UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Câmpus de Presidente Prudente

THAINÁ LANZA ARANHA

ESTUDO DO TEMPO DE INTERNAÇÃO HOSPITALAR PELO SUS POR CAUSAS EXTERNAS NO BRASIL NO ANO DE 2019

PRESIDENTE PRUDENTE

2024

THAINÁ LANZA ARANHA

ESTUDO DO TEMPO DE INTERNAÇÃO HOSPITALAR PELO SUS POR CAUSAS EXTERNAS NO BRASIL NO ANO DE 2019

Relatório Final do Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Estatística da FCT/Unesp para aproveitamento na disciplina Trabalho de Conclusão de Curso. Orientador: Prof. Dr. Mário Hissamitsu Tarumoto.

PRESIDENTE PRUDENTE

2024

A662e

Aranha, Thainá Lanza

Estudo do tempo de internação hospitalar pelo SUS por causas externas no Brasil no ano de 2019 / Thainá Lanza Aranha. -- Presidente Prudente, 2024

49 f. : il., tabs.

Trabalho de conclusão de curso (Bacharelado - Estatística) - Universidade Estadual Paulista (UNESP), Faculdade de Ciências e Tecnologia, Presidente Prudente

Orientador: Mario Hissamitsu Tarumoto

1. Acidentes. 2. Análise de Sobrevivência. 3. DATASUS - Departamento de Informática do Sistema Único de Saúde. 4. Internação. 5. Estatística. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Dados fornecidos pelo autor(a).

TERMO DE APROVAÇÃO

Thainá Lanza Aranha

ESTUDO DO TEMPO DE INTERNAÇÃO HOSPITALAR PELO SUS POR CAUSAS EXTERNAS NO BRASIL NO ANO DE 2019

Relatório de Final de Trabalho de Conclusão de Curso aprovado como requisito para obtenção de créditos na disciplina Trabalho de Conclusão do curso de graduação em Estatística da Faculdade de Ciências e Tecnologia da Unesp, pela seguinte banca examinadora:

gov.br Documento assinado digitalmente
MARIO HISSAMITSU TARUMOTO
Data: 17/12/2024 14:36:33-0300
Verifique em https://validar.iti.gov.br

Orientador:_____
Prof. Dr. Mário Hissamitsu Tarumoto
Departamento de Estatística

Edilson Ferreira Flores
Assinado de forma digital por Edilson Ferreira Flores
Dados: 2024.12.17 12:15:13 -03'00'

_____
Prof. Dr. Edilson Ferreira Flores
Departamento de Estatística

Presidente Prudente, 07 de dezembro de 2024

RESUMO

As internações hospitalares por causas externas no Brasil representam um grave problema de saúde pública, especialmente devido às sequelas físicas temporárias ou permanentes que impactam principalmente a população jovem, ativa e masculina. Este estudo analisou as internações ocorridas em 2019, utilizando dados do DATASUS e do Sistema de Informações Hospitalares do SUS (SIHSUS), com o objetivo de compreender padrões regionais, causas predominantes e o tempo de permanência hospitalar. Foram empregadas técnicas de análise exploratória e modelos de regressão em análise de sobrevivência, com abordagens paramétrica e semiparamétrica, para identificar fatores associados ao tempo de internação.

Os resultados apontaram que a distribuição Log-normal foi a mais adequada para modelar os tempos de internação na maioria dos estados, destacando-se pela flexibilidade em captar padrões específicos dos dados. O estudo identificou as principais causas de internação, incluindo acidentes de transporte, quedas, lesões autoprovocadas e agressões, com predomínio de internações entre homens de 15 a 39 anos. Internações prolongadas foram associadas a lesões mais graves, enquanto a maioria dos casos demandou períodos curtos de hospitalização (1 a 7 dias).

As análises ressaltam a importância de iniciativas preventivas, como campanhas de conscientização sobre segurança no trânsito, programas de prevenção de violência e cuidados para populações vulneráveis, como crianças e idosos. Além disso, destaca-se a necessidade de fortalecer a infraestrutura hospitalar e os serviços de emergência, particularmente em regiões com maior vulnerabilidade. Este trabalho reforça o papel da análise estatística como ferramenta essencial para compreender padrões de saúde pública e subsidiar políticas voltadas à redução de internações por causas externas e à melhoria do sistema de saúde.

Palavras-chave: Acidentes, Análise de Sobrevivência, DATASUS, Internação, Modelagem Estatística.

ABSTRACT

Hospital admissions due to external causes in Brazil represent a serious public health problem, especially due to temporary or permanent physical sequelae that mainly impact the young, active, male population. This study analyzed hospital admissions that occurred in 2019, using data from DATASUS and the SUS Hospital Information System (SIHSUS), with the aim of understanding regional patterns, predominant causes, and length of hospital stay. Exploratory analysis techniques and regression models in survival analysis, with parametric and semiparametric approaches, were used to identify factors associated with length of hospital stay.

The results indicated that the log-normal distribution was the most appropriate for modeling length of hospital stay in most states, standing out for its flexibility in capturing specific patterns in the data. The study identified the main causes of hospital admissions, including traffic accidents, falls, self-inflicted injuries, and assaults, with a predominance of hospital admissions among men aged 15 to 39. Longer hospital stays were associated with more severe injuries, while most cases required short hospital stays (1 to 7 days).

The analyses highlight the importance of preventive initiatives, such as awareness campaigns on road safety, violence prevention programs, and care for vulnerable populations, such as children and the elderly. In addition, the need to strengthen hospital infrastructure and emergency services is highlighted, particularly in regions with greater vulnerability. This work reinforces the role of statistical analysis as an essential tool for understanding public health patterns and supporting policies aimed at reducing hospitalizations due to external causes and improving the health system.

Keywords: Accidents, Survival Analysis, DATASUS, Hospitalization, Statistical Modeling.

ÍNDICIE DE ILUSTRAÇÕES

FIGURE 1: IDADE DOS PACIENTES (REGIÃO NORTE) 27
FIGURA 2: IDADE DOS PACIENTES (REGIÃO NORDESTE) 29
FIGURA 3: IDADE DOS PACIENTES (REGIÃO CENTRO-OESTE)..... 30
FIGURA 4: IDADE DOS PACIENTES (REGIÃO SUDESTE) 31
FIGURA 5: IDADE DOS PACIENTES (REGIÃO SUL)..... 32
FIGURA 6: DIAS DE PERMANÊNCIA NO HOSPITAL (REGIÃO NORTE) 34
FIGURA 7: DIAS DE PERMANÊNCIA NO HOSPITAL (REGIÃO NORDESTE)..... 35
FIGURA 8: DIAS DE PERMANÊNCIA NO HOSPITAL (REGIÃO CENTRO-OESTE) . 36
FIGURA 9: DIAS DE PERMANÊNCIA NO HOSPITAL (REGIÃO SUDESTE)..... 37
FIGURA 10: DIAS DE PERMANÊNCIA NO HOSPITAL (REGIÃO SUL)..... 37

ÍNDICE DE TABELA

TABELA 1 - OBSERVAÇÕES DE CADA ESTADO 20
TABELA 2: MEDIDAS DESCRITIVAS DAS OBSERVAÇÕES (SEXO) 22
TABELA 3: MEDIDAS DESCRITIVAS DAS OBSERVAÇÕES (IDADE) 23
TABELA 4: MEDIDAS DESCRITIVAS DAS OBSERVAÇÕES (DIAS DE PERMANÊNCIA) 24
TABELA 5: ESTIMADORES DE MÁXIMA VEROSSIMILHANÇA PARA OS PARÂMETROS DE INTERESSE DE ACORDO COM AS DISTRIBUIÇÕES ESTUDADAS 38

Sumário

1 INTRODUÇÃO 9
2 DESENVOLVIMENTO DO TRABALHO 11
2.1 Modelo de Regressão Linear Simples 11
2.2 Modelo de Regressão Linear Multipla 12
2.3 Análise de Sobrevivência 13
2.3.1 Distribuição Gamma 14
2.3.2 Distribuição Weibull 16
2.3.3 Distribuição Exponencial 17
2.3.4 Distribuição Log Normal 18
3 RESULTADOS 20
3.1 Organização dos dados 20
3.2 Análise Descritiva 21
3.2.1 Sexo dos Pacientes 25
3.2.2 Idade dos pacientes 25
3.2.3 Dias de Permanência no Hospital 32
3.2.5 Modelos Gamma, Weibull, Exponencial e Log-Normal 38
3.2.6 Causas 44
4 CONSIDERAÇÕES FINAIS 45
REFERÊNCIAS 47

1 INTRODUÇÃO

As internações hospitalares por causas externas, ocorrem quando uma pessoa sofre um acidente ou algum tipo de violência. Elas apresentam um grande problema quando se trata de sequelas físicas, temporárias ou permanentes, ocorrendo geralmente com a população mais jovem, ativa e do sexo masculino. No Brasil a principal faixa etária é entre 10 e 39 anos, com alguns casos na faixa etária mais de 60 anos que são considerados idosos. Como também, as internações por causas externas representaram 10% de todas as internações hospitalares realizadas pelo SUS e por hospitais particulares em 2019, sendo as duas causas com maior frequência as internações por quedas (34,9%) e acidentes de transporte terrestre (15,5%), de acordo com Mascarenhas e Barros (2015).

O perfil que gera as internações por causas externas da população brasileira tem sido um dos fatores mais estudado por pesquisas que abordam as características epidemiológicas das vítimas, a magnitude e a tendência desses agravos, bem como a qualidade dos dados. Tais análises utilizam os dados disponibilizados pelo Departamento de Informática do Sistema Único de Saúde (DATASUS), em particular, o Sistema de Internação Hospitalar do SUS (SIHSUS) que fornece, informações que são extremamente analisadas e amplamente divulgadas para o planejamento de intervenção.

O presente estudo tem o intuito de analisar as causas externas que levaram crianças e adultos a serem internados, considerando todas as internações no Brasil no ano de 2019, abordando os motivos que levaram a tal situação. Um estudo preliminar com uma amostra de dados obtidos no DATASUS, observou-se que de todas as situações de internação, percebe-se que tanto para o sexo masculino quanto feminino geram situações graves que provocam vários tipos de traumas. Uma das formas de se verificar a gravidade ou não da internação pode estar relacionado ao tempo que o paciente fica internado. Desta forma, o objeto principal de estudo deste projeto é o tempo de internação, sendo o início do período de estudo, a data de internação e o término do estudo, a alta do paciente. Desta forma, no caso de saída do paciente do hospital por outras causas, por exemplo, a morte do paciente, será considerado como censura. O estudo do tema “análise de sobrevivência” se

justifica como uma das técnicas que poderia ser considerada para a análise, considerando a presença de informações incompletas (censura). A comparação poderá ser feita por meio da construção de modelos de regressão em análise de sobrevivência, que poderá ser utilizada para estudar as causas que levam a um tempo maior de internação.

O objetivo geral deste trabalho é analisar o tempo de internação de pacientes no sistema SUS ocasionados por causas externas no Brasil no ano de 2019. Esta análise será realizada por meio da construção de modelos de regressão em análise de sobrevivência. Desta forma, será realizado as análises exploratórias dos dados do Sistema Único de Saúde, através do DATASUS, em específico com os dados de Internação hospitalar constante na base de dados Sistema de Internações Hospitalares do SUS (SIHSUS), como também um estudo das técnicas não paramétricas no contexto de análise de sobrevivência, o estudo de alguns modelos paramétricos adequados para a construção de modelos de regressão utilizado em estudos na área da saúde.

2 DESENVOLVIMENTO DO TRABALHO

Esse projeto tem como objetivo analisar os casos de internações hospitalares por causas externas, focando na distribuição dos casos por municípios e as principais causas de ocorrência no Brasil. A base de dados foi obtida através do Departamento de Informática do Sistema Único de Saúde (DATASUS), em particular, a base de dados do SIHSUS.

A análise foi feita a partir do software Rstudio e/ou Python, para a realização dos gráficos e filtrar os dados, utilizando a base do DATASUS no ano de 2019. Foram aplicadas técnicas estatísticas, como análise exploratória, construção de modelos de regressão em análise de sobrevivência, com a abordagem paramétrica e semiparamétrica.

A modelagem estatística é um ramo da estatística que se refere ao conjunto de técnicas probabilísticas para representar a realidade dos casos de internações hospitalares no Brasil, de forma mais simplificada. Esse método é amplamente utilizado para criar uma representação visual ou até mesmo um esquema, que define as possíveis ocorrências de casos de internação, no contexto a ser estudado.

A análise de sobrevivência é utilizada quando o tempo for o objeto de interesse, como no caso desse projeto. Em particular, o objeto de estudo será o tempo de internação de pacientes hospitalizados tendo como causa principal, acidentes ou algum tipo de violência, sendo considerado como início do estudo, a data de internação e o evento de interesse, a alta hospitalar. Desta forma, foram estudados alguns aspectos para aplicação dos modelos de regressão, iniciando com os modelos de regressão linear simples e múltipla. Posteriormente são apresentadas as principais distribuições de probabilidade utilizados na área de análise de sobrevivência. Assim, será apresentado os modelos de regressão em análise de sobrevivência.

2.1 Modelo de Regressão Linear Simples

Sabendo que um conjunto de n pares de valores para duas variáveis, X_i e Y_i (com i=1, 2, ..., n), se admitirmos que Y é uma função linear de X, podemos estabelecer uma regressão linear simples. O modelo estatístico dado será:

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

onde α e β são parâmetros, X é a variável explanatória e Y é a variável dependente.

O coeficiente angular da reta (β) é também denominado coeficiente de regressão, e o coeficiente linear da reta (α) é conhecido como termo constante da equação de regressão e $\varepsilon \sim N\ (0, \sigma^2)$

2.2 Modelo de Regressão Linear Multipla

Trata-se de quando é admitido que o valor da variável dependente é função linear de duas ou mais variáveis explanatórias. Logo, o modelo estatístico de uma regressão linear múltipla com k variáveis explanatórias é:

$$\mathrm{Y}_j = \alpha + \beta\,\mathrm{X}_{1j} + \beta_2 \mathrm{X}_{2j} + \cdots + \beta_k \mathrm{X}_{kj} + \varepsilon_j \qquad j = 1, 2, \ldots, n$$

ou

$$\mathrm{Y}_j = \alpha + \sum_{i=1}^{k} \beta_i\ \mathrm{X}_{ij} + \varepsilon_j \quad (1)$$

Fazendo a notação matricial do modelo, ficará:

$$Y = X\beta + \varepsilon$$

onde

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad X = \begin{bmatrix} 1 & X_{11} & \ldots & X_{k1} \\ 1 & X_{12} & \ldots & X_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & \ldots & X_{kn} \end{bmatrix}$$

$$\beta = \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \qquad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Será mantido as seguintes pressuposições que foram apresentadas no item (1) de regressão linear simples, seguindo os critérios abaixo:

I) a variável dependente (Yj) é função linear das variáveis explanatórias X_{ij}, em que $i = 1, \ldots, k\ e\ j = 1, \ldots, n$

II) os valores das variáveis explanatórias são fixos;

III) $E(\varepsilon_j) = 0$, ou seja, $E(\varepsilon) = \mathbf{0}$, onde esse **0** representa um vetor de zeros e portanto, $\varepsilon \sim N(\mathbf{0}, \mathbf{\Sigma})$, de tal forma que $\mathbf{\Sigma}$ é a matriz de variância e covariância.

IV) os erros são homocedásticos, isto é, $E(\varepsilon_j^2) = \sigma^2$;

V) os erros são não-correlacionados entre si, isto é, $E(\varepsilon_j\, \varepsilon_h) = 0$ para j ≠ h;

Realizando a combinação das pressuposições IV e V teremos $E(\varepsilon\varepsilon') = I\sigma^2$

Seja $p = k + 1$ o número de parâmetros a serem estimados (α, $\beta_1, \beta_3, \ldots, \beta_k$). Se dispor de apenas p observações, a determinação dos parâmetros se reduz a um problema matemático de resolução de um sistema de p equações com p incógnitas, não sendo possível fazer qualquer análise estatística. Devemos, portanto, ter $n > p$. Além disso, veremos que, para obter as estimativas de mínimos quadrados dos parâmetros, a matriz $X'X$ deve ser não singular, ou seja, seu posto deve ser igual a p. Isso significa que o posto de X deve ser igual a p. Nas deduções que se seguem, admitiremos que essas condições são observadas, isto é, admitiremos que X tem posto $p = k + 1 < n$.

Assim como na regressão linear simples, as suposições I, II e III são essenciais para demonstrar que os estimadores de mínimos quadrados são não tendenciosos. As cinco primeiras suposições permitem demonstrar que esses estimadores são os melhores estimadores lineares não tendenciosos (conforme o teorema de Gauss-Markov).

2.3 Análise de Sobrevivência

A análise de sobrevivência é uma área da estatística que cresceu significativamente nas últimas duas décadas do século XX, impulsionada pelo avanço das técnicas estatísticas e pela melhoria dos computadores. Esse crescimento é evidente pelo aumento de aplicações dessa análise na medicina e em diversas áreas cuja variável resposta é o tempo de ocorrência de um evento. Em análise de sobrevivência, a variável resposta é geralmente o tempo

até a ocorrência de um evento de interesse, chamado tempo de falha, que pode ser o tempo até a morte do paciente, a cura ou a recidiva de uma doença.

Neste trabalho, o ajuste dos modelos de sobrevivência foi realizado utilizando a distribuição Gama, Weibull, Exponencial e Log Normal que será detalhada nos tópicos a seguir.

2.3.1 Distribuição Gamma

Sabendo que X é uma variável aleatória que tem por finalidade mostrar o tempo até um evento específico. A distribuição Gamma é parametrizada por dois parâmetros: a forma e a escala. A função densidade de probabilidade da distribuição Gamma é dada por:

$$f(x|k,\theta) = \frac{1}{\theta^k \Gamma(k)} x^{k-1} \exp\{-x/\theta\}$$

onde $\theta > 0$ e $k > 0$ são os parâmetros de escala e a forma da distribuição, respectivamente. E $\Gamma(k)$ é a função gama definida como:

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

No entanto, a função de sobrevivência $S(x)$ para a distribuição Gamma é dada pela probabilidade de que a variável aleatória X seja maior que um valor específico x: $S(x|k,\theta) = 1 - \frac{\gamma(k,\theta x)}{\Gamma(k)}$, onde $\gamma(k,\theta x)$ é a função gama incompleta inferior.

É categorizado uma função de risco $h(x)$ quando a taxa de falha é imediata no tempo x, dada pela razão entre a função densidade de probabilidade e a função de sobrevivência:

$$h(x|k,\theta) = \frac{\frac{1}{\theta^k \Gamma(k)} exp\left(-\frac{x}{\theta}\right)}{1 - \frac{\gamma\left(k, \frac{x}{\theta}\right)}{\Gamma(k)}}$$

Contudo, para fazer uma estimação de máxima verossimilhança de forma pontual, será necessário considerar uma amostra aleatória $x_1, x_2, \dots, x_n$ obtida da distribuição Gamma, a função de verossimilhança para os parâmetros k e θ é o produto das densidades de cada uma das observações na amostra, tem se que:

$$L(k, \theta|x) = \prod_{i=1}^{n} \frac{1}{\theta^k \Gamma(k)} x_i^{k-1} exp\left(\frac{x_i}{\theta}\right)$$

Desta forma, a log-verossimilhança será dada por:

$$l(k, \theta|x) = -nlog\Gamma(k) - nklog\theta + (k-1)\sum_{i=1}^{n} logx_i - \frac{1}{\theta}\sum_{i=1}^{n} x_i$$

Os estimadores de máxima verossimilhança $\hat{k}$ e $\hat{\theta}$ são obtidos resolvendo as equações de verossimilhança $\frac{\partial l}{\partial k} = 0$ e $\frac{\partial l}{\partial \theta} = 0$ respectivamente.

Além da estimação de verossimilhança pontual, também se possui a de intervalos de confiança para os parâmetros do modelo que pode ser gerada pela aproximação normal assintótica dos estimadores de máxima verossimilhança. Desta forma, pode-se supor que os estimadores $\hat{k}$ e $\hat{\theta}$ seguem uma distribuição normal conjunta, os intervalos de confiança para k e θ com coeficiente de confiança $100(1-\alpha)\%$ são dados por:

$$\hat{k} - z_{\alpha/2}\sqrt{Var(\hat{k})} < k < \hat{k} + z_{\alpha/2}\sqrt{Var(\hat{k})}$$

$$\hat{\theta} - z_{\alpha/2}\sqrt{Var(\hat{\theta})} < \theta < \hat{\theta} + z_{\alpha/2}\sqrt{Var(\hat{\theta})}$$

Tendo que $z_{\alpha/2}$ é o quantil da distribuição normal padrão referente ao nível de confiança que se deseja.

2.3.2 Distribuição Weibull

Sabendo que X uma variável aleatória que representa o tempo até a falha. A distribuição Weibull é parametrizada por dois parâmetros: forma e escala. A função densidade de probabilidade da distribuição Weibull é dada por:

$$f(x|\beta,\lambda) = \frac{\beta}{\lambda}\left(\frac{x}{\lambda}\right)^{\beta-1} exp\left(-\left(\frac{x}{\lambda}\right)^{\beta}\right)$$

tendo que o $\beta > 0$ e $\lambda > 0$ são os parâmetros de forma e escala da distribuição Weibull, simultaneamente.

No entanto, a função de sobrevivência $S(x)$ para a distribuição Weibull tem-se que é a probabilidade de que a variável aleatória X seja maior que um valor específico x: $S(x) = P(X > x) = \exp\left(-\left(\frac{x}{\lambda}\right)^{\beta}\right)$. Contudo, também é categorizado uma função de risco $h(x)$ quando a taxa de falha é imediata no tempo x, dada que pela razão entre a função densidade de probabilidade e a função de sobrevivência, teremos:

$$h(x|k,\theta) = \frac{\beta}{\lambda}\left(\frac{x}{\lambda}\right)^{\beta-1}$$

Contudo, para fazer uma estimação de máxima verossimilhança de forma pontual, será necessário considerar uma amostra aleatória $x_1, x_2, \dots, x_n$ obtida da distribuição Weibull, a função de verossimilhança para os parâmetros β e λ é o produto das densidades de cada uma das observações na amostra, dada por:

$$L(\beta,\lambda|x) = \prod_{i=1}^{n} \frac{\beta}{\lambda}\left(\frac{x_i}{\lambda}\right)^{\beta-1} exp\left(-\left(\frac{x_i}{\lambda}\right)^{\beta}\right)$$

Desta forma, a log-verossimilhança será dada por:

$$l(\beta,\lambda|x) = -nlog\beta - n\beta log\lambda + (\beta-1)\sum_{i=1}^{n} logx_i - \sum_{i=1}^{n}\left(\frac{x_i}{\lambda}\right)^{\beta}$$

Logo, os estimadores de máxima verossimilhança $\hat{\beta}$ e $\hat{\lambda}$ são obtidos resolvendo as equações de verossimilhança $\frac{\partial l}{\partial \beta} = 0$ e $\frac{\partial l}{\partial \lambda} = 0$ simultaneamente.

Além da estimação de verossimilhança pontual, também se possui a de estimação com intervalos de confiança para os parâmetros do modelo que pode ser gerada pela aproximação normal também assintótica dos estimadores de máxima verossimilhança. Desta forma, pode-se supor que os estimadores $\hat{\beta}$ e $\hat{\lambda}$ seguem uma distribuição normal conjunta e os intervalos de confiança para β e λ com coeficiente de confiança $100(1-\alpha)\%$ será:

$$\hat{\beta} - z_{\alpha/2}\sqrt{Var(\hat{\beta})} < \beta < \hat{\beta} + z_{\alpha/2}\sqrt{Var(\hat{\beta})}$$

$$\hat{\lambda} - z_{\alpha/2}\sqrt{Var(\hat{\lambda})} < \lambda < \hat{\lambda} + z_{\alpha/2}\sqrt{Var(\hat{\lambda})}$$

tendo que o $z_{\alpha/2}$ é o quantil da distribuição normal padrão referente ao nível de confiança que se deseja.

2.3.3 Distribuição Exponencial

A distribuição da exponencial, tem por finalidade ser um dos modelos probabilísticos mais simples para descrever o tempo de falha. Logo, essa distribuição apresenta um único parâmetro e é a única que se caracteriza por ter uma função de taxa de falha constante.

Desta forma, a função de densidade de probabilidade para a variável aleatória tempo de falha T com distribuição exponencial será dada por:

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

em que o parâmetro $\alpha > 0$ é o tempo de vida médio. Logo, o parâmetro α tem a mesma unidade do tempo de falha t. Com isso, pode se concluir que se t, é medido em horas, α também será fornecido em horas. Observe que a Distribuição exponencial é um caso particular da distribuição Weibill.

2.3.4 Distribuição Log Normal

Assim como a distribuição de Weibull, a distribuição log-normal é amplamente utilizada para caracterizar tempos de vida de produtos e indivíduos. Isso inclui casos de fadiga de metal, semicondutores, diodos e isolamento elétrico. Além disso, é frequentemente utilizada para descrever situações clínicas, como o tempo de vida de pacientes que sofreram acidentes por causas externas ou que têm leucemia, entre outros.

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

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

Em que μ é a média do logaritmo do tempo de falha, e σ é o desvio-padrão. Existe uma relação entre as distribuições log-normal e normal, semelhante à relação entre as distribuições de Weibull e do valor extremo. Essa relação facilita a apresentação e análise de dados provenientes da distribuição log-normal. Como o nome sugere, o logaritmo de uma variável com distribuição log-normal possui uma distribuição normal com média μ e desvio-padrão σ. Isso significa que dados provenientes de uma distribuição log-normal podem ser analisados como se fossem provenientes de uma distribuição normal, desde que se considere o logaritmo dos dados em vez dos valores originais. As funções de sobrevivência e de taxa de falha de uma variável log-normal não possuem uma forma analítica explícita e são, portanto, representadas, respectivamente, por:

$$S(t) = \Phi\left(\frac{-\log(t) + \mu}{\sigma}\right) \quad \text{e } \lambda(t) = \frac{f(t)}{S(t)}$$

Observando as funções de taxa de falha não são monótonas como as das distribuições Weibull, os percentis para a distribuição log-normal podem ser obtidos a partir da tabela da normal padrão, utilizando a expressão:

$$t_p = \exp\{z_p\sigma + \mu\}$$

como z_p o $100\rho\%$ percentil da distribuição normal padrão. Logo a média e a variância da log-normal são dados, respectivamente, por

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

3 RESULTADOS

3.1 Organização dos dados

Os dados foram obtidos através do DATASUS e analisados pelo programa Rstudio e/ou Python, onde todo o processo de tratamento, modelagem e análise estatística foi realizado.

Desta forma, foi selecionado 26 estados do Brasil e um Distrito Federal no ano de 2019, para fazer a análise exploratória de dados e testar o programa computacional.

Essa base possui um total de aproximadamente 13 milhões de observações e 113 variáveis para cada estado como apresentado na Tabela 1, mas que foi selecionado apenas as variáveis de interesse para que tivesse um resultado mais preciso. Logo, as variáveis de interesse são o sexo, idade, dias de permanência na unidade hospitalar e a causa da internação dos pacientes. A fim de analisar o tempo de internação de pacientes, se há uma tendência entre o sexo masculino ou feminino a passar por acidentes por causas externas e em qual faixa etária tende a acontecer com maior frequência através do sistema SIHSUS no ano de 2019.

Tabela 1 - Observações de cada estado

ESTADOS	OBSERVAÇÕES	VARIÁVEIS
ACRE	44.889	113
ALAGOAS	177.680	113
AMAPÁ	40.876	113
AMAZONAS	196.792	113
BAHIA	850.173	113
CEARÁ	531.283	113
DISTRITO FEDERAL	230.592	113
ESPIRÍTO SANTO	247.180	113
GOIÁS	350.564	113
MARANHÃO	465.752	113
MATO GROSSO	196.702	113
MATO GROSSO DO SUL	178.522	113
MINAS GERAIS	1.292.500	113

PARÁ	484.950	113
PARAIBA	191.653	113
PARANÁ	896.060	113
PERNAMBUCO	584.616	113
PIAUÍ	220.790	113
RIO DE JANEIRO	754.725	113
RIO GRANDE DO NORTE	185.353	113
RIO GRANDE DO SUL	771.586	113
RONDÔNIA	122.900	113
RORAIMA	44.872	113
SÃO PAULO	2.606.482	113
SANTA CATARINA	513.237	113
SERGIPE	94.923	113
TOCANTINS	81.159	113

Fonte: Elaborado pela Autora

3.2 Análise Descritiva

Refere-se a uma abordagem fundamental para entender e resumir conjuntos de dados. Esta etapa pode direcionar possíveis técnicas de análise estatística que serão realizadas. Ela envolve a coleta, apresentação e caracterização dos dados de maneira a fornecer uma visão clara sobre as características principais do conjunto de dados sem fazer inferências ou predições. Logo, será apresentado a tabela de cada estado referente a moda, mediana, frequência masculino e feminino, proporção masculino e feminino.

Observando as tabelas 2 e 3 apresentadas a seguir, nota-se que o para todos os estados o tempo mínimo do paciente internado por causas externa é de 01 dia, considerando assim algo que não seja “grave” e o dia que um paciente ficou mais internado, foi um total de 315 dias no estado de São Paulo no ano de 2019. A base do DATASUS consta que 01 é o sexo masculino e 03 o sexo feminino, onde a idade mínima de internação foi de pacientes com 0 anos de idade, ou seja, bebês até pessoas idosas de 99 (não considerado ou não informado pelo sistema do SIHSUS) tendem a sofrer mais algum tipo de acidente por causa externa.

As causas mais evidentes serão apresentadas no item 3.2.6, assim teremos conhecimento do que mais impulsiona o tipo de internação.

Na Tabela 2 tem-se as medidas descritivas das observações com relação a variável sexo, tendo que FM (Frequência Masculino), FF (Frequência Feminino), PM (Proporção Masculino) e PF (Proporção do Feminino).

Tabela 2: Medidas Descritivas das Observações (Sexo)

Estados	**FM**	**FF**	**PM**	**PF**
Acre	3231	1137	0.739698	0.260302198
Amazonas	10206	4359	0.700721	0.299279094
Amapá	1739	718	0.707774	0.292226292
Pará	39305	16053	0.710015	0.289985187
Rondônia	11331	4805	0.702219	0.297781358
Roraima	2728	1021	0.727661	0.27233929
Tocantins	8215	3138	0.723597	0.276402713
Alagoas	10182	5177	0.662934	0.337066215
Bahia	56328	24253	0.699023	0.300976657
Ceará	33762	15213	0.689372	0.310627871
Maranhão	27998	12240	0.69581	0.304190069
Paraíba	10344	4966	0.675637	0.324363161
Pernambuco	39332	18768	0.676971	0.32302926
Piauí	18387	7191	0.71886	0.281140042
Rio Grande do Norte	13829	6309	0.686712	0.313288311
Sergipe	5723	2352	0.708731	0.29126935
Goiás	33925	15703	0.683586	0.316414121
Mato Grosso	20351	8457	0.706436	0.293564288
Mato Grosso do Sul	15651	7458	0.677269	0.322731403
Distrito Federal	14023	6351	0.688279	0.311720821
Minas Gerais	91358	45259	0.668716	0.331283808
Espírito Santo	18132	8169	0.689403	0.310596555

Rio de Janeiro	51201	26256	0.661025	0.338975173
São Paulo	162560	83798	0.659853	0.340147265
Santa Catarina	35487	17618	0.668242	0.331757838
Rio Grande do Sul	39729	23539	0.627948	0.372052222
Paraná	64016	33602	0.655781	0.344219304

Fonte: Elaborado pela Autora

Tabela 3: Medidas Descritivas das Observações (IDADE)

Estados	**Mínimo**	**Média**	**Máximo**	**Desvio Padrão**
Acre	0	34.89	99	19.36781
Amazonas	0	34.96	99	20.88934
Amapá	0	35.56	99	21.54909
Pará	0	35.46	99	20.40972
Rondônia	0	36.27	99	20.30603
Roraima	0	34.45	99	18.87445
Tocantins	0	36.93	99	20.9235
Alagoas	0	38.78	99	21.60006
Bahia	0	38.74	99	21.54682
Ceará	0	40.07	99	22.14424
Maranhão	0	35.89	99	21.6242
Paraíba	0	40.94	99	22.46774
Pernambuco	0	40.42	99	22.1401
Piauí	0	39.15	99	20.97548
Rio Grande do Norte	0	42.38	99	22.77455
Sergipe	0	40.05	99	21.24629
Goiás	0	40.78	99	20.85435
Mato Grosso	0	37.85	99	20.22501
Mato Grosso do Sul	0	39.61	99	21.8625

Distrito Federal	0	37.63	99	20.96087
Minas Gerais	0	43.30	99	22.38053
Espírito Santo	0	41.42	99	20.80616
Rio de Janeiro	0	43.05	99	22.48483
São Paulo	0	42.98	99	22.17603
Santa Catarina	0	45.23	99	21.32793
Rio Grande do Sul	0	41.22	99	23.01716
Paraná	0	41.92	99	22.02783

Fonte: Elaborado pela Autora

Tabela 4: Medidas Descritivas das Observações (DIAS DE PERMANÊNCIA)

Estados	**Mínimo**	**Média**	**Máximo**	**Desvio Padrão**
Acre	1	5.982	147	9.35391
Amazonas	1	7.124	86	8.024236
Amapá	1	10.74	196	15.67245
Pará	1	4.728	305	5.740402
Rondônia	1	6.609	243	8.87073
Roraima	1	10.07	170	14.02792
Tocantins	1	9.32	245	10.77028
Alagoas	1	5.231	120	6.500822
Bahia	1	5.018	156	6.212139
Ceará	1	7.149	279	9.559642
Maranhão	1	5.552	305	6.793086
Paraíba	1	7.49	101	7.401982
Pernambuco	1	6.092	125	6.732112
Piauí	1	5.253	288	6.277718
Rio Grande do Norte	1	6.98	107	8.080678
Sergipe	1	7.03	136	9.286049
Goiás	1	4.476	274	5.360355

Mato Grosso	1	4.574	155	5.595260
Mato Grosso do Sul	1	4.525	106	5.508964
Distrito Federal	1	7.431	118	9.12912
Minas Gerais	1	4.857	123	5.835117
Espírito Santo	1	5.629	90	6.103159
Rio de Janeiro	1	7.774	208	9.765871
São Paulo	1	5.127	315	6.735569
Santa Catarina	1	4.205	267	5.347397
Rio Grande do Sul	1	5.938	151	7.565715
Paraná	1	3.837	170	4.621891

Fonte: Elaborado pela Autora

3.2.1 Sexo dos Pacientes

A análise da tabela 1 no item 3.2, evidencia que as internações por causas externas no Brasil em 2019 foram majoritariamente masculinas, com proporção média nacional acima de 65%. Essa prevalência reflete padrões sociais e comportamentais que expõem mais os homens a fatores de risco. Estados como Acre (0,739) e Amapá (0,707) destacam-se com as maiores proporções masculinas, enquanto Rio Grande do Sul (0,627) e Paraná (0,655) têm uma distribuição mais equilibrada.

No entanto, por mais que os homens sejam maioria, há uma participação proporcionalmente maior de mulheres nesses estados, o que pode ser devido a diferentes padrões regionais de comportamento e exposição a causas externas no Brasil através de cada estado.

3.2.2 Idade dos pacientes

Analisando a tabela 2 no item 3.2, pode-se notar que em todos os estados, a idade mínima dos pacientes internados é 0 anos (bebês ou crianças), enquanto a idade máxima é de 99 anos. Logo, isso indica que as causas externas afetam todas as faixas etárias.

A maior parte dos estados apresentam médias entre 35 e 45 anos, indicando que adultos jovens e de meia-idade são os mais afetados por causas

externas. O desvio padrão varia de aproximadamente 18,74 (Roraima) a 23,48 (Rio de Janeiro), indicando dispersões significativas nas idades.

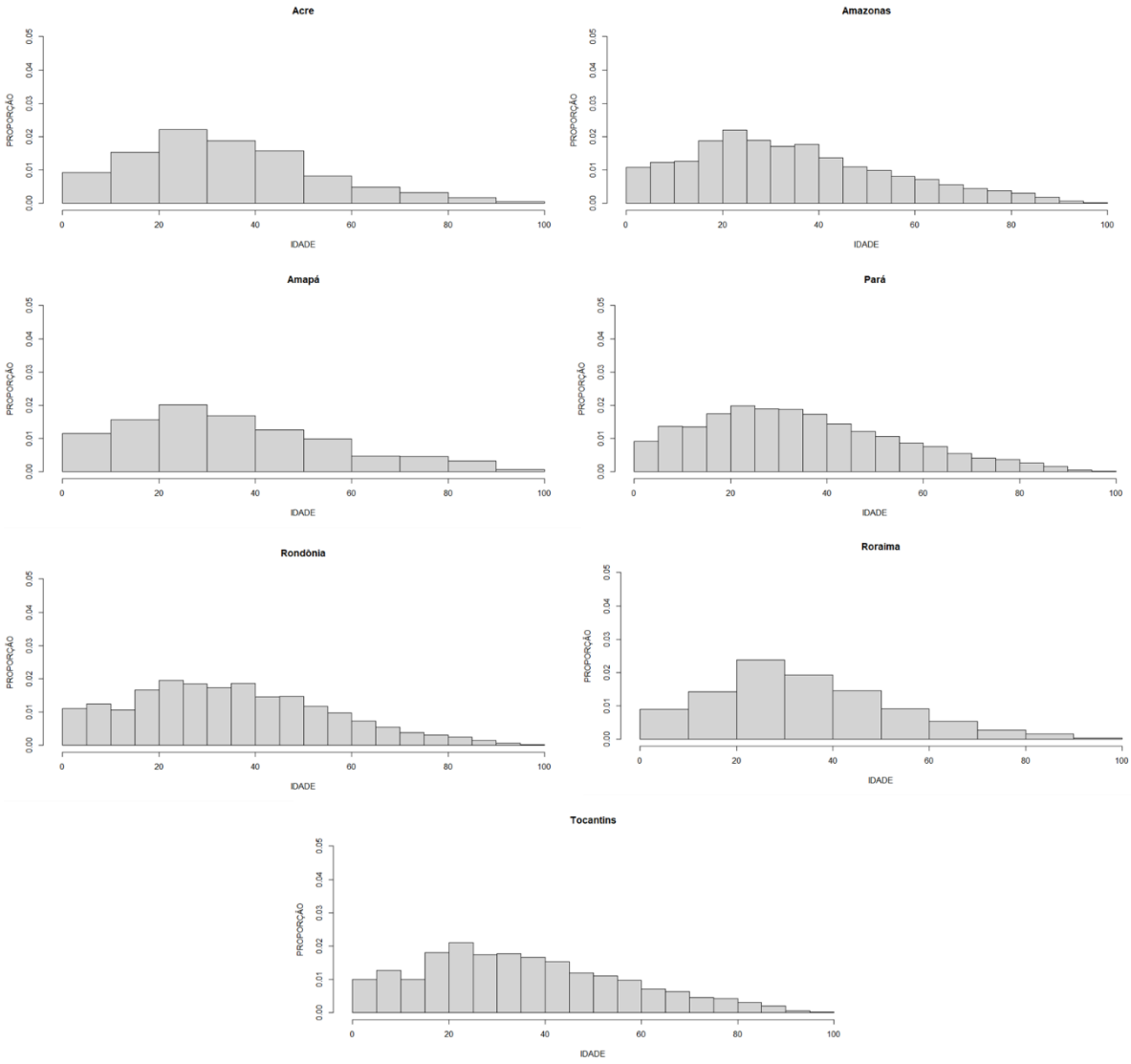
Estados com idade média mais baixa, como Acre e Roraima, podem estar lidando com causas externas que afetam populações mais jovens, como acidentes envolvendo crianças e jovens. Já os estados com idade média mais alta, como Santa Catarina, podem estar enfrentando causas externas mais relacionadas a acidentes envolvendo adultos e idosos.

As causas externas afetam todas as faixas etárias, mas o impacto é maior em adultos jovens e de meia-idade. O desvio padrão alto em vários estados sugere que as internações por causas externas não estão restritas a um grupo etário específico, mas abrangem uma ampla faixa de idades.

A região norte, de forma geral, apresenta idades médias entre 34,48 e 36,48. A menor média nacional está no estado do Acre, sugerindo que as causas externas nessa região podem impactar populações mais jovens.

Esse impacto acontece, pois segundo o Portal Quatro Rodas, homens sofrem mais acidentes que as mulheres em todas as idades e em todos os estados do Brasil, porém são entre os mais jovens (entre 18 e 34 anos) que observamos o maior número de casos para ambos os sexos. De acordo com a UNICAMP, os dias da semana em que ocorrem mais acidentes é sexta-feira e sábado, onde foram considerados os dias mais perigosos.

Figure 1: Idade dos Pacientes (Região Norte)



Fonte: Elaborado pela Autora

Já na região nordeste do país, a predominância da faixa etária de 15 a 40 anos, pois como nos demais são os mais tendenciosos a sofrerem acidentes de trânsito por imprudência, impulsividade, quedas, acidentes domésticos entre outros.

A gerente de Educação para o Trânsito (Gedut), Mirian Bastos, destaca-se que o motivo pelo qual homens costumam se envolver em mais acidentes do que mulheres está relacionado ao comportamento dos motoristas. Pois, já é algo considerado como cultural devido ao comportamento de competitividade, falta de respeito e/ou atenção as regras de convivência. Logo, sendo as

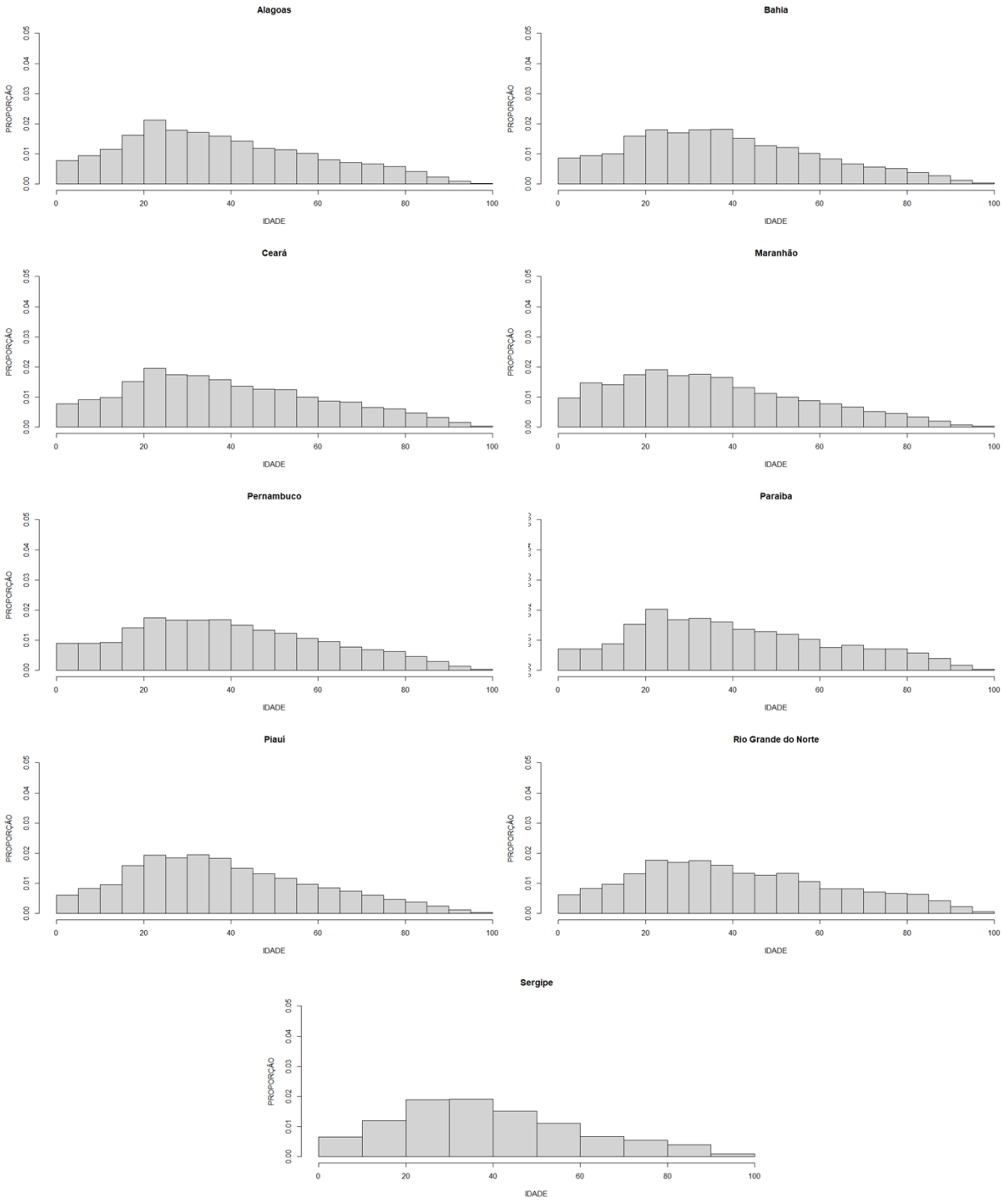
mulheres mais prudentes e atentas as normas de segurança, porém também existe casos de mulheres que causam acidentes por falta de atenção.

Cinco fatores de risco foram estabelecidos como os principais a serem enfatizados: velocidade, bebida alcoólica, capacete, cinto de segurança e cadeirinha de criança. Esses cinco são fatores de risco proximais (NADANOVSKY, SANTOS, 2021). A OMS preconiza que todos os países devem criar leis específicas para enfrentar esses cinco fatores de risco proximais no trânsito: limitar velocidade; impedir ingestão de bebida alcoólica; usar capacete; utilizar cinto de segurança; usar cadeirinha de contenção de criança (WHO, 2018).

No entanto, o Ministério da Saúde, retrata que o mais comum, são idosos e geralmente são vítimas de queda. Crianças as que a gente mais recebe, são vítimas de queimadura, perfuração por objeto cortante, ingestão de produto de limpeza e queimadura em olhos por produto de limpeza, uso de medicamento, manipulação de produtos inflamáveis, acidentes elétricos com tomadas.

No entanto o estado do Maranhão chama a atenção devido a um indicie um pouco mais alto do que os demais histogramas na faixa etária menor que 15 anos, normalmente isso ocorre por acidentes com brinquedos que estão danificados, parques infantis, brincadeiras nas ruas ou em área de construção sem a supervisão de adulto.

Figura 2: Idade dos Pacientes (Região Nordeste)



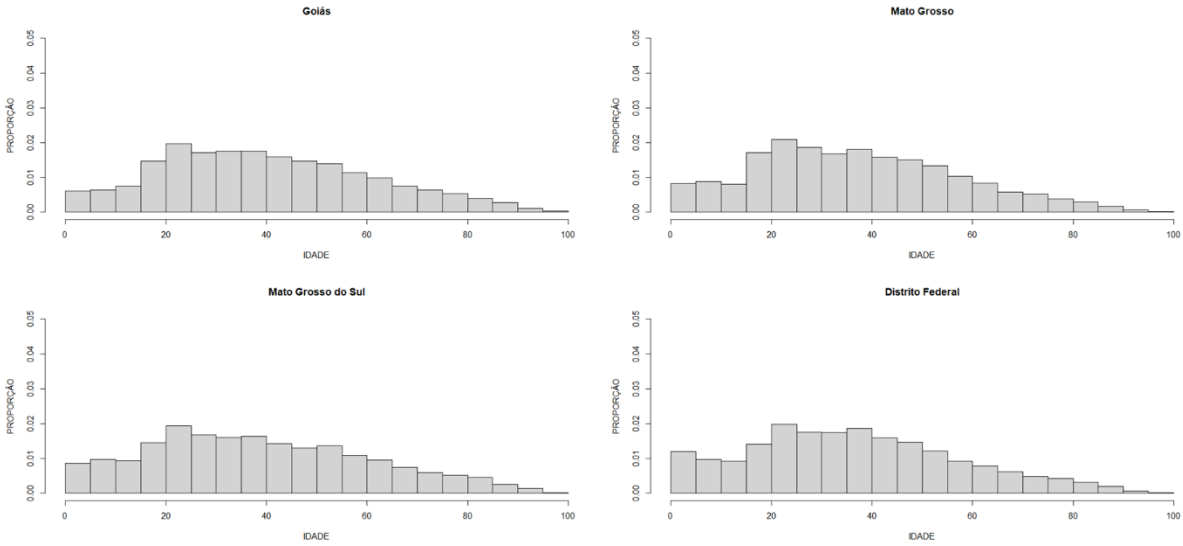
Fonte: Elaborado pela Autora

Na região centro-oeste, a faixa etária que mais sofre acidentes por causas externas é também a de adultos jovens, especialmente entre 15 e 39 anos. Esse padrão é comum em muitas regiões do Brasil, refletindo fatores

como o aumento de acidentes de trânsito, violência e acidentes no ambiente de trabalho, que tendem a afetar mais intensamente essa faixa etária.

Os esportes radicais são atividades que envolvem alto nível de adrenalina e risco. Eles são caracterizados por desafiar os limites físicos e mentais dos praticantes como por exemplo o paraquedismo, rapel, escalada, rafting e muitos outros. Essas atividades atraem tanto os moradores locais quanto os turistas, que buscam aventura e diversão na região Centro-Oeste do Brasil (LUCA VERO).

Figura 3: Idade dos Pacientes (Região Centro-Oeste)



Fonte: Elaborado pela Autora

No entanto a região sudeste, também prevalece a mesma faixa etária das regiões apresentadas anteriormente. Os motivos pelo qual acontece, também são os mesmos pelo fato de ser uma região extremamente populosa e com muitas indústrias.

Porém, o que chama a atenção é que nesta região o caso de pessoas na meia idade (40 a 59 anos) e idosos (mais de 60), observa-se um aumento significativo de acidentes por causas externas. A partir dos 40 anos, muitas pessoas já possuem uma rotina de trabalho intensa, o que as leva a estar no trânsito com frequência. Isso pode aumentar o risco de acidentes,

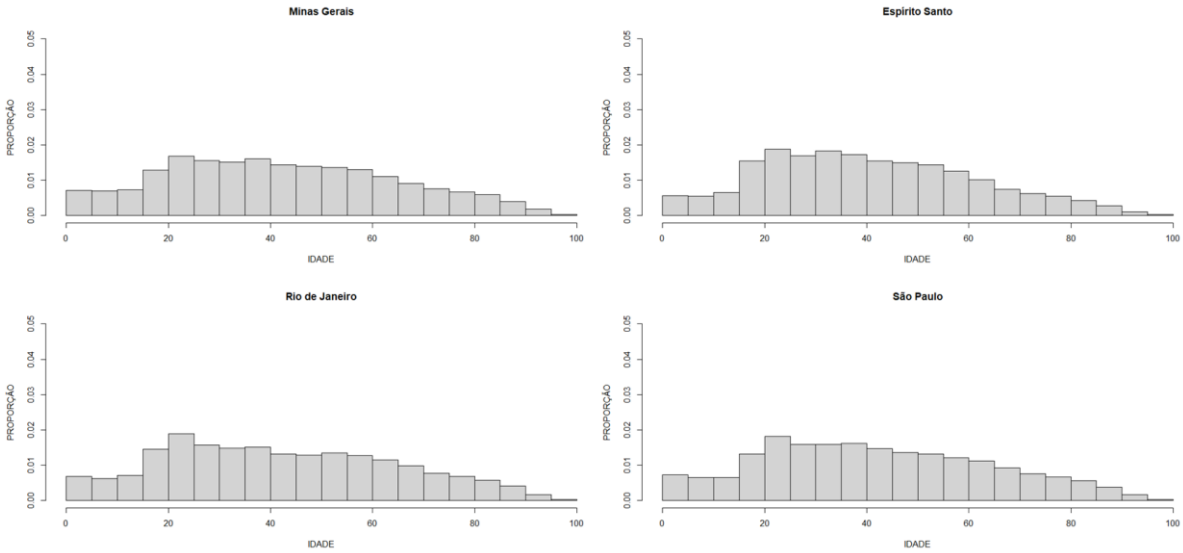
especialmente considerando a alta densidade de veículos e o trânsito intenso, como por exemplo São Paulo.

O aumento no uso de motocicletas tanto como meio de transporte quanto como atividade profissional (motoboys, por exemplo) pode impactar mais a faixa etária intermediária. A maior vulnerabilidade de quem usa moto pode explicar uma parte dos acidentes.

A Organização Mundial da Saúde, fizeram um chamado de alerta para a todos os países, para que fosse intensificado os serviços de saúde e apoio mental. Pois, esse também foi um dos fatores que contribuiu também para o aumento de estresse, ansiedade e depressão.

Onde, pode influenciar a acidentes por causas externas como por exemplo acidentes domésticos, podar uma árvore, excesso de álcool, utilização de ferramentas sem proteção de EPI (Equipamentos de Proteção Individual), alguns tipos de doenças como por exemplo o diabete e problemas cardiovasculares que podem afetar a coordenação motora do indivíduo.

Figura 4: Idade dos Pacientes (Região Sudeste)

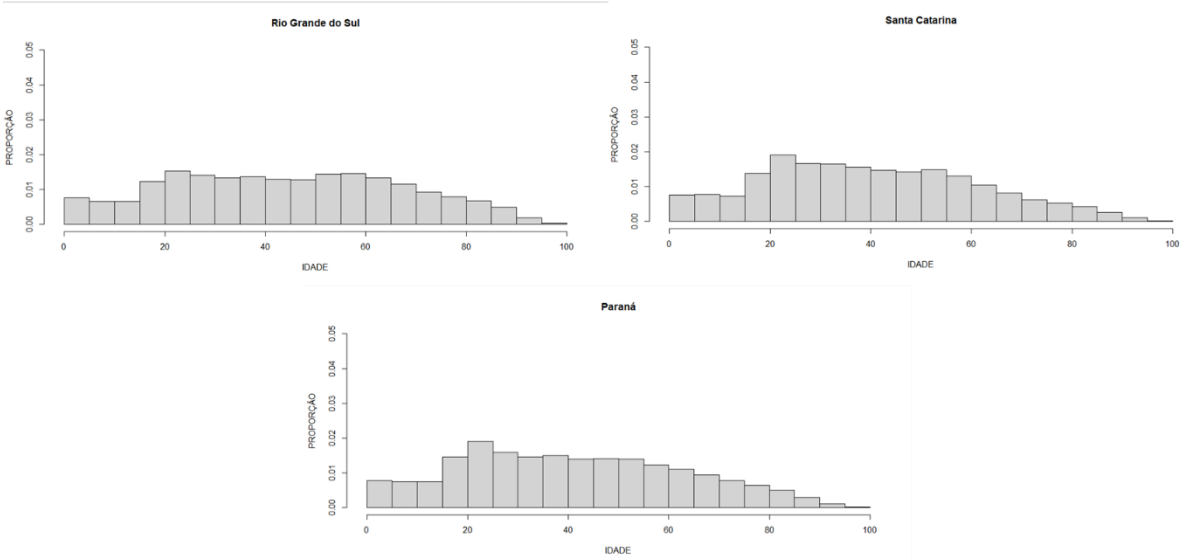


Fonte: Elaborado pela Autora

Por fim, a região sul, apresenta casos de crianças, que por sua imaturidade, curiosidade e intenso crescimento e desenvolvimento, encontra-se muitas vezes propensa a acidentes, e indefesa e vulnerável a violências.

Somando-se a esses fatores, a alta incidência de causas externas em crianças e adolescentes tem despertado em todo o mundo a necessidade de estudos desses eventos na população infanto-juvenil, segundo a Revista Brasileira de Epidemiologia. Nesta região em específico, os causos de idosos se dá por conta da mobilidade e equilíbrio, problemas de saúde, quedas e entre outros fatores.

Figura 5: Idade dos Pacientes (Região Sul)



Fonte: Elaborado pela Autora

3.2.3 Dias de Permanência no Hospital

Os dias de permanência hospitalar de pacientes vítimas de acidentes por causas externas é crucial para otimizar a gestão hospitalar, melhorar a qualidade do atendimento, identificar padrões de lesões, e planejar políticas de saúde pública mais eficazes. Esse tipo de estudo também contribui para a formulação de estratégias de prevenção, além de oferecer uma visão mais abrangente dos impactos sociais, econômicos e de saúde dos acidentes na população.

O tempo de internação pode ser um indicador da gravidade dos acidentes e das condições de saúde do paciente. Pacientes com lesões mais severas (como traumatismos cranianos, fraturas complexas, queimaduras, etc.)

tendem a ficar mais tempo internados, enquanto lesões menos graves podem exigir um tempo de internação mais curto. Para isso, será utilizado a variável dias de permanência como tempo de internação, morte como variável de censura: se houve morte, considera-se 0 caso contrário, 1.

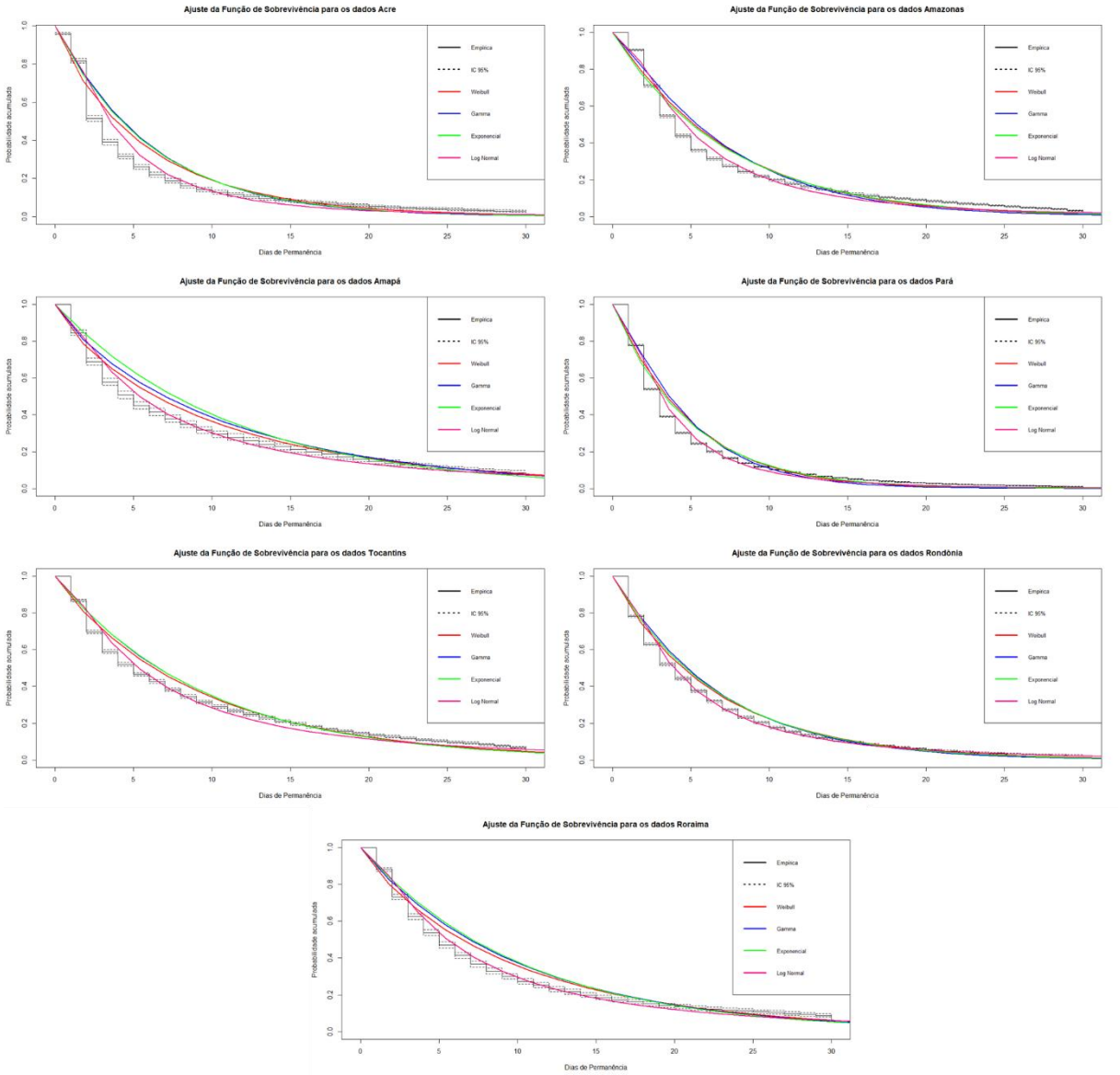
Na região norte, os gráficos mostram o ajuste da função de sobrevivência (Kaplan-Meier) para o Acre, Amazonas, Amapá, Pará, Rondônia, Roraima e Tocantins referente ao tempo de permanência de pessoas internadas no hospital por causas externas. Nestes gráficos já foram colocados também os gráficos ajustados das distribuições Weibull, Gama, Exponencial e Log-normal.

De modo geral, em todos os estados, há um número muito maior de internações com curta duração (1 a 7 dias). Essa tendência sugere que a maioria das causas externas requer internações de curta duração, possivelmente para tratamentos emergenciais que não necessitam de hospitalizações prolongadas.

Embora em números bem menores, internações prolongadas (mais de 30 dias) são observadas em todos os estados. Essas podem estar relacionadas a complicações mais graves ou reabilitações.

Em muitos casos, as curvas Weibull e Log-normal costumam ser boas opções para ajustar dados de sobrevivência, especialmente quando há variação significativa na taxa de risco ao longo do tempo. Porém, nos gráficos apresentados da região norte, nota-se que a log-normal foi a que mais se ajustou aos dados estudados para o tempo de permanência hospitalar (figura 11).

Figura 6: Dias de Permanência no Hospital (Região Norte)



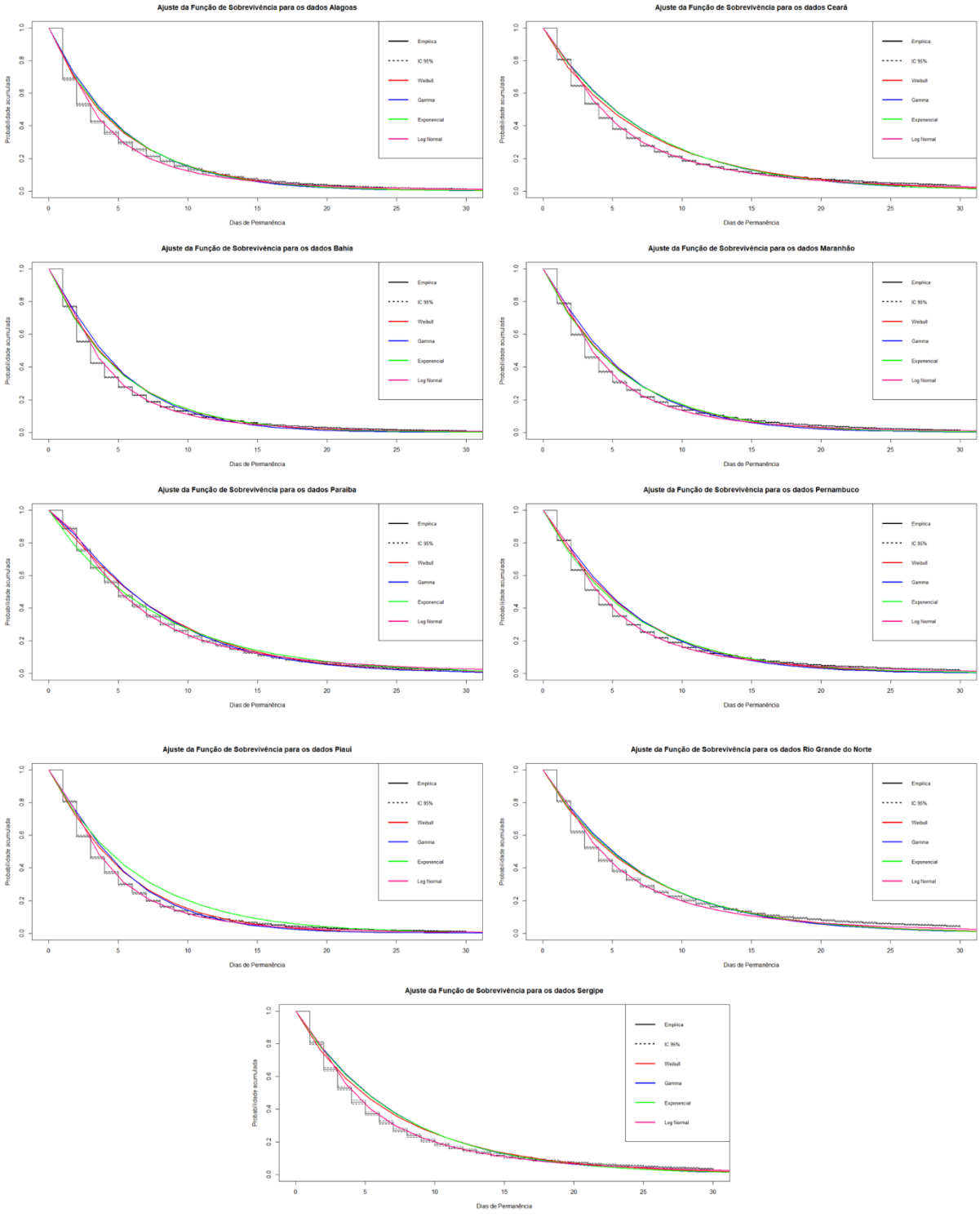
Fonte: Elaborado pela Autora

Já na região nordeste, gráficos para os estados de Alagoas, Bahia, Ceará, Maranhão, Paraíba, Pernambuco, Piauí, Rio Grande do Norte e Sergipe, a maioria das internações também ocorre em períodos curtos, de 1 a 7 dias.

De maneira geral, a Weibull parece ser a distribuição mais adequada para modelar os dados de sobrevivência na maioria dos estados. Porém, analisando o tempo de permanência em um hospital, onde há muitos pacientes

que saem rapidamente (valores baixos) e alguns poucos que permanecem por longos períodos (valores altos), a distribuição Log-normal foi também a mais adequada, pois reflete bem essa assimetria e concentração em valores menores.

Figura 7: Dias de Permanência no Hospital (Região Nordeste)

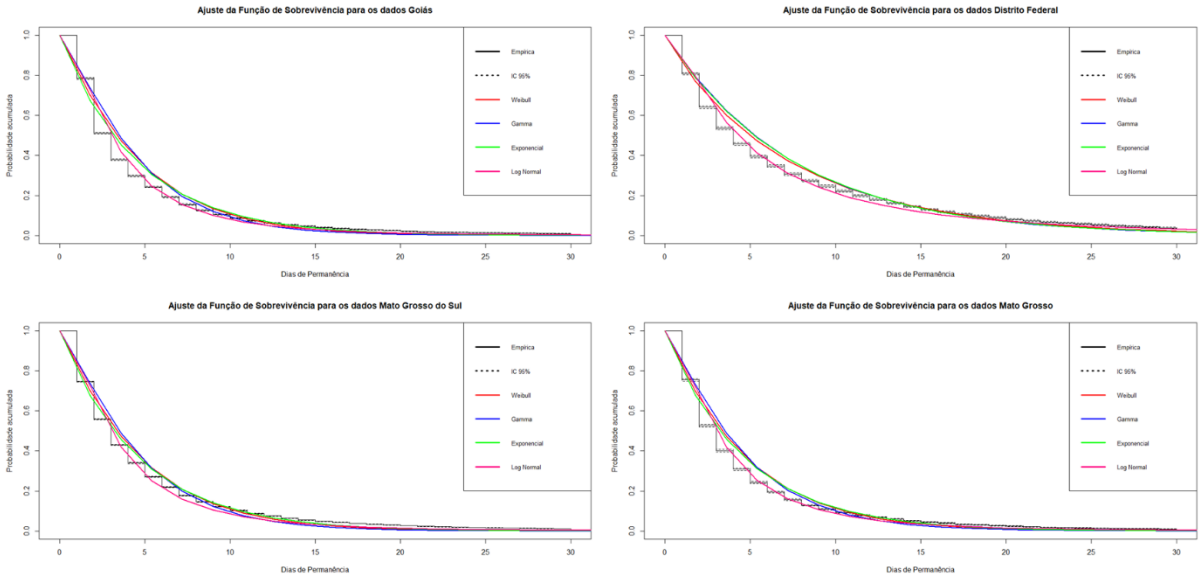


Fonte: Elaborado pela Autora

Na região do centro-oeste, notável a distribuição da frequência do tempo de permanência em internações nos estados: Goiás, Mato Grosso, Mato Grosso do Sul e Distrito Federal.

Logo, a Log-Normal mostra excelente aderência à curva empírica em quase toda a extensão do gráfico, especialmente nos valores intermediários e nas caudas (valores maiores de dias de permanência). A Weibull e a Gamma também se ajustam bem em algumas regiões, mas frequentemente divergem nas extremidades da curva. A Exponencial geralmente subestima a cauda dos dados empíricos, o que pode indicar que não captura bem a variabilidade nos valores maiores.

Figura 8: Dias de Permanência no Hospital (Região Centro-Oeste)

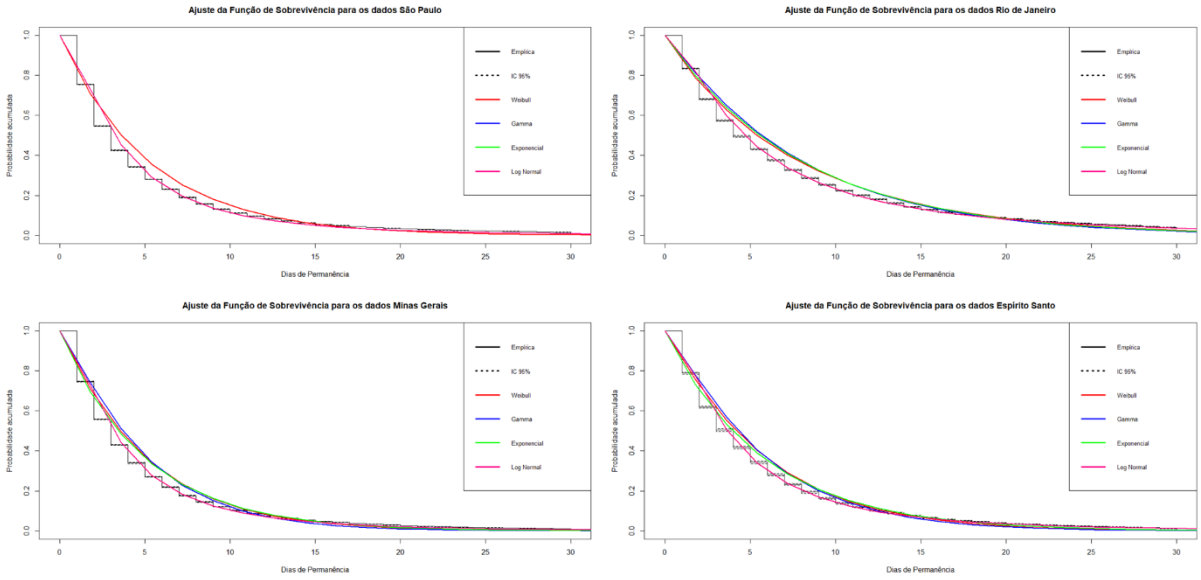


Fonte: Elaborado pela Autora

Já a região sudeste, onde concentra-se grande parte da população, temos que os estados de São Paulo, Minas Gerais, Espírito Santo e Rio de Janeiro, irão apresentar o gráfico de distribuição do tempo de internação.

A distribuição da Log-Normal se destaca novamente como o melhor ajuste para modelar os tempos de permanência nos estados apresentados, por apresentar uma proximidade empírica, como reflete bem a assimetria à direita observada nos dados, capturando adequadamente os valores maiores de permanência.

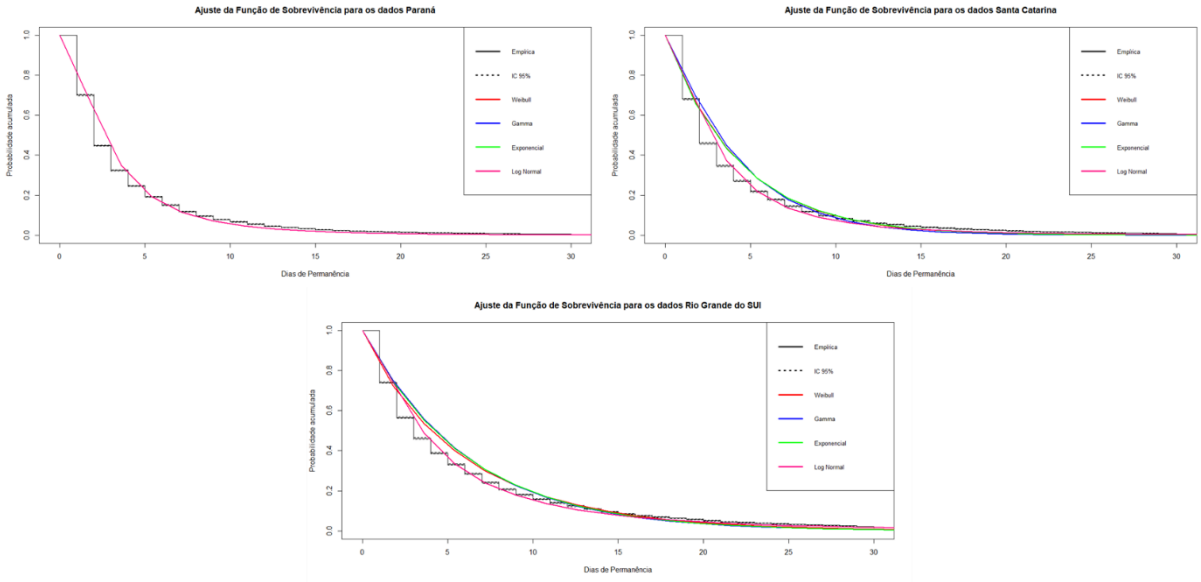
Figura 9: Dias de Permanência no Hospital (Região Sudeste)



Fonte: Elaborado pela Autora

Por fim, temos a região sul composta por Santa Catarina, Paraná e Rio Grande do Sul. Nos gráficos, a curva da distribuição Log-normal parece estar muito próxima da curva empírica de Kaplan-Meier em quase todas as faixas (início, meio e fim dos tempos de permanência), logo considera-se como melhor ajuste.

Figura 10: Dias de Permanência no Hospital (Região Sul)



Fonte: Elaborado pela Autora

3.2.5 Modelos Gamma, Weibull, Exponencial e Log-Normal

Foram apresentados os dados das distribuições anteriormente referente ao dia de permanência, sexo e idade como forma de estudo. Para isso, é necessário ter conhecimentos dos parâmetros na qual está sendo utilizado para cada distribuição.

Logo, teremos a tabela 4 que apresenta o estimador de máxima verossimilhança dos parâmetros das distribuições Gamma, Weibull, Exponencial e Log-normal, para cada estado em estudo.

Tabela 5: Estimadores de Máxima Verossimilhança para os parâmetros de interesse de acordo com as distribuições estudadas

Estado	Distribuição	Parâmetro	Estimador	Erro Padrão
AC	Gamma	θ	1,039989018957820	0.02017
		k	0,171204940146721	0.00425
	Weibull	β	0,921143855355364	0.00976
		λ	5,788376786083960	0.10375
	Exponencial	α	0,164565251072163	0.00256
	Log-normal	μ	1,251136896080280	0.0146
		σ	0,941986408204502	0.0104
AL	Gamma	θ	1,06644775057917	0.01162
		k	0,199904090696132	0.00278
	Weibull	β	0,963706313997722	0.00605
		λ	5,23990529700812	0.05014
	Exponencial	α	0,18733615683618	0.00163
	Log-normal	μ	1,13766867750377	0.00870
		σ	1,00405158478122	0.00621
AM	Gamma	θ	1,21886435470155	0.01298
		k	0,167775253344944	0.00221
	Weibull	β	1,04081031487354	0.00635
		λ	7,41249501970565	0.06360
	Exponencial	α	0,137351330619793	0.00116

	Log-normal	μ	1,51978082780806	0.00784
		σ	0,934287907793403	0.00558
AP	Gamma	θ	0,829385932166685	0.02079
		k	0,0753510901399277	0.00256
	Weibull	β	0,840935502559115	0.0126
		λ	9,89854257083083	0.2560
	Exponencial	α	0,0910178254495183	0.00187
	Log-normal	μ	1,68687199568901	0.0242
		σ	1,18282852546514	0.0172
BA	Gamma	θ	1,22748991296905	0.00568
		k	0,240596960006217	0.00138
	Weibull	β	1,02808190038133	0.00264
		λ	5,17920535244188	0.01950
	Exponencial	α	0,195603643683917	0.000716
	Log-normal	μ	1,17032216235293	0.00333
		σ	0,9156192307809	0.00237
CE	Gamma	θ	1,01814030381068	0.00601
		k	0,138842007461236	0.00106
	Weibull	β	0,936731882501388	0.00318
		λ	7,08812435070932	0.03802
	Exponencial	α	0,136334283788951	0.000649
	Log-normal	μ	1,4269843327387	0.00489
		σ	1,03607741259914	0.00349
DF	Gamma	θ	1,01883968788866	0.00930
		k	0,135251179715469	0.00158
	Weibull	β	0,949507941704191	0.00508
		λ	7,33623984038567	0.05999
	Exponencial	α	0,132747724711042	0.000974
	Log-normal	μ	1,45403309009621	0.00771
		σ	1,05572013826547	0.00548
ES	Gamma	θ	1,25606909440192	0.01029
		k	0,219473355287551	0.00221

	Weibull	β	1,06771592373771	0.00502
		λ	5,89772166813716	0.03774
	Exponencial	α	0,174346062048122	0.00113
	Log-normal	μ	1,29713929744982	0.00606
		σ	0,942237372261673	0.00432
GO	Gamma	θ	1,35543241768422	0.00809
		k	0,298902742388743	0.00216
	Weibull	β	1,07536304659285	0.0035
		λ	4,69640817445824	0.0216
	Exponencial	α	0,219922590051254	0.00103
	Log-normal	μ	1,10012276095695	0.00398
		σ	0,855854268578224	0.00284
MA	Gamma	θ	1,19348517510211	0.00774
		k	0,212037968095794	0.00171
	Weibull	β	1,02223607907064	0.00374
		λ	5,69533590554332	0.03033
	Exponencial	α	0,177411254485033	0.000913
	Log-normal	μ	1,25395180802083	0.00483
		σ	0,942769684646222	0.00343
MG	Gamma	θ	1,24845550474086	0.00446
		k	0,252889844654208	0.00111
	Weibull	β	1,04168309221262	0.00207
		λ	5,0410970382772	0.01444
	Exponencial	α	0,202105371565876	0.00057
	Log-normal	μ	1,14604701216349	0.00256
		σ	0,914039964575095	0.00183
MS	Gamma	θ	1,33681395133132	0.01156
		k	0,292142023531714	0.00307
	Weibull	β	1,06999722124818	0.00505
		λ	4,72744173022143	0.03177
	Exponencial	α	0,218086675758635	0.00148
	Log-normal	μ	1,10253879806512	0.00588

		σ	0,86902473368993	0.00418
MT	Gamma	θ	1,2954740886382	0.01011
		k	0,280031175428507	0.00267
	Weibull	β	1,05364205125117	0.00451
		λ	4,74781233214694	0.02933
	Exponencial	α	0,215779867395104	0.00133
	Log-normal	μ	1,09863747180759	0.00539
		σ	0,881502788223902	0.00383
PA	Gamma	θ	1,28652589789038	0.00734
		k	0,269008645496797	0.00188
	Weibull	β	1,04953826795694	0.00329
		λ	4,89981713331367	0.02223
	Exponencial	α	0,20874289737055	0.000938
	Log-normal	μ	1,12842275457023	0.00394
		σ	0,881173543817724	0.00280
PB	Gamma	θ	1,38007792070192	0.01462
		k	0,180636005800491	0.00232
	Weibull	β	1,14426188066556	0.00699
		λ	8,06248265935993	0.06174
	Exponencial	α	0,130412829447639	0.00108
	Log-normal	μ	1,63035760269134	0.00764
		σ	0,925466045666591	0.00545
PE	Gamma	θ	1,21076714754187	0.00662
		k	0,194854798782075	0.00132
	Weibull	β	1,04597595776419	0.0033
		λ	6,35050549230264	0.0278
	Exponencial	α	0,16060383294575	0.000697
	Log-normal	μ	1,36027362906725	0.00413
		σ	0,957549620278653	0.00294
PI	Gamma	θ	1,3014928922416	0.01061
		k	0,244117118400315	0.00243
	Weibull	β	1,06743889584605	0.00483

		λ	5,50019301578854	0.03500
	Exponencial	α	0,16060383294575	0.0012
	Log-normal	μ	1,24296814430366	0.00574
		σ	0,898983819554559	0.00409
PR	Gamma	θ	NA	NA
		k	NA	NA
	Weibull	β	NA	NA
		λ	NA	NA
	Exponencial	α	NA	NA
	Log-normal	μ	0,952606676947829	0.00283
		σ	0,844397321303325	0.00202
RJ	Gamma	θ	1,05696714883749	0.004917
		k	0,131722131114808	0.000786
	Weibull	β	0,967213542285877	0.00262
		λ	7,90573293074041	0.03219
	Exponencial	α	0,124491572226153	0.000466
	Log-normal	μ	1,54104969169878	0.00387
		σ	1,04379782072746	0.00277
RN	Gamma	θ	1,05688254569847	0.00976
		k	0,148825519563901	0.00175
	Weibull	β	0,972848832571311	0.0053
		λ	7,01026970954081	0.0565
	Exponencial	α	0,140756431096179	0.00104
	Log-normal	μ	1,41795200862588	0.00764
		σ	1,03681937057491	0.00544
RO	Gamma	θ	1,06627656152016	0.01095
		k	0,159541210370255	0.00208
	Weibull	β	0,962872304483398	0.00562
		λ	6,5569181453773	0.05922
	Exponencial	α	0,149562155802717	0.00123
	Log-normal	μ	1,36272172816562	0.00831
		σ	1,01564121987359	0.00591

RR	Gamma	θ	0,941864562772246	0.01940
		k	0,091638157233918	0.00248
	Weibull	β	0,902093513807593	0.0108
		λ	9,70240839413588	0.1899
	Exponencial	α	0,0973482584360717	0.00163
	Log-normal	μ	1,71288201654063	0.0182
		σ	1,09524782862072	0.0129
RS	Gamma	θ	1,03696566461449	0.00544
		k	0,170652664813261	0.00115
	Weibull	β	0,950232448046211	0.00289
		λ	5,92401703469065	0.02774
	Exponencial	α	0,164496799343572	0.000693
	Log-normal	μ	1,25134683631887	0.00429
		σ	1,02604880022527	0.00307
SC	Gamma	θ	1,20549372282936	0.00704
		k	0,282078474162735	0.00205
	Weibull	β	1,00992523817442	0.00326
		λ	4,30275936697403	0.02087
	Exponencial	α	0,233534835726472	0.00108
	Log-normal	μ	0,983904646616706	0.00415
		σ	0,903251999233124	0.00296
SE	Gamma	θ	1,01935334938368	0.01466
		k	0,140014637842411	0.00261
	Weibull	β	0,938129772071942	0.0078
		λ	7,05179986066543	0.0918
	Exponencial	α	0,137302394450836	0.00159
	Log-normal	μ	1,42158416509606	0.01192
		σ	1,03955985870049	0.00856
SP	Gamma	θ	NA	NA
		k	NA	NA
	Weibull	β	0,992914011863567	0.00148
		λ	5,24188759538228	0.01187

	Exponencial	α	NA	NA
	Log-normal	μ	1,16824538824222	0.00199
		σ	0,941940467678726	0.00142
TO	Gamma	θ	0,995268561579712	0.01220
		k	0,104841715792378	0.00166
	Weibull	β	0,94886411420918	0.007
		λ	9,25197831636364	0.102
	Exponencial	α	0,10534608861813	0.00104
	Log-normal	μ	1,67103851682504	0.01080
		σ	1,09833718484254	0.00769

3.2.6 Causas

As principais causas de internações por causas externas incluem diferentes tipos de acidentes e eventos adversos. Entre elas estão os acidentes de transporte, que envolvem pedestres, ciclistas, motociclistas, acidentes aéreos, entre outros. Também se destacam as lesões autoprovocadas, como autointoxicação por drogas, produtos de limpeza, sedativos, analgésicos, além de afogamentos e exposição a fumaça ou fogo. As agressões representam outra causa significativa, podendo incluir violência sexual, negligência e abandono, agressões por meio de força física ou pela utilização de objetos em movimento. Complicações médicas também aparecem como um importante causa de internação, envolvendo acidentes durante a prestação de cuidados médicos e cirúrgicos, reações anormais a medicamentos ou complicações cardíacas. Por fim, há as internações relacionadas à morbidade e mortalidade, que incluem sequelas de lesões autoprovocadas, agressões ou quedas, muitas vezes resultando em complicações graves.

4 CONSIDERAÇÕES FINAIS

O presente estudo analisou o tempo de internação hospitalar por causas externas no Brasil, com base nos dados fornecidos pelo DATASUS para o ano de 2019. A análise permitiu identificar padrões importantes sobre as internações, considerando variáveis como sexo, idade, dias de permanência e principais causas de internação. Os resultados ressaltam a relevância de estudar internações hospitalares como um indicador importante para o planejamento de políticas públicas e intervenções direcionadas à saúde da população.

Os modelos estatísticos empregados evidenciaram que a distribuição Log-normal foi a mais adequada para modelar os dados de permanência hospitalar na maioria dos estados brasileiros, apresentando flexibilidade para captar as características específicas dos dados. Ainda que a distribuição Weibull tenha se mostrado uma alternativa viável em alguns casos, a Log-normal demonstrou maior precisão nos ajustes. Por outro lado, os modelos Exponencial e Gamma apresentaram limitações significativas na adequação aos dados analisados.

Entre as principais causas de internação, destacam-se acidentes de transporte, lesões autoprovocadas, agressões e complicações médicas. A distribuição por faixa etária revelou que jovens adultos são os mais impactados, mas também foi possível observar uma frequência relevante de internações em crianças e idosos, especialmente relacionadas a quedas e acidentes domésticos. Em termos de gênero, o sexo masculino predominou em quase todas as regiões, possivelmente devido a comportamentos de risco e maior exposição a situações perigosas.

Essas informações destacam a importância de iniciativas preventivas para reduzir as internações por causas externas. Campanhas de conscientização sobre segurança no trânsito, programas de prevenção de violência e cuidados domiciliares para idosos podem contribuir significativamente para a redução desses índices. Além disso, o estudo evidencia a necessidade de investimentos em infraestrutura hospitalar e em serviços de emergência para regiões com maior vulnerabilidade, garantindo atendimento ágil e eficaz para reduzir os impactos dessas internações.

Em suma, este trabalho reforça a importância da aplicação de análises estatísticas em dados de saúde pública, como forma de identificar padrões e subsidiar ações estratégicas para a melhoria do sistema de saúde e da qualidade de vida da população. Estudos futuros podem ampliar a abordagem, considerando variáveis adicionais como por exemplo o custo para hospitais particulares e públicos, além de explorar outros modelos estatísticos para maior precisão na análise de internações hospitalares.

REFERÊNCIAS

COLOSIMO, E; GIOLO, S.R. Análise de Sobrevivência Aplicada. Edgard Blucher, São Paulo. 2006. Acessado em 20 de setembro de 2024.

CONCLINICA. *CID-10: Códigos de Doenças e Problemas de Saúde*. Disponível em: https://conclinica.com.br/cid10/v01-y98/. Acesso em: 26 de outubro de 2024.

ELANDT-JOHNSON, R.C.; JOHNSON, N.L. Survival Models and Data Analysis. Wiley, 2014. Acessado em 15 de agosto de 2024.

HOFFMANN, Rodolfo. Análise de Regressão: uma introdução ao a econometria. USP – Piracicaba, 2016. Disponível em: https://www.esalq.usp.br/biblioteca/sites/default/files/Analise_Regress%C3%A3o.pdf. Acessado em 28 de maio de 2024.

HOSMER, David W; LEMESHOW, Stanley; MAY, Susanne. Applied survival analysis: regression modeling of time to event data, Second edition. 2011. Disponível em: http://www.uop.edu.pk/ocontents/survival-analysis-self-learning-book.pdf. Acessado em 26 de setembro de 2024.

KLEINBAUM, David G.; KLEIN, Mitchel. Survival analysis: a self-learning text. Third Edition, 2012. Disponível em: http://www.uop.edu.pk/ocontents/survival-analysis-self-learningbook.pdf. Acessado em 20 de outubro de 2024.

LAWLESS, J.F. Statistical models and methods for lifetime data, John Wiley & Sons. 1982.

LEE, E.T.; WANG, J. Statistical Methods for Survival Data Analysis. 2003.

MASCARENHAS, Márcio D.M; BARROS, Maralisa B.A. REVISTA BRASILEIRA DE EPIDEMIOLOGIA. Caracterização das internações hospitalares por causas externas no sistema público de saúde, Brasil, 2015. Disponível em: https://www.scielo.br/j/rbepid/a/jfjWmLcsdhG66syYCBHdMDr/?lang=pt. Acesso em: 05 de outubro de 2023.

MINISTÉRIO DA SAÚDE. Plano Nacional de Saúde: 2020-2023. Disponível em: https://bvsms.saude.gov.br/bvs/publicacoes/plano_nacional_saude_2020_2023.pdf. Acesso em: 05 de outubro de 2023.

ORGANIZAÇÃO PAN-AMERICANA DA SAÚDE (OPAS). Pandemia de COVID-19 desencadeia aumento de 25% na prevalência de ansiedade e depressão. Disponível em: https://www.paho.org/pt/noticias/2-3-2022-pandemia-covid-19-desencadeia-aumento-25-na-prevalencia-ansiedade-e-depressao-em. Acesso em: 26 nov. 2024.

PÁGINA SEGUINTE. A aventura nos esportes radicais da região Centro-Oeste. Disponível em: https://paginaseguinte.com/2023/07/24/a-aventura-nos-esportes-radicais-da-regiao-centro-oeste/. Acesso em: 26 nov. 2024.

QUATRO RODAS ABRIL. Motoristas homens provocam mais acidentes do que mulheres no Brasil. Disponível em: https://quatrorodas.abril.com.br/noticias/motoristas-homens-provocam-mais-acidentes-do-que-mulheres-no-brasil. Acesso em: 26 nov. 2024.

SECRETARIA DE ESTADO DA SAÚDE DO PIAUÍ. Boletim de causas externas: acidentes e violências no Piauí, 2023. Disponível em: https://www.saude.pi.gov.br/uploads/document/file/1722/BOLETIM_CAUSAS_EXTERNAS_ATT_PIAU%C3%8D__2023.pdf. Acesso em: 26 nov. 2024.

SENADO FEDERAL. Acidentes domésticos: o perigo mora em casa. Disponível em: https://www12.senado.leg.br/radio/1/reportagem-especial/2024/06/24/acidentes-domesticos-o-perigo-mora-em-casa. Acesso em: 26 nov. 2024.

TRANSALVADOR. Acidentes de trânsito matam mais homens que mulheres em Salvador. Disponível em: https://transalvador.salvador.ba.gov.br/acidentes-de-transito-matam-mais-homens-que-mulheres-em-salvador/. Acesso em: 26 nov. 2024.