



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Câmpus de Presidente Prudente

FERNANDO ZHEN YE CHEN

**MODELAGEM DO TEMPO DE READMISSÃO NO HOSPITAL DE PACIENTES
TABAGISTAS: UMA ABORDAGEM UTILIZANDO O MODELO DE FRAÇÃO DE
CURA DE NÃO-MISTURA GAMA**

PRESIDENTE PRUDENTE

2025

FERNANDO ZHEN YE CHEN

**MODELAGEM DO TEMPO DE READMISSÃO NO HOSPITAL DE PACIENTES
TABAGISTAS: UMA ABORDAGEM UTILIZANDO O MODELO DE FRAÇÃO DE
CURA DE NÃO-MISTURA GAMA**

Relatório Final para 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. Ricardo Puziol de Oliveira

PRESIDENTE PRUDENTE

2025

C518m	Chen, Fernando Zhen Ye Modelagem do tempo de readmissão no hospital de pacientes tabagistas : uma abordagem utilizando o modelo de fração de cura de não-mistura gama / Fernando Zhen Ye Chen. -- Presidente Prudente, 2025 43 p. : il., tabs. Trabalho de conclusão de curso (Bacharelado - Estatística) - Universidade Estadual Paulista (UNESP), Faculdade de Ciências e Tecnologia, Presidente Prudente Orientador: Ricardo Puziol de Oliveira 1. Análise de sobrevivência. 2. Kaplan-Meier. 3. Modelo de fração de cura. 4. Modelo gama. I. Título.
-------	--

TERMO DE APROVAÇÃO

Fernando Zhen Ye Chen

MODELAGEM DO TEMPO DE READMISSÃO NO HOSPITAL DE PACIENTES
TABAGISTAS: UMA ABORDAGEM UTILIZANDO O MODELO DE FRAÇÃO DE CURA
DE NÃO-MISTURA GAMA

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:



Documento assinado digitalmente

RICARDO PUZIOL DE OLIVEIRA

Data: 01/12/2025 22:47:41-0300

Verifique em <https://validar.it.gov.br>

Orientador: _____

Prof. Dr. Ricardo Puziol de Oliveira
Departamento de Estatística



Documento assinado digitalmente

FERNANDO ANTONIO MOALA

Data: 02/12/2025 22:34:34-0300

Verifique em <https://validar.it.gov.br>

Prof. Dr. Fernando Antônio Moala
Departamento de Estatística

Presidente Prudente, 27 de novembro de 2025.

AGRADECIMENTOS

Gostaria de expressar minha sincera gratidão a todos que me apoiaram durante esta jornada acadêmica.

Em primeiro lugar, aos meus pais, Chen e Ming, por serem minha base fundamental e meu maior suporte em toda a minha caminhada. Às minhas irmãs, pelo carinho que sempre me manteve conectado ao lar, mesmo de longe.

Um agradecimento muito especial à minha namorada, Tomomi, pelo constante incentivo para seguir em frente e pelo indispensável suporte emocional, especialmente durante a reta final da graduação.

Ao meu orientador, Prof. Dr. Ricardo, pela paciência, dedicação e valiosas orientações que foram essenciais para a realização deste trabalho.

Minha gratidão também à Unesp e à sua biblioteca, por proporcionarem um ambiente acadêmico tão enriquecedor e plural, e por disponibilizarem um acervo que foi fundamental para a minha formação.

Por fim, agradeço aos amigos que conheci durante a graduação, pelo companheirismo e pelo apoio incondicional em cada etapa desta jornada.

RESUMO

Este trabalho tem como objetivo analisar os fatores associados à readmissão hospitalar, com ênfase na influência do tabagismo, por meio da aplicação de um modelo de fração de cura do tipo não-mistura baseado na distribuição Gama. Os dados utilizados neste trabalho são dados obtidos do estudo de John et al. (2015), referentes a 143 fumantes internados em um hospital comunitário com 180 leitos localizado em La Chaux-de-Fonds, Suíça, que inclui fumantes que no momento de admissão aos departamentos de medicina interna, cirurgia ou terapia intensiva, apresentaram pelo menos uma mancha de tabaco nos dedos. As internações ocorreram entre março de 2006 e dezembro de 2007, representando cerca de 6% dos fumantes admitidos no período. O modelo apresentou boa aderência aos dados e pode auxiliar na identificação de pacientes com maior risco, contribuindo para estratégias de intervenção em saúde pública voltadas ao controle do tabagismo.

Palavras-chaves: Análise de sobrevivência; Modelo Gama; Kaplan-Meier; Modelo de Fração de Cura.

ABSTRACT

This study aims to analyze the factors associated with hospital readmission, with an emphasis on the influence of smoking, by applying a non-mixed type cure fraction model based on the Gamma distribution. The data used in this work is from the study by John et al. (2015), which refers to 143 smokers admitted to a 180-bed community hospital located in La Chaux-de-Fonds, Switzerland, which includes smokers who, at the time of admission to the internal medicine, surgery or intensive care departments, had at least one tobacco stain on their fingers. The admissions took place between March 2006 and December 2007, representing around 6% of smokers admitted during the period. The model showed good adherence to the data and can help identify patients at greater risk, contributing to public health intervention strategies aimed at tobacco control.

Keywords: Survival analysis; Gamma model; Kaplan-Meier; Cure fraction model.

LISTA DE FIGURAS

Figura 1. Curva de sobrevivência estimada via Kaplan-Meier referente ao tempo até a readmissão no hospital.....	28
Figura 2. Matriz de correlação de Spearman entre as covariáveis do estudo.	29
Figura 3. Curva de sobrevivência estimada via Kaplan-Meier versus curva de sobrevivência ajustada pelo modelo de fração de cura de não-mistura gama referente ao tempo até a readmissão no hospital.....	30
Figura 4. Resíduos de Cox-Snell para o modelo de fração de cura de não-mistura gama proposto.	34

LISTA DE TABELAS

Tabela 1. Características dos pacientes tabagistas readmitidos e não-readmitidos ao longo do período de acompanhamento do hospital La Chaux-de-Fonds, no estudo de John et al. (2015).	22
Tabela 2. Resultado do estudo de simulação Monte Carlo para $\alpha=0.5$, $\beta=0.5$ e $p=0.2$	26
Tabela 3. Resultado do estudo de simulação Monte Carlo para $\alpha=1.0$, $\beta=1.0$ e $p=0.4$	27
Tabela 4. Resultado do estudo de simulação Monte Carlo para $\alpha=1.5$, $\beta=1.5$ e $p=0.6$	27
Tabela 5. Estimativas de parâmetros e intervalos de confiança para 95% do modelo de regressão de fração de cura de não-mistura gama.....	32

SUMÁRIO

1 INTRODUÇÃO	8
2 REFERENCIAL TEÓRICO	12
2.1. TRABALHOS RELACIONADOS	12
2.2. ESTUDOS DE SOBREVIVÊNCIA	13
2.3. MECANISMOS DE CENSURA.....	15
2.4. ESTIMADOR KAPLAN-MEIER.....	17
2.5. MODELO GAMA.....	18
2.6. MODELOS DE FRAÇÃO DE CURA DE NÃO-MISTURA	20
2.7. RESÍDUOS DE COX-SNELL.....	21
3 MATERIAIS E MÉTODOS	22
3.1. BASE DE DADOS.....	22
3.2. MODELO DE FRAÇÃO DE CURA DE NÃO-MISTURA GAMA	23
4 ESTUDO DE SIMULAÇÃO MONTE CARLO.....	25
5 RESULTADOS.....	28
5.1. ANÁLISE EXPLORATÓRIA DE DADOS	28
5.2. ANÁLISE PELO MODELO DE FRAÇÃO DE CURA GAMA	30
6. CONCLUSÃO	35
REFERÊNCIAS.....	36

1 INTRODUÇÃO

O tabaco é uma planta cultivada há séculos, atualmente comercializada no mundo todo, majoritariamente usado no fumo em cigarros, charutos ou cachimbos, em pesticidas sob a forma de tartarato de nicotina e em alguns remédios. A planta é rica em nicotina, um composto psicoativo responsável pela dependência química e associado a doenças crônicas, incluindo câncer, enfermidades cardiovasculares e respiratórias (PROCHASKA, 2023).

O impacto do consumo de tabaco sobre a saúde pública é expressivo e persistente. Atualmente, a chamada epidemia do tabaco constitui uma das maiores ameaças à saúde global, sendo responsável por mais de 8 milhões de óbitos por ano. Estima-se que cerca de 80% dos 1,3 bilhão de usuários de tabaco no mundo vivem em países de baixa e média renda, onde doenças e mortes relacionadas ao tabaco são mais críticas e severas. Todas as formas de uso de tabaco são prejudiciais, não existindo nenhum nível seguro de exposição e consumo ao tabaco (ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2023).

Apesar dos esforços de controle, a redução do tabagismo ao longo das últimas décadas não tem sido homogênea. A prevalência global do hábito diminuiu de 22,7% em 2007 para 17% em 2021; no entanto, essa tendência não se refletiu em todos os países. Em muitas nações, especialmente aquelas com crescimento econômico acelerado e forte atuação da indústria do tabaco por meio de estratégias de marketing agressivas, observam-se taxas de consumo em ascensão. Além disso, os países que produzem e exportam tabaco são, em sua maioria, nações em desenvolvimento que, ao suprirem a demanda global, acabam por internalizar os custos sanitários e sociais do consumo, piorando os indicadores de desigualdade (THE TOBACCO ATLAS, 2023).

As consequências clínicas do tabagismo vão além das doenças crônicas amplamente conhecidas. Estudos recentes têm demonstrado que o tabagismo atua como preditor independente de readmissão hospitalar precoce em diversos contextos, por exemplo, em pacientes com doenças psiquiátricas, principalmente esquizofrenia e bipolaridade (KAGABO, 2019) e complicações após a artroplastia de ombro (WHITE, 2022). Um sinal clínico de relevância crescente é a presença de manchas de tabaco nos dedos, observadas com maior frequência em fumantes de cigarros sem filtro ou em indivíduos que iniciaram o hábito precocemente. Esse marcador físico tem sido

associado a maior incidência de doenças tabaco-relacionadas e, potencialmente, a piores desfechos clínicos. Dessa forma, a observação de manchas de tabaco nos dedos pode constituir um indicativo clínico de risco elevado de readmissão hospitalar e mortalidade (JOHN ET. AL, 2015).

Além das implicações clínicas, observa-se uma preocupante transformação no perfil do consumo de produtos derivados do tabaco, notadamente entre adolescentes e adultos jovens. O aumento expressivo do uso de cigarros eletrônicos, dispositivos eletrônicos que vaporizam soluções aromatizadas, geralmente contendo nicotina tem contribuído para a chamada renormalização do ato de fumar (JONES, 2020). De princípio, estes foram anunciados pelas empresas de tabaco como um método para ajudar indivíduos a parar de fumar, entretanto essa afirmação não apresenta sustentação científica, uma vez que estudos compararam o uso de cigarros eletrônicos, placebos e adesivos de nicotina não demonstraram diferenças significativas nas taxas de abstinência (HUEY, 2020).

Em resposta à essa complexidade crescente da crise global do tabaco, a Organização Mundial da Saúde implementou, em 2005, a Convenção-Quadro para o Controle do Tabaco (FCTC), adotada por 182 países membros da Organização das Nações Unidas. Esta convenção internacional reúne diretrizes e compromissos voltados à redução do consumo e da oferta de produtos de tabaco, por meio de uma abordagem integrada que combina medidas legislativas, sanitárias, educacionais e econômicas (THE TOBACCO ATLAS, 2023). A implementação efetiva dessas medidas é essencial para conter o avanço da epidemia tabágica e mitigar seus impactos sobre a saúde pública global.

Considerando esse panorama, o presente trabalho tem como objetivo investigar os fatores associados à readmissão hospitalar, com especial atenção à influência do tabagismo sobre a probabilidade de um paciente ser readmitido após a alta. Para atingir tal finalidade, foi considerada uma base dados obtidos de forma secundária do estudo de John et al. (2015), referentes à 143 fumantes admitidos em um hospital comunitário de 180 leitos em La Chaux-de-Fonds, Suíça, que incluiu fumantes que apresentaram pelo menos uma mancha de tabaco nos dedos no momento da admissão aos departamentos de medicina interna, cirurgia ou unidade de terapia intensiva, independentemente de seu diagnóstico inicial, e que foram hospitalizados no período entre março de 2006 e dezembro de 2007 (6% dos fumantes).

Com o intuito de captar adequadamente a heterogeneidade na probabilidade de readmissão e mitigar a presença de um possível platô na curva de sobrevivência estimada por Kaplan-Meier, adotou-se uma abordagem estatística baseada em modelo de fração de cura do tipo não-mistura. Tais modelos permitem distinguir, de maneira implícita, a parcela de pacientes que não apresentarão o evento de interesse (readmissão), mesmo com longos períodos de acompanhamento (*follow-up*). Como função de sobrevivência base (baseline), foi adotada a distribuição Gama, devido a sua flexibilidade para modelagem de dados de tempo até o evento e aplicabilidade em contextos clínicos.

Além da contribuição metodológica, este estudo também visa contribuir com o debate sobre políticas públicas relacionadas ao controle do tabagismo, com ênfase na promoção de ambientes mais saudáveis e no fortalecimento dos sistemas de saúde. A proposta está alinhada aos Objetivos de Desenvolvimento Sustentável (ODS) da Organização das Nações Unidas, especialmente ao ODS 3, que visa assegurar uma vida saudável e promover o bem-estar para todos. Em particular, a meta 3.5 destaca a importância da prevenção e do tratamento do uso nocivo de substâncias, enquanto a meta 3.a reforça o compromisso com a implementação da FCTC em todos os países signatários.

1.1 OBJETIVOS

1.1.1 OBJETIVO GERAL

O presente estudo tem como objetivo geral investigar os fatores associados à readmissão hospitalar, com ênfase na influência do tabagismo, por meio de técnicas de análise de sobrevivência. Para tal, será proposto um modelo de fração de cura com o intuito de mitigar a presença de um possível platô na curva de sobrevivência sob abordagem de não-mistura, usando a distribuição gama como *baseline* para a construção do modelo. A proposta visa não apenas estimar o tempo até a readmissão, mas também quantificar a proporção de indivíduos que não apresentariam nova hospitalização ao longo do tempo (isto é, “curados”). Além disso, por meio dessa modelagem, busca-se identificar variáveis que influenciam significativamente a dinâmica da readmissão, contribuindo para a compreensão dos mecanismos subjacentes ao processo e subsidiando estratégias de intervenção.

1.1.2 OBJETIVOS ESPECÍFICOS

- Realizar uma análise exploratória dos dados, com o intuito de identificar fatores clínicos e demográficos associados à readmissão hospitalar, considerando variáveis como idade, sexo, condições clínicas pré-existentes, histórico de tabagismo, tipo de internação, tempo de permanência e procedimentos realizados durante a hospitalização.
- Estimar a função de sobrevivência e a curva de readmissão hospitalar por meio do estimador de Kaplan-Meier, avaliando padrões gerais de tempo até a readmissão.
- Desenvolver um modelo de fração de cura sob abordagem de não-mistura baseado na distribuição Gama, incorporando estrutura de regressão nos parâmetros relevantes, a fim de avaliar a associação entre os fatores identificados e a probabilidade de readmissão hospitalar.
- Avaliar a adequação e a capacidade preditiva do modelo proposto, por meio de análises de resíduos (como os resíduos de Cox-Snell) e de medidas de ajuste, discriminando entre indivíduos suscetíveis e não suscetíveis à readmissão.
- Discutir as implicações clínicas e em saúde pública dos achados, com ênfase na relação entre tabagismo e readmissão hospitalar.

2 REFERENCIAL TEÓRICO

2.1. TRABALHOS RELACIONADOS

Em seu estudo, Cartmell et al. (2018) investigou o impacto de um serviço de tratamento para dependência de tabaco (TDTS) baseado em evidências nas taxas de readmissão hospitalar em 30, 90 e 180 dias. Os pesquisadores analisaram dados de registros eletrônicos de saúde, informações do programa de cessação tabágica e um banco de dados estadual de utilização hospitalar, utilizando modelos de regressão logística ponderados por escores de propensão inversa para ajustar possíveis vieses. Os resultados mostraram que os pacientes expostos ao TDTS apresentaram uma redução significativa de 23% nas chances de readmissão em 30 dias ($OR = 0,77$; $p = 0,031$). Embora as reduções em 90 e 180 dias não tenham alcançado significância estatística, a tendência foi consistente ($OR = 0,87$ e $0,86$, respectivamente). O estudo concluiu que intervenções estruturadas para cessação do tabagismo durante a hospitalização, especialmente quando combinadas com acompanhamento pós-alta, como consultas à beira do leito e chamadas automatizadas, podem contribuir para a redução de readmissões hospitalares a curto prazo.

Já o estudo de Arena et al. (2024) investigou as diferenças entre não fumantes, ex-fumantes e fumantes atuais em relação ao tempo de internação hospitalar, readmissões (7 e 28 dias) e custos de readmissão em pacientes submetidos a cirurgias eletivas. Utilizando dados administrativos de 24.818 pacientes de sete hospitais na Austrália Ocidental (2016–2019), os resultados mostraram que fumantes atuais e ex-fumantes tiveram tempo de internação hospitalar significativamente maior (6,1 e 6,2 dias, respectivamente) em comparação com não fumantes (4,7 dias). Além disso, as chances de readmissão em 7 e 28 dias foram significativamente maiores para fumantes atuais ($OR = 1,37$ e $OR = 1,53$) e ex-fumantes ($OR = 1,29$ e $OR = 1,35$), com custos de readmissão também mais elevados para fumantes atuais ($aRR = 1,51$ em 7 dias; $aRR = 1,38$ em 28 dias).

Enquanto no estudo de Robins et al. (2024) avaliaram a implementação de um serviço hospitalar de tratamento para dependência de tabaco baseado no Modelo de Ottawa em dois hospitais de Londres, analisando a adesão, cessação do tabagismo e desfechos clínicos em 2.067 pacientes fumantes. O serviço incluiu identificação sistemática do status tabágico, aconselhamento personalizado, farmacoterapia durante a internação e suporte por telefone por seis meses pós-alta. Os resultados

revelaram que 79,4% dos pacientes aceitaram a intervenção, com taxas de cessação de 35,1% entre os contactados em seis meses (7,8% do total). Diagnósticos de câncer, diabetes e transtornos mentais aumentaram o risco de mortalidade e readmissão. O estudo destacou variações significativas entre os hospitais, atribuídas a diferenças na implementação do programa, como tempo de internação e estratégias de seguimento. Logo, o estudo concluiu que realizar intervenções personalizadas, levando em conta características demográficas e clínicas são essenciais na melhora do resultado, especialmente em populações com comorbidades.

Conforme no estudo de White et al. (2022) que investigaram o impacto do uso de tabaco nas taxas de reinternação e complicações após cirurgias de artroplastia do ombro. Os pesquisadores analisaram dados do National Readmissions Database (NRD) de 2016 a 2018, utilizando códigos ICD-10 para identificar usuários de tabaco e avaliar desfechos como reinternações em 30 e 90 dias, complicações cirúrgicas e médicas. Os resultados revelaram que os usuários de tabaco eram significativamente mais jovens (média de 62,9 anos vs. 69,9 anos; $p < 0,01$) e apresentaram taxas mais elevadas de reinternação em 30 dias e 90 dias, além de maior risco de complicações como deslocamento e infecções protéticas. Embora a análise multivariada não tenha identificado o tabaco como preditor independente de complicações para a coorte total, pacientes submetidos a artroplastia reversa tiveram risco aumentado de reinternação. O estudo concluiu que, embora a artroplastia do ombro seja viável para pacientes tabagistas, o uso de tabaco está associado a piores desfechos no curto prazo, reforçando a importância de intervenções pré e pós-operatórias, como programas de cessação do tabagismo, para mitigar esses riscos.

2.2. ESTUDOS DE SOBREVIVÊNCIA

Dados de sobrevivência, dados de tempo de vida, dados de tempo até a falha ou dados de tempo até o evento são termos usados para descrever dados que medem o tempo até a ocorrência de algum evento e surgem em diversos campos aplicados, como medicina, biologia, saúde pública, epidemiologia, engenharia, economia, demografia, entre muitos outros. Segundo Tutz et al. (2016), em todos esses campos, o foco está na modelagem do tempo até que ocorre um evento específico, ou seja, dados de tempo até o evento, bem como na previsão da probabilidade de resposta, sobrevivência ou tempo médio de vida, comparações das distribuições de

sobrevivência e a identificação de fatores de risco e/ou previsões relacionadas às respostas fornecidas pelos tempos de vida.

O evento pode ser morte, como o termo sugere, mas também pode ser qualquer circunstância bem definida. Por exemplo, na pesquisa médica, o evento pode ser alguma ocorrência que altere a vida, como a cura de uma doença, o tempo de remissão para um tipo específico de câncer, o tempo até que indivíduos expostos sejam infectados por uma doença, entre outros; na análise de confiabilidade, o evento pode estar relacionado a quebra, reparo de máquinas ou tempo de vida útil de componentes de um sistema de engenharia, entre outros; em outras áreas pode ser o tempo até a liberdade condicional de condenados (criminologia); duração do primeiro casamento (sociologia); tempo de assinatura de jornal ou revista (marketing); e pedidos de indenização trabalhista (seguros); entre muitas outras aplicações (COX, 1972; MALLER AND ZHOU, 1996; KLEIN AND MOESCHBERGER, 1997; COLLETT, 2015).

É válido destacar também que um estudo que gera dados de tempo de sobrevivência é necessariamente longitudinal por natureza (podendo ser um estudo de coorte observacional ou um ensaio clínico randomizado), em contraste com um estudo transversal, nenhum dos dados são obtidos em um ponto fixo da linha do tempo. Suponha, por exemplo, que um paciente seja selecionado aleatoriamente pela população e indique seu tempo até o evento de falha como T . Segundo Li e Ma (2013), é importante ajustar uma distribuição de probabilidade para a variável desejada T a fim de responder questões práticas como “por quanto tempo um paciente com câncer de pulmão pode sobreviver com 90% de probabilidade?” ou “o tratamento A será significativamente mais benéfico para pacientes com doença mental crônica do que o tratamento B?”.

Para responder a essas questões, há duas funções de grande interesse na análise de dados de sobrevivência: a função de sobrevivência e a função de risco. A função de sobrevivência pode ser interpretada e definida como a probabilidade de observar um tempo de sobrevivência maior que um valor fixo t , isto é:

$$S(t) = P(T \geq t) = 1 - F(t) \quad (2.1)$$

onde $F(t)$, denotado como função de distribuição acumulada, é a probabilidade de um indivíduo selecionado aleatoriamente morrer antes do tempo t . Segundo Lawless

(1982), a função de sobrevivência não é crescente e contínua à esquerda, isto é, $S(0) = 1$ e $S(\infty) = 0$.

Outra função importante relacionada à descrição da distribuição de sobrevivência é a função de risco. Ela pode ser definida como a probabilidade de morrer no instante t , dado que o tempo de sobrevivência seja pelo menos t , ou seja:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t \mid \geq t)}{\Delta t} \quad (2.2)$$

onde Δt é um incremento infinitesimal de tempo. Em algumas aplicações práticas, $h(t)$ também é chamada de taxa instantânea de falha e pode ser expressa como $h(t) = f(t)/S(t)$, onde $f(t)$ é a função de densidade de probabilidade (ou de massa, no caso discreto). O risco acumulado, $H(t)$, também pode ser obtido em termos da função de sobrevivência como:

$$S(t) = \exp\{-H(t)\} \quad (2.3)$$

Em resumo, conhecendo uma das quatro funções $f(t)$, $S(t)$, $H(t)$ e $h(t)$, podemos derivar as outras três. Essas funções podem servir ao mesmo propósito de descrever a distribuição do tempo de sobrevivência. Convencionalmente, $f(t)$ e $S(t)$ são usados para formar funções de verossimilhança para estimativa de parâmetros e testes de hipóteses, enquanto $h(t)$, é geralmente usado para apresentar modelos de regressão, como o modelo de riscos proporcionais de Cox (COX, 1972).

2.3. MECANISMOS DE CENSURA

Segundo Klein e Moeschberger (1997), Giolo e Colosimo (2006) e Collett (2015), a principal característica dos dados de sobrevivência é a presença de observações censuradas ou truncadas. O tempo de sobrevivência de um paciente é censurado quando o ponto de interesse não foi observado para esse paciente. Em geral, isso ocorre porque, por algum motivo, o acompanhamento do paciente foi interrompido ou porque o paciente faleceu por uma causa diferente da pesquisa científica. Na literatura, três tipos comuns de esquemas de censura são considerados:

- **Censura Tipo I:** Muitos estudos têm fundos limitados, e os pesquisadores não podem esperar até que todos os indivíduos desenvolvam o evento de interesse. Assim, é aceitável observar por um período fixo, por exemplo, nove meses ou cinco anos. Os tempos de sobrevivência dos indivíduos que desenvolveram o desfecho de interesse durante o período fixo são observações não censuradas. Os tempos de sobrevivência dos indivíduos que ainda não experimentaram o evento de interesse no final do estudo não são conhecidos exatamente, mas são registrados como tendo a duração total do estudo. Na censura tipo I, o tempo de censura é sempre igual ao comprimento total do estudo, o que simplifica cálculos de acompanhamento como a estimação de parâmetros (GIOLO E COLOSIMO, 2006).
- **Censura Tipo II:** Aparece geralmente em estudos laboratoriais com animais. O evento de interesse é frequentemente o desfecho fatal. O estudo prossegue até que uma proporção ou número fixo de indivíduos morra, por exemplo, 45% da amostra ou 120 indivíduos. Nesse caso, o tempo de censura é sempre igual ao maior tempo de sobrevivência não censurado.
- **Censura Aleatória:** Em muitos estudos clínicos e epidemiológicos, o período de acompanhamento é fixo em calendário, e pacientes são incluídos no estudo em momentos diferentes. Alguns podem desenvolver o evento antes do final do estudo, outros podem se retirar durante o período e se perder no acompanhamento, e outros podem nunca desenvolver o evento até o final do estudo. Assim, os tempos de censura não são idênticos e se comportam como variáveis aleatórias.

O tipo mais comum de censura apresentado em muitos estudos é uma censura aleatória, que ocorre quando um paciente é retirado no decorrer do estudo sem ter ocorrido a falha, ou também, por exemplo, se o paciente morrer por uma razão diferente da estudada. Neste tipo de censura, o mecanismo pode ser dividido em à direita, à esquerda e intervalar. O mais comum é a censura à direita, que corresponde ao fato do tempo de ocorrência do evento de interesse está à direita do tempo registrado. Esse esquema de censura, em particular, pode ser definido usando duas verificações examinadas, sendo T variável a associada ao tempo até a falha de um

paciente e C uma variável pesquisada independente de T , relacionada ao tempo de censura. Nesse caso, o tempo observado e o esquema de censura são dados por:

$$t = \min(T, C) \quad \delta = \begin{cases} 1, & T \leq C \\ 0, & T > C \end{cases} \quad (2.4)$$

Por outro lado, o esquema de censura à esquerda também pode ser definido usando duas variáveis avaliadas, como no esquema de censura à direita. Nesse caso, t é igual a T se o tempo de vida para coletado e δ indica se o tempo exato de vida T foi coletado ($\delta = 1$) ou não ($\delta = 0$). Portanto, para a censura à esquerda, os dados coletados são dados por:

$$t = \max(T, C) \quad \delta = \begin{cases} 1, & T \geq C \\ 0, & T < C \end{cases} \quad (2.5)$$

É importante ressaltar que, ao não considerar a presença de observações censuradas, as técnicas clássicas, como a análise de regressão padrão com erros normais e dados geralmente transformados, poderiam ser usadas na análise de dados de sobrevivência. Entretanto, essas técnicas, geralmente baseadas na suposição de normalidade ou métodos não paramétricos, podem ser inconvenientes na presença de investigações censuradas ou taxas de cura. Para lidar com essa limitação, novas técnicas têm sido apresentadas na literatura. As mais populares são descritas pelos modelos de longo prazo (também chamados de modelos de taxa de cura) (MALLER E ZHOU, 1996) e modelos de fragilidade (WIENKE, 2011; PRICE E MANATUNGA, 2001)

2.4. ESTIMADOR KAPLAN-MEIER

Em estudos de sobrevivência, observa-se com frequência que a curva de sobrevivência relativa tende a atingir um platô após determinado período de acompanhamento. De acordo com Lambert et al. (2007), esse efeito de platô ocorre quando a taxa de mortalidade dos indivíduos doentes se iguala à taxa de mortalidade esperada na população geral, ou seja, quando a mortalidade em excesso se torna nula. Tal comportamento sugere a presença de uma fração de cura na população-alvo, um subgrupo de pacientes considerados efetivamente livres do risco associado

à doença. A identificação dessa possível fração de pacientes “curados” demanda o uso de abordagens analíticas adequadas, sendo a análise não-paramétrica de Kaplan–Meier uma etapa inicial fundamental. A função de sobrevivência dos pacientes pode, então, ser estimada por meio desse estimador.

Para definir tal estimador, sejam $t_1 < \dots < t_m$ os instantes observados de ocorrência do evento (óbito). Para cada tempo t_j , define-se d_j como número de óbitos observados nesse instante e r_j como o número de pacientes em risco imediatamente antes de t_j . O estimador de Kaplan-Meier para a função de sobrevivência, denotado por $\hat{S}(t)$, é dado por:

$$\hat{S}(t) = \prod_{t_{(j)} \leq t} \left(1 - \frac{d_j}{r_j}\right) \quad (2.6)$$

Esse procedimento resulta em uma curva de sobrevivência que descreve, de forma empírica, a probabilidade de um indivíduo não apresentar o desfecho até determinado tempo.

2.5. MODELO GAMA

A distribuição gama, que inclui a distribuição exponencial como caso particular, foi utilizada, a primeira vez sob o contexto de análise de sobrevivência, por Brown e Flood (1947) para descrever o tempo de vida de copos de vidro circulando em uma cafeteria e, também, por Birnbaum e Saunders (1958) para descrever o tempo de vida de materiais eletrônicos. Na área médica, sua utilização é mais recente, sendo vista em 1979 no trabalho de Vaupel et. al (1979) que introduziu o conceito de fragilidade com base na distribuição gama e, alguns anos depois, no trabalho de Abadi et. al (2012) que introduziu essa distribuição para a análise paramétrica dos tempos de sobrevivência de pacientes com câncer de mama. Mas estes trabalhos não foram os únicos, a distribuição gama, junto a distribuição Weibull, é uma das distribuições mais utilizadas no contexto de análise de sobrevivência, devido sua flexibilidade em modelar diferentes formas de risco ao longo do tempo, e modelagem de dados assimétricos, que são comuns em análise de sobrevivência. A função densidade de probabilidade da distribuição gama é dada por:

$$f(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp\left\{-\left(\frac{x}{\beta}\right)\right\} \quad (2.7)$$

em que β é o parâmetro de escala, α é o parâmetro de forma, e $\Gamma(\cdot)$ é a função gama completa, dada por $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$. Para essa distribuição, a função de distribuição acumulada é dada por:

$$F(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)} \gamma\left(\alpha, \frac{x}{\beta}\right) \quad (2.8)$$

em que $\gamma(\cdot)$ representa a função gama incompleta (inferior), dada por $\gamma\left(\alpha, \frac{x}{\beta}\right) = \int_0^{\frac{x}{\beta}} x^{\alpha-1} e^{-x} dx$. Isso implica que a função de sobrevivência dessa distribuição é dada por:

$$S(x; \alpha, \beta) = 1 - \frac{1}{\Gamma(\alpha)} \gamma\left(\alpha, \frac{x}{\beta}\right) \quad (2.9)$$

Dentre as propriedades dessa distribuição, uma das mais importantes é o fato que à medida que β aumenta, ela se aproxima de uma distribuição normal com média $\mu = \alpha\beta$ e variância $\sigma^2 = \alpha\beta^2$. Além dessa propriedade, pode-se destacar também o coeficiente de assimetria que, no caso da distribuição gama, depende apenas do parâmetro α e é igual a $2/\sqrt{\alpha}$. Por outro lado, em relação aos momentos, tem-se que a função geradora de momentos da distribuição gama é dada por:

$$M(t) = (1 - \beta t)^{-\alpha} \quad (2.10)$$

desde que $t < 1/\beta$, o que implica que a média é $\alpha\beta$ e a variância $\alpha\beta^2$. Já o logaritmo da função de verossimilhança dessa distribuição, para uma amostra aleatória $x_1, \dots, x_n$, é dado por:

$$\ell(\alpha, \beta) = (1 - \alpha) \sum_{i=1}^n \ln x_i - \sum_{i=1}^n \frac{x_i}{\beta} - n\alpha \ln \beta - n \ln \Gamma(\alpha) \quad (2.11)$$

que implica que o estimador de máxima verossimilhança de β tem forma analítica e é definido por $\hat{\beta} = \bar{x}/\hat{\alpha}$, porém, para α , o estimador de máxima verossimilhança é obtido por métodos numéricos, como o método de Newton-Raphson.

2.6. MODELOS DE FRAÇÃO DE CURA DE NÃO-MISTURA

Em estudos de sobrevida, observa-se com frequência que a curva de sobrevivência relativa tende a atingir um platô após determinado período de acompanhamento. De acordo com Lambert et al. (2007), esse efeito de platô ocorre quando a taxa de mortalidade dos indivíduos doentes se iguala à taxa de mortalidade esperada na população geral, ou seja, quando a mortalidade em excesso se torna nula. Tal comportamento sugere a presença de uma fração de cura na população-alvo, ou seja, um subgrupo de pacientes considerados efetivamente livres do risco associado à doença (isto é, “curados”).

Abordagens tradicionais, como o estimador de Kaplan–Meier e o modelo de riscos proporcionais de Cox, nos quais fornecem estimativas da probabilidade de sobrevida e permitem a inclusão de múltiplas covariáveis explicativas, cujo não contemplam de modo explícito a hipótese de que determinados pacientes possam apresentar uma probabilidade muito baixa, ou nula, de evoluírem para o desfecho dentro do período de acompanhamento (follow-up), que caracteriza a presença de uma fração de cura na população-alvo (TSODIKOV et al., 2003, THERNEAU & GRAMBSCH, 2000; MALLER & ZHOU, 1996). Diante deste cenário, os chamados modelos de fração de cura (cure models) surgem exatamente para atender a essa necessidade: identificar e quantificar a proporção de indivíduos que não apresentam o evento de interesse durante o acompanhamento (ACHCAR et al., 2012). Em termos práticos, esses modelos classificam a população em suscetíveis (com risco efetivo de falecer no período analisado) e “curados” (que, por motivos diversos, têm risco praticamente nulo de evolução para o óbito ao longo do tempo avaliado). De acordo com Vahidpour (2016), existem duas abordagens para a construção desses modelos de fração de cura, o método de mistura e o método de não-mistura, sendo este último menos comum. Neste estudo, sob uma ótica mais inovadora, será considerado a abordagem de não-mistura para a construção do modelo de fração de cura, pois, diferente dos modelos de mistura, os modelos de não-mistura definem uma assíntota

para o risco cumulativa, permitindo uma interpretação mais prática do modelo. Para estes modelos, a função de sobrevivência é descrita por:

$$S(t) = \rho^{F_0(t)} = \exp\{\ln(\rho) F_0(t)\} \quad (2.12)$$

onde $t \in \mathbb{R}_+$, $\rho \in (0,1)$ é a probabilidade de cura e que $F_0(t) = 1 - S_0(t)$ denota a função distribuição acumulada para o grupo não-curado (ou suscetível) na população, chamada de baseline. No presente estudo, essa baseline será definida com base na distribuição gama introduzida na Seção 2.2. Da Equação 2.7, obtém que a função de densidade de probabilidade é descrita por:

$$f(t) = -\log(\rho) f_0(t) \exp\{\log(\rho) F_0(t)\} \quad (2.13)$$

com $t \in \mathbb{R}_+$

2.7. RESÍDUOS DE COX-SNELL

Para atestar a qualidade de ajuste do modelo proposto, em particular, na presença de covariáveis, será considerada uma análise de resíduos baseada nos resíduos de Cox-Snell, cuja definição deve ser adaptada à estrutura de modelos com cura, visto que apenas uma parcela da população está em risco de experimentar o desfecho. No caso dos modelos de não-mistura, a construção do resíduo de Cox-Snell deve basear-se na componente suscetível, isto é, na função $F_0(t)$, estimada a partir do modelo. Sendo assim, os resíduos de Cox-Snell sob o modelo de fração de cura de não-mistura são definidos como:

$$r_{C_i}^* = F_{0,i}(t_i) = \frac{\log\{\hat{S}_i(t_i)\}}{\log(\rho_i)} \quad (2.14)$$

em que $\hat{S}_i(t)$ é a função de sobrevivência estimada.

3 MATERIAIS E MÉTODOS

3.1. BASE DE DADOS

Para a condução do estudo, conforme dito anteriormente, foi considerada uma base de dados obtidos de forma secundária do estudo de John et al. (2015), referentes à 143 fumantes admitidos em um hospital comunitário de 180 leitos em La Chaux-de-Fonds, Suíça, que incluiu fumantes que apresentaram pelo menos uma mancha de tabaco nos dedos no momento da admissão aos departamentos de medicina interna, cirurgia ou unidade de terapia intensiva, independentemente de seu diagnóstico inicial, e que foram hospitalizados no período entre março de 2006 e dezembro de 2007 (6% dos fumantes). Como grupo controle, foram considerados fumantes admitidos nos mesmos departamentos e no mesmo hospital até janeiro de 2010, porém sem a presença de manchas de tabaco nos dedos, sendo 26,57% dos dados censurados. A Tabela 1 apresenta um resumo das características dos pacientes na base de dados.

Tabela 1. Características dos pacientes tabagistas readmitidos e não-readmitidos ao longo do período de acompanhamento do hospital La Chaux-de-Fonds, no estudo de John et al. (2015).

	Variáveis Contínuas	
	Não-Readmitidos (105)	Readmitidos (38)
	Média ± Desvio-Padrão	Média ± Desvio-Padrão
Idade	40 ± 10	57,10 ± 12,49
Peso	71,35 ± 17,36	72,57 ± 17,36
Altura	170,07 ± 8,49	170,52 ± 8,49
Tempo de Uso de Tabaco	43,27 ± 11,34	38,10 ± 11,34
Maços por Dia	1,30 ± 0,48	1,30 ± 0,48
	Variáveis Categóricas	
	Não-Readmitidos (105)	Readmitidos (38)
	n (%)	n (%)
Sexo		
Masculino	70 (66,67%)	23 (60,53%)
Feminino	35 (33,33%)	15 (39,47%)

Doença Cardiovascular		
Não	53 (50,48%)	19 (50%)
Sim	52 (49,52%)	19 (50%)
Derrame Cerebral		
Não	93 (88,57%)	34 (89,47%)
Sim	12 (11,43%)	4 (10,53%)
Doença Arterial		
Não	85 (80,95%)	32 (84,21%)
Sim	20 (19,05%)	6 (15,79%)
Câncer por Tabaco		
Não	89 (87,76%)	30 (78,95%)
Sim	16 (15,24%)	8 (21,05%)
Hipertensão		
Não	50 (47,62%)	18 (47,37%)
Sim	55 (52,38%)	20 (52,63%)
Depressão		
Não	71 (67,62%)	30 (78,95%)
Sim	34 (32,38%)	8 (21,05%)
Tosse		
Não	41 (39,05%)	17 (44,74%)
Sim	64 (60,95%)	21 (55,26%)
Ortopneia		
Não	92 (87,62%)	36 (94,74%)
Sim	13 (12,38%)	2 (5,26%)

Fonte: Elaboração própria com base nos dados do estudo.

3.2 MODELO DE FRAÇÃO DE CURA DE NÃO-MISTURA GAMA

Seja T uma variável aleatória correspondente ao tempo até a ocorrência do evento de interesse. Seguindo a formulação do modelo de não-mistura introduzida na Seção 2.6 e adotando como baseline a distribuição gama (Seção 2.5) com densidade

expressão na Equação 2.7 e distribuição acumulada expressa na Equação 2.8, pode-se construir o modelo de fração de cura de não-mistura gama da seguinte forma:

$$S(t) = \exp\left\{\ln(\rho) \frac{1}{\Gamma(\alpha)} \gamma\left(\alpha, \frac{t}{\beta}\right)\right\} \quad (3.1)$$

em que $\gamma(\cdot)$ representa a função gama incompleta (inferior). Observe que, se $t \rightarrow \infty$, então $S(t) \rightarrow \rho$, ou seja, a função de sobrevivência, neste caso, tem uma assíntota na taxa de cura ρ .

Para a estimação dos parâmetros do modelo proposto, foi considerado o método da máxima verossimilhança. Neste caso, seja $\{(t_i, \delta_i)\}_{i=1}^n$, onde t_i é o tempo observado e $\delta_i \in \{0,1\}$ é o indicador de censura, uma amostra aleatória iid. A função verossimilhança de $\theta = (\alpha, \beta, \rho)$ é descrita por:

$$L(\theta) = \prod_{i=1}^n \left[-\log(\rho) \frac{t_i^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha} \exp\left\{-\left(\frac{t_i}{\beta}\right)\right\} \exp\left\{\log(\rho) \frac{1}{\Gamma(\alpha)} \gamma\left(\alpha, \frac{t_i}{\beta}\right)\right\} \right]^{\delta_i} \times \left[\exp\left\{\ln(\rho) \frac{1}{\Gamma(\alpha)} \gamma\left(\alpha, \frac{t_i}{\beta}\right)\right\} \right]^{1-\delta_i} \quad (3.2)$$

Com base na verossimilhança acima, para determinar o estimador do vetor de parâmetros θ , deve-se maximizar L , ou, de forma equivalente, $\ell(\theta) = \log L(\theta)$. No entanto, dada a complexidade da verossimilhança e o fato de que a distribuição gama só possui forma fechada para um de seus estimadores, é necessário fazer o uso de métodos numéricos como Newton-Raphson ou Nelder-Mead para determinar os estimadores de máxima verossimilhança de θ .

No presente trabalho, esse processo será realizado por meio da função *optim* do ambiente R (versão 4.3.0), usando o método numérico de Nelder-Mead para a maximização de $\ell(\theta)$. Sob condições de regularidade (ver LEHMANN & CASELLA, 1998, pp. 461–463), a distribuição assintótica do estimador de máxima verossimilhança $\hat{\theta}$ é uma distribuição normal multivariada com média θ e matriz de covariância $\Sigma(\hat{\theta})$, a qual pode ser estimada de forma consistente pelo inverso da matriz de informação de Fisher, avaliada em $\theta = \hat{\theta}$, dada por:

$$\Sigma(\hat{\theta}) = \left[-\frac{\partial \ell(\theta)}{\partial \theta \partial \theta^\top} \right]^{-1} \quad (3.3)$$

em que,

$$\begin{aligned} \frac{\partial \ell(\theta)}{\partial \alpha} &= \sum_{i=1}^n \log t_i - n \psi(\alpha) - n \log \beta - \sum_{i=1}^n \frac{\psi(\alpha)}{\Gamma(\alpha)^2} + \left(\sum_{i=1}^n \frac{\partial \gamma\left(\alpha, \frac{t_i}{\beta}\right)}{\partial \alpha} \right) n \\ \frac{\partial \ell(\theta)}{\partial \beta} &= -\frac{n\alpha}{\beta} + \sum_{i=1}^n \frac{t_i}{\beta^2} + \left(\sum_{i=1}^n \frac{\partial \gamma\left(\alpha, \frac{t_i}{\beta}\right)}{\partial \beta} \right) \left(\sum_{i=1}^n \delta_i + n - \sum_{i=1}^n \delta_i \right) \end{aligned} \quad (3.4)$$

$$\frac{\partial \ell(\theta)}{\partial \rho} = -\frac{n}{\rho \log \rho} + \frac{n}{\rho} + \frac{n - \sum_{i=1}^n \delta_i}{\rho}$$

onde δ_i é o indicador de censura. Neste estudo, em particular, a matriz de informação de Fisher foi estimada numericamente por meio da função *optim*, com o argumento *hessian = TRUE*, no ambiente R (versão 4.3.0). A inversa dessa matriz foi então utilizada para estimar a matriz de variância-covariância assintótica dos estimadores, o que possibilitou a construção de intervalos de confiança com base na aproximação normal multivariada dos estimadores de máxima verossimilhança.

4 ESTUDO DE SIMULAÇÃO MONTE CARLO

Para avaliar as propriedades dos estimadores de máxima verossimilhança do modelo proposto, foi conduzido um estudo de simulação Monte Carlo. Para tal, foi considerada a geração de 1000 amostras pseudoaleatórias com tamanhos 25, 50, 75 e 100, a partir do modelo gama com fração de cura e aproximadamente 23% de censura à direita, que reflete nossa censura empírica. O método da transformação inversa foi utilizado para gerar as amostras pseudoaleatórias, e os valores nominais considerados para os parâmetros do modelo foram: $\alpha = 0.5$, $\beta = 0.5$ e $\rho = 0.2$ (Caso 1); $\alpha = 1.0$, $\beta = 1.0$ e $\rho = 0.4$ (Caso 2) e $\alpha = 1.5$, $\beta = 1.5$ e $\rho = 0.6$ (Caso 3).

Como estimador pontual, adotou-se o estimador de máxima verossimilhança, cujo desempenho foi avaliado em termos de viés (B) e erro quadrático médio (EQM). Assim, a partir das amostras dos parâmetros do modelo, obtiveram-se estimativas de Monte Carlo aproximadas para essas medidas, dadas por:

$$\hat{B}(\hat{\epsilon}) = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\epsilon}_i - \epsilon) \quad (3.5)$$

E,

$$\widehat{EQM}(\hat{\epsilon}) = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\epsilon}_i - \epsilon)^2 \quad (3.6)$$

onde $\epsilon = (\alpha, \beta, \rho)$ representa os parâmetros do modelo proposto e $\{\hat{\epsilon}_i\}_{i=1}^{1000}$ é a sequência das estimativas obtidas para cada amostra. Os resultados são apresentados nas Tabelas 2, 3 e 4, e mostram que os vieses das estimativas apresentaram magnitude reduzida e tenderam a zero à medida que o tamanho amostral aumenta, evidenciando a consistência do estimador. Observou-se também que os EQMs diminuíram com o aumento de n .

Tabela 2. Resultado do estudo de simulação Monte Carlo para $\alpha=0.5$, $\beta=0.5$ e $\rho=0.2$.

n	Parâmetro	B (Viés)	EQM
25	α	0.1700	0.0696
	β	0.6295	4.1450
	ρ	-0.0010	0.0085
50	α	0.1203	0.0317
	β	0.6727	2.9100
	ρ	-0.0084	0.0053
75	α	0.1081	0.0213
	β	0.5111	0.9338
	ρ	-0.0106	0.0039
100	α	0.0998	0.0161
	β	0.5096	0.8617
	ρ	-0.0126	0.0030

Fonte: Elaboração própria com base nos dados do estudo.

Tabela 3. Resultado do estudo de simulação Monte Carlo para $\alpha=1.0$, $\beta=1.0$ e $\rho=0.4$.

<i>n</i>	Parâmetro	B (Viés)	EQM
25	α	0.3207	0.3563
	β	0.0819	0.2601
	ρ	0.0029	0.0099
50	α	0.2059	0.1256
	β	0.1138	0.1362
	ρ	0.0010	0.0048
75	α	0.1849	0.0855
	β	0.1125	0.0969
	ρ	-0.0006	0.0032
100	α	0.1457	0.0551
	β	0.1376	0.0805
	ρ	0.0005	0.0023

Fonte: Elaboração própria com base nos dados do estudo.

Tabela 4. Resultado do estudo de simulação Monte Carlo para $\alpha=1.5$, $\beta=1.5$ e $\rho=0.6$.

<i>n</i>	Parâmetro	B (Viés)	EQM
25	α	1.3754	17.6676
	β	1.0792	267.4401
	ρ	-0.0108	0.0194
50	α	0.3913	1.0209
	β	0.7432	84.0980
	ρ	-0.0139	0.0108
75	α	0.2764	0.4066
	β	0.2593	1.8607
	ρ	-0.0063	0.0059
100	α	0.1978	0.2758
	β	0.2257	0.8011
	ρ	-0.0045	0.0043

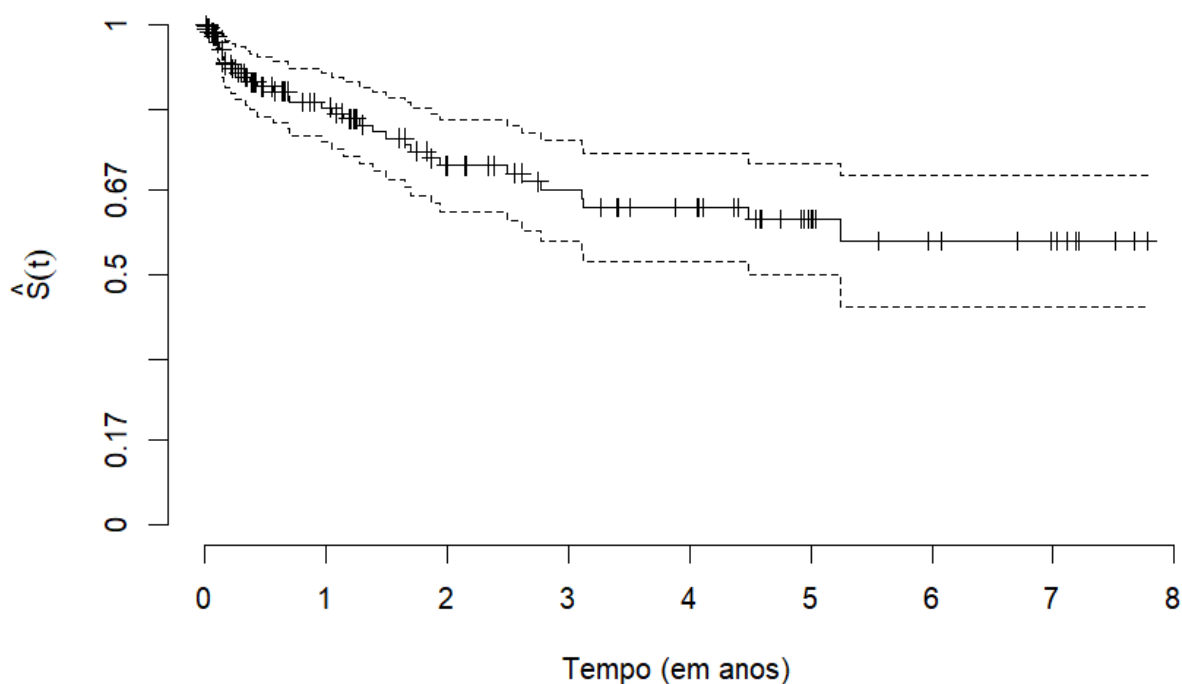
Fonte: Elaboração própria com base nos dados do estudo.

5 RESULTADOS

5.1. ANÁLISE EXPLORATÓRIA DE DADOS

Para uma análise preliminar dos dados, considerou-se a curva de sobrevivência estimada pelo Kaplan-Meier, junto a uma análise de correlação das variáveis, com base no coeficiente de correlação de Spearman. A Figura 1 apresenta a curva de sobrevivência estimada pelo Kaplan-Meier, que reflete a probabilidade de sobrevida dos pacientes readmitidos no hospital ao longo do tempo de acompanhamento do estudo. A linha sólida representa a estimativa pontual da função de sobrevivência, enquanto a faixa sombreada corresponde ao intervalo de confiança de 95%, fornecendo uma medida da incerteza associada às estimativas em cada ponto temporal. Os resultados evidenciam um declínio acentuado na probabilidade de sobrevivência nos primeiros anos, com uma estabilidade da curva após 5 anos de acompanhamento, formando um platô que sugere a presença de um subgrupo de pacientes que não experimenta o evento de interesse (readmissão), isto é, uma fração de “cura”).

Figura 1. Curva de sobrevivência estimada via Kaplan-Meier referente ao tempo até a readmissão no hospital.

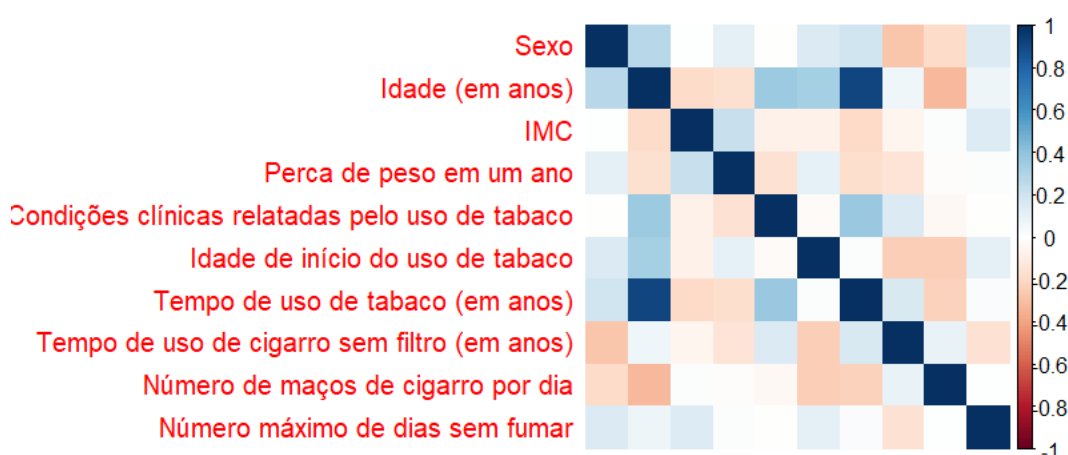


Fonte: Elaboração própria com base nos dados do estudo.

A Figura 2 exibe a matriz de correlação entre as covariáveis do estudo, calculada a partir do coeficiente de correlação de Spearman, o qual é apropriado para avaliar associações monotônicas, inclusive em presença de relações não lineares ou de variáveis ordinais. Os coeficientes variam entre -1 e 1, sendo representados por um gradiente de cores, onde tons azulados indicam correlações negativas e tons avermelhados, correlações positivas.

Da Figura 2, observa-se que as variáveis idade e tempo de uso de tabaco apresentam a correlação mais elevada entre os pares analisados, o que é coerente com a expectativa de que indivíduos mais velhos tendem a relatar maior tempo de uso de cigarro durante a vida. Além disso, existe uma correlação negativa moderada entre idade de início do uso de tabaco e tempo de uso de tabaco, o que reflete a relação inversa esperada: quanto mais precoce o início, maior tende a ser o tempo total de exposição ao cigarro. Além disso, a matriz de correlação não revela sinais evidentes de multicolinearidade severa entre as covariáveis. Essa constatação é reforçada pela ausência de correlações muito próximas de ± 1 , o que sugere que as variáveis contribuem com informações relativamente distintas para a construção do modelo de fração de cura proposto.

Figura 2. Matriz de correlação de Spearman entre as covariáveis do estudo.

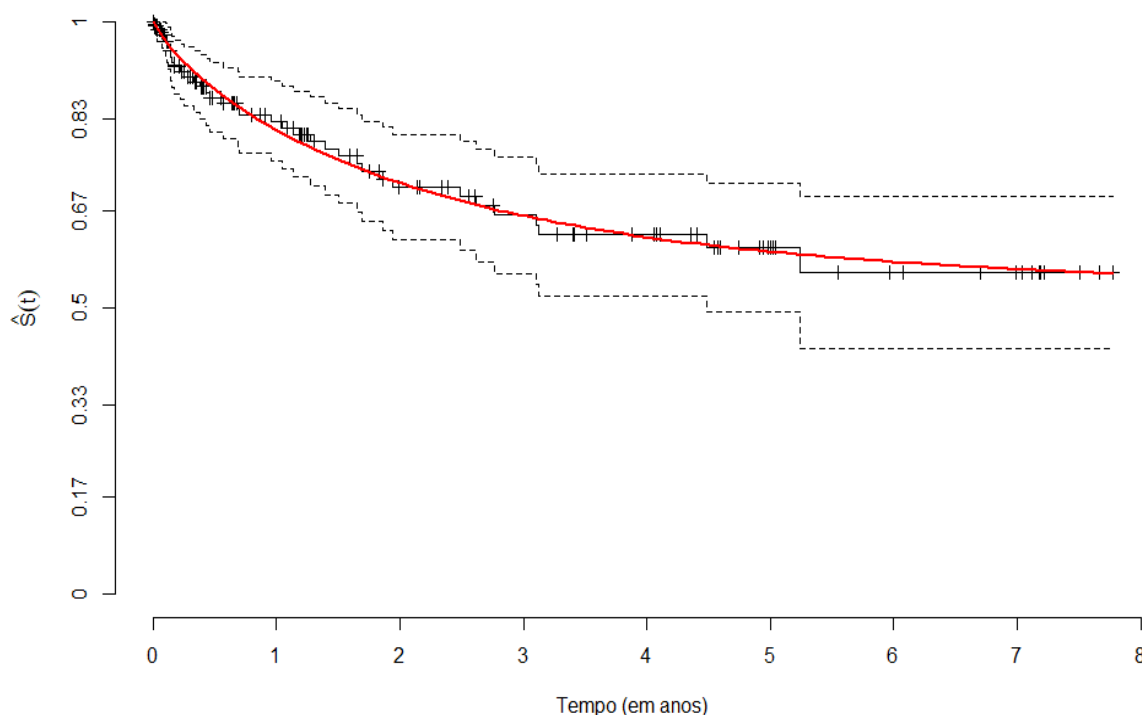


Fonte: Elaboração própria com base nos dados do estudo.

5.2. ANÁLISE PELO MODELO DE FRAÇÃO DE CURA GAMA

Na análise inicial, foi realizada a estimação dos parâmetros dos modelos propostos sem considerar o uso de estruturas de regressão, com o intuito de avaliar se o modelo proposto se adequava a curva de sobrevivência empírica obtida pelo estimador Kaplan-Meier. O resultado dessa análise mostrou que o processo de estimação dos parâmetros do modelo proposto convergiu de forma satisfatória, e a fração de cura estimada foi de 56,7%. A Figura 3 ilustra a comparação gráfica entre o modelo de fração de cura de não-mistura gama proposto e a curva de sobrevivência estimada via Kaplan-Meier, onde observa-se que a curva paramétrica do modelo proposto adere bem à trajetória empírica dos dados, sugerindo que a distribuição gama (com fração de cura) captura adequadamente a dinâmica de sobrevivência da coorte.

Figura 3. Curva de sobrevivência estimada via Kaplan-Meier versus curva de sobrevivência ajustada pelo modelo de fração de cura de não-mistura gama referente ao tempo até a readmissão no hospital.



Fonte: Elaboração própria com base nos dados do estudo.

Prosseguindo com a análise, incorporou-se uma estrutura de regressão para o parâmetro ρ , o qual representa a fração de cura do modelo proposto, com o objetivo de avaliar as influências dos fatores de risco sobre o comportamento dos pacientes curados e quais níveis de risco podem levar os pacientes à morte. Sendo assim, a estrutura de regressão adotada para ρ_i , associada ao i -ésimo paciente, é definida pela função logística descrita por:

$$\log\left(\frac{\rho_i}{1-\rho_i}\right) = \beta_0 + \sum_{j=1}^{10} \beta_j X_{ji} \quad (4.1)$$

em que os preditores X_{ji} representam os fatores de risco individuais, listados a seguir:

- x_{1i} : Sexo (0 = Masculino; 1 = Feminino)
- x_{2i} : Idade (em anos)
- x_{3i} : IMC
- x_{4i} : Perca de peso em um ano
- x_{5i} : Condições clínicas relacionadas pelo uso de tabaco
- x_{6i} : Idade de início do uso de tabaco
- x_{7i} : Tempo de uso de tabaco (em anos)
- x_{8i} : Tempo de uso de cigarro sem filtro (em anos)
- x_{9i} : Número de maços de cigarro por dia
- x_{10i} : Número máximo de dias sem fumar

A estimação dos parâmetros foi realizada no ambiente R, por meio da função *optim*, utilizando o método de Nelder-Mead para a maximização da função de log-verossimilhança. O argumento *hessian = TRUE* foi empregado para obter a matriz Hessiana no ponto de máxima verossimilhança, permitindo a estimação da matriz de variância-covariância dos estimadores e, conseqüentemente, a construção de intervalos de confiança (IC) assintóticos.

Os resultados da análise são exibidos na Tabela 5, que inclui as estimativas pontuais dos coeficientes de regressão β_j , os erros-padrão assintóticos associados, e os respectivos ICs ao nível de 95%. A significância de cada coeficiente foi avaliada com base na presença ou ausência do valor zero dentro dos ICs, em que a exclusão

do valor zero do intervalo sugere que a covariável (ou fator de risco) associada exerce influência sobre a probabilidade de cura. Por outro lado, a inclusão do valor zero no intervalo de confiança implica ausência de evidência estatística suficiente para afirmar que o efeito da covariável sobre a fração de cura difere de zero, dentro do nível de significância adotado.

Tabela 5. Estimativas de parâmetros e intervalos de confiança para 95% do modelo de regressão de fração de cura de não-mistura gama.

Parâmetro	Estimativa	Erro-Padrão	I.C 95%
β_0 (Intercepto)	1.2040	2.5158	(-3.7268, 6.1349)
β_1 (Sexo – 1 vs 0)	1.0582	0.7250	(-0.3628, 2.4792)
β_2 (Idade)	-0.0140	0.1637	(-0.3347, 0.3068)
β_3 (IMC)*	-0.2048	0.0714	(-0.3447, -0.0648)
β_4 (Perca de peso em um ano)	-0.0155	0.0368	(-0.0876, 0.0566)
β_5 (Condições clínicas pelo uso de tabaco – 1 vs 0)	-1.4019	0.9884	(-3.3391, 0.5353)
β_6 (Idade de início do uso de tabaco)	-0.0039	0.1687	(-0.3346, 0.3268)
β_7 (Tempo de uso de tabaco)	-0.0768	0.1679	(-0.4059, 0.2524)
β_8 (Tempo de uso de cigarro sem filtro)	-0.0070	0.0174	(-0.0411, 0.0270)
β_8 (Número de maços de cigarro por dia)	-0.7331	0.5355	(-1.7827, 0.3165)
β_8 (Número máximo de dias sem fumar)	< 0.0001	0.0015	(-0.0029, 0.0030)
α	1.2935	0.1829	(0.9350, 1.6521)
β	1.3985	0.3361	(0.7398, 2.0572)

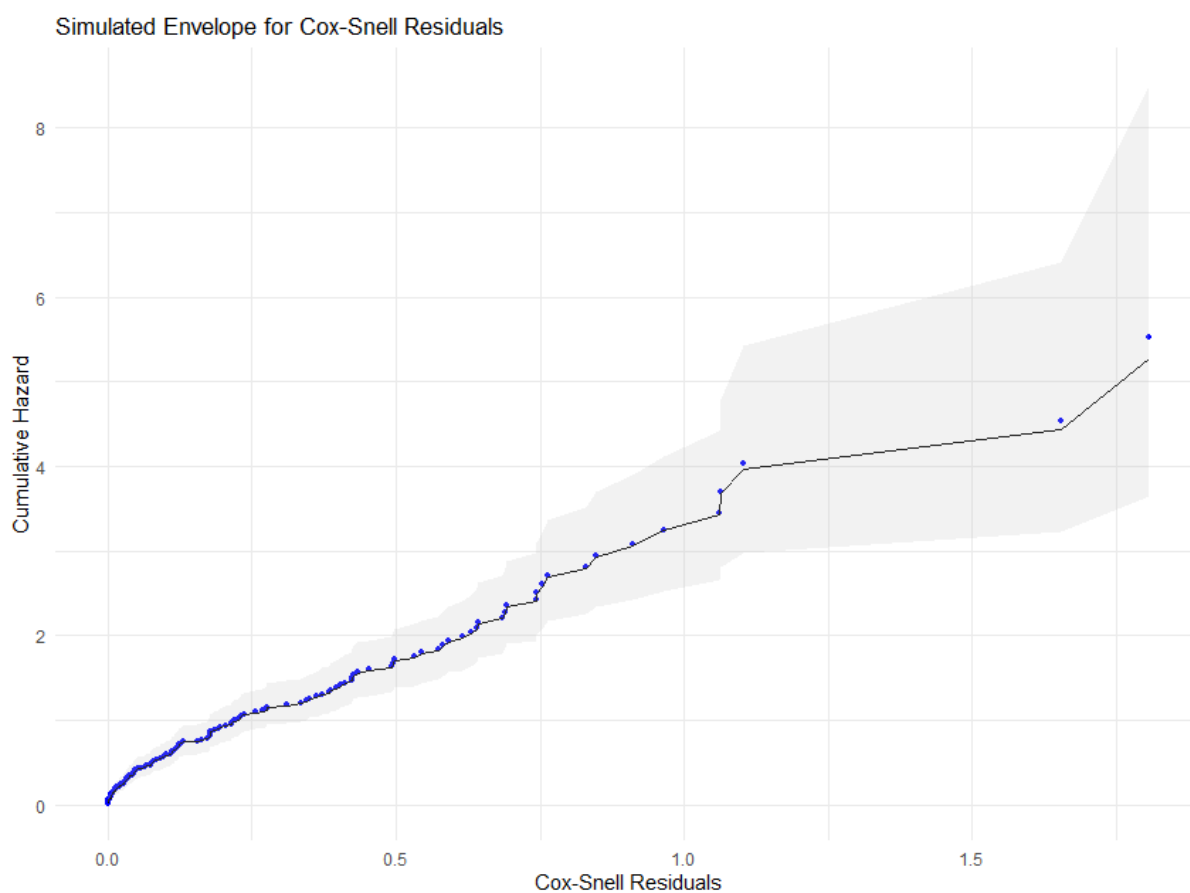
Fonte: Elaboração própria com base nos dados do estudo.

Dos resultados da Tabela 5, observa-se que apenas o IMC foi significativo, indicando que indivíduos com IMC mais alto apresentam maior risco de readmissão hospitalar ao longo do tempo. No estudo de Azodi et al. (2006), os resultados mostraram que um IMC elevado foi significativamente associado a um aumento no tempo médio de internação, sendo 4,7% maior para pacientes com sobrepeso e 7,0% maior para pacientes com obesidade em comparação com aqueles com peso normal.

A falta de significância das outras variáveis pode estar associada a um viés de seleção ou um perfil atípico de pacientes no conjunto de dados considerado no estudo. Embora alguns fatores demográficos sejam frequentemente suspeitos de influenciar os riscos de internação hospitalar relacionados ao uso do tabaco, alguns estudos relataram resultados nulos. No estudo de Kagabo et al (2019), analisaram que o sexo do paciente não era um preditor significativo de readmissão em uma coorte psiquiátrica após o controle do status de tabagismo.

A Figura 4 ilustra o gráfico dos resíduos de Cox-Snell obtidos a partir do modelo de fração de cura de não-mistura com distribuição Gama ajustado aos dados, em que se observa que o modelo se adere aos dados, sem violações graves de seus pressupostos.

Figura 4. Resíduos de Cox-Snell para o modelo de fração de cura de não-mistura gama proposto.



Fonte: Elaboração própria com base nos dados do estudo.

6. CONCLUSÃO

O presente estudo teve como objetivo principal analisar os fatores associados à readmissão hospitalar, com ênfase na influência do tabagismo, por meio da aplicação de um modelo de fração de cura do tipo não-mistura baseado na distribuição Gama. Inicialmente, a análise sem estrutura de regressão demonstrou que o modelo proposto é adequado para representar a dinâmica de sobrevivência observada, conforme evidenciado pela boa concordância entre a curva estimada pelo modelo paramétrico e a curva empírica de Kaplan–Meier. A fração de cura estimada foi de 56,7%, sugerindo que uma proporção significativa dos pacientes não foi readmitida no período de observação.

Ao incorporar uma estrutura de regressão para modelar a fração de cura em função de variáveis clínicas e comportamentais, foi possível avaliar a influência de potenciais fatores de risco sobre a probabilidade de não-readmissão. Dentre os preditores analisados, apenas o índice de massa corporal apresentou associação estatisticamente significativa, indicando que pacientes com maiores valores de IMC tendem a ter maior risco de readmissão hospitalar ao longo do tempo. As demais covariáveis não apresentaram significância estatística, o que pode ser atribuído, em parte, a limitações da amostra, como viés de seleção ou características particulares da população estudada. Adicionalmente, a análise dos resíduos de Cox-Snell confirmou a adequação do modelo ajustado, uma vez que não foram observadas violações relevantes dos pressupostos de modelagem, reforçando a validade dos resultados obtidos.

REFERÊNCIAS

- ABADI, A. et al.** Breast Cancer Survival Analysis: Applying the Generalized Gamma Distribution under Different Conditions of the Proportional Hazards and Accelerated Failure Time Assumptions. *International Journal of Preventive Medicine*, v. 30, n. 1, p. 57-76, 2012. Disponível em: https://www.researchgate.net/publication/231614473_Breast_Cancer_Survival_Analysis_Applying_the_Generalized_Gamma_Distribution_under_Different_Conditions_of_the_Proportional_Hazards_and_Accelerated_Failure_Time_Assumptions. Acesso em: 29 mai. 2025.
- ACHCAR, J. A.; et al.** Cure fraction models using mixture and non-mixture models. *Tatra Mountains*, v. 51, n. 1, p. 1-9, 2012. Disponível em: <https://sciendo.com/article/10.2478/v10127-012-0001-4?tab=abstract>. Acesso em: 06 jun. 2025.
- AMY, P. et al.** Global statistics on alcohol, tobacco and illicit drug use: 2017 status report. *National Library of Medicine*, v. 113. DOI: <https://doi.org/10.1111/add.14234>. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/add.14234>. Acesso em: 16 set. 2024.
- ARENA, G. et al.** Hospital length of stay and readmission after elective surgery: a comparison of current and former smokers with non-smokers. *BMC Health Services Research*, [S. l.], v. 24, n. 85, 2024. Disponível em: <https://doi.org/10.1186/s12913-024-10566-3>. Acesso em: 3 jun. 2025.
- AZODI, O. Sadr et al.** The impact of tobacco use and body mass index on the length of stay in hospital and the risk of post-operative complications among patients undergoing total hip replacement. *The Journal of Bone and Joint Surgery. British Volume*, [S. l.], v. 88-B, n. 10, p. 1316-1320, out. 2006. Disponível em: <https://doi.org/10.1302/0301-620X.88B10.17957>. Acesso em: 07 out. 2025.
- BENÍTEZ, P. N. et al.** Survival Analysis and COX Regression. Elsevier, vol. 39. DOI: 10.1016/j.aller.2011.07.007. Disponível em: <https://www.elsevier.es/en-revista-allergologia-et-immunopathologia-105-linkresolver-survival-analysis-cox-regression-S0301054611002679>. Acesso em: 23 set. 2024.
- BIRNBAUM, Z. W.; SAUNDERS, S. C.** A Statistical Model for Life-Length of Materials. *Journal of the American Statistical Association*, [S. l.], v. 53, n. 281, p. 151-160, 1958. Disponível em: <https://www.jstor.org/stable/2282575>. Acesso em: 6 jun. 2025.

BROWN, G. W.; FLOOD, M. M. Tumbler Mortality. *Journal of the American Statistical Association*, [S. l.], v. 42, n. 240, p. 562-574, dez. 1947. Disponível em: <http://www.jstor.org/stable/2280012>.

CARTER, B. D. et al. Smoking and Mortality – Beyond Established Causes. *The New England Journal of Medicine*, vol. 372. DOI: 10.1056/NEJMsa1407211. Disponível em: <https://www.nejm.org/doi/full/10.1056/NEJMsa1407211>. Acesso em: 16 out. 2024.

CARTMELL, K. B. et al. Effect of an Evidence-Based Inpatient Tobacco Dependence Treatment Service on 30, 90 and 180-Day Hospital Readmission Rates. *Medical Care*, v. 56, n. 4, p. 358-363, abr. 2018. DOI: <https://doi.org/10.1097/MLR.0000000000000884>. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC5851827/>. Acesso em: 03 jun. 2025.

COLLETT, D. *Modelling Survival Data in Medical Research*. 3. ed. New York: Chapman and Hall/CRC, 2015.

COLOSIMO, E. A.; GIOLO, S. R. *Análise de sobrevivência aplicada*. 1. ed. São Paulo: Edgard Blücher, 2006.

COX, D. R. Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, v. 34, n. 2, p. 187–220, 1972.

HUEY, S. W. et al. Smoke screen: The teen vaping epidemic uncovers a new concerning addiction. *National Library of Medicine*. DOI: 10.1097/JXX.0000000000000234. Disponível em: https://journals.lww.com/jaanp/abstract/2020/04000/smoke_screen__the_teen_vaping_epidemic_uncovers_a.4.aspx. Acesso em: 13 out. 2024.

JOHN, G. et al. Tobacco-Stained Fingers and Its Association with Death and Hospital Admission: A Retrospective Cohort Study. *National Library of Medicine*, [S. l.], v. 10, n. 9, p. e0138211, 2015. DOI: <https://doi.org/10.1371/journal.pone.0138211>. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4573751/>. Acesso em: 23 set. 2024.

JOHN, G. et al. Tobacco-stained fingers: a clue for smoking-related disease or harmful alcohol use? A case–control study. *BMJ Open*, [S. l.], v. 3, n. 11, p. e003304, 2013. DOI: 10.1136/bmjopen-2013-003304. Disponível em: <https://bmjopen.bmj.com/content/3/11/e003304>. Acesso em: 16 out. 2024.

JONES, K. et al. The vaping epidemic in adolescents. *Missouri Medicine*, [S. l.], v. 117, n. 1, p. 56-59, 2020. Disponível em:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7023954/#b3-ms117_p0056. Acesso em: 13 out. 2024.

KAGABO, R. *et al.* Association between smoking, and hospital readmission among inpatients with psychiatric illness at an academic inpatient psychiatric facility, 2000–2015. Elsevier, [S. l.], v. 9, p. 100181, 2019. DOI: <https://doi.org/10.1016/j.abrep.2019.100181>. Disponível: <https://www.sciencedirect.com/science/article/pii/S2352853219300471>. Acesso em: 15 out. 2024.

KALBFLEISCH, J. D.; PRENTICE, R. L. The Statistical Analysis of Failure Time Data. 2. ed. Hoboken: John Wiley & Sons, 2002.

KLEIN, J. P.; MOESCHBERGER, M. L. Survival analysis: techniques for censored and truncated data. 2. ed. New York: Springer, 2003.

LAMBERT, P. A. Modeling of the Cure Fraction in Survival Studies. The Stata Journal, [S. l.], v. 7, n. 3, p. 351-375, 2007. Disponível em: <https://doi.org/10.1177/1536867X0700700304>. Acesso em: 05. out. 2025.

LAWLESS, J. F. Statistical Models and Methods for Lifetime Data. 2. ed. Hoboken: Wiley, 1982.

LEE, E. T.; WANG, J. W. Statistical Methods for Survival Data Analysis. 4. ed. Hoboken: John Wiley & Sons, 2013.

LEHMANN, E. L.; CASELLA, G. Theory of point estimation. 2. ed. New York: Springer, 1998.

LI, J.; MA, S. Survival Analysis in Medicine and Genetics. 1. ed. New York: Chapman and Hall/CRC, 2013.

MALLER, R. A.; ZHOU, X. Survival Analysis with Long-Term Survivors. New York: Wiley, 1996.

MINKA, T. P. Estimating a Gamma Distribution, 2002. Disponível em: <https://tminka.github.io/papers/minka-gamma.pdf>. Acesso em: 28 mai. 2025.

OLIVEIRA, R. P. Multivariate lifetime models to evaluate long-term survivors in medical studies. 2019. Tese (Doutorado em Saúde Pública) - Faculdade de Medicina de Ribeirão Preto, University of São Paulo, Ribeirão Preto, 2019. doi:10.11606/T.17.2019.tde-08012020-110425. Acesso em: 19 set. 2024.

MEEKER, W. Q.; ESCOBAR, L. A. Statistical Methods for Reliability Data. New York: Wiley, 1998.

PRICE, D. L.; MANATUNGA, A. K. Modelling survival data with a cured fraction using frailty models. *Statistics in medicine*, [S. l.], v. 20, n. 9-10, p. 1515-1527, 2001. Disponível em: [10.1002/sim.687](https://doi.org/10.1002/sim.687) Acesso em: 05 out. 2025.

PROCHASKA, J. J. Smoking and other tobacco use. *MSD Manual*, 2023. Disponível em: <https://www.msdmanuals.com/professional/special-subjects/tobacco-use/smoking-and-other-tobacco-use>. Acesso em: 13 out. 2024.

ROBINS, J. et al. Evaluation of a hospital-initiated tobacco dependence treatment service: uptake, smoking cessation, readmission and mortality. *BMC Medicine*, v. 22, n. 139, 2024. DOI: [10.1186/s12916-024-03353-8](https://doi.org/10.1186/s12916-024-03353-8). Disponível em: <https://bmcmmedicine.biomedcentral.com/articles/10.1186/s12916-024-03353-8>. Acesso em: 05 jun. 2025.

THE TOBACCO ATLAS. The tobacco epidemic continues. The tobacco atlas, 2024. Disponível em: <https://tobaccoatlas.org/>. Acesso em: 16 set. 2024.

THERNEAU, T. M.; GRAMBSCH, P. M. Modeling Survival Data: Extending the Cox Model. New York: Springer, 2000.

TSODIKOV, A. D. et al. Estimating cure rates from survival data: An alternative to two-component mixture models. *JSTOR*, v. 98, n. 464, p. 1063-1078, 2003. Disponível em: <https://www.jstor.org/stable/30045351>. Acesso em: 06 jun. 2025.

TUTZ, G.; Schmid, M. Modeling Discrete Time-to-Event Data. Switzerland: Springer, 2016.

VAHIDPOUR, M. Cure Rate Models. 2016. Tese (Mestrado em matemática aplicada) - École Polytechnique de Montréal, 2016. Disponível em: <https://publications.polymtl.ca/2454/>. Acesso em: 17 set. 2024.

VAUPEL, J. W. et al. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, v. 16, n. 3, p. 439-454, ago. 1979. DOI: doi.org/10.2307/2061224. Disponível em: <https://link.springer.com/article/10.2307/2061224>. Acesso em: 06 jun. 2025.

WHITE, C. A. et al. Impact of tobacco usage on readmission and complication rates following shoulder replacement surgery: A study of 164,527 patients. *Sage Journals*, [S. l.], v. 15, n. 1, p. 66-74, 2022. DOI: <https://doi.org/10.1177/17585732221102393>. Disponível em: <https://journals.sagepub.com/doi/abs/10.1177/17585732221102393?journalCode=sela>. Acesso em: 05 jun. 2025.

WIENKE, Andreas. Frailty Models in Survival Analysis. Boca Raton: Chapman & Hall/CRC, 2011.