



MICAÍAS MOZZER DE OLIVEIRA

**Estudo do tempo de sobrevida de pacientes internados pela Covid-19 no Brasil.**

MARINGÁ – PARANÁ  
2022

MICAÍAS MOZZER DE OLIVEIRA

**Estudo do tempo de sobrevida de pacientes internados  
pela Covid-19 no Brasil.**

Dissertação apresentada ao Programa de Pós-graduação em Bioestatística do centro de ciências exatas da Universidade Estadual de Maringá como requisito parcial para obtenção do título de mestre em Bioestatística.

Orientador: Prof. Dr. Brian Alvarez Ribeiro de Melo

Universidade Estadual de Maringá - UEM

Departamento de Estatística - DES

Programa de Pós-Graduação em Bioestatística

MARINGÁ – PARANÁ

2022

Dados Internacionais de Catalogação-na-Publicação (CIP)  
(Biblioteca Central - UEM, Maringá - PR, Brasil)

O48e

Oliveira, Micaías Mozzer de

Estudo do tempo de sobrevida de pacientes internados pela Covid-19 no Brasil /  
Micaías Mozzer de Oliveira. -- Maringá, PR, 2022.  
54 f.: il., figs., tabs.

Orientador: Prof. Dr. Brian Alvarez Ribeiro de Melo.  
Dissertação (Mestrado) - Universidade Estadual de Maringá, Centro de Ciências  
Exatas, Departamento de Estatística, Programa de Pós-Graduação em Bioestatística,  
2022.

1. Análise de sobrevivência. 2. Covid-19 - Pesquisa médica - Métodos estatísticos. 3.  
Tempo de sobrevida. I. Melo, Brian Alvarez Ribeiro de, orient. II. Universidade Estadual de  
Maringá. Centro de Ciências Exatas. Departamento de Estatística. Programa de Pós-  
Graduação em Bioestatística. III. Título.

CDD 23.ed. 519.5

Ademir Henrique dos Santos - CRB-9/1065

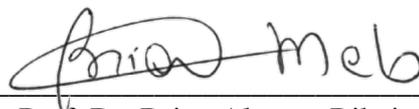
# MICAÍAS MOZZER DE OLIVEIRA

## Estudo do tempo de sobrevida de pacientes internados pela Covid-19 no

### Brasil

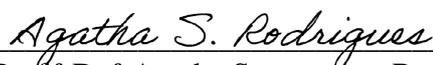
Dissertação apresentada ao Programa de Pós-Graduação em Bioestatística do Centro de Ciências Exatas da Universidade Estadual de Maringá, como requisito parcial para a obtenção do título de Mestre em Bioestatística.

#### BANCA EXAMINADORA



---

Prof. Dr. Brian Alvarez Ribeiro de Melo  
Universidade Estadual de Maringá – PBE/UEM



---

Prof.<sup>a</sup> Dr.<sup>a</sup> Agatha Sacramento Rodrigues  
Universidade Federal do Espírito Santo - DasLab/UFES



---

Prof. Dr. Willian Luis de Oliveira  
Universidade Estadual de Maringá – PBE/UEM

Maringá, 19 de maio de 2022.

# Resumo

Este estudo teve como objetivo avaliar o tempo entre a internação e o óbito dos pacientes acometidos pela COVID-19, e investigar quais fatores estão associados a esse tempo de sobrevida. Os dados foram obtidos por meio do Sistema de Informação de Vigilância Epidemiológica da Gripe, o qual contém todos os casos de internação por Síndrome Respiratória Aguda Grave no Brasil. Inicialmente, o tempo de sobrevida foi avaliado de forma descritiva por meio do estimador de Kaplan-Meier, que permite calcular o tempo mediano de sobrevida e a probabilidade de um paciente sobreviver a determinados períodos de internação. Em seguida, foi construído um modelo de regressão paramétrico para investigar a relação do tempo de sobrevida com diversos fatores de forma conjunta e os resultados interpretados em termos do tempo mediano de sobrevida. Entre os possíveis fatores associados, foram considerados as variáveis cadastrais (sexo, idade, região, raça e escolaridade), as comorbidades e os sintomas apresentados. Os resultados evidenciaram que pacientes idosos e com escolaridade baixa apresentam menor tempo de sobrevida e, conseqüentemente, maior risco de óbito. Com relação às comorbidades, as doenças neurológicas e hepáticas são os principais fatores de risco.

**Palavras-chave:** Análise de Sobrevivência. COVID-19. Tempo de sobrevida.

# Abstract

This study aimed to evaluate the time between hospitalization and death of patients affected by COVID-19, and to investigate which factors are associated with this survival time. Data were obtained through the Influenza Epidemiological Surveillance Information System, which contains all cases of hospitalization for Severe Acute Respiratory Syndrome in Brazil. Initially, survival time was descriptively evaluated using the Kaplan-Meier estimator, which allows the calculation of the median survival time and the probability of a patient surviving certain periods of hospitalization. Then, a parametric regression model was built to investigate the survival time relationship with several factors simultaneously and the results were interpreted in terms of median survival time. Among the possible associated factors, the registry variables (gender, age, region, race and education), comorbidities and symptoms were considered. The results showed that elderly patients with low education have a shorter survival time and, consequently, a higher risk of death. Regarding comorbidities, neurological and liver diseases are the main risk factors.

**Keywords:** Survival analysis. COVID-19. Survival time.

---

## Lista de ilustrações

---

|  |    |
|--|----|
| Figura 1 – Curva de sobrevivência estimada por Kaplan-Meier, dos pacientes em geral. . . . .   | 33 |
| Figura 2 – Gráficos quantil-quantil (direita) e resíduos vs ordem (esquerda) dos resíduos NRSP para o modelo Log-logístico múltiplo ajustado aos tempos de sobrevida dos pacientes acometidos pela Covid-19. . . . . | 43 |
| Figura 3 – Gráficos quantil-quantil dos resíduos NRSP dos modelos múltiplos, ajustados aos tempos de sobrevida dos pacientes acometidos pela Covid-19. . . . .   | 55 |
| Figura 4 – Gráficos Resíduos vs Ordem dos resíduos NRSP dos modelos múltiplos, ajustados aos tempos de sobrevida dos pacientes acometidos pela Covid-19. . . . .   | 56 |

---

## Lista de tabelas

---

|  |    |
|--|----|
| Tabela 1 – Frequências absoluta e relativa (%) de pacientes hospitalizados pela Covid-19, aos fatores associados. . . . .                  | 31 |
| Tabela 2 – Avaliação dos fatores associados ao tempo de sobrevida de pacientes internados pela Covid-19 por sócio-demográficos. . . . .    | 35 |
| Tabela 3 – Avaliação dos fatores associados ao tempo de sobrevida de pacientes internados pela Covid-19 por sintomas. . . . .              | 36 |
| Tabela 4 – Avaliação dos fatores associados ao tempo de sobrevida de pacientes internados pela Covid-19 por comorbidades. . . . .          | 37 |
| Tabela 5 – Estimação dos critérios AIC e BIC e logaritmo das funções de verossimilhança para os modelos selecionados. . . . .              | 38 |
| Tabela 6 – Resultados do modelo simples Log-Logístico de sobrevida dos pacientes internados pela Covid-19 aos fatores associados. . . . .  | 40 |
| Tabela 7 – Resultados do modelo múltiplo Log-Logístico de sobrevida dos pacientes internados pela Covid-19 aos fatores associados. . . . . | 42 |

---

# Sumário

---

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introdução</b>                                     | <b>10</b> |
| 1.1      | Exemplos de Análise de Sobrevivência na área da Saúde | 11        |
| 1.2      | Sequências do Trabalho                                | 12        |
| <b>2</b> | <b>Conceitos Básicos de Análise de Sobrevivência</b>  | <b>13</b> |
| 2.1      | Introdução  | 13        |
| 2.2      | Tempo de Falha  | 14        |
| 2.2.1    | O tempo inicial                                       | 14        |
| 2.2.2    | A escala de medida                                    | 14        |
| 2.2.3    | O evento de interesse                                 | 14        |
| 2.2.4    | Censura e Truncamento                                 | 14        |
| 2.3      | Funções Básicas de Sobrevivência                      | 16        |
| 2.3.1    | Função de Sobrevivência - $S(t)$                      | 16        |
| 2.3.2    | Função Taxa de Falha - $h(t)$                         | 16        |
| 2.3.3    | Função de Taxa de Falha Acumulada                     | 17        |
| 2.4      | Estimador de Kaplan-Meier                             | 18        |
| 2.5      | O Método da Máxima Verossimilhança                    | 19        |
| 2.5.1    | Função de verossimilhança na Análise de Sobrevivência | 20        |
| <b>3</b> | <b>Materiais e Métodos</b>                            | <b>22</b> |
| 3.1      | Banco de Dados  | 22        |
| 3.2      | Métodos estatísticos para análise de sobrevida        | 23        |
| 3.3      | Modelos de Regressão Paramétricos                     | 24        |
| 3.3.1    | Distribuição Weibull                                  | 25        |
| 3.3.2    | Distribuição Log-Logística                            | 26        |
| 3.4      | Seleção de Variáveis                                  | 27        |
| 3.5      | Adequação do Modelo Ajustado                          | 28        |
| 3.5.1    | Resíduos NRSP   | 28        |
| <b>4</b> | <b>Resultados</b>                                     | <b>30</b> |
| 4.1      | Análise Descritiva                                    | 30        |
| 4.2      | Ajuste dos Modelos de Regressão Paramétricos          | 38        |
| 4.3      | Interpretação dos Resultados                          | 38        |

|          |                             |           |
|----------|-----------------------------|-----------|
| 4.3.1    | Estimativas Brutas          | 38        |
| 4.3.2    | Modelo Múltiplo             | 41        |
| 4.4      | Análise de Resíduos         | 43        |
| <b>5</b> | <b>Discussão</b>            | <b>44</b> |
| <b>6</b> | <b>Considerações Finais</b> | <b>46</b> |
|          | <b>Referências</b>          | <b>47</b> |
|          | <b>Anexos</b>               | <b>53</b> |
|          | <b>ANEXO A ANEXO</b>        | <b>55</b> |

---

## Capítulo 1

---

# Introdução

---

Em dezembro de 2019, surgiu na província de Hubei, em Wuhan, China (HUANG *et al.*, 2020; LI *et al.*, 2020) um surto de uma doença ocasionada por um novo coronavírus e caracterizada por provocar pneumonia. O novo coronavírus (COVID-19) tornou-se um dos grandes desafios no século XXI e ocasiona diretamente e/ou indiretamente a saúde e a economia da população mundial.

No início do surto, os primeiros casos notificados estavam relacionados a um mercado de frutos do mar e animais vivos, em Wuhan (BOGOCH *et al.*, 2020). No Brasil, a primeira notificação da COVID-19 ocorreu no estado de São Paulo, em fevereiro de 2020.

As pessoas infectadas pelo vírus podem apresentar tosse, dores de garganta, dificuldades em respirar, febre e outras manifestações clínicas. Contudo, existem pessoas assintomáticas, as quais possuem uma importância epidemiológica, pois tem um potencial transmissor, segundo Surveillances (2020). Dentre os pacientes infectados pela COVID-19, muitos necessitam de internação em enfermaria ou UTI.

Em geral, pacientes que passam por longos períodos de internação acarretam problemas relacionados à mobilidade, levando à atrofia muscular, o que causa lesões nos nervos periféricos e trombozes, segundo a Dra. Fernanda Netto (UNIFOR – Fortaleza, Ceará, Brasil). A doutora ainda explica que pode ocorrer fraqueza na musculatura respiratória, levando a outras infecções. A presença de doenças relacionadas ao coração também afeta a função dos músculos respiratórios, em que pacientes podem apresentar fraqueza e fadiga da musculatura respiratória (AZEREDO; MACHADO, 1999). De acordo com Needham (2008), os pacientes sobreviventes de doença severa crítica tem significativa alteração neuromuscular e complicações que são prejudiciais ao desenvolvimento físico, funcional, bem como a qualidade de vida após a alta hospitalar. Por causa da atrofia e fraqueza muscular, os

pacientes perdem diversos movimentos, devido à dificuldade em se alimentar, em andar, em falar. Nesse cenário, o uso da fisioterapia, e também da reabilitação pós alta, ter uma alimentação adequada irá ajudar na recuperação do paciente, mesmo sendo um processo demorado (TV, 2021).

Em vista do exposto, o presente trabalho busca avaliar o tempo de sobrevida dos pacientes internados pelo novo coronavírus, COVID-19, no Brasil em 2020. Mais especificamente, temos como objetivo principal modelar o tempo de sobrevida através de um modelo de regressão paramétrico para avaliar diversos fatores associados a este tempo. Para isso, utiliza-se dados disponibilizados pelo Sistema de Informação da Vigilância Epidemiológica da Gripe (SIVEP-Gripe), que contém todos os casos de internação por Síndrome Respiratória Aguda Grave no Brasil.

## 1.1 Exemplos de Análise de Sobrevivência na área da Saúde

A análise de sobrevivência está fortemente ligada a estudos clínicos. Em McArthur *et al.* (1993), por exemplo, um estudo ao qual utilizou dados do Multicenter AIDS Cohort Study em Baltimore - Los Angeles, mostra que dos 492 pacientes homossexuais com AIDS, 64 desenvolveram demência. Essa mesma pesquisa, revelou que 15% deste estudo desenvolveu demência até o óbito, com tempo mediano de 6 meses.

A pesquisa de Viana *et al.* (1994), envolvendo crianças com leucemia, desenvolvido pelo Grupo Cooperativo para tratamento de Leucemias Agudas, teve como objetivo entender melhor os tipos de fatores que causam impacto no tempo de sobrevida de uma criança brasileira com Leucemia Linfoblástica Aguda (LLA); o estudo é composto por 128 crianças (< 15 anos) durante 1988 a 1992, em alguns hospitais de Belo Horizonte. O tempo de falha foi o tempo a partir da remissão até a recidiva ou óbito. Foram registrados que das 128 crianças, apenas 120 entraram em remissão na qual formou o banco de dados.

Em mesmo cenário da dissertação, os autores Sousa *et al.* (2020), avaliaram o tempo até entre o início dos sintomas e o óbito pela COVID-19 no estado de Ceará, Brasil, desde a primeira notificação do vírus até 14 de abril de 2020 e registraram que pessoas com comorbidades (cardiovascular, neurológica, pneumopatia) apresentaram maior risco de morte e menor probabilidade de sobrevivência. No estudo, foram analisados 2070 casos de pacientes com o vírus e o tempo de falha é a morte da COVID-19. Dos analisados, apenas 24% são idosos acima de 60 anos com tempo mediano de 44 anos, em que 19,9% dos casos de óbitos foram os idosos e 5,4% dos pacientes foram internados em UTI, com uma taxa de mortalidade de 18,9%. Também, em Muse *et al.* (2021) foi realizado uma pesquisa para determinar os fatores que afetam o tempo dos pacientes infectados pelo vírus, em que o

modelo Log-logístico se ajustou melhor aos dados do que os demais modelos paramétricos usados. Os resultados mostraram que a idade apresentou um desempenho significativo no tempo de sobrevivência desses pacientes.

De acordo com Hueda-Zavaleta *et al.* (2021), em sua pesquisa de quase 6 meses envolvendo mortalidade pela COVID-19 na cidade de Tacna, Peru, foi registrado uma predominância em pacientes do sexo masculino (74,10%) sendo 8 dias o tempo de internação pelo vírus. Como a internação pode provocar aumento na taxa de mortalidade, o mesmo estudo afirmou que 36,2% dos pacientes internados tiveram óbito durante o acompanhamento.

Um estudo no norte do México, com objetivo de descrever as características dos pacientes internados com COVID-19 e determinar as comorbidades associadas a mortalidade mostrou que no período de 5 meses, de 17.479 pacientes apenas 6,3% levaram ao óbito. O que mais significou esse impacto de mortalidade foram as comorbidades diabetes mel-litus, obesidade, pressão arterial e doença renal crônica; o fator que mais contribuiu para o risco de morte foram os pacientes idosos acima de 60 anos. Para mais detalhes, ver Salinas-Aguirre *et al.* (2021).

Vale ressaltar que, indiferente das regiões geográficas, a presença das comorbidades e dos sintomas em pacientes com COVID-19 se torna o maior fator de risco de mortalidade, com isso, esses fatores aumentam consideravelmente o risco em pacientes com mais de 60 anos de idade.

## 1.2 Sequências do Trabalho

A estruturação do trabalho dá-se da seguinte maneira: o Capítulo 2 apresenta uma breve descrição da fundamentação teórica e revisão dos principais conceitos de Análise de Sobrevivência. O Capítulo 3 descreve os dados utilizados na análise e disserta sobre o uso dos modelos de regressão para dados censurados. O banco de dados é construído através de uma ficha de registro individual, onde são consideradas as variáveis cadastrais, as comorbidades e os sintomas apresentados, além de outros dados clínicos epidemiológicos e de atendimento. O Capítulo 4 exhibe os resultados das análises da proporção de óbitos e do tempo de sobrevida em função dos fatores associados de forma simples e múltipla, através de análises descritivas e de um modelo de regressão paramétrico para o tempo, respectivamente. No Capítulo 5 é realizada uma discussão sobre os resultados encontrados e no Capítulo 6 são apresentadas as conclusões deste estudo.

---

## Capítulo 2

---

# Conceitos Básicos de Análise de Sobrevivência

---

## 2.1 Introdução

A análise de sobrevivência (AS) ou análise de sobrevida é um conjunto de processos estatísticos que vem sendo muito utilizados desde as últimas décadas do século passado, como por exemplo, em pesquisas biomédicas quando o tempo entre exposição e evento é de interesse clínico. Na AS estuda-se o tempo de duração até a ocorrência de um ou mais eventos de interesse, o qual denomina-se tempo de falha. O termo análise de sobrevivência é utilizado basicamente em situações médicas envolvendo dados censurados. Para o ramo das engenharias, esse tipo de estudo é comum em equipamentos ou produtos submetidos em testes para estimar seu tempo de vida. Os engenheiros denominam esta área de análise de confiabilidade. Alguns exemplos são encontrados em Nelson (1990), Meeker, Escobar e Lu (1998).

Na área da saúde, geralmente, a AS é aplicada em estudos nos quais o desfecho é o tempo até o óbito do paciente, ou até a cura ou recidiva de uma doença. Alguns exemplos podem ser encontrados em: Sue-Ling *et al.* (1993) que aborda o uso de esforços mais vigorosos à ressecção total em pacientes com câncer gástrico; Bustamante-Teixeira, Faerstein e Latorre (2002), que aborda esses conceitos na área da oncologia e Trevisol *et al.* (2012), que realiza um estudo de sobrevida em pacientes submetidos à angioplastia coronariana com stent.

Uma das principais características da AS é a presença de censuras, isto é, uma observação parcial da resposta. Isso ocorre em situações em que, por ocorrência da doença, o

acompanhamento do paciente foi interrompido, o mesmo mudou-se de cidade, o estudo foi finalizado ou o paciente veio a óbito devido a outras causas diferente da estudada.

## 2.2 Tempo de Falha

A variável resposta ou tempo de falha é constituída por três elementos: o tempo inicial, a escala de medida e o evento de interesse (falha). Estes elementos serão definidos a seguir.

### 2.2.1 O tempo inicial

Os pacientes devem ser comparados no tempo que ingressaram ao estudo. De fato, o tempo de início deve ser definido precisamente. Como por exemplo, neste trabalho, o tempo de início será a data do diagnóstico da doença.

### 2.2.2 A escala de medida

A escala de medida se concentra em definir se o tempo será medido em horas, dias, semanas, etc. É importante utilizar a mesma escala de tempo para todos os indivíduos na pesquisa. Em algumas áreas, como engenharia, é comumente usado outros tipos de escalas de medidas, como o número de ciclos, a quilometragem de um veículo ou qualquer outra medida de carga. Neste trabalho, a escala de medida é definida em dias.

### 2.2.3 O evento de interesse

O evento de interesse deve ser definido com precisão antes da realização da análise dos dados. Em situações médicas, a falha pode ser a morte ou a recidiva de uma doença específica.

Neste estudo será considerado apenas uma única falha, a saber, o óbito por Covid-19. Em outras situações, podem ocorrer duas ou mais falhas, as denominadas riscos competitivos (PRENTICE *et al.*, 1978).

### 2.2.4 Censura e Truncamento

Suponhamos que o evento de interesse numa pesquisa seja a morte por doença cardíaca e que, pode acontecer que o evento não tenha ocorrido devido a outra causa, por exemplo, morte de câncer pulmonar. Casos assim, são chamadas de censuras ou observações incompletas ou parciais, onde o evento de interesse não é aquele determinado até

o término do experimento. Em outras palavras, dizemos que uma variável aleatória (v.a.) é censurada quando não conseguimos observar seu valor exato.

A censura pode acontecer em três motivos:

- O estudo chegar ao fim sem o registro do evento de interesse;
- Perda de seguimento, isto é, pacientes podem não estar mais disponíveis na pesquisa até o final de um período;
- O indivíduo é removido do estudo por alguma outra causa, relacionado com o tempo de vida.

Quando um tempo de falha é censurado, a informação obtida sobre este tempo depende do tipo de censura. Assim, considera-se:

- Censura à direita: ocorre quando o evento de interesse não é observado até o término do estudo, isto é, o tempo de falha será superior ao tempo que é registrado. É o tipo de censura mais comum;
- Censura à esquerda: ocorre quando temos a incerteza do momento da ocorrência de falha, mas sabemos que tenha ocorrido antes do tempo observado, ou seja, o tempo de registro será maior que o tempo de falha;
- Censura intervalar: em geral, ocorre em estudos nos quais os indivíduos são acompanhados por visitas periódicas, em que, o evento de interesse é registrado em um determinado intervalo de tempo.

Além dessa classificação, a causa da censura pode estar ligada ao planejamento do experimento que visa estudar o tempo de falha. Nesse sentido, a censura pode ser:

- Tipo I: Estudo terminado após um período pré-determinado de tempo e o evento de interesse não é encontrado no indivíduo ao final do estudo. Para os indivíduos que não falharam, o tempo registrado será o tempo do término do estudo;
- Tipo II: Estudo terminado após ocorrer o evento de interesse em uma quantidade pré-determinada de indivíduos. Para os indivíduos que não falharam o tempo (censurado) registrado será o mesmo da última falha;
- Tipo aleatório: ocorre quando uma observação é removida no decorrer do estudo sem ter ocorrido o evento de interesse ou o indivíduo é removido por outra causa desconhecida.

Formalmente, o tempo de falha e a indicação de censura são construídas da seguinte forma: suponha duas variáveis aleatórias  $T$  e  $C$ , representando os tempos de falha e censura, respectivamente. Em geral assumimos que a variável aleatória  $C$  é independente de  $T$  e o tempo observado é o mínimo destas duas variáveis aleatórias, isto é,  $t = \min(T, C)$ . Além disso, observamos a variável  $\delta$ , que indica se o tempo observado  $t$  é um tempo de falha ou censura. Os dados observados para o  $i$ -ésimo indivíduo são representadas pelo par  $(t_i, \delta_i)$ ,  $i = 1, \dots, n$ . Assim:

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

Segundo Chaieb, Rivest e Abdous (2006) a censura e o truncamento são diferentes. Os indivíduos censurados estão inclusos na pesquisa, mas seus tempos correspondentes de falha não estão presentes. Já os indivíduos truncados, são ditos perdidos durante a pesquisa, ou seja, ocorre a exclusão dos mesmos no estudo.

## 2.3 Funções Básicas de Sobrevida

Na AS o tempo de falha é considerado uma variável aleatória positiva e existem algumas funções específicas para auxiliar no estudo deste tempo. Estas funções são comentadas abaixo.

### 2.3.1 Função de Sobrevida - $S(t)$

A função de sobrevida corresponde a probabilidade de que o indivíduo não tenha apresentado falha até o tempo pré-estabelecido  $t$ , ou seja  $S(t) = P(T > t)$ ,  $t > 0$ . Em contraste, a função de distribuição acumulada  $F(t)$  de uma variável aleatória é definida como a probabilidade do evento ocorrer em um indivíduo até o tempo  $t$ , isto é,  $F(t) = P(T \leq t)$ . Escrevendo  $S(t)$  em função de  $F(t)$ , temos a seguinte relação:

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

Na ausência de censura, um estimador não-paramétrico para a função  $S(t)$  é dada por

$$\hat{S}(t) = \frac{\text{Nº de observações sob risco no tempo } t}{\text{Nº total de observações}}.$$

### 2.3.2 Função Taxa de Falha - $h(t)$

A função de risco, conhecida também por função intensidade ou taxa de falha, é caracterizada como a taxa instantânea de morte de um indivíduo que tenha sobrevivido até o

tempo  $t$ , dada por:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P[(t \leq T \leq t + \Delta t) | T \geq t]}{\Delta t} = \frac{-S'(t)}{S(t)} = \frac{f(t)}{S(t)}, \quad t \geq 0.$$

A função de taxa de falha é mais informativa do que a função de sobrevivência. Desta maneira, a modelagem da função taxa de falha é um importante método para dados de sobrevivência e sua modelagem é importante para AS, pois pode apresentar forma crescente, decrescente, constante ou não-monótona.

De acordo com Carvalho *et al.* (2011),  $h(t)$  é uma taxa que não representa uma probabilidade, podendo assim assumir valores positivos, sem qualquer restrição. Sua dimensão é dada como o inverso do tempo,  $h(t) = \text{tempo}^{-1}$ .

Existe uma relação entre a função de probabilidade e a função de sobrevivência, dada por:

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} \log S(t). \quad (2.3.1)$$

### 2.3.3 Função de Taxa de Falha Acumulada

Outra função interessante na análise de sobrevivência é a função de taxa de falha acumulada, com capacidade de mostrar a taxa de falha acumulada dos indivíduos sob estudo. A função pode ser definida por:

$$H(t) = \int_0^t h(u) du. \quad (2.3.2)$$

Substituindo a Equação 2.3.1 em 2.3.2, temos:

$$\begin{aligned} H(t) &= \int_0^t \frac{f(u)}{1 - F(u)} \\ &= -\log [1 - F(u)] \Big|_0^t \\ &= -\log [1 - F(t)] + \log [1 - F(0)] \\ &= -\log [1 - F(t)] \\ &= -\log [S(t)]. \end{aligned}$$

Assim, a função de taxa acumulada e sua função de sobrevivência são, respectivamente:

$$\begin{aligned} H(t) &= -\log [S(t)] \\ S(t) &= e^{-H(t)}. \end{aligned}$$

A característica desta função é não ter uma interpretação direta na evolução de risco de morte ao longo do tempo, porém, é bastante útil na avaliação da função taxa de falha,  $h(t)$ . O motivo disto é na estimação não-paramétrica em que  $H(t)$  possui um estimador com ótimas propriedades e  $h(t)$  tende a ser difícil de ser estimada (NELSON, 1972).

## 2.4 Estimador de Kaplan-Meier

O estimador de Kaplan-Meier (K-M), proposto em 1958 por Kaplan e Meier, ou ainda o estimador de produto-limite, é utilizado com muita frequência pois é um dos estimadores empíricos nas análises de confiabilidade. Apresenta a principal solução para estimar a função de sobrevivência  $S(t)$  e taxa de falha acumulada  $H(t)$ .

Pelo método de K-M (KAPLAN; MEIER, 1958; KLEINBAUM, 2005), os intervalos de tempo não são fixos na análise, mas são definidos após o aparecimento de falha (por exemplo, o óbito). Na mesma situação, o número de óbitos em cada intervalo de tempo deve ser um. Segundo Colton (1979), é um método empírico não-paramétrico, ou seja, é independente da distribuição de probabilidade.

Para determinarmos o estimador, suponha que existem  $n$  pacientes no estudo e  $k \leq n$  falhas distintas e ordenadas nos tempos  $(t_1 \leq t_2 \leq \dots \leq t_k)$ . Os indivíduos que sobreviveram ao tempo pré-determinado  $t$  são ajustados pela censura, isto é, são alocados no cálculo da função de probabilidade de sobrevivência  $S(t)$  até o momento de serem considerados como perda. Diante dessas informações, a expressão geral do estimador de K-M para a função de sobrevivência no tempo  $t$ , é:

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

onde  $n_j$  é o número de indivíduos sob risco em  $t_j$ ,  $j = 1, \dots, k$  e  $d_j$  é o número de falhas em  $t_j$ .

O estimador de K-M possui algumas propriedades importantes, tais como, ser o estimador de máxima verossimilhança de  $S(t)$  e assintoticamente convergir em um processo gaussiano. Mais detalhes sobre as propriedades teóricas desse estimador são apresentadas em Kaplan e Meier (1958).

Para trabalharmos com o estimador K-M, é útil saber calcular a variância assintótica estimada de  $\hat{S}(t)$ , denotada por  $\widehat{Var}[\hat{S}(t)]$ . Essa variância pode ser escrita como:

$$\widehat{Var}[\hat{S}(t)] = [\hat{S}(t)]^2 \sum_{j:t_j < t} \frac{d_j}{n_j(n_j - d_j)}.$$

Com a variância estimada, podemos determinar os intervalos de confiança (IC) para  $S(t)$ . Assim, sabendo que para um  $t$  fixo,  $\hat{S}(t)$  segue, assintoticamente, distribuição Normal e considerando um nível de confiança  $(1 - \alpha)100\%$ , o intervalo aproximado de confiança é dado por:

$$IC [S(t); 1 - \alpha] = \hat{S}(t) \pm z_{\alpha/2} \sqrt{\widehat{Var}(\hat{S}(t))}.$$

O IC definido acima pode apresentar valores negativos ou maiores que 1, o que não faz sentido uma vez que  $S(t)$  é uma probabilidade. Desse modo, é necessário o uso de uma transformação em  $S(t)$  (COLOSIMO; GIOLO, 2006). Seja  $\hat{U}(t)$  o estimador transformador, temos:

$$\begin{aligned} 0 \leq \hat{S}(t) \leq 1 &\Leftrightarrow \\ -\infty \leq \log[\hat{S}(t)] \leq 0 &\Leftrightarrow \\ 0 \leq -\log[\hat{S}(t)] \leq \infty &\Leftrightarrow \\ -\infty \leq \log[-\log[\hat{S}(t)]] \leq \infty, & \end{aligned}$$

assim,  $\hat{U}(t) = \log[-\log[\hat{S}(t)]]$ .

A variância assintótica estimada de  $\hat{U}(t)$  é

$$\widehat{Var}(\hat{U}(t)) = \sum_{j:t_j < t} \frac{d_j}{n_j(n_j - d_j)} \left[ \sum_{j:t_j < t} \log\left(1 - \frac{d_j}{n_j}\right) \right]^{-2}.$$

Assim, com nível de  $100(1 - \alpha)\%$  de confiança, o intervalo aproximado para  $S(t)$  que possa assumir valores entre 0 e 1 é dado por:

$$IC [S(t); 1 - \alpha] = [\hat{S}(t)]^{\exp\left\{\pm z_{\alpha/2} \sqrt{\widehat{Var}(\hat{U}(t))}\right\}}.$$

## 2.5 O Método da Máxima Verossimilhança

Em 1922, o estatístico e geneticista Ronald Fisher (FISHER, 1925; FISHER, 1956), determinou um método de suma importância para realizar a estimação dos parâmetros, denominado por máxima verossimilhança. Segundo ele, este método é ideal para encontrar valores dos parâmetros em dados suficientemente grandes. De modo geral, os valores dos parâmetros são encontrados de forma a maximizar a probabilidade descrita pelo modelo estudado, na qual produz os dados que foram observados.

Dado uma amostra aleatória  $(t_1, t_2, \dots, t_n)$  de uma população com função de probabilidade  $f(t, \theta)$  de parâmetro genérico  $\theta$ , a função de verossimilhança é expressa por

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

Em muitas situações, o problema central da estimação é encontrar um vetor  $\hat{\theta}$  que possa maximizar a probabilidade de uma amostra específica  $(t_1, t_2, \dots, t_n)$ . Em outras palavras, a estimação por máxima verossimilhança se dá por meio de encontrar o vetor  $\hat{\theta}$  que maximize a função de verossimilhança. Neste caso,  $\hat{\theta}$  é chamado de Estimador de Máxima Verossimilhança (EMV) de  $\theta$ .

Na maioria dos casos, é ideal trabalhar com o logaritmo natural da função de verossimilhança,  $\log(L(\theta))$ , pois isso facilita nos cálculos e geram os mesmos valores da maximização desta função original. Nesse caso, o EMV é a solução do seguinte sistema de equação:

$$U(\theta) = \frac{\partial}{\partial \theta} \log [L(\theta)] = 0,$$

sendo  $U(\theta)$  o vetor escore.

### 2.5.1 Função de verossimilhança na Análise de Sobrevida

Para a construção da função de verossimilhança, consideramos um modelo paramétrico indexado por algum vetor de parâmetros  $\theta$ , isto é,  $T \sim F(\theta)$ . Similarmente, assumimos que  $C \sim G(\lambda)$ . Considerando a variável aleatória  $Y = \min(T, C)$  como a observação do tempo de vida de um indivíduo e utilizando a mesma ideia da definição de censura, quando há ocorrência da morte de um indivíduo ( $\delta = 1$ ) no experimento em um intervalo suficientemente pequeno,  $[t, t + dt)$  então, podemos admitir entre as v.a.'s  $T$  e  $C$  uma independência,

$$P[t \leq Y < t + dt, \delta = 1] = P[t \leq T < t + dt] P[C \geq t + dt]$$

e quando o tempo é censurado ( $\delta = 0$ ), tem-se:

$$P[t \leq Y < t + dt, \delta = 0] = P[t \leq C < t + dt] P[T \geq t + dt].$$

Caso  $T$  e  $C$  sejam variáveis aleatórias contínuas, com funções densidades  $f(t, \theta)$  e  $g(t, \lambda)$  e suas funções de sobrevivências  $S(t, \theta)$  e  $1 - G(t, \lambda)$ , dividindo-se por  $dt$  nas duas igualdades acima, com  $dt$  tendendo a  $0^+$  temos

$$L(\theta, \lambda) = \{f(t; \theta) [1 - G(t; \lambda)]\}^\delta \{g(t; \lambda) S(t; \theta)\}^{1-\delta}.$$

Selecionando aleatoriamente uma amostra de dimensão  $n$ ,  $((t_1, \delta_1), (t_2, \delta_2), \dots, (t_n, \delta_n))$ , temos a seguinte função de verossimilhança  $L(\boldsymbol{\theta}, \boldsymbol{\lambda})$ :

$$L(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \left\{ \prod_{i=1}^n [f(t_i; \boldsymbol{\theta})]^{\delta_i} [S(t_i; \boldsymbol{\theta})]^{1-\delta_i} \right\} \left\{ \prod_{i=1}^n [g(t_i; \boldsymbol{\lambda})]^{1-\delta_i} [1 - G(t_i; \boldsymbol{\lambda})]^{\delta_i} \right\}.$$

A contribuição das observações não-censuradas para a função de verossimilhança  $L(\boldsymbol{\theta})$  é a própria função de densidade  $f(t)$ , enquanto para as observações censuradas é a função de sobrevivência  $S(t)$ . Se a distribuição dos tempos de censura não depender dos parâmetros de interesse  $(\boldsymbol{\theta})$ , a verossimilhança é constante em relação às funções associadas à distribuição do tempo de censura, e podemos expressá-la da seguinte forma:

$$L(\boldsymbol{\theta}) \propto \prod_{i=1}^n [f(t_i; \boldsymbol{\theta})]^{\delta_i} [S(t_i; \boldsymbol{\theta})]^{1-\delta_i},$$

ou equivalentemente,

$$L(\boldsymbol{\theta}) \propto \prod_{i=1}^n [h(t_i; \boldsymbol{\theta})]^{\delta_i} S(t_i; \boldsymbol{\theta}).$$

---

## Capítulo 3

---

# Materiais e Métodos

---

Para execução das análises nesta dissertação, serão utilizados dados de pacientes internados por Síndrome Respiratória Aguda Grave (SRAG) organizado pelo Ministério da Saúde (MS).

### 3.1 Banco de Dados

Neste trabalho, apresenta-se um estudo transversal utilizando dados disponíveis no Sistema de Informação da Vigilância Epidemiológica da Gripe, SIVEP-Gripe (SIVEP-GRIFE, 2021), um sistema oficial desde 2009 que atribui os registros de casos e óbitos por Síndrome Respiratória Aguda Grave (SRAG) no Brasil, sendo atualizado diariamente sobre novos casos notificados no país (<<https://opendatasus.saude.gov.br/dataset/bd-srag-2021>>).

Em 2020, a infecção humana causada pelo novo Coronavírus - COVID-19, começou a se alastrar exponencialmente em todo o Brasil e, por conta disto, foi incorporada na rede das Influenzas e outros tipos de vírus respiratórios. Nesse estudo, consideramos os casos notificados durante o período de 01 de janeiro de 2020 até 31 de dezembro de 2020, sendo que a última atualização da base foi em 18 de janeiro de 2021, totalizando 480788 pacientes.

O banco de dados da Sivep-Gripe disponibiliza uma ficha de registro individual informando inúmeras variáveis explicativas. Algumas covariáveis foram consideradas no estudo, tais como: sócio-demográficos, por exemplo, a idade, o sexo, a raça/cor e escolaridade; as comorbidades (fatores de risco), por exemplo, puérpera, asma, diabete, obesidade, cardiovascular, pneumatopatia, outros tipos de comorbidades e entre outros; também foram considerados os sintomas, por exemplo, febre, tosse, dor de garganta, saturação

baixa ( $< 95\%$ ), vômito, fadiga, outros tipos de sintomas e entre outros.

A variável resposta será o tempo de internação calculado como o número de dias entre a data de entrada/internação e a data de saída, levando ao óbito ou alta (censura) de cada paciente. Foram selecionados aqueles pacientes que chegaram a ser internados, acima de 17 anos de idade com a seguinte codificação dos casos: óbito (1) ou censura (0). Vale frisar que aqueles pacientes que tiveram alta, ou alguma informação ignorada ou óbito por outras causas são considerados como censura.

A realização da pesquisa foi dispensada previamente à aprovação do Comitê de Ética de Pesquisa, pois a base de dados é de caráter público. De acordo com a Resolução 466/12 do Conselho de Pesquisa Brasileira, os pesquisadores relataram o total compromisso ético no manuseio, análise e na publicação dos dados visando que a pesquisa não necessite de aprovação.

## 3.2 Métodos estatísticos para análise de sobrevida

Para a análise do banco de dados descrito, foram utilizadas metodologias da literatura estatística no contexto de análise de sobrevivência, na qual a variável resposta de interesse é o tempo de internação até o óbito. Para esta análise estatística, utilizou-se o *Software R* (R Core Team, 2020).

Na realização dos estudos descritivos, é usado o estimador não-paramétrico Kaplan-Meier, descrito anteriormente, para a obtenção de medidas como: o tempo médio, o tempo mediano e os percentis de interesse, bem como avaliar quais serão as covariáveis significativas a serem inseridas nos modelos de regressão. Também foi utilizado o teste Log-rank (MANTEL *et al.*, 1966), que é a diferença entre o número observado de falhas em cada grupo e o correspondente número esperado de falhas sob a hipótese nula, para comparar as curvas de Kaplan-Meier de dois grupos e utilizando seu p-valor para determinar se essas curvas podem ou não ser consideradas diferentes com significância estatística.

Para a escolha dessas covariáveis, que serão associadas ao tempo que o paciente teve óbito, serão ajustados alguns modelos de regressão de sobrevivência tais como: modelo exponencial (COLOSIMO; GIOLO, 2006), modelo Weibull (COLOSIMO; GIOLO, 2006), modelo log-normal (COLOSIMO; GIOLO, 2006), modelo gama (BROWN; FLOOD, 1947), modelo log-logístico (BRÜEDERL; DIEKMANN, 1995), gama generalizada (STACY; MIHRAM, 1965) ou gompertz (POLLARD; VALKOVICS, 1992). Em todas as análises, foi adotado um nível de significância de 5%.

### 3.3 Modelos de Regressão Paramétricos

Na AS é muito comum usar as distribuições paramétricas em tempo de sobrevida. Em muitas situações, temos interesse em avaliar a relação do tempo de sobrevida com diversos fatores associados. Essa avaliação é realizada através da construção de modelos de regressão. Para isso, observamos além do tempo de sobrevida e da variável indicadora de censura, um vetor de covariáveis,  $\mathbf{x}$ , para cada indivíduo. Formalmente, a amostra é composta por  $(t_1, \delta_1, \mathbf{x}_1), \dots, (t_n, \delta_n, \mathbf{x}_n)$ .

A inclusão destas covariáveis é feita em um dos parâmetros da distribuição, por meio de um preditor linear e uma função de ligação. Esta função relaciona um componente aleatório ao componente sistemático, isto é, relaciona um parâmetro da distribuição de probabilidades do tempo de falha ao preditor linear,  $\eta$ .

Considere que o tempo de falha,  $T_i$ , seja uma v.a não negativa seguindo uma distribuição de probabilidade indexada por um parâmetro  $\theta_i$ , isto é,  $T_i \sim F(\theta_i)$ ,  $i = 1, 2, \dots, n$ . Então, podemos construir o modelo de regressão associando o parâmetro  $\theta_i$  às covariáveis por meio da expressão:

$$g(\theta_i) = \eta_i = \mathbf{x}_i^\top \boldsymbol{\beta} = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip},$$

onde

- $g(\cdot)$  é chamada de função de ligação;
- $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^\top$  é o vetor de parâmetros de dimensão  $p + 1$ ;
- $\mathbf{x}_i = (1, x_{i1}, x_{i2}, \dots, x_{ip})^\top$  é o vetor de dimensão  $p + 1$  já especificada, com  $i = 1, 2, \dots, n$ .

Mais informações a respeito da estruturação e exemplos podem ser encontradas em alguns artigos: Nelder e Wedderburn (1972), Williams (1987) e Lindsey (1997).

A interpretação dos coeficientes de regressão, é realizada em função da razão de risco e/ou razão dos tempos medianos de sobrevida, quando possível.

Considere, por exemplo, que os pacientes de um estudo são aleatorizados para receber um tratamento padrão ou um novo tratamento. Sejam  $h_1(t)$  e  $h_0(t)$  suas funções de risco num tempo  $t$  para os pacientes durante o tratamento novo e tratamento padrão, respectivamente. Tomando a razão de  $h_1(t)$  por  $h_0(t)$  e se o resultado desta razão de riscos não for associado ao tempo  $t$ , dizemos que a razão de risco é proporcional ao longo do tempo. Em outras palavras, um modelo é de riscos proporcionais, quando a razão das taxas de falhas entre dois pacientes diferentes for constante no tempo  $t$ .

Diversos modelos probabilísticos podem ser considerados para estudar o tempo de sobrevida  $T_i$ . Em seguida, a fim de exemplificar a construção e interpretação destes modelos, apresentamos os modelos de regressão Weibull e Log-logístico, sendo o primeiro um modelo de riscos proporcionais e o segundo não.

### 3.3.1 Distribuição Weibull

Proposto por Fisher e Tippet em 1928, a distribuição de probabilidade Weibull foi desenvolvida pelo físico sueco Waloddi Weibull, no ano de 1939 no estudo de resistência de materiais. Na ciência florestal, seu uso se tornou mais frequente na caracterização da estrutura diamétrica de florestas equiâneas e inequiâneas (CLUTTER *et al.*, 1983; BATISTA, 1989). Entretanto, a distribuição Weibull é flexível pois tem a capacidade de descrever curvas com diferentes formas sendo, uma principal vantagem e características em comparação as demais funções de probabilidade.

Na literatura estatística da análise de sobrevida, a função de taxa de falha é monótona, podendo ser decrescente, crescente ou constante.

Considere  $T$  uma variável aleatória seguindo distribuição Weibull  $(\alpha, \gamma)$ . Neste caso, sua função densidade de probabilidade (fdp), função de sobrevivência e função de risco são, respectivamente:

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

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

e

$$h(t) = \frac{f(t)}{S(t)} = \left(\frac{\gamma}{\alpha^\gamma}\right) t^{\gamma-1},$$

em que  $\gamma$  e  $\alpha$  são parâmetros de forma e escala, respectivamente. Também é importante dizer que  $t, \alpha$  e  $\gamma > 0$ .

A inserção de covariáveis no modelo de regressão Weibull pode ser realizada da seguinte forma:

$$\log(\alpha_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}.$$

Para a interpretação dos coeficientes de regressão, considere a existência de uma única covariável binária  $x_1 = (0, 1)$ , indicando a ausência ou presença de um fator associado ao tempo de sobrevida. Assim, o modelo pode ser definido por:

$$\log(\alpha_i) = \beta_0 + \beta_1 x_{i1} \rightarrow \log(\alpha_i) = \begin{cases} \beta_0 + \beta_1 & , x_{i1} = 1 \\ \beta_0 & , x_{i1} = 0 \end{cases}.$$

A razão de riscos dos pacientes que apresentam o fator associado em relação aos que não apresentam é:

$$\frac{h(t|x_1 = 1)}{h(t|x_1 = 0)} = \frac{\gamma t^{\gamma-1} (e^{\beta_0})^\gamma}{(e^{\beta_0+\beta_1})^\gamma \gamma t^{\gamma-1}} = (e^{-\beta_1})^\gamma.$$

Como a razão de riscos não depende do tempo  $t$ , temos que o modelo de regressão Weibull, com as covariáveis inseridas no parâmetro de escala é um modelo de riscos proporcionais.

Se  $x_1$  é uma covariável binária, então a taxa/risco de falha do grupo  $\{x_1 = 1\}$  é  $(e^{-\beta_1})^\gamma$  vezes o risco do grupo  $\{x_1 = 0\}$ .

### 3.3.2 Distribuição Log-Logística

A distribuição Log-Logística (LL) é um modelo de probabilidade utilizada em diversas áreas como confiabilidade, ciência atuarial, economia, hidrologia, engenharia e análise de sobrevivência. Especificamente, muitos trabalhos realizados em áreas clínicas/epidemiológicas utiliza bastante esta distribuição. Alguns desses trabalhos podem ser vistos em Tan e Jr (1993), Faradmal *et al.* (2014) e Hung *et al.* (2014). O matemático francês Verhulst (1838) foi o pioneiro a utilizar a distribuição log-logístico para modelar o crescimento de populações biológicas, como bactérias e células cancerosas. Apresenta uma forma semelhante à distribuição Log-Normal, entretanto, é mais adequada na aplicação de observações censuradas em análise descritiva. Alguns exemplos de aplicações podem ser encontrados em Bennett (1983), Singh, Lee e George (1988) e Diekmann (1989).

Considere uma v.a  $T$  seguindo distribuição Log-Logística  $(\alpha, \gamma)$ . Neste caso, a função densidade de probabilidade é dada por:

$$f(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1} \left[ 1 + \left( \frac{t}{\alpha} \right)^\gamma \right]^{-2},$$

onde  $t > 0$ ,  $\alpha > 0$  o parâmetro de escala e  $\gamma > 0$  o parâmetro de forma. Na AS, o modelo Log-Logístico tem uma vantagem por apresentar expressões simples para as funções de sobrevivência,  $S(t)$ , e de risco  $h(t)$ . Essas funções estão representadas abaixo.

$$S(t) = \frac{1}{1 + \left( \frac{t}{\alpha} \right)^\gamma},$$

e

$$h(t) = \frac{f(t)}{S(t)} = \frac{\gamma \left( \frac{t}{\alpha} \right)^{\gamma-1}}{\alpha \left[ 1 + \left( \frac{t}{\alpha} \right)^\gamma \right]}.$$

De maneira análoga, segue no modelo de regressão Log-Logístico a relação:

$$\log(\alpha_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}.$$

A razão de riscos dos pacientes que apresentam o fator associado em relação aos que não apresentam é:

$$\frac{h(t|x_1 = 1)}{h(t|x_1 = 0)} = \frac{\gamma \left( \frac{t}{\exp(\beta_0 + \beta_1 x_1)} \right)^{\gamma-1}}{\exp(\beta_0 + \beta_1 x_1) \left[ 1 + \left( \frac{t}{\exp(\beta_0 + \beta_1 x_1)} \right)^\gamma \right]} \frac{\exp(\beta_0) \left[ 1 + \left( \frac{t}{\exp(\beta_0)} \right)^\gamma \right]}{\gamma \left( \frac{t}{\exp(\beta_0)} \right)^{\gamma-1}}.$$

Como podemos observar, os riscos não são proporcionais pois há dependência do tempo e, também, fica difícil de termos uma precisão melhor da razão de risco. Por outro lado, podemos utilizar a razão dos tempos medianos de sobrevivência para interpretar os coeficientes de regressão. Isto é feito através da seguinte relação:

$$\begin{aligned} \text{Med}(T_i) &= \alpha_i \\ \frac{\text{Med}(T|x_1 = 1)}{\text{Med}(T|x_1 = 0)} &= \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} = e^{\beta_1}. \end{aligned}$$

Se  $x_1$  é uma covariável binária (0, 1), então o tempo mediano de sobrevida do grupo  $\{x_1 = 1\}$  é  $(e^{\beta_1})$  vezes o tempo do grupo  $\{x_1 = 0\}$ .

### 3.4 Seleção de Variáveis

Segundo Bozdogan (1987), a escolha do modelo apropriado, do ponto de vista estatístico, é o tópico mais importante para a análise de dados. O interesse é encontrar um modelo mais parcimonioso, isto é, o modelo que contenha o mínimo possível de parâmetros a serem estimados e que explique bem como é o comportamento da variável resposta. Dentre os critérios para a seleção de modelos, os mais utilizados são o Teste de Razão de Verossimilhanças (TRV), o Critério de Informação de Akaike (AIC) e o Critério de Informação Bayesiano (BIC). O AIC é um estimador de erro de predição e, portanto, da qualidade relativa dos modelos estatísticos para um determinado conjunto de dados (SWEET, 2017) que, quanto menor o valor do mesmo, melhor é a qualidade do modelo (AKAIKE, 1974); o BIC é outro critério para seleção de modelo entre um conjunto finito de modelos que aquele com menor valor, é o modelo mais apropriado (SCHWARZ, 1978); já o TRV ou log-verossimilhança consiste em comparar um modelo paramétrico restrito a um modelo irrestrito (LARRY, 2004).

Na sequência, foram ajustados modelos de regressão simples e múltipla para o tempo de sobrevida em função das variáveis descritas na Tabela 1. Devido ao grande número de

fatores e a fim de determinar o conjunto de variáveis que melhor expliquem o tempo de sobrevida, foram utilizadas técnicas de seleção automática, a saber, *stepwise* e *backward* (JR; LEMESHOW; STURDIVANT, 2013; MONTGOMERY; PECK; VINING, 2021) utilizando o AIC como critério. A técnica *stepwise* é uma ferramenta automática usada nos estágios exploratórios da construção de modelos para identificar um subconjunto útil de preditores. O processo adiciona sistematicamente a variável mais significativa ou remove a variável menos significativa durante cada etapa. Já a técnica *backward* é caracterizada por incorporar todas as variáveis e após percorrer várias etapas, uma variável por vez é removida. Se uma das etapas não houver remoção de alguma variável, o procedimento é interrompido e as variáveis restantes definem o modelo final.

## 3.5 Adequação do Modelo Ajustado

Após ajustar o modelo de regressão é fundamental avaliar a adequação dos dados ao modelo ajustado. Em modelos de regressão linear essa avaliação é feita através da análise de construção dos gráficos dos resíduos, os quais serão normalmente distribuídos quando o modelo ajustado for adequado. Essa característica fez com que diversas técnicas fossem desenvolvidas para avaliar a normalidade dos resíduos de um modelo. Contudo, em modelos mais complexos, a construção de resíduos normalmente distribuídos não é tão direta quanto em regressão linear. Para a AS, os autores Li, Wu e Feng (2021) desenvolveram novos resíduos para avaliar os modelos ajustados na presença de dados censurados. Esses resíduos são chamados de resíduos da Probabilidade de Sobrevivência Aleatorizada Normalizada (NRSP). A seguir, apresentamos brevemente os resíduos NRSP.

### 3.5.1 Resíduos NRSP

O uso de resíduos de probabilidade de sobrevida aleatorizadas normalizadas torna-se importante para realizar diagnósticos de modelos de regressão com dados censurados. A ideia central desses resíduos é similar à dos Resíduos Quantílicos Aleatorizados (DUNN; SMYTH, 1996) que, baseia no teorema da inversa da função distribuição acumulada com a diferença de que, quando um tempo de falha,  $t$ , for censurado, a função de sobrevivência  $S(t)$  será substituída por um número aleatório entre 0 e  $S(t)$ . Aplicando a função quantil nas probabilidades de sobrevivência estimadas aleatorizadas, obtemos os resíduos NRSP.

Esses resíduos nos permitem realizar diagnósticos de não linearidade além da verificação da qualidade de ajuste geral do modelo. Os resíduos NRSP podem estar mais centrados em torno do 0 do que distribuídos como  $N(0, 1)$ . Apesar disto, esse viés é muito pequeno quando o tamanho da amostra é muito maior que o número de parâmetros no mo-

delo. Quando um modelo é especificamente correto, a distribuição condicional de NRSPs dado o vetor de covariáveis  $x_i$  é seguido próximo a uma distribuição normal padrão. Consequentemente, a maioria das ferramentas de diagnóstico para modelos de regressão linear pode ser aplicada aos resíduos NRSP para diagnosticar o ajuste de modelos de regressão com dados censurados.

Formalmente, o resíduo NRSP é definido da seguinte forma

$$r_i^{NRSP}(t_i, \delta_i, U_i) = \Phi^{-1} \left( S_i^R(t_i, \delta_i, U_i) \right),$$

onde,  $S_i^R(t_i, \delta_i, U_i)$  é o valor aleatório entre 0 e  $S_i(t_i)$  quando  $t_i$  for uma censura e  $\Phi(\cdot)$  é a função de distribuição acumulada da Normal Padrão.

---

## Capítulo 4

---

# Resultados

---

Neste capítulo são abordados os resultados de todas as análises realizadas no banco de dados. Na Seção 4.1, apresentamos as frequências relativas e absolutas de cada variável inserida na análise. Posteriormente, é apresentado as probabilidades de sobrevivência dos pacientes em um período de dias. Na Seção 4.2 apresentamos a seleção do melhor modelo de regressão paramétrico, dentre os modelos considerados. Na Seção 4.3 apresentamos a discussão dos resultados do modelo de regressão simples e múltiplo escolhido em função das variáveis explicativas. Na Seção 4.4 discutimos a análise dos resíduos para o modelo de regressão considerado.

### 4.1 Análise Descritiva

Inicialmente foi observada a quantidade de pacientes hospitalizados por fatores associados, de modo geral. As observações perdidas, isto é, ignoradas ou não preenchidas foram descartadas para a análise, sendo que foi observado um total de 304.829 censuras. A região com maior índice de hospitalização foi a Sudeste com 50,70%. Os sintomas mais notificados que os pacientes adquiriram foram a falta de ar (dispneia), com 71,53%, seguindo com a tosse (70,49%) e febre (61,79%). O sintoma menos presenciado foi a dor abdominal com apenas 3,18% dos pacientes hospitalizados. Já entre os fatores de risco/comorbidades temos a doença cardiovascular crônica com maior frequência entre os pacientes, com 36,17%; as puérperas (até 45 dias do parto) e síndrome de down são os fatores menos registrados, de 0,29% e 0,26%, respectivamente. É importante dizer que 28,76% dos pacientes hospitalizados tiveram outro tipo de fator de comorbidade não esclarecido. Na Tabela 1 é possível observar todas essas informações com os demais fatores associados.

Tabela 1 – Frequências absoluta e relativa (%) de pacientes hospitalizados pela Covid-19, aos fatores associados.

| Variável                 | Categoria        | Frequências |              |
|--------------------------|------------------|-------------|--------------|
|                          |                  | Absoluta    | Relativa (%) |
| Região                   | Norte            | 35514       | 7,39         |
|                          | Nordeste         | 83647       | 17,40        |
|                          | Centro-Oeste     | 45151       | 9,39         |
|                          | Sudeste          | 243757      | 50,70        |
|                          | Sul              | 72683       | 15,12        |
| Faixa Etária             | Adulto           | 213684      | 44,44        |
|                          | Idoso            | 267104      | 55,56        |
| Sexo                     | F                | 210959      | 43,88        |
|                          | M                | 269770      | 56,12        |
| Raça/Cor                 | Branca           | 94553       | 56,22        |
|                          | Preta            | 10869       | 6,46         |
|                          | Amarela          | 1852        | 1,10         |
|                          | Parda            | 60208       | 35,80        |
|                          | Indígena         | 713         | 0,42         |
| Escolaridade             | Sem              | 12458       | 7,04         |
|                          | Fund. Incompleto | 48683       | 27,51        |
|                          | Fund. Completo   | 33043       | 18,67        |
|                          | Médio            | 55064       | 31,11        |
|                          | Superior         | 27735       | 15,67        |
| Internado em UTI         | Não              | 267332      | 61,23        |
|                          | Sim              | 169284      | 38,77        |
| Febre                    | Não              | 183698      | 38,21        |
|                          | Sim              | 297090      | 61,79        |
| Tosse                    | Não              | 141891      | 29,51        |
|                          | Sim              | 338897      | 70,49        |
| Dor de Garganta          | Não              | 398654      | 82,92        |
|                          | Sim              | 82134       | 17,08        |
| Dispneia                 | Não              | 136862      | 28,47        |
|                          | Sim              | 343926      | 71,53        |
| Desconforto Respiratório | Não              | 203345      | 42,29        |
|                          | Sim              | 277443      | 57,71        |

Continua...

| Variável          | Categoria | Frequências |              |
|-------------------|-----------|-------------|--------------|
|                   |           | Absoluta    | Relativa (%) |
| Saturação baixa   | Não       | 193782      | 40,30        |
|                   | Sim       | 287006      | 59,70        |
| Diarreia          | Não       | 415446      | 86,41        |
|                   | Sim       | 65342       | 13,59        |
| Vômito            | Não       | 443578      | 92,26        |
|                   | Sim       | 37210       | 7,74         |
| Dor abdominal     | Não       | 465487      | 96,82        |
|                   | Sim       | 15301       | 3,18         |
| Fadiga            | Não       | 421328      | 87,63        |
|                   | Sim       | 59460       | 12,37        |
| Perda de Paladar  | Não       | 451211      | 93,85        |
|                   | Sim       | 29577       | 6,15         |
| Perda de Olfato   | Não       | 451442      | 93,90        |
|                   | Sim       | 29346       | 6,10         |
| Puérpera          | Não       | 479390      | 99,71        |
|                   | Sim       | 1398        | 0,29         |
| Cardiovascular    | Não       | 306899      | 63,83        |
|                   | Sim       | 173889      | 36,17        |
| Hematológica      | Não       | 477034      | 99,22        |
|                   | Sim       | 3754        | 0,78         |
| Síndrome de Down  | Não       | 479546      | 99,74        |
|                   | Sim       | 1242        | 0,26         |
| Hepática          | Não       | 476185      | 99,04        |
|                   | Sim       | 4603        | 0,96         |
| Asma              | Não       | 467988      | 97,34        |
|                   | Sim       | 12800       | 2,66         |
| Diabetes          | Não       | 352401      | 73,30        |
|                   | Sim       | 128387      | 26,70        |
| Neurológica       | Não       | 460391      | 95,76        |
|                   | Sim       | 20397       | 4,24         |
| Outra Pneumopatia | Não       | 460946      | 95,87        |
|                   | Sim       | 19842       | 4,13         |
| Imunodepressão    | Não       | 467438      | 97,22        |
|                   | Sim       | 13350       | 2,78         |

Continua...

| Variável                   | Categoria | Frequências |              |
|----------------------------|-----------|-------------|--------------|
|                            |           | Absoluta    | Relativa (%) |
| Renal                      | Não       | 459685      | 95,61        |
|                            | Sim       | 21103       | 4,39         |
| Obesidade                  | Não       | 450885      | 93,78        |
|                            | Sim       | 29903       | 6,22         |
| Outro fator de comorbidade | Não       | 342503      | 71,24        |
|                            | Sim       | 138285      | 28,76        |

Em seguida, avaliamos o tempo de sobrevivência de maneira descritiva, utilizando o estimador de K-M para a função de sobrevivência, na qual é possível observar a probabilidade de um paciente sobreviver um determinado número de dias. Na Figura 1, temos o comportamento da função de sobrevivência em geral, ou seja, sem considerar nenhum dos fatores associados aos pacientes. Por meio dessas estimativas, constatou-se que a probabilidade de um paciente sobreviver ao primeiro dia de internação é de 97,9%, caindo para 87,5% no sétimo dia e chegando a 33,5% no trigésimo dia. Podemos observar que metade dos pacientes vem a óbito em até 21 dias de internação, sendo o tempo mediano de sobrevida de 21 dias e 63,40% dos tempos de sobrevida foram censuradas, isto é, são tempos de pacientes que tiveram alta ou óbito por outra causa que não a Covid-19. Por outro lado, 25% dos pacientes sobreviveram por mais de 38 dias evidenciando que a maior parte dos óbitos está concentrada nos primeiros dias após a internação e, ao passar do tempo, a taxa de mortalidade diminui.

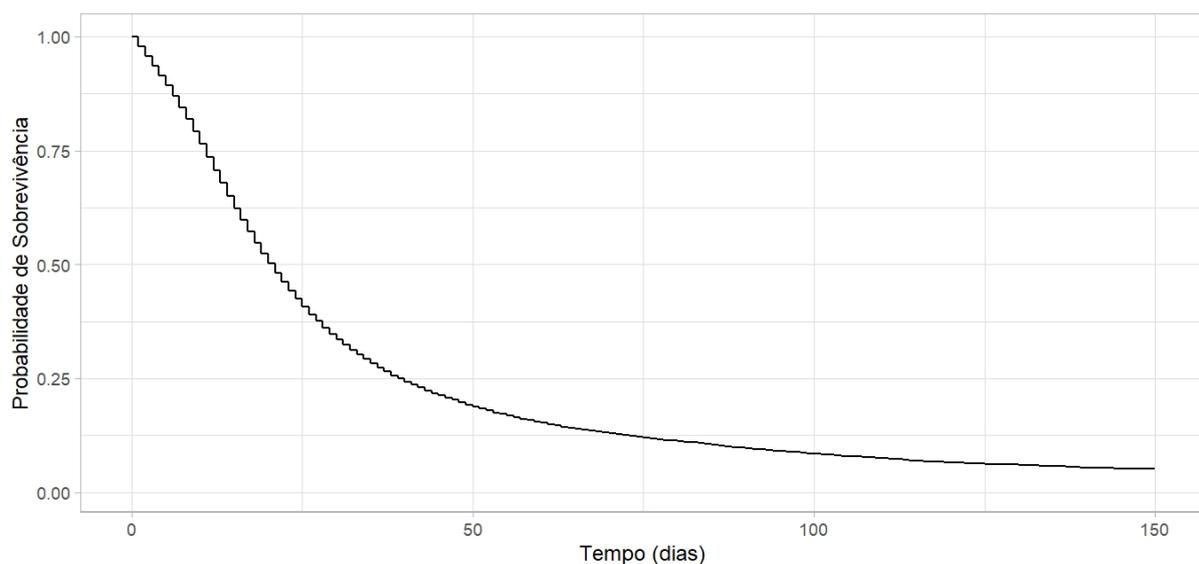


Figura 1 – Curva de sobrevivência estimada por Kaplan-Meier, dos pacientes em geral.

Nas Tabelas 2, 3 e 4 são apresentadas as probabilidades de um paciente sobreviver por mais de 1, 7, 15 e 30 dias, segundo os níveis dos fatores associados, além da quantidade de óbitos, do tempo mediano de sobrevida e do p-valor do teste Log-rank para cada fator associado dividido em variáveis sócio-demográficas, sintomas e comorbidades, respectivamente. Observamos na Tabela 2, que o maior número de casos de mortes foi a região Sudeste com 85.648 óbitos, que corresponde a 35,14% dos casos internados na região. O tempo mediano de sobrevida desta região é de 21 dias e a probabilidade de um paciente sobreviver além de 30 dias é de 34,09%. Como o p-valor do teste Log-rank é inferior a 5%, o tempo de sobrevida dos pacientes internados por Covid-19 difere segundo as regiões do Brasil. Ainda, de acordo com o teste Log-rank, não existem evidências o suficiente para afirmar que o tempo de sobrevida de homens e mulheres seja diferente e o mesmo ocorre quando comparamos pacientes com e sem síndrome de down, como podemos observar na Tabela 4. Nos demais fatores associados, em pelo menos um nível, o tempo de sobrevida difere dos demais.

Para a Tabela 3 podemos observar que mais de 50% dos pacientes internados na UTI não sobreviveram, enquanto que entre os que foram internados apenas na enfermaria a proporção de óbitos é de 21,77%. A probabilidade do paciente internado sobreviver por mais de 30 dias é de 27,63%, sendo que o tempo mediano de sobrevida é 18 dias. Notou-se também, que pacientes com saturação baixa tem a maior proporção de óbitos entre os sintomas, representando 42,65% dos pacientes que tiveram óbito e a probabilidade de que um paciente desse grupo sobreviva por mais de um mês é inferior a 31%.

Com a Tabela 4 podemos observar que pacientes com doença hepática crônica e neurológica crônica tem o mesmo tempo mediano de sobrevida de 16 dias. Entre as comorbidades a maior proporção de óbitos está em pacientes com doenças renais, sendo que nesse grupo 59,25% dos pacientes faleceram e a probabilidade de que um paciente desse grupo sobreviva por mais de 30 dias é inferior a 27%.

Tabela 2 – Avaliação dos fatores associados ao tempo de sobrevivida de pacientes internados pela Covid-19 por sócio-demográficos.

| Características     | Óbitos |       | Probabilidade de sobrevivida |        |         |         | Mediana | p-valor |
|---------------------|--------|-------|------------------------------|--------|---------|---------|---------|---------|
|                     | n      | %     | 1º Dia                       | 7º Dia | 15º Dia | 30º Dia |         |         |
| <b>Região</b>       |        |       |                              |        |         |         |         |         |
| Norte               | 15112  | 42,55 | 96,88                        | 77,87  | 55,58   | 30,83   | 18      |         |
| Nordeste            | 38174  | 45,64 | 96,28                        | 78,30  | 56,05   | 30,28   | 18      |         |
| Centro-Oeste        | 14661  | 32,47 | 98,77                        | 87,52  | 66,03   | 34,37   | 22      | <0,001  |
| Sudeste             | 85648  | 35,14 | 98,15                        | 85,72  | 63,24   | 34,09   | 21      |         |
| Sul                 | 22344  | 30,74 | 99,00                        | 89,69  | 68,49   | 36,58   | 23      |         |
| <b>Faixa Etária</b> |        |       |                              |        |         |         |         |         |
| Adulto              | 41227  | 19,29 | 98,78                        | 91,48  | 76,38   | 50,25   | 31      |         |
| Idoso               | 134732 | 50,44 | 97,28                        | 79,49  | 54,44   | 26,07   | 17      | <0,001  |
| <b>Sexo</b>         |        |       |                              |        |         |         |         |         |
| F                   | 74747  | 35,43 | 97,13                        | 84,19  | 62,58   | 34,19   | 21      |         |
| M                   | 101195 | 37,51 | 97,92                        | 84,79  | 62,21   | 33,09   | 20      | 0,090   |
| <b>Raça/Cor</b>     |        |       |                              |        |         |         |         |         |
| Branca              | 32472  | 34,34 | 98,42                        | 86,53  | 64,18   | 33,79   | 21      |         |
| Preta               | 4669   | 42,96 | 96,89                        | 81,00  | 57,46   | 29,39   | 18      |         |
| Amarela             | 683    | 36,88 | 98,08                        | 84,64  | 63,49   | 32,57   | 20      | <0,001  |
| Parda               | 24877  | 41,32 | 97,26                        | 81,38  | 58,56   | 31,17   | 19      |         |
| Indígena            | 302    | 42,36 | 97,80                        | 81,30  | 56,10   | 28,80   | 16      |         |
| <b>Escolaridade</b> |        |       |                              |        |         |         |         |         |
| Sem                 | 7260   | 58,28 | 95,58                        | 70,65  | 44,30   | 20,46   | 14      |         |
| Fund Incompleto     | 23019  | 47,28 | 97,35                        | 79,66  | 54,58   | 26,17   | 17      |         |
| Fund Completo       | 13157  | 39,82 | 97,63                        | 83,74  | 60,47   | 31,16   | 19      | <0,001  |
| Médio               | 15728  | 28,56 | 98,47                        | 88,79  | 68,71   | 37,77   | 23      |         |
| Superior            | 6287   | 22,67 | 99,07                        | 92,76  | 75,41   | 43,22   | 26      |         |

Tabela 3 – Avaliação dos fatores associados ao tempo de sobrevida de pacientes internados pela Covid-19 por sintomas.

| Características          | Óbitos |        | Probabilidade de sobrevida |        |         |         |         |    |        | Mediana | p-valor |
|--------------------------|--------|--------|----------------------------|--------|---------|---------|---------|----|--------|---------|---------|
|                          | n      | %      | 1º Dia                     | 7º Dia | 15º Dia | 30º Dia | 30º Dia |    |        |         |         |
| Internado em UTI         | Não    | 58197  | 21,77                      | 98,25  | 88,08   | 70,55   | 43,28   | 26 | <0,001 |         |         |
|                          | Sim    | 99617  | 58,85                      | 98,45  | 81,53   | 56,16   | 27,63   | 18 |        |         |         |
| Febre                    | Não    | 74634  | 40,63                      | 97,16  | 81,92   | 58,85   | 30,42   | 19 | <0,001 |         |         |
|                          | Sim    | 101325 | 34,11                      | 98,38  | 86,13   | 64,58   | 35,61   | 22 |        |         |         |
| Tosse                    | Não    | 61305  | 43,21                      | 96,70  | 80,40   | 57,58   | 30,11   | 19 | <0,001 |         |         |
|                          | Sim    | 114654 | 33,83                      | 98,42  | 86,26   | 64,44   | 35,12   | 22 |        |         |         |
| Dor de Garganta          | Não    | 151169 | 37,92                      | 97,81  | 84,01   | 61,57   | 32,81   | 20 | <0,001 |         |         |
|                          | Sim    | 24790  | 30,18                      | 98,39  | 87,07   | 66,48   | 37,65   | 23 |        |         |         |
| Dispneia                 | Não    | 39406  | 28,79                      | 98,05  | 88,00   | 68,38   | 39,99   | 24 | <0,001 |         |         |
|                          | Sim    | 136553 | 39,70                      | 97,86  | 83,21   | 60,21   | 31,41   | 20 |        |         |         |
| Desconforto Respiratório | Não    | 61186  | 30,09                      | 98,22  | 87,80   | 67,73   | 39,04   | 23 | <0,001 |         |         |
|                          | Sim    | 114773 | 41,37                      | 97,69  | 82,21   | 58,78   | 30,12   | 19 |        |         |         |
| Saturação baixa          | Não    | 53554  | 27,64                      | 98,16  | 88,13   | 68,40   | 39,88   | 24 | <0,001 |         |         |
|                          | Sim    | 122405 | 42,65                      | 97,75  | 82,26   | 58,95   | 30,32   | 19 |        |         |         |
| Diarreia                 | Não    | 156911 | 37,77                      | 97,79  | 83,93   | 61,55   | 33,05   | 20 | <0,001 |         |         |
|                          | Sim    | 19048  | 29,15                      | 98,74  | 88,38   | 67,93   | 37,05   | 23 |        |         |         |
| Vômito                   | Não    | 164116 | 37,00                      | 97,88  | 84,38   | 62,07   | 33,29   | 20 | <0,001 |         |         |
|                          | Sim    | 11843  | 31,83                      | 98,36  | 86,26   | 66,08   | 36,94   | 22 |        |         |         |
| Dor abdominal            | Não    | 171348 | 36,81                      | 97,89  | 88,44   | 62,24   | 33,49   | 21 | <0,001 |         |         |
|                          | Sim    | 4611   | 30,14                      | 98,69  | 87,31   | 66,33   | 35,31   | 22 |        |         |         |
| Fadiga                   | Não    | 157144 | 37,30                      | 97,80  | 84,05   | 61,90   | 33,69   | 21 | <0,001 |         |         |
|                          | Sim    | 18815  | 31,64                      | 98,75  | 87,94   | 65,71   | 32,06   | 21 |        |         |         |
| Perda de Paladar         | Não    | 169416 | 37,55                      | 97,83  | 84,06   | 61,81   | 33,28   | 20 | <0,001 |         |         |
|                          | Sim    | 6543   | 22,12                      | 99,28  | 91,93   | 72,24   | 38,06   | 24 |        |         |         |
| Perda de Olfato          | Não    | 169411 | 37,53                      | 97,83  | 84,08   | 61,82   | 33,30   | 20 | <0,001 |         |         |
|                          | Sim    | 6548   | 22,31                      | 99,23  | 91,73   | 72,12   | 37,68   | 24 |        |         |         |

Tabela 4 – Avaliação dos fatores associados ao tempo de sobrevida de pacientes internados pela Covid-19 por comorbidades.

| Características                   | Óbitos |        | Probabilidade de sobrevida |        |         |         | Mediana | p-valor |        |
|-----------------------------------|--------|--------|----------------------------|--------|---------|---------|---------|---------|--------|
|                                   | n      | %      | 1º Dia                     | 7º Dia | 15º Dia | 30º Dia |         |         |        |
| <b>Puérpera</b>                   | Não    | 175705 | 36,65                      | 97,91  | 84,51   | 62,33   | 33,52   | 21      | <0,001 |
|                                   | Sim    | 253    | 18,10                      | 98,70  | 91,90   | 77,60   | 46,70   | 29      |        |
| <b>Cardiovascular</b>             | Não    | 98404  | 32,06                      | 97,83  | 85,55   | 64,77   | 35,66   | 22      | <0,001 |
|                                   | Sim    | 77555  | 44,60                      | 98,07  | 82,84   | 58,84   | 29,58   | 19      |        |
| <b>Hematológica</b>               | Não    | 174101 | 36,50                      | 97,91  | 84,56   | 62,39   | 33,57   | 21      | <0,001 |
|                                   | Sim    | 1858   | 49,59                      | 98,03  | 80,86   | 58,37   | 31,23   | 19      |        |
| <b>Síndrome de Down</b>           | Não    | 175407 | 36,58                      | 97,92  | 84,54   | 62,37   | 33,54   | 21      | 0,300  |
|                                   | Sim    | 552    | 44,44                      | 97,02  | 81,43   | 58,63   | 34,35   | 19      |        |
| <b>Hepática</b>                   | Não    | 173391 | 36,41                      | 97,92  | 84,60   | 62,48   | 33,66   | 21      | <0,001 |
|                                   | Sim    | 2568   | 55,79                      | 97,50  | 77,09   | 52,19   | 24,85   | 16      |        |
| <b>Asma</b>                       | Não    | 172013 | 36,76                      | 97,90  | 84,41   | 62,20   | 33,42   | 21      | <0,001 |
|                                   | Sim    | 3946   | 30,83                      | 98,58  | 88,73   | 68,44   | 38,42   | 23      |        |
| <b>Diabetes</b>                   | Não    | 117614 | 33,38                      | 97,95  | 85,57   | 64,25   | 35,32   | 22      | <0,001 |
|                                   | Sim    | 58345  | 45,44                      | 97,81  | 81,81   | 57,89   | 29,60   | 19      |        |
| <b>Neurológica</b>                | Não    | 163906 | 35,60                      | 97,93  | 84,95   | 63,00   | 34,01   | 21      | <0,001 |
|                                   | Sim    | 12053  | 59,09                      | 97,54  | 75,78   | 50,60   | 25,65   | 16      |        |
| <b>Outra Pneumopatia</b>          | Não    | 164823 | 35,76                      | 97,93  | 84,80   | 62,82   | 33,99   | 21      | <0,001 |
|                                   | Sim    | 11136  | 56,12                      | 97,66  | 78,69   | 53,70   | 26,28   | 17      |        |
| <b>Imunodepressão</b>             | Não    | 169185 | 36,19                      | 97,91  | 84,63   | 62,47   | 33,61   | 21      | <0,001 |
|                                   | Sim    | 6774   | 50,74                      | 98,02  | 80,93   | 58,91   | 31,59   | 19      |        |
| <b>Renal</b>                      | Não    | 163455 | 35,56                      | 97,92  | 84,85   | 62,90   | 33,99   | 21      | <0,001 |
|                                   | Sim    | 12504  | 59,25                      | 97,88  | 78,02   | 53,05   | 26,78   | 17      |        |
| <b>Obesidade</b>                  | Não    | 164560 | 36,50                      | 97,89  | 84,36   | 62,21   | 33,51   | 21      | <0,001 |
|                                   | Sim    | 11399  | 38,12                      | 98,37  | 87,01   | 64,52   | 34,13   | 21      |        |
| <b>Outro fator de comorbidade</b> | Não    | 114497 | 33,43                      | 97,87  | 85,30   | 63,96   | 35,35   | 21      | <0,001 |
|                                   | Sim    | 61462  | 44,45                      | 98,04  | 82,69   | 58,92   | 30,05   | 19      |        |

## 4.2 Ajuste dos Modelos de Regressão Paramétricos

A Tabela 5 mostra os valores de AIC, BIC e log-verossimilhança dos modelos ajustados ao banco de dados mais utilizados na literatura.

Tabela 5 – Estimação dos critérios AIC e BIC e logaritmo das funções de verossimilhança para os modelos selecionados.

| Distribuição         | AIC            | BIC            | log-verossimilhança |
|----------------------|----------------|----------------|---------------------|
| Exponencial          | 1588795        | 1588806        | -794396,6           |
| Weibull              | 1577537        | 1577559        | -788766,5           |
| Gama                 | 1573510        | 1573532        | -786752,8           |
| Gama Generalizada    | 1568203        | 1568236        | -784098,6           |
| Log-Normal           | 1572565        | 1572587        | -786280,3           |
| <b>Log-Logística</b> | <b>1566898</b> | <b>1566920</b> | <b>-783447,0</b>    |
| Gompertz             | 1588503        | 1588526        | -794249,7           |

O modelo Log-logístico foi o que apresentou o menor valor do AIC, significando ser o modelo mais parcimonioso para seleção das covariáveis. O modelo múltiplo final, encontrado por ambas as técnicas (stepwise e backward), contém todos os fatores exceto a dor abdominal, síndrome de down e puérpera e os resultados estão apresentados na Tabela 7. Para o modelo múltiplo, nota-se que todas as variáveis foram significativas ( $p$ -valor  $< 0,05$ ), ou seja, todas impactam o tempo de sobrevivência dos pacientes e foram incluídos em análise 158358 pacientes.

## 4.3 Interpretação dos Resultados

### 4.3.1 Estimativas Brutas

A Tabela 6 apresenta os resultados do modelo Log-logístico ajustado em função de cada fator associado individualmente. Apresentamos os valores estimados dos coeficientes de regressão  $\beta$  e a razão estimada dos tempos medianos de sobrevivência  $e^\beta$ , o intervalo de 95% de confiança estimado para  $e^\beta$  (LI = Limite Inferior; LS = Limite Superior) e o  $p$ -valor, associado ao teste de Wald para avaliar a significância de cada coeficiente ( $\beta$ ) no modelo, indicando se essa razão é diferente de 1, sendo que quando o  $p$ -valor for maior que 0,05, não temos evidências significativas de que  $e^\beta$  seja diferente de 1 e, portanto, não temos evidências de que o tempo de sobrevivência seja influenciado pelo respectivo fator.

Avaliando os resultados do modelo log-logístico, notamos que o tempo mediano de sobrevivência dos pacientes internados na Região Norte é 33% menor que na Região Sul (nível de referência - ausente na tabela). Nas demais regiões, como Centro-Oeste, Nor-

deste e Sudeste, temos evidência de que o tempo mediano de vida nesses lugares diferem ( $p$ -valor  $< 0,001$  para ambos). Para a variável escolaridade, observe que temos evidências de diferenças significativas no tempo de sobrevida mediano entre pacientes dos grupos sem escolaridade (nível de referência) com os demais grupos. Mais especificamente, o tempo mediano de sobrevida de pacientes que completaram o ensino fundamental é 53% maior que dos pacientes sem escolaridade. E mais, pacientes com o ensino superior que foram internados, o tempo mediano de sobrevida é mais de duas vezes daqueles comparados com o grupo sem escolaridade.

Para as variáveis perda de paladar e olfato, os pacientes que foram notificados com a presença desses sintomas, tem quase o mesmo tempo mediano de sobrevida com diferença de 1%, daqueles pacientes que não tiveram esses sintomas. Observa-se também, pacientes puérperas apresentam tempo de sobrevida mediano 50% maior que pacientes sem essa características.

Tabela 6 – Resultados do modelo simples Log-logístico de sobrevida dos pacientes internados pela Covid-19 aos fatores associados.

| Características/Efeito   |                  | Simples |             |      |      |         |
|--------------------------|------------------|---------|-------------|------|------|---------|
|                          |                  | $\beta$ | $e^{\beta}$ | LI   | LS   | p-valor |
| Região                   | Centro-Oeste     | -0,07   | 0,93        | 0,91 | 0,94 | <0,001  |
|                          | Sudeste          | -0,15   | 0,86        | 0,85 | 0,88 | <0,001  |
|                          | Nordeste         | -0,38   | 0,68        | 0,67 | 0,69 | <0,001  |
|                          | Norte            | -0,39   | 0,67        | 0,66 | 0,69 | <0,001  |
| Faixa Etária             | Idoso            | -0,63   | 0,53        | 0,53 | 0,54 | 0,000   |
| Sexo                     | M                | 0,00    | 1,00        | 1,00 | 1,01 | 0,137   |
| Raça/Etnia               | Amarela          | -0,06   | 0,94        | 0,91 | 0,98 | 0,002   |
|                          | Parda            | -0,18   | 0,84        | 0,83 | 0,85 | <0,001  |
|                          | Preta            | -0,20   | 0,82        | 0,80 | 0,83 | <0,001  |
|                          | Indígena         | -0,21   | 0,81        | 0,76 | 0,87 | <0,001  |
| Escolaridade             | Fund. Incompleto | 0,27    | 1,31        | 1,28 | 1,35 | <0,001  |
|                          | Fund. Completo   | 0,42    | 1,53        | 1,49 | 1,57 | <0,001  |
|                          | Médio            | 0,65    | 1,91        | 1,86 | 1,95 | <0,001  |
|                          | Superior         | 0,85    | 2,34        | 2,28 | 2,41 | <0,001  |
| Uti                      | Sim              | -0,33   | 0,72        | 0,71 | 0,72 | 0,000   |
| Febre                    | Sim              | 0,18    | 1,19        | 1,18 | 1,20 | 0,000   |
| Tosse                    | Sim              | 0,22    | 1,24        | 1,23 | 1,25 | 0,000   |
| Dor de garganta          | Sim              | 0,14    | 1,15        | 1,14 | 1,16 | <0,001  |
| Dispneia                 | Sim              | -0,23   | 0,80        | 0,79 | 0,80 | 0,000   |
| Desconforto Respiratório | Sim              | -0,25   | 0,78        | 0,77 | 0,78 | 0,000   |
| Saturação baixa          | Sim              | -0,26   | 0,77        | 0,76 | 0,77 | 0,000   |
| Diarreia                 | Sim              | 0,18    | 1,20        | 1,18 | 1,21 | <0,001  |
| Vômito                   | Sim              | 0,10    | 1,10        | 1,08 | 1,12 | 0,000   |
| Dor abdominal            | Sim              | 0,11    | 1,12        | 1,09 | 1,14 | <0,001  |
| Fadiga                   | Sim              | 0,11    | 1,12        | 1,11 | 1,13 | <0,001  |
| Perda de paladar         | Sim              | 0,32    | 1,38        | 1,35 | 1,40 | <0,001  |
| Perda de olfato          | Sim              | 0,31    | 1,37        | 1,34 | 1,39 | 0,000   |
| Puérpera                 | Sim              | 0,41    | 1,50        | 1,36 | 1,65 | <0,001  |
| Cardiovascular           | Sim              | -0,14   | 0,87        | 0,86 | 0,87 | <0,001  |
| Hematológica             | Sim              | -0,12   | 0,89        | 0,85 | 0,92 | 0,000   |
| Síndrome de Down         | Sim              | -0,09   | 0,92        | 0,85 | 0,99 | 0,011   |
| Hepática                 | Sim              | -0,27   | 0,76        | 0,73 | 0,79 | <0,001  |
| Asma                     | Sim              | 0,18    | 1,20        | 1,17 | 1,23 | <0,001  |
| Diabetes Mellitus        | Sim              | -0,16   | 0,85        | 0,84 | 0,86 | <0,001  |
| Neurológica              | Sim              | -0,32   | 0,73        | 0,72 | 0,74 | <0,001  |
| Outra Pneumopatia        | Sim              | -0,23   | 0,79        | 0,78 | 0,81 | <0,001  |
| Imunodepressão           | Sim              | -0,10   | 0,90        | 0,88 | 0,92 | <0,001  |
| Renal                    | Sim              | -0,25   | 0,78        | 0,77 | 0,79 | <0,001  |
| Obesidade                | Sim              | 0,09    | 1,09        | 1,07 | 1,11 | <0,001  |
| Outra comorbidade        | Sim              | -0,13   | 0,88        | 0,88 | 0,89 | <0,001  |

As análises feitas até aqui são individuais, com intuito de que cada fator associado aos pacientes estão sendo avaliados de forma independente dos demais.

### 4.3.2 Modelo Múltiplo

A Tabela 7 apresenta os resultados do modelo múltiplo Log-logístico após a seleção de variáveis. Avaliando esses resultados, notamos que o tempo mediano de sobrevida dos pacientes internados na região Norte é 44% menor do que na região Sul (nível de referência), enquanto que na região Centro-Oeste o tempo foi 20% menor quando comparado à região de referência. Para a raça, observa-se que não temos evidências de que o tempo de sobrevivência mediano das pessoas amarelas e brancas (nível de referência) sejam diferentes, visto que  $p\text{-valor} = 0,198$ . Na variável escolaridade, pacientes com ensino superior apresentam um tempo mediano de sobrevida 73% maior dque pacientes sem escolaridade.

Nota-se que o tempo mediano de sobrevida do grupo de pacientes internados em UTI é 21% menor em relação aos pacientes internados em enfermaria. Entre os sintomas, o maior efeito observado foi o da tosse, sendo que, pacientes com tosse apresentaram um tempo mediano de sobrevida 17% maior que pacientes sem tosse. Por outro lado, pacientes com saturação baixa apresentaram um tempo mediano 15% menor que pacientes sem esse sintoma.

Para as comorbidades, a asma foi a única com um efeito positivo, sendo que pacientes com asma apresentaram um tempo mediano de sobrevida de 12% maior que pacientes sem asma. Entretanto, pacientes com doença neurológica crônica apresentaram tempo mediano de sobrevida 19% menor que pacientes sem essa comorbidade. Além disso, não temos evidências suficientes para afirmar que o tempo de sobrevida seja influenciado pelas variáveis puérperas, síndrome de down e dores abdominais.

Tabela 7 – Resultados do modelo múltiplo Log-logístico de sobrevida dos pacientes internados pela Covid-19 aos fatores associados.

| Características/Efeito     |                  | Múltiplos |             |      |      |         |
|----------------------------|------------------|-----------|-------------|------|------|---------|
|                            |                  | $\beta$   | $e^{\beta}$ | LI   | LS   | p-valor |
| Região                     | Centro-Oeste     | -0,22     | 0,80        | 0,77 | 0,82 | <0,001  |
|                            | Sudeste          | -0,26     | 0,77        | 0,76 | 0,79 | <0,001  |
|                            | Nordeste         | -0,43     | 0,65        | 0,63 | 0,67 | <0,001  |
|                            | Norte            | -0,57     | 0,56        | 0,55 | 0,58 | <0,001  |
| Faixa Etária               | Idoso            | -0,46     | 0,63        | 0,62 | 0,64 | 0,000   |
| Sexo                       | M                | -0,03     | 0,97        | 0,96 | 0,98 | <0,001  |
| Raça/Etnia                 | Amarela          | -0,04     | 0,96        | 0,90 | 1,02 | 0,100   |
|                            | Parda            | -0,06     | 0,94        | 0,93 | 0,96 | <0,001  |
|                            | Preta            | -0,11     | 0,90        | 0,87 | 0,92 | <0,001  |
|                            | Indígena         | -0,18     | 0,84        | 0,75 | 0,93 | <0,001  |
| Escolaridade               | Fund. Incompleto | 0,16      | 1,17        | 1,14 | 1,20 | <0,001  |
|                            | Fund. Completo   | 0,23      | 1,26        | 1,23 | 1,30 | <0,001  |
|                            | Médio            | 0,34      | 1,41        | 1,37 | 1,45 | <0,001  |
|                            | Superior         | 0,55      | 1,73        | 1,68 | 1,88 | <0,001  |
| Internados em UTI          | Sim              | -0,24     | 0,79        | 0,78 | 0,80 | <0,001  |
| Febre                      | Sim              | 0,08      | 1,09        | 1,07 | 1,10 | <0,001  |
| Tosse                      | Sim              | 0,16      | 1,17        | 1,16 | 1,19 | <0,001  |
| Dor de Garganta            | Sim              | 0,05      | 1,05        | 1,03 | 1,07 | <0,001  |
| Dispneia                   | Sim              | -0,15     | 0,86        | 0,85 | 0,88 | <0,001  |
| Desconforto Respiratório   | Sim              | -0,12     | 0,88        | 0,87 | 0,90 | <0,001  |
| Saturação baixa            | Sim              | -0,16     | 0,85        | 0,83 | 0,86 | <0,001  |
| Diarreia                   | Sim              | 0,10      | 1,10        | 1,08 | 1,13 | <0,001  |
| Vômito                     | Sim              | 0,03      | 1,03        | 1,01 | 1,06 | 0,007   |
| Dor abdominal              | Sim              | *         | *           | *    | *    | *       |
| Fadiga                     | Sim              | 0,06      | 1,06        | 1,04 | 1,08 | <0,001  |
| Perda de Paladar           | Sim              | 0,14      | 1,15        | 1,10 | 1,20 | <0,001  |
| Perda de Olfato            | Sim              | 0,07      | 1,07        | 1,02 | 1,12 | 0,002   |
| Puérpera                   | Sim              | *         | *           | *    | *    | *       |
| Cardiovascular             | Sim              | -0,02     | 0,98        | 0,97 | 1,00 | 0,012   |
| Hematológica               | Sim              | -0,09     | 0,91        | 0,86 | 0,98 | 0,003   |
| Síndrome de Down           | Sim              | *         | *           | *    | *    | *       |
| Hepática                   | Sim              | -0,19     | 0,83        | 0,78 | 0,88 | <0,001  |
| Asma                       | Sim              | 0,11      | 1,12        | 1,07 | 1,17 | <0,001  |
| Diabetes                   | Sim              | -0,05     | 0,95        | 0,94 | 0,96 | <0,001  |
| Neurológica                | Sim              | -0,21     | 0,81        | 0,79 | 0,84 | <0,001  |
| Outra Pneumopatia          | Sim              | -0,09     | 0,92        | 0,89 | 0,94 | <0,001  |
| Imunodepressão             | Sim              | -0,12     | 0,89        | 0,86 | 0,92 | <0,001  |
| Renal                      | Sim              | -0,12     | 0,88        | 0,86 | 0,91 | <0,001  |
| Obesidade                  | Sim              | -0,03     | 0,97        | 0,94 | 0,99 | 0,005   |
| Outro fator de comorbidade | Sim              | -0,07     | 0,94        | 0,92 | 0,95 | <0,001  |

## 4.4 Análise de Resíduos

A análise de resíduos é apresentada na Figura 2, na qual são apresentados os gráficos quantil-quantil (direita) e resíduos versus a ordem (esquerda) para os resíduos NRSP do modelo Log-logístico múltiplo. Podemos observar que a curva dos resíduos alinha-se bem com uma reta de  $45^\circ$ , com um desvio da reta na cauda superior. No gráfico da esquerda, observamos que os resíduos estão em sua maioria distribuídos no intervalo  $(-3, 3)$  e embora alguns ultrapassem esses limites, esse comportamento é esperado devido ao tamanho da amostra.

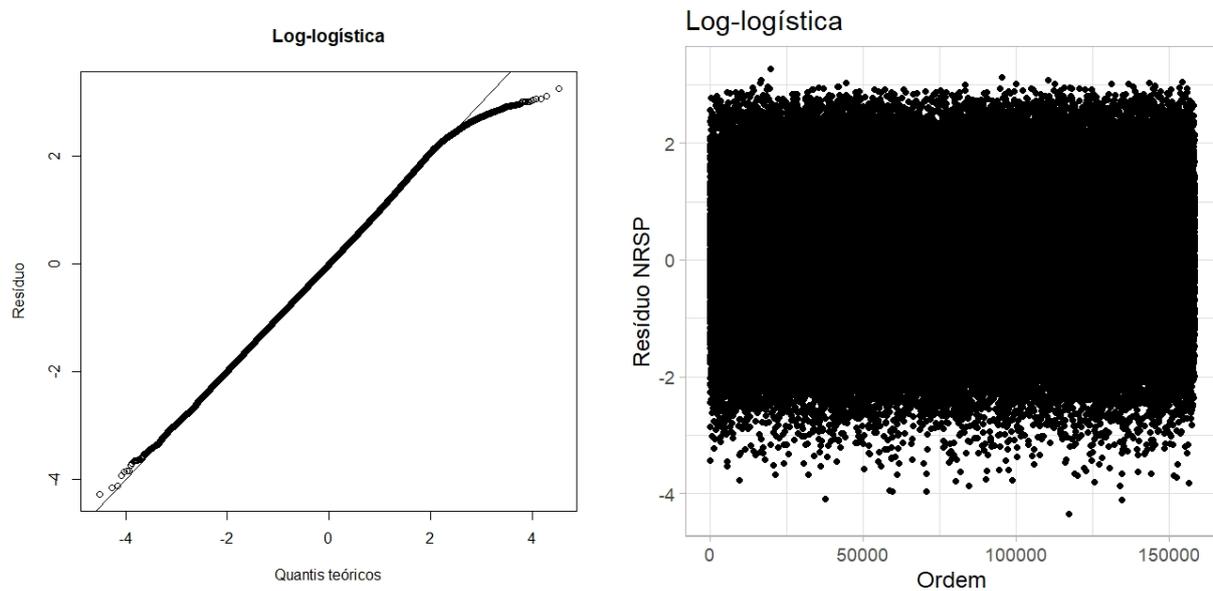


Figura 2 – Gráficos quantil-quantil (direita) e resíduos vs ordem (esquerda) dos resíduos NRSP para o modelo Log-logístico múltiplo ajustado aos tempos de sobrevivência dos pacientes acometidos pela Covid-19.

---

## Capítulo 5

---

# Discussão

---

Embora o Brasil tenha sido considerado o primeiro país da América do Sul a apresentar um caso confirmado de COVID-19, isto ocorreu várias semanas após a maioria dos países do hemisfério Norte (ORGANIZATION *et al.*, 2020a; ORGANIZATION *et al.*, 2020b).

A presença de doença cardiovascular é um fator de risco entre pacientes diagnosticados com COVID-19, uma vez que o tempo de sobrevivência desses pacientes é 2% menor quando comparados com pacientes que não apresentam essa comorbidade. Este resultado está de acordo com outros estudos que também encontraram associações significativas entre comorbidade cardiovascular e chance de óbito pela COVID-19 (DU *et al.*, 2020; SHI *et al.*, 2020; YAO *et al.*, 2020; SOUSA *et al.*, 2020; ZHANG *et al.*, 2021). Por outro lado, ao avaliar pacientes com asma nota-se que o tempo mediano de sobrevivência desses pacientes é maior em relação a pacientes sem asma ( $p$ -valor < 0,001). Esse resultado difere de outros encontrados na literatura (SOUSA *et al.*, 2020; MOWLA *et al.*, 2020; HALVATSIOTIS *et al.*, 2020) no qual não foram encontradas relações significativas entre a mortalidade pela COVID-19 e a presença de asma.

Em um grande estudo publicado pelo Centro Chinês de Controle e Prevenção de Doenças (TEAM, 2020) 44672 casos confirmados de infecção pela COVID-19 foram analisados e as comorbidades mais frequentes nos pacientes que evoluíram para óbito foram hipertensão arterial (39,7%), doença cardiovascular (22,7%) e diabetes mellitus (19,7%), enquanto que em nossa análise as comorbidades mais frequentes foram doença cardiovascular (44,08%), diabetes mellitus (33,16%) e doença renal (7,11%).

A proporção de mortes entre os indivíduos acima de 60 anos é de 50,44%, de modo que, esse grupo representa 76% das mortes ocorridas entre pessoas acima dos 18 anos, o que corrobora o resultado apresentado no site (<<https://www.gov.br/saude/pt-br/coronavirus/>

boletins-epidemiologicos/boletim-epidemiologico-covid-19-no-17.pdf/view>) no qual indivíduos com mais de 60 anos representam quase 70% dos óbitos pela COVID-19. Resultados similares também são apresentados em Sousa *et al.* (2020), Mowla *et al.* (2020), Covid *et al.* (2020) e Hueda-Zavaleta *et al.* (2021). Esses resultados são reafirmados pelo nosso estudo, em que a única comorbidade que não é um fator de risco é a asma. Na análise de regressão, observa-se que pacientes idosos são um grupo de risco, pois apresentam um tempo mediano de sobrevida 37% menor que adultos com menos de 60 anos.

Vários estudos relataram um maior risco de morrer em idosos e pessoas com comorbidades, especialmente doenças cardiovasculares (JORDAN; ADAB; CHENG, 2020; RODRIGUEZ *et al.*, 2020), não diferindo dos achados do presente estudo. Dentre as limitações nesse estudo, são encontradas diversas informações ausentes, o que limita os dados existentes e dificulta o monitoramento do paciente. Outras limitações estão envolvidas na resposta do paciente ou no julgamento do profissional de saúde sem nenhum teste confirmatório adicional; por exemplo, a obesidade só foi objeto de interpretação do profissional de saúde, sendo considerada apenas em estágios avançados da mesma.

---

## Capítulo 6

---

---

# Considerações Finais

---

Este estudo avaliou os casos de internação pela COVID-19, notificados em 2020, no Brasil, utilizando-se dos dados disponibilizados pela SIVEP-Gripe. Através dos resultados, constatou-se que a maioria dos pacientes hospitalizados pela COVID-19 são do sexo masculino, as comorbidades mais frequentes foram doenças neurológicas e cardiovasculares, e os sintomas mais comuns foram tosse e dispneia.

Sendo o Brasil um país de dimensões continentais, optou-se por incluir a região do paciente entre os fatores associados e observou-se diferenças significativas, de modo que pacientes internados nas regiões Norte e Nordeste apresentam maior risco de óbito. Em relação as características dos pacientes, diversos resultados são similares aos encontrados em outros estudos, destacando-se o impacto da faixa etária, com idosos (acima de 60 anos) apresentando um tempo de sobrevida muito inferior aos adultos, efeito similar ao da presença de doença cardiovascular e/ou diabetes. A presença de doença renal também apareceu como um fator de risco de grande impacto neste trabalho. Vale ressaltar a importância das aplicações estratégicas voltadas ao anteparo da população vulnerável a COVID-19.

Para futuras investigações que acercam o Brasil seria muito interessante a atualização e inclusão de novos dados, como o andamento da vacinação e as complicações ocorridas durante esse tempo, além de realizar análises de sensibilidade para os dados perdidos. Com isso, os possíveis resultados e conclusões serão mais detalhados e teremos informações muito úteis a respeito dos fatores de risco e, conseqüentemente, os prognósticos de pacientes infectados pela COVID-19 no Brasil.

---

## Referências

---

- AKAIKE, H. A new look at the statistical model identification. *IEEE transactions on automatic control*, leee, v. 19, n. 6, p. 716–723, 1974.
- AZEREDO, C. A. C.; MACHADO, M. d. G. R. Fisioterapia respiratória moderna. In: *Fisioterapia respiratória moderna*. [S.l.: s.n.], 1999. p. xiii–325.
- BATISTA, J. L. F. *Função Weibull como modelo para a distribuição de diâmetros de espécies arbóreas tropicais*. Tese (Doutorado) — Universidade de São Paulo, 1989.
- BENNETT, S. Log-logistic regression models for survival data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, Wiley Online Library, v. 32, n. 2, p. 165–171, 1983.
- BOGOCH, I. I.; WATTS, A.; THOMAS-BACHLI, A.; HUBER, C.; KRAEMER, M. U.; KHAN, K. Pneumonia of unknown aetiology in wuhan, china: potential for international spread via commercial air travel. *Journal of travel medicine*, Oxford University Press, v. 27, n. 2, p. taaa008, 2020.
- BOZDOGAN, H. Model selection and akaike's information criterion (aic): The general theory and its analytical extensions. *Psychometrika*, Springer, v. 52, n. 3, p. 345–370, 1987.
- BROWN, G. W.; FLOOD, M. M. Tumbler mortality. *Journal of the American Statistical Association*, Taylor & Francis, v. 42, n. 240, p. 562–574, 1947.
- BRÜEDERL, J.; DIEKMANN, A. The log-logistic rate model: Two generalizations with an application to demographic data. *Sociological Methods & Research*, SAGE PERIODICALS PRESS, v. 24, n. 2, p. 158–186, 1995.
- BUSTAMANTE-TEIXEIRA, M. T.; FAERSTEIN, E.; LATORRE, M. d. R. Técnicas de análise de sobrevivência. *Cadernos de Saúde Pública*, SciELO Brasil, v. 18, n. 3, p. 579–594, 2002.
- CARVALHO, M. S.; ANDREOZZI, V. L.; CODEÇO, C. T.; CAMPOS, D. P.; BARBOSA, M. T. S.; SHIMAKURA, S. E. *Análise de sobrevivência: teoria e aplicações em saúde*. [S.l.]: SciELO-Editora FIOCRUZ, 2011.

- CHAIEB, L. L.; RIVEST, L.-P.; ABDOUS, B. Estimating survival under a dependent truncation. *Biometrika*, Oxford University Press, v. 93, n. 3, p. 655–669, 2006.
- CLUTTER, J. L.; FORTSON, J. C.; PIENAAR, L. V.; BRISTER, G. H.; BAILEY, R. L. *et al.* *Timber management: A quantitative approach*. [S.l.]: John Wiley & Sons, Inc., 1983.
- COLOSIMO, E. A.; GIOLO, S. R. *Análise de sobrevivência aplicada*. São Paulo: Editora Blucher, 2006.
- COLTON, T. *Statistica in medicina: uso delle ta-voles di sopravvivenza*. *Piccin Editore, Padova*, 1979.
- COVID, C.; TEAM, R.; COVID, C.; TEAM, R.; COVID, C.; TEAM, R.; BIALEK, S.; BOUNDY, E.; BOWEN, V.; CHOW, N. *et al.* Severe outcomes among patients with coronavirus disease 2019 (covid-19)—united states, february 12–march 16, 2020. *Morbidity and mortality weekly report*, Centers for Disease Control and Prevention, v. 69, n. 12, p. 343, 2020.
- DIEKMANN, A. Diffusion and survival models for the process of entry into marriage. *Journal of Mathematical Sociology*, Taylor & Francis, v. 14, n. 1, p. 31–44, 1989.
- DU, R.-H.; LIANG, L.-R.; YANG, C.-Q.; WANG, W.; CAO, T.-Z.; LI, M.; GUO, G.-Y.; DU, J.; ZHENG, C.-L.; ZHU, Q. *et al.* Predictors of mortality for patients with covid-19 pneumonia caused by sars-cov-2: a prospective cohort study. *European Respiratory Journal*, Eur Respiratory Soc, v. 55, n. 5, 2020.
- DUNN, P. K.; SMYTH, G. K. Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, Taylor & Francis, v. 5, n. 3, p. 236–244, 1996.
- FARADMAL, J.; SOLTANIAN, A. R.; ROSHANAIEI, G.; KHODABAKHSHI, R.; KASAEIAN, A. Comparison of the performance of log-logistic regression and artificial neural networks for predicting breast cancer relapse. *Asian Pacific Journal of Cancer Prevention*, Asian Pacific Journal of Cancer Prevention, v. 15, n. 14, p. 5883–5888, 2014.
- FISHER, R. A. Theory of statistical estimation. In: CAMBRIDGE UNIVERSITY PRESS. *Mathematical proceedings of the Cambridge philosophical society*. [S.l.], 1925. v. 22, n. 5, p. 700–725.
- FISHER, R. A. *Statistical methods and scientific inference*. Hafner Publishing Co., 1956.
- HALVATSIOTIS, P.; KOTANIDOU, A.; TZANNIS, K.; JAHAJ, E.; MAGIRA, E.; THEODORAKOPOULOU, M.; KONSTANDOPOULOU, G.; GKEKA, E.; POURZITAKI, C.; KAPRAVELOS, N. *et al.* Demographic and clinical features of critically ill patients with covid-19 in greece: The burden of diabetes and obesity. *Diabetes research and clinical practice*, Elsevier, v. 166, p. 108331, 2020.
- HUANG, C.; WANG, Y.; LI, X.; REN, L.; ZHAO, J.; HU, Y.; ZHANG, L.; FAN, G.; XU, J.; GU, X. *et al.* Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. *The lancet*, Elsevier, v. 395, n. 10223, p. 497–506, 2020.

- HUEDA-ZVALETA, M.; COPAJA-CORZO, C.; BARDALES-SILVA, F.; FLORES-PALACIOS, R.; BARRETO-ROCCHETTI, L.; BENITES-ZAPATA, V. A. Factors associated with mortality due to covid-19 in patients from a public hospital in tacna, peru. *Revista Peruana de Medicina Experimental y Salud Pública*, SciELO Public Health, v. 38, p. 214–223, 2021.
- HUNG, H.-F.; WANG, Y.-C.; YEN, A. M.-F.; CHEN, H.-H. Stochastic model for hepatitis b virus infection through maternal (vertical) and environmental (horizontal) transmission with applications to basic reproductive number estimation and economic appraisal of preventive strategies. *Stochastic Environmental Research and Risk Assessment*, Springer, v. 28, n. 3, p. 611–625, 2014.
- JORDAN, R. E.; ADAB, P.; CHENG, K. *Covid-19: risk factors for severe disease and death*. [S.l.]: British Medical Journal Publishing Group, 2020.
- JR, D. W. H.; LEMESHOW, S.; STURDIVANT, R. X. *Applied logistic regression*. [S.l.]: John Wiley & Sons, 2013. v. 398.
- KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, Taylor & Francis, v. 53, n. 282, p. 457–481, 1958.
- KLEINBAUM, D. G. Km. survival analysis: A self-learning text. *Statistics for biology and health*, 2005.
- LARRY, W. *All of statistics: a concise course in statistical inference*. [S.l.]: Springer, 2004.
- LI, L.; WU, T.; FENG, C. Model diagnostics for censored regression via randomized survival probabilities. *Statistics in Medicine*, Wiley Online Library, v. 40, n. 6, p. 1482–1497, 2021.
- LI, Q.; GUAN, X.; WU, P.; WANG, X.; ZHOU, L.; TONG, Y.; REN, R.; LEUNG, K. S.; LAU, E. H.; WONG, J. Y. *et al.* Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia. *New England journal of medicine*, Mass Medical Soc, 2020.
- LINDSEY, J. K. Generalized linear modelling. *Applying generalized linear models*, Springer, p. 1–26, 1997.
- MANTEL, N. *et al.* Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep*, v. 50, n. 3, p. 163–170, 1966.
- MCARTHUR, J. C.; HOOVER, D.; BACELLAR, H.; MILLER, E.; COHEN, B.; BECKER, J.; GRAHAM, N.; MCARTHUR, J.; SELNES, O.; JACOBSON, L. *et al.* Dementia in aids patients: incidence and risk factors. *Neurology*, AAN Enterprises, v. 43, n. 11, p. 2245–2245, 1993.
- MEEKER, W. Q.; ESCOBAR, L. A.; LU, C. J. Accelerated degradation tests: modeling and analysis. *Technometrics*, Taylor & Francis, v. 40, n. 2, p. 89–99, 1998.

- MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. *Introduction to linear regression analysis*. [S.l.]: John Wiley & Sons, 2021.
- MOWLA, S. G. M.; AZAD, K. A. K.; KABIR, A.; BISWAS, S.; ISLAM, M. R.; BANIK, G. C.; KHAN, M. M. H.; ROHAN, K. I.; ALAM, M. A. Clinical profile of 100 confirmed covid-19 patients admitted in dhaka medical college hospital, dhaka, bangladesh. *Journal of Bangladesh College of Physicians and Surgeons*, p. 29–36, 2020.
- MUSE, A. H.; TOLBA, A. H.; FAYAD, E.; ALI, O. A. A.; NAGY, M.; YUSUF, M. Modelling the covid-19 mortality rate with a new versatile modification of the log-logistic distribution. *Computational Intelligence and Neuroscience*, Hindawi, v. 2021, 2021.
- NEEDHAM, D. M. Mobilizing patients in the intensive care unit: improving neuromuscular weakness and physical function. *Jama*, American Medical Association, v. 300, n. 14, p. 1685–1690, 2008.
- NELDER, J. A.; WEDDERBURN, R. W. Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, Wiley Online Library, v. 135, n. 3, p. 370–384, 1972.
- NELSON, W. Theory and applications of hazard plotting for censored failure data. *Technometrics*, Taylor & Francis, v. 14, n. 4, p. 945–966, 1972.
- NELSON, W. Accelerated life testing: statistical models, data analysis and test plans. *Accelerated life testing: statistical models data analysis and test plants*, v. 41, 1990.
- ORGANIZATION, W. H. *et al.* *Coronavirus disease 2019 (COVID-19) situation report-51*. 2020.
- ORGANIZATION, W. H. *et al.* *Coronavirus disease 2019 (covid-19): situation report, 73*. World Health Organization, 2020.
- POLLARD, J. H.; VALKOVICS, E. J. The gompertz distribution and its applications. *Genus*, JSTOR, p. 15–28, 1992.
- PRENTICE, R. L.; KALBFLEISCH, J. D.; JR, A. V. P.; FLOURNOY, N.; FAREWELL, V. T.; BRESLOW, N. E. The analysis of failure times in the presence of competing risks. *Biometrics*, JSTOR, p. 541–554, 1978.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2020.
- RODRIGUEZ, C. S. E.; TOVAR, A. R. L.; JARAMILLO, G. A. U.; RODRÍGUEZ, E. A. B.; PINILLA, E. A. I.; ROJAS, J. N. M. Supervivencia y mortalidad por covid-19 en bogotá, colombia durante marzo y julio de 2020. *Research, Society and Development*, v. 9, n. 11, p. e81291110049–e81291110049, 2020.
- SALINAS-AGUIRRE, J.; SÁNCHEZ-GARCÍA, C.; RODRÍGUEZ-SANCHEZ, R.; RODRÍGUEZ-MUÑOZ, L.; DÍAZ-CASTAÑO, A.; BERNAL-GÓMEZ, R. Características clínicas y comorbilidades asociadas a mortalidad en pacientes con covid-19 en coahuila (méxico). *Revista Clínica Española*, Elsevier, 2021.

- SCHWARZ, G. Estimating the dimension of a model. *The annals of statistics*, JSTOR, p. 461–464, 1978.
- SHI, Q.; ZHANG, X.; JIANG, F.; ZHANG, X.; HU, N.; BIMU, C.; FENG, J.; YAN, S.; GUAN, Y.; XU, D. *et al.* Clinical characteristics and risk factors for mortality of covid-19 patients with diabetes in wuhan, china: a two-center, retrospective study. *Diabetes care*, Am Diabetes Assoc, v. 43, n. 7, p. 1382–1391, 2020.
- SINGH, K. P.; LEE, C. M.-S.; GEORGE, E. O. On generalized log-logistic model for censored survival data. *Biometrical Journal*, Wiley Online Library, v. 30, n. 7, p. 843–850, 1988.
- SIVEP-GRIFE. 2021. <<https://opendatasus.saude.gov.br/dataset/bd-srag-2021>>. Acessado em 18/01/2021.
- SOUSA, G.; GARCES, T.; CESTARI, V.; FLORÊNCIO, R.; MOREIRA, T.; PEREIRA, M. Mortality and survival of covid-19. *Epidemiology & Infection*, Cambridge University Press, v. 148, 2020.
- STACY, E. W.; MIHRAM, G. A. Parameter estimation for a generalized gamma distribution. *Technometrics*, Taylor & Francis, v. 7, n. 3, p. 349–358, 1965.
- SUE-LING, H.; JOHNSTON, D.; MARTIN, I.; DIXON, M.; LANSDOWN, M.; MCMAHON, M.; AXON, A. Gastric cancer: a curable disease in britain. *British Medical Journal*, British Medical Journal Publishing Group, v. 307, n. 6904, p. 591–596, 1993.
- SURVEILLANCES, V. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (covid-19)—china, 2020. *China CDC weekly*, v. 2, n. 8, p. 113–122, 2020.
- SWEET, T. M. *A review of statistical rethinking: a Bayesian course with examples in R and Stan*. [S.l.]: SAGE Publications Sage CA: Los Angeles, CA, 2017.
- TAN, W.-Y.; JR, R. H. B. A stochastic model of the hiv epidemic and the hiv infection distribution in a homosexual population. *Mathematical Biosciences*, Elsevier, v. 113, n. 1, p. 115–143, 1993.
- TEAM, E. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (covid-19)—china, 2020. *China CDC weekly*, Chinese Center for Disease Control and Prevention, v. 2, n. 8, p. 113, 2020.
- TREVISOL, D. J.; BIANCHI, B. R.; SAKAE, T. M.; VINHOLES, D. B.; TREVISOL, F. S. Análise de sobrevida em pacientes submetidos à angioplastia coronariana com stent em um hospital da região sul de santa catarina. *Scientia Medica (Porto Alegre)*, v. 22, n. 2, p. 91–96, 2012.
- TV, R. *O HOSPITAL: Luta pela vida na pandemia de coronavírus*. São Paulo: [s.n.], 2021. <<https://www.youtube.com/watch?v=TGm1HFfF5E0>>. Acessado em 09/11/2021.
- VERHULST, P.-F. Notice sur la loi que la population suit dans son accroissement. *Corresp. Math. Phys.*, v. 10, p. 113–126, 1838.

VIANA, M. B.; MURAO, M.; RAMOS, G.; OLIVEIRA, H. M.; CARVALHO, R. D.; BASTOS, M. D.; COLOSIMO, E. A.; SILVESTRINI, W. S. Malnutrition as a prognostic factor in lymphoblastic leukaemia: a multivariate analysis. *Archives of disease in childhood*, BMJ Publishing Group Ltd, v. 71, n. 4, p. 304–310, 1994.

WILLIAMS, D. Generalized linear model diagnostics using the deviance and single case deletions. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, Wiley Online Library, v. 36, n. 2, p. 181–191, 1987.

YAO, Q.; WANG, P.; WANG, X.; QIE, G.; MENG, M.; TONG, X.; BAI, X.; DING, M.; LIU, W.; LIU, K. *et al.* Retrospective study of risk factors for severe sars-cov-2 infections in hospitalized adult patients. *Pol Arch Intern Med*, v. 130, n. 5, p. 390–399, 2020.

ZHANG, L.; HAN, C.; ZHANG, S.; DUAN, C.; SHANG, H.; BAI, T.; HOU, X. Diarrhea and altered inflammatory cytokine pattern in severe coronavirus disease 2019: impact on disease course and in-hospital mortality. *Journal of gastroenterology and hepatology*, Wiley Online Library, v. 36, n. 2, p. 421–429, 2021.

# Anexos

**ANEXO A**

---

**ANEXO**

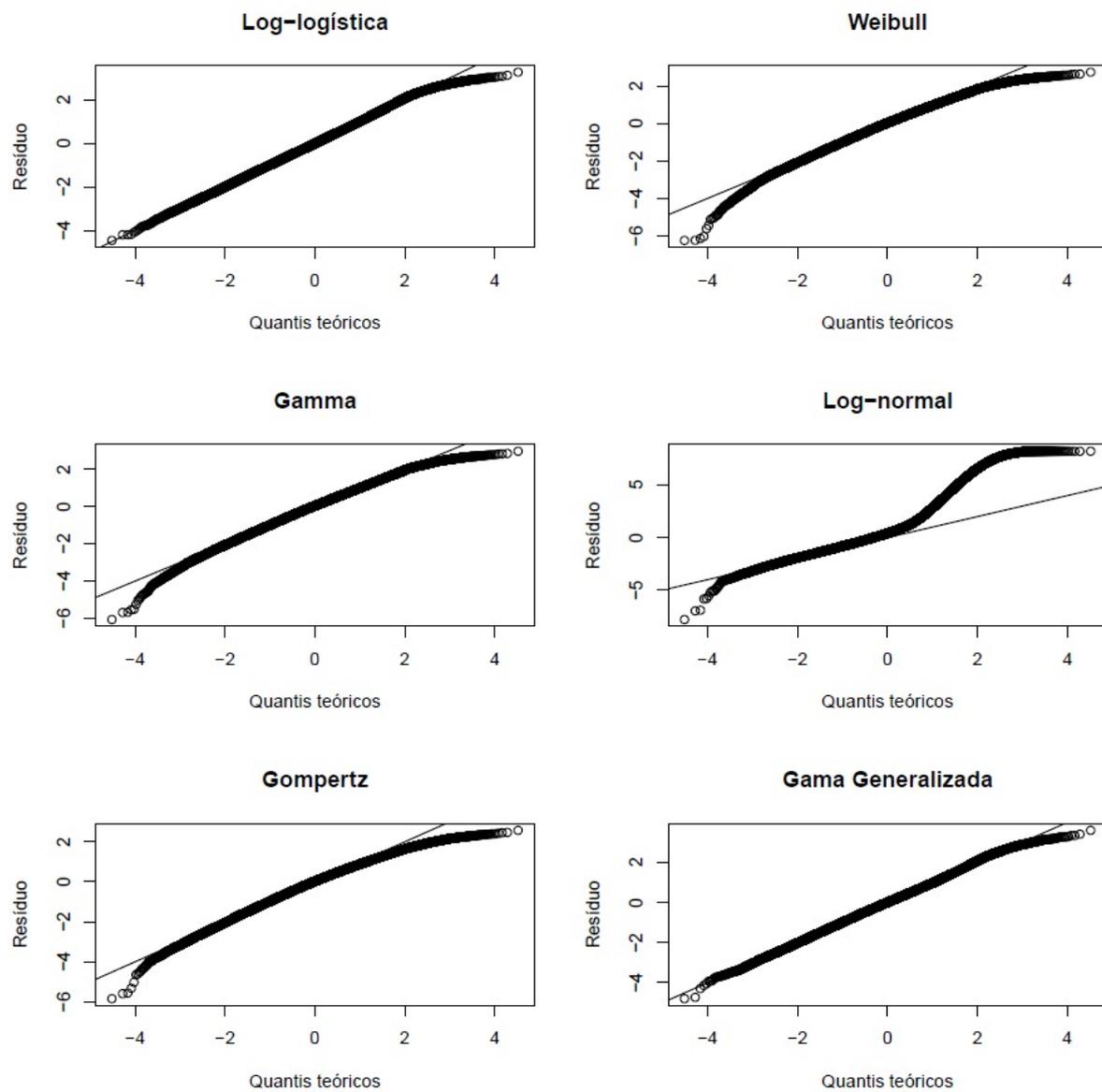


Figura 3 – Gráficos quantil-quantil dos resíduos NRSP dos modelos múltiplos, ajustados aos tempos de sobrevivência dos pacientes acometidos pela Covid-19.

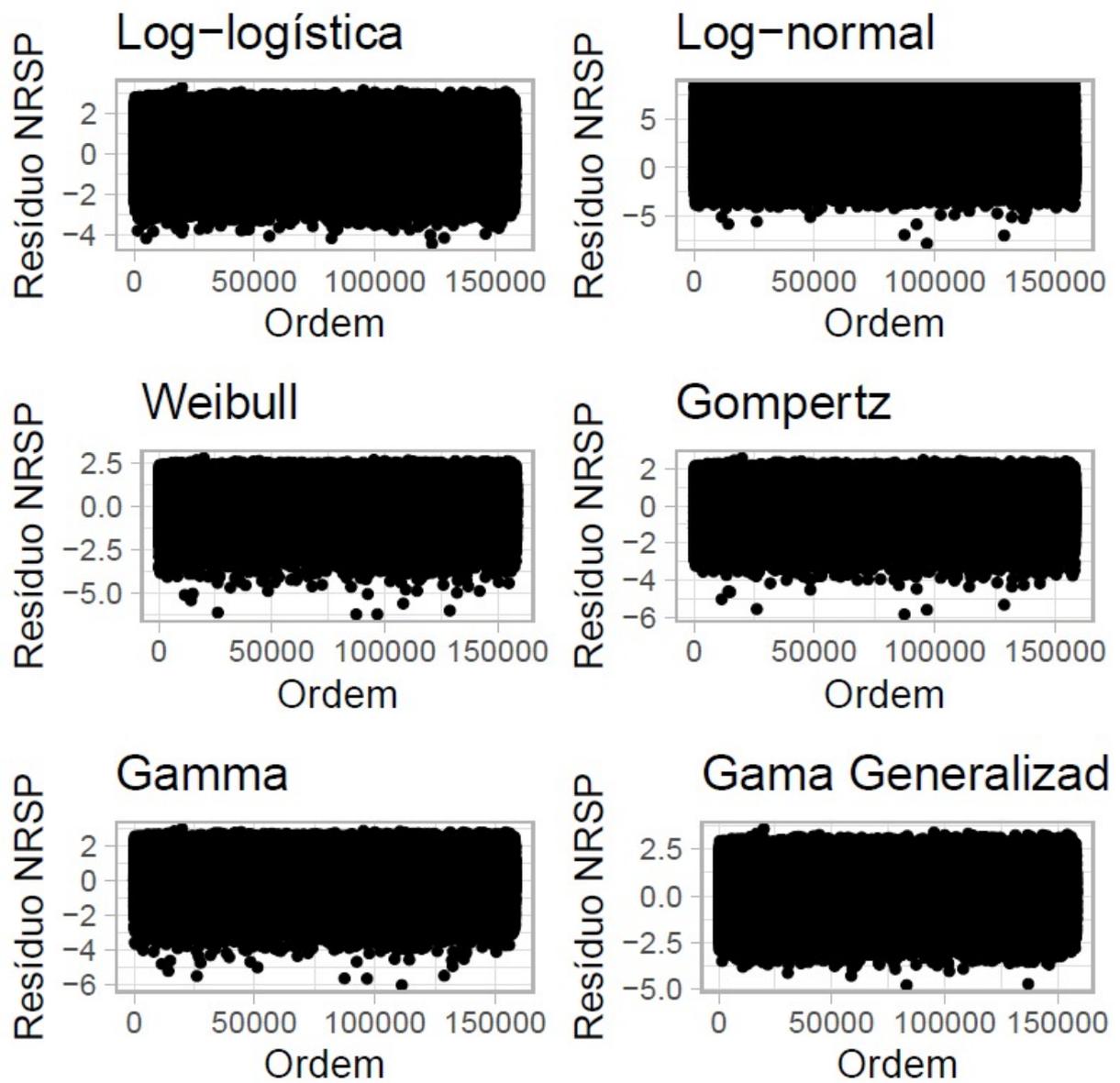


Figura 4 – Gráficos Resíduos vs Ordem dos resíduos NRSP dos modelos múltiplos, ajustados aos tempos de sobrevivência dos pacientes acometidos pela Covid-19.