunos do currículo velho têm maior probabilidade de evadir do que alunos do atual currículo, mantidas as demais variáveis constantes. Ainda da Tabela 8, verifica-se que alunos com maior média de créditos têm menor probabilidade de evadir do curso, mantidas as demais variáveis constantes. E que alunos com maior soma de créditos reprovados têm menor probabilidade de evadir do curso, mantidas as demais variáveis constantes.

Já os alunos com maior soma de créditos trancados têm menor probabilidade de evadir do curso, mantidas as demais variáveis constantes. E alunos que reprovaram qualquer disciplina nos dois primeiros anos do curso têm menor probabilidade de evadir do curso do que alunos que não reprovaram nos dois primeiros anos, mantidas as demais variáveis constantes. Na análise também foi possível verificar o efeito de interação entre as variáveis média de créditos por semestre com reprovou nos 2 primeiros anos. E o efeito de interação entre soma de créditos reprovados com reprovou nos 2 primeiros anos.

O primeiro efeito de interação entre as variáveis explicativas indica que quanto maior a média de créditos por semestre do estudante que reprovou qualquer disciplina nos dois primeiros anos do curso, maior a probabilidade de evadir em relação aos alunos que não reprovaram nos dois primeiros anos, mantidas as demais variáveis constantes. O segundo efeito de interação entre as variáveis explicativas indica que quanto maior a soma de créditos reprovados do estudante que reprovou qualquer disciplina nos dois primeiros anos do curso, maior a probabilidade de evadir em relação aos alunos que não reprovaram nos dois primeiros anos, mantidas as demais variáveis constantes.

Conforme descrito na Seção 3.3, uma importante etapa da análise de dados é verificar a qualidade do modelo de regressão ajustado pelo resíduo de Cox-Snell. A Figura 2 mostra que a função de sobrevivência dos resíduos estimadas por Kaplan-Meier e pela distribuição exponencial padrão estão bem próximas. Isso significa que o modelo de regressão log-normal descrito na Tabela 8 é um modelo robusto para os dados, o modelo está bem ajustado aos dados. Portanto, todas interpretações são válidas.

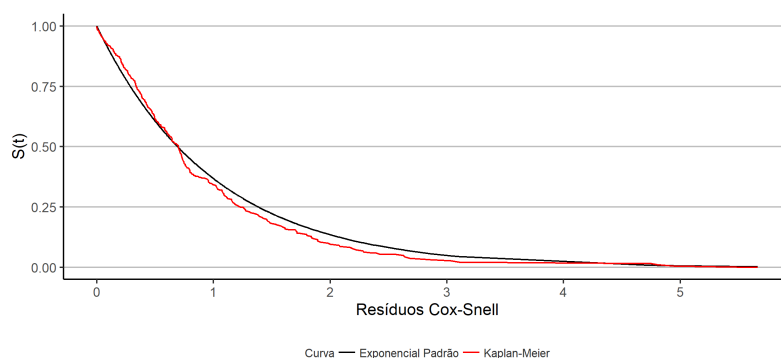


Figura 2. Resíduos de Cox-Snell do modelo final

5. Limitações

A pesquisa sobre evasão escolar é um assunto complexo que envolve dados acadêmicos e outros elementos externos, como dados educacionais antecedentes, diversidade, senso

de pertencimento e o ambiente universitário. Neste artigo, foram utilizados apenas dados acadêmicos disponibilizados pela UnB. Embora esses dados sejam importantes, incluir na análise dados de assistência estudantil, participação em projetos de extensão e de iniciação científica, enriqueceriam ainda mais os resultados e discussões encontradas.

6. Conclusões

A motivação inicial para este trabalho surgiu de um problema real, a evasão escolar na Licenciatura da Computação da UnB. Então, os autores querem responder: "Quais fatores acadêmicos e sociais impactam na probabilidade de evadir do curso de Licenciatura em Computação da Universidade de Brasília?". Para isso, foi ajustado o modelo de regressão log-normal aos dados.

Através do modelo é possível entender que os alunos com IRA alto têm maior probabilidade de sobreviver à evasão. Este resultado indica que o Departamento de Ciência da Computação deve propor mais apoio aos estudantes por meio de atividade de monitoria para as disciplinas e incentivar os alunos a frequentá-las. Com isso, os alunos podem tirar suas dúvidas e aumentar seu desempenho nas disciplinas. Consequentemente, aumentará o IRA. Outra importante estratégia é o Departamento de Ciência da Computação acompanhar todos os alunos com IRA inferior a 2,5 e propor estratégias para aumentar o IRA desses estudantes.

O resultado do modelo que alunos cotistas têm maior probabilidade de evadir do que alunos não cotistas, reforça a reflexão da necessidade de acompanhamento da Coordenação do curso de LC para alunos com IRA baixo. E outra estratégia que pode ser implementada tanto para aumentar IRA, quanto diminuir a evasão, é oferecer disciplinas estratégicas durante o Verão acadêmico e incentivar os alunos a cursarem uma disciplina no verão que o auxilie à avançar na grade curricular da LC. A conclusão que o currículo atual auxilia na redução da evasão é um importante resultado para o Departamento de Ciências da Computação. E reforça a necessidade da atualização dos currículos.

As demais interpretações obtidos neste artigo evidenciam a necessidade de acompanhando de perto da coordenação do curso de LC a alguns estudantes, em especial nos dois primeiros anos do curso. Pois, dependendo do desempenho acadêmico do estudante algumas estratégias, como trancar determinada disciplina, pode auxiliar ele a melhorar o rendimento acadêmico e com isso concluir o curso.

Com esses resultados, o Departamento de Ciência da Computação pode criar políticas para auxiliar os alunos a persistirem no curso de licenciatura. Com os resultados encontrados neste trabalho é fundamental dar continuidade nesta pesquisa. Modelos de regressão com fração de cura podem ser propostos para acomodar melhor os estudantes que ainda estão cursando a LC. Também é importante uma análise mais específica para homens e mulheres. E por fim, é essencial que mais informações acadêmicas sejam agregadas aos dados.

Referências

Bland, J. M. and Altman, D. G. (1998). Survival probabilities (the kaplan-meier method). *Bmj*, 317(7172):1572–1580.

- Calixto, C. (2015). Análise das causas de evasão discente no curso de licenciatura em computação: um estudo da ufpb virtual no formato uab. *Revista Tecnologias na Educação*, 7(12):1–13.
- Cox, D. R. and Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30(2):248–265.
- 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.
- Duran, R. a. (2024). Evasão nos cursos de graduação em computação no brasil. In *ACM Transactions on Computing Education*, pages 1–33. SBC.
- Fonseca, J. R., Friswell, M. I., Mottershead, J. E., and Lees, A. W. (2005). Uncertainty identification by the maximum likelihood method. *Journal of Sound and Vibration*, 288(3):587–599.
- Fritsch, R., da Rocha, C. S., and Vitelli, R. F. (2015). A evasão nos cursos de graduação em uma instituição de ensino superior privada. *Revista Educação em Questão*, 52(38):81–108.
- Kalbfleisch, J. D. and Prentice, R. L. (2011). *The statistical analysis of failure time data*. John Wiley & Sons.
- Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481.
- Lawless, J. F. (2011). *Statistical models and methods for lifetime data*. John Wiley & Sons.
- Prietch, S. S. and Pazeto, T. A. (2010). Estudo sobre a evasão em um curso de licenciatura em informática e considerações para melhorias. *WEIBASE, Maceió/AL*.
- 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:e50121–e50121.
- Souza, O., Morais, P., and Júnior, F. S. (2015). Um estudo sobre a evasão no curso de licenciatura em informática do ifrn–campus natal–zona norte. In *Anais do XXIII Workshop sobre Educação em Computação*, pages 216–225. SBC.
- Vasconcelos, V. and Andrade, E. (2018). Análise da evasão de alunos na licenciatura em computação. In *Anais do XXVI Workshop sobre Educação em Computação*, Porto Alegre, RS, Brasil. SBC.