



**UNIVERSIDADE ESTADUAL PAULISTA**  
**“JÚLIO DE MESQUITA FILHO”**  
Faculdade de Ciências e Tecnologia  
Câmpus de Presidente Prudente

Programa de Pós-Graduação em Matemática Aplicada e Computacional

# Modelos de sobrevivência para estudo do tempo de internação hospitalar até a alta do paciente

Dina Rubi Ramos Eugenio

Orientador: Prof. Dr. Mário Hissamitsu Tarumoto

Presidente Prudente - SP  
Setembro - 2021

**UNIVERSIDADE ESTADUAL PAULISTA**

Faculdade de Ciências e Tecnologia de Presidente Prudente

Programa de Pós-Graduação em Matemática Aplicada e Computacional

**Modelos de sobrevivência para estudo do  
tempo de internação hospitalar até a alta do  
paciente**

Dina Rubi Ramos Eugenio

Orientador: Prof. Dr. Mário Hissamitsu Tarumoto

Dissertação apresentada ao Programa de Pós-Graduação em Matemática Aplicada e Computacional da Faculdade de Ciências e Tecnologia da UNESP como requisito para obtenção do título de Mestre em Matemática Aplicada e Computacional.

Presidente Prudente - SP

Setembro - 2021

E87m Eugenio, Dina Rubi Ramos  
Modelos de sobrevivência para estudo do tempo de internação hospitalar até a alta do paciente / Dina Rubi Ramos Eugenio. -- Presidente Prudente, 2021  
84 f. : il., tabs.

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências e Tecnologia, Presidente Prudente  
Orientador: Mário Hissamitsu Tarumoto

1. Distribuições de Probabilidade Gama-G. 2. Distribuição de Probabilidade Odd-Weibull. 3. Estimador de Máxima Verossimilhança. 4. Dados de Sobrevivência. 5. Internação Hospitalar. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências e Tecnologia, Presidente Prudente. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.



**UNIVERSIDADE ESTADUAL PAULISTA**

**Câmpus de Presidente Prudente**

**CERTIFICADO DE APROVAÇÃO**

**TÍTULO DA DISSERTAÇÃO:** Modelos de sobrevivência para estudo do tempo de internação hospitalar até a alta do paciente

**AUTORA: DINA RUBI RAMOS EUGENIO**

**ORIENTADOR: MÁRIO HISSAMITSU TARUMOTO**

Aprovada como parte das exigências para obtenção do Título de Mestra em MATEMÁTICA APLICADA E COMPUTACIONAL, pela Comissão Examinadora:

Prof. Dr. MÁRIO HISSAMITSU TARUMOTO (Participação Virtual)  
Departamento de Estatística / Faculdade de Ciências e Tecnologia de Presidente Prudente

Profa. Dra. JULIANA BETINI FACHINI GOMES (Participação Virtual)  
Departamento de Estatística / Universidade de Brasília

Profa. Dra. ELIZABETH MIE HASHIMOTO (Participação Virtual)  
Campus Londrina / Universidade Tecnológica Federal do Paraná

Presidente Prudente, 30 de julho de 2021

# Dedicatória

---

Dedico este trabalho de pesquisa principalmente a Deus, por ser o inspirador e me dar forças para continuar neste processo de obtenção de um dos meus desejos mais desejados.

Aos meus pais, pelo amor, trabalho e sacrifício em todos esses anos, graças a vocês consegui chegar aqui e ser o que sou. É um orgulho e privilégio ser filha deles, eles são os melhores pais.

Aos meus irmãos, por sempre estarem presentes, me acompanharem e pelo apoio moral que me deram ao longo desta fase da minha vida.

Ao meu noivo Yhon, por tornar minha vida cada dia mais feliz.

A todas as pessoas que me apoiaram e fizeram com que o trabalho fosse bem sucedido, especialmente aquelas que abriram as portas para mim e compartilharam seus conhecimentos.

---

# Agradecimentos

Agradeço a Deus por abençoar a minha vida, por me guiar ao longo da minha existência, por ser o suporte e a força nos momentos de dificuldade e fraqueza.

Aos meus pais: Nicolas Lucio Ramos Cáceres e Marcelina Eugenio Mamani, por serem os principais promotores dos meus sonhos, por confiarem e acreditarem em mim, pelos conselhos, valores e princípios que me inculcaram. Aos meus irmãos Jaime, Raúl e Roy que estão sempre lá me apoiando, cuidando de mim quando eu preciso deles.

Ao meu noivo Yhon, ofereço um agradecimento mais do que especial, por ter me aturado nos momentos de estresse e me incentivando durante todo o período de desenvolvimento deste trabalho.

Agradeço ao meu orientador Prof. Dr. Mario Hissamitsu Tarumoto pela paciência, dedicação, incentivo e excelente orientação no desenvolvimento deste trabalho.

À minha família por sempre me apoiar em todas as decisões e projetos que tenho, eles são uma parte muito importante da minha vida.

Aos professores do Programa de Pós-Graduação em Matemática Aplicada e Computacional da UNESP pelos conhecimentos transmitidos.

Aos funcionários da Seção de Pós-Graduação, em especial à Cinthia, pela atenção e apoio oferecido.

A todos os meus colegas da Pós-Graduação que me ajudaram.

À professora Patricia por me dar apoio em tudo.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

*"Statisticians are the modern explorers".*

***D.J. Hand***

# Resumo

---

O objetivo principal deste trabalho é estudar e analisar os tempos de internação hospitalar até a alta do paciente, com base nos dados fornecidos pelo Sistema de Informação Hospitalar do SUS (SIH/SUS) que armazena os dados das internações hospitalares no âmbito do SUS, informados mensalmente por todas as instituições de saúde que, após análise e aprovação, são encaminhados ao DATASUS para processamento. Para atingir o comportamento dos tempos de internação hospitalar a partir de uma distribuição os modelos de sobrevivência Odd-Weibull e Gama-G foram estudados. Os modelos Gama-G tem se mostrado ótimos para obter as melhores inferências para dados de tempos de internação hospitalar. No entanto, os Estimadores de Máxima Verossimilhança dos modelos Odd-Weibull e Gama-G não têm o comportamento estável e para muitos casos não é possível a sua estimação, apenas para amostras muito grandes. As estimativas dos parâmetros obtidas a partir dos modelos de dois parâmetros são consideradas bons valores iniciais para serem utilizados nos procedimentos de estimação dos parâmetros dos modelos Gama-G e Odd-Weibull, considerando dados censurados sob a abordagem Clássica. Além disso, utilizando o mesmo conceito proposto por Zografos & Balakrishnan (2009) foi derivada a distribuição Gama-Log-Normal. No entanto, entre estes modelos, o modelo Gama-Log-Logística, proposto por Hashimoto (2013) se mostrou mais adequado para ajustar os dados de tempos de internação em estudo. Para este modelo, foi construído o modelo de regressão com a inclusão das covariáveis sexo e idade, bem como o efeito de interação. A aplicação deste modelo mostrou que, em geral, as mulheres possuem um tempo de internação menor e os mais idosos possuem tempo de internação maior, além disso, o efeito de interação foi significativo, indicando que os homens mais idosos ficam menos tempo internados.

**Palavras-Chave:** *Distribuições de Probabilidade Gama-G, Distribuição de Probabilidade Odd-Weibull, Estimador de Máxima Verossimilhança, Dados de Sobrevivência, Internação Hospitalar.*

# Abstract

---

The main objective of this work is to study and analyze hospital stay times until the patient's discharge, based on data provided by the SUS Hospital Information System (SIH/SUS) that stores data on hospital admissions within the scope of SUS, reported monthly by all health institutions which, after analysis and approval, are sent to DATASUS for processing. To achieve the behavior of hospital stays from a distribution, the Odd-Weibull and Gamma-G survival models were studied. Gamma-G models have proven to be great for obtaining the best inferences for data on hospital stays. However, the Maximum Likelihood Estimators of the Odd-Weibull and Gamma-G models don't have a stable behavior and in many cases its estimation isn't possible, only for very large samples. The parameter estimates obtained from the two-parameter models are considered good initial values to be used in the parameter estimation procedures of the Gamma-G and Odd-Weibull models, considering data censored under the Classical approach. Furthermore, using the same concept proposed by Zografos & Balakrishnan (2009), the Gamma-Log-Normal distribution was derived. However, among these models, the Gamma-Log-Logistics model, proposed by Hashimoto (2013) proved to be more adequate to adjust the hospital stay data under study. For this model, a regression model was built with the inclusion of the covariates sex and age, as well as the interaction effect. The application of this model showed that, in general, women have a shorter hospital stay and the elderly have a longer hospital stay, in addition, the interaction effect was significant, indicating that older men spend less time in hospital.

**Keywords:** *Gamma-G Probability Distributions, Odd-Weibull Probability Distribution, Maximum Likelihood Estimator, Survival Data, Hospital Internment.*

---

# Lista de Figuras

|      |   |    |
|------|---|----|
| 1.1  | TTT - plot de diferentes distribuições, que indicam a forma da função de risco. . . . .                                       | 18 |
| 1.2  | Ilustração dos mecanismos de censura (a) do tipo I, (b) do tipo II e (c) do tipo aleatória. . . . .                           | 19 |
| 1.3  | Funções densidade de probabilidade e acumulada da distribuição Weibull para diferentes valores de $\alpha$ . . . . .          | 21 |
| 1.4  | Funções de sobrevivência e de risco da distribuição Weibull para diferentes valores de $\alpha$ . . . . .                     | 21 |
| 1.5  | Funções densidade de probabilidade e acumulada da distribuição Logística para diferentes valores de $\sigma$ . . . . .        | 22 |
| 1.6  | Funções de sobrevivência e de risco da distribuição Logística com diferentes valores de $\sigma$ . . . . .                    | 23 |
| 1.7  | Funções densidade de probabilidade e acumulada da distribuição Log-Logística para diferentes valores de $\alpha$ . . . . .    | 24 |
| 1.8  | Funções de sobrevivência e de risco da distribuição Log-Logística para diferentes valores de $\alpha$ . . . . .               | 24 |
| 1.9  | Funções densidade de probabilidade e acumulada da distribuição Log-Normal para diferentes valores de $\sigma$ . . . . .       | 25 |
| 1.10 | Funções de sobrevivência e de risco da distribuição Log-Normal com diferentes valores de $\sigma$ . . . . .                   | 26 |
| 1.11 | Funções densidade de probabilidade e acumulada da distribuição Gama para diferentes valores de $\phi$ . . . . .               | 27 |
| 1.12 | Funções de sobrevivência e de risco da distribuição Gama para diferentes valores de $\phi$ . . . . .                          | 27 |
| 2.1  | Funções densidade de probabilidade e acumulada da distribuição Odd-Weibull para diferentes valores de $\phi$ . . . . .        | 32 |
| 2.2  | Funções de sobrevivência e de risco da distribuição Odd-Weibull para diferentes valores de $\phi$ . . . . .                   | 33 |
| 2.3  | Funções densidade de probabilidade e acumulada da distribuição Gama-Weibull para diferentes valores de $\phi$ . . . . .       | 34 |
| 2.4  | Funções de sobrevivência e de risco da distribuição Gama-Weibull para diferentes valores de $\phi$ . . . . .                  | 35 |
| 2.5  | Funções densidade de probabilidade e acumulada da distribuição Gama-Log-Logística para diferentes valores de $\phi$ . . . . . | 36 |

---

|     |  |    |
|-----|--|----|
| 2.6 | Funções de sobrevivência e de risco da distribuição Gama-Log-Logística para diferentes valores de $\phi$ . . . . .   | 37 |
| 2.7 | Funções densidade de probabilidade e acumulada da distribuição Gama-Log-Normal para diferentes valores de $\phi$ . . . . .   | 38 |
| 2.8 | Funções de sobrevivência e de risco da distribuição Gama-Log-Normal para diferentes valores de $\phi$ . . . . .  | 39 |
| 4.1 | Funções densidade de probabilidade e acumulada da distribuição Gama-Log-Logística para diferentes valores de $\phi$ . . . . .  | 51 |
| 4.2 | Funções de sobrevivência e de risco da distribuição Gama-Log-Logística para diferentes valores de $\phi$ . . . . .   | 52 |
| 5.1 | Curva de Kaplan-Meier para os tempos de internação hospitalar. . . . .   | 58 |
| 5.2 | Curvas de sobrevivência estimada de Kaplan-Meier e dos modelos Weibull, Logística, Log-Logística, Log-Normal e Gama para os tempos de internação hospitalar. . . . .                                       | 60 |
| 5.3 | Curvas de sobrevivência estimadas por meio do ajuste dos modelos Odd-Weibull, Gama-Weibull, Gama-Log-Logística e Gama-Log-Normal <i>versus</i> a curva de sobrevivência estimada por Kaplan-Meier. . . . . | 64 |
| 5.4 | Gráficos das sobrevivências estimadas por Kaplan-Meier <i>versus</i> as sobrevivências estimadas pelos modelos Odd-Weibull e Gama-Weibull. . . . .   | 65 |
| 5.5 | Gráficos das sobrevivências estimadas por Kaplan-Meier <i>versus</i> as sobrevivências estimadas pelos modelos Gama-Log-Logística e Gama-Log-Normal. . . . .   | 65 |
| 5.6 | TTT-plot para os tempos de internação. . . . .   | 67 |
| 5.7 | Curvas de sobrevivência para o tempo de internação hospitalar, para a covariável sexo, estimadas pelo método de Kaplan-Meier . . . . .   | 68 |
| 5.8 | Curvas de sobrevivência estimadas pelo modelo de regressão Gama-Log-Logística para dois pacientes do sexo feminino e dois pacientes do sexo masculino internados no hospital. . . . .                      | 71 |
| 5.9 | Riscos estimados pelo modelo de regressão Gama-Log-Logística para dois pacientes do sexo feminino e dois pacientes do sexo masculino internados no hospital. . . . .                                       | 72 |

---

## Lista de Tabelas

|     |   |    |
|-----|---|----|
| 5.1 | Pacientes internados no período de janeiro a dezembro de 2020 (DATASUS).  | 56 |
| 5.2 | Estatísticas descritivas dos dos pacientes internados no hospital no período de janeiro a dezembro de 2020 (DATASUS).   | 57 |
| 5.3 | Estimativas da sobrevivência para os tempos de internação usando-se o estimador de Kaplan-Meier e os modelos Weibull, Logística, Log-Logística, Log-Normal e Gama.            | 59 |
| 5.4 | Estimadores de máxima verossimilhança dos parâmetros dos modelos ajustados para os tempos de internação hospitalar e os erros padrões (entre parênteses).                     | 61 |
| 5.5 | Estimativas da sobrevivência para os tempos de internação usando-se o estimador de Kaplan-Meier e os modelos Odd-Weibull, Gama-Weibull, Gama-Log-Logística e Gama-Log-Normal. | 63 |
| 5.6 | Estatísticas AIC e BIC de diferentes distribuições de probabilidade, considerando-se os dados da internação hospitalar.   | 66 |
| 5.7 | Estimadores de máxima verossimilhança dos parâmetros dos modelos Gama-Log-Logística para os dados de tempos de internação.  | 69 |
| 5.8 | Estatísticas AIC e BIC para os dados da internação hospitalar.  | 70 |

---

# Sumário

|   |           |
|---|-----------|
| <b>Resumo</b>   | <b>7</b>  |
| <b>Abstract</b>   | <b>8</b>  |
| <b>Lista de Figuras</b>   | <b>9</b>  |
| <b>Lista de Tabelas</b>   | <b>10</b> |
| <b>Introdução</b>   | <b>13</b> |
| <b>1 Revisão bibliográfica</b>  | <b>16</b> |
| 1.1 Dados de sobrevivência . . . . .  | 16        |
| 1.2 Gráfico TTT - plot . . . . .  | 17        |
| 1.3 Censura . . . . .   | 17        |
| 1.4 Representação dos dados de sobrevivência . . . . .                      | 20        |
| 1.5 Alguns modelos paramétricos . . . . .                                   | 20        |
| 1.5.1 Distribuição Weibull . . . . .  | 20        |
| 1.5.2 Distribuição Logística . . . . .                                      | 22        |
| 1.5.3 Distribuição Log-Logística . . . . .                                  | 23        |
| 1.5.4 Distribuição Log-Normal . . . . .                                     | 25        |
| 1.5.5 Distribuição Gama . . . . .   | 26        |
| 1.6 Análise clássica . . . . .  | 28        |
| 1.6.1 Estimador de máxima verossimilhança . . . . .                         | 28        |
| 1.7 Critérios de seleção de modelos . . . . .                               | 30        |
| <b>2 Distribuições de probabilidade generalizadas</b>                       | <b>31</b> |
| 2.1 Distribuição Odd-Weibull . . . . .                                      | 31        |
| 2.2 Distribuições de probabilidade Gama-G . . . . .                         | 33        |
| 2.2.1 Geradora de distribuição gama . . . . .                               | 33        |
| 2.2.2 Distribuição Gama-Weibull . . . . .                                   | 33        |
| 2.2.3 Distribuição Gama-Log-Logística . . . . .                             | 35        |
| 2.2.4 Distribuição Gama-Log-Normal . . . . .                                | 37        |
| <b>3 Estimação dos Parâmetros dos Modelos</b>                               | <b>40</b> |
| 3.1 Estimadores dos Parâmetros da Distribuição Odd-Weibull . . . . .        | 40        |
| 3.2 Estimadores dos Parâmetros da Distribuição Gama-Weibull . . . . .       | 42        |
| 3.3 Estimadores dos Parâmetros da Distribuição Gama-Log-Logística . . . . . | 44        |

|          |  |           |
|----------|--|-----------|
| 3.4      | Estimadores dos Parâmetros da Distribuição Gama-Log-Normal . . . . .   | 46        |
| <b>4</b> | <b>Modelo de regressão Gama-Log-Logística</b>                          | <b>50</b> |
| 4.1      | Distribuição Gama-Log-Logística . . . . .                              | 50        |
| 4.2      | Estimadores dos Parâmetros do modelo de regressão Gama-Log-Logística . | 53        |
| <b>5</b> | <b>Aplicação</b>   | <b>56</b> |
| 5.1      | Pacientes do Sistema de Informações Hospitalares do SUS . . . . .      | 56        |
| <b>6</b> | <b>Considerações Finais e Propostas Futuras</b>                        | <b>73</b> |
| 6.1      | Considerações Finais . . . . .   | 73        |
| 6.2      | Propostas Futuras . . . . .  | 73        |
| <b>A</b> | <b>Programas Utilizados</b>  | <b>74</b> |
| A.1      | Análise clássica considerando dados censurados . . . . .               | 74        |
| A.2      | Análise de regressão considerando dados censurados . . . . .           | 78        |
|          | <b>Referências</b>   | <b>82</b> |

---

# Introdução

O Sistema de Informação Hospitalar do SUS (SIH/SUS) disponibiliza informações sobre internações hospitalares. O instrumento que rege o SIH é a autorização de internação hospitalar (AIH). O SIH armazena dados de internações hospitalares no âmbito do SUS, informados mensalmente por todas as instituições públicas ou particulares conveniadas de saúde que realizam internações e consolidados por municípios e estados plenos que, após análise e aprovação, são encaminhados ao DATASUS para processamento (DATASUS, 2021). O SIH também fornece aos administradores, relatórios com informações para pagamento pela produção aos fornecedores. Monitora o desempenho dos hospitais em relação às metas estabelecidas nos contratos entre gestores e hospitais. Garante uma ferramenta de apoio às ações locais de controle, avaliação e auditoria. Além disso, auxilia no conhecimento ou construção do perfil de morbidade e mortalidade hospitalar. Por fim, avalia a qualidade da assistência à saúde oferecida a uma população.

Neste trabalho, o interesse é estudar os tempos de internação de pacientes internados nos hospitais do estado de São Paulo, no período de janeiro a dezembro de 2020. Os tempos são analisados em dias, desde o início da internação até a alta do paciente.

Portanto, modelos flexíveis serão considerados para facilitar a modelagem dos dados de tempos de internação hospitalar. Vários modelos de sobrevivência paramétricos muito populares como Weibull (Weibull *et al.*, 1951) tem sido amplamente utilizados nas últimas décadas para modelar dados em estudos de confiabilidade, engenharia e biológicos bem como médicos, no entanto, apresenta algumas dificuldades na modelagem de alguns tipos de dados. Consequentemente, um progresso significativo foi feito em direção à generalização de alguns modelos de sobrevivência bem conhecidos como podem ser vistos em Lawless (2003). Neste contexto, as distribuições probabilísticas ou modelos probabilísticos estudadas aqui têm sido utilizadas em análise de sobrevivência para modelar tempos de falha de produtos, materiais e situações clínicas. Neste caso particular, serão usadas para modelar os tempos de internação hospitalar para efeitos de comparação. Dentre elas, podem ser citadas as distribuições Weibull, Gama, Log-Logística, Log-Normal (Lawless, 2011), Gama-Weibull (Stacy *et al.*, 1962), Gama-Log-Logística (Hashimoto, 2013), Odd-Weibull (Cooray, 2006) entre outras.

A necessidade de formas estendidas da distribuição Gama surge em muitas áreas aplicadas (Hashimoto, 2013). Estas distribuições denominadas distribuições Gama-G, tem se mostrado muito flexível para modelar dados de sobrevivência pois acomoda várias formas da função de risco, sendo estas: constante, crescente, decrescente, forma de banheira (unimodal). Contém sub-modelos como Weibull, Gama, Log-Normal e Log-Logística.

A distribuição Odd-Weibull é outra extensão da distribuição Weibull (Cooray, 2006), conveniente para modelar e ajustar diferentes conjuntos de dados, especialmente na presença de censura devido à sua flexibilidade em acomodar a função de risco dependendo dos valores dos seus parâmetros. Por estas razões, a distribuição é uma boa alternativa para resolver os problemas na modelagem de dados de sobrevivência.

Para realizar a análise dos dados, utilizando-se dados reais obtidos do SIH/SUS como os tempos de internação hospitalar. Considera-se as distribuições Gama-Weibull, Gama-Log-Logística, Gama-Log-Normal e Odd-Weibull para analisar tais conjuntos de dados e logo comparar os resultados obtidos com outros modelos como as distribuições Weibull, Logística, Log-Logística, Log-Normal e Gama, através do critério AIC (Akaike, 1974).

Este trabalho apresenta 6 capítulos, dispostos da seguinte forma: no Capítulo 1, foi realizada uma revisão bibliográfica, apresentando alguns conceitos básicos de análise de sobrevivência, análise clássica e critérios de seleção de modelos; no Capítulo 2 é apresentado algumas distribuições de probabilidade generalizadas; no Capítulo 3 é apresentado o método de estimação dos parâmetros da distribuição escolhida no trabalho, considerando-se observações censuradas. O Capítulo 4 apresenta o modelo de regressão Gama-Log-Logística. Uma aplicação dos conceitos apresentados nos capítulos anteriores a um conjunto de dados reais obtidas do SIH/SUS é ilustrada no Capítulo 5. O Capítulo 6 trata das considerações finais, bem como possibilidades de pesquisas futuras. Em seguida, são apresentadas as referências bibliográficas. E por fim, alguns dos programas desenvolvidos no software R podem ser visualizados nos Apêndices.

# Revisão bibliográfica

Este capítulo, desenvolve alguns conceitos preliminares que são básicos para o desenvolvimento de capítulos posteriores. Algumas das referências nas quais o trabalho se baseia são encontradas em Colosimo & Giolo (2006); Kleinbaum & Klein (2010); Lee & Wang (2003); Lawless (2003); Liu (2012).

## Análise de sobrevivência

Análise de sobrevivência é uma área da Estatística em que o interesse é estimar o tempo até a ocorrência de um determinado evento, por exemplo, a morte de um paciente com alguma doença, o retorno de uma doença, ou o abandono de uma instituição financeira por um cliente, entre outros.

Aqui segue alguns conceitos importantes usados na análise de sobrevivência.

### 1.1 Dados de sobrevivência

Considere  $T$  uma variável contínua não-negativa, que representa, em geral, o resultado do acompanhamento de indivíduos no decorrer do tempo, até a ocorrência de um determinado evento de interesse, ver Colosimo & Giolo (2006). Seja  $f(\cdot)$  a função de densidade de probabilidade e  $F(\cdot)$  a função de distribuição acumulada. Define-se então a função de sobrevivência, denotada por  $S(\cdot)$  como a probabilidade de  $T$  ser maior que  $t$ :

$$S(t) = P(T > t) = \int_t^{\infty} f(u)du = 1 - F(t), \quad \text{para todo } t > 0.$$

$S(t)$  é uma função monótona decrescente e representa a probabilidade do evento de interesse não ocorrer em pelo menos  $t$  unidades de tempo.  $S(\cdot)$  possui as seguintes propriedades  $S(0) = 1$  e  $\lim_{t \rightarrow \infty} S(t) = 0$ .

Uma função importante na análise de sobrevivência é a função risco (taxa de falha) associada a  $T$  denotada por  $h(\cdot)$  e definida por:

$$h(t) = \lim_{\Delta t \rightarrow \infty} \frac{P[t < T \leq t + \Delta t | T > t]}{\Delta t},$$

sendo  $h(t) \geq 0$  e  $\int_0^\infty h(t)dt = \infty$ .

Algumas relações matemáticas entre as funções definidas acima podem ser estabelecidas:

$$S(t) = \exp \left[ - \int_0^t h(t)dt \right],$$

$$f(t) = - \frac{\partial S(t)}{\partial t}$$

e

$$h(t) = \frac{f(t)}{S(t)}.$$

Desta forma, o conhecimento de uma função, por exemplo  $F(t)$ , implica no conhecimento das outras funções,  $f(t)$ ,  $h(t)$ , e  $S(t)$ .

## 1.2 Gráfico TTT - plot

Barlow & Campo (1975) propuseram uma técnica gráfica para verificar o comportamento da função de risco, chamada de TTT-plot (tempo total em teste). O gráfico é construído com o plot consecutivo das quantidades  $[r/n, G(r/n)]$ , sendo  $G(r/n)$  a função dada por

$$G(r/n) = \frac{\sum_{i=1}^r Y_{i:n} - (n-r)Y_{r:n}}{\sum_{i=1}^r Y_{i:n}},$$

onde  $r = 1, \dots, n$ ,  $i = 1, \dots, n$  sendo  $Y_{i:n}$  as estatísticas de ordem da amostra.

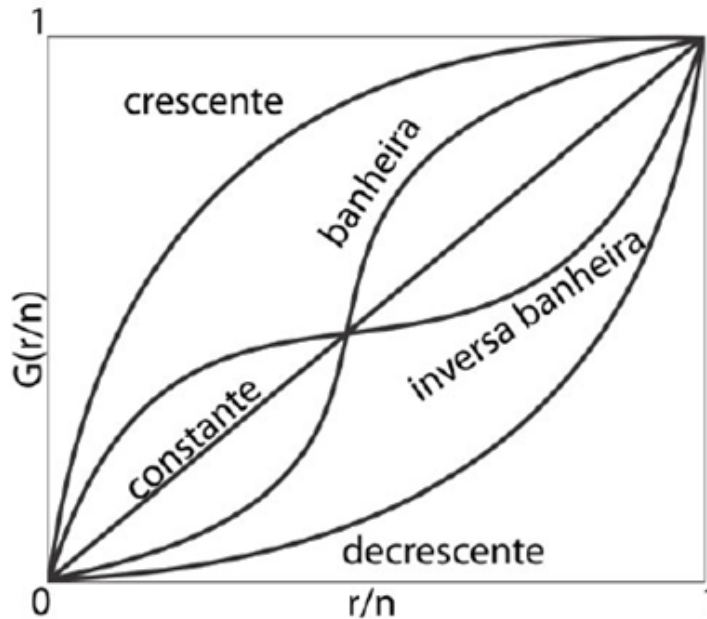
Utilizando-se a curva TTT em escala empírica, pode-se detectar o tipo de função de risco que os dados de tempo de vida possuem. Uma reta diagonal indica que os dados possuem uma função de risco constante. Se a curva for convexa, o risco é decrescente. Já a côncava indica que o risco é crescente. A forma convexa (côncava) e depois côncava (convexa) indica que a função tem risco em forma de banheira ou unimodal.

A Figura 1.1 a seguir mostra de forma mais clara, como verificar o comportamento da função de risco, utilizando-se o TTT-plot.

## 1.3 Censura

A principal característica nos dados de sobrevivência é a presença de censura, que é identificada como a observação parcial da resposta, ou seja, quando por algum motivo o acompanhamento do objeto em estudo é interrompido e a falha ou evento de interesse não é observado. Contudo, esta informação, embora parcial, deve ser considerada para possibilitar um estudo mais preciso do acontecimento de interesse. Os tipos de censura mais conhecidos são: censura à direita, censura à esquerda e censura intervalar. Neste

Figura 1.1: TTT - plot de diferentes distribuições, que indicam a forma da função de risco.



Fonte: Ramos (2014)

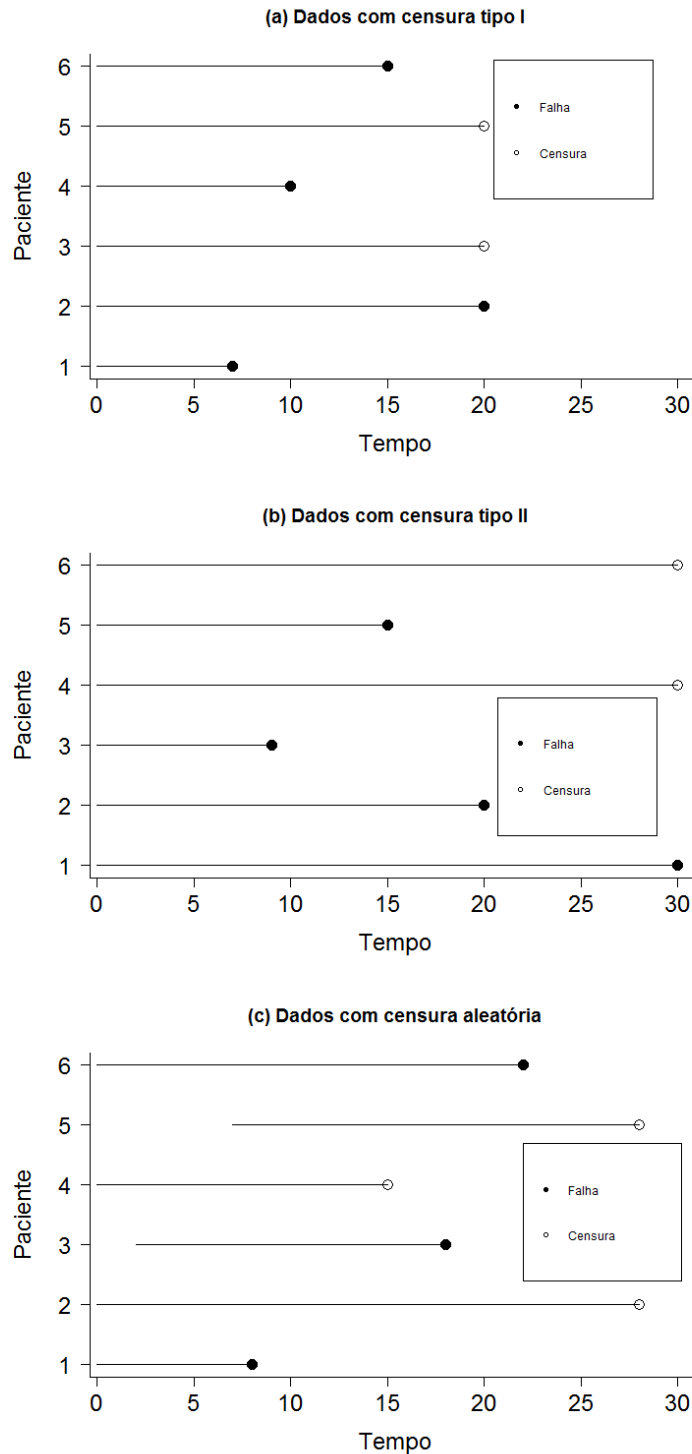
trabalho, considerou-se a censura à direita, onde o tempo de ocorrência do evento de interesse está à direita do tempo de censura observado.

Existem três mecanismos de censura à direita diferenciados entre eles e amplamente utilizados, são citados em Lawless (2011).

- **Censura tipo I:** Ocorre quando o tempo para o término do estudo é fixado antes do início e as observações que não apresentaram a falha até este instante são censuradas.
- **Censura tipo II:** Acontece quando uma quantidade de falhas a ser observada é previamente fixada. Quando essa quantidade é atingida, o estudo acaba e as observações que não falharam são ditas censuradas.
- **Censura aleatória:** Ocorre quando o componente em estudo é eliminado do experimento por ter ocorrido outro evento que não seja a falha de interesse. Este tipo de censura ocorre também com muita frequência em pesquisas médicas, quando um paciente é retirado do experimento, sem que tenha ocorrido a falha, ou que, por alguma razão, venha a morrer por uma causa diferente da estudada.

A Figura 1.2 ilustra os tipos de censura descritos. Em (a) ilustra-se a censura tipo I, no qual é estabelecido um tempo final do estudo em 20 unidades de tempo. Em (b) observa-se a censura do tipo II, onde o estudo acaba quando ocorre a  $n$ -ésima falha, e neste caso foi na 4ª falha. Em (c) tem-se a ilustração da censura aleatória, na qual o indivíduo pode entrar e/ou sair do estudo em qualquer momento.

Figura 1.2: Ilustração dos mecanismos de censura (a) do tipo I, (b) do tipo II e (c) do tipo aleatória.



Fonte: autoria própria.

## 1.4 Representação dos dados de sobrevivência

Os dados de sobrevivência para o indivíduo  $i$  ( $i = 1, \dots, n$ ) sob estudo são representados, em geral, pelo par  $(t_i, \delta_i)$  sendo  $t_i$  o tempo de falha ou de censura e  $\delta_i$  a variável indicadora de falha ou censura, isto é,

$$\delta_i = \begin{cases} 1 & \text{se } t_i \text{ é um tempo de falha,} \\ 0 & \text{se } t_i \text{ é um tempo censurado.} \end{cases}$$

Desta forma, a variável resposta em análise de sobrevivência é representada por duas colunas no banco de dados.

Na presença de covariáveis medidas no  $i$ -ésimo indivíduo, tais como  $x_i = (\text{sexo}_i, \text{idade}_i, \text{tratamento}_i)$ , os dados ficam representados por  $(t_i, \delta_i, x_i)$ .

## 1.5 Alguns modelos paramétricos

Em análise de sobrevivência o uso de modelos probabilísticos tem sido amplamente utilizado na literatura. Segundo Colosimo & Giolo (2006), os modelos Exponencial, Weibull e Log-Normal destacam-se na adaptação a diversas situações práticas. Neste contexto, apresenta-se o uso de distribuições de probabilidade na análise estatística de dados de sobrevivência. Tais distribuições, denominadas modelos probabilísticos ou paramétricos, descrevem adequadamente várias situações reais.

Aqui segue algumas distribuições importantes usadas na análise de sobrevivência.

### 1.5.1 Distribuição Weibull

A distribuição Weibull é frequentemente usada em estudos biomédicos e industriais e foi proposta por Weibull *et al.* (1951). Essa distribuição é bastante utilizada, pois possui diferentes formas e sua propriedade básica é que a função risco é monótona, deste modo, ela é crescente, decrescente ou constante.

Uma variável aleatória  $T$  com distribuição Weibull denotada por  $Weibull(\alpha, \mu)$ , possui função densidade de probabilidade (f.d.p.) dada por:

$$f(t|\boldsymbol{\theta}) = \mu\alpha (\mu t)^{\alpha-1} \exp\{-(\mu t)^\alpha\}, \quad (1.1)$$

sendo  $t > 0$ ,  $\boldsymbol{\theta} = (\alpha, \mu)^\top$ . O parâmetro de forma é  $\alpha > 0$  e o parâmetro de escala é  $\mu > 0$ .

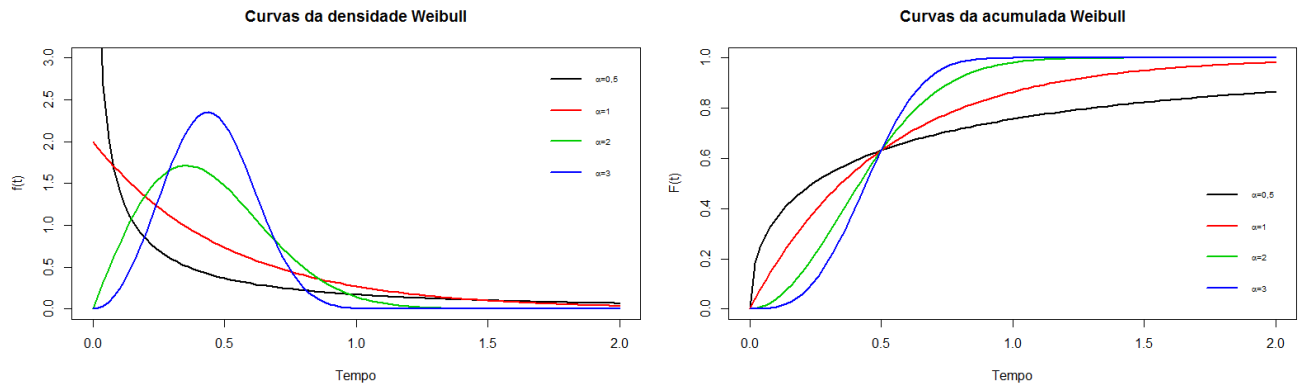
E sua correspondente função distribuição acumulada (F.D.A.) é dada por:

$$F(t|\boldsymbol{\theta}) = 1 - \exp\{-(\mu t)^\alpha\}, \quad (1.2)$$

para todo  $t > 0$ ,  $\alpha > 0$  e  $\mu > 0$ .

Pode-se observar na Figura 1.3 diferentes formas para a f.d.p. e F.D.A. da distribuição Weibull (1.1) para  $\mu = 2$  e diferentes valores de  $\alpha$ . Por meio dessa figura, observa-se que à medida que o valor do parâmetro  $\alpha$  aumenta a forma da função densidade de probabilidade é deslocada à direita.

Figura 1.3: Funções densidade de probabilidade e acumulada da distribuição Weibull para diferentes valores de  $\alpha$ .



Fonte: autoria própria.

Fonte: autoria própria.

A média e a variância da distribuição Weibull (1.1) são dadas respectivamente por:

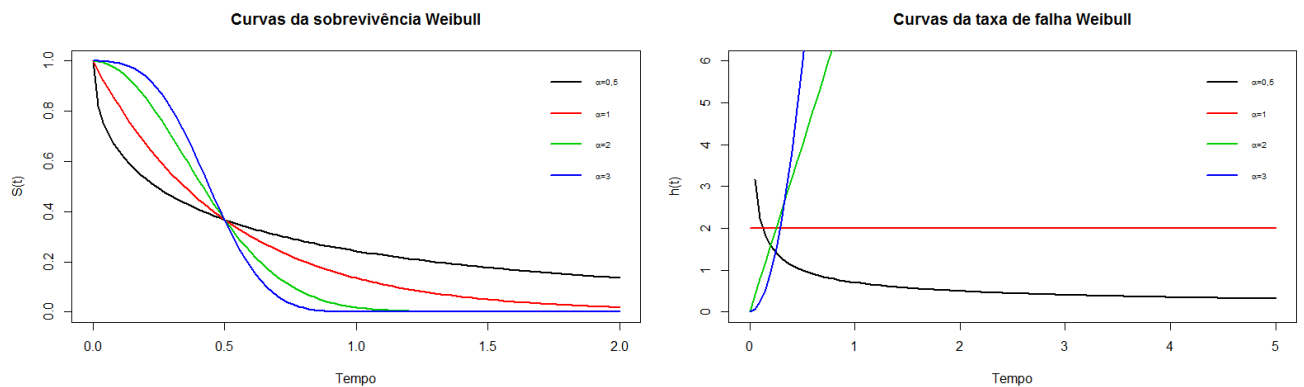
$$E(T) = \frac{1}{\mu} \Gamma \left( 1 + \frac{1}{\alpha} \right) \quad \text{e} \quad \text{Var}(T) = \frac{1}{\mu^2} \left\{ \Gamma \left( 1 + \frac{2}{\alpha} \right) - \left[ \Gamma \left( 1 + \frac{1}{\alpha} \right) \right]^2 \right\}.$$

A função de sobrevivência e a função risco são representadas por:

$$S(t|\boldsymbol{\theta}) = \exp \{ - (\mu t)^\alpha \} \quad \text{e} \quad h(t|\boldsymbol{\theta}) = \mu \alpha (\mu t)^{\alpha-1}.$$

Observa-se na Figura 1.4 diferentes formas para a função de sobrevivência e a função de risco da distribuição Weibull.

Figura 1.4: Funções de sobrevivência e de risco da distribuição Weibull para diferentes valores de  $\alpha$ .



Fonte: autoria própria.

Fonte: autoria própria.

### 1.5.2 Distribuição Logística

A função de crescimento logístico foi proposta inicialmente para estudos demográficos, isto é, para estudos de crescimento populacional humano. No livro Johnson *et al.* (1995), também cita aplicações como dados de produção agrícola e crescimento de mortalidades. A definição é apresentada a seguir.

Seja  $T$  uma variável aleatória contínua, com uma distribuição de probabilidade logística denotado por  $L(\mu, \sigma)$ , diz-se que  $T$  tem distribuição Logística com parâmetros de localização  $\mu$  e de escala  $\sigma$ , se a f.d.p. é dada por:

$$f(t|\boldsymbol{\theta}) = \frac{\exp\left(\frac{t-\mu}{\sigma}\right)}{\sigma \left[1 + \exp\left(\frac{t-\mu}{\sigma}\right)\right]^2}, \quad (1.3)$$

sendo  $t > 0$ ,  $\boldsymbol{\theta} = (\mu, \sigma)^\top$ . O parâmetro de localização é  $\mu \in \mathbb{R}$  e o parâmetro de escala é  $\sigma > 0$ .

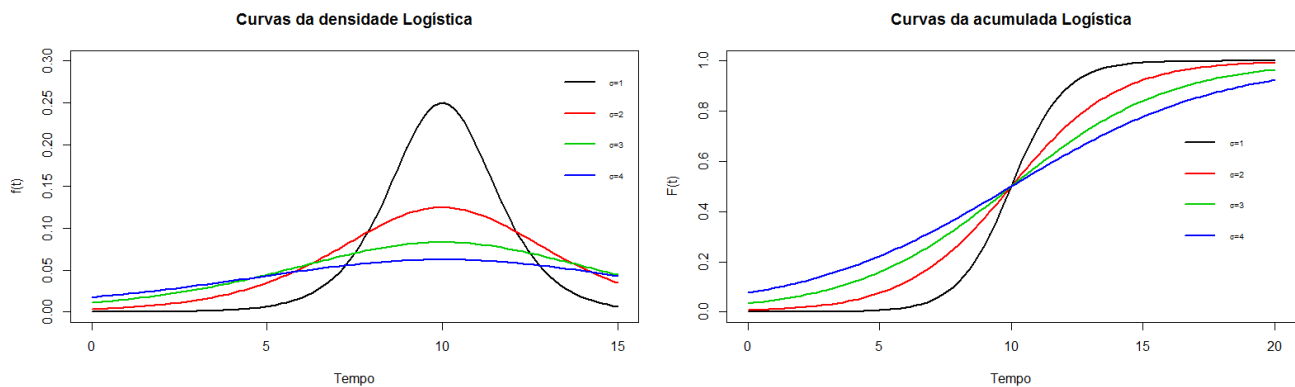
E sua correspondente F.D.A. é dada por:

$$F(t|\boldsymbol{\theta}) = 1 - \left[1 + \exp\left(\frac{t-\mu}{\sigma}\right)\right]^{-1},$$

para todo  $t > 0$ ,  $\mu \in \mathbb{R}$  e  $\sigma > 0$ .

Pode-se observar na Figura 1.5 diferentes formas para a f.d.p. e a F.D.A. da distribuição Logística (1.3) para  $\mu = 10$  e diferentes valores de  $\sigma$ . Por meio dessa figura, observa-se que à medida que o valor do parâmetro  $\sigma$  aumenta a forma da função densidade de probabilidade é mais achatada.

Figura 1.5: Funções densidade de probabilidade e acumulada da distribuição Logística para diferentes valores de  $\sigma$ .



Fonte: autoria própria.

Fonte: autoria própria.

A média e a variância da distribuição Logística (1.3) são dadas respectivamente por:

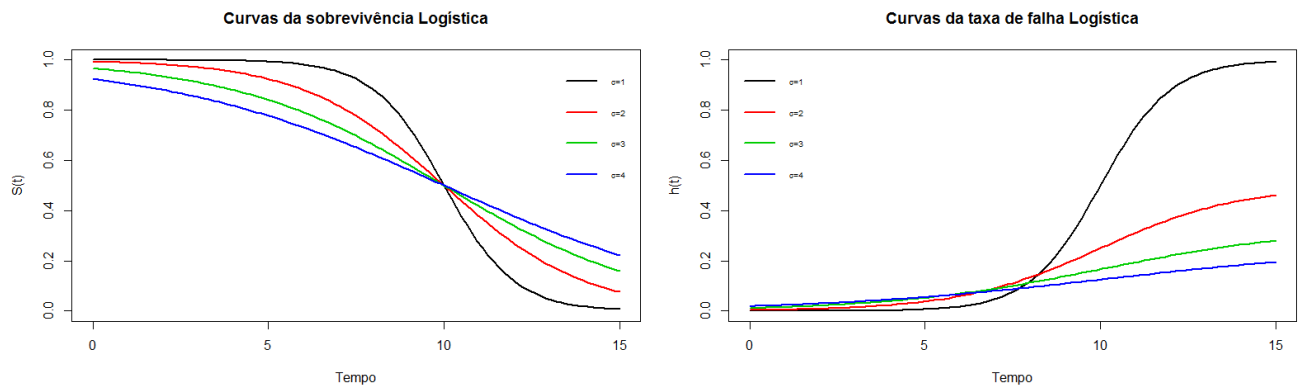
$$E(T) = \mu \quad \text{e} \quad \text{Var}(T) = \frac{\sigma^2 \pi^2}{3}.$$

As funções de sobrevivência e de risco são dadas respectivamente por:

$$S(t|\boldsymbol{\theta}) = \left[ 1 + \exp\left(\frac{t-\mu}{\sigma}\right) \right]^{-1} \quad \text{e} \quad h(t|\boldsymbol{\theta}) = \frac{\exp\left(\frac{t-\mu}{\sigma}\right)}{\sigma \left[ 1 + \exp\left(\frac{t-\mu}{\sigma}\right) \right]}.$$

Observa-se na Figura 1.6 diferentes formas para a função de sobrevivência e a função de risco da distribuição Logística.

Figura 1.6: Funções de sobrevivência e de risco da distribuição Logística com diferentes valores de  $\sigma$ .



Fonte: autoria própria.

Fonte: autoria própria.

### 1.5.3 Distribuição Log-Logística

Esta distribuição foi apresentada por Bennett (1983) e acaba sendo uma boa alternativa ao modelo Weibull por ser simples, mais flexível quanto à forma da taxa de falha.

Seja  $Y$  uma variável aleatória não negativa com distribuição Logística, sua função densidade de probabilidade é a expressão definida em (1.3). Considere a seguinte transformação:  $T = \log Y$ . Desta forma

$$f(t|\boldsymbol{\theta}) = \alpha \mu^\alpha t^{\alpha-1} [1 + (\mu t)^\alpha]^{-2}, \quad (1.4)$$

sendo  $t > 0$ ,  $\boldsymbol{\theta} = (\alpha, \mu)^\top$ . O parâmetro de forma é  $\alpha > 0$  e o parâmetro de escala é  $\mu > 0$ .

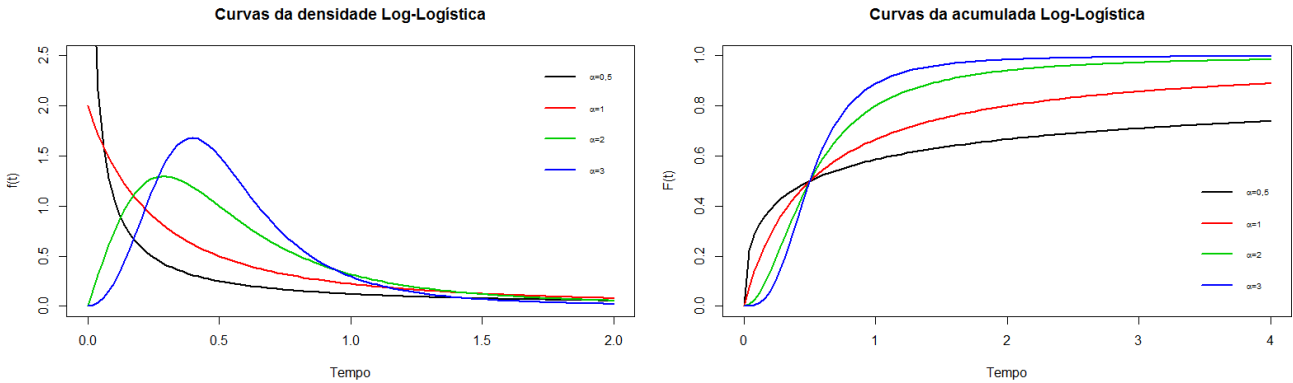
E sua correspondente F.D.A. é dada por:

$$F(t|\boldsymbol{\theta}) = 1 - [1 + (\mu t)^\alpha]^{-1}, \quad (1.5)$$

para todo  $t > 0$ ,  $\alpha > 0$  e  $\mu > 0$ .

A Figura (1.7) apresenta as curvas de densidade e acumulada da distribuição Log-Logística (1.4) para  $\mu = 2$  e diferentes valores de  $\alpha$ . Por meio dessa figura, observa-se que à medida que o valor do parâmetro  $\alpha$  aumenta a forma da função densidade de probabilidade é deslocada à direita.

Figura 1.7: Funções densidade de probabilidade e acumulada da distribuição Log-Logística para diferentes valores de  $\alpha$ .



Fonte: autoria própria.

Fonte: autoria própria.

A média e a variância da distribuição Log-Logística (1.4) são dadas respectivamente por:

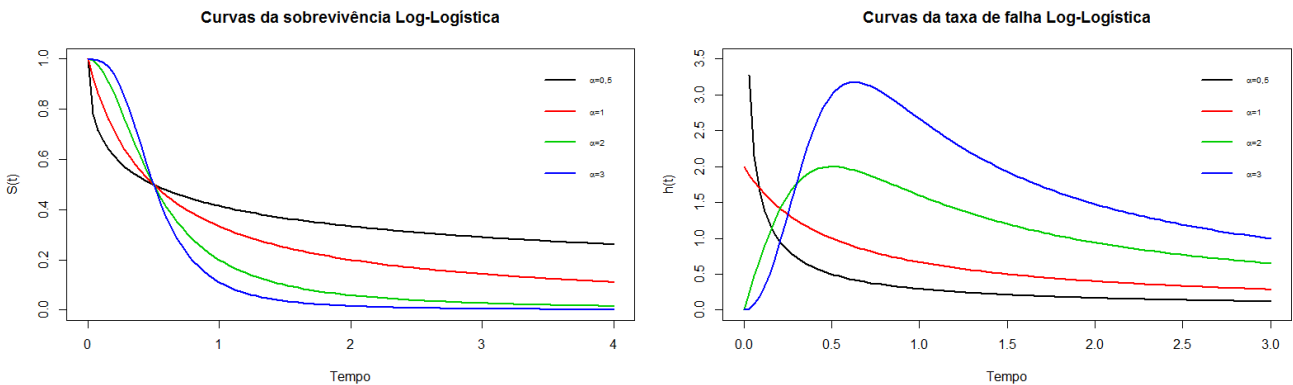
$$E(T) = \frac{\pi \csc(\pi/\alpha)}{\mu\alpha} \quad \text{e} \quad Var(T) = \frac{2\pi \csc(2\pi/\alpha)}{\mu^2\alpha} - \left(\frac{\pi \csc(\pi/\alpha)}{\mu\alpha}\right)^2.$$

A função de sobrevivência e a função risco são representadas por:

$$S(t|\theta) = [1 + (\mu t)^\alpha]^{-1} \quad \text{e} \quad h(t|\theta) = \mu\alpha (\mu t)^{\alpha-1} [1 + (\mu t)^\alpha]^{-1}.$$

Na Figura 1.8, observa-se as curvas de risco e sobrevivência da variável aleatória  $T$  com distribuição Log-Logística.

Figura 1.8: Funções de sobrevivência e de risco da distribuição Log-Logística para diferentes valores de  $\alpha$ .



Fonte: autoria própria.

Fonte: autoria própria.

### 1.5.4 Distribuição Log-Normal

A distribuição Log-Normal tem sido utilizada como modelo em diversas aplicações em engenharia, medicina entre outras. Se  $Y$  possui uma distribuição Normal com média  $\mu$  e variância  $\sigma^2$  e  $T = \exp(Y)$ , então diz-se que  $T$  possui uma distribuição Log-Normal e sua f.d.p. é dada por:

$$f(t|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\sigma t} \exp \left[ -\frac{1}{2} \left( \frac{\log(t) - \mu}{\sigma} \right)^2 \right], \quad (1.6)$$

sendo  $t > 0$ ,  $\boldsymbol{\theta} = (\mu, \sigma)^\top$ . O parâmetro de localização é  $\mu \in \mathbb{R}$  e o parâmetro de escala é  $\sigma > 0$ .

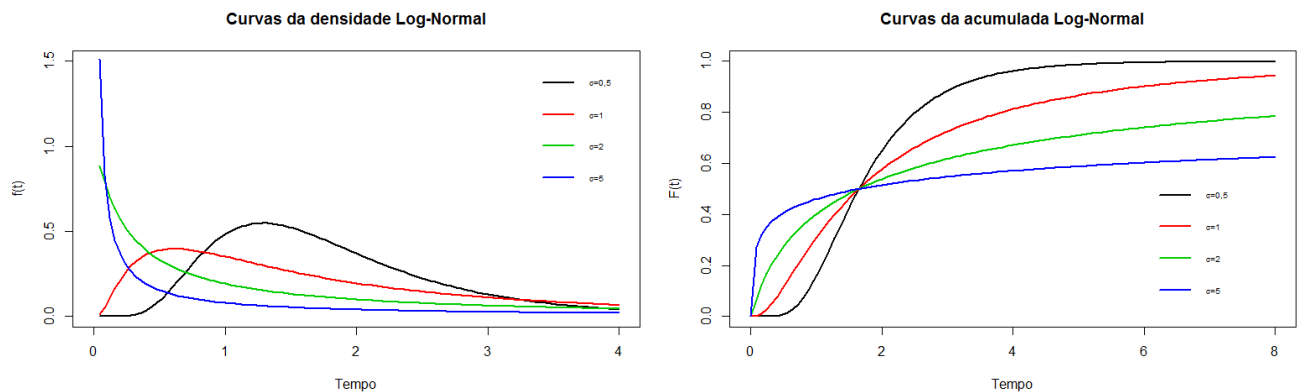
E sua correspondente F.D.A. é dada por:

$$F(t|\boldsymbol{\theta}) = 1 - \Phi \left( \frac{-\log(t) + \mu}{\sigma} \right), \quad (1.7)$$

para todo  $t > 0$ ,  $\sigma > 0$  e  $\mu \in \mathbb{R}$ .

Pode-se ver na Figura 1.9 diferentes formas para a f.d.p. e F.D.A. da distribuição Log-Normal (1.6) para  $\mu = 0,5$  e diferentes valores de  $\sigma$ . Por meio dessa figura, observa-se que à medida que o valor do parâmetro  $\sigma$  aumenta a forma da função densidade de probabilidade é deslocada à esquerda.

Figura 1.9: Funções densidade de probabilidade e acumulada da distribuição Log-Normal para diferentes valores de  $\sigma$ .



Fonte: autoria própria.

Fonte: autoria própria.

A média e variância da distribuição Log-Normal (1.6) são dadas respectivamente por

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

As funções de sobrevivência e de risco são dadas respectivamente por

$$S(t|\boldsymbol{\theta}) = \Phi \left( \frac{-\log(t) + \mu}{\sigma} \right)$$

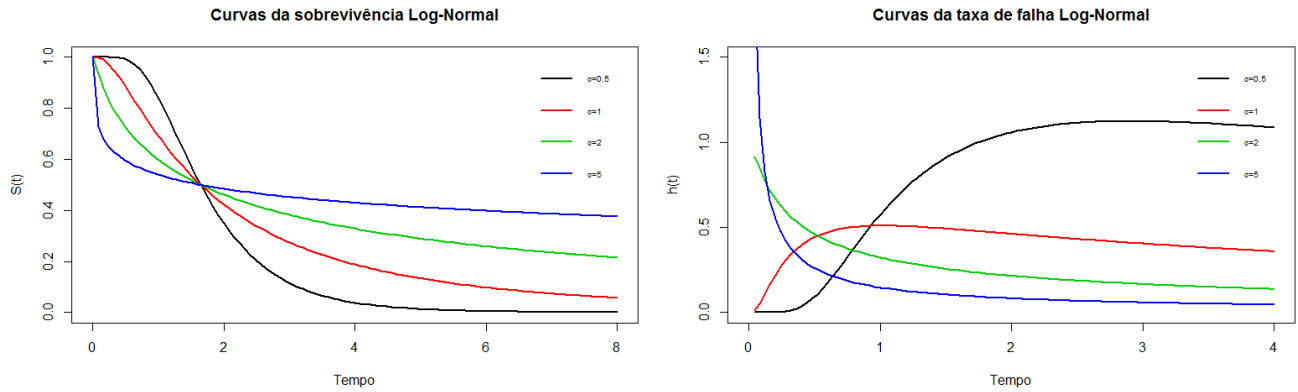
e

$$h(t|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\sigma t \left[ \Phi \left( \frac{-\log(t) + \mu}{\sigma} \right) \right]} \exp \left[ -\frac{1}{2} \left( \frac{\log(t) - \mu}{\sigma} \right)^2 \right],$$

sendo  $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp \left( \frac{-u^2}{2} \right) du$ .

Observa-se na Figura 1.10 diferentes formas para a função de sobrevivência e a função de risco da distribuição Log-Normal.

Figura 1.10: Funções de sobrevivência e de risco da distribuição Log-Normal com diferentes valores de  $\sigma$ .



Fonte: autoria própria.

Fonte: autoria própria.

### 1.5.5 Distribuição Gama

A distribuição Gama é comumente utilizada para se resolver problemas em análise de confiabilidade, previsões do tempo, para modelar seguros de crédito, entre outros.

Seja  $T$  uma variável aleatória com distribuição de probabilidade Gama denotado por  $Gama(\phi, \mu)$ , cuja função densidade de probabilidade é dada por:

$$f(t|\boldsymbol{\theta}) = \frac{\mu^\phi}{\Gamma(\phi)} t^{\phi-1} \exp(-\mu t), \quad (1.8)$$

sendo  $t > 0$ ,  $\boldsymbol{\theta} = (\phi, \mu)^\top$ . O parâmetro de forma é  $\phi > 0$  e o parâmetro de escala é  $\mu > 0$ . Além disso,  $\Gamma(\phi)$  a função gama definida como

$$\Gamma(\phi) = \int_0^\infty e^{-x} x^{\phi-1} dx.$$

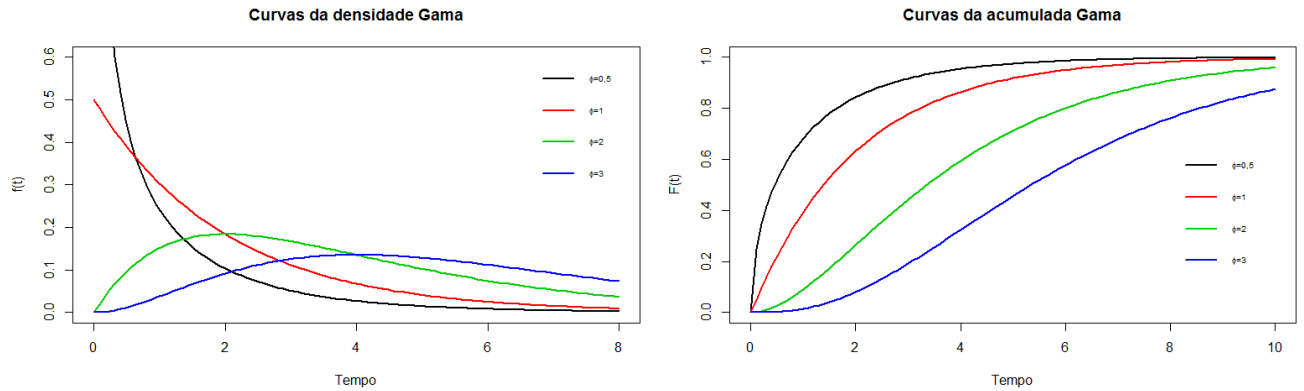
E sua correspondente F.D.A. é dada por:

$$F(t|\boldsymbol{\theta}) = 1 - \frac{\Gamma[\mu t, \phi]}{\Gamma(\phi)}.$$

para todo  $t > 0$ ,  $\phi > 0$  e  $\mu > 0$ .

Pode-se observar na Figura 1.11 diferentes formas para a f.d.p. e F.D.A. distribuição Gama (1.8) para  $\mu = 0,5$  e diferentes valores de  $\phi$ . Por meio dessa figura, observa-se que à medida que o valor do parâmetro  $\phi$  aumenta a forma da função densidade de probabilidade é deslocada à direita.

Figura 1.11: Funções densidade de probabilidade e acumulada da distribuição Gama para diferentes valores de  $\phi$ .



Fonte: autoria própria.

Fonte: autoria própria.

A média e a variância da distribuição Gama (1.8) são dadas respectivamente por:

$$E(T) = \frac{\phi}{\mu} \quad \text{e} \quad Var(T) = \frac{\phi}{\mu^2}.$$

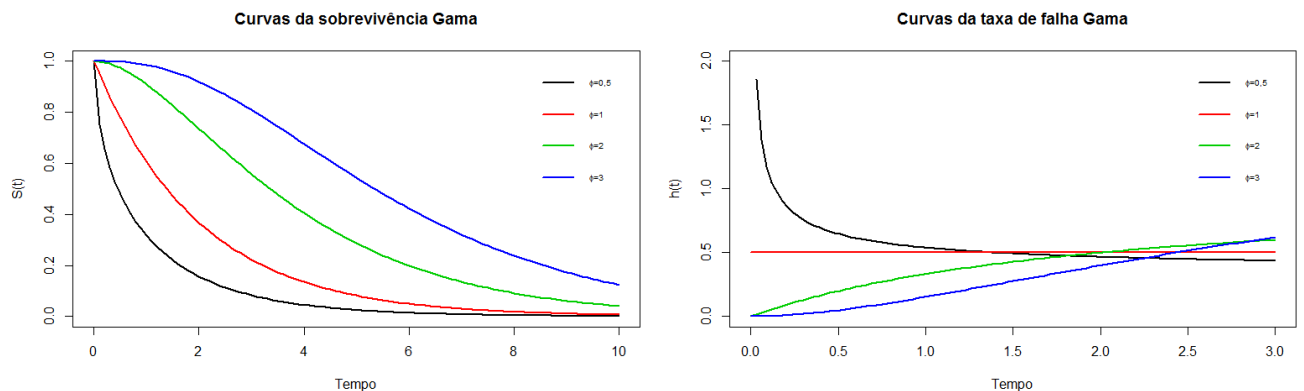
As funções de sobrevivência e de risco são dadas respectivamente por:

$$S(t|\boldsymbol{\theta}) = \frac{\Gamma[\mu t, \phi]}{\Gamma(\phi)} \quad \text{e} \quad h(t|\boldsymbol{\theta}) = \frac{\mu t^{\phi-1} \exp(-\mu t)}{\Gamma[\mu t, \phi]},$$

sendo  $\Gamma[y, x] = \int_x^\infty w^{y-1} e^{-w} dw$  denominada de função gama incompleta superior.

Observa-se na Figura 1.12 diferentes formas para a função de sobrevivência e a função de risco da distribuição Gama.

Figura 1.12: Funções de sobrevivência e de risco da distribuição Gama para diferentes valores de  $\phi$ .



Fonte: autoria própria.

Fonte: autoria própria.

## 1.6 Análise clássica

Ao se realizar uma análise estatística, empregam-se, comumente, dois contextos: o clássico (ou frequentista) e o Bayesiano.

Na Análise Clássica, o vetor de parâmetro  $\boldsymbol{\theta}$  desconhecido é tratado como uma constante fixa. Uma forma de se avaliar esse vetor parâmetro é encontrar um estimador pontual para  $\boldsymbol{\theta}$ . Sendo  $\mathbf{t} = (t_1, t_2, \dots, t_n)$  um vetor de  $n$  observações geradas por um modelo probabilístico  $f(\mathbf{t}|\boldsymbol{\theta})$ , é então escolhido algum método de estimação, tais como estimador de máxima verossimilhança, mínimos quadrados, método dos momentos, entre outros. Imaginando-se todos os possíveis conjuntos de dados  $t_1, t_2, \dots, t_n$  que poderiam ser gerados por  $f(\mathbf{t}|\boldsymbol{\theta})$  para um dado valor de  $\boldsymbol{\theta}$ , obtêm-se então a correspondente distribuição amostral de  $\hat{\boldsymbol{\theta}}$ .

### 1.6.1 Estimador de máxima verossimilhança

Dentre os métodos estatísticos de inferência Clássica, o método de máxima verossimilhança é preferível devido às suas melhores propriedades assintóticas. No método de máxima verossimilhança, os estimadores são obtidos a partir da maximização da função de verossimilhança (ver, por exemplo, Casella & Berger (2021)).

A função de verossimilhança de  $\boldsymbol{\theta}$ , correspondente à amostra aleatória observada, é dada por:

$$L(\boldsymbol{\theta}, \mathbf{t}) = \prod_{i=1}^n f(t_i|\boldsymbol{\theta}), \quad (1.9)$$

que resume a informação a respeito de  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ , contida nos dados amostrais  $\mathbf{t}$ .

Na presença de censura, altera-se a função de verossimilhança do modelo probabilístico. No caso em que não há censura presente na amostra, a função de verossimilhança é o produto das densidades completas (1.9). Na presença de censuras à direita, não se sabe exatamente quando a falha ocorre, sabe-se somente que o intervalo do tempo de falha está  $(t, \infty)$ , assim, a contribuição na função de verossimilhança também estará neste intervalo, então  $P[T > t] = S(t)$ . A função de verossimilhança para dados com censuras à direita será escrita na forma:

$$L(\boldsymbol{\theta}, \mathbf{t}) = \prod_{i=1}^d f(t_i|\boldsymbol{\theta}) \prod_{i=d+1}^n S(t_i|\boldsymbol{\theta}), \quad (1.10)$$

em que  $d$  é o número de observações que não foram censuradas.

A equação (1.10) pode ser escrita também como:

$$L(\boldsymbol{\theta}, \mathbf{t}) = \prod_{i=1}^n f(t_i|\boldsymbol{\theta})^{\delta_i} S(t_i|\boldsymbol{\theta})^{1-\delta_i} = \prod_{i=1}^n h(t_i|\boldsymbol{\theta})^{\delta_i} S(t_i|\boldsymbol{\theta}), \quad (1.11)$$

sendo  $\delta_i$  a variável indicadora da censura.

Para um modelo com número de parâmetros  $k$ , se a função de verossimilhança for diferenciável em  $\theta_i$ , as equações de verossimilhança são obtidas como a solução do sistema de equações:

$$\frac{\partial}{\partial \theta_i} \log(L(\boldsymbol{\theta}, \mathbf{t})) = 0, \quad i = 1, 2, \dots, k. \quad (1.12)$$

As soluções deste sistema (1.12) fornecem os estimadores de máxima verossimilhança. Em muitos casos, métodos numéricos como os de Newton-Raphson são necessários para se encontrar tais soluções.

O algoritmo de Newton-Raphson é representado através das atualizações sucessivas de

$$\boldsymbol{\theta}^{(n+1)} = \boldsymbol{\theta}^{(n)} - [J(\boldsymbol{\theta}^{(n)})]^{-1} G(\boldsymbol{\theta}^{(n)}),$$

sendo  $G(\boldsymbol{\theta})$ , o vetor contendo as derivadas parciais de  $\log(L(\boldsymbol{\theta}, \mathbf{t}))$  e  $J(\boldsymbol{\theta})$  a matriz de segundas derivadas parciais de  $\log(L(\boldsymbol{\theta}, \mathbf{t}))$  em relação a  $\boldsymbol{\theta}$ .

Os estimadores de máxima verossimilhança de  $\boldsymbol{\theta}$  são viciados para amostras pequenas. No caso de grandes amostras, tornam-se não-viciados e assintoticamente eficientes. Tais estimadores têm distribuição conjunta assintoticamente normal dada por:

$$(\hat{\boldsymbol{\theta}}) \sim N_k [\boldsymbol{\theta}, I^{-1}(\boldsymbol{\theta})] \quad \text{para } n \rightarrow \infty,$$

em que  $I(\boldsymbol{\theta})$  é a matriz de Informação de Fisher resultante de

$$I(\boldsymbol{\theta}) = \begin{bmatrix} I_{11}(\boldsymbol{\theta}) & \cdots & I_{1k}(\boldsymbol{\theta}) \\ \vdots & \ddots & \vdots \\ I_{1k}(\boldsymbol{\theta}) & \cdots & I_{kk}(\boldsymbol{\theta}) \end{bmatrix}$$

sendo  $I_{ij}(\boldsymbol{\theta})$ , a quantidade denominada Informação de Fisher de  $\boldsymbol{\theta}$  em  $i$  e  $j$  dada por:

$$I_{ij}(\boldsymbol{\theta}) = E \left[ \left( \frac{\partial}{\partial \theta_i \partial \theta_j} \log(f(\mathbf{t}|\boldsymbol{\theta})) \right)^2 \right], \quad i, j = 1, 2, \dots, k.$$

Quando o tamanho amostral é grande, pode-se construir intervalos de confiança aproximados para os parâmetros individuais  $\theta_i$ , com coeficiente de confiança  $100(1 - \gamma)\%$ , através das distribuições marginais dadas por:

$$(\theta_i) \sim N [\boldsymbol{\theta}, I_{ii}^{-1}(\boldsymbol{\theta})] \quad \text{para } n \rightarrow \infty.$$

Com estes resultados é possível construir intervalos de confiança individuais aproximados.

O estimador de máxima verossimilhança possui uma propriedade muito importante conhecida como propriedade de invariância (Casella & Berger, 2021). Isto é, suponha que exista pelo menos uma transformação um a um que leva  $\boldsymbol{\theta}$  a  $\tau(\boldsymbol{\theta})$ , utilizando-se a propriedade de invariância do estimador de máxima verossimilhança, pode-se afirmar que se  $\hat{\boldsymbol{\theta}}$  é um estimador de máxima verossimilhança de  $\boldsymbol{\theta}$  para qualquer função de  $\tau(\boldsymbol{\theta})$ , o estimador de máxima verossimilhança de  $\tau(\boldsymbol{\theta})$  é  $\tau(\hat{\boldsymbol{\theta}})$ .

## 1.7 Critérios de seleção de modelos

Em problemas que envolvem incerteza, utilizar critérios que auxiliem na comparação e seleção de modelos paramétricos é de fundamental importância em uma análise de dados. Neste tópico serão discutidos dois diferentes critérios de comparação de modelos: o AIC e o BIC.

Akaike (1974) propõe um método baseado na medida de Informação de Kullback-Leibler. Seja  $k$  o número de parâmetros a serem estimados,  $n$  o número de observações de  $\mathbf{t}$  e  $\hat{\boldsymbol{\theta}}$  uma estimativa de  $\boldsymbol{\theta}$ , o critério de informação de Akaike (AIC) é obtido por:

$$AIC = -2 \log \left( L(\hat{\boldsymbol{\theta}}; \mathbf{t}) \right) + 2k.$$

Dado um conjunto de modelos candidatos para  $\mathbf{t}$ , ajustados os dados, o preferido será o que fornecer o menor AIC. Além de selecionar um ótimo ajuste, o critério penaliza a adição de parâmetros, desencorajando *overfitting*, ou seja, a seleção de um modelo extremamente complexo e com muitos parâmetros que tenham um pobre desempenho preditivo.

Outro critério muito utilizado é o Critério de Informação Bayesiana (BIC), desenvolvido por Schwarz *et al.* (1978). Este critério possui estrutura semelhante ao AIC. O BIC é obtido através de resultados assintóticos e da suposição de que os dados pertencem à família exponencial, a equação é dada por:

$$BIC = -2 \log \left( L(\hat{\boldsymbol{\theta}}; \mathbf{t}) \right) + k \log(n).$$

Semelhante ao AIC, o critério penaliza a adição de parâmetros. Dos modelos candidatos, o adequado será o que possuir o menor BIC.

# Distribuições de probabilidade generalizadas

Neste capítulo, se desenvolve alguns modelos de sobrevivência paramétricos generalizados. Vários modelos de sobrevivência paramétricos muito populares como Weibull, Logística, Log-Logística, Log-Normal e Gama, apresentam algumas dificuldades na modelagem de alguns tipos de dados onde as taxas de risco podem ser em forma de banheira ou unimodal. Conseqüentemente, um progresso significativo foi feito em direção a generalização de alguns modelos de sobrevivência bem conhecidos.

Essa nova classe de modelos generalizados são flexíveis e facilitam a modelagem dos dados. Algumas das referências são Stacy *et al.* (1962); Cooray (2006); Zografos & Balakrishnan (2009).

## 2.1 Distribuição Odd-Weibull

A distribuição Odd-Weibull é uma generalização do modelo Weibull, introduzida por Cooray (2006) e estudada por Ortega *et al.* (2014), tem se mostrado útil para modelagem de dados de sobrevivência.

Uma variável aleatória  $T$  com distribuição Odd-Weibull denotada por  $OW(\phi, \alpha, \mu)$ , possui f.d.p. dada por:

$$f(t|\boldsymbol{\theta}) = \left(\frac{\alpha\phi}{t}\right) (\mu t)^\alpha \exp [(\mu t)^\alpha] \{\exp [(\mu t)^\alpha] - 1\}^{\phi-1} \left\{1 + \{\exp [(\mu t)^\alpha] - 1\}^\phi\right\}^{-2}, \quad (2.1)$$

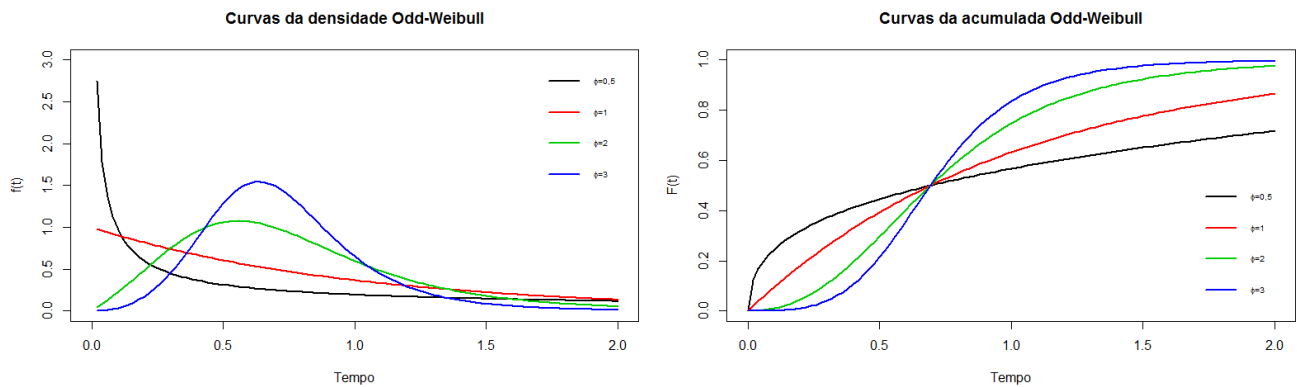
sendo  $t > 0$ ,  $\boldsymbol{\theta} = (\phi, \alpha, \mu)^\top$ . Os parâmetros de forma são  $\phi > 0$ ,  $\alpha > 0$  e  $\mu > 0$  é o parâmetro de escala. A distribuição Odd-Weibull é assintoticamente equivalente à distribuição Log-Logística proposta por Bennett (1983) quando  $\mu$  tende ao infinito.

A F.D.A. é dada por:

$$F(t|\boldsymbol{\theta}) = 1 - \left\{ 1 + \{\exp[(\mu t)^\alpha] - 1\}^\phi \right\}^{-1}, \quad t > 0.$$

A Figura 2.1 ilustra algumas formas da função densidade de probabilidade (2.1) para  $\alpha = 1$ ,  $\mu = 1$  e diferentes valores de  $\phi$ . Por meio dessa figura, observa-se que à medida que o valor do parâmetro  $\phi$  aumenta a forma da função densidade de probabilidade é deslocada à direita.

Figura 2.1: Funções densidade de probabilidade e acumulada da distribuição Odd-Weibull para diferentes valores de  $\phi$ .



Fonte: autoria própria.

Fonte: autoria própria.

A função de quantil correspondente à Equação (2.1) é

$$Q(u) = \frac{1}{\mu} \log^{1/\alpha} \left[ 1 + \left( \frac{u}{1+u} \right)^{1/\phi} \right], \quad 0 < u < 1.$$

A função de sobrevivência e a função risco são representadas por:

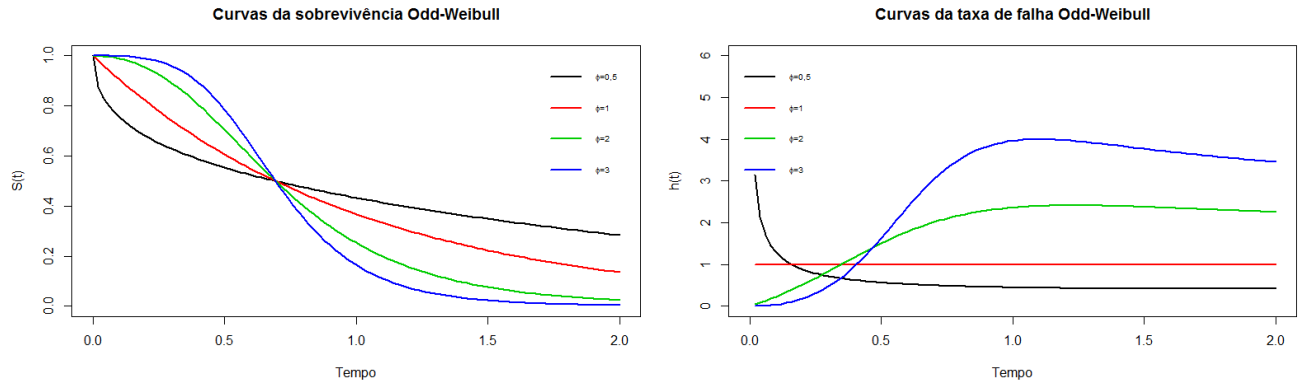
$$S(t|\boldsymbol{\theta}) = \left\{ 1 + \{\exp[(\mu t)^\alpha] - 1\}^\phi \right\}^{-1}$$

e

$$h(t|\boldsymbol{\theta}) = \left( \frac{\alpha\phi}{t} \right) (\mu t)^\alpha \exp[(\mu t)^\alpha] \{\exp[(\mu t)^\alpha] - 1\}^{\phi-1} \left\{ 1 + \{\exp[(\mu t)^\alpha] - 1\}^\phi \right\}^{-1}.$$

Na Figura 2.2, observa-se as curvas de sobrevivência e risco da variável aleatória  $T$  com distribuição Odd-Weibull.

Figura 2.2: Funções de sobrevivência e de risco da distribuição Odd-Weibull para diferentes valores de  $\phi$ .



Fonte: autoria própria.

Fonte: autoria própria.

## 2.2 Distribuições de probabilidade Gama-G

### 2.2.1 Geradora de distribuição gama

A função geradora de distribuições de probabilidade proposta por Zografos & Balakrishnan (2009) é definida da seguinte forma: se  $G(x)$  é uma função de distribuição acumulada de uma variável aleatória contínua, então, sua F.D.A. da função geradora de distribuição gama é dada por:

$$F(x|\phi) = \frac{1}{\Gamma(\phi)} \gamma \{-\log[1 - G(x)], \phi\}, \quad (2.2)$$

em que  $\gamma(w, \phi) = \int_0^w e^{-u} u^{\phi-1} du$  é a função gama incompleta, Johnson *et al.* (2005).

Derivando a equação (2.2), obtém-se a sua f.d.p. expressada como

$$f(x|\phi) = \frac{g(x)}{\Gamma(\phi)} \{-\log[1 - G(x)]\}^{\phi-1}, \quad (2.3)$$

em que  $g(x) = dG(x)/dx$ ,  $\phi > 0$  é o parâmetro de forma que flexibiliza a distribuição a ser modificada e o suporte de  $f(x|\phi)$  é definido por  $g(x)$ .

### 2.2.2 Distribuição Gama-Weibull

A distribuição Gama-Weibull ou Gama Generalizada de três parâmetros introduzida por Stacy *et al.* (1962) e estudada por Hashimoto (2013) e Ramos (2014) entre outros, tem se mostrado um ótimo modelo que leva a excelentes inferências para dados de sobrevivência. É obtida substituindo as expressões definidas em (1.1) e (1.2) na função geradora de distribuições gama (2.3).

Uma variável aleatória  $T$  com distribuição Gama-Weibull denotada por  $GW(\phi, \alpha, \mu)$ , possui f.d.p. dada por:

$$f(t|\boldsymbol{\theta}) = \frac{\alpha\mu^{\alpha\phi}}{\Gamma(\phi)} t^{\alpha\phi-1} \exp\{-(\mu t)^\alpha\}, \quad (2.4)$$

sendo  $t > 0$ ,  $\boldsymbol{\theta} = (\phi, \alpha, \mu)^\top$ . Os parâmetros de forma são  $\phi > 0$  e  $\alpha > 0$  e  $\mu > 0$  é o parâmetro de escala.

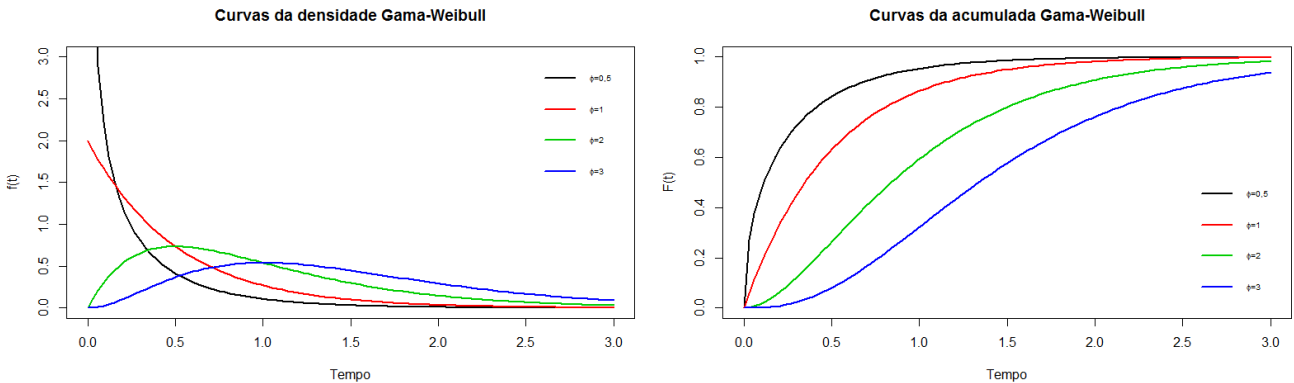
A F.D.A. é dada por:

$$F(t|\boldsymbol{\theta}) = 1 - \int_{(\mu t)^\alpha}^{\infty} \frac{1}{\Gamma(\phi)} w^{\phi-1} e^{-w} dw = 1 - \frac{\Gamma[\phi, (\mu t)^\alpha]}{\Gamma(\phi)}, \quad t > 0,$$

sendo  $\Gamma[y, x] = \int_x^\infty w^{y-1} e^{-w} dw$  denominada de função gama incompleta superior.

Pode-se observar na Figura 2.3 diferentes formas para a f.d.p. e F.D.A. distribuição Gama-Weibull (2.4) para  $\mu = 2$ ,  $\alpha = 1$  e diferentes valores de  $\phi$ . Por meio dessa figura, observa-se que à medida que o valor do parâmetro  $\phi$  aumenta a forma da função densidade de probabilidade é deslocada à direita.

Figura 2.3: Funções densidade de probabilidade e acumulada da distribuição Gama-Weibull para diferentes valores de  $\phi$ .



Fonte: autoria própria.

Fonte: autoria própria.

A média, a variância e a função quantílica da distribuição Gama-Weibull (2.4) são dadas, respectivamente, por:

$$E(T) = \frac{1}{\mu\Gamma(\phi)} \Gamma\left(\frac{1}{\alpha} + \phi\right),$$

$$Var(T) = \frac{1}{\mu^2\Gamma(\phi)} \left[ \Gamma\left(\phi + \frac{2}{\alpha}\right) - \frac{\Gamma(\phi + 1/\alpha)^2}{\Gamma(\phi)} \right],$$

$$Q(u) = \frac{1}{\mu} Z^{1/\alpha},$$

em que  $Z = Q_\phi(u)$  denota a função quantílica da gama com parâmetro  $\phi$ .

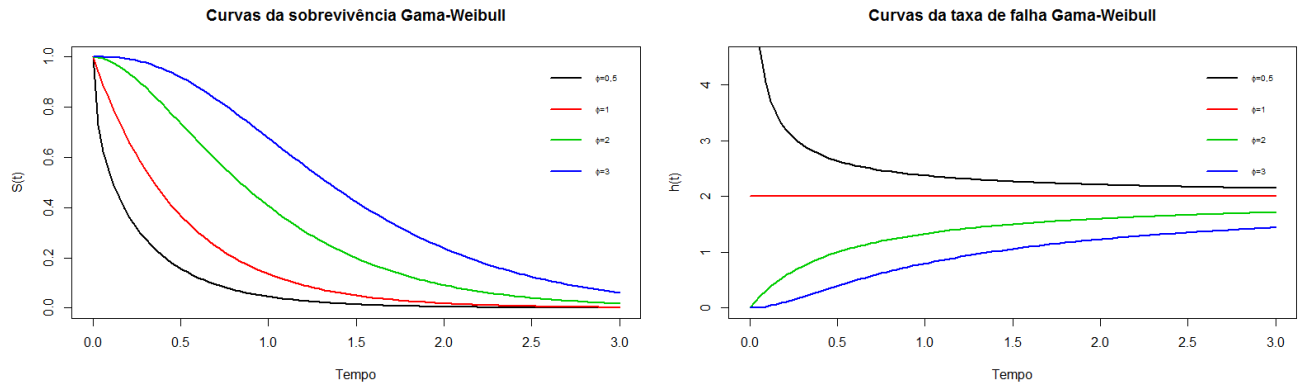
A função de sobrevivência e a função risco são representadas, respectivamente, por:

$$S(t|\boldsymbol{\theta}) = \frac{\Gamma[\phi, (\mu t)^\alpha]}{\Gamma(\phi)},$$

$$h(t|\boldsymbol{\theta}) = \frac{\alpha \mu^{\alpha\phi} t^{\alpha\phi-1} \exp\{- (\mu t)^\alpha\}}{\Gamma[\phi, (\mu t)^\alpha]}.$$

Observa-se na Figura 2.4 diferentes formas para a função de sobrevivência e a função de risco da distribuição Gama-Weibull.

Figura 2.4: Funções de sobrevivência e de risco da distribuição Gama-Weibull para diferentes valores de  $\phi$ .



Fonte: autoria própria.

Fonte: autoria própria.

### 2.2.3 Distribuição Gama-Log-Logística

A distribuição Gama-Log-Logística foi proposta por Hashimoto (2013), tem se mostrado muito flexível para descrever o comportamento dos dados observados. É obtida substituindo as expressões definidas em (1.4) e (1.5) na função geradora de distribuições gama (2.3).

Uma variável aleatória  $T$  com distribuição Gama-Log-Logística denotada por  $GLL(\phi, \alpha, \mu)$ , possui f.d.p. dada por:

$$f(t|\boldsymbol{\theta}) = \frac{\alpha \mu^\alpha}{\Gamma(\phi)} t^{\alpha-1} [1 + (\mu t)^\alpha]^{-2} \{\log [1 + (\mu t)^\alpha]\}^{\phi-1}, \quad (2.5)$$

sendo  $t > 0$ ,  $\boldsymbol{\theta} = (\phi, \alpha, \mu)^\top$ . Os parâmetros de forma são  $\phi > 0$  e  $\alpha > 0$  e  $\mu > 0$  o parâmetro de escala.

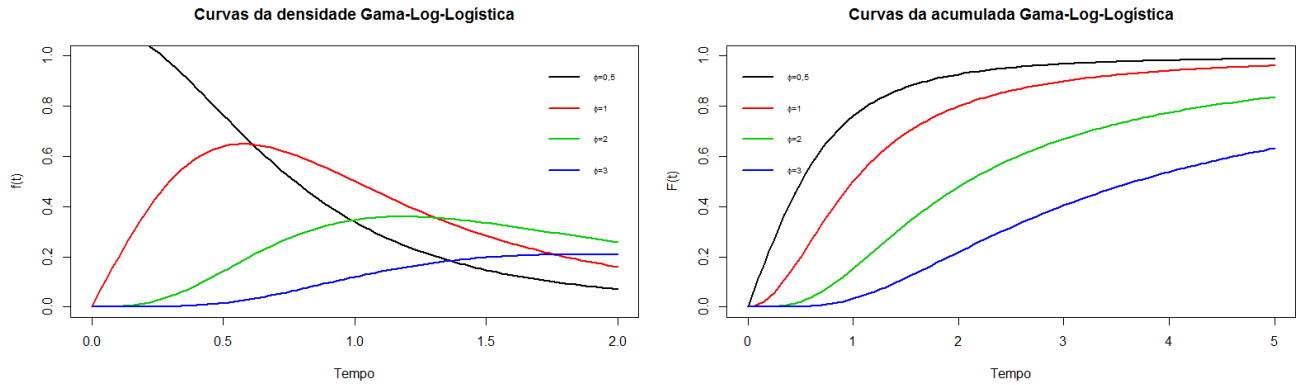
A F.D.A. é dada por:

$$F(t|\boldsymbol{\theta}) = 1 - \frac{\Gamma[\phi, \log\{1 + (\mu t)^\alpha\}]}{\Gamma(\phi)}, \quad t > 0,$$

sendo  $\Gamma[y, x] = \int_x^\infty w^{y-1} e^{-w} dw$  denominada de função gama incompleta superior.

A Figura 2.5 apresenta as curvas de densidade e acumulada da distribuição Gama-Log-Logística (2.5) para  $\mu = 1$ ,  $\alpha = 2$  e diferentes valores de  $\phi$ . Por meio dessa figura, observa-se que à medida que o valor do parâmetro  $\phi$  aumenta a forma da função densidade de probabilidade é deslocada à direita.

Figura 2.5: Funções densidade de probabilidade e acumulada da distribuição Gama-Log-Logística para diferentes valores de  $\phi$ .



Fonte: autoria própria.

Fonte: autoria própria.

A média, a variância e a função quantílica da distribuição Gama-Log-Logística (2.5) são dadas respectivamente por:

$$E(T) = \frac{1}{\mu\Gamma(\phi)} \int_0^1 \left(\frac{1-u}{u}\right)^{1/\alpha} \left[\log\left(\frac{1}{u}\right)\right]^{\phi-1} du,$$

$$Var(T) = \frac{1}{\mu^2\Gamma(\phi)} \left\{ \int_0^1 \left(\frac{1-u}{u}\right)^{2/\alpha} \left[\log\left(\frac{1}{u}\right)\right]^{\phi-1} du - \frac{1}{\Gamma(\phi)} \left[ \int_0^1 \left(\frac{1-u}{u}\right)^{1/\alpha} \left[\log\left(\frac{1}{u}\right)\right]^{\phi-1} du \right]^2 \right\}$$

e

$$Q(u) = \frac{1}{\mu} [\exp(Z) - 1]^{1/\alpha}, \tag{2.6}$$

em que  $Z = Q_\phi(u)$  denota a função quantílica da gama com parâmetro  $\phi$ .

A função de sobrevivência e a função risco são representadas por:

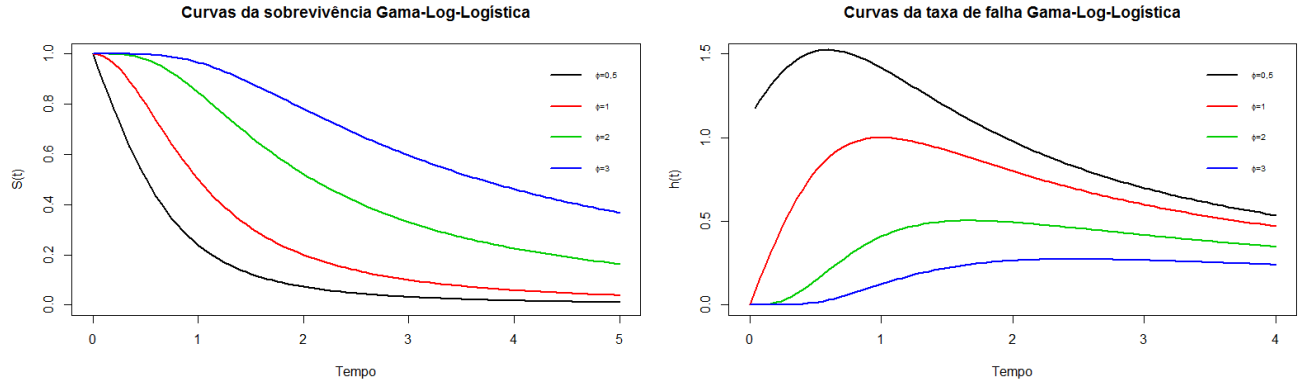
$$S(t|\boldsymbol{\theta}) = \frac{\Gamma[\phi, \log\{1 + (\mu t)^\alpha\}]}{\Gamma(\phi)}$$

e

$$h(t|\boldsymbol{\theta}) = \frac{\alpha\mu^\alpha}{\Gamma(\phi)} t^{\alpha-1} [1 + (\mu t)^\alpha]^{-2} \{\log[1 + (\mu t)^\alpha]\}^{\phi-1} \left\{ \frac{\Gamma[\phi, \log\{1 + (\mu t)^\alpha\}]}{\Gamma(\phi)} \right\}^{-1}.$$

Na Figura 2.6, apresenta-se as curvas de sobrevivência e risco da variável aleatória  $T$  com distribuição Gama-Log-Logística.

Figura 2.6: Funções de sobrevivência e de risco da distribuição Gama-Log-Logística para diferentes valores de  $\phi$ .



Fonte: autoria própria.

Fonte: autoria própria.

## 2.2.4 Distribuição Gama-Log-Normal

A distribuição Gama-Log-Normal é apresentada como uma alternativa para modelar os dados de sobrevivência. É obtida substituindo as expressões definidas em (1.6) e (1.7) na função geradora de distribuições gama (2.3).

Uma variável aleatória  $T$  com distribuição Gama-Log-Normal denotada por  $GLN(\phi, \alpha, \mu)$ , possui f.d.p. dada por:

$$f(t|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma t}\Gamma(\phi)} \exp\left(-\frac{1}{2}\left(\frac{\log(t) - \mu}{\sigma}\right)^2\right) \left\{-\log\left[\Phi\left(\frac{-\log(t) + \mu}{\sigma}\right)\right]\right\}^{\phi-1}, \quad (2.7)$$

sendo  $t > 0$ ,  $\boldsymbol{\theta} = (\phi, \mu, \sigma)^\top$ . O parâmetro de forma é  $\phi > 0$ , o parâmetro de locação é  $-\infty < \mu < +\infty$  e o parâmetro de escala é  $\sigma > 0$ .

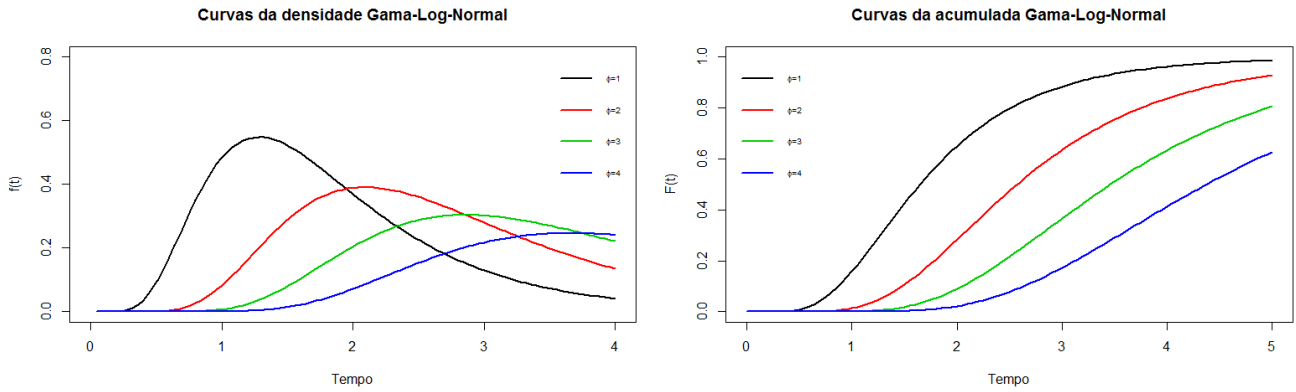
A F.D.A. é dada por:

$$F(t|\boldsymbol{\theta}) = 1 - \frac{\Gamma\left[\phi, -\log\left\{\Phi\left(\frac{-\log(t) + \mu}{\sigma}\right)\right\}\right]}{\Gamma(\phi)}, \quad t > 0,$$

sendo  $\Gamma[y, x] = \int_x^\infty w^{y-1}e^{-w}dw$  denominada de função gama incompleta superior.

A Figura 2.7 apresenta as curvas de densidade e acumulada da distribuição Gama-Log-Normal (2.7) para  $\mu = 0, 5$ ,  $\alpha = 0, 5$  e diferentes valores de  $\phi$ . Por meio dessa figura, observa-se que à medida que o valor do parâmetro  $\phi$  aumenta a forma da função densidade de probabilidade é deslocada à direita.

Figura 2.7: Funções densidade de probabilidade e acumulada da distribuição Gama-Log-Normal para diferentes valores de  $\phi$ .



Fonte: autoria própria.

Fonte: autoria própria.

A função quantílica da distribuição Gama-Log-Normal (2.7) é dada por:

$$Q(u) = \exp(\mu - \sigma \Phi^{-1}(e^{-Z})),$$

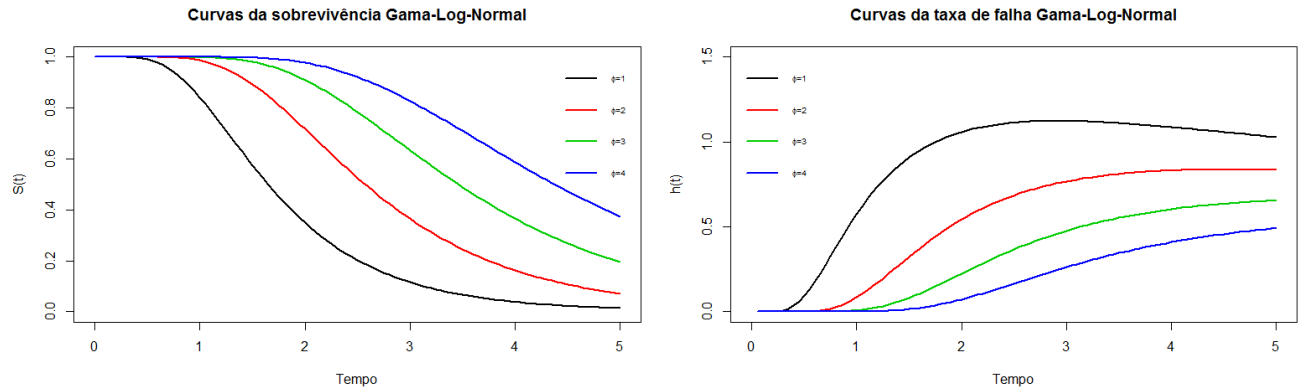
em que  $Z = Q_\phi(u)$  denota a função quantílica da gama com parâmetro  $\phi$ .

A função de sobrevivência e a função risco são representadas, respectivamente, por:

$$S(t|\boldsymbol{\theta}) = \frac{\Gamma\left[\phi, -\log\left\{\Phi\left(\frac{-\log(t) + \mu}{\sigma}\right)\right\}\right]}{\Gamma(\phi)},$$

$$h(t|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma t}\Gamma(\phi)} \exp\left(-\frac{1}{2}\left(\frac{\log(t) - \mu}{\sigma}\right)^2\right) \left\{-\log\left[\Phi\left(\frac{-\log(t) + \mu}{\sigma}\right)\right]\right\}^{\phi-1} \left\{\frac{\Gamma\left[\phi, -\log\left\{\Phi\left(\frac{-\log(t) + \mu}{\sigma}\right)\right\}\right]}{\Gamma(\phi)}\right\}^{-1}.$$

Na Figura 2.8, apresenta-se as curvas de sobrevivência e risco da variável aleatória  $T$  com distribuição Gama-Log-Normal.

Figura 2.8: Funções de sobrevivência e de risco da distribuição Gama-Log-Normal para diferentes valores de  $\phi$ .

Fonte: autoria própria.

Fonte: autoria própria.

## Estimação dos Parâmetros dos Modelos

Os modelos probabilísticos apresentados no capítulo anterior são caracterizados por quantidades desconhecidas, denominadas parâmetros. Em estudo que envolve tempos de falha, os parâmetros devem ser estimados a partir das observações amostrais, para que o modelo fique determinado e, assim, seja possível responder às perguntas de interesse. Para mais detalhes sobre o assunto é sugerido Colosimo & Giolo (2006); Hashimoto (2013); Ramos (2014); Cooray (2006); Carrasco *et al.* (2008).

O método a ser usado para a estimação dos parâmetros que incorpora censuras no seu processo de estimação é o método de máxima verossimilhança, pois ele é relativamente simples de ser entendido e tem propriedades ótimas para grandes amostras.

Na presença de observações censuradas, a função de verossimilhança é escrita como a equação (1.11), substituindo-se adequadamente a função densidade de probabilidade e a função de sobrevivência do modelo que se deseja ajustar.

### 3.1 Estimadores dos Parâmetros da Distribuição Odd-Weibull

Seja  $(t_1, \delta_1), (t_2, \delta_2), \dots, (t_n, \delta_n)$  um conjunto de dados de tamanho  $n$  com distribuição Odd-Weibull, em que  $t_i$  é a variável resposta e  $\delta_i$  é o indicador de censura ( $0 =$  tempo de censura,  $1 =$  tempo de falha), para  $i = 1, \dots, n$ . Então, considerando observações com censura, a função de verossimilhança é dada por:

$$\begin{aligned} L(\boldsymbol{\theta}; \mathbf{t}) &= \prod_{i=1}^n f(t_i|\boldsymbol{\theta})^{\delta_i} S(t_i|\boldsymbol{\theta})^{1-\delta_i} \\ &= \prod_{i=1}^n \left\{ \left( \frac{\alpha\phi}{t_i} \right) (\mu t_i)^\alpha \exp [(\mu t_i)^\alpha] \{ \exp [(\mu t_i)^\alpha] - 1 \}^{\phi-1} \left( 1 + \{ \exp [(\mu t_i)^\alpha] - 1 \}^\phi \right)^{-2} \right\}^{\delta_i} \\ &\quad \left\{ \left\{ 1 + \{ \exp [(\mu t_i)^\alpha] - 1 \}^\phi \right\}^{-1} \right\}^{1-\delta_i}, \end{aligned}$$

em que  $\boldsymbol{\theta} = (\phi, \alpha, \mu)^\top$  é o vetor de parâmetros desconhecidos.

Assim,

$$L(\boldsymbol{\theta}, \mathbf{t}) = (\alpha\phi\mu^\alpha)^d \left\{ \prod_{i=1}^n t_i^{\delta_i(\alpha-1)} \right\} \exp \left[ \mu^\alpha \sum_{i=1}^n \delta_i t_i^\alpha \right] \left\{ \prod_{i=1}^n (\exp [(\mu t_i)^\alpha] - 1)^{\delta_i(\phi-1)} \right\} \\ \prod_{i=1}^n \left( 1 + \{ \exp [(\mu t_i)^\alpha] - 1 \}^\phi \right)^{-2\delta_i} \prod_{i=1}^n \left( 1 + \{ \exp [(\mu t_i)^\alpha] - 1 \}^\phi \right)^{\delta_i-1}, \quad (3.1)$$

em que  $d$  é o número de observações que não foram censuradas.

Tomando-se o logaritmo da função de verossimilhança (3.1), segue que:

$$\ell(\boldsymbol{\theta}, \mathbf{t}) = d \log(\alpha\phi\mu^\alpha) + (\alpha - 1) \sum_{i=1}^n \delta_i \log(t_i) + \mu^\alpha \sum_{i=1}^n \delta_i t_i^\alpha + (\phi - 1) \sum_{i=1}^n \delta_i \log(\exp [(\mu t_i)^\alpha] - 1) \\ - \sum_{i=1}^n (\delta_i + 1) \log \left( 1 + \{ \exp [(\mu t_i)^\alpha] - 1 \}^\phi \right). \quad (3.2)$$

A função escore é obtida, calculando-se a derivada do logaritmo da função de verossimilhança (3.2) em relação aos parâmetros do vetor  $\boldsymbol{\theta}$ . Assim, tem-se que

$$\frac{\partial}{\partial \phi} \ell(\boldsymbol{\theta}, \mathbf{t}) = \frac{d}{\phi} + \sum_{i=1}^n \delta_i \log(\exp [(\mu t_i)^\alpha] - 1) \\ - \sum_{i=1}^n \frac{(\delta_i + 1) (\exp [(\mu t_i)^\alpha] - 1)^\phi \log(\exp [(\mu t_i)^\alpha] - 1)}{1 + \{ \exp [(\mu t_i)^\alpha] - 1 \}^\phi},$$

$$\frac{\partial}{\partial \alpha} \ell(\boldsymbol{\theta}, \mathbf{t}) = \frac{d}{\alpha} + d \log(\mu) + \sum_{i=1}^n \delta_i \log(t_i) + \mu^\alpha \sum_{i=1}^n \delta_i t_i^\alpha \log(\mu t_i) \\ + (\phi - 1) \sum_{i=1}^n \frac{\delta_i \exp [(\mu t_i)^\alpha] (\mu t_i)^\alpha \log(\mu t_i)}{\exp [(\mu t_i)^\alpha] - 1} \\ - \phi \sum_{i=1}^n \frac{(\delta_i + 1) \{ \exp [(\mu t_i)^\alpha] - 1 \}^{\phi-1} \exp [(\mu t_i)^\alpha] (\mu t_i)^\alpha \log(\mu t_i)}{1 + \{ \exp [(\mu t_i)^\alpha] - 1 \}^\phi},$$

$$\frac{\partial}{\partial \mu} \ell(\boldsymbol{\theta}, \mathbf{t}) = \frac{\alpha}{\mu} + \alpha \mu^{\alpha-1} \sum_{i=1}^n \delta_i t_i^\alpha + \frac{\alpha(\phi - 1)}{\mu} \sum_{i=1}^n \frac{\delta_i \exp [(\mu t_i)^\alpha] (\mu t_i)^\alpha}{\exp [(\mu t_i)^\alpha] - 1} \\ - \frac{\alpha\phi}{\mu} \sum_{i=1}^n \frac{(\delta_i + 1) \{ \exp [(\mu t_i)^\alpha] - 1 \}^{\phi-1} \exp [(\mu t_i)^\alpha] (\mu t_i)^\alpha}{1 + \{ \exp [(\mu t_i)^\alpha] - 1 \}^\phi}.$$

Igualando-se as funes escore a zero, obtm-se as equaces de verossimilhana dadas por:

$$\begin{aligned} \sum_{i=1}^n \frac{(\delta_i + 1) \left( \exp \left[ (\hat{\mu}t_i)^{\hat{\alpha}} \right] - 1 \right)^{\hat{\phi}} \log \left( \exp \left[ (\hat{\mu}t_i)^{\hat{\alpha}} \right] - 1 \right)}{1 + \left\{ \exp \left[ (\hat{\mu}t_i)^{\hat{\alpha}} \right] - 1 \right\}^{\hat{\phi}}} &= \frac{d}{\hat{\phi}} + \sum_{i=1}^n \delta_i \log \left( \exp \left[ (\hat{\mu}t_i)^{\hat{\alpha}} \right] - 1 \right), \\ \hat{\phi} \sum_{i=1}^n \frac{(\delta_i + 1) \left\{ \exp \left[ (\hat{\mu}t_i)^{\hat{\alpha}} - 1 \right] \right\}^{\hat{\phi}-1} \exp \left[ (\hat{\mu}t_i)^{\hat{\alpha}} \right] (\hat{\mu}t_i)^{\hat{\alpha}} \log(\hat{\mu}t_i)}{1 + \left\{ \exp \left[ (\hat{\mu}t_i)^{\hat{\alpha}} \right] - 1 \right\}^{\hat{\phi}}} &= \frac{d}{\hat{\alpha}} + d \log(\hat{\mu}) + \sum_{i=1}^n \delta_i \log(t_i) \\ &+ \hat{\mu}^{\hat{\alpha}} \sum_{i=1}^n \delta_i t_i^{\hat{\alpha}} \log(\hat{\mu}t_i) + (\hat{\phi} - 1) \sum_{i=1}^n \frac{\delta_i \exp \left[ (\hat{\mu}t_i)^{\hat{\alpha}} \right] (\hat{\mu}t_i)^{\hat{\alpha}} \log(\hat{\mu}t_i)}{\exp \left[ (\hat{\mu}t_i)^{\hat{\alpha}} \right] - 1}, \\ \frac{\hat{\alpha}\hat{\phi}}{\hat{\mu}} \sum_{i=1}^n \frac{(\delta_i + 1) \left\{ \exp \left[ (\hat{\mu}t_i)^{\hat{\alpha}} \right] - 1 \right\}^{\hat{\phi}-1} \exp \left[ (\hat{\mu}t_i)^{\hat{\alpha}} \right] (\hat{\mu}t_i)^{\hat{\alpha}}}{1 + \left\{ \exp \left[ (\hat{\mu}t_i)^{\hat{\alpha}} \right] - 1 \right\}^{\hat{\phi}}} &= \frac{\hat{\alpha}}{\hat{\mu}} + \hat{\alpha}\hat{\mu}^{\hat{\alpha}-1} \sum_{i=1}^n \delta_i t_i^{\hat{\alpha}} \\ &+ \frac{\hat{\alpha}(\hat{\phi} - 1)}{\hat{\mu}} \sum_{i=1}^n \frac{\delta_i \exp \left[ (\hat{\mu}t_i)^{\hat{\alpha}} \right] (\hat{\mu}t_i)^{\hat{\alpha}}}{\exp \left[ (\hat{\mu}t_i)^{\hat{\alpha}} \right] - 1}, \end{aligned}$$

cujas solues fornecem os estimadores de mxima verossimilhana. Mtodos numricos como o de Newton-Rapshon so necessrios para encontrar a soluo deste sistema no linear. Neste trabalho, foi utilizado a linguagem de programao R verso 3.5.1.

## 3.2 Estimadores dos Parmetros da Distribuio Gama-Weibull

Seja  $(t_1, \delta_1), (t_2, \delta_2), \dots, (t_n, \delta_n)$  um conjunto de dados de tamanho  $n$  com distribuio Gama-Weibull, em que  $t_i$   a varivel resposta e  $\delta_i$   o indicador de censura ( $0 =$  tempo de censura,  $1 =$  tempo de falha), para  $i = 1, \dots, n$ . Ento, considerando observaes com censura, a funo de verossimilhana  dada por:

$$\begin{aligned} L(\boldsymbol{\theta}; \mathbf{t}) &= \prod_{i=1}^n f(t_i | \boldsymbol{\theta})^{\delta_i} S(t_i | \boldsymbol{\theta})^{1-\delta_i} \\ &= \prod_{i=1}^n \left\{ \frac{\alpha \mu^{\alpha \phi}}{\Gamma(\phi)} t_i^{\alpha \phi - 1} \exp \left\{ -(\mu t_i)^{\alpha} \right\} \right\}^{\delta_i} \left\{ \frac{\Gamma[\phi, (\mu t_i)^{\alpha}]}{\Gamma(\phi)} \right\}^{1-\delta_i}, \end{aligned}$$

em que  $\boldsymbol{\theta} = (\phi, \alpha, \mu)^{\top}$   o vetor de parmetros desconhecidos.

Assim,

$$L(\boldsymbol{\theta}, \mathbf{t}) = \frac{\alpha^d \mu^{d\alpha\phi}}{\Gamma(\phi)^d} \left\{ \prod_{i=1}^n t_i^{\delta_i(\alpha\phi-1)} \right\} \exp \left\{ -\mu^\alpha \sum_{i=1}^n \delta_i t_i^\alpha \right\} \prod_{i=1}^n \left\{ \frac{\Gamma[\phi, (\mu t_i)^\alpha]}{\Gamma(\phi)} \right\}^{1-\delta_i}, \quad (3.3)$$

em que  $d$  é o número de observações que não foram censuradas.

O logaritmo da função de verossimilhança (3.3) é dado por:

$$\begin{aligned} \ell(\boldsymbol{\theta}, \mathbf{t}) = & d \log(\alpha) + d\alpha\phi \log(\mu) - d \log(\Gamma(\phi)) + (\alpha\phi - 1) \sum_{i=1}^n \delta_i \log(t_i) \\ & - \mu^\alpha \sum_{i=1}^n \delta_i t_i^\alpha + \sum_{i=1}^n (1 - \delta_i) \log \left\{ \frac{\Gamma[\phi, (\mu t_i)^\alpha]}{\Gamma(\phi)} \right\}. \end{aligned} \quad (3.4)$$

Objetivando-se encontrar as equações de verossimilhança, serão utilizados os seguintes resultados:

$$\frac{\partial}{\partial k} \Gamma[k, x] = \int_x^\infty w^{k-1} \log(w) e^{-w} dw = \Psi[k, x] \quad \text{e} \quad \frac{\partial}{\partial x} \Gamma[k, x] = -x^{k-1} e^{-x}.$$

A função escore é obtida, calculando-se a derivada do logaritmo da função de verossimilhança (3.4) em relação aos parâmetros do vetor  $\boldsymbol{\theta}$ . Assim, tem-se que:

$$\frac{\partial}{\partial \phi} \ell(\boldsymbol{\theta}, \mathbf{t}) = d\alpha \log(\mu) - d\psi(\phi) + \alpha \sum_{i=1}^n \delta_i \log(t_i) + \sum_{i=1}^n \left\{ (1 - \delta_i) \left[ \frac{\Psi[\phi, (\mu t_i)^\alpha]}{\Gamma[\phi, (\mu t_i)^\alpha]} \right] \right\},$$

$$\begin{aligned} \frac{\partial}{\partial \alpha} \ell(\boldsymbol{\theta}, \mathbf{t}) = & d/\alpha + d\phi \log(\mu) + \phi \sum_{i=1}^n \delta_i \log(t_i) - \mu^\alpha \sum_{i=1}^n \delta_i t_i^\alpha \log(\mu t_i) \\ & - \sum_{i=1}^n \left\{ (1 - \delta_i) \left[ \frac{(\mu t_i)^{\alpha(\phi-1)} e^{-(\mu t_i)^\alpha} \log(\mu t_i)}{\Gamma[\phi, (\mu t_i)^\alpha]} \right] \right\}, \end{aligned}$$

$$\frac{\partial}{\partial \mu} \ell(\boldsymbol{\theta}, \mathbf{t}) = \frac{d\alpha\phi}{\mu} - \alpha\mu^{\alpha-1} \sum_{i=1}^n \delta_i t_i^\alpha - \sum_{i=1}^n \left\{ (1 - \delta_i) \left[ \frac{\alpha\mu^{\alpha-1} t_i^\alpha (\mu t_i)^{\alpha(\phi-1)} e^{-(\mu t_i)^\alpha}}{\Gamma[\phi, (\mu t_i)^\alpha]} \right] \right\},$$

em que  $\psi(\cdot)$  denota a função digama.

Igualando-se as funções escore a zero, obtém-se as equações de verossimilhança dadas por:

$$\sum_{i=1}^n \left\{ (1 - \delta_i) \left[ \frac{\Psi[\hat{\phi}, (\hat{\mu} t_i)^{\hat{\alpha}}]}{\Gamma[\hat{\phi}, (\hat{\mu} t_i)^{\hat{\alpha}}]} \right] \right\} = -d\hat{\alpha} \log(\hat{\mu}) + d\psi(\hat{\phi}) - \hat{\alpha} \sum_{i=1}^n \delta_i \log(t_i),$$

$$\sum_{i=1}^n \left\{ (1 - \delta_i) \left[ \frac{(\hat{\mu}t_i)^{\hat{\phi}-1} e^{-(\hat{\mu}t_i)^{\hat{\phi}}} \log(\hat{\mu}t_i)}{\Gamma[\hat{\phi}, (\hat{\mu}t_i)^{\hat{\phi}}]} \right] \right\} = \frac{d}{\hat{\alpha}} + d\hat{\phi} \log(\hat{\mu}) + \hat{\phi} \sum_{i=1}^n \delta_i \log(t_i) - \hat{\mu}^{\hat{\alpha}} \sum_{i=1}^n \delta_i t_i^{\hat{\alpha}} \log(\hat{\mu}t_i),$$

$$\sum_{i=1}^n \left\{ (1 - \delta_i) \left[ \frac{\hat{\alpha} \hat{\mu}^{\hat{\alpha}-1} t_i^{\hat{\alpha}} (\hat{\mu}t_i)^{\hat{\phi}-1} e^{-(\hat{\mu}t_i)^{\hat{\phi}}}}{\Gamma[\hat{\phi}, (\hat{\mu}t_i)^{\hat{\phi}}]} \right] \right\} = \frac{d\hat{\alpha}\hat{\phi}}{\hat{\mu}} - \hat{\alpha} \hat{\mu}^{\hat{\alpha}-1} \sum_{i=1}^n \delta_i t_i^{\hat{\alpha}},$$

cujas soluções fornecem os estimadores de máxima verossimilhança. Métodos numéricos como o de Newton-Rapshon são necessários para encontrar a solução deste sistema não linear. Neste trabalho, foi utilizado a linguagem de programação R versão 3.5.1.

### 3.3 Estimadores dos Parâmetros da Distribuição Gama-Log-Logística

Seja  $(t_1, \delta_1), (t_2, \delta_2), \dots, (t_n, \delta_n)$  um conjunto de dados de tamanho  $n$  com distribuição Gama-Log-Logística, em que  $t_i$  é a variável resposta e  $\delta_i$  é o indicador de censura ( $0 =$  tempo de censura,  $1 =$  tempo de falha), para  $i = 1, \dots, n$ . Então, considerando observações com censura, a função de verossimilhança é dada por:

$$\begin{aligned} L(\boldsymbol{\theta}; \mathbf{t}) &= \prod_{i=1}^n f(t_i|\boldsymbol{\theta})^{\delta_i} S(t_i|\boldsymbol{\theta})^{1-\delta_i} \\ &= \prod_{i=1}^n \left\{ \frac{\alpha \mu^\alpha}{\Gamma(\phi)} t_i^{\alpha-1} [1 + (\mu t_i)^\alpha]^{-2} \{\log [1 + (\mu t_i)^\alpha]\}^{\phi-1} \right\}^{\delta_i} \\ &\quad \left\{ \frac{\Gamma[\phi, \log \{1 + (\mu t_i)^\alpha\}]}{\Gamma(\phi)} \right\}^{1-\delta_i}, \end{aligned}$$

em que  $\boldsymbol{\theta} = (\phi, \alpha, \mu)^\top$  é o vetor de parâmetros desconhecidos.

Assim,

$$\begin{aligned} L(\boldsymbol{\theta}, \mathbf{t}) &= \frac{\alpha^d \mu^{d\alpha}}{\Gamma(\phi)^d} \left\{ \prod_{i=1}^n t_i^{\delta_i(\alpha-1)} \right\} \left\{ \prod_{i=1}^n [1 + (\mu t_i)^\alpha]^{-2\delta_i} \right\} \left\{ \prod_{i=1}^n (\log [1 + (\mu t_i)^\alpha])^{\delta_i(\phi-1)} \right\} \\ &\quad \prod_{i=1}^n \left\{ \frac{\Gamma[\phi, \log \{1 + (\mu t_i)^\alpha\}]}{\Gamma(\phi)} \right\}^{1-\delta_i}, \end{aligned} \quad (3.5)$$

em que  $d$  é o número de observações que não foram censuradas.

O logaritmo da função de verossimilhança (3.5) é dado por:

$$\begin{aligned} \ell(\boldsymbol{\theta}, \mathbf{t}) = & d \log(\alpha) + d \log(\mu) - d \log(\Gamma(\phi)) + (\alpha - 1) \sum_{i=1}^n \delta_i \log(t_i) \\ & - 2 \sum_{i=1}^n \delta_i \log[1 + (\mu t_i)^\alpha] + (\phi - 1) \sum_{i=1}^n \delta_i \log(\log[1 + (\mu t_i)^\alpha]) \\ & + \sum_{i=1}^n (1 - \delta_i) \log \left\{ \frac{\Gamma[\phi, \log\{1 + (\mu t_i)^\alpha\}]}{\Gamma(\phi)} \right\}. \end{aligned} \quad (3.6)$$

Objetivando-se encontrar as equações de verossimilhança, serão utilizados os seguintes resultados:

$$\frac{\partial}{\partial k} \Gamma[k, x] = \int_x^\infty w^{k-1} \log(w) e^{-w} dw = \Psi[k, x] \quad \text{e} \quad \frac{\partial}{\partial x} \Gamma[k, x] = -x^{k-1} e^{-x}.$$

A função escore é obtida, calculando-se a derivada do logaritmo da função de verossimilhança (3.6) em relação aos parâmetros do vetor  $\boldsymbol{\theta}$ . Assim, tem-se que:

$$\frac{\partial}{\partial \phi} \ell(\boldsymbol{\theta}, \mathbf{t}) = -d\psi(\phi) + \sum_{i=1}^n \delta_i \log(\log[1 + (\mu t_i)^\alpha]) + \sum_{i=1}^n \left\{ (1 - \delta_i) \left[ \frac{\Psi[\phi, \log\{1 + (\mu t_i)^\alpha\}]}{\Gamma[\phi, \log\{1 + (\mu t_i)^\alpha\}]} \right] \right\},$$

$$\begin{aligned} \frac{\partial}{\partial \alpha} \ell(\boldsymbol{\theta}, \mathbf{t}) = & \frac{d}{\alpha} + d \log(\mu) + \sum_{i=1}^n \delta_i \log(t_i) - 2 \sum_{i=1}^n \delta_i \left[ \frac{(\mu t_i)^\alpha \log(\mu t_i)^\alpha}{1 + (\mu t_i)^\alpha} \right] \\ & + (\phi - 1) \sum_{i=1}^n \delta_i \frac{(\mu t_i)^\alpha \log(\mu t_i)^\alpha}{(1 + (\mu t_i)^\alpha) \log\{1 + (\mu t_i)^\alpha\}} - \sum_{i=1}^n \left\{ \frac{(1 - \delta_i) (\log\{1 + (\mu t_i)^\alpha\})^{\phi-1} (\mu t_i)^\alpha \log(\mu t_i)^\alpha}{(1 + (\mu t_i)^\alpha)^2 \Gamma[\phi, \log\{1 + (\mu t_i)^\alpha\}]} \right\}, \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial \mu} \ell(\boldsymbol{\theta}, \mathbf{t}) = & \frac{d\alpha}{\mu} - 2 \sum_{i=1}^n \left\{ \frac{\delta_i \alpha \mu^{\alpha-1} t_i^\alpha}{1 + (\mu t_i)^\alpha} \right\} + (\phi - 1) \sum_{i=1}^n \left\{ \frac{\delta_i \alpha \mu^{\alpha-1} t_i^\alpha}{(1 + (\mu t_i)^\alpha) \log(1 + (\mu t_i)^\alpha)} \right\} \\ & - \sum_{i=1}^n \left\{ \frac{(1 - \delta_i) \alpha \mu^{\alpha-1} t_i^\alpha \{\log(1 + (\mu t_i)^\alpha)\}^{\phi-1}}{(1 + (\mu t_i)^\alpha)^2 \Gamma[\phi, \log\{1 + (\mu t_i)^\alpha\}]} \right\}, \end{aligned}$$

em que  $\psi(\cdot)$  denota a função digama.

Igualando-se as funções escore a zero, obtém-se as equações de verossimilhança dadas por:

$$\sum_{i=1}^n \left\{ (1 - \delta_i) \left[ \frac{\Psi[\hat{\phi}, \log\{1 + (\hat{\mu} t_i)^{\hat{\alpha}}\}]}{\Gamma[\hat{\phi}, \log\{1 + (\hat{\mu} t_i)^{\hat{\alpha}}\}]} \right] \right\} = d\psi(\hat{\phi}) - \sum_{i=1}^n \delta_i \log(\log[1 + (\hat{\mu} t_i)^{\hat{\alpha}}]),$$

$$\sum_{i=1}^n \left\{ \frac{(1 - \delta_i) (\log \{1 + (\hat{\mu}t_i)^{\hat{\alpha}}\})^{\hat{\phi}-1} (\hat{\mu}t_i)^{\hat{\alpha}} \log(\hat{\mu}t_i)}{(1 + (\hat{\mu}t_i)^{\hat{\alpha}})^2 \Gamma[\hat{\phi}, \log \{1 + (\hat{\mu}t_i)^{\hat{\alpha}}\}]} \right\} = \frac{d}{\hat{\alpha}} + d \log(\hat{\mu}) + \sum_{i=1}^n \delta_i \log(t_i) \\ - 2 \sum_{i=1}^n \delta_i \left[ \frac{(\hat{\mu}t_i)^{\hat{\alpha}} \log(\hat{\mu}t_i)^{\hat{\alpha}}}{1 + (\hat{\mu}t_i)^{\hat{\alpha}}} \right] + (\hat{\phi} - 1) \sum_{i=1}^n \delta_i \frac{(\hat{\mu}t_i)^{\hat{\alpha}} \log(\hat{\mu}t_i)^{\hat{\alpha}}}{(1 + (\hat{\mu}t_i)^{\hat{\alpha}}) \log \{1 + (\hat{\mu}t_i)^{\hat{\alpha}}\}},$$

$$\sum_{i=1}^n \left\{ \frac{(1 - \delta_i) \hat{\alpha} \hat{\mu}^{\hat{\alpha}-1} t_i^{\hat{\alpha}} \{\log(1 + (\hat{\mu}t_i)^{\hat{\alpha}})\}^{\hat{\phi}-1}}{(1 + (\hat{\mu}t_i)^{\hat{\alpha}})^2 \Gamma[\hat{\phi}, \log \{1 + (\hat{\mu}t_i)^{\hat{\alpha}}\}]} \right\} = \frac{d\hat{\alpha}}{\hat{\mu}} - 2 \sum_{i=1}^n \left\{ \frac{\delta_i \hat{\alpha} \hat{\mu}^{\hat{\alpha}-1} t_i^{\hat{\alpha}}}{1 + (\hat{\mu}t_i)^{\hat{\alpha}}} \right\} \\ + (\hat{\phi} - 1) \sum_{i=1}^n \left\{ \frac{\delta_i \hat{\alpha} \hat{\mu}^{\hat{\alpha}-1} t_i^{\hat{\alpha}}}{(1 + (\hat{\mu}t_i)^{\hat{\alpha}}) \log(1 + (\hat{\mu}t_i)^{\hat{\alpha}})} \right\}.$$

cujas soluções fornecem os estimadores de máxima verossimilhança. Métodos numéricos como o de Newton-Rapshon são necessários para encontrar a solução deste sistema não linear. Neste trabalho, foi utilizado a linguagem de programação R versão 3.5.1.

### 3.4 Estimadores dos Parâmetros da Distribuição Gama-Log-Normal

Seja  $(t_1, \delta_1), (t_2, \delta_2), \dots, (t_n, \delta_n)$  um conjunto de dados de tamanho  $n$  com distribuição Gama-Log-Normal, em que  $t_i$  é a variável resposta e  $\delta_i$  é o indicador de censura (0 = tempo de censura, 1 = tempo de falha), para  $i = 1, \dots, n$ . Então, considerando observações com censura, a função de verossimilhança é dada por:

$$L(\boldsymbol{\theta}; \mathbf{t}) = \prod_{i=1}^n f(t_i | \boldsymbol{\theta})^{\delta_i} S(t_i | \boldsymbol{\theta})^{1-\delta_i} \\ = \prod_{i=1}^n \left\{ \frac{1}{\sqrt{2\pi}\sigma t_i \Gamma(\phi)} \exp\left(-\frac{1}{2} \left(\frac{\log(t_i) - \mu}{\sigma}\right)^2\right) \left\{ -\log \left[ \Phi\left(\frac{-\log(t_i) + \mu}{\sigma}\right) \right] \right\}^{\phi-1} \right\}^{\delta_i} \\ \left\{ \frac{\Gamma\left[\phi, -\log\left\{\Phi\left(\frac{-\log(t_i) + \mu}{\sigma}\right)\right\}\right]}{\Gamma(\phi)} \right\}^{1-\delta_i},$$

em que  $\boldsymbol{\theta} = (\phi, \alpha, \mu)^\top$  é o vetor de parâmetros desconhecidos.

Assim,

$$L(\boldsymbol{\theta}, \mathbf{t}) = \frac{1}{(\sqrt{2\pi})^d \sigma^d \Gamma(\phi)^d} \left\{ \prod_{i=1}^n \frac{1}{t_i^{\delta_i}} \right\} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \delta_i \left( \frac{\log(t_i) - \mu}{\sigma} \right)^2 \right\} \\ \prod_{i=1}^n \left\{ -\log \left[ \Phi \left( \frac{-\log(t_i) + \mu}{\sigma} \right) \right] \right\}^{\delta_i(\phi-1)} \\ \prod_{i=1}^n \left\{ \frac{\Gamma \left[ \phi, -\log \left\{ \Phi \left( \frac{-\log(t_i) + \mu}{\sigma} \right) \right\} \right]}{\Gamma(\phi)} \right\}^{1-\delta_i}, \quad (3.7)$$

em que  $d$  é o número de observações que não foram censuradas.

O logaritmo da função de verossimilhança (3.7) é dado por:

$$\ell(\boldsymbol{\theta}, \mathbf{t}) = -d \log(\sqrt{2\pi}) - d \log(\sigma) - d \log(\Gamma(\phi)) - \sum_{i=1}^n \delta_i \log(t_i) - \frac{1}{2} \sum_{i=1}^n \delta_i \left( \frac{\log(t_i) - \mu}{\sigma} \right)^2 \\ + (1 - \phi) \sum_{i=1}^n \delta_i \log \left( \log \left[ \Phi \left( \frac{-\log(t_i) + \mu}{\sigma} \right) \right] \right) \\ + \sum_{i=1}^n (1 - \delta_i) \log \left\{ \frac{\Gamma \left[ \phi, -\log \left\{ \Phi \left( \frac{-\log(t_i) + \mu}{\sigma} \right) \right\} \right]}{\Gamma(\phi)} \right\}. \quad (3.8)$$

Objetivando-se encontrar as equações de verossimilhança, serão utilizados os seguintes resultados:

$$\frac{\partial}{\partial k} \Gamma[k, x] = \int_x^\infty w^{k-1} \log(w) e^{-w} dw = \Psi[k, x] \quad \text{e} \quad \frac{\partial}{\partial x} \Gamma[k, x] = -x^{k-1} e^{-x}.$$

A função escore é obtida, calculando-se a derivada do logaritmo da função de verossimilhança (3.8) em relação aos parâmetros do vetor  $\boldsymbol{\theta}$ . Assim, tem-se que:

$$\frac{\partial}{\partial \phi} \ell(\boldsymbol{\theta}, \mathbf{t}) = -d\psi(\phi) - \sum_{i=1}^n \delta_i \log \left( \log \left[ \Phi \left( \frac{-\log(t_i) + \mu}{\sigma} \right) \right] \right) \\ + \sum_{i=1}^n \left\{ (1 - \delta_i) \frac{\Psi \left[ \phi, -\log \left\{ \Phi \left( \frac{-\log(t_i) + \mu}{\sigma} \right) \right\} \right]}{\Gamma \left[ \phi, -\log \left\{ \Phi \left( \frac{-\log(t_i) + \mu}{\sigma} \right) \right\} \right]} \right\}.$$

$$\begin{aligned} \frac{\partial}{\partial \mu} \ell(\boldsymbol{\theta}, \mathbf{t}) = & \\ & \frac{1}{\sigma} \sum_{i=1}^n \delta_i \left( \frac{\log(t_i) - \mu}{\sigma} \right) + \frac{1 - \phi}{\sqrt{2\pi}\sigma} \sum_{i=1}^n \delta_i \frac{\exp \left\{ -\frac{1}{2} \left( \frac{-\log(t_i) + \mu}{\sigma} \right)^2 \right\}}{\log \left[ \Phi \left( \frac{-\log(t_i) + \mu}{\sigma} \right) \right] \Phi \left( \frac{-\log(t_i) + \mu}{\sigma} \right)} \\ & + \frac{1}{\sqrt{2\pi}\sigma} \sum_{i=1}^n (1 - \delta_i) \frac{\left( -\log \left\{ \Phi \left( \frac{-\log(t_i) + \mu}{\sigma} \right) \right\} \right)^{\phi-1} \exp \left\{ -\frac{1}{2} \left( \frac{-\log(t_i) + \mu}{\sigma} \right)^2 \right\}}{\Gamma \left[ \phi, -\log \left\{ \Phi \left( \frac{-\log(t_i) + \mu}{\sigma} \right) \right\} \right]}. \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial \sigma} \ell(\boldsymbol{\theta}, \mathbf{t}) = & \\ & -\frac{d}{\sigma} + \frac{1}{\sigma} \sum_{i=1}^n \delta_i \left( \frac{\log(t_i) - \mu}{\sigma} \right)^2 - \frac{1 - \phi}{\sqrt{2\pi}\sigma} \sum_{i=1}^n \delta_i \frac{\exp \left\{ -\frac{1}{2} \left( \frac{-\log(t_i) + \mu}{\sigma} \right)^2 \right\} \left( \frac{\log(t_i) - \mu}{\sigma} \right)}{\log \left[ \Phi \left( \frac{-\log(t_i) + \mu}{\sigma} \right) \right] \Phi \left( \frac{-\log(t_i) + \mu}{\sigma} \right)} \\ & + \frac{1}{\sqrt{2\pi}\sigma} \sum_{i=1}^n (1 - \delta_i) \frac{\left( -\log \left\{ \Phi \left( \frac{-\log(t_i) + \mu}{\sigma} \right) \right\} \right)^{\phi-1} \exp \left\{ -\frac{1}{2} \left( \frac{-\log(t_i) + \mu}{\sigma} \right)^2 \right\} \left( \frac{\log(t_i) - \mu}{\sigma} \right)}{\Gamma \left[ \phi, -\log \left\{ \Phi \left( \frac{-\log(t_i) + \mu}{\sigma} \right) \right\} \right]}, \end{aligned}$$

em que  $\psi(\cdot)$  denota a função digama.

Igualando-se as funções escore a zero, obtém-se as equações de verossimilhança dadas por:

$$\begin{aligned} \sum_{i=1}^n \left\{ (1 - \delta_i) \frac{\Psi \left[ \hat{\phi}, -\log \left\{ \Phi \left( \frac{-\log(t_i) + \hat{\mu}}{\hat{\sigma}} \right) \right\} \right]}{\Gamma \left[ \hat{\phi}, -\log \left\{ \Phi \left( \frac{-\log(t_i) + \hat{\mu}}{\hat{\sigma}} \right) \right\} \right]} \right\} = \\ d\psi(\hat{\phi}) + \sum_{i=1}^n \delta_i \log \left( \log \left[ \Phi \left( \frac{-\log(t_i) + \hat{\mu}}{\hat{\sigma}} \right) \right] \right), \end{aligned}$$

$$\begin{aligned}
& \frac{1}{\sqrt{2\pi}\hat{\sigma}} \sum_{i=1}^n (1 - \delta_i) \frac{\left(-\log \left\{ \Phi \left( \frac{-\log(t_i) + \hat{\mu}}{\hat{\sigma}} \right) \right\}\right)^{\hat{\phi}-1} \exp \left\{ -\frac{1}{2} \left( \frac{-\log(t_i) + \hat{\mu}}{\hat{\sigma}} \right)^2 \right\}}{\Gamma \left[ \hat{\phi}, -\log \left\{ \Phi \left( \frac{-\log(t_i) + \hat{\mu}}{\hat{\sigma}} \right) \right\} \right]} = \\
& - \frac{1}{\hat{\sigma}} \sum_{i=1}^n \delta_i \left( \frac{\log(t_i) - \hat{\mu}}{\hat{\sigma}} \right) - \frac{1 - \hat{\phi}}{\sqrt{2\pi}\hat{\sigma}} \sum_{i=1}^n \delta_i \frac{\exp \left\{ -\frac{1}{2} \left( \frac{-\log(t_i) + \hat{\mu}}{\hat{\sigma}} \right)^2 \right\}}{\log \left[ \Phi \left( \frac{-\log(t_i) + \hat{\mu}}{\hat{\sigma}} \right) \right] \Phi \left( \frac{-\log(t_i) + \hat{\mu}}{\hat{\sigma}} \right)}, \\
& \frac{1}{\sqrt{2\pi}\hat{\sigma}} \sum_{i=1}^n (1 - \delta_i) \frac{\left(-\log \left\{ \Phi \left( \frac{-\log(t_i) + \hat{\mu}}{\hat{\sigma}} \right) \right\}\right)^{\hat{\phi}-1} \exp \left\{ -\frac{1}{2} \left( \frac{-\log(t_i) + \hat{\mu}}{\hat{\sigma}} \right)^2 \right\} \left( \frac{\log(t_i) - \hat{\mu}}{\hat{\sigma}} \right)}{\Gamma \left[ \hat{\phi}, -\log \left\{ \Phi \left( \frac{-\log(t_i) + \hat{\mu}}{\hat{\sigma}} \right) \right\} \right]} = \\
& \frac{d}{\hat{\sigma}} - \frac{1}{\hat{\sigma}} \sum_{i=1}^n \delta_i \left( \frac{\log(t_i) - \hat{\mu}}{\hat{\sigma}} \right)^2 + \frac{1 - \hat{\phi}}{\sqrt{2\pi}\hat{\sigma}} \sum_{i=1}^n \delta_i \frac{\exp \left\{ -\frac{1}{2} \left( \frac{-\log(t_i) + \hat{\mu}}{\hat{\sigma}} \right)^2 \right\} \left( \frac{\log(t_i) - \hat{\mu}}{\hat{\sigma}} \right)}{\log \left[ \Phi \left( \frac{-\log(t_i) + \hat{\mu}}{\hat{\sigma}} \right) \right] \Phi \left( \frac{-\log(t_i) + \hat{\mu}}{\hat{\sigma}} \right)},
\end{aligned}$$

cujas soluções fornecem os estimadores de máxima verossimilhança. Métodos numéricos como o de Newton-Raphson são necessários para encontrar a solução deste sistema não linear. Neste trabalho, foi utilizado a linguagem de programação R versão 3.5.1..

# Modelo de regressão Gama-Log-Logística

Neste capítulo, se apresenta o modelo de regressão porque em muitas áreas do saber os estudos envolvem covariáveis que podem estar relacionadas como o tempo de sobrevivência. Certamente, essas covariáveis devem ser incluídas na análise estatística dos dados porque a função das covariáveis é acelerar ou desacelerar o tempo de vida.

## 4.1 Distribuição Gama-Log-Logística

A distribuição Gama-Log-Logística que foi estudada na seção 2.2.3 pertence aos modelos de tempo de vida acelerado. Ao considerar a reparametrização  $\mu = \exp(-m)$ , obtém-se a nova distribuição de probabilidade.

Se  $T$  é uma variável aleatória com distribuição Gama-Log-Logística definida em (2.5), então

$$f(t|\boldsymbol{\theta}) = \frac{\alpha \exp(-m\alpha)}{\Gamma(\phi)} t^{\alpha-1} [1 + (\exp(-m)t)^\alpha]^{-2} \{\log [1 + (\exp(-m)t)^\alpha]\}^{\phi-1}, \quad (4.1)$$

em que  $t > 0$ ,  $\boldsymbol{\theta} = (\phi, \alpha, m)^\top$ .

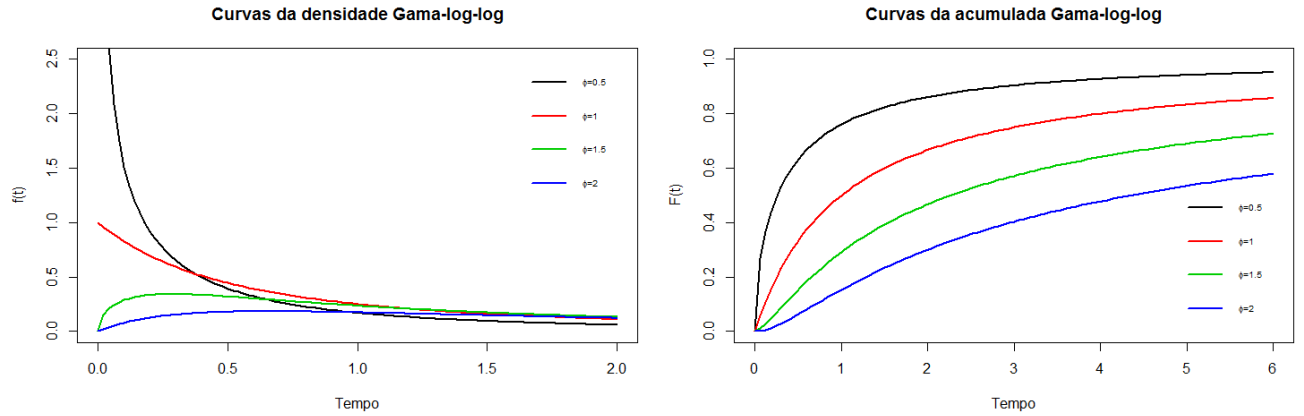
A F.D.A. é dada por:

$$F(t|\boldsymbol{\theta}) = 1 - \frac{\Gamma[\phi, \log\{1 + (\exp(-m)t)^\alpha\}]}{\Gamma(\phi)}, \quad t > 0,$$

sendo  $\Gamma[y, x] = \int_x^\infty w^{y-1} e^{-w} dw$  denominada de função gama incompleta superior.

A Figura 4.1 ilustra algumas formas da função densidade de probabilidade (4.1) para  $m = 0$ ,  $\alpha = 1$  e diferentes valores de  $\phi$ . Por meio dessa figura, observa-se que à medida que o valor do parâmetro  $\phi$  aumenta a forma da função densidade de probabilidade é mais achatada.

Figura 4.1: Funções densidade de probabilidade e acumulada da distribuição Gama-Log-Logística para diferentes valores de  $\phi$ .



Fonte: autoria própria.

Fonte: autoria própria.

A média e a variância da distribuição Gama-Log-Logística (4.1), que são expressos, respectivamente, por:

$$E(T) = \frac{\exp(m)}{\Gamma(\phi)} \int_0^1 \left(\frac{1-u}{u}\right)^{1/\alpha} \left[\log\left(\frac{1}{u}\right)\right]^{\phi-1} du$$

e

$$Var(T) = \frac{\exp(2m)}{\Gamma(\phi)} \left\{ \int_0^1 \left(\frac{1-u}{u}\right)^{2/\alpha} \left[\log\left(\frac{1}{u}\right)\right]^{\phi-1} du - \frac{1}{\Gamma(\phi)} \left[ \int_0^1 \left(\frac{1-u}{u}\right)^{1/\alpha} \left[\log\left(\frac{1}{u}\right)\right]^{\phi-1} du \right]^2 \right\}$$

A função de sobrevivência e a função risco são representadas por:

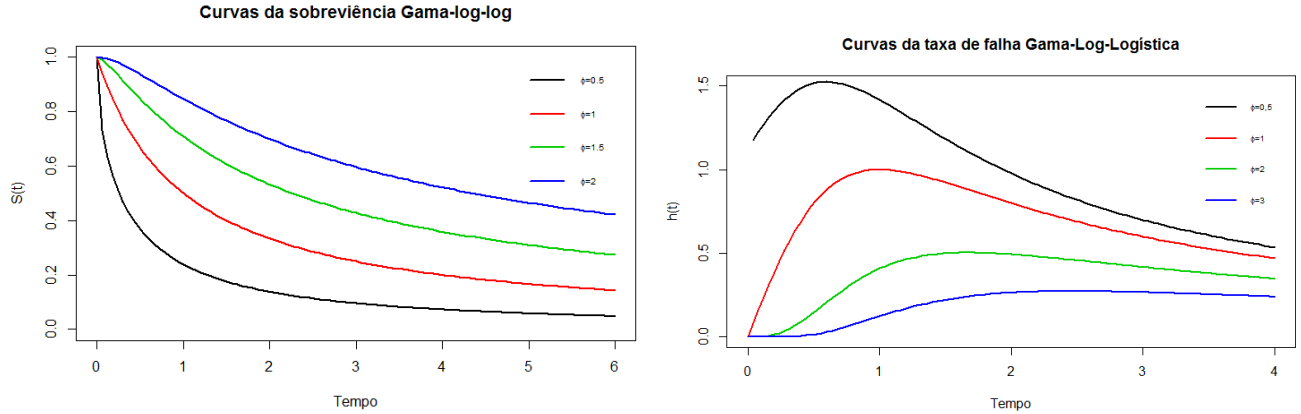
$$S(t|\boldsymbol{\theta}) = \frac{\Gamma[\phi, \log\{1 + (\exp(-m)t)^\alpha\}]}{\Gamma(\phi)}$$

e

$$h(t|\boldsymbol{\theta}) = \frac{\alpha \exp(-m\alpha) t^{\alpha-1} [1 + (\exp(-m)t)^\alpha]^{-2} \{\log[1 + (\exp(-m)t)^\alpha]\}^{\phi-1}}{\Gamma[\phi, \log\{1 + (\exp(-m)t)^\alpha\}]}$$

Na Figura 4.2, apresentamos as curvas de sobrevivência e risco da variável aleatória  $T$  com a nova distribuição Gama-Log-Logística.

Figura 4.2: Funções de sobrevivência e de risco da distribuição Gama-Log-Logística para diferentes valores de  $\phi$ .



Fonte: autoria própria.

Fonte: autoria própria.

Após ter definido a distribuição da variável resposta  $T$ , na qual pertence à família de distribuições que se caracteriza pelo fato de ter um parâmetro de forma  $\alpha$  ( $\alpha > 0$ ) e parâmetro de locação  $m$  ( $-\infty < m < +\infty$ ) que depende do vetor de covariáveis  $\mathbf{x}$ , denotado por  $m(\mathbf{x}) = \mathbf{x}^\top \boldsymbol{\beta}$ , em que  $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)^\top$  é o vetor de parâmetros desconhecidos a serem estimados e  $\mathbf{x}^\top = (x_0, \dots, x_p)$  é o vetor formado por observações das  $(p + 1)$  covariáveis.

Então, o modelo de regressão que relaciona a variável resposta  $T$  a um vetor de covariáveis  $\mathbf{x}$  é um modelo escrito da seguinte forma

$$T = \exp(\mathbf{x}^\top \boldsymbol{\beta}) \exp(z/\alpha),$$

em que  $T$  é os tempos de sobrevivência e  $z$  é uma variável aleatória que assume uma distribuição Log-Gama-Log-Logística independente de  $\mathbf{x}$ , como segue

$$f(z) = \frac{1}{\Gamma(\phi)} \exp(z) [1 + \exp(z)]^{-2} \{\log [1 + \exp(z)]\}^{\phi-1}, \quad -\infty < z < +\infty,$$

em que  $\phi > 0$  é o parâmetro de forma desconhecido.

Portanto, a função densidade de probabilidade, a função de sobrevivência e a função risco de  $T|x$  assumem as formas

$$f(t|x) = \frac{\alpha \exp(-\mathbf{x}^\top \boldsymbol{\beta} \alpha)}{\Gamma(\phi)} t^{\alpha-1} \left[1 + \left(\exp(-\mathbf{x}^\top \boldsymbol{\beta}) t\right)^\alpha\right]^{-2} \left\{\log \left[1 + \left(\exp(-\mathbf{x}^\top \boldsymbol{\beta}) t\right)^\alpha\right]\right\}^{\phi-1}, \quad t > 0,$$

$$S(t|x) = \frac{\Gamma[\phi, \log \{1 + (\exp(-\mathbf{x}^\top \boldsymbol{\beta}) t)^\alpha\}]}{\Gamma(\phi)}, \quad t > 0$$

e

$$h(t|x) = \frac{\alpha \exp(-\mathbf{x}^\top \boldsymbol{\beta} \alpha) t^{\alpha-1} \left[1 + \left(\exp(-\mathbf{x}^\top \boldsymbol{\beta}) t\right)^\alpha\right]^{-2} \left\{\log \left[1 + \left(\exp(-\mathbf{x}^\top \boldsymbol{\beta}) t\right)^\alpha\right]\right\}^{\phi-1}}{\Gamma[\phi, \log \{1 + (\exp(-\mathbf{x}^\top \boldsymbol{\beta}) t)^\alpha\}]},$$

respectivamente.

## 4.2 Estimadores dos Parâmetros do modelo de regressão Gama-Log-Logística

Para estimar os parâmetros do modelo de regressão foi utilizado o método da máxima verossimilhança, descrito no Capítulo 2, pois permite levar em conta a informação das observações censuradas. Assim, para um conjunto de dados que consiste de  $n$  observações,  $(t_1, \delta_1, x_1), \dots, (t_n, \delta_n, x_n)$ , em que  $\delta_i$  é o indicador de censura e  $x_i$  é o vetor de covariáveis associados a  $i$ -ésima observação, para  $i = 1, \dots, n$ . Então, considerando observações com censura, a função de verossimilhança é dada por:

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{i=1}^n f(t_i|x_i)^{\delta_i} S(t_i|x_i)^{1-\delta_i} \\ &= \prod_{i=1}^n \left\{ \frac{\alpha \exp(-x_i^\top \beta \alpha)}{\Gamma(\phi)} t^{\alpha-1} \left[ 1 + \left( \exp(-x_i^\top \beta) t \right)^\alpha \right]^{-2} \left\{ \log \left[ 1 + \left( \exp(-x_i^\top \beta) t \right)^\alpha \right] \right\}^{\phi-1} \right\}^{\delta_i} \\ &\quad \left\{ \frac{\Gamma[\phi, \log \{1 + (\exp(-x_i^\top \beta) t)^\alpha\}]}{\Gamma(\phi)} \right\}^{1-\delta_i}. \end{aligned}$$

Assim,

$$\begin{aligned} L(\boldsymbol{\theta}) &= \frac{\alpha^d}{\Gamma(\phi)^d} \left\{ \prod_{i=1}^n \exp(-\delta_i x_i^\top \beta \alpha) \right\} \left\{ \prod_{i=1}^n t_i^{\delta_i(\alpha-1)} \right\} \left\{ \prod_{i=1}^n \left[ 1 + \left( \exp(-x_i^\top \beta) t_i \right)^\alpha \right]^{-2\delta_i} \right\} \\ &\quad \left\{ \prod_{i=1}^n \left( \log \left[ 1 + \left( \exp(-x_i^\top \beta) t_i \right)^\alpha \right] \right)^{\delta_i(\phi-1)} \right\} \prod_{i=1}^n \left\{ \frac{\Gamma[\phi, \log \{1 + (\exp(-x_i^\top \beta) t_i)^\alpha\}]}{\Gamma(\phi)} \right\}^{1-\delta_i}. \end{aligned} \quad (4.2)$$

O logaritmo da função de verossimilhança (4.2) é dado por:

$$\begin{aligned} \ell(\boldsymbol{\theta}) &= d \log(\alpha) - d \log(\Gamma(\phi)) - \alpha \sum_{i=1}^n \delta_i x_i^\top \beta + (\alpha - 1) \sum_{i=1}^n \delta_i \log(t_i) \\ &\quad - 2 \sum_{i=1}^n \delta_i \log \left\{ 1 + \left[ \exp(-x_i^\top \beta) t_i \right]^\alpha \right\} + (\phi - 1) \sum_{i=1}^n \delta_i \log \left( \log \left\{ 1 + \left[ \exp(-x_i^\top \beta) t_i \right]^\alpha \right\} \right) \\ &\quad + \sum_{i=1}^n (1 - \delta_i) \log \left( \frac{\Gamma[\phi, \log \{1 + [\exp(-x_i^\top \beta) t_i]^\alpha\}]}{\Gamma(\phi)} \right), \end{aligned} \quad (4.3)$$

em que  $\boldsymbol{\theta} = (\phi, \alpha, \beta^\top)^\top$  é o vetor de parâmetros desconhecidos,  $d$  é o número de observações que não foram censuradas e  $\Gamma[\cdot, \cdot]$  é a função gama incompleta superior.

Para encontrar a equação de verossimilhança, os seguintes resultados serão usados:

$$\frac{\partial}{\partial y} \Gamma[y, x] = \int_x^\infty w^{y-1} \log(w) e^{-w} dw = \Psi[y, x] \quad \text{e} \quad \frac{\partial}{\partial x} \Gamma[y, x] = -x^{y-1} e^{-x}.$$

A função escore é obtida, calculando-se a derivada do logaritmo da função de verossimilhança (4.3) em relação aos parâmetros do vetor  $\boldsymbol{\theta}$ . Assim, tem-se que:

$$\begin{aligned} \frac{\partial}{\partial \phi} \ell(\boldsymbol{\theta}) &= -n\psi(\phi) + \sum_{i=1}^n \delta_i \log \left( \log \left\{ 1 + \left[ \exp(-x_i^\top \beta) t_i \right]^\alpha \right\} \right) \\ &\quad + \sum_{i=1}^n \frac{(1 - \delta_i) \Psi \left[ \phi, \log \left\{ 1 + \left[ \exp(-x_i^\top \beta) t_i \right]^\alpha \right\} \right]}{\Gamma \left[ \phi, \log \left\{ 1 + \left[ \exp(-x_i^\top \beta) t_i \right]^\alpha \right\} \right]}, \\ \frac{\partial}{\partial \alpha} \ell(\boldsymbol{\theta}) &= \frac{d}{\alpha} - \sum_{i=1}^n \delta_i x_i^\top \beta + \sum_{i=1}^n \delta_i \log(t_i) - 2 \sum_{i=1}^n \frac{\delta_i \left[ \exp(-x_i^\top \beta) t_i \right]^\alpha \log \left( \exp(-x_i^\top \beta) t_i \right)}{1 + \left[ \exp(-x_i^\top \beta) t_i \right]^\alpha} \\ &\quad + (\phi - 1) \sum_{i=1}^n \frac{\delta_i \left[ \exp(-x_i^\top \beta) t_i \right]^\alpha \log \left[ \exp(-x_i^\top \beta) t_i \right]^\alpha}{\left( 1 + \left[ \exp(-x_i^\top \beta) t_i \right]^\alpha \right) \log \left\{ 1 + \left[ \exp(-x_i^\top \beta) t_i \right]^\alpha \right\}} \\ &\quad - \sum_{i=1}^n \frac{(1 - \delta_i) \left( \log \left\{ 1 + \left[ \exp(-x_i^\top \beta) t_i \right]^\alpha \right\} \right)^{\phi-1} \left[ \exp(-x_i^\top \beta) t_i \right]^\alpha \log \left( \exp(-x_i^\top \beta) t_i \right)}{\left( 1 + \left[ \exp(-x_i^\top \beta) t_i \right]^\alpha \right)^2 \Gamma \left[ \phi, \log \left\{ 1 + \left[ \exp(-x_i^\top \beta) t_i \right]^\alpha \right\} \right]}, \\ \frac{\partial}{\partial \beta_j} \ell(\boldsymbol{\theta}) &= -\alpha \sum_{i=1}^n \delta_i x_{ij}^\top + 2\alpha \sum_{i=1}^n \frac{\delta_i x_{ij}^\top \left[ \exp(-x_i^\top \beta) t_i \right]^\alpha}{1 + \left[ \exp(-x_i^\top \beta) t_i \right]^\alpha} \\ &\quad - (\phi - 1) \alpha \sum_{i=1}^n \frac{\delta_i x_{ij}^\top \left[ \exp(-x_i^\top \beta) t_i \right]^\alpha}{\left( 1 + \left[ \exp(-x_i^\top \beta) t_i \right]^\alpha \right) \log \left( 1 + \left[ \exp(-x_i^\top \beta) t_i \right]^\alpha \right)} \\ &\quad + \alpha \sum_{i=1}^n \frac{(1 - \delta_i) x_{ij}^\top \left[ \exp(-x_i^\top \beta) t_i \right]^\alpha \left( \log \left\{ 1 + \left[ \exp(-x_i^\top \beta) t_i \right]^\alpha \right\} \right)^{\phi-1}}{\left( 1 + \left[ \exp(-x_i^\top \beta) t_i \right]^\alpha \right)^2 \Gamma \left[ \phi, \log \left\{ 1 + \left[ \exp(-x_i^\top \beta) t_i \right]^\alpha \right\} \right)}. \end{aligned}$$

Igualando-se as funções escore a zero, obtêm-se as equações de verossimilhança dadas por:

$$\begin{aligned} &-n\psi(\hat{\phi}) + \sum_{i=1}^n \delta_i \log \left( \log \left\{ 1 + \left[ \exp(-x_i^\top \hat{\beta}) t_i \right]^{\hat{\alpha}} \right\} \right) \\ &\quad + \sum_{i=1}^n \frac{(1 - \delta_i) \Psi \left[ \hat{\phi}, \log \left\{ 1 + \left[ \exp(-x_i^\top \hat{\beta}) t_i \right]^{\hat{\alpha}} \right\} \right]}{\Gamma \left[ \hat{\phi}, \log \left\{ 1 + \left[ \exp(-x_i^\top \hat{\beta}) t_i \right]^{\hat{\alpha}} \right\} \right]} = 0, \\ \frac{d}{d\hat{\alpha}} &- \sum_{i=1}^n \delta_i x_i^\top \hat{\beta} + \sum_{i=1}^n \delta_i \log(t_i) - 2 \sum_{i=1}^n \frac{\delta_i \left[ \exp(-x_i^\top \hat{\beta}) t_i \right]^{\hat{\alpha}} \log \left( \exp(-x_i^\top \hat{\beta}) t_i \right)^{\hat{\alpha}}}{1 + \left[ \exp(-x_i^\top \hat{\beta}) t_i \right]^{\hat{\alpha}}} \\ &\quad + (\hat{\phi} - 1) \sum_{i=1}^n \frac{\delta_i \left[ \exp(-x_i^\top \hat{\beta}) t_i \right]^{\hat{\alpha}} \log \left[ \exp(-x_i^\top \hat{\beta}) t_i \right]^{\hat{\alpha}}}{\left( 1 + \left[ \exp(-x_i^\top \hat{\beta}) t_i \right]^{\hat{\alpha}} \right) \log \left\{ 1 + \left[ \exp(-x_i^\top \hat{\beta}) t_i \right]^{\hat{\alpha}} \right\}} \\ &\quad - \sum_{i=1}^n \frac{(1 - \delta_i) \left( \log \left\{ 1 + \left[ \exp(-x_i^\top \hat{\beta}) t_i \right]^{\hat{\alpha}} \right\} \right)^{\hat{\phi}-1} \left[ \exp(-x_i^\top \hat{\beta}) t_i \right]^{\hat{\alpha}} \log \left( \exp(-x_i^\top \hat{\beta}) t_i \right)}{\left( 1 + \left[ \exp(-x_i^\top \hat{\beta}) t_i \right]^{\hat{\alpha}} \right)^2 \Gamma \left[ \hat{\phi}, \log \left\{ 1 + \left[ \exp(-x_i^\top \hat{\beta}) t_i \right]^{\hat{\alpha}} \right\} \right]} = 0, \end{aligned}$$

$$\begin{aligned}
& - \hat{\alpha} \sum_{i=1}^n \delta_i x_{ij}^\top + 2\hat{\alpha} \sum_{i=1}^n \frac{\delta_i x_{ij}^\top \left[ \exp(-x_i^\top \hat{\beta}) t_i \right]^{\hat{\alpha}}}{1 + \left[ \exp(-x_i^\top \hat{\beta}) t_i \right]^{\hat{\alpha}}} \\
& - (\hat{\phi} - 1) \hat{\alpha} \sum_{i=1}^n \frac{\delta_i x_{ij}^\top \left[ \exp(-x_i^\top \hat{\beta}) t_i \right]^{\hat{\alpha}}}{\left( 1 + \left[ \exp(-x_i^\top \hat{\beta}) t_i \right]^{\hat{\alpha}} \right) \log \left( 1 + \left[ \exp(-x_i^\top \hat{\beta}) t_i \right]^{\hat{\alpha}} \right)} \\
& + \hat{\alpha} \sum_{i=1}^n \frac{(1 - \delta_i) x_{ij}^\top \left[ \exp(-x_i^\top \hat{\beta}) t_i \right]^{\hat{\alpha}} \left( \log \left\{ 1 + \left[ \exp(-x_i^\top \hat{\beta}) t_i \right]^{\hat{\alpha}} \right\} \right)^{\hat{\phi}-1}}{\left( 1 + \left[ \exp(-x_i^\top \hat{\beta}) t_i \right]^{\hat{\alpha}} \right)^2 \Gamma \left[ \hat{\phi}, \log \left\{ 1 + \left[ \exp(-x_i^\top \hat{\beta}) t_i \right]^{\hat{\alpha}} \right\} \right]} = 0.
\end{aligned}$$

em que  $\psi(\cdot)$  é a função digama,  $\Psi[y, x] = \int_x^\infty w^{y-1} \log(w) e^{-w} dw$  e  $j = 0, 1, \dots, p$ .

Portanto, para encontrar as estimativas dos parâmetros, foi utilizada a linguagem de programação R versão 3.5.1. Pois, nesse caso, não é possível resolver analiticamente o sistema de equações.

No caso do tamanho da amostra ser grande e sob certas condições gerais de regularidade da função de verossimilhança, os intervalos de confiança e testes de hipóteses podem ser obtidos utilizando o fato de que a distribuição assintótica dos estimadores de máxima verossimilhança é a distribuição normal com média  $\boldsymbol{\theta}$  e matriz de covariância igual ao inverso da matriz de informação. Portanto, a matriz de covariância assintótica é igual a  $\mathbf{I}^{-1}(\boldsymbol{\theta})$  com  $\mathbf{I}(\boldsymbol{\theta}) = E \left[ \ddot{L}(\boldsymbol{\theta}) \right]$ , em que

$$\ddot{L}(\boldsymbol{\theta}) = - \left( \frac{\partial^2 \log [L(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right).$$

Como não é possível determinar a matriz de informação de Fisher  $\mathbf{I}(\boldsymbol{\theta})$  devido a presença de observações censuradas, uma forma alternativa é utilizar a matriz de informação observada  $\ddot{L}(\boldsymbol{\theta})$ , avaliada no estimador de máxima verossimilhança  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ . A aproximação normal assintótica de  $\hat{\boldsymbol{\theta}}$  é expressa como

$$\hat{\boldsymbol{\theta}}^\top \sim N_{(p+2)} \left[ \boldsymbol{\theta}^\top; \ddot{L}(\boldsymbol{\theta})^{-1} \right],$$

em que  $\ddot{L}(\boldsymbol{\theta})$  é a matriz de informação observada  $(p+2)(p+2)$ .

Assim, o intervalo de confiança assintótico para  $\beta_j$ , com  $j = 1, \dots, p$ , considerando um nível de confiança de  $(1 - \alpha)100\%$ , é expresso por:

$$\hat{\beta}_j \pm z_{\alpha/2} \sqrt{\widehat{Var}(\hat{\beta}_j)},$$

em que  $z_{\alpha/2}$  é o valor que excede com probabilidade  $\alpha/2$  a distribuição normal padrão.

## Aplicação

Neste capítulo, as técnicas estatísticas apresentadas nos capítulos anterior são aplicadas ao conjunto de dados provenientes do departamento de informática do Sistema Unico de Saúde do Brasil (DATASUS), que é o órgão da Secretaria de Gestão Estratégica e Participativa do Ministério da Saúde, com a responsabilidade de coletar, processar e disseminar informações sobre saúde. Particularmente, analisa-se dados provenientes por atendimentos das internações hospitalares do SIH/SUS, ou seja, o número de dias de internação por paciente especificamente na região de São Paulo.

Para as análises e tratamento do banco de dados foi utilizado o software livre R, versão 3.5.1. Tal software permite a utilização de códigos abertos, propiciando um amplo espaço para análises estatísticas, gerenciamento de banco de dados e também análises gráficas.

### 5.1 Pacientes do Sistema de Informações Hospitalares do SUS

Para a realização das análises, foi utilizado um banco de dados fornecido pelo SIH/SUS que contém informações sobre o número de dias em que o paciente ficou internado no hospital. As informações foram coletadas em um período de 1 ano, compreendendo no período de janeiro a dezembro de 2020. Foram observados 153550 pacientes apresentado na Tabela 5.1. A variável resposta foi determinada como sendo o tempo (em dias) da data de internação até receber a alta (data de saída). Para esses pacientes, censura (1 indica que o paciente foi dado de alta e 0 indica que o paciente morreu), sexo (3 indica que o paciente foi feminino e 1 indica que o paciente foi masculino) e idade foram registrados, encontram-se também na Tabela 5.1.

Tabela 5.1: Pacientes internados no período de janeiro a dezembro de 2020 (DATASUS).

| Paciente | Tempo | Censura | Sexo | Idade |
|----------|-------|---------|------|-------|
| 1        | 3     | 1       | 3    | 23    |
| 2        | 3     | 1       | 3    | 28    |
| 3        | 3     | 1       | 3    | 19    |
| 4        | 2     | 1       | 3    | 21    |
| 5        | 2     | 1       | 1    | 34    |
| ⋮        | ⋮     | ⋮       | ⋮    | ⋮     |

| Paciente | Tempo | Censura | Sexo | Idade |
|----------|-------|---------|------|-------|
| ⋮        | ⋮     | ⋮       | ⋮    | ⋮     |
| 49       | 5     | 0       | 3    | 59    |
| ⋮        | ⋮     | ⋮       | ⋮    | ⋮     |
| 153546   | 3     | 1       | 1    | 44    |
| 153547   | 5     | 1       | 3    | 70    |
| 153548   | 3     | 1       | 3    | 10    |
| 153549   | 1     | 1       | 1    | 11    |
| 153550   | 19    | 1       | 3    | 3     |

Fonte: autoria própria.

Note que, a Tabela 5.1 mostra o banco de dados obtido no SIH/SUS, onde temos a internação dos pacientes até receber alta ou não. Por exemplo, o primeiro paciente foi hospitalizado por 3 dias até a alta e o paciente 49 foi hospitalizado por 5 dias até morrer.

Na Tabela 5.2, mostra-se as estatísticas descritivas dos pacientes no SIH/SUS em 2020 no estado de São Paulo; observa-se que dos 153550 pacientes, o 7% (10558) não apresentaram o evento de interesse. Além disso, sabe-se que o tempo mediano de internação é de 3 dias. O tempo máximo observado foi de 335 dias e tempo mínimo observado foi de 1 dia.

Tabela 5.2: Estatísticas descritivas dos dos pacientes internados no hospital no período de janeiro a dezembro de 2020 (DATASUS).

| Tempo                | Censura    | Sexo      | Idade                 |
|----------------------|------------|-----------|-----------------------|
| Mínimo : 1           | 0 : 10558  | 1 : 65488 | Mínimo : 0            |
| Primeiro quartil : 2 | 1 : 142992 | 3 : 88062 | Primeiro quartil : 25 |
| Mediana : 3          |            |           | Mediana : 43          |
| Média : 5,533        |            |           | Média : 44,2          |
| Terceiro quartil : 6 |            |           | Terceiro quartil : 64 |
| Máximo : 335         |            |           | Máximo : 99           |

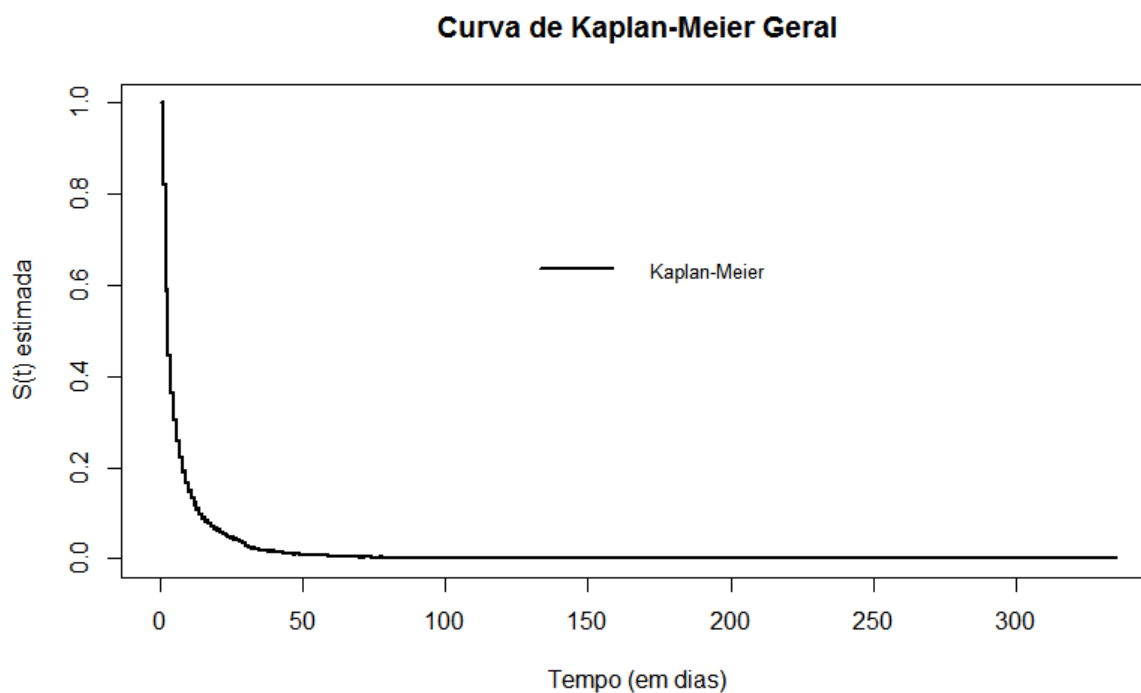
Fonte: autoria própria.

O objetivo é modelar o tempo de internação hospitalar dos pacientes no SIH/SUS em 2020 no estado de São Paulo, cada paciente foi acompanhado por 1 ano até a alta (evento de interesse) ou até a perda de acompanhamento (óbito).

O interesse de estudo é verificar se o tempo mediano de internação hospitalar se encontra dentro do prazo máximo (3 dias), estipulado pelo SIH/SUS.

Inicialmente, para verificar o comportamento dos tempos de internação hospitalar, a curva de sobrevivência global (sem considerar covariáveis) foi obtida por meio do estimador de Kaplan-Meier, pode ser visualizada na Figura 5.1.

Figura 5.1: Curva de Kaplan-Meier para os tempos de internação hospitalar.



Fonte: autoria própria.

Por meio da Figura 5.1, pode-se observar, para cada instante de tempo, a probabilidade do paciente estar internado além daquele instante de tempo. Essa curva é decrescente com o tempo.

Na Tabela 5.3 são apresentadas as estimativas das funções de sobrevivência para os tempos de internação hospitalar usando-se os modelos Weibull, Logística, Log-Logística, Log-Normal, Gama e também o Kaplan-Meier. Note que, por exemplo, as probabilidades de um paciente ter alta após de 5 dias é de 30% (Kaplan-Meier), 43% (Weibull), 45% (Logística), 31% (Log-Logística), 8% (Log-Normal) e 44% (Gama).

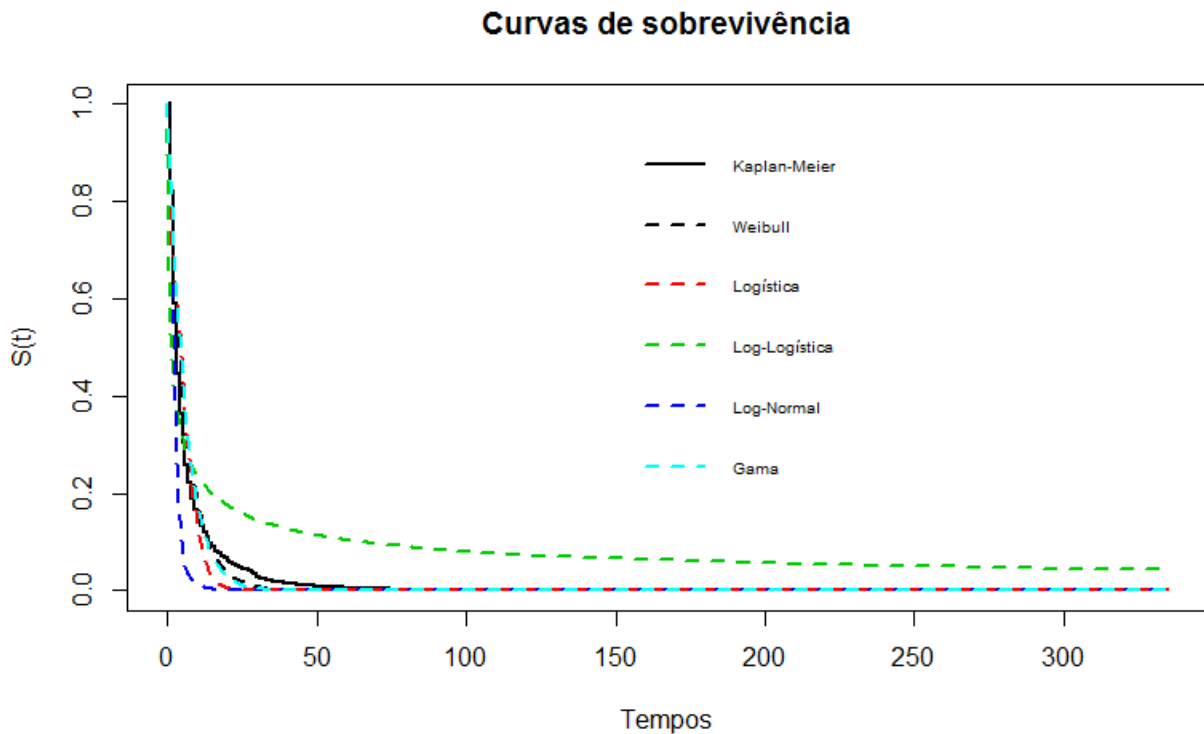
Tabela 5.3: Estimativas da sobrevivência para os tempos de internação usando-se o estimador de Kaplan-Meier e os modelos Weibull, Logística, Log-Logística, Log-Normal e Gama.

| Tempos | Kaplan-Meier | Weibull | Logística | Log-Logística | Log-Normal | Gama |
|--------|--------------|---------|-----------|---------------|------------|------|
| 1      | 0,82         | 0,84    | 0,75      | 0,53          | –          | 0,87 |
| 2      | 0,59         | 0,70    | 0,69      | 0,43          | 0,63       | 0,74 |
| 3      | 0,44         | 0,59    | 0,61      | 0,38          | 0,28       | 0,62 |
| 4      | 0,36         | 0,50    | 0,53      | 0,34          | 0,15       | 0,53 |
| 5      | 0,30         | 0,43    | 0,45      | 0,31          | 0,08       | 0,44 |
| 6      | 0,26         | 0,36    | 0,37      | 0,29          | 0,05       | 0,37 |
| 7      | 0,22         | 0,31    | 0,30      | 0,28          | 0,03       | 0,31 |
| 8      | 0,19         | 0,26    | 0,23      | 0,26          | 0,02       | 0,26 |
| 9      | 0,17         | 0,22    | 0,18      | 0,25          | 0,02       | 0,22 |
| 10     | 0,15         | 0,19    | 0,13      | 0,24          | 0,01       | 0,18 |
| 11     | 0,13         | 0,16    | 0,10      | 0,23          | 0,01       | 0,15 |
| 12     | 0,12         | 0,13    | 0,07      | 0,22          | 0,01       | 0,13 |
| 13     | 0,11         | 0,11    | 0,05      | 0,21          | 0,00       | 0,10 |
| 14     | 0,10         | 0,10    | 0,04      | 0,21          | 0,00       | 0,09 |
| 15     | 0,09         | 0,08    | 0,03      | 0,20          | 0,00       | 0,07 |
| 16     | 0,08         | 0,07    | 0,02      | 0,19          | 0,00       | 0,06 |
| 17     | 0,08         | 0,06    | 0,01      | 0,19          | 0,00       | 0,05 |
| 18     | 0,07         | 0,05    | 0,01      | 0,18          | 0,00       | 0,04 |
| 19     | 0,07         | 0,04    | 0,01      | 0,18          | 0,00       | 0,03 |
| 20     | 0,06         | 0,04    | 0,01      | 0,18          | 0,00       | 0,03 |
| 21     | 0,06         | 0,03    | 0,00      | 0,17          | 0,00       | 0,02 |
| 22     | 0,05         | 0,03    | 0,00      | 0,17          | 0,00       | 0,02 |
| 23     | 0,05         | 0,02    | 0,00      | 0,16          | 0,00       | 0,02 |
| 24     | 0,05         | 0,02    | 0,00      | 0,16          | 0,00       | 0,01 |
| 25     | 0,05         | 0,02    | 0,00      | 0,16          | 0,00       | 0,01 |
| 26     | 0,04         | 0,01    | 0,00      | 0,16          | 0,00       | 0,01 |
| 27     | 0,04         | 0,01    | 0,00      | 0,15          | 0,00       | 0,01 |
| 28     | 0,04         | 0,01    | 0,00      | 0,15          | 0,00       | 0,01 |
| 29     | 0,03         | 0,01    | 0,00      | 0,15          | 0,00       | 0,01 |
| 30     | 0,03         | 0,01    | 0,00      | 0,15          | 0,00       | 0,00 |
| ⋮      | ⋮            | ⋮       | ⋮         | ⋮             | ⋮          | ⋮    |
| 121    | 0,00         | 0,00    | 0,00      | 0,04          | 0,00       | 0,00 |

Fonte: autoria própria.

A Figura 5.2 mostra a curva de Kaplan-Meier (linha contínua), assim como as curvas dos modelos Weibull, Logística, Log-Logística, Log-Normal e Gama (linhas pontilhadas) da função de sobrevivência para os tempos de internação hospitalar. Observa-se que o ajuste dos modelos Log-Logística e Log-Normal apresentados na Figura 5.2 não tem um ajuste satisfatório. Em relação aos modelos Weibull, Logística e Gama, nota-se que estes modelos fornecem uma melhor aproximação para a curvas de Kaplan-Meier pelo menos nos primeiros dias de estudo.

Figura 5.2: Curvas de sobrevivência estimada de Kaplan-Meier e dos modelos Weibull, Logística, Log-Logística, Log-Normal e Gama para os tempos de internação hospitalar.



Fonte: autoria própria.

Neste cenário, outros modelos que melhor se ajustem aos dados devem ser usados. Assim, a variável resposta pode ser considerada como tendo uma distribuição Odd-Weibull, Gama-Weibull, Gama-Log-Logística ou Gama-Log-Normal.

Então, para determinar as estimativas dos parâmetros dos modelos Odd-Weibull, Gama-Weibull, Gama-Log-Logística e Gama-Log-Normal, utiliza-se método de máxima verossimilhança. Para este fim maximiza-se o logaritmo da função de verossimilhança (3.4), (3.6), (3.8) e (3.2) usando métodos numéricos de maximização. A programação computacional foi feita no software R versão 3.5.1 usando a função *nlm*.

Os estimadores de máxima verossimilhança dos parâmetros dos modelos foram obtidas considerando como valores iniciais  $\phi = 1,0000$ ,  $\alpha = 0,9732$  e  $\mu = 0,1703$ ; onde  $\alpha = 0,9732$  e  $\mu = 0,1703$  foram obtidos a partir do ajuste da distribuição Weibull no software R versão 3.5.1. Os estimadores de máxima verossimilhança dos parâmetros dos modelos ajustados, os erros padrões (entre parênteses) são apresentados na Tabela 5.4.

Tabela 5.4: Estimadores de máxima verossimilhança dos parâmetros dos modelos ajustados para os tempos de internação hospitalar e os erros padrões (entre parênteses).

| Modelos            | $\hat{\phi}$                      | $\hat{\alpha}$                   | $\hat{\mu}$                      | $\hat{\sigma}$                   |
|--------------------|-----------------------------------|----------------------------------|----------------------------------|----------------------------------|
| Weibull            | -                                 | 0,9732(3,4360*10 <sup>-6</sup> ) | 0,1703(8,0476*10 <sup>-6</sup> ) | -                                |
| Logística          | -                                 | -                                | 4,3789(1,6454*10 <sup>-4</sup> ) | 3,0141(5,5141*10 <sup>-6</sup> ) |
| Log-Logística      | -                                 | 0,5514(4,7644*10 <sup>-6</sup> ) | 0,8275(6,1719*10 <sup>-6</sup> ) | -                                |
| Log-Normal         | -                                 | -                                | 1,2730(6,0796*10 <sup>-6</sup> ) | 0,9551(3,5502*10 <sup>-6</sup> ) |
| Gama               | 1,1257(1,3756*10 <sup>-5</sup> )  | -                                | 0,1904(6,5110*10 <sup>-7</sup> ) | -                                |
| Odd-Weibull        | 1,1916(1,9358*10 <sup>-5</sup> )  | 0,9273(4,3931*10 <sup>-6</sup> ) | 0,1467(3,4991*10 <sup>-7</sup> ) | -                                |
| Gama-Weibull       | 13,0804(6,7425*10 <sup>-3</sup> ) | 0,2829(7,3124*10 <sup>-7</sup> ) | 2159,38(1,0619*10 <sup>4</sup> ) | -                                |
| Gama-Log-Logística | 5,9255(0,0052)                    | 2,5639(0,0002)                   | 2,7950(0,0016)                   | -                                |
| Gama-Log-Normal    | 4,8892(0,0331)                    | -                                | -1,4653(0,0080)                  | 1,1790(1,8525*10 <sup>-5</sup> ) |

Fonte: autoria própria.

As expressões das estimativas das funções de sobrevivência para os modelos Odd-Weibull, Gama-Weibull, Gama-Log-Logística e Gama-Log-Normal são, respectivamente,

$$\hat{S}(t|\hat{\theta})_{ow} = \left\{ 1 + \left\{ \exp \left[ (0,1467t)^{0,9273} \right] - 1 \right\}^{1,1916} \right\}^{-1},$$

$$\hat{S}(t|\hat{\theta})_{gw} = \frac{\Gamma [13,0804, (2159,38t)^{0,2829}]}{\Gamma(13,0804)},$$

$$\hat{S}(t|\hat{\theta})_{gll} = \frac{\Gamma [5,9255, \log \{ 1 + (2,7950t)^{2,5639} \}]}{\Gamma(5,9255)},$$

$$\hat{S}(t|\hat{\theta})_{gln} = \frac{\Gamma \left[ 4,8892, -\log \left\{ \Phi \left( \frac{-\log(t) - 1,4653}{1,1790} \right) \right\} \right]}{\Gamma(4,8892)}.$$

Os valores que aparecem nas expressões apresentadas são os estimadores de máxima verossimilhança dos parâmetros de cada um dos modelos, apresentadas na Tabela 5.4.

Os valores estimados para essas funções, por exemplo, no tempo  $t = 10$  dias são, respectivamente,

$$\hat{S}(t|\hat{\theta})_{ow} = \left\{ 1 + \left\{ \exp \left[ (0,1467t)^{0,9273} \right] - 1 \right\}^{1,1916} \right\}^{-1} = 0,2019.$$

$$\hat{S}(t|\hat{\theta})_{gw} = \frac{\Gamma [13,0804, (2159,38t)^{0,2829}]}{\Gamma(13,0804)} = 0,1480.$$

$$\hat{S}(t|\hat{\theta})_{gll} = \frac{\Gamma [5,9255, \log \{1 + (2,7950t)^{2,5639}\}]}{\Gamma(5,9255)} = 0,1400.$$

$$\hat{S}(t|\hat{\theta})_{gln} = \frac{\Gamma \left[ 4,8892, -\log \left\{ \Phi \left( \frac{-\log(t) - 1,4653}{1,1790} \right) \right\} \right]}{\Gamma(4,8892)} = 0,1389.$$

Observe que os valores obtidos por meio dos modelos Gama-Log-Logística (0,1400) e Gama-Log-Normal (0,1389) são bem próximas. O mesmo não é observado para os modelos Gama-Weibull e Odd-Weibull, que apresenta um valor estimado ligeiramente diferente dos obtidos para os outros dois modelos.

Na Tabela 5.5 são apresentadas as estimativas das funções de sobrevivência para os tempos de internação hospitalar usando-se os modelos Odd-Weibull, Gama-Weibull, Gama-Log-Logística e Gama-Log-Normal e também o Kaplan-Meier para os tempos de internação hospitalar. Note que, por exemplo, as probabilidades de um paciente ter alta após de 5 dias é de 30% (Kaplan-Meier), 47% (Odd-Weibull), 38% (Gama-Weibull), 32% (Gama-Log-Logística) e 36% (Gama-Log-Normal).

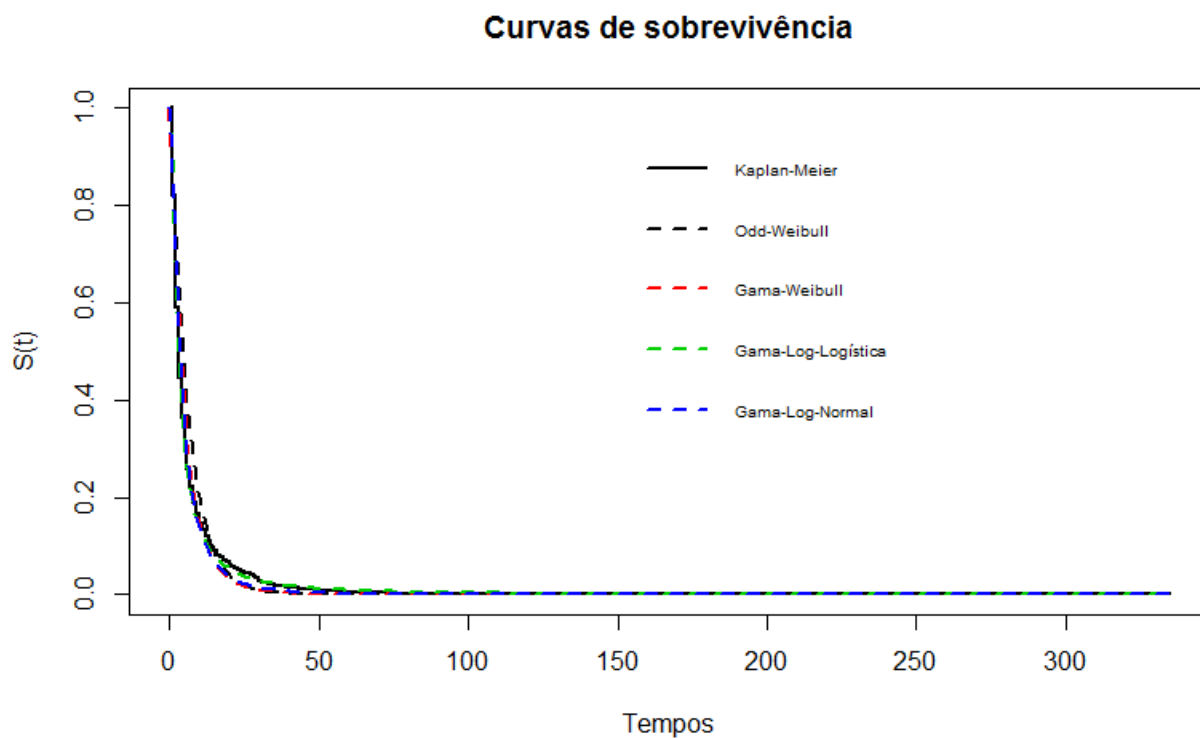
Tabela 5.5: Estimativas da sobrevivência para os tempos de internação usando-se o estimador de Kaplan-Meier e os modelos Odd-Weibull, Gama-Weibull, Gama-Log-Logística e Gama-Log-Normal.

| Tempos | Kaplan-Meier | Odd-Weibull | Gama-Weibull | Gama-Log-Logística | Gama-Log-Normal |
|--------|--------------|-------------|--------------|--------------------|-----------------|
| 1      | 0,82         | 0,88        | 0,90         | 0,94               | 0,91            |
| 2      | 0,59         | 0,76        | 0,73         | 0,70               | 0,73            |
| 3      | 0,44         | 0,65        | 0,59         | 0,52               | 0,57            |
| 4      | 0,36         | 0,55        | 0,47         | 0,40               | 0,45            |
| 5      | 0,30         | 0,47        | 0,38         | 0,32               | 0,36            |
| 6      | 0,26         | 0,39        | 0,31         | 0,26               | 0,29            |
| 7      | 0,22         | 0,33        | 0,26         | 0,22               | 0,24            |
| 8      | 0,19         | 0,28        | 0,21         | 0,19               | 0,20            |
| 9      | 0,17         | 0,24        | 0,18         | 0,16               | 0,16            |
| 10     | 0,15         | 0,20        | 0,15         | 0,14               | 0,14            |
| 11     | 0,13         | 0,17        | 0,12         | 0,12               | 0,12            |
| 12     | 0,12         | 0,15        | 0,11         | 0,11               | 0,10            |
| 13     | 0,11         | 0,12        | 0,09         | 0,10               | 0,09            |
| 14     | 0,10         | 0,11        | 0,08         | 0,09               | 0,08            |
| 15     | 0,09         | 0,09        | 0,07         | 0,08               | 0,07            |
| 16     | 0,08         | 0,08        | 0,06         | 0,07               | 0,06            |
| 17     | 0,08         | 0,07        | 0,05         | 0,07               | 0,05            |
| 18     | 0,07         | 0,06        | 0,04         | 0,06               | 0,05            |
| 19     | 0,07         | 0,05        | 0,04         | 0,06               | 0,04            |
| 20     | 0,06         | 0,04        | 0,03         | 0,05               | 0,04            |
| 21     | 0,06         | 0,04        | 0,03         | 0,05               | 0,03            |
| 22     | 0,05         | 0,03        | 0,03         | 0,05               | 0,03            |
| 23     | 0,05         | 0,03        | 0,02         | 0,04               | 0,03            |
| 24     | 0,05         | 0,02        | 0,02         | 0,04               | 0,02            |
| 25     | 0,05         | 0,02        | 0,02         | 0,04               | 0,02            |
| 26     | 0,04         | 0,02        | 0,02         | 0,04               | 0,02            |
| 27     | 0,04         | 0,01        | 0,01         | 0,03               | 0,02            |
| 28     | 0,04         | 0,01        | 0,01         | 0,03               | 0,02            |
| 29     | 0,03         | 0,01        | 0,01         | 0,03               | 0,02            |
| 30     | 0,03         | 0,01        | 0,01         | 0,03               | 0,01            |
| ⋮      | ⋮            | ⋮           | ⋮            | ⋮                  | ⋮               |
| 121    | 0,00         | 0,00        | 0,00         | 0,00               | 0,00            |

Fonte: autoria própria.

A Figura 5.3 mostra as estimativas de Kaplan-Meier da função de sobrevivência (linhas contínuas), assim como as estimativas dos modelos Odd-Weibull, Gama-Weibull, Gama-Log-Logística e Gama-Log-Normal (linhas pontilhadas) para os tempos de internação hospitalar. Observa-se que o ajuste dos modelos Odd-Weibull, Gama-Weibull, Gama-Log-Logística e Gama-Log-Normal apresentados na Figura 5.3 tem um ajuste satisfatório. Note, a partir desta figura, que o modelo Gama-Log-Logística (linha pontilhada verde) apresenta um ajuste mais do que satisfatório, pois a curva Gama-Log-Logística está quase sobre a curva de Kaplan-Meier.

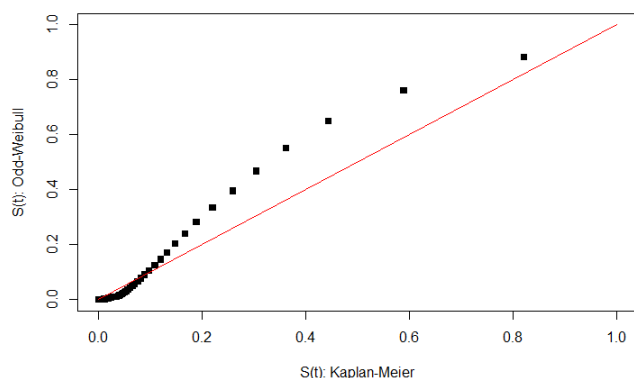
Figura 5.3: Curvas de sobrevivência estimadas por meio do ajuste dos modelos Odd-Weibull, Gama-Weibull, Gama-Log-Logística e Gama-Log-Normal *versus* a curva de sobrevivência estimada por Kaplan-Meier.



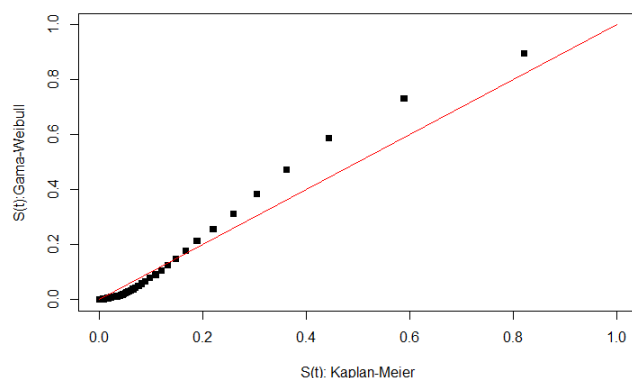
Fonte: autoria própria.

Para a escolha de um dos modelos, foram construídos os gráficos das estimativas das sobrevivências obtidas pelo método de Kaplan-Meier *versus* as estimativas das sobrevivências obtidas a partir dos modelos Odd-Weibull, Gama-Weibull, Gama-Log-Logística e Gama-Log-Normal, respectivamente. Esses gráficos encontram-se nas Figuras 5.4 e 5.5.

Figura 5.4: Gráficos das sobrevivências estimadas por Kaplan-Meier *versus* as sobrevivências estimadas pelos modelos Odd-Weibull e Gama-Weibull.

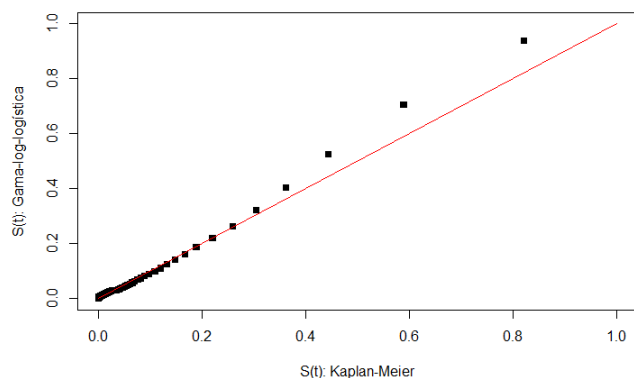


Fonte: autoria própria.

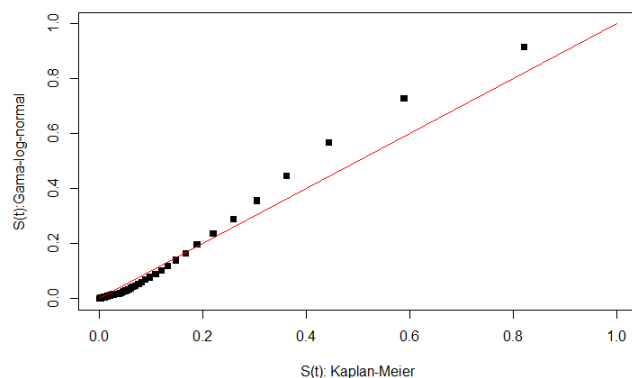


Fonte: autoria própria.

Figura 5.5: Gráficos das sobrevivências estimadas por Kaplan-Meier *versus* as sobrevivências estimadas pelos modelos Gama-Log-Logística e Gama-Log-Normal.



Fonte: autoria própria.



Fonte: autoria própria.

A partir dos gráficos apresentados nas Figuras 5.4 e 5.5, é possível observar que o modelo Odd-Weibull parece não ser adequado para esses dados, pois a curva se apresenta um tanto afastada da reta  $y = x$ . Por outro lado, os modelos Gama-Weibull, Gama-Log-Logística e Gama-Log-Normal acompanham um pouco mais de perto a reta  $y = x$ , indicando ser um desses modelos, possivelmente, adequado para os dados sob estudo.

A Tabela 5.6 apresenta os resultados dos Critérios AIC e BIC de diferentes distribuições de probabilidade, considerando-se os tempos de internação hospitalar.

Tabela 5.6: Estatísticas AIC e BIC de diferentes distribuições de probabilidade, considerando-se os dados da internação hospitalar.

| Critérios          | AIC      | BIC      |
|--------------------|----------|----------|
| Weibull            | 795371,2 | 795391,1 |
| Logística          | 932226,6 | 932246,5 |
| Log-Logística      | 752658,5 | 752678,4 |
| Log-Normal         | 749783,8 | 749803,7 |
| Gama               | 794334,0 | 794353,8 |
| Odd-Weibull        | 793462,0 | 793491,8 |
| Gama-Weibull       | 759629,3 | 759659,1 |
| Gama-Log-Logística | 732981,4 | 733011,2 |
| Gama-Log-Normal    | 746145,4 | 746175,2 |

Fonte: autoria própria.

Então, pode-se concluir através da Tabela 5.6, que a distribuição Gama-Log-Logística apresenta os menores valores dos critérios AIC e BIC, indicando que seria o modelo mais adequado, dentre os modelos ajustados aos dados da internação hospitalar. Assim, a distribuição Gama-Log-Logística foi considerada para analisar o conjunto de dados de tempo de internação hospitalar. Portanto, o modelo é dado por:

$$\hat{S}(t|\hat{\theta})_{gll} = \frac{\Gamma [5, 9255, \log \{1 + (2, 7950t)^{2,5639}\}]}{\Gamma(5, 9255)}, \quad t > 0, \quad (5.1)$$

em que  $T$  é a variável aleatória que representa o tempo de internação hospitalar dos pacientes.

Então, é possível obter as seguintes informações sobre o conjunto de dados do tempo de internação hospitalar dos pacientes a partir do modelo ajustado (5.1):

- O tempo mediano, denotado por  $\hat{t}_{0,5}$ , pode ser estimado por meio da função quantílica (2.6). Dessa forma, tem-se

$$\hat{t}_{0,5} = \frac{1}{2, 7950} \left[ \exp(\hat{Q}_{5,9255}(0, 5)) - 1 \right]^{1/2,5639} = 3, 1683;$$

em que  $\hat{Q}_{5,9255}(0, 5) = 5, 5957$  é a função quantílica da distribuição gama de um parâmetro.

Logo, o tempo mediano de internação hospitalar dos pacientes é aproximadamente de 3 dias.

- b. A probabilidade de um paciente ter alta depois de 10 dias é estimada por meio da função de sobrevivência (5.1). Assim,

$$\hat{S}(t|\hat{\theta})_{gll} = \frac{\Gamma [5, 9255; \log \{1 + (2, 7950t)^{2,5639}\}]}{\Gamma(5, 9255)} = 0,1400.$$

Portanto, a probabilidade do paciente ter alta após de 10 dias é de 0,14, ou seja, a probabilidade dos pacientes ficarem internados menos de 10 dias é 0,86.

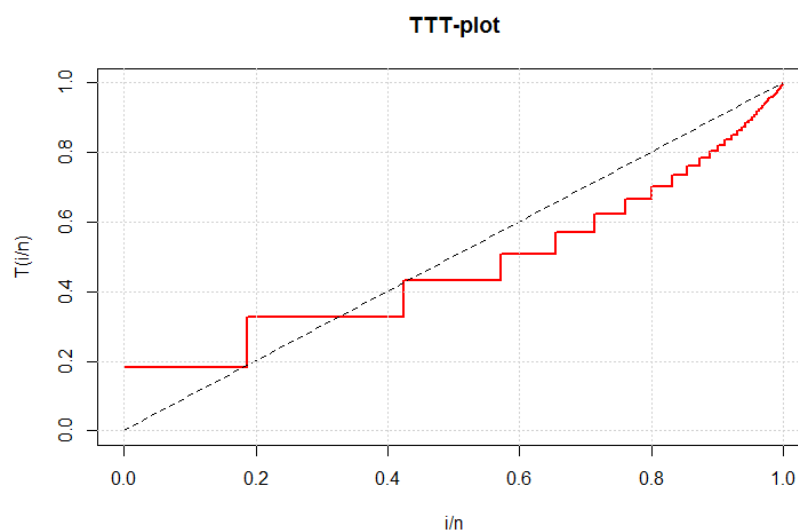
Agora, deseja-se verificar se o sexo e a idade do paciente exercem alguma influência no tempo de internação, bem como se existe interação entre essas covariáveis.

Dessa forma, a variável resposta foi determinada como sendo o tempo (em dias) da data de internação até a receber alta (data de saída). No entanto, os pacientes que não tiveram alta, ou seja, faleceram nesse período, foram considerados como tempos censurados. Assim, as seguintes variáveis foram observados:

- i.  $t$ : Tempo de internação (em dias);
- ii.  $x_1$ : Idade (em anos);
- iii.  $x_2$ : Sexo (1 =feminino, 0 =masculino).

Primeiramente, para verificar o comportamento da função de risco ou taxa de falha dos dados de tempos de internação foi construído a curva TTT, representada na Figura 5.6. A partir dessas figuras, observa-se que o TTT-plot indica que os tempos de internação dos pacientes apresentam uma taxa de falha unimodal. Portanto, para a análise dos dados, é utilizado o modelo de regressão Gama-Log-Logística, pois o modelo Gama-Log-Logística se ajusta melhor aos dados de tempos de internação com taxa de falha unimodal. .

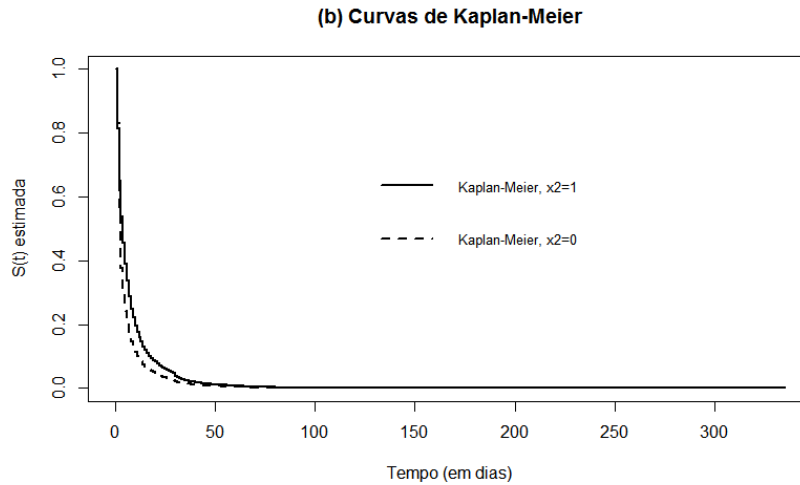
Figura 5.6: TTT-plot para os tempos de internação.



Fonte: autoria própria.

Para cada covariável apresentada na 5.1 foi utilizado o estimador de Kaplan-Meier com o propósito de verificar a associação de cada uma delas com o tempo de internação hospitalar. As curvas de sobrevivência estimadas para as covariáveis da Tabela 5.1 encontram-se na Figura 5.7.

Figura 5.7: Curvas de sobrevivência para o tempo de internação hospitalar, para a covariável sexo, estimadas pelo método de Kaplan-Meier



Fonte: autoria própria.

A análise das curvas apresentadas na Figura 5.7 sugere a presença de diferença significativa para a covariável sexo. Para essa covariável, as curvas não estão próximas uma da outra o que pode indicar que as mesmas diferem entre si ao longo do tempo. Por outro lado, devido ao grande número de observações disponíveis (amostra grande), e também o fato de que a idade é considerada como sendo um fator de prognóstico usual, a covariável idade é considerada.

Dessa forma, de acordo com que foi observado nas Figuras 5.6 e 5.7, pode-se considerar os modelos Gama-Log-Logística, como segue:

1.  $t_i = \exp(\beta_0) \exp(z_i/\alpha)$ ,  $i = 1, \dots, 153550$ ,
2.  $t_i = \exp(\beta_0 + \beta_1 x_{1i}) \exp(z_i/\alpha)$ ,  $i = 1, \dots, 153550$ ,
3.  $t_i = \exp(\beta_0 + \beta_2 x_{2i}) \exp(z_i/\alpha)$ ,  $i = 1, \dots, 153550$ ,
4.  $t_i = \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}) \exp(z_i/\alpha)$ ,  $i = 1, \dots, 153550$ ,
5.  $t_i = \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i}) \exp(z_i/\alpha)$ ,  $i = 1, \dots, 153550$ ,

em que  $t_i$  denota os tempos de internação,  $x_1$  denota a idade e  $x_2$  denota o sexo de cada paciente.

Para maximizar a função (4.3) e assim encontrar os estimadores de máxima verossimilhança dos parâmetros dos modelos propostos acima, foi utilizado o pacote estatístico R versão 3.5.1. com os valores iniciais obtidos a partir do ajuste do modelo de regressão Log-Logística no pacote estatístico R versão 3.5.1. Dessa forma, na Tabela 5.7 encontra-se as estimativas dos parâmetros dos modelos, os erros padrões e as significâncias dos parâmetros.

Tabela 5.7: Estimadores de máxima verossimilhança dos parâmetros dos modelos Gama-Log-Logística para os dados de tempos de internação.

| Modelos | Parâmetro       | Estimativa | Erro padrão      | p-valor | Intervalo de confiança(95%)         |
|---------|-----------------|------------|------------------|---------|-------------------------------------|
| 1       | $\hat{\phi}$    | 5,9257     | 0,0051           | –       | (5,9156;5,9359)                     |
|         | $\hat{\alpha}$  | 2,5639     | 0,0002           | –       | (2,5634;2,5644)                     |
|         | $\hat{\beta}_0$ | -1,0279    | 0,0002           | 0,0001  | (-1,0283;-1;0274)                   |
| 2       | $\hat{\phi}$    | 6,4013     | $7,9735*10^{-3}$ | –       | (6,3857;6,4169)                     |
|         | $\hat{\alpha}$  | 2,6827     | $4,0790*10^{-4}$ | –       | (2,6819;2,6835)                     |
|         | $\hat{\beta}_0$ | -1,3508    | $3,1739*10^{-4}$ | 0,0001  | (-1,3514;-1,3502)                   |
|         | $\hat{\beta}_1$ | 0,0056     | $8.7269*10^{-9}$ | 0,0001  | $(5,6444*10^{-3};5,6445*10^{-3})$   |
| 3       | $\hat{\phi}$    | 6,1094     | $6,0815*10^{-3}$ | –       | (6,0975;6,1214)                     |
|         | $\hat{\alpha}$  | 2,6046     | $3,1938*10^{-4}$ | –       | (2,6040;2,6052)                     |
|         | $\hat{\beta}_0$ | -0,9981    | $2,3119*10^{-4}$ | 0,0001  | (-0,9985;-0,9976)                   |
|         | $\hat{\beta}_2$ | -0,1120    | $1,8092*10^{-5}$ | 0,0001  | $(-1,1207*10^{-1};-1,1200*10^{-1})$ |
| 4       | $\hat{\phi}$    | 6,5822     | $9,0045*10^{-3}$ | –       | (6,5645;6,5998)                     |
|         | $\hat{\alpha}$  | 2,7220     | $4,5168*10^{-4}$ | –       | (2,7211;2,7229)                     |
|         | $\hat{\beta}_0$ | -1,3269    | $3,3683*10^{-4}$ | 0,0001  | (-1,3275;-1,3262)                   |
|         | $\hat{\beta}_1$ | 0,0054     | $8,8770*10^{-9}$ | 0,0001  | $(5,4178*10^{-3};5,4178*10^{-3})$   |
|         | $\hat{\beta}_2$ | -0,0806    | $1,8423*10^{-5}$ | 0,0001  | (0,0053;0,0054)                     |
| 5       | $\hat{\phi}$    | 6,6203     | $9,2146*10^{-3}$ | –       | (6,6022;6,6384)                     |
|         | $\hat{\alpha}$  | 2,7299     | $4,6013*10^{-4}$ | –       | (2,7290;2,7308)                     |
|         | $\hat{\beta}_0$ | -1,3104    | $3,4858*10^{-4}$ | 0,0001  | (-1,3110;-1,3097)                   |
|         | $\hat{\beta}_1$ | 0,0049     | $1,8278*10^{-8}$ | 0,0001  | $(4,9330*10^{-3};4,9331*10^{-3})$   |
|         | $\hat{\beta}_2$ | -0,1228    | $9,0465*10^{-5}$ | 0,0001  | (-0,1230;-0,1226)                   |
|         | $\hat{\beta}_3$ | 0,0009     | $3,4811*10^{-8}$ | 0,0001  | $(9,2991*10^{-3};9,3004*10^{-3})$   |

Fonte: autoria própria.

Por meio da Tabela 5.7, observa-se que as covariáveis  $x_1$  e  $x_2$  influenciam na variabilidade do tempo de internação, ao nível de 5% de significância, em todos os modelos de regressão.

Para dar continuidade a análise dos dados, considerou-se as estatísticas AIC e BIC definida na Seção 1.7 para a escolha do modelo mais adequado para descrever o comportamento do tempo de internação. A Tabela 5.8 apresenta os resultados das estatísticas AIC e BIC, considerando-se os tempos de internação hospitalar.

Tabela 5.8: Estatísticas AIC e BIC para os dados da internação hospitalar.

| Modelos | Critérios |          |
|---------|-----------|----------|
|         | AIC       | BIC      |
| 1       | 732981,4  | 733011,2 |
| 2       | 729320,6  | 729360,4 |
| 3       | 732289,5  | 732329,3 |
| 4       | 729957,7  | 730007,4 |
| 5       | 728946,8  | 729006,4 |

Fonte: autoria própria.

Então, pode-se concluir a partir dos valores das estatísticas AIC e BIC na Tabela 5.8, que o último modelo de tempo de vida acelerado ( AIC = 728946, 8 e BIC = 729006, 4) é mais adequado do que os outros modelos. Pois o modelo apresenta os menores valores das estatísticas AIC e BIC dentre os modelos ajustados aos dados da internação hospitalar.

Logo, o modelo final ajustado aos dados de tempo de internação é expresso por:

$$t_i = \exp(-1,3104 + 0,0049x_{1i} - 0,1228x_{2i} + 0,0009x_{1i}x_{2i}), \quad i = 1, \dots, 153550 \quad (5.2)$$

Note, para o modelo ajustado (5.2), que  $\hat{\beta}_1$  é positivo, o que implica que quanto maior o valor de  $x_1$  (idade), maior a probabilidade do paciente estar internado. Observe que  $\hat{\beta}_2$  é negativo, o que implica que pacientes do sexo feminino ( $x_2 = 1$ ) apresentam menor probabilidade de internação do que a dos pacientes do sexo masculino ( $x_2 = 0$ ). Por outro lado, tem-se que  $\hat{\beta}_3$  é positivo, o que implica que existe o efeito de interação entre sexo e idade, assim os pacientes mais velhos do sexo feminino tem maior chance de ficar internado por mais tempo.

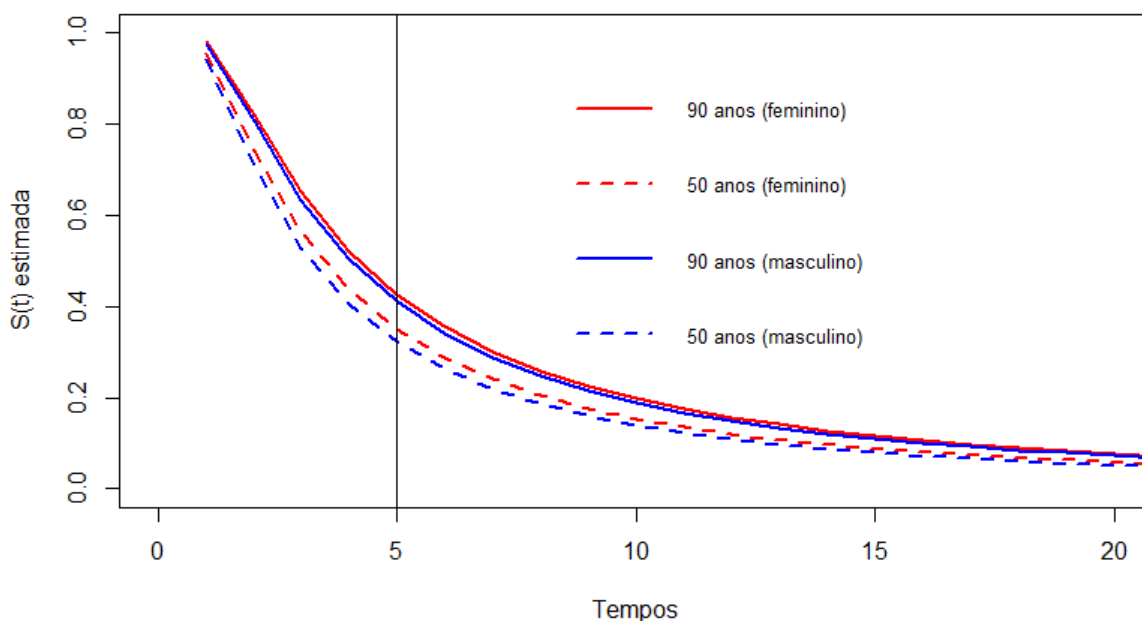
A função de sobrevivência estimada, por este modelo, para um paciente internado no hospital é obtida por:

$$\hat{S}(t|x_1, x_2) = \frac{\Gamma \left[ 6,6203, \log \left\{ 1 + [\exp(1,3104 - 0,0049x_1 + 0,1228x_2 - 0,0009x_1x_2)t]^{2,7299} \right\} \right]}{\Gamma(6,6203)} \quad (5.3)$$

para  $t > 0$ , em que  $x_1$  é a variável idade em anos observada para este paciente internado e  $x_2$  indica se o paciente pertence ao sexo feminino ou masculino. Para pacientes do sexo feminino, tem-se  $x_2 = 1$ . Em caso contrário,  $x_2 = 0$ .

Na Figura 5.8, observa-se as curvas de sobrevivência estimadas para dois pacientes pertencentes ao sexo feminino e dois outros pacientes pertencentes ao sexo masculino, um com  $x_1 = 90$  e outro com  $x_1 = 50$ , são apresentadas. Note que, o efeito de interação entre sexo e idade pode ser visto na diferença da distancia entre a curva de sobrevivência estimada do paciente com 90 e 50 anos do sexo feminino e masculino. Esta distancia entre a curva de sobrevivência do feminino com masculino com 50 anos é maior do que a distancia entre a curva de sobrevivência do feminino com masculino com 90 anos. Também pode-se observar o efeito da idade, a curva de sobrevivência do paciente com 90 anos fica acima da curva de sobrevivência do paciente com 50 anos, isto é, pacientes com idade avançada tem a tendencia de ficar mais tempo internados.

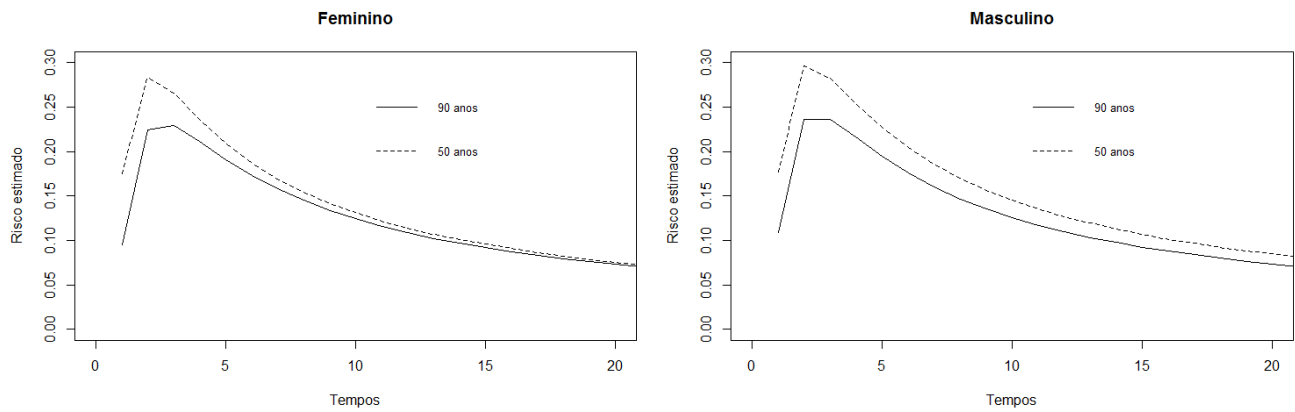
Figura 5.8: Curvas de sobrevivência estimadas pelo modelo de regressão Gama-Log-Logística para dois pacientes do sexo feminino e dois pacientes do sexo masculino internados no hospital.



Fonte: autoria própria.

Os correspondentes riscos estimados dos pacientes considerados na Figura 5.8 encontram-se na Figura 5.9. Estes são unimodal ao longo do tempo, o que é uma característica do modelo Gama-Log-Logística. Pode-se notar que a taxa instantânea de falha do paciente com  $x_1 = 50$ , em relação ao paciente com  $x_1 = 90$ , é maior tanto do sexo feminino quanto do sexo masculino. Portanto, quanto menor a idade do paciente, maior o risco de falha. Comparativamente, os pacientes do sexo masculino apresentam risco de falha estimado maior do que o referido risco estimado para os pacientes do sexo feminino.

Figura 5.9: Riscos estimados pelo modelo de regressão Gama-Log-Logística para dois pacientes do sexo feminino e dois pacientes do sexo masculino internados no hospital.



Fonte: autoria própria.

Fonte: autoria própria.

---

# Considerações Finais e Propostas Futuras

## 6.1 Considerações Finais

- Por meio do presente trabalho pode-se constatar que os modelos Odd-Weibull e Gama-G são modelos muito flexíveis para modelar os dados de sobrevivência, pois acomodam várias formas da função de risco, sendo estas: constante, crescente, decrescente, forma de banheira (unimodal).
- Neste trabalho, a fim de se obter bons valores iniciais para serem utilizados nos procedimentos de estimação dos parâmetros das distribuições de probabilidade Odd-Weibull e Gama-G considerando-se dados censurados sob o enfoque clássico, utiliza-se as estimativas dos modelos de dois parâmetros.
- Para a análise dos dados de tempos de internação hospitalar SIH/SUS, os modelos provenientes dos modelos Gama-G, se mostraram os mais adequados, em particular o modelo Gama-Log-Logística (Hashimoto, 2013).
- O modelo de regressão permitiu mostrar que o tempo de internação em geral é maior em homens comparado às mulheres e além disso, pessoas com idade maior tem mais tempo de internação.
- O efeito de interação significativo indica que apesar do sexo feminino ter tempo de internação menor, com o aumento da idade esta diferença tende a diminuir.

## 6.2 Propostas Futuras

- Pretende-se em um futuro breve, estudar melhor as propriedades, bem como a função geradora de momentos da distribuição Gama-Log-Normal, que também mostrou bons ajustes aos dados desta natureza.
- Pretende-se também, realizar análise de resíduos no modelo de regressão. Os resíduos mais utilizados como o resíduo martingale e o resíduo componente do desvio que verificam a adequação do modelo ajustado e auxiliam a encontrar observações atípicas.

# Programas Utilizados

## A.1 Análise clássica considerando dados censurados

Algoritmo desenvolvido no R para calcular as estimativas de máxima verossimilhança (EMV) dos parâmetros dos modelos.

```
##### Limpando dados do R
rm(list=ls())

# Pacotes necessários

library(read.dbc)
library(stringr)
library(dplyr)
library(tidyr)
library(lubridate)
library(survival)
library(zipfR)

## Leitura da base de dados
dados18 = read.dbc(file.choose())

# Calculo do tempo total internado
dados= dados18 %>% mutate(tempototal=ymd(dados18$DT_SAIDA)-ymd(dados18$DT_INTER))

# Filtragem dos dados somente para pacientes cuja variavel "DIAS_PERM"
# é igual ao tempo total calculado
dados1 = dados %>% filter(DIAS_PERM == tempototal)

## Criando uma variavel com TRUE or FALSE para as palavras
#que comecem com F (Transtornos mentais e comportamentais)
Trans_M_C <- str_detect(dados1$DIAG_PRINC ,pattern="^F")
head(Trans_M_C)
table(Trans_M_C)
```

```

# Define as variáveis
tempo<-dados1$DIAS_PERM
cens<-1-dados1$MORTE
sexo<-factor(dados1$SEXO)
idade<-dados1$IDADE
#####

dd<-data.frame(Trans_M_C,tempo,cens,sexo,idade)
str(dd)

# Não Transtornos mentais e comportamentais e tempo > 0
dd1<-dd %>% filter(Trans_M_C == "FALSE",tempo > 0)

#Análise não paramétrica Kaplan-Meier
ekm1<-survfit(Surv(dd1$tempo,dd1$cens)~1)
summary(ekm1)

plot(ekm1,xlab = "Tempo (em dias)",ylab = "S(t) estimada",lwd = 2)
legend(120,0.8,lty = 1,legend = "Kaplan-Meier",bty = "n",cex = 0.8,lwd = 2)
title("Curva de Kaplan-Meier")

#####
#####
#####
#####

##### Log da Verossimilhança -- GAMA #####
log_lik <- function(theta)
{
fi0 <- theta[1]
alfa <- 1
mi0 <- theta[2]
lf<- sum(delta*log(alfa*(mi0^(alfa*fi0))/gamma(fi0)))+
(alfa*fi0-1)*sum(delta*log(t))-(mi0^alfa)*sum(delta*(t^alfa))+
sum((1-delta)*log((pgamma((mi0*t)^alfa,fi0,lower=FALSE)*gamma(fi0))/gamma(fi0)))
lf1 =-(lf)
return(lf1)
}

#####
t<-dd1$tempo
delta<-dd1$cens
#-----
# AJUSTE
v0=c(1,0.1814143)
r1=nlm(log_lik, v0, hessian = TRUE)

```

```

r1
#-----
##### Log da Verossimilhanca -- G-Weibull#####
log_like <- function(theta)
{
fi <- theta[1]
alfa <- theta[2]
mi <- theta[3]
lf<- sum(delta*log(alfa*(mi^(alfa*fi))/gamma(fi)))+
(alfa*fi-1)*sum(delta*log(t))-(mi^alfa)*sum(delta*(t^alfa))+
sum((1-delta)*log((pgamma((mi*t)^alfa,fi,lower=FALSE)*gamma(fi))/gamma(fi)))
lf1 =-(lf)
return(lf1)
}

#=====#
t<-dd1$tempo
delta<-dd1$cens
#-----
# AJUSTE
v0=c(1,0.9732,0.1703)
r2=nlm(log_like, v0, hessian = TRUE)
r2
#-----
##### Log da Verossimilhanca -- G-log-logistica#####
log_like1 <- function(theta)
{
fi1 <- theta[1]
alf1 <- theta[2]
mi1 <- theta[3]
lf<- sum(delta*log(alf1*(mi1^(alf1))/gamma(fi1)))+
(alf1-1)*sum(delta*log(t))-
2*sum(delta*log(1+(mi1*t)^alf1))+(fi1-1)*sum(delta*log(log(1+(mi1*t)^alf1)))+
sum((1-delta)*log((pgamma(log(1+(mi1*t)^alf1),fi1,lower=FALSE))))
lf1 =-(lf)
return(lf1)
}

#=====#
t<-dd1$tempo
delta<-dd1$cens
#----
# AJUSTE
v0=c(1,0.9732,0.1703)
r3=nlm(log_like1, v0, hessian = TRUE)
r3
#-----

```

```
##### Log da Verossimilhanca -- G-log-normal#####
log_like2 <- function(theta)
{
  fii <- theta[1]
  mii <- theta[2]
  si <- theta[3]
  lf<- sum(delta*log(1/(sqrt(2*pi)*si*t* gamma(fii))))-
  0.5*sum(delta*((log(t)-mii)/si)^2)+
  (fii-1)*sum(delta*log(-log(pnorm((-log(t)+mii)/si))))+
  sum((1-delta)*log((pgamma(-log(pnorm((-log(t)+mii)/si)),fii,lower=FALSE)))
  lf1 =-(lf)
  return(lf1)
}

#=====#
t<-dd1$tempo
delta<-dd1$cens
#-----
# AJUSTE
v0=c(1,0.9732,0.1703)
r4=nlm(log_like2, v0, hessian = TRUE)
r4
#-----
##### Log da Verossimilhanca -- Odd-Weibull #####
log_like3 <- function(theta)
{
  fi2 <- theta[1]
  alf2 <- theta[2]
  mi2 <- theta[3]
  lf<- d*log(alf2)+(d*alf2)*log(mi2)+d*log(fi2)+
  (alf2-1)*sum(delta*log(t))+sum(delta*((t*mi2)^alf2))+
  (fi2-1)*sum(delta*log(exp((t*mi2)^alf2))-1)-
  sum((1+delta)*log(1+(exp((t*mi2)^alf2)-1)^fi2))
  lf1 =-(lf)
  return(lf1)
}

#=====#
d<-nrow(filter(dd1,cens==1))
t<-dd1$tempo
delta<-dd1$cens
#----
# AJUSTE
v0=c(1,0.9732,0.1703)
r5=nlm(log_like3, v0, hessian = TRUE)
r5
#-----
```

## A.2 Análise de regressão considerando dados censurados

Algoritmo desenvolvido no R para calcular as estimativas de máxima verossimilhança (EMV) dos parâmetros do modelo de regressão Gama-Log-Logística.

```
rm(list=ls())
# Pacotes

library(stringr)
library(read.dbc)
library(dplyr)
library(tidyr)
library(lubridate)
library(survival)
library(zipfR)

## Leitura da base de dados
dados20 = read.dbc(file.choose())

# Calculo do tempo total internado
dados= dados20 %>% mutate(tempototal=ymd(dados20$DT_SAIDA)-ymd(dados20$DT_INTER))

# Filtragem dos dados somente para pacientes cuja variavel "DIAS_PERM"
# é igual ao tempo total calculado
dados1 = dados %>% filter(DIAS_PERM == tempototal)

## Criando uma variavel com TRUE or FALSE para as palavras
#que comecem com F (Transtornos mentais e comportamentais)
Trans_M_C <- str_detect(dados1$DIAG_PRINC ,pattern="^F")

# Define as variáveis
tempo<-dados1$DIAS_PERM
cens<-1-dados1$MORTE
#sexo<-factor(dados1$SEXO)
sexo<-dados1$SEXO
idade<-dados1$IDADE

#####

dd<-data.frame(Trans_M_C,tempo,cens,sexo,idade)
str(dd)

dd$sexo=as.numeric(dd$sexo)
sexo=dd$sexo-1
dd<-data.frame(Trans_M_C,tempo,cens,sexo,idade)
str(dd)
```

```

# Não Transtornos mentais e comportamentais e tempo > 0
dd1<-dd %>% filter(Trans_M_C == "FALSE",tempo > 0)

#####

# dados
t<-dd1$tempo
delta<-dd1$cens
d=sum(delta)
x1<-dd1$idade
x2<-dd1$sexo
x3<-x1*x2
#-----
#-----
#TTT-plot

library(AdequacyModel)
TTT(t, col = "red", lwd = 2.5, grid = TRUE, lty = 2)
title("TTT-plot")

#-----
#-----
#Curva de Kaplan-Meier (Tempo)
ekm<-survfit(Surv(t,delta)~1)
summary(ekm)

plot(ekm,xlab = "Tempo (em dias)",ylab = "S(t) estimada",lwd=2)
legend(120,0.8,lty = 1,legend = "Kaplan-Meier",lwd=2,
bty = "n",cex = 0.8)
title("(a) Curva de Kaplan-Meier ")

#-----
#-----
#Curva de Kaplan-Meier (~dd1$sexo)
ekm1<-survfit(Surv(t,delta)~dd1$sexo)
summary(ekm1)

plot(ekm1,xlab = "Tempo (em dias)",ylab = "S(t) estimada",lty = c(1,2),lwd=2)
legend(120,0.8,lty = c(1,2),c("Kaplan-Meier, x2=0", "Kaplan-Meier, x2=1"),
lwd=2, bty = "n",cex = 0.8)
title("(b) Curvas de Kaplan-Meier")

#####
#####  MODELO DE TEMPO DE VIDA ACELERADO #####
#           t=exp(x'B)exp(az)
#####

```

```
##### Log da Verossimilhanca -- G-log-logistica#####
```

```
log_like0<- function(theta)
{
fi <- theta[1]
alf <- theta[2]
b0<- theta[3]
xb <- b0
lf<- d*log(alf)- d*log(gamma(fi))-alf*sum(delta*(xb))+
(alf-1)*sum(delta*log(t))-2*sum(delta*log(1+(exp(-xb)*t)^alf))+
(fi-1)*sum(delta*log(log(1+(exp(-xb)*t)^alf)))+
sum((1-delta)*log(pgamma(log(1+(exp(-xb)*t)^alf),fi,lower=FALSE)))
lf1 =-(lf)
return(lf1)
}
#.....
```

```
theta=c(1,1.027442,1.769778)
r0=nlm(log_like0,theta,hessian=T)
r0
#=====
```

```
log_like1<- function(theta)
{
fi <- theta[1]
alf <- theta[2]
b0<- theta[3]
b1<- theta[4]
xb <- b0+b1*x1
lf<- d*log(alf)- d*log(gamma(fi))-alf*sum(delta*(xb))+
(alf-1)*sum(delta*log(t))-2*sum(delta*log(1+(exp(-xb)*t)^alf))+
(fi-1)*sum(delta*log(log(1+(exp(-xb)*t)^alf)))+
sum((1-delta)*log(pgamma(log(1+(exp(-xb)*t)^alf),fi,lower=FALSE)))
lf1 =-(lf)
return(lf1)
}
#.....
```

```
theta=c(1,1.01818,1.395013364,0.008374329)
r1=nlm(log_like1,theta,hessian=T)
r1
#=====
```

```
log_like2<- function(theta)
{
fi <- theta[1]
alf <- theta[2]
b0<- theta[3]
```

```

b2<- theta[4]
xb <- b0+b2*x2
lf<- d*log(alf)- d*log(gamma(fi))-alf*sum(delta*(xb))+
(alf-1)*sum(delta*log(t))-2*sum(delta*log(1+(exp(-xb)*t)^alf))+
(fi-1)*sum(delta*log(log(1+(exp(-xb)*t)^alf)))+
sum(((1-delta)*log(pgamma(log(1+(exp(-xb)*t)^alf),fi,lower=FALSE)))
lf1 =-(lf)
return(lf1)
}
#.....
theta=c(1,1.015453,1.9553655,-0.3331068)
r2=nlm(log_like2,theta,hessian=T)
r2
#=====
log_like3<- function(theta)
{
fi <- theta[1]
alf <- theta[2]
b0<- theta[3]
b1<- theta[4]
b2<- theta[5]
xb <- b0+b1*x1+b2*x2
lf<- d*log(alf)- d*log(gamma(fi))-alf*sum(delta*(xb))+
(alf-1)*sum(delta*log(t))-2*sum(delta*log(1+(exp(-xb)*t)^alf))+
(fi-1)*sum(delta*log(log(1+(exp(-xb)*t)^alf)))+
sum(((1-delta)*log(pgamma(log(1+(exp(-xb)*t)^alf),fi,lower=FALSE)))
lf1 =-(lf)
return(lf1)
}
#.....
theta=c(1,1.006401,1.586099974,0.007954165,-0.308531661)
r3=nlm(log_like3,theta,hessian=T)
r3

#5°
#=====
log_like4<- function(theta)
{
fi <- theta[1]
alf <- theta[2]
b0<- theta[3]
b1<- theta[4]
b2<- theta[5]
b3<- theta[6]
xb <- b0+b1*x1+b2*x2+b3*x3
lf<- d*log(alf)- d*log(gamma(fi))-alf*sum(delta*(xb))+
(alf-1)*sum(delta*log(t))-2*sum(delta*log(1+(exp(-xb)*t)^alf))+

```

```
(fi-1)*sum(delta*log(log(1+(exp(-xb)*t)^alf)))+
sum((1-delta)*log(pgamma(log(1+(exp(-xb)*t)^alf),fi,lower=FALSE)))
lf1 =-(lf)
return(lf1)
}
#.....
theta=c(1,1.004327,1.721835557,0.004978534,-0.529956548,0.005110581)
r4=nlm(log_like4,theta,hessian=T)
r4

#=====
```

# Referências

---

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, **19**(6), 716–723.
- Barlow, R. E. & Campo, R. (1975). Total time on test processes and applications to failure data analysis. In *Reliability and fault tree analysis*.
- Bennett, S. (1983). Log-logistic regression models for survival data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **32**(2), 165–171.
- Carrasco, J. M., Ortega, E. M. & Cordeiro, G. M. (2008). A generalized modified weibull distribution for lifetime modeling. *Computational Statistics and Data Analysis*, **53**(2), 450–462.
- Casella, G. & Berger, R. L. (2021). *Statistical inference*. Cengage Learning.
- Colosimo, E. A. & Giolo, S. R. (2006). *Análise de sobrevivência aplicada*.. Editora Blucher.
- Cooray, K. (2006). Generalization of the weibull distribution: the odd weibull family. *Statistical Modelling*, **6**(3), 265–277.
- DATASUS (2021). [www2.datasus.gov.br/DATASUS/index.php](http://www2.datasus.gov.br/DATASUS/index.php), 09 de fevereiro de 2021.
- Hashimoto, E. M. (2013). *Modelo de regressão gama-G em análise de sobrevivência*.. Ph.D. thesis, Universidade de São Paulo.
- Johnson, N. L., Kotz, S. & Balakrishnan, N. (1995). *Continuous univariate distributions, volume 2*., volume 289. John wiley and sons.
- Johnson, N. L., Kemp, A. W. & Kotz, S. (2005). *Univariate discrete distributions*., volume 444. John Wiley and Sons.
- Kleinbaum, D. G. & Klein, M. (2010). *Survival analysis*.. Springer.
- Lawless, J. F. (2003). Statistical models and methods for lifetime data. *Copyright© 2003 John Wiley and Sons, Inc.*, **3**, 577.
- Lawless, J. F. (2011). *Statistical models and methods for lifetime data*., volume 362. John Wiley and Sons.
- Lee, E. T. & Wang, J. (2003). *Statistical methods for survival data analysis*., volume 476. John Wiley and Sons.
- Liu, X. (2012). *Survival analysis: models and applications*.. John Wiley and Sons.

- Ortega, E. M., Cordeiro, G. M., Hashimoto, E. M. & Cooray, K. (2014). A log-linear regression model for the odd weibull distribution with censored data. *Journal of Applied Statistics*, **41**(9), 1859–1880.
- Ramos, P. L. (2014). *Aspectos computacionais para inferência na distribuição gama generalizada*. Master's thesis, Universidade Estadual Paulista (UNESP).
- Schwarz, G. *et al.* (1978). Estimating the dimension of a model. *Annals of statistics*, **6**(2), 461–464.
- Stacy, E. W. *et al.* (1962). A generalization of the gamma distribution. *The Annals of mathematical statistics*, **33**(3), 1187–1192.
- Weibull, W. *et al.* (1951). A statistical distribution function of wide applicability. *Journal of applied mechanics*, **18**(3), 293–297.
- Zografos, K. & Balakrishnan, N. (2009). On families of beta-and generalized gamma-generated distributions and associated inference. *Statistical methodology*, **6**(4), 344–362.