



**ARTHUR OLIVEIRA COSTA**

**MODELOS DE FRAGILIDADE E MULTIESTADO NA  
AVALIAÇÃO DO TEMPO DE INTERNAMENTOS  
HOSPITALARES POR COVID-19 EM MINAS GERAIS**

**LAVRAS – MG**

**2021**

**ARTHUR OLIVEIRA COSTA**

**MODELOS DE FRAGILIDADE E MULTIESTADO NA AVALIAÇÃO DO TEMPO DE  
INTERNAMENTOS HOSPITALARES POR COVID-19 EM MINAS GERAIS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

Profa. Dra. Carla Regina Guimarães Brighenti

Orientadora

**LAVRAS – MG**

**2021**

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca  
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Costa, Arthur Oliveira.

Modelos de Fragilidade e Multiestado na Avaliação do Tempo  
de Internamentos Hospitalares por Covid-19 em Minas Gerais /  
Arthur Oliveira Costa. - 2021.

51 p. : il.

Orientador(a): Carla Regina Guimarães Brighenti.

Dissertação (mestrado acadêmico) - Universidade Federal de  
Lavras, 2021.

Bibliografia.

1. Análise de Sobrevivência. 2. Modelos Multiestado. 3.  
Coronavírus. I. Brighenti, Carla Regina Guimarães. II. Título.

**ARTHUR OLIVEIRA COSTA**

**MODELOS DE FRAGILIDADE E MULTIESTADO NA AVALIAÇÃO DO TEMPO DE  
INTERNAMENTOS HOSPITALARES POR COVID-19 EM MINAS GERAIS  
MODELS OF FRAGILITIES IN THE EVALUATION OF THE TIME OF HOSPITAL  
STAYS BY COVID-19 IN MINAS GERAIS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

APROVADA em 26 de fevereiro de 2021.

Prof. Dr. Tiago Almeida de Oliveira UEPB

Prof. Dr. Renato Ribeiro de Lima UFLA

Profa. Dra. Carla Regina Guimarães Brighenti  
Orientadora

**LAVRAS – MG  
2021**

*Dedico este trabalho a minha mãe, Maria do Socorro, ao meu pai Josenildo Costa, por me encorajarem a conquistar os meus sonhos e que sempre me apoiaram em todos os momentos difíceis.*

## **AGRADECIMENTOS**

A Deus, por tudo.

A Universidade Federal de Lavras e ao departamento de Estatística, juntamente com seus docentes e funcionários.

A professora Carla Regina Guimarães Brighenti, pelas horas de dedicação, comprometimento e ensinamentos, por acreditar e tornar possível este trabalho, por ser minha orientadora, muito obrigado.

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

Enfim, a todos que fazem parte da minha vida e torcem pelo meu sucesso.

*“Tudo o que um sonho precisa para ser  
realizado é alguém que acredite que  
ele possa ser realizado.”  
(Roberto Shinyashiki)*

## RESUMO

A partir de dezembro de 2019, vários pacientes foram internados com pneumonia de causa desconhecida e descobriram que a infecção era causada por um novo tipo de vírus da família dos Coronavírus, denominado de SARS-CoV-2 e a doença causada por ele de Covid-19 (*Coronavirus 19 disease*), o que provocou uma pandemia. Desde então, várias pesquisas e bancos de dados têm sido disponibilizados e análises estatísticas têm sido propostas, visando tanto aspectos clínicos quanto administrativos, devido ao aumento da demanda por internamentos. Neste trabalho objetivou-se ajustar modelos de análise de sobrevivência incluindo efeito de fragilidade e modelos multiestado ao tempo de internamentos de pacientes portadores de Covid-19 até sua recuperação, no Estado de Minas Gerais. Os dados foram obtidos dos arquivos do Sistema de Informações Hospitalares do Sistema Único de Saúde, nos meses de março a outubro de 2020, no Estado de Minas Gerais. Foram obtidas variáveis associadas a cada um dos 24946 pacientes diagnosticados com Covid-19. Modelos de fragilidade foram ajustados para analisar o tempo de internamento considerando a gestão pública ou privada dos hospitais, considerando as covariáveis sexo, idade, ocupação em unidade de terapia intensiva (UTI) e presença de diagnóstico secundário, sendo o óbito considerado um dado censurado. Dos internamentos dos pacientes portadores do Covid-19, 53,54% eram do sexo masculino, 33,81% necessitaram de UTI, 8,86% dos pacientes tiveram pelo menos um diagnóstico secundário, sendo a média de idade dos pacientes igual a 61,35 anos. No caso dos pacientes que foram a óbito a média de idade aumentou para de 70,29 anos. Considerando o modelo de fragilidade ajustado pela a distribuição log-normal, estimou-se que o risco de recuperação (alta hospitalar) diminui 1,33% a cada acréscimo de um ano a idade do paciente. Pacientes com maiores tempos de internamentos foram aqueles em que houve presença de diagnóstico secundário, utilização da UTI e do sexo masculino. No modelo multiestado observou-se que a variável idade todas as transições foram significativas. Além disso, pacientes que foram atendidos em hospitais de gestão pública tiveram tempos de internamentos maiores do que pacientes que foram atendidos em hospitais de gestão privada. Portanto, o estudo de modelos de fragilidades e multiestado na avaliação do tempo de internamentos hospitalares, tornam-se um recurso auxiliar importante para os órgãos competentes.

**Palavras-chave:** Hospitalização. Infecções por Coronavírus. Unidade de Terapia Intensiva. Administração Hospitalar. Análise de Sobrevida.



## ABSTRACT

Since December, 2019, many patients were interned due a pneumonia of unknown cause and discovered that the infection was caused by a new virus type from the family Crown virus, denominated SARS-CoV-2 and the disease originated by the virus was called Covid-19, the one which caused the pandemic. Since then, various researches and data basis have been made available and the statistic analysis have been proposed, focusing either on the clinic and the administrative aspects due the internment demand raises. Therefore on this paper, it was an objective to adjust models of surviving analysis including the fragility effect and multi-state models on the internment time of patients carriers of Covid-19, in the state of Minas Gerais. The obtained data variables were associated to each of the 24.946 patients diagnosed on Covid-19. Fragility models were adjusted to analyze the internment time considering the public and private management of the hospitals, considering the public or private management of hospitals, besides the covariates of gender, age, necessity of hospitalization in Intensive Care Unit (ICU) and presence of secondary diagnosis, the death was considered a censored data. From the hospitalized patients carriers of Covid-19, 53,54% were male, 33,81% needed the ICU, 8,86% of the patients had at least a secondary diagnosis, being the patient's average age equal to 61,35 years. In the case of patients that died the average age raised to 70,29 years. Considering the fragility model adjusted to the log-normal distribution, it was estimated that the recovering risk decreases 1,33% to each year raise on the patient's age. Patients with more hospitalization times were the ones in which there was the presence of secondary diagnosis, necessity of ICU and they were from the male gender. In the multi-state model, it was observed that the variable 'age', all the transitions were significant. Besides that, patients which were treated in public hospitals had longer internment times than the patients that were treated private hospitals. Therefore, the fragility and multi-state models study in the evaluation of the hospital internment times, turns to be an important auxiliary resource to the competent organs.

**Keywords:** Hospitalization. Coronavirus infections. Intensive care unit. Hospital administration. Survival Analysis.

## LISTA DE FIGURAS

Figura 2.1 – Esquema do modelo multiestado “enfermidade-morte”. . . . .	29
Figura 3.1 – Esquema do modelo multiestado “enfermidade-morte”, para os casos de internamentos hospitalares por Covid-19 no Estado. . . . .	35
Figura 4.1 – Mapa de intervalos de classes para o número de internamentos hospitalares por Covid-19 no Estado da Minas Gerais no período de março a outubro de 2020. . . . .	39
Figura 4.2 – Mapa de quintis para o número de internamentos hospitalares por Covid-19 no Estado da Minas Gerais no período de março a outubro de 2020. . . . .	39
Figura 4.3 – Curvas de sobrevivência estimada pelo método de <i>Kaplan-Meier</i> para a variável SEXO. . . . .	41
Figura 4.4 – Curvas de sobrevivência estimada pelo método de <i>Kaplan-Meier</i> para a variável UTI. . . . .	42
Figura 4.5 – Curvas de sobrevivência estimada pelo método de <i>Kaplan-Meier</i> para a variável DIAGSEC. . . . .	42
Figura 4.6 – Curvas de sobrevivência estimada pelo método de <i>Kaplan-Meier</i> para a variável GESTÃO. . . . .	43
Figura 4.7 – Curvas de sobrevivência estimadas para pacientes diagnosticado com Covid-19 das gestões públicas e privadas em Minas Gerais no período de março a outubro de 2020. . . . .	45
Figura 4.8 – Representação do modelo multiestado para as frequências associadas a cada transição. . . . .	47
Figura 4.9 – Representação do modelo multiestado para as probabilidades associadas a cada transição. . . . .	47

## LISTA DE TABELAS

Tabela 3.1 – Códigos de transição na análise de modelo de multiestado. . . . .	36
Tabela 4.1 – Municípios com maior número de internamentos hospitalares por Covid-19 no Estado de Minas Gerais no período de março a outubro de 2020. . . . .	38
Tabela 4.2 – Médias de idade, frequências e porcentagens dos dados de internamentos hospitalares por Covid-19 em Minas Gerais no período de março a outubro de 2020. . . . .	41
Tabela 4.3 – Modelos de Cox sem fragilidade, fragilidade semi-paramétrico log-normal, fragilidade semi-paramétrico gama e os respectivos AIC, TRV e variância.	43
Tabela 4.4 – Estimativas, teste da razão de verossimilhança (TRV), AIC, razão de risco (RR) e teste associado à fragilidade obtidos para o modelo semi-paramétrico de fragilidade log-normal ajustado aos dados de internamentos hospitalares por Covid-19 em Minas Gerais no período de março a outubro de 2020. . .	44
Tabela 4.5 – Modelo de fragilidade paramétrica gama com diferentes valores de AIC para cada distribuição. . . . .	45
Tabela 4.6 – Estimativas e teste associado à fragilidade obtidos para o modelo paramétrico de fragilidade Gama e risco de base com distribuição t ajustado aos dados de internamentos hospitalares por Covid-19 em Minas Gerais no período de março a outubro de 2020. . . . .	46
Tabela 4.7 – Frequências das transições na análise de modelo de multiestado. . . . .	46
Tabela 4.8 – Estimativas, razão de risco (RR), erro padrão(EP) da estimativa e o valor crítico de Z obtidos para o modelo semi-paramétrico de Cox ajustado aos dados de internamentos hospitalares por Covid-19 em Minas Gerais no período de março a outubro de 2020. . . . .	48

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>10</b>
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>	<b>13</b>
<b>2.1</b>	<b>A Covid-19 e seus internamentos</b>	<b>13</b>
<b>2.2</b>	<b>Análise de Sobrevivência</b>	<b>14</b>
<b>2.3</b>	<b>Modelos paramétricos</b>	<b>17</b>
<b>2.4</b>	<b>Modelo de regressão de Cox</b>	<b>19</b>
<b>2.5</b>	<b>Modelos com fragilidade</b>	<b>20</b>
<b>2.5.1</b>	<b>Distribuições de probabilidade para fragilidade</b>	<b>22</b>
<b>2.5.2</b>	<b>Fragilidade semi-paramétrica</b>	<b>25</b>
<b>2.6</b>	<b>Modelos Multiestado</b>	<b>28</b>
<b>2.7</b>	<b>Razões de riscos</b>	<b>29</b>
<b>3</b>	<b>MÉTODOS</b>	<b>31</b>
<b>3.1</b>	<b>Caracterização dos leitos hospitalares no Estado de Minas Gerais</b>	<b>31</b>
<b>3.2</b>	<b>Fontes dos dados</b>	<b>31</b>
<b>3.3</b>	<b>Análise descritiva</b>	<b>32</b>
<b>3.4</b>	<b>Ajuste de modelos semi-paramétricos com fragilidade</b>	<b>33</b>
<b>3.5</b>	<b>Ajuste de modelos paramétricos com fragilidade</b>	<b>35</b>
<b>3.6</b>	<b>Ajuste de modelos multiestado</b>	<b>35</b>
<b>3.7</b>	<b>Diagnóstico dos modelos de fragilidade</b>	<b>36</b>
<b>3.8</b>	<b>Uso de Pacotes para Análise Estatística</b>	<b>37</b>
<b>4</b>	<b>RESULTADOS E DISCUSSÃO</b>	<b>38</b>
<b>4.1</b>	<b>Análise descritiva</b>	<b>38</b>
<b>4.2</b>	<b>Modelos Semi-paramétricos</b>	<b>43</b>
<b>4.3</b>	<b>Modelos de fragilidades paramétricos</b>	<b>45</b>
<b>4.4</b>	<b>Modelo multiestado</b>	<b>46</b>
<b>5</b>	<b>CONCLUSÃO</b>	<b>49</b>
	<b>REFERÊNCIAS</b>	<b>50</b>

## 1 INTRODUÇÃO

Quando se deseja avaliar o tempo de internamento até a recuperação e ocorre o óbito tem-se um dado incompleto, ou seja, uma observação parcial da resposta, pois o evento de interesse não foi atingido, obtendo-se neste caso um dado de “sobrevivência” censurado (COLOSIMO; GIOLO, 2006).

Por dados de sobrevivência referimo-nos a todo o tipo de dados que resulta da observação de tempos desde um início bem especificado até à ocorrência de um evento de interesse (FERNANDES, 2010). A análise de sobrevivência abrange uma classe de modelos e métodos estatísticos destinados à avaliar esses dados.

Em análise de sobrevivência, existem os modelos semi-paramétricos, também denominados simplesmente de modelo de regressão de Cox (COX, 1972), que tem sido bastante utilizado em estudos clínicos. Além da flexibilidade, este modelo permite incorporar facilmente covariáveis dependentes do tempo, que ocorrem com frequência em várias áreas de aplicação (COLOSIMO; GIOLO, 2006). Dessa forma, é possível explicar o risco do indivíduo falhar sob o efeito de covariáveis.

Em alguns casos, a suposição de independência é violada devido ao agrupamento, ou seja, a presença de grupos de indivíduos quem têm fatores de riscos não observáveis em comum. Exemplos são os ensaios clínicos multicêntricos, onde pacientes de um mesmo hospital têm um risco possivelmente mais similares em relação a pacientes de outro hospital, devido a fatores sociais, climáticos e geográficos, por exemplo. Pode-se então relacionar o tipo de administração do hospital, se público ou privado, a um risco dependente, por exemplo, a fatores sociais. Quando o modelo visa, além de avaliar os riscos de falha, descrever a existência de uma variável aleatória não observada que agrupa determinados indivíduos, o denominamos de modelo de fragilidade (COLOSIMO; GIOLO, 2006). Na forma mais simples, o efeito aleatório (fragilidade) do grupo possui um efeito na função de risco para todos os indivíduos no mesmo grupo (KLEIN; MOESCHBERGER, 1997).

Outra situação é a mudança de estado de uma variável que dá origem então a multiestados. Os modelos multiestado podem ser considerados como uma generalização da análise de sobrevivência clássica onde além do evento de interesse, os estados intermediários são identificados. Dessa forma é observada uma sequência de eventos, originando mais do que uma observação por indivíduo. Esta modelagem oferece uma ferramenta flexível para o estudo dos efeitos de variáveis preditoras (covariáveis) para as transições entre diferentes estados (FER-

NANDES, 2010). Este é o caso, por exemplo, de pacientes de Covid-19, doença infecciosa causada pelo vírus SARS-CoV-2, em que os casos mais leves da Covid-19 podem progredir da internação em leito hospitalar diretamente para a alta, sem passar pela internação em unidades de terapia intensiva (UTI) ou mesmo vir a óbito sem chegar a ter internação em UTI.

Diante do exposto, estimar o tempo de internação e de permanência na UTI é uma necessidade e uma estratégia para melhorar a eficiência no setor hospitalar. A constante pressão exercida sobre as equipes de UTIs por fontes pagadoras é uma realidade inegável. Além disso, estimar o tempo de internação em função de diferentes variáveis pode melhorar o diálogo com pacientes e familiares, modulando assim as suas expectativas<sup>1</sup>.

---

<sup>1</sup> [https://www.epimedolutions.com/material\\_educativo/a-importancia-da-estimativa-do-tempo-de-permanencia-na-uti/](https://www.epimedolutions.com/material_educativo/a-importancia-da-estimativa-do-tempo-de-permanencia-na-uti/)

**Objetivo Principal**

Ajustar modelos de análise de sobrevivência incluindo efeito de fragilidade e modelo multiestado no tempo de internação de pacientes portadores de Covid-19, até sua recuperação.

**Objetivos Secundários**

- Verificar a existência de associação no tempo de permanência de pacientes que foram atendidos em hospitais de gestão pública ou privada;
- Avaliar variáveis como idade, sexo e utilização de UTI nos modelos ajustados, verificando sua adequação;
- Modelar o tempo de internação considerando as mudanças ocorridas durante a permanência no hospital considerando o tempo de usos de leitos e UTI.

## 2 REFERENCIAL TEÓRICO

### 2.1 A Covid-19 e seus internamentos

No final de dezembro de 2019, hospitais em *Wuhan*, China, identificaram vários pacientes com pneumonia de causa desconhecida (JIANG et al., 2020). Identificou-se que a infecção foi causada por um novo tipo de vírus da família dos Coronavírus, sendo esse vírus nomeado de SARS-CoV-2 e a doença causada por ele de Covid-19 (*Coronavirus 19 disease*)<sup>2</sup>.

A transmissão do vírus se dá por contato com gotículas originárias das vias respiratórias de pessoas infectadas, geradas ao falar, tossir ou espirrar. Além disso, o SARS-CoV-2 pode ser transmitido quando uma pessoa sadia toca objetos ou superfícies contaminadas e, assim, toca em mucosas (olhos, nariz ou boca) (BAI et al., 2020);(LAI et al., 2020).

Considerando os elevados níveis de transmissão, no dia 11 de março de 2020 a Organização Mundial da Saúde (OMS) caracterizou a Covid-19 como uma pandemia devido à sua rápida disseminação por países como Itália, Espanha e, posteriormente Estados Unidos. No Brasil, em 26 de fevereiro de 2020, foi confirmado o primeiro caso de Covid-19 no Estado de São Paulo e o primeiro óbito em 17 de março de 2020, no mesmo Estado. Nos meses de abril, maio e junho, o número de casos e óbitos aumentou exponencialmente e, então, em 20 de junho de 2020, o Brasil era o segundo país do mundo com o maior número de casos confirmados, com mais de 1 milhão de casos, e também o segundo país com mais mortes confirmadas, cerca de 50 mil<sup>3</sup>.

Nesse contexto, Minas Gerais, segundo Estado mais populoso do país, se destacou inicialmente por apresentar uma situação aparentemente controlada. Enquanto os estados vizinhos da região Sudeste acumularam mais de 300 mil casos e quase 22 mil óbitos pela Covid-19, Minas Gerais foi um dos Estados com menor taxa de contaminado por 100 mil habitantes<sup>4</sup>.

Os estudos relativos ao tempo de internamentos são usualmente descritivos. No entanto, a idade dos pacientes, sexo ou o diagnóstico de outras doenças podem afetar fortemente a duração da permanência dos mesmos nos hospitais. Além disso, a necessidade de utilizar uma unidade de terapia intensiva (UTI) pode prolongar a permanência do paciente no hospital, ou estar relacionada a ocorrência de óbito. Esses pacientes vão para a UTI quando seu quadro respiratório se agrava, ou seja na prática, quando a frequência respiratória fica muito alta e a

<sup>2</sup> <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>

<sup>3</sup> <http://saude.gov.br/images/pdf/2020/June/25/Boletim-epidemiologico-COVID-19-2.pdf>

<sup>4</sup> <http://saude.gov.br/images/pdf/2020/June/25/Boletim-epidemiologico-COVID-19-2.pdf>



saturação de oxigênio muito baixa ou quando a infecção do pulmão causada pelo Coronavírus desencadeia problemas cardiovasculares.

Machado, Martins e Leite (2016) analisaram se a mortalidade hospitalar ajustada varia segundo fonte de pagamento das internações, natureza jurídica e arranjo de financiamento dos hospitais, aplicando regressão logística multinível. Estes autores observaram que as maiores chances de óbito foram em pacientes do sexo masculino, em faixas etárias mais avançadas, com comorbidade, que utilizaram unidade de terapia intensiva, e tinham o Sistema Único de Saúde (SUS) como fonte de pagamento.

De fevereiro a junho de 2020, o total de leitos de UTI disponíveis no Brasil aumentou cerca de 20 mil unidades. Atualmente, o País conta 66,7 mil leitos desse tipo, ou seja, quase 45% a mais do que no início do ano.

Em fevereiro de 2020, o Cadastro Nacional de Estabelecimentos de Saúde apontava no Brasil que, das 46 mil unidades de UTI, metade delas estava disponível para o SUS, e a outra metade era reservada à saúde privada ou suplementar (planos de saúde), que atende a aproximadamente 22% da população.

Isto reflete uma desigualdade no acesso aos leitos de UTI disponíveis entre os serviços públicos e privados. Se considerados os dois segmentos, a quantidade de leitos de UTI representa, em média, no Brasil, atualmente 2,2 leitos para cada grupo de 10 mil habitantes. Proporcionalmente, no entanto, o SUS conta apenas com 1,1 leitos de UTI para cada grupo de 10 mil habitantes, enquanto a rede “não SUS” tem 5 leitos para cada 10 mil beneficiários de planos de saúde. Dessa forma, um paciente particular ou de plano de saúde teria cinco vezes mais chances de obter um leito desse tipo do que um que depende exclusivamente da assistência do SUS <sup>5</sup>.

O tema é uma necessidade do setor para identificar pacientes de alto risco de longa permanência e implementar planos de cuidados específicos para estes ou estimar e planejar a capacidade de atendimento e fluxo da UTI (vagas para cirurgias eletivas e transferências de pacientes).

## **2.2 Análise de Sobrevida**

A Análise de Sobrevida reúne um conjunto de técnicas e métodos estatísticos úteis na análise do tempo de vida de indivíduos em que, geralmente, a variável resposta é o tempo até

---

<sup>5</sup> <https://medicinasa.com.br/numero-leitos-uti/>

a ocorrência de um evento de interesse. Estes eventos são chamados de falha. O tempo de falha é o tempo decorrido a partir de um instante inicial até a ocorrência do evento de interesse. Uma falha pode ser a morte de um indivíduo no estudo ou uma remissão da doença (COLOSIMO; GIOLO, 2006).

Existem casos em que os tempos de sobrevivência não são observados, isso é, indivíduos podem não ser acompanhados até a ocorrência de um evento de interesse, no período de observação. Esta característica dos dados de sobrevivência é denominada censura e diferencia a Análise de Sobrevivência dos demais campos da Estatística (HOSMER; LEMESHOW; MAY, 2008). Censura é a observação parcial da resposta que foi interrompida por algum motivo, impedindo a observação do tempo de falha do indivíduo (COLOSIMO; GIOLO, 2006).

Existem três tipos de censura em estudos de sobrevivência (CARVALHO et al., 2011):

- Censura à direita: ocorre quando o verdadeiro tempo entre o início e o evento excede o tempo observado. Ignorar esse aspecto pode levar a superestimação do risco do evento.
- Censura à esquerda: ocorre quando o exato momento de ocorrência do evento é desconhecido. Entretanto, sabe-se que ele ocorreu em algum momento anterior ao da observação.
- Censura intervalar: ocorre quando o exato momento de ocorrência do evento é desconhecido. Contudo, sabe-se que ocorreu entre duas datas conhecidas.

Existem três mecanismos conhecidos de censura à direita: a censura do tipo I em que o estudo será terminado após um período pré-estabelecido de tempo. Assim, as observações cujo evento de interesse não foi observado até este tempo são ditas censuradas; a censura do tipo II na qual o estudo será terminado após ter ocorrido o evento de interesse para um número pré-estabelecido de observações e a censura aleatória que ocorre se a observação for retirada no decorrer do estudo sem ter ocorrido o evento de interesse ou se o evento de interesse ocorrer por uma razão diferente da estudada.

As censuras também são classificadas como não informativas e informativas. No primeiro caso, as censuras ocorrem ao acaso e não estão relacionados com o desfecho. Por outro lado, censuras informativas ocorrem quando estão relacionadas ao desfecho, como, por exemplo, se os indivíduos abandonam o estudo devido a efeitos adversos a um dos tratamentos (CARVALHO et al., 2011).

Outra forma de dados incompletos é denominada truncamento. Esta ocorre quando indivíduos não são incluídos no estudo por motivos relacionados à ocorrência de um determinado

desfecho. Existem dois tipos de truncamento, que estão relacionados com a janela de tempo de observação pré-estabelecida (CARVALHO et al., 2011):

- Dado truncado à esquerda: o critério de seleção inclui somente os indivíduos que sofreram o evento após o tempo de início da janela temporal de observação.
- Dado truncado à direita: o critério de seleção dos indivíduos inclui somente aqueles que sofreram o evento de interesse.

Portanto, para realizar uma análise de sobrevivência é preciso ter informações sobre o tempo de falha e a censura.

A análise de sobrevivência tem sido desenvolvida e aplicada em diferentes áreas, com diferentes denominações, como por exemplo análise do tempo até o evento (*time to event analysis*), na engenharia de qualidade (*reliability analysis, failure time analysis*), em economia (*transition analysis, duration analysis*), em sociologia (*event history analysis*) ou na pesquisa clínica e epidemiológica (*survival analysis* ou *lifetime data analysis*).

### **Função de sobrevivência**

Esta é uma das principais funções probabilísticas usadas para descrever estudos de sobrevivência. A função de sobrevivência é definida como a probabilidade de uma observação não falhar até um certo tempo  $t$ , ou seja, a probabilidade de uma observação sobreviver ao tempo  $t$ . Em termos probabilísticos, isto é escrito como:

$$S(t) = P(T > t) = 1 - F(t) = \int_t^{\infty} f(u)du.$$

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

### **Função de taxa de falha ou de risco**

A função de taxa de falha  $\lambda(t)$  é bastante útil para descrever a distribuição do tempo de vida de pacientes. Ela descreve a forma em que a taxa instantânea de falha muda com o tempo.

A função de taxa de falha de  $T$  é, então, definida como:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t},$$

a função de taxa de falha ou de risco pode ser definida também como:

$$\lambda(t) = -\frac{d}{dt}[\log S(t)].$$

A função de taxa de falha é mais informativa do que a função de sobrevivência. Diferentes funções de sobrevivência podem ter formas semelhantes, enquanto as respectivas funções de taxa de falha podem diferir drasticamente (COLOSIMO; GIOLO, 2006).

### 2.3 Modelos paramétricos

Nesta seção serão apresentados modelos paramétricos para análise de sobrevivência, os quais podem ser úteis quando a distribuição até o evento é suposta conhecida (SILVA, 2005). Quando esses modelos oferecem um bom ajuste aos dados, tendem a produzir estimativas mais precisas sobre as quantidades de interesse, pois se baseiam num número menor de parâmetros. Ao mesmo tempo, se o modelo paramétrico for escolhido de forma incorreta, pode conduzir a estimadores consistentes de quantidades erradas (KLEIN; MOESCHBERGER, 1997).

A seguir são apresentadas as distribuições mais usuais em análise de sobrevivência.

#### Distribuição Exponencial

A distribuição mais simples para estudos de sobrevivência é a exponencial (LEE; WANG, 2003). No final dos anos 1940, era utilizada principalmente em estudos de confiabilidade de sistemas eletrônicos e, posteriormente, em estudos de tempo de vida e em remissão de doenças crônicas e infecciosas (CARVALHO et al., 2011).

A função de densidade de probabilidade para a variável aleatória tempo de falha  $T$  com distribuição exponencial é dada por:

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

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

As funções de sobrevivência  $S(t)$  e de taxa de falha  $\lambda(t)$  são dadas, respectivamente, por:

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

e

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

Como pode-se observar, a função de risco  $\lambda(t)$  para a distribuição Exponencial é constante. Assim, na prática, dificilmente a hipótese de que a função de risco seja constante, ou que os tempos de sobrevivência sejam distribuídos exponencialmente é válida (COLLET, 2004).

### **Distribuição de Weibull**

A distribuição de Weibull é a distribuição de probabilidade do tempo mais utilizada na análise de sobrevivência, devido à simplicidade e flexibilidade da função de risco e sobrevivência.

Para uma variável aleatória  $T$  com distribuição de Weibull, tem-se a função densidade de probabilidade dada por:

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

em que  $\gamma$  é o parâmetro de forma e  $\alpha$  é o de escala, ambos positivos. O parâmetro  $\alpha$  tem a mesma unidade de medida de  $t$  e  $\gamma$  não tem unidade.

Para esta distribuição, as funções de sobrevivência e de risco são, respectivamente,

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

e

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

para  $t \geq 0$ ,  $\alpha > 0$ . Observe que, quando  $\gamma = 1$ , tem-se a distribuição Exponencial. Sendo assim, a distribuição Exponencial é um caso particular da distribuição de Weibull.

### **Distribuição Log-Normal**

Assim como a distribuição de Weibull, a distribuição log-normal é muito utilizada para caracterizar tempos de vida de produtos e indivíduos.

A função densidade de uma variável aleatória  $T$  com distribuição log-normal é dada por:

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

em que  $\mu$  é a média do logaritmo do tempo de falha e  $\sigma$  é o desvio padrão.

As funções de sobrevivência e de taxa de falha de uma variável log-normal não apresentam uma forma analítica explícita e são, desse modo, representadas, respectivamente, por:

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

e

$$\lambda = \frac{f(t)}{S(t)},$$

em que  $\Phi(\cdot)$  é a função de distribuição acumulada de uma normal padrão.

Existem várias outras distribuições possíveis, tais como a gama e a log-logística.

## 2.4 Modelo de regressão de Cox

O modelo de regressão de Cox (1972) permite a análise de dados provenientes de estudos de tempo de vida em que a resposta é o tempo até a ocorrência de um evento de interesse.

Considere  $p$  covariáveis, de modo que  $\mathbf{x}$  seja um vetor com os componentes  $\mathbf{x} = (x_1, \dots, x_p)'$ .

A expressão geral do modelo de regressão de Cox é dada por:

$$\lambda(t) = \lambda_0(t)g(\mathbf{x}'\boldsymbol{\beta}),$$

em que  $g$  é uma função não-negativa que deve ser especificada, tal que  $g(0) = 1$ . Este modelo é composto pelo produto de dois componentes, um não-paramétrico e outro paramétrico. O componente não-paramétrico,  $\lambda_0(t)$ , não é especificado e é uma função não-negativa do tempo. O componente paramétrico é frequentemente usado na seguinte forma:

$$g(\mathbf{x}'\boldsymbol{\beta}) = \exp\{\mathbf{x}'\boldsymbol{\beta}\} = \exp\{\beta_1 x_1 + \dots + \beta_p x_p\},$$

em que  $\boldsymbol{\beta}$  é o vetor de parâmetros associado às covariáveis. Esta forma garante que  $\lambda(t)$  seja sempre não-negativa.

Para a estimação do vetor de parâmetros desconhecidos  $\boldsymbol{\beta}$ , Cox (1975) propôs a função de verossimilhança parcial,  $L(\boldsymbol{\beta})$ , a qual é dada por:

$$L(\boldsymbol{\beta}) = \prod_{j \in \mathfrak{S}} \left( \frac{\exp\{\boldsymbol{\beta}'x_j\}}{\sum_{j \in R_i} \exp\{\boldsymbol{\beta}'x_j\}} \right) = \prod_{i=1}^n \left( \frac{\exp\{\boldsymbol{\beta}'x_i\}}{\sum_{j \in R_i} \exp\{\boldsymbol{\beta}'x_j\}} \right)^{\delta_i}.$$

Os indivíduos censurados entram na função de verossimilhança parcial  $L(\boldsymbol{\beta})$  por meio do conjunto de risco  $R_i$  (o conjunto dos indivíduos em risco anterior ao tempo  $t_i$ ) e contribuem, para esta função somente enquanto permanecem em risco. Os parâmetros  $\boldsymbol{\beta}$  são estimados maximizando-se o logaritmo da função  $L(\boldsymbol{\beta})$ , isto é, resolvendo-se o sistema de funções escores dadas por:

$$\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \delta_i \left[ x_i - \frac{\sum_{j \in R_i} x_j \exp\{\boldsymbol{\beta}'x_j\}}{\sum_{j \in R_i} \exp\{\boldsymbol{\beta}'x_j\}} \right] = 0$$

em que  $\ell(\boldsymbol{\beta}) = \log L(\boldsymbol{\beta})$ .

Versões paramétricas do modelo de Cox são obtidas especificando-se uma função paramétrica para a função de risco de base  $\lambda_0(t)$ , ou seja, assumindo-se uma distribuição de probabilidade para os tempos  $t_i$ ,  $i = 1, \dots, n$ . Distribuições usuais nas versões paramétricas são, dentre outras, a Exponencial, Weibull e a Log-Normal.

## 2.5 Modelos com fragilidade

O modelo de fragilidade é caracterizado pela utilização de uma extensão do modelo de Cox, ou seja, de uma variável aleatória não observável, que representa as informações que não podem ou não foram observadas, como fatores ambientais, sociais e genéticos ou informações que por algum motivo não foram consideradas no planejamento.

Uma das formas encontradas para incorporar esse efeito aleatório, que é chamado de variável de fragilidade, é introduzi-lo na modelagem da função de risco com o objetivo de controlar a heterogeneidade não observável das unidades em estudo, inclusive a dependência das unidades que partilham os mesmos fatores de risco.

Um modelo de fragilidade fornece uma maneira conveniente de lidar com observações correlacionadas ou heterogeneidade não observada. Uma fragilidade é um efeito não observado, suposto aleatório, compartilhado pelas unidades observacionais ou experimentais dentro de um grupo (um indivíduo, uma ninhada, um centro de investigação, etc.) (KLEIN; MOESCHBERGER, 1997).

O fato de considerar o efeito aleatório introduzido na função de risco faz com que o modelo englobe dois tipos de fonte de variação para os dados. A primeira delas, que gera a heterogeneidade entre as observações, é causada por covariáveis individuais não observadas. A segunda fonte de variação é proveniente das covariáveis comuns a indivíduos de um mesmo grupo ou família, sendo a fragilidade compartilhada. Num modelo de fragilidade compartilhada, o efeito aleatório está associado a grupos de indivíduos, ao invés de estar associada a cada indivíduo isoladamente. A principal suposição do modelo de fragilidade compartilhada é de que todos os indivíduos de um mesmo grupo  $i$  partilham da mesma distribuição de fragilidade. Os tempos de vida são considerados condicionalmente independentes com relação à fragilidade compartilhada.

Como extensão do modelo de Cox clássico, a inclusão de efeitos aleatórios no modelo de fragilidade é realizada a partir de uma variável aleatória não observável  $Z$  que independe do tempo, com  $E[Z] = 1$  e  $Var[Z] = \theta$ , ou seja,

$$\lambda_{ij}(t) = z_i \lambda_0(t) \exp \{x'_{ij} \boldsymbol{\beta}\},$$

em que  $\lambda_0(t)$  é a função de risco de base,  $x'_{ij}$  são os preditores e  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$  são os respectivos coeficientes de regressão (CARVALHO et al., 2011).

Versões paramétricas desse modelo apresentado são obtidas pela especificação de uma função paramétrica para o risco de base  $\lambda_0(t)$  como, por exemplo, a de Weibull e a exponencial, dentre outras, igual no modelo de Cox.

Um grupo  $i$  com fragilidade  $Z_i > 1$  tende a experimentar algum determinado evento com taxa mais rápida que no modelo de Cox básico. Do mesmo modo, aqueles que apresentarem fragilidades  $Z_i < 1$  tendem a ter tempos maiores até a ocorrência do evento. Outra consideração a ser ressaltada é que, quando a variância  $\theta$  se aproxima de zero, o modelo se reduz ao modelo de Cox básico (CARVALHO et al., 2011).

Assim, a fragilidade representa um efeito aleatório que descreve o risco comum, isto é, a fragilidade compartilhada por indivíduos dentro de um mesmo grupo. Esses grupos apresentam fragilidades diferentes, onde grupos com valores grandes de fragilidade deverão experimentar o evento de interesse em tempos menores do que aqueles com valores de fragilidade pequenos.



### 2.5.1 Distribuições de probabilidade para fragilidade

Várias distribuições de probabilidade para as fragilidades têm sido propostas, incluindo a gama, a estável positiva (*positive stable*), inversa gaussiana e a log-normal. Essas distribuições são úteis para modelar esse efeito aleatório incluso no modelo. Suas características são descritas a seguir.

#### Fragilidade Gama

A mais comumente usada tem sido a distribuição Gama por ser não negativa, extremamente flexível e algebricamente conveniente. Para, por exemplo, diferentes escolhas dos parâmetros de forma e escala, a distribuição Gama varia desde uma exponencial até uma curva em forma de um sino, similar à Normal.

Uma fragilidade Gama é uma variável aleatória  $Z$ , correspondente a  $Gama(\mu, \theta)$  onde a média  $\mu$  é fixada em 1 e a variância  $\theta$ , apresenta função densidade de probabilidade

$$f(z) = \frac{\theta^{-\frac{1}{\theta}} z^{\frac{1}{\theta}-1} \exp\{-z/\theta\}}{\Gamma(1/\theta)}, \quad \theta > 0, \quad (2.1)$$

em que  $\Gamma(\cdot)$  é a função gama.

A transformada de Laplace associada é dada por

$$\mathcal{L}(s) = (1 + \theta s)^{-\frac{1}{\theta}}, \quad s \geq 0.$$

Para a distribuição Gama, o tau de *Kendall* (HOUGAARD, 2000), o qual é uma medida de associação entre quaisquer dois eventos de tempo de um mesmo grupo no caso multivariado, pode ser calculado como

$$\tau = \frac{\theta}{\theta + 2} \in (0, 1).$$

#### Fragilidade estável positiva

Hougaard (2000) introduz a distribuição estável positiva como uma família com dois parâmetros: de escala  $\delta > 0$  e o índice  $\alpha < 1$ . Estabelecendo  $\delta = \alpha$ , a distribuição de fragilidade estável positiva  $PS^*(\nu)$  é obtida, com  $\nu = 1 - \alpha$ .

A função densidade de probabilidade associada é dada por

$$f(z) = -\frac{1}{\pi z} \sum_{k=1}^{\infty} \frac{\Gamma(k(1-\nu)+1)}{k!} (-z^{\nu-1})^k \sin((1-\nu)k\pi), \quad \nu \in (0, 1),$$

em que a média e variância são indefinidas. Portanto, o parâmetro de heterogeneidade  $\nu$  não correspondem à variância do termo de fragilidade.

Em contraste a função densidade de probabilidade, a transformada de Laplace associada tem uma forma muito simples,

$$\mathcal{L}(s) = \exp\{-s^{1-\nu}\}, \quad s \geq 0.$$

Com dados agrupados, o tau de *Kendall* para distribuição de fragilidade estável positiva é

$$\tau = \nu \in (0, 1).$$

### Fragilidade Inversa Gaussiana

A distribuição fragilidade Inversa Gaussiana  $IG^*(\theta)$  tem densidade

$$f(z) = \frac{1}{\sqrt{2\pi\theta}} z^{-\frac{3}{2}} \exp\left\{-\frac{(z-1)^2}{2\theta z}\right\}, \quad \theta > 0,$$

em que a média e a variância são 1 e  $\theta$ , respectivamente. Para a transformada de Laplace temos

$$\mathcal{L}(s) = \exp\left\{\frac{1}{\theta}(1 - \sqrt{1 + 2\theta s})\right\}, \quad s \geq 0.$$

Com dados multivariados, o tau de *Kendall* (HOUGAARD, 2000), para distribuição de fragilidade inversa Gaussiana é dada por

$$\tau = \frac{1}{2} - \frac{1}{\theta} + 2 \frac{\exp\{2/\theta\}}{\theta^2} \int_{2/\theta}^{\infty} \frac{\exp\{-z\}}{z} dz \in (0, 1/2).$$

### Fragilidade Log-Normal

McGilchrist (1993) desenvolveu a metodologia para ajustar modelos de fragilidades em paralelo com a teoria dos modelos clássicos mistos. Assim ele propôs o modelo  $\lambda_{ij}(t) = \lambda_0(t) \exp(x'_{ij}\beta + w_i)$ , em que  $w_i$  é o valor atual do efeito aleatório  $W_i$  seguindo uma distribuição normal com média zero e variância  $\gamma$ , em que  $Z_i = \exp(w_i)$  A fragilidade correspondente

tem uma distribuição lognormal (DUCHATEAU; JANSSEN, 2008), com função densidade de probabilidade dada por

$$f(z) = \frac{1}{z\sqrt{2\pi\gamma}} \exp\left\{-\frac{(\log z)^2}{2\gamma}\right\},$$

com  $\gamma > 0$ . A média e a variância da fragilidade são dadas, respectivamente, por

$$E(Z) = \exp\{\gamma/2\}$$

e

$$\text{Var}(Z) = \exp\{2\gamma\} - \exp\{\gamma\}.$$

No modelo de fragilidade Log-Normal a avaliação da transformada de Laplace não é explícita, devido a complicações na obtenção das suas funções de sobrevivência e de risco. Outra medida que também não possui uma forma fechada para este modelo é o tau de *Kendall*, sendo este obtido através da transformada de Laplace.

### **Estimação Paramétrica**

De acordo com Munda et al. (2012), na análise de dados de sobrevivência agrupados com a presença de censura à direita, a observação por indivíduos  $j \in J_i = \{1, \dots, n_i\}$  do grupo  $i \in I = \{1, \dots, n\}$  é o par  $\mathbf{u}_{ij} = (y_{ij}, \delta_{ij})$ , em que  $y_{ij} = \min(t_{ij}, c_{ij})$ , ou seja, o mínimo entre o tempo de sobrevivência  $t_{ij}$  e o tempo de censura  $c_{ij}$ , onde  $\delta_{ij} = I(t_{ij} \leq c_{ij})$  é o indicador de falha. É também possível a inclusão da informação da covariável nessa estrutura, ou seja,  $\mathbf{u}_{ij} = (y_{ij}, \delta_{ij}, \mathbf{x}_{ij})$ , em que  $\mathbf{x}_{ij}$  denota o vetor de covariáveis para a  $ij$ -ésima observação. Se a truncagem à esquerda também se faz presente, tempos truncados  $\tau_{ij}$  são representados pelo vetor  $\boldsymbol{\tau}$ .

No ajuste paramétrico, a estimação é baseada na verossimilhança marginal, em que as fragilidades tem sido integradas sob a média da verossimilhança condicional com a respectiva distribuição da fragilidade. Sob a suposição de censura à direita não informativa e independência entre o tempo de censura e o tempo de sobrevivência das variáveis aleatórias, dada a informação da covariável, a log-verossimilhança marginal dos dados observados  $\mathbf{u} = \{\mathbf{u}_{ij}; i \in I, J \in J_i\}$  pode ser escrita como

$$\begin{aligned}
\ell_{\text{marg}}(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{u} | \boldsymbol{\tau}) &= \sum_{i=1}^n \left\{ \left[ \sum_{j=1}^{n_i} \delta_{ij} (\log(\lambda_0(y_{ij})) + \mathbf{x}'_{ij} \boldsymbol{\beta}) \right] \right. \\
&+ \log \left[ (-1)^{d_i} \mathcal{L}^{d_i} \left( \sum_{j=1}^{n_i} \Lambda_0(y_{ij}) \exp(\mathbf{x}'_{ij} \boldsymbol{\beta}) \right) \right] \\
&\left. - \log \left[ \mathcal{L} \left( \sum_{j=1}^{n_i} \Lambda_0(\tau_{ij}) \exp(\mathbf{x}'_{ij} \boldsymbol{\beta}) \right) \right] \right\}, \tag{2.2}
\end{aligned}$$

com  $d_i = \sum_{j=1}^{n_i} \delta_{ij}$  o número de eventos no  $i$ -ésimo grupo, e  $\mathcal{L}^{(q)}(\cdot)$  a  $q$ -ésima derivada da transformação de Laplace da distribuição da fragilidade.

As estimativas de  $\boldsymbol{\theta}$ ,  $\boldsymbol{\beta}$  e  $\boldsymbol{\gamma}$  são obtidas maximizando a log-verossimilhança, o que pode ser facilmente feito quando for possível calcular as derivadas de ordem superior da transformação de Laplace  $\mathcal{L}^{(q)}(\cdot)$ , até  $q = \max\{d_1, \dots, d_n\}$  (MUNDA et al., 2012).

### 2.5.2 Fragilidade semi-paramétrica

McGilchrist e Aisbett (1991) usaram a expressão de representação alternativa para o modelo para poder derivar a verossimilhança parcial penalizada. O modelo é representado da seguinte forma:

$$\lambda_{ij}(t) = \lambda_0(t) \exp \{x'_{ij} \boldsymbol{\beta} + w_i\},$$

em que  $w_i = \log z_i$ , sendo que  $w_i$  denota o efeito aleatório presente no modelo com variância  $\gamma$  e  $z_i$  denota a fragilidade com variância  $\theta$ .

#### Fragilidade Gama

Consideramos uma distribuição Gama uni-paramétrica para as fragilidades, com média 1 e a variância  $\theta$ , cuja a densidade pode ser vista em (2.1). Além disso, a modelagem é baseada no efeito aleatório  $W = \log(Z)$ , ao invés das fragilidades. Dada a função densidade para os  $Z_i$ 's em (2.1), então a função densidade para os  $W_i$ 's corresponde a

$$f(w) = \frac{\theta^{-\frac{1}{\theta}} (\exp\{w\})^{\frac{1}{\theta}-1} \exp(-\exp\{w\}/\theta)}{\Gamma(1/\theta)}, \quad \theta > 0. \tag{2.3}$$

### Fragilidade Log-Normal

Se considerarmos a situação em que os efeitos aleatórios  $w_i$  seguem uma distribuição normal com média  $\mu$  e variância  $\gamma$ . As fragilidades  $z_i$  seguem dessa forma uma distribuição log-normal com os parâmetros  $\mu$  e  $\gamma$  representada pela função densidade de probabilidade dada pela equação (2.4), ver (DUCHATEAU; JANSSEN, 2008).

$$f(z) = \frac{1}{z\sqrt{2\pi\gamma}} \exp\left\{-\frac{1}{2\gamma}(\log z - \mu)^2\right\}, \quad (2.4)$$

com

$$E(Z) = \exp\{\mu + \gamma/2\}$$

e

$$Var(Z) = \exp\{2\mu + \gamma\}(\exp\{\gamma\} - 1).$$

### Estimação semi-paramétrica

A estimação dos parâmetros nos modelos de fragilidade semi-paramétricos com distribuições da componente de fragilidade Gama e Log-Normal é feita por meio da verossimilhança parcial penalizada aproximada. No modelo Gama também é possível a estimação por meio do algoritmo EM (Esperança Maximização). Duchateau e Janssen (2008) mostram que as estimativas geradas pelo método da verossimilhança parcial penalizada são as mesmas encontradas pelo algoritmo EM.

De acordo com Duchateau e Janssen (2008) na abordagem da verossimilhança parcial penalizada, a verossimilhança completa dos dados consiste de duas partes. A primeira parte é a verossimilhança condicional dos dados dado as fragilidades, enquanto que a segunda parte corresponde à distribuição das fragilidades.

Nesta abordagem a segunda parte da verossimilhança é considerada como um termo de penalização. Se o valor atual do efeito aleatório está longe da média zero, o valor absoluto do logaritmo da função densidade avaliada será tipicamente grande e o termo de penalidade tem tipicamente uma contribuição negativa para verossimilhança completa dos dados. Adicionalmente, tomando os efeitos aleatórios como outro conjunto de parâmetros na primeira parte da

verossimilhança, esta verossimilhança pode ser transformada em uma expressão de verossimilhança parcial (DUCHATEAU; JANSSEN, 2008). Como resultado tem-se

$$L_{ppl}(\boldsymbol{\gamma}, \boldsymbol{\beta}, \mathbf{w}) = L_{part}(\boldsymbol{\beta}, \mathbf{w}) - L_{pen}(\boldsymbol{\gamma}, \mathbf{w}), \quad (2.5)$$

em que  $L_{ppl}(\boldsymbol{\gamma}, \boldsymbol{\beta}, \mathbf{w})$  é a verossimilhança completa dos dados,  $L_{part}(\boldsymbol{\beta}, \mathbf{w})$  é a verossimilhança condicional dos dados dado as fragilidades e  $L_{pen}(\boldsymbol{\gamma}, \mathbf{w})$  é a função de penalidade. Com  $\eta_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + w_i$  e  $\boldsymbol{\eta} = (\eta_{11}, \dots, \eta_{nn_i})$ , temos que

$$L_{part}(\boldsymbol{\beta}, \mathbf{w}) = \sum_{i=1}^n \sum_{j=1}^{n_i} \delta_{ij} \left[ \eta_{ij} - \log \left( \sum_{qw \in R(y_{ij})} \exp(\eta_{qw}) \right) \right] \quad (2.6)$$

e

$$L_{pen}(\boldsymbol{\gamma}, \mathbf{w}) = - \sum_{i=1}^n \log f_W(w_i),$$

onde  $f_W(w_i)$  é a função densidade para os efeitos aleatórios  $w_i$ 's.

### **Função de penalização no modelo com efeito aleatório Log-Normal**

Para os efeitos aleatórios  $w_i, i = 1, \dots, n$ , tendo uma densidade Log-Normal com média zero e variância  $\gamma$ , temos que a função de penalidade no modelo é dada por

$$l_{pen}(\boldsymbol{\gamma}, \mathbf{w}) = \frac{1}{2} \sum_{i=1}^n \left( \frac{w_i^2}{\gamma} + \log(2\pi\gamma) \right).$$

Segundo Duchateau e Janssen (2008), a maximização da log-verossimilhança parcial penalizada é feita por meio de um algoritmo iterativo em que é avaliado em dois passos: "*inner loop*" e "*outer loop*". No "*inner loop*" é usado o processo de Newton-Raphson para maximizar, para um valor provisório de  $\boldsymbol{\gamma}$ ,  $l_{ppl}(\boldsymbol{\gamma}, \boldsymbol{\beta}, \mathbf{w})$  para  $\boldsymbol{\beta}$  e  $\mathbf{w}$  (melhores preditores linear imparciais). No "*outer loop*", a maximização restrita da verossimilhança do estimador de  $\boldsymbol{\gamma}$  é obtida usando os preditores. O processo é repetido até a convergência.

### **Função de penalização no modelo com efeito aleatório Gama**

Considerando que os efeitos aleatórios são representados pela função densidade  $f(w)$  dada em (2.3), a verossimilhança parcial penalizada para o modelo com fragilidade gama pode

ser escrita como observado em (2.5), com a mesma expressão para a primeira parte da verossimilhança parcial dada em (2.6), mas agora com uma função de penalização dada por

$$l_{pen}(\boldsymbol{\theta}, \mathbf{w}) = -\frac{1}{\theta} \sum_{i=1}^n (w_i - \exp(w_i)).$$

A maximização da log-verossimilhança no modelo com fragilidade Gama é similar á maximização vista para o modelo com efeitos aleatórios seguindo uma distribuição Normal no passo "*inner loop*". O que os diferencia é apenas a função de penalização. No modelo Gama, para um valor fixado de  $\boldsymbol{\theta}$ , as estimativas de  $\boldsymbol{\beta}$  e  $\mathbf{w}$  são obtidas.

A abordagem da verossimilhança parcial penalizada para o modelo com fragilidade gama usa um "*outer loop*" diferente do usado no modelo com efeito aleatório com densidade Normal. No modelo Gama, o "*outer loop*" é baseado na maximização de uma versão perfil da verossimilhança marginal para  $\boldsymbol{\theta}$ . Mais detalhes sobre a estimação dos parâmetros no modelo gama são encontrados em Duchateau e Janssen (2008).

## 2.6 Modelos Multiestado

Um modelo multiestado é um processo estocástico  $(X(t), t \in [0, \infty))$  em que  $X(t)$  denota uma variável aleatória que pode assumir um dos valores no conjunto finito de estados  $S = 1, \dots, N$ . Este processo possui a informação das diferentes transições que ocorrem a um indivíduo ao longo do tempo, assim como o tempo de transição. Os processos multiestado são completamente caracterizados por intensidades de transição ou por probabilidades de transição entre estados  $h$  e  $j$ ,  $(h, j \in S)$ . Os modelos multiestado são representados por diagramas que indicam o número finito de estados clínicos que um paciente pode ocupar.

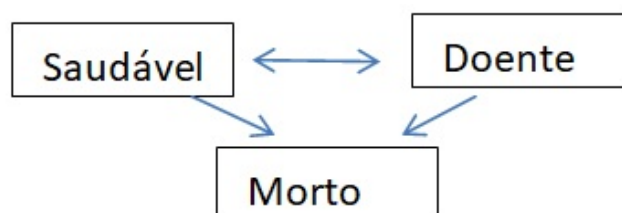
Os estados podem ser considerados como transientes ou absorventes (MEIRA-MACHADO; UNA-ALVAREZ; CADARSO-SUAREZ, 2006). Um estado absorvente não permite transições a partir dele sendo que quando um indivíduo atinge esse estado permanece nele para sempre. A morte é o exemplo mais comum de um estado absorvente. Um estado diz-se transiente quando é possível ocorrer transições a partir dele.

A complexidade de um modelo multiestado depende do número de estados definidos e do número de transições permitidas entre eles (MEIRA-MACHADO et al., 2009). Exemplos comuns de estados são condições como 'saudável', 'doente', 'doente com complicação' e 'morto'.

O modelo de mortalidade é o caso mais simples entre os modelos multiestado e ocorre quando existe apenas uma possibilidade de transição entre os estados inicial ‘vivo’ e um estado absorvente ‘morto’ (ANDERSEN; KEIDING, 2002). O modelo de três estados progressivos é aquele não qual o estado inicial ‘vivo’ é particionado em dois estados transientes. Cada indivíduo em estudo só poderá transitar para o estado mais avançado subsequente (*Vivo* → *Doente* → *Morto*).

Já no modelo enfermidade-morte, cada indivíduo pode transitar entre os estados de ‘saudável’ e ‘doente’, até entrar no estado absorvente de ‘morte’. Está presente a possibilidade de reversibilidade do modelo, ou seja, é possível ocorrer a transição.

Figura 2.1 – Esquema do modelo multiestado “enfermidade-morte”.



## 2.7 Razões de riscos

Para dados independentes, isto é, para  $\theta = 0$ , tem-se que a razão entre as funções de risco de dois indivíduos com vetores de covariáveis  $x_1$  e  $x_2$  não depende de  $t$ , ou seja, o risco do indivíduo com vetor de covariáveis  $x_1$  falhar, relativo ao risco do indivíduo com vetor de covariáveis  $x_2$  é dado por:

$$RR(t) = \frac{\lambda_0(t) \exp\{\boldsymbol{\beta}'x_1\}}{\lambda_0(t) \exp\{\boldsymbol{\beta}'x_2\}} = \exp\{\boldsymbol{\beta}'(x_1 - x_2)\}.$$

Devido à suposição de proporcionalidade dos riscos, esta razão depende somente das estimativas de  $\boldsymbol{\beta}$ . Para  $\theta \neq 0$ , ou seja, para dados correlacionados, o vetor  $\boldsymbol{\beta}$  tem interpretação diferente daquela em que os dados são independentes. A interpretação só é a mesma se forem comparados dois indivíduos de um mesmo grupo, ou seja, dois indivíduos com fragilidades iguais (KLEIN, 1992).

Agora, se forem comparados dois indivíduos com os mesmos valores das covariáveis mas pertencentes a grupos distintos, o risco relativo será a razão entre as fragilidades, isto é:



$$RR(t) = \frac{z_k \lambda_0(t) \exp\{\boldsymbol{\beta}' x_1\}}{z_j \lambda_0(t) \exp\{\boldsymbol{\beta}' x_1\}} = \frac{z_k}{z_j}.$$

Finalmente, se forem comparados dois indivíduos com covariáveis diferentes,  $x_1$  e  $x_2$ , pertencentes a grupos distintos, grupos 1 e 2, tem-se, de acordo com Klein (1992), que

$$RR(t) = \exp\{(x_1 - x_2)' \boldsymbol{\beta}\} \left[ \frac{1 + \theta \hat{\Lambda}_0(t) \exp\{x_2' \boldsymbol{\beta}\}}{1 + \theta \hat{\Lambda}_0(t) \exp\{x_1' \boldsymbol{\beta}\}} \right].$$

A razão dos riscos, diferentemente das anteriores, depende agora do tempo  $t$ . Esta razão tende a 1 quando  $t \rightarrow \infty$ , independente de quais sejam os valores das covariáveis. Ainda, conforme  $\theta$  cresce, a razão converge para 1 mais rapidamente.

### 3 MÉTODOS

#### 3.1 Caracterização dos leitos hospitalares no Estado de Minas Gerais

O Estado de Minas Gerais possui 853 municípios, sendo a capital Belo Horizonte o município mais populoso, com cerca de 2.512.070 habitantes em 2019. Para o ano de 2020, a população estimada do Estado é de 21.292.666 pessoas. No censo de 2010 a densidade demográfica foi registrada em 33,41 habitantes/ $km^2$  com uma concentração de 85,29% nas áreas urbanas. Em relação aos grupos etários, no censo de 2010, a faixa de 0 a 14 anos correspondia a 22,42%, enquanto os de 15 a 64, 69,43%, e os que possuíam mais de 65 anos perfaziam 8,15% da população mineira total. De acordo com os dados do portal do CFM (Conselho Federal de Medicina), no ano de 2018, o total de leitos de UTI disponíveis em Belo Horizonte era de 1595, sendo que destes, 846 mantidos pelo SUS. Em junho de 2020, no Estado de Minas Gerais o total de leitos de UTI era igual a 4.568, sendo 2.879 leitos do SUS. Dados consolidados pela AMIB (Associação de Medicina Intensiva Brasileira), por meio do projeto UTIs Brasileiras, apontam que dos cerca de 16 mil leitos adultos existentes no Sistema Único de Saúde (SUS), a taxa de ocupação média é de 95%. Quanto aos leitos exclusivos para tratamento da Covid-19, há 1.871 para adultos, 31 leitos pediátricos. Deste total apenas 418 são do SUS para adultos<sup>6</sup>.

#### 3.2 Fontes dos dados

Os dados utilizados nesta pesquisa foram obtidos dos arquivos reduzidos da base de dados do Sistema de Informações Hospitalares do Sistema Único de Saúde (SIH-SUS), do Departamento de Informática do SUS (DATASUS), relativa ao ano de 2020 (MS, 2020), dos meses de março a outubro no Estado de Minas Gerais, com informações hospitalares obtidas dos formulários de Autorização de Internação Hospitalar (AIH).

Os dados obtidos por meio do SIH (Sistema de informações hospitalares) permitem o acompanhamento das ações e serviços de saúde executados por pessoas naturais ou jurídicas, de direito público ou privado provendo informações dos pacientes cuja atenção é custeada por planos e seguros privados de assistência à saúde. Os dados estão no formato .dbc (DBF comprimido), contendo os dados das AIHs (Autorização de Internação Hospitalar), uma AIH em cada linha. Cada pacote .dbc corresponde as AIHs processadas pelo DataSus em um determinado

<sup>6</sup> <https://portal.cfm.org.br/images/stories/pdf/leitos%20de%20uti%20covid%20junho%202020%20-%20estados.pdf>

mês. É importante salientar que o mês no qual a AIH foi processada não necessariamente corresponde ao mês em que o paciente foi internado ou teve alta hospitalar, pois as contas hospitalares são apresentadas em diferentes momentos, por diversos motivos (atraso no processamento no hospital, reapresentação de contas, etc.).

Os pacotes do DataSus foram então importados para o RStudio, utilizando-se o pacote *read.dbc* (disponível no CRAN), descompactados e concatenados em um data frame. Este banco contém as informações sobre as internações SUS no Estado de Minas Gerais entre, março e outubro de 2020. Posteriormente foram filtrados todos os registros incluindo apenas Minas Gerais, eliminadas variáveis que não são de interesse para estudo e adicionadas variáveis que não estavam no banco original como nome do hospital (proveniente da lista de estabelecimento do CNES no DataSus), nome da cidade de origem do paciente (proveniente da lista de municípios do Brasil) e corrigidas distorções.

Utilizou-se a AIH 1, que contém registro de identificação do paciente, procedimentos, diagnose e tratamento realizado, acompanhado dos respectivos códigos das afecções diagnosticadas. Foi construído um banco de dados com os códigos de diagnóstico principal, segundo a CID-10 (OMS, 1995), incluídos nos capítulos XIX e XX, (B34.2) infecção por Covid-19 de localização não especificada e (B97.2) Covid-19, como causa de doenças classificadas em outros capítulos. Totalizou-se 24946 observações, onde cada observação corresponde às variáveis associadas a cada um dos pacientes diagnosticados com Covid-19. As variáveis são UF\_ZI (município gestor), MUNIC\_RES (código do IBGE referente ao município de residência do paciente), SEXO (masculino ou feminino), UTI (pacientes que precisaram ou não de uma unidade de terapia intensiva), UTI\_MES\_TO (total de dias de UTI durante a internação), DIAG\_PRINC (código do diagnóstico principal, segundo o CID-10), DIAGSEC (presença ou ausência de diagnóstico secundário além do diagnóstico principal da Covid-19), GESTAO (indica o tipo de gestão do hospital), IDADE, DIAS\_PERM (dias de permanência) e MORTE (indica se o paciente teve saída com morte ( 0: não; 1: sim)).

### 3.3 Análise descritiva

Na primeira etapa foram realizadas análises descritivas. Primeiro elaborou-se uma tabela dos municípios com maior número de internamentos hospitalares por Covid-19 no Estado, depois um mapa de intervalos de classes e um mapa de quintis para ocorrência de internações registradas por meio do código do IBGE referente ao município de residência do paciente. Pos-

teriormente elaborou-se uma tabela de médias de idade e frequências das demais covariáveis que foram utilizadas na predição dos modelos.

Para a análise de sobrevivência, iniciou-se com a estimação das curvas de sobrevivência pelo método de *Kaplan-Meier* e pelo teste de *log-rank* para comparar a igualdade das curvas entre os níveis das variáveis consideradas no estudo.

### 3.4 Ajuste de modelos semi-paramétricos com fragilidade

Foram realizadas análises para avaliar os efeitos de várias covariáveis na cura dos pacientes, incluindo a investigação simultânea de diferentes covariáveis na sobrevivência dos pacientes, estimação do risco relativo, testes de significância e modelos de regressão.

A variável resposta é o tempo de internamento de pacientes portadores de Covid-19 até sua recuperação ( $T$ ) e foram contabilizados em dias, os óbitos foram considerados dados censurados ocorridos durante o período de estudo. Foram ajustados os modelos semi-paramétricos, incluindo as fragilidades e realizando a estimação dos efeitos de covariáveis.

Para o ajuste da variável aleatória no modelo de fragilidade foi considerado o tipo de gestão dos hospitais, dividido em dois grupos: gestão pública ou privada. Além disso, foram ainda consideradas outras 4 covariáveis: IDADE; SEXO (masculino ou feminino); UTI (pacientes que precisaram ou não de uma unidade de terapia intensiva) e DIAGSEC (presença ou ausência de diagnóstico secundário além do diagnóstico principal da Covid-19).

Foi utilizado o modelo de regressão clássico de Cox (1972), considerando 4 covariáveis, de modo que  $\mathbf{x}$  seja um vetor com os componentes  $\mathbf{x} = (x_1, \dots, x_4)'$ .

O componente não-paramétrico, risco de base  $\lambda_0(t)$ , é uma função não-negativa do tempo. O componente paramétrico  $g$  é uma função não-negativa frequentemente usado na seguinte forma multiplicativa:

$$\lambda(t) = \lambda_0(t)g(\mathbf{x}'\boldsymbol{\beta}) = \lambda_0(t) \exp\{\mathbf{x}'\boldsymbol{\beta}\} = \lambda_0(t) \exp\{\beta_1x_1 + \dots + \beta_4x_4\},$$

em que  $\boldsymbol{\beta}$  é o vetor de parâmetros associado às variáveis independentes. Esta forma garante que a função de risco  $\lambda(t)$ , seja sempre não-negativa, sendo o método de estimação de parâmetros baseado na verossimilhança parcial (COX, 1975).

Dada essa estrutura de dados, foi analisada a sobrevida com modelos de risco proporcional de Cox por efeitos mistos (fragilidade), sendo o efeito não observado, suposto aleatório, compartilhado pelas unidades observacionais ou experimentais dentro de um grupo. Neste caso, o tipo de gestão do hospital com censura correspondente a ocorrência do óbito.

A inclusão de efeitos aleatórios no modelo de fragilidade foi realizada a partir de uma variável aleatória não observável  $Z$  que independe do tempo, com  $E[Z] = 1$  e  $Var[Z] = \theta$ , ou seja,

$$\lambda_{ij}(t) = z_i \lambda_0(t) \exp \{x'_{ij} \boldsymbol{\beta}\}, \quad (3.1)$$

para  $j = 1, \dots, n_i$ , os  $n_i$  tempos até os pacientes da  $i$ -ésima gestão receberem alta hospitalar ( $i = 1$  - gestão pública,  $2$  - gestão privada),  $\mathbf{x}$  o vetor de variáveis independentes,  $\lambda_0(t)$  uma função de risco de base conhecida para os modelos paramétricos e desconhecida para os modelos semi-paramétricos,  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_4)$  um vetor de coeficientes de regressão desconhecidos.

Um grupo  $i$  com fragilidade  $Z_i > 1$  tende a experimentar algum determinado evento com taxa mais rápida que no modelo de Cox básico. Do mesmo modo, aqueles que apresentarem fragilidades  $Z_i < 1$  tendem a ter tempos maiores até a ocorrência do evento. Outra consideração a ser ressaltada é que, quando a variância  $\theta$  se aproxima de zero, o modelo se reduz ao modelo de Cox básico (CARVALHO et al., 2011).

Em (3.1) tem-se de forma equivalente, considerando  $z_i = \exp\{w_i\}$ , ou seja,

$$\lambda_{ij}(t) = \lambda_0(t) \exp\{x'_{ij} \boldsymbol{\beta} + w_i\}$$

para os valores das fragilidades, assumidas serem uma amostra independente de variáveis aleatórias  $z_i$ , seguindo uma distribuição Gama ou Log-normal.

A estimação dos parâmetros nos modelos de fragilidade semi-paramétricos com distribuições da componente de fragilidade Gama e Log-normal foi feita por meio da verossimilhança parcial penalizada aproximada. No modelo Gama também é possível a estimação por meio do algoritmo EM (DUCHATEAU; JANSSEN, 2008).

A razão de riscos (RR) para comparação entre dois indivíduos foi calculada a partir da equação (3.2), sendo que, no caso de mesmos valores das covariáveis, mas pertencentes a

grupos distintos, ou seja, hospital de gestão pública ou privada, o risco relativo de falha é dado pela razão entre as fragilidades, dada por:

$$RR(t) = \frac{z_k \lambda_0(t) \exp \{ \boldsymbol{\beta}' x_1 \}}{z_j \lambda_0(t) \exp \{ \boldsymbol{\beta}' x_1 \}}. \quad (3.2)$$

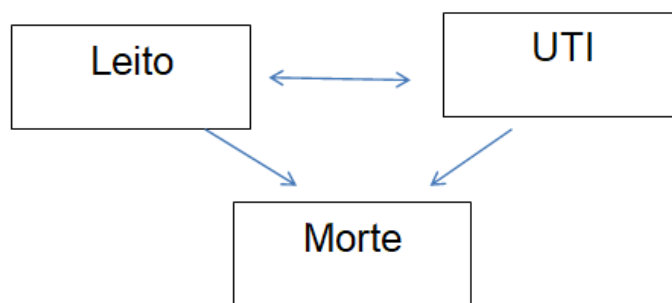
### 3.5 Ajuste de modelos paramétricos com fragilidade

Várias distribuições de probabilidade para as fragilidades foram utilizadas; a Gama, a Estável Positiva, Inversa Gaussiana e a Log-Normal. Para o risco de base foram testadas as distribuições Exponencial, Weibull, Log-Normal, Gaussiana, Logistic, t e a Extreme.

### 3.6 Ajuste de modelos multiestado

Foi utilizado o modelo multiestado enfermidade-morte, considerando-se o modelo com transições entre os estados “leito hospitalar” (doença em estado médio), a internação em UTI (doença em estado grave) e a morte.

Figura 3.1 – Esquema do modelo multiestado “enfermidade-morte”, para os casos de internamentos hospitalares por Covid-19 no Estado.



As transições,

1. primeira transição: Leito para UTI;
2. segunda transição: Leito para Morte;
3. terceira transição: UTI para Morte.

A primeira etapa em uma análise de modelo de multiestado é configurar a tabela com os códigos de transição. Nesta tabela estão especificadas quais transições diretas são possíveis (aquelas com - são impossíveis) e atribui números às transições para referência posterior (Tabela 3.1).

Tabela 3.1 – Códigos de transição na análise de modelo de multiestado.

de	para		
	LEITO	UTI	MORTE
LEITO	-	1	2
UTI	-	-	3
MORTE	-	-	-

Para o conjunto de dados original de Covid-19, cada linha corresponde a um único paciente. Para os dados de formato longo, cada linha corresponderá a uma transição pela qual um paciente está em risco.

Depois de preparar os dados em formato longo, a estimativa dos efeitos das covariáveis é feita usando a regressão de Cox um recurso do pacote *mstate* (WREEDE et al., 2011). Considerando o modelo de Markov, onde assumi-se diferentes efeitos das covariáveis para diferentes transições.

### 3.7 Diagnóstico dos modelos de fragilidade

A escolha do melhor modelo de regressão é uma difícil tarefa, pois raramente existe um único melhor modelo. Por conseguinte, é necessário estabelecer uma estratégia que seja um compromisso entre a seleção de um modelo que proporcione um bom ajuste aos dados e, por outro lado, seja fácil de interpretar suavizando os dados.

Foram utilizados os seguintes processos:

- foi utilizado o procedimento stepwise, que tem como a abordagem geral definir uma equação de regressão com um conjunto de variáveis e então acrescentar (forward) ou eliminar variáveis (backward), considerando o nível de significância de 0,05;
- posteriormente foi utilizado o critério de informação de Akaike, que se baseia na função log-verossimilhança. A ideia básica no Critério de Informação de Akaike (AIC) é ajustar o modelo mais parcimonioso possível, ou seja, que tenha um menor número de parâmetros em comparação com o modelo contendo todos os parâmetros (modelo saturado), mas que

consiga explicar ou descrever o fenômeno tão bem ou até mesmo melhor que o modelo saturado (MOORE, 2016). O objetivo é encontrarmos um modelo tal que a quantidade abaixo seja minimizada.

$$AIC = -2\ell(\hat{\beta}) + 2p,$$

em que,  $\ell(\hat{\beta})$  é a log-verossimilhança do modelo e  $p$  é o número de parâmetros. Um valor baixo para AIC é considerado como representativo de um melhor ajustamento e, assim, na seleção de modelos deve-se ter como objetivo a minimização de AIC.

Para testar a existência de associação entre as observações, ou seja, testar a hipótese nula  $H_0 : Var(Z) = \theta = 0$ , estatísticas de teste comumente usadas são: a de Wald, a da razão de verossimilhanças e a estatística escore. Assintoticamente, tais estatísticas têm distribuição  $\chi_1^2$ .

A variância da variável de fragilidade é uma maneira de quantificar a fragilidade. Se  $\theta = 0$  indicará que todas as fragilidades serão iguais a 1, ficando reduzido no modelo de Cox e valores grandes de  $\theta$  indicarão um alto grau de heterogeneidade entre os grupos e forte associação dentro dos grupos (COLOSIMO; GIOLO, 2006).

$$W_\theta = (\hat{\theta} - \theta_0)' I(\hat{\theta})(\hat{\theta} - \theta_0),$$

em que  $I(\hat{\theta})$  é a matriz de informação observada. Sob  $H_0$  e para  $\hat{\theta}$  de dimensão 1, tem-se:

$$W_\theta = \frac{\hat{\theta}^2}{\hat{Var}(\hat{\theta})}.$$

### 3.8 Uso de Pacotes para Análise Estatística

As análises estatísticas foram realizadas utilizando o software R 3.5.3 (R Core Team, 2019), por meio dos pacotes *survival* (THERNEAU; LUMLEY, 2015). Para facilitar a análise e visualização dos resultados da análise de sobrevivência foi utilizado o pacote *survminer* (KASSAMBARA et al., 2017). O pacote *flexsurv* (JACKSON, 2016) foi usado para definir e prever a partir de modelos de multiestados paramétricos. A principal função de modelagem, *flexsurvreg*, usa a tradicional sintaxe de *survreg* do pacote de sobrevivência de referência (THERNEAU; LUMLEY, 2015). Ele também se incorpora com o pacote *mstate* (WREEDE et al., 2011). Ele contém funções para preparação de dados, análise descritiva, estimativa de risco e previsão para modelos multiestados.



## 4 RESULTADOS E DISCUSSÃO

Neste capítulo mostraremos todos os resultados deste trabalho, separados por quatro seções. Na primeira seção estão as análises descritivas, onde extraímos resultados importantes do tempo de internamentos hospitalares por Covid-19 no Estado de Minas Gerais. Na segunda e terceira seções estão as predições dos modelos de fragilidade semi-paramétricos e paramétricos respectivamente. Na última seção está o modelo multiestado, onde podemos enxergar as transições de estado dos pacientes no modelo de Cox.

### 4.1 Análise descritiva

Inicialmente elaborou-se uma tabela dos municípios com maior número de internamentos hospitalares por Covid-19 no Estado de Minas Gerais, no período de março a outubro (Tabela 4.1). Pode-se observar nesta tabela que, como esperado, a capital Belo Horizonte é o município que tem o maior número de internamentos hospitalares por Covid-19. A alta densidade populacional da capital pode ser um dos motivos para transmissão do Covid-19.

Tabela 4.1 – Municípios com maior número de internamentos hospitalares por Covid-19 no Estado de Minas Gerais no período de março a outubro de 2020.

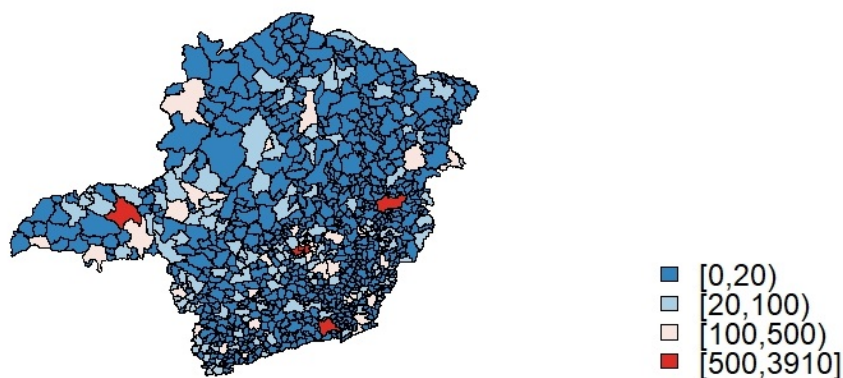
<b>Municípios</b>	<b>Número de Internamentos</b>	<b>População Estimada</b>
Belo Horizonte	3910	2521564
Uberlândia	1323	699097
Contagem	959	668949
Governador Valadares	621	281046
Juiz de Fora	585	573285
Total	7398	4743941

Os cinco municípios citados na Tabela 4.1 concentram 29,66% dos internamentos do Estado, sendo que a população correspondente a esses municípios equivale a 22,28% do total. Isto indica um número de internamentos maior que o esperado, provavelmente, devido aos internamentos dos municípios do interior.

Na análise com métodos descritivos de visualização do número de internamentos hospitalares por Covid-19 no Estado de Minas Gerais, no período de março à outubro, como primeira análise exploratória foi construído o mapa de intervalos de classes (Figura 4.1) e observou-se que as 5 cidades (Belo Horizonte, Uberlândia, Contagem, Governador Valadares e Juiz de Fora) que estão representadas pela cor vermelha no mapa, tiveram maiores números de internamentos

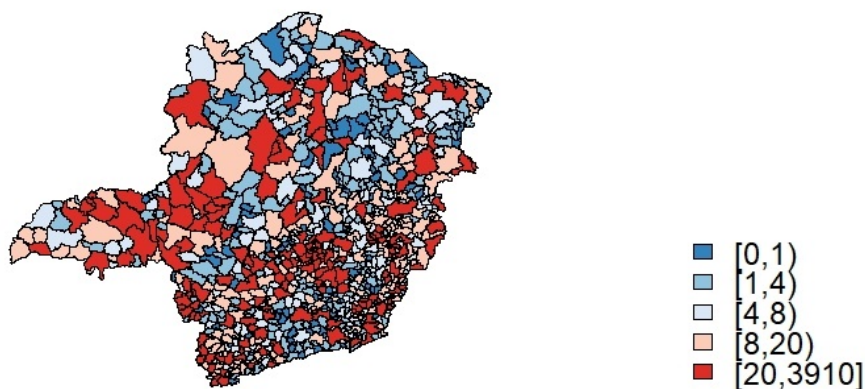
hospitalares por Covid-19 entre [500;3910]. Esse número elevado de internamentos hospitalares nessas cidades tem relação com a alta densidade populacional por serem as maiores cidades populacionais do Estado.

Figura 4.1 – Mapa de intervalos de classes para o número de internamentos hospitalares por Covid-19 no Estado da Minas Gerais no período de março a outubro de 2020.



Na Figura 4.2 temos o mapa de quintis onde as classes foram separadas em percentuais, observa-se que 80% dos municípios, representados pelas quatro primeiras classes que estão representados pelas cores azuis e bege, possuem um baixo número de internamentos hospitalares por Covid-19 entre [0;20) enquanto que os municípios representados pela cor vermelha apresenta um alto número de internamentos hospitalares por Covid-19 entre [20;3910]. Assim, dos 853 municípios de Minas Gerais, 80% deles (682 municípios) possuem menos que 20 internamentos por Covid-19, acumulados no período de março a outubro de 2020.

Figura 4.2 – Mapa de quintis para o número de internamentos hospitalares por Covid-19 no Estado da Minas Gerais no período de março a outubro de 2020.



Na Tabela 4.2 pode-se observar que no período de estudo (março à outubro) foram registrados 24946 pacientes diagnosticados com Covid-19, sendo que 86,74% precisaram ser atendidos nos hospitais públicos do Estado, enquanto que 13,26% foram atendidos em hospitais privados. A maioria é do sexo masculino com 53,54%, sendo 46,46% dos pacientes do sexo feminino, 33,81% dos pacientes necessitaram de uma unidade de terapia intensiva (UTI), enquanto que 66,19% dos pacientes não necessitaram. No diagnóstico secundário só 8,86% dos pacientes foram diagnosticados com uma ou mais outras doenças, portanto 91,14% tiveram seu diagnóstico de internação exclusivamente devido a ocorrência de Covid-19. Quanto à idade dos pacientes obteve uma média de 61,35 anos no geral, sendo que este valor aumenta para 70,29 anos, quando são considerados apenas os pacientes que foram a óbito.

Segundo a OMS, 80% dos pacientes com Covid-19 apresentam sintomas leves e sem complicações, 15% evoluem para hospitalização que necessita de oxigenoterapia e 5% precisam ser atendidos em unidade de terapia intensiva (UTI) (ORGANIZATION et al., 2020). O nosso estudo indicou resultados diferentes da OMS, uma vez que dos pacientes que precisaram ser internados (33,81%) foram atendidos em unidade de terapia intensiva. Segundo o informe epidemiológico da Secretaria de Estado de Saúde, no dia 31 de julho eram confirmados 127.106 casos no Estado de Minas Gerais, sendo que 27.289(21,47%) estavam em acompanhamento, 97.048(76,35%) eram casos recuperados, a taxa de letalidade da Covid-19 era de 2,18% <sup>7</sup>. O informe epidemiológico está de acordo com a OMS, pois os números de recuperados estão semelhantes aos números de sintomas leves. No caso de pacientes com necessidade de internação, obteve-se uma taxa de letalidade bem superior sendo igual a 20,08%.

Barman et al. (2020) estimou o tempo médio de recuperação de pacientes com 60 anos ou mais em 25 dias  $IC_{95\%}[17, 22; 32, 78]$ . Por outro lado, o tempo de recuperação de pacientes com menos de 60 anos foi estimado em 21 dias  $IC_{95\%}[12, 82; 29, 32]$ . Esse estudo revela ainda que 50% dos pacientes tiveram alta em 25 dias, cerca de 40% em 20 dias, enquanto apenas 4% tiveram alta hospitalar em 10 dias. Portanto o tempo de recuperação desse estudo tem um tempo muito superior ao tempo de internação do Estado de Minas Gerais do nosso estudo.

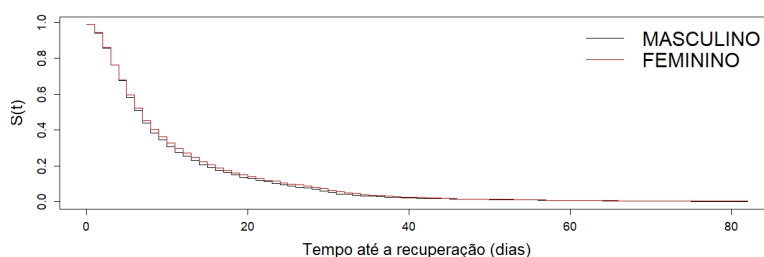
<sup>7</sup> <https://www.saude.mg.gov.br/component/gmg/story/13307-informe-epidemiologico-coronavirus-31-7-2020>

Tabela 4.2 – Médias de idade, frequências e porcentagens dos dados de internamentos hospitalares por Covid-19 em Minas Gerais no período de março a outubro de 2020.

Variável	Categoria	Pacientes	Gestão Pública	Gestão Privada	Eventos (alta)	Censuras (óbito)
<b>IDADE (ANOS)</b>		61,35	61,56	59,99	59,10	70,29
<b>SEXO</b>	MASCULINO	53,54%	11643	1712	10547	2808
	FEMININO	46,46%	9996	1595	9390	2201
<b>UTI</b>	SIM	33,81%	7465	968	4726	3707
	NÃO	66,19%	14174	2339	15211	1302
<b>DIAGNÓSTICO SECUNDÁRIO</b>	SIM	08,86%	2027	183	1405	805
	NÃO	91,14%	19612	3124	18532	4204
<b>TOTAL</b>		24946 (100%)	21639 (86,74%)	3307 (13,26%)	19937 (79,92%)	5009 (20,08%)

Inicialmente, para verificar a significância das variáveis dicotômicas, realizou-se a análise não paramétrica de *Kaplan-Meier*, construindo as curvas de sobrevivência conforme as figuras apresentadas a seguir. Para a variável sexo, obteve-se um tempo mediano de internamento semelhante para homens e mulheres, sendo este igual a 7 dias. No entanto, ao realizar o teste de log-rank, a hipótese de igualdade entre os tempos de recuperação foi rejeitada ( $p < 0,05$ ), indicando diferença significativa entre as curvas (Figura 4.3).

Figura 4.3 – Curvas de sobrevivência estimada pelo método de *Kaplan-Meier* para a variável SEXO.

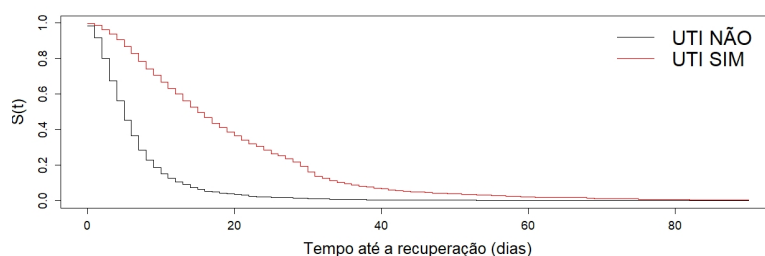


Quanto à necessidade de atendimento em unidade de terapia intensiva, obteve-se diferença significativa entre as curvas ( $p < 0,05$ ), sendo o tempo mediano dos internados em UTI foi igual a 15 dias, enquanto os internados sem necessidade de UTI foi de apenas 5 dias (Figura 4.4).

A gestão adequada dos leitos é fundamental, uma vez que o paciente com Covid-19 pode se tornar um paciente de longa permanência na UTI. Na média, o tempo de permanência de um paciente em uma UTI no hospital público é em torno de 6,5 dias. No caso do paciente com

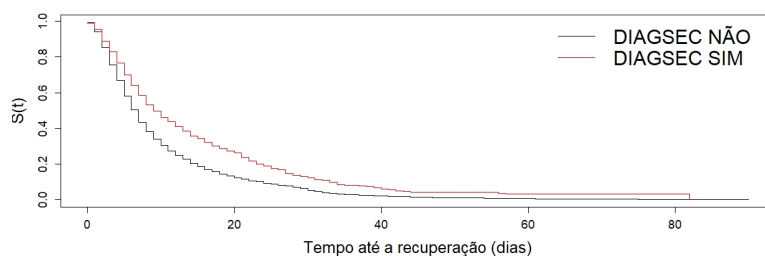
Covid-19, ele permanece em média 14 dias<sup>8</sup>. O tempo mediano de permanência de pacientes diagnosticado com Covid-19 que necessita de UTI está próximo com a média do comunicado da AMIB que é de 15 dias.

Figura 4.4 – Curvas de sobrevivência estimada pelo método de *Kaplan-Meier* para a variável UTI.



Conforme indica a Figura 4.5, os pacientes que foram diagnosticados com uma ou mais doenças apresentam um tempo necessário de recuperação maior do que os pacientes que não possuíam outras doenças, sendo o tempo mediano igual a 9 dias para os primeiros e de 7 dias para os demais ( $p < 0,05$ ).

Figura 4.5 – Curvas de sobrevivência estimada pelo método de *Kaplan-Meier* para a variável DIAGSEC.



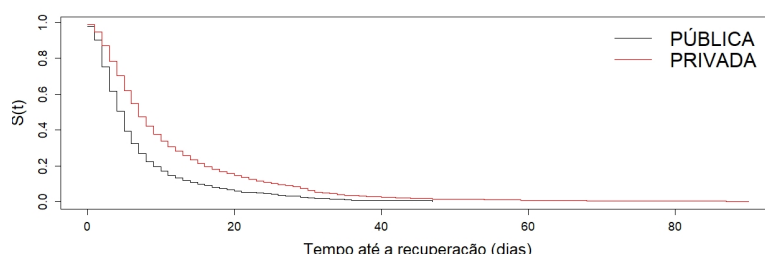
O tipo de gestão do hospital também foi significativo para o tempo de internamentos, indicando que pacientes atendidos em hospitais públicos apresentaram um tempo de recuperação maior do que os pacientes que foram atendidos em hospitais privados ( $p < 0,05$ ). O tempo mediano dos pacientes que foram atendidos em hospitais públicos foi de 7 dias, enquanto que em hospitais privados foi de apenas 5 dias (Figura 4.6).

Silva et al. (2013) afirma que o indicador tempo de permanência hospitalar permite avaliar a eficiência de uma determinada unidade hospitalar e serve como base para mensurar o número de leitos necessários para o atendimento da população de uma área específica. Os mesmos autores ao avaliar a internação de pacientes submetidos à cirurgia de revascularização

<sup>8</sup> <http://www.somiti.org.br/arquivos/site/comunicacao/noticias/2020/covid-19/comunicado-da-amib-sobre-o-avanco-do-covid-19-e-a-necessidade-de-leitos-em-utis-no-futuro.pdf>

do miocárdio, concluíram que aqueles que tinham como fonte pagadora o SUS tiveram tempo de permanência menor que os demais.

Figura 4.6 – Curvas de sobrevivência estimada pelo método de *Kaplan-Meier* para a variável GESTÃO.



A técnica não paramétrica permite descrever o comportamento das covariáveis avaliadas. No entanto não é possível realizar a inclusão conjunta das covariáveis na análise de Kaplan Meier. Dessa forma, torna-se adequado utilizar um modelo semi-paramétrico para construir modelos que possam acomodar as variáveis conjuntamente, realizando a estimação dos coeficientes.

Ao testar os modelos de Cox, ajustando as quatro covariáveis, sem inclusão do efeito aleatório (gestão hospitalar), obteve-se um resultado significativo ( $p < 0,05$ ) para as covariáveis. Posteriormente foram também ajustados os modelos de fragilidade semi-paramétrico gama e log-normal.

## 4.2 Modelos Semi-paramétricos

O AIC dos modelos de fragilidade semi-paramétricos são iguais e os TRV não diferem muito, assim optamos pelo o modelo que forneceu uma maneira conveniente de lidar com as observações correlacionadas, este modelo foi o de fragilidade semi-paramétrico Log-Normal, que teve a variância menor em relação ao modelo de fragilidade semi-paramétrico Gama (Tabela 4.3).

Tabela 4.3 – Modelos de Cox sem fragilidade, fragilidade semi-paramétrico log-normal, fragilidade semi-paramétrico gama e os respectivos AIC, TRV e variância.

Modelos	AIC	TRV	Variância
COX	358700,1	8257	-
LOG-NORMAL	358088,0	8874	0,14
GAMA	358088,0	8871	2,00

Analisando os resultados apresentados na Tabela 4.4, pode-se observar que o efeito de idade foi significativo ( $p < 0,05$ ), sendo que, ao aumentarmos em 1 ano a idade, o risco de recuperação (alta hospitalar) do paciente diminui em 1,33%. O efeito de UTI também foi significativo ( $p < 0,05$ ) e o risco de recuperação dos pacientes que precisaram ser atendidos em uma UTI diminui 73,23% em relação ao risco dos pacientes que não precisaram ser atendidos em uma UTI. O efeito da covariável DIAGSEC foi significativo ( $p < 0,05$ ) e portanto, o risco de recuperação dos pacientes que receberam um diagnóstico secundário de outras doenças diminui 20,80% em relação ao risco dos pacientes que não foram diagnosticados com outras doenças. A covariável SEXO foi significativa ( $p < 0,05$ ), sendo que o risco de recuperação dos pacientes do sexo masculino diminui 3,77% em relação aos pacientes do sexo feminino.

Tabela 4.4 – Estimativas, teste da razão de verossimilhança (TRV), AIC, razão de risco (RR) e teste associado à fragilidade obtidos para o modelo semi-paramétrico de fragilidade log-normal ajustado aos dados de internamentos hospitalares por Covid-19 em Minas Gerais no período de março a outubro de 2020.

<b>Covariável</b>	<b>Estimativa</b>	<b>Erro-Padrão</b>	$\chi^2$	<b>Valor-p</b>	<b>RR</b>
IDADE	-0,0134	0,0004	1190,19	< 0,001	0,9867
UTISIM	-1,3179	0,0174	5748,89	< 0,001	0,2677
DIAGSECSIM	-0,2332	0,0277	70,64	< 0,001	0,7920
SEXOM	-0,0384	0,0142	7,31	< 0,006	0,9623
fragilidade log-normal	-	-	694,40	< 0,001	-
TRV	8874	-	-	< 0,001	-
AIC	358088	-	-	-	-
$\gamma$	0,1457	-	-	-	-

Com o modelo selecionado e os coeficientes obtidos, pode-se apresentar o modelo semi-paramétrico de fragilidade Log-Normal para cada tipo de gestão, conforme equações (4.1) e (4.2), para a gestão pública e privada respectivamente.

$$\lambda_{1j}(t) = 0,7636\lambda_0(t) \exp\{-0,0134IDADE - 1,3179UTI - 0,2332DIAGSEC - 0,0384SEXO\} \quad (4.1)$$

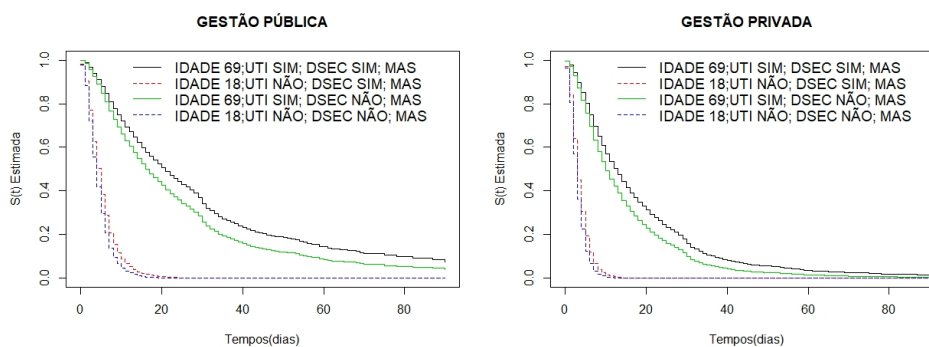
$$\lambda_{2j}(t) = 1,3096\lambda_0(t) \exp\{-0,0134IDADE - 1,3179UTI - 0,2332DIAGSEC - 0,0384SEXO\} \quad (4.2)$$

O risco de recuperação dos pacientes que foram atendidos em uma administração pública diminui 41,69% em relação aos pacientes que foram atendidos em uma administração privada.

$$RR = \frac{0,7636}{1,3096} = 0,5831$$

Na Figura 4.7 pode-se observar claramente que as curvas de sobrevivência representadas pela cor preta (IDADE 69; UTI SIM; DSEC SIM; MAS) e verde (IDADE 69; UTI SIM; DSEC NÃO; MAS) apresentam decréscimos mais acentuados ao longo do tempo na gestão privada. Já as curvas de sobrevivência representadas pela cor vermelha (IDADE 18; UTI NÃO; DIAGSEC SIM; MAS) e azul (IDADE 18; UTI NÃO; DIAGSEC NÃO; MAS) tiveram tempos menores do que as curvas das cores pretas e verdes para as duas gestões. Logo, pacientes da gestão privada tendem a ter um tempo de recuperação menor do que os pacientes da gestão pública.

Figura 4.7 – Curvas de sobrevivência estimadas para pacientes diagnosticado com Covid-19 das gestões públicas e privadas em Minas Gerais no período de março a outubro de 2020.



### 4.3 Modelos de fragilidades paramétricos

Na Tabela 4.5 temos os valores de AIC do modelo de fragilidade Gama com diferentes distribuições. A distribuição que melhor se ajustou aos dados foi a distribuição t, por ter o menor valor de AIC.

Tabela 4.5 – Modelo de fragilidade paramétrica gama com diferentes valores de AIC para cada distribuição.

Distribuições	AIC
Gaussian	139847,9
Logistic	135091,1
t	133182,8
Extreme	160379,5



Pode-se observar na Tabela 4.6 que as fragilidades não foram significativas ao nível 5% de significância. Assim, o modelo paramétrico de fragilidade Gama e risco de base com distribuição t, apesar de ter um AIC menor que os modelos de fragilidades semi-paramétrico, não se ajustou bem aos dados.

Tabela 4.6 – Estimativas e teste associado à fragilidade obtidos para o modelo paramétrico de fragilidade Gama e risco de base com distribuição t ajustado aos dados de internamentos hospitalares por Covid-19 em Minas Gerais no período de março a outubro de 2020.

<b>Covariável</b>	<b>Estimativa</b>	<b>RR</b>	<b>Erro-Padrão</b>	<b>Z</b>	<b>Valor-p</b>
Intercepto	1,8954	-	1,1876	1,60	0,11
IDADE	0,0563	1,0579	0,0019	29,38	< 0,001
DIAGSECSIM	1,3258	3,7652	0,1354	9,79	< 0,001
UTISIM	7,9689	2889,7	0,0958	83,19	< 0,001
fragilidade privada	-1,5439	-	1,1849	-1,30	0,19
fragilidade pública	0,5802	-	1,1819	0,49	0,62
log(scale)	1,4883	-	0,0068	218,24	< 0,001
scale	4,43	-	-	-	-

De modo geral, observamos que entre os modelos de fragilidades, o que melhor se ajustou aos dados de internamentos hospitalares por Covid-19 foi o modelo semi-paramétrico de fragilidade Log-Normal, sendo este, portanto, o melhor modelo para explicar o fenômeno estudado, onde verificamos que as variáveis idade, UTI, diagnóstico secundário e sexo foram significativos e que influenciam no tempo de internamentos hospitalares.

#### 4.4 Modelo multiestado

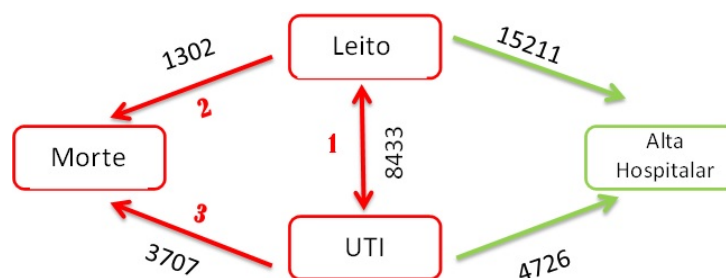
De acordo com a tabela de frequências das transições consideradas, observa-se que dos 24946 internamentos hospitalares por Covid-19, 8433 pacientes precisaram de uma UTI, 1302 pacientes foram a óbitos e 15211 tiveram alta hospitalar. No entanto dos 8433 pacientes que necessitaram de uma UTI, 3707 foram a óbito e 4727 receberam alta hospitalar (Tabela 4.7).

Tabela 4.7 – Frequências das transições na análise de modelo de multiestado.

<b>de</b>	<b>para</b>				
	<b>LEITO</b>	<b>UTI</b>	<b>MORTE</b>	<b>Sem eventos</b>	<b>Entradas</b>
LEITO	0	8433	1302	15211	24946
UTI	0	0	3707	4727	8433
MORTE	0	0	0	5009	5009

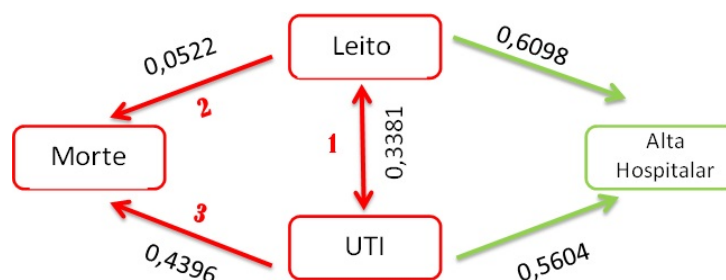
Na figura 4.8 cada seta representada pela cor vermelho corresponde a uma transição e os números (1,2 e 3) são os códigos de transição. O número de eventos ou frequências associados a cada transição é mostrado próximo à seta correspondente.

Figura 4.8 – Representação do modelo multiestado para as frequências associadas a cada transição.



Já na figura 4.9 cada seta representada pela cor vermelho corresponde a uma transição e os números (1,2 e 3) também são os códigos de transição, agora as probabilidades associadas a cada transição são mostradas próximo à seta correspondente.

Figura 4.9 – Representação do modelo multiestado para as probabilidades associadas a cada transição.



Na Tabela 4.8 pode-se observar na variável SEXO que só na primeira transição de leito para UTI foi significativo ( $p < 0,05$ ). Assim, o risco de ocorrer a primeira transição nos pacientes do sexo masculino aumenta 6,17% em relação aos pacientes do sexo feminino.

Para a variável IDADE todas as transições foram significativas ( $p < 0,05$ ). Na primeira transição, de leito para UTI, se aumentarmos em um 1 ano a idade, o risco de transição do paciente diminui em 0,4%. Na segunda transição, de leito para morte, se aumentarmos em um 1 ano a idade, o risco de transição do paciente aumenta em 5% e na terceira transição, de UTI para morte, se aumentarmos em um 1 ano a idade, o risco de transição do paciente aumenta em 0,71%.

A primeira e a terceira transições foram significativas ( $P < 0,05$ ), na variável GESTÃO. Portanto, o risco de ocorrer a primeira transição, de leito para UTI, nos pacientes da gestão pública diminui 42,27% em relação a gestão privada. Já o risco de ocorrer a terceira transição, de UTI para morte nos pacientes da gestão pública, diminui 17,24% em relação a gestão privada.

O efeito da variável diagnóstico secundário na segunda e a terceira transições foram significativas ( $P < 0,05$ ). Assim, o risco na segunda transição, de leito para morte dos pacientes com diagnóstico secundário, aumenta 89,39% em relação aos pacientes que não tem outro diagnóstico além da Covid-19. Na terceira transição, de UTI para morte o risco dessa transição dos pacientes com diagnóstico secundário, diminui 17,75% em relação aos pacientes que não tem outro diagnóstico além da Covid-19.

Tabela 4.8 – Estimativas, razão de risco (RR), erro padrão(EP) da estimativa e o valor crítico de Z obtidos para o modelo semi-paramétrico de Cox ajustado aos dados de internamentos hospitalares por Covid-19 em Minas Gerais no período de março a outubro de 2020.

<b>Covariável</b>	<b>Transição</b>	<b>Estimativa</b>	<b>RR</b>	<b>EP(Estimativa)</b>	<b>Z</b>	<b>Valor-p</b>
SEXOM	1	0,0599	1,0617	0,0253	2,368	0,0179
SEXOM	2	0,0180	1,0182	0,0575	0,314	0,7534
SEXOM	3	0,0254	1,0257	0,0475	0,534	0,5933
IDADE	1	-0,0040	0,9960	0,0008	-5,148	< 0,001
IDADE	2	0,0488	1,0500	0,0022	22,327	< 0,001
IDADE	3	0,0071	1,0071	0,0018	4,023	< 0,001
GESTAOSUS	1	-0,5494	0,5773	0,0390	-14,087	< 0,001
GESTAOSUS	2	-0,0905	0,9135	0,0971	-0,932	0,3514
GESTAOSUS	3	-0,1892	0,8276	0,0795	-2,381	0,0172
DIAGSECSIM	1	-0,0104	0,9897	0,0432	-0,239	0,8108
DIAGSECSIM	2	0,6386	1,8939	0,0777	8,222	< 0,001
DIAGSECSIM	3	-0,1954	0,8225	0,0622	-3,143	0,0017

A resolução Nº 2.156/2016 do Conselho Federal de Medicina estabelece critérios para a admissão e a alta de pacientes em UTI, que devem ser seguidos por todos os médicos, tanto da rede pública quanto privada. A norma tem o objetivo de melhorar o fluxo de atendimento médico diante da carência de leitos de UTI no país, especialmente nos hospitais públicos. Sendo a taxa mediana de permanência das UTIs públicas é de 9,2 dias enquanto que nas particulares é de 5,6<sup>9</sup>.

Na prática, o médico intensivista avalia se o paciente atende aos requisitos para ocupar um leito de UTI ou se deve dar lugar a outra pessoa em situação mais crítica. A admissão de um paciente na UTI deverá ser priorizada de acordo com uma escala de 1 a 5, em que quanto mais baixo o número, maior prioridade o paciente terá, considerando, inclusive, as suas chances de recuperação. No número 5 da escala, estão pacientes com doença incurável e na fase terminal, ou ainda moribundo, sem possibilidade de recuperação.

Apesar de a UTI representar uma área de cuidado intensivo, a internação de pacientes em seus leitos por período de tempo prolongado pode ser sinônimo de malefícios, tanto para o enfermo, como para o próprio hospital. A geração de despesas excessivas, a possibilidade de rejeição da admissão de pacientes críticos devido a falta de leitos (resultando em aumento da morbidade intrahospitalar) (ANS, 2013), o maior risco de exposição a infecções e a alta probabilidade de desenvolvimento de problemas psicológicos, como ansiedade e depressão, são algumas das adversidades causadas pela internação por tempo prolongado na UTI. Sendo assim, a média de permanência de UTI por pacientes de Covid-19, que é o foco deste estudo, representa um meio de investigação para esses problemas, os quais podem ser gerados quando ela se apresenta elevada.

<sup>9</sup> [https://www.sopati.com.br/lermais\\_materias.php?cd\\_materias=506&friurl=-Resolucao-do-CFM-define-criterios-de-admissao-de-pacientes-em-UTIs-#.X5SkJ1hKjIU](https://www.sopati.com.br/lermais_materias.php?cd_materias=506&friurl=-Resolucao-do-CFM-define-criterios-de-admissao-de-pacientes-em-UTIs-#.X5SkJ1hKjIU)

## 5 CONCLUSÃO

Diante do exposto, pode-se concluir que um efeito aleatório foi identificado, indicando a presença de grupos de indivíduos que têm fatores de riscos não observáveis em comum, ou seja, pacientes de uma mesma gestão têm risco mais similares em relação às pacientes de outra gestão.

No modelo de multiestado, os estados intermediários foram identificados. Portanto, foi observada uma sequência de eventos, originando mais do que uma observação por indivíduo. Assim, identificamos nas variáveis preditoras efeitos para as transições entre diferentes estados.

O estudo de modelos de fragilidades e multiestado na avaliação do tempo de internamentos hospitalares por Covid-19 em Minas Gerais, torna-se um recurso auxiliar importante para os órgãos competentes, para uma administração mais adequada dos leitos pois, pacientes podem ficar sem leitos ou ocupar leitos de UTI por um prolongado tempo, causando malefícios tanto para o paciente como para a gestão hospitalar e assim, causando impactos na saúde coletiva do Estado de Minas Gerais.

## REFERÊNCIAS

- ANDERSEN, P. K.; KEIDING, N. Multi-state models for event history analysis. **Statistical methods in medical research**, Sage Publications Sage CA: Thousand Oaks, CA, v. 11, n. 2, p. 91–115, 2002.
- BAI, Y. et al. Presumed asymptomatic carrier transmission of covid-19. **Jama**, American Medical Association, v. 323, n. 14, p. 1406–1407, 2020.
- BARMAN, M. P. et al. Covid-19 pandemic and its recovery time of patients in india: A pilot study. **Diabetes & Metabolic Syndrome: Clinical Research & Reviews**, Elsevier, v. 14, n. 5, p. 1205–1211, 2020.
- CARVALHO, M. S. et al. **Análise de sobrevivência: teoria e aplicações em saúde**. 2. ed. Rio Janeiro: FIOCRUZ, 2011.
- COLLET, D. **Modelling Survival Data in Medical Research**. [S.l.]: London: Chapman & Hall/CRC, 2004.
- COLOSIMO, E. A.; GIOLO, S. R. **Análise de Sobrevivência Aplicada**. 1. ed. São Paulo: Edgard Blucher, 2006.
- COX, D. R. Regression models and life tables. **Journal of the Royal Statistical Society. Series B**, v. 39, n. 2, p. 1–38, 1972.
- COX, D. R. Partial likelihood. **Biometrika**, Oxford University Press, v. 62, n. 2, p. 269–276, 1975.
- DUCHATEAU, L.; JANSSEN, P. **The frailty model**. 1. ed. New York: Springer Science & Business Media, 2008.
- FERNANDES, A. M. R. **Análise de dados em modelos multiestado**. Tese (Doutorado), 2010.
- HOSMER, D. W.; LEMESHOW, S.; MAY, S. **Applied survival analysis: regression modeling of time to event data**. [S.l.]: Wiley-Interscience Hoboken, 2008.
- HOUGAARD, P. Shared frailty models. In: **Analysis of multivariate survival data**. [S.l.]: Springer-Verlag, 2000.
- JACKSON, C. H. flexsurv: a platform for parametric survival modeling in r. **Journal of statistical software**, Europe PMC Funders, v. 70, 2016.
- JIANG, F. et al. Review of the clinical characteristics of coronavirus disease 2019 (covid-19). **Journal of general internal medicine**, Springer, p. 1–5, 2020.
- KASSAMBARA, A. et al. Package ‘survminer’. **Drawing Survival Curves using ‘ggplot2’.(R package version 0.3. 1.)**, 2017.
- KLEIN, J. P. Semiparametric estimation of random effects using the cox model based on the em algorithm. **Biometrics**, v. 48, p. 795–806, Sep 1992.
- KLEIN, J. P.; MOESCHBERGER, M. L. **Survival analysis: techniques for censored and truncated data**. [S.l.]: New York: Springer-Verlag, 1997.
- LAI, C.-C. et al. Severe acute respiratory syndrome coronavirus 2 (sars-cov-2) and corona virus disease-2019 (covid-19): the epidemic and the challenges. **International journal of antimicrobial agents**, Elsevier, p. 105924, 2020.
- LEE, E. T.; WANG, J. **Statistical methods for survival data analysis**. [S.l.]: John Wiley & Sons, 2003. v. 476.

MACHADO, J. P.; MARTINS, M.; LEITE, I. d. C. O arranjo público privado e a mortalidade hospitalar por fontes de pagamento. **Revista de Saúde Pública**, SciELO Public Health, v. 50, p. 42, 2016.

MCGILCHRIST, C.; AISBETT, C. Regression with frailty in survival analysis. **Biometrics**, JSTOR, v. 47, p. 461–466, 1991.

MCGILCHRIST, C. A. Reml estimation for survival models with frailty. **Biometrics**, JSTOR, p. 221–225, 1993.

MEIRA-MACHADO, L.; UNA-ALVAREZ, J. D.; CADARSO-SUAREZ, C. Nonparametric estimation of transition probabilities in a non-markov illness–death model. **Lifetime Data Analysis**, Springer, v. 12, n. 3, p. 325–344, 2006.

MEIRA-MACHADO, L. et al. Multi-state models for the analysis of time-to-event data. **Statistical methods in medical research**, SAGE Publications Sage UK: London, England, v. 18, n. 2, p. 195–222, 2009.

MOORE, D. F. **Applied survival analysis using R**. [S.l.]: Springer, 2016.

MUNDA, M. et al. Parfm: parametric frailty models in r. **Journal of Statistical Software**, Foundation for Open Access Statistics, v. 51, n. 11, p. 1–20, 2012.

ORGANIZATION, W. H. et al. **Oxygen sources and distribution for COVID-19 treatment centres: interim guidance, 4 April 2020**. [S.l.], 2020.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2019. Disponível em: <<https://www.R-project.org/>>.

SILVA, G. S. d. et al. Avaliação do tempo de permanência hospitalar em cirurgia de revascularização miocárdica segundo a fonte pagadora. **Revista da Associação Médica Brasileira**, SciELO Brasil, v. 59, n. 3, p. 248–253, 2013.

SILVA, S. L. C. d. Abordagem bayesiana para modelos paramétricos na análise de sobrevivência. 2005.

THERNEAU, T. M.; LUMLEY, T. Package ‘survival’. **R Top Doc**, v. 128, p. 112, 2015.

WREEDE, L. C. de et al. mstate: an r package for the analysis of competing risks and multi-state models. **Journal of statistical software**, v. 38, n. 7, p. 1–30, 2011.