

**UNIVERSIDADE DE SÃO PAULO**

Instituto de Ciências Matemáticas e de Computação

**Novas Modelagens de Risco Aditivo com Fragilidade para  
Análise de Dados de Sobrevivência**

**Felipe Rodrigues da Silva**

Tese de Doutorado do Programa Interinstitucional de Pós-Graduação em  
Estatística (PIPGEs)



SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: \_\_\_\_\_

**Felipe Rodrigues da Silva**

## Novas Modelagens de Risco Aditivo com Fragilidade para Análise de Dados de Sobrevida

Tese apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP e ao Departamento de Estatística – DEs-UFSCar, como parte dos requisitos para obtenção do título de Doutor em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística. *VERSÃO REVISADA*

Área de Concentração: Estatística

Orientadora: Profa. Dra. Vera Tomazella

**USP – São Carlos**  
**Março de 2024**

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi  
e Seção Técnica de Informática, ICMC/USP,  
com os dados inseridos pelo(a) autor(a)

R696n RODRIGUES, FELIPE  
Novas Modelagens de Risco Aditivo com  
Fragilidade para Análise de Dados de Sobrevivência  
/ FELIPE RODRIGUES; orientadora Vera Tomazella. --  
São Carlos, 2024.  
102 p.

Tese (Doutorado - Programa Interinstitucional de  
Pós-graduação em Estatística) -- Instituto de Ciências  
Matemáticas e de Computação, Universidade de São  
Paulo, 2024.

1. Análise de sobrevivência. 2. Eventos  
recorrentes. 3. Fração de cura. 4. Fragilidade. 5.  
Modelo de risco aditivo. I. Tomazella, Vera,  
orient. II. Título.

**Felipe Rodrigues da Silva**

**New Additive Risk Modeling with Frailty for Survival Data  
Analysis**

Thesis submitted to the Institute of Mathematics and Computer Science – ICMC-USP and to the Department of Statistics – DEs-UFSCar – in accordance with the requirements of the Statistics Interagency Graduate Program, for the degree of Doctor in Statistics. *FINAL VERSION*

Concentration Area: Statistics

Advisor: Profa. Dra. Vera Tomazella

**USP – São Carlos  
March 2024**



# AGRADECIMENTOS

---

---

Em primeiro lugar agradeço a Deus autor e sustentador de todas as coisas.

Também quero agradecer minha esposa por todo apoio durante o doutorado e nossa filha que nasceu durante essa jornada.

Agradeço em especial toda minha família, amigos e irmãos de Teresina por todas as orações.

Aos irmão de Araraquara, meu muito obrigado.

Agradecimentos à Professora Vera Tomazella<sup>1</sup> pela paciência e bons conselhos e todos aqueles que contribuíram para que a produção desse trabalho fosse possível.

Agradeço a todos os professores que ajudaram na minha formação e aos amigos do doutorado.

Agradecimentos especiais são direcionados à Universidade Federal do Piauí<sup>2</sup> por autorizar meu afastamento para o Doutorado.

---

<sup>1</sup> Orientadora

<sup>2</sup> <<http://www.ufpi.br/>>





# RESUMO

RODRIGUES, F. S. **Novas Modelagens de Risco Aditivo com Fragilidade para Análise de Dados de Sobrevida**. 2024. 102 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2024.

A análise de sobrevivência emerge como uma valiosa área da estatística para examinar o tempo até a ocorrência de eventos de interesse. Diversos modelos foram concebidos e aplicados em diversas áreas tais como: Medicina, Engenharia, Biomedicina e Ciências Sociais. O modelo proposto por [Cox \(1972\)](#) destaca-se como um dos mais reconhecidos e empregados na análise de dados de sobrevivência. No entanto, é importante notar que esse modelo parte da suposição de que os riscos são proporcionais, uma premissa que nem sempre é justificável. Um modelo alternativo aos modelos de riscos proporcionais de Cox é o modelo de risco aditivo que inicialmente foi proposto por [Aalen \(1980\)](#). No modelo aditivo o efeito das covariáveis é inserido aditivamente à função de risco de base. Em muitas situações existem fatores não observados no estudo que influenciam o tempo de sobrevivência, dessa forma para dados de sobrevivência univariados um efeito aleatório, denominado por [Aalen \(1978\)](#) e [Clayton \(1978\)](#) como termo de fragilidade, pode ser inserido de forma aditiva ou multiplicativa para estimar essa heterogeneidade não observada. Neste contexto, o termo de fragilidade inserido aditivamente na modelagem de risco para análise de dados univariados e dados de eventos recorrentes foi estudado e aplicado a dados reais. Além disso, uma proposta de estimador para as fragilidades individuais foi apresentada. Também um modelo de fração de cura com fragilidade aditiva foi proposto e aplicado a dados reais, onde esse modelo é aplicável a estudos em que existem indivíduos que são considerados imunes, curados ou não suscetíveis ao evento de interesse. Uma nova modelagem de risco aditivo alternativa também foi proposta baseada em [Gupta \(2016\)](#). A abordagem de máxima verossimilhança foi utilizada para estimação dos parâmetros dos modelos estudados, e estudos via simulação de Monte Carlo foram desenvolvidos para avaliar o comportamento dos estimadores de máxima verossimilhança.

**Palavras-chave:** Análise de sobrevivência, eventos recorrentes, fração de cura, fragilidade, modelo de risco aditivo.



# ABSTRACT

RODRIGUES, F. S. **New Additive Risk Modeling with Frailty for Survival Data Analysis** . 2024. 102 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2024.

Survival analysis emerges as a valuable statistical area for examining the time until the occurrence of events of interest. Several models were designed and applied in different areas such as: Medicine, Engineering, Biomedicine and Social Sciences. The model proposed by [Cox \(1972\)](#) stands out as one of the most recognized and used in the analysis of survival data. However, it is important to note that this model assumes that risks are proportional, an assumption that is not always reasonable. An alternative model to Cox proportional hazards models is the additive hazard model that was initially proposed by [Aalen \(1980\)](#). In the additive model, the effect of the covariates is inserted additively into the base hazard function. In many situations there are factors not observed in the study that influence survival time, so for univariate survival data a random effect, called [Aalen \(1978\)](#) and [Clayton \(1978\)](#) as a frailty term, can be entered additively or multiplicatively to estimate this unobserved heterogeneity. In this context, the additively inserted frailty term for risk modeling in univariate data analysis and recurring event data was studied and applied to real data. Furthermore, a proposal for an estimator for individual frailties was presented. Also a cure fraction model with additive frailty was proposed and applied to real data, where this model is applicable to studies in which there are individuals who are considered immune, cured or not susceptible to the event of interest. A new alternative additive risk modeling was also proposed based on [Gupta \(2016\)](#). The maximum likelihood estimation approach was used to estimate the parameters of the models studied, and studies via Monte Carlo simulation were developed to evaluate the behavior of maximum likelihood estimators.

**Keywords:** Additive hazard model, cure fraction, frailty, recurrent events, survival analysis.



# LISTA DE ILUSTRAÇÕES

---

---

Figura 1 – Curvas de sobrevivência da Equação 2.26 sem covariáveis com risco de base de uma distribuição Weibull( $\mu, \gamma$ ) para diferentes valores de $\mu, \gamma, a$ e $b$ versus tempo. . . . .	31
Figura 2 – Viés, EQM, REQM, EAM, e DP de $(\hat{a}, \hat{b}, \hat{\mu}, \hat{\beta})$ versus $n$ para o cenário 1. . . . .	34
Figura 3 – Viés, EQM, REQM, EAM, e DP de $(\hat{a}, \hat{b}, \hat{\mu}, \hat{\beta})$ versus $n$ para o cenário 2. . . . .	35
Figura 4 – Viés, EQM, REQM, EAM, e DP de $(\hat{a}, \hat{b}, \hat{\mu}, \hat{\beta})$ versus $n$ para o cenário 3. . . . .	35
Figura 5 – Viés, EQM, REQM, EAM, e DP de $(\hat{a}, \hat{b}, \hat{\mu}, \hat{\beta})$ versus $n$ para o cenário 4. . . . .	36
Figura 6 – Curva de sobrevivência estimada pelo estimador Kaplan-Meier para dados de câncer de pâncreas. . . . .	37
Figura 7 – Curva de sobrevivência estimada pelo MFA-Gama-Exponencial para dados de câncer de pâncreas. . . . .	38
Figura 8 – Fragilidades individuais dos 355 pacientes do estudo. . . . .	38
Figura 9 – Curvas de sobrevivência Kaplan-Meier estratificada por grupo . . . . .	39
Figura 10 – Curvas de sobrevivências estimadas pelo MFA-Gama-Exponencial com as curvas estimadas por Kaplan-Meier. . . . .	40
Figura 11 – Curva de sobrevivência estimada por Kaplan-Meier para os dados de câncer de pulmão. . . . .	42
Figura 12 – Box-plot das fragilidades individuais para os indivíduos que falharam e não falharam durante o tempo de acompanhamento. . . . .	43
Figura 13 – Gráfico dos tempos recorrentes de detecção de tumores cancerígenos. . . . .	54
Figura 14 – Função média amostral acumulada para os grupos de ratos tratados e não tratados. . . . .	54
Figura 15 – Curvas de sobrevivência estimadas para os grupos de ratos tratados e não tratados. . . . .	55
Figura 16 – Comparação das curvas de sobrevivências estimadas pelo modelo aditivo com fragilidade Gama( $1/\alpha, 1/\alpha$ ) e risco de base da distribuição Weibull( $\mu, \gamma$ ) com as curvas estimadas pelo estimador não paramétrico para dados recorrentes. . . . .	56
Figura 17 – Comparação das curvas de sobrevivências estimadas pelo modelo aditivo com fragilidade Gama( $1/\alpha, 1/\alpha$ ) e risco de base da distribuição Exponencial( $\mu$ ) com as curvas estimadas pelo estimador não paramétrico para dados recorrentes. . . . .	57
Figura 18 – Comparação das curvas de sobrevivências estimadas pelo modelo aditivo com fragilidade Gama( $a, b$ ) e risco de base da distribuição Exponencial( $\mu$ ) com as curvas estimadas pelo estimador não paramétrico para dados recorrentes. . . . .	61

Figura 19 – Curva de sobrevivência do modelo (4.2). . . . .	66
Figura 20 – Curvas de sobrevivência do modelo (4.7) com risco de base de uma distribuição Weibull( $\mu, \gamma$ ) para diferentes valores dos parâmetros $a, b, \mu$ e $\gamma$ versus tempo. . . . .	68
Figura 21 – Viés, EQM, REQM, EAM e DP de $(\hat{a}, \hat{b}, \hat{\mu})$ versus $n$ para o SMMAF com $a = 0.01, b = 0.1, \mu = 0.03$ . As linhas pretas e vermelhas representam o cenário onde $p_0 = 0.2$ e $0.5$ , respectivamente . . . . .	70
Figura 22 – Viés, EQM, REQM, EAM e DP de $(\hat{a}, \hat{b}, \hat{\mu})$ versus $n$ para o SMMAF com $a = 0.001, b = 0.04, \mu = 0.03$ . As linhas pretas e vermelhas representam o cenário onde $p_0 = 0.2$ e $0.5$ , respectivamente . . . . .	71
Figura 23 – Viés, EQM, REQM, EAM e DP de $(\hat{a}, \hat{b}, \hat{\mu})$ versus $n$ para o SMMAF com $a = 0.01, b = 0.5, \mu = 0.03$ . As linhas pretas e vermelhas representam o cenário onde $p_0 = 0.2$ e $0.5$ , respectivamente . . . . .	71
Figura 24 – Curvas de sobrevivência estimadas por Kaplan-Meier para os dados de melanoma . . . . .	72
Figura 25 – Curvas de sobrevivência estimadas pelo modelo (4.7) com funções de risco de base das distribuições Exponencial e Weibull, respectivamente . . . . .	73
Figura 26 – Curvas de sobrevivência estimadas por Kaplan-Meier para cada categoria N . . . . .	75
Figura 27 – Curvas de sobrevivência estimadas pelo modelo (4.14) (linhas tracejadas) com funções de risco de base das distribuições Exponencial e Weibull, respectivamente . . . . .	77
Figura 28 – Curvas de sobrevivência estimadas por Kaplan-Meier para cada grupo de NLR . . . . .	77
Figura 29 – Curvas de sobrevivência estimadas pelo modelo (4.10) (linhas tracejadas) com funções de risco de base das distribuições Exponencial e Weibull, respectivamente . . . . .	79
Figura 30 – Viés, EQM, REQM, EAM, e DP de $(\hat{a}, \hat{b}, \hat{\mu}, \hat{\beta})$ versus $n$ para o cenário 1 da Tabela 22. . . . .	86
Figura 31 – Viés, EQM, REQM, EAM, e DP de $(\hat{a}, \hat{b}, \hat{\mu}, \hat{\beta})$ versus $n$ para o cenário 2 da Tabela 22. . . . .	86
Figura 32 – Viés, EQM, REQM, EAM, e DP de $(\hat{a}, \hat{b}, \hat{\mu}, \hat{\beta})$ versus $n$ para o cenário 3 da Tabela 22. . . . .	87
Figura 33 – Viés, EQM, REQM, EAM, e DP de $(\hat{a}, \hat{b}, \hat{\mu}, \hat{\beta})$ versus $n$ para o cenário 4 da Tabela 22. . . . .	87
Figura 34 – Viés, EQM, REQM, EAM, e DP de $(\hat{a}, \hat{b}, \hat{\mu}, \hat{\beta})$ versus $n$ para o cenário 5 da Tabela 22. . . . .	88
Figura 35 – Curvas de sobrevivência estimadas pelo modelo (5.7) (linhas contínuas) com função de risco de base da distribuição Weibull para os dados de leucemia. . . . .	89

# LISTA DE TABELAS

---

---

Tabela 1	– Valores dos parâmetros do modelo (2.30) sob diferentes cenários. . . . .	34
Tabela 2	– Estimativas de máxima verossimilhança (EMV), erros padrão (EP), intervalos de confiança (IC) de 95%, AIC e BIC para os modelo aditivo com fragilidades Gama-Exponencial e Gama-Weibull para os dados de câncer de pâncreas. . . . .	37
Tabela 3	– Estatísticas descritivas dos dados de leucemia. . . . .	39
Tabela 4	– Estimativas de máxima verossimilhança, erros padrão, intervalos de confiança de 95%, AIC e BIC para os modelos aditivo com fragilidade Gama e riscos de base da distribuição Exponencial( $\mu$ ) e da distribuição Weibull( $\mu, \gamma$ ) para os dados de leucemia. . . . .	40
Tabela 5	– Fragilidade individual estimada dos 33 pacientes com leucemia aguda divididos por grupo com base na Tabela 4. . . . .	41
Tabela 6	– Estimativas de máxima verossimilhança, erros padrão, intervalos de confiança de 95%, AIC e BIC para os modelos aditivo com fragilidade Gama e riscos de base da distribuição Exponencial( $\mu$ ) e da distribuição Weibull( $\mu, \gamma$ ) para os dados de câncer de pulmão. . . . .	42
Tabela 7	– EQM, EAM e Viés dos estimadores do modelo (3.3) sem covariável, com fragilidade Gama( $1/\alpha, 1/\alpha$ ) e risco de base de uma distribuição Exponencial( $\mu$ ) para diferentes valores de $\alpha, \mu$ e $n$ , com percentual de censura (PC) variando de 0% a 20%. . . . .	52
Tabela 8	– Valores dos parâmetros do modelo (3.3) sob diferentes cenários. . . . .	52
Tabela 9	– EQM e Viés dos estimadores do modelo (3.3) com fragilidade Gama( $1/\alpha, 1/\alpha$ ) e risco de base de uma distribuição Exponencial( $\mu$ ) para $\mu = 0.03$ e diferentes valores de $\alpha, \beta$ e $n$ , com percentual de censura (PC) variando de 0% a 20%. . . . .	53
Tabela 10	– Estimativas de máxima verossimilhança, erros padrão, limite inferior (LI) e limite superior (LS) dos intervalos de confiança de 95% para os parâmetros do modelo de fragilidade aditivo Gama-Weibull. . . . .	55
Tabela 11	– Estimativas de máxima verossimilhança, erros padrão, limite inferior (LI) e limite superior (LS) dos intervalos de confiança de 95% para os parâmetros do modelo de fragilidade aditivo Gama-Exponencial. . . . .	56
Tabela 12	– Teste da Razão de verossimilhança para comparar os modelos. . . . .	57
Tabela 13	– Fragilidades individuais dos 48 ratos e seus respectivos numeros de recorrências divididos por grupo. . . . .	58

Tabela 14 – Estimativas de máxima verossimilhança, erros padrão, intervalos de confiança de 95%, AIC e BIC para os modelos aditivo com fragilidade Gama e riscos de base da distribuição Exponencial( $\mu$ ) e da distribuição Weibull( $\mu, \gamma$ ) para os dados de câncer em ratos. . . . .	61
Tabela 15 – Fragilidades individuais dos 48 ratos e seus respectivos numeros de recorrências divididos por grupo. . . . .	62
Tabela 16 – Viés, EQM, REQM, EAM e DP de $\hat{p}_0$ no modelo (4.7) com risco de base de uma distribuição exponencial com $\mu = 0.03$ para diferentes valores de $a, b, p_0$ e $n$ , com percentual de censura (PC) variando de 30% a 60%. . . . .	70
Tabela 17 – Estimativas de máxima verossimilhança, erros padrão, intervalos de confiança de 95%, AIC e BIC para os parâmetros do modelo (4.7) com riscos de base da distribuição Exponencial( $\mu$ ) e da distribuição Weibull( $\mu, \gamma$ ). . . . .	72
Tabela 18 – Estimativas de máxima verossimilhança, erros padrão, intervalos de confiança de 95%, AIC e BIC para os parâmetros do modelo (4.14) com riscos de base da distribuição Exponencial( $\mu$ ) e da distribuição Weibull( $\mu, \gamma$ ). . . . .	76
Tabela 19 – Estimativa de máxima verossimilhança através da função de ligação logística para a fração de cura de cada categoria da covariável N no modelo (4.14) com risco de base das distribuições Exponencial e Weibull. . . . .	76
Tabela 20 – Estimativas de máxima verossimilhança, erros padrão, intervalos de confiança de 95%, AIC e BIC para os parâmetros do modelo (4.10) com riscos de base da distribuição Exponencial( $\mu$ ) e da distribuição Weibull( $\mu, \gamma$ ) para os dados de ISC. . . . .	78
Tabela 21 – Estimativa de máxima verossimilhança (EMV) através da função de ligação logística para a fração de cura de cada categoria da covariável NLR no modelo (4.14) com risco de base das distribuições Exponencial e Weibull. . . . .	79
Tabela 22 – Valores dos parâmetros do modelo (5.2) sob diferentes cenários . . . . .	85
Tabela 23 – Estimativas de máxima verossimilhança, erros padrão, intervalos de confiança de 95%, AIC e BIC para os modelos aditivo com fragilidade Gama (5.2) e riscos de base da distribuição Exponencial( $\mu$ ) e da distribuição Weibull( $\mu, \gamma$ ). . . . .	89
Tabela 24 – Fragilidade individual estimada dos 33 pacientes com leucemia aguda divididos por grupo com base na Tabela 23. . . . .	90
Tabela 25 – Estimativas de máxima verossimilhança (EMV), erros padrão (EP), intervalos de confiança (IC) de 95%, AIC e BIC para os modelos aditivo com fragilidade Gama e riscos de base da distribuição Exponencial( $\mu$ ) e da distribuição Weibull( $\mu, \gamma$ ) . . . . .	91
Tabela 26 – Dados dos tempos de vida dos 33 pacientes com leucemia aguda divididos por grupo. . . . .	99
Tabela 27 – Dados dos tempos recorrentes de detecção de tumores em 48 ratos. . . . .	100



# SUMÁRIO

---

---

1	INTRODUÇÃO . . . . .	17
1.1	Objetivo da pesquisa . . . . .	19
1.2	Contribuição do trabalho . . . . .	20
1.3	Organização do trabalho . . . . .	20
2	MODELAGEM DE RISCO ADITIVO COM FRAGILIDADE PARA DADOS UNIVARIADOS . . . . .	22
2.1	Modelo Aditivo de Aalen . . . . .	22
2.1.1	<i>Abordagem semi-paramétrica e paramétrica</i> . . . . .	24
2.2	Modelo de fragilidade aditivo (MFA): Uma abordagem paramétrica	25
2.2.1	<i>Modelo de fragilidade aditivo com covariáveis</i> . . . . .	28
2.2.2	<i>Função de Verossimilhança</i> . . . . .	28
2.2.3	<i>Modelo Aditivo com fragilidade Gama</i> . . . . .	29
2.3	Estudo de Simulação . . . . .	33
2.4	Aplicação para dados Reais . . . . .	36
2.4.1	<i>Dados de câncer de pâncreas</i> . . . . .	37
2.4.2	<i>Dados de Leucemia</i> . . . . .	39
2.4.3	<i>Dados de câncer de pulmão</i> . . . . .	41
2.5	Conclusão . . . . .	43
3	MODELAGEM DE RISCO ADITIVO COM FRAGILIDADE PARA DADOS DE EVENTOS RECORRENTES . . . . .	45
3.1	Modelo aditivo com fragilidade para dados de eventos recorrentes .	45
3.1.1	<i>Modelo de fragilidade aditivo com covariáveis</i> . . . . .	46
3.1.2	<i>Modelo Aditivo com fragilidade Gama</i> . . . . .	47
3.1.2.1	<i>Fragilidades individuais</i> . . . . .	49
3.1.3	<i>Estimador não paramétrico da função de sobrevivência para dados de eventos recorrentes.</i> . . . . .	49
3.2	Estudo de simulação . . . . .	51
3.3	Aplicação para dados reais . . . . .	53
3.3.1	<i>Aplicação para dados câncer</i> . . . . .	53
3.3.2	<i>Outra parametrização para a fragilidade Gama</i> . . . . .	58
3.4	Conclusão . . . . .	62

4	MODELAGEM DE RISCO ADITIVO COM FRAGILIDADE E FRACÇÃO DE CURA . . . . .	64
4.1	Modelagem . . . . .	65
4.1.1	<i>Modelo de mistura padrão</i> . . . . .	65
4.1.2	<i>Modelo de mistura padrão com fragilidade aditiva (MMPFA)</i> . . . . .	66
4.1.3	<i>Inferência</i> . . . . .	67
4.1.4	<i>Modelo de mistura padrão com fragilidade aditiva Gama</i> . . . . .	67
4.1.5	<i>Simulação</i> . . . . .	69
4.1.6	<i>Aplicação</i> . . . . .	71
4.1.7	<i>Modelo de mistura padrão com fragilidade aditiva na presença de covariáveis</i> . . . . .	73
4.2	Aplicação . . . . .	75
4.2.1	<i>Dados de câncer de mama</i> . . . . .	75
4.2.2	<i>Dados de infecção sítio cirúrgico</i> . . . . .	77
4.2.3	<i>Conclusão</i> . . . . .	79
5	MODELAGEM DE RISCO ADITIVO ALTERNATIVA COM FRAGILIDADE PARA DADOS UNIVARIADOS . . . . .	80
5.1	Modelo de risco aditivo alternativo com fragilidade para dados univariados . . . . .	81
5.2	Estudo de Simulação . . . . .	85
5.3	Aplicação . . . . .	88
5.3.1	<i>Dados Leucemia</i> . . . . .	88
5.3.2	<i>Dados de câncer de pulmão</i> . . . . .	90
5.3.3	<i>Conclusão</i> . . . . .	92
6	CONCLUSÕES E TRABALHOS FUTUROS . . . . .	93
	REFERÊNCIAS . . . . .	95
	APÊNDICE A            CONJUNTOS DE DADOS . . . . .	99
A.1	Dados dos tempos de vida dos 33 pacientes com leucemia aguda divididos por grupo. . . . .	99
A.2	Dados dos tempos recorrentes de detecção de tumores em 48 ratos	100
	APÊNDICE B            DESENVOLVIMENTOS ALGÉBRICOS . . . . .	101
B.1	Distribuição da fragilidade $V$ dado $T=t$ na parametrização $Gama(a,b)$	102

---

## INTRODUÇÃO

---

A análise de sobrevivência é uma área da estatística útil para estudar o tempo até a ocorrência de um ou mais eventos de interesse. Esse tipo de análise teve início na área médica e ganhou mais notoriedade a partir do final do século XX. Desde então muitos modelos têm sido propostos nas mais diversas áreas, tais como: Medicina, engenharias, biomedicina, ciências sociais, dentre outras. Uma característica particular dos dados de sobrevivência é que eles podem apresentar uma observação incompleta do tempo até o evento de interesse, denominada de censura.

Um modelo bastante flexível e extensivamente usado em análise de sobrevivência por incorporar o efeito de covariáveis fixas é o modelo de riscos proporcionais de Cox (COX, 1972) que pode ser generalizado para incorporar o efeito de covariáveis dependentes do tempo (COLOSIMO; NOGUEIRA, 1996). Entretanto esse modelo que incorpora o efeito das covariáveis de forma multiplicativa na função de risco de base supõe que os riscos são proporcionais, suposição que muitas vezes não é razoável. Para superar essa limitação alguns modelos foram propostos na literatura, dentre eles as chamadas extensões do modelo de Cox, como o modelo de Cox estratificado, onde assume-se que as taxas de falha são proporcionais somente em cada estrato (COLOSIMO, 1997). Aalen (1989) citou algumas limitações do modelo de Cox, dentre elas podemos destacar que as suposições do modelo de Cox podem não ser verificadas na prática, e outra seria que mesmo quando as propriedades de proporcionalidade são satisfeitas não há garantia da adequação do modelo de Cox. Apesar disso o modelo de Cox é largamente utilizado na literatura, algumas vezes sem que suas propriedades sejam checadas.

Um modelo alternativo aos modelos de risco multiplicativo é o modelo de risco aditivo que inicialmente foi proposto por Aalen (1980). No modelo aditivo o efeito das covariáveis é introduzida na função de risco de forma aditiva. Lin e Ying (1994), Lin e Ying (1997) e Martinussen, Scheike e Zucker (2011) demonstraram que em algumas situações, descritas a seguir, é mais apropriado modelar dados de sobrevivência através do modelo de risco aditivo. De

acordo com [Lin e Ying \(1994\)](#) o modelo de risco aditivo é uma alternativa prática ao modelo de riscos proporcionais e a escolha de qual modelagem utilizar é uma questão empírica. Ainda segundo [Lin e Ying \(1994\)](#) as duas modelagens focam em aspectos bastante diferentes da associação entre fatores de risco e o desfecho do estudo, por exemplo, enquanto o modelo aditivo se baseia na diferença dos riscos o modelo multiplicativo se baseia na razão dos riscos, portanto segundo esses autores a escolha de qual modelagem utilizar é uma questão prática, onde o mais parcimonioso é preferível. [Lin e Ying \(1997\)](#) destacaram que do ponto de vista da saúde pública é mais importante avaliar a diferença de riscos do que a razão de risco na descrição da associação entre o fator de risco e a ocorrência da doença, como argumentaram [Breslow e Day \(1980\)](#). [Martinussen, Scheike e Zucker \(2011\)](#) estudaram um modelo de risco aditivo com covariáveis que mudam no tempo, situação em que os modelos de risco aditivo são mais apropriados que o modelo de Cox.

Para dados de eventos recorrentes, isto é eventos de interesse que podem ocorrer mais de uma vez para o mesmo indivíduo sem que este saia do estudo, é razoável assumir que os tempos são dependentes e que os riscos não são proporcionais, dessa forma uma alternativa para modelar esse tipo de dados são os chamados modelos de risco aditivo. Esses dados também são chamados de multivariados e para medir a dependência entre os tempos de recorrência é utilizado um efeito aleatório no modelo, denominado por [Aalen \(1978\)](#) e [Clayton \(1978\)](#) como termo de fragilidade que pode ser inserido de forma aditiva ou multiplicativa ([VAUPEL; MANTON; STALLARD, 1979](#)). Uma abordagem paramétrica dos modelos aditivos com fragilidade foi apresentada por [Tomazella \(2003\)](#). Essa abordagem pode ser vista como uma extensão do modelo de Aalen ([AALEN, 1980](#)).

Em geral, os termos de fragilidades são considerados independentes, sendo algumas vezes identicamente distribuídos e outras vezes pertencentes somente à mesma família de distribuições de fragilidade. Para dados de sobrevivência univariados o termo de fragilidade pode ser usado para descrever a influência de covariáveis não observadas. Como esse efeito aleatório é não negativo algumas distribuições como Gama, log-normal, estáveis positivas foram propostas, sendo a distribuição Gama a mais utilizada para modelar a fragilidade ([NIELSEN \*et al.\*, 1992](#)). A suposição padrão para a fragilidade é usar uma distribuição Gama pois essa distribuição é mais tratável algebricamente, entretanto [Hougaard \(1995\)](#) estudou outras distribuições para a variável de fragilidade, tal como a Gaussiana Inversa.

Apesar dos modelos aditivos serem uma abordagem alternativa interessante aos modelos multiplicativos, eles têm sido pouco estudados em comparação com os modelos multiplicativos. Dentre os estudos para essa modelagem nos últimos anos destacamos [Martinussen e Scheike \(2002\)](#) que estudaram métodos não paramétricos para analisar dados de *status* atual através do modelo de risco aditivo, [Tomazella, Louzada-Neto e Silva \(2006\)](#) que analisaram dados de eventos recorrentes considerando Processo de Poisson homogêneos com fragilidade aditiva Gama sob uma abordagem bayesiana, [Gupta e Gupta \(2009\)](#) que estudaram um modelo geral de

fragilidade aditiva e [Gupta \(2016\)](#) que estudou propriedades de um modelo de fragilidade aditiva geral.

Algumas modelagens aditivas com fragilidade foram propostas na literatura, dentre elas [Rocha \(1995\)](#) propôs um modelo de fragilidade linear assumindo que a variável de fragilidade é incluída aditivamente à função de risco de base  $\lambda_0(t)$ . [Silva \(2001\)](#) estudou esse modelo sob uma abordagem bayesiana e [Tomazella \(2003\)](#) estudou esse modelo com covariáveis incorporadas aditivamente à função de risco de base. [Martinussen, Scheike e Zucker \(2011\)](#) estudaram o modelo de riscos aditivos de Aalen com estrutura de fragilidade multiplicativa sobreposta. Um modelo de fragilidade aditiva geral foi proposto por [Gupta \(2016\)](#) e mais recentemente [Wang et al. \(2018\)](#) estudaram outro modelo aditivo para dados de eventos recorrentes.

Em análise de sobrevivência quando uma parcela da população em estudo nunca apresentará o evento de interesse, mesmo se acompanhada por um tempo suficientemente grande, geralmente são utilizados os conhecidos modelos de longa duração, também chamados de modelos de sobrevivência com fração de cura. Esses modelos permitem estimar a proporção de curados em uma determinada população diagnosticada com uma doença, por exemplo. Um exemplo clássico é quando um paciente é "curado" de câncer e não tem recidiva, nesse caso esse paciente está imune (não suscetível) ao evento de interesse e faz parte da fração de curados na população em estudo. Neste contexto, o modelo de mistura padrão proposto por [Boag \(1949\)](#) e [Berkson e Gage \(1952\)](#) é o mais conhecido.

Muitos estudos a respeito de modelos que envolvem misturas de distribuições e fração de cura vêm sendo desenvolvidos na literatura. Por exemplo, [Farewell \(1977\)](#) abordou o modelo de mistura Weibull e investigou como o fator de risco afeta o tempo de desenvolvimento de uma doença, [Longini e Halloran \(1996\)](#) propuseram o modelo de mistura padrão com fragilidade e [Price e Manatunaga \(2001\)](#) estudaram o modelo de mistura padrão com diferentes distribuições para a fragilidade. [Gonzales, Tomazella e Taconelli \(2013\)](#) apresentaram uma abordagem paramétrica para o modelo de mistura padrão com fragilidade Gama incluída de forma multiplicativa na função de risco de base na presença de covariáveis.

## 1.1 Objetivo da pesquisa

O objetivo desse trabalho pode ser resumido nos seguintes tópicos:

- Investigar os principais modelos aditivos com fragilidade já existentes;
- Avaliar o comportamento dos estimadores de máxima verossimilhança para os modelos estudados;
- Propor estimadores para as fragilidades individuais e aplicar a dados reais;
- Propor um modelo de longa duração com fragilidade aditiva;

- Propor uma modelagem de fragilidade aditiva alternativa ao modelo proposto por [Rocha \(1995\)](#).

## 1.2 Contribuição do trabalho

As principais contribuições desse trabalho podem ser resumidas nas seguintes:

- Proposta de estimadores para a fragilidade individual a partir do modelo originalmente proposto por [Rocha \(1995\)](#);
- Proposta de estimador pontual para fragilidade individual a partir do modelo para dados de eventos recorrentes estudado por [Tomazella \(2003\)](#);
- Proposta de uma modelagem de mistura padrão com fragilidade aditiva, e aplicação do modelo proposto à dados reais;
- Desenvolvimento de um estudo via simulação de Monte Carlo para avaliar o comportamento dos estimadores de máxima verossimilhança para os parâmetros do modelo proposto no [Capítulo 4](#);
- Proposta de um modelo de fragilidade aditivo alternativo para dados univariados, e aplicação à dados reais;
- Estudo via simulação de Monte Carlo para avaliar o comportamento dos estimadores de máxima verossimilhança para os parâmetros do modelo alternativo apresentado no [Capítulo 5](#);
- Proposta de estimadores pontuais para as fragilidades individuais do modelo apresentado no [Capítulo 5](#).

## 1.3 Organização do trabalho

No [Capítulo 2](#) apresentamos a modelagem aditiva paramétrica para dados univariados, descrevemos o modelo de fragilidade aditiva originalmente proposto por [Rocha \(1995\)](#). Fizemos aplicações desse modelo à dados reais e apresentamos um estudo via simulação de Monte Carlo para avaliar os estimadores dos parâmetros do modelo.

No [Capítulo 3](#), discutimos brevemente a modelagem de fragilidade aditiva para dados de eventos recorrentes. Descrevemos o modelo estudado por [Tomazella \(2003\)](#) e obtivemos alguns resultados teóricos para esse modelo, assumindo uma distribuição Gama para a variável de fragilidade. Além disso, realizamos uma aplicação desse modelo a dados reais e apresentamos um estudo numérico via simulação de Monte Carlo para avaliar os estimadores de máxima verossimilhança

No [Capítulo 4](#) apresentamos uma breve discussão sobre o modelo de mistura padrão, propomos o modelo de mistura padrão com fragilidade aditiva e fizemos uma aplicação desse modelo à dados reais assumindo uma distribuição Gama para a variável de fragilidade, além de um estudo simulado para avaliar o comportamento dos estimadores de máxima verossimilhança.

No [Capítulo 5](#) apresentamos um modelo de fragilidade aditivo alternativo, fizemos um estudo do comportamento dos estimadores de máxima verossimilhança para esse modelo, aplicamos essa modelagem em dados reais e comparamos com o modelo estudado no [Capítulo 2](#). Todas as análises numéricas e gráficas apresentadas nesse trabalho foram feitas utilizando o *software* [R-Development-Core-Team \(2023\)](#) na versão 4.1.2.

# MODELAGEM DE RISCO ADITIVO COM FRAGILIDADE PARA DADOS UNIVARIADOS

---

---

Aalen (1980) propôs um modelo de risco aditivo que é uma modelagem alternativa ao modelo de Cox. Esse modelo de Aalen tem vantagens práticas, especialmente quando as covariáveis têm efeitos variando no tempo. Um modelo aditivo pode ser utilizado quando é razoável supor que os efeitos das covariáveis contribuem aditivamente para o risco condicional.

## 2.1 Modelo Aditivo de Aalen

Considerando situações onde não impõe-se que as funções de risco sejam proporcionais, temos a segunda classe de modelos de risco em análise de sobrevivência, conhecida por modelos aditivos. Assim, o modelo proposto por Aalen (1980) quando o efeito das covariáveis é expresso aditivamente na função de risco, é definido pela seguinte função de risco na unidade  $i$

$$\lambda(t; \mathbf{x}_i) = \lambda_0(t) + G(\mathbf{x}_i, \boldsymbol{\beta}), \quad (2.1)$$

em que  $\lambda_0(\cdot)$  é a função de risco de base,  $G(\cdot)$  é uma função positiva e  $\boldsymbol{\beta}$  é um vetor  $p$  dimensional de parâmetro desconhecidos de efeitos associados a  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$ .

A forma mais geral da função de risco para o modelo aditivo (2.1) é dada da seguinte forma:

$$\lambda(t|\mathbf{x}_i(t)) = \beta_0(t) + \sum_{j=1}^p \beta_j(t)x_{ij}(t),$$

em que  $x_{ij}(t)$  são as covariáveis que podem depender do tempo, com  $i = 1, \dots, n$  e  $j = 1, \dots, p$ , onde  $n$  é o número de indivíduos no estudo e  $p$  é o número de covariáveis no modelo. Os  $\beta_j(t)$  são funções do tempo desconhecidas que medem a influência das respectivas covariáveis e  $\beta_0(t)$  pode ser visto como uma função de risco base.



Algumas características do modelo aditivo de Aalen são:

- O efeito das covariáveis é aditivo na função de risco;
- A modelagem proposta por Aalen possibilita monitorar covariáveis dependentes do tempo;
- Essa modelagem ainda permite que as covariáveis variem com o tempo;
- Nenhuma forma paramétrica particular é assumida para os  $\beta_j(t)$ ;
- Os  $\beta_j(t)$  também são chamados de coeficientes de risco ou funções de regressão;
- Esse modelo proposto por Aalen é totalmente não-paramétrico, pois ajusta funções e não parâmetros, por isso é muito flexível;
- O modelo aditivo de Aalen tem a desvantagem de possibilitar a estimação de valores negativos para a função de risco.

A estimação do modelo aditivo de Aalen não é muito simples, na verdade estimar os  $\beta_j(t)$  diretamente é uma tarefa difícil, para mais detalhes ver (COLOSIMO; GIOLO, 2006). Uma alternativa utilizada na prática é estimar as funções de risco acumulada, dada por:

$$B_j(t) = \int_0^t \beta_j(u) du,$$

que são elementos do vetor coluna  $B_j(t)$ . O estimador proposto por Aalen para estimar  $B(t)$ , é dado por:

$$\hat{B}(t) = \sum_{t_i < t} Z(t_i) I(t_i),$$

em que,  $t_1 < t_2 < \dots < t_k$  são os tempos de falha ordenados,  $\hat{B}(t)$  é o vetor dos  $\hat{B}_j(t)$ ,  $I(t_i)$  é um vetor com o  $i$ -ésimo elemento igual a 1 se o evento ocorre para o indivíduo  $i$  no tempo  $t$  e é zero, caso contrário e  $Z(t_i)$  é a inversa generalizada de  $X(t_i)$  que é uma matriz  $n \times (p + 1)$  composta por vetores do tipo  $\mathbf{x}_i(t) = (1, x_{i1}(t), \dots, x_{ip}(t))'$  na  $i$ -ésima linha se a falha não ocorreu para o indivíduo  $i$  e ele ainda estiver no estudo, e contém apenas zeros na linha  $i$  se o  $i$ -ésimo indivíduo não estiver em risco no tempo  $t$ .

Os  $\hat{B}_j(t)$  podem ser vistos como funções empíricas que descrevem a influência da  $j$ -ésima covariável, e a inclinação da função de regressão acumulada  $\hat{B}_j(t)$  indica como a covariável influencia a função de risco ao longo do tempo. Por exemplo, se a inclinação for positiva em relação ao eixo do tempo, então naquele período de tempo o aumento dos valores da covariável influenciou no aumento do risco, se a inclinação for negativa em relação ao eixo do tempo, então naquele período de tempo o aumento dos valores da covariável influenciou na diminuição do risco, já quando a função de regressão acumulada se comporta de maneira contante (sem inclinações), então a covariável relacionada tem efeito constante no risco.

Para todas as covariáveis fixadas no tempo inicial do estudo, temos que o estimador da função de risco acumulada para um indivíduo é dado por:

$$\hat{\Lambda}(t|\mathbf{x}) = \hat{B}_0(t) + \sum_{j=1}^p \hat{B}_j(t)x_j,$$

sendo os valores de  $t$  tais que a matriz  $X(t)$  é não-singular. A função de sobrevivência estimada é dada pela seguinte relação:

$$\hat{S}(t|\mathbf{x}) = \exp\{-\hat{\Lambda}(t|\mathbf{x})\}.$$

### 2.1.1 Abordagem semi-paramétrica e paramétrica

O modelo de risco aditivo de Aalen é totalmente não-paramétrico, já o modelo de risco aditivo proposto por Lin e Ying (1994) tem uma abordagem semi-paramétrica, assumindo para função de risco de base uma função não-negativa e considerando as funções de regressão  $\beta_j(t)$  constantes ao longo do tempo, dessa forma temos o seguinte modelo:

$$\lambda(t|\mathbf{x}) = \lambda_0(t) + \mathbf{x}'\beta,$$

sendo  $\lambda_0(t)$  a função de risco de base, e  $\beta$  é um vetor  $p$ -dimensional de parâmetros desconhecidos de efeitos associados a  $\mathbf{x}$ . Para esse modelo a interpretação dos parâmetros é mais direta do que no modelo original de Aalen, entretanto ainda mantém a desvantagem de admitir a estimação da função de risco com valores negativos.

No modelo aditivo paramétrico atribui-se distribuição de probabilidade para o risco de base, por exemplo: Exponencial, Weibull, dentre outras. Nesse trabalho consideramos os critérios AIC (AKAIKE, 1974) e BIC (SCHWARZ, 1978) para selecionar o modelo mais parcimonioso dentre os modelos utilizados. Diferente do modelo de Cox, o modelo aditivo incorpora as covariáveis aditivamente ao risco de base, ou seja, o modelo de risco fica da seguinte forma

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

em que  $\lambda_0(\cdot)$  é a função de risco base,  $g(\cdot)$  é uma função conhecida,  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$  é um vetor de covariáveis do indivíduo  $i$  ( $i = 1, \dots, n$ ) e  $\beta = (\beta_1, \dots, \beta_p)$  é um vetor de parâmetros desconhecidos associado às covariáveis. Assumindo que a função  $g(\cdot) = \mathbf{x}'_i\beta$  é a função identidade temos que

$$\lambda(t|\mathbf{x}_i) = \lambda_0(t) + \mathbf{x}'_i\beta. \quad (2.2)$$

Note que essa modelagem assumindo que a função  $g(\cdot)$  é a função identidade pode apresentar valor negativo para a função de risco, o que é uma desvantagem em relação a abordagem multiplicativa que assume uma função exponencial para  $g(\cdot)$  multiplicando o risco de base. Por se tratar de uma abordagem paramétrica devemos supor uma distribuição de probabilidade para o risco de base, como por exemplo a distribuição Exponencial( $\mu$ ) que tem risco constante ( $\lambda_0(t) = \mu$ ). Neste caso a equação (2.2) pode ser reescrita como  $\lambda(t|\mathbf{x}_i) = \mu + \mathbf{x}'_i\beta$ .

## 2.2 Modelo de fragilidade aditivo (MFA): Uma abordagem paramétrica

O termo de fragilidade é um efeito aleatório não-negativo que capta possíveis variações não explicadas pelas covariáveis, denotado por  $V$ . Nos modelos de fragilidade aditivos esse efeitos aleatórios são introduzidos de forma aditiva, em que esse termo de fragilidade mede a heterogeneidade não observada entre os indivíduos e a dependência entre os indivíduos que compartilham os mesmos fatores de risco. A variável aleatória  $V$  pode seguir diversas distribuições de probabilidade, sendo as distribuições Gama, Gaussiana inversa e a família de distribuições estáveis as mais utilizadas. Nessa modelagem a variância da fragilidade é vista como uma medida de heterogeneidade em toda a população, em que valores pequenos da variância levam a riscos parecidos entre os indivíduos e variância grande significa alta variação nos riscos dos indivíduos.

O modelo de fragilidade aditivo geral proposto por Rocha (1995) é representado por:

$$\lambda(t|v) = \lambda_0(t) + v. \quad (2.3)$$

O termo  $v$  é a fragilidade introduzida aditivamente a função de risco do modelo (2.3) e ajuda a explicar a heterogeneidade dos indivíduos e o efeito de covariáveis não observadas no estudo por algum motivo, como por exemplo por dificuldade de mensuração da variável, indisponibilidade, dentre outras. O termo de fragilidade ajuda a indicar um risco aumentado de falha e no contexto de dados de eventos recorrentes também ajuda a medir a dependência entre os tempos de recorrência.

A função de sobrevivência condicional é dada por:

$$S(t|v) = \exp\{-\Lambda(t|v)\}, \text{ em que } \Lambda(t|v) = \int_0^t \lambda(u|v)du.$$

Considerando a função de risco com fragilidade inserida aditivamente dada em (2.3) temos que a função de risco acumulada condicional é dada por,

$$\Lambda(t|v) = \int_0^t [\lambda_0(u) + v]du = vt + \int_0^t \lambda_0(u)du = vt + \Lambda_0(t).$$

Logo, temos que

$$S(t|v) = \exp\{-vt - \Lambda_0(t)\}. \quad (2.4)$$

Pela relação entre as funções de sobrevivência, função de risco e função de densidade de probabilidade temos que a função de densidade de probabilidade condicional é

$$\begin{aligned} f(t|v) &= S(t|v)\lambda(t|v) \\ &= \exp\{-vt - \Lambda_0(t)\}(\lambda_0(t) + v). \end{aligned} \quad (2.5)$$

Considerando, por exemplo, a função de risco de base de uma distribuição *Exponencial*( $\mu$ ) na [Equação 2.5](#), temos que

$$\begin{aligned} f(t|v) &= \exp\{-vt - \mu t\}(\mu + v) \\ &= (\mu + v)\exp\{-t(\mu + v)\}, \end{aligned} \quad (2.6)$$

ou seja, pela [Equação 2.6](#) temos que  $T|v \sim \text{Exponencial}(\mu + v)$ . Logo  $E[T|v] = (\mu + v)^{-1}$ .

Como a função de verossimilhança usual não comporta funções condicionais é interessante obter as funções não condicionais para desenvolver o método de máxima verossimilhança. A função de sobrevivência não condicional do modelo (2.3) é dada por:

$$S(t) = \int_0^{\infty} S(t|v)f(v)dv, \quad (2.7)$$

em que  $f(v)$  é a função de densidade da variável de fragilidade. Portanto pela [Equação 2.7](#) temos a função de sobrevivência não condicional dada por

$$S(t) = \int_0^{\infty} \exp\{-\Lambda_0(t)\}\exp\{-vt\}f(v)dv,$$

e a transformada de Laplace é dada por:

$$L(t) = \int_0^{\infty} \exp\{-vt\}f(v)dv, \quad t \geq 0. \quad (2.8)$$

Logo,

$$S(t) = \exp\{-\Lambda_0(t)\}L(t) \quad (2.9)$$

A função de risco não condicional é dada pela seguinte relação com  $S(t)$  dada em (2.9):

$$\begin{aligned} \lambda(t) &= -\frac{d}{dt}[\log(S(t))] \\ &= -\frac{d}{dt}[\log(\exp\{-\Lambda_0(t)\}L(t))] \\ &= -\frac{d}{dt}[-\Lambda_0(t) + \log(L(t))] \\ &= \frac{d}{dt}[\Lambda_0(t)] - \frac{d}{dt}[\log(L(t))] \\ &= \lambda_0(t) - \frac{L'(t)}{L(t)}, \end{aligned} \quad (2.10)$$

em que  $L'(t)$  é a derivada da transformada de Laplace em relação a  $t$ .

A partir do modelo (2.3) ainda podemos encontrar a distribuição condicional da variável de fragilidade  $V$  dado  $T > t$ , ou seja, a distribuição da fragilidade dos indivíduos que sobreviveram até o tempo  $t$ . Note que podemos escrever a função de densidade de probabilidade  $f(v|T > t)$  em termos da função de sobrevivência condicional dada em (2.4), da função de sobrevivência marginal dada em (2.9) e da densidade da variável de fragilidade  $f(v)$ , da seguinte forma

$$f(v|T > t) = \frac{S(t|v)f(v)}{S(t)} = \frac{\exp\{-vt - \Lambda_0(t)\}f(v)}{\exp\{-\Lambda_0(t)\}L(t)} = \frac{\exp\{-vt\}f(v)}{L(t)}. \quad (2.11)$$

Note ainda que pela [Equação 2.11](#),

$$\begin{aligned} E[V|T > t] &= \int_0^\infty v f(v|T > t) dv \\ &= \int_0^\infty v \frac{\exp\{-vt\}f(v)}{L(t)} dv \\ &= \frac{\int_0^\infty v \exp\{-vt\}f(v) dv}{L(t)} \\ &= \frac{-\int_0^\infty -v \exp\{-vt\}f(v) dv}{L(t)} \\ &= \frac{-\int_0^\infty \frac{d}{dt} [\exp\{-vt\}] f(v) dv}{L(t)} \\ &= \frac{-\frac{d}{dt} [\int_0^\infty \exp\{-vt\} f(v) dv]}{L(t)} \\ &= \frac{-L'(t)}{L(t)}. \end{aligned} \quad (2.12)$$

Ou seja, utilizando a [Equação 2.12](#) podemos reescrever a [Equação 2.10](#) como

$$\lambda(t) = \lambda_0(t) + E[V|T > t].$$

Agora perceba que  $L(t)$  como representada em (2.8) pode ser vista como a função geradora de momentos da variável de fragilidade  $V$  aplicada no ponto  $-t$ , ou seja,  $L(t) = M_V(-t) = E[\exp\{-Vt\}]$ . Dessa forma podemos dizer que  $E[V|T > t] = -\frac{d}{dt} \log(M_V(-t))$  e  $Var[V|T > t] = \frac{d^2}{dt^2} \log(M_V(-t))$ . Portanto, a partir da distribuição da fragilidade é possível obtermos a média e a variância da fragilidade para os indivíduos sobreviventes no  $t$ . Para os indivíduos que falharam no tempo  $t$  podemos calcular a distribuição condicional da variável de fragilidade  $V$  dado  $T = t$ , da seguinte forma

$$\begin{aligned} f(v|T = t) &= \frac{f(t|v)f(v)}{f(t)} = \frac{S(t|v)\lambda(t|v)f(v)}{S(t)\lambda(t)} \\ &= \frac{\exp\{-vt - \Lambda_0(t)\}(\lambda_0(t) + v)f(v)}{\exp\{-\Lambda_0(t)\}L(t) (\lambda_0(t) + E[V|T > t])} \\ &= \frac{\exp\{-vt\}(\lambda_0(t) + v)f(v)}{L(t) (\lambda_0(t) + E[V|T > t])} \end{aligned} \quad (2.13)$$

Para estimar as fragilidades  $v_i$  de cada indivíduo  $i$  ( $i = 1, \dots, n$ ), podemos utilizar a distribuição condicional da fragilidade para os indivíduos que falharam ( $T = t$ ) e para os que

sobreviveram ao tempo  $t$ . A estimativa pode ser obtida calculando a fragilidade condicional média para cada indivíduo que falhou no tempo  $t$  e para cada indivíduo que sobreviveu ao tempo  $t$ , respectivamente, como

$$\hat{v}_i = \hat{E}[V|T = t_i] \quad (2.14)$$

$$\hat{v}_i = \hat{E}[V|T > t_i] \quad (2.15)$$

### 2.2.1 Modelo de fragilidade aditivo com covariáveis

O modelo de fragilidade aditivo com covariáveis  $\mathbf{x}$  inseridas aditivamente através de uma função  $g$  identidade é dado pela seguinte expressão:

$$\lambda(t|v, \mathbf{x}_i) = \lambda_0(t) + \mathbf{x}'_i \beta + v. \quad (2.16)$$

Seguindo os mesmos passos do modelo aditivo geral dado em (2.3) temos que a função de sobrevivência condicional é dada por:

$$S(t|v, \mathbf{x}_i) = \exp\{-\Lambda_0(t) - \mathbf{x}'_i \beta t - vt\}.$$

As funções de sobrevivência e de risco não condicionais do modelo (2.16) são expressas respectivamente por:

$$S(t|\mathbf{x}_i) = \exp\{-\Lambda_0(t)\} \exp\{-\mathbf{x}'_i \beta t\} L(t) \quad (2.17)$$

$$\lambda(t|\mathbf{x}_i) = \lambda_0(t) + \mathbf{x}'_i \beta - \frac{L'(t)}{L(t)}. \quad (2.18)$$

### 2.2.2 Função de Verossimilhança

Assumindo que temos  $n$  indivíduos independentes, e que os dados observados são da forma  $\mathcal{D} = \{(t_i, \delta_i), i = 1, \dots, n\}$ , em que  $t_i = \min\{\tau_i, C_i\}$  é o tempo observado do indivíduo  $i$ , com  $\tau_i$  denotando os tempos de falha e  $C_i$  denotando os tempos de censura. O termo  $\delta_i$ , chamado de indicador de censura, é igual a 1 se  $\{\tau_i \leq C_i\}$  e assume o valor zero caso contrário. Outra suposição usual em análise de sobrevivência é assumir que os tempos de sobrevivência e censura são independentes. Sob o mecanismo de censura não informativa temos que a função de verossimilhança é dada por

$$L(\theta|\mathcal{D}) = \prod_{i=1}^n \left\{ [\lambda(t_i)]^{\delta_i} S(t_i) \right\}, \quad (2.19)$$

em que  $\theta$  é o vetor de parâmetros do modelo. A partir de (2.19) podemos obter a função de verossimilhança para o modelo aditivo com fragilidade dado em (2.3) utilizando as funções (2.9)

e (2.10) e assim estimarmos os parâmetros via método de máxima verossimilhança (MV). A função de verossimilhança para o modelo (2.3) é dada por :

$$L(\theta|\mathcal{D}) = \prod_{i=1}^n \left\{ \left[ \lambda_0(t_i) - \frac{L'(t_i)}{L(t_i)} \right]^{\delta_i} \exp\{-\Lambda_0(t_i)\} L(t_i) \right\}. \quad (2.20)$$

O logaritmo da função de verossimilhança em (2.20) denotado por  $l(\theta|\mathcal{D}) = \log[L(\theta|\mathcal{D})]$  é dado por:

$$l(\theta|\mathcal{D}) = \sum_{i=1}^n \delta_i \left\{ \log \left[ \lambda_0(t_i) - \frac{L'(t_i)}{L(t_i)} \right] \right\} - \sum_{i=1}^n \Lambda_0(t_i) + \sum_{i=1}^n \log[L(t_i)], \quad (2.21)$$

e as estimativas de MV para os parâmetros são obtidas numericamente maximizando a função dada em (2.21). Nesse trabalho utilizamos o pacote `maxLik` (HENNINGSEN; TOOMET, 2011) do *software* R-Development-Core-Team (2023) para obtemos as estimativas de MV. Os escores de do vetor  $\theta$  são obtidos derivando a Equação 2.21 com relação a cada componente de  $\theta$ .

Para obtermos a função de verossimilhança para o modelo (2.16) e assim estimarmos os parâmetros via método de máxima verossimilhança (MV), podemos utilizar as funções de sobrevivência e risco não condicionais dadas em (2.17) e (2.18), respectivamente. Para o modelo aditivo com fragilidade dado em (2.16) temos a seguinte função de verossimilhança:

$$L(\theta|\mathcal{D}) = \prod_{i=1}^n \left\{ \left[ \lambda_0(t_i) + \mathbf{x}'_i \beta - \frac{L'(t_i)}{L(t_i)} \right]^{\delta_i} \exp\{-\Lambda_0(t_i) - \mathbf{x}'_i \beta t_i\} L(t_i) \right\}, \quad (2.22)$$

em que  $\delta_i = 1$  se  $t_i$  for um tempo de falha e é zero caso contrário,  $\theta$  é o vetor de parâmetros do modelo e  $\mathcal{D} = \{(t_i, \mathbf{x}_i, \delta_i), i = 1, \dots, n\}$ . O logaritmo da função de verossimilhança em (2.22) denotado por  $l(\theta|\mathcal{D}) = \log[L(\theta|\mathcal{D})]$  é dado por:

$$l(\theta|\mathcal{D}) = \sum_{i=1}^n \delta_i \left\{ \log \left[ \lambda_0(t_i) + \mathbf{x}'_i \beta - \frac{L'(t_i)}{L(t_i)} \right] \right\} - \sum_{i=1}^n [\Lambda_0(t_i) + \mathbf{x}'_i \beta t_i] + \sum_{i=1}^n \log[L(t_i)], \quad (2.23)$$

e as estimativas de MV para os parâmetros são obtidas numericamente maximizando a função dada em (2.23). Os escores de do vetor  $\theta$  são obtidos derivando a Equação 2.23 com relação a cada componente de  $\theta$ .

### 2.2.3 Modelo Aditivo com fragilidade Gama

Inicialmente iremos utilizar a distribuição Gama( $a, b$ ) para a variável de fragilidade do modelo dado na Equação 2.16, por esta ser mais tratável algebricamente, em que

$$f(v) = \frac{b^a}{\Gamma(a)} v^{a-1} \exp\{-vb\}, \quad a > 0, \quad b > 0. \quad (2.24)$$

Na parametrização Gama( $a, b$ ) temos que  $E(V) = a/b$  e  $Var(V) = a/b^2$ , nesse caso utilizamos o método delta, descrito em Casella e Berger (2002), para estimar os erros padrão

dessas quantidades. Para garantir a identificabilidade dos modelos de fragilidade multiplicativos [Elbers e Ridder \(1982\)](#) propuseram que o valor esperado da variável de fragilidade deveria ser igual a um. Para os modelos de fragilidade aditivo não temos essa garantia e além disso do ponto de vista prático não parece razoável supor que para modelagem aditiva o valor esperado da variável de fragilidade seja igual a um, dessa forma utilizaremos nesse trabalho a parametrização em que o valor esperado da variável de fragilidade não seja fixado.

A função de sobrevivência não condicional com fragilidade Gama dada em [2.24](#) a partir da equação [\(2.17\)](#) é de tal forma que:

$$\begin{aligned}
 S(t|\mathbf{x}_i) &= \exp\{-\Lambda_0(t)\} \exp\{-\mathbf{x}'_i\beta t\} L(t) \\
 &= \exp\{-\Lambda_0(t)\} \exp\{-\mathbf{x}'_i\beta t\} \int_0^\infty \exp\{-vt\} \frac{b^a}{\Gamma(a)} v^{a-1} \exp\{-vb\} dv \\
 &= \frac{b^a}{\Gamma(a)} \exp\{-\Lambda_0(t)\} \exp\{-\mathbf{x}'_i\beta t\} \int_0^\infty \exp\{-v(t+b)\} v^{a-1} dv
 \end{aligned} \tag{2.25}$$

Observando o integrando na expressão [\(2.25\)](#) como núcleo de uma distribuição Gama  $(a, t+b)$ , obtemos que  $L(t) = \left(\frac{b}{t+b}\right)^a$ , logo:

$$S(t|\mathbf{x}_i) = \exp\{-\Lambda_0(t) - \mathbf{x}'_i\beta t\} \left(\frac{b}{t+b}\right)^a, \tag{2.26}$$

e pela [Equação 2.18](#) a função de risco não condicional para fragilidade dada em [2.24](#) é

$$\lambda(t|\mathbf{x}_i) = \lambda_0(t) + \mathbf{x}'_i\beta + \left(\frac{a}{t+b}\right). \tag{2.27}$$

Pela [Equação 2.11](#) temos que a função de densidade de probabilidade de  $V$  dado  $T > t$  é dado por

$$\begin{aligned}
 f(v|T > t) &= \exp\{-vt\} \frac{b^a}{\Gamma(a)} v^{a-1} \exp\{-vb\} \left(\frac{t+b}{b}\right)^a \\
 &= \frac{(t+b)^a}{\Gamma(a)} \exp\{-v(t+b)\} v^{a-1},
 \end{aligned}$$

ou seja,  $V|T > t \sim \text{Gama}(a, t+b)$  e portanto  $E[V|T > t] = a/(t+b)$  e  $\text{Var}[V|T > t] = a/(t+b)^2$ .

Podemos obter ainda a distribuição da fragilidade  $V$  dado  $T = t$  a partir da seguinte expressão

$$\begin{aligned}
 f(v|T = t) &= \frac{S(t|v)\lambda(t|v)f(v)}{S(t)\lambda(t)} \\
 &= \frac{(\lambda_0(t) + \mathbf{x}'_i\beta)}{(a/(t+b)) + \lambda_0(t) + \mathbf{x}'_i\beta} \frac{(t+b)^a}{\Gamma(a)} \exp\{-v(t+b)\} v^{a-1} \\
 &\quad + \frac{a/(t+b)}{(a/(t+b)) + \lambda_0(t) + \mathbf{x}'_i\beta} \frac{(t+b)^{a+1}}{\Gamma(a+1)} \exp\{-v(t+b)\} v^{(a+1)-1} \\
 &= w_1 \times f_{G_1} + w_2 \times f_{G_2}.
 \end{aligned} \tag{2.28}$$



Em que  $f_{G_1}$  é a função de densidade de uma distribuição Gama( $a, t + b$ ) e  $f_{G_2}$  é a função de densidade de uma distribuição Gama( $a + 1, t + b$ ), com  $w_1 + w_2 = 1$ . O valor esperado de  $V|T = t$  pode ser calculado a partir de (2.28) da seguinte maneira

$$\begin{aligned}
 E[V|T = t] &= \int_0^\infty v f(v|T = t) dv = \int_0^\infty v (w_1 \times f_{G_1} + w_2 \times f_{G_2}) dv \\
 &= \int_0^\infty v \times w_1 \times f_{G_1} dv + \int_0^\infty v \times w_2 \times f_{G_2} dv \\
 &= w_1 \int_0^\infty v \times f_{G_1} dv + w_2 \int_0^\infty v \times f_{G_2} dv \\
 &= w_1 \left( \frac{a}{t + b} \right) + w_2 \left( \frac{a + 1}{t + b} \right). \tag{2.29}
 \end{aligned}$$

De acordo com as equações (2.14) e (2.15) podemos então estimar as fragilidades individuais com a parametrização Gama( $a, b$ ) por:  $\hat{v}_i = \hat{w}_1 \left( \frac{\hat{a}}{t_i + \hat{b}} \right) + \hat{w}_2 \left( \frac{\hat{a} + 1}{t_i + \hat{b}} \right)$  para os indivíduos que falharam no tempo  $t$  e por  $\hat{v}_i = \frac{\hat{a}}{t_i + \hat{b}}$  para os indivíduos censurados.

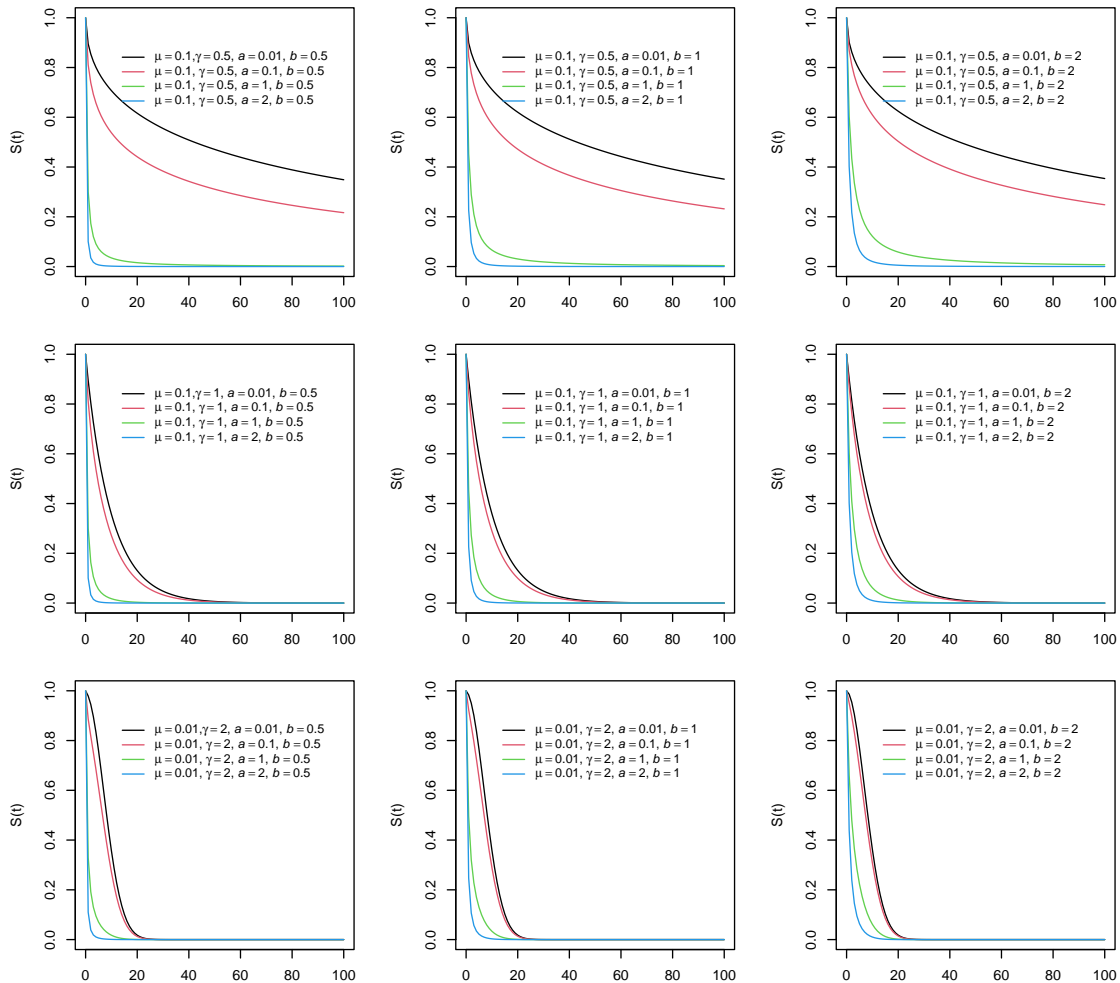


Figura 1 – Curvas de sobrevivência da Equação 2.26 sem covariáveis com risco de base de uma distribuição Weibull( $\mu, \gamma$ ) para diferentes valores de  $\mu, \gamma, a$  e  $b$  versus tempo.

Para o risco de base também é possível escolher várias distribuições, dentre elas a distribuição Weibull( $\mu, \gamma$ ). Na [Figura 1](#) apresentamos curvas de sobrevivência da [Equação 2.26](#) sem covariáveis com risco de base de uma distribuição Weibull( $\mu, \gamma$ ) para diferentes valores de  $\mu, \gamma, a$  e  $b$ . Ainda na [Figura 1](#) observamos que quando o parâmetro  $a$  cresce as curvas de sobrevivência caem rapidamente para zero, isso acontece porque o valor médio da variável de fragilidade cresce, ou seja, nesses cenários os indivíduos estão mais fragilizados.

Para distribuição Weibull( $\mu, \gamma$ ) temos função de densidade  $f(t) = \gamma\mu t^{\gamma-1} \exp[-\mu t^\gamma]$ , função de risco  $\lambda(t) = \gamma\mu t^{\gamma-1}$  e função de risco acumulado  $\Lambda(t) = \mu t^\gamma$ . Quando  $\gamma = 1$  temos a distribuição Exponencial( $\mu$ ) com  $\lambda(t) = \mu$  constante. Nesse caso a função de sobrevivência não condicional dada em [\(2.26\)](#) fica da forma:

$$S(t|\mathbf{x}_i) = \exp\{-\mu t - \mathbf{x}'_i \beta t\} \left(\frac{b}{t+b}\right)^a, \quad (2.30)$$

e a função de risco não condicional é expressa como:

$$\lambda(t|\mathbf{x}_i) = \mu + \mathbf{x}'_i \beta + \left(\frac{a}{t+b}\right).$$

A partir das funções  $S(t|\mathbf{x}_i)$  e  $\lambda(t|\mathbf{x}_i)$  é possível obter a função de verossimilhança e assim estimar os parâmetros via método de MV. Para o modelo aditivo com fragilidade Gama( $a, b$ ) e risco base de uma distribuição Exponencial( $\mu$ ) temos a seguinte função de verossimilhança:

$$L(\mu, a, b, \beta | \mathcal{D}) = \prod_{i=1}^n \left[ \left(\frac{a}{t_i+b}\right) + \mu + \mathbf{x}'_i \beta \right]^{\delta_i} \left(\frac{b}{t_i+b}\right)^a \exp[-\mu t_i - \mathbf{x}'_i \beta t_i], \quad (2.31)$$

em que  $\delta_i = 1$  se  $t_i$  for um tempo de falha e é zero caso contrário. O logaritmo da função de verossimilhança em [\(2.31\)](#) denotado por  $l(\mu, a, b, \beta | \mathcal{D}) = \log[L(\mu, a, b, \beta | \mathcal{D})]$  é dado por:

$$l(\mu, a, b, \beta | \mathcal{D}) = \sum_{i=1}^n \delta_i \log \left[ \left(\frac{a}{t_i+b}\right) + \mu + \mathbf{x}'_i \beta \right] + a \sum_{i=1}^n \log \left[ \frac{b}{t_i+b} \right] - \sum_{i=1}^n [\mu t_i + \mathbf{x}'_i \beta t_i], \quad (2.32)$$

e as estimativas de MV para os parâmetros são obtidas numericamente maximizando a função dada em [\(2.32\)](#). Os escores de  $\mu, a, b$  e  $\beta$  são obtidos derivando a [Equação 2.32](#) com relação a  $\mu, a, b$  e  $\beta$  respectivamente, e são dados por

$$\begin{aligned} \frac{\partial l(\mu, a, b, \beta | \mathcal{D})}{\partial \mu} &= \sum_{i=1}^n \frac{\delta_i}{(a/(t_i+b)) + \mu + \mathbf{x}'_i \beta} - \sum_{i=1}^n t_i, \\ \frac{\partial l(\mu, a, b, \beta | \mathcal{D})}{\partial a} &= \sum_{i=1}^n \delta_i \frac{(t_i+b)^{-1}}{(a/(t_i+b)) + \mu + \mathbf{x}'_i \beta} + \sum_{i=1}^n \log \left[ \frac{b}{t_i+b} \right], \\ \frac{\partial l(\mu, a, b, \beta | \mathcal{D})}{\partial b} &= - \sum_{i=1}^n \delta_i \frac{a(t_i+b)^{-2}}{(a/(t_i+b)) + \mu + \mathbf{x}'_i \beta} + \frac{a}{b} \sum_{i=1}^n \frac{t_i}{t_i+b}, \\ \frac{\partial l(\mu, a, b, \beta | \mathcal{D})}{\partial \beta} &= \sum_{i=1}^n \delta_i \frac{\mathbf{x}'_i \mathbf{A}}{(a/(t_i+b)) + \mu + \mathbf{x}'_i \beta} - \sum_{i=1}^n \mathbf{x}'_i \mathbf{A} t_i, \end{aligned} \quad (2.33)$$

em que,  $\mathbf{A} = (1, \dots, 1)_{p \times 1}$ . Os estimadores de MV dos parâmetros  $\mu, a, b$  e  $\beta$  são obtidos por métodos numéricos e não possuem forma fechada. A matriz observada de Fisher  $\mathcal{J}_{\mu, a, b, \beta} = J(\mu, a, b, \beta)$  é dada pelo negativo da matriz de derivadas das equações (2.33), em que

$$J(\mu, a, b, \beta) = \begin{bmatrix} \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial \mu^2} & \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial a \partial \mu} & \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial b \partial \mu} & \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial \beta \partial \mu} \\ \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial \mu \partial a} & \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial a^2} & \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial b \partial a} & \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial \beta \partial a} \\ \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial \mu \partial b} & \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial a \partial b} & \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial b^2} & \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial \beta \partial b} \\ \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial \mu \partial \beta} & \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial a \partial \beta} & \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial b \partial \beta} & \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial \beta^2} \end{bmatrix},$$

em que os elementos da matriz  $J(\mu, a, b, \beta)$  estão apresentados em [Equação B.1](#).

A distribuição assintótica conjunta para os estimadores de MV dos parâmetros do modelo (2.16) com fragilidade Gama( $a, b$ ) e risco base de uma distribuição Exponencial( $\mu$ ) segue aproximadamente a seguinte distribuição

$$\sqrt{n} \begin{pmatrix} \hat{\mu} - \mu \\ \hat{a} - a \\ \hat{b} - b \\ \hat{\beta} - \beta \end{pmatrix} \xrightarrow{D} N_4(\mathbf{0}, \mathcal{F}_{\mu, a, b, \beta}^{-1}).$$

De forma geral a estimação intervalar assintótica de 95% de confiança para o vetor de parâmetros  $\theta$  dos modelos propostos neste trabalho podem ser obtida através da seguinte expressão

$$\left[ \hat{\theta} - z_{0.975} \sqrt{\mathcal{F}^{-1}(\hat{\theta})}; \hat{\theta} + z_{0.975} \sqrt{\mathcal{F}^{-1}(\hat{\theta})} \right], \quad (2.34)$$

em que  $z_{0.975}$  é o quantil  $1 - 0.05/2$  da distribuição Normal padrão e  $\hat{\theta}$  é o estimador de MV para  $\theta$ .

## 2.3 Estudo de Simulação

Nesse estudo os tempos do modelo de fragilidade aditivo Gama-Exponencial foram gerados pelo método da inversão, para avaliar o comportamento assintótico dos estimadores de máxima verossimilhança para os parâmetros do modelo com uma covariável utilizamos 500 réplicas de Monte Carlo e avaliamos o viés, o Erro Quadrático Médio (EQM), a Raiz do Erro Quadrático Médio (REQM), o Erro Absoluto Médio (EAM) e Desvio padrão (DP) para 4 cenários mostrado na [Tabela 1](#).

Avaliamos os cenários para diferentes tamanhos de amostra  $n = 50, 100, 500$  e  $1000$ . O algoritmo seguiu a seguinte estrutura

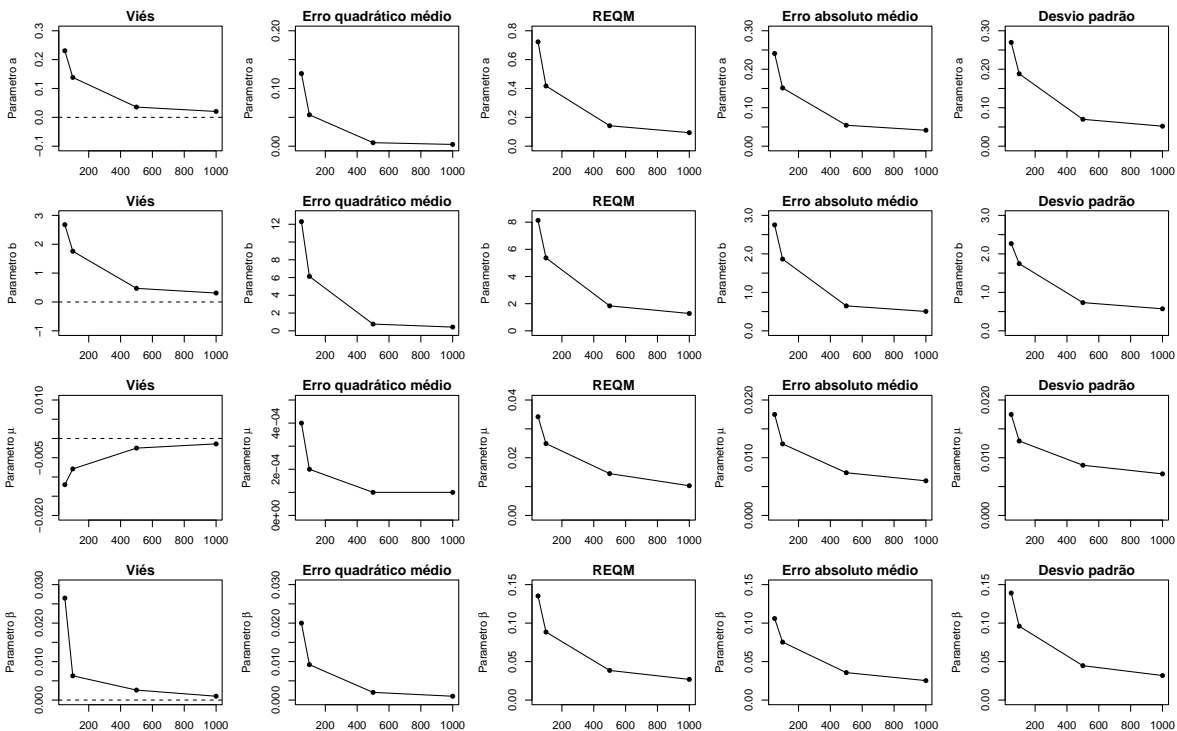
1. determinamos os valores dos parâmetros,  $a, b, \mu$  e  $\beta$ ;

2. geramos  $t'_i$  de  $F(t) = u$ , em que  $u \sim U(0, 1)$ ;
3. geramos  $Z_i \sim U(0, t_{MAX})$ ,  $t_{MAX} = \max(t'_i)$ ;
4.  $t_i = \min(t'_i, z_i)$ ;
5. Se  $t_i < z_i$ ,  $\delta_i = 1$ , se  $t_i \geq z_i$ ,  $\delta_i = 0$ .

Tabela 1 – Valores dos parâmetros do modelo (2.30) sob diferentes cenários.

Parâmetros	Cenários			
	1	2	3	4
$a$	0.05	0.05	0.05	0.5
$b$	0.8	0.8	1.8	1
$\mu$	0.05	0.05	0.05	0.05
$\beta$	0.5	1	0.5	0.1

Pelos resultados resumidos nas Figura 2, Figura 3, Figura 4 e Figura 5 percebemos que para os parâmetros  $a, b, \mu$  e  $\beta$  do modelo de fragilidade aditivo com fragilidade Gama( $a, b$ ) e risco de base Exponencial( $\mu$ ), tanto o viés como o EQM dos estimadores diminuem a medida que o tamanho amostral cresce, ou seja, em geral as propriedades assintóticas dos estimadores de máxima verossimilhança são verificadas. Além disso outras medidas complementares como o REQM, EAM e DP também se aproximam de zero quando  $n$  aumenta, o que sugere que as estimativas variam menos em torno da média para grandes amostras.

Figura 2 – Viés, EQM, REQM, EAM, e DP de  $(\hat{a}, \hat{b}, \hat{\mu}, \hat{\beta})$  versus  $n$  para o cenário 1.

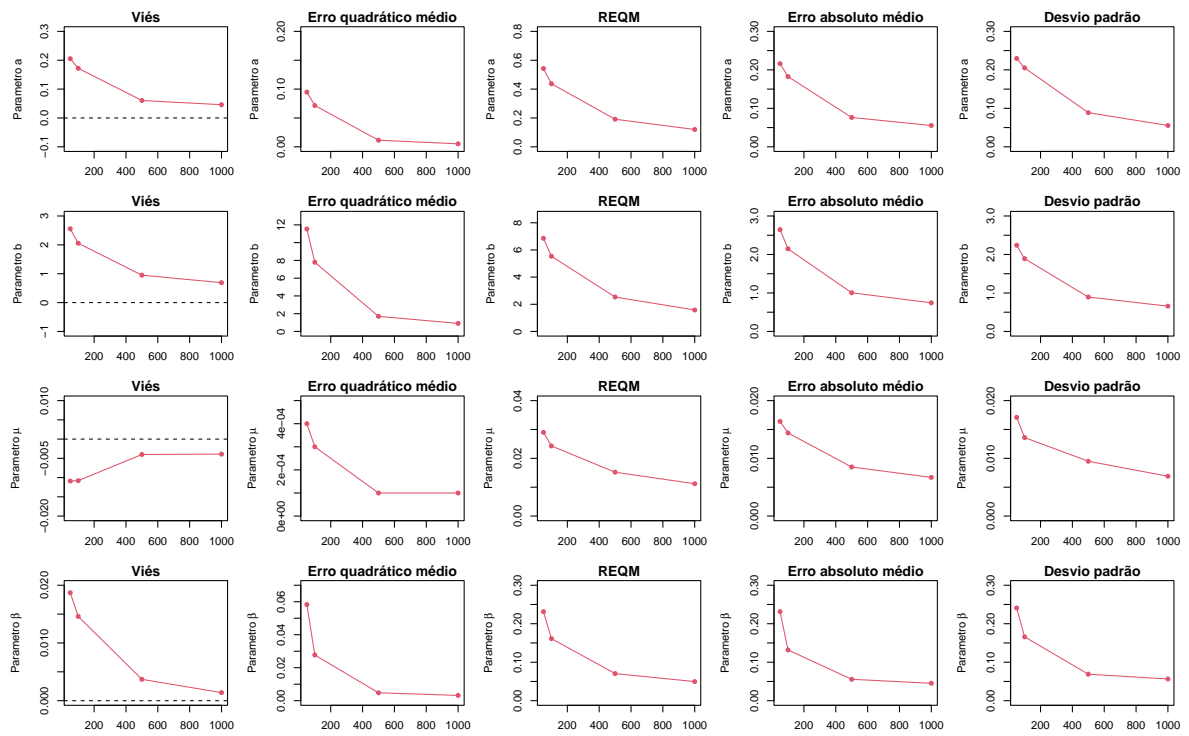


Figura 3 – Viés, EQM, REQM, EAM, e DP de  $(\hat{a}, \hat{b}, \hat{\mu}, \hat{\beta})$  versus  $n$  para o cenário 2.

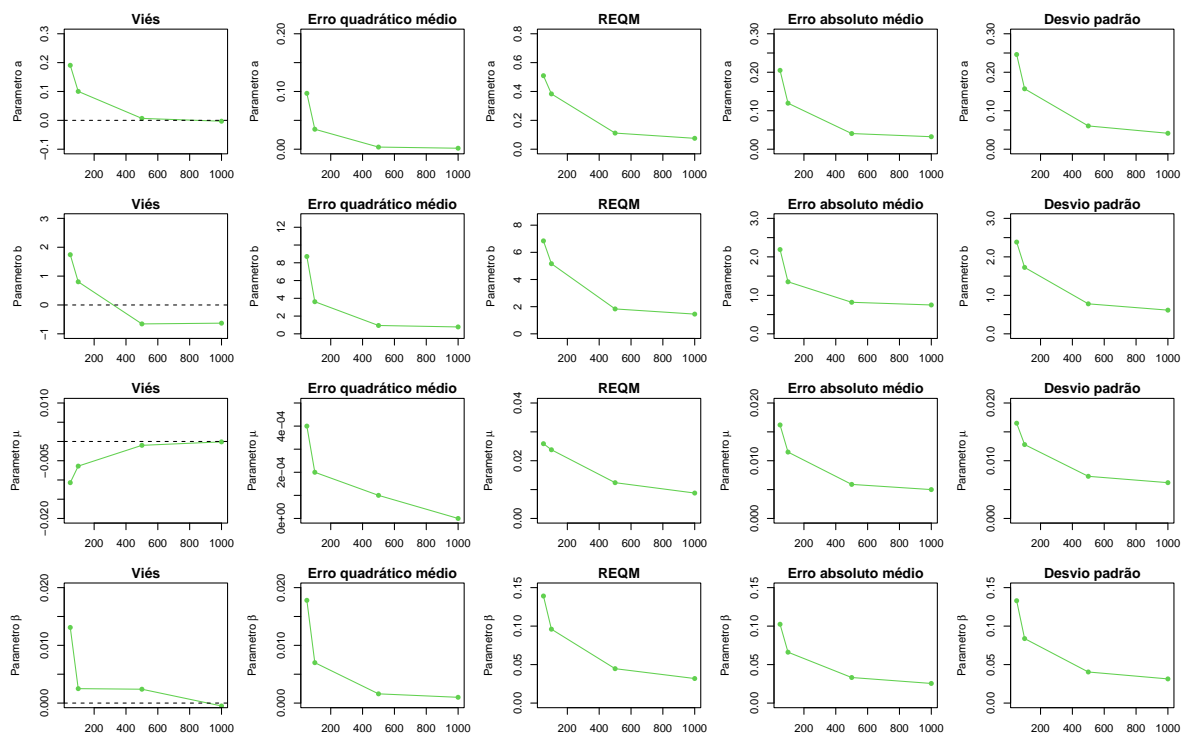


Figura 4 – Viés, EQM, REQM, EAM, e DP de  $(\hat{a}, \hat{b}, \hat{\mu}, \hat{\beta})$  versus  $n$  para o cenário 3.

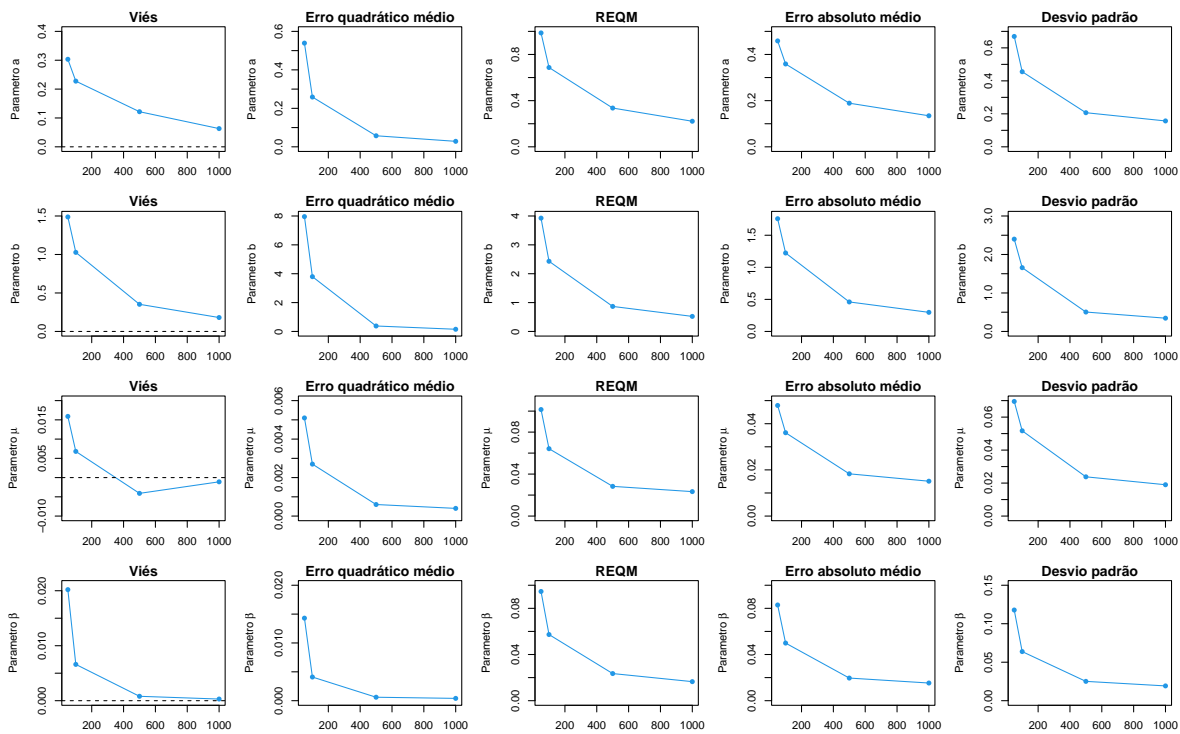


Figura 5 – Viés, EQM, REQM, EAM, e DP de  $(\hat{a}, \hat{b}, \hat{\mu}, \hat{\beta})$  versus  $n$  para o cenário 4.

Entretanto, para o parâmetro  $b$  da fragilidade  $\text{Gama}(a, b)$  percebemos que os valores das medidas REQM, EAM e DP não ficaram tão próximos de zero como para os demais parâmetros do modelo mesmo para  $n = 1000$ , sugerindo que para uma estimação mais precisa desse parâmetro é necessário uma amostra maior.

Em geral para todos os cenários simulados os estimadores de máxima verossimilhança comportaram-se de maneira similar e apresentaram resultados dentro do esperado, ou seja, a medida que o tamanho amostral cresce viés, EQM, REQM, EAM e DP descerem, portanto as estimativas obtidas parecem fornecer um desempenho razoável na estimação dos parâmetros do modelo de fragilidade aditivo (2.16). Por outro lado as probabilidades de cobertura não ficaram tão próximas dos valores nominais de 95% como esperado. Uma possível solução seria utilizar mais réplicas de Monte Carlo para que os valores das probabilidades de cobertura se aproximassem mais dos valores nominais.

## 2.4 Aplicação para dados Reais

Nessa seção apresentamos três aplicações do modelo de fragilidade aditivo com variável de fragilidade seguindo uma distribuição Gama e assumindo as distribuições Exponencial e Weibull para o risco de base. Nesse conjunto de dados, apenas algumas variáveis explicativas estão disponíveis no estudo, entretanto muitos outros fatores não observados podem influenciar o tempo de sobrevivência dos indivíduos. Geralmente, na coleta de dados em estudos retrospectivos,

fatores como por exemplo: estilo de vida, antecedentes pessoais ou familiares de câncer, ocupação e fatores genéticos são desconhecidos e não podem ser incluídos na análise. Dessa forma, se covariáveis clinicamente importantes não forem incluídas no planejamento do estudo a quantidade de heterogeneidade não observável aumentará, e portanto a inclusão de um termo de fragilidade pode ser importante para minimizar este problema.

### 2.4.1 Dados de câncer de pâncreas

Os dados dessa aplicação estão disponíveis em [Abu-Shawar et al. \(2020\)](#) e são referentes a 355 pacientes com câncer de pâncreas acompanhados de julho de 2006 a dezembro de 2018 em um estudo retrospectivo. Dentre os pacientes cerca de 59% são homens e 41% são mulheres. Nesse estudo 88.5% dos pacientes apresentaram o evento de interesse (morte) e 11.5% foram censurados, o tempo mediano de vida foi de 208 dias com intervalo de confiança de 95% (180; 249) dias e a respectiva curva de Kaplan-Meier é dada na [Figura 6](#).

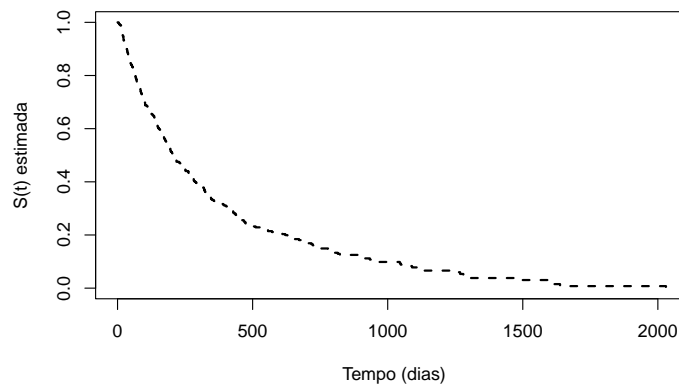


Figura 6 – Curva de sobrevivência estimada pelo estimador Kaplan-Meier para dados de câncer de pâncreas.

Aplicamos o modelo (2.3) com uma distribuição Gama( $a, b$ ) para a variável de fragilidade e riscos de base da distribuição Exponencial( $\mu$ ) e da distribuição Weibull( $\mu, \gamma$ ) aos dados.

Tabela 2 – Estimativas de máxima verossimilhança (EMV), erros padrão (EP), intervalos de confiança (IC) de 95%, AIC e BIC para os modelo aditivo com fragilidades Gama-Exponencial e Gama-Weibull para os dados de câncer de pâncreas.

Parâmetros	MFA-Gama-Exponencial			MFA-Gama-Weibull		
	EMV	EP	IC	EMV	EP	IC
$a$	$7.4 \times 10^{-9}$	0.0002	(0.0000; 0.0003)	$8.3 \times 10^{-9}$	0.0002	(0.0000; 0.0005)
$b$	$8.7 \times 10^{-2}$	0.0380	(0.0122; 0.1612)	$1.2 \times 10^{-1}$	0.0306	(0.0606; 0.1806)
$\mu$	$2.8 \times 10^{-3}$	0.0002	(0.0025; 0.0031)	$2.9 \times 10^{-3}$	0.0002	(0.0025; 0.0032)
$\gamma$	-	-	-	$9.2 \times 10^{-1}$	0.0405	(0.8366; 0.9953)
AIC		4328.52			4326.40	
BIC		4340.14			4341.89	

Os resultados obtidos estão apresentados na [Tabela 2](#), onde também comparamos os modelos em termos de AIC e BIC. Pelos critérios do AIC e BIC apresentados na [Tabela 2](#) o modelo de fragilidade aditivo Gama-Exponencial é melhor que o modelo com risco de base Weibull. A partir do modelo (2.3) com risco de base Exponencial estimado na [Tabela 2](#) obtivemos a curva de sobrevivência na [Figura 7](#).

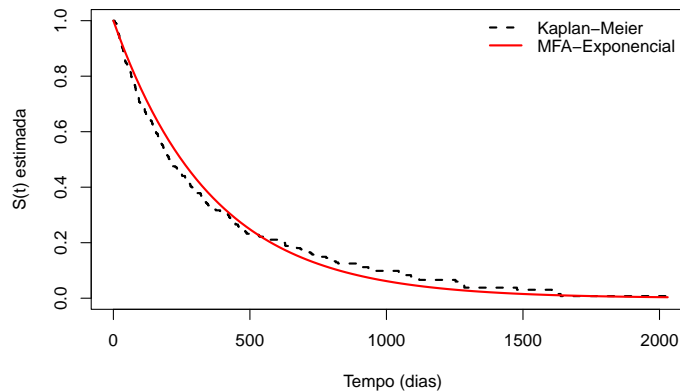


Figura 7 – Curva de sobrevivência estimada pelo MFA-Gama-Exponencial para dados de câncer de pâncreas.

Para calcular as fragilidades individuais apresentadas na [Figura 8](#) utilizamos as estimativas da [Tabela 2](#) e aplicamos na [Equação 2.29](#) para os indivíduos do estudo que falharam e calculamos  $\hat{E}[V|T > t_i]$  para os pacientes que não apresentaram o evento de interesse. A média das fragilidades individuais estimadas ( $1.87 \times 10^{-9}$ ) obtida é bem próxima da estimativa do valor esperado da variável de fragilidade ( $\hat{a}/\hat{b}$ ).

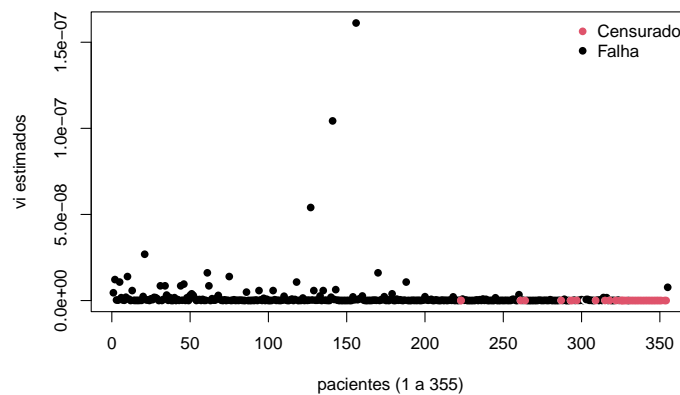


Figura 8 – Fragilidades individuais dos 355 pacientes do estudo.

Em geral os indivíduos que não falharam, ou sobreviveram mais tempo, apresentaram menor fragilidade, como vemos nos pontos vermelhos da [Figura 8](#). O paciente que apresentou a



maior fragilidade foi o indivíduo que sobreviveu apenas 4 dias, enquanto o paciente que morreu após 2030 dias de acompanhamento foi o que apresentou a menor fragilidade.

## 2.4.2 Dados de Leucemia

Os dados dessa aplicação foram estudados inicialmente por Feigl e Zelen (1965) e estão disponíveis na Tabela 26. Esses dados são de um estudo realizado com 33 pacientes que tinham leucemia aguda, e foram observados semanalmente. Dentre as covariáveis observadas está a variável indicadora da ausência do antígeno Calla na superfície dos blatos, denotada por (Ag-), ou seja,  $x = 1$  para o grupo (Ag-) e  $x = 0$  para o grupo (Ag+) que apresentaram o antígeno Calla na superfície dos blatos. O grupo Ag+ tinha 17 pacientes e o grupo Ag- contava com 16 pacientes.

Tabela 3 – Estatísticas descritivas dos dados de leucemia.

Grupos	n	Tempo mediano	LI (95%)	LS (95%)
Ag+	17	56	22	121
Ag-	16	7.5	4	43

De acordo com Tabela 3 o grupo Ag+ apresentou um tempo mediano de sobrevivência de 56 semanas, enquanto o grupo Ag- apresentou apenas 7.5 semanas de tempo mediano de sobrevivência. Na Figura 9 percebemos que as curvas de sobrevivência estimadas pelo estimador de Kaplan e Meier (1958) (K-M) parecem indicar que os pacientes do grupo Ag+ tem maior sobrevida que os pacientes do grupo Ag-. O teste Log-Rank (MANTEL, 1966) foi aplicado e obtivemos  $p$ -valor= 0.004, ou seja, rejeitamos a hipótese nula de que as curvas de sobrevivência são iguais a 1 de 5% de significância.

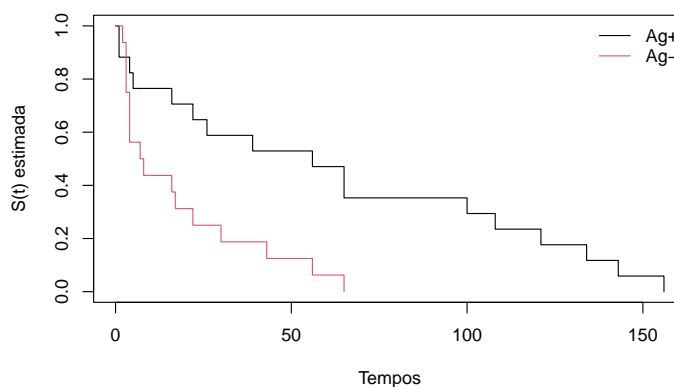


Figura 9 – Curvas de sobrevivência Kaplan-Meier estratificada por grupo

Para a aplicação aos dados da [Tabela 26](#) utilizamos a modelagem paramétrica com fragilidades Gama( $a, b$ ) com riscos de base da distribuição Exponencial( $\mu$ ) e da distribuição Weibull( $\mu, \gamma$ ) e comparamos os modelos utilizados em termos do AIC e BIC. As estimativas de máxima verossimilhança dos modelos abordados estão disponíveis na [Tabela 4](#).

Tabela 4 – Estimativas de máxima verossimilhança, erros padrão, intervalos de confiança de 95%, AIC e BIC para os modelos aditivo com fragilidade Gama e riscos de base da distribuição Exponencial( $\mu$ ) e da distribuição Weibull( $\mu, \gamma$ ) para os dados de leucemia.

Parametros	MFA-Gama-Exponencial			MFA-Gama-Weibull		
	EMV	EP	IC	EMV	EP	IC
$\beta$	0.0295	0.0134	(0.0032; 0.0557)	0.0489	0.0176	(0.0145; 0.0834)
$a$	0.0648	0.0760	(0.0000; 0.2138)	0.0905	0.0163	(0.0585; 0.1224)
$b$	1.8722	2.9915	(0.0000; 7.7354)	2.2018	2.8216	(0.0000; 7.7321)
$\mu$	0.0127	0.0042	(0.0045; 0.0210)	0.0127	0.0043	(0.0044; 0.0211)
$\gamma$	-	-	-	1.3438	0.6289	(0.1112; 2.5764)
AIC		305.46			306.62	
BIC		311.45			314.10	

Considerando os valores obtidos pelos critérios AIC e BIC apresentados na [Tabela 4](#), percebemos que o modelo aditivo com fragilidade Gama e risco de base da distribuição Exponencial fornece um ajuste melhor aos dados analisados do que o modelo Gama-Weibull, e portanto iremos modelar os dados a partir do modelo MFA-Gama-Exponencial.

Através das estimativas de máxima verossimilhança apresentadas na [Tabela 4](#) podemos concluir que como  $\hat{\beta} = 0.0295$  e o respectivo intervalo de confiança para o parâmetro  $\beta$  não inclui o valor zero, então a covariável grupo é significativa, ou seja, ser do grupo Ag- aumenta o risco de morte em pacientes com leucemia aguda.

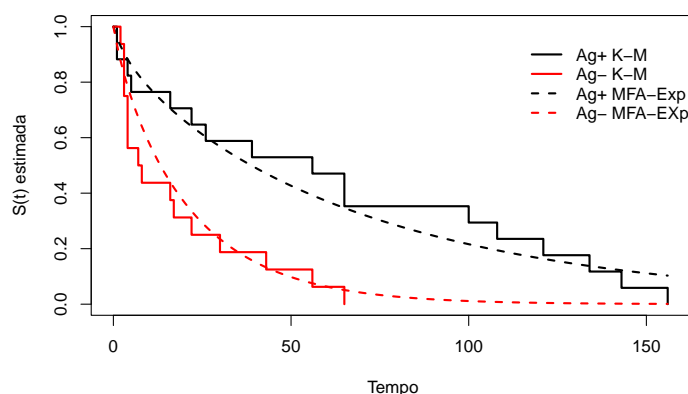


Figura 10 – Curvas de sobrevivências estimadas pelo MFA-Gama-Exponencial com as curvas estimadas por Kaplan-Meier.

Na [Figura 10](#) observamos que o modelo aditivo com fragilidade Gama( $a, b$ ) e função de risco de base Exponencial( $\mu$ ) parece se ajustar bem aos dados, pois se aproxima da curva

estimada não parametricamente pelo estimador K-M.

Tabela 5 – Fragilidade individual estimada dos 33 pacientes com leucemia aguda divididos por grupo com base na Tabela 4.

$i$	Tempos (Grupo Ag+)	$\hat{v}_i$	$i$	Tempos (Grupo Ag-)	$\hat{v}_i$
1	65	0.0020	18	56	0.0016
2	156	0.0006	19	65	0.0013
3	100	0.0011	20	17	0.0074
4	134	0.0007	21	7	0.0239
5	16	0.0160	22	16	0.0081
6	108	0.0010	23	22	0.0052
7	121	0.0009	24	3	0.0625
8	4	0.0901	25	4	0.0463
9	39	0.0043	26	2	0.0900
10	143	0.0007	27	3	0.0625
11	56	0.0025	28	8	0.0202
12	26	0.0079	29	4	0.0463
13	22	0.0101	30	3	0.0625
14	1	0.2451	31	30	0.0035
15	1	0.2451	32	4	0.0463
16	5	0.0713	33	43	0.0022
17	65	0.0020	-	-	-

Para calcular as fragilidades individuais apresentadas na Tabela 5 utilizamos as estimativas da Tabela 4 e aplicamos na Equação 2.29, pois todos os indivíduos do estudo falharam. Observando as fragilidades individuais estimadas apresentadas na Tabela 5 percebemos que em geral, com exceção dos pacientes 14 e 15, os indivíduos mais frágeis estão no Grupo Ag-, isto é, os indivíduos que não apresentaram o antígeno calla na superfície dos blastos são mais frágeis do que os indivíduos do Grupo Ag+. Quando calculamos a média das fragilidades individuais estimadas (0.0361) obtemos um valor bem próximo da estimativa do valor esperado da variável de fragilidade ( $\hat{a}/\hat{b} = 0.0346$ ).

### 2.4.3 Dados de câncer de pulmão

De acordo com o instituto nacional de câncer do Brasil (INCA), o câncer de pulmão é o terceiro mais comum em homens e o quarto em mulheres no Brasil, segundo as estimativas 2023. Em mortalidade é o primeiro entre os homens e o segundo entre as mulheres segundo estimativas mundiais de 2020, que apontou incidência de 2.2 milhões de casos novos, sendo 1.4 milhão em homens e 770 mil em mulheres. Ainda de acordo com o INCA a taxa de sobrevivência relativa em cinco anos para câncer de pulmão é de 18% e apenas 16% dos cânceres são diagnosticados em estágio inicial (câncer localizado), para o qual a taxa de sobrevivência de cinco anos é de 56%.

No estudo apresentado por Gazon *et al.* (2022), foram incluídos pacientes com diagnóstico de câncer de pulmão no estado de São Paulo, Brasil, entre 2000 e 2014. O acompanhamento foi realizado até 2018. O evento de interesse foi definido como óbito por câncer. O tempo até

o evento foi definido como o período desde o diagnóstico de câncer de pulmão até a morte por câncer. Os dados de pacientes que não morreram de câncer de pulmão durante o período de acompanhamento foram censurados à direita. O principal objetivo foi avaliar os efeitos das covariáveis medidas no início do estudo, incluindo sexo, idade no diagnóstico, estágio clínico e cirurgia, e detectar uma possível heterogeneidade não observada. Pela função de sobrevivência apresentada na Figura 11 percebemos que a taxa de sobrevivência em 5 anos é próxima de 10% e tempo mediano de sobrevivência é de 0.73 ano.

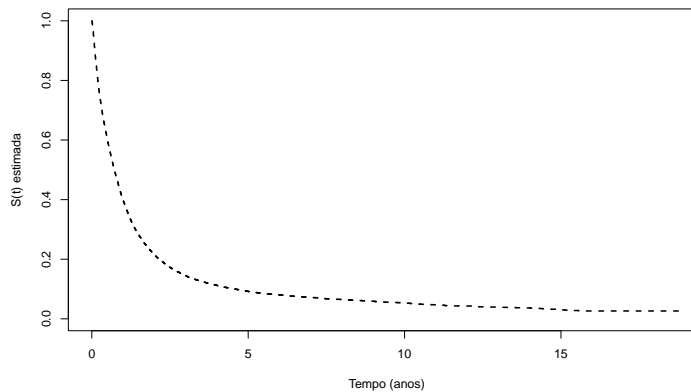


Figura 11 – Curva de sobrevivência estimada por Kaplan-Meier para os dados de câncer de pulmão.

As covariáveis utilizadas nessa aplicação estão categorizadas da seguinte maneira: Sexo (0: Masculino, 1: Feminino); Idade (0: < 60 anos, 1:  $\geq$  60 anos); cirurgia (0: Não, 1: Sim) e estágio clínico (0: estágio I(baseline), 1: estágio II, 2: estágio III e 3: estágio IV), sendo que quanto maior o estágio clínico mais avançado está o câncer. Aplicando o modelo de fragilidade aditivo Gama (2.26) com riscos de base Exponencial e Weibull obtivemos no modelo com as covariáveis sexo, idade no diagnóstico, estágio clínico e cirurgia os seguintes resultados

Tabela 6 – Estimativas de máxima verossimilhança, erros padrão, intervalos de confiança de 95%, AIC e BIC para os modelos aditivo com fragilidade Gama e riscos de base da distribuição Exponencial( $\mu$ ) e da distribuição Weibull( $\mu, \gamma$ ) para os dados de câncer de pulmão.

Parâmetros	MFA-Gama-Exponencial			MFA-Gama-Weibull		
	EMV	EP	IC	EMV	EP	IC
$a$	0.3192	0.0205	(0.2791; 0.3594)	0.7184	0.0839	(0.5538; 0.8829)
$b$	0.5599	0.0403	(0.4810; 0.6388)	0.9174	0.0943	(0.7326; 1.1022)
$\beta_1$ (sexo)	-0.0416	0.0053	(-0.0519; -0.0313)	-0.0460	0.0051	(-0.0560; -0.0360)
$\beta_2$ (idade)	0.0324	0.0049	(0.0229; 0.0420)	0.0403	0.0047	(0.0311; 0.0496)
$\beta_3$ (cirurgia)	-0.2824	0.0089	(-0.2998; -0.2650)	-0.2795	0.0089	(-0.2969; -0.2621)
$\beta_4$ (estágio II)	0.0220	0.0068	(0.0086; 0.0354)	0.0259	0.0067	(0.0127; 0.0391)
$\beta_5$ (estágio III)	0.1359	0.0089	(0.1185; 0.1532)	0.1336	0.0087	(0.1165; 0.1506)
$\beta_6$ (estágio IV)	0.5308	0.0120	(0.5073; 0.5544)	0.5243	0.0120	(0.5009; 0.5478)
$\mu$	0.3048	0.0108	(0.2838; 0.3259)	0.1800	0.0166	(0.1474; 0.2125)
$\gamma$	-	-	-	1.2757	0.0593	( 1.1596; 1.3919)
AIC		65549.22			65470.85	
BIC		65624.27			65554.23	

De acordo com os valores obtidos pelos critérios AIC e BIC apresentados na [Tabela 6](#) percebemos que o modelo aditivo com fragilidade Gama e risco de base da distribuição Weibull forneceu um ajuste melhor aos dados analisados do que o modelo Gama-Exponencial, e portanto iremos modelar os dados a partir do modelo MFA-Gama-Weibull.

De acordo com a [Tabela 6](#) podemos concluir que todas as covariáveis são significativas. Pacientes do sexo feminino tem maior tempo de sobrevivência ( $\beta_1 = -0.0460$ ), como esperado pacientes mais jovens também tem melhor prognóstico ( $\beta_2 = 0.0403$ ), bem como pessoas que se submeteram à cirurgia ( $\beta_3 = -0.2795$ ). Por fim quanto maior o estágio clínico menor o tempo de vida do paciente.

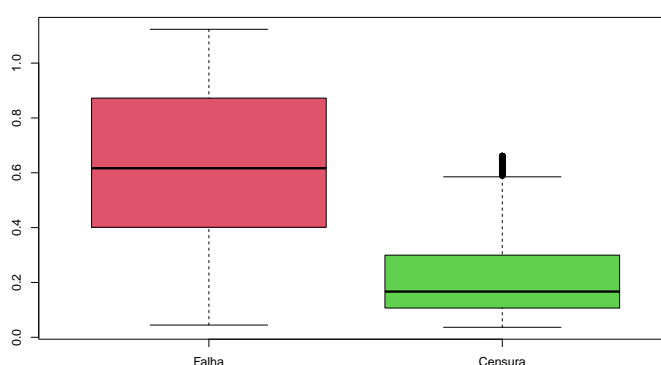


Figura 12 – Box-plot das fragilidades individuais para os indivíduos que falharam e não falharam durante o tempo de acompanhamento.

Calculamos a média e a variância da variável de fragilidade e obtivemos 0.7831 e 0.8536, respectivamente. Em seguida, utilizamos o método delta para calcular os respectivos erros padrão dessas quantidades e obtivemos os intervalos de confiança assintóticos (0.7384; 0.8277) e (0.6934; 1.0138), respectivamente. Nesse caso podemos admitir a presença de heterogeneidade não observada, o que justifica a inclusão do termo de fragilidade no modelo. Como esperado em geral as fragilidades individuais dos pacientes que falharam são maiores em relação aqueles que não apresentaram o evento de interesse durante o tempo de acompanhamento. A média das fragilidades individuais também ficou próxima da estimativa do valor esperado da variável de fragilidade.

## 2.5 Conclusão

Os modelos de fragilidade aditivos são alternativas interessantes aos modelos de risco multiplicativo, entretanto essa modelagem aditiva apresentada tem a vantagem de poder estimar valores negativos para a função de risco. A inclusão do termo de fragilidade no modelo aditivo faz sentido, sobretudo em estudos onde fatores não observáveis podem influenciar nos tempos de vida.

O estudo de simulação desenvolvido mostrou que em geral as estimativas de máxima verossimilhança melhoram em termos de Viés, EQM, REQM, EAM e DP a medida que o tamanho da amostra aumenta, entretanto mesmo para amostras maiores ( $n = 1000$ ) o comportamento das estimativas para o parâmetro  $b$  ainda não é tão bom, sobretudo em termos de Viés e EQM, sugerindo que para estimativas mais consistentes são necessários bancos de dados maiores, o que muitas vezes não se observa na prática.

O estimador proposto para calcular as fragilidades individuais a partir do modelo aditivo estudado nesse capítulo mostrou-se coerente, no sentido que atribui maior fragilidade para os indivíduos que apresentaram o evento de interesse mais rapidamente, ou seja, os indivíduos que não falharam, ou sobreviveram mais tempo, apresentaram menor fragilidade individual, enquanto aqueles que sobreviveram pouco tempo são considerados mais frágeis. Além disso a média das fragilidades individuais estimadas se aproxima bem da estimativa do valor esperado para a variável de fragilidade.

---

# MODELAGEM DE RISCO ADITIVO COM FRAGILIDADE PARA DADOS DE EVENTOS RECORRENTES

---

Diferente dos modelos de risco compartilhado em que cada indivíduo do mesmo grupo tem um único tempo, nos dados de eventos recorrentes cada indivíduo pode ser visto como uma espécie de grupo, em que cada indivíduo pode apresentar múltiplos tempos de falha, entretanto os tempos de recorrência são ordenados e nesse caso faz sentido truncar os tempos até um tempo máximo de estudo. Esse tempo total pode ser diferente para cada indivíduo, ou pode ser o mesmo para todos, dependendo do tipo de estudo. Dessa forma estudar uma modelagem aditiva para esse tipo de dados é muito útil quando o interesse é analisar o tempo de vida de pacientes que apresentam múltiplas falhas do mesmo tipo, por exemplo: ataques cardíacos, ataques epiléticos, ou ainda múltiplas falhas de tipos diferentes, como múltiplas sequelas, sintomas de piora de uma doença.

## 3.1 Modelo aditivo com fragilidade para dados de eventos recorrentes

Em estudos com eventos recorrentes os dados são multivariados, ou seja, os tempos de falhas recorrentes do indivíduo  $i$  são da forma  $0 \leq t_{i1} < t_{i2} < \dots < t_{i,m_i} \leq T_i$ , em que  $T_i$  é o tempo de vida total e  $t_{ij}$  é o tempo de vida observado do indivíduo  $i$  na  $j$ -ésima falha, sendo  $i = 1, 2, 3, \dots, n$  o índice que identifica o indivíduo e  $j = 1, 2, 3, \dots, m_i$  o índice que representa o evento recorrente do indivíduo  $i$ . Na modelagem para esse tipo de dados o termo de fragilidade também mede a dependência entre os tempos de falha do indivíduo, os quais são assumidos condicionalmente independentes dado a fragilidade.

Os dados observados são da forma  $\mathcal{D} = \{(t_{ij}, \delta_{ij}), i = 1, \dots, n, j = 1, \dots, m_i\}$ , em que

$t_{ij}$  é o tempo de vida observado do indivíduo  $i$  na  $j$ -ésima falha e  $\delta_{ij} = 1$  se  $t_{ij} \leq T_i$  e  $\delta_{ij} = 0$  se  $t_{ij} > T_i$ . Outra suposição usual é que os tempos entre as recorrências tem mesma função de sobrevivência. Quando esses tempos seguem uma distribuição exponencial temos um processo de Poisson homogêneo com função de intensidade constante, caso contrário quando a função de intensidade não é constante temos um processo de Poisson não homogêneo.

O modelo aditivo com fragilidade para dados de eventos recorrentes pode ser representado por:

$$\lambda_i(t|v_i) = \lambda_0(t) + v_i, \quad (3.1)$$

sendo  $v_i$  uma variável aleatória não negativa e  $\lambda_0(t)$  é a função de risco de base. A função  $S(t)$  pode ser expressa em termos da função de sobrevivência condicional  $S(t|v_i)$  da seguinte forma

$$\begin{aligned} S(t) &= \int_0^\infty S(t|v_i) f(v_i) dv_i \\ &= \int_0^\infty \exp[-\Lambda(t|v_i)] f(v_i) dv_i \\ &= \int_0^\infty \exp\left[-\int_0^t \lambda(u|v_i) du\right] f(v_i) dv_i \\ &= \int_0^\infty \exp\left[-\int_0^t \lambda_0(u) + v_i du\right] f(v_i) dv_i \\ &= \int_0^\infty \exp[-\Lambda_0(t) - v_i t] f(v_i) dv_i \\ &= \exp[-\Lambda_0(t)] \int_0^\infty \exp[-v_i t] f(v_i) dv_i \\ &= \exp[-\Lambda_0(t)] L(t), \end{aligned}$$

em que,  $L(t)$  é a transformada de Lapalce dada em (2.8). Dessa forma a função de intensidade acumulada não condicional para o modelo dado em (3.1) é dada por

$$\Lambda(t) = -\log(S(t)) = \Lambda_0(t) - \log(L(t)), \quad (3.2)$$

e portanto

$$\lambda_i(t) = \lambda_0(t) - \frac{L'(t)}{L(t)}.$$

### 3.1.1 Modelo de fragilidade aditivo com covariáveis

Incluindo covariáveis no modelo (3.1) também de forma aditiva temos a seguinte equação:

$$\lambda_i(t|v_i, \mathbf{x}_i) = \lambda_0(t) + \mathbf{x}_i' \boldsymbol{\beta} + v_i, \quad (3.3)$$

e podemos usar a mesma estratégia para deduzir a função de intensidade não condicional, que é dada por:

$$\lambda_i(t|\mathbf{x}_i) = \lambda_0(t) + \mathbf{x}_i' \boldsymbol{\beta} - \frac{L'(t)}{L(t)}. \quad (3.4)$$



Note que a função de intensidade acumulada  $\int_0^{T_i} \lambda_i(u|\mathbf{x}_i) du$  é obtida integrando a [Equação 3.4](#) do tempo inicial até o tempo total do estudo. Dessa forma podemos obter a função de verossimilhança pela seguinte expressão

$$\begin{aligned} L(\theta|\mathcal{D}) &= \prod_{i=1}^n \prod_{j=1}^{m_i} \left\{ [\lambda_i(t_{ij}|\mathbf{x}_i)]^{\delta_{ij}} \right\} \times \prod_{i=1}^n \exp \left[ - \int_0^{T_i} \lambda_i(u|\mathbf{x}_i) du \right] \\ &= \prod_{i=1}^n \prod_{j=1}^{m_i} \left\{ \left[ \lambda_0(t_{ij}) + \mathbf{x}'_i \beta - \frac{L'(t_{ij})}{L(t_{ij})} \right]^{\delta_{ij}} \right\} \\ &\quad \times \prod_{i=1}^n \exp \left\{ -\Lambda_0(T_i) - \mathbf{x}'_i \beta T_i \right\} L(T_i), \end{aligned} \quad (3.5)$$

em que  $\theta$  é o vetor de parâmetros do modelo e  $\mathcal{D} = \{(t_{ij}, \mathbf{x}_i, \delta_{ij}), i = 1, \dots, n, j = 1, \dots, m_i\}$ . A função de log verossimilhança correspondente é dada por:

$$\begin{aligned} l(\theta|\mathcal{D}) &= \sum_{i=1}^n \sum_{j=1}^{m_i} \delta_{ij} \log \left( \lambda_0(t_{ij}) + \mathbf{x}'_i \beta - \frac{L'(t_{ij})}{L(t_{ij})} \right) \\ &\quad + \sum_{i=1}^n \left( -\Lambda_0(T_i) - \mathbf{x}'_i \beta T_i \right) + \log(L(T_i)), \end{aligned} \quad (3.6)$$

em que,  $\delta_{ij} = 1$  se  $t_{ij}$  é observado e  $\delta_{ij} = 0$  se  $t_{ij}$  é censurado. Para o procedimento de estimação pelo método de máxima verossimilhança utilizamos a função log verossimilhança dada em (3.6).

Além de estimarmos os parâmetros do modelo (3.3) temos o interesse de estimarmos as fragilidades de cada indivíduo em estudo, dessa forma utilizamos a ideia desenvolvida por Wang, Klein e Moeschberger (1995) e utilizada em um contexto mais geral por Munda, Rotolo e Legrand (2012), e propomos o seguinte estimador para as fragilidades individuais do modelo (3.3)

$$\hat{v}_i = \frac{E[V^{d_i+1} \exp\{-VT_i\}]}{E[V^{d_i} \exp\{-VT_i\}]}, \quad (3.7)$$

em que,  $d_i = \sum_{j=1}^{m_i} \delta_{ij}$  é o número de eventos recorrentes para o indivíduo  $i$  e  $T_i$  é o tempo de vida total do estudo. Escrevendo os valores esperados apresentados em (3.7) em termos de derivadas da transformada de Laplace temos que  $E[V^{d_i+1} \exp\{-VT_i\}] = (-1)^{d_i+1} L^{(d_i+1)}(T_i)$  e  $E[V^{d_i} \exp\{-VT_i\}] = (-1)^{d_i} L^{(d_i)}(T_i)$ , ou seja,

$$\hat{v}_i = \frac{-L^{(d_i+1)}(T_i)}{L^{(d_i)}(T_i)}. \quad (3.8)$$

As estimativas das fragilidades individuais podem ser obtidas como função de (3.8).

### 3.1.2 Modelo Aditivo com fragilidade Gama

Assumindo a distribuição Gama( $1/\alpha, 1/\alpha$ ) para a variável de fragilidade do modelo (3.3) temos que a transformada de Laplace é dada por  $L(t) = (1 + \alpha t)^{-1/\alpha}$  e a partir de (3.4) temos a seguinte função de intensidade marginal:

$$\lambda_i(t|\mathbf{x}_i) = \lambda_0(t) + \mathbf{x}'_i \beta + (1 + \alpha t)^{-1} \quad (3.9)$$

e a função de sobrevivência é expressa como:

$$\begin{aligned}
S(t|\mathbf{x}_i) &= \exp[-\Lambda(t|\mathbf{x}_i)] = \exp\left[-\int_0^t \lambda_i(u|\mathbf{x}_i)du\right] \\
&= \exp\left[-\int_0^t \lambda_0(u) + \mathbf{x}'_i\beta + (1 + \alpha u)^{-1}du\right] \\
&= \exp\left[-\mathbf{x}'_i\beta t - \Lambda_0(t) - \int_0^t (1 + \alpha u)^{-1}du\right] \\
&= \exp\left[-\mathbf{x}'_i\beta t - \Lambda_0(t) - \frac{\log(1 + \alpha t)}{\alpha}\right] \\
&= \exp[-\mathbf{x}'_i\beta t - \Lambda_0(t)] (1 + \alpha t)^{-1/\alpha}.
\end{aligned} \tag{3.10}$$

A partir da [Equação 3.5](#) obtemos a seguinte função de verossimilhança:

$$\begin{aligned}
L(\theta|\mathcal{D}) &= \prod_{i=1}^n \prod_{j=1}^{m_i} [\lambda_0(t_{ij}) + \mathbf{x}'_i\beta + (1 + \alpha t_{ij})^{-1}]^{\delta_{ij}} \times \\
&\quad \prod_{i=1}^n \exp[-\mathbf{x}'_i\beta T_i - \Lambda_0(T_i)] (1 + \alpha T_i)^{-1/\alpha},
\end{aligned} \tag{3.11}$$

em que  $\mathcal{D} = \{(t_{ij}, \mathbf{x}_i, \delta_{ij}), i = 1, \dots, n, j = 1, \dots, m_i\}$ . A função de log verossimilhança correspondente é dada por:

$$\begin{aligned}
l(\theta|\mathcal{D}) &= \sum_{i=1}^n \sum_{j=1}^{m_i} \delta_{ij} \log\left(\lambda_0(t_{ij}) + \mathbf{x}'_i\beta + (1 + \alpha t_{ij})^{-1}\right) \\
&\quad + \sum_{i=1}^n (-\Lambda_0(T_i) - \mathbf{x}'_i\beta T_i) - \frac{1}{\alpha} \log(1 + \alpha T_i),
\end{aligned} \tag{3.12}$$

em que,  $\delta_{ij} = 1$  se  $t_{ij}$  é observado e  $\delta_{ij} = 0$  se  $t_{ij}$  é censurado. Para completar a modelagem paramétrica devemos supor uma distribuição para o risco de base, como por exemplo: Exponencial, Weibull, dentre outras. Para obter as estimativas de máxima verossimilhança numericamente descrevemos a seguir a função de verossimilhança assumindo o risco de base de uma distribuição Weibull.

$$\begin{aligned}
L(\mu, \gamma, \alpha, \beta|\mathcal{D}) &= \prod_{i=1}^n \prod_{j=1}^{m_i} [\gamma \mu t_{ij}^{\gamma-1} + \mathbf{x}'_i\beta + (1 + \alpha t_{ij})^{-1}]^{\delta_{ij}} \times \\
&\quad \prod_{i=1}^n \exp[-\mathbf{x}'_i\beta T_i - \mu T_i^\gamma] (1 + \alpha T_i)^{-1/\alpha}
\end{aligned}$$

A função de log verossimilhança correspondente é dada por:

$$\begin{aligned}
l(\mu, \gamma, \alpha, \beta|\mathcal{D}) &= \sum_{i=1}^n \sum_{j=1}^{m_i} \delta_{ij} \log\left(\gamma \mu t_{ij}^{\gamma-1} + \mathbf{x}'_i\beta + (1 + \alpha t_{ij})^{-1}\right) \\
&\quad + \sum_{i=1}^n (-\mu T_i^\gamma - \mathbf{x}'_i\beta T_i) - \frac{1}{\alpha} \log(1 + \alpha T_i)
\end{aligned} \tag{3.13}$$

Para um risco de base constante de uma distribuição Exponencial( $\mu$ ) temos:

$$L(\mu, \alpha, \beta | \mathcal{D}) = \prod_{i=1}^n \prod_{j=1}^{m_i} [\mu + \mathbf{x}'_i \beta + (1 + \alpha t_{ij})^{-1}]^{\delta_{ij}} \times \prod_{i=1}^n \exp[-\mathbf{x}'_i \beta T_i - \mu T_i] (1 + \alpha T_i)^{-1/\alpha}$$

A função de log verossimilhança correspondente é dada por:

$$l(\mu, \alpha, \beta | \mathcal{D}) = \sum_{i=1}^n \sum_{j=1}^{m_i} \delta_{ij} \log(\mu + \mathbf{x}'_i \beta + (1 + \alpha t_{ij})^{-1}) + \sum_{i=1}^n (-\mu T_i - \mathbf{x}'_i \beta T_i) - \frac{1}{\alpha} \log(1 + \alpha T_i) \quad (3.14)$$

### 3.1.2.1 Fragilidades individuais

Utilizando o resultado apresentado em [Munda, Rotolo e Legrand \(2012\)](#) temos que para fragilidade  $V \sim \text{Gama}(1/\alpha, 1/\alpha)$ ,  $L^{(d_i)}(T_i) = (-1)^{d_i} (1 + \alpha T_i)^{-d_i} \left[ \prod_{k=0}^{d_i-1} (1 + k\alpha) \right] (1 + \alpha T_i)^{-1/\alpha}$  e  $L^{(d_i+1)}(T_i) = (-1)^{d_i+1} (1 + \alpha T_i)^{-(d_i+1)} \left[ \prod_{k=0}^{d_i} (1 + k\alpha) \right] (1 + \alpha T_i)^{-1/\alpha}$ , para  $d_i \geq 1$ . Dessa forma podemos estimar a fragilidade individual para dados de eventos recorrentes a partir da [Equação 3.8](#) para o modelo aditivo paramétrico com fragilidade  $\text{Gama}(1/\alpha, 1/\alpha)$  pela seguinte expressão

$$\hat{v}_i = (1 + \hat{\alpha} T_i)^{-1} (1 + d_i \hat{\alpha}) \quad (3.15)$$

Nesse trabalho consideramos que o tempo de truncamento  $T_i$  é o mesmo para todos indivíduos e ainda que os indivíduos só entram no estudo após apresentarem a primeira falha. Se o indivíduo  $i$  não apresenta tempos recorrentes então  $d_i = 0$ , ou seja, podemos estimar sua fragilidade individual por  $\hat{v}_i = (1 + \hat{\alpha} T_i)^{-1}$ .

### 3.1.3 Estimador não paramétrico da função de sobrevivência para dados de eventos recorrentes.

Para estimar a função de sobrevivência para tempos recorrentes [Wang e Chang \(1999\)](#) propuseram inicialmente um estimador de  $S(t)$  para tempos de recorrências sem censura  $(t_{i1}, t_{i2}, \dots, t_{im_i})$ . Definindo

$$U_i(t) = \frac{1}{m_i^*} \sum_{j=1}^{m_i^*} I(T_{ij} > t),$$

em que  $m_i^* = m_i - 1$  se  $m_i \geq 2$ ,  $m_i^* = 1$  se  $m_i = 1$  e  $I(\cdot)$  é a função indicadora.

Note que  $U_i(t)$  pode ser entendido como a proporção de eventos que não ocorreram antes do tempo  $t$  para o indivíduo  $i$ . Pode ser verificado que, condicionados a  $(m_i, X_i, C_i)$ , os tempos  $T_{i1}, \dots, T_{im_i^*}$  são identicamente distribuídos e isso implica que  $S(t) = E[U_i(t)]$ . Definindo

$a_i = a(C_i)$ , uma função positiva tal que  $E(a_i^2) < \infty$ ,  $i = 1, \dots, n$ , o estimador da função de sobrevivência para eventos recorrentes não censurados pode ser definido por:

$$\hat{S}_n(t) = \frac{\sum_{i=1}^n a_i U_i(t)}{\sum_{i=1}^n a_i}, \quad (3.16)$$

em que  $a_i$  é um peso que depende do tempo de censura  $C_i$  e pondera  $U_i(t)$ . Para generalizar o estimador  $\hat{S}_n(t)$  dado na equação [Equação 3.16](#) para tempos recorrentes com censura  $(t_{i1}, t_{i2}, \dots, t_{i, m_i-1}, t_{i, m_i}^+)$  renomea-se os tempos de recorrência observados e de censura da seguinte forma

$$y_{ij} = \begin{cases} t_{ij}, & \text{se } j = 1, \dots, m_i - 1 \\ t_{i, m_i}^+, & \text{se } j = m_i \end{cases}$$

A massa total do risco definido no tempo  $t$ , é dada por:

$$R^*(t) = \sum_{i=1}^n \left[ \frac{a_i}{m_i^*} \sum_{j=1}^{m_i^*} I(y_{ij} \geq t) \right],$$

e a massa de recorrência no tempo  $t$ , dada por:

$$d^*(t) = \sum_{i=1}^n \left[ \frac{a_i I(m_i \geq 2)}{m_i^*} \sum_{j=1}^{m_i^*} I(y_{ij} = t) \right]$$

Denotando por  $y_1^*, y_2^*, \dots, y_k^*$  os tempos não censurados distintos e ordenados, o estimador generalizado da função de sobrevivência para eventos recorrentes proposto por [Wang e Chang \(1999\)](#) é expresso pelo produto limite:

$$\hat{S}_n(t) = \prod_{y_i^* \leq t} \left\{ 1 - \frac{d^*(y_i^*)}{R^*(y_i^*)} \right\}. \quad (3.17)$$

Note que  $\hat{S}_n(t)$  é não crescente em  $t$  e satisfaz  $0 \leq \hat{S}_n(t) \leq 1$ , além disso o estimador da [Equação 3.17](#) se reduz ao estimador apresentado na [Equação 3.16](#) se os dados não são censurados. Embora o estimador de [Kaplan e Meier \(1958\)](#) obtido pelos tempos de recorrência agrupados seja geralmente enviesados, o estimador de Kaplan-Meier derivado dos primeiros tempos de recorrência é apropriado para estimar  $S(t)$ . [Wang e Chang \(1999\)](#) estudaram o comportamento do estimador proposto  $\hat{S}_n(t)$  a partir da escolha dos  $a_i$ 's, segundo eles, fazendo escolhas apropriadas para  $a_i$  o estimador  $\hat{S}_n(t)$  pode ser melhor que o estimador de Kaplan-Meier. Para  $a_i = 1$  (ou uma constante positiva qualquer) o estimador  $\hat{S}_n(t)$  é sempre melhor. Fica claro que, nessa abordagem de [Wang e Chang \(1999\)](#), o comportamento de  $\hat{S}_n(t)$  depende da escolha de  $a_i$ . Ainda de acordo com [Wang e Chang \(1999\)](#) a eficiência do estimador  $\hat{S}_n(t)$  depende do quanto os tempos são correlacionados. Se os tempos são fracamente correlacionados a eficiência de  $\hat{S}_n(t)$  melhora, mas para tempos altamente correlacionados não há ganhos significativos ao usar  $\hat{S}_n(t)$  para estimar  $S(t)$ .

## 3.2 Estudo de simulação

Nesse estudo os tempos do modelo de fragilidade aditivo Gama-Exponencial foram gerados pelo método da inversão, e o método de avaliação utilizou 1000 réplicas de Monte Carlo para os critérios de EQM, EAM e Viés para alguns cenários variando os valores do parâmetro  $\alpha$  e  $\mu = 0.03$ , tamanhos de amostra  $n = 50, 100$  e proporção de censura 0.0, 0.1 e 0.2. O algoritmo seguiu os seguintes passos

1. Definimos os parâmetros do modelo a ser gerado;
2. Geramos o número de recorrências de cada indivíduo da amostra;
3. Geramos os tempos a partir da função de distribuição;
4. Calculamos as estimativas de máxima verossimilhança;
5. Repetimos os passos 2 e 3 até o número máximo de réplicas;
6. Por fim calculamos as medidas de EQM, EAM e Viés a partir da média dos valores obtidos em cada réplica.

Para valores maiores de  $\alpha$  o EQM, EAM e Viés crescem com a proporção de censura, entretanto as medidas de EQM, EAM e Viés não caem sistematicamente com o aumento de  $n$ . Para o parâmetro  $\mu$ , devido sua pequena magnitude, temos valores de EQM, EAM e Viés próximos de zero e não percebemos ganhos significativos com o aumento do tamanho da amostra, sobretudo para valores grandes de  $\alpha$ .

Pelos resultados numéricos apresentados na [Tabela 7](#) percebemos que a medida que aumentamos o tamanho da amostra o EQM do estimador para o parâmetro  $\alpha = 10$  diminui, assim como o EAM e o Viés. Entretanto, para  $\alpha = 50$  e  $\alpha = 100$  isso não acontece. Esse comportamento para valores grandes do parâmetro de fragilidade  $\alpha$  parece indicar uma limitação da abordagem clássica para estimar os parâmetros do modelo aditivo em cenários de grande variabilidade entre os indivíduos. Em geral quanto maior o valor do  $\alpha$  verdadeiro maiores os valores de EQM, EAM e Viés. Já para o parâmetro  $\mu$  as medidas de EQM, EAM e Viés não se alteram significativamente quando o tamanho da amostra aumenta. Os valores nominais do EQM, EAM e Viés para as estimativas do parâmetro  $\mu$  apresentam valores nominais baixos devido a magnitude das estimativas serem bem próximas a zero. Como esperado observamos que a medida que o percentual de censura aumenta as estimativas pioram, tanto em termos do EQM e EAM como em termos do Viés.

Tabela 7 – EQM, EAM e Viés dos estimadores do modelo (3.3) sem covariável, com fragilidade Gama( $1/\alpha, 1/\alpha$ ) e risco de base de uma distribuição Exponencial( $\mu$ ) para diferentes valores de  $\alpha$ ,  $\mu$  e  $n$ , com percentual de censura (PC) variando de 0% a 20%.

PC	n	$\alpha$	$\mu$	$\hat{\alpha}$			$\hat{\mu}$		
				EQM	EAM	Viés	EQM	EAM	Viés
0%	50	10	0.03	23.6651	4.0077	1.8737	0.0000	0.0042	-0.0002
		50	0.03	46.9722	6.2492	6.1781	0.0002	0.0151	0.0151
		100	0.03	49.9823	6.4106	6.3471	0.0003	0.0173	0.0173
	100	10	0.03	19.3208	3.5647	-0.0284	0.0001	0.0047	-0.0023
		50	0.03	47.3491	6.2301	6.1887	0.0002	0.0149	0.0149
		100	0.03	52.6826	6.6627	6.6339	0.0003	0.0172	0.0172
5%	50	10	0.03	34.6707	5.0707	4.6636	0.0002	0.0131	-0.0131
		50	0.03	48.6403	6.3357	6.3065	0.0000	0.0040	-0.0039
		100	0.03	52.4957	6.6767	6.6665	0.0000	0.0027	-0.0023
	100	10	0.03	28.3444	4.4608	3.8928	0.0002	0.0134	-0.0134
		50	0.03	51.7418	6.6655	6.6431	0.0000	0.0039	-0.0038
		100	0.03	53.8497	6.7719	6.7435	0.0000	0.0023	-0.0021
10%	50	10	0.03	46.8201	6.1324	6.0548	0.0004	0.0200	-0.0200
		50	0.03	54.4140	6.7668	6.7493	0.0002	0.0131	-0.0131
		100	0.03	57.1353	7.0198	6.9939	0.0001	0.0118	-0.0118
	100	10	0.03	41.2634	5.7137	5.6504	0.0004	0.0200	-0.0200
		50	0.03	55.3854	6.9503	6.9388	0.0002	0.0132	-0.0132
		100	0.03	57.6395	7.0651	7.0394	0.0001	0.0118	-0.0118
20%	50	10	0.03	48.3938	6.3672	6.3320	0.0005	0.0233	-0.0233
		50	0.03	56.7655	6.9988	6.9710	0.0003	0.0177	-0.0177
		100	0.03	59.1366	7.1283	7.1157	0.0003	0.0166	-0.0166
	100	10	0.03	47.0713	6.2470	6.1950	0.0005	0.0233	-0.0233
		50	0.03	56.1443	6.9860	6.9687	0.0003	0.0177	-0.0177
		100	0.03	60.0186	7.2270	7.2028	0.0003	0.0166	-0.0166

Um estudo simulado com 500 réplicas de Monte Carlo também foi feito para o modelo (3.3) com fragilidade Gama( $1/\alpha, 1/\alpha$ ) e risco de base de uma distribuição Exponencial( $\mu$ ). Nesse estudo observamos as estimativas de máxima verossimilhança em termo do EQM e do viés, variando os parâmetros  $\alpha$  e  $\beta$ , o tamanho da amostra e o percentual de censura. Os valores escolhidos para os parâmetros no estudo de simulação foram baseados nas estimativas obtidas na aplicação da [Subseção 3.3.1](#).

Tabela 8 – Valores dos parâmetros do modelo (3.3) sob diferentes cenários.

Parâmetros	Cenários			
	1	2	3	4
$\alpha$	10	10	50	50
$\beta$	0.01	-0.01	0.01	-0.01
$\mu$	0.03	0.03	0.03	0.03

De acordo com a [Tabela 9](#) observamos que em geral o EQM e o viés das estimativas para o parâmetro  $\alpha$  diminuem quando o tamanho da amostra aumenta. Quando o percentual de censura aumenta o EQM e o viés das estimativas para o parâmetro  $\alpha$  tendem a aumentar. Para as estimativas de  $\mu$  e  $\beta$  o EQM e o viés não se alteram muito quando o valor de  $n$  é maior, entretanto os valores do EQM e viés das estimativas para os parâmetros  $\mu$  e  $\beta$  crescem para os cenários de maior censura.

Tabela 9 – EQM e Viés dos estimadores do modelo (3.3) com fragilidade  $\text{Gama}(1/\alpha, 1/\alpha)$  e risco de base de uma distribuição Exponencial( $\mu$ ) para  $\mu = 0.03$  e diferentes valores de  $\alpha$ ,  $\beta$  e  $n$ , com percentual de censura (PC) variando de 0% a 20%.

PC	n	$\alpha$	$\beta$	$\hat{\alpha}$		$\hat{\mu}$		$\hat{\beta}$	
				EQM	Viés	EQM	Viés	EQM	Viés
0%	50	10	0.01	27.7622	4.4886	0.0000	0.0013	0.0002	-0.0107
		10	-0.01	26.6493	4.4556	0.0000	0.0009	0.0001	0.0086
		50	0.01	31.0242	4.9501	0.0002	0.0128	0.0001	-0.0096
		50	-0.01	32.0764	5.0356	0.0002	0.0129	0.0001	0.0065
	250	10	0.01	27.6014	4.5841	0.0000	0.0011	0.0001	-0.0103
		10	-0.01	28.1328	4.5483	0.0000	0.0010	0.0001	0.0089
		50	0.01	31.0817	4.8771	0.0002	0.0128	0.0001	-0.0093
		50	-0.01	31.2646	4.9521	0.0002	0.0128	0.0001	0.0067
20%	50	10	0.01	32.1474	5.0026	0.0006	-0.0236	0.0001	-0.0106
		10	-0.01	33.1303	5.1272	0.0006	-0.0238	0.0001	-0.0100
		50	0.01	31.4792	4.9797	0.0004	-0.0186	0.0001	-0.0099
		50	-0.01	31.1186	4.9788	0.0003	-0.0185	0.0001	0.0093
	250	10	0.01	28.3982	4.6860	0.0006	-0.0240	0.0001	-0.0105
		10	-0.01	31.8196	5.0948	0.0006	-0.0240	0.0001	0.0102
		50	0.01	31.0293	4.9556	0.0003	-0.0184	0.0001	-0.0098
		50	-0.01	31.0458	4.9244	0.0003	-0.0184	0.0001	0.0092

Em geral as estimativas para o parâmetro  $\beta$  acertam o sinal, entretanto com valores muito próximos a zero. Além disso, as probabilidades de cobertura não ficaram tão próximas dos valores nominais de 95% como esperado. Dessa forma há indícios de que o método clássico não é tão robusto para estimar os parâmetros do modelo (3.3). Uma alternativa seria uma abordagem bayesiana estudada por [Tomazella \(2003\)](#).

### 3.3 Aplicação para dados reais

#### 3.3.1 Aplicação para dados câncer

O conjunto de dados apresentados por [Gail, Santner e Brown \(1980\)](#) e disponível na [Tabela 27](#) é resultado de um estudo com 48 ratos livres de câncer. Os ratos receberam um composto carcinógeno e após 60 dias foram divididos aleatoriamente em dois grupos (Grupo

controle e Grupo tratamento). Um período de acompanhamento de 122 dias começou após a randomização, durante o qual eles foram examinados duas vezes por semana para avaliação do desenvolvimento de novos tumores (falha). Os ratos foram divididos em um grupo de 23 que recebeu o tratamento e em outro de 25 ratos que formaram o grupo controle. Nesse estudo a única covariável foi o tratamento (sim=1, não=-1). Os dados apresentados na [Tabela 27](#) são referentes aos tempos decorridos em dias até a ocorrência de novos tumores nos 48 ratos estudados.

Uma análise inicial pode ser feita através da [Figura 13](#), onde os ratos do grupo tratamento apresentam menos recorrências de tumores em comparação com os ratos do grupo controle que apresentam mais recorrências do evento de interesse durante o tempo de estudo, isso sugere que o tratamento tem contribuído para a diminuição no número de tumores nos ratos do estudo.

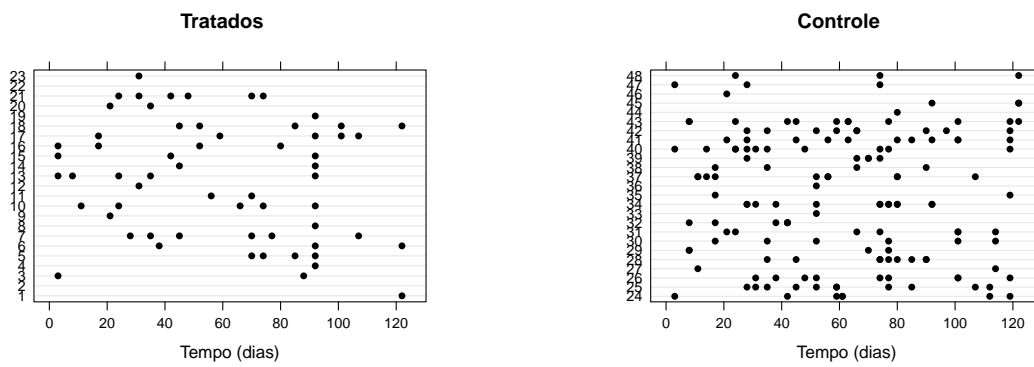


Figura 13 – Gráfico dos tempos recorrentes de detecção de tumores cancerígenos.

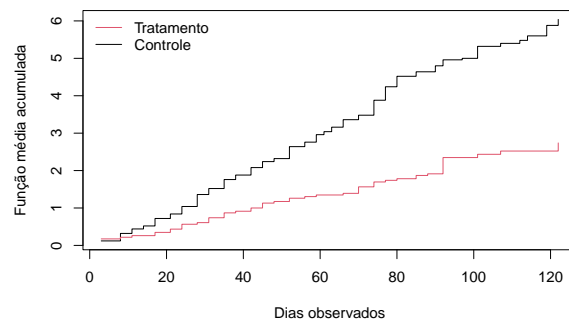


Figura 14 – Função média amostral acumulada para os grupos de ratos tratados e não tratados.

A [Figura 14](#) apresenta a função média amostral acumulada do grupo tratamento e do grupo controle ao longo do estudo. Notamos que a função média acumulada do grupo controle cresce de maneira linear ao longo do estudo, ou seja, a medida que o tempo passa as recorrências de tumores aumentam continuamente para o grupo que não recebeu tratamento. Já para o grupo tratado percebemos que o crescimento do número de tumores tende a estabilizar ao longo do tempo, dando indícios do efeito do tratamento.



Através da [Figura 15](#), que representa as curvas de sobrevivência obtidas pelo estimador não paramétrico para dados de eventos recorrentes proposto por [Wang e Chang \(1999\)](#) e calculado utilizando o pacote `newTestSurvRec` ([MANRIQUE, 2018](#)) do *software* `R-Development-Core-Team (2023)` para cada grupo, percebemos que o grupo tratado parece ter uma maior sobrevida, o que parece está relacionado com a quantidade de tumores detectados em cada grupo.

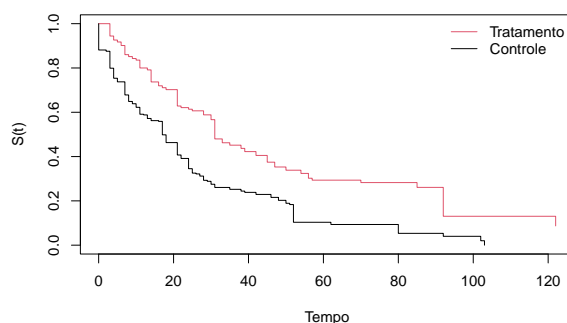


Figura 15 – Curvas de sobrevivência estimadas para os grupos de ratos tratados e não tratados.

Para ajustar os dados da [Tabela 27](#) utilizamos inicialmente o modelo aditivo com fragilidade Gama( $1/\alpha, 1/\alpha$ ) com função de risco de base de uma distribuição Weibull( $\mu, \gamma$ ). Através da maximização da função de log verossimilhança dada na equação (3.13) obtivemos as seguintes estimativas de máxima verossimilhança para os parâmetros do modelo aditivo com fragilidade Gama( $1/\alpha, 1/\alpha$ ) e distribuição Weibull( $\mu, \gamma$ ) para o risco de base.

Tabela 10 – Estimativas de máxima verossimilhança, erros padrão, limite inferior (LI) e limite superior (LS) dos intervalos de confiança de 95% para os parâmetros do modelo de fragilidade aditivo Gama-Weibull.

Parâmetros	EMV	EP	LI (95%)	LS (95%)
$\beta$	-0.014	0.0025	-0.0179	-0.0098
$\alpha$	82.19	43.32	10.71	153.67
$\mu$	0.028	0.0091	0.0125	0.0427
$\gamma$	1.051	0.0673	0.9397	1.1617

De acordo com os resultados da [Tabela 10](#) podemos concluir que a covariável tratamento é significativa pois o intervalo de confiança  $[-0.0179; -0.0098]$  para  $\beta$  não contém zero, ou seja, o risco de desenvolver novos tumores diminui para os ratos que estão no grupo tratado ( $\hat{\beta} = -0.014$ ). Além disso há heterogeneidade entre os ratos ( $\hat{\alpha} = 82.19$ ) e uma dependência entre as recorrências de tumores em cada rato. Como o intervalo de confiança para o parâmetro  $\gamma$  inclui o valor 1, então há uma indicação de que podemos utilizar a distribuição exponencial para o risco de base.

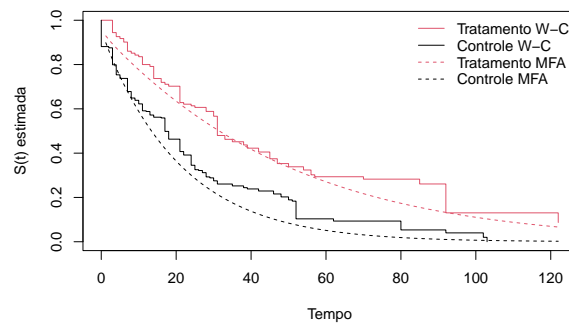


Figura 16 – Comparação das curvas de sobrevivências estimadas pelo modelo aditivo com fragilidade Gama( $1/\alpha$ ,  $1/\alpha$ ) e risco de base da distribuição Weibull( $\mu$ ,  $\gamma$ ) com as curvas estimadas pelo estimador não paramétrico para dados recorrentes.

Utilizando o estimativas apresentadas na Tabela 10 obtivemos as curvas de sobrevivência estimadas expressas na Figura 16, em que W-C são as curvas estimadas pelo estimador não paramétrico de Wang e Chang (1999) e MFA são as funções de sobrevivência estimadas pelo modelo de fragilidade aditivo para dados de eventos recorrentes assumindo o risco base de uma distribuição Weibull( $\mu$ ,  $\gamma$ ). Pela Figura 16 pode-se perceber que as curvas de sobrevivência estimadas pelo modelo aditivo para os dados recorrentes de tumores em ratos, com risco de base de uma distribuição Weibull, ficam próximas das curvas estimadas pelo estimador não paramétrico para a sobrevivência de dados recorrentes.

Utilizando agora o risco de base de uma distribuição Exponencial( $\mu$ ) e a mesma distribuição Gama( $1/\alpha$ ,  $1/\alpha$ ) para a fragilidade, obtivemos as seguintes estimativas de MV para os parâmetros  $\alpha$ ,  $\beta$  e  $\mu$  do modelo maximizando (3.14) numericamente. A partir dos resultados da Tabela 11 podemos concluir que a covariável tratamento é significativa pois o intervalo de confiança  $[-0.0180; -0.0099]$  para o parâmetro  $\beta$  não contém zero, ou seja, o risco de desenvolver novos tumores diminui para os ratos que estão no grupo tratado ( $\hat{\beta} = -0.014$ ).

Tabela 11 – Estimativas de máxima verossimilhança, erros padrão, limite inferior (LI) e limite superior (LS) dos intervalos de confiança de 95% para os parâmetros do modelo de fragilidade aditivo Gama-Exponencial.

Parâmetros	EMV	EP	LI (95%)	LS (95%)
$\beta$	-0.014	0.0024	-0.0180	-0.0099
$\alpha$	68.79	31.91	16.14	121.4
$\mu$	0.035	0.0025	0.0311	0.0392

Através do valor de  $\hat{\alpha} = 68.79$  observamos que há heterogeneidade entre os ratos, ou seja, existem fatores não observados que podem influenciar no tempo de vida dos ratos e há ainda uma dependência entre as ocorrências de tumores nos ratos, isto é, a medida que novos tumores surgem o rato fica mais frágil.

Aplicando o teste da razão de verossimilhança (TRV) para comparar o modelo aditivo com fragilidade  $Gama(1/\alpha, 1/\alpha)$  e risco de base da distribuição Weibull( $\mu, \gamma$ ) (modelo maior) com o modelo aditivo com fragilidade  $Gama(1/\alpha, 1/\alpha)$  e risco de base da distribuição Exponencial( $\mu$ ) (modelo menor) obtivemos os seguintes resultados:

Tabela 12 – Teste da Razão de verossimilhança para comparar os modelos.

Teste	Estatística de teste	Graus de liberdade	p-valor
TRV	1.5134	1	0.2186

Portanto pelo p-valor dado na [Tabela 12](#) não rejeitamos a hipótese nula de que o modelo menor é mais adequado a um nível de 5% de significância. Dessa forma iremos utilizar o modelo com função de risco de base de uma distribuição Exponencial( $\mu$ ) cuja estimativas de máxima verossimilhança estão na [Tabela 11](#).

Na [Figura 17](#) percebe-se que as curvas de sobrevivência estimadas pelo modelo aditivo para os dados recorrentes de tumores em ratos, com risco de base de uma distribuição Exponencial, também ficam próximas das curvas estimadas pelo estimador não paramétrico para a sobrevivência de dados recorrentes, assim como na [Figura 18](#).

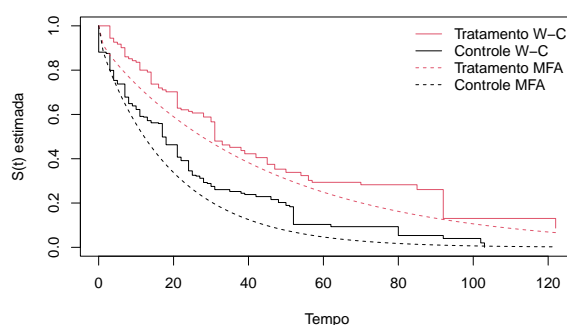


Figura 17 – Comparação das curvas de sobrevivências estimadas pelo modelo aditivo com fragilidade  $Gama(1/\alpha, 1/\alpha)$  e risco de base da distribuição Exponencial( $\mu$ ) com as curvas estimadas pelo estimador não paramétrico para dados recorrentes.

Calculando as fragilidades individuais a partir da [Equação 3.15](#) utilizando as estimativas da [Tabela 11](#) obtemos os resultados da [Tabela 13](#). Para os dados da [Tabela 27](#) analisados percebemos que quanto maior o número de recorrências (presença de tumores) maior a fragilidade do rato, o que era esperado devido a natureza do evento em estudo, ou seja, quanto mais tumores no rato mais frágil ele fica e portanto maior o risco de desenvolvimento de novos tumores.

Tabela 13 – Fragilidades individuais dos 48 ratos e seus respectivos numeros de recorrências divididos por grupo.

$i$	$d_i$ (Grupo Tratado)	$\hat{v}_i$	$i$	$d_i$ (Grupo Controle)	$\hat{v}_i$
1	1	0.0083	24	7	0.0575
2	0	0.0001	25	11	0.0903
3	2	0.0165	26	9	0.0739
4	1	0.0083	27	2	0.0165
5	4	0.0329	28	9	0.0739
6	3	0.0247	29	4	0.0329
7	6	0.0493	30	6	0.0493
8	1	0.0083	31	7	0.0575
9	1	0.0083	32	6	0.0493
10	5	0.0411	33	1	0.0083
11	2	0.0165	34	13	0.1067
12	1	0.0083	35	2	0.0165
13	5	0.0411	36	1	0.0083
14	2	0.0165	37	10	0.0821
15	3	0.0247	38	4	0.0329
16	4	0.0329	39	5	0.0411
17	5	0.0411	40	11	0.0903
18	5	0.0411	41	11	0.0903
19	1	0.0083	42	9	0.0739
20	2	0.0165	43	12	0.0985
21	6	0.0493	44	1	0.0083
22	0	0.0001	45	3	0.0247
23	1	0.0083	46	1	0.0083
-	-	-	47	3	0.0247
-	-	-	48	3	0.0247

Ainda pela [Tabela 13](#) percebemos que os ratos não tratados (Grupo Controle) são em geral mais frágeis do que os ratos do grupo que recebeu tratamento, o que também era esperado devido ao valor negativo da estimativa do parâmetro  $\beta$ . O rato que apresentou mais recorrências de tumores (rato 34) foi o que apresentou maior fragilidade dentre todos os ratos do estudo e os menos frágeis foram exatamente os ratos 2 e 22.

Entretanto utilizando a parametrização  $\text{Gama}(1/\alpha, 1/\alpha)$  a média das fragilidades individuais ([Tabela 13](#)) não se aproxima da média assumida para variável de fragilidade ( $E(V) = 1$ ). Dessa forma assumimos outra parametrização para a variável de fragilidade de tal forma que não definimos um valor específico para  $E(V)$ , permitindo que os dados estimem esse valor.

### 3.3.2 Outra parametrização para a fragilidade Gama

Assumindo a distribuição  $\text{Gama}(a, b)$  para a variável de fragilidade do modelo (3.3) temos que a transformada de Laplace é dada por  $L(t) = \left(\frac{b}{t+b}\right)^a$  e a partir de (3.4) temos a

seguinte função de intensidade marginal:

$$\lambda_i(t|\mathbf{x}_i) = \lambda_0(t) + \mathbf{x}'_i\boldsymbol{\beta} + \frac{a}{t+b} \quad (3.18)$$

e a função de sobrevivência é expressa como:

$$\begin{aligned} S(t|\mathbf{x}_i) &= \exp[-\Lambda(t|\mathbf{x}_i)] = \exp\left[-\int_0^t \lambda_i(u|\mathbf{x}_i)du\right] \\ &= \exp[-\mathbf{x}'_i\boldsymbol{\beta}t - \Lambda_0(t)] \left(\frac{b}{t+b}\right)^a. \end{aligned}$$

A partir da [Equação 3.5](#) obtemos a seguinte função de verossimilhança:

$$\begin{aligned} L(\boldsymbol{\theta}|\mathcal{D}) &= \prod_{i=1}^n \prod_{j=1}^{m_i} \left[ \lambda_0(t_{ij}) + \mathbf{x}'_i\boldsymbol{\beta} + \frac{a}{t_{ij}+b} \right]^{\delta_{ij}} \times \\ &\quad \prod_{i=1}^n \exp[-\mathbf{x}'_i\boldsymbol{\beta}T_i - \Lambda_0(T_i)] \left(\frac{b}{T_i+b}\right)^a, \end{aligned} \quad (3.19)$$

em que  $\boldsymbol{\theta}$  é o vetor de parâmetros do modelo e  $\mathcal{D} = \{(t_{ij}, \mathbf{x}_i, \delta_{ij}), i = 1, \dots, n, j = 1, \dots, m_i\}$ . A função de log verossimilhança correspondente é dada por:

$$\begin{aligned} l(\boldsymbol{\theta}|\mathcal{D}) &= \sum_{i=1}^n \sum_{j=1}^{m_i} \delta_{ij} \log \left( \lambda_0(t_{ij}) + \mathbf{x}'_i\boldsymbol{\beta} + \frac{a}{t_{ij}+b} \right) \\ &\quad + \sum_{i=1}^n (-\Lambda_0(T_i) - \mathbf{x}'_i\boldsymbol{\beta}T_i) + a \log \left( \frac{b}{T_i+b} \right), \end{aligned} \quad (3.20)$$

em que  $\delta_{ij} = 1$  se  $t_{ij}$  é observado e  $\delta_{ij} = 0$  se  $t_{ij}$  é censurado. Para completar a modelagem paramétrica devemos escolher uma distribuição para o risco de base e a partir da [Equação 3.20](#) podemos obter as estimativas de máxima verossimilhança numericamente.

Utilizando o resultado apresentado em [Munda, Rotolo e Legrand \(2012\)](#) temos que para fragilidade  $V \sim \text{Gama}(a, b)$ :

$$L^{(d_i)}(T_i) = (-1)^{d_i} (T_i + b)^{-d_i} \left[ \prod_{k=0}^{d_i-1} (a+k) \right] \left( \frac{b}{T_i+b} \right)^a$$

e

$$L^{(d_i+1)}(T_i) = (-1)^{d_i+1} (T_i + b)^{-(d_i+1)} \left[ \prod_{k=0}^{d_i} (a+k) \right] \left( \frac{b}{T_i+b} \right)^a,$$

para  $d_i \geq 1$ . Dessa forma podemos estimar a fragilidade individual para dados de eventos recorrentes a partir da [Equação 3.8](#) para o modelo aditivo paramétrico com fragilidade  $\text{Gama}(a, b)$  pela seguinte expressão

$$\hat{v}_i = (T_i + \hat{b})^{-1} (\hat{a} + d_i) \quad (3.21)$$

Nesse trabalho consideramos que o tempo de truncamento  $T_i$  é o mesmo para todos indivíduos e ainda que os indivíduos só entram no estudo após apresentarem a primeira falha.

Se o indivíduo  $i$  não apresenta tempos recorrentes então  $d_i = 0$ , ou seja, podemos estimar sua fragilidade individual por  $\hat{v}_i = \frac{\hat{a}}{T_i + \hat{b}}$ , que coincide com o caso univariado do [Capítulo 2](#).

Para ajustar os dados da [Tabela 27](#) utilizamos inicialmente o modelo aditivo com fragilidade Gama( $a, b$ ) com função de risco de base de uma distribuição Weibull( $\mu, \gamma$ ), temos que de acordo com (3.19) a função de verossimilhança é dada por:

$$L(\mu, \gamma, a, b, \beta | \mathcal{D}) = \prod_{i=1}^n \prod_{j=1}^{m_i} [\gamma \mu^\gamma t_{ij}^{\gamma-1} + \mathbf{x}'_i \beta + a/(t_{ij} + b)]^{\delta_{ij}} \times \prod_{i=1}^n \exp[-\mathbf{x}'_i \beta T_i - (\mu T_i)^\gamma] \left( \frac{b}{T_i + b} \right)^a$$

A função de log verossimilhança correspondente é dada por:

$$l(\mu, \gamma, a, b, \beta | \mathcal{D}) = \sum_{i=1}^n \sum_{j=1}^{m_i} \delta_{ij} \log \left( \gamma \mu^\gamma t_{ij}^{\gamma-1} + \mathbf{x}'_i \beta + a/(t_{ij} + b) \right) + \sum_{i=1}^n \left( -(\mu T_i)^\gamma - \mathbf{x}'_i \beta T_i \right) + a \log \left( \frac{b}{T_i + b} \right) \quad (3.22)$$

Para um risco de base constante de uma distribuição Exponencial( $\mu$ ) temos:

$$L(\mu, a, b, \beta | \mathcal{D}) = \prod_{i=1}^n \prod_{j=1}^{m_i} [\mu + \mathbf{x}'_i \beta + a/(t_{ij} + b)]^{\delta_{ij}} \times \prod_{i=1}^n \exp[-\mathbf{x}'_i \beta T_i - \mu T_i] \left( \frac{b}{T_i + b} \right)^a$$

A função de log verossimilhança correspondente é dada por:

$$l(\mu, a, b, \beta | \mathcal{D}) = \sum_{i=1}^n \sum_{j=1}^{m_i} \delta_{ij} \log \left( \mu + \mathbf{x}'_i \beta + a/(t_{ij} + b) \right) + \sum_{i=1}^n \left( -\mu T_i - \mathbf{x}'_i \beta T_i \right) + a \log \left( \frac{b}{T_i + b} \right) \quad (3.23)$$

A partir dos dados da [Tabela 27](#) obtivemos as estimativas de MV para os parâmetros do modelo aditivo com fragilidade Gama( $a, b$ ) e riscos de base das distribuições Exponencial( $\mu$ ) e Weibull( $\mu, \gamma$ ), respectivamente. De acordo com os resultados da [Tabela 14](#) podemos concluir que o modelo Gama-Exponencial é mais adequado pelos critérios AIC e BIC, além disso o intervalo de confiança para o parâmetro  $\gamma$  inclui o valor 1, então há uma indicação de que podemos utilizar a distribuição exponencial para o risco de base. A covariável tratamento é significativa pois o intervalo de confiança  $[-0.0215; -0.0117]$  para  $\beta$  não contém zero, ou seja, o risco de desenvolver novos tumores diminui para os ratos que estão no grupo tratado ( $\hat{\beta} = -0.0166$ ). Além disso o valor esperado estimado para a variável de fragilidade é de ( $\hat{a}/\hat{b} = 0.0239$ ) e a varância estimada da variável de fragilidade é ( $\hat{a}/\hat{b}^2 = 0.0127$ ).

Tabela 14 – Estimativas de máxima verossimilhança, erros padrão, intervalos de confiança de 95%, AIC e BIC para os modelos aditivo com fragilidade Gama e riscos de base da distribuição Exponencial( $\mu$ ) e da distribuição Weibull( $\mu, \gamma$ ) para os dados de câncer em ratos.

Parametros	MFA-Gama-Exponencial			MFA-Gama-Weibull		
	EMV	EP	IC	EMV	EP	IC
$\beta$	-0.0166	0.0025	(-0.0215; -0.0117)	-0.0070	0.0024	(-0.0117; -0.0023)
$a$	0.0448	0.0449	(0.0000; 0.1328)	0.0383	0.0605	(0.0000; 0.1569)
$b$	1.8764	0.7441	(0.4180; 3.3348)	2.2867	0.9153	(0.4928; 4.0806)
$\mu$	0.0355	0.0028	( 0.0300; 0.0410)	0.0283	0.0046	(0.0192 ; 0.0373)
$\gamma$	-	-	-	1.1142	0.1029	(0.9124; 1.3159)
AIC		1810.58			1817.79	
BIC		1818.06			1827.15	

Utilizando as estimativas do modelo apresentadas na [Tabela 14](#) obtivemos as curvas de sobrevivência estimadas expressas na [Figura 18](#), em que W-C são as curvas estimadas pelo estimador não paramétrico de [Wang e Chang \(1999\)](#) e MFA são as funções de sobrevivência estimadas pelo modelo aditivo dado na [Equação 3.19](#) aplicando o risco acumulado de base da distribuição Exponencial( $\mu$ ). Pela [Figura 18](#) pode-se perceber que as curvas de sobrevivência estimadas pelo modelo aditivo para os dados recorrentes de tumores em ratos, com risco de base de uma distribuição Exponencial, ficam próximas das curvas estimadas pelo estimador não paramétrico para a sobrevivência de dados recorrentes.

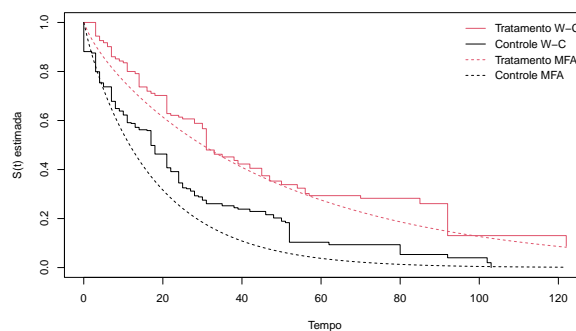


Figura 18 – Comparação das curvas de sobrevivências estimadas pelo modelo aditivo com fragilidade Gama( $a, b$ ) e risco de base da distribuição Exponencial( $\mu$ ) com as curvas estimadas pelo estimador não paramétrico para dados recorrentes.

Com base no estimador proposto para as fragilidades individuais ([Equação 3.21](#)) obtivemos os resultados da [Tabela 15](#), utilizando as estimativas para os parâmetros do modelo MFA-Gama-Exponencial da [Tabela 14](#). Note que os valores nominais das fragilidades individuais estimadas na [Tabela 15](#) são similares as estimativas da [Tabela 13](#), entretanto a média das fragilidades individuais estimadas pelo modelo com  $V \sim Gama(a, b)$  fica próxima ao valor de  $\hat{E}(V) = 0.0239$ , o que faz mais sentido.

Tabela 15 – Fragilidades individuais dos 48 ratos e seus respectivos numeros de recorrências divididos por grupo.

$i$	$d_i$ (Grupo Tratado)	$\hat{v}_i$	$i$	$d_i$ (Grupo Controle)	$\hat{v}_i$
1	1	0.0087	24	7	0.0587
2	0	0.0004	25	11	0.0920
3	2	0.0170	26	9	0.0753
4	1	0.0087	27	2	0.0170
5	4	0.0337	28	9	0.0753
6	3	0.0254	29	4	0.0337
7	6	0.0503	30	6	0.0503
8	1	0.0087	31	7	0.0587
9	1	0.0087	32	6	0.0503
10	5	0.0420	33	1	0.0087
11	2	0.0170	34	13	0.1086
12	1	0.0087	35	2	0.0170
13	5	0.0420	36	1	0.0087
14	2	0.0170	37	10	0.0836
15	3	0.0254	38	4	0.0337
16	4	0.0337	39	5	0.0420
17	5	0.0420	40	11	0.0920
18	5	0.0420	41	11	0.0920
19	1	0.0087	42	9	0.0753
20	2	0.0170	43	12	0.1003
21	6	0.0503	44	1	0.0087
22	0	0.0004	45	3	0.0254
23	1	0.0087	46	1	0.0087
-	-	-	47	3	0.0254
-	-	-	48	3	0.0254

### 3.4 Conclusão

O modelo aditivo para dados de eventos recorrentes estudado nesse capítulo é uma alternativa aos modelos de risco multiplicativo. Para eventos recorrentes a suposição de dependência entre os tempos do mesmo indivíduo é razoável e faz sentido inserir o termo de fragilidade para medir essa dependência, bem como a heterogeneidade não observada.

Para os dados disponíveis na [Tabela 27](#) estudados nesse trabalho ajustamos um modelo de fragilidade Gama com risco de base exponencial, ou seja, utilizamos um modelo simples e conseguimos um ajuste relativamente próximo as curvas de sobrevivência estimadas pelo estimador não paramétrico para dados de eventos recorrentes. Pela modelagem aditiva com fragilidade concluímos que o tratamento aplicado nos ratos diminuiu a intensidade do processo cancerígeno, ou seja, diminuiu o aparecimento de novos tumores nos ratos tratados. Ainda através do modelo utilizado concluímos que existem fatores não observados no estudo que influenciam no tempo de vida dos ratos e que os tempos recorrentes são dependentes, ou seja, quanto mais frágil o rato está, maior a intensidade do processo cancerígeno, confirmando a importância da



inclusão do termo de fragilidade ao modelo.

Como esperado, observamos no estudo simulado que à medida que o percentual de censura aumenta, as estimativas de máxima verossimilhança para os parâmetros do modelo aditivo com fragilidade pioram, tanto em termos de EQM e EAM como em termos de Viés. Ainda pelos resultados numéricos via simulação de Monte Carlo verificamos que para valores grandes do parâmetro de fragilidade  $\alpha$ , o método de estimação clássico apresenta dificuldades para estimar os parâmetros do modelo aditivo.

O estimador proposto para calcular as fragilidades individuais dos pacientes a partir do modelo estudado nessa capítulo mostrou-se bem intuitivo, no sentido que atribui maior fragilidade para os ratos que apresentaram maior recorrência do evento de interesse, visto que na aplicação desse capítulo os eventos de interesse debilitavam mais os animais, ou seja, quanto maior a intensidade do processo mais frágil é o rato. Ainda em relação a estimação das fragilidades individuais percebemos que nas aplicações utilizadas nesse trabalho fez mais sentido a parametrização Gama( $a, b$ ) para a variável de fragilidade, pois nesse caso a média das fragilidades individuais estimadas ficaram próximas a  $\hat{E}(V)$ .

---

## MODELAGEM DE RISCO ADITIVO COM FRAGILIDADE E FRAÇÃO DE CURA

---

---

Com o avanço da medicina um número cada vez maior de pacientes passou a experimentar, através de tratamentos mais eficazes, maior tempo de vida e até mesmo a cura de doenças antes consideradas altamente letais. Em análise de sobrevivência quando a função de sobrevivência não vai para zero (imprópria), depois de um longo período de acompanhamento, geralmente são utilizados os conhecidos modelos de longa duração, também chamados de modelos de sobrevivência com fração de cura. Esses modelos permitem estimar a proporção de curados em uma determinada população diagnosticada com uma doença, por exemplo. [Boag \(1949\)](#) e [Berkson e Gage \(1952\)](#) propuseram o modelo de mistura padrão, que consiste de uma mistura de distribuições paramétricas, sendo uma degenerada na fração de curados e a outra poderada por uma função de sobrevivência própria para a parte da população formada pelos não curados.

Incorporar ao modelo de longa duração um efeito aleatório que capta a heterogeneidade não observada pelas covariáveis comuns aos indivíduos é uma maneira de incluir no modelo duas fontes de variação, apesar dos modelos de sobrevivência com fração de cura assumirem homogeneidade para os indivíduos que não foram curados, ou seja, para os que apresentaram o evento de interesse no estudo, é razoável esperar que exista uma certa heterogeneidade não observada influenciada por fatores de risco não presentes no estudo.

[Longini e Halloran \(1996\)](#) propuseram o modelo de mistura padrão com fragilidade e [Price e Manatunaga \(2001\)](#) estudaram o modelo de mistura padrão com diferentes distribuições para a fragilidade. [Gonzales, Tomazella e Taconelli \(2013\)](#) apresentaram uma abordagem paramétrica para o modelo de mistura padrão com fragilidade Gama incluída de forma multiplicativa na função de risco de base na presença de covariáveis.

Nesse trabalho iremos propor uma modelagem paramétrica considerando o modelo de mistura padrão com fragilidade aditiva. Nesta proposta é assumido uma distribuição Gama para

a variável de fragilidade do modelo proposto e o modelo de fração de cura considerado será o modelo de mistura padrão. A inclusão de covariáveis ao modelo será feita de duas maneiras: somente no termo da fração de cura ou no termo da fração de cura e também no termo da função de sobrevivência.

## 4.1 Modelagem

### 4.1.1 Modelo de mistura padrão

O modelo de mistura padrão proposto por [Berkson e Gage \(1952\)](#) é um dos modelos mais comum na análise de sobrevivência para ajustar dados de longa duração. Considerando  $T$  uma variável aleatória não negativa e contínua, representando o tempo de vida, sabemos que

$$P(T > t | M_i = 1) = S(t) \text{ e } P(T > t | M_i = 0) = 1, \quad (4.1)$$

em que,  $M_i$  é uma variável não observável de Bernoulli, tal que  $M_i = 0$  se o indivíduo  $i$  não está em risco e  $M_i = 1$  se o indivíduo  $i$  está em risco, com  $i = 1, \dots, n$ . Assumindo que  $P(M_i = 0) = p_0$  e  $P(M_i = 1) = 1 - p_0$ , sendo  $p_0 \in (0, 1)$  a proporção de curados ou imunes, a probabilidade de o tempo de vida ser maior que um determinado tempo  $t$ , independente do grupo a que ele pertença é dada por

$$S_{pop}(t) = P(T > t) = P(T > t | M_i = 0)P(M_i = 0) + P(T > t | M_i = 1)P(M_i = 1) \quad t \geq 0.$$

Assim, a função de sobrevivência populacional é dada por

$$S_{pop}(t) = p_0 + (1 - p_0)S(t), \quad t \geq 0. \quad (4.2)$$

em que  $S(\cdot)$  representa a função de sobrevivência própria associada aos indivíduos em risco.

A função de sobrevivência  $S_{pop}(t)$  possui as seguintes propriedades:

- Se  $p_0 = 0$ , então  $S_{pop}(t) = S(t)$ ;
- $S_{pop}(0) = 1$ ;
- $S_{pop}(t)$  é decrescente;
- $\lim_{t \rightarrow \infty} S_{pop}(t) = p_0$

A última propriedade retrata o fato de a função de sobrevivência populacional ser imprópria, pois a curva de sobrevivência estabiliza em  $p_0$ , justamente a probabilidade de cura da população. Na [Figura 19](#) apresentamos de maneira esquemática a curva de sobrevivência do modelo com fração de cura, onde  $p_0$  indica a proporção de não susceptíveis ao evento de interesse.

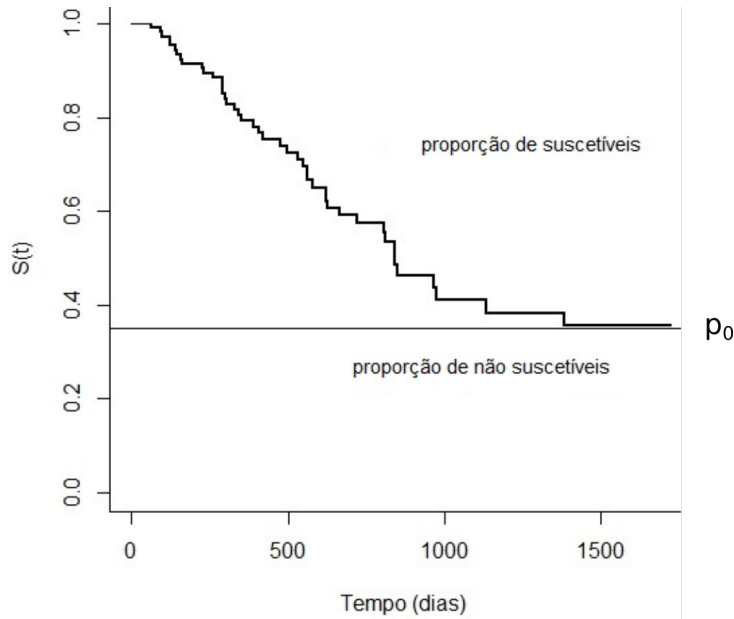


Figura 19 – Curva de sobrevivência do modelo (4.2).

#### 4.1.2 Modelo de mistura padrão com fragilidade aditiva (MMPFA)

Para modelo de fragilidade aditivo geral expresso na (2.3) podemos verificar que quando o tempo é muito grande a função de sobrevivência não condicional vai para zero, ou seja, o modelo estudado no Capítulo 2 não capta a fração de cura. Dessa forma é interessante estudarmos um modelo aditivo que possa estimar a proporção de curados, quando existir. De maneira geral podemos escrever o modelo de mistura padrão com fragilidade aditiva a partir da Equação 4.2 substituindo  $S(t)$  pela Equação 2.9, ou seja, o modelo de mistura padrão com fragilidade aditiva estudado nesse trabalho segue a seguinte equação:

$$S_{pop}(t) = p_0 + (1 - p_0) \exp\{-\Lambda_0(t)\}L(t) \quad (4.3)$$

em que  $\Lambda_0(t)$  é a função de risco base acumulada, que depende da escolha de distribuições paramétricas tais como: Exponencial, Weibull, Gompertz entre outras, e  $L(t)$  é a transformada de Laplace que depende da escolha para a distribuição da variável de fragilidade.

Pela relação da função de densidade de probabilidade com a função de sobrevivência temos que  $f(t) = -\frac{d}{dt}S(t)$ , logo da equação (4.2) obtemos que  $f_{pop}(t) = -\frac{d}{dt}S_{pop}(t) = (1 - p_0)f(t)$ , em que  $f(t)$  é a função densidade de probabilidade dos indivíduos susceptíveis (em risco). Aplicando esse resultado podemos reescrever  $f_{pop}(t)$  da seguinte forma:

$$f_{pop}(t) = (1 - p_0)f(t) = (1 - p_0)\lambda(t)S(t) \quad t \geq 0. \quad (4.4)$$

Para o modelo de fragilidade aditivo dado em (2.3) temos que (4.4) pode ser reescrita da seguinte forma

$$f_{pop}(t) = (1 - p_0)f(t) = (1 - p_0) \left( \lambda_0(t) - \frac{L'(t)}{L(t)} \right) \exp\{-\Lambda_0(t)\}L(t) \quad t \geq 0. \quad (4.5)$$

### 4.1.3 Inferência

Para determinar a função de verossimilhança, consideremos para a  $i$ -ésima observação os dados observados  $\mathcal{D} = (t_i, \delta_i)$ , em que  $t_i$  denota o tempo observado,  $\delta_i$  é a variável indicadora de censura, com  $\delta_i = 1$  se  $t_i$  é não censurado e  $\delta_i = 0$  caso contrário. Assim, a função de verossimilhança baseada nos dados observados supondo censura não informativa a partir de (4.5) e (4.3) é dada por

$$\begin{aligned}
 L(\theta|\mathcal{D}) &= \prod_{i=1}^n [f_{pop}(t_i)]^{\delta_i} [S_{pop}(t_i)]^{1-\delta_i} \\
 &= \prod_{i=1}^n [(1-p_0)\lambda(t_i)S(t_i)]^{\delta_i} \times [p_0 + (1-p_0)S(t_i)]^{1-\delta_i} \\
 &= \prod_{i=1}^n \left[ (1-p_0) \left( \lambda_0(t_i) - \frac{L'(t_i)}{L(t_i)} \right) \exp\{-\Lambda_0(t_i)\} L(t_i) \right]^{\delta_i} \\
 &\quad \times [p_0 + (1-p_0) \exp\{-\Lambda_0(t_i)\} L(t_i)]^{1-\delta_i}.
 \end{aligned} \tag{4.6}$$

A estimação do vetor de parâmetros  $\theta$  do modelo será feita através da maximização do logaritmo da função de verossimilhança,  $l(\theta|\mathcal{D}) = \log(L(\theta|\mathcal{D}))$ . Para o parâmetro da proporção de cura ( $p_0$ ) iremos utilizar a função logito para fazer a ligação com o vetor de covariáveis e para estimá-lo basta substituir os parâmetros da regressão pelas estimativas obtidas na função de ligação utilizada. Os erros padrão das estimativas para a fração de cura serão obtidos através do método delta.

### 4.1.4 Modelo de mistura padrão com fragilidade aditiva Gama

Assumindo que a variável da fragilidade segue uma distribuição  $Gama(a, b)$  e substituindo a sua respectiva transformada de Laplace no modelo de mistura padrão com fragilidade (4.3) temos que o modelo de mistura padrão com fragilidade aditiva  $Gama(a, b)$  é dado por:

$$S_{pop}(t) = p_0 + (1-p_0) \exp\{-\Lambda_0(t)\} \left( \frac{b}{t+b} \right)^a \quad t \geq 0. \tag{4.7}$$

em que  $p_0$  é a proporção de curados na população e  $\Lambda_0(t)$  é a função de risco base acumulada, que pode assumir diferentes formas tais como:  $\Lambda_0(t) = \mu t$ ,  $\Lambda_0(t) = \mu t^\gamma$  e  $\Lambda_0(t) = \lambda \exp(\gamma t)$  para as distribuições Exponencial, Weibull e Gompertz, respectivamente, dentre outras.

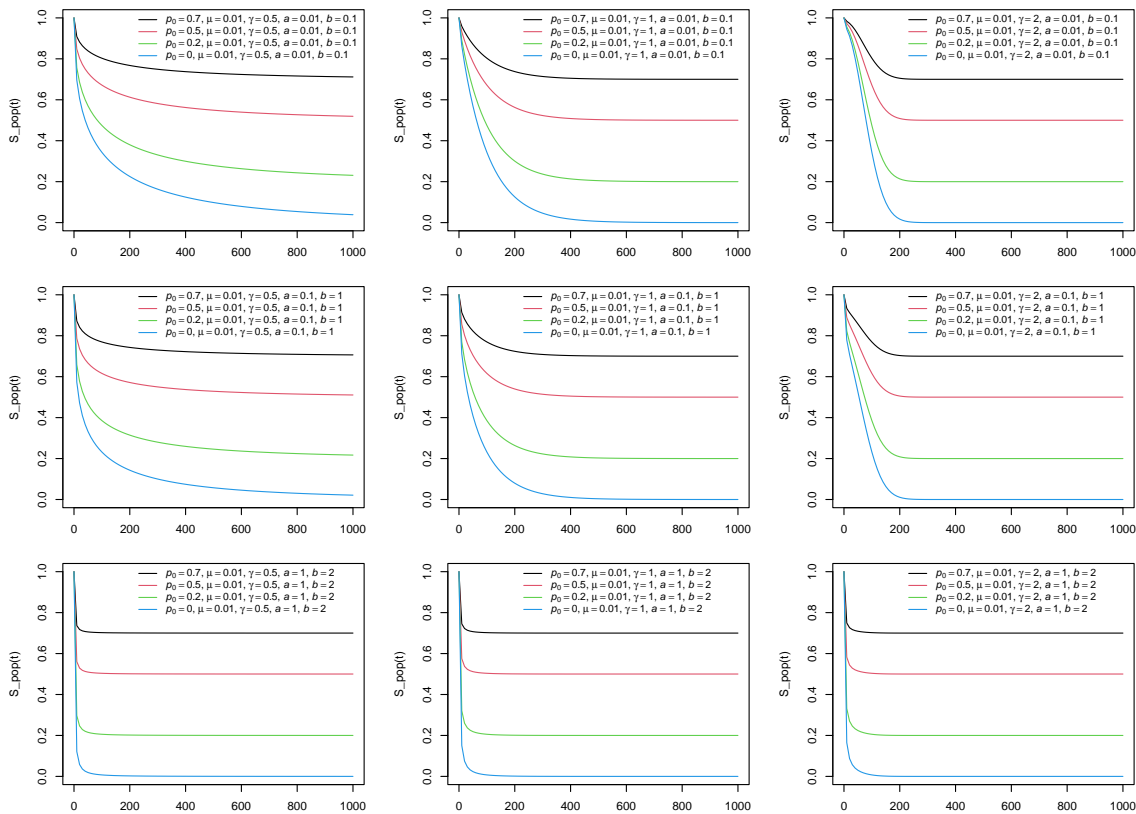


Figura 20 – Curvas de sobrevivência do modelo (4.7) com risco de base de uma distribuição Weibull( $\mu, \gamma$ ) para diferentes valores dos parâmetros  $a, b, \mu$  e  $\gamma$  versus tempo.

Na Figura 20 percebemos que a medida que o parâmetro  $\gamma$  da função de risco de base da distribuição Weibull cresce a função de sobrevivência decai mais rapidamente para a proporção de curados. Para os modelos com risco de base da distribuição Weibull( $\mu, \gamma$ ) as funções de sobrevivência são mais flexíveis pois apresentam mais um parâmetro, como vemos na Figura 20.

A função de verossimilhança do modelo de mistura padrão com fragilidade aditivo  $Gama(a, b)$  é expressa pela seguinte equação

$$L(\theta|\mathcal{D}) = \prod_{i=1}^n \left[ (1-p_0) \left( \lambda_0(t_i) + \frac{a}{t_i + b} \right) \exp\{-\Lambda_0(t_i)\} \left( \frac{b}{t_i + b} \right)^a \right]^{\delta_i} \times \left[ p_0 + (1-p_0) \exp\{-\Lambda_0(t_i)\} \left( \frac{b}{t_i + b} \right)^a \right]^{1-\delta_i} \quad (4.8)$$

Para completar a modelagem paramétrica é necessário assumir uma distribuição para o tempo de sobrevivência dos indivíduos em risco e a partir dessa suposição obter uma forma paramétrica para a função de risco de base ( $\lambda_0(t)$ ). Algumas distribuições podem ser utilizadas para essa finalidade, por exemplo: Exponencial, Weibull, Gompertz, dentre outras.

### 4.1.5 Simulação

Neste estudo foram gerados os tempos do modelo de fragilidade aditiva com distribuição de fragilidade gama( $a, b$ ) e uma taxa de risco basal exponencial com  $\mu = 0.03$  pelo método de inversão. Avaliamos o comportamento das estimativas de máxima verossimilhança para  $p_0 = 0,2$  e  $p_0 = 0.5$  e para diferentes valores de  $a$  e  $b$  com tamanhos amostrais  $n = 100, n = 300$  e  $n = 500$  e taxas de censura de 30% a 60%. Cada amostra foi replicada 5.000 vezes. O viés, o erro quadrático médio (EQM), a raiz do erro quadrático médio (REQM), o erro médio absoluto (EAM) e o desvio padrão das estimativas (DP) foram calculados para os parâmetros  $a, b, \mu$  e  $p_0$ . Todos os cálculos foram realizados em [R-Development-Core-Team \(2023\)](#).

A descrição da geração de dados é descrita a seguir.

1. determinar os valores dos parâmetros  $a, b, \mu$  e  $p_0$ ;
2. gerar  $t'_i$  de  $F(t) = u$ , onde  $u \sim U(0, 1)$ ;
3. gerar  $M_i \sim \text{Bernoulli}(1 - p_0)$  e  $Z_i \sim U(0, t_{MAX}), t_{MAX} = \max(t'_i)$ ;
4. se  $M_i = 0$  definir  $y_i = \infty$ ; se  $M_i = 1$ , considere  $y_i = t'_i$ ;
5.  $t_i = \min(y_i, z_i)$ ;
6. se  $t_i < z_i$ ,  $\delta_i = 1$ , se  $t_i \geq z_i$ ,  $\delta_i = 0$ .

De acordo com a [Tabela 16](#) observamos que em geral o viés, EQM, REQM, EAM e DP dos estimadores para o parâmetro  $p_0$  diminuem à medida que o tamanho da amostra aumenta. Quando as taxas de censura aumentam, o EQM, REQM, EAM e DP das estimativas do parâmetro  $p_0$  tendem a aumentar. o viés das frações de cura é muito pequeno, mesmo para os menores valores amostrais.

Em geral, um tamanho de amostra maior faz com que o viés, EQM, REQM, EAM e DP de  $a$  e  $\mu$  se tornem próximos de zero, como apresentamos na [Figura 21](#), [Figura 22](#) e [Figura 23](#), porém mesmo para  $n = 500$  o viés, EQM, REQM, EAM e DP de  $b$  não se aproximam tanto de zero. Além disso, conforme [Figura 21](#), [Figura 22](#) e [Figura 23](#), notamos que os valores das medidas calculadas estão mais próximos de zero para cenários com menor fração de cura, representados na [Figura 21](#), [Figura 22](#) e [Figura 23](#) pela linha contínua preta.

Em geral, os valores das probabilidades de cobertura variaram em torno do valor nominal de 95%, entretanto em alguns cenários as probabilidades de cobertura para os parâmetros  $a$  e  $b$  não ficaram próximas do valor nominal.

Tabela 16 – Viés, EQM, REQM, EAM e DP de  $\hat{p}_0$  no modelo (4.7) com risco de base de uma distribuição exponencial com  $\mu = 0.03$  para diferentes valores de  $a, b, p_0$  e  $n$ , com percentual de censura (PC) variando de 30% a 60%.

PC	$p_0$	$n$	$a$	$b$	$\hat{p}_0$				
					Viés	EQM	REQM	EAM	DP
35.50%	0.2	100	0.01	0.1	-0.0141	0.0049	0.0698	0.0532	0.0792
35.79%			0.01	0.5	-0.0128	0.0051	0.0715	0.0542	0.0790
35.57%			0.001	0.04	-0.0127	0.0050	0.0704	0.0538	0.0767
32.75%		300	0.01	0.1	-0.0019	0.0010	0.0322	0.0251	0.0322
32.89%			0.01	0.5	-0.0028	0.0011	0.0336	0.0256	0.0337
32.67%			0.001	0.04	-0.0032	0.0010	0.0318	0.0249	0.0327
31.68%		500	0.01	0.1	-0.0005	0.0006	0.0238	0.0188	0.0229
31.79%			0.01	0.5	-0.0019	0.0006	0.0241	0.0188	0.0231
31.62%			0.001	0.04	-0.0015	0.0005	0.0234	0.0185	0.0231
59.75%	0.5	100	0.01	0.1	-0.0218	0.0096	0.0979	0.0653	0.1420
59.95%			0.01	0.5	-0.0167	0.0087	0.0935	0.0618	0.1121
59.78%			0.001	0.04	-0.0182	0.0091	0.0954	0.0640	0.1119
57.92%		300	0.01	0.1	-0.0016	0.0018	0.0430	0.0304	0.0385
58.06%			0.01	0.5	-0.0010	0.0019	0.0431	0.0303	0.0399
57.91%			0.001	0.04	-0.0012	0.0018	0.0429	0.0307	0.0375
57.29%		500	0.01	0.1	-0.0007	0.0007	0.0272	0.0211	0.0264
57.42%			0.01	0.5	-0.0005	0.0007	0.0271	0.0212	0.0269
57.24%			0.001	0.04	-0.0012	0.0008	0.0285	0.0215	0.0269

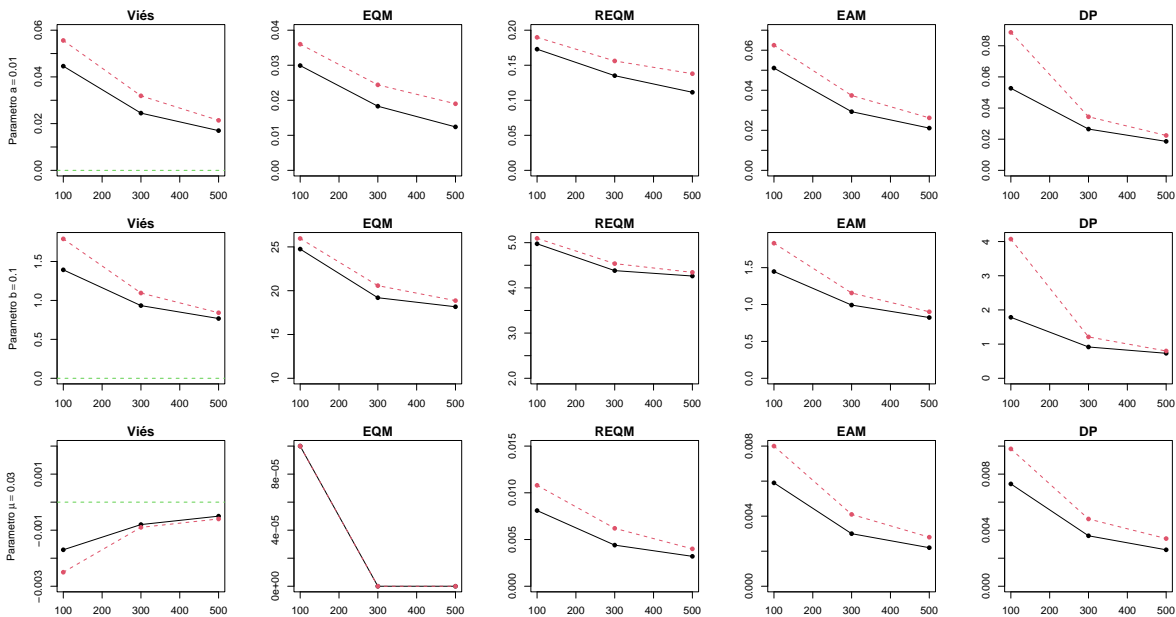


Figura 21 – Viés, EQM, REQM, EAM e DP de  $(\hat{a}, \hat{b}, \hat{\mu})$  versus  $n$  para o SMMAF com  $a = 0.01, b = 0.1, \mu = 0.03$ . As linhas pretas e vermelhas representam o cenário onde  $p_0 = 0.2$  e  $0.5$ , respectivamente



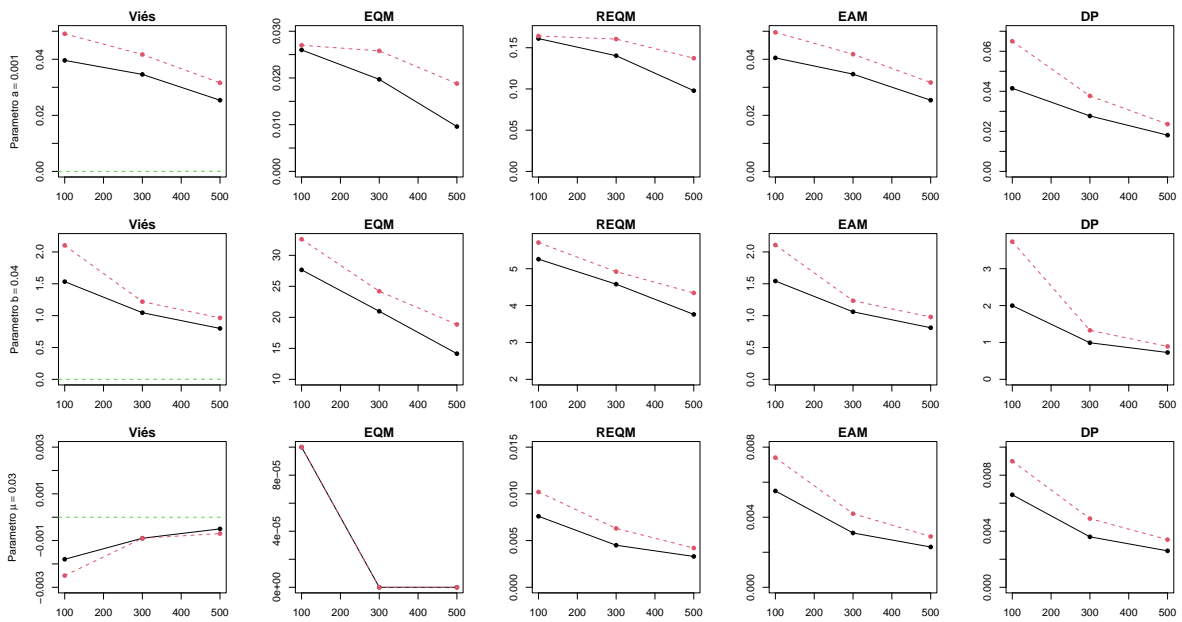


Figura 22 – Viés, EQM, REQM, EAM e DP de  $(\hat{a}, \hat{b}, \hat{\mu})$  versus  $n$  para o SMMAF com  $a = 0.001, b = 0.04, \mu = 0.03$ . As linhas pretas e vermelhas representam o cenário onde  $p_0 = 0.2$  e  $0.5$ , respectivamente

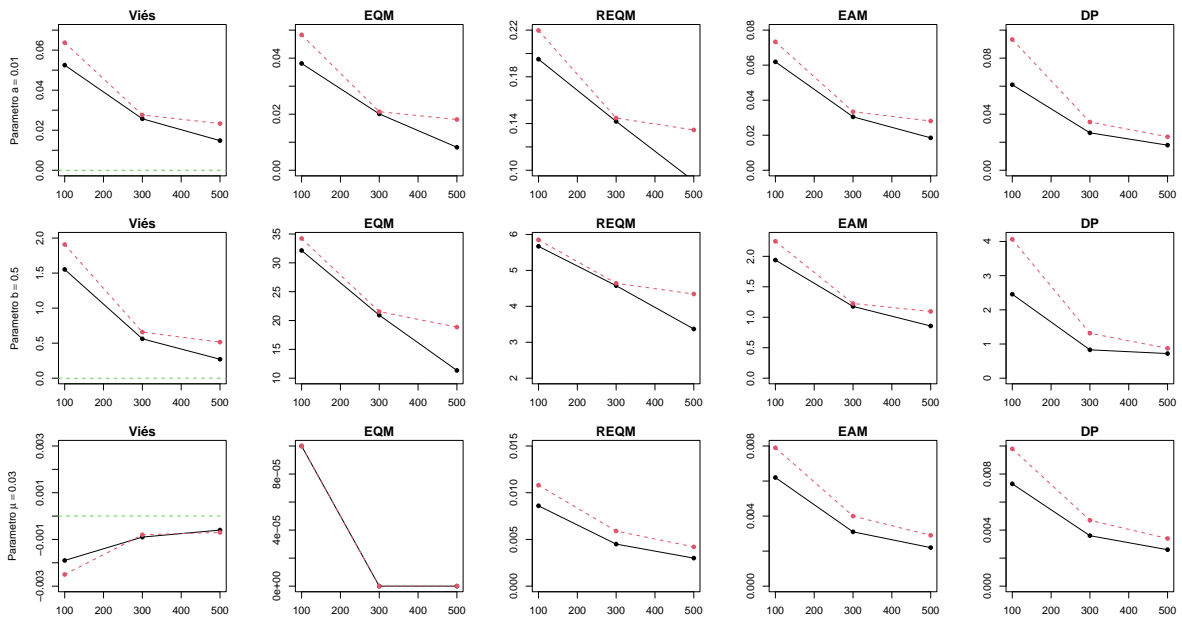


Figura 23 – Viés, EQM, REQM, EAM e DP de  $(\hat{a}, \hat{b}, \hat{\mu})$  versus  $n$  para o SMMAF com  $a = 0.01, b = 0.5, \mu = 0.03$ . As linhas pretas e vermelhas representam o cenário onde  $p_0 = 0.2$  e  $0.5$ , respectivamente

### 4.1.6 Aplicação

A aplicação será feita aos dados dos tempos de vida de 417 pacientes diagnosticados com câncer de pele (melanoma) e que foram incluídos no estudo de 1991 a 1995 e acompanhados até

o ano de 1998. Nesse conjunto de dados mais da metade dos tempos são de censura (55,64%), ou seja, mais da metade dos pacientes não apresentaram o evento de interesse durante o estudo (óbito devido ao melanoma).

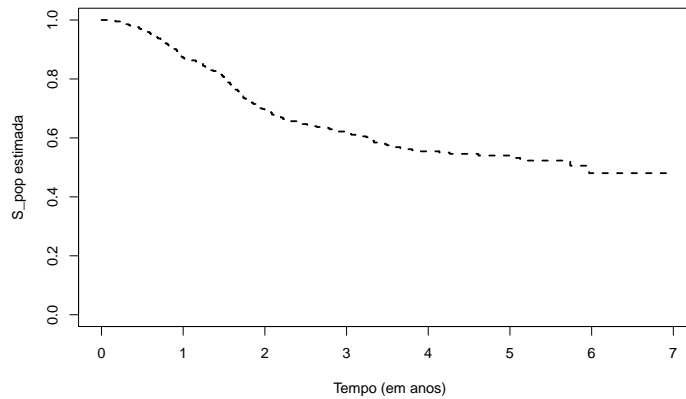


Figura 24 – Curvas de sobrevivência estimadas por Kaplan-Meier para os dados de melanoma

Como é conhecido esse tipo de câncer tem uma alta proporção de cura quando detectado precocemente, por isso é razoável esperar que parte desses indivíduos que não foram a óbito durante o estudo foram curados e, portanto não apresentarão esse evento ao longo da vida. Na Figura 24 temos a curva de sobrevivência estimada pelo estimador de Kaplan-Meier, onde a função estabiliza em um patamar de 50%, indicando que há uma fração de curados nesses pacientes observados. Na Tabela 17 temos as estimativas de máxima verossimilhança para os parâmetros do modelo da Equação 4.7, com fragilidade  $Gama(a, b)$ , seus respectivos erros padrão, além do AIC e BIC de cada modelo como critérios para comparação de modelos.

Tabela 17 – Estimativas de máxima verossimilhança, erros padrão, intervalos de confiança de 95%, AIC e BIC para os parâmetros do modelo (4.7) com riscos de base da distribuição Exponencial( $\mu$ ) e da distribuição Weibull( $\mu, \gamma$ ).

Parâmetros	Exponencial			Weibull		
	EMV	EP	IC	EMV	EP	IC
$a$	0.0039	0.0034	(0.0000; 0.0107)	0.0059	0.0274	(0.0000; 0.0595)
$b$	4.5064	0.5370	(3.4540; 5.5589)	1.2266	0.3813	(0.4793; 1.9739)
$\mu$	0.2905	0.0502	(0.1922; 0.3888)	0.4401	0.0293	(0.3827; 0.4975)
$\gamma$	-	-	-	1.5750	0.1076	(1.3641; 1.7859)
$p_0$	0.3819	0.0602	(0.2638; 0.4999)	0.5040	0.0294	(0.4464; 0.5616)
AIC		1098.324		1067.410		
BIC		1114.456		1087.576		

Pelos critérios AIC e BIC o modelo de mistura padrão com fragilidade aditiva  $Gama(a, b)$  e com risco de base de uma distribuição Weibull( $\lambda, \gamma$ ) apresentou um melhor ajuste. Nos dois modelos avaliados a estimativa do valor esperado da fragilidade foi pequena e para o modelo

selecionado pelos critérios AIC e BIC a estimativa de  $\hat{a}/\hat{b}$  da fragilidade foi de 0.0048 e a variância estimada ( $\hat{a}/\hat{b}^2$ ) da variável de fragilidade é de 0.0039. A existência de heterogeneidade não observada pode ser contornada com a inclusão de covariáveis, mas em muitas situações práticas, como nessa aplicação, fatores não observáveis como fatores genéticos podem exercer bastante influência no tempo de vida dos pacientes, o que justifica a inclusão do termo de fragilidade na modelagem.

Na Figura 25 temos que os ajustes das curvas de sobrevivência para os modelos de mistura padrão com fragilidade aditiva Gama( $a, b$ ) e com risco de base de uma distribuição Exponencial e Weibull ficaram relativamente próximos à curva de Kaplan-meier, entretanto percebemos que o modelos de mistura padrão com fragilidade aditiva Gama( $a, b$ ) e com risco de base de uma distribuição Weibull apresentou um melhor ajuste em relação ao modelo com risco de base de uma distribuição Exponencial.

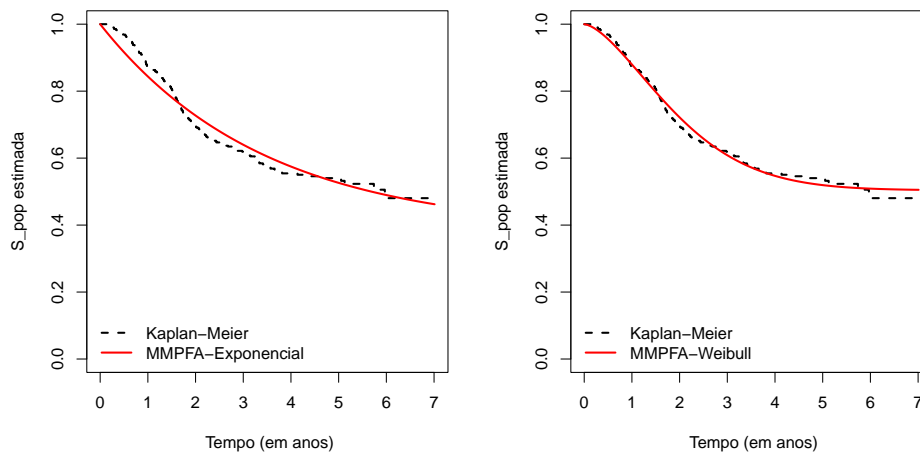


Figura 25 – Curvas de sobrevivência estimadas pelo modelo (4.7) com funções de risco de base das distribuições Exponencial e Weibull, respectivamente

#### 4.1.7 Modelo de mistura padrão com fragilidade aditiva na presença de covariáveis

Nessa modelagem as covariáveis podem ser introduzidas no parâmetro da fração de cura, ou na função de sobrevivência, ou ainda nos dois termos. As covariáveis que podem ter efeito na fração de cura e na função de sobrevivência podem ser as mesmas. Dessa forma o modelo de mistura padrão com covariáveis pode ser escrito da seguinte forma

$$S_{pop}(t|\mathbf{x}) = p_0(\mathbf{x}) + (1 - p_0(\mathbf{x}))S(t|\mathbf{x}) \quad (4.9)$$

e o modelo de mistura padrão com fragilidade aditiva na presença de covariáveis pode ser expresso como

$$S_{pop}(t|\mathbf{x}) = p_0(\mathbf{x}) + (1 - p_0(\mathbf{x})) \exp\{-\Lambda_0(t) - \mathbf{x}'\beta t\}L(t) \quad (4.10)$$

Para ligar a fração de cura às covariáveis podemos utilizar as funções de ligação logito, probito ou complemento log-log, visto que essas funções assumem valores no intervalo (0,1) e são usadas para modelar probabilidades. Nesse trabalho utilizamos somente a função logito dada por:

$$p_0(\mathbf{x}) = \frac{\exp\{\mathbf{b}\mathbf{x}\}}{1 + \exp\{\mathbf{b}\mathbf{x}\}}. \quad (4.11)$$

Consideramos para a  $i$ -ésima observação os dados observados  $\mathcal{D} = (t_i, \delta_i, x_i)$ , em que  $t_i$  denota o tempo observado,  $\delta_i$  é a variável indicadora de censura, com  $\delta_i = 1$  se  $t_i$  é não censurado e  $\delta_i = 0$  caso contrário e  $\mathbf{x}_i$  representa o vetor de covariáveis que tem efeito na distribuição latente e na taxa de cura. De maneira similar a Equação 4.6 obtemos a seguinte função de verossimilhança

$$\begin{aligned} L(\boldsymbol{\theta}|\mathcal{D}) &= \prod_{i=1}^n [f_{pop}(t_i|\mathbf{x}_i)]^{\delta_i} [S_{pop}(t_i|\mathbf{x}_i)]^{1-\delta_i} \\ &= \prod_{i=1}^n [(1 - p_0(\mathbf{x}_i))\lambda(t_i|\mathbf{x}_i)S(t_i|\mathbf{x}_i)]^{\delta_i} \times [p_0(\mathbf{x}_i) + (1 - p_0(\mathbf{x}_i))S(t_i|\mathbf{x}_i)]^{1-\delta_i} \\ &= \prod_{i=1}^n \left[ (1 - p_0(\mathbf{x}_i)) \left( \lambda_0(t_i) + \mathbf{x}_i'\boldsymbol{\beta} - \frac{L'(t_i)}{L(t_i)} \right) \exp\{-\Lambda_0(t_i) - \mathbf{x}_i'\boldsymbol{\beta}t_i\}L(t_i) \right]^{\delta_i} \\ &\quad \times [p_0(\mathbf{x}_i) + (1 - p_0(\mathbf{x}_i)) \exp\{-\Lambda_0(t_i) - \mathbf{x}_i'\boldsymbol{\beta}t_i\}L(t_i)]^{1-\delta_i} \end{aligned} \quad (4.12)$$

Rodrigues *et al.* (2009) consideraram covariáveis só na fração de cura. Seguindo essa mesma ideia podemos reescrever a Equação 4.9 da seguinte maneira

$$S_{pop}(t|\mathbf{x}) = p_0(\mathbf{x}) + (1 - p_0(\mathbf{x}))S(t) \quad (4.13)$$

e o modelo de mistura padrão com fragilidade aditiva na presença de covariáveis dado em (4.10) pode ser reescrito como

$$S_{pop}(t|\mathbf{x}) = p_0(\mathbf{x}) + (1 - p_0(\mathbf{x})) \exp\{-\Lambda_0(t)\}L(t) \quad (4.14)$$

Assim, a função de verossimilhança baseada nos dados observados supondo censura não informativa é dada por

$$\begin{aligned} L(\boldsymbol{\theta}|\mathcal{D}) &= \prod_{i=1}^n [f_{pop}(t_i|\mathbf{x}_i)]^{\delta_i} [S_{pop}(t_i|\mathbf{x}_i)]^{1-\delta_i} \\ &= \prod_{i=1}^n [(1 - p_0(\mathbf{x}_i))\lambda(t_i)S(t_i)]^{\delta_i} \times [p_0(\mathbf{x}_i) + (1 - p_0(\mathbf{x}_i))S(t_i)]^{1-\delta_i} \\ &= \prod_{i=1}^n \left[ (1 - p_0(\mathbf{x}_i)) \left( \lambda_0(t_i) - \frac{L'(t_i)}{L(t_i)} \right) \exp\{-\Lambda_0(t_i)\}L(t_i) \right]^{\delta_i} \\ &\quad \times [p_0(\mathbf{x}_i) + (1 - p_0(\mathbf{x}_i)) \exp\{-\Lambda_0(t_i)\}L(t_i)]^{1-\delta_i} \end{aligned}$$

Para completar a abordagem paramétrica é necessário supor uma distribuição para a fragilidade e para o risco de base.

## 4.2 Aplicação

Nesta seção, ilustramos algumas aplicações do modelo proposto. Utilizamos dois conjuntos de dados reais que são estudos clínicos médicos relacionados à ocorrência de câncer apresentados a seguir. Para cada aplicação ajustada, calculamos as estimativas de máxima verossimilhança dos parâmetros, erro padrão, intervalos de confiança de 95%, critérios AIC e BIC e estimativa de máxima verossimilhança da fração de cura, e o erro padrão para a fração de cura obtido através do método delta.

### 4.2.1 Dados de câncer de mama

Os dados dessa aplicação são de um estudo em mulheres diagnosticadas com câncer de mama triplo negativo, que receberam o tratamento de quimioterapia neoadjuvante de 2001 a 2013 no A. C. Camargo Cancer Center, São Paulo, Brasil. Nesse estudo o evento de interesse era a morte devido ao câncer de mama, o tempo de vida considerado foi desde a data do diagnóstico até o óbito e os tempos de censura foram definidos da data do diagnóstico até o fim do estudo para as pacientes que não apresentaram o evento de interesse, ou da data do diagnóstico até à morte por outra causa. Esses dados também foram analisados com outra modelagem por [Scudilio et al. \(2018\)](#).

O estudo contou com 78 pacientes, dentre as quais 32,1% morreram devido ao câncer de mama e 67,9% não apresentaram o evento de interesse. Dentre as covariáveis observadas no estudo utilizamos nessa aplicação a covariável que capta as características dos linfonodos das cadeias de drenagem linfática do órgão em que o tumor está, denotada por N, em que N0 representa que os linfonos próximos não contêm câncer e N1, N2 e N3 representam o grau de espalhamento do câncer para os linfonodos, de forma que N0 significa que não houve espalhamento do câncer para os linfonodos e N1, N2 e N3 significa que houve espalhamento. Sendo assim categorizamos N=0 quando não houve espalhamento (N0) e N=1 caso contrário (N1, N2 e N3).

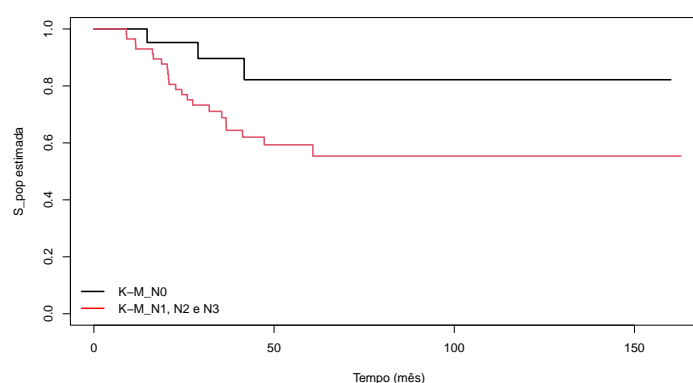


Figura 26 – Curvas de sobrevivência estimadas por Kaplan-Meier para cada categoria N

Para os 21 pacientes na categoria N0 apenas 14,3% veio à óbito, enquanto que 38,6% sofreram o evento de interesse para a outra categoria (N1, N2 e N3). Na Figura 26 vemos as curvas de sobrevivência estimadas pelo estimador não paramétrico de Kaplan-Meier, onde a função de sobrevivência para os pacientes onde não houve espalhamento do câncer decai mais lentamente e se estabiliza em um patamar mais elevado do que a curva de sobrevivência dos indivíduos que apresentaram espalhamento do câncer. Esse comportamento é esperado, pois os casos de espalhamento indicam um estágio mais avançado da doença.

De acordo com as estimativas de máxima verossimilhança apresentadas na Tabela 18 o modelo de mistura padrão com fragilidade Gama( $a, b$ ) e risco de base de uma distribuição Weibull( $\mu, \gamma$ ) apresenta um melhor ajuste em termos de AIC e BIC, o que também é possível observar na Figura 27, pois se ajustou melhor a curva de sobrevivência estimada não parametricamente.

Tabela 18 – Estimativas de máxima verossimilhança, erros padrão, intervalos de confiança de 95%, AIC e BIC para os parâmetros do modelo (4.14) com riscos de base da distribuição Exponencial( $\mu$ ) e da distribuição Weibull( $\mu, \gamma$ ).

Exponencial			
Parâmetros	EMV	EP	IC
$a$	$5.13 \times 10^{-8}$	$2.94 \times 10^{-7}$	(0; $6.28 \times 10^{-7}$ )
$b$	0.0463	0.0450	(0; 0.1344)
$\mu$	0.0196	0.0084	(0.0051; 0.0369)
$\gamma$	-	-	-
$b_0$	1.3022	0.7115	(-0.0924; 2.6968)
$b_1$	-1.6543	0.8178	(-3.2572; -0.0514)
AIC	303.6137		
BIC	315.3973		
Weibull			
Parâmetros	EMV	EP	IC
$a$	$7.12 \times 10^{-8}$	0.0000	( $7.12 \times 10^{-8}$ ; $7.12 \times 10^{-8}$ )
$b$	0.0464	0.0136	(0.0198; 0.0730)
$\mu$	0.0309	0.0034	(0.0242; 0.0376)
$\gamma$	2.1940	0.3593	(1.4897; 2.8982)
$b_0$	1.5615	0.6462	(0.2949; 2.8281)
$b_1$	-1.3499	0.7083	(-2.7380; 0.0383)
AIC	292.2958		
BIC	306.4361		

Tabela 19 – Estimativa de máxima verossimilhança através da função de ligação logística para a fração de cura de cada categoria da covariável N no modelo (4.14) com risco de base das distribuições Exponencial e Weibull.

Paâmetros	Exponencial		Weibull	
	EMV	EP	EMV	EP
$p_0$	0.7862	0.1196	0.8266	0.0926
$p_1$	0.4129	0.1518	0.5527	0.0752

Avaliando as estimativas das frações de cura apresentadas na [Tabela 19](#) obtidas pela função logito dada em (4.11) temos que a proporção de curado no grupo que não teve espalhamento do câncer é 50% maior que no grupo onde houve espalhamento do câncer.

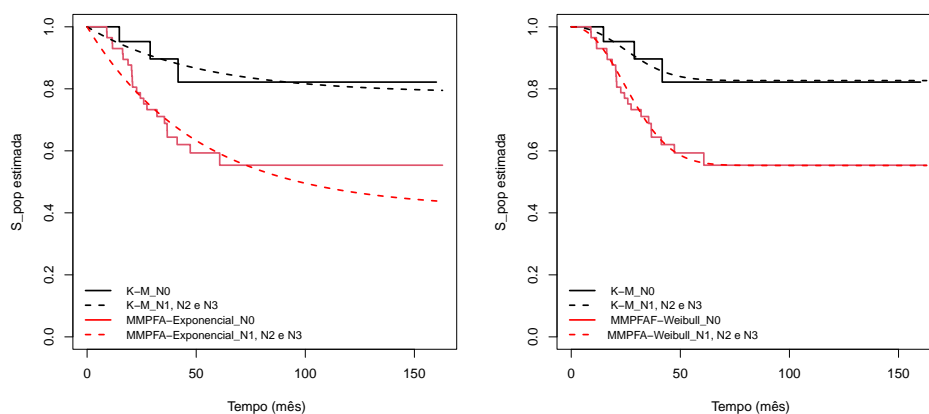


Figura 27 – Curvas de sobrevivência estimadas pelo modelo (4.14) (linhas tracejadas) com funções de risco de base das distribuições Exponencial e Weibull, respectivamente

#### 4.2.2 Dados de infecção sítio cirúrgico

Nessa aplicação foram utilizados dados de um estudo retrospectivo real com prontuários de 2011 a 2019 para identificar pacientes que foram reoperados para desbridamento de infecções sítio cirúrgica (ISC) após ressecção de tumor cerebral. Dos 81 casos de ISC, 57 foram submetidos a desbridamento. A relação neutrófilo-linfócito (NLR) foi estudada como possível preditor de sobrevida. Nesse estudo, o evento de interesse foi definido como óbito. O conjunto de dados contém portanto 57 pacientes, dos quais 38 (66.7%) morreram, 39 (68.4%) eram pacientes do sexo masculino e 36 (63.2%) pacientes tinham  $NLR > 3.5$ . Entre os pacientes do grupo ( $NLR \leq 3.5$ ) 47.6% faleceram, enquanto 77.7% dos pacientes com ( $NLR > 3.5$ ) faleceram.

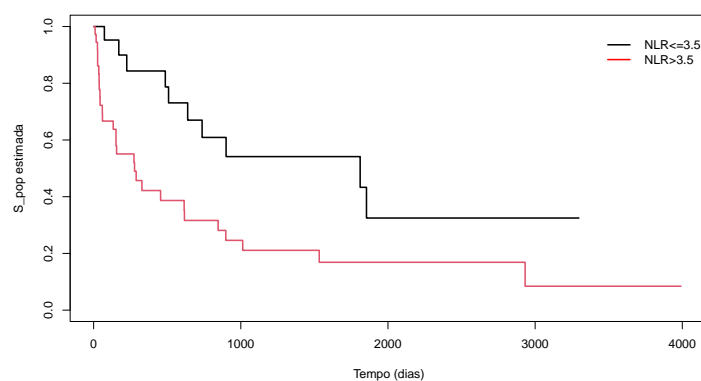


Figura 28 – Curvas de sobrevivência estimadas por Kaplan-Meier para cada grupo de NLR

Telles *et al.* (2022) apresentaram um estudo para analisar se o desbridamento precoce melhorou a sobrevida em comparação à antibioticoterapia isolada em pacientes com infecções pós-craniotomia após neurocirurgias oncológicas. Neste estudo é possível observar que valores menores ou iguais a 3.5 para a relação neutrófilo-linfócito (NLR) nos pacientes predizem melhor prognóstico de sobrevida.

A Figura 28 apresenta a função de sobrevivência estimada pelo estimador de Kaplan-Meier para diferentes grupos de NLR (0:  $NLR \leq 3.5$  e 1:  $NLR > 3.5$ ). As curvas de sobrevivência não decaem a zero, caracterizando dados de longa duração, principalmente para indivíduos com bom prognóstico. As estimativas de máxima verossimilhança para os parâmetros do modelo da Equação 4.10, com distribuição de fragilidade Gama, seus respectivos erros padrão, além do AIC e BIC de cada modelo são apresentados na Tabela 20.

Tabela 20 – Estimativas de máxima verossimilhança, erros padrão, intervalos de confiança de 95%, AIC e BIC para os parâmetros do modelo (4.10) com riscos de base da distribuição Exponencial( $\mu$ ) e da distribuição Weibull( $\mu, \gamma$ ) para os dados de ISC.

Exponencial			
Parâmetros	EMV	EP	IC
$a$	$7.93 \times 10^{-7}$	$6.13 \times 10^{-5}$	(0; 0.0001)
$b$	0.0817	0.0844	(0; 0.2470)
$\mu$	0.0008	0.0005	(0; 0.0017)
$\gamma$	-	-	-
$b_0$	-1.1197	1.2551	(-3.5797; 1.3403)
$b_1$	-0.9035	1.3992	(-3.6459; 1.8389)
$\beta$	0.0015	0.0007	(0.0001; 0.0028)
AIC	602.6312		
BIC	614.8896		
Weibull			
Parâmetros	EMV	EP	IC
$a$	$1.27 \times 10^{-6}$	0.0029	(0; 0.0057)
$b$	0.5285	0.2905	(0; 1.0979)
$\mu$	0.0010	0.0007	(0; 0.0024)
$\gamma$	0.7337	0.1365	(0.4661; 1.0013)
$b_0$	-0.9394	1.0349	(-2.9678; 1.0890)
$b_1$	-1.2204	1.2836	(-3.7362; 1.2953)
$\beta$	0.0009	0.0008	(-0.0007; 0.0025)
AIC	602.3172		
BIC	616.6186		

O MMPFA com risco de base de uma distribuição Exponencial apresentou melhor ajuste pelos critérios BIC, pois segundo os critérios AIC foram muito próximos. O intervalo de confiança assintótico de 95% para o parâmetro  $\gamma$  dado na Tabela 20 contém o valor 1 (0.4661; 1.0013), indicando que o MMPFA com função de risco de base exponencial parece ser uma escolha mais razoável do que o modelo função de risco de base de uma Weibull. A estimativa de máxima verossimilhança obtida para o parâmetro  $\beta$  ( $\beta = 0.0015$ ) indica menor sobrevida



para os pacientes do grupo 1, ou seja, indivíduos com  $RNL > 3,5$  possuem menor tempo de vida estimado, conforme mostra a [Figura 29](#).

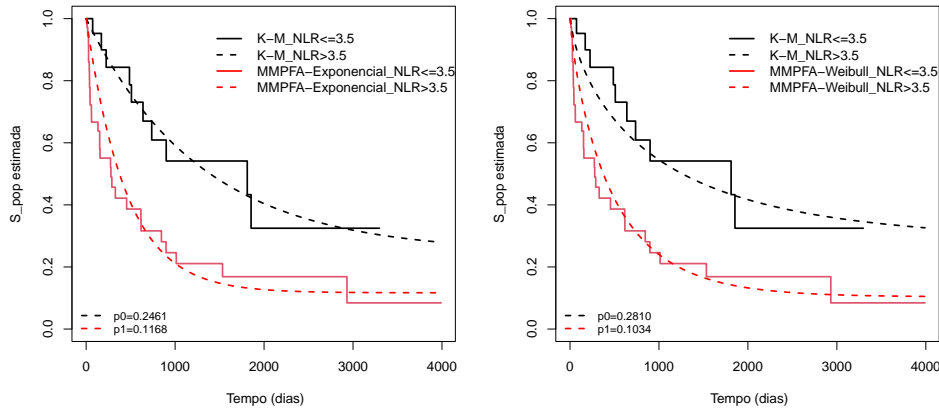


Figura 29 – Curvas de sobrevivência estimadas pelo modelo (4.10) (linhas tracejadas) com funções de risco de base das distribuições Exponencial e Weibull, respectivamente

Tabela 21 – Estimativa de máxima verossimilhança (EMV) através da função de ligação logística para a fração de cura de cada categoria da covariável NLR no modelo (4.14) com risco de base das distribuições Exponencial e Weibull.

Parameters	Exponencial		Weibull	
	EMV	EP	EMV	EP
$p_0$	0.2461	0.2328	0.2810	0.2091
$p_1$	0.1168	0.0636	0.1034	0.0676

A partir da função de ligação logito obtivemos as estimativas para as frações de cura de cada grupo da covariável NLR na [Tabela 21](#) e através do método delta obtivemos os erros padrão correspondentes. Notamos que o valor estimado de  $p_0$  é menor para o modelo com função de risco de base Exponencial quando comparado ao modelo com função de risco base de uma distribuição Weibull enquanto para  $p_1$  acontece o contrário, mas em ambos os modelos a fração estimada de curados do grupo com  $NLR \leq 3.5$  é mais que o dobro do grupo com  $NLR > 3.5$ .

### 4.2.3 Conclusão

Nesse capítulo consideramos uma modelagem paramétrica utilizando o modelo de mistura padrão com fragilidade aditiva. Nessa proposta assumimos uma distribuição Gama para a variável de fragilidade. Procedemos com a inclusão de covariáveis no termo da fração de cura através da função logística e fizemos aplicações a dados reais onde obtivemos bons ajustes, o que demonstra um potencial dessa abordagem. Além disso, no estudo de simulação apresentado obtivemos resultados que mostram um comportamento aceitável das estimativas de máxima verossimilhança no que se refere às propriedades assintóticas dos estimadores.

---

## MODELAGEM DE RISCO ADITIVO ALTERNATIVA COM FRAGILIDADE PARA PARA DADOS UNIVARIADOS

---

A modelagem estudada no [Capítulo 2](#) inclui as covariáveis no modelo aditivo apresentado na [Equação 2.3](#) de forma que não garante valores positivos para a função de risco. Para resolver essa desvantagem iremos propor nesse capítulo uma modelagem de fragilidade aditiva que garanta que os risco sempre serão positivos. [Gupta \(2016\)](#) estudou um modelo geral de fragilidade aditiva e apresentou alguns casos especiais e exemplos. O modelo geral apresentado por [Gupta \(2016\)](#) é descrito pela seguinte expressão:

$$\lambda(t|v) = \lambda_0(t) + A(v)\phi(t), \quad (5.1)$$

em que  $\lambda_0(t)$  é a função de risco de base,  $\phi(t)$  é uma função conhecida e  $A(v)$  é uma função da variável de fragilidade. A função de risco não condicional de (5.1) é dada por  $\lambda(t) = \lambda_0(t) + \phi(t)E[A(V)|T > t]$ . Como casos particulares ele apresenta  $\lambda(t|v) = \lambda_0(t) + v\phi(t)$  ([MUKHERJEE; ROY, 1998](#)) e  $\lambda(t|v) = \lambda_0(t) + \beta_0v$  ([KIM; LEE, 1998](#)).

Nesse trabalho propomos o seguinte modelo

$$\lambda(t|v) = \lambda_0(t) + \exp(\mathbf{x}'\beta)v, \quad (5.2)$$

em que modelo (5.2) seria um caso particular do modelo proposto por [Gupta \(2016\)](#). Nesse modelo  $\lambda_0(t)$  é o risco de base comum a todos os indivíduos no estudo,  $V$  é a variável de fragilidade,  $\exp(\cdot)$  é uma função positiva das covariáveis adicionada à função de risco de base e  $\beta$  é o vetor de coeficientes da regressão. Com essa modelagem garantimos que a função de risco é sempre positiva, o que já seria uma vantagem em relação ao modelo do [Capítulo 2](#).

## 5.1 Modelo de risco aditivo alternativo com fragilidade para dados univariados

Dado o modelo (5.2) temos que a função de risco acumulada condicional segue a seguinte expressão

$$\Lambda(t|v, \mathbf{x}) = \Lambda_0(t) + \exp(\mathbf{x}'\beta)vt. \quad (5.3)$$

Seguindo os mesmos passos do modelo aditivo dado em (2.3) temos que a função de sobrevivência condicional é dada por:

$$S(t|v, \mathbf{x}) = \exp\{-\Lambda_0(t) - \exp(\mathbf{x}'\beta)vt\}.$$

As funções de sobrevivência e de risco não condicionais do modelo (5.2) são expressas respectivamente por:

$$S(t|\mathbf{x}) = \exp\{-\Lambda_0(t)\}L_V(\exp(\mathbf{x}'\beta)) \quad (5.4)$$

e

$$\lambda(t|\mathbf{x}) = \lambda_0(t) - \frac{L'_V(\exp(\mathbf{x}'\beta))}{L_V(\exp(\mathbf{x}'\beta))}. \quad (5.5)$$

em que  $L_V(\cdot)$  é a transformada de Laplace da distribuição da variável  $V$  e  $L'_V(\cdot)$  é a derivada de  $L_V(\cdot)$  em relação a  $t$ .

A função de sobrevivência não condicional com fragilidade Gama dada em 2.24 a partir da equação (5.4) é de tal forma que:

$$\begin{aligned} S(t|\mathbf{x}) &= \exp\{-\Lambda_0(t)\}L_V(\exp(\mathbf{x}'\beta)) \\ &= \exp\{-\Lambda_0(t)\} \int_0^\infty \exp\{-vt\exp(\mathbf{x}'\beta)\} \frac{b^a}{\Gamma(a)} \exp\{-vb\} v^{a-1} dv \\ &= \frac{b^a}{\Gamma(a)} \exp\{-\Lambda_0(t)\} \int_0^\infty \exp\{-v(\exp(\mathbf{x}'\beta) + b)\} v^{a-1} dv \end{aligned} \quad (5.6)$$

Observando o integrando na expressão (5.6) como núcleo de uma distribuição Gama com parâmetros  $(a, \exp(\mathbf{x}'\beta) + b)$ , obtemos que  $L_V(t) = \left(\frac{b}{\exp(\mathbf{x}'\beta) + b}\right)^a$ , logo:

$$S(t|\mathbf{x}) = \exp\{-\Lambda_0(t)\} \left(\frac{b}{\exp(\mathbf{x}'\beta) + b}\right)^a, \quad (5.7)$$

e pela Equação 5.5 a função de risco não condicional para fragilidade dada em 2.24 é

$$\lambda(t|\mathbf{x}) = \lambda_0(t) + \frac{a\exp(\mathbf{x}'\beta)}{\exp(\mathbf{x}'\beta) + b}. \quad (5.8)$$

Pela [Equação 2.11](#) temos que a função de densidade de probabilidade de  $V$  dado  $T > t$  é dado por

$$\begin{aligned} f(v|T > t) &= \frac{\exp\{-\Lambda_0(t) - \exp(\mathbf{x}'\beta)vt\} \frac{b^a}{\Gamma(a)} \exp\{-vb\} v^{a-1}}{\exp\{-\Lambda_0(t)\} \left(\frac{b}{t\exp(\mathbf{x}'\beta)+b}\right)^a} \\ &= \frac{(t\exp(\mathbf{x}'\beta) + b)^a \exp\{-\Lambda_0(t)\} b^a}{b^a \exp\{-\Lambda_0(t)\} \Gamma(a)} \exp\{-v(t\exp(\mathbf{x}'\beta) + b)\} v^{a-1} \\ &= \frac{(t\exp(\mathbf{x}'\beta) + b)^a}{\Gamma(a)} \exp\{-v(t\exp(\mathbf{x}'\beta) + b)\} v^{a-1}, \end{aligned}$$

ou seja,  $V|T > t \sim \text{Gama}(a, t\exp(\mathbf{x}'\beta) + b)$  e portanto  $E[V|T > t] = a/(t\exp(\mathbf{x}'\beta) + b)$  e a variância de  $V|T > t$  é dada por  $\text{Var}[V|T > t] = a/(t\exp(\mathbf{x}'\beta) + b)^2$ . Podemos obter ainda a distribuição da fragilidade  $V$  dado  $T = t$  a partir da [Equação 2.13](#) da seguinte maneira

$$\begin{aligned} f(v|T = t) &= \frac{S(t|v)\lambda(t|v)f(v)}{S(t)\lambda(t)} \\ &= \frac{\exp\{-\Lambda_0(t) - \exp(\mathbf{x}'\beta)vt\} (\lambda_0(t) + \exp(\mathbf{x}'\beta)v) v^{a-1} \frac{b^a}{\Gamma(a)} \exp\{-vb\}}{\exp\{-\Lambda_0(t)\} \left(\frac{b}{t\exp(\mathbf{x}'\beta)+b}\right)^a \left(\lambda_0(t) + \frac{a\exp(\mathbf{x}'\beta)}{t\exp(\mathbf{x}'\beta)+b}\right) \Gamma(a)} \\ &= (\lambda_0(t) + \exp(\mathbf{x}'\beta)v) \frac{\exp\{-\exp(\mathbf{x}'\beta)vt\} v^{a-1} \frac{b^a}{\Gamma(a)} \exp\{-vb\}}{\left(\frac{b}{t\exp(\mathbf{x}'\beta)+b}\right)^a \left(\lambda_0(t) + \frac{a\exp(\mathbf{x}'\beta)}{t\exp(\mathbf{x}'\beta)+b}\right) \Gamma(a)} \\ &= \frac{\lambda_0(t)}{\left(\lambda_0(t) + \frac{a\exp(\mathbf{x}'\beta)}{t\exp(\mathbf{x}'\beta)+b}\right)} \frac{(t\exp(\mathbf{x}'\beta) + b)^a}{\Gamma(a)} v^{a-1} \exp\{-v(t\exp(\mathbf{x}'\beta) + b)\} \\ &\quad + \frac{\exp(\mathbf{x}'\beta)}{\left(\lambda_0(t) + \frac{a\exp(\mathbf{x}'\beta)}{t\exp(\mathbf{x}'\beta)+b}\right)} \frac{(t\exp(\mathbf{x}'\beta) + b)^a}{\Gamma(a)} v^{a+1-1} \exp\{-v(t\exp(\mathbf{x}'\beta) + b)\} \\ &= w_1 \times \frac{(t\exp(\mathbf{x}'\beta) + b)^a}{\Gamma(a)} v^{a-1} \exp\{-v(t\exp(\mathbf{x}'\beta) + b)\} \\ &\quad + \frac{\exp(\mathbf{x}'\beta)}{\left(\lambda_0(t) + \frac{a\exp(\mathbf{x}'\beta)}{t\exp(\mathbf{x}'\beta)+b}\right)} \frac{a}{t\exp(\mathbf{x}'\beta) + b} \frac{(t\exp(\mathbf{x}'\beta) + b)^{a+1}}{\Gamma(a+1)} v^{a+1-1} \\ &\quad \times \exp\{-v(t\exp(\mathbf{x}'\beta) + b)\} \\ &= w_1 \times f_{G_1} + w_2 \times f_{G_2}. \end{aligned}$$

Em que  $f_{G_1}$  é a função de densidade de uma distribuição  $\text{Gama}(a, t\exp(\mathbf{x}'\beta) + b)$  e  $f_{G_2}$  é a função de densidade de uma distribuição  $\text{Gama}(a+1, t\exp(\mathbf{x}'\beta) + b)$ , com  $w_1 + w_2 = 1$ . O valor esperado de  $V|T = t$  pode ser calculado da seguinte maneira

$$\begin{aligned} E[V|T = t] &= \int_0^\infty v f(v|T = t) dv = \int_0^\infty v (w_1 \times f_{G_1} + w_2 \times f_{G_2}) dv \\ &= \int_0^\infty v \times w_1 \times f_{G_1} dv + \int_0^\infty v \times w_2 \times f_{G_2} dv \\ &= w_1 \int_0^\infty v \times f_{G_1} dv + w_2 \int_0^\infty v \times f_{G_2} dv \\ &= w_1 \left( \frac{a}{t\exp(\mathbf{x}'\beta) + b} \right) + w_2 \left( \frac{a+1}{t\exp(\mathbf{x}'\beta) + b} \right). \end{aligned} \tag{5.9}$$

De acordo com as equações  $E[V|T > t]$  e  $E[V|T = t]$  dado em (5.9) podemos então estimar as fragilidades individuais com a parametrização Gama( $a, b$ ) por:

$$\hat{v}_i = \hat{w}_1 \left( \frac{\hat{a}}{t_i \exp(\mathbf{x}_i' \hat{\beta}) + \hat{b}} \right) + \hat{w}_2 \left( \frac{\hat{a} + 1}{t_i \exp(\mathbf{x}_i' \hat{\beta}) + \hat{b}} \right) \quad (5.10)$$

para os indivíduos que falharam no tempo  $t$  e por  $\hat{v}_i = \left( \frac{\hat{a}}{t_i \exp(\mathbf{x}_i' \hat{\beta}) + \hat{b}} \right)$  para os indivíduos censurados.

Para o risco de base também é possível escolher várias distribuições, dentre elas a distribuição Weibull( $\mu, \gamma$ ) com função de densidade  $f(t) = \gamma \mu^\gamma t^{\gamma-1} \exp[-(\mu t)^\gamma]$ , função de risco  $\lambda(t) = \gamma \mu^\gamma t^{\gamma-1}$  e função de risco acumulado  $\Lambda(t) = (\mu t)^\gamma$ . Quando  $\gamma = 1$  temos a distribuição Exponencial( $\mu$ ) com  $\lambda(t) = \mu$  constante. Nesse caso a função de sobrevivência não condicional dada em (5.7) fica da forma:

$$S(t|\mathbf{x}) = \exp\{-\mu t\} \left( \frac{b}{t \exp(\mathbf{x}'\beta) + b} \right)^a,$$

e a função de risco marginal é expressa como:

$$\lambda(t|\mathbf{x}) = \mu + \frac{a \exp(\mathbf{x}'\beta)}{t \exp(\mathbf{x}'\beta) + b}.$$

A partir das funções  $S(t|\mathbf{x})$  e  $\lambda(t|\mathbf{x})$  é possível obter a função de verossimilhança e assim estimar os parâmetros via método de MV. Para o modelo aditivo com fragilidade Gama( $a, b$ ) e risco base de uma distribuição Exponencial( $\mu$ ) temos a seguinte função de verossimilhança:

$$L(\mu, a, b, \beta | \mathcal{D}) = \prod_{i=1}^n \left\{ \left[ \mu + \frac{a \exp(\mathbf{x}_i' \beta)}{t_i \exp(\mathbf{x}_i' \beta) + b} \right]^{\delta_i} \exp\{-\mu t_i\} \left( \frac{b}{t_i \exp(\mathbf{x}_i' \beta) + b} \right)^a \right\}, \quad (5.11)$$

em que  $\delta_i = 1$  se  $t_i$  for um tempo de falha e é zero caso contrário. O logaritmo da função de verossimilhança em (5.11) denotado por  $l(\mu, a, b, \beta | \mathcal{D}) = \log[L(\mu, a, b, \beta | \mathcal{D})]$  é dado por:

$$\begin{aligned} l(\mu, a, b, \beta | \mathcal{D}) &= \sum_{i=1}^n \delta_i \log \left[ \mu + \frac{a \exp(\mathbf{x}_i' \beta)}{t_i \exp(\mathbf{x}_i' \beta) + b} \right] \\ &\quad + a \sum_{i=1}^n \log \left[ \frac{b}{t_i \exp(\mathbf{x}_i' \beta) + b} \right] - \sum_{i=1}^n \mu t_i, \end{aligned} \quad (5.12)$$

e as estimativas de MV para os parâmetros são obtidas numericamente maximizando a função dada em (5.12). Os escores de  $\mu, a, b$  e  $\beta$  são obtidos derivando a Equação 5.12 com relação a  $\mu, a, b$  e  $\beta$  respectivamente, e são dados por

$$\begin{aligned}
\frac{\partial l(\mu, a, b, \beta | \mathcal{D})}{\partial \mu} &= \sum_{i=1}^n \frac{\delta_i}{\left[ \mu + \frac{a \exp(\mathbf{x}_i' \beta)}{t_i \exp(\mathbf{x}_i' \beta) + b} \right]} - \sum_{i=1}^n t_i, \\
\frac{\partial l(\mu, a, b, \beta | \mathcal{D})}{\partial a} &= \sum_{i=1}^n \frac{\delta_i \exp(\mathbf{x}_i' \beta) (t_i \exp(\mathbf{x}_i' \beta) + b)}{\left[ \mu + \frac{a \exp(\mathbf{x}_i' \beta)}{t_i \exp(\mathbf{x}_i' \beta) + b} \right] (t_i \exp(\mathbf{x}_i' \beta) + b)^2} + \sum_{i=1}^n \log \left[ \frac{b}{t_i \exp(\mathbf{x}_i' \beta) + b} \right], \\
\frac{\partial l(\mu, a, b, \beta | \mathcal{D})}{\partial b} &= - \sum_{i=1}^n \frac{\delta_i a \exp(\mathbf{x}_i' \beta)}{\left[ \mu + \frac{a \exp(\mathbf{x}_i' \beta)}{t_i \exp(\mathbf{x}_i' \beta) + b} \right] (t_i \exp(\mathbf{x}_i' \beta) + b)^2} \\
&\quad + a \sum_{i=1}^n \frac{(t_i \exp(\mathbf{x}_i' \beta) + b) - b}{\left[ \frac{b}{t_i \exp(\mathbf{x}_i' \beta) + b} \right] (t_i \exp(\mathbf{x}_i' \beta) + b)^2}, \\
\frac{\partial l(\mu, \alpha, \beta | \mathcal{D})}{\partial \beta} &= \sum_{i=1}^n \frac{\delta_i \mathbf{x}_i' \mathbf{A} a \exp(\mathbf{x}_i' \beta)}{\left[ \mu + \frac{a \exp(\mathbf{x}_i' \beta)}{t_i \exp(\mathbf{x}_i' \beta) + b} \right]} - a \sum_{i=1}^n \frac{b t_i \exp(\mathbf{x}_i' \beta) \mathbf{x}_i' \mathbf{A}}{\left[ \frac{b}{t_i \exp(\mathbf{x}_i' \beta) + b} \right] (t_i \exp(\mathbf{x}_i' \beta) + b)^2}, \quad (5.13)
\end{aligned}$$

em que,  $\mathbf{A} = (1, \dots, 1)_{p \times 1}$ . Os estimadores de MV dos parâmetros  $\mu, a, b$  e  $\beta$  são obtidos por métodos numéricos e não possuem forma fechada. A matriz observada de Fisher  $\mathfrak{J}_{\mu, a, b, \beta} = J(\mu, a, b, \beta)$  é dada pelo negativo da matriz de derivadas das equações (5.13) e a matriz  $J(\mu, a, b, \beta)$  é expressa da seguinte forma

$$J(\mu, a, b, \beta) = \begin{bmatrix} \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial \mu^2} & \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial a \partial \mu} & \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial b \partial \mu} & \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial \beta \partial \mu} \\ \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial \mu \partial a} & \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial a^2} & \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial b \partial a} & \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial \beta \partial a} \\ \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial \mu \partial b} & \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial a \partial b} & \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial b^2} & \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial \beta \partial b} \\ \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial \mu \partial \beta} & \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial a \partial \beta} & \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial b \partial \beta} & \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial \beta^2} \end{bmatrix},$$

A distribuição assintótica conjunta para os estimadores de MV dos parâmetros do modelo (5.2) com fragilidade Gama( $a, b$ ) e risco base de uma distribuição Exponencial( $\mu$ ) segue aproximadamente a seguinte distribuição

$$\sqrt{n} \begin{pmatrix} \hat{\mu} - \mu \\ \hat{a} - a \\ \hat{b} - b \\ \hat{\beta} - \beta \end{pmatrix} \xrightarrow{D} N_4(\mathbf{0}, \mathcal{F}_{\mu, a, b, \beta}^{-1}).$$

De forma geral a estimação intervalar assintótica de 95% de confiança para o vetor de parâmetros  $\theta = (\mu, a, b, \beta)$  pode ser obtida através da seguinte expressão

$$\left[ \hat{\theta} - z_{0.975} \sqrt{\mathcal{F}^{-1}(\hat{\theta})}; \hat{\theta} + z_{0.975} \sqrt{\mathcal{F}^{-1}(\hat{\theta})} \right], \quad (5.14)$$

em que  $z_{0.975}$  é o quantil  $1 - 0.05/2$  da Normal padrão e  $\hat{\theta}$  é o estimador de MV para  $\theta$ .

## 5.2 Estudo de Simulação

Nesse estudo os tempos do modelo de fragilidade aditivo Gama-Exponencial ((5.2)) foram gerados pelo método da inversão. Para avaliar o comportamento assintótico das EMV para os parâmetros do modelo utilizamos 1000 réplicas de Monte Carlo e avaliamos o viés, o Erro Quadrático Médio (EQM), a Raiz do Erro Quadrático Médio (REQM), o Erro Absoluto Médio (EAM) e Desvio padrão (DP) para 5 cenários mostrado na [Tabela 22](#)

Tabela 22 – Valores dos parâmetros do modelo (5.2) sob diferentes cenários

Parâmetros	Cenários				
	1	2	3	4	5
$a$	1	1	1	0.5	0.5
$b$	100	10	2	1	2
$\mu$	0.01	0.01	0.1	0.01	0.01
$\beta$	2.2	2.2	1.2	0.5	0.5

Avaliamos os cenários para diferentes tamanhos de amostra  $n = 100, 500$  e  $1000$ . O algoritmo para gerar os dados seguiu a seguinte estrutura

1. determinamos os valores dos parâmetros,  $a, b, \mu$  e  $\beta$ ;
2. geramos  $t'_i$  de  $F(t) = u$ , em que  $u \sim U(0, 1)$ ;
3. geramos  $Z_i \sim U(0, t_{MAX})$ ,  $t_{MAX} = \max(t'_i)$ ;
4.  $t_i = \min(t'_i, z_i)$ ; se  $t_i < z_i$ ,  $\delta_i = 1$ , se  $t_i \geq z_i$ ,  $\delta_i = 0$ .

Pelos resultados simulados apresentados nas [Figura 30](#), [Figura 31](#), [Figura 32](#), [Figura 33](#) e [Figura 34](#) percebemos que para os parâmetros  $a, b, \mu$  e  $\beta$  do modelo de fragilidade aditivo com fragilidade Gama( $a, b$ ) e risco de base Exponencial( $\mu$ ), tanto o viés como o EQM das estimativas diminuem a medida que o tamanho amostral cresce, ou seja, em geral as propriedades assintóticas dos estimadores de máxima verossimilhança são verificadas. Além disso outras medidas complementares como o REQM, EAM e DP também se aproximam de zero quando  $n$  aumenta, o que sugere que as estimativas variam menos em torno da média para grandes amostras.

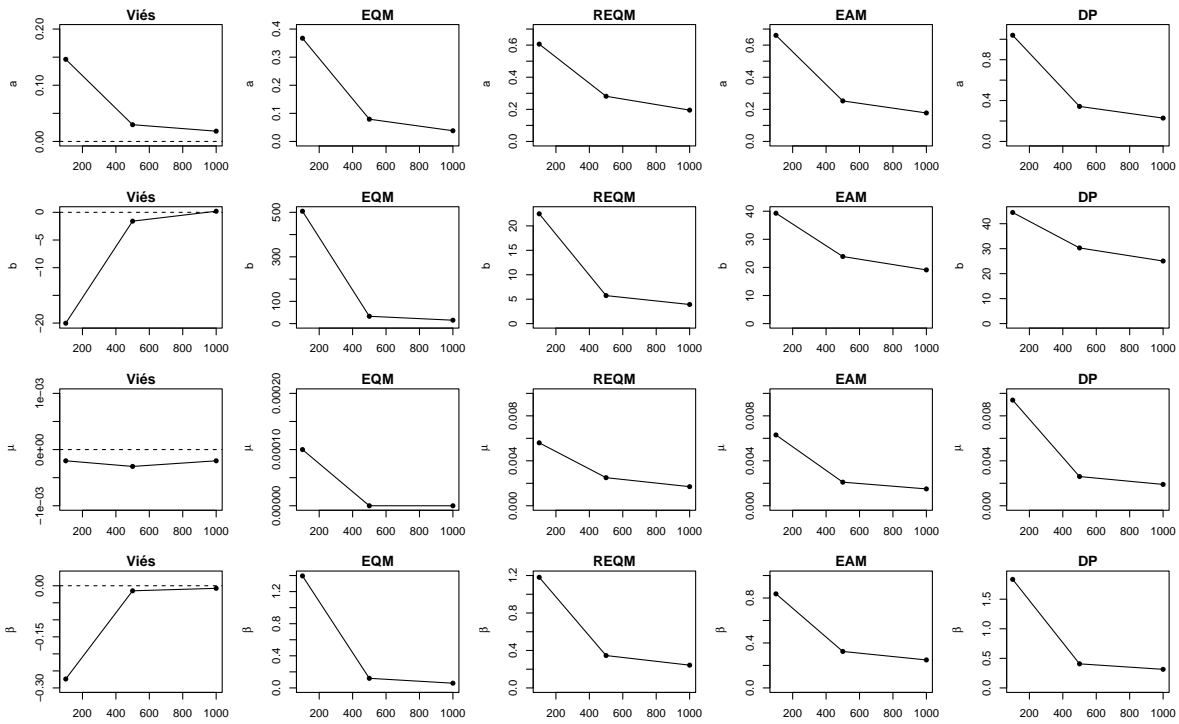


Figura 30 – Viés, EQM, REQM, EAM, e DP de  $(\hat{a}, \hat{b}, \hat{\mu}, \hat{\beta})$  versus  $n$  para o cenário 1 da Tabela 22.

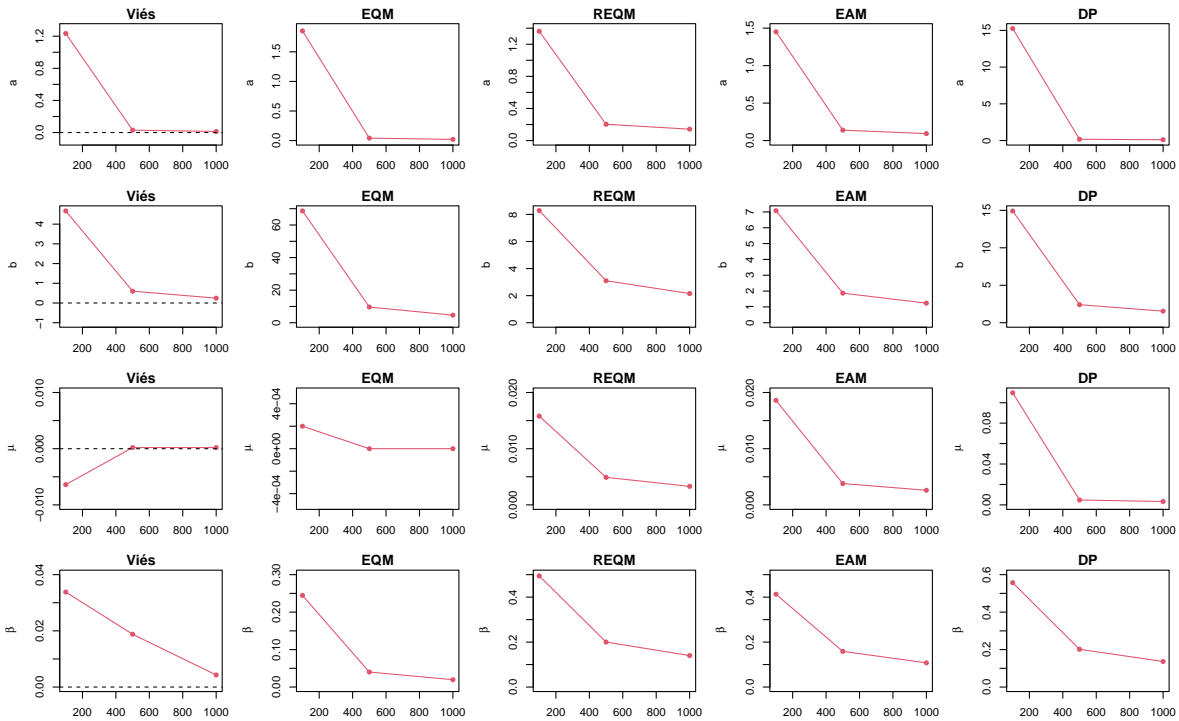


Figura 31 – Viés, EQM, REQM, EAM, e DP de  $(\hat{a}, \hat{b}, \hat{\mu}, \hat{\beta})$  versus  $n$  para o cenário 2 da Tabela 22.

Entretanto, para o parâmetro  $b$  da fragilidade  $\text{Gama}(a, b)$  percebemos que os valores das medidas calculadas não ficaram tão próximos de zero como para os demais parâmetros do



modelo mesmo para  $n = 1000$ , sugerindo que para uma estimação mais precisa desse parâmetro é necessário uma amostra maior.

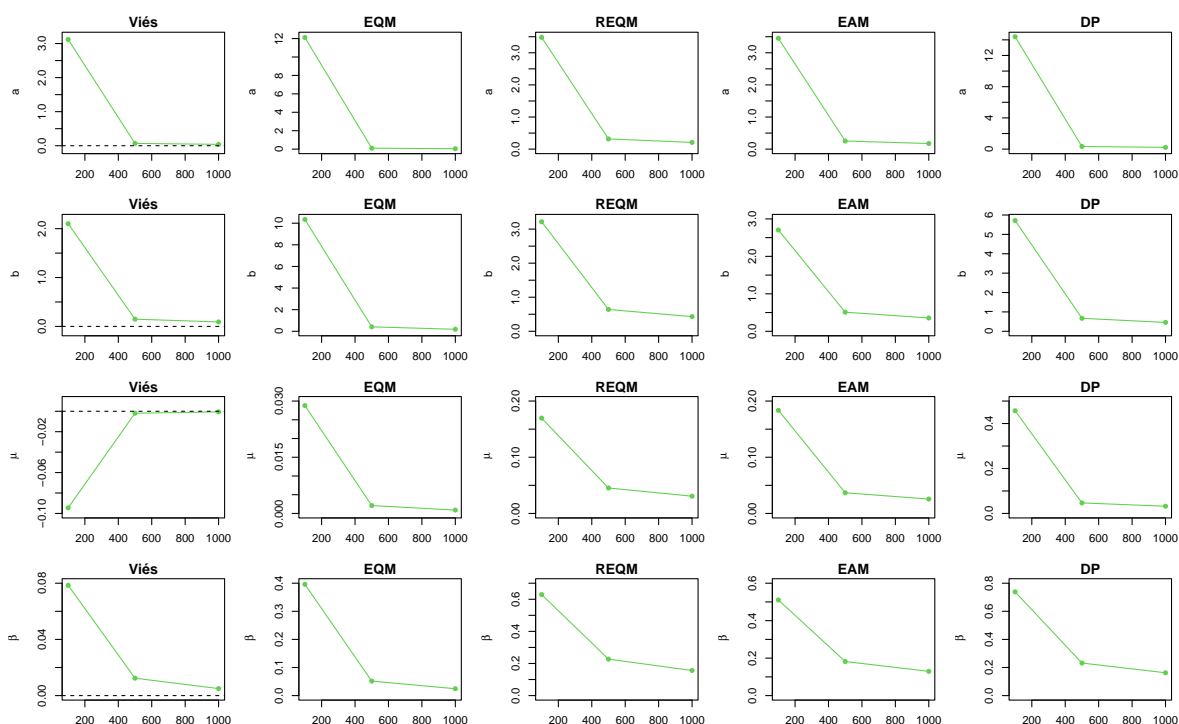


Figura 32 – Viés, EQM, REQM, EAM, e DP de  $(\hat{a}, \hat{b}, \hat{\mu}, \hat{\beta})$  versus  $n$  para o cenário 3 da Tabela 22.

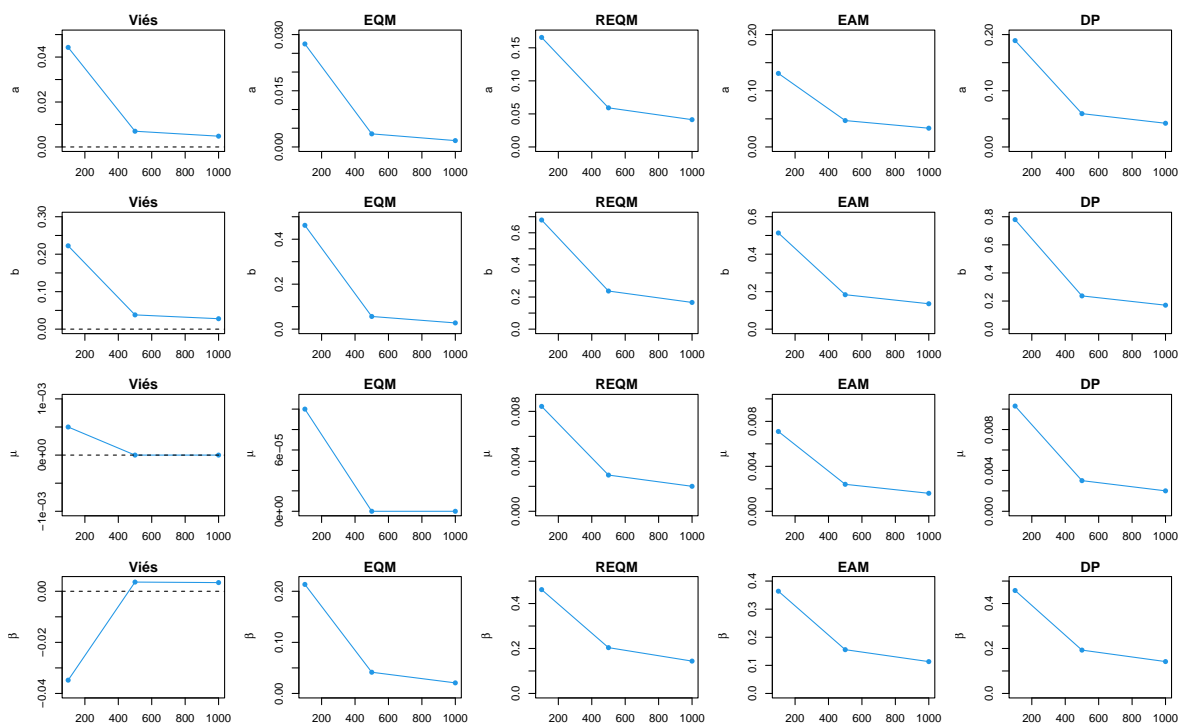


Figura 33 – Viés, EQM, REQM, EAM, e DP de  $(\hat{a}, \hat{b}, \hat{\mu}, \hat{\beta})$  versus  $n$  para o cenário 4 da Tabela 22.

Em geral para todos os cenários simulados os estimadores de máxima verossimilhança comportaram-se de maneira similar e apresentaram resultados dentro do esperado, ou seja, as estimativas obtidas parecem fornecer um desempenho razoável na estimação do modelo de fragilidade aditivo com fragilidade  $\text{Gama}(a, b)$  e risco de base Exponencial( $\mu$ ). Além disso, as probabilidades de cobertura ficaram bem próximas dos valores nominais de 95% em todos os cenários estudados.

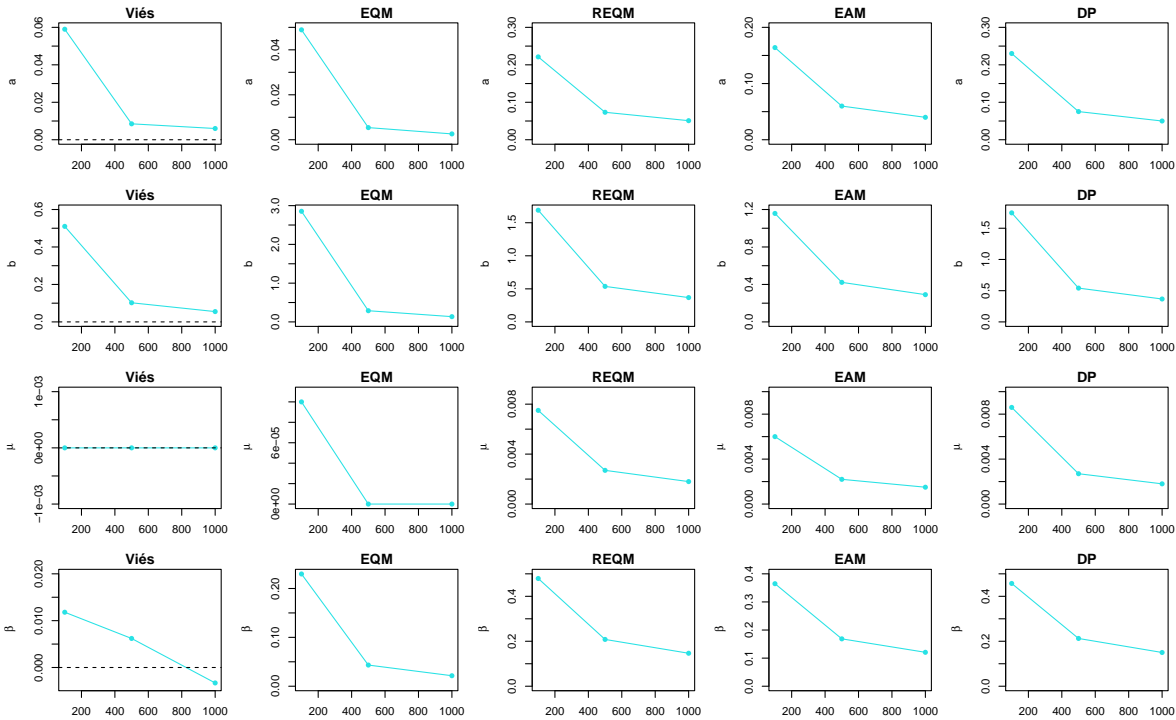


Figura 34 – Viés, EQM, REQM, EAM, e DP de  $(\hat{a}, \hat{b}, \hat{\mu}, \hat{\beta})$  versus  $n$  para o cenário 5 da Tabela 22.

## 5.3 Aplicação

Nessa seção iremos utilizar alguns dos bancos de dados já abordados no Capítulo 2 para ilustrar a modelagem (5.2) e para compararmos com o modelo apresentado no Capítulo 2 em termos de AIC e BIC.

### 5.3.1 Dados Leucemia

Inicialmente vamos aplicar aos dados de Leucemia que tratam de um estudo realizado com 33 pacientes, observados semanalmente, que tinham leucemia aguda. Nesse estudo a covariável indicadora da ausência do antígeno Calla na superfície dos blatos foi observada e a função de sobrevivência estimada por Kaplan-Meier foi ilustrada na Figura 9.

Aplicamos o modelo (5.2) aos dados de Leucemia da Tabela 26 e obtivemos os seguintes resultados (Tabela 23) para a modelagem paramétrica com fragilidades  $\text{Gama}(a, b)$  e riscos de

base da distribuição Exponencial( $\mu$ ) e da distribuição Weibull( $\mu, \gamma$ ) e comparamos os modelos utilizados em termos do AIC e BIC.

Tabela 23 – Estimativas de máxima verossimilhança, erros padrão, intervalos de confiança de 95%, AIC e BIC para os modelos aditivo com fragilidade Gama (5.2) e riscos de base da distribuição Exponencial( $\mu$ ) e da distribuição Weibull( $\mu, \gamma$ ).

Parametros	MFA-Gama-Exponencial			MFA-Gama-Weibull		
	EMV	EP	IC	EMV	EP	IC
$\beta$	2.2435	1.1991	(-0.1068; 4.5938)	1.4175	0.5246	(0.3894; 2.4457)
$a$	0.9890	0.9563	(0.0000; 2.8634)	1.2616	0.3077	(0.6585; 1.8647)
$b$	108.43	6.0408	(96.59 120.27)	52.26	5.6872	(41.12; 63.41)
$\mu$	0.0107	0.0076	(0.0000; 0.0256)	0.0070	0.0006	(0.0059; 0.0081)
$\gamma$	-	-	-	9.6435	8.6318	(0.0000; 26.5615)
AIC		307.41			303.10	
BIC		313.40			310.58	

Pelos critérios AIC e BIC apresentados na Tabela 23, percebemos que o modelo de fragilidade aditiva Gama e risco de base da distribuição Weibull fornece um melhor ajuste aos dados dentre os modelos testados, inclusive quando comparado com os modelos do Capítulo 2, que apresentaram tanto AIC como BIC maiores. A Figura 35 ilustra que as curvas estimadas pelo modelo (5.7) conseguem se ajustar bem as curvas estimadas por Kaplan-Meier (linhas tracejadas).

A estimativa de máxima verossimilhança para o parâmetro  $\beta$  é de 1.4175 e o respectivo intervalo de confiança (0.3894;2.4457) não inclui o valor zero, então a covariável grupo é significativa, e nesse caso ser do grupo Ag+ aumenta a probabilidade de pacientes com leucemia aguda sobreviver.

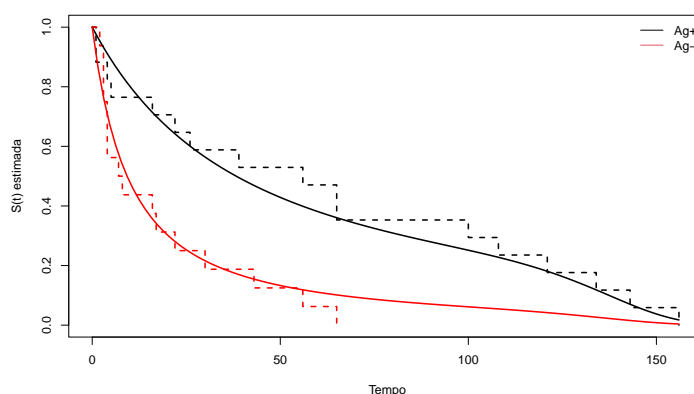


Figura 35 – Curvas de sobrevivência estimadas pelo modelo (5.7) (linhas contínuas) com função de risco de base da distribuição Weibull para os dados de leucemia.

Para calcular as fragilidades individuais apresentadas na Tabela 24 utilizamos as estimativas da Tabela 23 e aplicamos na Equação 5.10, pois todos os indivíduos do estudo falharam.

Observando as fragilidades individuais estimadas apresentadas na [Tabela 24](#) percebemos que em geral, com exceção dos pacientes 14 e 15, os indivíduos mais frágeis estão no Grupo Ag-, ou seja, os indivíduos que não apresentaram o antígeno calla na superfície dos blastos são mais frágeis do que os indivíduos do Grupo Ag+. Quando calculamos a média das fragilidades individuais estimadas (0.0237) obtemos um valor bem próximo da estimativa do valor esperado da variável de fragilidade ( $\hat{a}/\hat{b} = 0.0241$ ).

Tabela 24 – Fragilidade individual estimada dos 33 pacientes com leucemia aguda divididos por grupo com base na [Tabela 23](#).

$i$	Tempos (Grupo Ag+)	$\hat{v}_i$	$i$	Tempos (Grupo Ag-)	$\hat{v}_i$
1	65	0.0192	18	56	0.0080
2	156	0.0063	19	65	0.0070
3	100	0.0131	20	17	0.0185
4	134	0.0076	21	7	0.0279
5	16	0.0331	22	16	0.0191
6	108	0.0114	23	22	0.0158
7	121	0.0091	24	3	0.0350
8	4	0.0402	25	4	0.0329
9	39	0.0248	26	2	0.0374
10	143	0.0069	27	3	0.0350
11	56	0.0209	28	8	0.0265
12	26	0.0289	29	4	0.0329
13	22	0.0305	30	3	0.0350
14	1	0.0425	31	30	0.0128
15	1	0.0425	32	4	0.0329
16	5	0.0395	33	43	0.0098
17	65	0.0192	-	-	-

### 5.3.2 *Dados de câncer de pulmão*

Nesse banco de dados descrito no [Capítulo 2](#) foram incluídos pacientes com diagnóstico de câncer de pulmão no estado de São Paulo, Brasil, entre 2000 e 2014. O acompanhamento foi realizado até 2018. O evento de interesse foi definido como óbito por câncer. O tempo até o evento foi definido como o período desde o diagnóstico de câncer de pulmão até a morte por câncer. Os dados de pacientes que não morreram de câncer de pulmão durante o período de acompanhamento foram censurados à direita. O principal objetivo foi avaliar os efeitos das covariáveis medidas no início do estudo, incluindo sexo, idade no diagnóstico, estágio clínico e cirurgia, e detectar uma possível heterogeneidade não observada.

Aplicamos o modelo (5.2) aos dados de câncer de pulmão e obtivemos os seguintes resultados ([Tabela 25](#)) para a modelagem paramétrica com fragilidades Gama( $a, b$ ) e riscos de base da distribuição Exponencial( $\mu$ ) e da distribuição Weibull( $\mu, \gamma$ ) e comparamos os modelos utilizados em termos do AIC e BIC.

Tabela 25 – Estimativas de máxima verossimilhança (EMV), erros padrão (EP), intervalos de confiança (IC) de 95%, AIC e BIC para os modelos aditivo com fragilidade Gama e riscos de base da distribuição Exponencial( $\mu$ ) e da distribuição Weibull( $\mu, \gamma$ )

MFA-Gama-Exponencial			
Parametros	EMV	EP	IC
$a$	2.7818	0.0929	(2.5997; 2.9640)
$b$	7.3247	0.3731	(6.5935; 8.0560)
$\beta_1$ (sexo)	-0.2557	0.0164	(-0.2878; -0.2237)
$\beta_2$ (idade)	0.1711	0.0162	(0.1393; 0.2029)
$\beta_3$ (cirurgia)	-0.8302	0.0274	(-0.8840; -0.7765)
$\beta_4$ (estágio II)	0.5706	0.0489	(0.4747; 0.6665)
$\beta_5$ (estágio III)	0.9524	0.0450	(0.8641; 1.0406)
$\beta_6$ (estágio IV)	1.5172	0.0490	(1.4211; 1.6133)
$\mu$	$1.7 \times 10^{-8}$	0.0090	(0.0000; 0.0177)
$\gamma$	-	-	-
AIC		64159.08	
BIC		64234.12	
MFA-Gama-Weibull			
Parâmetros	EMV	EP	IC
$a$	1.7680	0.0313	(1.7066; 1.8294)
$b$	2.1665	0.0568	(2.0553; 2.2778)
$\beta_1$ (sexo)	-0.3360	0.0171	(-0.3696; -0.3024)
$\beta_2$ (idade)	0.0405	0.0166	(0.0079; 0.0731)
$\beta_3$ (cirurgia)	-1.0406	0.0234	(-1.0865; -0.9947)
$\beta_4$ (estágio II)	0.1109	0.0429	(0.0269; 0.1950)
$\beta_5$ (estágio III)	0.4220	0.0286	(0.3659; 0.4781)
$\beta_6$ (estágio IV)	1.0816	0.0270	(1.0286; 1.1346)
$\mu$	$2.5 \times 10^{-7}$	$4.0 \times 10^{-6}$	(0.0000; $8.0 \times 10^{-6}$ )
$\gamma$	0.8956	0.3027	(0.3023; 1.4890)
AIC		64941.66	
BIC		65025.04	

Pelos critérios AIC e BIC apresentados na [Tabela 25](#) temos que o modelo aditivo com fragilidade Gama e risco de base da distribuição Exponencial forneceu um ajuste melhor aos dados analisados do que o modelo Gama-Weibull, e por isso modelamos os dados a partir do modelo MFA-Gama-Exponencial. Comparando ainda os modelos estudados nessa capítulo com o modelo do [Capítulo 2](#), percebemos que em termos de AIC e BIC o modelo (5.2) também se mostrou melhor com ambos os riscos de base utilizados nessa aplicação.

De acordo com os resultados da [Tabela 25](#) podemos concluir que todas as covariáveis são significativas. Através do modelo ajustado percebemos que mulheres tem maior tempo de sobrevivência ( $\beta_1 = -0.2557$ ). Como esperado pacientes mais jovens também tem melhor prognóstico ( $\beta_2 = 0.1711$ ), assim como pessoas submetidas à cirurgia ( $\beta_3 = -0.8302$ ). Por fim quanto maior o estágio clínico menor o tempo de vida esperado do paciente.

Calculamos a média e a variância da variável de fragilidade e obtivemos 0.3798 e 0.0518,

respectivamente. Em seguida, utilizamos o método delta para calcular os respectivos erros padrão dessas quantidades e obtivemos os intervalos de confiança assintóticos (0.3400; 0.4195) e (0.0418; 0.0619), respectivamente. Nesse caso podemos admitir a presença de heterogeneidade não observada, o que justifica a inclusão do termo de fragilidade no modelo. Como esperado em geral as fragilidades individuais dos pacientes que falharam são maiores em relação aqueles que não apresentaram o evento de interesse durante o tempo de acompanhamento.

### **5.3.3 Conclusão**

Nesse capítulo consideramos uma modelagem de fragilidade aditiva alternativa ao modelo estudado no [Capítulo 2](#). Nessa proposta assumimos uma forma que garante que o risco estimado sempre será positivo, eliminando uma desvantagem do modelo do [Capítulo 2](#). Fizemos duas aplicações a dados reais onde obtivemos melhores ajustes em termos de AIC e BIC comparado as aplicações com o modelo do [Capítulo 2](#). Além disso, no estudo de simulação apresentado obtivemos resultados que mostram um comportamento aceitável das estimativas de máxima verossimilhança no que se refere às propriedades assintóticas dos estimadores para o modelo proposto.

---

## CONCLUSÕES E TRABALHOS FUTUROS

---

Nos modelos estudados nesse trabalho a inclusão do termo de fragilidade aditivamente à função de risco de base possibilitou quantificar a heterogeneidade não observável, quando covariáveis importantes foram omitidas do estudo. Esses modelos se mostraram úteis nas aplicações aos dados reais utilizados nesse trabalho, entretanto apresentaram algumas dificuldades na estimação clássica para cenários de grande variabilidade e inconsistências quando as covariáveis são incluídas aditivamente ao modelo do [Capítulo 2](#) através da função identidade, podendo levar à estimativas negativas para a função de risco. Dessa forma se fez necessário o estudo de um modelo ([Capítulo 5](#)) que superasse essas dificuldades e inconsistências.

Os estimadores propostos para estimar as fragilidades individuais se mostraram intuitivos, e quando utilizados em dados reais forneceram estimativas condizentes com a média da distribuição da variável de fragilidade, quando assumimos  $V \sim Gama(a, b)$ .

No [Capítulo 4](#) fizemos um estudo da modelagem de mistura padrão com fragilidade aditiva. Essa abordagem se mostrou útil nas aplicações aos dados reais utilizados nesse trabalho e tem por construção uma estrutura capaz de estimar a fração de cura, muito aplicável em diversos tipos de estudos médicos.

No [Capítulo 5](#) o modelo de fragilidade aditivo alternativo se mostrou mais adequado, visto que supera algumas deficiências do modelo do [Capítulo 2](#). Do ponto de vista da estimação dos parâmetros do modelo alternativo, os estimadores de máxima verossimilhança apresentaram resultados dentro do esperado. Além disso, do ponto de vista prático o modelo alternativo apresentado no [Capítulo 5](#) apresentou melhores resultados nas aplicações em dados reais em relação ao modelo do [Capítulo 2](#).

Portanto, com base nesse estudo podemos sugerir o modelo de fragilidade aditivo alternativo apresentado nesse trabalho como uma boa opção na modelagem de dados de sobrevivência. E ainda, concluímos que os estimadores para as fragilidades individuais propostos são úteis para identificar os indivíduos mais fragilizados por fatores não observados.

Como continuação desse trabalho pretendemos:

- Fazer um estudo de sensibilidade e identificabilidade para o modelo de fragilidade aditivo;
- Propor um modelo de fração de cura com fragilidade aditiva utilizando o modelo do [Capítulo 5](#);
- Fazer um estudo comparativo do modelo do [Capítulo 5](#) com outros modelos de fragilidade aditivo;
- Estudar uma nova abordagem para dados de eventos recorrentes baseada no modelo do [Capítulo 5](#).



## REFERÊNCIAS

---

AALEN, O. O. Nonparametric inference for a family of counting processes. **The Annals of Statistics**, JSTOR, n. 6, p. 701–726, 1978. Citado nas páginas 7, 9 e 18.

\_\_\_\_\_. A model for nonparametric regression analysis of counting processes. **Mathematical Statistics and Probability. Lecture Notes in Statistics**, Springer, n. 2, p. 1–25, 1980. Citado nas páginas 7, 9, 17, 18 e 22.

\_\_\_\_\_. A linear regression model for the analysis of lifetimes. **Statistics in Medicine**, n. 8, p. 907–925, 1989. Citado na página 17.

ABU-SHAWER, O.; ABU-SHAWER, M.; SHURMAN, A.; LATTOUF, A.; HAIMOUR, A.; HAMDAN, O.; MANSOUR, R.; ALTAMIMI, T.; AL-HUSSAINI, M. The clinical value of peripheral immune cell counts in pancreatic cancer. **PLoS ONE**, v. 15, n. 6, 2020. Citado na página 37.

AKAIKE, H. A new look at the statistical model identification. **IEEE Transaction on Automatic Control**, v. 19, n. 6, p. 716–723, 1974. Citado na página 24.

BERKSON, J.; GAGE, R. P. Survival curve for cancer patients following treatment. **Journal of the American Statistical Association**, v. 47, p. 501–515, 1952. Citado nas páginas 19, 64 e 65.

BOAG, J. W. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. **Journal of the Royal Statistical Society. Series B (Methodological)**, v. 11, p. 15–53, 1949. Citado nas páginas 19 e 64.

BRESLOW, N. E.; DAY, N. E. The design and analysis of case-control studies. In: \_\_\_\_\_. **Statistical Methods in Cancer Research**. Lyon: IARC, 1980. Vol I. Citado na página 18.

CASELLA, G.; BERGER, R. L. Statistical inference. In: \_\_\_\_\_. **Second edition**. [S.l.]: Thomson Learning, 2002. Citado na página 29.

CLAYTON, D. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. **Biometrika**, v. 65, n. 1, p. 141–151, 1978. Citado nas páginas 7, 9 e 18.

COLOSIMO, E. A. A note on the stratified proportional hazards model. **International Journal of Mathematical and Statistical Sciences**, v. 6, 1997. Citado na página 17.

COLOSIMO, E. A.; GIOLO, S. R. Análise de sobrevivência aplicada. In: \_\_\_\_\_. São Paulo: Blucher, 2006. Citado na página 23.

COLOSIMO, E. A.; NOGUEIRA, M. L. G. O modelo de regressão de cox com covariável dependente do tempo: Uma aplicação envolvendo oacientes infectados pelo hiv. **Revista Brasileira de Estatística**, v. 53, p. 5–13, 1996. Citado na página 17.

COX, D. R. Regression models and life tables (with discussion). **Journal of the Royal Statistical Society, Series B**, v. 34, n. 2, p. 187–220, 1972. Citado nas páginas 7, 9 e 17.

- ELBERS, C.; RIDDER, G. True and spurious duration dependence: The identificability of the proportional hazard model. **Review of Economic Studies XLIX**, p. 403–409, 1982. Citado na página 30.
- FAREWELL, V. A model for binary variable with time-censored observations. **Biometrika**, v. 38, p. 43–46, 1977. Citado na página 19.
- FEIGL, P.; ZELEN, M. Estimation of exponential survival probabilities with concomitant information. **Biometrics**, v. 21, n. 4, p. 826–838, 1965. Citado na página 39.
- GAIL, M.; SANTNER, T.; BROWN, C. An analysis of comparative carcinogenesis experiments based on multiple times to tumor. **Biometrics**, v. 36, p. 255–266, 1980. Citado na página 53.
- GAZON, A. B.; MILANI, E. A.; MOTA, A. L.; LOUZADA, F.; TOMAZELLA, V. L. D.; CALSAVARA, V. F. Nonproportional hazards model with a frailty term formodeling subgroups with evidence of long-term survivors: application to a lung cancer dataset. **Biometrical Journal**, v. 64, p. 105–130, 2022. Citado na página 41.
- GONZALES, J. F. B.; TOMAZELLA, V. L. D.; TACONELLI, J. P. Parametric estimation of the mixture model with gamma frailty in the presence of covariates. **Rev. Bras. Biom.**, v. 31, n. 2, p. 233–247, 2013. Citado nas páginas 19 e 64.
- GUPTA, R. Properties of additive frailty model in survival analysis. **Metrika**, v. 79, n. 1, p. 1–17, 2016. Citado nas páginas 7, 9, 19 e 80.
- GUPTA, R. C.; GUPTA, R. D. General frailty model and stochastic orderings. **Journal of Statistical Planning and Inference**, v. 139, p. 3277–3287, 2009. Citado na página 18.
- HENNINGSEN, A.; TOOMET, O. maxlik: A package for maximum likelihood estimation in R. **Computational Statistics**, v. 26, n. 3, p. 443–458, 2011. Citado na página 29.
- HOUGAARD, P. Frailty models for survival data. **Lifetime Data Analysis**, v. 1, n. 3, p. 255–273, 1995. Citado na página 18.
- KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. **Journal of the American Statistical Association**, v. 53, p. 457–481, 1958. Citado nas páginas 39 e 50.
- KIM, J.; LEE, S. Two sample goodness of fit test for additive risk models with censored observations. **Biometrika**, v. 85, p. 593–603, 1998. Citado na página 80.
- LIN, D.; YING, Z. Semiparametric analysis of the additive risk model. **Biometrika**, n. 81, p. 61–71, 1994. Citado nas páginas 17, 18 e 24.
- \_\_\_\_\_. Additive hazard regression models for survival data. In: \_\_\_\_\_. **Proceedings of the First Seattle Symposium in Biostatistics. Lecture Notes in Statistics**. New York: Springer, 1997. Citado nas páginas 17 e 18.
- LONGINI, I. M.; HALLORAN, M. E. A frailty mixture model for estimating vaccine efficacy. **Applied Statistics**, v. 45, p. 165–173, 1996. Citado nas páginas 19 e 64.
- MANRIQUE, C. M. M. **newTestSurvRec: Statistical Tests to Compare Curves with Recurrent Events**. 2018. Disponível em: <<https://CRAN.R-project.org/package=newTestSurvRec>>. Acesso em: 29/10/2020. Citado na página 55.

- MANTEL, N. Evaluation of survival data and two new rank order statistics arising in its consideration. **Cancer Chemotherapy Reports**, v. 50, p. 163–170, 1966. Citado na página 39.
- MARTINUSSEN, T.; SCHEIKE, T. H. Efficient estimation in additive hazards regression with current status data. **Biometrika**, v. 89, p. 649–658, 2002. Citado na página 18.
- MARTINUSSEN, T.; SCHEIKE, T. H.; ZUCKER, D. M. The aalen additive gamma frailty hazards model. **Biometrika**, v. 98, p. 831–843, 2011. Citado nas páginas 17, 18 e 19.
- MUKHERJEE, S. P.; ROY, D. Bivariate distributions through cox's regression model. In: \_\_\_\_\_. New Delhi: Norosa Publishing Co, 1998. Citado na página 80.
- MUNDA, M.; ROTOLO, F.; LEGRAND, C. parfm: Parametric frailty models in r. **Journal of Statistical Software**, v. 51, n. 11, p. 1–20, 2012. Disponível em: <<https://www.jstatsoft.org/index.php/jss/article/view/v051i11>>. Citado nas páginas 47, 49 e 59.
- NIELSEN, G.; GILL, R.; ANDERSEN, P.; SORENSEN, T. A counting process approach to maximum likelihood estimation in frailty models. **Scandinavian Journal of Statistics**, v. 19, n. 1, p. 25–43, 1992. Citado na página 18.
- PRICE, D. L.; MANATUNAGA, A. K. Modelling survival data with a cured fraction using frailty models. **Statistics in Medicine**, v. 20, p. 1515–1527, 2001. Citado nas páginas 19 e 64.
- R-DEVELOPMENT-CORE-TEAM. **R: A Language and Environment for Statistical Computing**. 2023. Disponível em: <<http://www.R-project.org/>> Acesso em: 29/09/2023. Citado nas páginas 21, 29, 55 e 69.
- ROCHA, C. S. **Modelos com fragilidade em análise de sobrevivência**. Tese (Doutorado) — Universidade de Lisboa, Lisboa, 1995. Citado nas páginas 19, 20 e 25.
- RODRIGUES, J.; CASTRO, M. de; CANCHO, V.; BALAKRISHNAN, N. Com-poisson cure rate survival models and an application to a cutaneous melanoma data. **Journal of Statistical Planning and Inference**, v. 139, p. 3605–3611, 2009. Citado na página 74.
- SCHWARZ, G. Estimating the dimensional of a model. **Annals of Statistics**, v. 6, n. 2, p. 461–464, 1978. Citado na página 24.
- SCUDILIO, J.; CALSAVARA, V. F.; ROCHA, R.; LOUZADA, F.; TOMAZELLA, V.; RODRIGUES, A. S. Defective models induced by gamma frailty term for survival data with cured fraction. **Journal of Applied Statistics**, p. 484–507, 2018. Citado na página 75.
- SILVA, G. L. **Análise Bayesiana de modelos de sobrevivência com fragilidade**. Tese (Doutorado) — Universidade Técnica de Lisboa - Instituto Superior Técnico, Lisboa, 2001. Citado na página 19.
- TELLES, J. P.; YAMAKI, V. N.; CARACANTE, R. A.; MARTINS, V. H. B.; PAIVA, W. S.; TEIXEIRA, M. J.; FIGUEIREDO, E. G.; NEVILLE, I. S. Early versus delayed debridement for surgical site infection after oncological neurosurgery. **Surgical Neurology International**, v. 13, n. 283, 2022. Citado na página 78.
- TOMAZELLA, V.; LOUZADA-NETO, F.; SILVA, G. Bayesian modeling of recurrent events data with an additive gamma frailty distribution and a homogeneous poisson process. **Journal of Statistical Theory and Applications**, v. 5, p. 417–429, 2006. Citado na página 18.

TOMAZELLA, V. L. D. **Modelagem de dados de eventos recorrentes via processo de Poisson com termo de fragilidade**. Tese (Doutorado) — Universidade de São Paulo - ICMC, São Carlos, 2003. Citado nas páginas 18, 19, 20 e 53.

VAUPEL, J.; MANTON, K.; STALLARD, E. The impact of heterogeneity individual frailty on the dynamics of mortality. **Demography**, v. 16, n. 1, p. 439–454, 1979. Citado na página 18.

WANG, M.-C.; CHANG, S.-H. Nonparametric estimation of a recurrent survival function. **Journal of the American Statistical Association**, v. 94, n. 445, p. 146–153, 1999. Citado nas páginas 49, 50, 55, 56 e 61.

WANG, S.; WANG, C.; WANG, P.; SUN, J. Semiparametric analysis of the additive hazards model with informatively interval-censored failure time data. **Computational Statistics and Data Analysis**, v. 125, p. 1–9, 2018. Citado na página 19.

WANG, S. T.; KLEIN, J. P.; MOESCHBERGER, M. L. Semi-parametric estimation of covariate effects using the positive stable frailty model. **Applied Stochastic Models and Data Analysis**, v. 11, p. 121–133, 1995. Citado na página 47.

## CONJUNTOS DE DADOS

### A.1 Dados dos tempos de vida dos 33 pacientes com leucemia aguda divididos por grupo.

Tabela 26 – Dados dos tempos de vida dos 33 pacientes com leucemia aguda divididos por grupo.

id	Tempos (Grupo Ag+)	id	Tempos (Grupo Ag-)
1	65	18	56
2	156	19	65
3	100	20	17
4	134	21	7
5	16	22	16
6	108	23	22
7	121	24	3
8	4	25	4
9	39	26	2
10	143	27	3
11	56	28	8
12	26	29	4
13	22	30	3
14	1	31	30
15	1	32	4
16	5	33	43
17	65	-	-

## A.2 Dados dos tempos recorrentes de detecção de tumores em 48 ratos

Tabela 27 – Dados dos tempos recorrentes de detecção de tumores em 48 ratos.

id	Tempos (Grupo Tratado)	id	Tempos (Grupo Controle)
1	122	24	3, 42, 59, 61,61,112,119,122+
2	122+	25	28,31,35,45,52,59,59,77,85,107,112,122+
3	3,88,122+	26	31,38,48,52,74,77,101,101,119,122+
4	92,122+	27	11,114,122+
5	70,74,85,92,122+	28	35,45,74,74,77,80,85,90,90,122+
6	38,92,122	29	8,8,70,77,122+
7	28,35,45,70,77,107,122+	30	17,35,52,77,101,114,122+
8	92,122+	31	61,24,66,74,101,101,114,122+
9	21,122+	32	8,17,38,42,42,42,122+
10	11,24,66,74,92,122+	33	52,122+
11	56,70,122+	34	28,28, 31,38,52,74,74,77,77,80,80,92,92,122+
12	31,122+	35	17,119,122+
13	3,8,24,35,92,122+	36	52,122+
14	45,92,122+	37	11,11,14,17,52,56,56,80,80,107,122+
15	3,42,92,122+	38	17,35,66,90,122+
16	3,17,52,80,122+	39	28,66,70,70,74,122+
17	17,59,92,101,107,122+	40	3,14,24,24,28,31,35,48,74,77,119,122+
18	45,52,85,101,122	41	21,28,45,56,63,80,85,92,101,101,119,122+
19	92,122+	42	28,35,52,59,66,66,90,97,119,122+
20	21,35,122+	43	8,8,24,42,45,59,63,63,77,101,119,122
21	24,31,42,48,70,74,122+	44	80,122+
22	122+	45	92,122,122
23	31,122+	46	21, 122+
-	-	47	3,28,74,122+
-	-	48	24,74,122

<sup>1</sup>O símbolo + indica censura, ou seja, nenhum tumor detectado no último dia de estudo.

## DESENVOLVIMENTOS ALGÉBRICOS

---

$$\begin{aligned}
\frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial \mu^2} &= \sum_{i=1}^n \frac{\delta_i}{[a/(t_i + b) + \mu + \mathbf{x}'_i \beta]^2}, \\
\frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial a^2} &= - \sum_{i=1}^n \frac{\delta_i (t_i + b)^{-2}}{[a/(t_i + b) + \mu + \mathbf{x}'_i \beta]^2}, \\
\frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial b^2} &= - \sum_{i=1}^n \delta_i \left[ \frac{2a(t_i + b)^{-3} [a/(t_i + b) + \mu + \mathbf{x}'_i \beta] + a^2 (t_i + b)^{-4}}{[a/(t_i + b) + \mu + \mathbf{x}'_i \beta]^2} \right] + \frac{a}{b^2} \sum_{i=1}^n \frac{t_i (t_i + 2b)}{(t_i + b)^2}, \\
\frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial \beta^2} &= \sum_{i=1}^n \frac{\delta_i \mathbf{x}'_i \mathbf{x}_i}{[a/(t_i + b) + \mu + \mathbf{x}'_i \beta]^2}, \\
\frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial \mu \partial a} &= \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial a \partial \mu} = - \sum_{i=1}^n \frac{\delta_i (t_i + b)^{-1}}{[a/(t_i + b) + \mu + \mathbf{x}'_i \beta]^2}, \\
\frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial \mu \partial b} &= \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial b \partial \mu} = \sum_{i=1}^n \frac{\delta_i a (t_i + b)^{-2}}{[a/(t_i + b) + \mu + \mathbf{x}'_i \beta]^2}, \\
\frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial \mu \partial \beta} &= \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial \beta \partial \mu} = \sum_{i=1}^n \frac{\delta_i \mathbf{x}'_i \mathbf{A}}{[a/(t_i + b) + \mu + \mathbf{x}'_i \beta]^2}, \\
\frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial a \partial b} &= - \sum_{i=1}^n \delta_i \left[ \frac{(t_i + b)^{-2} [a/(t_i + b) + \mu + \mathbf{x}'_i \beta] + a (t_i + b)^{-3}}{[a/(t_i + b) + \mu + \mathbf{x}'_i \beta]^2} \right] + \frac{1}{b} \sum_{i=1}^n \frac{t_i}{t_i + b} \\
&= \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial b \partial a} \\
\frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial a \partial \beta} &= \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial \beta \partial a} = - \sum_{i=1}^n \frac{\delta_i \mathbf{x}'_i \mathbf{A} (t_i + b)^{-1}}{[a/(t_i + b) + \mu + \mathbf{x}'_i \beta]^2}, \\
\frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial b \partial \beta} &= \frac{\partial^2 l(\mu, a, b, \beta | \mathcal{D})}{\partial \beta \partial b} = - \sum_{i=1}^n \frac{\delta_i \mathbf{x}'_i \mathbf{A} a (t_i + b)^{-2}}{[a/(t_i + b) + \mu + \mathbf{x}'_i \beta]^2}.
\end{aligned} \tag{B.1}$$

## B.1 Distribuição da fragilidade $V$ dado $T=t$ na parametrização $Gama(a, b)$

$$\begin{aligned}
f(v|T=t) &= \frac{f(t|v)f(v)}{f(t)} = \frac{S(t|v)\lambda(t|v)f(v)}{S(t)\lambda(t)} \\
&= \frac{\exp\{-\mathbf{x}'_i\beta t - vt - \Lambda_0(t)\}(\lambda_0(t) + \mathbf{x}'_i\beta + v)}{\left(\frac{b}{t+b}\right)^a \exp\{-\Lambda_0(t) - \mathbf{x}'_i\beta t\} \left(\frac{a}{t+b} + \lambda_0(t) + \mathbf{x}'_i\beta\right) \Gamma(a)} \frac{b^a}{\Gamma(a)} \exp\{-vb\} v^{a-1} \\
&= \frac{\exp\{-vt\}(\lambda_0(t) + \mathbf{x}'_i\beta + v)}{b^a \left(\frac{a}{t+b} + \lambda_0(t) + \mathbf{x}'_i\beta\right)} \frac{(t+b)^a b^a}{\Gamma(a)} \exp\{-vb\} v^{a-1} \\
&= \frac{(\lambda_0(t) + \mathbf{x}'_i\beta + v)}{\left(\frac{a}{t+b} + \lambda_0(t) + \mathbf{x}'_i\beta\right)} \frac{(t+b)^a}{\Gamma(a)} \exp\{-v(t+b)\} v^{a-1} \\
&= \frac{(\lambda_0(t) + \mathbf{x}'_i\beta)}{\left(\frac{a}{t+b} + \lambda_0(t) + \mathbf{x}'_i\beta\right)} \frac{(t+b)^a}{\Gamma(a)} \exp\{-v(t+b)\} v^{a-1} \\
&\quad + \frac{(t+b)^a}{\left(\frac{a}{t+b} + \lambda_0(t) + \mathbf{x}'_i\beta\right) \Gamma(a)} \exp\{-v(t+b)\} v^a \\
&= \frac{(\lambda_0(t) + \mathbf{x}'_i\beta)}{\left(\frac{a}{t+b} + \lambda_0(t) + \mathbf{x}'_i\beta\right)} \frac{(t+b)^a}{\Gamma(a)} \exp\{-v(t+b)\} v^{a-1} \\
&\quad + \frac{(a(t+b))}{(a(t+b))} \frac{(t+b)^a}{\left(\frac{a}{t+b} + \lambda_0(t) + \mathbf{x}'_i\beta\right) \Gamma(a)} \exp\{-v(t+b)\} v^a \\
&= \frac{(\lambda_0(t) + \mathbf{x}'_i\beta)}{\left(\frac{a}{t+b} + \lambda_0(t) + \mathbf{x}'_i\beta\right)} \frac{(t+b)^a}{\Gamma(a)} \exp\{-v(t+b)\} v^{a-1} \\
&\quad + \frac{\left(\frac{a}{t+b}\right)}{\left(\frac{a}{t+b} + \lambda_0(t) + \mathbf{x}'_i\beta\right)} \frac{(t+b)^{a+1}}{\Gamma(a+1)} \exp\{-v(t+b)\} v^{(a+1)-1} \\
&= w_1 \times f_{G_1} + w_2 \times f_{G_2}, \tag{B.2}
\end{aligned}$$

em que

$$\begin{aligned}
w_1 &= \frac{(\lambda_0(t) + \mathbf{x}'_i\beta)}{\left(\frac{a}{t+b} + \lambda_0(t) + \mathbf{x}'_i\beta\right)}, \\
w_2 &= \frac{\left(\frac{a}{t+b}\right)}{\left(\frac{a}{t+b} + \lambda_0(t) + \mathbf{x}'_i\beta\right)},
\end{aligned}$$

e  $G_1 \sim Gama(a, t+b)$  e  $G_2 \sim Gama(a+1, t+b)$



