



UNIVERSIDADE  
E D U A R D O  
MONDLANE

**FACULDADE DE CIÊNCIAS**

**Departamento de Matemática e Informática**

Trabalho de Licenciatura em Estatística

**Análise dos Factores Determinantes no Tempo de Empregabilidade  
entre Graduados Universitários em Moçambique**

**Autora:** Filónia Paulino Guambe

Maputo, Abril de 2026



UNIVERSIDADE  
E D U A R D O  
MONDLANE

**FACULDADE DE CIÊNCIAS**

**Departamento de Matemática e Informática**

Trabalho de Licenciatura em Estatística

**Análise dos Factores Determinantes no Tempo de Empregabilidade  
entre Graduados Universitários em Moçambique**

**Autora:** Filónia Paulino Guambe

**Supervisor:** Miranda Albino Martins Muaualo, Ph.D, UFRJ.

Maputo, Abril de 2026

## **Declaração de Honra**

Declaro por minha honra que o presente Trabalho de Licenciatura é resultado da minha investigação e que o processo foi concebido para ser submetido apenas para a obtenção do grau de Licenciada em Estatística, na Faculdade de Ciências da Universidade Eduardo Mondlane.

Maputo, Abril de 2026

---

(Filónia Paulino Guambe)

## **Agradecimentos**

Em primeiro lugar quero agradecer a Deus, por tudo o que tem feito por mim, tem me ensinado a confiar Nele, até mesmo na ciência, entregando tudo em Suas mãos.

Aos meus pais, tenho muito a dizer sobre meu pai, que me apoiou nos momentos mais difíceis da vida, especialmente para que eu não desistisse da faculdade. Seu apoio foi inestimável; jamais teria conseguido chegar até aqui sem ele. Que Deus o tenha em Sua luz perpétua.

Aos meus irmãos e sobrinhos, não posso citar todos, mas cada um foi indispensável no meu apoio moral, físico e educacional. Meu sincero agradecimento à Mana Natália, Tio Chico, Tio Gabito, Mano Pai, Nano, Belcia e Natilazia pelo papel fundamental que desempenharam na minha vida.

Ao meu esposo, que sempre me deu apoio e confiança para estudar em grupo, mesmo quando eu era a única mulher entre vários homens. Sou grata por sua compreensão e por todas as noites em que me afastei para me dedicar aos estudos.

Ao minha família cristã, que sempre intercede por mim, especialmente à célula de Luís Cabral, com destaque para as diáconisas Judite, Joaquina, Cacilda e a menina Rose.

Também sou grata às minhas amigas intercessoras: Ratifa, Rosinha, Nilza, Arselia, Ivon e Dinisse César, por suas orações e apoio constante.

Ao meu supervisor Doutor Miranda Albino Martins Muualo, pelos ensinamentos, apoio e por me dar a oportunidade de crescimento e força para continuar sempre em frente. O mesmo agradecimento estende-se para os meus Docentes por não pouparem esforços para transmitir o conhecimento.

A todos os meus colegas, em especial ao Valdemiro Américo, Rufino Mambo, José Vidal, Caeson e Francisco, pelo companheirismo, ensinamentos em todas as etapas da vida, apoio incondicional e cordial amizade.

Muito obrigado a todos!!!!

## Resumo

O presente trabalho aborda a transição dos graduados universitários de escola para o mercado de trabalho com objectivo de analisar os factores determinantes do tempo de empregabilidade entre graduados universitários em Moçambique. Os dados foram obtidos no Centro de Estudos de Economia e Gestão da UEM. A análise foi feita mediante a aplicação de técnicas de análise de sobrevivência, aplicando os testes de Kaplan-Meier, log-rank, e os modelos Weibull, Exponencial e Log-Normal. Os resultados mostraram que da amostra de 11673 graduados, 54.8% eram do sexo masculino e a maior parte eram da faixa etária compreendida entre 18 à 22 anos. O modelo de regressão Log-Normal foi o modelo que apresentou melhor ajuste aos dados. Os factores determinante do tempo de empregabilidade entre os graduados universitários foram idade, curso e tipo de universidade.

**Palavras-chave:** Empregabilidade, Análise de Sobrevivência, modelo de regressão Log-Normal.

## **Abstract**

This paper addresses the transition of university graduates from school to the labour market with the aim of analysing the factors determining the length of employability among university graduates in Mozambique. The data were collected from the Center for Economic and Management Studies. The analysis was carried out using survival analysis techniques, including the Kaplan-Meier test, log-rank test, and the Weibull, Exponential, and Log-Normal regression models. The results showed that, out of a sample of 11,673 graduates, 54.8% were male, and the majority were aged between 18 and 22 years. The Log-Normal regression model provided the best fit to the data. The main determinants of time to employment among university graduates were age, field of study, and type of university.

**Keywords:** Employability, Survival Analysis, Log-Normal Regression Model.

## **Dedicatória**

Dedico esta monografia ao meu pai.  
**(Paulino Guambe)**

*“O mundo não está ameaçado pelas pessoas más, e sim por aquelas que permitem a maldade.”*

*(Albert Einstein)*

# Índice

Lista de Figuras . . . . .	ix
Lista de Tabelas . . . . .	x
<b>1 INTRODUÇÃO</b>	<b>1</b>
1.1 Contextualização . . . . .	1
1.2 Definição do problema . . . . .	2
1.3 Objectivos . . . . .	2
1.3.1 Objectivo geral . . . . .	2
1.3.2 Objectivos específicos . . . . .	2
1.4 Relevância de estudo . . . . .	3
1.4.1 Estrutura do trabalho . . . . .	4
<b>2 Revisão da Literatura</b>	<b>5</b>
2.1 Conceitos básicos . . . . .	5
2.2 Factores Determinantes do Tempo de Empregabilidade . . . . .	6
2.3 Estudos Relacionados . . . . .	7
2.4 Análise de Sobrevivência . . . . .	8
2.4.1 Descrição do Tempo de Sobrevivência . . . . .	10
2.4.2 Função Densidade de Probabilidade . . . . .	10
2.4.3 Função de Sobrevivência . . . . .	11
2.4.4 Função Taxa de Falha . . . . .	11
2.4.5 Função de Risco Acumulado . . . . .	11
2.4.6 Relações entre as Funções de Sobrevivência, Risco e Risco Acumulado . . . . .	12
2.4.7 Estimador de Kaplan-Meier . . . . .	12
2.4.8 Testes para a Comparação das Curvas de Sobrevivência . . . . .	13
2.4.9 Teste Log-Rank . . . . .	13
2.4.10 Teste de Gehan . . . . .	14
2.4.11 Teste de Tarone-Ware . . . . .	15
2.4.12 Modelos Paramétricos de Análise de Sobrevivência . . . . .	15
2.4.13 Estimação dos Coeficientes dos Modelos de Sobrevivência . . . . .	18
2.4.14 Comparação de modelos . . . . .	19
2.4.15 Análise de Resíduos . . . . .	21

<b>3</b>	<b>MATERIAL E MÉTODOS</b>	<b>23</b>
3.1	Classificação de Pesquisa . . . . .	23
3.1.1	Quanto à Natureza . . . . .	23
3.1.2	Quanto à abordagem . . . . .	23
3.1.3	Quanto aos objectivos . . . . .	23
3.1.4	Quanto aos procedimentos . . . . .	24
3.2	População, amostra e amostragem . . . . .	24
3.3	Material . . . . .	24
3.4	Métodos . . . . .	25
<b>4</b>	<b>RESULTADOS E DISCUSSÃO</b>	<b>28</b>
4.1	Caracterização da amostra em estudo . . . . .	28
4.2	Resultados da Estimação não Paramétrica . . . . .	28
4.3	Estimação dos coeficientes do modelo que ajusta os dados . . . . .	30
4.3.1	Seleção do Modelo . . . . .	30
4.3.2	Interpretação dos Parâmetros do Modelo Log-Normal . . . . .	31
4.3.3	Análise dos Resíduos do Modelo Log-Normal . . . . .	32
4.4	Discussão dos resultados . . . . .	34
<b>5</b>	<b>CONCLUSÕES E RECOMENDAÇÕES</b>	<b>35</b>
5.1	Conclusões . . . . .	35
5.2	Recomendações . . . . .	35
	<b>REFERÊNCIAS</b>	<b>37</b>
	<b>Apêndices</b>	<b>40</b>

# Lista de Figuras

2.1	Ilustração de alguns mecanismos de censura em que (●) representa falha e (○) é a censura. <b>Fonte:</b> Colosimo & Giolo (2006) . . . . .	9
4.1	Distribuição dos graduados por sexo e por idade . . . . .	28
4.2	Curva de sobrevivência global para os graduados . . . . .	29
4.3	Curvas da função de sobrevivência estimadas por Kaplan-Meier para os factores Sexo e Região . . . . .	30
4.4	Resíduos padronizados $e_i^*$ estimados pelo método de Kaplan-Meier e pelo modelo Log-Normal e a respectiva curva de sobrevivência estimada. . . . .	32
4.5	Resíduos de Cox-Snell estimados pelo método de Kaplan-Meier e pelo modelo exponencial padrão e a respectiva curva de sobrevivência. . . . .	33

# Lista de Tabelas

2.1	Tabela de contingência gerada no tempo $t_j$ para dois grupos . . . . .	14
3.1	Descrição das variáveis em estudo . . . . .	25
4.1	Estatísticas de comparação das curvas de sobrevivência . . . . .	30
4.2	Critério de informação de Akaike e Critério de informação Bayesiano para selecção dos modelos paramétricos de sobrevivência . . . . .	31
4.3	Estimativas dos coeficientes do Modelo Log-Normal . . . . .	31
5.1	Proporção de estudantes graduados por área de estudo e sexo . . . . .	40
5.2	Estimativas dos coeficientes do Modelo Exponencial . . . . .	40
5.3	Estimativas dos coeficientes do Modelo Weibull . . . . .	41

# CAPÍTULO 1

## INTRODUÇÃO

### 1.1 Contextualização

Actualmente, a empregabilidade dos graduados de ensino superior é um tema central nas discussões sociais e económicas. O que se observa frequentemente é que muitos estudantes universitários, após concluírem seus cursos, enfrentam dificuldades significativas na transição entre a universidade e o mercado de trabalho. Esse desafio tornou-se ainda mais relevante diante das profundas transformações pelas quais passam as sociedades contemporâneas, que impactam substancialmente essa transição (Monteiro et al., 2019). No passado, um diploma superior era quase uma garantia de emprego e de ascensão às posições mais elevadas na carreira profissional. Entretanto, a ausência dessa relação direta transformou a questão da empregabilidade dos graduados em um problema social e político, conforme (Alves, 2009).

Em Moçambique a transição da escola para o mercado de trabalho representa um dos desafios mais significativos enfrentados pelos graduados universitários. Estudos indicam que o tempo necessário para a inserção no mercado de trabalho é influenciado por uma série de factores, como a adequação da formação académica às demandas do mercado, a aquisição de habilidades práticas e a rede de contactos profissionais (Baum & Payea, 2004). Em um contexto de alta competitividade e limitações económicas, essa busca por emprego muitas vezes se prolonga, impactando a trajetória profissional dos jovens e o desenvolvimento socioeconómico do país (Tomlinson, 2012).

A empregabilidade dos graduados está intrinsecamente ligada a uma série de variáveis, incluindo a qualidade da formação académica, as habilidades adquiridas, a rede de contactos profissionais, e as condições económicas vigentes (Yorke & Knight, 2006). Em Moçambique, esses factores podem ser ainda mais complexos, dado o contexto de um mercado de trabalho em desenvolvimento, marcado por desafios como a disparidade regional, a informalidade e as limitações no sector privado (Mahdi et al., 2018).

É nesse contexto que este trabalho pretende estudar os factores determinantes no tempo de empregabilidade entre graduados universitários em Moçambique, identificando as barreiras enfrentadas e as oportunidades que podem acelerar a inserção desses jovens no mercado de trabalho.

## 1.2 Definição do problema

A elevada taxa de desemprego dos jovens em Moçambique, especialmente entre os graduados universitários, é uma questão de crescente preocupação. Apesar do aumento no número de instituições de ensino superior e do número de graduados anualmente, o mercado de trabalho formal no país tem mostrado uma capacidade limitada de absorver essa força de trabalho qualificada. Essa situação levanta importantes questões sobre a eficácia do sistema educacional em preparar os estudantes para o mercado de trabalho e sobre os factores que prolongam o tempo de empregabilidade dos recém-formados.

A transição para o mercado de trabalho não é apenas uma questão de oferta e demanda, mas envolve uma complexa interação de factores que vão desde as habilidades e competências adquiridas durante a formação, até a rede de contactos profissionais e as condições económicas do país. Em Moçambique, essas questões são exacerbadas pela desigualdade regional, onde graduados em áreas urbanas podem ter melhores oportunidades em comparação com seus pares em regiões rurais, e pela alta informalidade do mercado de trabalho.

Esses desafios tornam-se ainda mais críticos diante da pressão para que os jovens contribuam para o desenvolvimento econômico do país, o que só é possível se forem adequadamente inseridos no mercado de trabalho em tempo hábil. A falta de oportunidades adequadas pode levar à frustração, subemprego, ou até à migração de talentos, o que agrava a situação socioeconômica local.

Portanto, mediante ao problema surge a seguinte questão: **Quais são os factores determinantes no tempo de empregabilidade dos graduados universitários em Moçambique?**

## 1.3 Objectivos

### 1.3.1 Objectivo geral

Analisar os factores determinantes no tempo de empregabilidade entre graduados universitários em Moçambique.

### 1.3.2 Objectivos específicos

- Descrever o perfil socioeconómico e demográfico dos graduados em estudo;

- Identificar os factores determinantes do tempo de empregabilidade entre graduados universitários em Moçambique;
- Estimar as curvas de sobrevivência entre graduados universitários em Moçambique; e
- Comparar o tempo de empregabilidade entre diferentes regiões de Moçambique.

## 1.4 Relevância de estudo

A relevância deste estudo reside na sua capacidade de abordar um tema crítico para o desenvolvimento socioeconómico de Moçambique: a empregabilidade dos graduados universitários. Em um contexto onde a taxa de desemprego juvenil é alarmante, entender os factores que influenciam o tempo de inserção no mercado de trabalho é fundamental para a formulação de políticas públicas eficazes e para a melhoria do sistema educacional.

De acordo com dados do [Instituto Nacional de Estatística – Moçambique \(2022\)](#), a taxa de desemprego geral no país é de cerca de 18.4%. Esses números destacam a urgência de investigar as causas desse fenómeno e de desenvolver estratégias que melhorem a empregabilidade dos graduados.

Além disso, ao identificar as barreiras que os graduados enfrentam, o estudo pode contribuir para a criação de estratégias que facilitem a transição dos jovens para o mercado de trabalho. A análise das variáveis que afectam a empregabilidade, como a qualidade da formação académica e a rede de contactos profissionais, pode fornecer percepções valiosas para instituições educacionais, empresas e formuladores de políticas.

Para os graduados, a pesquisa também é relevante, pois pode ajudá-los a compreender melhor o mercado de trabalho e a se prepararem de forma mais eficaz para os desafios que enfrentarão. Ao destacar as oportunidades que podem acelerar a inserção profissional, o estudo pode inspirar acções que promovam a capacitação e o desenvolvimento de habilidades práticas, essenciais para a empregabilidade.

Por fim, ao contribuir para a discussão sobre a relação entre educação e mercado de trabalho em Moçambique, este estudo pode servir como um ponto de partida para futuras pesquisas e intervenções que visem melhorar a situação da empregabilidade dos graduados, promovendo assim o desenvolvimento económico e social do país. E aos estudantes da área e as gerações vindouras, que serva de literatura para os próximos trabalhos, com especial ênfase em análise de sobrevivência.

### **1.4.1 Estrutura do trabalho**

O presente trabalho é composto por cinco capítulos, conforme descrito a seguir: No Capítulo 1- Introdução: apresenta-se a contextualização do problema em estudo, a formulação da questão de pesquisa, a definição do problema, os objectivos do estudo, bem como a sua relevância científica e social. O Capítulo 2-Revisão de Literatura: aborda a problemática da empregabilidade, os factores associados, os conceitos fundamentais da análise de sobrevivência, incluindo a estimação paramétrica da função de sobrevivência, testes não paramétricos, análise de resíduos. No Capítulo 3- Materiais e Métodos: descreve-se a fonte de dados utilizada, as variáveis do estudo, o processo de amostragem adoptado e a metodologia aplicada, detalhando os procedimentos seguidos para alcançar os objectivos propostos. O Capítulo 4-Resultados e Discussão: apresenta os resultados obtidos por meio da análise estatística, com destaque para a aplicação da técnica multivariada de análise de sobrevivência. Além disso, realiza-se a discussão dos achados à luz de estudos anteriores, com base na literatura existente. Por fim, o Capítulo 5-Conclusões e Recomendações sintetiza as principais conclusões do trabalho, discute suas limitações, apresenta recomendações práticas e aponta sugestões para pesquisas futuras.

# CAPÍTULO 2

## Revisão da Literatura

### 2.1 Conceitos básicos

[Rueda et al. \(2004\)](#) afirmam que a empregabilidade pode ser definida como o conjunto de competências e habilidades que possibilitam a uma pessoa conquistar e manter um trabalho. Trata-se das iniciativas adoptadas pelos indivíduos para desenvolver capacidades e adquirir conhecimentos que favoreçam sua inserção, seja em empregos formais ou informais. Além disso, aspectos como características pessoais, crenças e atitudes têm um papel significativo, podendo influenciar diretamente os resultados obtidos na busca por oportunidades de trabalho impactando, assim, sua empregabilidade.

O mercado de trabalho é considerado um dos principais vectores para o desenvolvimento socioeconómico de um país. Em Moçambique, o problema de inserção dos graduados universitários no mercado de trabalho tem crescido pela limitada capacidade de absorção do mercado formal. Alguns autores definem empregabilidade como sendo a capacidade de um indivíduo obter e manter um emprego. A empregabilidade depende de diversos factores, entre eles a qualificação académica, o contexto económico e as redes de contacto profissionais. Estudos recentes em Moçambique mostram que a capacidade de inserção no mercado de trabalho está fortemente associada ao tipo de curso frequentado, às habilidades práticas adquiridas, bem como ao acesso a redes formais e informais de contacto ([Medola et al., 2020](#); [R. Santos et al., 2024](#)).

Conforme [Tomlinson \(2012\)](#), a empregabilidade pode ser influenciada por variáveis macroeconómicas e individuais. No contexto dos países com economias em desenvolvimento, como a de Moçambique, o crescimento económico e a criação de empregos formais têm sido insuficientes para acompanhar o aumento no número de graduados, especialmente em áreas urbanas, onde a concentração de formados é maior. Além disso, a disparidade regional também exerce grande influência, visto que graduados em regiões urbanas enfrentam uma situação distinta daqueles que vivem em áreas rurais ([Mahdi et al., 2018](#)).

Outro conceito central para a análise do tempo de empregabilidade é a transição da escola para o mercado de trabalho. [Baum & Payea \(2004\)](#) argumentam que essa transição é influenciada pela adequação entre a formação acadêmica e as exigências do mercado, além de factores como habilidades práticas e experiência de trabalho anterior. Em Moçambique, essa transição é particularmente difícil devido ao elevado nível de informalidade do mercado de trabalho, onde muitas oportunidades são encontradas no sector informal ([Alves, 2009](#)).

## 2.2 Factores Determinantes do Tempo de Empregabilidade

De acordo com [Baum & Payea \(2004\)](#), a formação acadêmica tem um papel crucial na empregabilidade dos jovens graduados. O autor afirma que indivíduos com diplomas de ensino superior possuem uma maior chance de inserção no mercado de trabalho e melhores condições salariais. No entanto, esse cenário pode ser diferente em países em desenvolvimento, como Moçambique, onde a formação acadêmica pode não estar directamente alinhada às demandas do mercado de trabalho.

Estudos mostram que graduados de áreas como engenharia, saúde e ciências exatas tendem a ter um tempo de empregabilidade menor em comparação a graduados em ciências humanas e sociais ([Mahdi et al., 2018](#)). Isso ocorre porque há uma maior demanda por profissionais com formação técnica e prática no mercado de trabalho moçambicano, enquanto a oferta de vagas para outras áreas é mais limitada.

As redes de contactos são outro factor crítico para a empregabilidade. A teoria do capital social, proposta por [Bourdieu \(1986\)](#), sugere que a posse de uma ampla rede de contactos profissionais pode acelerar o processo de inserção no mercado de trabalho. [Yorke & Knight \(2006\)](#) afirmam que indivíduos que participam de actividades extracurriculares e mantêm um bom relacionamento com colegas e professores têm maior probabilidade de conseguir emprego rapidamente, uma vez que essas redes actuam como facilitadoras no processo de indicação para vagas de trabalho.

Em Moçambique, estudos indicam que a informalidade do mercado de trabalho torna a rede de contactos ainda mais essencial. Segundo [Mahdi et al. \(2018\)](#), muitos empregos são obtidos através de conexões pessoais, uma vez que o mercado formal é limitado e não consegue absorver toda a demanda de graduados.

Outro factor importante é a experiência prática adquirida durante ou após o curso universitário. [Baum & Payea \(2004\)](#) destacam que graduados com experiência anterior de trabalho, estágios ou participação em programas de desenvolvimento de carreira tendem a conseguir emprego mais rapidamente do que aqueles sem experiência. No contexto moçambicano, ([Mahdi et al., 2018](#)) observaram que, devido à elevada competitividade no mercado de trabalho, a experiência prévia pode ser um diferencial significativo para a empregabilidade.

As condições econômicas de um país influenciam directamente o tempo de empregabilidade dos graduados. Segundo [Tomlinson \(2012\)](#), em países com economias instáveis, o processo de transição para o mercado de trabalho pode ser prolongado. Em Moçambique, a desigualdade regional também é um factor a ser considerado. Graduados em áreas urbanas, como Maputo, têm maior facilidade de encontrar emprego em comparação com aqueles em regiões rurais, onde as oportunidades de trabalho formal são mais escassas ([Monteiro et al., 2019](#)).

Estudos como os de [Fragoso et al. \(2019\)](#) têm contribuído significativamente para a compreensão dos desafios enfrentados por jovens graduados em busca de emprego. Esses estudos apontam que, além das dificuldades econômicas, as barreiras institucionais, como a falta de políticas públicas voltadas para a juventude, prolongam o tempo de empregabilidade. Outros estudos, como o de [Mahdi et al. \(2018\)](#), reforçam a importância de melhorar a qualidade da educação e promover programas de estágios e desenvolvimento de carreira para facilitar a inserção dos jovens no mercado de trabalho.

## 2.3 Estudos Relacionados

Estudos recentes têm utilizado a análise de sobrevivência para investigar os determinantes do tempo de empregabilidade entre graduados universitários. Por exemplo, [Wobse et al. \(2022\)](#) aplicaram o modelo de tempo de falha acelerado (Accelerated Failure Time - AFT) para analisar o tempo de espera até o primeiro emprego entre graduados da Wolkite University, na Etiópia. O estudo revelou que a mediana do tempo de espera foi de 17 meses, sendo influenciada por factores como a faculdade de origem, a média acadêmica, a região de residência e os métodos de busca de emprego. Além disso, os autores identificaram que graduados que utilizaram redes de contacto para procurar emprego tiveram tempos de espera mais curtos em comparação com aqueles que dependeram exclusivamente de anúncios públicos.

De forma semelhante, [Alemu & Yismaw \(2022\)](#) investigaram o tempo de espera entre a graduação e o primeiro emprego entre graduados da Debre Markos University (Burie campus), também na Etiópia. Utilizando o estimador de Kaplan-Meier, o modelo de riscos proporcionais de Cox e o modelo AFT com distribuição log-normal, os autores identificaram como factores significativos a média final, idade, género, estágio realizado e número de candidaturas enviadas.

Conforme [Neves & Lima \(2010\)](#) analisaram a duração da procura do primeiro emprego com mais de três meses de duração, utilizando o estimador de Kaplan-Meier. Observaram que a transição média para o primeiro emprego foi de 20,4 meses, e também mostraram que quanto mais elevado o nível de escolaridade mais curtos são os tempos de procura. Após dois anos, mais de 40% dos jovens com até o 3.º ciclo do ensino básico ainda procuravam emprego, enquanto essa proporção

cafa para cerca de 15% entre os que possuíam ensino superior. Os autores identificaram, a partir do modelo de riscos proporcionais de Cox, que jovens com ensino superior completo apresentaram risco de 73,1% maior de encontrar emprego em comparação com aqueles com apenas o ensino básico.

Resultados semelhantes foram encontrados por [Oliveira et al. \(2006\)](#), que, aplicando a análise de sobrevivência ao contexto brasileiro, mostraram que a entrada no mercado de trabalho tende a ocorrer precocemente e, em muitos casos, antecede a saída da escola. Nesse estudo, a educação dos pais especialmente da mãe reduziu significativamente o risco de abandono escolar e de transição precoce para a fecundidade, indicando que níveis mais elevados de escolaridade familiar e individual estão associados a trajetórias laborais mais estáveis e a transições mais tardias.

Apesar da crescente aplicação, há ainda escassez de estudos semelhantes em países lusófonos, como Moçambique, onde os desafios de inserção no mercado de trabalho para graduados universitários podem apresentar particularidades relevantes.

Assim, o presente estudo busca preencher essa lacuna, aplicando técnicas de análise de sobrevivência para investigar os factores que influenciam o tempo de empregabilidade entre graduados universitários em Moçambique.

## 2.4 Análise de Sobrevivência

De acordo com [R. O. Santos \(2017\)](#) & [Sousa \(2019\)](#), a análise de sobrevivência é uma das áreas da estatística com maior implementação nos últimos anos. Na análise de sobrevivência a variável resposta é o tempo até a ocorrência de um evento de interesse, por exemplo, a morte de um indivíduo ou a falha de um equipamento.

Conforme [Campos \(2016\)](#), na análise de sobrevivência a principal característica de dados de sobrevivência é a presença de censura, que é a observação parcial da resposta. Devido a um acontecimento no qual algumas informações foram interrompidas de serem estudadas, ou pelo facto do estudo terminar e seu tempo de falha durar além da época de estudo, ou por uma interrupção antes do final do estudo. De uma forma geral, pode-se dizer que a análise de sobrevivência corresponde a modelos de regressão que incorporam "censura" na variável dependente.

De acordo com [Colosimo & Giolo \(2006\)](#), o tempo de falha é o tempo do início da observação até um evento final do estudo ou perda do tratamento. A data do diagnóstico ou do início do tratamento da doença também são outras escolhas possíveis.

Segundo [Campos \(2016\)](#), o tempo de início deve ser bem definido para cada paciente ou indi-

viduo em estudo. A escala de medida é quase sempre o tempo real ou “relógio”, mas apresentam outras alternativas. É muito importante definir de forma clara a falha em estudos de sobrevivência. O tempo de sobrevivência ou tempo de falha vai do tempo inicial até a ocorrência do evento falha usando a escala de medida já definida.

De acordo com [Sousa \(2019\)](#), dados censurados são frequentes em estudos de sobrevivência de longa duração, pois muitas vezes o estudo é concluído antes que todos os indivíduos apresentem falha. Assim, é comum que haja observações incompletas. Essas observações são chamadas de censuras e podem ocorrer por diversos motivos, como a perda de seguimento do indivíduo, a morte por uma razão diferente da estudada ou a não ocorrência do evento de interesse. As observações censuradas são essenciais para o estudo, pois oferecem dados sobre o tempo de vida dos pacientes. Além disso, excluir essas censuras ao calcular estatísticas relevantes pode levar a conclusões distorcidas.

Segundo [Colosimo & Giolo \(2006\)](#), uma análise de sobrevivência envolve uma resposta temporal e frequentemente adota uma perspectiva de longo prazo. Ela pode ser finalizada antes que o evento de interesse ocorra para todos os casos da amostra. Uma característica importante desse tipo de estudo é a presença de observações incompletas ou parciais do tempo de sobrevivência, conhecidas como censura. Ainda segundo [Colosimo & Giolo \(2006\)](#), existem três tipos de censura: Censura do tipo I, Censura do tipo II e Censura aleatória. A Figura 2.1 ilustra os mecanismos de censura.

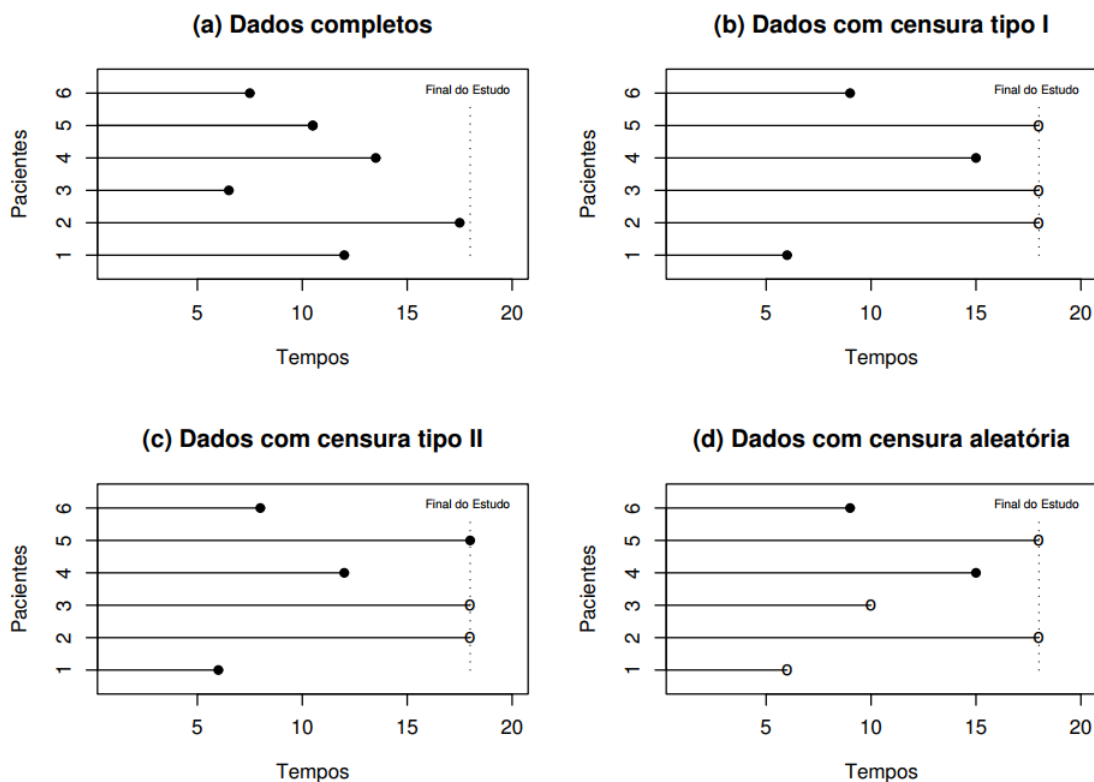


Figura 2.1: Ilustração de alguns mecanismos de censura em que (●) representa falha e (○) é a censura. **Fonte:** [Colosimo & Giolo \(2006\)](#)

- **Censura do Tipo I:** Censura do Tipo I ou censura à esquerda ocorre quando o estudo é terminado após um período pré-estabelecido de tempo.
- **Censura do Tipo II:** É como nos outros casos aquela onde o estudo será terminado após ter ocorrido o evento de interesse em um número pré-estabelecido de indivíduos.
- **Censura aleatório:** É a mais comum na prática e ocorre quando no estudo é retirada uma observação sem ter ocorrido o evento de interesse. Ou seja, censura aleatória acontece quando há perda de seguimento.

### 2.4.1 Descrição do Tempo de Sobrevivência

Seja  $T$  a variável aleatória não negativa, geralmente contínua, que representa o tempo até a falha ou censura. Segundo [Colosimo & Giolo \(2006\)](#),  $T$  é definido por meio da função de sobrevivência ou pela função taxa de falha. Essas duas funções, assim como outras relacionadas, são amplamente utilizadas na análise de dados de sobrevivência. A variável indicadora de falha ou censura deve ser incluída no estudo para fins analíticos e é expressa por

$$\delta_i = \begin{cases} 1, & \text{se o } i\text{-ésimo indivíduo falhou} \\ 0, & \text{se o } i\text{-ésimo indivíduo foi censurado} \end{cases}$$

A variável  $\delta_i$  representa, juntamente com o tempo de falha  $t_i$ , os dados de sobrevivência para o indivíduo  $i$  ( $i = 1, \dots, n$ ). Na presença de um vector de covariáveis  $\mathbf{x}_i$ ,  $i = 1, \dots, n$ , os dados de sobrevivência são representados por  $(t_i, \delta_i, \mathbf{x}_i)$ .

### 2.4.2 Função Densidade de Probabilidade

De acordo com [Magalhães \(2006\)](#), a distribuição de uma variável aleatória contínua é definida por sua função de densidade de probabilidade. A probabilidade de essa variável assumir um valor dentro de um intervalo específico é determinada pela integral de sua função  $f(t)$  ao longo desse intervalo. A função densidade é definida da seguinte maneira:

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t)}{\Delta t}$$

Para que uma função  $f(t)$  seja considerada uma função de densidade de probabilidade de uma variável aleatória absolutamente contínua, ela deve atender às seguintes condições:

- $f(t)$  é não-negativa, ou seja,  $f(t) \geq 0 \forall t \in \mathcal{R}$ . Em especial, se  $t$  é uma variável não-negativa.
- A área entre a função densidade e o eixo  $t$  é igual a 1, ou seja, se  $t$  é uma variável aleatória não-negativa, tem-se:

$$\int_{-\infty}^{\infty} f(t) dt = 1$$

### 2.4.3 Função de Sobrevivência

De acordo com Colosimo & Giolo (2006) a função de sobrevivência é definida como a probabilidade de uma observação não falhar além de um certo tempo  $t$ , ou seja, a probabilidade de uma observação sobreviver para além do instante  $t$  e representa-se por

$$S(t) = P(T \geq t). \quad (2.1)$$

Em consequência, a função de distribuição acumulada, isto é,  $F(t) = 1 - S(t)$ , é definida como a probabilidade de uma observação não sobreviver além do tempo  $t$ .

Segundo Colosimo & Giolo (2006), a Função de Sobrevivência possui as seguintes propriedades:

1.  $S(0) = 1$  i.e. todos os indivíduos estão vivos no começo do experimento;
2.  $S(t)$  é contínua à direita;
3.  $\lim_{t \rightarrow \infty} S(t) = 0$ , i.e. eventualmente todos os indivíduos morrem; e
4.  $S(t)$  é monótona decrescente e continua.

### 2.4.4 Função Taxa de Falha

De acordo com (Colosimo & Giolo, 2006), a função taxa de falha é também chamada função de risco e representa a taxa de falha instantânea no tempo  $t$  condicional à sobrevivência até o tempo  $t$ , e também é chamada de função *hazard* e representa a taxa instantânea de um indivíduo sofrer um evento entre o tempo  $t$  e  $t+\Delta t$ , dado que ele sobreviveu até o tempo  $t$ . Como  $\Delta t$ , é o incremento de tempo considerado, pode ser infinitamente pequeno, a função de risco expressa o risco instantâneo de ocorrência de um evento, dado que até então o evento não tenha ocorrido. Matematicamente pode ser expressa pela fórmula:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{[(t \leq T < t + \Delta t) | T \geq t]}{\Delta t}. \quad (2.2)$$

### 2.4.5 Função de Risco Acumulado

De acordo com R. O. Santos (2017), a função de risco acumulado  $A(t)$  mede o risco de ocorrência do evento até um determinado tempo  $t$ . É de realçar que o risco acumulado também é uma taxa e está restrito ao intervalo  $[0, 1]$ , matematicamente, isso significa a soma de todos os riscos em todos os tempos até o tempo  $t$ . Como  $T$  é uma variável aleatória contínua, esta é uma soma infinitesimal e pode ser calculada a partir de uma integral da função de risco:

$$A(t) = \int_0^t \lambda(u) du. \quad (2.3)$$

De acordo com [Colosimo & Giolo \(2006\)](#), a função de risco acumulado,  $\Lambda(t)$ , não tem uma interpretação direta mas pode ser útil na avaliação da função de maior interesse que é a de taxa de falha. Isto acontece essencialmente na estimação não-paramétrica em que  $\Lambda(t)$  apresenta um estimador com propriedades óptima e é difícil de ser estimada.

#### 2.4.6 Relações entre as Funções de Sobrevivência, Risco e Risco Acumulado

Segundo [Carvalho et al. \(2011\)](#), a função de sobrevivência, de risco e de risco acumulado são relacionadas entre si. Assim, ao estimar a função de sobrevivência  $S(t)$ , obtém-se também as funções de risco  $\lambda(t)$  e de risco acumulado  $\Lambda(t)$  a partir do logaritmo de  $S(t)$ . Por outro lado, se a função de risco é estimada, pode-se obter a função de risco acumulado, e a partir dela, a função de sobrevivência  $S(t)$ . Essas funções são matematicamente equivalente e representam diferentes formas de descrever o mesmo fenómeno. As relações são expressas por:

$$\lambda(t) = -\frac{d}{dt}(\log S(t)) \quad (2.4)$$

onde  $f(t)$  é a função densidade de probabilidade.

$$\Lambda(t) = \int_0^t \lambda(u)du = -\log(S(t)) \quad (2.5)$$

$$S(t) = \exp\{-\Lambda(t)\} = \exp\left\{-\int_0^t \lambda(u)du\right\}. \quad (2.6)$$

#### 2.4.7 Estimador de Kaplan-Meier

Segundo [R. O. Santos \(2017\)](#), o estimador não paramétrico de Kaplan-Meier é o mais conhecido e utilizado para estimar a função de sobrevivência, proposto por Kaplan e Meier (1958). Também é conhecido como estimador limite-produto. O estimador de Kaplan-Meier, divide o tempo de falha em intervalos, cujos limites correspondem ao tempo de falha da amostra. As informações da probabilidade estimada de sobrevivência para um determinado tempo são ilustradas através da curva de Kaplan-Meier.

Ainda segundo [R. O. Santos \(2017\)](#), esta função designa-se por função de sobrevivência empírica e, com base numa amostra de dimensão  $n$ , define-se do seguinte modo.

$$\hat{S}(t) = \frac{\text{número de observações } > t}{n}; \quad t \geq 0. \quad (2.7)$$

O estimador de Kaplan-Meier, na sua construção, considera tantos intervalos quantos forem o número de falhas distintas. Assumindo:

- $t_1 < t_2 < \dots < t_k$ , os  $k$  tempos distintos e ordenados de falha;

- $d_j$ , o número de falhas em  $t_j, j=1, \dots, k$  e
- $n_j$  o número de indivíduos sob risco em  $t_j$ , ou seja, os indivíduos que não falharam e não foram censurados até o instante imediatamente anterior a  $t_j$ .

De acordo com [Colosimo & Giolo \(2006\)](#), o estimador é definido pela expressão 2.8

$$\hat{S}(t) = \prod_{j:t_j < t} \left( \frac{n_j - d_j}{n_j} \right) = \prod_{j:t_j < t} \left( 1 - \frac{d_j}{n_j} \right). \quad (2.8)$$

As principais propriedades do estimador de Kaplan-Meier são as seguintes:

1. não-viciado para amostras grandes;
2. é fracamente consistente;
3. converge assintoticamente para um processo gaussiano; e
4. é estimador de máxima verosimilhança.

## 2.4.8 Testes para a Comparação das Curvas de Sobrevivência

Conforme [Andrade & Cabete \(2012\)](#), após determinar as estimativas das funções utilizadas na análise de sobrevivência, é comum realizar a comparação entre dois ou mais grupos de indivíduos. Esses grupos podem ser definidos de acordo com diferentes categorias de uma característica particular. Quando os indivíduos são distribuídos em dois grupos,  $G_A$  e  $G_B$ , frequentemente testa-se a hipótese relativa às funções de sobrevivência correspondentes. Diversos testes não paramétricos estão disponíveis para essa finalidade, destacando-se os testes log-rank, Gehan Breslow e Tarone-Ware.

## 2.4.9 Teste Log-Rank

De acordo com [Colosimo & Giolo \(2006\)](#), após calcular a estimativa de Kaplan-Meier para as funções de sobrevivência e representá-las graficamente, utiliza-se testes não paramétricos para comparar as curvas de sobrevivência de diferentes grupos. O teste log-rank é amplamente aplicado para comparar funções de sobrevivência e é particularmente útil quando a razão das funções de risco entre os grupos é aproximadamente constante.

Considerando dois grupos de indivíduos, em que  $S_i(t)$  representa a função de sobrevivência do  $i$ -ésimo grupo (A e B), as hipóteses de teste são:

$$H_0 : S_A(t) = S_B(t) \quad \text{vs} \quad H_1 : S_A(t) \neq S_B(t). \quad (2.9)$$

Sejam  $t_1 < t_2 < \dots < t_k$  os tempos de falha distintos da amostra formada pelos  $m+n$  indivíduos, onde  $d_j$  é o número de falhas em  $t_j$  ( $j = 1, \dots, k$ ), e  $d_{ij}$  representa o número de falhas no grupo

A e B em  $t_j$ . Denota-se  $n_j$  como o número de indivíduos em risco imediatamente antes de  $t_j$ , considerando os dois grupos. Os dados podem ser organizados em uma tabela de contingência  $2 \times 2$ , como mostrado na Tabela 2.1.

Tabela 2.1: Tabela de contingência gerada no tempo  $t_j$  para dois grupos

	Grupos		Total
	A	B	
Falha	$d_{Aj}$	$d_{Bj}$	$d_j$
Não falha	$n_{Aj} - d_{Aj}$	$n_{Bj} - d_{Bj}$	$n_j - d_j$
<b>Total</b>	$n_{Aj}$	$n_{Bj}$	$n_j$

**Fonte:** Aparício (2013).

Sob a hipótese nula,  $H_0$ , a distribuição de  $d_{Aj}$ , condicional ao número de falhas em  $t_j$  e ao total de indivíduos em risco, segue uma distribuição hipergeométrica:

$$p(d_{Aj} | d_j, n_j) = \frac{\binom{d_j}{d_{Aj}} \binom{n_j - d_j}{n_{Aj} - d_{Aj}}}{\binom{n_j}{n_{Aj}}}. \quad (2.10)$$

As médias e variâncias condicionais da variável aleatória  $d_{Aj}$  são:

$$e_{Aj} = \frac{n_{Aj} d_j}{n_j}, \quad v_{Aj} = \frac{n_{Aj} n_{Bj} d_j (n_j - d_j)}{n_j^2 (n_j - 1)}. \quad (2.11)$$

Definindo  $U = \sum_{j=1}^k (d_{Aj} - e_{Aj})$ , tem-se que sob  $H_0$ ,  $E(U) = 0$  e  $\text{var}(U) = \sum_{j=1}^k v_{Aj}$ . A estatística do teste é:

$$T = \frac{U^2}{\text{var}(U)}, \quad (2.12)$$

que segue uma distribuição qui-quadrado com 1 grau de liberdade ( $\chi_1^2$ ).

#### 2.4.10 Teste de Gehan

Conforme [Gonçalves \(2020\)](#), o teste de Gehan é uma extensão do teste de Mann-Whitney-Wilcoxon para dados censurados. Este teste avalia a hipótese nula de que não há diferenças entre as funções de sobrevivência dos grupos. Considere-se  $m$  e  $n$  as dimensões das amostras dos grupos 1 e 2. As estatísticas do teste são dadas por:

$$U_{kj} = \begin{cases} +1 & \text{se } (t_k > t_j, \delta_j = 1) \text{ ou } (t_k = t_j, \delta_k = 0, \delta_j = 1), \\ -1 & \text{se } (t_k < t_j, \delta_k = 1) \text{ ou } (t_k = t_j, \delta_k = 1, \delta_j = 0), \\ 0 & \text{caso contrário.} \end{cases} \quad (2.13)$$

A estatística final é:

$$W_G = \frac{U_G^2}{V_G}, \quad (2.14)$$

com distribuição qui-quadrado com 1 grau de liberdade.

### 2.4.11 Teste de Tarone-Ware

Conforme explicado por [Anchisi \(2011\)](#), os testes de log-rank e Gehan fazem parte de uma classe de métodos não paramétricos em que a estatística de teste pode ser representada de forma geral como:

$$TW = \frac{\left[ \sum_{j=1}^r w_j (D_{1j} - e_{1j}) \right]^2}{\sum_{j=1}^r w_j^2 v_{1j}},$$

em que  $w_j$  representa constantes conhecidas que atribuem pesos distintos às diferentes parcelas das somas. Sob a hipótese nula  $H_0$ , a distribuição assintótica dessa estatística segue uma distribuição qui-quadrado com um grau de liberdade.

No caso do teste log-rank, os pesos são definidos como  $w_j = 1$ , enquanto para o teste de Gehan, utilizam-se  $w_j = n_j$ . O teste de Tarone-Ware propõe uma abordagem intermediária, em que os pesos adotados são  $w_j = \sqrt{n_j}$ .

### 2.4.12 Modelos Paramétricos de Análise de Sobrevivência

#### Modelo Exponencial

Segundo [Carvalho et al. \(2011\)](#), o modelo exponencial é um modelo adequado para estudos onde, o tempo até a falha é bem detalhado através de uma distribuição de probabilidade exponencial, foi o mais amplamente utilizado em estudos de confiabilidade em sistemas eletrônicos e posteriormente em estudos de tempo de vida, e em remissão de doenças crônicas e infecciosas no final dos anos 1940.

De acordo com [Colosimo & Giolo \(2006\)](#), a distribuição exponencial é um dos modelos probabilísticos mais simples usados para descrever o tempo de falha. Esta distribuição apresenta um único parâmetro e é a única que se caracteriza por ter uma função de taxa de falha constante. Ela tem sido extensivamente usada como um modelo para o tempo de vida de certos produtos e materiais e tem descrito adequadamente o tempo de vida de óleos isolantes e dielectricos, entre outros.

De acordo com [Colosimo & Giolo \(2006\)](#), a função densidade de probabilidade para a variável aleatória tempo de falha  $t$  com distribuição exponencial é dada por:

$$f(t) = \frac{1}{\alpha} \exp \left\{ - \left( \frac{t}{\alpha} \right) \right\}, \quad t \geq 0, \quad (2.15)$$

Onde o parâmetro  $\alpha \geq 0$  é o tempo médio de vida. O parâmetro  $\alpha$  tem a mesma unidade do tempo de falha  $t$ . Isto é, se  $t$  é medido em horas,  $\alpha$  também será medido em horas. As funções de

sobrevivência  $S(t)$  e de risco  $\lambda(t)$  são dadas, respectivamente, por:

$$S(t) = \exp \left\{ - \left( \frac{t}{\alpha} \right) \right\} \quad (2.16)$$

e

$$\lambda(t) = \frac{1}{\alpha} \quad \text{para } t \geq 0. \quad (2.17)$$

No modelo de regressão exponencial, usado quando se assume que o risco é constante ao longo do tempo, o parâmetro  $\alpha$  depende das covariáveis  $x$  da seguinte forma:

$$\lambda(t | x) = \alpha(x) = \exp(x\beta); \quad (2.18)$$

e sua função de sobrevivência para o modelo exponencial fica:

$$S(t | x) = \exp(-\exp(x\beta)t).$$

## Modelo Weibull

Segundo [Magalhães \(2006\)](#), a distribuição Weibull foi proposta por Weibull em 1939, e é utilizada para situações em que o evento de interesse (tempo até falha) é bem definido por meio da distribuição de probabilidade Weibull, pois, ele possui uma grande variedade de formas, todas com propriedades básicas: função de taxa de falha é monótona, isto é, crescente, decrescente ou constante e, por isso, consegue se adaptar a diversas situações práticas.

A função densidade de probabilidade do modelo Weibull para a variável aleatória tempo até a falha  $t$ , é dada pela expressão:

$$f(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1} \exp \left\{ - \left( \frac{t}{\alpha} \right)^\gamma \right\}, \quad t \geq 0, \quad (2.19)$$

Onde o parâmetro de forma  $\gamma$  e o parâmetro de escala  $\alpha$  tem a mesma unidade de medida e  $t$  e  $\gamma$  não tem medida. Para esta distribuição, as funções de sobrevivência e de risco são, respectivamente,

$$S(t) = \exp \left\{ - \left( \frac{t}{\alpha} \right)^\gamma \right\} \quad (2.20)$$

e

$$\lambda(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1}, \quad (2.21)$$

para  $t, \alpha$  e  $\gamma \geq 0$ . Note que quando  $\gamma = 1$  tem-se a distribuição exponencial e, sendo assim, a distribuição exponencial é um caso particular da distribuição de Weibull.

O parâmetro  $\gamma$  determina a forma da função taxa de falha da seguinte maneira:

- Se  $\gamma < 1 \rightarrow$  a função da taxa de falha decresce,

- Se  $\gamma > 1 \rightarrow$  a função da taxa de falha cresce,
- Se  $\gamma = 1 \rightarrow$  a função da taxa de falha é constante.

## Modelo Log-Normal

Segundo [Serranho \(2015\)](#), no modelo log-normal é considerado a regressão linear da média  $\mu$  da variável  $Y = \ln(T)$ , matematicamente é representado por

$$\mu = \beta(x) + \epsilon \quad (2.22)$$

em que  $\beta(x)$  é a combinação linear dos betas e  $\epsilon$  segue uma distribuição normal com média nula. A sua função de sobrevivência é dada por

$$s(t/x) = \Phi\left(\frac{\beta(x) - \ln t}{\sigma}\right) \quad (2.23)$$

## Distribuição Gama

A distribuição gama é menos utilizada que a distribuição Weibull, embora se ajuste bem a diversos dados de tempo de vida e inclua a distribuição exponencial como um caso particular. ( $k = 1$ ).

A função densidade de probabilidade é

$$f(t; a, k) = \frac{1}{\Gamma(k)a^k} t^{k-1} \exp\left[-\left(\frac{t}{a}\right)\right]; \quad t \geq 0, \quad (2.24)$$

em que  $k > 0$  é o parâmetro de forma e  $a > 0$  é o parâmetro de escala.

Segundo as suas função de sobrevivência e de risco são dadas por:

$$S(t; a, k) = \int_t^\infty \frac{1}{\Gamma(k)a^k} u^{k-1} \exp\left[-\left(\frac{u}{a}\right)\right] du = 1 - I(k, t/a) \quad (2.25)$$

## Distribuição Gama Generalizada

Segundo [Gonçalves \(2020\)](#), a distribuição Gama Generalizada foi proposta por Stacy (1962), podendo ser caracterizada por três parâmetros,  $\sigma$ ,  $r$  e  $\mu$ . A função densidade de probabilidade da distribuição Gama Generalizada em relação a variável aleatória  $T$  é dada por:

$$f(t; \mu, \sigma, r) = \frac{r}{\Gamma(\mu)\sigma^{r\mu}} t^{r\mu-1} \exp\left\{-\left(\frac{t}{\sigma}\right)^r\right\},$$

onde  $\sigma > 0$  é o parâmetro de escala,  $\mu > 0$  e  $r > 0$  são os parâmetros de forma, que dependem do vector de covariáveis  $\mathbf{x}$ , ou seja  $\sigma = \exp(\mathbf{x}^T \boldsymbol{\beta}_\sigma)$ ,  $\mu = \exp(\mathbf{x}^T \boldsymbol{\beta}_\mu)$  e  $r = \mathbf{x}^T \boldsymbol{\beta}_r$ , onde  $\boldsymbol{\beta}_m =$

$(\beta_0, \beta_1, \dots, \beta_q)$  é o vector de parâmetros desconhecidos,  $m = \mu, \sigma$  e  $r$ , e  $\mathbf{x} = (x_1, x_2, \dots, x_q)$ , sendo  $q$  o número de variáveis explicativas.

Esta é uma importante distribuição visto que possui como casos particulares: a distribuição Exponencial quando  $\gamma = k = 1$ ; a distribuição Gama quando  $\gamma = 1$  e a distribuição Weibull quando  $k = 1$ .

A Esperança e a Variância da distribuição Gama Generalizada são dadas, respectivamente por,

$$E[T] = \frac{\sigma \Gamma\left(\frac{r\mu+1}{r}\right)}{\Gamma(\mu)} \quad \text{e} \quad \text{Var}[T] = \frac{\sigma^2}{\Gamma(\mu)} \left\{ \Gamma\left(\frac{r\mu+2}{r}\right) - \frac{[\Gamma\left(\frac{r\mu+1}{r}\right)]^2}{\Gamma(\mu)} \right\}.$$

As funções de sobrevivência  $S(t)$  e risco  $\lambda(t)$  para a Distribuição Gama Generalizada são dadas respectivamente por:

$$\begin{aligned} S(t) &= P(T \geq t) \\ &= 1 - P(T < t) \\ &= 1 - \int_0^t \frac{r}{\Gamma(\mu)\sigma^{r\mu}} s^{r\mu-1} \exp\left\{-\left(\frac{s}{\sigma}\right)^r\right\} ds \end{aligned}$$

e

$$\lambda(t) = \frac{f(t)}{S(t)} = \frac{t^{r\mu-1}}{S(t)}.$$

onde  $F(t)$  é a função de distribuição Gama Generalizada acumulada.

### 2.4.13 Estimação dos Coeficientes dos Modelos de Sobrevivência

Em análise de sobrevivência utiliza-se o método de máxima verossimilhança que consiste em encontrar o valor do vector de parâmetros  $\theta$  que maximize a função de verossimilhança. Esse estimador é uma ótima opção pois incorpora as censuras, é relativamente simples e possui ótimas propriedades para grandes amostras (Colosimo & Giolo, 2006).

De início, considera-se uma amostra de observações  $t_1, t_2, \dots, t_n$  da população, em que todas as observações são não censuradas. A função de verossimilhança para o vector de parâmetros  $\theta$  é expressa por

$$L(\theta) = \prod_{i=1}^n f(t_i, \theta), \tag{2.26}$$

em que  $f(t_i)$  é a função densidade de probabilidade.

Agora, é preciso dividir as observações entre as censuradas e as não censuradas, uma vez que os dados considerados são dados de tempos de vida. Ou seja, sejam  $r$  observações não censuradas,

$(1, 2, \dots, r)$  e  $n - r$  observações censuradas,  $(r + 1, r + 2, \dots, n)$ . Assim, a expressão para a função de verossimilhança que abrange todos os tipos de censura é expressa por:

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

em que  $S(t_i)$  é a função de sobrevivência e  $\delta_i$  é a variável indicadora de falha.

É comum trabalhar com o logaritmo da função de verossimilhança que é dado por

$$\ell(\theta) = \sum_{i \in F} \ln[f(y_i)] + \sum_{i \in C} \ln[S(y_i)], \quad (2.28)$$

em que  $F$  denota o conjunto de observações não censuradas e  $C$  o conjunto de observações censuradas.

#### 2.4.14 Comparação de modelos

Conforme [Campos \(2016\)](#), a escolha do modelo que descreve de forma adequada um conjunto de dados de sobrevivência é uma etapa muito importante da análise. Se o modelo utilizado não é adequado aos dados os resultados podem não retratar correctamente a realidade, invalidando toda a análise realizada.

#### Critério de informação de Akaike

De acordo com [Akaike \(1973\)](#), na validação do desempenho dos modelos, o Critério de informação de Akaike tem sido bastante utilizado para diferentes estruturas de covariâncias. O autor relacionou a discrepância, que existe entre o modelo verdadeiro e o modelo aproximado, com a máxima verossimilhança, que é o que possui melhor ajuste para os dados observados.

Ainda segundo [Akaike \(1973\)](#), o princípio do Critério de informação de Akaike é seleccionar uma combinação de variáveis exploratórias o modelo para a função de correlação que minimize o valor do AIC. Ou seja, quanto menor o valor de AIC melhor será o ajuste do modelo.

Este critério irá penalizar a log verossimilhança de cada modelo, de forma a representar o número de parâmetros que estão a ser estimados, permitindo a comparação de modelos. A estimativa do AIC para um determinado modelo é dada por:

$$AIC = -2l(\hat{\theta}) + 2k \quad (2.29)$$

onde,  $l(\hat{\theta})$  é o logaritmo da função verosimilhança do modelo e  $k$  é número de parâmetros do modelo .

## Teste da razão de verosimilhanças

Segundo [Carvalho et al. \(2011\)](#), o teste da razão de verosimilhança (RV) é um teste usado para comparar modelos aninhados, isto é, modelo geral e obtenção do valor do logaritmo de sua função de verosimilhança  $\log L(\hat{\theta}_G)$  e o modelo reduzido com menos parâmetros do que o geral e obtenção do valor do logaritmo da sua função de verosimilhança  $\log L(\hat{\theta}_R)$ . Baseado na função de verosimilhança, esse teste envolve a comparação dos valores do logaritmo da função de verosimilhança maximizada sem restrição e sob a  $H_0$ . A estatística é dada por

$$RV = -2[\log L(\hat{\theta}_G) - \log L(\hat{\theta}_R)], \quad (2.30)$$

A estatística RV segue assintoticamente a distribuição  $\chi^2$  com número de graus de liberdade igual à diferença no número de covariáveis dos modelos.

O modelo proposto não será adequado quando for rejeitada a  $H_0$ , ou seja, quando o  $p$  - valor for menor que o nível de significância estabelecido, geralmente 5%, se  $p$ - valor  $> 0,05$ , não há evidências estatísticas a 5% de que o modelo proposto não seja adequado.

## Método gráfico do ajuste do modelo

De acordo com [Carvalho et al. \(2011\)](#), o método gráfico é uma técnica auxiliar na escolha da família de distribuições paramétricas. É portanto uma ferramenta exploratória para a escolha de possíveis modelos.

Segundo [Colosimo & Giolo \(2006\)](#), sugerem dois métodos gráficos para auxiliar na escolha dos modelos candidatos, onde o método 1 consiste na comparação da função de sobrevivência do modelo proposto com o estimador de Kaplan-Meier e o método 2 consiste na linearização da função de sobrevivência.

Segundo [R. O. Santos \(2017\)](#), o segundo procedimento baseia-se na ideia de linearização da função de tal forma que se o modelo proposto for adequado o resultado será uma recta. Alterações na linearidade, é um indicativo de que o modelo sugerido não se ajusta adequadamente aos dados.

No caso do modelo de Weibull, espera-se encontrar rectas paralelas, e no caso do modelo exponencial espera-se que a inclinação das rectas seja igual a um. Pode-se também fazer um gráfico dos valores estimados de sobrevivência via Kaplan-Meier versus os valores estimados de sobrevivência pelas distribuições paramétricas. Uma recta traçada no ângulo de 45 graus indica um bom

ajuste entre os modelos.

De acordo com [Carvalho et al. \(2011\)](#), outra análise gráfica útil consiste em comparar as curvas de sobrevivência geradas pelo estimador não paramétrico, de Kaplan-Meier com as estimadas parametricamente. Quanto mais próximo o modelo paramétrico estiver da curva de sobrevivência do estimador de Kaplan-Meier, melhor o ajuste do modelo. Os métodos gráficos apresentados nesta secção, devem ser usados para descartar modelos claramente inapropriados, e não para demonstrar que um modelo paramétrico é melhor, pois, para isso a análise de resíduos seria mais apropriada.

## Medida Global de Qualidade do Ajuste

Segundo [Carvalho et al. \(2011\)](#), para avaliar a qualidade do ajuste de um modelo paramétrico, pode-se utilizar uma medida global de qualidade do ajuste denominada função desvio (ou deviance). Em que, a hipótese nula de que o modelo se ajusta aos dados pode ser testada pela estatística da função desvio

$$D = 2 \left[ l(\tilde{\theta})_G - l(\hat{\theta})_R \right], \quad (2.31)$$

onde  $l(\hat{\theta})_R$  e  $l(\tilde{\theta})_G$  são, o logaritmo da função de verosimilhança do modelo reduzido nas estimativas de máxima verosimilhança e o logaritmo da função de verosimilhança do modelo geral.

Um modelo geral é aquele que ajusta  $n$  parâmetros para todas  $n$  observações. Para calcular a deviance, os parâmetros de perturbação são mantidos fixos em ambos modelos. Por parâmetros de perturbação entende-se, aqueles comuns aos dois ajustes, e nos quais não há interesse no estimador. Quanto maior for a deviance do modelo, pior a qualidade do ajuste.

Ainda segundo [Carvalho et al. \(2011\)](#), a estatística  $D$  segue assintoticamente, a distribuição qui-quadrado ( $\chi^2$ ) com  $n - p - 1$  graus de liberdade, ou seja, o número de observações menos o número de covariáveis do modelo menos um. Nesse caso, quanto menor a estatística, maior o  $p$ -valor e melhor a qualidade de ajuste do modelo.

### 2.4.15 Análise de Resíduos

De acordo com [Colosimo & Giolo \(2006\)](#), uma avaliação da adequação do modelo ajustado é parte fundamental da análise dos dados. No modelo de regressão linear usual, uma análise gráfica dos resíduos é usada para esta finalidade. Nos modelos de regressão para análise de sobrevivência, a definição de resíduos não é tão clara e, desse modo, diversos resíduos têm sido propostos na literatura para obter o ajuste do modelo.

Técnicas gráficas, que fazem o uso dos diferentes resíduos propostos são, em particular, bastante utilizadas para examinar diferentes aspectos do modelo. Um desses aspectos é o de avaliar, por meio dos resíduos, a distribuição dos erros. Os resíduos de Cox-Snell e resíduos martingale devem

ser utilizadas como um meio de rejeitar modelos claramente inapropriados e não para "provar" que um particular modelo paramétrico está correcto, mesmo porque, em muitas aplicações, dois ou mais modelos paramétricos podem fornecer ajustes razoáveis bem como estimativas similares das quantidades de interesse.

## Resíduo de Cox-Snell

Na análise de sobrevivência, a interpretação dos resíduos possui relevância na verificação do ajuste do modelo proposto. Nesta etapa duas propriedades devem ser observadas: a distribuição e aleatoriedade dos resíduos.

Conforme [Gonçalves \(2020\)](#), os resíduos de Cox-Snell são usualmente utilizados para averiguar o ajustamento global do modelo de tempo de vida acelerado. Estes resíduos são definidos por

$$\hat{e}_i = \hat{\Lambda}(t_i | \mathbf{x}_i),$$

onde para o  $i$ -ésimo indivíduo,  $t_i$  é o instante,  $x_i$  o vector de covariáveis e  $\hat{\Lambda}$  a função de risco acumulado estimado para o modelo.

Segundo [Colosimo & Giolo \(2006\)](#), quando o modelo possui bom ajuste é possível verificar que o gráfico dos resíduos de Cox-Snell contra o logaritmo negativo da função de sobrevivência empírica construída pelo estimador Kaplan-Meier gera uma recta com um angulo de  $45^\circ$ .

## Resíduos Martingale

Segundo [Campos \(2016\)](#), os resíduos martingale  $m_i$ , para modelos paramétricos onde todas as covariáveis são fixas, são baseados no processo de contagem individual e definidos por:

$$\hat{m}_i = \hat{\delta}_i - \hat{e}_i, \quad (2.32)$$

onde  $\hat{\delta}_i$  é a variável indicadora de censura e  $\hat{e}_i$  os resíduos de Cox-Snell. Esses resíduos, são vistos como uma estimativa do número de falhas em excesso observadas nos dados mas não predito pelo modelo. São usados, em geral, para examinar a melhor forma funcional por exemplo linear, quadrática, etc, para uma dada covariável em um modelo de regressão assumido para os dados sob estudo. Se a curva suavizada obtida no gráfico: resíduo  $\times$  variável for linear, nenhuma transformação na variável é necessária, caso contrário pode-se fazer a inclusão de um termo quadrático da covariável no modelo ou sugerir alguma transformação da mesma como por exemplo a transformação log-log.

# CAPÍTULO 3

## MATERIAL E MÉTODOS

### 3.1 Classificação de Pesquisa

#### 3.1.1 Quanto à Natureza

Quanto à natureza a pesquisa é aplicada. Segundo [Nascimento & Sousa \(2016\)](#) pesquisa aplicada dedica-se na geração de conhecimento sólido para resolução de um determinado problema. Ela é preticamente aplicado em problemas práticos, Por exemplo, estudar o efeito dos estilos de liderança no clima organizacional em certa empresa para melhorar as relações interpessoais no ambiente de trabalho.

#### 3.1.2 Quanto à abordagem

Quanto à abordagem a pesquisa é quantitativa. Conforme [Richardson \(1999\)](#), a abordagem quantitativa é caracterizada pelo emprego da quantificação tanto nas modalidades de recolha de informações quanto no tratamento delas por meio de técnicas estatísticas. Neste estudo, pretende-se estimar modelos estatísticos, tais como curvas de sobrevivência, para avaliar o tempo de inserção no mercado de trabalho entre graduados, o que confirma a orientação quantitativa da investigação.

#### 3.1.3 Quanto aos objectivos

Com relação aos seus objectivos, a presente pesquisa é de natureza explicativa, pois busca identificar os factores determinantes que influenciam o tempo de empregabilidade dos graduados universitários em Moçambique. Conforme [Gil \(2008\)](#), a pesquisa explicativa tem como principal finalidade esclarecer os factores que contribuem para a ocorrência dos fenômenos, indo além da simples descrição ou análise exploratória. Neste sentido, o estudo busca compreender a relação entre variáveis como escolaridade, região, rede de contatos, entre outras, com o tempo necessário para a inserção no mercado de trabalho

### 3.1.4 Quanto aos procedimentos

Quanto aos procedimentos técnicos, a pesquisa caracteriza-se como um levantamento. De acordo com Lakatos & Marconi (2010), o levantamento é uma investigação directa dos fenômenos mediante a interrogação das pessoas cujos comportamentos se deseja conhecer. Assim, serão utilizados questionários estruturados para recolha de dados junto a graduados universitários, com o intuito de descrever suas características e identificar padrões associados à empregabilidade.

## 3.2 População, amostra e amostragem

Neste estudo, a população-alvo é composta por estudantes finalistas universitários de 2017, totalizando aproximadamente 17.977 graduados de seis universidades: UP, UEM, USTM, A Politécnica, UCM (campus da Beira), e UniZambeze. A amostra base foi realizada entre março e novembro de 2017, em Maputo e Beira, cobrindo 27 faculdades e 106 cursos, com 2.174 finalistas entrevistados (1.024 mulheres e 1.150 homens). De seguida incluiu-se seis rondas de inquéritos telefónicos trimestrais entre março de 2018 e setembro de 2019, com um acompanhamento de até 2.100 participantes. Portanto, dos 2.100 participantes da baseline foram variando ao longo das rondas, de 2.051 na Ronda 1 até 1.841 na Ronda 6. Objectivo foi de analisar sua trajectória de transição da universidade para o mercado de trabalho. O acompanhamento longitudinal permitiu observar aspectos como inserção profissional, condições de emprego, evolução da carreira e factores que influenciam a empregabilidade dos graduados.

Para esse estudo, foi extraído um subconjunto da base de dados contendo 11.673 observações, que representam os registos válidos e consistentes utilizados no presente estudo. Essa restrição se deve a critérios de qualidade dos dados, como a remoção de duplicações, preenchimento incompleto de variáveis-chave e inconsistências entre as rondas.

## 3.3 Material

Utilizou-se uma base de dados secundária, referente a transição de escola para o mercado de trabalho dos graduados universitários em Moçambique. A base de dados foi obtida mediante um pedido do mesmo no Centro de Estudo de Economia e Gestão da Universidade Eduardo Mondlane (CEEG-UEM) através do link disponível online<sup>1</sup>. Na Tabela 3.1, são apresentadas as 7 covariáveis do estudo que foram utilizadas nas técnicas de Análise de Sobrevivência. A variável de interesse seleccionada foi o tempo de empregabilidade, ou seja, o tempo até conseguir o primeiro emprego, que é definido como o período desde a graduação até o momento em que o graduado obteve o primeiro emprego .

---

<sup>1</sup><https://igmozambique.wider.unu.edu/database/school-work-transition-higher-education-and-technical-and-vocational-education-and>

As covariáveis, como idade do graduado, curso de graduação, e instituição de ensino, foram categorizadas conforme as recomendações de estudos anteriores e práticas comuns na análise de dados educacionais e laborais. Para a classificação das covariáveis nível de educação dos pais, participação em atividades extracurriculares, uso de plataformas de emprego, e tempo de procura de emprego semanal, foram utilizados os mesmos pontos de corte observados em outros estudos similares.

Para o processamento dos dados utilizou-se o programa R versão 4.0.4. O relatório será produzido com recurso ao programa  $\text{\LaTeX}$ .

Tabela 3.1: Descrição das variáveis em estudo

Variáveis	Descrição	Codificação
Idade	Idade do graduado	1-18-22
		2-23-25
		3-26-55
Sexo	Sexo do graduado	0-Masculino
		1-Feminino
Estado_civil	Es casado (a)?	0-Não
		1-Sim
		1-Cabo Delgado
		2-Gaza
		3-Inhambane
		4-Manica
Província	Qual é a provincia de residência?	5-Maputo Cidade
		6-Maputo Provincia
		7-Nampula
		8-Niassa
		9-Sofala
		10-Tete
		11-Zambezia
		1-Educação
		2-Línguas e Humanitárias
		3-Ciências Sociais
		4-Ciências Naturais
Curso	Nome do curso de graduação	5-Engenharia
		6-Agricultura
		7-Saúde
		1-UEM
		2-UCM
		3-UNIZAMBEZE
		4-USTM
Universidade	Nome da universidade	5-UP
		6-A POLITECNICA
Nota final	Nota final de conclusão de curso	

### 3.4 Métodos

Neste trabalho foram aplicados métodos não paramétricos, como Kaplan-Meier, Log-rank, Gehan-Breslow e Taranoe-Ware devido a sua importante para descrever os dados de sobrevivência. Essas técnicas, não permitem a inclusão de covariáveis, o que torna os resultados/as análise não muito robustos. Portanto, para analisar os efeitos dessas covariáveis, é necessário a aplicação dos modelos paramétricos de sobrevivência. Dessa forma, foram usados os modelos exponencial, weibull e log-normal (são os mais usados pela sua simplicidade(Colosimo & Giolo, 2006)), gama, gama

generalizado (como detalhado na secção 2.4). Todas as hipóteses foram testadas a um nível de significância de 5% e para efeitos de avaliação da regra de decisão, usou-se o p-valor associado à estatística do teste.

### Modelo de Regressão Exponencial

Para analisar a relação entre as covariáveis e o tempo até o evento, utilizou-se o modelo de regressão exponencial. A sua forma logaritimizado é dada da seguinte forma:

$$Y = x\beta + W \quad (3.1)$$

Onde o  $x = (1, x_1, \dots, x_{p-1})$  é o vector de covariáveis,  $\beta$  é o vector dos parâmetros a serem estimados e,  $W$  tem distribuição de valor extremo padrão. De acordo com Campos (2016), as funções densidade de probabilidade e de sobrevivência de  $Y$  dado  $x$  é dada pelas seguintes equações:

$$f(y/x) = \exp[y - x\beta - \exp(y - x\beta)] \quad (3.2)$$

e

$$s(y/x) = \exp[-\exp(y - x\beta)] \quad (3.3)$$

### Modelo de Regressão Weibull

Além do modelo exponencial, ajustou-se também o modelo de regressão weibull que é considerado o caso particular do modelo exponencial, e também é o mais simples para descrever o tempo de falha. Além disso, o modelo tem ampla aplicabilidade, pois a sua função taxa de falha é monótona. O modelo weibull é considerado o modelo exponencial linearizado, com o termo de erro seguindo uma distribuição normal com variância  $\sigma^2$ . O modelo é representado da seguinte forma:

$$Y = \ln(t) = x\beta + \sigma W \quad (3.4)$$

As funções densidade de probabilidade e de sobrevivência de  $Y$  dado  $x$  são expressas da seguinte forma:

$$f(y/x) = \frac{1}{\sigma} \exp\left[\frac{(y - x\beta)}{\sigma} - \exp\left(\frac{(y - x\beta)}{\sigma}\right)\right] \quad (3.5)$$

e

$$s(y/x) = \frac{1}{\sigma} \exp\left[-\exp\left(\frac{(y - x\beta)}{\sigma}\right)\right] \quad (3.6)$$

### Modelo de Regressão Log-Normal

Adicionalmente, considerou-se o modelo log-normal adequado para situações em que os tempos seguem uma distribuição assimétrica, matematicamente é representado por

$$T \sim \text{LogNorm}(\mu, \sigma) \quad (3.7)$$

em que os parâmetros  $\mu$  e  $\sigma$ , são a média e o desvio padrão, respectivamente. A sua função densidade de probabilidade e sua função de sobrevivência é representada da seguinte forma

$$f(t) = \frac{1}{\sqrt{2\pi t\sigma}} e^{-\frac{(\ln(t)-\mu)^2}{2\sigma^2}}, t > 0 \quad (3.8)$$

e a função distribuição de probabilidade é dada por

$$F(t) = \Phi\left(\frac{\ln t - \mu}{\sigma}\right), t > 0, \quad (3.9)$$

onde  $\Phi$  representa a função de distribuição normal padrão. Portanto, a função de sobrevivência é expressa pela seguinte expressão

$$s(t) = \Phi\left(\frac{\mu - \ln t}{\sigma}\right), t > 0, \quad (3.10)$$

A função taxa de risco é dada pela seguinte relação 3.8 e 3.10

$$\lambda(t) = \frac{f(t)}{s(t)} \quad (3.11)$$

O valor esperado e a variância são dados respectivamente por

$$E[T] = e^{\mu + \frac{\sigma^2}{2}}, \text{var}(T) = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1), \quad (3.12)$$

# CAPÍTULO 4

## RESULTADOS E DISCUSSÃO

### 4.1 Caracterização da amostra em estudo

De acordo com a Figura 4.1, que apresenta a distribuição dos graduados por sexo e faixa etária, os dados indicam que a amostra é composta predominantemente por graduados do sexo masculino, correspondendo a 54,8%. Em relação à idade, observa-se uma predominância de graduados mais jovens, com 35,5% na faixa de 18 a 22 anos, seguidos por 33,3% na faixa de 23 a 25 anos. Por fim, os graduados na faixa etária de 26 a 55 anos representam 31,3%, evidenciando a presença não significativa desse grupo na amostra.

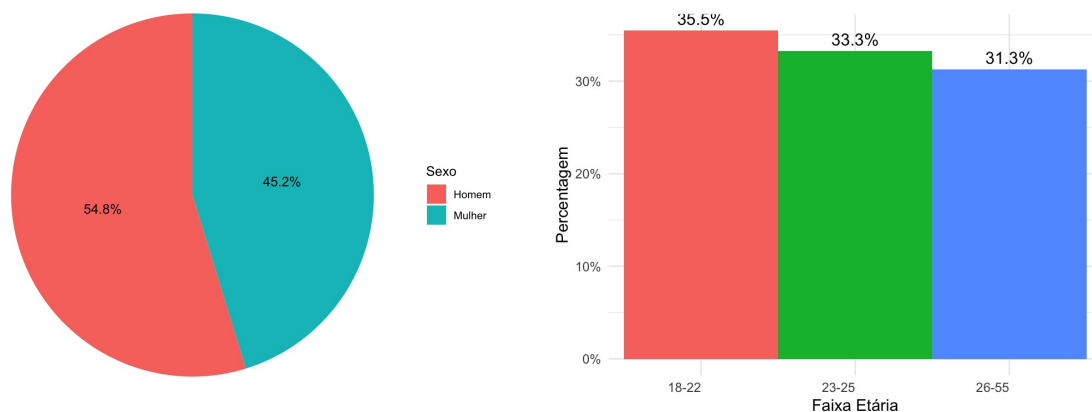


Figura 4.1: Distribuição dos graduados por sexo e por idade

### 4.2 Resultados da Estimação não Paramétrica

Na Figura 4.2, observa-se a curva de sobrevivência que ilustra o tempo até a obtenção do primeiro emprego pelos graduados universitários. A curva inicia com uma probabilidade de sobrevivência de 100% (1,0), indicando que, no início do estudo, nenhum dos graduados havia conseguido o primeiro emprego. Conforme o tempo avança, a curva apresenta uma redução gradativa, sugerindo que alguns graduados obtiveram o primeiro emprego ao longo das rondas (trimestral).

Ao final do período de análise (ronda 6), a probabilidade de sobrevivência alcança cerca de 75%, o que indica que apenas 25% dos graduados conseguiram o primeiro emprego. Por outro lado, a maioria (75%) ainda não havia entrado no mercado de trabalho até o final da última ronda.

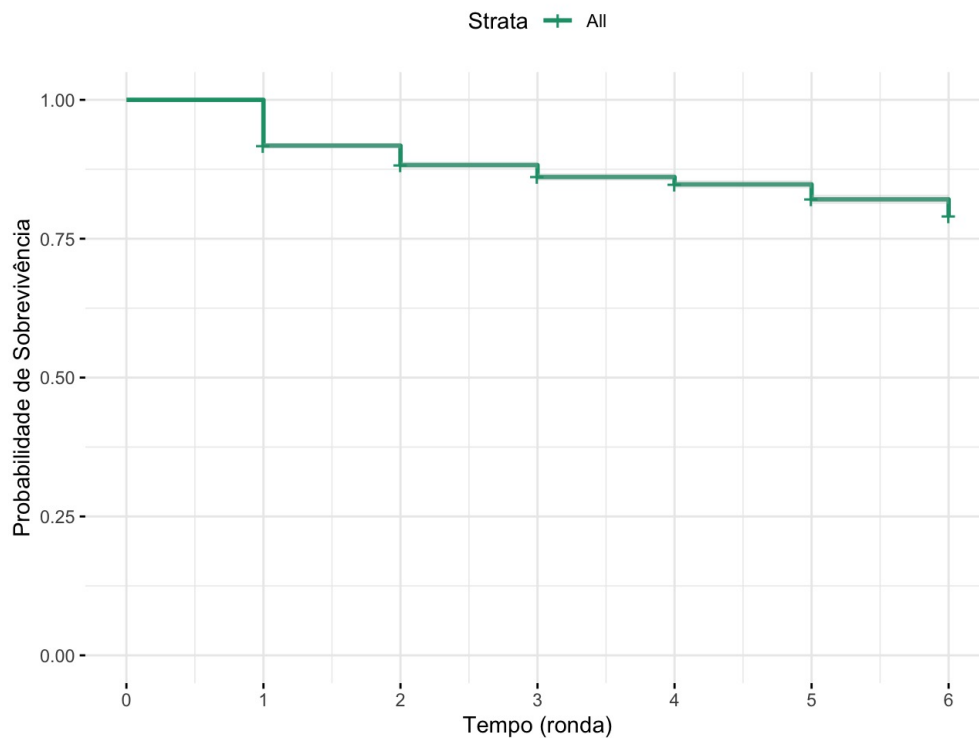


Figura 4.2: Curva de sobrevivência global para os graduados

Na Figura 4.3, observa-se curva de sobrevivência que descreve o tempo até do primeiro emprego por sexo e região, qualquer que seja a ronda considerada, a probabilidade do tempo até obtenção do primeiro emprego é sempre superior para o grupos das mulheres comparativamente aos dos homens, com exceção da ronda 6, onde a sua probabilidade é aproximadamente igual. Além disso, observamos que não existe diferença significativa entre as curvas de sobrevivência ao nível de significância de 5%. Para a região de residência, observou-se também que as curvas de sobrevivência não diferem entre si, ou seja, independente da sua região de residência a probabilidade do tempo até obtenção do primeiro empregado é aproximadamente igual.

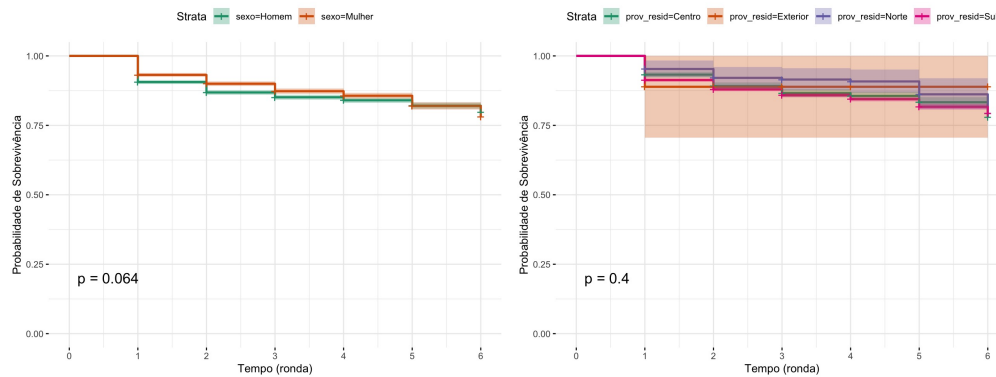


Figura 4.3: Curvas da função de sobrevivência estimadas por Kaplan-Meier para os factores Sexo e Região

De acordo com a Tabela 4.1, observa-se que houve diferença significativa entre o sexo dos graduados, indicando que existe diferença no tempo de empregabilidade, seu p-valor é menor do que o nível de significância de 5%, essa diferença foi notório na estatística de Gehan-Breslow e Tarone-Ware. Para os graduados casados e não casados, a sua curva de sobrevivência diferem entre si. Não houve diferenças significativas para as variáveis Tipo de universidade, area de estudo e região, ou seja, as sua curvas de sobrevivência não diferem entre si.

Tabela 4.1: Estatísticas de comparação das curvas de sobrevivência

	log-rank	Gehan-Breslow	Tarone-Ware
Sexo	-1.853(0.064)	3.735(0.000)	3.003 (0.003)
Casado	-1.290 (0.197)	2.781 (0.005)	2.151 (0.032)
Tipo da universidade	-2.493 (0.869)	5.185 (0.394)	2.381 (0.794)
Area de estudo	-2.493 (0.869)	10.25 (0.114)	6.225 (0.399)
Região	-3.183 (0.364)	7.011 (0.072)	5.574 (0.134)

## 4.3 Estimação dos coeficientes do modelo que ajusta os dados

### 4.3.1 Selecção do Modelo

Na Tabela 4.2 foi aplicado o critério de informação de Akaike e Bayesiabo, de modo a verificar qual modelo probabilístico melhor explica o tempo de sobrevivência até obtenção do primeiro emprego, pelo valor de critério de informação de Akaike e Baysiano, constatou-se que o modelo Log-Normal é o melhor modelo que se ajusta aos dados comparativamente aos modelo de Weibull e modelo de Exponencial, visto que apresenta o menor valor de AIC e BIC

Tabela 4.2: Critério de informação de Akaike e Critério de informação Bayesiano para selecção dos modelos paramétricos de sobrevivência

	AIC	BIC
Modelo Log-Normal	2038.307	2157.694
Modelo Weibull	2046.186	2165.573
Modelo Exponencial	2252.808	2366.225

Após a selecção do melhor modelo de sobrevivência, estimou-se os coeficientes do modelo com base nas variáveis em estudo. Na Tabela 4.3 são apresentadas as estimativas, erro padrão e os valores de probabilidade obtidos para o modelo Log-Normal ajustado através do pacote GAMLSS. Os parâmetros foram seleccionados considerando-se um nível de significância de 5%.

Tabela 4.3: Estimativas dos coeficientes do Modelo Log-Normal

	$\hat{\beta}$	$\exp(\hat{\beta})$	Desv. Padrão	Z-valor	P-valor
Constante	2.342	0.096	0.385	6.090	<b>0.000</b>
Sexo (Mulher)	-0.096	1.101	0.057	-1.690	0.091
Idade	0.027	0.973	0.007	3.740	<b>0.000</b>
Casado(a)	0.002	0.998	0.102	0.020	0.985
Curso:Línguas e ciências humanas	0.127	0.880	0.138	0.930	0.355
Curso:Ciências Sociais	0.088	0.916	0.089	0.980	0.325
Curso:Ciências Naturais	0.291	0.748	0.125	2.320	<b>0.020</b>
Curso:Engenharia	0.148	0.862	0.118	1.260	0.208
Curso:Agricultura	0.145	0.865	0.142	1.020	0.306
Curso:Saúde	0.135	0.874	0.151	0.890	0.373
Universidade:UCM	-0.384	1.469	0.163	-2.350	<b>0.019</b>
Universidade:UNIZAMBEZE	-0.143	1.154	0.133	-1.080	0.282
Universidade:USTM	0.070	0.932	0.125	0.560	0.575
Universidade:UP	0.231	0.793	0.089	2.610	<b>0.009</b>
Universidade:A POLITECNIA	0.063	0.939	0.167	0.380	0.707
Região: Exterior	2.586	0.075	293.8	0.010	0.993
Região: Norte	0.019	0.981	0.161	0.120	0.904
Região: Sul	-0.144	1.154	0.115	-1.240	0.214
Média Final	-0.033	1.034	0.021	-1.590	0.111

### 4.3.2 Interpretação dos Parâmetros do Modelo Log-Normal

Para o modelo seleccionado, foram obtidas as estimativas, como a Tabela 4.3 ilustra, verifica-se que as covariáveis idade, curso e universidade são significativas a um nível de significância de 5%.

De acordo com a Tabela 4.3, sendo o coeficiente do factor idade  $\hat{\beta}_2 = 0.027$ , pode-se esperar que para cada um ano que aumenta de idade o risco de um graduado obter seu primeiro emprego aumenta em  $\exp(0.027) = 0.973$  vezes, mantendo-se constante os restantes factores.

Controlando o efeito dos restantes factores, pode-se esperar que a taxa de risco em obter o primeiro emprego a um graduado que cursou ciências naturais é  $\exp(0.291) = 0.748$  vezes maior em relação a um graduado que cursou educação.

Em relação a universidade, quer dizer que a taxa de risco de obter o primeiro emprego um graduado que frequentou a Universidade Católica de Moçambique(UCM) é  $\exp(-0.384) = 1.469$  vezes menor em relação a um graduado que frequentou Universidade Eduardo Mondlane(UEM), mantendo-se outros factores constante. Por outro lado, observa-se que frequentou a Universidade Pedagógica (UP) a chance aumenta em 0.793 vezes mais em relação ao da UEM.

### 4.3.3 Análise dos Resíduos do Modelo Log-Normal

Para avaliar a qualidade de ajuste global do modelo Log-Normal foram obtidos os gráficos dos resíduos padronizados  $e_i^*$  estimados pelo método de Kaplan-Meier e pelo modelo exponencial (Figura 4.4) e a respectiva curva de sobrevivência, os gráficos dos resíduos de Cox-Snell estimados pelo método de Kaplan-Meier e pelo modelo exponencial padrão e as respectivas curvas de sobrevivência (Figura 4.5).

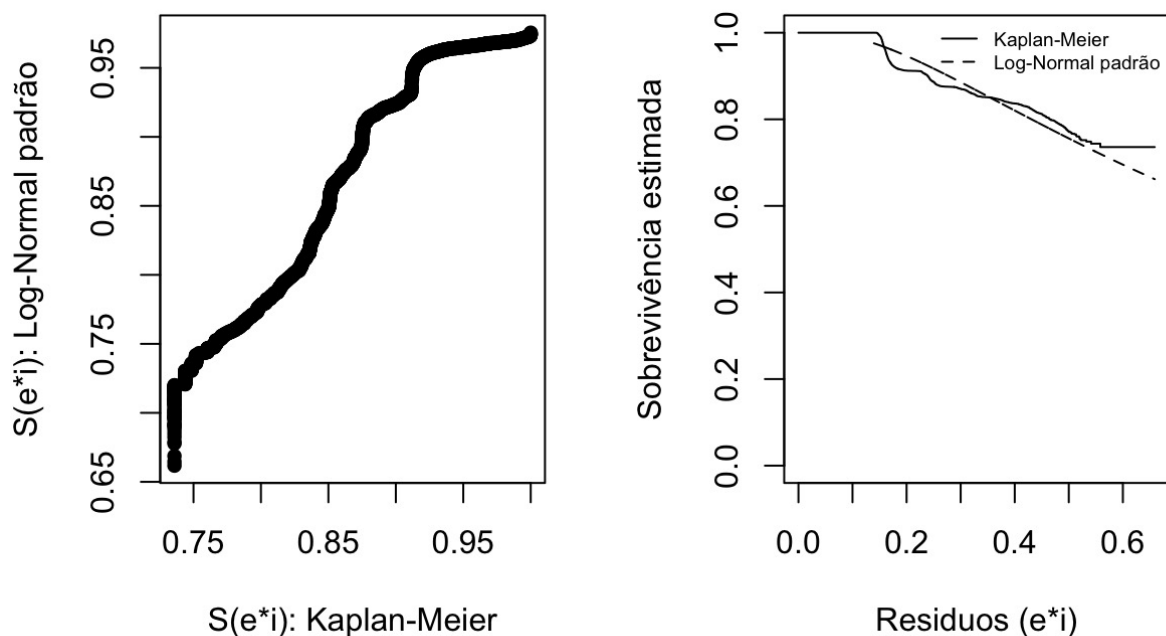


Figura 4.4: Resíduos padronizados  $e_i^*$  estimados pelo método de Kaplan-Meier e pelo modelo Log-Normal e a respectiva curva de sobrevivência estimada.

Em ambas as Figuras 4.4, pode-se perceber que os resíduos padronizados apresentam comporta-

mento mais próximo de uma recta, indicando assim um bom ajuste desta distribuição em relação a exponencial que apresenta uma curvatura maior.

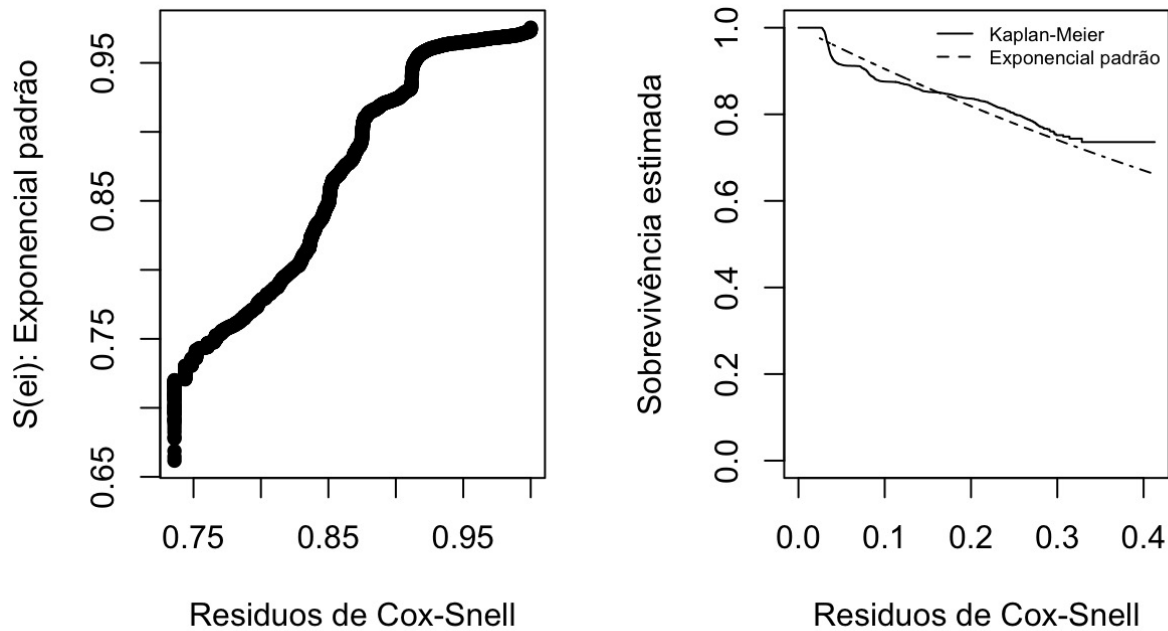


Figura 4.5: Resíduos de Cox-Snell estimados pelo método de Kaplan-Meier e pelo modelo exponencial padrão e a respectiva curva de sobrevivência.

A Figura 4.5, apresenta a função de sobrevivência do resíduo de Cox-Snell para o modelo Log-Normal com a função de sobrevivência da distribuição exponencial com média 1. Pode-se observar que a função de sobrevivência dos resíduos se ajustou bem à função de sobrevivência exponencial padrão, indicando um bom ajuste do modelo Log-Normal.

## 4.4 Discussão dos resultados

Este trabalho teve como objectivo principal analisar os factores determinantes do tempo de empregabilidade entre graduados universitário em Moçambique, verificou-se que muitos graduados enfrentam dificuldades significativas para encontrar o primeiro emprego. O trabalho destacou que apenas 25% dos graduados conseguiram seu primeiro emprego.

Quanto ao tempo de sobrevivência até a obtenção do primeiro emprego, verificou-se que os homens apresentam uma maior probabilidade de alcançar esse marco em comparação às mulheres, esses resultados são semelhantes com [Alemu & Yismaw \(2022\)](#). Essa diferença reflete possíveis desigualdades de género no mercado de trabalho, que podem estar associadas a factores culturais, sociais ou estruturais.

Em relação à região de residência, não foi identificada nenhuma diferença significativa no tempo de sobrevivência. Ou seja, morar nas regiões Centro, Norte, Sul ou na diáspora não influencia as chances de obter o primeiro emprego.

Por fim, ao testar as curvas de sobrevivência utilizando os testes de Log-rank, Gehan-Breslow e Tarone-Ware, os resultados mostraram uma consistência entre os métodos. Notou-se, em particular, que a curva de sobrevivência estratificada por sexo apresenta uma diferença estatisticamente significativa, corroborando a ideia de que o género é um factor relevante no tempo até a obtenção do primeiro emprego. Além disso, observou-se que casado(a) e tipo da universidade, as suas curvas mostram-se estatisticamente significativas.

Com base nos resultados obtidos pelos modelos paramétricos, verificou-se que usando o modelo log-normal, a variável resultado é explicada pelos três factores. Observou-se que ele apresenta melhores resultados, portanto, o modelo paramétrico log-normal é adequado para explicar, facilita a análise estatística dos dados, o que concorda com [Colosimo & Giolo \(2006\)](#).

Ao analisar os possíveis factores de risco associados ao tempo de sobrevivência dos graduados até a obtenção do primeiro emprego, os resultados do modelo log-normal indicaram que a idade do graduado, o curso e a universidade foram significativos.

Em relação a idade, o estudo verificou que graduados de idades mais avançadas apresentam maior chance em obter primeiro emprego, resultados similar ao de estudo feito por ([Alemu & Yismaw, 2022](#)). Além disso, os resultados do estudo mostram que graduado que frequentam ciências naturais tem maior chance de inserção no mercado de trabalho. Resultados semelhantes foram observados no estudo de [Mahdi et al. \(2018\)](#). [Wobse et al. \(2022\)](#) no seu estudo concluiu que a universidade de origem está associado a maior chance de empregabilidade, resultados similar ao deste estudo.

# CAPÍTULO 5

## CONCLUSÕES E RECOMENDAÇÕES

### 5.1 Conclusões

O trabalho teve como objectivo principal analisar os factores determinante do tempo de empregabilidade entre graduado universitários em Moçambique. Com os resultados do estudo concluiu-se que:

- Dos 11 673 graduados em estudo, verificou-se que 54.8% eram do sexo masculino contra 45.2% do sexo feminino, onde maioritadamente eram da faixa etária de 18 à 22 anos de idade.
- O estudo identificou como factores determinante no tempo de empregabilidade dos graduados universitários idade, curso de ciências naturais e a universidade pedagogica. Como factores de proteção no tempo de empregabilidade o sexo feminino e a universidade catolica de Moçambique.
- Com base no modelo estimado obteve-se os seguintes resultados: os homens apresentam uma maior probabilidade de obter primeiro emprego em comparação às mulheres.
- E quanto a região de residência, não foi identificada nenhuma diferença significativa no tempo de sobrevivência, ou seja, independentemente da região de residência, o tempo até obtenção do primeiro emprego foi aproximadamente igual.

### 5.2 Recomendações

Diante dos resultados encontrados no presente estudo recomenda-se:

- Programas de estágios e mentoria específicos para mulheres e Políticas de sensibilização sobre igualdade de género em recrutamento.
- Informação sobre oportunidades de trabalho para os finalistas universitários e de reforçar a transparência nos processos de recrutamento.

- Sugere-se a realização de cursos na área de Ciências Naturais e a criação de programas regionais de emprego jovem.

# REFERÊNCIAS

- Akaike, H. (1973). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. doi: 10.1109/TAC.1974.1100705
- Alemu, M., & Yismaw, B. (2022). Understanding waiting time from graduation to first employment: Survival analysis-based evidence from ethiopia. *Education Research International*, 2022, 1-11. Retrieved from <https://doi.org/10.1155/2022/9001679> doi: 10.1155/2022/9001679
- Alves, N. (2009). *Inserção profissional e formas identitárias: o caso dos licenciados da universidade de lisboa*. Educa.
- Anchisi, M. (2011). *Métodos não paramétricos para análise de dados de sobrevivência* (Dissertação de Mestrado). Universidade de Lisboa, Lisboa.
- Andrade, A. B., & Cabete, B. (2012). *Análise de sobrevivência com acontecimentos múltiplos: Aplicação ao estudo do tempo até à ocorrência de enfarte do miocárdio* (Dissertação de Mestrado). Universidade de Lisboa.
- Baum, S., & Payea, K. (2004). *Education pays: The benefits of higher education for individuals and society*. College Board.
- Bourdieu, P. (1986). The forms of capital. In *Handbook of theory and research for the sociology of education*.
- Campos, J. D. S. (2016). *Factores explicativos para a evasão no ensino superior através da análise de sobrevivência: o caso da ufpe* (Dissertação de Mestrado). Universidade Federal de Pernambuco.
- Carvalho, M. S., Andreozzi, V. L., Codeço, C. T., Campos, D. P., Barbosa, M. T. S., & Shimamura, S. E. (2011). *Análise de sobrevivência: teoria e aplicações em saúde*. SciELO-Editora FIOCRUZ.
- Colosimo, E. A., & Giolo, S. R. (2006). *Análise de sobrevivência aplicada*. Edgard Blucher.
- Fragoso, A., Valadas, S. T., & Paulos, L. (2019). Ensino superior e empregabilidade: percepções de estudantes e graduados, empregadores e académicos. *Educação & Sociedade*, 40, e0186612.

- Gil, A. C. (2008). *Métodos e técnicas de pesquisa social* (6th ed.). São Paulo: Atlas.
- Gonçalves, D. O. (2020). *Modelos paramétricos de sobrevivência aplicados a dados de câncer* (Dissertação de Mestrado). Universidade Federal de Alfenas-MG, Alfenas, MG, Brasil.
- Instituto Nacional de Estatística – Moçambique, I. (2022). *Relatório final do inquérito sobre orçamento familiar – iof 2022* (Tech. Rep.). INE.
- Lakatos, E. M., & Marconi, M. d. A. (2010). *Fundamentos de metodologia científica* (7th ed.). São Paulo: Atlas.
- Magalhães, M. N. (2006). *Probabilidade e variáveis aleatórias* (2nd ed.). EDUSP.
- Mahdi, S., Massingue, A. C. A., Massarongo Chivulele, F. A. P., Baez Ramirez, J. E., Walker, I. D., Mucavele Macule, A. N., ... Casal, J. (2018). *Mozambique economic update: Shifting to more inclusive growth* (Tech. Rep.). The World Bank.
- Medola, E., Fox, L., et al. (2020). *Transição para o mercado de trabalho dos estudantes universitários em moçambique* (Tech. Rep.). International Growth Centre (IGC). Retrieved from <https://www.theigc.org/sites/default/files/2020-02/Medola-et-al-2020-policy-brief-Mozambique-PT.pdf>
- Monteiro, A., Gonçalves, C., & Santos, P. J. (2019). Jovens: Do ensino superior para o mercado de trabalho.
- Nascimento, F. P. d., & Sousa, F. (2016). Classificação da pesquisa. natureza, método ou abordagem metodológica, objetivos e procedimentos. *Metodologia da Pesquisa Científica: teoria e prática—como elaborar TCC. Brasília: Thesaurus.*
- Neves, S., & Lima, F. (2010). Transição escola–mercado de trabalho: duração da procura do 1º emprego. *INE, Estatísticas do Emprego, 2º T.*
- Oliveira, E. L. d., Rios-Neto, E. G., & Oliveira, A. M. H. C. d. (2006). Transições dos jovens para o mercado de trabalho, primeiro filho e saída da escola: o caso brasileiro. *Revista brasileira de Estudos de População, 23*, 109–127.
- Richardson, R. J. (1999). *Pesquisa social: Métodos e técnicas*. São Paulo: Atlas.
- Rueda, F. J. M., Martins, L. J., & Campos, K. C. L. (2004). Empregabilidade: o que os alunos universitários entendem sobre isto? *Revista Psicologia: Teoria e Prática, 6*(2), 63–73.
- Santos, R., Jones, S., & Xirinda, G. (2024). *Informed job entry: Does labour market information speed job-taking in mozambique?* (WIDER Working Paper 2024/16). UNU–WIDER. Retrieved from <https://www.wider.unu.edu/sites/default/files/Publications/Working-paper/PDF/wp2024-16-informed-job-entry-Mozambique.pdf>  
doi: 10.35188/UNU-WIDER/2024/474-8

- Santos, R. O. (2017). *Formação de doutores para atividades de caráter acadêmico via modelo de riscos proporcionais de cox e regressão logística* (Dissertação de Mestrado). Universidade de Brasília.
- Serranho, P. (2015). Notas de análise de sobrevivência: uma introdução com r.
- Sousa, G. C. V. d. (2019). *Modelos de regressão em análise de sobrevivência: uma aplicação na modelagem do tempo de vida de micrurus corallinus em cativeiro* (Unpublished master's thesis). Universidade de São Paulo, Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba. (Versão revisada de acordo com a resolução CoPGr 6018 de 2011)
- Tomlinson, M. (2012). Graduate employability: A review of conceptual and empirical themes. *Higher education policy*, 25, 407–431.
- Wobse, B. A., Menuta, Y. H., & Liga, A. D. (2022). Modeling of waiting time to first employment of graduates at wolkite university, ethiopia: Application of accelerated failure time models. *Cogent Education*, 9(1), 2143032. Retrieved from <https://doi.org/10.1080/2331186X.2022.2143032> doi: 10.1080/2331186X.2022.2143032
- Yorke, M., & Knight, P. T. (2006). Curricula for economic and social gain. *Higher Education*, 51, 565–588.

# Apêndices

## Outros resultados

Tabela 5.1: Proporção de estudantes graduados por área de estudo e sexo

Área de Formação	M	H	M+H
Educação	14.6%	16.2%	30.8%
Letras e Humanidades	0.7%	0.9%	1.6%
Ciências sociais, gestão e direito	20.9%	23.6%	44.5%
Ciências naturais	1.4%	2.6%	4.0%
Engenharias, indústrias e construção	2.5%	5.4%	7.9%
Agricultura	0.9%	1.2%	2.1%
Saúde e bem-estar	2.7%	3.1%	5.7%
Serviços	0.7%	2.7%	3.4%
<b>Total</b>	<b>44.3%</b>	<b>55.7%</b>	<b>100.0%</b>

Tabela 5.2: Estimativas dos coeficientes do Modelo Exponencial

	$\hat{\beta}$	$\exp(\hat{\beta})$	Desv. padrão	Z	P-valor
Constante	3.190	0.041	0.973	3.280	<b>0.001</b>
sexoMulher	-0.291	1.338	0.140	-2.080	<b>0.038</b>
Idade	0.075	0.928	0.020	3.820	<b>0.000</b>
Casado(a)	0.004	0.996	0.264	0.020	0.987
Curso:Línguas e ciências humanas	0.189	0.828	0.338	0.560	0.576
Curso:Ciências Sociais	0.215	0.806	0.225	0.960	0.339
Curso:Ciências Naturais	0.614	0.541	0.321	1.910	0.056
Curso:Engenharia	0.420	0.657	0.309	1.360	0.174
Curso:Agricultura	0.330	0.719	0.354	0.930	0.350
Curso:Saúde	0.313	0.731	0.359	0.870	0.384
Universidade:UCM	-0.837	2.309	0.381	-2.200	<b>0.028</b>
Universidade:UNIZAMBEZE	-0.387	1.473	0.324	-1.190	0.232
Universidade:USTM	-0.137	1.147	0.288	-0.480	0.634
Universidade:UP	0.427	0.652	0.232	1.840	0.066
Universidade:A POLITECNICA	0.059	0.942	0.407	0.150	0.884
Região:Exterior	12.00	0.000	1810	0.010	0.995
Região:Norte	-0.140	1.150	0.369	-0.380	0.705
Região:Sul	-0.331	1.392	0.278	-1.190	0.234
Média Final	-0.066	1.068	0.052	-1.270	0.206

Tabela 5.3: Estimativas dos coeficientes do Modelo Weibull

	$\hat{\beta}$	Desv. Padrão	Z-valor	P-valor
Constante	2.090	0.367	5.71	<b>0.000</b>
Sexo (Mulher)	-0.112	0.052	-2.15	<b>0.031</b>
Idade	0.028	0.008	3.73	<b>0.000</b>
Casado(a)	-0.007	0.099	-0.07	0.943
Curso:Línguas e ciências humanas	0.096	0.125	0.76	0.446
Curso:Ciências Sociais	0.095	0.084	1.14	0.254
Curso:Ciências Naturais	0.260	0.120	2.17	<b>0.030</b>
Curso:Engenharia	0.157	0.115	1.360	0.173
Curso:Agricultura	0.139	0.131	1.060	0.289
Curso:Saúde	0.086	0.130	0.660	0.509
Universidade: UCM	-0.286	0.140	-2.05	<b>0.040</b>
Universidade:UNIZAMBEZE	-0.134	0.120	-1.12	0.264
Universidade: USTM	-0.0296	0.107	-0.28	0.782
Universidade:UP	0.193	0.086	2.240	<b>0.025</b>
Universidade: A POLITECNICA	0.022	0.151	0.150	0.882
Região: Exterior	6.740	15100	0.000	0.100
Região: Norte	-0.018	0.137	-0.13	0.897
Região: Sul	-0.114	0.103	-1.110	0.267
Média Final	-0.027	0.0197	-1.350	0.178
Log(escala)	-0.992	0.056	-17.610	<b>0.000</b>

## CÓDIGO R

```
#import dataset

library(haven)
data_report <- read_dta("Desktop/All info/UEM//Monografia/base de dados/d

# cleaning data
library(dplyr)
attach(data_report)

data_clean<-data_report %>% select(id,round_main,work_status,
female, age, married, prov_resid, std_area_id, university,
english, job_search, experience, firstjob)

attach(data_clean)

data_clean$firstjob<-ifelse(is.na(data_clean$firstjob), 0,
```

```

ifelse(data_clean$firstjob==1, 1, 0))

data_clean<-data_clean %>% filter(data_clean$round_main<=6)

data_clean$prov_resid<-ifelse(data_clean$prov_resid %in% c(2, 3, 5, 6),
"Sul", ifelse(data_clean$prov_resid %in% c(4, 9, 10, 11), "Centro",
ifelse(data_clean$prov_resid %in% c(7, 8, 1), "Norte", "Exterior")))

# caracterização da amostra

library(ggplot2)

#install.packages("data.table")
library(data.table)

write.table(data_clean, file = "writetable.data.csv")

# survival analyze
# tempo até conseguir o primeiro emprego
library(survival)
library(survminer)

km_employee <- survfit(Surv(round_main, firstjob) ~ 1,
data = na.omit(data_clean))

ggsurvplot(
km_employee,
data = na.omit(data_clean),
conf.int = TRUE,          # Exibe intervalo de confiança
xlab = "Tempo (ronda)",   # Renomeia o eixo X
ylab = "Probabilidade de Sobrevivência",
palette = "Dark2",       # Escolhe uma paleta de cores
ggtheme = theme_minimal() # Aplica um tema minimalista
)

# tempo até conseguir o primeiro emprego por sexo

data_clean<-data_clean %>% mutate(female=ifelse(female==1,

```

```

"Mulher", "Homem"))

data_clean<-data_clean %>% rename(female, sexo=female)

sex <- survfit(Surv(round_main, firstjob) ~ sexo,
data = na.omit(data_clean))

ggsurvplot(
sex,
data = na.omit(data_clean),
pval = TRUE, conf.int = TRUE,
xlab = "Tempo (ronda)", # Renomeia o eixo X
ylab = "Probabilidade de Sobrevida",
palette = "Dark2",
ggtheme = theme_minimal()
)

region <- survfit(Surv(round_main, firstjob) ~ prov_resid,
data = na.omit(data_clean))

ggsurvplot(
region,
data = na.omit(data_clean),
pval = TRUE, conf.int = TRUE,
xlab = "Tempo (ronda)", # Renomeia o eixo X
ylab = "Probabilidade de Sobrevida",
palette = "Dark2",
ggtheme = theme_minimal()
)

# comparação de curva de sobrevivência
library(survival)
library(coin)

attach(data_clean)
logrank_test(Surv(round_main, firstjob) ~ factor(sexo),
type = "logrank")
logrank_test(Surv(round_main, firstjob) ~ factor(sexo),

```

```

type = "Gehan-Breslow")
logrank_test(Surv(round_main, firstjob) ~ factor(sexo),
type = "Tarone-Ware")

logrank_test(Surv(round_main, firstjob) ~ factor(married),
type = "logrank")
logrank_test(Surv(round_main, firstjob) ~ factor(married),
type = "Gehan-Breslow")
logrank_test(Surv(round_main, firstjob) ~ factor(married),
type = "Tarone-Ware")

logrank_test(Surv(round_main, firstjob) ~ factor(university),
type = "logrank")
logrank_test(Surv(round_main, firstjob) ~ factor(university),
type = "Gehan-Breslow")
logrank_test(Surv(round_main, firstjob) ~ factor(university),
type = "Tarone-Ware")

logrank_test(Surv(round_main, firstjob) ~ factor(std_area_id),
type = "logrank")
logrank_test(Surv(round_main, firstjob) ~ factor(std_area_id),
type = "Gehan-Breslow")
logrank_test(Surv(round_main, firstjob) ~ factor(std_area_id),
type = "Tarone-Ware")

logrank_test(Surv(round_main, firstjob) ~ factor(prov_resid),
type = "logrank")
logrank_test(Surv(round_main, firstjob) ~ factor(prov_resid),
type = "Gehan-Breslow")
logrank_test(Surv(round_main, firstjob) ~ factor(prov_resid),
type = "Tarone-Ware")

# modelos parametricos
library(gamlss)
library(gamlss.cens)

Modeloexp<-gamlss(Surv(round_main, firstjob) ~ sexo + age+ married+
as.factor(std_area_id) + as.factor(university)+english+job_search+
as.factor(prov_resid), data=na.omit(data_clean),

```

```

family=cens(EXP, type="right"))
summary(Modeloexp)

ModeloLOGNO<-gamlss(Surv(round_main, firstjob) ~ sexo + age+ married+
as.factor(std_area_id) + as.factor(university)+english+job_search+
as.factor(prov_resid), data=na.omit(data_clean),
family=cens(LOGNO, type="right"))
summary(ModeloLOGNO)

ModeloWei<-gamlss(Surv(round_main, firstjob) ~ sexo + age+ married+
as.factor(std_area_id) + as.factor(university)+english+job_search+
as.factor(prov_resid), data=na.omit(data_clean),
family=cens(WEI3, type="right"))
summary(ModeloWei)

# Selecao do Modelo( Criterio AIC)
AIC(Modeloexp, ModeloLOGNO, ModeloWei)
BIC(Modeloexp, ModeloLOGNO, ModeloWei)

# Valores ajustados do modelo Weibull
fitted_WEI <- predict(ModeloLOGNO, type = "response")

# Gráfico de resíduos vs. valores ajustados
plot(fitted_WEI, resid_deviance_WEI, main = "Resíduos vs.
Valores Ajustados (Log-Normal)",
xlab = "Valores Ajustados", ylab = "Resíduos Deviance",
col = "blue", pch = 20)
abline(h = 0, col = "red", lwd = 2)

#====New analysi====#

library(flexsurv)

# exponencial
fit_exp <- flexsurvreg(Surv(round_main,firstjob)~1, dist='exponential',
data = data_clean)
fit_exp

```

```

lambda<-1/fit_exp$res[1]

# LogNormal

fit_log<-flexsurvreg(Surv(round_main, firstjob)~1, dist = "lognorm",
data = data_clean)
fit_log

# weibull
fit_wei <- flexsurvreg(Surv(round_main,firstjob)~1, dist='weibull',
data = data_clean)
fit_wei

lambda = fit_wei$res[1]
alpha = fit_wei$res[2]

# modelo gama
fit_gamma <- flexsurvreg(Surv(round_main, firstjob)~1, dist = 'gamma',
data = data_clean)
fit_gamma

# modelo gama
fit_gamma <- flexsurvreg(Surv(round_main, firstjob)~1, dist = 'gamma',
data = data_clean)
fit_gamma

# modelo generalizado (gama generalizada)
fit_gammagen <- flexsurvreg(Surv(round_main, firstjob)~1,
dist = 'gengamma', data = data_clean)
fit_gammagen

# selecção do modelo
Modelo = c("Gama Generalizado", "Exponencial", "Log-Normal", "Weibull",
"gamma")

```

```

Verossimilhanca = c(fit_gammagen$loglik, fit_exp$loglik, fit_log$loglik,
TRV = 2*(fit_gammagen$loglik-Verossimilhanca)
valor_p = pchisq(TRV,df=2,lower.tail=FALSE) %>% round(2)

resultado = data.frame(Modelo=Modelo,
Verossimilhanca = Verossimilhanca,
TRV=TRV,
valor_p=valor_p)

ggpubr::ggtexttable(resultado, rows=NULL)

# Modelos com regressores

# exponencial
fit_exp <- flexsurvreg(Surv(round_main,firstjob)~sexo + age+english,
dist='exponential', data = data_clean)
fit_exp

# LogNormal

fit_log<-flexsurvreg(Surv(round_main, firstjob)~sexo + age+english,
dist = "lognorm", data = data_clean)
fit_log

# weibull

fit_wei <- flexsurvreg(Surv(round_main,firstjob)~sexo + age+english,
dist='weibull', data = data_clean)
fit_wei

# modelo gama

fit_gamma <- flexsurvreg(Surv(round_main, firstjob)~sexo + age+english,
dist = 'gamma', data = data_clean)
fit_gamma

# modelo gama

fit_gamma <- flexsurvreg(Surv(round_main, firstjob)~sexo + age+english,

```

```

dist = 'gamma', data = data_clean)
fit_gamma

# modelo generalizado (gama generalizada)
fit_gammagen <- flexsurvreg(Surv(round_main, firstjob)~sexo + age+
english, dist = 'gengamma', data = data_clean)
fit_gammagen

# selecao do modelo

Modelo = c("Gama Generalizado", "Exponencial",
"Log-Normal", "Weibull", "gamma")
Verossimilhanca = c(fit_gammagen$loglik, fit_exp$loglik, fit_log$loglik,
fit_wei$loglik, fit_gamma$loglik)
TRV = 2*(fit_gammagen$loglik-Verossimilhanca)
valor_p = pchisq(TRV,df=2,lower.tail=FALSE) %>% round(2)

resultado = data.frame(Modelo=Modelo,
Verossimilhanca = Verossimilhanca,
TRV=TRV,
valor_p=valor_p)

resultado

#=====Adequação do modelo ajustado=====#

# determinação de residuos ei* (Log-normal padrão)

reg_Log<-survreg(Surv(round_main, firstjob)~sexo + age+english,
dist="lognorm", data=data_clean)

summary(reg_Log)

residuos<-predict(reg_Log, type = "response")
padr.log<-(log(round_main)-log(residuos))/reg_Log$scale
padr.orig<-exp(padr.log)
padr.km<-survfit(Surv(padr.orig, firstjob)~1)

```

```

res<-padr.km$time
sln<-pnorm(-log(res))

# win.graph(width=12, height=6)

par(mfrow=c(1,2))

plot(padr.km$surv, sln, pch=16, ylab = "S(e*i): Log-Normal padrão",
xlab = "S(e*i): Kaplan-Meier")
plot(padr.km, conf.int = F, mark.time = F, pch=16, xlab = "Resíduos (e*i)
, ylab = "Sobrevivência estimada")
lines(res, sln, lty=2)
legenda=c("Kaplan-Meier", "Log-Normal padrão")
legend(x="topright",legend=legenda, lty=1:2, cex=0.6, bty="n")
par(mfrow=c(1,1))

# determinação de resíduos de cox-snell

snell<--log(1-pnorm(padr.log))
snell.km<-survfit(Surv(snell,firstjob)~1)
t<-snell.km$time
s.t<-snell.km$surv
s.exp<-exp(-t)

par(mfrow=c(1,2))

plot(s.t, s.exp, pch=16, ylab = "S(ei): Exponencial padrão",
xlab = "Resíduos de Cox-Snell")
plot(snell.km, conf.int = F, mark.time = F, pch=16, xlab = "Resíduos de
Cox-Snell", ylab = "Sobrevivência estimada")
lines(t, s.exp, lty=4)
legenda=c("Kaplan-Meier", "Exponencial padrão")
legend(x="topright",legend=legenda, lty=1:2, cex=0.6, bty="n")
par(mfrow=c(1,1))

```