



UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS I - CAMPINA GRANDE
CENTRO DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA
CURSO DE BACHARELADO EM ESTATÍSTICA

RIVALDO BEZERRA DE AQUINO FILHO

UM ESTUDO SOBRE A EVASÃO NO CURSO DE GESTÃO PÚBLICA
DA UFCG-CDSA VIA ANÁLISE DE SOBREVIVÊNCIA

CAMPINA GRANDE - PB

2024

RIVALDO BEZERRA DE AQUINO FILHO

**UM ESTUDO SOBRE A EVASÃO NO CURSO DE GESTÃO PÚBLICA
DA UFCG-CDSA VIA ANÁLISE DE SOBREVIVÊNCIA**

Trabalho de Conclusão de Curso apresentado ao Departamento de Estatística da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de Bacharel em Estatística.

Orientador: Profa. Dra. Ana Patricia Bastos Peixoto

Coorientador: Profa. Ms. Débora de Sousa Cordeiro

CAMPINA GRANDE - PB

2024

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

A657e Aquino Filho, Rivaldo Bezerra de.
Um estudo sobre a evasão no curso de gestão pública da UFCG-VDSA via análise de sobrevivência [manuscrito] / Rivaldo Bezerra de Aquino Filho. - 2024.
37 p. : il. colorido.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2024.

"Orientação : Profa. Dra. Ana Patricia Bastos Peixoto, Coordenação do Curso de Estatística - CCT. "

"Coorientação: Profa. Ma. Débora de Sousa Cordeiro , Departamento de Estatística - CCT. "

1. Evasão acadêmica. 2. Análise de sobrevivência. 3. Políticas públicas. I. Título

21. ed. CDD 519.5

RIVALDO BEZERRA DE AQUINO FILHO

UM ESTUDO SOBRE A EVASÃO NO CURSO DE GESTÃO PÚBLICA
DA UFCG-CDSA VIA ANÁLISE DE SOBREVIVÊNCIA

Trabalho de Conclusão de Curso apresentado ao Departamento de Estatística da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de Bacharel em Estatística.

Aprovado em: 02/07/2024.

BANCA EXAMINADORA

Ana Patricia Bastos Peixoto

Profa. Dra. Ana Patricia Bastos Peixoto
Universidade Estadual da Paraíba (UEPB)

Cleanderson Romualdo Fidelis

Prof. Ms. Cleanderson Romualdo Fidelis
Universidade Federal do Acre (UFAC)

Lucas Cardoso Pereira

Prof. Ms Lucas Cardoso Pereira
Universidade Estadual da Paraíba (UEPB)

A minha avó, Gentila, pelo exemplo de vida e acolhida.

AGRADECIMENTOS

Primeiramente agradeço a Deus por saúde e força na minha jornada acadêmica.

A minha esposa, Jaqueline Matias, por todo amor, companheirismo e incentivo e paciência para concluir esse trabalho.

A minha mãe Maria Goretti, meu irmão Raphael Pereira e minha irmã Liliann Rose, pelo carinho e compreensão.

A todos os professores do curso de Estatística , que ajudaram na minha formação, em especial aos professores Cleanderson Fidelis, Pedro Monteiro e Maria Vitória que me aturaram em diversas disciplinas.

Aos colegas de curso, principalmente a João Vitor, Hillnara Ferreira e Jonas Gomes pelo coleguismo e ajuda nos estudos.

A minha orientadora, professora Ana Patricia Bastos Peixoto, pela paciência, contribuições e fé no neste trabalho.

A minha coorientadora, professora Débora de Sousa Cordeiro, por todas contribuições no trabalho.

Agradeço à Universidade Estadual da Paraíba pela oferta do curso de Estatística.

“Às vezes as questões são complicadas e as respostas são simples..”
(Theodor Seuss Geisel)

RESUMO

O trabalho aborda a importância da avaliação na gestão pública, enfatizando seu papel crucial na medição e aprimoramento do desempenho das políticas implementadas, com foco na evasão de alunos do curso de Tecnologia em Gestão Pública da Universidade Federal de Campina Grande. O principal objetivo é compreender o fenômeno da evasão para estabelecer metas que ajudem a minimizá-lo. A pesquisa utiliza métodos quantitativos, com ênfase na Análise de Sobrevivência, um campo de estudo da Estatística que investiga o tempo até a ocorrência de um evento, neste caso, a evasão. Como resultado, observou-se que a maior taxa de evasão ocorre do primeiro para o segundo período e do oitavo para o nono período, com probabilidades de evasão de 25,8% e 17,4%, respectivamente. Outro achado é que os fatores sexo e idade não apresentaram diferenças significativas, mas o tipo de ingresso por cota mostrou-se relevante, com o grupo de alunos cotistas apresentando maior risco de evasão. Além disso, o tempo mediano para evasão foi identificado no sétimo período, um período após o tempo mínimo esperado para a conclusão do curso.

Palavras-chave: evasão acadêmica; análise de sobrevivência; políticas públicas.

ABSTRACT

This study addresses the importance of evaluation in public management, emphasizing its crucial role in measuring and improving the performance of implemented policies, with a focus on student dropout rates in the Technology in Public Management program at the Federal University of Campina Grande. The main objective is to understand the phenomenon of dropout rates in order to establish goals that help minimize it. The research employs quantitative methods, with a focus on Survival Analysis, a statistical technique that investigates the time until the occurrence of an event, in this case, dropout. As a result, it was observed that the highest dropout rate occurs from the first for the second period and from the eighth to the ninth period, with probabilities of evasion 25.8% and 17.4%, respectively. Another significant finding is that factors such as gender and age did not show significant differences, but the type of admission through quotas proved to be relevant, with quota students showing a higher risk of dropout. Additionally, the median time for evasion was identified in the seventh period, a period following the minimum expected time for course completion.

Keywords: academic evasion; survival analysis; public policy.

LISTA DE ILUSTRAÇÕES

Figura 1 – Evasão por período	26
Figura 2 – Graduados por período	26
Figura 3 – Curva de Sobrevivência da evasão dos alunos.	27
Figura 4 – Curva de Sobrevivência em relação ao gênero	29
Figura 5 – Curva de Sobrevivência de acordo com a Cota	30
Figura 6 – Curva de Sobrevivência em relação a idade.	31
Figura 7 – Comparativo das Curvas de Sobrevivência	32
Figura 8 – Linearização das Curvas de Sobrevivência	33

LISTA DE TABELAS

Tabela 1 – Teste de <i>logrank</i>	17
Tabela 2 – Tabela de matrícula por Sexo	25
Tabela 3 – Tabela de matrícula por Cota	25
Tabela 4 – Tabela de matrícula por Idade	25
Tabela 5 – Estimativa da Taxa de Falha nos Intervalos	28
Tabela 6 – Estimativa de Falha nos Intervalos	28
Tabela 7 – Estimativas para Função de Sobrevivência	32

SUMÁRIO

1	INTRODUÇÃO	11
2	FUNDAMENTAÇÃO TEÓRICA	13
2.1	Métodos Quantitativos	13
2.2	Análise de Sobrevivência	13
2.2.1	<i>Conceitos Básicos</i>	<i>13</i>
2.3	Técnicas Não-Paramétricas	15
2.3.1	<i>Estimador de Kaplan-Meier</i>	<i>15</i>
2.3.2	<i>Teste de logrank</i>	<i>16</i>
2.4	Técnicas Paramétricas	17
2.4.1	<i>Modelo Exponencial</i>	<i>18</i>
2.4.2	<i>Modelo Weibull</i>	<i>18</i>
2.4.3	<i>Modelo Log-normal</i>	<i>19</i>
2.4.4	<i>Modelo Gamma Generalizado</i>	<i>20</i>
2.4.5	<i>Método da Máxima Verosimilhança</i>	<i>20</i>
2.4.6	<i>Escolha do Modelo Probabilístico</i>	<i>21</i>
2.4.6.1	<i>Métodos Gráficos</i>	<i>21</i>
2.4.6.2	<i>Teste da Razão de Verosimilhanças</i>	<i>22</i>
3	MATERIAIS E MÉTODOS	23
4	RESULTADOS E DISCUSSÃO	25
5	CONCLUSÕES	35
	REFERÊNCIAS	36

1 INTRODUÇÃO

A avaliação na gestão pública possui um papel importante, pois permite medir e avaliar o desempenho das políticas implementadas, ou seja, verificar se os objetivos estabelecidos foram alcançados e identificar possíveis áreas de melhoria. Isto auxilia na tomada de decisões embasadas por evidências, na responsabilização e transparência, na melhoria do uso de recursos e na busca pela melhoria contínua e aprendizado organizacional. Segundo Silveira (2018), o principal objetivo da administração pública é prestar serviços à sociedade. Para alcançar esse objetivo, é necessário buscar uma maior eficiência das instituições e órgãos públicos e uma das formas de alcançar isso é através de uma avaliação detalhada, cuidadosa e minuciosa dos resultados obtidos.

Alguns estudos, como os de Costa, Bispo e Pereira (2018), Júnior et al. (2022) e Lopes et al. (2023), indicam diversos fatores que podem levar à evasão, entre eles a idade, sexo e tipo de admissão. Considerando esses fatores, é crucial observá-los atentamente para desenvolver políticas de gestão mais assertivas e eficazes na redução desses riscos. Analisar detalhadamente os motivos que levam à evasão permite implementar estratégias direcionadas e específicas, diminuindo as chances de evasão dos alunos.

O estudo de evasão, que se refere à desvinculação por reprovação, ao abandono ou à desistência dos estudos por parte dos alunos antes da conclusão do curso, é importante para identificar possíveis falhas no sistema educacional, planejar programas de suporte estudantil, evitar o desperdício de recursos financeiros e, principalmente, pensar no retorno social de incluir no mercado profissionais qualificados. Para Coimbra, Silva e Costa (2021), a evasão é uma das principais preocupações do Ministério da Educação, visto como um alvo a ser combatido ou um índice a ser reduzido, sobretudo por representar, de alguma forma, o fracasso institucional.

Um campo da Estatística que pode ser utilizado para a análise da evasão é a Análise de Sobrevivência, muito utilizada em diversas áreas, como na saúde, engenharia, economia, entre outras. De acordo com Colosimo e Giolo (2006), a Análise de Sobrevivência consiste em um conjunto de métodos estatísticos que visa estudar o tempo até a ocorrência de um evento de interesse, denominando esse tempo como tempo de falha. Neste caso, o evento de interesse é a evasão, e o tempo de falha será o tempo (em semestres) até a evasão do aluno. Dessa forma, utilizaremos esse conjunto de métodos para analisar o tempo até a evasão dos alunos no curso Superior de Tecnologia em Gestão Pública da UFCG.

Essa abordagem permite uma compreensão mais profunda dos fatores que influenciam a permanência e a desistência dos alunos, contribuindo para a formulação de políticas mais eficazes e direcionadas para diminuir a taxa de evasão e aumentar a taxa de sucesso acadêmico dos estudantes.

Neste trabalho, pretende-se estudar a evasão dos alunos do curso de Tecnologia em Gestão Pública da Universidade Federal de Campina Grande (UFCG). Segundo Ferreira et al. (2017), o campus de Sumé surgiu como resultado do programa REUNI (Reestruturação e

Expansão das Universidades Federais), e o curso de Gestão Pública foi um dos selecionados para o campus com o objetivo de promover o desenvolvimento regional. Assim, a escolha deste curso para o estudo proporciona uma visão sobre os impactos e resultados dessa política de expansão educacional.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo tem como objetivo principal apresentar os principais conceitos que fundamentam esta pesquisa. Nele, será apresentado, os conceitos de métodos quantitativos, Análise de Sobrevivência e as técnicas não-paramétricas e paramétricas para estimar a função de sobrevivência.

2.1 Métodos Quantitativos

Métodos quantitativos são abordagens de pesquisa que utilizam dados numéricos e técnicas estatísticas rigorosas para descrever, explicar, prever e controlar fenômenos. Eles envolvem a medição objetiva de variáveis e a utilização de instrumentos padronizados, como questionários e testes, permitindo a obtenção de resultados generalizáveis e replicáveis. Essas características fornecem uma base sólida para a tomada de decisões informadas e a formulação de políticas (GIL, 2019).

Segundo Gil (2019), a utilização de métodos quantitativos ajuda na obtenção de dados precisos e relevantes, permitindo a medição objetiva de variáveis, análises estatísticas e generalização dos resultados. Na gestão pública, esses métodos são especialmente importantes para identificar padrões, avaliar políticas e otimizar recursos.

2.2 Análise de Sobrevivência

De acordo com Sousa-Ferreira (2022), a Análise de Sobrevivência é um campo antigo da Estatística, emergindo no século XVII para abordar desafios nas Ciências Biomédicas. Seu objetivo principal é analisar o tempo de vida, desde um ponto inicial até a ocorrência de um evento específico, como recidiva de doença, falha de equipamento, casamento, conclusão de estudos ou dissolução de uma empresa. Isso moldou sua terminologia e aplicação em diversas áreas, incluindo a área de avaliação de políticas públicas.

2.2.1 Conceitos Básicos

Para a apresentação dos conceitos básicos e técnicas não-paramétricas, foram utilizados como base os conceitos de Colosimo e Giolo (2006).

Dois conceitos básicos no estudo de Análise de Sobrevivência são o de tempo de falha e censura. O tempo de falha é definido como o tempo até a ocorrência de um evento de interesse. A censura, por sua vez, ocorre quando não se tem a informação completa sobre o tempo de falha para todas as observações do estudo. As censuras podem ser de 3 tipos:

- i) **Tipo I:** ocorre quando o estudo tem o tempo de duração pré-estabelecido;
- ii) **Tipo II:** ocorre quando o estudo termina após uma certa determinada quantidade de falhas ocorrerem;

- iii) **Aleatória:** ocorre quando o indivíduo sai do estudo por motivos que não têm ligação com o evento de interesse.

Considerando o tempo de falha, pode-se representar os dados de sobrevivência da seguinte forma: Seja T uma variável aleatória, não negativa, representando o tempo de falha e t_1, \dots, t_n uma amostra aleatória de T . Os dados de sobrevivência são representados pelo par (t_i, δ_i) , em que t_i representa o tempo de falha do indivíduo i e δ_i representa a variável indicadora de falha, ou seja,

$$\delta_i = \begin{cases} 1, & \text{se ocorreu falha} \\ 0, & \text{caso contrário,} \end{cases} \quad (2.1)$$

com $i = 1, \dots, n$.

Após a definição dos conceitos previamente mencionados, serão abordadas algumas funções básicas e outros conceitos relevantes, no que se segue. Desse modo, A função de sobrevivência é definida como a probabilidade de uma observação não falhar até um certo tempo t , ou seja, a probabilidade de uma observação sobreviver ao tempo t . Em termos probabilísticos, isto é escrito como

$$S(t) = P(T \geq t).$$

Podemos obter a função de distribuição acumulada a partir da função de sobrevivência, uma vez que

$$\begin{aligned} P(T \leq t) &= 1 - P(T > t) \\ &= 1 - S(t). \end{aligned}$$

Função taxa de falha ou função de risco

A probabilidade da falha ocorrer em um intervalo de tempo $[t, t + \Delta t]$ é dada por:

$$\begin{aligned} P(t < T < t + \Delta t) &= P(T \leq t + \Delta t) - P(T \leq t) \\ &= F(t + \Delta t) - F(t) \\ &= 1 - S(t + \Delta t) - 1 + S(t) \\ &= S(t) - S(t + \Delta t). \end{aligned} \quad (2.2)$$

A taxa de falha no intervalo $[t, t + \Delta t)$ é definida como a probabilidade da falha ocorrer neste intervalo, dado que não ocorreu antes de t dividida pelo comprimento do intervalo. Ou seja,

$$\frac{S(t) - S(t + \Delta t)}{\Delta t S(t)}.$$

Segue então que, a taxa de falha instantânea, também denominada de função de risco ou taxa de falha de T , é definida por:

$$\begin{aligned} \lambda(t) &= \lim_{\Delta t \rightarrow 0} \frac{S(t) - S(t + \Delta t)}{\Delta t S(t)} \\ &= \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}. \end{aligned} \quad (2.3)$$

Observe que, com a função taxa de falha podemos calcular o risco de um indivíduo sofrer um evento de interesse em um certo tempo, sabendo que ele não sofreu o evento até aquele momento.

Tempo Médio

O tempo médio de vida é obtido pela área sob a curva da função de sobrevivência dada por:

$$t_m = E(t) = \int_0^{\infty} S(t) dt. \quad (2.4)$$

Tempo Mediano

O tempo mediano, também conhecido como tempo mediano de sobrevivência, representa o tempo em que metade da população do estudo já experimentou o evento de interesse e a outra metade ainda não, ou seja, é o tempo t tal que a probabilidade de sobrevivência é 50%. Matematicamente, é definido como o tempo t para o qual a função de sobrevivência $S(t)$ é igual a 0,5.

2.3 Técnicas Não-Paramétricas

A Análise de Sobrevivência utiliza diversas técnicas não-paramétricas para estimar a função de sobrevivência, sendo o estimador de Kaplan-Meier e o estimador de Nelson-Aalen os mais conhecidos. Essas técnicas são especialmente úteis quando as características da população em estudo não são completamente conhecidas ou quando os dados apresentam censura.

Neste trabalho, o estimador de Kaplan-Meier é especialmente utilizado, sendo considerado, segundo Lima (2018), o modelo mais utilizado e simples para a estimação da função de sobrevivência. Além disso, é observado que as estimativas de Kaplan-Meier são geralmente muito próximas às do estimador de Nelson-Aalen, sendo esta última tendendo a ser sempre maiores ou iguais a de Kaplan-Meier.

2.3.1 Estimador de Kaplan-Meier

Segundo Colosimo e Giolo (2006), o estimador de Kaplan-Meier, também conhecido por estimador limite-produto, é uma adaptação da função de sobrevivência empírica, quando não existem censuras, isto é

$$\hat{S}(t) = \frac{\text{n}^\circ \text{ de observações que não falharam até o tempo } t}{\text{n}^\circ \text{ total de observações no estudo}}.$$

$\hat{S}(t)$ é uma função escada com degraus nos tempos observados de falha de tamanho $1/n$, em que n é o tamanho da amostra. Se existirem empates em um certo tempo t , o tamanho do degrau fica multiplicado pelo número de empates.

O estimador de Kaplan-Meier, na sua construção, considera tantos intervalos de tempo quantos forem o número de falhas distintas. Os limites dos intervalos de tempo são os tempos de falha da amostra.

Nesse sentido, o estimador de Kaplan-Meier é definido como sendo:

$$\hat{S}(t) = \prod_{i:t_{(i)} \leq t} \left(\frac{n_i - d_i}{n_i} \right) = \prod_{i:t_{(i)} \leq t} \left(1 - \frac{d_i}{n_i} \right), \quad (2.5)$$

em que, n_i : número de indivíduos que podem falhar (sob risco) no tempo t_i , isto é, indivíduos que não apresentaram o evento de interesse e também não foram censurados até o instante imediatamente anterior a t_i ; d_i : número de indivíduos que falharam no tempo t_i ; $t_1 < t_2 < \dots < t_k$, são k tempos distintos e ordenados de falhas.

2.3.2 Teste de logrank

Na pesquisa de Lopes et al. (2023), foi realizada uma revisão da literatura onde foram identificados alguns fatores que podem influenciar na evasão. Com base nisso, três fatores foram adotados nesta pesquisa para levantamento de hipóteses: sexo, admissão por sistema de cotas e idade.

Além das estimativas de sobrevivência, a comparação entre duas ou mais funções de sobrevivência também se beneficia de técnicas não paramétricas. Testes como o *logrank* e o *Wilcoxon* são empregados para avaliar diferenças estatisticamente significativas entre grupos populacionais ou para determinar o impacto de diferentes intervenções. Nesta pesquisa, o teste de *logrank* foi utilizado para realizar essas comparações.

Para comparação de curvas de sobrevivência o teste de *logrank* é um dos mais utilizados, segundo Serranho e Ramos (2017). Ele ainda afirma, que isso se deve pois ele permite testar igualmente em todo o intervalo, ou seja, ele testa a hipótese nula de que não há diferença entre os grupos em termos do tempo até a ocorrência de um evento (como a evasão de alunos) ao longo de todo o intervalo de tempo estudado. As definições de Colosimo e Giolo (2006) serão empregadas para descrever este teste.

O teste *logrank* é adequado quando as taxas de falhas dos grupos são aproximadamente constantes, ou seja, quando as populações possuem taxas de falhas proporcionais. Segundo Chiesa, Moraes e Rigão (2002), a estatística do teste é baseada na diferença entre o número de falhas observadas em cada grupo e o número esperado de falhas sob a hipótese nula.

Considere inicialmente a comparação das duas funções de sobrevivência $S_1(t)$ e $S_2(t)$. Sejam $t_1 < t_2 < \dots < t_k$ os tempos de falhas distintos combinados das duas amostras individuais. Suponha que, exatamente no tempo t_j , ocorrem d_j falhas e que n_j indivíduos estejam sob risco no tempo imediatamente inferior a t_j na amostra combinada. Denote d_{ij} como o número de falhas da amostra i e n_{ij} indivíduos da amostra i , com $i = 1, 2$ e $j = 1, \dots, k$. Em cada tempo de falha t_j , os dados estarão dispostos em tabelas de formato 2×2 com d_{ij} falhas e $n_{ij} - d_{ij}$ sobreviventes na coluna i .

Tabela 1 – Teste de *logrank*.

	Grupo 1	Grupo 2	Total
Falha	d_{1j}	d_{2j}	d_j
Não Falha	$n_{1j} - d_{1j}$	$n_{2j} - d_{2j}$	$n_{2j} - d_{2j}$
Total	n_{1j}	n_{2j}	n_j

Fonte: (Colosimo e Giolo, 2006)

Uma explicação simplificada é que, ao condicionar a experiência de falha e censura até o tempo t_j (fixando as marginais das colunas) e o número de falhas no tempo t_j (fixando as marginais da linha), a distribuição de d_{2j} é Hipergeométrica

$$\frac{\binom{n_{1j}}{d_{1j}} \binom{n_{2j}}{d_{2j}}}{\binom{n_j}{d_j}}. \quad (2.6)$$

Desta forma, a média e a variância de d_{2j} são dadas, respectivamente, por

$$w_{2j} = n_{2j}d_j n_j^{-1}$$

$$(V_j)_2 = n_{2j}(n_j - n_{2j})d_j(n_j - d_j)n_j^{-2}(n_j - 1)^{-1}.$$

A estatística $d_{2j} - w_{2j}$ tem média zero e variância $(V_j)_2$. E, dessa forma, se as k tabelas de contingência forem independentes, um teste aproximado para verificar a igualdade das curvas de sobrevivência pode ser baseado na estatística:

$$T = \frac{[\sum_{j=1}^k (d_{2j} - w_{2j})]^2}{\sum_{j=1}^k (V_j)_2},$$

que sob a hipótese nula $H_0 : S_1(t) = S_2(t)$, a estatística de teste, segundo Colosimo e Giolo (2006), segue uma distribuição Qui-quadrado com um grau de liberdade para grandes amostras e para todo t no período de acompanhamento.

2.4 Técnicas Paramétricas

Diferente dos modelos não-paramétricos, que não assumem nenhuma forma específica para a distribuição dos tempos de sobrevivência, os modelos paramétricos partem do princípio de que os tempos de sobrevivência seguem uma distribuição estatística bem definida. Isso significa que, antes de ajustar o modelo, o pesquisador deve escolher uma distribuição teórica (como Exponencial, Weibull, Log-Normal ou Gamma) que ele acredita que descreva adequadamente os dados de sobrevivência.

Segundo Wey, Connett e Rudser (2015), a utilização combinada de modelos paramétricos e não-paramétricos permite uma avaliação mais robusta dos dados de sobrevivência. Esta abordagem ajuda a garantir que, se os resultados forem consistentes entre os dois tipos de modelos,

a confiança nas conclusões é fortalecida. A robustez das conclusões é aumentada porque os modelos não-paramétricos, que não fazem suposições rígidas sobre a forma da distribuição dos dados, podem complementar os modelos paramétricos, que assumem distribuições específicas, proporcionando uma validação cruzada das inferências

Neste trabalho, serão apresentados quatro modelos paramétricos para o ajuste de dados de sobrevivência: Exponencial, Weibull, Log-Normal e Gama Generalizado. Também será discutido o Teste da Razão de Verossimilhança, utilizado para determinar o modelo mais adequado aos dados de sobrevivência. Para a apresentação dos modelos paramétricos foram utilizados os conceitos de Serranho (2015).

2.4.1 Modelo Exponencial

O modelo exponencial é baseado na distribuição exponencial para variável T , isto é, $T \sim Exp(1/\lambda)$.

Em que o parâmetro λ corresponde à taxa de falha (constante) do modelo exponencial. Pelas propriedades da distribuição exponencial, temos a sua função densidade de probabilidade na forma

$$f(t) = \frac{1}{\lambda} e^{-\frac{t}{\lambda}}, \quad t \geq 0, \quad (2.7)$$

a função de distribuição de probabilidade,

$$F(t) = 1 - \frac{1}{\lambda} e^{-\frac{t}{\lambda}}, \quad t \geq 0, \quad (2.8)$$

e o valor esperado e a variância, dados por

$$E(t) = \lambda \quad \text{e} \quad Var(t) = \lambda^2. \quad (2.9)$$

Pela relação entre a função de sobrevivência e a função distribuição de probabilidade de T , temos que

$$S(t) = 1 - F(t) = e^{-\frac{t}{\lambda}}, \quad t \geq 0. \quad (2.10)$$

Neste sentido, a função taxa de falha é constante e dada por

$$\lambda(t) = \frac{1}{\lambda}, \quad t \geq 0. \quad (2.11)$$

2.4.2 Modelo Weibull

O modelo de Weibull é uma generalização do modelo exponencial com a introdução de mais um parâmetro, isto é a variável T tem distribuição $T \sim Wei(\alpha, \gamma)$, em que o parâmetro $\gamma > 0$ se chama o parâmetro de forma e $\alpha > 0$ é o parâmetro de escala. A função densidade de probabilidade é dada por

$$f(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1} e^{-\left(\frac{t}{\alpha}\right)^\gamma}, \quad t \geq 0. \quad (2.12)$$

Para $\alpha = \frac{1}{\lambda}$ e $\gamma = 1$ a expressão anterior se reduz à função densidade de probabilidade apresentada em (2.7). Este aspeto fica ainda mais evidente ao considerarmos a função distribuição de probabilidade da Weibull dada por

$$F(t) = 1 - e^{-(\frac{t}{\alpha})^\gamma}, \quad t \geq 0. \quad (2.13)$$

Pela relação entre a função de sobrevivência e a função distribuição de probabilidade de T , temos que, para o modelo Weibull a função de sobrevivência é dada por

$$S(t) = e^{-(\frac{t}{\alpha})^\gamma}, \quad t \geq 0, \quad (2.14)$$

e, portanto, a função taxa de falha é dada por

$$\lambda(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1}, \quad t \geq 0. \quad (2.15)$$

Para a distribuição Weibull, o valor esperado e variância são dados por:

$$E(t) = \alpha \left(1 + \frac{1}{\gamma}\right) \quad \text{e} \quad \text{Var}(t) = \alpha^2 \left\{ \Gamma\left(1 + \frac{2}{\gamma}\right) - \Gamma\left(1 + \frac{1}{\gamma}\right)^2 \right\}, \quad (2.16)$$

em que a função Gama Γ é definida por

$$\Gamma = \int_0^\infty u^{x-1} e^{-u} du. \quad (2.17)$$

2.4.3 Modelo Log-normal

No modelo Log-normal a variável tempo de sobrevivência T é a distribuição Log-normal $T \sim \text{LogNorm}(\mu, \sigma)$, com parâmetros μ e σ que são, respetivamente, a média e desvio padrão da variável aleatória $X = \ln(T)$. Assim, a sua função densidade de probabilidade é dada por

$$f(t) = \frac{1}{\sqrt{2\pi t} \sigma} e^{-\frac{(\ln(t)-\mu)^2}{2\sigma^2}}, \quad t > 0, \quad (2.18)$$

a função distribuição de probabilidade da forma

$$F(t) = \Phi\left(\frac{\ln(t) - \mu}{\sigma}\right), \quad t > 0, \quad (2.19)$$

em que Φ é a função de distribuição da normal padrão. Assim, temos a função de sobrevivência

$$S(t) = \Phi\left(\frac{\mu - \ln(t)}{\sigma}\right). \quad (2.20)$$

E a função taxa de falha dada pela relação

$$\lambda(t) = \frac{f(t)}{S(t)}. \quad (2.21)$$

O valor esperado e a variância são dados respetivamente por

$$E(T) = e^{\mu + \frac{\sigma^2}{2}} \quad \text{e} \quad \text{Var}(T) = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1). \quad (2.22)$$

2.4.4 Modelo Gamma Generalizado

A distribuição Gama generalizada é também utilizada em várias aplicações de Análise de Sobrevivência, uma vez que esta se adapta a vários comportamentos distintos consoante a escolha dos parâmetros da distribuição, incluindo a distribuição Exponencial, Weibull e Gama. Diz-se que a variável T que denota o tempo de sobrevivência tem uma distribuição Gama generalizada $T \sim GamaGen(\gamma, \kappa, \alpha)$, se a função densidade de probabilidade associada for da forma

$$f(t) = \frac{\gamma}{\Gamma(\kappa)\alpha^{\gamma\kappa}} t^{\gamma\kappa-1} e^{-\left(\frac{t}{\alpha}\right)^\gamma}, \quad t > 0, \quad (2.23)$$

em que Γ é a função Gama. A função distribuição de probabilidade de uma variável Gama generalizada é dada por

$$F(t) = \frac{g(\kappa, \left(\frac{t}{\alpha}\right)^\gamma)}{\Gamma(\kappa)}, \quad (2.24)$$

em que g é a função gama incompleta inferior, dada por

$$g(s, x) = \int_0^x t^{s-1} e^{-t} dt. \quad (2.25)$$

O valor esperado e a variância são dados respetivamente por

$$E(T) = \frac{\alpha\Gamma\left(\frac{\gamma\kappa+1}{\kappa}\right)}{\Gamma(\kappa)} \quad e \quad Var(T) = \frac{\alpha^2}{\Gamma(\kappa)} \left\{ \Gamma\left(\frac{\gamma\kappa+2}{\gamma}\right) - \frac{\left[\Gamma\left(\frac{\gamma\kappa+1}{\gamma}\right)\right]^2}{\Gamma(\kappa)} \right\}. \quad (2.26)$$

Vale notar que a distribuição Gama generalizada abarca várias das distribuições já abordadas, no qual

- se $\gamma = \kappa = 1$ então $T \sim Exp(1/\alpha)$;
- se $\kappa = 1$ então $T \sim Wei(\gamma, \alpha)$;
- se $\kappa \rightarrow \infty$ então a distribuição Gama generalizada tende para uma Log-normal;

2.4.5 Método da Máxima Verossimilhança

De acordo com Serranho (2015), após a escolha do modelo paramétrico, a estimativa dos parâmetros adequados ao conjunto de dados coletados é necessária. Indica-se que o método mais utilizado para essa finalidade é o método da máxima verossimilhança. Desta forma, o método será apresentado, fundamentando-se nos conceitos de Serranho (2015).

Caso não haja censuras, o método da máxima verossimilhança passa simplesmente por considerar a função de verossimilhança (sem censuras)

$$\mathcal{L}(t_i, \theta) = \prod_{i=1}^n f(t_i, \theta)$$

em que os valores $t_i, i = 1, 2, \dots, n$ são os tempos de falha recolhidos e os valores $\theta = (\theta_1, \theta_2, \dots, \theta_m)$ são os parâmetros a estimar. Dessa forma, tenta-se escolher os valores dos parâmetros θ que

maximizem a função anterior, uma vez que maximizando o produto dos valores $f(t_i, \theta)$ o parâmetro θ maximiza a probabilidade dos tempos t_i recolhidos serem os obtidos por amostragem aleatória da variável com parâmetro θ . Assim, o procedimento passa por considerar o logaritmo da função anterior

$$\mathcal{M}(t_i, \theta) := \ln(\mathcal{L}(t_i, \theta)) = \sum_{i=1}^n \ln(f(t_i, \theta))$$

que transforma o produtório num somatório, sem alterar o argumento máximo da função, facilitando os cálculos das derivadas seguintes. De seguida, há que determinar quais os pontos em que a primeira derivada em θ dessa função \mathcal{M} se anula e verificar se a segunda derivada é negativa nesses pontos, por forma a garantir que estamos na presença de um máximo da função de sobrevivência.

No caso de existirem censuras, temos de introduzir essa informação na função de verosimilhança. Se no caso de falha no instante t_i , queremos maximizar a probabilidade desta ocorrer em t_i e logo queremos maximizar $f(t_i, \theta)$ em função de θ para os t_i recolhido, no caso de uma censura em t_i , queremos maximizar a função de sobrevivência em t_i por forma a traduzir a informação de que naquele instante ainda temos informação de existir um sobrevivente (desconhecendo o instante posterior em que este falha). Assim, o método da máxima verosimilhança passa por considerar a função de verosimilhança (com censuras)

$$\mathcal{L}(t_i, \theta) = \prod_{i=1}^n [f(t_i, \theta)]^{\delta_i} \cdot [S(t_i, \theta)]^{1-\delta_i} \quad (2.27)$$

em que δ_i assume os valores 1 ou 0 consoante o instante t_i é uma falha ou uma censura, respectivamente.

Da relação $\lambda(t) = f(t)/S(t)$ podemos reescrever a expressão anterior como

$$\mathcal{L}(t_i, \theta) = \prod_{i=1}^n [\lambda(t_i, \theta)]^{\delta_i} \cdot S(t_i, \theta). \quad (2.28)$$

pelo que a função logaritmo da função de verosimilhança é dada por

$$\mathcal{M}(t_i, \theta) = \ln(\mathcal{L}(t_i, \theta)) = \sum_{i=1}^n [\delta_i \ln(\lambda(t_i, \theta)) + \ln(S(t_i, \theta))]. \quad (2.29)$$

2.4.6 Escolha do Modelo Probabilístico

Para Colosimo e Giolo (2006), a seleção do modelo adequado é fundamental na análise paramétrica de dados de tempo de vida. Assim, para a escolha dos modelos, serão apresentados dois métodos utilizados por Colosimo e Giolo (2006): os métodos gráficos e o teste da razão de verossimilhança.

2.4.6.1 Métodos Gráficos

O primeiro método gráfico a ser apresentado envolve a comparação da função de sobrevivência do modelo proposto com o estimador de Kaplan-Meier. Neste procedimento, ajustam-se

os modelos propostos ao conjunto de dados e, a partir das estimativas dos parâmetros de cada modelo, estimam-se suas respectivas funções de sobrevivência. Para o conjunto de dados, obtém-se também a estimativa de Kaplan-Meier para a função de sobrevivência ($\hat{S}(t)$).

Finalmente, comparam-se graficamente as funções de sobrevivência estimadas para cada modelo proposto com ($\hat{S}(t)$). O modelo (ou os modelos) adequado será aquele em que sua curva de sobrevivência se aproximar daquela do estimador de Kaplan-Meier.

O segundo método gráfico consiste na linearização da função de sobrevivência. A ideia básica é construir gráficos que sejam aproximadamente lineares se o modelo proposto for apropriado. Violações da linearidade podem ser rapidamente verificadas visualmente.

O gráfico utilizado é baseado em uma transformação que lineariza a função de sobrevivência do modelo proposto. Se o modelo for adequado, o resultado final será uma linha reta.

2.4.6.2 Teste da Razão de Verossimilhanças

Para o teste de hipóteses

$$\begin{cases} H_0 : \theta = \theta_0, \\ H_1 : \theta \neq \theta_0. \end{cases}$$

o teste da razão das verossimilhanças compara os valores dos logaritmos da função de verossimilhança para o valor estimado $\hat{\theta}$ e o valor de teste θ_0 . Assim, sob a hipótese nula, temos a estatística de teste

$$TRV = -2 \ln \left(\frac{\mathcal{L}(\hat{\theta})}{\mathcal{L}(\theta_0)} \right) \sim \chi_{m,1-\alpha}^2.$$

Reescrevendo a estatística de teste com a notação $\mathcal{M}(\theta) = \ln(L(\theta))$ utilizada anteriormente, temos de forma equivalente a estatística de teste

$$TRV = 2(\mathcal{M}(\hat{\theta}) - \mathcal{M}(\theta_0)) \sim \chi_{m,1-\alpha}^2.$$

A hipótese nula é rejeita se $TRV > \chi_{m,1-\alpha}^2$.

O teste de razão de verossimilhanças serve também para estudar se a generalização de modelos é necessária. Por exemplo, dado um conjunto de dados, este teste serve para verificar se o modelo Exponencial é suficiente, ou se a sua generalização pelo modelo de Weibull apresenta uma diferença significativa.

3 MATERIAIS E MÉTODOS

A natureza do estudo é de cunho quantitativo, pois busca quantificar o tempo até a evasão acadêmica e identificar fatores que podem influenciar essa evasão. O enfoque quantitativo se destaca pela sua objetividade e fornece uma compreensão clara e objetiva dos padrões de evasão, permitindo a análise estatística rigorosa dos dados coletados.

Os participantes da pesquisa são estudante do curso *Superior de Tecnologia em Gestão Pública*, do campus do Centro de Desenvolvimento Sustentável do Semiárido (CDSA), da Universidade Federal de Campina Grande (UFCG). A amostra inclui 549 estudantes, e os dados foram coletados a partir da plataforma Fala.Br (<https://falabr.cgu.gov.br/web/home>), um canal oficial de comunicação da Controladoria-Geral da União (CGU), através de uma solicitação de informações sobre alunos evadidos entre os períodos 2010.1 e 2022.1.

Os dados utilizados são referentes a matrículas dos alunos, em alguns casos o mesmo aluno pode ter evadido mais de uma vez. Para os dados de censura consideramos o tipo de evasão “graduado” (censura aleatória). Dessa forma, tem-se que o quantitativo de evasões é de 332 estudantes e o quantitativo de dados censurados é de 217 estudantes. Uma observação importante é que os "graduados" estão presentes em praticamente todos os períodos. Isso ocorre porque alguns estudantes ingressam como graduados de outros cursos ou ingressam mais de uma vez no mesmo curso, podendo dispensar algumas disciplinas do currículo.

Consta, no estudo, alguns fatores que podem contribuir para o nosso evento de interesse, que é a evasão. Os fatores disponíveis no nosso estudo são: **Sexo**, **Cota** e **Idade**. Esses fatores se configuram como potenciais contribuintes para a evasão acadêmica:

- i) **Sexo**: Será investigada a possível influência do gênero na probabilidade de evasão, considerando as diferentes trajetórias e desafios que homens e mulheres podem enfrentar no ambiente acadêmico.
- ii) **Admissão por Sistema de Cotas**: A pesquisa examinará como o acesso ao ensino superior através de cotas sociais, raciais ou indígenas impacta a permanência no curso, identificando possíveis obstáculos e oportunidades.
- iii) **Idade**: A análise buscará compreender como a idade do estudante no momento da ingresso no curso influencia o risco de evasão, considerando os diferentes níveis de maturidade, responsabilidades e desafios que cada faixa etária enfrenta.

Inicialmente, foi realizada uma análise descritiva para proporcionar uma compreensão detalhada dos dados, considerando os fatores selecionados e as informações gerais sobre o tempo até a evasão. Esta etapa serve para delinear o perfil dos estudantes do curso levando em consideração os dados de evasão. Em seguida, utilizou-se a Análise de Sobrevivência, uma conjunto de técnicas estatísticas, apresentadas no começo do trabalho para alcançar o objetivo proposto. A Análise de Sobrevivência é particularmente adequada para este tipo de estudo, pois permite modelar o tempo até a ocorrência de eventos, como a evasão acadêmica,

No decorrer do trabalho serão apresentadas estimativas para a função de sobrevivência utilizando os estimadores não-paramétricos de Kaplan-Meier e estimadores paramétricos para uma avaliação mais robusta do modelo. O uso combinado de estimadores não-paramétricos e paramétricos permite comparar diferentes abordagens e verificar a consistência dos resultados.

O primeiro passo foi mensurar o tempo até a evasão utilizando o estimador não-paramétrico de Kaplan-Meier. Isso proporciona uma visão geral das estimativas de tempo até a evasão, incluindo o tempo mediano, o tempo médio até a evasão e as taxas de falha nos intervalos. Como já apresentado anteriormente, a análise de Kaplan-Meier é amplamente utilizada em estudos de sobrevivência devido à sua simplicidade e eficácia em estimar a função de sobrevivência a partir de dados censurados.

No segundo passo, foram verificados os fatores que podem contribuir para a evasão. Para isso, foram selecionados três fatores: sexo, admissão por sistema de cotas e idade. Foi realizada uma análise gráfica dos fatores e o teste de Log-Rank para confirmar a existência de diferenças significativas entre os grupos dentro desses fatores. O teste de Log-Rank é um método estatístico utilizado para comparar curvas de sobrevivência entre dois ou mais grupos, avaliando se as diferenças observadas são estatisticamente significativas.

Por último, serão apresentadas estimativas da função de sobrevivência por meio de modelos paramétricos, que assumem que os tempos de sobrevivência seguem uma distribuição estatística bem definida (como exponencial, Weibull, log-normal ou Gamma). Para a escolha do modelo paramétrico mais adequado, realizou-se uma análise gráfica comparativa entre as curvas de sobrevivência estimadas por Kaplan-Meier e os modelos paramétricos, além da aplicação do Teste da Razão de Verossimilhança.

Além disso, o software *R* na versão 4.2.1 e a interface *RStudio* foram utilizados, juntamente com os pacotes *ggplot2*, *survival*, *survminer* e *flexsurv*, para estimar a função de sobrevivência e gerar gráficos. Esses pacotes são para análise de dados estatísticos e fornecem ferramentas poderosas para a visualização e interpretação dos resultados.

Dessa forma, este estudo oferece uma análise detalhada sobre a evasão acadêmica no curso de gestão pública, utilizando técnicas estatísticas para compreender os fatores que influenciam esse fenômeno e fornecer insights valiosos para a implementação de políticas educacionais mais eficazes.

4 RESULTADOS E DISCUSSÃO

Inicialmente, serão apresentados os fatores selecionados que podem ter impacto na evasão e também do dados preliminares da pesquisa, permitindo uma compreensão do contexto e dos elementos que influenciam o objeto de estudo. Entre os fatores considerados estão variáveis sexo, cota e idade. Esses fatores foram selecionados com base em uma revisão da literatura e em estudos anteriores que identificaram sua relevância na análise da evasão. Por exemplo, o estudo de Costa, Bispo e Pereira (2018) levantou fatores que impactam na evasão e também utilizou a Análise de Sobrevivência como base para o estudo.

Em termos descritivos, a variável sexo é qualitativa, com as categorias M (masculino) e F (feminino). A variável cota também é qualitativa, com as categorias Sim e Não. A variável idade, por sua vez, é qualitativa e categorizada em Maior ou igual a 25 e Menor que 25. As Tabelas 2, 3 e 4 apresentam, respectivamente, as tabelas de frequência com os quantitativos de matrículas destrinchados pelos grupos dos fatores sexo, cota e idade.

Tabela 2 – Tabela de matrícula por Sexo

Sexo	Frequência Absoluta	Frequência Relativa	Frequência Relativa %
Masculino	287	0,5228	52,28%
Feminino	262	0,4772	47,72%
Total	549	1	100%

Fonte: Elaborado pelo próprio autor, 2024.

Tabela 3 – Tabela de matrícula por Cota

Cota	Frequência Absoluta	Frequência Relativa	Frequência Relativa %
Sim	159	0,2896	28,96%
Não	390	0,7104	71,04%
Total	549	1	100%

Fonte: Elaborado pelo próprio autor, 2024.

Tabela 4 – Tabela de matrícula por Idade

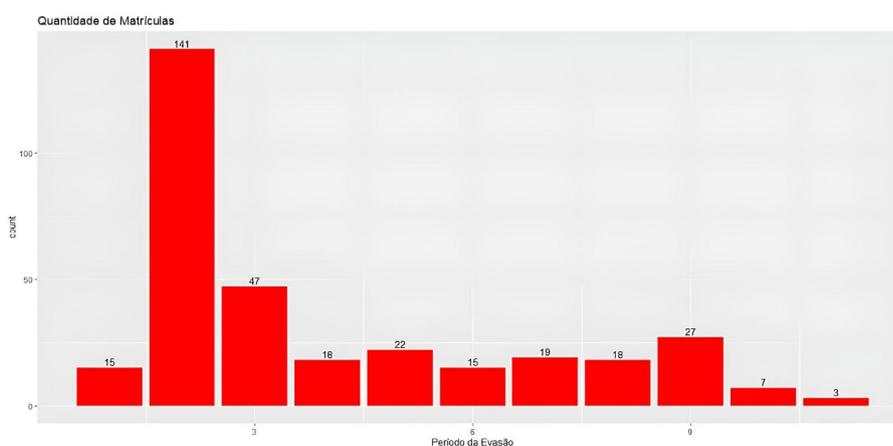
Idade	Frequência Absoluta	Frequência Relativa	Frequência Relativa %
25+	236	0,4299	42,99%
25-	313	0,5701	57,01%
Total	549	1	100%

Fonte: Elaborado pelo próprio autor, 2024.

Temos em termos de quantidade de matrículas por sexo M = 287 e F = 262, já em termos de admissão por sistema de Cotas, Não = 390 e Sim = 159 e por fim considerando Idade, onde 25+ indica idade maior ou igual a 25 anos e 25- indica idade menor que 25 anos, temos 25+ = 236 e 25- = 313. Em termos de sexo temos números de alunos quase equivalente, já em termos de idade temos uma pequena vantagem para o quantitativo de alunos abaixo de 25 anos e por fim em termo de admissão por cotas temos, onde os alunos cotistas representam aproximadamente 30% dos alunos matriculados.

Na Figura 1, a seguir, temos o gráfico do número de evadidos por período:

Figura 1 – Evasão por período

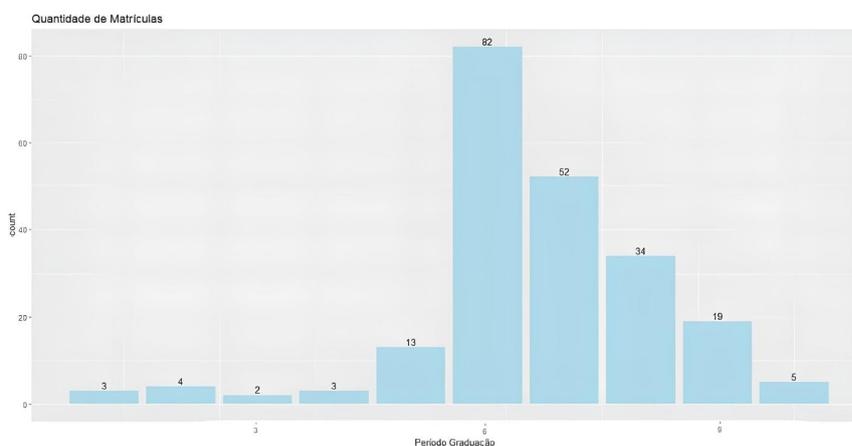


Fonte: Elaborado pelo autor, 2024.

É importante notar que os períodos de maior evasão são o segundo período (141), o terceiro período (47) e o nono período (27). Assim, nota-se uma tendência dos alunos desistirem já no primeiro ano. Outro ponto interessante, é que o "tempo limite" para os alunos concluírem o curso é no nono período e muitos alunos que não se graduaram o oitavo acabam desistindo no nono.

A Figura 2 apresenta o gráfico de censura, que neste caso representa os alunos graduados.

Figura 2 – Graduados por período

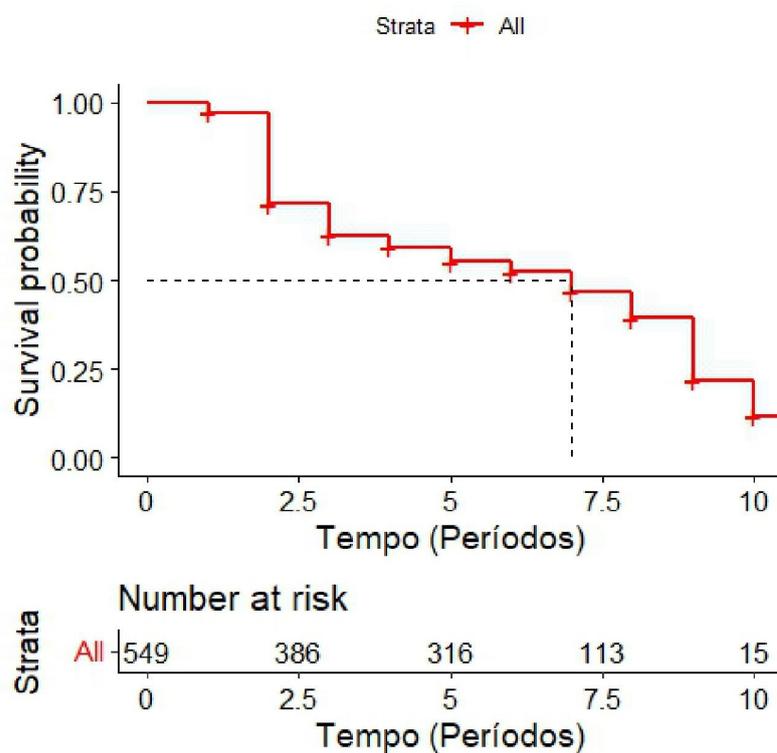


Fonte: Elaborado pelo autor, 2024.

Já com esse gráfico notamos que, os períodos com maior número de graduados são o sexto (82), o sétimo (52) e o oitavo (34). Neste sentido, tem-se que a maioria dos alunos não conseguem concluir o curso no tempo mínimo.

Para analisar a evasão dos alunos, primeiramente utilizou-se o método não paramétrico de Kaplan-Meier, obtendo-se as seguintes informações, apresentadas na Figura 3, referente ao gráfico da curva de sobrevivência utilizando o estimador de Kaplan-Meier da evasão dos alunos.

Figura 3 – Curva de Sobrevivência da evasão dos alunos.



Fonte: Elaborado pelo autor, 2024.

Temos que, de acordo com a figura 3, a estimativa de tempo mediano até a evasão é no 7 período, indicando que 50% dos alunos não evadiram até o sétimo período, também foi calculado o tempo médio até a evasão é de aproximadamente 6,18. Outro dado importante, é que no segundo período há uma queda expressiva na curva de sobrevivência indicando uma maior taxa de evasão, o mesmo é percebido, mas de forma um pouco menos atenuada, no nono período. Por interpolação, foi calculado a probabilidade de um aluno não evadir até o tempo médio, o valor foi $\hat{S}(6,18) = 0,513$. O próximo resultado, apresentado na Tabela 5, é o cálculo da estimativa da taxa de falha nos intervalos.

Tabela 5 – Estimativa da Taxa de Falha nos Intervalos

Intervalo	Taxa de Falha
[1,2)	0,266
[2,3)	0,122
[3,4)	0,0534
[4,5)	0,0696
[5,6)	0,053
[6,7)	0,103
[7,8)	0,159
[8,9)	0,445
[9,10)	0,467
[10,11)	1

Fonte: Elaborado pelo próprio autor, 2024.

Analisando a Tabela 5, observa-se que nos 5 últimos intervalos, que representam os períodos, a probabilidade de ocorrer a evasão dado que não ocorreu antes é superior 10%. Assim, temos que depois do sexto período a possibilidade do aluno evadir aumenta consideravelmente.

Analisando apenas a probabilidade da falha ocorrer no intervalo, tem-se os resultados na Tabela 6, a seguir:

Tabela 6 – Estimativa de Falha nos Intervalos

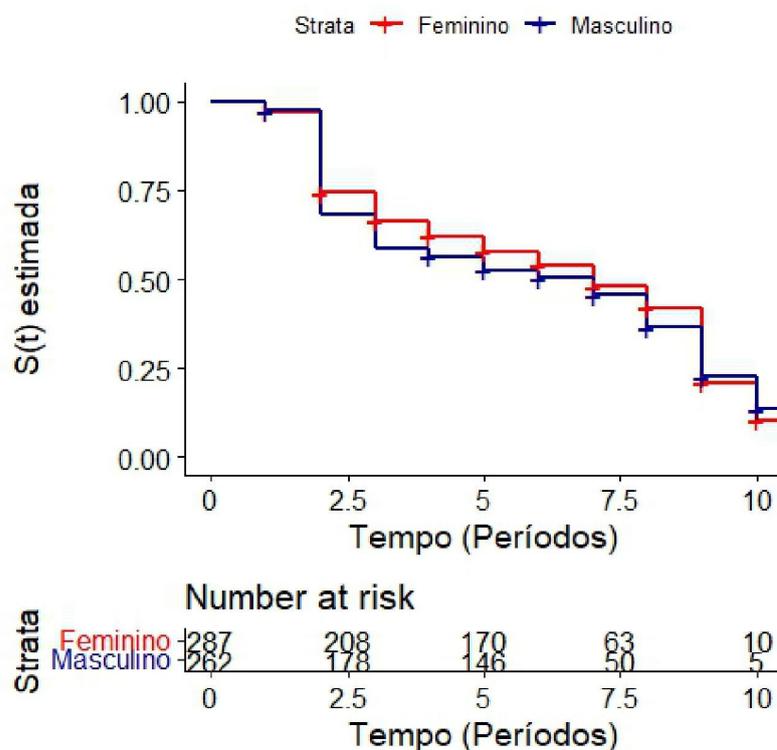
Intervalo	Falha
[1,2)	0,258
[2,3)	0,087
[3,4)	0,033
[4,5)	0,041
[5,6)	0,029
[6,7)	0,054
[7,8)	0,075
[8,9)	0,174
[9,10)	0,106
[10,11)	0,117

Fonte: Elaborado pelo próprio autor, 2024.

Observa-se que, a probabilidade de falha é maior nos intervalos [1,2) e [8,9), com probabilidade de aproximadamente 0,258 e 0,174 respectivamente. Desta forma, temos que entre o período 1 e o período 2 há uma chance de 25,8% de que a evasão ocorra e entre o período 8 e o período 9 há uma chance de 17,4% de que a evasão ocorra.

Também se utilizou o modelo de estimador de Kaplan-Meier para fazer comparações entre grupos. Na variável sexo, sua curva de sobrevivência é apresentada na Figura 4 e obteve-se os seguintes resultados.

Figura 4 – Curva de Sobrevivência em relação ao gênero

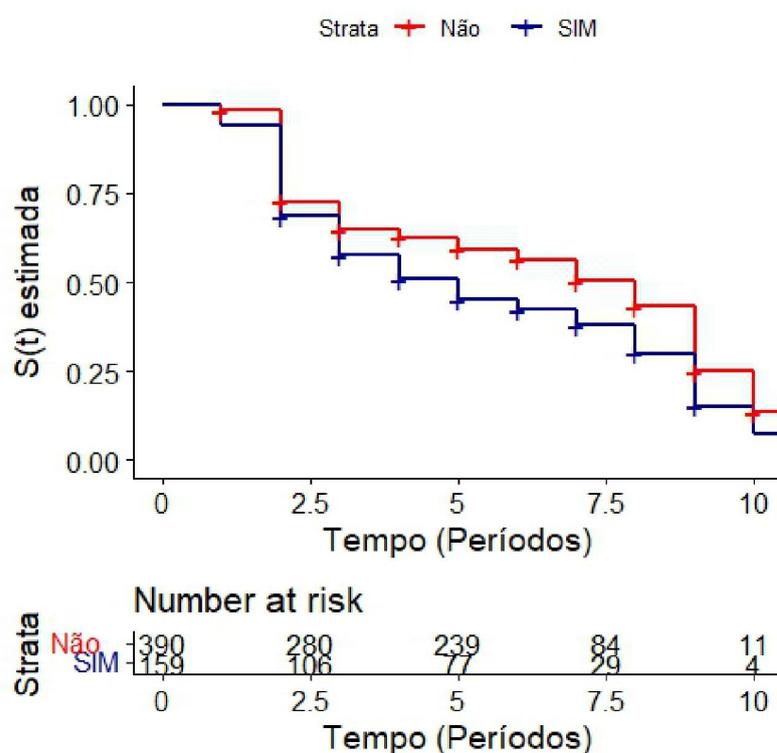


Fonte: Elaborado pelo autor, 2024.

O tempo mediano estimado tanto para o sexo feminino quanto para o masculino continuou 7 (sétimo período), já o tempo médio de evasão no sexo feminino foi de aproximadamente 6,334, enquanto no sexo masculino foi de 6,019. Dessa forma, esses valores indicam que, em média, as mulheres tendem a evadir ligeiramente mais tarde do que os homens.

Graficamente, analisando a Figura 4, tem-se que não é possível notar uma diferença significativa entre os dois grupos, do segundo ao nono período o grupo feminino aparece com menor risco de evasão, já a partir do tempo 9, por uma pequena diferença, a probabilidade do aluno evadir é mais baixa no grupo masculino. Nota-se ainda que os dois grupos possuem curvas de sobrevivência muito próximas, podendo indicar que não há diferença significativa entre as estimativas das suas funções de sobrevivência, ou seja, a probabilidade de sobrevivência em qualquer ponto no tempo é quase a mesma para ambos os grupos. Assim, realizou-se o teste de log-rank e obteve-se um p-valor = 0,4, podendo-se concluir que não há evidências que exista diferença significativa entre os grupos a 0,05 de significância. As mesmas comparações foram realizadas para o grupo de cotistas e não cotistas, obtendo-se os seguintes resultados.

Figura 5 – Curva de Sobrevivência de acordo com a Cota

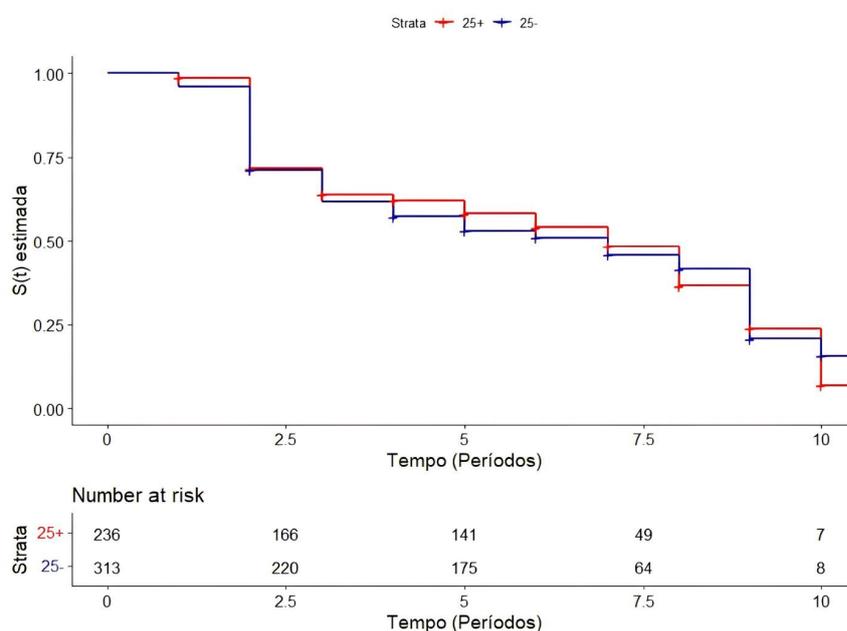


Fonte: Elaborado pelo autor, 2024.

O tempo mediano estimado até a evasão é no período 5 para cotista e para o não cotista foi no período 8. Isso indica que a metade dos alunos cotistas evadiu até o período 5, enquanto isso ocorreu no período 8 com os alunos não cotistas. Já o tempo médio de evasão aluno que não entrou por cota foi de aproximadamente 6,466, enquanto dos alunos cotistas foi de aproximadamente 5,496, indicando também que os alunos cotistas evadem mais cedo do que os não cotistas. Dessa forma, com base no tempo mediano e no tempo médio, temos que o grupo de alunos cotistas tem probabilidade de evadir mais cedo do que os não cotistas. É possível notar uma diferença mais significativa entre estes grupos, analisando a Figura 5, em comparação com os grupos separados por sexo, onde os alunos não cotistas tem menos risco de evasão ao longo dos períodos.

Para averiguar se existe diferença significativa, realizou-se o teste de logrank e obteve-se um p-valor = 0,003, podendo-se concluir que há evidências que existe diferença significativa entre os grupos de não cotistas e cotistas a 0,05 de significância. Assim, há justificativa para a necessidade de intervenções específicas, onde pode-se considerar a implementação de políticas de suporte e programas de apoio direcionados aos alunos cotistas para melhorar suas taxas de retenção. Por último, foram realizadas comparações para o grupo idade, obtendo-se os seguintes resultados.

Figura 6 – Curva de Sobrevivência em relação a idade.



Fonte: Elaborado pelo autor, 2024.

O tempo mediano estimado tanto para os alunos com idades maiores ou iguais 25 anos quanto para para os alunos com idades menores de 25 foi no período 7, já o tempo médio de evasão para os alunos com idades maiores ou iguais 25 anos foi de aproximadamente 6,244, enquanto para os menores de 25 foi de 6,142. Dessa forma, esses valores indicam que, em média, os alunos com idades maiores ou iguais 25 anos tendem a evadir ligeiramente mais tarde do que os alunos com idades menores de 25.

Graficamente, não é possível notar uma diferença significativa entre os dois grupos, do intervalo de tempo de primeiro ao oitavo período o grupo como idades maiores ou iguais 25 aparece com menor risco de evasão, já a partir do oitavo ao nono período, por uma pequena diferença, a probabilidade do aluno não evadir é mais alta no grupo menores de 25. Também nota-se que os dois grupos possuem curvas de sobrevivência muito próximas, podendo indicar que não há evidências que exista diferença significativa entre as estimativas das suas funções de sobrevivência, ou seja, os alunos de ambos os grupos têm uma probabilidade muito parecida de evadir ao longo do tempo, indicando que a idade pode não ser um fator determinante significativo na evasão.

Assim, realizou-se o teste de log-rank e obteve-se um p-valor = 0,8, podendo-se concluir que não há diferença significativa entre os grupos a 0,05 de significância.

O próximo passo foi utilizar modelos paramétricos para estimar a função de sobrevivência. A seguir, segue a Tabela 7 com as estimativas das funções de sobrevivência, tendo como base a estimativa por Kaplan-Meier e os modelos paramétricos Exponencial, Weibull, Log-Normal e o Gama Generalizado. Esta tabela apresentará as estimativas das funções de sobrevivência para diferentes modelos, permitindo uma comparação entre elas, servindo como base para avaliar os

modelos paramétricos e ajudando a identificar qual modelo oferece a melhor representação da função de sobrevivência.

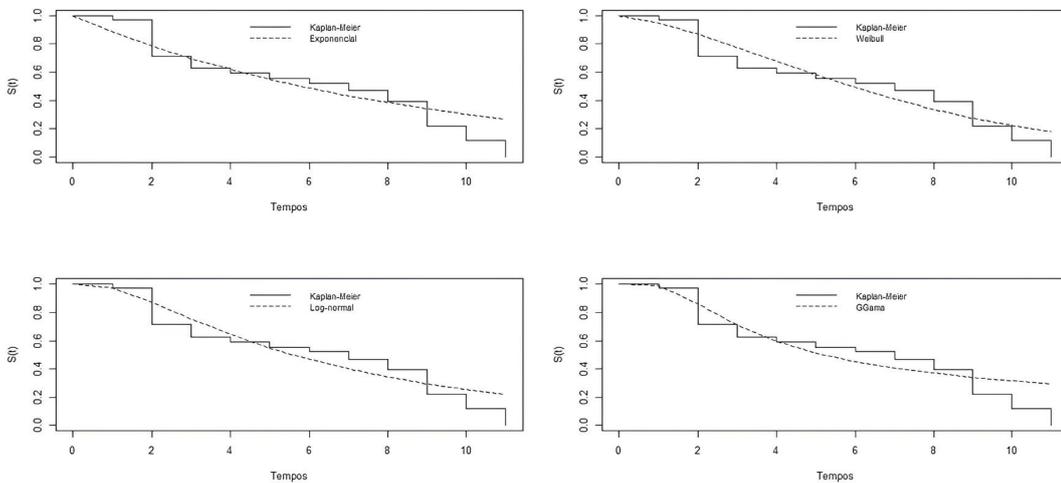
Tabela 7 – Estimativas para Função de Sobrevivência

time	Kapla-Meier	Exponencial	Weibull	Log-Normal	Gama Generalizada
1	0,973	0,887	0,951	0,973	0,994
2	0,714	0,787	0,870	0,875	0,863
3	0,627	0,698	0,775	0,757	0,709
4	0,594	0,620	0,677	0,645	0,596
5	0,552	0,550	0,582	0,548	0,514
6	0,523	0,488	0,492	0,466	0,453
7	0,469	0,433	0,410	0,398	0,407
8	0,394	0,384	0,338	0,341	0,370
9	0,220	0,341	0,275	0,294	0,339
10	0,117	0,302	0,221	0,255	0,314
11	0,000	0,268	0,176	0,221	0,293

Fonte: Elaborado pelo autor, 2024.

A seguir, na Figura 7, são apresentados os gráficos de comparação entre as estimativas das curvas de sobrevivência de Kaplan-Meier e os modelos paramétricos.

Figura 7 – Comparativo das Curvas de Sobrevivência

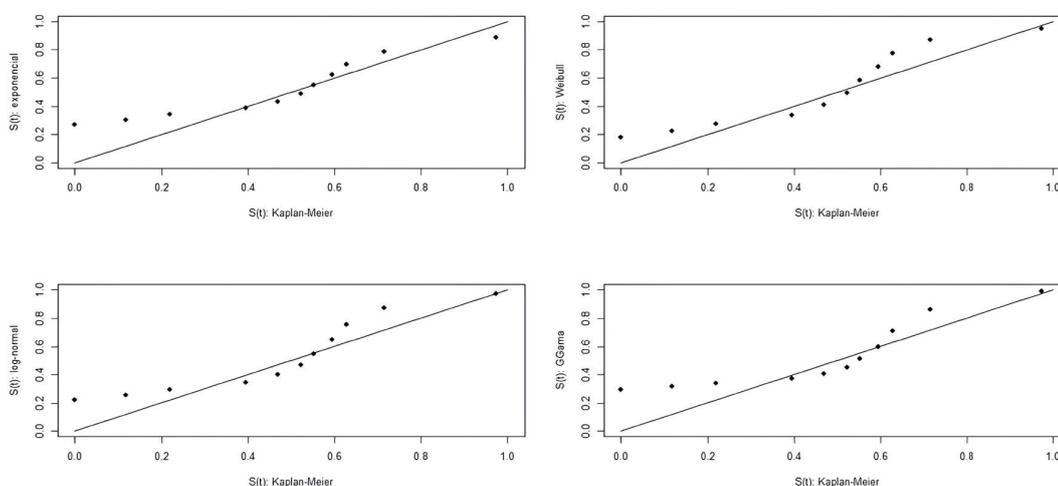


Fonte: Elaborado pelo autor, 2024.

Através dos comparativos gráficos apresentados na figura a curva de sobrevivência de Kaplan-Meier se adequou relativamente bem as curvas paramétricas, onde os melhores sobreposições forma percebidas nos modelos Weibull e Log-normal, onde o modelo Gama Generalizado tem o melhor ajuste nos períodos iniciais.

Para melhor avaliar os modelos, utilizou-se um método de linearização. Com esse método, obteve-se os seguintes resultados gráficos.

Figura 8 – Linearização das Curvas de Sobrevivência



Fonte: Elaborado pelo autor, 2024.

Com os gráficos da Figura 8, tem-se ainda indicação de melhor ajuste nos modelos Weibull e Log-normal, com uma vantagem para o modelos Weibull.

Através do teste do Teste da Razão de Verossimilhança, teve-se como resultado que o modelo que mais adequado aos dados é o modelo Gama Generalizado a 0,05 de significância. Esse modelo foi comparado com os modelos Exponencial (p -valor $< 0,0001$), Weibull (p -valor = 0,0001), Log-Normal (p -valor = 0,0001).

Apesar do Teste da razão de Verossimilhança indicar o modelo Gama Generalizado como o mais adequado, sugere-se que o modelo Weibull como melhor modelo, isso se deve que graficamente ele teve uma aproximação semelhante ao modelo recomendado e aparentemente uma melhor linearização. Além disso, o modelo Weibull é um modelo mais simples, precisando de menos parâmetros para descrever os dados.

Com o modelo escolhido, a adequação sugere que a natureza dos dados pode ser capturada de forma eficaz utilizando a estrutura do modelo Weibull. Indicando que a distribuição Weibull é uma boa representação da função de sobrevivência dos dados, permitindo simplificações e inferências adicionais.

Com os resultados obtidos, podemos fazer algumas observações. No estudo realizado por Costa, Bispo e Pereira (2018) no curso de Administração, que também é um curso de Ciências Sociais Aplicadas, constatou-se que o sexo teve uma diferença significativa, onde as mulheres têm menor probabilidade de evadir. No entanto, neste estudo, não observamos essa indicação.

Já no estudo conduzido por Júnior et al. (2022), observa-se outra disparidade em relação à realidade apresentada aqui. Em seu estudo, ao analisar alguns cursos das áreas de Ciências Sociais Aplicadas, constatou-se que alunos com mais de 25 anos têm maior chance de evasão, o que contrasta com os dados deste estudo, onde não foi observada diferença significativa.

No que se refere à admissão por sistemas de cotas, os dados apresentados sugerem uma

diferença significativa, com alunos cotistas apresentando maior risco de evasão. Este resultado não corrobora com o estudo realizado por Pinheiro Pinheiro, Pereira e Xavier (2021), que indicava a inexistência de diferença significativa.

Outro dado importante é que a taxa de evasão tendem a se concentrar mais no primeiro ano. Este dado, corrobora com a pesquisa apresentada por Salles (2023).

5 CONCLUSÕES

A estimativa de falha é maior do primeiro para o segundo período e do oitavo para o nono, com esse resultado pode-se levantar duas hipóteses a primeira é que é comum se ter um maior quantitativo de evasão no segundo período e que o papel de políticas sobre os ingressos tem que ser mais bem implementado e a outra é que o curso de gestão pública tem como prazo mínimo o sexto período, dessa forma tem-se que avaliar as causas de evasão ocorridas no nono período (prazo limite). Algumas hipóteses podem ser elaboradas, como a de dificuldade de escrita no TCC ou até dificuldade de continuar no curso após esse prazo mínimo.

Também foi possível notar que o maior quantitativo de evasão se deu no segundo período e que o maior quantitativo de graduados se deu no sexto período. A estimativa de tempo mediano até a evasão é no 7 período, ou seja, um período após o tempo mínimo de conclusão do curso. Ainda se nota que a probabilidade do aluno não evadir após o nono período é inferior a 22%.

As estimativas da função de sobrevivência entre os grupos masculino e feminino não teve diferença significativa, em termos de idade, também não houve diferença significativa, porém entre os grupos cotista e não cotista, houve diferença, sendo o grupo cotista com mais risco de evasão. O modelo Weibull foi o mais adequado para estimar a função de sobrevivência.

Tem-se como indicação de trabalhos futuros, a utilização de outras variáveis que possam influenciar o tempo de sobrevivência e comparativos com outros cursos da área das ciências sociais aplicadas.

REFERÊNCIAS

- CHIESA, E. M. F.; MORAES, A. B.; RIGÃO, M. H. Análise de sobrevivência através de um método não paramétrico. *Santa Maria*, v. 61, 2002. Citado na página 16.
- COIMBRA, C. L.; SILVA, L. B. e.; COSTA, N. C. D. A evasão na educação superior: definições e trajetórias. *Educação e Pesquisa*, SciELO Brasil, v. 47, p. e228764, 2021. Citado na página 11.
- COLOSIMO, E. A.; GIOLO, S. R. *Análise de sobrevivência aplicada*. [S.l.]: Editora Blucher, 2006. Citado 6 vezes nas páginas 11, 13, 15, 16, 17 e 21.
- COSTA, F. J. d.; BISPO, M. d. S.; PEREIRA, R. d. C. d. F. Dropout and retention of undergraduate students in management: a study at a brazilian federal university. *RAUSP Management Journal*, SciELO Brasil, v. 53, p. 74–85, 2018. Citado 3 vezes nas páginas 11, 25 e 33.
- FERREIRA, K. F. C. et al. Transformações socioespaciais na cidade de sumé/pb a partir da implantação do cdsa/ufcg. Universidade Federal de Campina Grande, 2017. Citado na página 11.
- GIL, A. C. *Métodos e técnicas de pesquisa social*. [S.l.]: 7. ed. Editora Atlas SA, 2019. Citado na página 13.
- JÚNIOR, O. d. G. F. et al. Uma experiência com business intelligence para apoiar a gestão acadêmica em uma universidade federal brasileira. *Revista Ibérica de Sistemas e Tecnologias de Informação*, Associação Ibérica de Sistemas e Tecnologias de Informacao, n. 46, p. 5–20, 2022. Citado 2 vezes nas páginas 11 e 33.
- LIMA, J. G. Modelo de regressão log-logístico discreto para dados da política de zoneamento e uso do solo na presença de observações censuradas. 2018. Citado na página 15.
- LOPES, R. et al. Fatores associados à evasão de calouros no ensino superior: um estudo com dados da universidade federal do recôncavo da bahia. *Revista Brasileira de Educação*, SciELO Brasil, v. 28, p. e280042, 2023. Citado 2 vezes nas páginas 11 e 16.
- PINHEIRO, D. C.; PEREIRA, R. D.; XAVIER, W. S. Impactos das cotas no ensino superior: um balanço do desempenho dos cotistas nas universidades estaduais. *Revista Brasileira de Educação*, SciELO Brasil, v. 26, p. e260020, 2021. Citado na página 34.
- SALLES, S. *Evasão na graduação da USP é de 17%, mas número varia por curso – Jornal da USP*. 2023. <<https://jornal.usp.br/diversidade/evasao-na-graduacao-da-usp-e-de-17-mas-numero-varia-muito-por-curso/>>. (Accessed on 06/27/2024). Citado na página 34.
- SERRANHO, P. Notas de análise de sobrevivência: uma introdução com r. 2015. Citado 2 vezes nas páginas 18 e 20.
- SERRANHO, P.; RAMOS, M. d. R. Bioestatística com spss: notas de apoio. 2017. Citado na página 16.
- SILVEIRA, A. B. Gestão da qualidade no serviço público. *Gestão pública-Unisul Virtual*, 2018. Citado na página 11.

SOUSA-FERREIRA, I. Análise de sobrevivência em r. *Boletim SPE*, Sociedade Portuguesa de Estatística, p. 10–14, 2022. Citado na página 13.