



**UEPB**

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

**VITORIA SOARES DE SOUZA**

**MÉTODOS DE ANÁLISE DE SOBREVIVÊNCIA APLICADOS A DADOS DE  
PACIENTES COM CÂNCER DE OVÁRIO NO ESTADO DA PARAÍBA**

**CAMPINA GRANDE - PB  
2024**

VITORIA SOARES DE SOUZA

**MÉTODOS DE ANÁLISE DE SOBREVIVÊNCIA APLICADOS A DADOS DE  
PACIENTES COM CÂNCER DE OVÁRIO NO ESTADO DA PARAÍBA**

Trabalho de Conclusão de Curso (Artigo) apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de bacharel em Estatística.

**Orientador:** Prof. Dr. Tiago Almeida de Oliveira.  
**Coorientador:** Profa. Ma. Débora de Sousa Cordeiro

**CAMPINA GRANDE - PB  
2024**

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

S729m Souza, Vitoria Soares de.

Métodos de análise de sobrevivência aplicados a dados de pacientes com câncer de ovário no estado da Paraíba [manuscrito] / Vitoria Soares de Souza. - 2024.

36 f. : il. color.

Digitado.

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

"Orientação : Prof. Dr. Tiago Almeida de Oliveira, Departamento de Estatística - CCT".

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

1. Câncer de ovário. 2. Sobrevida. 3. Fatores de risco. 4. Análise estatística. I. Título

21. ed. CDD 310

VITORIA SOARES DE SOUZA

MÉTODOS DE ANÁLISE DE SOBREVIVÊNCIA APLICADOS A DADOS DE PACIENTES  
COM CÂNCER DE OVÁRIO NO ESTADO DA PARAÍBA

Trabalho de Conclusão de Curso (Artigo) apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de bacharel em Estatística.

Trabalho aprovado em 22 de Novembro de 2024.

**BANCA EXAMINADORA**



---

Prof. Dr. Tiago Almeida de Oliveira  
(Orientador)  
Universidade Estadual da Paraíba (UEPB)



---

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



---

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

Dedico este trabalho, primeiramente, a Deus, à minha família, professores e amigos, que estiveram ao meu lado ao longo desta trajetória e contribuíram para o meu crescimento acadêmico.

“Quando for a hora certa, Eu o Senhor seu Deus, farei acontecer.”  
(Isaias 60:22)

## LISTA DE ILUSTRAÇÕES

Figura 1	– Ilustração do avanço do câncer de ovário. . . . .	11
Figura 2	– Exemplo de censura do Tipo I com (•) indicando a falha e (◊) a censura. . .	13
Figura 3	– Exemplo de censura do Tipo II com (•) indicando a falha e (◊) a censura. . .	13
Figura 4	– Exemplo de censura aleatória com (•) indicando a falha e (◊) a censura. . .	14
Figura 5	– Frequência dos casos de câncer de ovário segundo a Faixa Etária, Raça/Cor, Meio de diagnóstico e Extensão. . . . .	24
Figura 6	– Curva de sobrevivência de <i>Kaplan-Meier</i> por covariáveis. . . . .	25
Figura 7	– Análise dos resíduos de <i>Cox-Snell</i> das estimativas de sobrevivência dos modelos de regressão em comparação com <i>Kaplan-Meier</i> . . . . .	28
Figura 8	– Gráfico comparativo das estimativas de sobrevivência de <i>Kaplan-Meier</i> com as estimativas dos modelos Exponencial, Weibull e Log-normal. . . . .	28

## LISTA DE TABELAS

Tabela 1	– Descrição das variáveis utilizadas no estudo. . . . .	10
Tabela 2	– Contingências observadas em $t_j$ . . . . .	16
Tabela 3	– Frequência de tipo de óbito. . . . .	22
Tabela 4	– Medidas descritivas dos dados de câncer de ovário no estado da Paraíba. . .	23
Tabela 5	– Descrição da Morfologia e Frequências. . . . .	23
Tabela 6	– Aplicação dos testes <i>Log-rank</i> e <i>Wilcoxon</i> à covariável faixa etária. . . . .	26
Tabela 7	– Aplicação dos testes <i>Log-rank</i> e <i>Wilcoxon</i> à covariável raça/cor. . . . .	26
Tabela 8	– Aplicação dos testes <i>Log-rank</i> e <i>Wilcoxon</i> à covariável meio de diagnóstico. . . . .	27
Tabela 9	– Aplicação dos testes <i>Log-rank</i> e <i>Wilcoxon</i> à covariável extensão. . . . .	27
Tabela 10	– Log-verossimilhança, Testes de Razão de Verossimilhança e Critério de Informação de Akaike para Modelos de Distribuição. . . . .	29
Tabela 11	– Modelo de regressão paramétrico ajustado por covariáveis através da distribuição Weibull. . . . .	29

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> . . . . .	<b>9</b>
<b>2</b>	<b>MATERIAIS E MÉTODOS</b> . . . . .	<b>10</b>
2.1	<b>Material</b> . . . . .	<b>10</b>
2.2	<b>Métodos</b> . . . . .	<b>11</b>
2.2.1	<i>Conceitos de Análise de Sobrevida</i> . . . . .	<b>12</b>
2.2.2	<i>Censura</i> . . . . .	<b>12</b>
2.2.3	<i>Função de Sobrevida</i> . . . . .	<b>13</b>
2.2.4	<i>Função densidade de probabilidade</i> . . . . .	<b>14</b>
2.2.5	<i>Função de Risco ou Taxa de Falha</i> . . . . .	<b>14</b>
2.3	<b>Técnicas Não-Paramétricas</b> . . . . .	<b>15</b>
2.3.1	<i>Kaplan-Meier</i> . . . . .	<b>15</b>
2.3.2	<i>Teste Log-rank</i> . . . . .	<b>16</b>
2.3.3	<i>Teste de Wilcoxon</i> . . . . .	<b>17</b>
2.4	<b>Modelos Paramétricos</b> . . . . .	<b>17</b>
2.4.1	<i>Distribuição Exponencial</i> . . . . .	<b>18</b>
2.4.2	<i>Distribuição Weibull</i> . . . . .	<b>18</b>
2.4.3	<i>Distribuição Log-Normal</i> . . . . .	<b>19</b>
2.4.4	<i>Modelos Paramétricos de Regressão</i> . . . . .	<b>19</b>
2.4.5	<i>Adequação dos Modelos ajustados</i> . . . . .	<b>20</b>
2.4.6	<i>Estimativa dos Parâmetros dos Modelos</i> . . . . .	<b>21</b>
2.4.6.1	<i>Método da Máxima Verossimilhança</i> . . . . .	<b>21</b>
2.4.6.2	<i>Teste da Razão de Verossimilhança</i> . . . . .	<b>22</b>
<b>3</b>	<b>RESULTADOS E DISCUSSÃO</b> . . . . .	<b>22</b>
<b>4</b>	<b>CONCLUSÃO</b> . . . . .	<b>30</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>31</b>
	<b>APÊNDICE A – DESCRIÇÕES DAS MORFOLOGIAS</b> . . . . .	<b>34</b>

## MÉTODOS DE ANÁLISE DE SOBREVIVÊNCIA APLICADOS A DADOS DE PACIENTES COM CÂNCER DE OVÁRIO NO ESTADO DA PARAÍBA

Vitoria Soares de Souza\*  
 Tiago Almeida de Oliveira†  
 Débora de Sousa Cordeiro‡

### RESUMO

Este estudo investigou o tempo de sobrevivência de pacientes com câncer de ovário no estado da Paraíba, considerando fatores como idade, raça/cor, meio de diagnóstico e extensão da doença. A amostra incluiu 146 pacientes, dos quais 72,00% faleceram devido ao câncer, enquanto 28,00% tiveram casos censurados. A mediana de idade foi de 57,5 anos, com maior risco em mulheres a partir dos 50 anos ou mais. A média de sobrevida geral incluindo falhas e censuras foi de 516 dias, com mediana de 340 dias, indicando uma sobrevida curta para muitas pacientes. A maioria das pacientes tinha entre 31 e 60 anos aproximadamente 48,6%, e 63,7% se autodeclararam negras ou pardas, sugerindo desigualdades no acesso ao diagnóstico e tratamento. O tipo de tumor mais comum foi o adenocarcinoma, e 93,8% dos diagnósticos foram feitos por exames histológicos. As curvas de sobrevivência, analisadas pelo método *Kaplan-Meier*, mostraram que mulheres mais jovens e brancas apresentaram melhor taxa de sobrevivência. A análise estatística revelou que a faixa etária não teve impacto significativo, mas a raça/cor foi um fator relevante, com diferenças acentuadas devido a desigualdades no acesso a diagnóstico e tratamento entre os grupos. Segundo o critério AIC, o modelo Weibull foi o mais adequado para ajustar o modelo de regressão paramétrico. A presença de metástase se mostrou o principal fator de pior prognóstico no câncer de ovário, aumentando em cerca de 8 vezes de morte no tratamento. A falta de informações sobre a extensão do câncer também eleva o risco, possivelmente devido a diagnósticos tardios ou tratamentos inadequados. Pacientes com mais de 60 anos apresentaram uma redução de 16% no risco de falha, indicando uma resposta mais positiva ao tratamento. Os resultados destacam a importância do diagnóstico precoce e o impacto de fatores socioeconômicos e culturais nas desigualdades de sobrevida no câncer de ovário na Paraíba.

**Palavras-chaves:** câncer de ovário; sobrevida; fatores de risco; desigualdades de saúde.

### ABSTRACT

This study investigated the survival time of ovarian cancer patients in the state of Paraíba, considering factors such as age, race/color, diagnosis method, and disease extent. The sample included 146 patients, of whom 72.00% died due to cancer, while 28.00% had censored cases. The median age was 57.5 years, with higher risk in women aged 50 years or older. The overall survival mean, including failures and censures, was 516 days, with a median of 340 days, indicating short survival for many patients. Most patients were between 31 and 60 years old (approximately 48.6%), and 63.7% self-identified as Black or Brown, suggesting inequalities in access to diagnosis and treatment. The most common tumor type was adenocarcinoma, and 93.8% of diagnoses were made through histological exams. Survival curves, analyzed by the Kaplan-Meier method, showed that younger and White women had better survival rates. Statistical analysis revealed that age had no significant impact, but race/color was a relevant factor,

\* Vitoria Soares de Souza, Depto de Estatística, UEPB, Campina Grande, PB, vitoria.soares.souza@aluno.uepb.edu.br

† Prof. Dr. Tiago Almeida de Oliveira, Depto de Estatística, UEPB, Campina Grande, PB, tiagoestatistico@gmail.com

‡ Prof. Ma. Débora de Sousa Cordeiro, Depto de Estatística, UEPB, Campina Grande, PB, debora.cordeiro@servidor.uepb.edu.br

with pronounced differences due to inequalities in access to diagnosis and treatment among the groups. According to the AIC criterion, the Weibull model was the most appropriate for fitting the parametric regression model. The presence of metastasis was the main factor of worse prognosis in ovarian cancer, increasing the death risk by approximately 8 times during treatment. The lack of information on cancer extent also raised the risk, possibly due to late diagnoses or inadequate treatments. Patients over 60 years old showed a 16% reduction in failure risk, indicating a more positive response to treatment. The results highlight the importance of early diagnosis and the impact of socioeconomic and cultural factors on survival inequalities in ovarian cancer in Paraíba.

**Keywords:** ovarian cancer; survival; risk factors; Health disparities.

## 1 INTRODUÇÃO

O corpo humano é uma estrutura complexa composta por vários sistemas que trabalham em conjunto para manter a vida. No entanto, o corpo feminino possui diferenças significativas, exclusivas devido ao seu papel na reprodução e ao sistema hormonal (CONSOLARO; MARIA-ENGLER, 2014).

Os ovários, por sua vez, são órgãos localizados na região pélvica, encontrados um de cada lado e sustentados por ligamentos, que se conectam ao útero através das tubas uterinas. Geralmente, apresentam cerca de 3cm de comprimento, 1,5cm de largura e 1cm de espessura, com formato amendoado (SANTOS, 2024). Dentre as importantes funções que os ovários desempenham, destacam-se a produção de óvulos, estrogênio e progesterona, essenciais para as características sexuais secundárias (como o crescimento das mamas, o alargamento dos quadris, o crescimento de pelos púbicos, o ciclo menstrual, entre outros).

Nesse sentido, as mulheres apresentam doenças exclusivas do seu gênero, como o câncer de ovário, que é a segunda neoplasia ginecológica mais comum entre elas, ficando atrás apenas do câncer de colo de útero. Também conhecido como carcinoma ovariano, é um tipo de câncer que se forma nos tecidos dos ovários (SAÚDE, 2024). Esse câncer ocorre quando as células dos ovários sofrem mutações genéticas, resultando em crescimento descontrolado e formação de tumores malignos. Esses tumores podem se espalhar para outras partes do corpo, tornando o tratamento e a gestão da doença desafiadores. Segundo informações do Instituto Nacional do Câncer (INCA), o câncer de ovário é a oitava neoplasia mais comum entre as mulheres no Brasil. Estima-se que, em 2023, cerca de 7.310 mulheres serão diagnosticadas com essa doença no país (INCA, 2023).

Na fase inicial, o câncer de ovário não causa sintomas específicos. À medida que o tumor cresce, pode causar pressão, dor ou inchaço no abdômen, pelve, costas ou pernas, náusea, indigestão, gases, prisão de ventre, diarreia, falta de apetite, desejo de urinar com mais frequência, cansaço constante e/ou massa palpável no abdômen (ONCOGUIA, 2023). Dentre os fatores que aumentam os riscos de câncer de ovário, podem-se citar a idade, fatores reprodutivos, fatores hormonais, histórico familiar, fatores genéticos, terapia hormonal após a menopausa, tratamento para fertilidade, obesidade, entre outros (CRUZ, 2024).

É crucial enfatizar que o câncer de ovário é uma doença de alta gravidade, associada a uma taxa de sobrevivência relativamente baixa (FIOCRUZ, 2022). Nesse sentido, este trabalho busca estudar as pacientes acometidas pelo câncer de ovário no estado da Paraíba entre os anos de 2006 e 2020, utilizando análise de sobrevivência, com o objetivo de estimar o tempo (em dias) de sobrevida das mulheres desde o diagnóstico até o óbito. Além disso, será investigado se fatores como idade, raça/cor, meio de diagnóstico e extensão da doença afetam a sobrevida dessas pacientes. Para isso, serão comparadas as curvas de sobrevivência utilizando testes estatísticos

adequados para avaliar possíveis diferenças.

## 2 MATERIAIS E MÉTODOS

Nesta seção, serão explorados os recursos metodológicos empregados na criação deste trabalho. Será discutido a coleta e a descrição dos dados utilizados, os métodos estatísticos aplicados nas análises e como cada um desses métodos contribui para o avanço do estudo.

### 2.1 Material

A análise será realizada com base em um banco de dados público e gratuito fornecido pelo Instituto Nacional de Câncer (INCA), que inclui informações sobre mulheres do estado da Paraíba diagnosticadas com câncer de ovário. O conjunto de dados inclui informações de 146 pacientes e cobre o período desde o primeiro diagnóstico registrado em 2006 até 2020. O banco de dados original é composto por 17 variáveis, das quais foram selecionadas para o presente estudo apenas: idade, raça/cor, meio de diagnóstico, extensão, tempo em dias e status.

Para a interpretação dos dados coletados, foram aplicados métodos estatísticos utilizando o software R (R Core Team, 2023). Essa abordagem permitiu investigar tendências, identificar fatores de risco e compreender a dinâmica do câncer de ovário na Paraíba. O pacote empregado para o tratamento dos dados de sobrevivência foi o *survival* (THERNEAU, 2023).

Para compreender os dados, será realizada uma análise descritiva que inclui desde uma visão geral à aspectos específicos da sobrevivência (REIS; REIS, 2002). Será, então, estimado o tempo de vida das mulheres diagnosticadas com câncer de ovário na Paraíba, desde o diagnóstico até o óbito, utilizando técnicas de estimação não paramétricas. O estimador de Kaplan e Meier (1958) será utilizado para calcular as funções de sobrevivência, enquanto o teste *Log-rank* e *Wilcoxon* serão aplicados para a comparação das curvas de sobrevivência. As variáveis utilizadas estão detalhadas na Tabela 1.

Tabela 1 – Descrição das variáveis utilizadas no estudo.

Variável	Classificação
Idade	Até 30 anos; 31 a 60 anos; Acima de 60 anos.
Raça/Cor	Branca; Preta/parda; sem informação.
Meio de diagnóstico	Clínico; Histologia da metástase; H. do tumor primário; Sem informações.
Extensão	Localizado; Metástase; Sem informações.
Tempo em dias	1 a 3 846.
Status	Censura = 0; Óbito = 1.

Fonte: Elaborado pelo autor, (2024).

A distribuição etária dos casos de câncer de ovário será analisada por meio da estratificação dos dados em três categorias de idade, com base nas idades mínima e máxima observadas, que variam de 14 a 94 anos. As faixas etárias foram definidas como até 30 anos, de 31 a 60 anos e acima de 60 anos. Essa categorização facilita a comparação entre grupos etários, permitindo identificar possíveis diferenças na incidência do câncer de ovário e fornecendo subsídios para intervenções específicas conforme a idade.

Estudos epidemiológicos indicam que múltiplos fatores, incluindo determinantes socioeconômicos e características demográficas, como raça e cor da pele, podem influenciar tanto o risco de desenvolvimento quanto a resposta ao tratamento. No contexto do bando de dados, as

categorias de raça/cor comumente utilizadas — “Branca”, “Preta/Parda” e “Sem informação” — refletem a diversidade populacional do país e possibilitam análises direcionadas sobre possíveis disparidades em saúde (KABAD; BASTOS; SANTOS, 2012). A Figura 1 ilustra três estágios do câncer de ovário.

Figura 1 – Ilustração do avanço do câncer de ovário.



Fonte: Medimagen (2017)

No primeiro estágio, os ovários saudáveis funcionam adequadamente, apresentando células reprodutivas e hormonais em bom estado, sem indícios de câncer. No estágio de câncer localizado, as células malignas começam a se desenvolver nos ovários, frequentemente sem sintomas significativos, sendo detectadas em exames de rotina; os possíveis sinais incluem dor abdominal e alterações menstruais. O tratamento inicial geralmente envolve cirurgia, com a possibilidade de quimioterapia. Entretanto, no câncer metastático, a doença se espalha para órgãos como o abdômen, fígado e pulmões, causando sintomas mais intensos, como dor abdominal severa e perda de peso. O tratamento nessa fase é mais complexo, podendo incluir quimioterapia, terapia-alvo e imunoterapia (CONSCIENTE, 2024).

O tempo de sobrevida em dias foi calculado com base na diferença entre a data do óbito e a data do diagnóstico por câncer de ovário. Para a análise de sobrevida, focamos em casos de censura à direita. Os dados analisados foram de 1 a 3.846 dias (aproximadamente 10 anos).

A censura à direita ocorreu quando o evento (óbito) não foi observado até o final do estudo, seja porque a paciente estava viva ou porque não houve óbito registrado por câncer de ovário até o término do acompanhamento. Além disso, a censura aleatória foi considerada nos casos em que a paciente saiu do estudo sem apresentar falha, por motivos como desistência ou perda de contato.

Na variável “Status”, a ocorrência do óbito é indicada pelo valor “1” (falha), enquanto a censura, representando pacientes que não faleceram por câncer de ovário ou que foram retiradas do estudo, é indicada pelo valor “0”. Esse modelo permite uma análise focada no tempo até o evento, considerando adequadamente os dados censurados para reduzir viés nos resultados.

## 2.2 Métodos

Para a compreensão dos dados, foi realizada uma análise descritiva juntamente com técnicas de análise de sobrevivência em uma perspectiva geral. Utilizou-se o estimador *Kaplan-Meier* e o teste *log-rank* para comparar curvas de sobrevivência. Modelos paramétricos estimaram as funções de sobrevivência considerando distribuições de probabilidade, avaliados pelo teste

de razão de verossimilhança e Critério de Informação de *Akaike*(AIC). Por fim, aplicaram-se modelos de regressão paramétricos para identificar a influência das covariáveis estudadas.

### 2.2.1 *Conceitos de Análise de Sobrevida*

O estudo da sobrevivência é um campo de estudo amplamente aplicado na área da saúde, focado em examinar o período até que um determinado evento de interesse ocorra. Nesse contexto, é fundamental levar em conta dois aspectos principais: o tempo e a falha.

- O tempo: representa a variável dependente do modelo, sendo medido desde o início do estudo até a ocorrência do evento de interesse.
- A censura: ocorrem quando a informação sobre o tempo até o evento não está completamente disponível para todos os indivíduos. No âmbito clínico, isso pode incluir a perda de acompanhamento do paciente, o término do estudo sem que o paciente tenha vivenciado o evento de interesse ou a morte do paciente por motivos diferentes do estudo.

A censura ocorre nos dados de sobrevivência para o indivíduo  $i$ , em que  $i = 1, \dots, n$ , é geralmente representada pelo par  $(t_i, \delta_i)$ . Conforme descrito por Colosimo e Giolo (2006),

$$\delta_i = \begin{cases} 1, & \text{se ocorre a falha,} \\ 0, & \text{se ocorre a censura,} \end{cases}$$

em que,

- $t_i$  é o tempo até a falha ou censura ;
- $\delta_i$  é uma variável indicadora que denota se o evento foi uma falha (1) ou censura (0).

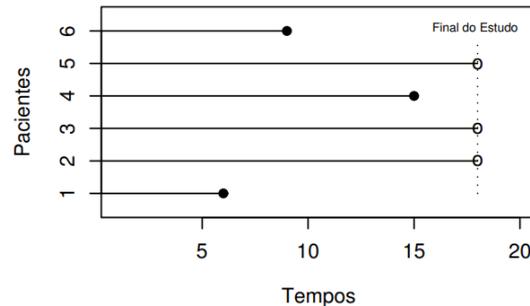
### 2.2.2 *Censura*

Os dados de sobrevivência têm como característica central a censura, que ocorre quando o evento de interesse não é completamente observado devido à interrupção do acompanhamento. Embora parcial, essa informação é relevante para uma análise mais precisa do fenômeno. A censura pode ocorrer de diversas formas, sendo os tipos mais comuns: censura à direita, à esquerda e intervalar (EUGENIO, 2021).

Existem três tipos principais de mecanismo de censura à direita, conforme descrito por (LAWLESS, 1982):

1. **Censura do tipo I:** acontece quando o período para o término do estudo é definido previamente, e as observações que não registraram a ocorrência da falha até esse momento são consideradas censuradas.

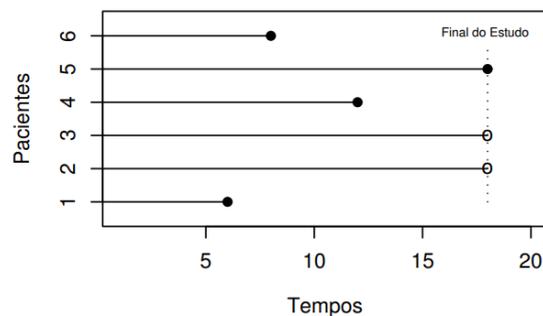
Figura 2 – Exemplo de censura do Tipo I com (•) indicando a falha e (○) a censura.



Fonte: Colosimo e Giolo (2006).

2. **Censura tipo II:** ocorre quando se estabelece previamente um número fixo de falhas a ser observado. Assim que esse número é alcançado, o estudo é encerrado, e as observações que não apresentaram falhas são classificadas como censuradas.

Figura 3 – Exemplo de censura do Tipo II com (•) indicando a falha e (○) a censura.



Fonte: Colosimo e Giolo (2006).

3. **Censura aleatória:** ocorre quando o elemento em análise é retirado do experimento devido à ocorrência de um evento diferente da falha de interesse. Esse tipo é comum em pesquisas médicas, quando um paciente é excluído do estudo sem apresentar a falha esperada ou falece por uma causa distinta da que está sendo investigada.

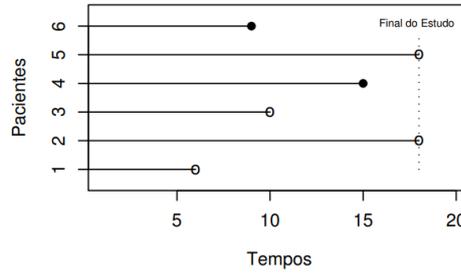
Esses tipos de mecanismo de censura representam casos de dados incompletos, permitindo uma maior análise das informações disponíveis.

### 2.2.3 Função de Sobrevivência

Segundo Colosimo e Giolo (2006), a variável aleatória não negativa  $T$ , que indica o tempo até a falha, é frequentemente descrita na análise de sobrevivência por meio da sua função de sobrevivência ou da função de taxa de falha (ou risco). A função de sobrevivência representa a probabilidade de que um evento não tenha ocorrido até um determinado tempo  $t$ , ou seja, a probabilidade de que uma observação continue viva além do tempo  $t$ . Ela é definida por:

$$S(t) = P(T > t). \quad (1)$$

Figura 4 – Exemplo de censura aleatória com (•) indicando a falha e (◦) a censura.



Fonte: Colosimo e Giolo (2006).

Observe que a função pode ser derivada a partir da função de distribuição acumulada. Nesse contexto, a função de distribuição acumulada representa a probabilidade de que um indivíduo não sobreviva até o tempo  $t$ . A função é estabelecida por:

$$S(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t). \quad (2)$$

#### 2.2.4 Função densidade de probabilidade

A função densidade, por sua vez, é a probabilidade de ocorrência de uma falha em um pequeno intervalo por unidade de tempo (CESAR, 2015). Assim, trata-se de uma variável aleatória contínua e positiva. Ela é representada da seguinte forma,

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t)}{\Delta t}. \quad (3)$$

#### 2.2.5 Função de Risco ou Taxa de Falha

A função de risco, também conhecida como taxa de falha, define-se como a probabilidade de uma falha ocorrer em um intervalo específico, dado que não ocorreu antes desse intervalo, conforme citam Colosimo e Giolo (2006), e é expressa por:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t \mid T > t)}{\Delta t}. \quad (4)$$

Quando  $\Delta t \rightarrow 0$ , a probabilidade de falha pode ser representada como,

$$\lambda(t) = -\frac{d \log S(t)}{dt}, \quad (5)$$

em que  $S(t)$  é a função de sobrevivência e  $f(t)$  a função de densidade de probabilidade. A taxa de falha também pode ser expressa como:

$$\lambda(t) = \frac{f(t)}{S(t)}. \quad (6)$$

Nesse sentido, à medida que a probabilidade de sobrevivência diminui, o risco correspondente aumenta, e vice-versa. Assim, a função de risco acumulado  $\Lambda(t)$  quantifica o risco acumulado de ocorrência do evento até o tempo  $t$  e é definida como:

$$\Lambda(t) = \int_0^t \lambda(u) du. \quad (7)$$

Vale salientar que a função de sobrevivência pode ser derivada a expressão de risco acumulado,

$$S(t) = \exp \left\{ - \int_0^t \lambda(u) du \right\}. \quad (8)$$

A função de risco pode assumir quatro diferentes formas ao longo do tempo. Em um cenário crescente, o risco aumenta com o passar do tempo, enquanto em um tipo decrescente, o risco diminui gradualmente. Já na forma conhecida como “banheira”, o risco inicialmente é decrescente, permanece constante sem variações ao longo do período observado e crescente depois de um certo ponto.

## 2.3 Técnicas Não-Paramétricas

A presença de dados censurados pode introduzir viés em estudos de sobrevivência, o que exige o desenvolvimento de métodos específicos para lidar com esses dados. Técnicas não paramétricas são particularmente adequadas para essas análises, pois permitem estimar a função de sobrevivência sem precisar assumir uma distribuição de probabilidade específica.

A análise não paramétrica é especialmente útil quando os dados não seguem uma distribuição normal. Métodos não paramétricos, como os de *Kaplan-Meier* e *Nelson-Aalen*, são importantes porque não requerem suposições sobre a forma da distribuição dos dados (BEWICK; CHEEK; BALL, 2004).

### 2.3.1 Kaplan-Meier

O método *Kaplan-Meier* é amplamente utilizado em estudos de sobrevivência, especialmente quando há dados censurados, ou seja, quando o tempo exato de um evento não é conhecido para todos os indivíduos. Esse método é particularmente relevante em pesquisas sobre câncer, como no caso do câncer de ovário, e permite identificar padrões temporais de mortalidade e progressão da doença, possibilitando uma avaliação mais precisa do prognóstico dos pacientes.

Desenvolvido por Kaplan e Meier (1958), o método é conhecido como estimador limite-produto e oferece estimativas da probabilidade de sobrevivência ao longo do tempo, mesmo com dados incompletos (STATISTICS, 2024). Conforme explicado por Colosimo e Giolo (2006), o método leva em consideração tanto o tempo em que a falha (evento) é observada quanto os momentos em que o evento ainda não ocorreu, sendo esse período denominado tempo de censura.

Para utilizar o método de *Kaplan-Meier*, é necessário dispor os tempos de ocorrência de falhas em uma sequência crescente,  $t_1 < t_2 < \dots < t_k$  em  $k$  tempos distintos. A função é estabelecida na forma:

$$\hat{S}(t) = \prod_{j:t_j < t} \left( \frac{n_j - d_j}{n_j} \right) = \prod_{j:t_j < t} \left( 1 - \frac{d_j}{n_j} \right), \quad (9)$$

em que:

- $t_j$  são os tempos em que ocorrem eventos (falhas) observáveis.
- $d_j$  é o número de falhas em  $t_j$ ,  $j = 1, 2, 3, \dots, k$ ;
- $n_j$  é o número de indivíduos sob risco.

A função  $\hat{S}(t)$  estima a probabilidade de um indivíduo sobreviver além do tempo  $t$ . Se  $\hat{S}(t)$  diminui conforme  $t$  aumenta, isso indica que a probabilidade de sobrevivência diminui com o tempo, o que é comum em análises de sobrevivência. Os valores de  $\hat{S}(t)$  variam de 0 a 1. Um

valor 1 indica uma probabilidade total de sobrevivência até o tempo  $t$ , enquanto um valor 0 sugere que todos os indivíduos passaram pelo evento antes ou no tempo  $t$ .

De acordo com Lima (2010), o estimador de *Kaplan-Meier* possui algumas propriedades importantes. Ele representa o estimador de máxima verossimilhança para a função de sobrevivência  $S(t)$ . Em amostras grandes, esse estimador é não-viciado e converge para um processo Gaussiano. Além disso, o estimador apresenta consistência fraca.

### 2.3.2 Teste Log-rank

As curvas de sobrevivência fornecem uma análise visual inicial sobre possíveis diferenças entre grupos ou categorias de uma variável, mas essa avaliação é apenas preliminar. Embora elas permitam observar o comportamento do tempo até o evento, a comparação visual, por si só, não garante a significância estatística das diferenças observadas.

Para confirmar essas diferenças, é necessário aplicar testes que avaliem a significância estatística entre os grupos. Nesse contexto, o teste *Log-rank* é amplamente utilizado em estudos de sobrevivência. Por se tratar de um teste não paramétrico, é adequado para comparar as curvas de dois ou mais grupos em relação ao tempo até a ocorrência de eventos (como óbito, recaída ou progressão da doença), sem exigir suposições sobre a forma da distribuição do tempo de sobrevivência (BEWICK; CHEEK; BALL, 2004).

Segundo Colosimo e Giolo (2006), o teste baseia-se na premissa de que, na ausência de diferenças significativas entre as curvas de sobrevivência dos grupos, a taxa de falha esperada em qualquer ponto no tempo será praticamente igual para todos os grupos. Isso implica que, caso as curvas se sobreponham de maneira similar, o comportamento da falha ao longo do tempo será consistente entre eles. Nesse sentido, o teste possui as seguintes hipóteses:

$$\begin{cases} H_0 : S_1(t) = S_2(t) = S_3(t) = \dots = S_n(t) \\ H_1 : S_i(t) \neq S_j(t) \text{ para algum } i \text{ e } j, \end{cases}$$

em que

- $n$  é o número total de grupos.
- $i$  e  $j$  são índices que variam de 1 a  $n$ , representando diferentes grupos.

Para a realização do teste, é necessário classificar em ordem crescente os tempos de falha das combinações de grupos, em que, será demonstrado na Tabela 2 o teste com base em dois grupos.

Tabela 2 – Contingências observadas em  $t_j$ .

	Grupos		Totais
	1	2	
Falha	$d_{1j}$	$d_{2j}$	$d_j$
Não Falha	$n_{1j} - d_{1j}$	$n_{2j} - d_{2j}$	$n_j - d_j$
Totais	$n_{1j}$	$n_{2j}$	$n_j$

Fonte: Elaborado pelo autor,(2024).

Então, se são fixadas as marginais da coluna e linha, percebe-se que  $d_{2j}$  segue distribuição Hipergeométrica, sob hipótese nula  $H_0$ . Logo a média e variância de  $d_{2j}$  são dadas, respectivamente, por:

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

Se for assumido que as  $k$  tabelas de contingência geradas são independentes, um teste aproximado para a igualdade das duas funções de sobrevivência pode ser baseado na estatística:

$$T = \frac{[\sum_{j=1}^K (d_{2j} - W_{2j})]}{\sum_{j=1}^K (V_j)^2}. \quad (10)$$

Quando comparamos apenas dois grupos, a estatística  $T$  segue distribuição Qui-quadrado ( $\chi^2$ ) com 1 grau de liberdade. Esse conceito pode ser generalizado para a comparação de  $k$  grupos simultaneamente.

### 2.3.3 Teste de Wilcoxon

O teste foi desenvolvido para comparar amostras emparelhadas; posteriormente, foi adaptado para lidar com dados censurados, conforme descrito por Gehan (1965) e Breslow (1970). Esta adaptação permite a comparação entre grupos sem a necessidade de atender aos pressupostos paramétricos, como a normalidade, o que é comum em estudos de sobrevivência, nos quais os dados podem estar incompletos devido à censura.

Diferente do teste *Log-rank*, o teste de *Wilcoxon* pondera os dados pelo número de indivíduos sob risco ( $w_j = n_j$ ), o que dá mais peso às observações do início do estudo, quando todos os indivíduos ainda estão sob risco. À medida que o tempo avança, os indivíduos deixam de estar sob risco conforme ocorrem falhas ou censuras.

Esse método leva em conta tanto a experiência de sobrevivência quanto o padrão de censura, o que significa que diferenças no padrão de censura entre os grupos comparados podem influenciar o resultado do teste. Em alguns casos, a rejeição ou não da hipótese nula pode ser influenciada não apenas pelas diferenças nas taxas de sobrevivência, mas também pelo padrão de censura observado (WILCOXON, 1945).

Essa dependência do padrão de censura pode, no entanto, tornar o teste menos robusto. Se o padrão de censura entre os grupos for significativamente diferente, o teste pode levar a conclusões enviesadas, pois as diferenças observadas podem refletir o padrão de censura, em vez das verdadeiras diferenças nas taxas de sobrevivência.

## 2.4 Modelos Paramétricos

As técnicas abordadas até o momento apresentam limitações para lidar com a influência de outros fatores sobre o tempo de falha, uma vez que não consideram variáveis adicionais que possam afetar o desfecho. Para superar essa limitação, utilizam-se modelos de regressão, que permitem a inclusão de covariáveis. No contexto de análise de sobrevivência, os modelos de regressão podem ser classificados em dois tipos principais: modelos paramétricos e semiparamétricos.

Os modelos paramétricos assumem distribuição específica para o tempo de sobrevivência (como Weibull, Exponencial, Log-normal) e ajustam o modelo com base nesses parâmetros (ZELLER; BESSEGATO, 2004). Já os modelos semiparamétricos, como o modelo de Cox, são mais flexíveis, pois não impõem uma forma prévia para a distribuição do tempo de sobrevivência, sendo amplamente utilizados quando não se deseja assumir uma distribuição específica.

### 2.4.1 Distribuição Exponencial

A distribuição Exponencial é um modelo fundamental utilizado para representar o tempo até a falha de um evento, caracterizando-se por uma taxa de falha constante (NAKANO; CAR-RASCO, 2006). Essa constância implica que todos os indivíduos que ainda não sofreram a falha apresentam o mesmo risco de falhar ao longo do tempo. Essa propriedade é conhecida como “falta de memória”. Quando consideramos  $T$  como a variável aleatória que representa o tempo até a falha, essa característica torna a distribuição Exponencial útil para a análise do tempo de sobrevivência. Algumas funções importantes são apresentadas a seguir:

1. Função densidade de probabilidade:

$$f(t) = \frac{1}{\alpha} \exp \left\{ - \left( \frac{t}{\alpha} \right) \right\}, \quad t \geq 0. \quad (11)$$

2. Função de sobrevivência:

$$S(t) = \exp \left\{ - \left( \frac{t}{\alpha} \right) \right\}. \quad (12)$$

3. Função de risco ou taxa de falha :

$$\lambda(t) = \frac{1}{\alpha}, \quad (13)$$

em que  $\alpha > 0$  é o tempo médio de vida onde é expresso na mesma unidade do tempo de falha  $t$ .

A distribuição Exponencial apresenta propriedades fundamentais para a análise do tempo até a falha. A esperança, que indica o valor médio esperado para esse tempo, é dada por  $E[T] = \alpha$ . Adicionalmente, a variância, que quantifica a dispersão dos tempos de falha em relação à média, é expressa por  $Var(T) = \alpha$ .

### 2.4.2 Distribuição Weibull

A distribuição de *Weibull* é amplamente empregada na modelagem do tempo de sobrevivência devido à sua flexibilidade (CARDIAL, 2017). Além disso, é frequentemente utilizada por apresentar uma variedade de formas, todas com propriedades básicas (SHAW, 1987). A função de densidade de probabilidade para o tempo da falha é dada por:

$$f(t) = \frac{\gamma}{a^\gamma} t^{\gamma-1} \exp \left\{ - \left( \frac{t}{a} \right)^\gamma \right\}, \quad t \geq 0, \quad (14)$$

em que,  $\gamma \geq 0$  é o parâmetro de forma e  $a \geq 0$  o de escala, em que  $\gamma$  está na mesma unidade de tempo do problema analisado. Tem-se também:

1. Função de sobrevivência:

$$S(t) = \exp \left\{ - \left( \frac{t}{a} \right)^\gamma \right\}. \quad (15)$$

2. Função de risco ou taxa de falha: apresenta um comportamento monótono, podendo apresentar um comportamento crescente, decrescente ou constante.

$$\lambda(t) = \frac{\gamma}{a^\gamma} t^{\gamma-1}. \quad (16)$$

A esperança da distribuição *Weibull*, representando o tempo médio até a falha, é dada por:

$$E[T] = \alpha \Gamma \left[ 1 + \left( \frac{1}{\gamma} \right) \right]. \quad (17)$$

Por outro lado, a variância, que mede a dispersão dos tempos de falha, é expressa por:

$$Var(T) = \alpha^2 \left[ \Gamma \left[ 1 + \left( \frac{2}{\gamma} \right) \right] - \Gamma \left[ 1 + \left( \frac{1}{\gamma} \right) \right]^2 \right]. \quad (18)$$

### 2.4.3 Distribuição Log-Normal

A distribuição Log-normal é comumente associada à análise da duração de produtos e à expectativa de vida de pacientes com determinadas doenças. Caracterizada por uma taxa de falhas que aumenta inicialmente, atinge um pico e posteriormente decresce, essa distribuição estatística, quando aplicada a uma variável aleatória  $T$ , permite modelar diversos fenômenos de tempo de vida (COLOSIMO; GIOLO, 2006).

1. Função densidade de probabilidade (FDP):

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2} \left( \frac{\log(t) - \mu}{\sigma} \right)^2 \right\}, \quad t \geq 0, \quad (19)$$

em que,  $\mu$  representa a média dos logaritmos dos tempos de falha e  $\sigma$  o desvio padrão associado a essa média.

2. Função de sobrevivência:

$$S(t) = \Phi \left( \frac{-\log(t) + \mu}{\sigma} \right), \quad (20)$$

sendo,  $\Phi(\cdot)$  a função de distribuição acumulada de uma normal padrão.

3. Função de risco ou taxa de falha:

$$\lambda(t) = \frac{f(t)}{S(t)}. \quad (21)$$

Na distribuição Log-normal, a esperança e a variância da variável aleatória  $T$  são expressas pelas seguintes fórmulas. A esperança é dada por:

$$E[T] = \exp\{\mu + \sigma^2/2\}. \quad (22)$$

Enquanto a variância é calculada por:

$$Var(T) = \exp\{2\mu + \sigma^2\} (\exp\{\sigma^2\} - 1). \quad (23)$$

### 2.4.4 Modelos Paramétricos de Regressão

Nesta subseção, apresentamos os três modelos paramétricos de regressão das distribuições mencionadas, que são utilizados para estimar o tempo de sobrevivência de objetos ou pessoas em estudo. Consideramos  $\beta'$  como os parâmetros a serem estimados em relação às covariáveis, assumindo que  $T$  segue a distribuição associada. Os modelos podem ser expressos da seguinte forma:

1. **Modelo Exponencial:**

$$T = \exp(\beta_0 + \beta_1 x) + \varepsilon. \quad (24)$$

2. **Modelo Weibull:**

$$Y = \log(T) = \beta_0 + \beta_1 x + \sigma v. \quad (25)$$

3. **Modelo Log-Normal:**

$$Y = \log(T) = \beta_0 + \beta_1 x + \mu \sigma. \quad (26)$$

### 2.4.5 Adequação dos Modelos ajustados

A avaliação da qualidade dos ajustes dos modelos utilizados é indispensável. Essa verificação permite identificar falhas que podem comprometer a eficácia do modelo para os dados analisados, além de ajudar na seleção do modelo mais adequado. Nos modelos paramétricos de regressão, uma abordagem comum para essa verificação é a análise de resíduos. A técnica gráfica significativa apresentada por Cox e Snell (1968) permite a avaliação do ajuste geral do modelo e é utilizada neste estudo.

Os resíduos são calculados por meio da seguinte expressão:

$$\hat{e}_i = \hat{\Lambda}(t_i | x_i), \quad (27)$$

em que  $\hat{\Lambda}(\cdot)$  representa a função de risco acumulado derivada do modelo ajustado. Para os modelos abordados, os resíduos de *Cox-Snell* são definidos da seguinte maneira:

1. **Exponencial:**

$$\hat{e}_i = \left[ t_i \exp \left\{ -x_i \hat{\beta} \right\} \right]. \quad (28)$$

2. **Weibull:**

$$\hat{e}_i = \left[ t_i \exp \left\{ -x_i \hat{\beta} \right\} \right]^{\hat{\gamma}}. \quad (29)$$

3. **Log-normal:**

$$\hat{e}_i = -\log \left[ 1 - \phi \left( \frac{\log(t_i) - x_i \hat{\beta}}{\hat{\sigma}} \right) \right]. \quad (30)$$

De acordo com Lawless (1982), os resíduos segue distribuição exponencial quando o modelo é apropriado, ou seja, quando suas estimativas se aproximam dos valores reais. Nesse cenário, as estimativas serão comparadas com as obtidas pelo método de *Kaplan-Meier*. Para selecionar os modelos de regressão paramétricos ajustados por covariáveis, foi utilizado o Critério de Informação de *Akaike* (AIC), uma das métricas mais comuns para avaliar a qualidade de um modelo (AKAIKE, 1974).

O AIC é utilizado para comparar modelos, de forma que o menor valor indica o modelo mais adequado entre aqueles avaliados (YAMANOUCHI, 2017). Sua fórmula é expressa por:

$$AIC = 2K - 2 \ln(L), \quad (31)$$

em que,  $K$  representa o número de parâmetros do modelo e  $L$  é o valor da função de verossimilhança máxima. Essa fórmula penaliza modelos com muitos parâmetros, favorecendo aqueles que alcançam um bom equilíbrio entre precisão e simplicidade usando uma menor quantidade de parâmetros.

### 2.4.6 Estimativa dos Parâmetros dos Modelos

O processo de estimação dos parâmetros dos modelos é realizado por meio de estatísticas inferenciais, que visam obter estimativas para os parâmetros de distribuições de probabilidade com base em uma amostra representativa. Esse método permite que os parâmetros estimados reflitam características da população, cujos valores verdadeiros são desconhecidos. Neste caso específico, será utilizado o método da máxima verossimilhança, uma técnica amplamente empregada na inferência estatística por sua capacidade de produzir estimativas consistentes e eficientes. A máxima verossimilhança busca encontrar os valores dos parâmetros que maximizam a probabilidade de observar os dados disponíveis, tornando as estimativas mais precisas para análises subsequentes.

#### 2.4.6.1 Método da Máxima Verossimilhança

De acordo com Colosimo e Giolo (2006), o método da máxima verossimilhança busca fornecer estimativas para os parâmetros desconhecidos de uma população, com o objetivo de obter os valores mais prováveis (ou próximos aos verdadeiros) com base em uma amostra. Este método é amplamente utilizado devido às suas propriedades de consistência e eficiência, fornecendo estimativas que se aproximam dos valores reais à medida que o tamanho da amostra aumenta. Considerando, por exemplo, uma amostra de observações não censuradas,  $t_1, \dots, t_n$ , de uma população descrita pela função de densidade:

$$f(t) = \frac{1}{\alpha} \exp\left(-\frac{t}{\alpha}\right), \quad (32)$$

em que as observações seguem, nesse caso, uma distribuição Exponencial com o parâmetro  $\alpha$  a ser estimado. A função de verossimilhança para um parâmetro genérico  $\theta$  de uma população é expressa por:

$$L(\theta) = \prod_{i=1}^n f(t_i; \theta), \quad (33)$$

em que  $\theta$  é o valor que maximiza a função  $L(\theta)$ . A função de verossimilhança é uma ferramenta essencial na estimativa dos parâmetros da população, já que os valores que a maximizam são os mais prováveis de gerar os dados observados.

Cada tipo de censura gera uma função de verossimilhança distinta. No caso de censura à direita, que inclui censuras aleatórias, de tipo I e de tipo II, a fórmula que descreve todos os tipos de censura é dada por:

$$L(\theta) = \prod_{i=1}^r f(t_i; \theta) \prod_{i=r+1}^n S(t_i; \theta), \quad (34)$$

em que,  $S(t)$  é a função de sobrevivência, dada por:

$$S(t) = \exp\left(-\frac{t}{\alpha}\right), \quad (35)$$

com  $S(t)$  definida conforme a distribuição da variável aleatória  $T$ .

As observações são agrupadas em dois conjuntos: as primeiras  $r$  observações, que são não censuradas ( $1, 2, \dots, r$ ), e as  $n - r$  observações restantes, que são censuradas ( $r + 1, r + 2, \dots, n$ ). Para cada tipo de censura, o ajuste da função de verossimilhança permite estimar os parâmetros da distribuição levando em consideração a incerteza introduzida pelas censuras.

#### 2.4.6.2 Teste da Razão de Verossimilhança

O Teste da Razão de Verossimilhança (TRV) é utilizado para avaliar a adequação de modelos, comparando-os por meio dos valores dos logaritmos das funções de verossimilhança maximizada. As hipóteses testadas são:

$$\begin{cases} H_0 : \text{O modelo é adequado.} \\ H_1 : \text{O modelo não é adequado.} \end{cases}$$

No contexto de sobrevivência, esse teste é frequentemente realizado através da distribuição Gama generalizada, que abrange modelos como o Exponencial, Weibull, Log-normal, entre outros. A estatística do teste é calculada por:

$$TRV = -2 \log \left( \frac{L(\hat{\theta}_M)}{L(\hat{\theta}_G)} \right) = 2 [\log L(\hat{\theta}_G) - \log L(\hat{\theta}_M)], \quad (36)$$

em que  $\log L(\hat{\theta}_G)$  e  $\log L(\hat{\theta}_M)$  são os logaritmos das verossimilhanças dos modelos generalizado e de interesse, respectivamente. Esse teste segue uma distribuição Qui-quadrado, com os graus de liberdade definidos pela diferença no número de parâmetros entre os modelos comparados.

### 3 RESULTADOS E DISCUSSÃO

O estudo visa analisar o tempo de sobrevivência, em dias, de pacientes diagnosticadas com câncer de ovário na Paraíba, desde o diagnóstico até o óbito. Fatores como idade, raça/cor, meio de diagnóstico e extensão serão investigados quanto ao impacto na curva de sobrevivência. Os dados envolvem 146 pacientes, e a análise considera as mortes por câncer de ovário e as observações censuradas, conforme apresentado na Tabela 3.

Tabela 3 – Frequência de tipo de óbito.

<b>Tipo</b>	<b>Frequência</b>	<b>Porcentagem (%)</b>
Câncer	105	72,00%
Censura	41	28,00 %

Fonte: Elaborado pelo autor, (2024).

O evento de interesse, isto é, o óbito devido ao câncer, ocorreu em 105 casos, representando 72,0% das pacientes. Em contrapartida, 41 observações foram censuradas, correspondendo a 28,0% dos casos. Censura, neste contexto, refere-se aos casos em que o falecimento não ocorreu por câncer de ovário ou as informações completas não estavam disponíveis.

Em sequência, realizou-se uma análise descritiva para resumir e descrever os dados, sintetizando as principais características do conjunto. A Tabela 4 apresenta os resultados descritivos das variáveis analisadas.

Com base nas informações da Tabela 4, é evidente que o risco de câncer de ovário aumenta com a idade. A maioria das pacientes está na faixa etária de 50 anos ou mais, especialmente considerando que a mediana é de 57,5 anos. É relevante notar que 25% das pacientes têm menos de 46 anos, com a paciente mais jovem apresentando apenas 14 anos. Essa situação pode indicar

Tabela 4 – Medidas descritivas dos dados de câncer de ovário no estado da Paraíba.

Medidas	Mín	1st Qu.	Mediana	Média	3rd Qu.	Máx
Idade	14,00	46,00	57,50	57,84	70,00	94,00
Tempo (dias)	2,00	42,25	340,00	515,21	656,00	3.846,00

Fonte: Elaborado pelo autor, (2024).

a presença de casos raros, ressaltando a importância de promover a conscientização e a prevenção também entre grupos mais jovens.

Além disso, é importante destacar que a incidência é significativamente maior em mulheres mais velhas, como cita ONCOGUIA (2023). A média geral de sobrevivência das pacientes incluindo falhas e censuras é de aproximadamente 516 dias e a mediana, de 340 dias, sugerindo que muitas têm sobrevivências mais curtas, o que pode impactar o tratamento. A Tabela 5 mostra a distribuição dos tipos de câncer de ovário por morfologia na Paraíba, conforme registrado no banco de dados.

Tabela 5 – Descrição da Morfologia e Frequências.

Morfologia	Frequência	Porcentagem (%)
Adenocarcinoma de Células Claras, SOE	2	1,37%
Adenocarcinoma Mucinoso	3	2,05%
Adenocarcinoma Papilar, SOE	2	1,37%
Adenocarcinoma, SOE	81	51,48%
Carcinoma de Células Gigantes	1	0,68%
Carcinoma Endometriode, SOE	4	2,74%
Carcinoma Escamocelular, SOE	2	1,37%
Carcinoma, SOE	24	16,44%
Carcinossarcoma, SOE	1	0,68%
Cistadenocarcinoma Papilar Seroso	1	0,68%
Cistadenocarcinoma Seroso, SOE	4	2,74%
Cistadenocarcinoma, SOE	4	2,74%
Cistoadenocarcinoma Papilar, SOE	2	1,37%
Neoplasia Maligna	13	8,90%
Sarcoma, SOE	1	0,68%
Tumor Papilar de Superfície Serosa, Malignidade Limítrofe	1	0,68%

Fonte: Elaborado pelo autor, (2024).

Os tipos mais comuns de tumores de ovário são adenocarcinoma, carcinoma e neoplasia maligna, todos com origem epitelial, indicando que se desenvolvem no tecido que reveste o ovário. Dentre esses, o adenocarcinoma, SOE (Sem Outra Especificação), é o mais prevalente, com um total de 81 ocorrências, seguido pelo carcinoma, SOE, com 24 ocorrências, e pela neoplasia maligna, com 13 ocorrências. A sigla "SOE" (Sem Outra Especificação), conforme indicado pelo INCA (2006), significa que esses tumores não foram classificados em subtipos específicos, possivelmente devido à falta de características histológicas adicionais ou de informações mais detalhadas.

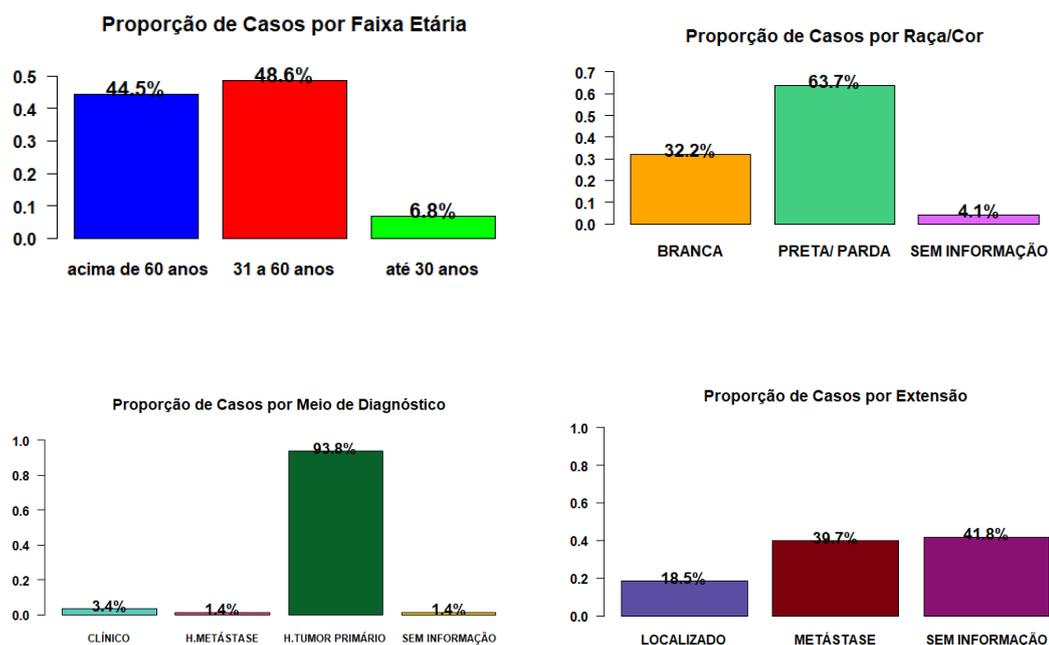
Além desses tipos mais comuns, existem subtipos de tumores de ovário menos frequentes

e com características específicas, como o adenocarcinoma mucinoso e o carcinoma escamocelular. Esses tumores apresentam uma ocorrência menor, sugerindo uma baixa prevalência na população estudada. Há ainda tumores considerados raros, como o sarcoma, SOE, e o carcinosarcoma, SOE, cada um com apenas uma ocorrência, evidenciando a raridade desses subtipos.

A diversidade de subtipos presentes reflete a ampla variedade histológica dos tumores de ovário, indicando que diferentes variações celulares e moleculares podem estar envolvidas no desenvolvimento dessas neoplasias. Esse aspecto destaca a complexidade no diagnóstico e tratamento dos tumores ovarianos, reforçando a importância de estudos aprofundados para melhor compreensão dos diferentes subtipos e de suas características específicas.

Conforme ilustrado na Figura 5, a distribuição etária entre as mulheres diagnosticadas apresenta uma concentração significativa na faixa de 31 a 60 anos, correspondendo a 48,6% do total. Esse grupo representa quase metade da população avaliada, sugerindo que a maioria dos diagnósticos ocorre nessa fase intermediária da vida. Em seguida, observa-se que 44,5% das pacientes têm mais de 60 anos, indicando uma alta prevalência de diagnósticos também em idades mais avançadas. Por outro lado, a menor proporção de diagnósticos ocorre entre mulheres mais jovens: apenas 6,8% das pacientes têm até 30 anos, o que pode apontar para uma menor incidência ou subdiagnóstico nessa faixa etária.

Figura 5 – Frequência dos casos de câncer de ovário segundo a Faixa Etária, Raça/Cor, Meio de diagnóstico e Extensão.



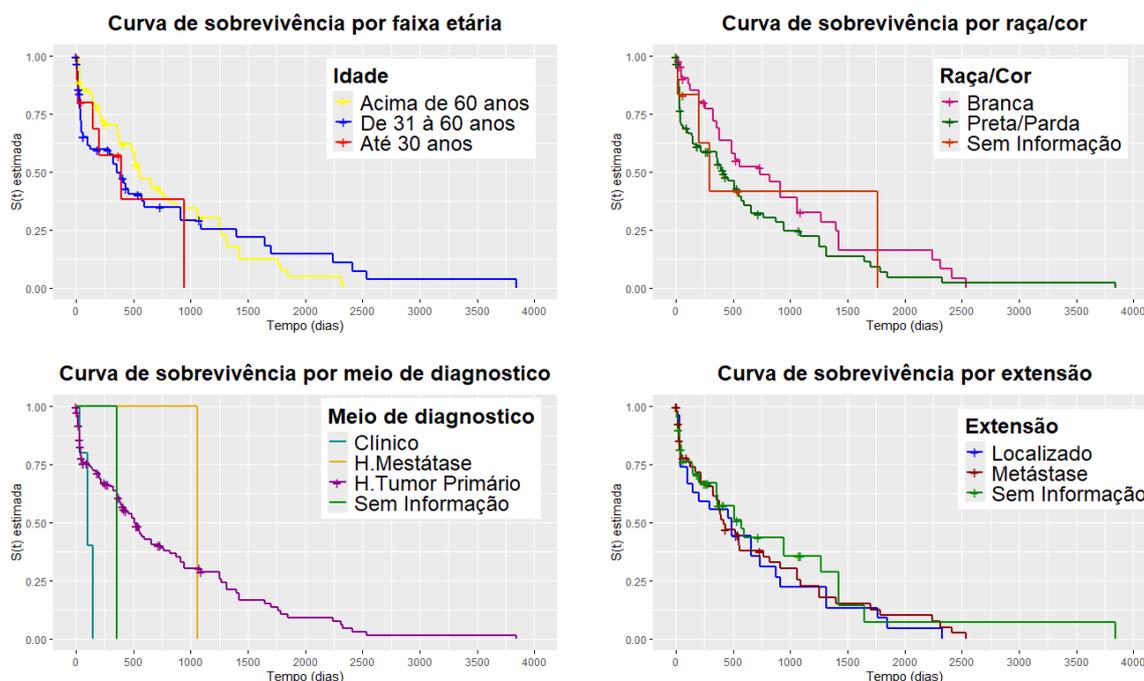
Fonte: Elaborado pelo autor, (2024).

No que se refere à distribuição étnica, a maioria das pacientes, 63,7%, se autodeclara como preta ou parda, o que pode refletir tanto características demográficas da população quanto possíveis desigualdades de saúde. Em contraste, 32,2% das diagnosticadas se identificam como brancas. Esses dados podem indicar diferenças de acesso ao diagnóstico e ao tratamento, além de fatores socioeconômicos e culturais que influenciam o perfil das pacientes diagnosticadas.

O diagnóstico precoce do câncer de ovário é crucial para o prognóstico das mulheres na Paraíba, pois o tratamento depende do estágio da doença. Conforme a Figura 5, 93,8% dos

casos são diagnosticados por exames histológicos, 3,4% clinicamente, e 39,7% das pacientes têm metástase. O método de *Kaplan-Meier* foi utilizado para estimar a sobrevivência, calculando a probabilidade de sobrevivência ao longo do tempo, mesmo com dados censurados, e analisando mortalidade e recorrência.

Figura 6 – Curva de sobrevivência de *Kaplan-Meier* por covariáveis.



Fonte: Elaborado pelo autor, (2024).

Na Figura 6, é exibida, inicialmente, a curva de sobrevivência por idade. Pacientes com até 30 anos apresentam uma taxa de sobrevivência de 90% nos primeiros 10 dias, caindo para 57,1% em 200 dias. Para aqueles entre 31 e 60 anos, a sobrevivência chega a 95,18% em 23 dias, mas cai para 38,73% em 363 dias e 10,04% após cerca de 1.763 dias, aproximadamente 5 anos. A estimativa de sobrevivência para pacientes com mais de 60 anos apresenta um declínio gradual. Inicialmente, a taxa é de 96,88% até o sexto dia, caindo para 65,2% após 59 dias, não havendo sobreviventes após aproximadamente 3.846 dias (cerca de 10 anos).

Em relação à covariável raça/cor, os grupos são: Branca, Preta/Parda e Sem Informação. O grupo de mulheres brancas mostram maior probabilidade de sobrevivência, enquanto o grupo das pacientes pretas/pardas apresentam uma queda mais rápida. A probabilidade de sobrevivência além de 500 dias é aproximadamente 42,89%. O grupo das pacientes sem informação sobre a raça apresenta maior incerteza e uma queda acentuada após 1.000 dias, indicando menor quantidade de dados disponíveis.

Estudos como o de Latado et al. (2009) aponta o impacto da raça/etnia nos desfechos de sobrevivência, a curva na Figura 6 corrobora essa evidência. No entanto, é importante enfatizar que houve uma grande proporção de mulheres autodeclaradas negras ou pardas nessa coorte, o que pode ter influenciado a significância estatística em relação ao impacto da etnia na sobrevivência das pacientes com câncer de ovário. Nesse caso, é fundamental observar a participação desproporcional de mulheres desses grupos étnicos, uma questão que deve ser considerada independentemente da doença em estudo.

As curvas de sobrevivência para os diferentes métodos de diagnóstico mostram seus respectivos grupos, sendo eles: Clínico, Histologia de Metástase, Histologia de Tumor Primário

e Sem Informação. O grupo Clínico inclui pacientes com uma rápida queda na sobrevivência nos primeiros dias, enquanto o grupo de Histologia de Metástase inclui pacientes que apresentam alta sobrevida inicial, mas que declina rapidamente após 800 dias. O grupo de Histologia de Tumor Primário comporta pacientes com uma diminuição mais lenta da sobrevida ao longo de 4.000 dias, apresentando uma probabilidade de 28,91% após 1.000 dias, indicando um prognóstico mais favorável.

Os resultados apresentados na Figura 6 mostram que as curvas de sobrevivência dos três grupos (localizado, metastático e sem informação) revelam diferenças sutis. Isso sugere que outros fatores, como tratamento e acesso à saúde, também desempenham um papel significativo na sobrevivência das pacientes. Além disso, a maneira como o câncer se manifesta pode evidenciar essas sutis variações entre as pacientes.

Na sequência, foi realizado o teste *Log-rank* e *Wilcoxon* para comparar as curvas de sobrevivência entre os grupos de cada fator mencionado, sendo eles: idade, raça/cor, meio de diagnóstico e extensão. Para cada um deles, as comparações foram feitas entre todas as combinações de grupos, em pares.

Tabela 6 – Aplicação dos testes *Log-rank* e *Wilcoxon* à covariável faixa etária.

<b>Grupos</b>	<b><i>Log-rank</i> P-valor</b>	<b><i>Wilcoxon</i> P-valor</b>
Até 30 anos X 31 a 60 anos	0,20	0,20
Até 30 anos X acima de 60 anos	0,70	1,00
31 a 60 anos X acima de 60 anos	0,60	0,08

Fonte: Elaborado pelo autor, (2024).

Na Tabela 6, observou-se que os p-valores são predominantemente altos, indicando que não há evidências suficientes para afirmar que a faixa etária influencia a sobrevivência das pacientes com câncer de ovário no estado da Paraíba. No entanto, o p-valor de 0.08 obtido no teste *Wilcoxon* entre os grupos "31 a 60 anos" e "acima de 60 anos" sugere uma tendência que merece atenção. Embora essa diferença não seja estatisticamente significativa, os resultados podem indicar que as pacientes acima de 60 anos apresentam uma taxa de sobrevivência inferior em comparação às pacientes na faixa etária de 31 a 60 anos.

Tabela 7 – Aplicação dos testes *Log-rank* e *Wilcoxon* à covariável raça/cor.

<b>Grupos</b>	<b><i>Log-rank</i> P-valor</b>	<b><i>Wilcoxon</i> P-valor</b>
Branca X Preta/Parda	0,06	< 0,01

Fonte: Elaborado pelo autor, (2024).

A Tabela 7 mostra que, embora o teste *Log-rank* não tenha identificado diferenças significativas entre os grupos por raça, o teste *Wilcoxon* sugere uma variação relevante na sobrevida. Essa diferença pode refletir o impacto de fatores como acesso a serviços de saúde, educação, emprego e renda na saúde e sobrevivência das pacientes.

Tabela 8 – Aplicação dos testes *Log-rank* e *Wilcoxon* à covariável meio de diagnóstico.

<b>Grupos</b>	<b><i>Log-rank</i> P-valor</b>	<b><i>Wilcoxon</i> P-valor</b>
Clínico X H.Metástase	< 0,01	0,06
Clínico X H.Tumor Primario	< 0,01	< 0,01
H.Metástase X H.Tumor Primario	0,80	0,40

Fonte: Elaborado pelo autor, (2024).

Na Tabela 8, são exibidas as diferenças na sobrevida entre pacientes com câncer diagnosticados clinicamente, aqueles com tumor primário e os que têm tumor metastático. Os resultados indicam que as pacientes diagnosticadas clinicamente têm uma sobrevida significativamente melhor em comparação aos pacientes com tumor primário, com um p-valor menor que 0,01. Isso significa que a diferença é muito pequena e não é provável que tenha ocorrido por acaso. Além disso, em comparação aos pacientes com metástase, a sobrevida das diagnosticados clinicamente também é melhor, com um p-valor abaixo de 0,01. No entanto, o teste utilizado para comparar os dois grupos de metástase e tumor primário não mostrou uma diferença significativa, pois o p-valor foi de 0,80 e 0,40, respectivamente. Isso significa que a sobrevida entre esses dois grupos é bastante similar.

Tabela 9 – Aplicação dos testes *Log-rank* e *Wilcoxon* à covariável extensão.

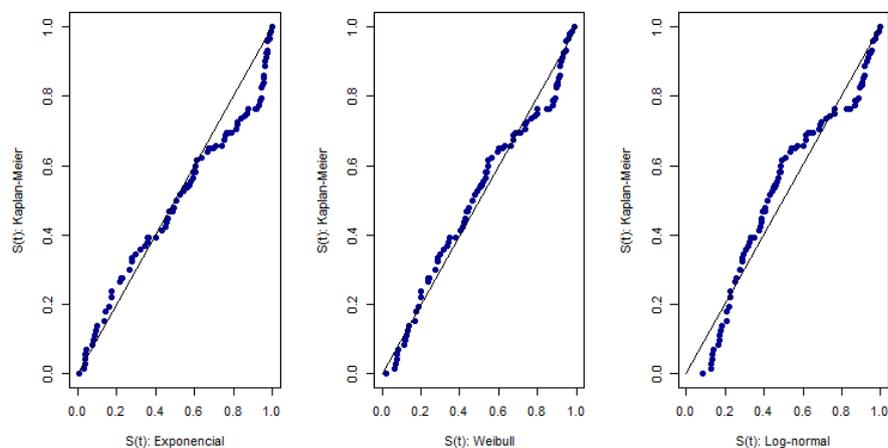
<b>Grupos</b>	<b><i>Log-rank</i> P-valor</b>	<b><i>Wilcoxon</i> P-valor</b>
Localizado X Metástase	0,70	0,80

Fonte: Elaborado pelo autor, (2024).

Os resultados da Tabela 9 mostram p-valores acima de 5%, indicando que não há diferença significativa nas taxas de sobrevivência entre os grupos. Essa conclusão aponta para a extensão da doença, seja localizada ou metastática, como um fator que não exerce um impacto significativo nas taxas de sobrevivência das pacientes avaliadas. Fatores como o tipo de câncer, grau de diferenciação, subtipos moleculares, tratamento recebido, adesão ao tratamento, idade, sexo e comorbidades influenciam significativamente a sobrevivência e o prognóstico dos pacientes.

Dessa forma, para a seleção do melhor modelo de regressão que avalie a influência dos fatores apresentados, na chance de sobrevivência das pacientes, foram feitas análises gráficas comparando as estimativas de *Kaplan-Meier* com os modelos probabilísticos. Os resíduos de *Cox-Snell* (Figura 7) indicaram bom ajuste tanto para a distribuição Exponencial quanto para a Weibull.

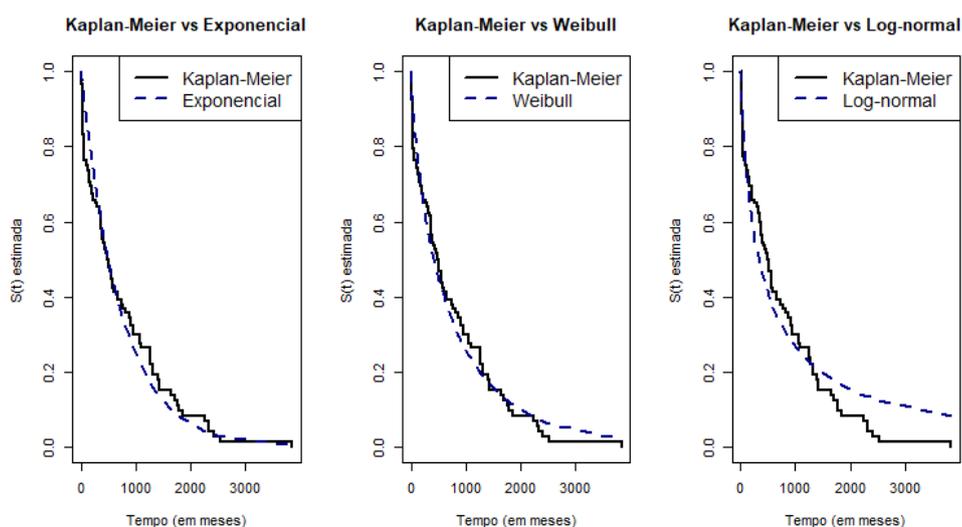
Figura 7 – Análise dos resíduos de *Cox-Snell* das estimativas de sobrevivência dos modelos de regressão em comparação com *Kaplan-Meier*.



Fonte: Elaborado pelo autor, (2024).

As curvas ajustadas com as três distribuições de probabilidade, juntamente com as estimativas de *Kaplan-Meier* mostram, na Figura 8, que as distribuições Exponencial e Weibull são as mais adequadas para os dados analisados, pois a curva azul pontilhada segue de forma consistente a curva de *Kaplan-Meier*.

Figura 8 – Gráfico comparativo das estimativas de sobrevivência de *Kaplan-Meier* com as estimativas dos modelos Exponencial, Weibull e Log-normal.



Fonte: Elaborado pelo autor, (2024).

Para validar os resultados das análises gráficas, foi realizado o teste da razão de verossimilhança com as seguintes hipóteses:

- i) o modelo Exponencial é adequado;

- ii) o modelo Weibull é adequado;
- iii) o modelo Log-normal é adequado.

O propósito deste teste é identificar qual distribuição proporciona o melhor ajuste aos dados, comparando sua capacidade de modelar a distribuição gama generalizada, considerando aspectos como a forma da distribuição e a adequação aos padrões observados.

Tabela 10 – Log-verossimilhança, Testes de Razão de Verossimilhança e Critério de Informação de Akaike para Modelos de Distribuição.

Modelo	$\log(L(\theta))$	TRV	P-valor	AIC
Gama Generalizada	-787,64	-	-	1581,281
Exponencial	-795,29	15,303	0,0005	1592,585
Weibull	-788,98	2,683	0,1014	1581,964
Log-normal	-794,69	14,097	0,0002	1593,378

Fonte: Elaborado pelo autor, (2024).

Conforme os resultados apresentados na Tabela 10, o modelo Gama Generalizada, como modelo de referência, apresenta o melhor ajuste, enquanto o modelo Weibull surge como alternativa viável devido à ausência de diferença estatisticamente significativa no ajuste. Em contrapartida, os modelos Exponencial e Log-normal mostraram-se menos adequados, conforme o TRV. O modelo Weibull, apresenta menor AIC (1581.964), oferecendo o melhor ajuste em comparação aos modelos Exponencial (AIC de 1592.585) e Log-normal (AIC de 1593.378). Embora o Exponencial também forneça um bom ajuste, seu AIC maior sugere menor eficiência que o modelo Weibull. O Log-normal, com o AIC mais alto, apresenta o ajuste menos satisfatório entre os três modelos. Portanto, com base no critério AIC, o modelo Weibull é o mais adequado entre os avaliados.

Tabela 11 – Modelo de regressão paramétrico ajustado por covariáveis através da distribuição Weibull.

Covariáveis	Estimativas ( $\hat{\beta}_i$ )	Exp(Estimativas)	P-valor
(Intercept)	4,933	138,826	< 0,010
IDADE_C acima de 60 anos	- 0,171	0,843	0,525
IDADE_C até 30 anos	- 0,531	0,588	0,304
Raca/Cor PRETA/ PARDA	- 0,466	0,628	0,105
Raca/Cor SEM INFORMAÇÃO	- 0,152	0,859	0,825
Meio de Diagnóstico H.METÁSTASE	2,083	8,028	0,058
Meio de Diagnóstico H.TUMOR PRIMÁRIO	1,903	6,707	< 0,010
Meio de Diagnóstico SEM INFORMAÇÃO	1,124	3,076	0,286
Extensão METÁSTASE	- 0,052	0,949	0,879
Extensão SEM INFORMAÇÃO	0,461	1,585	0,194

Fonte: Elaborado pelo autor, (2024).

Após a obtenção do modelo a partir da distribuição Weibull, foi realizada a seleção do melhor modelo por meio de critérios de seleção como *stepwise*, AIC e menor p-valor. Em todos os critérios adotados o melhor modelo foi o modelo completo com todas as covariáveis, embora

algumas não sejam significativas, pois mesmo com altos p-vaolres a retirada destas impactavam negativamente no modelo.

Nesse sentido, o modelo de regressão paramétrica ajustado por covariáveis através da distribuição Weibull para o câncer de ovário, revela importantes fatores de risco associados à evolução da doença. A covariável meio de diagnóstico demonstrou que o diagnóstico realizado por metástase está associado a um aumento de cerca de 8 vezes no risco de morte pelo câncer em relação a outros métodos de diagnóstico ( $\text{Exp}(\beta) = 8,028$ ). Esse achado reforça o impacto negativo das metástases no prognóstico, indicando uma maior gravidade da doença e um desfecho menos favorável.

Por outro lado, a extensão do câncer indica que a presença de metástase está associada a uma discreta redução no risco de morte ( $\text{Exp}(\beta) = 0,949$ ). Esse resultado pode parecer contraintuitivo e sugere que fatores adicionais podem influenciar essa relação. Por outro lado, a ausência de informações sobre a extensão do câncer está associada a um aumento de 58,5% no risco de morte das pacientes ( $\text{Exp}(\beta) = 1,585$ ), possivelmente indicando incertezas diagnósticas ou dificuldades no manejo clínico.

Em relação à idade, pacientes com mais de 60 anos apresentam um risco 16% menor ( $\text{Exp}(\beta) = 0,843$ ) em comparação com pacientes mais jovens. Esse resultado pode refletir diferenças na biologia do tumor ou na abordagem terapêutica entre diferentes faixas etárias.

Por fim, a análise destaca a importância de diagnósticos precoces e informações detalhadas sobre a extensão do câncer no manejo clínico do câncer de ovário, reforçando que fatores como metástases e diagnósticos tardios impactam significativamente no desfecho da doença.

#### 4 CONCLUSÃO

Os resultados deste estudo revelam importantes aspectos sobre a sobrevivência das pacientes diagnosticadas com câncer de ovário no estado da Paraíba. A análise dos dados, que abrange 146 pacientes, mostra que 71,91% dos óbitos ocorreram devido ao câncer, enquanto 28,08% dos casos foram censurados, destacando a importância de considerar esses casos no modelo de sobrevivência.

Em relação à morfologia dos tumores, o adenocarcinoma foi o tipo mais prevalente, seguido pelo carcinoma e pela neoplasia maligna. A diversidade de subtipos encontrados reflete a complexidade do câncer de ovário e a importância de diagnósticos precisos para direcionar tratamentos adequados. A distribuição etária dos casos indica que a maioria dos diagnósticos ocorre entre 31 e 60 anos, com uma alta prevalência de casos em mulheres acima de 60 anos.

As curvas de sobrevivência pelo método de Kaplan-Meier mostraram que idade, raça/cor, meio de diagnóstico e extensão da doença influenciam o prognóstico. Mulheres negras ou pardas tiveram menor sobrevivência, evidenciando desigualdades socioeconômicas e de acesso à saúde. O modelo Weibull identificou que a presença de metástase é o principal fator de pior prognóstico, aumentando o risco em oito vezes. Os resultados reforçam a importância do diagnóstico precoce e de políticas públicas para reduzir disparidades, melhorando o acesso ao tratamento e a sobrevivência das pacientes.

Por fim, este estudo contribui para a compreensão dos fatores que afetam a sobrevivência de pacientes com câncer de ovário na Paraíba, fornecendo informações importantes para o desenvolvimento de estratégias de prevenção e tratamento, especialmente para grupos mais vulneráveis. Para pesquisas futuras, recomenda-se o uso de dados que incluam informações sobre diagnósticos secundários e proporções similares nas variáveis analisadas. Além disso, o modelo de riscos de Cox pode ser utilizado para uma análise adicional dos fatores que impactam a sobrevida das pacientes.

## REFERÊNCIAS

- AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, v. 19, p. 716–723, 1974. Citado na página 20.
- BEWICK, V.; CHEEK, L.; BALL, J. Revisão estatística 12: Análise de sobrevivência. *Cuidados Críticos*, v. 8, p. 389, 2004. Disponível em: <<https://doi.org/10.1186/cc2955>>. Citado 2 vezes nas páginas 15 e 16.
- BRESLOW, N. Um teste generalizado de kruskal-wallis para comparar amostras  $k$  sujeitas a padrões desiguais de censura. *Biometrika*, v. 57, n. 3, 1970. Citado na página 17.
- CARDIAL, M. R. P. *Distribuição Weibull discreta exponenciada para dados com presença de censura: uma abordagem clássica e bayesiana*. Dissertação (Dissertação (Mestrado em Estatística)) — Universidade de Brasília, Departamento de Estatística, 2017. Citado na página 18.
- CESAR, K. A. *Análise Estatística de Sobrevivência: Um Estudo com Pacientes com Câncer de Mama*. Dissertação (Mestrado) — Universidade Católica de Brasília, 2015. Citado na página 14.
- COLOSIMO, E. A.; GIOLO, S. R. *Análise de Sobrevivência Aplicada*. São Paulo: Blucher, 2006. Citado 7 vezes nas páginas 12, 13, 14, 15, 16, 19 e 21.
- CONSCIENTE, M. *Tudo Sobre Câncer de Ovário: Tipos, Estágios e Metástase*. 2024. Acesso em: 04 nov. 2024. Disponível em: <<https://mulherconsciente.com.br/cancer-de-ovario/tudo-sobre-cancer-de-ovario/#metastase>>. Citado na página 11.
- CONSOLARO, M. E. L.; MARIA-ENGLER, S. S. *Citologia Clínica Cérvico-Vaginal*. [S.l.]: São Paulo, 2014. Citado na página 9.
- COX, D. R.; SNELL, E. J. A general definition of residuals. *Journal of the Royal Statistical Society. Series B (Methodological)*, v. 30, n. 2, p. 248–275, 1968. Citado na página 20.
- CRUZ, H. A. O. *Câncer de Ovário*. 2024. <<https://www.hospitaloswaldocruz.org.br/centro-especializado/oncologia/cancer-de-ovario/>>. Acesso em: 09 out. 2024. Citado na página 9.
- EUGENIO, D. R. R. *Modelos de sobrevivência para estudo do tempo de internação hospitalar até a alta do paciente*. 84 p. Dissertação (Mestrado) — Universidade Estadual Paulista (Unesp), Faculdade de Ciências e Tecnologia, Presidente Prudente, 2021. Dissertação (mestrado). Orientador: Mário Hissamitsu Tarumoto. Disponível em: <<https://repositorio.unesp.br/server/api/core/bitstreams/e7624dcb-60e8-4a58-80b2-e4d1b0bd73eb/content>>. Citado na página 12.
- FIOCRUZ, F. O. C. *Câncer de Ovário: Portal de Boas Práticas em Saúde da Mulher, da Criança e do Adolescente*. 2022. Acesso em: 24 out. 2024. Disponível em: <<https://portaldeboaspraticas.iff.fiocruz.br/wp-content/uploads/2022/10/cancerdeovario-220823163525-7f6c84d4.pdf>>. Citado na página 9.
- GEHAN, E. A. A generalized wilcoxon test for comparing arbitrarily censored samples. *Biometrika*, v. 52, n. 1/2, 1965. Citado na página 17.

- INCA. *Codificação de Tumores e Classificações Utilizadas em Registros de Câncer*. 2006. Acesso em: [data de acesso]. Disponível em: <<http://www1.saude.rs.gov.br/dados/1156340739752AulaCodificacaoCIDECID-O2006%5Binca%5D.pdf>>. Citado na página 23.
- INCA. *Câncer de ovário*. [S.l.], 2023. Disponível em: <<https://www.gov.br/inca/pt-br/assuntos/cancer/tipos/ovario>>. Citado na página 9.
- KABAD, J. F.; BASTOS, J. L.; SANTOS, R. V. Raça, cor e etnia em estudos epidemiológicos sobre populações brasileiras: revisão sistemática na base pubmed. *Physis: Revista de Saúde Coletiva*, Rio de Janeiro, v. 22, n. 3, p. 895–918, 2012. Citado na página 11.
- KAPLAN, E. L.; MEIER, P. Estimação não paramétrica a partir de observações incompletas. *Jornal da Associação Americana de Estatística*, v. 53, p. 457–485, 1958. Citado 2 vezes nas páginas 10 e 15.
- LATADO, A. L. et al. Existe evidência para tratar insuficiência cardíaca baseada na raça ou etnia? *Revista da Associação Médica Brasileira*, SciELO Brasil, v. 55, p. 110–116, 2009. Citado na página 25.
- LAWLESS, J. F. *Statistical Models and Methods for Lifetime Data*. 2. ed. New York: John Wiley and Sons, 1982. Citado 2 vezes nas páginas 12 e 20.
- LIMA, T. R. *Aula 3: Análise de Sobrevivência*. 2010. Acesso em 5 de setembro de 2024. Disponível em: <<http://www.de.ufpb.br/~tarciana/MDS/Aula3.pdf>>. Citado na página 16.
- MEDIMAGEN. *Câncer de ovário*. 2017. Blogue Medimagen. Publicado em: 1 de setembro de 2017. Acessado em: 8 de outubro de 2024. Disponível em: <<http://medimagen.mx/?p=295>>. Citado na página 11.
- NAKANO, E. Y.; CARRASCO, C. G. Uma avaliação do uso de um modelo contínuo na análise de dados discretos de sobrevivência. *Sociedade Brasileira de Matemática Aplicada e Computacional–SBMAC*, 2006. Citado na página 18.
- ONCOGUIA, I. *Fatores de Risco para Câncer de Ovário*. 2023. Acesso em: 26 set. 2024. Disponível em: <<https://www.oncoguia.org.br/conteudo/fatores-de-risco-para-cancer-de-ovario/10925/1126/>>. Citado 2 vezes nas páginas 9 e 23.
- REIS, E. A.; REIS, I. A. Análise descritiva de dados. *Relatório Técnico do Departamento de Estatística da UFMG*, v. 1, 2002. Citado na página 10.
- SANTOS, V. S. *Ovários*. [S.l.], 2024. Disponível em: <<https://brasilecola.uol.com.br/biologia/ovarios.htm>>. Citado na página 9.
- SAÚDE. *Glossário Temático Controle de Câncer*. [S.l.], 2024. Disponível em: <<https://www.inca.gov.br/sites/ufu.sti.inca.local/files//media/document//glossario-tematico-controle-de-cancer.pdf>>. Citado na página 9.
- SHAW, M. Weibull analysis of component failure data from accelerated testing. *Reliability Engineering*, Elsevier, v. 19, n. 3, p. 237–243, 1987. Citado na página 18.
- STATISTICS, L. *Kaplan-Meier Using SPSS Statistics*. 2024. Acesso em: 24 out. 2024. Disponível em: <<https://statistics.laerd.com/spss-tutorials/kaplan-meier-using-spss-statistics.php>>. Citado na página 15.

THERNEAU, T. M. *A Package for Survival Analysis in R*. [S.l.], 2023. R package. Disponível em: <<https://CRAN.R-project.org/package=survival>>. Citado na página 10.

WILCOXON, F. Individual comparisons by ranking methods. *Biometrics Bulletin*, v. 1, n. 6, p. 80–83, 1945. Citado na página 17.

YAMANOUCHI, T. K. *Seleção de modelos lineares mistos utilizando critérios de informação*. 57 p. Dissertação (Dissertação (Mestrado)) — Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2017. Versão revisada de acordo com a resolução CoPGr 6018 de 2011. Citado na página 20.

ZELLER, C. B.; BESSEGATO, L. F. *Análise de Sobrevivência*. 2004. Acesso em: 03 nov. 2024. Disponível em: <[https://www.bessegato.com.br/UFJF/est100\\_confabilidade/173\\_03\\_modelos\\_probabilisticos.pdf](https://www.bessegato.com.br/UFJF/est100_confabilidade/173_03_modelos_probabilisticos.pdf)>. Citado na página 17.

## APÊNDICE A – DESCRIÇÕES DAS MORFOLOGIAS

1. Adenocarcinoma de células claras: é um tumor maligno, que deriva de células glandulares epiteliais secretoras;
2. Adenocarcinoma mucinoso: é um tipo de câncer maligno que se forma nas células mucinosas do ovário, muitas vezes associado à produção de muco em seu interior;
3. Adenocarcinoma papilar: é um tipo de câncer de ovário que se caracteriza por apresentar crescimento tumoral em padrões de projeções papilares (semelhantes a pequenos dedos) sob o microscópio;
4. Adenocarcinoma: é o tipo mais comum de câncer de ovário e geralmente se desenvolve a partir das células que revestem a superfície externa do ovário;
5. Carcinoma de células gigantes: esse tipo de câncer é caracterizado pela presença de células tumorais de grande tamanho quando observadas sob um microscópio. O carcinoma de células gigantes pode afetar diferentes partes do corpo;
6. Carcinoma endometriode: é um tipo de câncer que se origina no revestimento interno do útero, conhecido como endométrio;
7. Carcinoma escamocelular: refere-se a um tipo de câncer que se origina nas células escamosas, que são células epiteliais planas e delgadas encontradas em várias partes do corpo, incluindo a pele e membranas mucosas, como aquelas que revestem as vias respiratórias, o esôfago, a boca e o colo do útero;
8. Carcinoma: refere-se a um tipo de câncer que se origina nas células epiteliais, que são as células que revestem a superfície interna e externa do corpo. Os carcinomas são o tipo mais comum de câncer e podem ocorrer em várias partes do corpo, incluindo a pele, os órgãos internos, as glândulas e as membranas mucosas;
9. Carcinosarcoma: também conhecido como tumor maligno misto, é um tipo agressivo de câncer que pode ocorrer em vários órgãos do corpo. Essa neoplasia é chamada de carcinosarcoma porque é uma combinação de dois tipos diferentes de células cancerosas: as células carcinomatosas, que são de origem epitelial, e as células sarcomatosas, que são de origem mesenquimal. Isso significa que o tumor contém características tanto de tumores epiteliais quanto de tumores de tecido conjuntivo, tornando-o um câncer complexo e desafiador de tratar;
10. Cistadenocarcinoma papilar seroso: é um tipo de câncer que se origina a partir das células epiteliais que revestem cistos no ovário, tende a formar estruturas papilares e é composto por células epiteliais serosas. Esse tipo de câncer de ovário é geralmente considerado agressivo;
11. Cistadenocarcinoma seroso: refere-se a um tipo de câncer de ovário que se origina nas células epiteliais que revestem o ovário e que produzem um fluido seroso. Este é um tipo comum de câncer de ovário e é conhecido por sua agressividade e potencial de disseminação;
12. Cistadenocarcinoma: é um tipo de câncer que se origina nas células epiteliais que revestem as estruturas císticas em vários órgãos, mais comumente nos ovários e pâncreas. Esse câncer é caracterizado pela formação de cistos ou estruturas císticas malignas;

13. Cistadenocarcinoma papilar: é um tipo de câncer que se origina nas células epiteliais que revestem as estruturas císticas;
14. Neoplasia maligna: refere-se a um crescimento anormal de células no corpo que é canceroso e tem a capacidade de se espalhar para outras partes do organismo, invadindo os tecidos vizinhos e, potencialmente, formando metástases em órgãos distantes;
15. Sarcoma: é um tipo de câncer que se desenvolve no tecido muscular ou no tecido conjuntivo do útero;
16. Tumor papilar de superfície serosa, malignidade limítrofe: este tipo se desenvolve na superfície de um órgão ou na cavidade serosa do corpo.

## AGRADECIMENTOS

Minha gratidão vai, em primeiro lugar, a Deus, cuja infinita sabedoria e força me sustentaram ao longo de toda esta jornada. Agradeço à minha amada família — meu pai, Luiz José de Souza; minha mãe, Josefa Soares de Souza; e meus irmãos, Sebastião, Maria Tatiana e Luis Fernando Soares de Souza — que estiveram ao meu lado em cada passo, oferecendo amor, paciência e incentivo incondicional.

Sou profundamente grata a todos os professores do curso da UEPB, que contribuíram de diversas maneiras para o meu crescimento acadêmico e pessoal, compartilhando conhecimentos e experiências.

Agradeço especialmente à minha turma de Estatística e a todos os amigos que fiz durante essa jornada, em especial Hellen, Davi, Elyda, Marcela, Jéssica, Gislânia, Kailca, Pedro, Josefer-son, Andreza, Gabriel e Maria Karolina, pela paciência e ajuda constantes. Agradeço também a Cristiane e Thaislane, que, com seus momentos descontraídos durante os lanches, trouxeram leveza e alegria aos meus dias de estudo.

Expresso minha gratidão ao meu orientador, Dr. Tiago Almeida de Oliveira, e à coorientadora, Ma. Débora de Sousa Cordeiro, por aceitarem o convite para me guiar neste trabalho, oferecendo orientações e compartilhando ensinamentos fundamentais ao longo de toda a trajetória.

Agradeço ao professor Me. Cleanderson Romualdo Fidelis, membro da banca, por ministrar a disciplina de Análise de Sobrevivência de maneira clara e objetiva, despertando em mim o interesse por esse campo.

Por fim, expresso minha gratidão à Universidade Estadual da Paraíba, que tem sido uma parte importante da minha vida desde agosto de 2019, proporcionando-me a oportunidade de concluir o curso de Bacharelado em Estatística.

Este trabalho é o resultado do apoio e carinho de todos que acreditaram em mim e me encorajaram a seguir em frente. A todos, meus sinceros agradecimentos.