

**Um estudo dos modelos de sobrevivência de longa duração
LIGcr e GEPGWcr**

Caroline Amantéa Stella

Dissertação de Mestrado do Programa Interinstitucional de
Pós-Graduação em Estatística (PIPGEs)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Caroline Amantéa Stella

Um estudo dos modelos de sobrevivência de longa duração LIGcr e GEPGWcr

Dissertação 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 Mestra em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística. *VERSÃO REVISADA*

Área de Concentração: Estatística

Orientador: Prof. Dr. Adriano Kamimura Suzuki

USP – São Carlos
Novembro de 2022

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

S824e Stella, Caroline Amantea
Um estudo dos modelos de sobrevivência de longa
duração LIGcr e GEPGWcr / Caroline Amantea Stella;
orientador Adriano Kamimura Suzuki. -- São Carlos,
2022.
121 p.

Dissertação (Mestrado - Programa
Interinstitucional de Pós-graduação em Estatística) --
Instituto de Ciências Matemáticas e de Computação,
Universidade de São Paulo, 2022.

1. Análise de Sobrevivência. 2. Fração de cura.
3. Modelo Logarítmico Inversa Gaussiana. 4. Modelo
Geometric Exponentiated Power Generalized Weibull.
5. Simulação. I. Suzuki, Adriano Kamimura , orient.
II. Título.

Caroline Amantéa Stella

A study of LIGcr and GEPGWcr long-term survival models

Master dissertation submitted to the Institute of Mathematics and Computer Sciences – ICMC-USP and to the Department of Statistics – DEs-UFSCar, in partial fulfillment of the requirements for the degree of the Master Interagency Program Graduate in Statistics.
FINAL VERSION

Concentration Area: Statistics

Advisor: Prof. Dr. Adriano Kamimura Suzuki

USP – São Carlos
November 2022



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa Interinstitucional de Pós-Graduação em Estatística

Folha de Aprovação

Defesa de Dissertação de Mestrado da candidata Caroline Amantéa Stella, realizada em 04/10/2022.

Comissão Julgadora:

Prof. Dr. Adriano Kamimura Suzuki (USP)

Profa. Dra. Gleici da Silva Castro Perdoná (USP)

Profa. Dra. Giovana Oliveira Silva (UFBA)

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa Interinstitucional de Pós-Graduação em Estatística.

Dedico este trabalho aos meus pais, Luiz Sérgio Stella e Valéria Aparecida Amantéa Stella e ao meu esposo Horácio Ansmirie Formenton Filho.

AGRADECIMENTOS

Agradeço primeiramente e imensamente a Deus por todo o cuidado, amor e misericórdia que sempre teve comigo, por todos os sonhos que colocou em meu coração, pela força que me deu nos meus momentos de fraqueza, por toda a calma nos momentos de tempestade, por toda a fé em mim quando nem eu me achava capaz e por ter colocado pessoas maravilhosas no meu caminho. Agradeço também a Virgem Maria, por todo o zelo, força, carinho que me fazia sentir e por ser uma inspiração para mim.

Agradeço imensamente aos meus pais Sérgio e Valéria pelo amor e carinho incondicional que tiveram e têm comigo, por sempre incentivarem os estudos, apoiarem meus sonhos e não medirem esforços para me ajudar a chegar até aqui. Agradeço a minha irmã Camille que sempre me ajuda em tudo o que pode. À minha avó Leontina e em memória dos meus avós Valter, Baptista e Alda por todo o amor, risadas e cuidado deles para comigo seja na terra ou onde estiverem.

Agradeço ao meu companheiro e esposo Horácio Ansmirie Formenton Filho pelo amor, carinho, paciência, apoio, cuidado, conselhos, força que sempre me deu e todo o suporte principalmente na etapa final desta dissertação.

Agradeço ao meu orientador, o professor Dr. Adriano Kamimura Suzuki, pelos conhecimentos compartilhados, por todo suporte neste trabalho, pela paciência, confiança e, por ser exemplo de dedicação e comprometimento como professor e pesquisador. Agradeço a todos os professores que passaram na minha vida, por cada ensinamento e cada forma diferente de ver o mundo. De forma especial ao professor Dr. Mário de Castro Andrade Filho e a professora Dra. Katiane Silva Conceição, que pelas disciplinas ministradas despertaram em mim o amor pela Estatística.

Agradeço aos meus amigos de Birigui, aos meus amigos do Ministério Universidades Renovada (MUR), aos meus colegas da graduação e pós, todos da minha família que rezaram por mim, ao Padre João e o Padre Jonas que sempre me acolheram na Paróquia N. Sra. de Fátima.

Agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Código de Financiamento 001, pelo auxílio financeiro para o desenvolvimento deste trabalho; à Universidade de São Paulo (USP); à Universidade Federal de São Carlos (UFSCAR) e ao Programa Interinstitucional de Pós Graduação em Estatística (PIPGES) pela oportunidade de aprendizado e infraestrutura, às secretarias da pós graduação e de forma especial a Monique da Conceição por toda a sua paciência e preocupação para com os alunos.

Muito obrigada a todas as pessoas e instituições que de alguma forma me ajudaram a estar aqui hoje!

*“Há três caminhos para o fracasso: não ensinar o que se sabe, não praticar o que se ensina, e
não perguntar o que se ignora.”
(São Beda)*

RESUMO

STELLA, C. A. **Um estudo dos modelos de sobrevivência de longa duração LIGcr e GEPGWcr.** 2022. 129 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2022.

Neste trabalho estudamos dois modelos de sobrevivência de longa-duração denominados modelo Logarítmico Inversa Gaussiana com fração de cura (LIGcr) e modelo *Geometric Exponentiated Power Generalized Weibull* com fração de cura (GEPGWcr). Ambos os modelos levam em consideração a existência de heterogeneidade não observada. Para isso, os dois modelos consideram a variável aleatória que representa a fragilidade sendo discreta e começando em zero. Para os dois modelos estudados, os parâmetros são estimados pelo método de Máxima Verossimilhança. Em um estudo de simulação, investigamos as médias das estimativas, os vieses, os erros quadráticos médios e as probabilidades de cobertura. Por fim, mostramos a aplicabilidade dos modelos a conjuntos de dados reais.

Palavras-chave: Análise de Sobrevivência, Fração de cura, Modelo Logarítmico Inversa Gaussiana, Modelo Geometric Exponentiated Power Generalized Weibull, Simulação.

ABSTRACT

STELLA, C. A. **A study of LIGcr and GEPGWcr long-term survival models.** 2022. 129 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2022.

In this work we study two long-term survival models denominated Logarithmic Inverse Gaussian cure rate (LIGcr) model and Geometric Exponentiated Power Generalized Weibull cure rate (GEPGWcr) model. Both models take into account the existence of unobserved heterogeneity. For this, both models consider a random variable that describes the frailty been discrete and starting on zero. For both models studied, the parameters are estimated by the maximum likelihood method. In a simulations study, we investigated the averages of the estimates, the biases, the mean squared errors and coverage probabilities. Finally, we show the applicability of the models to real data sets.

Keywords: Survival Analysis, Cure fraction, Logarithmic Inverse Gaussian Model, Geometric Exponentiated Power Generalized Weibull Model, Simulation.

LISTA DE ILUSTRAÇÕES

Figura 1	– Gráfico com as funções de densidade de probabilidade. Primeiro, no gráfico (a), fixando o parâmetro λ em 1,00 e variando os valores do parâmetro μ . Em seguida, no gráfico (b), variando o parâmetro λ e fixando o parâmetro μ em 1,00.	50
Figura 2	– Gráfico com as funções de sobrevivência. Primeiro, no gráfico (a), fixando o parâmetro λ em 1,00 e variando os valores do parâmetro μ . Em seguida, no gráfico (b), variando o parâmetro λ e fixando o parâmetro μ em 1,00. . . .	51
Figura 3	– Gráfico com as funções risco. Primeiro, no gráfico (a), fixando o parâmetro λ em 1,00 e variando os valores do parâmetro μ . Em seguida, no gráfico (b), variando o parâmetro λ e fixando o parâmetro μ em 1,00.	52
Figura 4	– Gráfico com a função de densidade de probabilidade, função de sobrevivência e função de risco para diferentes valores do parâmetro $p_{0(1)}$ denotado por p na legenda. Utilizamos diferentes valores para $p_{0(1)}$ e fixamos $\mu = 1,00$ e $\lambda = 2,00$	54
Figura 5	– Gráfico com a curva do estimador de KM para os dados de melanoma. . . .	71
Figura 6	– Gráfico com o Tempo Total em Teste considerando os indivíduos censurados para os dados de melanoma.	72
Figura 7	– Gráfico com a curva do estimador de KM juntamente com a curva de sobrevivência do modelo LIGcr ajustado (verde).	73
Figura 8	– Estimativa de Kaplan-Meier da função de sobrevivência juntamente com a estimativa da função de sobrevivência do modelo LIGcr considerando a inclusão da covariável úlcera no parâmetro $p_{0(1)}$. O modelo ajustado para os indivíduos sem úlcera está na cor roxa e o modelo ajustado para os os indivíduos com úlcera está na cor verde-escuro	74
Figura 9	– Estimativa de Kaplan-Meier da função de sobrevivência juntamente com a estimativa da função de sobrevivência do modelo LIGcr considerando a inclusão da covariável úlcera no parâmetro $p_{0(1)}$ e no parâmetro μ . O modelo ajustado para os indivíduos sem úlcera está na cor roxa e o modelo ajustado para os os indivíduos com úlcera está na cor verde-escuro	75
Figura 10	– Gráfico com as funções de densidade de probabilidade para diferentes valores dos parâmetros. No gráfico (a) fixando os parâmetros $\eta = 1,5$, $\psi = 0,5$ e $v = 1,8$ e variando os valores do parâmetro α . No gráfico (b), variando o parâmetro η e fixando os parâmetros $\alpha = 1,5$, $\psi = 0,5$ e $v = 0,8$	83

Figura 11 – Gráfico com as funções de densidade de probabilidade para diferentes valores dos parâmetros. No gráfico (a),fixando os parâmetros $\alpha = 1,2, \eta = 1,0$ e $\nu = 1,8$ e variando os valores do parâmetro ψ . No gráfico (b), variando os valores de ν e fixando os parâmetros $\alpha = 0,5, \eta = 0,8$ e $\psi = 1,2$	84
Figura 12 – Gráfico com as funções de sobrevivência. No gráfico (a), fixando os parâmetros $\eta = 1,5, \psi = 0,5$ e $\nu = 1,8$ e variando os valores do parâmetro α . Em segundo, no gráfico (b), variando o parâmetro η e fixando os parâmetros $\alpha = 1,5, \psi = 0,5$ e $\nu = 0,8$. No gráfico (c),fixando os parâmetros $\alpha = 1,2, \eta = 1,0$ e $\nu = 1,8$ e variando os valores do parâmetro ψ . No gráfico (d), variando os valores de ν e fixando os parâmetros $\alpha = 0,5, \eta = 0,8$ e $\psi = 1,2$	86
Figura 13 – No gráfico (a),fixando os parâmetros $\alpha = 1,2, \eta = 1,0$ e $\nu = 1,8$ e variando os valores do parâmetro ψ . No gráfico (b), variando os valores de ν e fixando os parâmetros $\alpha = 0,5, \eta = 0,8$ e $\psi = 1,2$	87
Figura 14 – Gráfico com as funções de risco para diferentes valores dos parâmetros. No gráfico (a), fixando os parâmetros $\eta = 1,5, \psi = 0,5$ e $\nu = 1,8$ e variando os valores do parâmetro α . No gráfico (b), variando o parâmetro η e fixando os parâmetros $\alpha = 1,5, \psi = 0,5$ e $\nu = 0,8$	89
Figura 15 – No gráfico (c),fixando os parâmetros $\alpha = 1,2, \eta = 1,0$ e $\nu = 1,8$ e variando os valores do parâmetro ψ . No gráfico (d), variando os valores de ν e fixando os parâmetros $\alpha = 0,5, \eta = 0,8$ e $\psi = 1,2$	90
Figura 16 – Gráfico com as funções de densidade de probabilidade, sobrevivência e risco do modelo GEPGW para diferentes valores da proporção de cura, ω e fixando os parâmetros $\alpha = 0,5, \eta = 3,2, \psi = 0,8$ e $\nu = 2,5$	92
Figura 17 – Gráfico com a curva de sobrevivência de KM.	104
Figura 18 – Gráfico com o Tempo Total em Teste considerando os indivíduos censurados.	104
Figura 19 – Gráfico com a curva de sobrevivência de KM juntamente com a curva de sobrevivência do modelo GEPGWcr ajustado (verde).	105
Figura 20 – Curva de sobrevivência de KM dicotomizada pela idade.	107
Figura 21 – Gráfico TTT.	108
Figura 22 – Curva de sobrevivência de KM juntamente com a curva de sobrevivência do modelo GEPGWcr ajustado considerando a covariável idade dicotomizada para o parâmetro $p_{0(2)}$. O modelo ajustado para os indivíduos com idade inferior a 70 anos está na cor roxa e o modelo ajustado para os os indivíduos com idade igual ou maior que 70 está na cor verde-escuro	109

Figura 23 – Curva de sobrevivência de KM juntamente com a curva de sobrevivência do modelo GEPGWcr ajustado considerando a inclusão de covariável idade dicotomizada no parâmetro da proporção de cura e no parâmetro ψ . O modelo ajustado para os indivíduos com idade inferior a 70 anos está na cor **roxa** e o modelo ajustado para os os indivíduos com idade igual ou maior que 70 está na cor **verde-escuro**. 110

LISTA DE TABELAS

Tabela 1 – Simulação do modelo LIGcr considerando $p_{0(1)} = 0,60$, $\mu = 1$ e $\lambda = 1$ para diferentes valores amostrais.	60
Tabela 2 – Simulação do modelo LIGcr considerando $p_{0(1)} = 0,45$, $\mu = 3$ e $\lambda = 4$ para diferentes valores amostrais.	61
Tabela 3 – Simulação do modelo LIGcr considerando $p_{0(1)} = 0,30$, $\mu = 2$ e $\lambda = 1$ para diferentes valores amostrais.	62
Tabela 4 – Simulação do modelo LIGcr considerando $\beta_0 = -1$, $\beta_1 = 4$, $\mu = 1$ e $\lambda = 1$ para diferentes valores amostrais.	63
Tabela 5 – Simulação do modelo LIGcr considerando $\beta_0 = -1,5$, $\beta_1 = 0,75$, $\mu = 2$ e $\lambda = 3$ para diferentes valores amostrais.	64
Tabela 6 – Simulação do modelo LIGcr considerando $\beta_0 = 0,5$, $\beta_1 = 0,7$, $\mu = 3$ e $\lambda = 4$ para diferentes valores amostrais.	65
Tabela 7 – Simulação do modelo LIGcr considerando $\beta_0 = 0,4$, $\beta_1 = 0,8$, $\beta_2 = 0,02$, $\beta_3 = 0,9$ e $\lambda = 4,00$ para diferentes valores amostrais.	66
Tabela 8 – Simulação do modelo LIGcr considerando $\beta_0 = 0,5$, $\beta_1 = -1$, $\beta_2 = 0,9$, $\beta_3 = 0,5$ e $\lambda = 2$ para diferentes valores amostrais.	67
Tabela 9 – Simulação do modelo LIGcr considerando $\beta_0 = -1,2$, $\beta_1 = -0,3$, $\beta_2 = 0,09$, $\beta_3 = 0,5$ e $\lambda = 3$ para diferentes valores amostrais.	68
Tabela 10 – Porcentagem de indivíduos para cada grupo nas covariáveis. Para a Censura, 0 são os indivíduos censurados e 1 os indivíduos que atingiram o evento de falha. Para úlcera = 0 representa os indivíduos sem úlcera e úlcera = 1 os indivíduos com úlcera. Para a variável Sexo, 0 representa o sexo feminino e 1 representa o sexo masculino.	69
Tabela 11 – Porcentagem dos indivíduos considerando a covariável úlcera em relação ao sexo.	70
Tabela 12 – Porcentagem dos indivíduos censurados e não censurados considerando o sexo dos indivíduos.	70
Tabela 13 – Porcentagem dos indivíduos considerando a presença ou ausência de úlcera considerando cada sexo.	70
Tabela 14 – Medidas resumo das variáveis tempo, idade, ano e espessura.	71
Tabela 15 – Estimativa de máxima verossimilhança para os parâmetros $p_{0(1)}$, μ e λ e os seus respectivos erros padrão (entre parênteses).	73

Tabela 16 – Estimativa de máxima verossimilhança e desvio padrão para os parâmetros β_0, β_1, μ e λ considerando a inclusão da covariável úlcera no modelo. . . .	74
Tabela 17 – Estimativa de máxima verossimilhança e desvio padrão para $\beta_0, \beta_1, \beta_2, \beta_3$ e λ considerando a inclusão da covariável úlcera nos parâmetros $p_{0(1)}$ e μ . . .	75
Tabela 18 – Estimativas dos critérios AIC e BIC para o modelo LIGcr para dois cenários: considerando a covariável úlcera apenas em $p_{0(1)}$ e considerando a covariável úlcera em $p_{0(1)}$ e μ	75
Tabela 19 – Simulação do modelo GEPGWcr considerando $\beta_0 = -1, \beta_1 = 1,5, \alpha = 1,5, \eta = 1, \psi = 1$ e $v = 1$ para diferentes valores amostrais.	97
Tabela 20 – Simulação do modelo GEPGWcr considerando $\beta_0 = -1, \beta_1 = 0,5, \alpha = 1,5, \eta = 1,2, \psi = 1$ e $v = 1,8$ para diferentes valores amostrais.	98
Tabela 21 – Simulação do modelo GEPGWcr considerando $\beta_0 = 0,3, \beta_1 = 0,2, \alpha = 2,5, \eta = 0,2, \psi = 0,8$ e $v = 2$ para diferentes valores amostrais.	99
Tabela 22 – Simulação do modelo GEPGWcr considerando $\beta_0 = -1,5, \beta_1 = 0,5, \alpha = 1,5, \eta = 1, \beta_3 = -1, \beta_4 = 0,5$ e $v = 1$ para diferentes valores amostrais. . .	100
Tabela 23 – Simulação do modelo GEPGWcr considerando $\beta_0 = -1,0, \beta_1 = 0,5, \alpha = 1,5, \eta = 1,2, \beta_3 = -1,5, \beta_4 = 1$ e $v = 1$ para para diferentes valores amostrais. 101	
Tabela 24 – Simulação do modelo GEPGWcr considerando $\beta_0 = 0,2, \beta_1 = 0,5, \alpha = 1,4, \eta = 0,5, \beta_3 = -2, \beta_4 = 0,5$ e $v = 1,2$ para diferentes valores amostrais. . .	102
Tabela 25 – Medidas descritivas do tempo de sobrevida dos pacientes com leucemia após o transplante.	103
Tabela 26 – Porcentagem geral de sobrevida dos pacientes.	103
Tabela 27 – Estimativas de máxima verossimilhança e (desvio padrão) dos parâmetros do modelo GEPGWcr.	105
Tabela 28 – Medidas resumo das variáveis tempo, idade e ano do diagnóstico	106
Tabela 29 – Porcentagem dos indivíduos considerando a covariável idade menor que 70 anos(0) e maior/igual a 70 anos(1) em relação ao estágio	106
Tabela 30 – Porcentagem dos indivíduos censurados e não censurados considerando o estágio do câncer de laringe	107
Tabela 31 – Porcentagem dos indivíduos considerando idade menor que 70 anos (0) e idade maior/igual que 70 anos (1) e a censura	107
Tabela 32 – Estimativas de máxima verossimilhanças e (desvios padrão) dos parâmetros do modelo GEPGWcr.	109
Tabela 33 – Estimativas de máxima verossimilhança e (desvios padrão) para os parâmetros do modelo.	110
Tabela 34 – Critérios AIC e BIC para o modelo GEPGWcr para dois cenários: considerando a covariável idade dicotomizada apenas em $p_{0(2)}$ e considerando a covariável idade dicotomizada em $p_{0(2)}$ e ψ	110

LISTA DE ABREVIATURAS E SIGLAS

AIC	<i>Akaike Information Criterion</i>
BIC	<i>Bayesian Information Criterion</i>
DP	Desvio Padrão
EPGW	<i>Eponentiated Power Generalized Weibull</i>
EQM	Erro Quadrático Médio
FGM	Função Geradora de Momentos
FGP	Função Geradora de Probabilidade
GEPGW _{cr}	Geométric <i>Exponentiated Power Generalized Weibull</i> com fração de cura
IG	Inversa Gaussiana
KM	Kaplan-Meier
LIG _{cr}	Logarítmico Inversa Gaussiana com fração de cura
MON	Número de Ordem Médio, <i>Mean-Order-Number</i>
PC	Probabilidade de Cobertura
PGW	<i>Power Generalized Weibull</i>
TTT	Tempo Total em Teste

SUMÁRIO

1	INTRODUÇÃO	25
1.1	Objetivos	27
1.2	Organização do trabalho	27
2	PREFÁCIO	29
2.1	Alguns Conceitos em Análise de Sobrevida	29
2.1.1	<i>O Tempo de Falha e Censuras</i>	29
2.1.2	<i>Função de Sobrevida e Função de Risco</i>	31
2.1.3	<i>A função de verossimilhança</i>	32
2.1.4	<i>Estimadores não paramétrico da sobrevida</i>	33
2.2	Modelos de Sobrevida	33
2.2.1	<i>Modelo de Regressão de Cox</i>	33
2.2.2	<i>Modelos de Longa Duração</i>	34
2.2.3	<i>Modelos com Fragilidade</i>	37
3	MODELO LOGARÍTMICO INVERSA GAUSSIANA	41
3.1	O Modelo LIGcr	41
3.1.1	<i>Inferência Clássica</i>	55
3.2	Estudo de Simulação e Aplicação a dados reais	57
3.2.1	<i>Dados Artificiais</i>	58
3.2.2	<i>Simulação</i>	60
3.2.3	<i>Aplicação a Dados Reais</i>	68
3.2.3.1	<i>Ajuste sem Covariáveis</i>	72
3.2.3.2	<i>Ajuste com Covariáveis</i>	73
4	MODELO GEOMETRIC EXPONENTIATED POWER GENERALIZED WEIBULL	77
4.1	O Modelo GEPGWcr	77
4.1.1	<i>Inferência Clássica</i>	93
4.2	Estudo de Simulação e Aplicação a Dados reais	95
4.2.0.1	<i>Ajuste com covariáveis</i>	105
5	CONSIDERAÇÕES FINAIS	111

REFERÊNCIAS	113
APÊNDICE A A FUNÇÃO LAMBERT	117
APÊNDICE B O GRÁFICO TTT	119
ANEXO A ANEXOS LIG	123
A.1 Demonstração da Proposição 1	123
A.2 FGM da distribuição Logarítmica começando em zero	123
A.3 Esperança e Variância da distribuição Logarítmica começando em zero	124
ANEXO B ANEXOS GEPGW	127
B.1 FGP da distribuição Geométrica começando em zero	127
B.2 A FGM da distribuição Geométrica começando em zero	127
B.3 Esperança e Variância da distribuição Geométrica começando em zero	128

INTRODUÇÃO

O tempo até a ocorrência de um determinado evento também conhecido como evento de interesse ou falha, é um tema que instiga muitos pesquisadores. Um cenário que existe nesse tipo de estudo é a possibilidade de não termos o tempo de vida de um indivíduo ou de um equipamento. Em casos que o tempo de falha não é observado devido a algum acontecimento inesperado, como perda de contato com paciente, morte do paciente (desde que não seja pelo evento de interesse), temos o que é conhecido como censura.

O fato de não termos sempre a observação esperada (o tempo até a falha) faz com que técnicas comuns em estatística que utilizam todos os tempos de falha não possam ser aplicadas, como é o caso de planejamento e experimento ou análise de regressão tradicional. Se, sempre que ocorresse a não observação do tempo de vida, retirássemos cada uma dessas observações do estudo, toda a informação coletada até aquele momento, ou seja, anterior a censura, seria informação perdida até aquele momento. Uma saída para esse problema é a utilização da metodologia de análise de sobrevivência, que permite a utilização de dados censurados no modelo sem descartá-los (COLOSIMO; GIOLO, 2006).

Apesar da metodologia de análise de sobrevivência permitir uma análise com dados censurados, ela tem algumas limitações quando aplicada a conjunto de dados reais no atual avanço da medicina e tratamentos. Em sua abordagem clássica, a análise de sobrevivência supõe que: se um indivíduo é acompanhado por um período de tempo suficientemente longo então este indivíduo atingirá o evento de interesse, a falha. Porém, essa suposição muitas vezes não se aplica aos dados reais fazendo com que as técnicas clássicas de análise de sobrevivência não sejam adequadas nesses cenários. O avanço da medicina torna esse tipo de cenário, indivíduos que não atingem a falha, ainda mais recorrente já que muitas doenças que antes levavam muitos pacientes a morte, como o câncer de mama, passaram a ter uma probabilidade de morte menor do que no passado. Dados com essa proporção de indivíduos que nunca atingem a falha atraem muitos pesquisadores. Exemplo disso são os trabalhos de: Ibrahim, Chen e Sinha (2001), Rodrigues,

Cancho e Castro (2008), Lopes e Bolfarine (2012), Martinez e Achcar (2014), Suzuki, Cancho e Louzada (2016), Ramos e Louzada (2017), Fernandes, Suzuki e Saraiva (2018) que apresentam modelos mais flexíveis considerando a proporção de indivíduos curados.

Modelos que se adequam as bases de dados em que os indivíduos nunca atingem a falha são conhecidos como modelos de longa duração (ou modelos de sobrevivência com fração de cura). Considerando t o tempo até a ocorrência do evento de interesse, nos modelos de sobrevivência com proporção de cura a função de sobrevivência, $S(t)$ para $t \rightarrow \infty$, não atinge zero, ou seja $\lim_{t \rightarrow \infty} S(t) \neq 0$. As funções de sobrevivência com essa característica são conhecidas como funções de sobrevivência impróprias.

Rodrigues, Cancho e Castro (2008) apresentam uma teoria unificada para os modelos de sobrevivência de longa duração por meio de uma composição da função geradora de probabilidade do número de causas de ocorrência do evento de interesse e da função de sobrevivência para o tempo dos pacientes em risco. Além de apresentarem a escrita do modelo de longa duração como modelo de mistura entre indivíduos curados e indivíduos suscetíveis.

Outro problema presente nos conjuntos de dados é a suposição de homogeneidade da população. Essa suposição pode fazer com que o melhor modelo ajustado, mesmo que se considere todas as covariáveis disponíveis, ainda não tenha resultados satisfatórios quando o modelo é utilizado. Isso ocorre porque os indivíduos são muito heterogêneos (tem diferentes fragilidades) e, o pesquisador não consegue observar a heterogeneidade da população com as covariáveis disponíveis. Quando consideramos que uma população é homogênea supomos que, dependendo das covariáveis, qualquer indivíduo tem o mesmo risco de sofrer o evento de interesse, e a função de sobrevivência e a função de risco para cada indivíduo é a mesma (WIENKE, 2010; SOUZA *et al.*, 2017; ZAVALETA, 2016). Na falta de alguma covariável importante para o modelo temos a heterogeneidade não observada, ou seja, supomos que os indivíduos são homogêneos quando eles são heterogêneos. Esse tipo de situação acaba impactando na variabilidade que é maior do que a esperada e, não podendo ser corrigida pelas variáveis coletadas (CARVALHO *et al.*, 2011). Como forma de solução desse tipo de problema, pesquisadores estudam a inclusão de um efeito aleatório (fragilidade) na intenção de captar as diferentes fragilidades dos indivíduos. Modelos com a inclusão desse efeito aleatório são também conhecidos como modelos com fragilidade.

A fragilidade (efeito aleatório) acaba alterando a estimativa pontual proporcionalmente a sua importância. Além disso, a fragilidade modifica o risco base de um indivíduo. Usualmente ela é denotada como uma variável aleatória contínua e não negativa (WIENKE, 2010; ZAVALETA, 2016; CARVALHO *et al.*, 2011; SOUZA *et al.*, 2017; CANCHO *et al.*, 2018). Ao usarmos uma variável aleatória contínua, positiva para denotar a fragilidade nós não consideramos que os indivíduos possam ter risco zero, ou seja, não consideramos que exista uma proporção de indivíduos curados. Uma forma de considerarmos a heterogeneidade não observada e indivíduos com risco zero é denotarmos a fragilidade por uma variável aleatória discreta que pode assumir

valor zero, assim temos uma probabilidade diferente de zero para o ponto de massa da distribuição em zero (ZAVALETA, 2016; SOUZA *et al.*, 2017; CANCHO *et al.*, 2018).

Essa foi a motivação para estudarmos os modelos apresentados neste trabalho. Consideramos a proporção de indivíduos curados e heterogeneidade não observada, assim alguns indivíduos assumiriam risco zero de morrer devido a falha. Por isso, consideramos a distribuição da variável aleatória que representa fragilidade, denotada por Z , sendo uma variável aleatória discreta e começando em zero, e o risco base de um indivíduo $h_0(t)$, passa a ser $Zh_0(t)$.

Suzuki *et al.* (2017) propuseram o modelo com esquemas de ativação latente em que a distribuição Logarítmica estaria associada com as possíveis causas que levam ao evento de interesse. A utilização dessa distribuição é na obtenção de um modelo que leva em consideração as possíveis causas que levam ao evento de interesse podendo ser de 0 a infinitas, neste trabalho estudamos inclusão de efeitos aleatórios devido a heterogeneidade da população que não pode ser explicada pelas covariáveis existentes permitindo que os paciente possam ter risco zero, ou seja, proporção de indivíduos curados.

1.1 Objetivos

Neste trabalho tivemos o objetivo de estudar modelos de longa duração na análise de sobrevivência além de estudar a inclusão de uma fragilidade para cada indivíduo considerando o grupo de indivíduos curados. Para isso estudamos dois modelos que assumem a distribuição da fragilidade (Z) como sendo discreta e começando em zero. As distribuições utilizadas para Z foram a Logarítmica e a Geométrica. Para as distribuições basais foram consideradas a distribuição Inversa Gaussiana para Z com distribuição Logarítmica e a distribuição *Exponentiated Power Generalized Weibull* para Z com distribuição Geométrica. A estimação dos parâmetros foi feita por meio de uma abordagem clássica utilizando o estimador de máxima verossimilhança. Realizamos um estudo de simulação para verificar o ajuste do modelo a conjuntos de dados artificiais. Como ilustração, ajustamos os modelos estudados a conjuntos de dados reais.

1.2 Organização do trabalho

O Capítulo 2 apresenta uma breve revisão de alguns conceitos básicos de Análise de Sobrevivência.

No Capítulo 3 e no Capítulo 4 descrevemos, respectivamente, a construção do modelo Logarítmico Inversa Gaussiana com fração de cura e do modelo Geométrico *Exponentiated Power Generalized Weibul* com fração de cura. Em ambos os capítulos, por meio de uma abordagem clássica, realizamos um estudo de simulação e uma aplicação a dados reais.

Por fim, o Capítulo 5 apresenta as considerações finais juntamente com as perspectivas de trabalhos futuros.

PREFÁCIO

2.1 Alguns Conceitos em Análise de Sobrevivência

Podemos aplicar análise de sobrevivência em muitas áreas como demografia, ciências sociais, engenharia, medicina, entre outras. Especialmente na área médica, que, com os estudos observam um determinado paciente por um período de tempo afim de observar se o paciente sobrevive a uma determinada doença ou tratamento e, se ele sobreviveu, quanto tempo foi a sobrevida desse paciente.

As características principais em dados de análise de sobrevivência são: a censura e o tempo de falha. Nesta seção apresentaremos as definições de censura e o tempo de falha assim como algumas outras definições e conceitos em análise de sobrevivência úteis para o entendimento do desenvolvimento deste trabalho.

2.1.1 *O Tempo de Falha e Censuras*

De uma forma geral definimos o tempo de falha como o tempo até a ocorrência do evento de interesse (a falha), constituído do tempo inicial e a escala de medida (COLOSIMO; GIOLO, 2006).

Não temos uma regra exata para definirmos o tempo inicial, porém normalmente o definimos como o início do estudo. Para a escala de medida, o usual é o tempo real o que também não é regra, já que para os estudos de engenharia essa escala de medida não é utilizada. Definimos a falha (evento de interesse) como o evento que se deseja observar, pode ser a morte do paciente, a recidiva da doença ou eventos recorrentes a um mesmo indivíduo (como acompanharmos o tempo até o indivíduo pegar a primeira gripe, segunda gripe, etc.) por isso a definição do tempo de falha é extremamente importante. De acordo com Colosimo e Giolo (2006), a falha pode ocorrer devido a uma ou mais causas, se temos mais de uma causa de morte do indivíduo, temos o que chamamos de riscos competitivos.

Já a censura é a presença da observação incompletas ou parcial, exemplo disso é a perda de acompanhamento do paciente no decorrer do estudo e a não ocorrência do evento de interesse até o termino do experimento. A censura, pode ser classificada de três modos:

- **Tipo I:** O estudo é terminado após um determinado período de tempo já definido no início do estudo e os indivíduos que não tiveram a falha são censurados.
- **Tipo II:** O estudo só termina após um número pré-determinado de indivíduos atingirem o evento de interesse.
- **Aleatória:** Um indivíduo é retirado no decorrer do estudo sem que a falha tenha ocorrido.

Além das classificações do tipo da censura, também temos os mecanismos da censura. Estes são definidos em três categorias, dessa forma o mecanismo pode ser (COLOSIMO; GIOLO, 2006):

- **À direita:** a ocorrência do evento de interesse acontece depois do início do estudo, ou seja, o tempo do início do estudo até o evento de interesse é maior que o tempo observado.
- **À esquerda:** a ocorrência do evento de interesse acontece antes do início do estudo, ou seja, o tempo de ocorrência do evento de interesse é desconhecido, com a informação de que o evento já ocorreu antes do tempo observado.
- **Intervalar:** o evento de interesse acontece em um intervalo de tempo, geralmente ocorre em estudos que o acompanhamento com os pacientes é feito periodicamente.

Outro ponto importante a ser lembrado é o truncamento, que muitas vezes é confundido com censura. O truncamento caracterizado como a exclusão de certos indivíduos do estudo devido a uma condição. O truncamento pode ser tanto à direita, quanto à esquerda.

Um exemplo de truncamento à direita é em estudos com pacientes de AIDS em que o interesse seria o desenvolvimento da AIDS em pacientes já infectados e a data da infecção seria o tempo de origem. Porém o número total de pacientes infectados que não desenvolveram a doença é desconhecido e dessa forma, esses pacientes não são incluídos no estudo. Estas observações são truncadas à direita, já que apenas os pacientes que tem a doença comprovada fazem parte da amostra.

Para o truncamento à esquerda, um exemplo seria a estimação da distribuição do tempo de vida de moradores de certa localidade em que as observações foram retiradas da base de dados da previdência local, com isso, apenas moradores que atingiram a aposentadoria fazem parte da amostra.

Neste estudo, nos basearemos no mecanismo de censura a direita, situação mais encontrada em estudos de medicina, engenharia e ciências sociais. Segundo Colosimo e Giolo

(2006), os resultados assintóticos em análise de sobrevivência não exigem o reconhecimento do mecanismo de censura, dessa forma temos que as técnicas estatísticas são utilizadas na análise dos três mecanismos de de censura.

Para o indicador de censura, neste trabalho, utilizaremos a notação usual em que t_i é o tempo de falha do i -ésimo indivíduo, para $i = 1, \dots, n$, n o tamanho da amostra e δ_i é o indicador de censura ou falha do i -ésimo indivíduo, apresentado na Equação 2.1:

$$\delta_i = \begin{cases} 1, & \text{se } t_i \text{ é um tempo de falha,} \\ 0, & \text{se } t_i \text{ é um tempo censurado.} \end{cases} \quad (2.1)$$

2.1.2 Função de Sobrevivência e Função de Risco

O tempo de falha, que descreve os tempos de vida de uma população, é uma variável aleatória contínua e não negativa denotada por T com função de densidade de probabilidade dada da forma $f(t; \mathbf{v})$ em que \mathbf{v} é o vetor de parâmetros de interesse. A função de sobrevivência ($S(t; \mathbf{v})$) da análise de sobrevivência em sua forma clássica é denotada como a probabilidade de uma observação não falhar até certo tempo t . Apresentamos a seguir a função de sobrevivência:

$$S(t; \mathbf{v}) = P(T > t; \mathbf{v}) = \int_t^{\infty} f(u; \mathbf{v}) du,$$

em que \mathbf{v} é o vetor de parâmetros da variável T .

Pode-se obter a probabilidade de uma observação não sobreviver ao tempo t em função da distribuição acumulada de T , da seguinte forma:

$$F(t; \mathbf{v}) = 1 - S(t; \mathbf{v}). \quad (2.2)$$

Se considerarmos um intervalo $[t_1, t_2)$ a probabilidade de falha, Equação 2.3, e a taxa de falha (rh), Equação 2.4, são definidas como:

$$P(t_1 \leq T < t_2; \mathbf{v}) = F(t_2; \mathbf{v}) - F(t_1; \mathbf{v}) = S(t_1; \mathbf{v}) - S(t_2; \mathbf{v}) \quad (2.3)$$

e,

$$\frac{P(T \in [t_1, t_2) | T \geq t_1)}{(t_2 - t_1)} = \frac{P(T \in [t_1, t_2))}{(t_2 - t_1)P(T \geq t_1; \mathbf{v})} = \frac{S(t_1; \mathbf{v}) - S(t_2; \mathbf{v})}{\Delta(t)S(t_1; \mathbf{v})}, \quad (2.4)$$

em que $\Delta(t) = t_2 - t_1$.

A taxa de falha é a probabilidade da falha ocorrer no intervalo definido, $[t_1, t_2)$, dado que não ocorreu antes de t_1 e é dividida pelo comprimento de $[t_1, t_2)$.

Se definirmos o intervalo como $[t, t + \Delta t)$, temos que a taxa de falha dada pela [Equação 2.4](#) e escrita da seguinte forma:

$$rh(t; \mathbf{v}) = \frac{S(t; \mathbf{v}) - S(t + \Delta t; \mathbf{v})}{\Delta t S(t; \mathbf{v})},$$

em que $\Delta t = t_2 - t_1$.

Para $\Delta t \rightarrow 0$ em $h(t; \mathbf{v})$, temos a taxa de falha instantânea no tempo t que é condicionada a sobrevivência até este tempo t . Além disso, para $\Delta t \rightarrow 0$, temos a função de risco que pode ser definida como ([RODRIGUES; CANCHO; CASTRO, 2008](#)):

$$\begin{aligned} h(t; \mathbf{v}) &= \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t S(t; \mathbf{v})} \\ &= \frac{1}{S(t; \mathbf{v})} \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t; \mathbf{v})}{\Delta t} = \frac{F'(t; \mathbf{v})}{S(t; \mathbf{v})} = \frac{f(t; \mathbf{v})}{S(t; \mathbf{v})}. \end{aligned} \quad (2.5)$$

A função de risco acumulada é dada pela equação:

$$H(t; \mathbf{v}) = \int_0^t h(u; \mathbf{v}) du.$$

Algumas relações interessantes entre a função de sobrevivência ([Subseção 2.1.2](#)), a função de risco ([Equação 2.5](#)) e a função de densidade acumulada ([Equação 2.2](#)) estão apresentadas a seguir:

$$h(t; \mathbf{v}) = \frac{F'(t; \mathbf{v})}{S(t; \mathbf{v})} = \frac{1}{S(t; \mathbf{v})} \frac{d(1 - S(t; \mathbf{v}))}{dt} = -\frac{d \log(S(t; \mathbf{v}))}{dt}$$

e,

$$f(t; \mathbf{v}) = h(t; \mathbf{v})S(t; \mathbf{v}) = -S'(t; \mathbf{v}).$$

2.1.3 A função de verossimilhança

Considere que o tempo de vida denotado por Y está sujeito a censura à direta.

Assim, em uma amostra de tamanho n e denotando por C_i o i -ésimo tempo de censura, temos que o i -ésimo tempo observado T_i é dado por $T_i = \min\{Y_i, C_i\}$. Além disso, há o interesse no indicador de censura $\delta_i = I(Y_i \leq C_i)$, em que $\delta_i = 1$ se T_i é o tempo de falha e $\delta_i = 0$ se for o tempo de censura, para $i = 1, \dots, n$.

Assim, a função de verossimilhança $L(\mathbf{D}; \boldsymbol{\vartheta})$ é dada por:

$$L(\mathbf{D}; \boldsymbol{\vartheta}) = \prod_{i=1}^n f(t_i; \mathbf{v})^{\delta_i} S(t_i; \mathbf{v})^{1-\delta_i},$$

em que \mathbf{v} é o vetor de parâmetros da distribuição de T , $\boldsymbol{\vartheta}$ é o vetor com todos os vetores de parâmetros, $\mathbf{D} = (\mathbf{t}, \boldsymbol{\delta}, \mathbf{x})$, \mathbf{t} é o vetor dos tempos, \mathbf{x} é o vetor das covariáveis e $\boldsymbol{\delta}$ o vetor das censuras.

O logaritmo da função de verossimilhança é dado por:

$$l(\boldsymbol{\vartheta}|\mathbf{D};) = \sum_{i=1}^n \delta_i \ln(f(t_i; \mathbf{v})) + \sum_{i=1}^n (1 - \delta_i) \ln(S(t_i; \mathbf{v})),$$

em que \mathbf{v} é o vetor de parâmetros da distribuição T , $\boldsymbol{\vartheta}$ é o vetor com todos os vetores de parâmetros, $\mathbf{D} = (\mathbf{t}, \boldsymbol{\delta}, \mathbf{x})$, \mathbf{t} é o vetor dos tempos, \mathbf{x} é o vetor das covariáveis e $\boldsymbol{\delta}$ o vetor das censuras.

2.1.4 Estimadores não paramétrico da sobrevivência

Em análise de sobrevivência temos alguns estimadores não paramétricos, tais como: estimador de Nelson-Aalen proposto por Nelson (1972) e retomado por Aalen (1978), estimador da tabela de vida ou atuarial (COLOSIMO; GIOLO, 2006) e, o estimador de Kaplan-Meier ((KAPLAN; MEIER, 1958)), que será utilizado neste trabalho.

O estimador de Kaplan-Meier (KM), proposto por Kaplan e Meier em 1958, é um estimador não paramétrico muito utilizado para estimações da função de sobrevivência que considera os tempos de falha da amostra estudada. Para a sua obtenção, as seguintes suposições são necessárias (KAPLAN; MEIER, 1958; COLOSIMO; GIOLO, 2006):

- $t_1 < t_2 < \dots < t_k$, para os k tempos de falha ordenados e distintos;
- d_j o número de falhas em $t_j, j = 1, \dots, k$;
- Os indivíduos que não falharam e não foram censurados até o instante imediatamente anterior a t_j é o número de indivíduos sob risco em t_j , que é representado por n_j .

Com isso, o estimador de KM é da forma:

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

2.2 Modelos de Sobrevivência

2.2.1 Modelo de Regressão de Cox

Normalmente utilizamos o termo função basal quando trabalhamos com modelo de riscos proporcionais, também conhecidos como modelo de Regressão de Cox (COX, 1972). Nesta abordagem supomos que a razão do risco de dois indivíduos diferentes é constante no tempo, ou

seja, a razão entre o risco dos dois indivíduos não depende do tempo. Além disso, o modelo de riscos proporcionais é composto pelo componente não paramétrico e o componente paramétrico. O componente não paramétrico é uma função não negativa do tempo, também conhecido como função basal $h_0(t)$. Já o componente paramétrico é uma função $g(\mathbf{x}'\boldsymbol{\beta})$ em que $g(\cdot)$ é uma função não negativa que deve ser especificada de forma que $g(0) = 1$, \mathbf{x} é o vetor de covariáveis e $\boldsymbol{\beta}$ é o vetor de parâmetros associados às covariáveis.

Para os riscos proporcionais na análise de sobrevivência se trabalhamos com a função de risco base (função de risco basal), $h_0(t; \mathbf{v})$, quando $\mathbf{x} = 0$, o resultante é o componente não paramétrico $h_0(t)$. Dessa forma, temos que o modelo de Regressão de Cox é dado por:

$$h(t|\mathbf{x}) = h_0(t)g(\mathbf{x}'\boldsymbol{\beta}).$$

Com a função de risco base podemos obter a função de risco acumulada base e a função de sobrevivência base, respectivamente, por meio das expressões:

$$H_0(t; \mathbf{v}) = \int_0^t h_0(u; \mathbf{v}) du$$

e,

$$S_0(t; \mathbf{v}) = \exp(-H_0(t; \mathbf{v})).$$

Com a função de sobrevivência e a função de risco é possível encontrar: a função de densidade acumulada, a função de densidade de probabilidade. Também, a partir dessas duas últimas funções, é possível encontrar a função de sobrevivência e a função de risco como mostramos nas relações acima.

2.2.2 Modelos de Longa Duração

Berkson e Gage (1952) apresentaram o modelo mais conhecido de mistura padrão, já Rodrigues, Cancho e Castro (2008) apresentam uma escrita unificada dos modelos de sobrevivência com longa duração.

Rodrigues, Cancho e Castro (2008) abordaram conjuntos de dados em que os indivíduos submetidos ao tratamento ficam curados (não atingindo o evento de falha) e indivíduos que atingem o evento de falha. Dividiram os indivíduos desses conjuntos em dois grupos: o grupo dos suscetíveis e o grupo dos curados.

Diferente da distribuição de sobrevivência definida na Subseção 2.1.2, nos modelos de longa duração, a função de sobrevivência tem algumas características diferentes. Quando $t \rightarrow \infty$, temos que $S(t; \boldsymbol{\gamma}) > 0$ é uma função de sobrevivência imprópria e definimos $\boldsymbol{\gamma}$ como o vetor de parâmetros da distribuição imprópria. Para $t \rightarrow \infty$, a função de sobrevivência resulta em $p_0 > 0$, também conhecida como proporção de cura (proporção de indivíduos curados).

A função de sobrevivência imprópria $S(t; \boldsymbol{\gamma})$ ainda pode ser escrita como modelo de mistura com uma função de sobrevivência própria $S(t; \boldsymbol{\nu})$, em que $\boldsymbol{\nu}$ que é o vetor de parâmetros da função de sobrevivência sem o parâmetro referente à proporção de cura.

É possível ter uma flexibilização na condição de normalização imposta a função densidade clássica. Seja

$$\int_0^\infty f(u; \boldsymbol{\gamma}) du = \Upsilon \leq 1,$$

em que $\boldsymbol{\gamma}$ é o vetor de parâmetros da distribuição imprópria. Podemos definir a classe de funções de sobrevivência de forma mais ampla. Denotamos essa forma por $S_{pop}(t; \boldsymbol{\gamma})$ e a expressamos por:

$$S_{pop}(t; \boldsymbol{\gamma}) = 1 - \Upsilon + \int_t^\infty f(u; \boldsymbol{\nu}) du,$$

para $t \geq 0$.

Se $\Upsilon = 1$ então $S_{pop}(t; \boldsymbol{\gamma}) = S(t; \boldsymbol{\gamma})$, ou seja, $S_{pop}(t; \boldsymbol{\gamma})$ contém a função de sobrevivência usual da análise de sobrevivência e $S_{pop}(t = 0; \boldsymbol{\gamma}) = 1$ com $\lim_{t \rightarrow \infty} S_{pop}(t; \boldsymbol{\gamma}) = 1 - \Upsilon$. Sendo $S_{pop}(t; \boldsymbol{\gamma})$ a função de sobrevivência imprópria, $p_0 = 1 - \Upsilon$.

Escrevendo a função de sobrevivência imprópria como modelo de mistura temos que:

$$S_{pop}(t; \boldsymbol{\gamma}) = p_0 + (1 - p_0)S(t; \boldsymbol{\nu}),$$

em que $S(t; \boldsymbol{\nu}) = \int_t^\infty \frac{f(u; \boldsymbol{\nu}) du}{\Upsilon}$ é a função de sobrevivência usual para pacientes não curados, ou seja, $S(t; \boldsymbol{\nu})$ é função de sobrevivência própria.

As mesmas relações da função de taxa de falha, função de densidade acumulada e função de densidade de probabilidade valem para as funções envolvendo a proporção de cura, como é possível ver abaixo:

$$\begin{aligned} h_{pop}(t; \boldsymbol{\gamma}) &= \frac{F'_{pop}(t; \boldsymbol{\gamma})}{S_{pop}(t; \boldsymbol{\gamma})} \\ &= \frac{1}{S_{pop}(t; \boldsymbol{\gamma})} \frac{d(1 - S_{pop}(t; \boldsymbol{\gamma}))}{dt} \\ &= - \frac{d \log(S_{pop}(t; \boldsymbol{\gamma}))}{dt}, \end{aligned}$$

$$f_{pop}(t; \boldsymbol{\gamma}) = h_{pop}(t; \boldsymbol{\gamma}) S_{pop}(t; \boldsymbol{\gamma}) = -S'_{pop}(t; \boldsymbol{\gamma}),$$

$$F_{pop}(t; \boldsymbol{\gamma}) = 1 - S_{pop}(t; \boldsymbol{\gamma})$$

e,

$$H_{\text{pop}}(t; \boldsymbol{\gamma}) = \int_0^t h_{\text{pop}}(u; \boldsymbol{\gamma}) du.$$

Modelos de Longa Duração Unificados

Rodrigues, Cancho e Castro (2008) apresentam o modelo de longa duração em dois estágios: o estágio de iniciação e o estágio de maturação. No estágio de iniciação temos o número de causas ou riscos do evento de interesse (falha). Já no estágio de maturação temos o tempo até a ocorrência do evento de interesse (a falha).

Para o estágio de iniciação, seja M uma variável aleatória que representa o número de riscos para ocorrer um determinado evento de interesse com função distribuição de probabilidade:

$$p_m = P[M = m],$$

para $m = 0, 1, 2, 3, \dots$

Para o estágio de maturação, dado $M = m$, e sejam W_j , $j = 0, 1, \dots, m$, em que os W_j são variáveis aleatórias independentes e contínuas com função de distribuição acumulada (f.d.a.) denotadas por $F(w) = 1 - S(w)$ e independentes da variável aleatória M .

As variáveis aleatórias M e W_j são variáveis latentes (não observadas) e cada W_j representa o tempo de ocorrência do evento devido ao j -ésimo risco ou causa. Dessa forma, temos que T é uma variável aleatória observável ou censura, o tempo de ocorrência da falha é escrito como:

$$T = \min\{W_0, W_1, W_2, \dots, W_M\}. \quad (2.6)$$

Com a suposição que $P[W_0 = \infty] = 1$, neste grupo temos indivíduos que são imunes a doença, ou seja, admitimos a possibilidade de p_0 , uma proporção da população não apresentar falha.

Com as definições: estágio de iniciação, estágio de maturação e, indicando S_{pop} como a função de sobrevivência da população da variável aleatória T , ou seja, $S_{\text{pop}} = P[\text{não ocorrer o evento de interesse no intervalo } (0, t]] = P[T > t]$, temos o Teorema 3 baseado nos Teoremas 1 e 2, todos enunciados e provados por Rodrigues, Cancho e Castro (2008).

Teorema 1. A função de sobrevivência da variável aleatória T correspondente a um modelo de longa duração em dois estágios é dada por:

$$S_{\text{pop}}(t; \boldsymbol{\gamma}) = A_p(S(t; \boldsymbol{\gamma})) = \sum_{m=0}^{\infty} p_m S(t; \mathbf{v})^m,$$

em que $A_p(\cdot)$ é a função geradora da sequência $p = p_m$.

Teorema 2. A função de sobrevivência de longa duração $S_{\text{pop}}(t; \boldsymbol{\gamma})$ é tal que:

$$\lim_{t \rightarrow \infty} S_{\text{pop}}(t; \boldsymbol{\gamma}) = P[M = 0] = p_0,$$

em que p_0 representa a proporção dos indivíduos que são imunes, ou seja, proporção de não ocorrências de falhas na população.

Teorema 3. A função de sobrevivência de longa duração em dois estágios é dada por:

$$S_{pop}(t; \boldsymbol{\gamma}) = p_0 + (1 - p_0)S^*(t; \boldsymbol{v}),$$

em que:

$$S^*(t; \boldsymbol{v}) = \sum_{m=1}^{\infty} p_m^* S(t; \boldsymbol{v})^m \quad \text{e} \quad p_m^* = \frac{p_m}{1 - p_0}.$$

Temos que a função de sobrevivência de longa duração em dois estágios faz parte da classe de funções de sobrevivências impróprias, assim é possível termos interpretação biológica do modelo de mistura padrão para riscos competitivos.

Com a definição de T_i dada pela [Equação 2.6](#), a definição de δ_i dada pela [Equação 2.1](#), utilizando a censura à direita, considerando que $\mathbf{t} = (t_1, t_2, \dots, t_n)^T$, $\mathbf{X} = (x_0, x_1, \dots, x_n)^T$, matriz $n \times (q+1)$, $\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_n)^T$, $\boldsymbol{\zeta}$ o vetor dos parâmetros da distribuição de M_i e dos coeficientes das covariáveis ($\boldsymbol{\beta}$), \boldsymbol{v} o vetor de parâmetros da distribuição e $\boldsymbol{\vartheta}$ o vetor de parâmetros dos vetores $\boldsymbol{\zeta}$ e \boldsymbol{v} . Denotamos os dados observados por $\mathbf{D}_{obs} = (n, \mathbf{X}, \mathbf{t}, \boldsymbol{\delta})$.

A densidade condicional de (t_i, δ_i) dado $M_i = m_i$, para o modelo de longa duração em dois estágios, é dada por :

$$f(t_i, \delta_i | m_i, \boldsymbol{\vartheta}) = S(t_i, \boldsymbol{v})^{m_i - \delta_i} m_i f(t_i; \boldsymbol{v})^{\delta_i} p_{m_i}(\boldsymbol{\zeta}).$$

A função verossimilhança unificada para dois estágios é definida como:

$$\begin{aligned} L(\boldsymbol{\vartheta}, \mathbf{D}_{obs}) &= \prod_{i=1}^n \{f_{pop}(t_i; \boldsymbol{\vartheta})\}^{\delta_i} \{S_{pop}(t_i; \boldsymbol{\vartheta})\}^{1-\delta_i} \\ &= \prod_{i \in \bar{C}} f_{pop}(t_i; \boldsymbol{\vartheta}) \prod_{i \in C} S_{pop}(t_i; \boldsymbol{\vartheta}), \end{aligned}$$

em que C é o conjunto de observações censuradas e \bar{C} é o conjunto de observações não censuradas.

Segundo [Rodrigues, Cancho e Castro \(2008\)](#) a função verossimilhança para dois estágios, com a censura tipo I ou censura aleatória tem as propriedades usuais da função verossimilhança.

2.2.3 Modelos com Fragilidade

A existência de uma heterogeneidade não observada acaba provocando uma maior variabilidade das observações do que é esperado. Eventualmente a variabilidade pode ser explicada com a inclusão de novas covariáveis, porém se o que temos é uma variabilidade genética não há uma variável que podemos utilizar para medir a variabilidade diretamente. Dessa forma temos, uma variabilidade entre os indivíduos e não há uma variável que consiga medir essa variabilidade

diretamente. Assim, é utilizado a inclusão de um efeito aleatório para cada paciente de forma que as estimativas dos efeitos das covariáveis sejam mais consistentes. Dessa forma mesmo faltando informações importantes as variáveis medidas podem estimar o parâmetro e o desvio padrão de forma mais precisa. Além disso, segundo [Carvalho et al. \(2011\)](#), a inclusão do efeito aleatório além de ajudar no problema da heterogeneidade não observada entre grupos também ajuda na estrutura de correlação intragrupos.

Um modelo que utiliza a fragilidade tem o efeito de modificar uma função de taxa de falha base de $h_0(t; \mathbf{v})$ para $Zh_0(t; \mathbf{v})$ para um indivíduo. No modelo de fragilidade consideramos que os riscos são proporcionais a uma determinada variável aleatória (a fragilidade) normalmente denotada por Z ([ZAVALETA, 2016](#)). Pensando em um modelo de Cox com riscos proporcionais incluindo a fragilidade e considerando que \mathbf{x} é o vetor de covariáveis e \mathbf{v} o vetor de parâmetros da distribuição, temos que o modelo passa a ser $h(t; \mathbf{v}) = Zh_0(t; \mathbf{v}) \exp(\mathbf{x}'\boldsymbol{\beta})$ ([CARVALHO et al., 2011](#); [WIENKE, 2010](#)).

Usualmente, a fragilidade é considerada uma variável aleatória contínua e não negativa. No entanto, ela também pode ser uma variável aleatória discreta e não negativa, assumindo que os riscos são proporcionais a ela para os casos que a fragilidade assumiria valores zero ([ZAVALETA, 2016](#)).

A fragilidade assumindo valores iguais a zero nos dá o indicativo que existem indivíduos com risco zero, ou seja, indivíduos que nunca atingirão o evento de interesse, dessa forma, caímos no cenário de dados com proporção de indivíduos curados. Para valores com fragilidade diferentes de zero quanto maior o valor de Z maior risco desse indivíduo atingir o evento de interesse, a falha.

Para a escrita de modelos de sobrevivência considerando a fragilidade, podemos escrever a função de sobrevivência condicionada a Z dada pela [Equação 2.7](#) apresentada a seguir:

$$\begin{aligned} S(t, \mathbf{v}|Z) &= Pr(T > t, \mathbf{v}|Z) = \exp\left(-Z \int_0^t h_0(u, \mathbf{v}) du\right) \\ &= \exp(-ZH_0(t; \mathbf{v})) = (e^{\{-H_0(t; \mathbf{v})\}})^Z \\ &= (S_0(t; \mathbf{v}))^Z. \end{aligned} \tag{2.7}$$

Para a obtenção da função de sobrevivência marginal basta aplicarmos o somatório para os casos que a fragilidade for discreta, ou a integral para os casos que a fragilidade for contínua, como é apresentado pelas equações a seguir:

$$S(t; \mathbf{v}) = \sum_{z=0}^{\infty} S(t|z, \mathbf{v}) p_z = E(S(t|Z, \mathbf{v})) = E((S_0(t; \mathbf{v}))^Z) = G_z[S_0(t; \mathbf{v})] \tag{2.8}$$

e,

$$S(t; \mathbf{v}) = \int S(t, \mathbf{v}|z) dF(z) = E(S(t, \mathbf{v}|Z)) = \mathcal{L}[H_0(t; \mathbf{v})],$$

em que $Pr[Z = z] = p_z$, $G_z[S_0(t; \mathbf{v})]$ é a função de geradora de probabilidade de Z (caso discreto) e \mathcal{L} é a transformada de Laplace em Z (caso contínuo).

Se, na [Equação 2.8](#) temos $Pr(Z = 0) > 0$, a função de sobrevivência é imprópria já que para $t \rightarrow \infty$ a função de sobrevivência é $G_z(0)$ sendo $G_z(0) = Pr(Z = 0) > 0$, o que é uma característica de modelos com fração de cura. Além disso, se $Pr(Z = 0) = 0$ a função de sobrevivência é própria, pois para $t \rightarrow \infty$ a função de sobrevivência é $G_z(0)$ e sendo $G_z(0) = Pr(Z = 0) = 0$. Para ambos os casos comentados a acima, quando $t \rightarrow 0$ temos que a função de sobrevivência é igual a $G_z(1) = 1$.

MODELO LOGARÍTMICO INVERSA GAUSSIANA

3.1 O Modelo LIGcr

A fim de obtermos nosso primeiro modelo estudado considerando longa duração e a fragilidade, temos que estudar uma distribuição discreta no modelo. Para a obter o modelo Logarítmico Inversa Gaussiana com fração de cura (LIGcr) consideramos a variável aleatória Z seguindo distribuição Logarítmica começando em zero. A distribuição Logarítmica usual não começa em zero. [Johnson, Kemp e Kotz \(2005\)](#) enunciam a distribuição Logarítmica como:

$$Pr[X = x; \theta] = \begin{cases} \frac{a\theta^x}{x}, x = 1, 2, 3, \dots \\ \frac{(x-1)\theta}{x} Pr[X = x-1], x = 2, 3, \dots \end{cases},$$

em que $0 < \theta < 1$ e $a = -[\ln(1 - \theta)]^{-1}$.

A Função Geradora de Probabilidade (FGP) é dada por:

$$G_x(s) = \frac{\ln(1 - \theta s)}{\ln(1 - \theta)}. \quad (3.1)$$

Fazendo a modificação de que $x = z + 1$ temos a distribuição Logarítmica começando em zero, também utilizada por [Suzuki et al. \(2017\)](#) onde propuseram um modelo considerando três esquemas de ativação latente com a variável aleatória do número de causas não observáveis do evento de interesse (M) seguindo uma distribuição Logarítmica discreta começando em zero.

Considerando que a distribuição da variável aleatória Z tem distribuição Logarítmica começando em zero, temos que sua função massa de probabilidade, denotada pela [Equação 3.2](#),

é apresentada a seguir:

$$P(Z = z; \theta) = -\frac{1}{\ln(1 - \theta)} \frac{\theta^{(z+1)}}{(z+1)}, \quad z = 0, 1, 2, \dots, \quad (3.2)$$

em que $0 < \theta < 1$.

A FGP da distribuição Logarítmica é apresentada na [Proposição 1](#):

Proposição 1. FGP da distribuição Logarítmica começando em zero:

$$G_z(s) = \frac{\ln(1 - \theta s)}{s \ln(1 - \theta)}, \quad (3.3)$$

para $|s| < 1$.

A prova da [Proposição 1](#) pode ser vista no [Apêndice A](#).

A Função Geradora de Momentos (FGM) da distribuição Logarítmica começando em zero, dada pela [Equação 3.4](#):

$$M_z(a) = \frac{\ln(1 - \theta e^a)}{e^a \ln(1 - \theta)}, \quad (3.4)$$

em que $0 < \theta < 1$.

Por meio da [Equação 3.4](#), é possível encontrar a esperança de Z ($E[Z] = M'_z(0)$) e a variância de Z ($Var[Z] = M''_z(0) - (M'_z(0))^2$), que são apresentadas, na [Equação 3.5](#) e [Equação 3.6](#), respectivamente.

$$\begin{aligned} E[Z] &= M'_z(0) \\ &= -\frac{\theta}{(1 - \theta) \ln(1 - \theta)} - 1. \end{aligned} \quad (3.5)$$

$$\begin{aligned} Var[Z] &= M''_z(0) - (M'_z(0))^2 \\ &= \frac{-\theta}{(1 - \theta)^2 \ln(1 - \theta)} \left(1 + \frac{\theta}{\ln(1 - \theta)} \right). \end{aligned} \quad (3.6)$$

Maiores detalhes dos cálculos para a obtenção da FGM, da esperança e da variância da distribuição Logarítmica começando em zero estão apresentados no [Apêndice A](#).

Definindo a função de sobrevivência basal como $S_0(t; \mathbf{v})$, a função de densidade basal como $f_0(t; \mathbf{v})$ e \mathbf{v} sendo o vetor de parâmetros da distribuição basal. Considerando a distribuição da variável aleatória Z sendo a Logarítmica denotada pela [Equação 3.2](#) com FGP dada pela

[Equação 3.3](#) e usando o resultado de análise de sobrevivência dado pela [Equação 2.8](#) temos que a função de sobrevivência do modelo estudado LIGcr é dada pela [Equação 3.7](#):

$$S_{\text{pop}(1)}(t; \boldsymbol{\gamma}_{(1)}) = G_z[S_0(t; \mathbf{v})] = \frac{\ln(1 - \theta S_0(t; \mathbf{v}))}{S_0(t; \mathbf{v}) \ln(1 - \theta)}, \quad (3.7)$$

para $|S_0(t; \mathbf{v})| < 1$. Com \mathbf{v} o vetor de parâmetros da distribuição basal, sendo a distribuição basal, nesse caso, a distribuição Inversa Gaussiana (IG); $S_{\text{pop}(1)}$ a distribuição de longa duração do modelo LIGcr (modelo estudado 1), em que $\boldsymbol{\gamma}_{(1)}$ é o seu vetor de parâmetros.

Temos que $S_0(t; \mathbf{v})$ é uma função de sobrevivência própria, ou seja, para $t \rightarrow 0$ temos que $S_0(t; \mathbf{v}) = 1$ e para $t \rightarrow \infty$ temos que $S_0(t; \mathbf{v}) = 0$.

Para estudar um pouco mais o comportamento da função $S_{\text{pop}(1)}(t; \boldsymbol{\gamma}_{(1)})$ aplicamos o limite para $t \rightarrow 0$ ([Equação 3.8](#)) e para $t \rightarrow \infty$ ([Equação 3.9](#)), mostrados a seguir:

$$\lim_{t \rightarrow 0} S_{\text{pop}(1)}(t; \boldsymbol{\gamma}_{(1)}) = \lim_{t \rightarrow 0} \frac{\ln(1 - \theta S_0(t; \mathbf{v}))}{S_0(t; \mathbf{v}) \ln(1 - \theta)} = 1 \quad (3.8)$$

e,

$$\lim_{t \rightarrow \infty} S_{\text{pop}(1)}(t; \boldsymbol{\gamma}_{(1)}) = \lim_{t \rightarrow \infty} \frac{\ln(1 - \theta S_0(t; \mathbf{v}))}{S_0(t; \mathbf{v}) \ln(1 - \theta)} = \frac{0}{0} \quad (\text{indeterminação}).$$

Aplicando a regra do L'Hopital, temos que:

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{(\ln(1 - \theta S_0(t; \mathbf{v})))'}{(S_0(t; \mathbf{v}) \ln(1 - \theta))'} &= \lim_{t \rightarrow \infty} \frac{\frac{1}{1 - \theta S_0(t; \mathbf{v})} (-\theta S_0'(t; \mathbf{v}))}{S_0'(t; \mathbf{v}) \ln(1 - \theta)} \\ &= \lim_{t \rightarrow \infty} \frac{(-\theta S_0'(t; \mathbf{v}))}{1 - \theta S_0(t; \mathbf{v})} \frac{1}{S_0'(t; \mathbf{v}) \ln(1 - \theta)} \\ &= \lim_{t \rightarrow \infty} \frac{-\theta}{(1 - \theta S_0(t; \mathbf{v})) \ln(1 - \theta)} \\ &= \frac{-\theta}{\ln(1 - \theta)} \lim_{t \rightarrow \infty} \frac{1}{(1 - \theta S_0(t; \mathbf{v}))} \\ &= \frac{-\theta}{\ln(1 - \theta)}. \end{aligned} \quad (3.9)$$

Na [Equação 3.9](#) observamos que a função de sobrevivência do modelo estudado é uma função de sobrevivência imprópria e tem uma proporção de cura sendo $-\theta / \ln(1 - \theta)$.

Utilizando as relações de análise de sobrevivência é possível encontrarmos a função densidade de probabilidade ([Equação 3.10](#)) e a função risco ([Equação 3.11](#)) relacionadas da função de sobrevivência do modelo estudado. Para isso utilizamos as relações: $f_{\text{pop}(1)}(t; \boldsymbol{\gamma}_{(1)}) = -S_{\text{pop}(1)}'(t; \boldsymbol{\gamma}_{(1)})$ e $h_{\text{pop}(1)}(t; \boldsymbol{\gamma}_{(1)}) = f_{\text{pop}(1)}(t; \boldsymbol{\gamma}_{(1)}) / S_{\text{pop}(1)}(t; \boldsymbol{\gamma}_{(1)})$.

A função de densidade de probabilidade associada a função de sobrevivência do modelo LIGcr é:

$$\begin{aligned}
 f_{\text{pop}(1)}(t; \boldsymbol{\gamma}_{(1)}) &= - \left(\frac{d}{dt} (S_{\text{pop}}(t; \boldsymbol{\gamma}_{(1)})) \right) = - \left(\frac{d}{dt} \left(\frac{\ln(1 - \theta S_0(t; \mathbf{v}))}{S_0(t; \mathbf{v}) \ln(1 - \theta)} \right) \right) \\
 &= - \left(\frac{(\ln(1 - \theta S_0(t; \mathbf{v})))' (S_0(t; \mathbf{v}) \ln(1 - \theta)) - \ln(1 - \theta S_0(t; \mathbf{v})) (S_0(t; \mathbf{v}) \ln(1 - \theta))'}{(S_0(t; \mathbf{v}) \ln(1 - \theta))^2} \right) \\
 &= - \left(\frac{\left(\frac{-\theta S'_0(t; \mathbf{v}) (S_0(t; \mathbf{v}) \ln(1 - \theta))}{1 - \theta S_0(t; \mathbf{v})} \right) - \ln(1 - \theta S_0(t; \mathbf{v})) (S'_0(t; \mathbf{v}) \ln(1 - \theta))}{(S_0(t; \mathbf{v}))^2 (\ln(1 - \theta))^2} \right) \\
 &= - \left(\frac{\ln(1 - \theta) \left(\frac{(-\theta S'_0(t; \mathbf{v})) S_0(t; \mathbf{v})}{1 - \theta S_0(t; \mathbf{v})} - S'_0(t; \mathbf{v}) \ln(1 - \theta S_0(t; \mathbf{v})) \right)}{(S_0(t; \mathbf{v}))^2 (\ln(1 - \theta))^2} \right) \\
 &= -S'_0(t; \mathbf{v}) \left(\frac{-\theta S_0(t; \mathbf{v})}{1 - \theta S_0(t; \mathbf{v})} - \ln(1 - \theta S_0(t; \mathbf{v})) \right) \frac{1}{(S_0(t; \mathbf{v}))^2 \ln(1 - \theta)}.
 \end{aligned}$$

Como $f(t; \mathbf{v}) = -S'(t; \mathbf{v})$, temos que:

$$\begin{aligned}
 f_{\text{pop}(1)}(t; \boldsymbol{\gamma}_{(1)}) &= \frac{f_0(t; \mathbf{v})}{\ln(1 - \theta) (S_0(t; \mathbf{v}))^2} \left(\frac{-\theta S_0(t; \mathbf{v})}{1 - \theta S_0(t; \mathbf{v})} - \ln(1 - \theta S_0(t; \mathbf{v})) \right) \\
 &= - \frac{f_0(t; \mathbf{v})}{\ln(1 - \theta) (S_0(t; \mathbf{v}))^2} \left(\frac{\theta S_0(t; \mathbf{v})}{1 - \theta S_0(t; \mathbf{v})} + \ln(1 - \theta S_0(t; \mathbf{v})) \left(\frac{1 - \theta S_0(t; \mathbf{v})}{1 - \theta S_0(t; \mathbf{v})} \right) \right) \\
 &= - \frac{f_0(t; \mathbf{v}) (\theta S_0(t; \mathbf{v}) + (1 - \theta S_0(t; \mathbf{v})) \ln(1 - \theta S_0(t; \mathbf{v})))}{\ln(1 - \theta) (S_0(t; \mathbf{v}))^2 (1 - \theta S_0(t; \mathbf{v}))}. \quad (3.10)
 \end{aligned}$$

A função de risco do modelo LIGcr é da forma:

$$h_{\text{pop}(1)}(t; \boldsymbol{\gamma}_{(1)}) = \frac{-f_0(t; \mathbf{v}) [\theta S_0(t; \mathbf{v}) + (1 - \theta S_0(t; \mathbf{v})) \ln(1 - \theta S_0(t; \mathbf{v}))]}{S_0(t; \mathbf{v}) \ln(1 - \theta) (1 - \theta S_0(t; \mathbf{v}))},$$

em que $f_0(t; \mathbf{v})$ e $S_0(t; \mathbf{v})$ são as funções de densidade e sobrevivência basais, respectivamente.

Note que na [Equação 3.7](#), [Equação 3.10](#) e [Equação 3.11](#) temos a função de sobrevivência, densidade e risco, respectivamente, do modelo LIGcr que são as mesmas equações de sobrevivência, densidade e risco no esquema de primeira ativação do modelo proposto por [Suzuki et al. \(2017\)](#). A diferença entre ambos é a interpretação em cada caso como comentado na [Capítulo 1](#). Uma outra diferença são as distribuições basais utilizadas. [Suzuki et al. \(2017\)](#) considera a distribuição Weibull e neste trabalho consideramos a Inversa Gaussiana.

Representação como Modelo de Mistura

[Rodrigues, Cancho e Castro \(2008\)](#) apresentam que a função de sobrevivência imprópria pode ser escrita como um modelo de mistura em que a função de sobrevivência imprópria seria a

proporção de cura somado a proporção de indivíduos em risco multiplicado por uma função de sobrevivência própria, como comentado na [Subseção 2.2.2](#).

Seguindo o resultado de [Rodrigues, Cancho e Castro \(2008\)](#) e considerando a proporção de cura do modelo LIGcr sendo $p_{0(1)}$, temos que a função de sobrevivência pode ser escrita da forma:

$$\begin{aligned} S_{\text{pop}}(t; \boldsymbol{\gamma}_{(1)}) &= p_{0(1)} + (1 - p_{0(1)})S_{(1)}^*(t; \mathbf{v}^*) \\ &= \frac{-\theta}{\ln(1 - \theta)} + \left[1 - \left(\frac{-\theta}{\ln(1 - \theta)} \right) \right] S_{(1)}^*(t; \mathbf{v}^*), \end{aligned}$$

em que $S_{(1)}^*(t; \mathbf{v}^*)$ é a equação de sobrevivência dos indivíduos em risco do modelo estudado LIGcr e \mathbf{v}^* é o vetor de parâmetros da função de sobrevivência dos indivíduos em risco $S_{(1)}^*$.

É possível obter a função de sobrevivência $S_{(1)}^*(t; \mathbf{v}^*)$ como mostramos na equação abaixo:

$$S_{\text{pop}(1)}(t; \boldsymbol{\gamma}_{(1)}) = \frac{\ln(1 - \theta S_0(t; \mathbf{v}))}{S_0(t; \mathbf{v}) \ln(1 - \theta)} = \frac{-\theta}{\ln(1 - \theta)} + \left[1 - \left(\frac{-\theta}{\ln(1 - \theta)} \right) \right] S_{(1)}^*(t; \mathbf{v}^*).$$

Isolando $S_{(1)}^*(t; \mathbf{v}^*)$, temos que a função de sobrevivência dos indivíduos em risco dada pela [Equação 3.11](#) é:

$$\begin{aligned} S_{(1)}^*(t; \mathbf{v}^*) &= \frac{\theta S_0(t; \mathbf{v}) + \ln(1 - \theta S_0(t; \mathbf{v}))}{S_0(t; \mathbf{v})(\theta + \ln(1 - \theta))} \\ &= \frac{\theta}{\theta + \ln(1 - \theta)} + \frac{\ln(1 - \theta S_0(t; \mathbf{v}))}{S_0(t; \mathbf{v})(\theta + \ln(1 - \theta))}. \end{aligned} \quad (3.11)$$

Podemos verificar se a função de sobrevivência é própria pois, sendo própria quando $t \rightarrow 0$, $S_{(1)}^*(t; \mathbf{v}^*) \rightarrow 1$ e quando $t \rightarrow \infty$, $S_{(1)}^*(t; \mathbf{v}^*) \rightarrow 0$. Apresentamos as contas dessa verificação nas equações a seguir:

Se $t \rightarrow 0$

$$\lim_{t \rightarrow 0} S_{(1)}^*(t; \mathbf{v}^*) = \frac{\theta S_0(t; \mathbf{v}) + \ln(1 - \theta S_0(t; \mathbf{v}))}{S_0(t; \mathbf{v})(\theta + \ln(1 - \theta))} = \frac{\theta + \ln(1 - \theta)}{\theta + \ln(1 - \theta)} = 1.$$

Se $t \rightarrow \infty$

$$\lim_{t \rightarrow \infty} S_{(1)}^*(t; \mathbf{v}^*) = \lim_{t \rightarrow \infty} \frac{\theta S_0(t; \mathbf{v}) + \ln(1 - \theta S_0(t; \mathbf{v}))}{S_0(t; \mathbf{v})(\theta + \ln(1 - \theta))} = \frac{0}{0} \text{ (indeterminação).}$$

Aplicando L'Hopital, temos que:

$$\begin{aligned}
\lim_{t \rightarrow \infty} S_{(1)}^*(t; \mathbf{v}^*) &= \lim_{t \rightarrow \infty} \frac{d(\theta S_0(t; \mathbf{v}) + \ln(1 - \theta S_0(t; \mathbf{v}))) / dt}{d(S_0(t; \mathbf{v})(\theta + \ln(1 - \theta))) / dt} \\
&= \lim_{t \rightarrow \infty} \frac{\frac{1}{1 - \theta S_0(t; \mathbf{v})} (-\theta S'_0(t; \mathbf{v})) + \theta S'_0(t; \mathbf{v})}{S'_0(t; \mathbf{v})(\theta + \ln(1 - \theta))} \\
&= \lim_{t \rightarrow \infty} -\frac{\theta^2 S_0(t; \mathbf{v})}{(1 - \theta S_0(t; \mathbf{v}))(\theta + \ln(1 - \theta))} \\
&= \lim_{t \rightarrow \infty} \frac{-\theta + \theta(1 - \theta S_0(t; \mathbf{v}))}{(1 - \theta S_0(t; \mathbf{v}))(\theta + \ln(1 - \theta))} \\
&= \frac{-\theta + \theta(1 - 0)}{\theta + \ln(1 - \theta)} \\
&= \frac{-\theta + \theta}{\theta + \ln(1 - \theta)} = 0.
\end{aligned}$$

Além disso, também podemos obter a função de densidade de probabilidade e a função de risco relacionadas com a função de sobrevivência $S_{(1)}^*(t; \mathbf{v}^*)$.

Para obtermos a função de densidade de probabilidade relacionada com a função de sobrevivência $S_{(1)}^*(t; \mathbf{v}^*)$ utilizamos a relação de análise de sobrevivência clássica, $f_{(1)}^*(t; \mathbf{v}^*) = -d(S_{(1)}^*(t; \mathbf{v}^*)) / dt$ como mostra a equação abaixo:

$$\begin{aligned}
f_{(1)}^*(t; \mathbf{v}^*) &= -\frac{d}{dt} \left(\frac{\theta}{\theta + \ln(1 - \theta)} + \frac{\ln(1 - \theta S_0(t; \mathbf{v}))}{S_0(t; \mathbf{v})(\theta + \ln(1 - \theta))} \right) \\
&= -\left(\frac{\frac{1}{1 - \theta S_0(t; \mathbf{v})} [-\theta S'_0(t; \mathbf{v})] [S_0(t; \mathbf{v})(\theta + \ln(1 - \theta))] - \ln(1 - \theta S_0(t; \mathbf{v})) S'_0(t; \mathbf{v}) [\theta + \ln(1 - \theta)]}{(S_0(t; \mathbf{v}))^2 (\theta + \ln(1 - \theta))^2} \right) \\
&= -\left(\frac{\frac{[-\theta S'_0(t; \mathbf{v})] [S_0(t; \mathbf{v})(\theta + \ln(1 - \theta))] - (1 - \theta S_0(t; \mathbf{v})) (\ln(1 - \theta S_0(t; \mathbf{v})) S'_0(t; \mathbf{v}) [\theta + \ln(1 - \theta)])}{1 - \theta S_0(t; \mathbf{v})}}{(S_0(t; \mathbf{v}))^2 (\theta + \ln(1 - \theta))^2} \right) \\
&= -\left[\frac{(-S'_0(t; \mathbf{v})) (\theta S_0(t; \mathbf{v}) + \ln(1 - \theta S_0(t; \mathbf{v}))) - \theta S_0(t; \mathbf{v}) \ln(1 - \theta S_0(t; \mathbf{v}))}{(S_0(t; \mathbf{v}))^2 (\theta + \ln(1 - \theta))^2 (1 - \theta S_0(t; \mathbf{v}))} \right] \\
&= \frac{f_0(t; \mathbf{v}) [\theta S_0(t; \mathbf{v}) (\ln(1 - \theta S_0(t; \mathbf{v})) - 1) - \ln(1 - \theta S_0(t; \mathbf{v}))]}{(S_0(t; \mathbf{v}))^2 (\theta + \ln(1 - \theta))^2 (1 - \theta S_0(t; \mathbf{v}))},
\end{aligned}$$

em que $f_0(t; \mathbf{v})$ é a função de densidade basal e $S_0(t; \mathbf{v})$ a função de sobrevivência basal.

Para a função de risco relacionada com a função de sobrevivência $S_{(1)}^*(t; \mathbf{v}^*)$ a relação utilizada foi $h_{(1)}^*(t; \mathbf{v}^*) = f_{(1)}^*(t; \mathbf{v}^*) / S_{(1)}^*(t; \mathbf{v}^*)$ como mostramos seguir:

$$\begin{aligned}
h_{(1)}^*(t; \mathbf{v}^*) &= \frac{\frac{f_0(t; \mathbf{v}) [\theta S_0(t; \mathbf{v}) (\ln(1 - \theta S_0(t; \mathbf{v})) - 1) - \ln(1 - \theta S_0(t; \mathbf{v}))]}{(S_0(t; \mathbf{v}))^2 (\theta + \ln(1 - \theta))^2 (1 - \theta S_0(t; \mathbf{v}))}}{\frac{\theta S_0(t; \mathbf{v}) + \ln(1 - \theta S_0(t; \mathbf{v}))}{S_0(t; \mathbf{v})(\theta + \ln(1 - \theta))}} \\
&= \frac{f_0(t; \mathbf{v})}{S_0(t; \mathbf{v})} \left[\frac{\theta S_0(t; \mathbf{v}) [\ln(1 - \theta S_0(t; \mathbf{v})) - 1] - \ln(1 - \theta S_0(t; \mathbf{v}))}{(1 - \theta S_0(t; \mathbf{v})) [\theta S_0(t; \mathbf{v}) + \ln(1 - \theta S_0(t; \mathbf{v}))]} \right],
\end{aligned}$$

em que $f_0(t; \mathbf{v})$ é a função de densidade basal e $S_0(t; \mathbf{v})$ a função de sobrevivência basal.

Proporção de Cura

A função de sobrevivência do modelo é uma função imprópria com a proporção de cura (ou proporção de indivíduos curados) dada pela [Equação 3.12](#) para o modelo LIGcr definida como $p_{0(1)}$ é apresentada a seguir:

$$p_{0(1)} = -\frac{\theta}{\ln(1 - \theta)}. \quad (3.12)$$

Note que a proporção de indivíduos curados ([Equação 3.12](#)) está no intervalo $[0, 1]$ como a [Equação 3.13](#) e [Equação 3.14](#) apresentam a seguir:

Se $\theta \rightarrow 0$

$$\lim_{\theta \rightarrow 0} \frac{-\theta}{\ln(1 - \theta)} = \frac{0}{0} (\text{indeterminação}).$$

Aplicando a técnica de L'Hopital, temos que:

$$\lim_{\theta \rightarrow 0} \frac{\frac{d(-\theta)}{d\theta}}{\frac{d(\ln(1 - \theta))}{d\theta}} = \lim_{\theta \rightarrow 0} \frac{-1}{-\frac{1}{\ln(1 - \theta)}} = \lim_{\theta \rightarrow 0} (1 - \theta) = 1. \quad (3.13)$$

Se $\theta \rightarrow 1$

$$\lim_{\theta \rightarrow 1} \frac{-\theta}{\ln(1 - \theta)} = \frac{-1}{-\infty} = 0. \quad (3.14)$$

A proporção de cura está escrita em função de θ . Para ficar mais fácil a interpretação desejamos isolar o parâmetro θ em função de p_0 , o que para o modelo LIGcr não é possível ser feito de forma explícita. No entanto, é possível obter por meio da função Lambert W proposta por ([CORLESS et al., 1996](#)). Então, utilizando a função Lambert W e o auxílio do *software* WolframAlpha ([Wolfram Alpha: Computational Intelligence, 2020](#)), a proporção de cura pode ser reescrita, esta é apresentada na [Equação 3.15](#):

$$p_{0(1)} = 1 + \frac{W(-\theta e^{-\theta})}{\theta}, \quad (3.15)$$

em que $W(\cdot)$ é a função W de Lambert, cuja a sua definição está apresentada no [Apêndice A](#).

Além disso, como em [Suzuki et al. \(2017\)](#), reescrevemos a função de sobrevivência e a função de densidade de probabilidade utilizando $p_{0(1)}$ dando maior interpretação para o modelo. A função de sobrevivência de longa duração e a função de densidade de probabilidade podem ser vistas na [Equação 3.16](#) e [Equação 3.17](#) a seguir:

$$S_{\text{pop}(1)}(t; \mathbf{y}_{(1)}) = -\frac{\ln(1 - W_0 S_0(t; \mathbf{v}))}{W_0 S_0(t; \mathbf{v})} p_{0(1)} \quad (3.16)$$

e,

$$f_{\text{pop}(1)}(t; \boldsymbol{\gamma}_{(1)}) = \frac{W_0 S_0(t; \boldsymbol{v}) + [1 - W_0 S_0(t; \boldsymbol{v})] \ln(1 - W_0 S_0(t; \boldsymbol{v}))}{(1 - W_0 S_0(t; \boldsymbol{v})) [W_0 S_0(t; \boldsymbol{v})]^2} p_{0(1)} f_0(t; \boldsymbol{v}), \quad (3.17)$$

em que $W_0 = 1 + p_{0(1)} W(-e^{-1/p_{0(1)}}/p_{0(1)})$, sendo $W(\cdot)$ a função Lambert W e $S_0(t; \boldsymbol{v})$ e $f_0(t; \boldsymbol{v})$ são as funções de sobrevivência e densidade basal, respectivamente.

Distribuição IG

A distribuição IG foi proposta em 1915 por Schrödinger em [Schrödinger \(1915\)](#), frequentemente chamada de distribuição Wald ([SUZUKI; CANCHO; LOUZADA, 2016](#)). Ela veio da motivação como distribuição do tempo de primeira passagem do movimento Browniano com desvio positivo ou processo Weiner. O nome de Inversa Gaussiana como conhecemos atualmente foi proposto por [Tweedie \(1945\)](#), como afirma [Tweedie et al. \(1957\)](#). [Suzuki, Cancho e Louzada \(2016\)](#) comentam que a distribuição IG tem aplicações úteis para estudos com tempo de vida e em vários outros campos devido ser distribuição de uma variável aleatória referente ao tempo de primeira passagem. A distribuição IG também tem algumas propriedades interessantes como comentam [Suzuki, Cancho e Louzada \(2016\)](#) e [Hanagal e Dabade \(2013\)](#). [Hanagal e Dabade \(2013\)](#) também comentam que esta distribuição é: unimodal; pertence a família exponencial; se assemelha a funções de densidade simétrica e possui muitas semelhanças com a distribuição Gaussiana padrão (Normal (0, 1)) sendo que a distribuição da IG tem uma maior flexibilidade na modelagem de falhas precoces predominantes na distribuição do tempo de vida, além de sua taxa de falha não ser monótona ([SUZUKI; CANCHO; LOUZADA, 2016](#)).

Seja X uma variável aleatória contínua positiva. Temos que X segue a distribuição IG com os parâmetros $\mu \geq 0$ e $\lambda \geq 0$ cuja função densidade de probabilidade é apresentada na [Equação 3.18](#):

$$f_{IG}(x; \lambda, \mu) = \left(\frac{\lambda}{2\pi x^3} \right)^{(1/2)} \exp \left\{ -\frac{\lambda(x - \mu)^2}{2\mu^2 x} \right\}, \quad (3.18)$$

em que $x > 0$.

Além disso, a esperança da função de distribuição IG é da forma $E[X] = \mu$ e a variância é da forma $Var[X] = \mu^3/\lambda$.

Seguindo a mesma notação de [Suzuki et al. \(2017\)](#), a função de sobrevivência é dada da forma:

$$S_{IG}(x; \lambda, \mu) = \Phi \left(\sqrt{\frac{\lambda}{x}} \left(1 - \frac{\lambda}{x} \right) \right) - e^{2\lambda/\mu} \Phi \left(-\sqrt{\frac{\lambda}{x}} \left(1 + \frac{\lambda}{x} \right) \right), \quad (3.19)$$

em que $x > 0$.

Com a [Equação 3.18](#) e a [Equação 3.19](#) é possível encontrar a função de risco com a relação $h(x; \lambda, \pi) = f(x; \lambda, \pi)/S(x; \lambda, \pi)$. Dessa forma, temos a função de risco escrita abaixo:

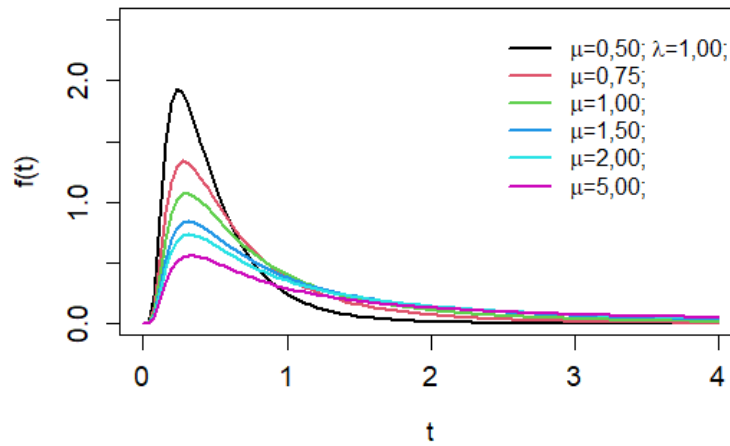
$$h_{IG}(x; \lambda, \mu) = \frac{(\lambda/(2\pi x^3))^{(1/2)} \exp\{(\lambda(x-\mu)^2)/(2\mu^2 x)\}}{\Phi\left(\sqrt{\lambda/x}(1 - (\lambda/x))\right) - e^{2\lambda/\mu} \Phi\left(-\sqrt{(\lambda/x)}(1 + (\lambda/x))\right)}.$$

Comportamento da distribuição IG

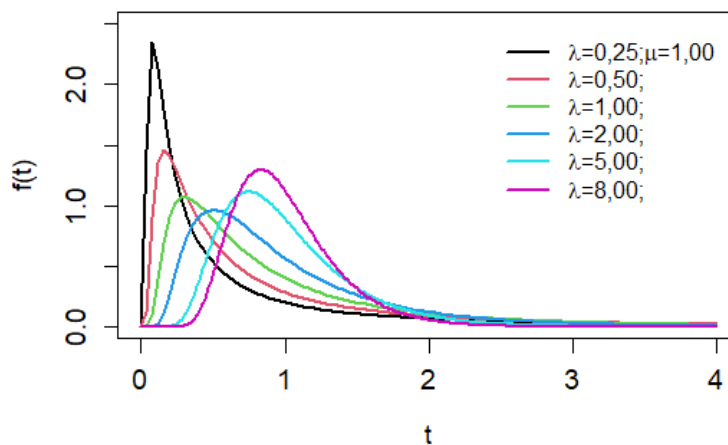
Para entendermos melhor a distribuição IG bem como suas funções de densidade, de sobrevivência e de risco, fizemos um estudo do comportamento dessas funções.

A [Figura 1](#) apresenta a função densidade de probabilidade com diferentes valores dos parâmetros μ e λ . A [Figura 1a](#) apresenta o comportamento da função de densidade de probabilidade para a distribuição IG para diferentes valores de μ e $\lambda = 1,00$ fixado. Já na [Figura 1b](#) temos o comportamento da função de densidade de probabilidade para a distribuição IG fixando μ em 2 e variando os valores de λ . Podemos observar que conforme o valor de λ aumenta a função de densidade fica mais simétrica e mais centrada na média, μ .

Figura 1 – Gráfico com as funções de densidade de probabilidade. Primeiro, no gráfico (a), fixando o parâmetro λ em 1,00 e variando os valores do parâmetro μ . Em seguida, no gráfico (b), variando o parâmetro λ e fixando o parâmetro μ em 1,00.



(a)



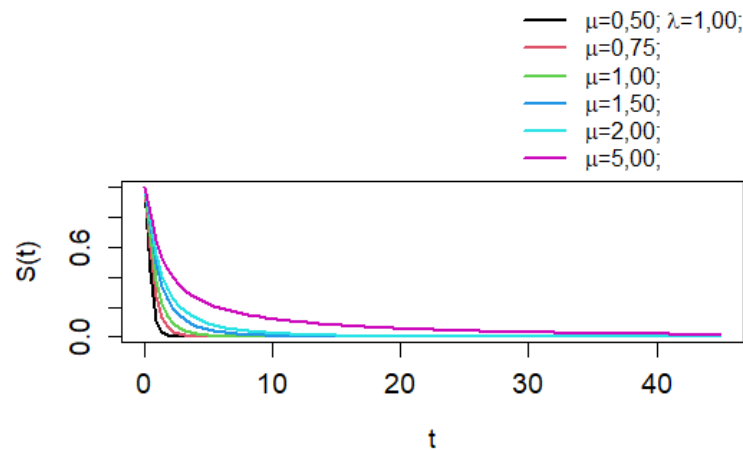
(b)

Fonte: Elaborada pela autora.

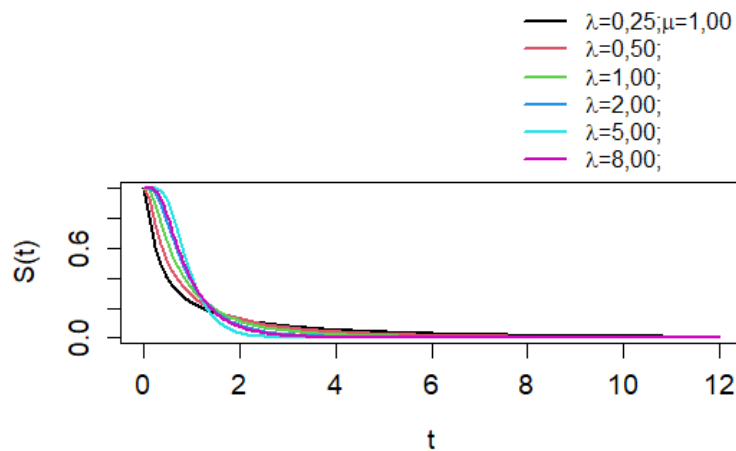
Na [Figura 2](#) temos os comportamentos da função de sobrevivência para diferentes valores dos parâmetros da distribuição IG. O gráfico da [Figura 2a](#) apresenta o comportamento da função de sobrevivência para diferentes valores de μ e $\lambda = 1,00$ fixado, observe que apesar de demorar para a função de sobrevivência atingir o zero, depois de $t = 40$ ela acaba atingindo o zero já que é uma função de sobrevivência própria. O gráfico apresentado na [Figura 2b](#) apresenta o comportamento da função de sobrevivência para diferentes valores de λ e fixando $\mu = 1,00$. Notamos que para valores de μ maiores a função de sobrevivência tende a demorar mais para atingir o zero, embora ela atinja em todos os casos. Podemos observar também que quando $\lambda = 2,00$ e $\mu = 1,00$ a função de sobrevivência é muito parecida com a função de sobrevivência

quando $\lambda = 8$ e $\mu = 1,00$.

Figura 2 – Gráfico com as funções de sobrevivência. Primeiro, no gráfico (a), fixando o parâmetro λ em 1,00 e variando os valores do parâmetro μ . Em seguida, no gráfico (b), variando o parâmetro λ e fixando o parâmetro μ em 1,00.



(a)

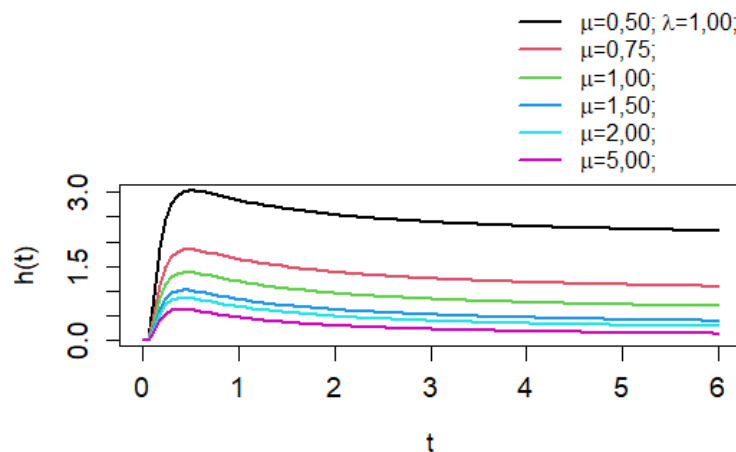


(b)

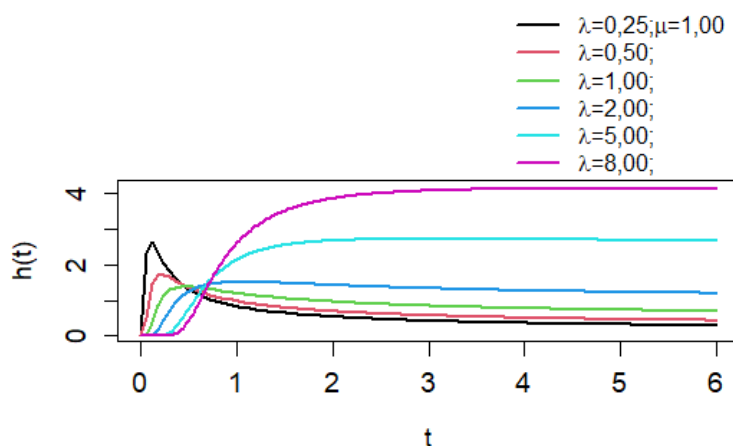
Fonte: Elaborada pela autora.

Por fim, a [Figura 3](#) apresenta o comportamento da função de risco para diferentes valores dos parâmetros da IG. Na [Figura 3a](#) temos o comportamento da função de risco considerando diferentes valores de μ e apenas um valor para λ , fixado em 1,00. O gráfico apresentado pela [Figura 3b](#) nos apresenta o comportamento da função de risco considerando diferentes valores de λ e $\mu = 1,00$. É possível percebermos que para maiores valores de λ a função de risco depois de um determinado tempo acaba por aumentar.

Figura 3 – Gráfico com as funções risco. Primeiro, no gráfico (a), fixando o parâmetro λ em 1,00 e variando os valores do parâmetro μ . Em seguida, no gráfico (b), variando o parâmetro λ e fixando o parâmetro μ em 1,00.



(a)



(b)

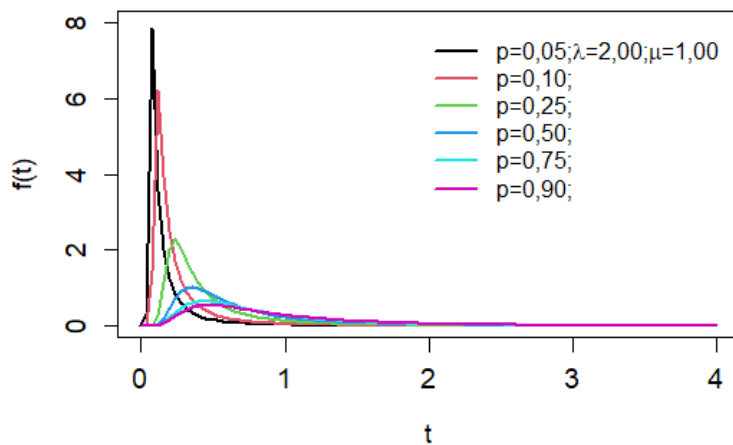
Fonte: Elaborada pela autora.

Comportamento do Modelo *LIGcr*

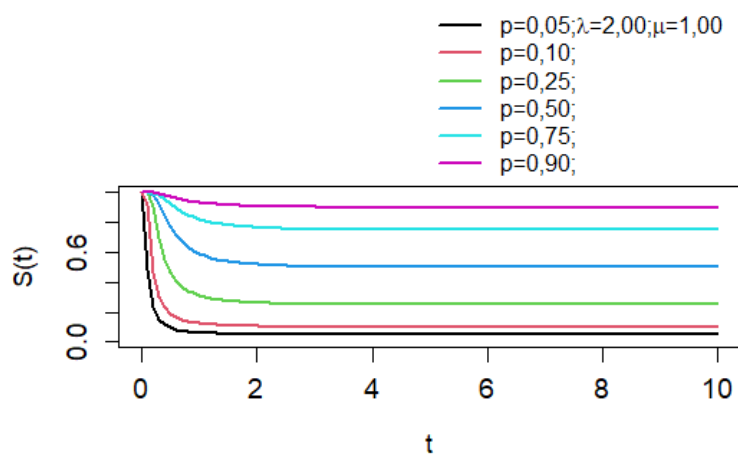
Na Figura 4 temos o comportamento das funções de sobrevivência, densidade de probabilidade e risco do modelo *LIGcr*, em que variamos apenas os valores de $p_{0(1)}$ nas três funções. O $p_{0(1)}$ é denotado por p na legenda do gráfico, os valores dos parâmetros foram $\mu = 1,00$ e $\lambda = 2,00$. O gráfico apresentado pela Figura 4a nos apresenta o comportamento da função de densidade do modelo *LIGcr* dado pela Equação 3.7, já no gráfico da Figura 4c temos sua função de risco. A Figura 4b nos apresenta o gráfico do comportamento da função de sobrevivência com o modelo *LIGcr*. Comparando a função de sobrevivência da distribuição IG apresentada na Figura 2, observamos que a distribuição IG não tem uma proporção de cura, enquanto que para o

modelo *LIGcr*, a função de sobrevivência tem proporção de cura, ou seja, espera-se que a função de sobrevivência nunca atinja o zero mas que ela se estabilize, como é possível observarmos na [Figura 4b](#).

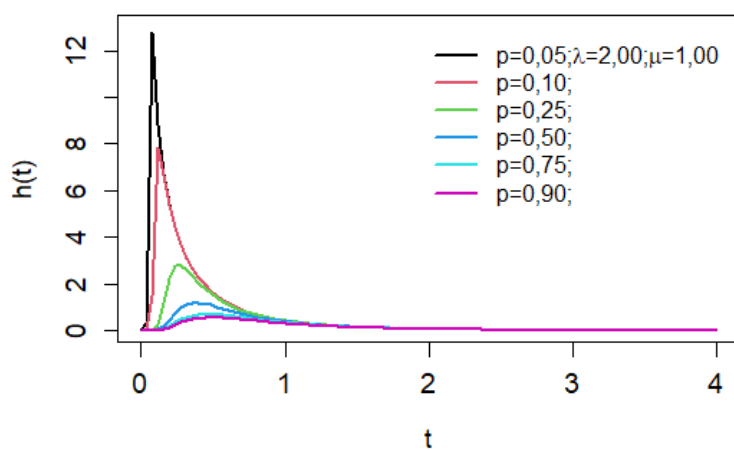
Figura 4 – Gráfico com a função de densidade de probabilidade, função de sobrevivência e função de risco para diferentes valores do parâmetro $p_{0(1)}$ denotado por p na legenda. Utilizamos diferentes valores para $p_{0(1)}$ e fixamos $\mu = 1,00$ e $\lambda = 2,00$.



(a)



(b)



(c)

Fonte: Elaborada pela autora.

3.1.1 Inferência Clássica

Considerando que o tempo de censura do i -ésimo indivíduo é denotado por C_i e que o tempo de falha do i -ésimo indivíduo é denotado por Y_i , temos que o tempo observado do i -ésimo indivíduo é dado pela variável $T_i = \min\{Y_i, C_i\}$. Sendo δ_i o indicador de censura do i -ésimo indivíduo, como definido na [Seção 2.1](#).

Se o indivíduo é censurado, o tempo observado é C_i e, se o indivíduo atinge a falha, o tempo observado é Y_i . Além disso, seja $\mathbf{t} = (t_1, t_2, \dots, t_n)^T$ o vetor de tempos observados, $\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_n)^T$ o vetor de censura dos indivíduos e $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$ o vetor de covariáveis do i -ésimo indivíduo e $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^T$ o vetor de covariáveis dos indivíduos. Temos que $\mathbf{D} = (\mathbf{t}, \boldsymbol{\delta}, \mathbf{x})$ é o vetor de dados completos e $\boldsymbol{\vartheta}_{(1)} = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}_{(1)}^T)$ o vetor de parâmetros com o vetor dos parâmetros das covariáveis, composto por $\boldsymbol{\beta}^T$ (o vetor de parâmetros das covariáveis) e $\boldsymbol{\gamma}_{(1)}^T = (\lambda, \mu, p_{0(1)})$ é o vetor dos parâmetros do modelo LIGcr e $p_{0(1)}$ a proporção de cura do mesmo.

Considerando que o vetor de parâmetros da distribuição basal do modelo LIGcr é definido por $\mathbf{v}_{(1)}$ temos que $\mathbf{v}_{(1)} = (\lambda, \mu)$. Dessa forma, a função de verossimilhança é dada por:

$$\begin{aligned}
 L(\boldsymbol{\vartheta}_{(1)}; \mathbf{D}) &= \prod_{i=1}^n (f_{\text{pop}(1)}(t_i; \boldsymbol{\vartheta}_{(1)}))^{\delta_i} (S_{\text{pop}(1)}(t_i; \boldsymbol{\vartheta}_{(1)}))^{1-\delta_i} \\
 &= \prod_{i=1}^n \left(\frac{W_0 S_0(t; \mathbf{v}_{(1)}) + [1 - W_0 S_0(t; \mathbf{v}_{(1)})] \ln(1 - W_0 S_0(t; \mathbf{v}_{(1)}))}{(1 - W_0 S_0(t; \mathbf{v}_{(1)})) [W_0 S_0(t; \mathbf{v}_{(1)})]^2} p_{0(1)} f_0(t; \mathbf{v}_{(1)}) \right)^{\delta_i} \\
 &\quad \times \left(-\frac{\ln(1 - W_0 S_0(t; \mathbf{v}_{(1)}))}{W_0 S_0(t; \mathbf{v}_{(1)})} p_{0(1)} \right)^{1-\delta_i} \\
 &= \prod_{i=1}^n \frac{p_{0(1)}}{w_0} \left(\frac{W_0 S_0(t_i; \mathbf{v}_{(1)}) + [1 - W_0 S_0(t_i; \mathbf{v}_{(1)})] \ln(1 - W_0 S_0(t_i; \mathbf{v}_{(1)})) f_0(t_i; \mathbf{v}_{(1)})}{(1 - W_0 S_0(t_i; \mathbf{v}_{(1)}))} \right)^{\delta_i} \\
 &\quad \times \left(-\frac{((\ln(1 - W_0 S_0(t_i; \mathbf{v}_{(1)})))^{1-\delta_i})}{(S_0(t_i; \mathbf{v}_{(1)}))^{1+\delta_i}} \right),
 \end{aligned}$$

em que $W_0 = 1 + p_{0(1)} W(-e^{-1/p_{0(1)}}/p_{0(1)})$ sendo $W(\cdot)$ a função W Lambert e $S_0(t; \mathbf{v}_{(1)})$ e $f_0(t; \mathbf{v}_{(1)})$ são as funções de sobrevivência e densidade basal, respectivamente. A proporção de cura $p_{0(1)}$ do modelo LIGcr, $\boldsymbol{\vartheta}_{(1)}$ o vetor de parâmetros do modelo LIGcr e $\mathbf{v}_{(1)}$ o vetor de parâmetros da distribuição IG.

Assim, aplicando o logaritmo, obtemos o logaritmo da função de verossimilhança, que é dado por:

$$\begin{aligned}
l(\boldsymbol{\vartheta}_{(1)}; \mathbf{D}) &= \sum_{i=1}^n \ln \left(\frac{p_{0(1)}}{W_0} \right) \\
&+ \sum_{i=1}^n \delta_i \ln \left(\frac{W_0 S_0(t_i; \mathbf{v}_{(1)}) + [1 - W_0 S_0(t_i; \mathbf{v}_{(1)})] \ln(1 - W_0 S_0(t_i; \mathbf{v}_{(1)})) f_0(t_i; \mathbf{v}_{(1)})}{(1 - W_0 S_0(t_i; \mathbf{v}_{(1)}))} \right) \\
&+ \sum_{i=1}^n (1 - \delta_i) \ln(-(\ln(1 - W_0 S_0(t_i; \mathbf{v}_{(1)}))) - \sum_{i=1}^n (1 + \delta_i) \ln(S_0(t_i; \mathbf{v}_{(1)})).
\end{aligned}$$

Substituindo $f_0(t_i; \mathbf{v}_{(1)})$ e $S_0(t_i; \mathbf{v}_{(1)})$ pelas suas respectivas expressões, obtemos:

$$\begin{aligned}
l(\boldsymbol{\vartheta}_{(1)}; \mathbf{D}) &= \sum_{i=1}^n \ln \left(\frac{p_{0(1)}}{W_0} \right) + \sum_{i=1}^n \delta_i \ln \left\{ W_0 \times \left[\Phi \left(\sqrt{\frac{\lambda}{t_i}} \left(1 - \frac{\lambda}{t_i} \right) \right) - e^{2\lambda/\mu} \Phi \left(-\sqrt{\frac{\lambda}{t_i}} \left(1 + \frac{\lambda}{t_i} \right) \right) \right] \right. \\
&+ \left. \left[1 - W_0 \times \left[\Phi \left(\sqrt{\frac{\lambda}{t_i}} \left(1 - \frac{\lambda}{t_i} \right) \right) - e^{2\lambda/\mu} \Phi \left(-\sqrt{\frac{\lambda}{t_i}} \left(1 + \frac{\lambda}{t_i} \right) \right) \right] \right] \right\} \\
&\times \ln \left(1 - W_0 \times \left[\Phi \left(\sqrt{\frac{\lambda}{t_i}} \left(1 - \frac{\lambda}{t_i} \right) \right) - e^{2\lambda/\mu} \Phi \left(-\sqrt{\frac{\lambda}{t_i}} \left(1 + \frac{\lambda}{t_i} \right) \right) \right] \right) \\
&\times \left(\frac{\lambda}{2\pi t_i^3} \right)^{(1/2)} \exp \left\{ -\frac{\lambda(t_i - \mu)^2}{2\mu^2 t_i} \right\} \\
&\times \left. \frac{1}{\left(1 - W_0 \times \left[\Phi \left(\sqrt{\frac{\lambda}{t_i}} \left(1 - \frac{\lambda}{t_i} \right) \right) - e^{2\lambda/\mu} \Phi \left(-\sqrt{\frac{\lambda}{t_i}} \left(1 + \frac{\lambda}{t_i} \right) \right) \right] \right)} \right\} \\
&+ \sum_{i=1}^n (1 - \delta_i) \ln \left\{ -\ln \left(1 - W_0 \times \left[\Phi \left(\sqrt{\frac{\lambda}{t_i}} \left(1 - \frac{\lambda}{t_i} \right) \right) - e^{2\lambda/\mu} \Phi \left(-\sqrt{\frac{\lambda}{t_i}} \left(1 + \frac{\lambda}{t_i} \right) \right) \right] \right) \right\} \\
&- \sum_{i=1}^n (1 + \delta_i) \ln \left(\Phi \left(\sqrt{\frac{\lambda}{t_i}} \left(1 - \frac{\lambda}{t_i} \right) \right) - e^{2\lambda/\mu} \Phi \left(-\sqrt{\frac{\lambda}{t_i}} \left(1 + \frac{\lambda}{t_i} \right) \right) \right),
\end{aligned}$$

em que $W_0 = 1 + p_{0(1)} W(-e^{-1/p_{0(1)}}/p_{0(1)})$ sendo $W(\cdot)$ a função W Lambert e $p_{0(1)}$ a proporção de cura do modelo LIGcr.

A partir da equação acima obtemos as derivadas de $\log(L(\boldsymbol{\vartheta}_{(1)}; \mathbf{D}))$ em relação a cada um dos parâmetros e igualamos a zero para encontrar os valores que maximizam o logaritmo da função de verossimilhança. Temos que a obtenção dos estimadores dos parâmetros: λ , μ e $p_{0(1)}$ pelas soluções das seguintes equações:

$$\left. \frac{\partial l(\boldsymbol{\vartheta}_{(1)}; \mathbf{D})}{\partial \lambda} \right|_{\lambda=\hat{\lambda}} = 0,$$

$$\left. \frac{\partial l(\boldsymbol{\vartheta}_{(1)}; \mathbf{D})}{\partial \mu} \right|_{\mu=\hat{\mu}} = 0$$

e,

$$\left. \frac{\partial l(\boldsymbol{\vartheta}_{(1)}; \mathbf{D})}{\partial p_{0(1)}} \right|_{p_{0(1)}=p_{0(1)}^{\hat{}}} = 0.$$

A função escore do modelo é dada da forma $U(\boldsymbol{\vartheta}_{(1)}) = \partial \log L(\boldsymbol{\vartheta}_{(1)}; \mathbf{D}) / (\partial \boldsymbol{\vartheta}_{(1)})$ e a matriz observada é dada por $\mathbf{I} = -\partial U^\top / (\partial \boldsymbol{\vartheta}_{(1)}) = -\partial^2 \log L(\boldsymbol{\vartheta}_{(1)}; \mathbf{D}) / (\partial \boldsymbol{\vartheta}_{(1)} \partial \boldsymbol{\vartheta}_{(1)}^\top)$. Note que as equações acima não possuem forma fechada. Assim, as estimativas dos parâmetros foi obtida numericamente com o auxílio da função *optim()* pelo *software R Development Core Team (2017a)* utilizando o método *L-BFGS-B*.

Usualmente, em modelos com proporção de indivíduos curados, as covariáveis são relacionadas no parâmetro da proporção de cura. Dessa forma, cada indivíduo (baseado nas covariáveis) tem uma proporção de cura diferente dos outros, ou um grupo de indivíduos tem proporção de cura diferente do(s) outro(s) grupo(s).

Para a inclusão das covariáveis no parâmetro da proporção de cura do modelo LIGcr, $p_{0(1)}$, utilizamos a função de ligação logística. Assim, considerando o vetor de parâmetros das covariáveis sendo dado por $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ e a proporção de cura do *i-ésimo* indivíduo é $p_{0(1)i} = \frac{-\theta_i}{\log(1-\theta_i)}$, temos que:

$$\log \left(\frac{p_{0(1)i}}{1 - p_{0(1)i}} \right) = \mathbf{x}_i^T \boldsymbol{\beta} \Rightarrow e^{\mathbf{x}_i^T \boldsymbol{\beta}} = \frac{p_{0(1)i}}{1 - p_{0(1)i}} \Rightarrow p_{0(1)i} = \frac{e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}}.$$

Também foi feita a inclusão de covariável no parâmetro μ utilizando a função de ligação exponencial. Considerando que o vetor de parâmetros das covariáveis é dado por $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$, então:

$$\log(\mu) = \mathbf{x}_i^T \boldsymbol{\beta} \Rightarrow e^{\mathbf{x}_i^T \boldsymbol{\beta}} = \mu.$$

3.2 Estudo de Simulação e Aplicação a dados reais

Realizamos um estudo de simulação com o objetivo de avaliar as estimativas dos parâmetros do modelo LIGcr. Consideramos dados artificiais sem a inclusão de covariáveis e com a inclusão de covariáveis.

Em cada simulação foi calculada a média, o viés, o Erro Quadrático Médio (EQM) e a Probabilidade de Cobertura (PC), que foi obtida por meio do intervalo assintótico de 95%.

Considerando essas métricas, para o vício e o EQM esperamos que quanto mais próximo de zero melhor o resultado. A PC utilizamos para avaliarmos a qualidade dos estimadores intervalares. No cálculo do Intervalo de Confiança utilizamos as variâncias assintóticas a partir da matriz de Informação Observada de Fisher.

Sendo N o total de réplicas, o EQM e o vício dos parâmetros podem ser obtidos, respectivamente, por $EQM = \frac{1}{N} \sum_{i=1}^N (\hat{\phi}_i - \phi_0)^2$, e o vício $= \frac{1}{N} \sum_{i=1}^N \hat{\phi}_i - \phi_0$, em que ϕ_0 é o vetor dos verdadeiros valores dos parâmetros e $\hat{\phi}$ é o vetor dos parâmetros estimados.

Nesta Seção também fizemos aplicação a um conjunto de dados reais. Além disso, calculamos os valores do *Akaike Information Criterion* (AIC) e *Bayesian Information Criterion* (BIC). Para o cálculo dos critérios utilizados temos que: n_p é o número de parâmetros estimados¹; n é o total de observações e $l(\hat{\boldsymbol{\theta}}, \mathbf{D})$ é o logaritmo da função de verossimilhança do modelo, assim temos que: $AIC = -2l(\hat{\boldsymbol{\theta}}, \mathbf{D}) + 2n_p$ e $BIC = -2l(\hat{\boldsymbol{\theta}}, \mathbf{D}) + \ln(n)n_p$ (SEBER; LEE, 2012). A análise dos critérios AIC e BIC é feita observando o valor do critério para cada modelo, o melhor modelo é aquele com o menor valor do critério.

3.2.1 Dados Artificiais

Para o cálculo da probabilidade de cobertura consideramos a normalidade assintótica (distribuição Normal(0, 1)) com 95% de confiança. Além disso, apresentamos a média da proporção do total de observações censuradas em cada cenário nas tabelas que apresentam os resultados das simulações. Assim, como Souza *et al.* (2017), os tempos de censura, estão associados a τ sendo este parâmetro controle da proporção de censura da população que não é curada. Nos cenários de simulação que tiveram a inclusão de covariáveis, simulamos a covariável com uma distribuição Binomial (1, 0, 5) sendo n o tamanho das amostras para cada cenário. Além disso, utilizamos a função de ligação logística no parâmetro da proporção de cura, $\log(p_{0i}) = \exp(\beta_0 + \beta_1 x_i)$ e a função de ligação exponencial no parâmetro μ . Em todo o estudo, foi utilizado o *software R Development Core Team* (2017a).

Algoritmo de Simulação com Covariáveis

O algoritmo abaixo considera a inclusão de covariáveis utilizando a função de ligação logística para a proporção de cura e a ligação exponencial para o parâmetro μ do modelo LIG e para o parâmetro ψ do modelo GEPGW.

Passo 1: Inicialize $j = 1$. Defina n (tamanho amostral) e r (número de réplicas desejadas);

Passo 2: Gere a covariável segundo a distribuição Binomial(1;0,5) para a simulação de cada indivíduo da amostra;

¹ Apesar da quantidade de parâmetros ser fixa no modelo apresentamos a forma genérica para a obtenção do estimador já que o modelo LIGcr quando ajustado utilizando covariáveis em $p_{0(1)}$ aumenta a quantidade de parâmetros, da mesma forma quando acrescentamos covariáveis em μ

Passo 3: Gerar n tempos de censura t_{aux} , em que cada tempo foi simulado de uma distribuição Uniforme(0, τ), sendo τ o parâmetro fixo que controla o percentual de observações censuradas;

Passo 4: Simule n valores de Z que segue uma distribuição Logarítmica (θ) começando em zero;

Passo 5: Para $i = 1, \dots, n$, se o tempo $z_i = 0$ então o tempo t_i é infinito. Caso contrário, gere t_{k_i} tempos em que $k_i = 1, 2, \dots, z_i$ e $t_{k_i} \sim IG(\lambda, \mu)$ e faça $t_i = \min(t_{k_i})$. Se $t_{aux_i} > t_i$ então considere o indivíduo como não censurado, caso contrário considere o indivíduo como censurado. Defina o tempo final do i -ésimo indivíduo como sendo $t_{i\text{final}} = \min(t_i, t_{aux_i})$ e faça $j = j + 1$;

Passo 6: Repita os passos de 2 a 5 até atingir o número de réplicas desejadas.

Para os resultados obtidos utilizamos o *software* [R Development Core Team \(2017b\)](#).

3.2.2 Simulação

Na simulação consideramos o modelo LIGcr. Os valores dos parâmetros μ e λ foram trocados junto com os valores de $p_{0(1)}$ mostrando diferentes cenários nas tabelas.

A [Tabela 1](#), [Tabela 2](#) e a [Tabela 3](#) apresentam os resultados das simulações sem a inclusão de covariáveis e considerando a proporção de indivíduos curados: 60%, 45% e 30%, respectivamente. Para cada cenário de simulação, consideramos diferentes valores para o parâmetro relacionado a proporção de cura como também para μ e λ .

Simulações sem Covariáveis

Neste trabalho realizamos um estudo de simulação com 1.000 réplicas considerando quatro tamanhos amostrais $n = 50, 150, 300$ e 500 .

A [Tabela 1](#) apresenta a média das estimativas dos parâmetros. Podemos observar que conforme o tamanho da amostra aumenta, menor é o vício e o EQM. Além disso, a PC que fica mais próxima do valor nominal de 95%. Também, podemos observar que a proporção de censura fica próxima de 63% sendo a proporção de cura 60%.

Tabela 1 – Simulação do modelo LIGcr considerando $p_{0(1)} = 0,60$, $\mu = 1$ e $\lambda = 1$ para diferentes valores amostrais.

Tamanho amostral (n)	Parâmetros	Sumário dos parâmetros				Censura
		Média	Vício	EQM	P. C.	
50	$p_{0(1)}$	0,5930	-0,0070	0,0047	0,9650	62,94%
	μ	1,2373	0,2373	0,4063	0,9730	
	λ	1,0885	0,0885	0,1148	0,9720	
150	$p_{0(1)}$	0,5971	-0,0029	0,0017	0,9580	63,15%
	μ	1,1282	0,1282	0,1290	0,9600	
	λ	1,0363	0,0363	0,0356	0,9490	
300	$p_{0(1)}$	0,5970	-0,0030	0,0008	0,9630	63,04%
	μ	1,0610	0,0610	0,0465	0,9580	
	λ	1,0224	0,0224	0,0157	0,9600	
500	$p_{0(1)}$	0,5992	-0,0008	0,0005	0,9650	63,22%
	μ	1,0292	0,0292	0,0187	0,9510	
	λ	1,0142	0,0142	0,0100	0,9440	

A [Tabela 2](#) apresenta os resultados da simulação considerando a proporção de cura em 45% com $\mu = 3$ e $\lambda = 4$. Observamos que apesar da proporção de censura ser mais baixa que a proporção de censura da [Tabela 1](#) a diferença entre a quantidade de dados censurados e a proporção de cura é menor na [Tabela 1](#) (aproximadamente 3%) do que na [Tabela 2](#) (aproximadamente 7%), o que possivelmente pode influenciar nos valores obtidos das estimativas dos parâmetros e, consequentemente influenciando o vício, o EQM e a PC. Assim, como na [Tabela 1](#), observamos

que conforme aumenta o tamanho amostral n menor o vício e o EQM. A PC também fica mais próxima do valor nominal de 95%.

Tabela 2 – Simulação do modelo LIGcr considerando $p_{0(1)} = 0,45$, $\mu = 3$ e $\lambda = 4$ para diferentes valores amostrais.

Tamanho amostral (n)	Parâmetros	Sumário dos parâmetros				Censura
		Média	Vício	EQM	P. C.	
50	$p_{0(1)}$	0,4503	0,0003	0,0052	0,9720	52,16%
	μ	3,1361	0,1361	1,6473	0,8720	
	λ	4,4835	0,4835	1,4984	0,9660	
150	$p_{0(1)}$	0,4501	0,0001	0,0021	0,9620	52,43%
	μ	3,0962	0,0962	0,6299	0,8990	
	λ	4,1742	0,1742	0,3991	0,9690	
300	$p_{0(1)}$	0,4487	-0,0013	0,0012	0,9530	52,30%
	μ	3,1195	0,1195	0,4570	0,9130	
	λ	4,0893	0,0893	0,1890	0,9590	
500	$p_{0(1)}$	0,4509	0,0009	0,0007	0,9500	52,43%
	μ	3,0463	0,0463	0,2681	0,9020	
	λ	4,0411	0,0411	0,1137	0,9410	

Já na [Tabela 3](#) observamos que a proporção de dados censurados fica em torno de 34% sendo a proporção de cura igual a 30%. A diferença entre a quantidade de censura e a proporção de cura é de aproximadamente 4%, que é maior que a diferença da [Tabela 1](#) e menor que a diferença da [Tabela 2](#). Podemos observar também que a PC apresenta valores próximos ao valor nominal de 95%.

Tabela 3 – Simulação do modelo LIGcr considerando $p_{0(1)} = 0,30$, $\mu = 2$ e $\lambda = 1$ para diferentes valores amostrais.

Tamanho amostral (n)	Parâmetros	Sumário dos parâmetros				Censura
		Média	Vício	EQM	P. C.	
50	$p_{0(1)}$	0,3043	0,0043	0,0035	0,9730	33,21%
	μ	2,2937	0,2937	2,1311	0,9420	
	λ	1,0769	0,0769	0,0543	0,9740	
150	$p_{0(1)}$	0,3040	0,0040	0,0013	0,9730	33,75%
	μ	2,1360	0,1360	0,9072	0,9270	
	λ	1,0256	0,0256	0,0164	0,9500	
300	$p_{0(1)}$	0,3006	0,0006	0,0008	0,9590	33,62%
	μ	2,1923	0,1923	0,8107	0,9140	
	λ	1,0117	0,0117	0,0074	0,9540	
500	$p_{0(1)}$	0,3018	0,0018	0,0004	0,9720	33,79%
	μ	2,1052	0,1052	0,4608	0,9170	
	λ	1,0052	0,0052	0,0041	0,9530	

Simulações com Covariáveis

Já a [Tabela 4](#), [Tabela 5](#) e a [Tabela 6](#) apresentam os resultados das simulações com a inclusão de uma covariável binária no parâmetro da proporção de cura, utilizando a função de ligação logística e considerando os valores de β_0 e β_1 como sendo: $\beta_0 = -1$ e $\beta_1 = 4$; $\beta_0 = -1,5$ e $\beta_1 = 0,75$; $\beta_0 = 0,5$ e $\beta_1 = 0,7$, respectivamente. Para cada grupo de indivíduos temos uma determinada proporção de cura que depende do valor dos parâmetros β_0 e β_1 .

A [Tabela 4](#) apresenta os resultados em que $\beta_0 = -1$, $\beta_1 = 4$, $\mu = 1$, $\lambda = 1$ com a proporção de censura em torno de 63%. Para os indivíduos pertencentes ao grupo $X = 1$ a proporção de cura é próxima de 95%, já para os indivíduos pertencentes ao grupo $X = 0$ a proporção de cura é próxima de 27%. Além disso, podemos perceber que as PC's se aproximam do valor nominal de 95% e o EQM dos parâmetros reduz conforme aumenta o tamanho amostral n .

Tabela 4 – Simulação do modelo LIGcr considerando $\beta_0 = -1$, $\beta_1 = 4$, $\mu = 1$ e $\lambda = 1$ para diferentes valores amostrais.

Tamanho amostral (n)	Parâmetros	Sumário dos parâmetros				Censura
		Média	Vício	EQM	P. C.	
50	β_0	-1,1063	-0,1063	0,1861	0,9680	61,59%
	β_1	3,7924	-0,2076	1,5055	0,9510	
	μ	1,2622	0,2622	0,6006	0,9930	
	λ	1,1637	0,1637	0,1546	0,9820	
150	β_0	-1,0178	-0,0178	0,0578	0,9660	63,01%
	β_1	4,0673	0,0673	0,3608	0,9670	
	μ	1,1349	0,1349	0,2553	0,9710	
	λ	1,0372	0,0372	0,0307	0,9680	
300	β_0	-1,0114	-0,0114	0,0306	0,9530	63,19%
	β_1	4,0769	0,0769	0,2540	0,9580	
	μ	1,0983	0,0983	0,1559	0,9550	
	λ	1,0215	0,0215	0,0150	0,9500	
500	β_0	-1,0093	-0,0093	0,0176	0,9610	63,23%
	β_1	4,0270	0,0270	0,1315	0,9510	
	μ	1,0516	0,0516	0,0763	0,9490	
	λ	1,0165	0,0165	0,0084	0,9630	

Podemos observar na [Tabela 5](#) que a proporção média de censura é em torno de 33%. Para os indivíduos pertencentes ao grupo de $X = 1$ a proporção de cura é aproximadamente 32%, já para os indivíduos pertencentes ao grupo de $X = 0$ a proporção de cura é aproximadamente 18%. Para este cenário de simulação utilizamos $\mu = 2$ e $\lambda = 3$. Como resultado, notamos uma redução do vício e do EQM, além das PC's ficarem mais próximas do valor nominal de 95% conforme aumenta o tamanho amostral n .

Tabela 5 – Simulação do modelo LIGcr considerando $\beta_0 = -1,5$, $\beta_1 = 0,75$, $\mu = 2$ e $\lambda = 3$ para diferentes valores amostrais.

Tamanho amostral (n)	Parâmetros	Sumário dos parâmetros				Censura
		Média	Vício	EQM	P. C.	
50	β_0	-1,5056	-0,0056	0,1343	0,9510	33,44%
	β_1	0,7904	0,0404	0,1345	0,9530	
	μ	1,9874	-0,0126	0,5739	0,8760	
	λ	3,3909	0,3909	0,8918	0,9750	
150	β_0	-1,5063	-0,0063	0,0424	0,9570	33,22%
	β_1	0,7575	0,0075	0,0410	0,9560	
	μ	2,0764	0,0764	0,4421	0,8970	
	λ	3,1219	0,1219	0,1820	0,9550	
300	β_0	-1,5052	-0,0052	0,0214	0,9530	33,29%
	β_1	0,7595	0,0095	0,0231	0,9470	
	μ	2,0722	0,0722	0,2638	0,8990	
	λ	3,0593	0,0593	0,0819	0,9520	
500	β_0	-1,5105	-0,0105	0,0134	0,9480	33,13%
	β_1	0,7514	0,0014	0,0130	0,9410	
	μ	2,0888	0,0888	0,2234	0,9260	
	λ	3,0445	0,0445	0,0470	0,9530	

Vamos considerar agora os valores de $\beta_0 = 0,5$, $\beta_1 = 0,7$, $\mu = 3$ e $\lambda = 4$. Dessa forma, os indivíduos que pertencem ao grupo $X = 0$ tem proporção de cura próxima de 62%. Já para os indivíduos pertencentes ao grupo $X = 1$ sua proporção de cura é de aproximadamente 77% e a proporção de censura é em torno de 78%. A [Tabela 6](#) apresenta os resultados obtidos para este cenário.

Tabela 6 – Simulação do modelo LIGcr considerando $\beta_0 = 0,5$, $\beta_1 = 0,7$, $\mu = 3$ e $\lambda = 4$ para diferentes valores amostrais.

Tamanho amostral (n)	Parâmetros	Sumário dos parâmetros				Censura
		Média	Vício	EQM	P. C.	
50	β_0	0,6088	0,1088	0,2505	0,9850	78,15%
	β_1	0,7654	0,0654	0,9393	0,9750	
	μ	2,5153	-0,4847	0,8925	0,7500	
	λ	5,5124	1,5124	7,6418	0,9950	
150	β_0	0,5617	0,0617	0,0846	0,9740	78,52%
	β_1	0,7206	0,0206	0,1724	0,9490	
	μ	2,7896	-0,2104	0,5695	0,8230	
	λ	4,5056	0,5056	1,4089	0,9720	
300	β_0	0,5274	0,0274	0,0424	0,9670	78,39%
	β_1	0,7042	0,0042	0,0797	0,9510	
	μ	2,9306	-0,0694	0,4727	0,8480	
	λ	4,3021	0,3021	0,6859	0,9690	
500	β_0	0,5139	0,0139	0,0278	0,9650	78,38%
	β_1	0,7036	0,0036	0,0516	0,9450	
	μ	2,9967	-0,0033	0,3647	0,8890	
	λ	4,2107	0,2107	0,3695	0,9630	

Realizamos também um outro estudo onde consideramos a inclusão de uma covariável binária no parâmetro $p_{0(1)}$ e no parâmetro μ conjuntamente. Para isso, em $p_{0(1)}$ foi utilizada a função logística e os parâmetros relacionados à covariável incluída são: β_0 e β_1 . Já para o parâmetro μ utilizamos a função de ligação exponencial e os parâmetros relacionados a covariável incluída são: β_2 e β_3 . Os resultados são apresentados na [Tabela 7](#), [Tabela 8](#) e [Tabela 9](#).

Podemos observar na [Tabela 7](#) que, conforme o tamanho amostral aumenta, menor fica o EQM e mais próximo a PC fica do valor nominal de 0,95. Além disso, apesar de ser o cenário tem a maior proporção de censura, é o que de uma forma geral apresentou os menores EQM's. Para os indivíduos que pertencem ao grupo $X = 0$, a proporção de cura é próxima de 60% enquanto que é próxima de 77% para os indivíduos que pertencem ao grupo $X = 1$.

Tabela 7 – Simulação do modelo LIGcr considerando $\beta_0 = 0,4$, $\beta_1 = 0,8$, $\beta_2 = 0,02$, $\beta_3 = 0,9$ e $\lambda = 4,00$ para diferentes valores amostrais.

Tamanho amostral (n)	Parâmetros	Sumário dos parâmetros				Censura
		Média	Vício	EQM	P. C.	
50	β_0	0,2959	-0,1041	0,1928	0,9670	71,74%
	β_1	0,7443	-0,0557	0,5905	0,9780	
	β_2	0,2874	0,2674	0,3911	0,9930	
	β_3	2,0153	1,1153	9,3591	0,9620	
	λ	4,1114	0,1114	1,3433	0,9690	
150	β_0	0,3365	-0,0635	0,0615	0,9550	72,52%
	β_1	0,7810	-0,0190	0,2192	0,9690	
	β_2	0,1592	0,1392	0,0261	0,9850	
	β_3	1,4696	0,5696	4,4257	0,9370	
	λ	3,9234	-0,0766	0,3880	0,9570	
300	β_0	0,3620	-0,0380	0,0268	0,9630	72,66%
	β_1	0,7827	-0,0173	0,1164	0,9630	
	β_2	0,1170	0,0970	0,0124	0,9740	
	β_3	1,1406	0,2406	1,7084	0,9220	
	λ	3,9029	-0,0971	0,2050	0,9640	
500	β_0	0,3666	-0,0334	0,0186	0,9470	72,77%
	β_1	0,8079	0,0079	0,0649	0,9510	
	β_2	0,0970	0,0770	0,0081	0,9630	
	β_3	0,9531	0,0531	0,4265	0,9370	
	λ	3,9121	-0,0879	0,1315	0,9650	

Na [Tabela 8](#), podemos observar que, de uma forma geral, a proporção de censura é próxima de 53%, a PC apresenta valores bem próximos do valor nominal de 0,95 e, para os indivíduos pertencentes ao grupo com a covariável $X = 1$ tem uma proporção de cura próxima de 38% enquanto que os indivíduos pertencentes ao grupo com a covariável $X = 0$ tem uma proporção de cura próxima de 62%.

Tabela 8 – Simulação do modelo LIGcr considerando $\beta_0 = 0,5$, $\beta_1 = -1$, $\beta_2 = 0,9$, $\beta_3 = 0,5$ e $\lambda = 2$ para diferentes valores amostrais.

Tamanho amostral (n)	Parâmetros	Sumário dos parâmetros				Censura
		Média	Vício	EQM	P. C.	
50	β_0	0,4521	-0,0479	0,1893	0,9610	51,63%
	β_1	-1,0427	-0,0427	0,3177	0,9740	
	β_2	1,4251	0,5251	5,1394	1,0000	
	β_3	2,5505	2,0505	31,3283	1,0000	
	λ	2,2970	0,2970	0,4858	0,9610	
150	β_0	0,4982	-0,0018	0,0619	0,9600	52,82%
	β_1	-1,0528	-0,0528	0,1278	0,9450	
	β_2	0,9986	0,0986	0,5817	0,9910	
	β_3	2,2184	1,7184	21,2083	0,9830	
	λ	2,0894	0,0894	0,0983	0,9570	
300	β_0	0,5052	0,0052	0,0347	0,9410	52,92%
	β_1	-1,0440	-0,0440	0,0734	0,9400	
	β_2	0,9315	0,0315	0,0525	0,9690	
	β_3	1,6916	1,1916	13,3920	0,9700	
	λ	2,0639	0,0639	0,0496	0,9400	
500	β_0	0,5023	0,0023	0,0188	0,9500	52,94%
	β_1	-1,0241	-0,0241	0,0402	0,9530	
	β_2	0,9201	0,0201	0,0301	0,9720	
	β_3	1,2967	0,7967	8,9427	0,9530	
	λ	2,0399	0,0399	0,0262	0,9520	

Já na [Tabela 9](#) podemos observar uma censura de aproximadamente 36%. Além disso, o EQM de β_2 apresentou um valor maior em relação aos resultados obtidos anteriormente com os mesmos tamanhos amostrais. A PC ficou próxima do valor nominal de 95%. Neste cenário para os indivíduos que pertencem ao grupo $X = 0$ temos que a porcentagem de indivíduos curados é aproximadamente 23%. Já para o grupo $X = 1$ a porcentagem de indivíduos curados é aproximadamente 18%.

Tabela 9 – Simulação do modelo LIGcr considerando $\beta_0 = -1,2$, $\beta_1 = -0,3$, $\beta_2 = 0,09$, $\beta_3 = 0,5$ e $\lambda = 3$ para diferentes valores amostrais.

Tamanho amostral (n)	Parâmetros	Sumário dos parâmetros				Censura
		Média	Vício	EQM	P. C.	
50	β_0	-1,2452	-0,0452	0,3165	0,9130	36,11%
	β_1	-0,3000	0,0000	0,4139	0,9340	
	β_2	0,9507	0,8607	13,3244	0,8670	
	β_3	1,4574	0,9574	35,1453	0,9850	
	λ	3,6801	0,6801	1,3513	0,9890	
150	β_0	-1,2735	-0,0735	0,1686	0,8820	36,29%
	β_1	-0,2940	0,0060	0,2413	0,8630	
	β_2	0,8019	0,7119	9,0978	0,8880	
	β_3	1,4909	0,9909	28,4174	0,9750	
	λ	3,3268	0,3268	0,3776	0,9710	
300	β_0	-1,2538	-0,0538	0,0895	0,9050	36,33%
	β_1	-0,2906	0,0094	0,1279	0,8970	
	β_2	0,4309	0,3409	3,7106	0,9040	
	β_3	1,1848	0,6848	15,3565	0,9650	
	λ	3,1863	0,1863	0,1662	0,9400	
500	β_0	-1,2388	-0,0388	0,0473	0,9520	36,39%
	β_1	-0,2980	0,0020	0,0837	0,9010	
	β_2	0,2401	0,1501	1,4297	0,9390	
	β_3	1,0629	0,5629	8,3174	0,9530	
	λ	3,1107	0,1107	0,0783	0,9560	

3.2.3 Aplicação a Dados Reais

Realizamos uma aplicação a dados reais utilizando a base de dados Melanoma contida no [R Development Core Team \(2017a\)](#). Os dados apresentam 205 pacientes da Dinamarca com melanoma maligno. O conjunto de dados é composto por 7 variáveis definidas abaixo:

- *Time*: indicando o tempo observado do paciente, em dias;
- *Status*: indicando se o paciente morreu do melanoma (1), se está vivo (2) e se morreu de outras causas (3);
- *Sex*: indicando o sexo do paciente, 1 se o indivíduo é homem e 0 se o indivíduo é mulher;
- *Age*: indicando a idade (em anos) do paciente no início do experimento;
- *Year*: o ano da operação;
- *Thickness*: indicando a espessura do tumor, em mm;
- *Ulcer*: indicando a presença de ulcera, 1 para presença e 0 para ausência.

Neste aplicação nós modificamos a variável *Status* ao considera-lá apenas sendo 0, se o indivíduo foi censurado e 1, se o indivíduo que atingiu a falha (morreu devido ao melanoma). Além disso, transformamos o tempo de dias para anos.

Análise Descritiva

Para um melhor conhecimento dos dados fizemos uma breve análise descritiva. A [Tabela 10](#) apresenta a porcentagem dos indivíduos classificados por variáveis dicotômicas. Para os indivíduos considerados do grupo 0 na variável úlcera temos a porcentagem dos indivíduos que tem a ausência de úlcera, os indivíduos classificados como 1 indicam a presença de úlcera. Os indivíduos classificados como 0 para a censura representam indivíduos censurados, já os indivíduos classificados como 1 representam os indivíduos não censurados, ou seja, que atingiram o evento de falha. Os indivíduos classificados como 0 para a variável sexo representam as mulheres, já os indivíduos classificados como 1 para a mesma variável representam os homens. Da [Tabela 10](#) é possível perceber a maioria dos pacientes com melanoma maligno são mulheres, além disso também observamos que a proporção de censura dessa base de dados é maior que 70%, indicando alto índice de censura. Podemos notar que os pacientes com melanoma maligno em sua maioria têm a presença de úlcera.

Tabela 10 – Porcentagem de indivíduos para cada grupo nas covariáveis. Para a Censura, 0 são os indivíduos censurados e 1 os indivíduos que atingiram o evento de falha. Para úlcera = 0 representa os indivíduos sem úlcera e úlcera = 1 os indivíduos com úlcera. Para a variável Sexo, 0 representa o sexo feminino e 1 representa o sexo masculino.

Variáveis	Valor		Total
	0	1	
Censura	72,2%	27,8%	100%
Úlcera	56,1%	43,9%	100%
Sexo	61,5%	38,5%	100%

Na [Tabela 11](#) é possível observamos que a maioria dos homens tem úlcera, o que difere para as mulheres em que a ausência de úlcera é maior quando comparamos as mulheres sem úlcera. Podemos observar na [Tabela 12](#) que tanto para o sexo feminino como para o sexo masculino a maioria dos indivíduos são censurados. Já na [Tabela 13](#) é possível observarmos que para os indivíduos que atingiram o evento de falha (morreram devido ao melanoma) a maioria tem úlcera, enquanto que para os indivíduos censurados a maioria deles não tem úlcera.

Tabela 11 – Porcentagem dos indivíduos considerando a covariável úlcera em relação ao sexo.

		Úlcera		Total
		Não	Sim	
Sexo	Feminino	79 (38,5 %)	47 (22,9%)	126 (61,4%)
	Masculino	36 (17,6%)	43 (21,0%)	79 (38,6%)
Total		115 (56,1%)	90 (43,9%)	205 (100%)

Tabela 12 – Porcentagem dos indivíduos censurados e não censurados considerando o sexo dos indivíduos.

		Censura		Total
		0	1	
Sexo	Feminino	98 (47,8%)	28 (13,7%)	126 (61,5%)
	Masculino	50 (24,4%)	29 (14,1%)	79 (38,5%)
Total		148 (72,2%)	57 (27,8%)	205 (100%)

Tabela 13 – Porcentagem dos indivíduos considerando a presença ou ausência de úlcera considerando cada sexo.

		Úlcera		Total
		Não	Sim	
Censura	0	99 (48,3%)	49 (23,9%)	148 (72,2%)
	1	16 (7,8%)	41 (20,0%)	57 (27,8%)
Total		115 (56,1 %)	90 (43,9%)	205 (100%)

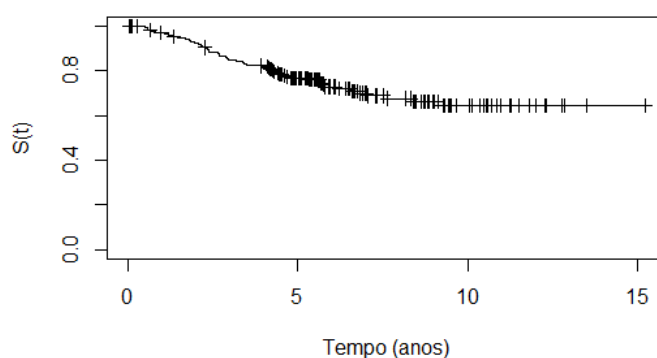
A [Tabela 14](#) apresenta as medidas resumo das variáveis tempo, idade, ano e espessura. Podemos perceber que o maior período do tempo é de aproximadamente 15 anos enquanto que o menor tempo é de aproximadamente 0,03 anos, aproximadamente 11 dias. Pela idade dos pacientes podemos perceber que apesar da menor idade ser de 4 anos e a maior idade ser 95 anos, cerca de 50% dos indivíduos tem idade entre 42 e 65 anos, além disso a média da idade dos indivíduos é de aproximadamente 52 anos. O ano de operação dos indivíduos variou entre 1962 e 1977, mas podemos perceber que cerca de 50% dos indivíduos foram operados entre 1968 e 1972. Também, que apesar da espessura variar muito, cerca de 75% dos pacientes tem o tumor com espessura menor ou igual a 3,5 mm.

Tabela 14 – Medidas resumo das variáveis tempo, idade, ano e espessura.

Variáveis	Medidas Resumo					
	Mínimo	25%	Mediana	Média	75%	Máximo
Tempo	0,0274	4,1781	5,4932	5,8981	8,3342	15,2466
Idade	4,0000	42,0000	54,0000	52,4600	65,0000	95,0000
Ano	1962,0000	1968,0000	1970,0000	1970,0000	1972,0000	1977,0000
Espessura	0,1000	0,9700	1,9400	2,9200	3,5600	17,4200

A [Figura 5](#) exibe a curva de sobrevivência obtida pelo estimador KM, na qual podemos observar um platô bem acima de zero, indicando que modelos que ignoram a possibilidade de cura não serão adequados para esses dados.

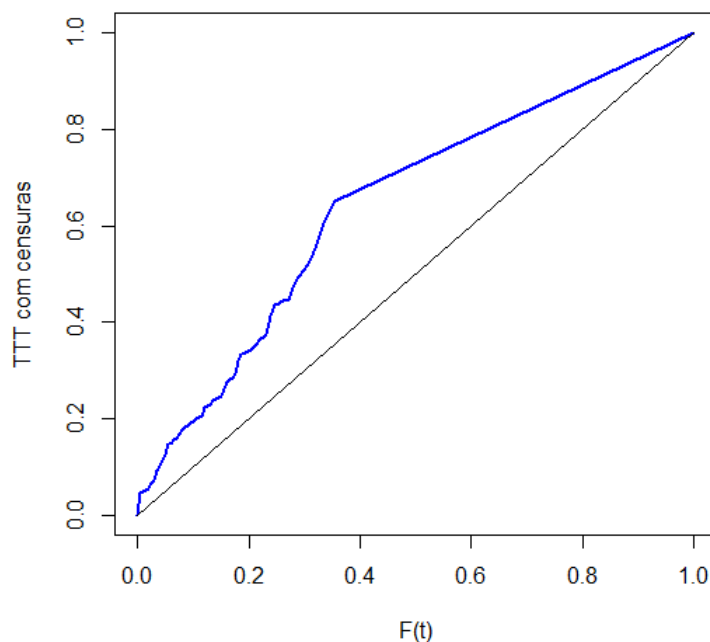
Figura 5 – Gráfico com a curva do estimador de KM para os dados de melanoma.



Fonte: Elaborada pela autora.

[Brunk et al. \(1972\)](#) apresentaram a ideia de um método de análise utilizando o gráfico Tempo Total em Teste (TTT) que fornece informações da função de risco. [Barlow e Campo \(1975\)](#) apresentaram a descrição do modelo do gráfico TTT e as propriedades da análise de envelhecimento. É comum o uso do gráfico TTT para o estudo de qual modelo ajustaria melhor a base de dados. Uma metodologia muito utilizada para o gráfico TTT é a que foi proposta por [Aarset \(1985\)](#) que utiliza apenas os tempos de sobrevivência completos. Como temos muitas observações censuradas na base de dados estudada, utilizamos a teoria do gráfico TTT proposta por [Sun e Kececloglu \(1999\)](#) que permite a incorporação das observações censuradas no cálculo do gráfico, dando informações mais precisas do comportamento da função de risco para dados com censura. A [Figura 6](#) nos mostra pelo gráfico TTT uma curvatura côncava, dessa forma temos indícios que para este conjunto de dados a função de risco dos indivíduos é crescente. A distribuição basal do modelo LIGcr é a IG, o qual acomoda risco crescente.

Figura 6 – Gráfico com o Tempo Total em Teste considerando os indivíduos censurados para os dados de melanoma.

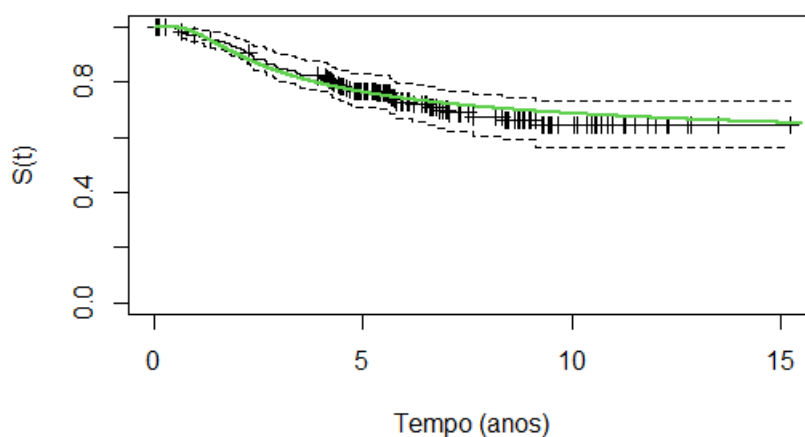


Fonte: Elaborada pela autora.

3.2.3.1 Ajuste sem Covariáveis

Nesta subseção, para ajustar o modelo LIGcr aos dados reais, não consideramos a inclusão de covariáveis. Na [Figura 7](#) podemos perceber que a curva do modelo ajustado captou bem a curva do estimador de KM. Além disso, apresentamos na [Tabela 15](#) o valor da estimativa de máxima verossimilhança de cada parâmetro e o Desvio Padrão (DP). Também, podemos perceber que tanto os parâmetros $p_{0(1)}$ e λ tiveram baixos DP.

Figura 7 – Gráfico com a curva do estimador de KM juntamente com a curva de sobrevivência do modelo LIGcr ajustado (**verde**).



Fonte: Elaborada pela autora.

Tabela 15 – Estimativa de máxima verossimilhança para os parâmetros $p_{0(1)}$, μ e λ e os seus respectivos erros padrão (entre parênteses).

	Parâmetro		
	$p_{0(1)}$	μ	λ
EMV	0,55 (0,10)	76,32 (508,02)	5,53 (0,83)

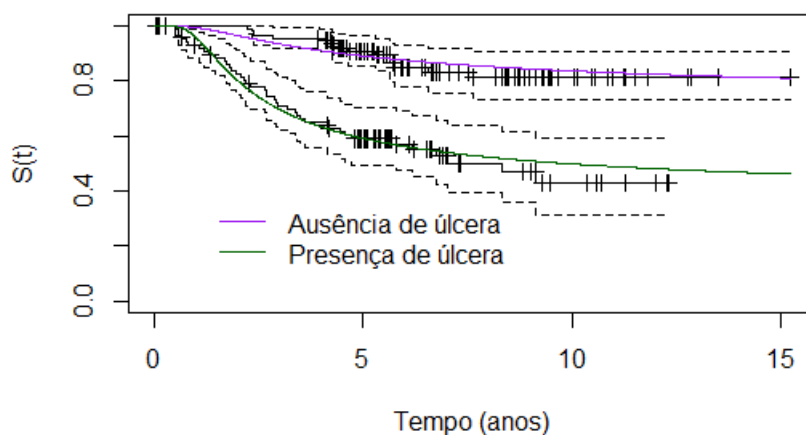
3.2.3.2 Ajuste com Covariáveis

Nesta subseção consideramos a inclusão de covariáveis no modelo LIGcr. Ajustamos o modelo considerando a inclusão da covariável úlcera. No primeiro cenário consideramos a inclusão da covariável úlcera apenas para a proporção de cura do modelo LIGcr como podemos observar pela [Figura 8](#) e [Tabela 16](#). No segundo cenário incluímos a covariável úlcera nos parâmetros $p_{0(1)}$ e μ do modelo LIG, como podemos observar na [Figura 9](#) e [Tabela 17](#). Além disso, apresentamos na [Tabela 18](#), uma tabela com o cálculo do critério AIC e do critério BIC para cada ajuste do modelo LIGcr.

A [Figura 8](#) apresenta a estimativa da função de sobrevivência pela curva de KM considerando o grupo de indivíduos com úlcera e o grupo de indivíduos sem úlcera junto com o ajuste do modelo para cada um dos grupos. Podemos perceber que para ambos os grupos o modelo ajustado captou bem as curvas de sobrevivência obtidas pelo estimador de KM com a sobrevivência de cada grupo. Com a inclusão da covariável úlcera em $p_{0(1)}$, o DP obtida para o parâmetro μ não foi menor. Observamos pelas curvas das sobrevivências KM que independentemente do tempo

as curvas dos grupos não se cruzam, dando o indicativo de riscos proporcionais. Notamos pelas estimativas dos parâmetros β 's que os indivíduos que não tem úlcera tem uma proporção de cura próxima de 71,44%, enquanto que os indivíduos com úlcera tem uma proporção de cura próxima de 35,9%. Assim, os indivíduos sem úlcera desse conjunto de dados tem uma maior probabilidade de sobreviver ao melanoma maligno.

Figura 8 – Estimativa de Kaplan-Meier da função de sobrevivência juntamente com a estimativa da função de sobrevivência do modelo LIGcr considerando a inclusão da covariável úlcera no parâmetro $p_{0(1)}$. O modelo ajustado para os indivíduos sem úlcera está na cor **roxa** e o modelo ajustado para os indivíduos com úlcera está na cor **verde-escuro**.



Fonte: Elaborada pela autora.

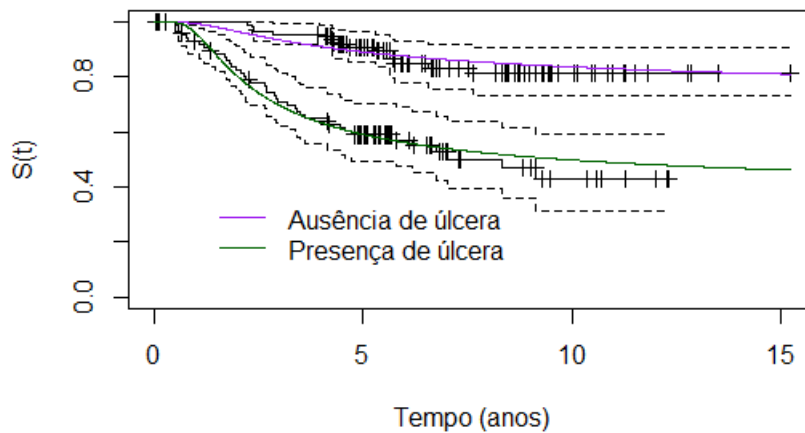
Tabela 16 – Estimativa de máxima verossimilhança e desvio padrão para os parâmetros β_0, β_1, μ e λ considerando a inclusão da covariável úlcera no modelo.

	Parâmetro			
	β_0	β_1	μ	λ
EMV	0,9167 (0,4945)	-1,4963 (0,3177)	86,0381 (566,9268)	6,3755 (0,9573)

Consideramos também um ajuste do modelo LIGcr com a inclusão da covariável úlcera nos parâmetros $p_{0(1)}$ e μ . A Figura 9 apresenta a estimativa da função de sobrevivência obtida pelo estimador de KM considerando o grupo de indivíduos que tem ausência de úlcera e o grupo de indivíduos que tem presença de úlcera juntamente com o ajuste do modelo LIGcr estimado para cada grupo. Para ambos os grupos, o modelo LIGcr que considerou a inclusão da covariável úlcera nos parâmetros $p_{0(1)}$ e μ teve um bom ajuste quando comparamos com a curva do estimador de KM. Apesar do DP dos parâmetros na função de ligação do parâmetro μ ,

(β_2 e β_3), terem o valor de DP bem abaixo do DP do parâmetro μ como na Tabela 16, podemos observar um leve aumento no DP dos parâmetros β_0 e β_1 .

Figura 9 – Estimativa de Kaplan-Meier da função de sobrevivência juntamente com a estimativa da função de sobrevivência do modelo LIGcr considerando a inclusão da covariável úlcera no parâmetro $p_{0(1)}$ e no parâmetro μ . O modelo ajustado para os indivíduos sem úlcera está na cor **roxa** e o modelo ajustado para os os indivíduos com úlcera está na cor **verde-escuro**.



Fonte: Elaborada pela autora.

Tabela 17 – Estimativa de máxima verossimilhança e desvio padrão para $\beta_0, \beta_1, \beta_2, \beta_3$ e λ considerando a inclusão da covariável úlcera nos parâmetros $p_{0(1)}$ e μ .

	Parâmetro				
	β_0	β_1	β_2	β_3	λ
EMV	0,8776 (0,6296)	-1,4218 (0,7776)	5,4687 (24,8219)	-1,5937 (25,4499)	6,3678 (0,9544)

Na Tabela 18 apresentamos os valores de AIC e BIC calculados para cada cenário do modelo, ou seja, com a covariável úlcera apenas em $p_{0(1)}$ e com a covariável úlcera em $p_{0(1)}$ e μ . Podemos observar menor valor de AIC e de BIC para o modelo com covariável apenas na proporção de cura.

Tabela 18 – Estimativas dos critérios AIC e BIC para o modelo LIGcr para dois cenários: considerando a covariável úlcera apenas em $p_{0(1)}$ e considerando a covariável úlcera em $p_{0(1)}$ e μ

Modelo	AIC	BIC
Covariável em $p_{0(1)}$	437,8176	451,1096
Covariável em $p_{0(1)}$ e μ	439,8062	456,4212

MODELO GEOMETRIC *EXPONENTIATED* *POWER GENERALIZED* WEIBULL

4.1 O Modelo GEPGWcr

Para o estudo do segundo modelo estudado considerando longa duração e fragilidade temos que para a construção do modelo Geométric *Exponentiated Power Generalized* Weibull com fração de cura (GEPGWcr) assumimos que a variável aleatória Z , que denota a fragilidade, segue uma distribuição Geométrica começando em zero. Considerando $\omega \in (0, 1)$, a função massa de probabilidade da distribuição é dada pela [Equação 4.1](#):

$$Pr[Z = z; \omega] = (1 - \omega)^z \omega, \quad z=0,1,2,\dots \quad (4.1)$$

A FGP da distribuição Geométrica começando em zero é dada por:

$$G_z(s) = \omega \frac{1}{1 - (s(1 - \omega))}, \quad (4.2)$$

em que $|s(1 - \omega)| < 1$. O cálculo para obter a FGP da distribuição Z está apresentado no [Apêndice B](#).

A FGM da distribuição Geométrica começando em zero é dada pela equação a seguir:

$$M_z(a) = \omega \frac{1}{1 - [e^a(1 - \omega)]}. \quad (4.3)$$

Por meio da [Equação 4.3](#) é possível encontrar a esperança de Z ($E[Z] = M'_z(0)$) e a variância de Z ($Var[Z] = M''_z(0) - (M'_z(0))^2$), que estão apresentadas, respectivamente, na [Equação 4.4](#) e [Equação 4.5](#):

$$E[Z] = \frac{1 - \omega}{\omega}. \quad (4.4)$$

e

$$\text{Var}[Z] = \frac{1 - \omega}{\omega^2}. \quad (4.5)$$

Maiores detalhes dos cálculos da FGM, da Esperança e da Variância da distribuição Geométrica começando em zero estão apresentados no [Apêndice B](#).

Utilizando o resultado de análise de sobrevivência ([Equação 2.8](#)) e a FGP da distribuição Z ([Equação 4.2](#)), obtemos a função de sobrevivência do modelo GEPGWcr, dada pela [Equação 4.6](#):

$$S_{\text{pop}(2)}(t; \boldsymbol{\gamma}_{(2)}) = G_z[S_0(t; \mathbf{v})] = \frac{\omega}{1 - [S_0(t; \mathbf{v})(1 - \omega)]}, \quad (4.6)$$

em que $|S_0(t; \mathbf{v})(1 - \omega)| < 1$, $\boldsymbol{\gamma}_{(2)}$ o vetor de parâmetros do modelo estudado 2 (GEPGWcr), \mathbf{v} o vetor de parâmetros da distribuição basal (EPGW) e, $S_0(t; \mathbf{v})$ é uma função de sobrevivência basal e própria, ou seja, para $t \rightarrow 0$, temos que $S_0(t; \mathbf{v}) = 1$ e para $t \rightarrow \infty$, temos que $S_0(t; \mathbf{v}) = 0$.

Para estudar o comportamento da função $S_{\text{pop}(2)}(t; \boldsymbol{\gamma}_{(2)})$ aplicamos o limite para $t \rightarrow 0$ ([Equação 4.7](#)) e para $t \rightarrow \infty$ ([Equação 4.8](#)) apresentadas a seguir:

$$\lim_{t \rightarrow 0} S_{\text{pop}(2)}(t; \boldsymbol{\gamma}_{(2)}) = \lim_{t \rightarrow 0} \frac{\omega}{1 - [S_0(t; \mathbf{v})(1 - \omega)]} = 1 \quad (4.7)$$

e,

$$\lim_{t \rightarrow \infty} S_{\text{pop}(2)}(t; \boldsymbol{\gamma}_{(2)}) = \lim_{t \rightarrow \infty} \frac{\omega}{1 - [S_0(t; \mathbf{v})(1 - \omega)]} = \omega. \quad (4.8)$$

Observamos na [Equação 4.8](#) que a função de sobrevivência é imprópria e tem uma proporção de cura como sendo ω .

Utilizando as relações existentes entre $S_{\text{pop}}(t; \boldsymbol{\gamma})$, $f_{\text{pop}}(t; \boldsymbol{\gamma})$ e $h_{\text{pop}}(t; \boldsymbol{\gamma})$ citadas na [Subseção 2.2.2](#) e, considerando $f_0(t; \mathbf{v})$ e $S_0(t; \mathbf{v})$ as funções de densidade e sobrevivência basal, temos a função de densidade e função de risco do modelo GEPGWcr dadas, respectivamente, pela [Equação 4.9](#) e [Equação 4.10](#):

$$\begin{aligned}
f_{\text{pop}(2)}(t; \boldsymbol{\gamma}_{(2)}) &= - \left(\frac{d}{dt} (S_{\text{pop}(2)}(t; \boldsymbol{\gamma}_{(2)})) \right) = - \left(\frac{d}{dt} \left(\frac{\omega}{1 - [S_0(t; \mathbf{v})(1 - \omega)]} \right) \right) \\
&= - \frac{d}{dt} (\omega(1 - [S_0(t; \mathbf{v})(1 - \omega)])^{-1}) \\
&= -\omega(-1)(1 - S_0(t; \mathbf{v})(1 - \omega))^{-2}(-S'_0(t; \mathbf{v}))(1 - \omega) \\
&= \frac{\omega f_0(t; \mathbf{v})(1 - \omega)}{(1 - S_0(t; \mathbf{v})(1 - \omega))^2}
\end{aligned} \tag{4.9}$$

e,

$$\begin{aligned}
h_{\text{pop}(2)}(t; \boldsymbol{\gamma}_{(2)}) &= \frac{f_{\text{pop}(2)}(t; \boldsymbol{\gamma}_{(2)})}{S_{\text{pop}(2)}(t; \boldsymbol{\gamma}_{(2)})} \\
&= \frac{\omega f_0(t; \mathbf{v})(1 - \omega)(1 - S_0(t; \mathbf{v})(1 - \omega))^{-2}}{\omega(1 - [S_0(t; \mathbf{v})(1 - \omega)])^{-1}} \\
&= \frac{\omega f_0(t; \mathbf{v})(1 - \omega)[1 - S_0(t; \mathbf{v})(1 - \omega)]}{\omega(1 - S_0(t; \mathbf{v})(1 - \omega))^2} \\
&= \frac{f_0(t; \mathbf{v})(1 - \omega)}{(1 - S_0(t; \mathbf{v})(1 - \omega))}.
\end{aligned} \tag{4.10}$$

Representação como Modelo de Mistura

Como a função de sobrevivência do modelo é imprópria, seguindo o resultado de [Rodrigues, Cancho e Castro \(2008\)](#), podemos escrevê-la como modelo de mistura e encontrar $S_{(2)}^*(t; \mathbf{v}^{**})$, que é a função de sobrevivência própria que compõe a mistura de $S_{\text{pop}(2)}(t; \boldsymbol{\gamma}_{(2)})$ e \mathbf{v}^{**} é o vetor de parâmetros da distribuição $S_{(2)}^*(t; \mathbf{v}^{**})$. A [Equação 4.11](#) mostra a função de sobrevivência própria $S_{(2)}^*(t; \mathbf{v}^{**})$. A seguir apresentamos o modelo escrito como modelo de mistura:

$$S_{\text{pop}(2)}(t; \boldsymbol{\gamma}_{(2)}) = \frac{\omega}{1 - S_0(t; \mathbf{v})(1 - \omega)} = p_{0(2)} + (1 - p_{0(2)})S_{(2)}^*(t; \mathbf{v}^{**}),$$

em que $p_{0(2)}$ é a fração de cura do modelo de mistura. Isolando o $S_{(2)}^*(t; \mathbf{v}^{**})$ na equação acima, temos que:

$$\begin{aligned}
S_{(2)}^*(t; \mathbf{v}^{**}) &= \frac{\frac{\omega}{1 - S_0(t; \mathbf{v})(1 - \omega)} - \omega}{1 - \omega} \\
&= \frac{\omega - \omega(1 - S_0(t; \mathbf{v})(1 - \omega))}{1 - S_0(t; \mathbf{v})(1 - \omega)} \frac{1}{1 - \omega} \\
&= \frac{\omega S_0(t; \mathbf{v})}{1 - S_0(t; \mathbf{v})(1 - \omega)},
\end{aligned} \tag{4.11}$$

em que $S_0(t; \mathbf{v})$ é a função de densidade basal escolhida.

Por meio da [Equação 4.11](#), podemos mostrar que $S_{(2)}^*(t; \mathbf{v}^{**})$ é uma função própria calculando os limites, como é mostrado pelas equações a seguir:

$$\lim_{t \rightarrow 0} S_{(2)}^*(t; \mathbf{v}^{**}) = \lim_{t \rightarrow 0} \frac{\omega S_0(t; \mathbf{v})}{1 - S_0(t; \mathbf{v})(1 - \omega)} = \frac{\omega}{1 - 1 + \omega} = 1$$

e,

$$\lim_{t \rightarrow \infty} S_{(2)}^*(t; \mathbf{v}^{**}) = \lim_{t \rightarrow \infty} \frac{\omega S_0(t; \mathbf{v})}{1 - S_0(t; \mathbf{v})(1 - \omega)} = \frac{\omega 0}{1 - 0(1 + \omega)} = 0.$$

A função de densidade e a função de risco relacionadas com a função de sobrevivência de $S_{(2)}^*(t; \mathbf{v}^{**})$ são apresentadas, respectivamente, por:

$$\begin{aligned} f_{(2)}^*(t; \mathbf{v}^{**}) &= -\frac{d}{dt} \left(\frac{\omega S_0(t; \mathbf{v})}{1 - S_0(t; \mathbf{v})(1 - \omega)} \right) \\ &= -\left[\frac{\omega S_0'(t; \mathbf{v})(1 - S_0(t; \mathbf{v}))(1 - \omega) - \omega S_0(t; \mathbf{v})(-S_0'(t; \mathbf{v})(1 - \omega))}{(1 - S_0(t; \mathbf{v})(1 - \omega))^2} \right] \\ &= -\left[\frac{\omega S_0'(t; \mathbf{v})(1 - \omega)(1 - S_0(t; \mathbf{v}) + S_0(t; \mathbf{v}))}{(1 - S_0(t; \mathbf{v})(1 - \omega))^2} \right] \\ &= \frac{\omega f_0(t; \mathbf{v})(1 - \omega)(1 - S_0(t; \mathbf{v}) + S_0(t; \mathbf{v}))}{(1 - S_0(t; \mathbf{v})(1 - \omega))^2} \end{aligned}$$

e,

$$\begin{aligned} h_{(2)}^*(t; \mathbf{v}^{**}) &= \frac{\frac{\omega f_0(t; \mathbf{v})(1 - \omega)(1 - S_0(t; \mathbf{v}) + S_0(t; \mathbf{v}))}{(1 - S_0(t; \mathbf{v})(1 - \omega))^2}}{\frac{\omega S_0(t; \mathbf{v})}{1 - S_0(t; \mathbf{v})(1 - \omega)}} \\ &= \frac{f_0(t; \mathbf{v})}{S_0(t; \mathbf{v})} \frac{(1 - \omega)}{S_0(t; \mathbf{v})(1 - S_0(t; \mathbf{v})(1 - \omega))} \\ &= h_0(t; \mathbf{v}) \frac{(1 - \omega)}{S_0(t; \mathbf{v})(1 - S_0(t; \mathbf{v})(1 - \omega))}. \end{aligned}$$

A Proporção de Cura

Na [Equação 4.8](#) mostramos que quando $t \rightarrow \infty$, a função de sobrevivência do modelo GEPGWcr tende a ω . Assim, consideramos que a proporção de cura do modelo GEPGWcr seja denotada por $p_{(2)}$ e escrita como $p_{(2)} = \omega$, em que ω é o parâmetro da distribuição geométrica. Logo, temos que a proporção de cura pertence ao intervalo $[0, 1]$. Assim, é possível escrever a

função de densidade e a função de sobrevivência do modelo GEPGWcr em função de $p_{0(2)}$, que são apresentadas, respectivamente, pela Equação 4.12 e Equação 4.13:

$$f_{\text{pop}(2)}(t; \boldsymbol{\gamma}_{(2)}) = \frac{p_{0(2)} f_0(t; \mathbf{v})(1 - p_{0(2)})}{(1 - S_0(t)(1 - p_{0(2)}))^2} \quad (4.12)$$

e,

$$S_{\text{pop}(2)}(t; \boldsymbol{\gamma}_{(2)}) = \frac{p_{0(2)}}{1 - [S_0(t; \mathbf{v})(1 - p_{0(2)})]}, \quad (4.13)$$

em que $f_0(t; \mathbf{v})$ e $S_0(t; \mathbf{v})$ são as funções de densidade e sobrevivência basal, respectivamente, que serão descritas a seguir.

A Distribuição Basal

Pena-Ramirez *et al.* (2018) apresentaram a distribuição *Eponentiated Power Generalized Weibull* (EPGW) considerando a distribuição *Power Generalized Weibull* (PGW), proposta por Bagdonavicius e Nikulin (2001), em um contexto para tempo de vida acelerado. A distribuição EPGW pertence a família exp-G (PENA-RAMIREZ *et al.*, 2018). Essa distribuição é bastante flexível, acomodando várias distribuições conhecidas de tempo de vida e ao mesmo tempo, alguns modelos especiais com características bastante convenientes para teste de bondade de ajuste das distribuições que são englobados pela EPGW. Com quatro parâmetros, sendo um parâmetro que flexibiliza a assimetria e a curtose (PENA-RAMIREZ *et al.*, 2018).

Considere T uma variável aleatória contínua e positiva, $\mathbf{v}_{(2)} = (\alpha, \eta, \psi, \nu)$ o vetor de parâmetros da distribuição EPGW, em que $\alpha > 0$, $\eta > 0$, $\psi > 0$ e $\nu > 0$. O parâmetro ψ é o parâmetro escala e α, η, ν são parâmetros de forma. Temos que a função densidade de probabilidade é dada pela Equação 4.14:

$$f_{EPGW}(t; \mathbf{v}_{(2)}) = \alpha \eta \psi \nu t^{\nu-1} \frac{(1 + \psi t^\nu)^{\alpha-1} \exp\{1 - (1 + \psi t^\nu)^\alpha\}}{[1 - \exp\{1 - (1 + \psi t^\nu)^\alpha\}]^{1-\eta}}. \quad (4.14)$$

A função de densidade acumulada e a função quantil da distribuição EPGW são dadas, respectivamente, pela Equação 4.15 e Equação 4.16:

$$F_{EPGW}(t; \mathbf{v}_{(2)}) = [1 - \exp\{1 - (1 + \psi t^\nu)^\alpha\}]^\eta \quad (4.15)$$

e,

$$Q_{EPGW}(u; \mathbf{v}_{(2)}) = \psi^{-1/\nu} \left\{ 1 - \log(1 - u^{1/\eta})^{1/\alpha} \right\}^{1/\nu}, \quad (4.16)$$

em que $u \in (0, 1)$ e $\mathbf{v}_{(2)}$ é o vetor de parâmetros da distribuição EPGW.

Para $t > 0, \alpha < 1, \eta < 1$, e $v < 1$ a distribuição EPGW é log-convexa; já para $t > 0, \alpha > 1, \eta > 1$, e $v > 1$ a distribuição é log-côncava (PENA-RAMIREZ *et al.*, 2018).

Podemos encontrar o s -ésimo momento da distribuição EPGW, apresentado na Equação 4.17:

$$\mu'_s = E(T^s) = \eta \psi^{-s/v} I_s(\alpha, \eta, v), \quad (4.17)$$

em que $I_s(\alpha, \eta, v) = \int_0^1 \{[1 - \log(1 - u)]^{s/v} u^{\eta-1} du$.

A função de risco da distribuição EPGW e a função de sobrevivência são dadas, respectivamente, pela Equação 4.18 e Equação 4.19:

$$\begin{aligned} h_{EPGW}(t; \mathbf{v}_{(2)}) &= \alpha \eta \psi v t^{v-1} (1 + \psi t^v)^{\alpha-1} \\ &\times \frac{\exp\{1 - (1 + \psi t^v)^\alpha\} [1 - \exp\{1 - (1 + \psi t^v)^\alpha\}]^{1-\eta}}{1 - [1 - \exp\{1 - (1 + \psi t^v)^\alpha\}]^\eta} \end{aligned} \quad (4.18)$$

e

$$S_{EPGW}(t; \mathbf{v}_{(2)}) = 1 - [1 - \exp\{1 - (1 + \psi t^v)^\alpha\}]^\eta, \quad (4.19)$$

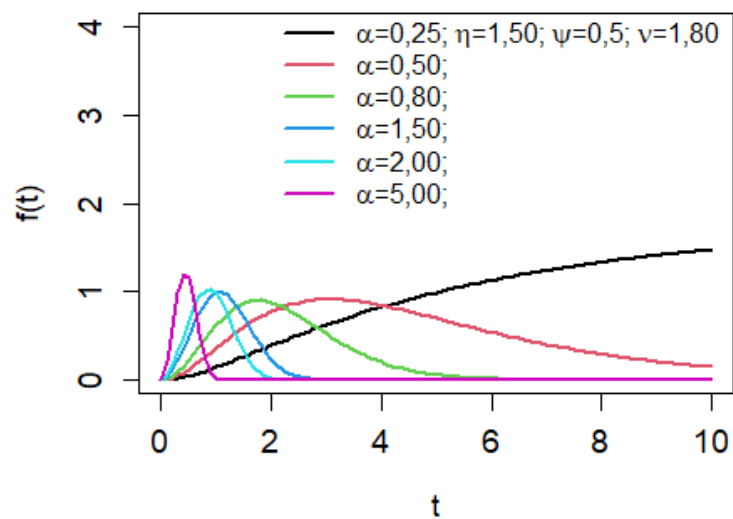
em que $\mathbf{v}_{(2)}$ é o vetor de parâmetros da distribuição EPGW

Comportamento da distribuição Exponentiated Power Generalized Weibull

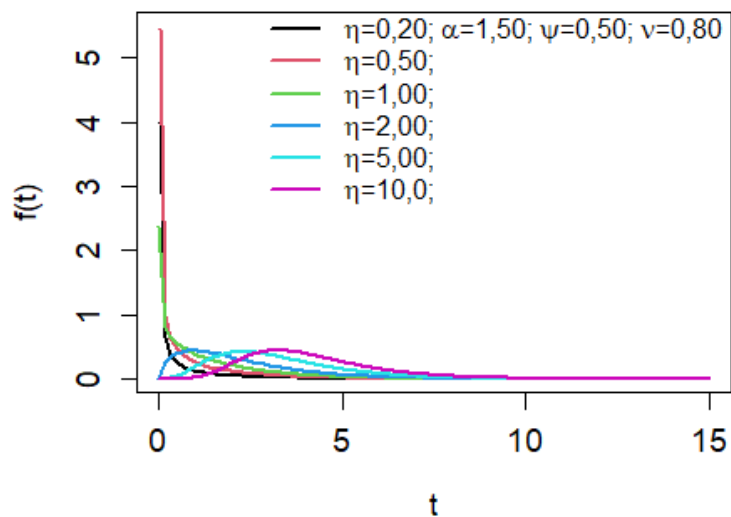
Neste seção apresentamos um estudo de comportamento das funções de densidade, risco e sobrevivência da distribuição EPGW.

As Figura 10 e Figura 11 apresentam as formas da função densidade de probabilidade da distribuição EPGW para diferentes valores dos seus parâmetros. Na Figura 10a fixamos os parâmetros $\eta = 1,5$; $\psi = 0,5$ e $v = 1,8$ e variamos os valores de α . Observamos que conforme menor o valor de α menos pesada é a cauda direita. Já na Figura 10b, apresenta o comportamento considerando os parâmetros $\alpha = 1,5$; $\psi = 0,5$ e $v = 0,8$ e variamos os valores de η . Podemos observar que a distribuição tende a ficar mais achatada com maiores valores de η . Na Figura 11a, os valores dos parâmetros foram $\alpha = 1,2$; $\eta = 1,5$ e $v = 1,8$ e variamos os valores de ψ . Notamos que, apesar da cauda direita da distribuição ser maior que a cauda esquerda em todos os casos desse cenário, tivemos uma simetria considerável. Além disso, é possível notar que o parâmetro ψ influencia na curtose da distribuição EPGW. Para a Figura 11b, os parâmetros foram $\alpha = 0,5$; $\eta = 0,8$ e $\psi = 1,2$ e variamos os valores de v . Observamos que o parâmetro v influencia na curtose, quanto mais baixo o seu valor, mais achatada fica a distribuição. Além disso, notamos que o parâmetro η também influencia na centralização da distribuição.

Figura 10 – Gráfico com as funções de densidade de probabilidade para diferentes valores dos parâmetros. No gráfico (a) fixando os parâmetros $\eta = 1,5$, $\psi = 0,5$ e $\nu = 1,8$ e variando os valores do parâmetro α . No gráfico (b), variando o parâmetro η e fixando os parâmetros $\alpha = 1,5$, $\psi = 0,5$ e $\nu = 0,8$.



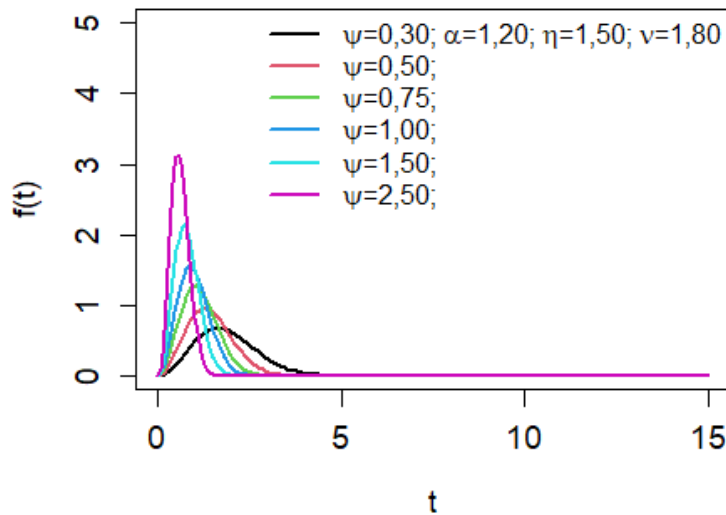
(a)



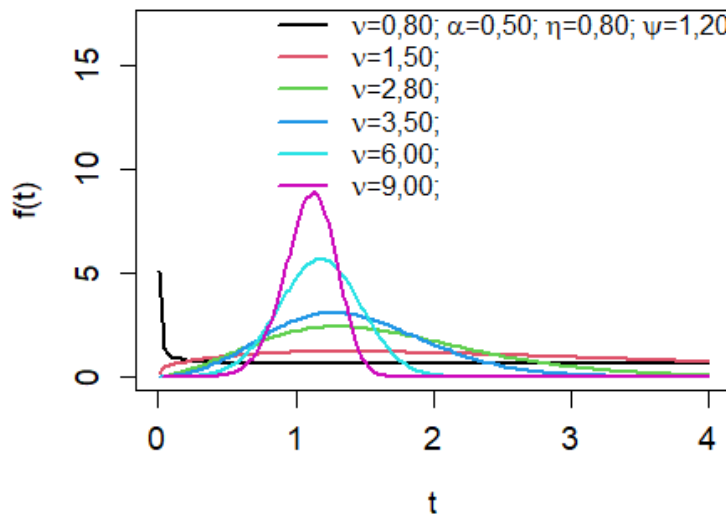
(b)

Fonte: Elaborada pela autora.

Figura 11 – Gráfico com as funções de densidade de probabilidade para diferentes valores dos parâmetros. No gráfico (a), fixando os parâmetros $\alpha = 1,2$, $\eta = 1,0$ e $v = 1,8$ e variando os valores do parâmetro ψ . No gráfico (b), variando os valores de v e fixando os parâmetros $\alpha = 0,5$, $\eta = 0,8$ e $\psi = 1,2$.



(a)



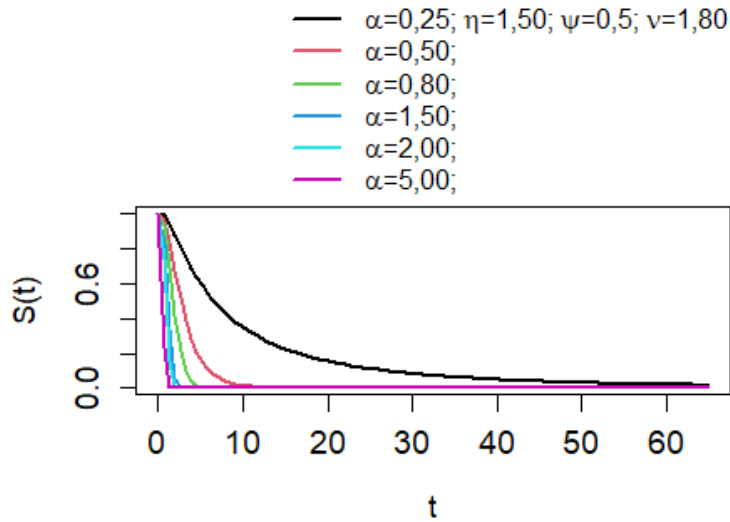
(b)

Fonte: Elaborada pela autora.

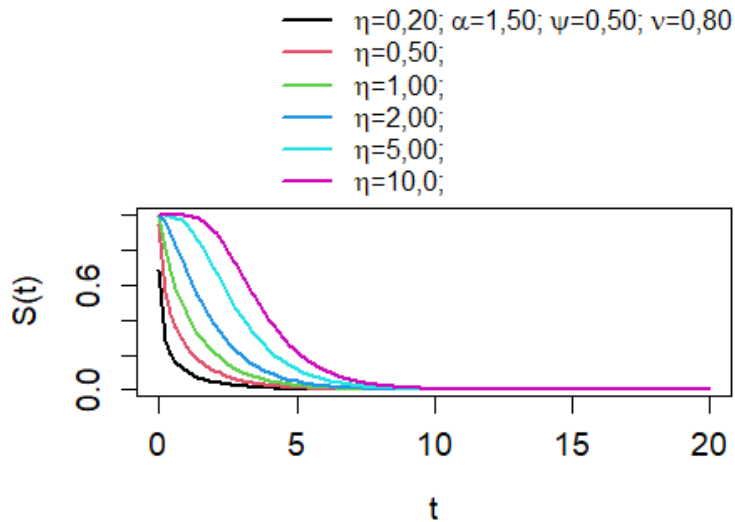
Nas [Figura 12](#) e [Figura 13](#) temos diferentes comportamentos da função de sobrevivência de EPGW. Observamos que a função de sobrevivência sempre vai para 0. Para a [Figura 12a](#), consideramos os parâmetros $\eta = 1,5$, $\psi = 0,5$ e $v = 1,8$ fixos e variamos os valores de α . Observamos que, apesar da curva de sobrevivência ter uma rápida queda na maioria dos casos, para pequenos valores de α a curva de sobrevivência tenha um decrescimento mais suave. A

Figura 12b nos apresenta o comportamento para $\alpha = 1,5$; $\psi = 0,5$ e $v = 0,8$ fixos. Nota-se que, apesar do aumento no parâmetro η influenciar o início da curva de sobrevivência fazendo com que ela demore mais para decrescer, ele não influencia tanto no comportamento decrescente da função de sobrevivência como o parâmetro α . Já na Figura 13a, os parâmetros foram $\alpha = 1,2$; $\eta = 1,5$ e $v = 1,8$, fixos. Notamos que, menores valores de ψ auxiliam a curva a ser mais suave apesar de não influenciar tanto no tempo da função de sobrevivência a atingir o zero. Na Figura 13b consideramos os parâmetros $\alpha = 0,5$; $\eta = 0,8$ e $\psi = 1,2$ são fixos e variamos os valores de v . Observa-se que, o parâmetro v também influencia no decrescimento da curva de sobrevivência. Para grandes valores a curva de sobrevivência aparenta decrescer mais rapidamente, já para menores valores de v , a curva da sobrevivência aparenta demorar mais para decrescer.

Figura 12 – Gráfico com as funções de sobrevivência. No gráfico (a), fixando os parâmetros $\eta = 1,5$, $\psi = 0,5$ e $\nu = 1,8$ e variando os valores do parâmetro α . Em segundo, no gráfico (b), variando o parâmetro η e fixando os parâmetros $\alpha = 1,5$, $\psi = 0,5$ e $\nu = 0,8$. No gráfico (c), fixando os parâmetros $\alpha = 1,2$, $\eta = 1,0$ e $\nu = 1,8$ e variando os valores do parâmetro ψ . No gráfico (d), variando os valores de ν e fixando os parâmetros $\alpha = 0,5$, $\eta = 0,8$ e $\psi = 1,2$.



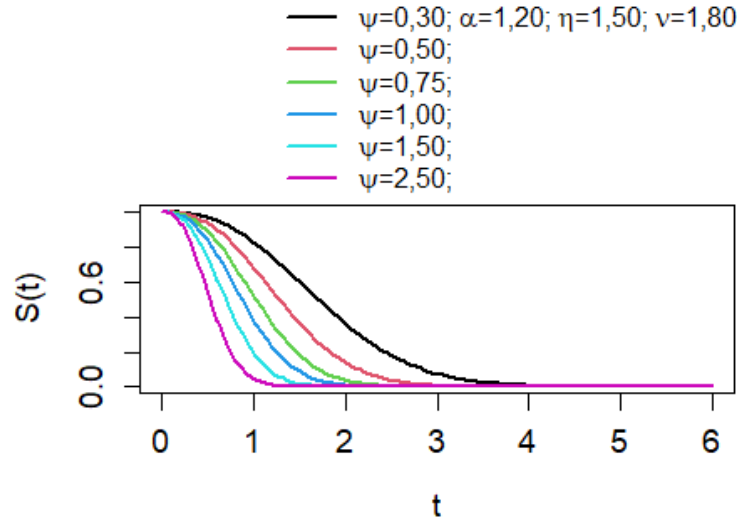
(a)



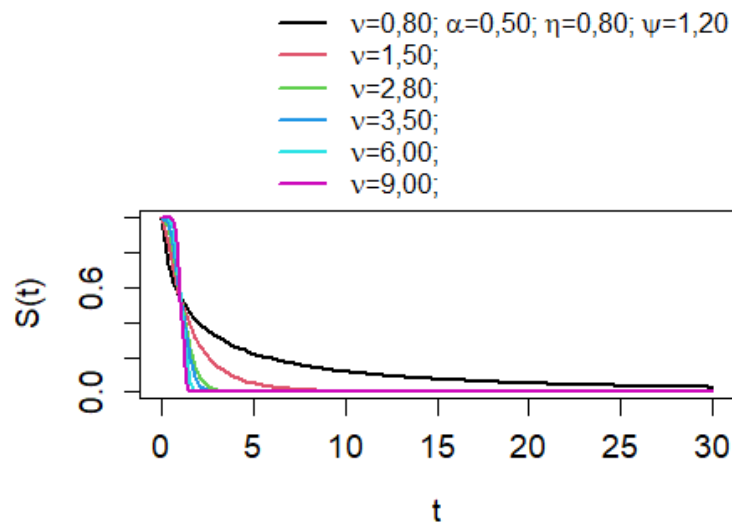
(b)

Fonte: Elaborada pela autora.

Figura 13 – No gráfico (a), fixando os parâmetros $\alpha = 1,2$, $\eta = 1,0$ e $\nu = 1,8$ e variando os valores do parâmetro ψ . No gráfico (b), variando os valores de ν e fixando os parâmetros $\alpha = 0,5$, $\eta = 0,8$ e $\psi = 1,2$.



(a)



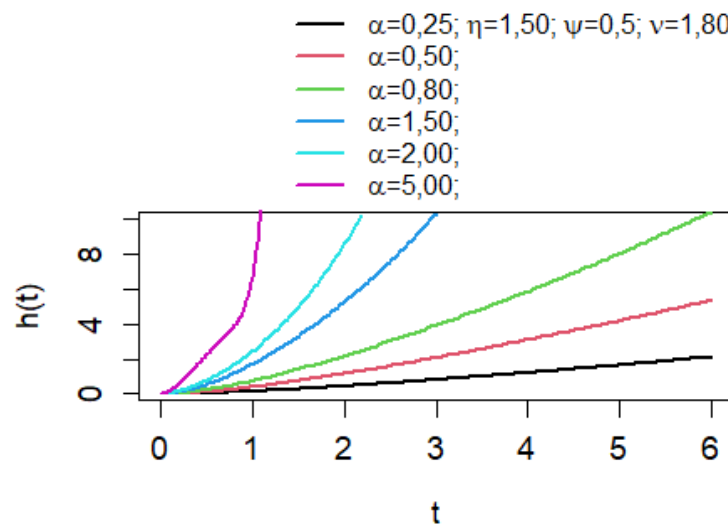
(b)

Fonte: Elaborada pela autora.

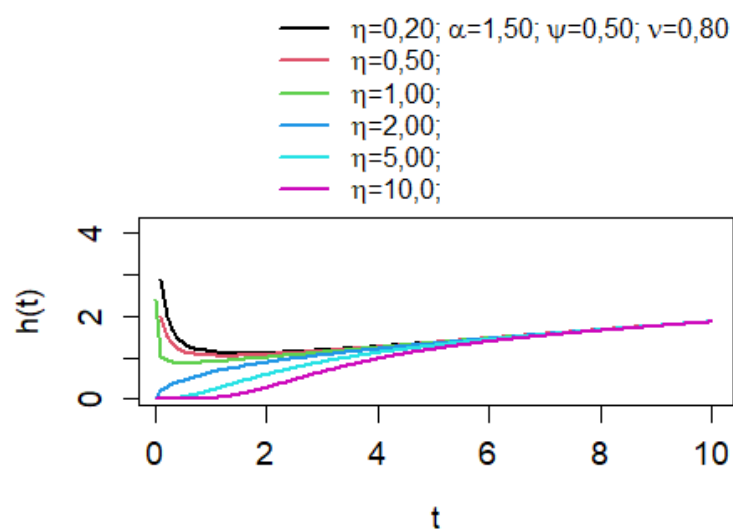
Nas [Figura 14](#) e [Figura 15](#) apresentamos o comportamento da função de risco de EPGW para diferentes configurações dos parâmetros. Na [Figura 14a](#) os parâmetros foram fixados em $\eta = 1,5$; $\psi = 0,5$ e $\nu = 1,8$ e os valores de α foram variados. Já para a [Figura 14b](#) temos fixos os parâmetros $\alpha = 1,5$; $\psi = 0,5$ e $\nu = 0,8$ fixos e os valores de η foram variados. Na [Figura 15a](#) consideramos os parâmetros $\alpha = 1,2$; $\eta = 1,5$ e $\nu = 1,8$ e os valores de ψ foram variados. E, a

Figura 15b os parâmetros $\alpha = 0,5$; $\eta = 0,8$ e $\psi = 1,2$ e os valores de ν foram variados. Nas Figura 14 e Figura 15 podemos notar que a função de risco da distribuição EPGW é bem flexível podendo se comportar de muitas formas como na Figura 14b e Figura 15b. Além disso, podemos observar na Figura 14b e Figura 15b que os parâmetros η e ν possivelmente são os que mais influenciaram na flexibilidade da função de risco. Já os parâmetros α e ψ influenciam mais na curvatura da função de risco como podemos observar na Figura 14a e Figura 15a, em que menores valores para esses parâmetros aparentam fazer com que a curvatura da função fique mais suave ou mais abrupta.

Figura 14 – Gráfico com as funções de risco para diferentes valores dos parâmetros. No gráfico (a), fixando os parâmetros $\eta = 1,5$, $\psi = 0,5$ e $\nu = 1,8$ e variando os valores do parâmetro α . No gráfico (b), variando o parâmetro η e fixando os parâmetros $\alpha = 1,5$, $\psi = 0,5$ e $\nu = 0,8$.



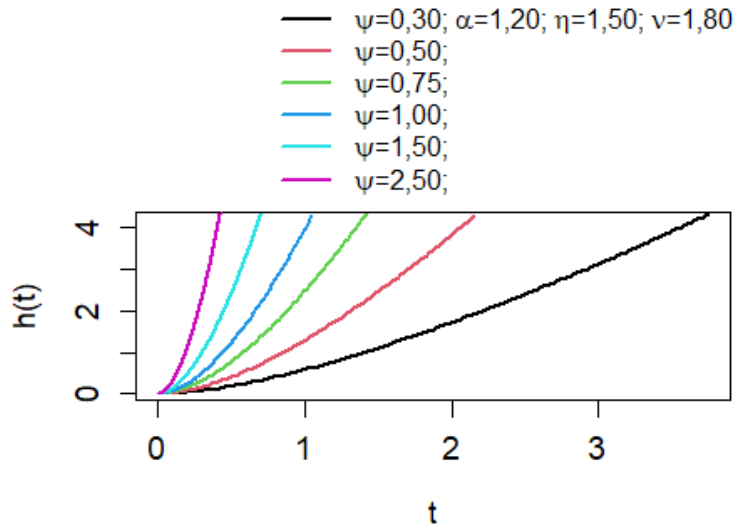
(a)



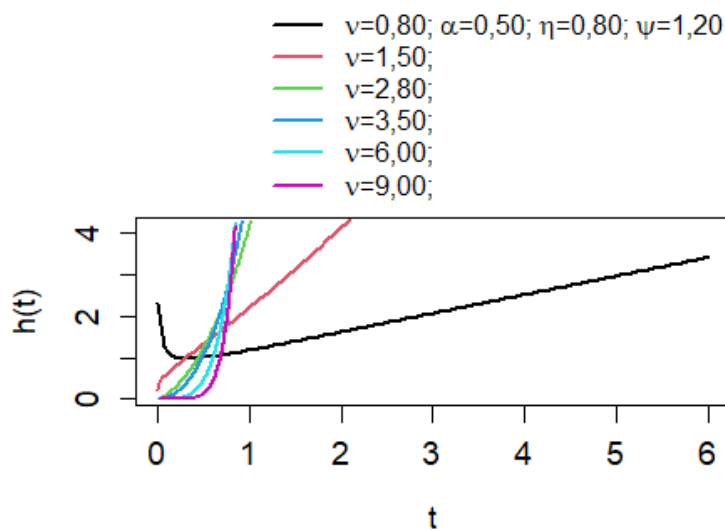
(b)

Fonte: Elaborada pela autora.

Figura 15 – No gráfico (c), fixando os parâmetros $\alpha = 1,2$, $\eta = 1,0$ e $v = 1,8$ e variando os valores do parâmetro ψ . No gráfico (d), variando os valores de v e fixando os parâmetros $\alpha = 0,5$, $\eta = 0,8$ e $\psi = 1,2$.



(a)



(b)

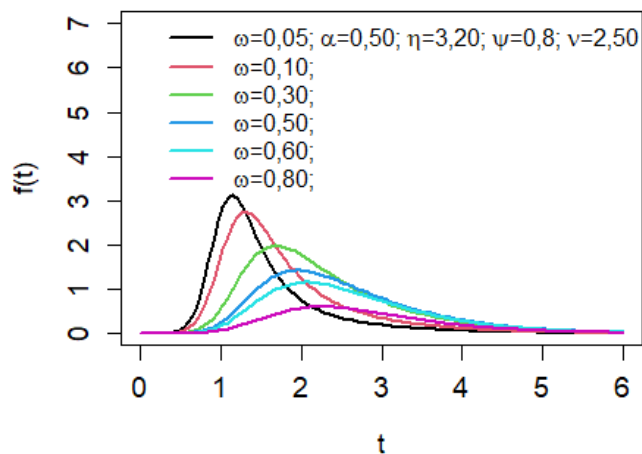
Fonte: Elaborada pela autora.

Comportamento do Modelo GEPGWcr

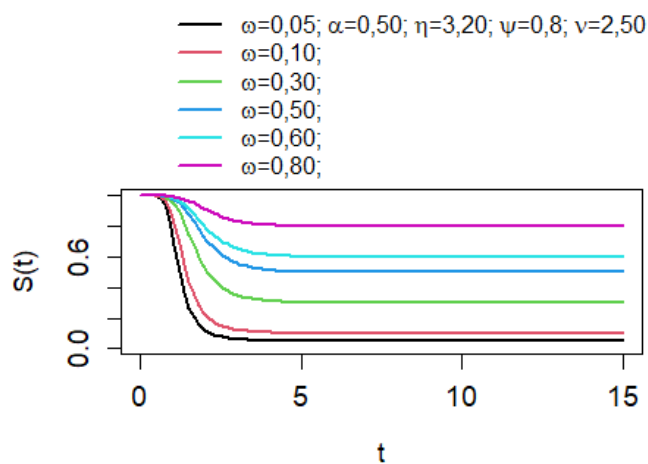
A Figura 16 apresenta o comportamento das funções de densidade de probabilidade (Equação 4.9), sobrevivência (Equação 4.6) e risco (Equação 4.10) do modelo GEPGWcr para diferentes valores da proporção de cura, ω . Os valores dos parâmetros α , η , ψ e v foram fixados em: 0,5; 3,2; 0,8 e 2,5, respectivamente. O comportamento da função de densidade e da função

de sobrevivência estão apresentados na [Figura 16a](#) e [Figura 16b](#), respectivamente. Podemos notar que, a partir do valor aproximadamente 5, a curva de sobrevivência se estabiliza mostrando a proporção de cura. Já na [Figura 16c](#) podemos observar os comportamentos da função de risco do modelo. É possível notar que a função de risco do modelo GEPGWcr é flexível, podendo apresentar assimetria unimodal, com diferentes valores de curtoses.

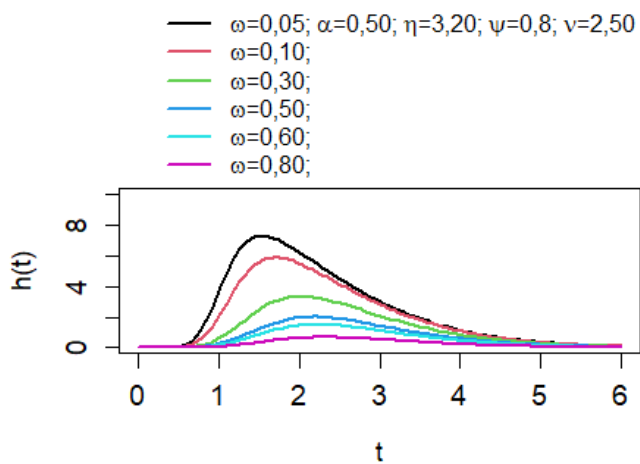
Figura 16 – Gráfico com as funções de densidade de probabilidade, sobrevivência e risco do modelo GEPGW para diferentes valores da proporção de cura, ω e fixando os parâmetros $\alpha = 0,5$, $\eta = 3,2$, $\psi = 0,8$ e $\nu = 2,5$



(a)



(b)



(c)

Fonte: Elaborada pela autora.

4.1.1 Inferência Clássica

Considerando que o tempo de censura do i -ésimo indivíduo é denotado por C_i e que o tempo de falha do i -ésimo indivíduo é denotado por Y_i , temos que o tempo observado do i -ésimo indivíduo é dado pela variável $T_i = \min\{Y_i, C_i\}$. Sendo δ_i o indicador de censura do i -ésimo indivíduo, como definido na [Seção 2.1](#).

Se o indivíduo é censurado, o tempo observado é C_i e, se o indivíduo atinge a falha, o tempo observado é Y_i . Seja $\mathbf{t} = (t_1, t_2, \dots, t_n)^T$ o vetor de tempos observados, $\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_n)^T$ o vetor de censura dos indivíduos e $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$ o vetor de covariáveis do i -ésimo indivíduo e $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^T$ o vetor de covariáveis dos indivíduos. Temos que $\mathbf{D} = (\mathbf{t}, \boldsymbol{\delta}, \mathbf{x})$ é o vetor de dados completos e $\boldsymbol{\vartheta}_{(2)} = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}_{(2)}^T)$ o vetor de parâmetros do modelo GEPGWcr com o vetor dos parâmetros das covariáveis, composto por $\boldsymbol{\beta}^T$ e $\boldsymbol{\gamma}_{(2)}^T = (\alpha, \eta, \psi, \nu, p_{0(2)})$ é o vetor dos parâmetros do modelo GEPGWcr, sendo $p_{0(2)}$ a proporção de cura do modelo GEPGWcr.

Considerando que o vetor de parâmetros da distribuição basal do modelo GEPGWcr é denotado por $\mathbf{v}_{(2)}$ e que distribuição basal do modelo é a EPGW, temos que $\mathbf{v}_{(2)} = c(\alpha, \eta, \psi, \nu)$. Dessa forma, a função de verossimilhança é dada por:

$$\begin{aligned} L(\boldsymbol{\vartheta}_{(2)}; \mathbf{D}) &= \prod_{i=1}^n (f_{\text{pop}(2)}(t_i; \boldsymbol{\vartheta}_{(2)}))^{\delta_i} (S_{\text{pop}(2)}(t_i; \boldsymbol{\vartheta}_{(2)}))^{1-\delta_i} \\ &= \prod_{i=1}^n \left(\frac{p_{0(2)} f_0(t_i; \mathbf{v}_{(2)}) (1 - p_{0(2)})}{(1 - S_0(t_i; \mathbf{v}_{(2)})(1 - p_{0(2)}))^2} \right)^{\delta_i} \times \left(\frac{p_{0(2)}}{1 - [S_0(t_i; \mathbf{v}_{(2)})(1 - p_{0(2)})]} \right)^{1-\delta_i}, \end{aligned}$$

em que $p_{0(2)} = \omega$ e $S_0(t; \mathbf{v}_{(2)})$ e $f_0(t; \mathbf{v}_{(2)})$ são as funções de sobrevivência e densidade basal, respectivamente.

O logaritmo da função de verossimilhança é dado por:

$$\begin{aligned} l(\boldsymbol{\vartheta}_{(2)}; \mathbf{D}) &= \sum_{i=1}^n \delta_i \ln(p_{0(2)} f_0(t_i; \mathbf{v}_{(2)}) (1 - p_{0(2)})) - \delta_i \ln(1 - S_0(t_i; \mathbf{v}_{(2)}) (1 - p_{0(2)}))^2 \\ &\quad + \sum_{i=1}^n (1 - \delta_i) \ln(p_{0(2)}) - (1 - \delta_i) \ln(1 - S_0(t_i; \mathbf{v}_{(2)}) (1 - p_{0(2)})) \\ &= n \ln(p_{0(2)}) - \sum_{i=1}^n \ln(1 - S_0(t_i; \mathbf{v}_{(2)}) (1 - p_{0(2)})) + \sum_{i=1}^n \delta_i \ln(p_{0(2)} f_0(t_i; \mathbf{v}_{(2)}) (1 - p_{0(2)})) \\ &\quad - \left(\sum_{i=1}^n \delta_i \ln(p_{0(2)}) + \sum_{i=1}^n \delta_i \ln(1 - S_0(t_i; \mathbf{v}_{(2)}) (1 - p_{0(2)})) \right) \\ &= n \ln(p_{0(2)}) + \sum_{i=1}^n \ln(1 - S_0(t_i; \mathbf{v}_{(2)}) (1 - p_{0(2)})) + \sum_{i=1}^n \delta_i \ln \left(\frac{f_0(t_i; \mathbf{v}_{(2)}) (1 - p_{0(2)})}{1 - S_0(t_i; \mathbf{v}_{(2)}) (1 - p_{0(2)})} \right). \end{aligned}$$

Substituindo as expressões de $f_0(t_i; \mathbf{v}_{(2)})$ e $S_0(t_i; \mathbf{v}_{(2)})$ obtemos:

$$\begin{aligned}
l(\boldsymbol{\vartheta}_{(2)}; \mathbf{D}) &= n \ln(p_{0(2)}) + \sum_{i=1}^n \ln(1 - [1 - \exp\{1 - (1 + \psi t_i^v)^\alpha\}]^\eta (1 - p_{0(2)})) \\
&+ \sum_{i=1}^n \delta_i \ln \left(\alpha \eta \psi v t^{v-1} \frac{(1 + \psi t^v)^{\alpha-1} \exp\{1 - (1 + \psi t^v)^\alpha\}}{[1 - \exp\{1 - (1 + \psi t^v)^\alpha\}]^{1-\eta}} (1 - p_{0(2)}) \right) \\
&+ \sum_{i=1}^n \delta_i \ln(1 - [1 - \exp\{1 - (1 + \psi t_i^v)^\alpha\}]^\eta (1 - p_{0(2)})).
\end{aligned}$$

Podemos obter as derivadas de $\log(L(\boldsymbol{\vartheta}_{(2)}; \mathbf{D}))$ em relação a cada um dos parâmetros e igualamos a zero para encontrar os valores que maximizam o logaritmo da função de verossimilhança. Temos que a obtenção dos estimadores dos parâmetros $p_{0(2)}$, α , η , ψ e v pelas soluções das seguintes equações:

$$\left. \frac{\partial l(\boldsymbol{\vartheta}_{(2)}; \mathbf{D})}{\partial p_{0(2)}} \right|_{p_{0(2)} = \hat{p}_{0(2)}} = 0,$$

$$\left. \frac{\partial l(\boldsymbol{\vartheta}_{(2)}; \mathbf{D})}{\partial \alpha} \right|_{\alpha = \hat{\alpha}} = 0,$$

$$\left. \frac{\partial l(\boldsymbol{\vartheta}_{(2)}; \mathbf{D})}{\partial \eta} \right|_{\eta = \hat{\eta}} = 0,$$

$$\left. \frac{\partial l(\boldsymbol{\vartheta}, \mathbf{D})}{\partial \psi} \right|_{\psi = \hat{\psi}} = 0$$

e,

$$\left. \frac{\partial l(\boldsymbol{\vartheta}_{(2)}; \mathbf{D})}{\partial v} \right|_{v = \hat{v}} = 0.$$

Note que as equações acima não possuem forma fechada. Assim, as estimativas dos parâmetros foi obtida numericamente com o auxílio da função *optim()* pelo *software R Development Core Team (2017a)* utilizando o método *L-BFGS-B*.

A função score do modelo é dada da forma $U(\boldsymbol{\vartheta}_{(2)}) = \partial \log L(\boldsymbol{\vartheta}_{(2)}; \mathbf{D}) / (\partial \boldsymbol{\vartheta}_{(2)})$, já a matriz observada é dada por $\mathbf{I} = -\partial U^\top / (\partial \boldsymbol{\vartheta}_{(2)}) = -\partial^2 \log L(\boldsymbol{\vartheta}_{(2)}; \mathbf{D}) / (\partial \boldsymbol{\vartheta}_{(2)} \partial \boldsymbol{\vartheta}_{(2)}^\top)$.

Usualmente, em modelos com proporção de indivíduos curados, as covariáveis colocadas no modelo são incluídas pelo parâmetro da proporção de cura. Além da inclusão de covariáveis no parâmetro da proporção de cura, também fizemos a inclusão de covariáveis no parâmetro ψ .

Para a inclusão das covariáveis no parâmetro da proporção de cura do modelo GEPGWcr utilizamos a função de ligação logística, da seguinte forma:

$$\log\left(\frac{p_{0(2)i}}{1 - p_{0(2)i}}\right) = \mathbf{x}_i^T \boldsymbol{\beta} \Rightarrow p_{0(2)i} = \frac{e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}}.$$

Já para a inclusão de covariáveis no parâmetro ψ , utilizamos a função de ligação exponencial, apresentada a seguir:

$$\log(\psi) = \mathbf{x}_i^T \boldsymbol{\beta} \Rightarrow e^{\mathbf{x}_i^T \boldsymbol{\beta}} = \psi.$$

4.2 Estudo de Simulação e Aplicação a Dados reais

Realizamos um estudo de simulação com o objetivo de avaliar as estimativas dos parâmetros do modelo GEPGWcr. Consideramos dados artificiais com a inclusão de covariáveis.

Analogamente ao estudo feito para o modelo LIGcr, calculamos a média das estimativas dos parâmetros, a média do vício, o EQM e a PC com valor nominal de 95%.

Dados Artificiais

Para o cálculo da probabilidade de cobertura consideramos a normalidade assintótica (distribuição Normal(0, 1)) com 95% de confiança. Simulamos a covariável com uma distribuição Binomial (1, 0, 5) sendo n o tamanho amostral para cada cenário. Além disso, utilizamos a função de ligação logística no parâmetro da proporção de cura, $\log(p_{02}) = \exp(\beta_0 + \beta_1 x_i)$ e a função de ligação exponencial no parâmetro ψ do modelo GEPGWcr. Em todo o estudo, foi utilizado o *software* R Development Core Team (2017a).

Simulação

Neste trabalho realizamos um estudo de simulação com 1.000 réplicas considerando quatro tamanhos amostrais $n = 50, 150, 300$ e 500 . Apresentamos três cenários considerando a inclusão de uma covariável binária no parâmetro da proporção de cura utilizando a função de ligação logística. Os resultados são apresentados na [Tabela 19](#), [Tabela 20](#) e [Tabela 21](#).

Na [Tabela 19](#) consideramos uma proporção de censura próxima de 48%. Para os indivíduos que pertencem ao grupo $X = 0$ temos que a proporção de cura é cerca de 27%. Já para os indivíduos que pertencem ao grupo $X = 1$ temos que a proporção de cura é cerca de 62%. Podemos observar que conforme o tamanho da amostra aumenta, o EQM reduz como podemos observar principalmente para o parâmetro ψ . Além da redução dos EQM's dos parâmetros, podemos observar também que a PC fica mais próxima do valor nominal de 95% conforme o tamanho da amostra aumenta.

Já para a [Tabela 20](#) temos uma proporção de censura próxima de 50%. A proporção de cura para indivíduos que pertencem ao grupo $X = 0$ é aproximadamente 27% e para os indivíduos que pertencem ao grupo $X = 1$ é aproximadamente 38% o que também acaba influenciando uma baixa proporção de censuras. Observamos que o EQM dos parâmetros de uma forma geral acabaram aumentando quando comparado com os resultados da [Tabela 19](#).

A [Tabela 21](#) nos apresenta um cenário com uma proporção de censura próxima de 63%. Neste cenário os indivíduos que fazem parte do grupo $X = 0$ tem uma proporção de cura próximo de 57% enquanto que os indivíduos que fazem parte do grupo $X = 1$ tem uma proporção de cura próxima de 62%. Podemos observar que apesar da PC não estar tão próximo do valor nominal, conforme aumentamos o tamanho da amostra, a PC se aproxima de 0,95. Também, os EQM's dos parâmetros em todos os tamanhos de amostra foram baixos.

Também estudamos a inclusão de uma covariável binária no parâmetro ψ conjuntamente com o parâmetro $p_{0(2)}$, utilizando a função de ligação exponencial e os parâmetros relacionados a covariável incluída são: β_3 e β_4 . Consideramos a mesma covariável binária no parâmetro da proporção de cura, os parâmetros relacionados a covariável incluída em $p_{0(2)}$ são: β_0 e β_1 . Apresentamos os resultados na [Tabela 22](#), [Tabela 23](#) e [Tabela 24](#).

Na [Tabela 22](#) utilizamos os valores: $\beta_0 = -1,5$, $\beta_1 = 0,5$, $\alpha = 1,5$, $\eta = 1$, $\beta_3 = -1$, $\beta_4 = 0,5$ e $v = 1$. Para os indivíduos pertencentes ao grupo $X = 0$ a proporção de cura é 18%. Já para os indivíduos que pertencem ao grupo $X = 1$ a proporção de cura é 27% aproximadamente e a proporção de dados censurados de uma forma geral é próxima de 33%. Observamos que a PC apresenta resultados mais próximos do valor nominal de 95%.

Na [Tabela 23](#) utilizamos $\beta_0 = -1,0$, $\beta_1 = 0,5$, $\alpha = 1,5$, $\eta = 0,5$, $\beta_3 = -1,5$, $\beta_4 = 1$ e $v = 1$. A proporção de cura para os indivíduos pertencentes ao grupo $X = 0$ é 27% aproximadamente. E, para os indivíduos do grupo $X = 1$, a proporção de cura é de aproximadamente 38%. Também podemos observar que, de uma forma geral, o EQM dos parâmetros apresentados na [Tabela 23](#) é menor do que o EQM dos parâmetros apresentados na [Tabela 22](#). Em contrapartida, os valores das PC ficam mais próximos do valor nominal 95% no estudo apresentado pela [Tabela 22](#).

Já para a [Tabela 24](#) os valores dos parâmetros foram $\beta_0 = 0,2$, $\beta_1 = 0,5$, $\alpha = 1,4$, $\eta = 0,5$, $\beta_3 = -2$, $\beta_4 = 0,5$ e $v = 1,2$. Podemos observar que a proporção de dados censurados é aproximadamente 65%. Os indivíduos pertencentes ao grupo $X = 0$ tem uma proporção de curados de 55% aproximadamente, já os indivíduos pertencentes ao grupo $X = 1$ tem uma proporção de aproximadamente 67% de curados. Como resultado, temos que a PC se aproxima do valor nominal de 95% conforme o tamanho amostral aumenta. E, os valores dos EQM's vão reduzindo conforme n aumenta.

Tabela 19 – Simulação do modelo GEPGWcr considerando $\beta_0 = -1$, $\beta_1 = 1,5$, $\alpha = 1,5$, $\eta = 1$, $\psi = 1$ e $\nu = 1$ para diferentes valores amostrais.

Tamanho amostral (n)	Parâmetros	Sumário dos parâmetros				Censura
		Média	Vício	EQM	P. C.	
50	β_0	-1,1341	-0,1341	0,4286	0,9640	47,70%
	β_1	1,5430	0,0430	0,3482	0,9510	
	α	2,0489	0,0489	1,1577	0,9120	
	η	2,3332	1,8332	28,4773	0,8480	
	ψ	2,5662	1,5662	105,0700	0,8560	
	ν	0,8934	-0,0066	0,5617	0,8520	
150	β_0	-1,0360	-0,0360	0,0800	0,9610	47,90%
	β_1	1,5187	0,0187	0,1132	0,9620	
	α	2,2012	0,2012	0,7824	0,8890	
	η	0,9889	0,4889	3,1499	0,8030	
	ψ	1,8077	0,8077	14,2452	0,8790	
	ν	0,9999	0,0999	0,3863	0,8360	
300	β_0	-1,0213	-0,0213	0,0331	0,9640	47,74%
	β_1	1,5023	0,0023	0,0497	0,9560	
	α	2,2667	0,2667	0,6530	0,9030	
	η	0,7055	0,2055	0,5267	0,8120	
	ψ	1,4200	0,4200	5,8907	0,8850	
	ν	0,9867	0,0867	0,2526	0,8440	
500	β_0	-1,0112	-0,0112	0,0190	0,9480	47,80%
	β_1	1,5101	0,0101	0,0326	0,9490	
	α	2,2876	0,2876	0,5938	0,9120	
	η	0,5597	0,0597	0,1032	0,8450	
	ψ	1,2344	0,2344	2,2333	0,9150	
	ν	0,9936	0,0936	0,1662	0,8800	

Tabela 20 – Simulação do modelo GEPGWcr considerando $\beta_0 = -1$, $\beta_1 = 0,5$, $\alpha = 1,5$, $\eta = 1,2$, $\psi = 1$ e $\nu = 1,8$ para diferentes valores amostrais.

Tamanho amostral (n)	Parâmetros	Sumário dos parâmetros				Censura
		Média	Vício	EQM	P. C.	
50	β_0	-1,2108	-0,2108	0,9363	0,9750	49,28%
	β_1	0,5327	0,0327	0,3545	0,9440	
	α	1,8293	0,3293	2,1016	0,9090	
	η	5,1960	3,9960	541,1244	0,9460	
	ψ	7,0350	6,0350	589,6213	0,9710	
	ν	2,2655	0,4655	2,7713	0,9770	
150	β_0	-1,0747	-0,0747	0,1270	0,9680	49,09%
	β_1	0,5147	0,0147	0,0983	0,9590	
	α	1,9743	0,4743	1,7660	0,8870	
	η	1,4442	0,2442	3,4261	0,9370	
	ψ	2,6225	1,6225	20,4186	0,9780	
	ν	2,5677	0,7677	2,4887	0,9770	
300	β_0	-1,0402	-0,0402	0,0425	0,9680	49,23%
	β_1	0,5211	0,0211	0,0518	0,9520	
	α	1,9180	0,4180	1,5499	0,8650	
	η	1,1488	-0,0512	0,9017	0,9230	
	ψ	2,1249	1,1249	12,0732	0,9800	
	ν	2,5789	0,7789	2,2500	0,9770	
500	β_0	-1,0163	-0,0163	0,0261	0,9550	49,25%
	β_1	0,5030	0,0030	0,0287	0,9590	
	α	1,7618	0,2618	1,3123	0,8350	
	η	1,0987	-0,1013	0,4243	0,8890	
	ψ	2,0119	1,0119	8,3931	0,9710	
	ν	2,4839	0,6839	1,9248	0,9680	

Tabela 21 – Simulação do modelo GEPGWcr considerando $\beta_0 = 0,3$, $\beta_1 = 0,2$, $\alpha = 2,5$, $\eta = 0,2$, $\psi = 0,8$ e $\nu = 2$ para diferentes valores amostrais.

Tamanho amostral (n)	Parâmetros	Sumário dos parâmetros				Censura
		Média	Vício	EQM	P. C.	
50	β_0	0,2780	-0,0220	0,1761	0,9620	62,08%
	β_1	0,1910	-0,0090	0,3401	0,9540	
	α	2,4899	-0,0101	0,6089	0,8970	
	η	1,2859	1,0859	9,0643	0,9100	
	ψ	1,3353	0,5353	4,9063	0,6990	
	ν	1,5277	-0,4723	1,1352	0,6700	
150	β_0	0,3066	0,0066	0,0564	0,9660	62,81%
	β_1	0,1746	-0,0254	0,1134	0,9430	
	α	2,6067	0,1067	0,3907	0,8990	
	η	0,4300	0,2300	0,7559	0,9490	
	ψ	1,0388	0,2388	2,0463	0,7300	
	ν	1,7038	-0,2962	0,4752	0,7080	
300	β_0	0,3030	0,0030	0,0290	0,9470	63,01%
	β_1	0,1927	-0,0073	0,0513	0,9600	
	α	2,5989	0,0989	0,2163	0,9110	
	η	0,2934	0,0934	0,0932	0,9590	
	ψ	0,9112	0,1112	0,6489	0,7430	
	ν	1,7612	-0,2388	0,3016	0,7240	
500	β_0	0,3016	0,0016	0,0161	0,9550	63,07%
	β_1	0,2019	0,0019	0,0293	0,9550	
	α	2,5994	0,0994	0,1795	0,9370	
	η	0,2598	0,0598	0,0172	0,9530	
	ψ	0,8447	0,0447	0,1485	0,7400	
	ν	1,7651	-0,2349	0,2215	0,7530	

Tabela 22 – Simulação do modelo GEPGWcr considerando $\beta_0 = -1,5$, $\beta_1 = 0,5$, $\alpha = 1,5$, $\eta = 1$, $\beta_3 = -1$, $\beta_4 = 0,5$ e $v = 1$ para diferentes valores amostrais.

Tamanho amostral (n)	Parâmetros	Sumário dos parâmetros				Censura
		Média	Vício	EQM	P. C.	
50	β_0	-5,5293	-4,0293	48,2596	0,8350	32,02%
	β_1	2,2821	1,7821	42,0111	0,8520	
	α	1,5663	0,0663	3,0575	0,7070	
	η	1,7996	0,7996	4,6304	0,8380	
	β_3	-3,8068	-2,8068	47,3273	0,8050	
	β_4	2,5336	2,0336	45,0037	0,8230	
	v	1,3112	0,3112	0,8969	0,8570	
150	β_0	-2,2984	-0,7984	7,8899	0,8710	33,36%
	β_1	1,1082	0,6082	7,8245	0,9150	
	α	1,9458	0,4458	1,4827	0,8630	
	η	1,2022	0,2022	1,3218	0,8470	
	β_3	-2,0752	-1,0752	13,9685	0,8750	
	β_4	1,4103	0,9103	12,7689	0,9120	
	v	1,4276	0,4276	0,9395	0,8810	
300	β_0	-1,7231	-0,2231	2,0223	0,9150	33,15%
	β_1	0,6793	0,1793	1,9456	0,9370	
	α	1,9248	0,4248	1,1492	0,8990	
	η	1,1043	0,1043	0,7656	0,8740	
	β_3	-1,4963	-0,4963	3,7275	0,9000	
	β_4	0,8249	0,3249	2,6765	0,9390	
	v	1,3672	0,3672	0,8091	0,8990	
500	β_0	-1,5922	-0,0922	0,8059	0,9230	32,95%
	β_1	0,5783	0,0783	0,8483	0,9520	
	α	1,9570	0,4570	1,0429	0,9330	
	η	1,0234	0,0234	0,2895	0,9170	
	β_3	-1,4078	-0,4078	1,7654	0,9530	
	β_4	0,6965	0,1965	1,3055	0,9470	
	v	1,2491	0,2491	0,4697	0,9380	

Tabela 23 – Simulação do modelo GEPGWcr considerando $\beta_0 = -1,0$, $\beta_1 = 0,5$, $\alpha = 1,5$, $\eta = 1,2$, $\beta_3 = -1,5$, $\beta_4 = 1$ e $v = 1$ para diferentes valores amostrais.

Tamanho amostral (n)	Parâmetros	Sumário dos parâmetros				Censura
		Média	Vício	EQM	P. C.	
50	β_0	-3,1998	-2,1998	36,7621	0,9000	40,04%
	β_1	2,6385	2,1385	37,1890	0,9090	
	α	2,1399	0,6399	4,4101	0,8320	
	η	1,7583	0,5583	2,8617	0,9160	
	β_3	-2,9194	-1,4194	25,4428	0,9200	
	β_4	2,8566	1,8566	21,4794	0,9200	
	v	1,4223	0,4223	1,3707	0,9500	
150	β_0	-1,1175	-0,1175	1,3809	0,9570	39,84%
	β_1	0,5987	0,0987	1,4543	0,9510	
	α	2,0004	0,5004	2,0521	0,8230	
	η	1,1088	-0,0912	0,5574	0,8930	
	β_3	-2,1799	-0,6799	3,0876	0,9600	
	β_4	1,5435	0,5435	2,3356	0,9660	
	v	1,5144	0,5144	1,1549	0,9450	
300	β_0	-1,0473	-0,0473	0,1770	0,9570	39,59%
	β_1	0,5273	0,0273	0,2179	0,9480	
	α	1,8762	0,3762	1,6382	0,8100	
	η	1,0285	-0,1715	0,3295	0,8720	
	β_3	-2,1608	-0,6608	1,3230	0,9590	
	β_4	1,4857	0,4857	1,0879	0,9590	
	v	1,4841	0,4841	0,9156	0,9530	
500	β_0	-1,0101	-0,0101	0,0403	0,9460	39,78%
	β_1	0,5029	0,0029	0,0618	0,9500	
	α	1,7589	0,2589	1,4132	0,7950	
	η	1,0180	-0,1820	0,2601	0,8610	
	β_3	-2,0849	-0,5849	1,0602	0,9590	
	β_4	1,4771	0,4771	1,0758	0,9590	
	v	1,4619	0,4619	0,8507	0,9570	

Tabela 24 – Simulação do modelo GEPGWcr considerando $\beta_0 = 0,2$, $\beta_1 = 0,5$, $\alpha = 1,4$, $\eta = 0,5$, $\beta_3 = -2$, $\beta_4 = 0,5$ e $v = 1,2$ para diferentes valores amostrais.

Tamanho amostral (n)	Parâmetros	Sumário dos parâmetros				Censura
		Média	Vício	EQM	P. C.	
50	β_0	-3,5843	-3,7843	48,9692	0,8790	63,68%
	β_1	2,8043	2,3043	42,9488	0,8610	
	α	2,2620	0,8620	8,1395	0,6620	
	η	1,4064	0,9064	4,1047	0,8310	
	β_3	-5,0337	-3,0337	63,4696	0,7630	
	β_4	2,9243	2,4243	60,7093	0,8080	
	v	1,2454	0,0454	0,9404	0,8120	
150	β_0	-0,1997	-0,3997	3,3256	0,9480	65,79%
	β_1	0,7719	0,2719	2,9424	0,9460	
	α	1,9592	0,5592	2,7940	0,7750	
	η	0,8406	0,3406	1,1182	0,8230	
	β_3	-2,3067	-0,3067	8,3577	0,8330	
	β_4	0,9249	0,4249	5,7254	0,9470	
	v	1,4279	0,2279	0,7830	0,8260	
300	β_0	0,0807	-0,1193	0,5743	0,9640	66,13%
	β_1	0,5856	0,0856	0,5466	0,9650	
	α	2,0761	0,6761	1,8374	0,8630	
	η	0,6750	0,1750	0,3264	0,8280	
	β_3	-2,4480	-0,4480	3,5645	0,8430	
	β_4	0,7137	0,2137	1,8322	0,9460	
	v	1,3728	0,1728	0,5381	0,8430	
500	β_0	0,1716	-0,0284	0,0487	0,9550	66,18%
	β_1	0,5056	0,0056	0,0580	0,9510	
	α	2,0389	0,6389	1,5643	0,8540	
	η	0,6073	0,1073	0,1943	0,8160	
	β_3	-2,4239	-0,4239	1,4988	0,8480	
	β_4	0,5806	0,0806	0,3907	0,9360	
	v	1,3719	0,1719	0,4712	0,8400	

Aplicação a Dados Reais

A aplicação a dados reais foi dividida em duas etapas: na primeira etapa apresentamos o ajuste do modelo GEPGWcr sem covariáveis para o conjunto de dados apresentado por [Ghitany M. E.; Toomet \(2002\)](#). Já na segunda etapa apresentamos o ajuste do modelo GEPGWcr com covariável no parâmetro da proporção de cura para o conjunto de dados LARYNX apresentado por [Klein e Moeschberger \(2003\)](#) e encontrado no pacote ‘KMsurv’ do [R Development Core Team \(2017a\)](#).

Sem considerar covariáveis

Nesta seção apresentamos o ajuste do modelo de sobrevivência GEPGWcr sem covariáveis para o conjunto de dados apresentado por [Ghitany M. E.; Toomet \(2002\)](#). O conjunto de dados é a sobrevida em meses de pacientes com leucemia que fizeram o transplante alogênico¹. Nesta base de dados de 50 pacientes, 22 pacientes foram a óbito depois do transplante. A [Tabela 25](#) mostra as medidas descritivas de tempo dos pacientes que receberam o transplante. Observamos que a média de meses de vida dos pacientes é de 18,74 meses, mais do que 1,5 ano. Observamos também que 75% dos pacientes vivem pelo menos 31,88 meses que é mais de 2,5 anos. A [Tabela 26](#) mostra de uma forma geral a sobrevida dos pacientes. Observamos que nos primeiros 6 meses quase 70% dos pacientes ainda estavam vivos e não foram censurados; depois de 24 meses cerca de 50% dos pacientes transplantados estão vivos e não foram censurados.

Tabela 25 – Medidas descritivas do tempo de sobrevida dos pacientes com leucemia após o transplante.

Min.	1º	Mediana	Média	3º	Max.
0.03	3,61	12,45	18,74	31,88	60,62

Tabela 26 – Porcentagem geral de sobrevida dos pacientes.

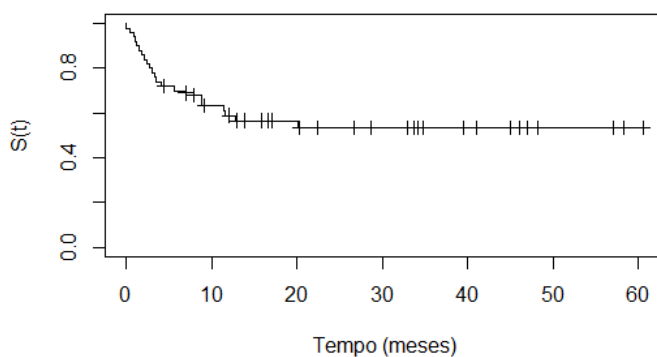
Meses	%
6	69,94
12	58,97
18	56,61
24	53,63

Para uma análise do comportamento dos indivíduos ao decorrer do tempo plotamos a curva de sobrevivência estimada pelo KM apresentado na [Figura 17](#). Além disso, para um maior conhecimento dos dados, utilizamos o gráfico TTT apresentado por [Brunk et al. \(1972\)](#). Utilizamos a metodologia apresentada por [Sun e Kececiloglu \(1999\)](#) que permite a incorporação de dados censurados nos seus cálculos. Mais detalhes sobre o gráfico TTT estão apresentados no [Apêndice B](#).

Na [Figura 18](#) observamos que curvatura do gráfico TTT é côncava assim, temos que a função de risco dos indivíduos é decrescente. Como a distribuição basal EPGW acomoda diferentes comportamento da função de risco, esperamos que o modelo GEPGW apresente um bom ajuste para os dados descritos acima.

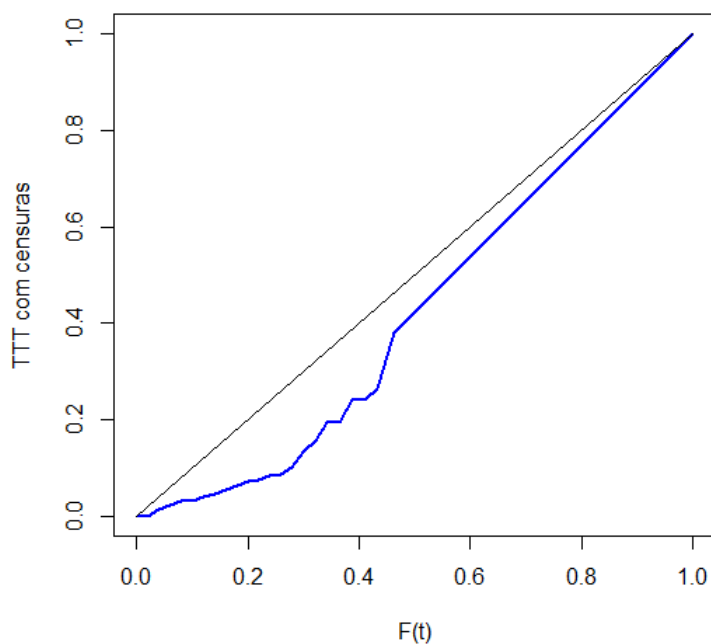
¹ As células precursoras da medula que o paciente irá receber provêm de outro indivíduo (doador) que não seja o próprio paciente. O transplante ocorre de acordo com o nível de compatibilidade do material sanguíneo. Este tipo de transplante também pode ser feito a partir de células precursoras de medula óssea obtidas do sangue de um cordão umbilical

Figura 17 – Gráfico com a curva de sobrevivência de KM.



Fonte: Elaborada pela autora.

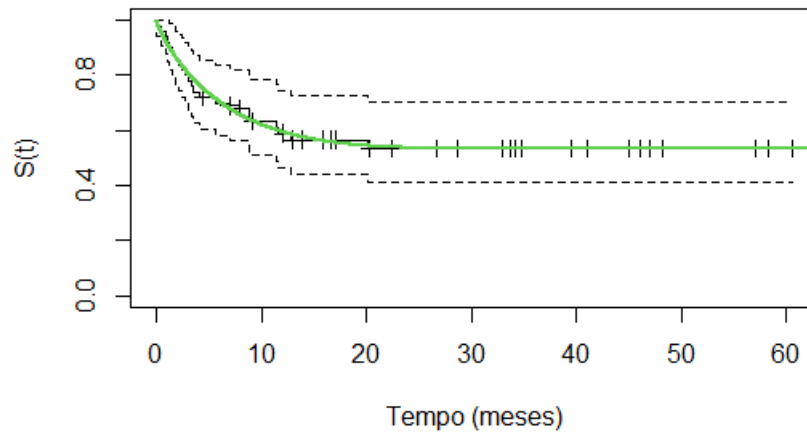
Figura 18 – Gráfico com o Tempo Total em Teste considerando os indivíduos censurados.



Fonte: Elaborada pela autora.

Ajustamos o modelo GEPGWcr aos dados reais apresentados de [Ghitany M. E.; Toomet \(2002\)](#). Na [Figura 19](#) podemos perceber que a curva do modelo ajustado captou bem a curva do estimador de KM. Além disso, também apresentamos na [Tabela 27](#) o valor do EMV de cada parâmetro juntamente com os DP's informados entre parênteses. Por meio das estimativas dos parâmetros, obtemos uma proporção de cura próxima de 53%.

Figura 19 – Gráfico com a curva de sobrevivência de KM juntamente com a curva de sobrevivência do modelo GEPGWcr ajustado (**verde**).



Fonte: Elaborada pela autora.

Tabela 27 – Estimativas de máxima verossimilhança e (desvio padrão) dos parâmetros do modelo GEPGWcr.

	Parametros				
	ω	α	η	ψ	ν
Estimativas	0,5327 (0,0742)	1,7500 (5,4814)	0,6363 (0,8745)	0,0159 (0,0511)	1,4168 (1,8666)

4.2.0.1 Ajuste com covariáveis

Nesta segunda etapa ajustamos o modelo GEPGWcr com a inclusão de covariáveis para o conjunto de dados LARYNX apresentado por [Klein e Moeschberger \(2003\)](#). Esse conjunto de dados pode ser encontrado no [R Development Core Team \(2017a\)](#) no pacote ‘KMsurv’.

Análise Descritiva

A base de dados é composta por 90 observações e 5 variáveis descritas a seguir:

- *Time*: indicando o tempo do paciente em meses;
- *Delta*: indicando a falha (1) se o paciente morreu do cancer de laringe ou a censura(0) se o paciente foi censurado;
- *Stage*: o estágio do câncer de laringe;

- *Age*: a idade do paciente;
- *Diagyr*: o ano que o câncer de laringe foi diagnosticado.

No ajuste da base de dados consideramos a variável idade para a inclusão de covariáveis na proporção de cura. Para isso, dicotomizamos a covariável em: (0) menor que 70 anos e (1) maior ou igual a 70 anos. Pela [Tabela 28](#), podemos observar que a base de dados tem até 10,70 meses, pouco menos de um ano. Além disso, podemos observar que pelo menos 50% da amostra tem idade até 65 anos e que os pacientes com câncer de laringe têm idade entre 41 e 86 anos.

Tabela 28 – Medidas resumo das variáveis tempo, idade e ano do diagnóstico

Variáveis	Medidas Resumo					
	Min	25%	50%	Média	75%	Máx.
tempo	0,10	2,00	4,00	4,20	6,20	10,70
idade	41	57	65	64,61	72	86
<i>diagyr</i>	1970	1973	1974	-	1976	1978

Observamos na [Tabela 29](#) que 65,55% dos pacientes tem menos do que 70 anos. Além disso, mais de 50% dos pacientes estão no estágio 1 ou 2 do câncer de laringe e 14% estão no estágio 4 da doença. Pela [Tabela 30](#) cerca de 30% dos indivíduos estão no estágio 1 ou 2 e são censurados e cerca de 30% dos indivíduos que vieram a óbito estão no estágio 3 ou 4. Além disso, 44,44% dos pacientes são censurados. Pela [Tabela 31](#), 24,44% dos pacientes atingiram a falha e tem idade igual ou superior a 70 anos contra 34,44% dos indivíduos que não atingiram a falha e a idade é inferior a 70 anos.

Tabela 29 – Porcentagem dos indivíduos considerando a covariável idade menor que 70 anos(0) e maior/igual a 70 anos(1) em relação ao estágio

		Idade		Total
		0	1	
Estágio	1	22 (24,44%)	11 (12,22%)	33 (36,66%)
	2	11 (12,22%)	6 (6,67%)	17 (18,89%)
	3	19 (21,11%)	8 (8,89%)	27 (30,00%)
	4	7 (7,78%)	6 (6,67%)	13 (14,45%)
Total		59 (65,55%)	31 (34,45%)	90 (100,00%)

Tabela 30 – Porcentagem dos indivíduos censurados e não censurados considerando o estágio do câncer de laringe

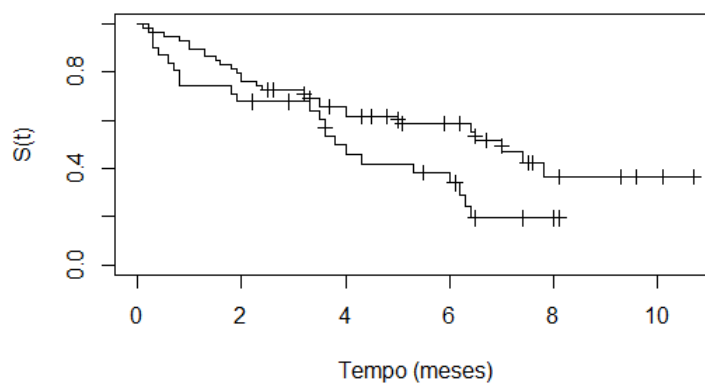
		Censura		Total
		0	1	
Estágio	1	18 (20,00%)	15 (16,67%)	33 (36,67%)
	2	10 (11,11%)	7 (7,78%)	17 (18,89%)
	3	10 (11,11%)	17 (18,89%)	27 (30,00%)
	4	2 (2,22%)	11 (12,22%)	13 (14,44%)
Total		40 (44,44%)	50 (55,56%)	90 (100,00%)

Tabela 31 – Porcentagem dos indivíduos considerando idade menor que 70 anos (0) e idade maior/igual que 70 anos (1) e a censura

		Idade		Total
		0	1	
Censura	0	31 (34,44%)	9 (10,00%)	40 (44,44%)
	1	28 (31,12%)	22 (24,44%)	50 (35,56%)
Total		69 (65,56%)	31 (34,44%)	90 (100,00%)

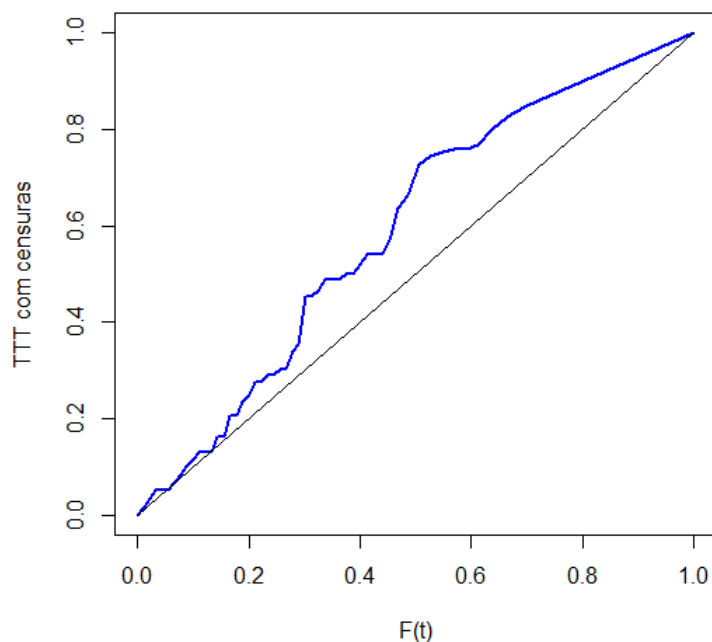
Pela [Figura 20](#) podemos observar que os pacientes com idade inferior a 70 anos tem uma sobrevida maior do que os pacientes com idade igual ou superior a 70 anos. Além disso, a proporção de cura fica próximo dos 35% e 20%, respectivamente. Para uma análise mais detalhada também utilizamos o gráfico TTT com censura apresentado na [Figura 21](#). Como a função de risco tem curvatura côncava temos que a função de risco é crescente.

Figura 20 – Curva de sobrevivência de KM dicotomizada pela idade.



Fonte: Elaborada pela autora.

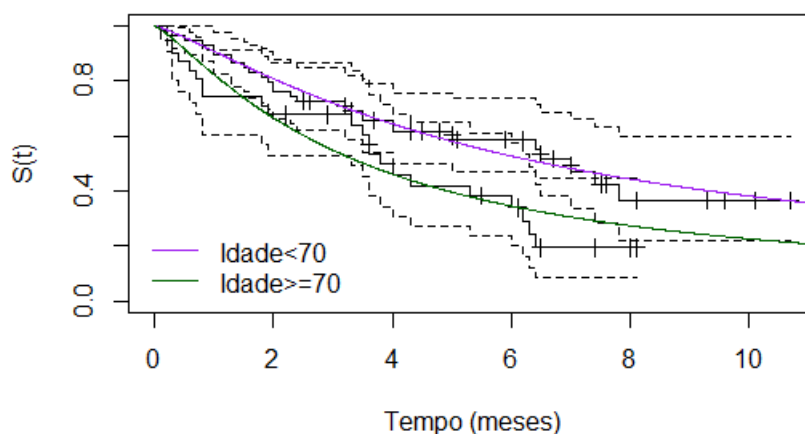
Figura 21 – Gráfico TTT.



Fonte: Elaborada pela autora.

Ajustamos o modelo GEPGWcr aos dados reais descritos acima com a inclusão da covariável idade dicotomizada em: menor que 70 anos e maior ou igual a 70 anos. Pela [Figura 22](#) podemos perceber que a curva do modelo ajustado captou bem a curva de sobrevivência de KM para os indivíduos com idade menor que 70 anos. Além disso, apresentamos na [Tabela 32](#) o valor das estimativas de máxima verossimilhança de cada parâmetro juntamente com os DP's informados entre parênteses. Com as estimativas dos parâmetros, obtemos a proporção de cura é de 13% (para menor que 70 anos) e de 6% (para maior que 70 anos).

Figura 22 – Curva de sobrevivência de KM juntamente com a curva de sobrevivência do modelo GEPGWcr ajustado considerando a covariável idade dicotomizada para o parâmetro $p_{0(2)}$. O modelo ajustado para os indivíduos com idade inferior a 70 anos está na cor **roxa** e o modelo ajustado para os indivíduos com idade igual ou maior que 70 está na cor **verde-escuro**.



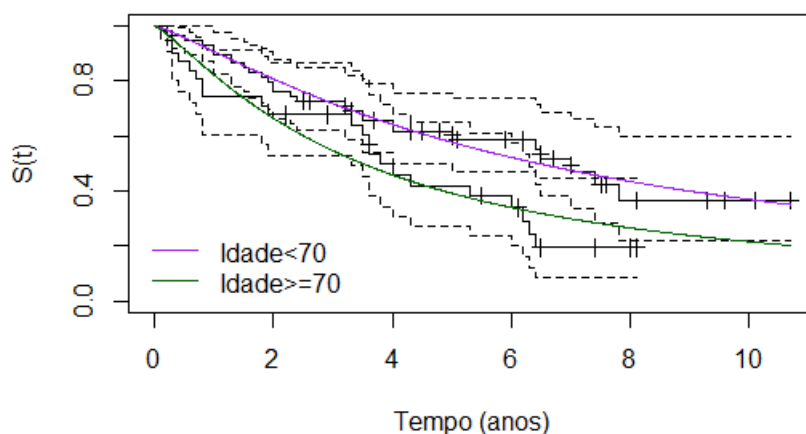
Fonte: Elaborada pela autora.

Tabela 32 – Estimativas de máxima verossimilhanças e (desvios padrão) dos parâmetros do modelo GEPGWcr.

	Parametros					
	β_0	β_2	α	η	ψ	ν
Estimativas	-1,8636 (2,1612)	-0,7490 (0,4050)	1,5369 (2,6756)	0,9619 (0,7157)	0,0087 (0,0094)	1,2891 (0,9585)

Em um outro estudo, relacionamos a covariável idade no parâmetro ψ . A [Figura 23](#) apresenta a curva de sobrevivência de KM juntamente com a curva de sobrevivência do modelo GEPGWcr ajustado. Além disso, a [Tabela 33](#) apresenta as estimativas dos parâmetros do modelo com a covariável idade dicotomizada relacionada no parâmetros da proporção de cura, $p_{0(2)}$, e no parâmetro ψ . Pela [Tabela 34](#) temos que o modelo com a covariável idade dicotomizada apenas em $p_{0(2)}$ tem AIC e BIC menor do que o modelo com a inclusão da covariável idade dicotomizada em $p_{0(2)}$ e em ψ . Além disso, o DP do primeiro caso é bem menor do que no segundo caso.

Figura 23 – Curva de sobrevivência de KM juntamente com a curva de sobrevivência do modelo GEPGWcr ajustado considerando a inclusão de covariável idade dicotomizada no parâmetro da proporção de cura e no parâmetro ψ . O modelo ajustado para os indivíduos com idade inferior a 70 anos está na cor **roxa** e o modelo ajustado para os indivíduos com idade igual ou maior que 70 está na cor **verde-escuro**.



Fonte: Elaborada pela autora.

Tabela 33 – Estimativas de máxima verossimilhança e (desvios padrão) para os parâmetros do modelo.

	Parametros						
	β_0	β_2	α	η	β_3	β_4	ν
Estimativas	-9,4031 (116,0878)	-6,6078 (119,6877)	0,1852 (3,9703)	1,4048 (5,0121)	-6,6323 (68,8824)	-4,17115 (97,02699)	0,8730 (3,1541)

Tabela 34 – Critérios AIC e BIC para o modelo GEPGWcr para dois cenários: considerando a covariável idade dicotomizada apenas em $p_{0(2)}$ e considerando a covariável idade dicotomizada em $p_{0(2)}$ e ψ .

Modelo	AIC	BIC
Covariável em $p_{0(2)}$	312,0632	327,0620
Covariável em $p_{0(2)}$ e ψ	313,7844	331,2831

CONSIDERAÇÕES FINAIS

Neste trabalho estudamos dois modelos de sobrevivência com fração de cura: os modelos LIGcr e GEPGWcr. Apresentamos a metodologia utilizada para a construção desses modelos. Para os dois modelos estudados, os parâmetros são estimados pelo método de Máxima Verossimilhança.

Em ambos os modelos, realizamos um estudo de simulação com o objetivo de investigar a performance do procedimento de estimação dos parâmetros dos modelos. Apresentamos as médias das estimativas, os vieses, os erros quadráticos médios e as probabilidades de cobertura. Como resultado desses estudo, foi verificado que, com diferentes tamanhos amostrais e diferentes configurações de parâmetros, as estimativas obtidas foram próximas do verdadeiro valor. Também, foi realizado um estudo de simulação do modelo GEPGWcr sem covariáveis.

Por fim, mostramos a aplicabilidade dos modelos a conjuntos de dados reais.

Como continuidade do trabalho, pretendemos fazer um estudo de análise dos resíduos para os modelos ajustados além de flexibilizar os modelos estudados utilizando outras distribuições basal. Por exemplo, a flexível distribuição Kumaraswamy complementary Weibull geometric (Kw–CWG) proposto por [Afify et al. \(2017\)](#), que contém vários submodelos como casos especiais. Além disso, podemos estender estes modelos para o caso multivariado (bivariado) considerando funções cópulas ([NELSEN, 2007](#)).

REFERÊNCIAS

- AALLEN, O. Nonparametric inference for a family of counting processes. **The Annals of Statistics**, JSTOR, p. 701–726, 1978. Citado na página 33.
- AARSET, M. V. The null distribution for a test of constant versus "bathtub" failure rate. **Scandinavian Journal of Statistics**, JSTOR, p. 55–61, 1985. Citado na página 71.
- AFIFY, A. Z.; CORDEIRO, G. M.; BUTT, N. S.; ORTEGA, E. M.; SUZUKI, A. K. A new lifetime model with variable shapes for the hazard rate. **Brazilian Journal of Probability and Statistics**, Brazilian Statistical Association, v. 31, n. 3, p. 516–541, 2017. Citado na página 111.
- BAGDONAVICIUS, V.; NIKULIN, M. **Accelerated life models: modeling and statistical analysis**. [S.l.]: CRC press, 2001. Citado na página 81.
- BARLOW, R. E.; CAMPO, R. A. **Total Time on Test Processes and Applications to Failure Data Analysis**. [S.l.], 1975. Disponível em: <<https://inis.iaea.org/search/searchsinglerecord.aspx?recordsFor=SingleRecord&RN=8291125>>. Acesso em: Nov. 2018. Citado nas páginas 71 e 119.
- BERKSON, J.; GAGE, R. P. Survival curve for cancer patients following treatment. **Journal of the American Statistical Association**, Taylor & Francis, v. 47, n. 259, p. 501–515, 1952. Citado na página 34.
- BOLKER, B.; BOLKER, M. B.; IMPORTS, M. Package 'emdbook'. 2020. Citado na página 117.
- BRUNK, H.; BARLOW, R.; BARTHOLOMEW, D.; BREMNER, J. **Statistical Inference under Order Restrictions. (The Theory and Application of Isotonic Regression)**. [S.l.], 1972. Citado nas páginas 71 e 103.
- CANCHO, V. G.; ZAVALITA, K. E.; MACERA, M. A.; SUZUKI, A. K.; LOUZADA, F. A bayesian cure rate model with dispersion induced by discrete frailty. **Communications for Statistical Applications and Methods**, Korean Statistical Society, v. 25, n. 5, p. 471–488, 2018. Citado nas páginas 26 e 27.
- CARVALHO, M. S.; ANDREOZZI, V. L.; CODEÇO, C. T.; CAMPOS, D. P.; BARBOSA, M. T. S.; SHIMAKURA, S. E. Fragilidade. In: _____. **Análise de Sobrevivência - Teoria e aplicações em saúde**. 2. ed. [S.l.]: Fiocruz, 2011. cap. 12, p. 363–395. Citado nas páginas 26 e 38.
- COLOSIMO, E. A.; GIOLO, S. R. **Análise de Sobrevivência Aplicada**. 1. ed. [S.l.]: EDGARD BLÜCHER, 2006. 367 p. Citado nas páginas 25, 29, 30, 31 e 33.
- CORLESS, R. M.; GONNET, G. H.; HARE, D. E.; JEFFREY, D. J.; KNUTH, D. E. On the lambertw function. **Advances in Computational mathematics**, Springer, v. 5, n. 1, p. 329–359, 1996. Citado nas páginas 47 e 117.

- COX, D. R. Regression models and life-tables. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 34, n. 2, p. 187–202, 1972. Citado na página 33.
- FERNANDES, P. G.; SUZUKI, A. K.; SARAIVA, E. F. O modelo lindley-weibull com proporção de cura: Uma abordagem bayesiana. **Revista Brasileira de Biometria**, v. 36, n. 4, p. 998–1022, 2018. Citado na página 26.
- GHITANY M. E.; TOOMET, O. Maximum likelihood estimation of burr xii distribution parameters under random censoring. **Journal of Applied Statistics**, v. 29, n. 7, p. 955–965, 2002. Citado nas páginas 102, 103 e 104.
- HANAGAL, D. D.; DABADE, A. D. Modeling of inverse gaussian frailty model for bivariate survival data. **Communications in Statistics-Theory and Methods**, Taylor and Francis, v. 42, n. 20, p. 3744–3769, 2013. Citado na página 48.
- IBRAHIM, J. G.; CHEN, M.; SINHA, D. **Bayesian Survival Analysis**. 1. ed. [S.l.]: New York: Springer-Verlag, 2001. 480 páginas. Citado na página 25.
- JOHNSON, N. L.; KEMP, A. W.; KOTZ, S. **Univariate discrete distributions**. 3. ed. [S.l.]: John Wiley and Sons, 2005. 646 páginas. Citado na página 41.
- KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. **Journal of the American statistical association**, Taylor & Francis, v. 53, n. 282, p. 457–481, 1958. Citado na página 33.
- KLEIN, J. P.; MOESCHBERGER, M. L. **Survival analysis: techniques for censored and truncated data**. Springer, v. 1230, 2003. Citado nas páginas 102 e 105.
- LOPES, C. M. C.; BOLFARINE, H. Random effects in promotion time cure rate models. **Computational Statistics and Data Analysis**, v. 56, n. 1, p. 75 – 87, 2012. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0167947311001745>>. Citado na página 26.
- MARTINEZ, E. Z.; ACHCAR, J. A. Bayesian bivariate generalized lindley model for survival data with a cure fraction. **Computer methods and programs in biomedicine**, Elsevier, v. 117, n. 2, p. 145–157, 2014. Citado na página 26.
- NELSEN, R. B. **An introduction to copulas**. [S.l.]: Springer Science & Business Media, 2007. Citado na página 111.
- NELSON, W. Theory and applications of hazard plotting for censored failure data. **Technometrics**, Taylor & Francis, v. 14, n. 4, p. 945–966, 1972. Citado na página 33.
- PENA-RAMIREZ, F. A.; GUERRA, R. R.; CORDEIRO, G. M.; MARINHO, P. R. The exponentiated power generalized weibull: Properties and applications. **Anais da Academia Brasileira de Ciências**, SciELO Brasil, v. 90, n. 3, p. 2553–2577, 2018. Citado nas páginas 81 e 82.
- R Development Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2017. Disponível em: <<http://www.R-project.org>>. Citado nas páginas 57, 58, 68, 94, 95, 102, 105 e 117.
- _____. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2017. Disponível em: <<http://www.R-project.org>>. Citado na página 59.

- RAMOS, P. L.; LOUZADA, F. A new long-term survival distribution. **Biostat Biometrics Open Acc**, v. 1, n. 4, p. 1–6, 2017. Citado na página 26.
- RODRIGUES, J.; CANCHO, V. G.; CASTRO, M. Teoria unificada de análise de sobrevivência. In: _____. [S.l.]: Edgard Blucher, 2008. Citado nas páginas 26, 32, 34, 36, 37, 44, 45 e 79.
- SCHRÖDINGER, E. Zur theorie der fall-und steigversuche an teilchen mit brownscher bewegung. **Physikalische Zeitschrift**, v. 16, p. 289–295, 1915. Citado na página 48.
- SEBER, G. A.; LEE, A. J. **Linear regression analysis**. 2. ed. [S.l.]: John Wiley & Sons, 2012. 557 p. Citado na página 58.
- SOUZA, D. de; CANCHO, V. G.; RODRIGUES, J.; BALAKRISHNAN, N. Bayesian cure rate models induced by frailty in survival analysis. **Statistical methods in medical research**, SAGE Publications Sage UK: London, England, v. 26, n. 5, p. 2011–2028, 2017. Citado nas páginas 26, 27 e 58.
- SUN, F.-B.; KECECLOGLU, D. A new method for obtaining the ttt plot for a censored sample. In: IEEE. **Annual Reliability and Maintainability. Symposium. 1999 Proceedings (Cat. No. 99CH36283)**. [S.l.], 1999. p. 112–117. Citado nas páginas 71, 103 e 119.
- SUZUKI, A. K.; BARRIGA, G. D. C.; LOUZADA, F.; CANCHO, V. G. A general long-term aging model with different underlying activation mechanisms: Modeling, bayesian estimation, and case influence diagnostics. **Communications in Statistics-Theory and Methods**, Taylor and Francis, v. 46, n. 6, p. 3080–3098, 2017. Citado nas páginas 27, 41, 44, 47 e 48.
- SUZUKI, A. K.; CANCHO, V. G.; LOUZADA, F. The poisson inverse gaussian regression model with cure rate: a bayesian approach and its case influence diagnostics. **Statistical Papers**, Springer, v. 57, n. 1, p. 133–159, 2016. Citado nas páginas 26 e 48.
- TWEEDIE, M. C. Inverse statistical variates. **Nature**, Nature Publishing Group, v. 155, n. 3937, p. 453, 1945. Citado na página 48.
- TWEEDIE, M. C. *et al.* Statistical properties of inverse gaussian distributions. i. **The Annals of Mathematical Statistics**, Institute of Mathematical Statistics, v. 28, n. 2, p. 362–377, 1957. Citado na página 48.
- WIENKE, A. **Frailty models in survival analysis**. [S.l.]: CRC press, 2010. 298 p. Citado nas páginas 26 e 38.
- Wolfram Alpha: Computational Intelligence. **Wolfram Alpha:Computational Intelligence**. [S.l.], 2020. Disponível em: <<https://www.wolframalpha.com>>. Citado na página 47.
- ZAVALETA, K. E. C. **Modelos série de potência com excesso de zeros observáveis e latentes**. Tese (Tese de Doutorado) — DEs – Universidade Federal de São Carlos, São Carlos – SP, 2016. Citado nas páginas 26, 27 e 38.

A FUNÇÃO LAMBERT

A função Lambert W é definida como o inverso de valores múltiplos da função $w \mapsto we^w$. Dessa forma, temos que $W(z)e^{W(z)} = z$. Essa função tem muitas aplicações em matemática pura e aplicada. Lambert, em 1758, solucionou a equação trinomial $x = q + x^m$ dando desenvolvimento em séries para x em potência q . Já a forma mais simétrica da função de Lambert W foi proposta por Euler (CORLESS *et al.*, 1996).

Corless *et al.* (1996) apresentam que a função de Lambert W pode ser utilizada em aplicações combinatórias, exponenciação iterada, solução de equações, solução de um problema de combustível de aviação, solução de um problema de cinética enzimática, solução de equações lineares de retardo com coeficiente constante, solução de similaridade para a equação de Richards, raízes assintóticas de trinômios, análise de algoritmos, aplicações pedagógicas, entre várias outras aplicações.

Neste trabalho, a utilização da função Lambert W foi na reescrita da proporção de cura do modelo LIGcr, ou seja, obtemos o parâmetro θ em função de p_0 , facilitando a interpretação do modelo.

O Software (R Development Core Team, 2017a) tem a função Lambert W implementada, que pode ser obtida pela função `LambertW` do pacote *emdbook* (BOLKER; BOLKER; IMPORTS, 2020).

O GRÁFICO TTT

Barlow e Campo (1975) apresentam a descrição do modelo do gráfico TTT. Um método de análise de dados para observações não negativas utilizado para tempos de sobrevivência que permite a análise mesmo para dados incompletos. O método utiliza um gráfico conhecido como gráfico TTT fornecendo informações da função de risco auxiliando na escolha de modelos probabilísticos. Sun e Kececloglu (1999) em seu trabalho apresentam a incorporação das observações censuradas nos cálculos utilizados para a montagem do TTT. Já que para a presença de observações censuradas as falhas ordenadas se tornam incertas refletindo nas estatísticas do TTT, sendo este não podendo ser obtido diretamente (SUN; KECECLOGLU, 1999). O gráfico TTT proposto por Sun e Kececloglu (1999) também é chamado de Número de Ordem Médio, *Mean-Order-Number* (MON).

Para o gráfico TTT, a análise leva em consideração a curva TTT e a linha diagonal que passa pela origem. Com isso, temos a interpretação do comportamento da função de risco. Se o gráfico que apresenta uma curvatura convexa e depois côncava, a função de risco tem forma de banheira, também conhecida como forma U; já quando o gráfico apresenta curvatura côncava e depois convexa, a função de risco tem forma unimodal. Quando a curvatura do gráfico é apenas côncava, a função de risco é monótona e crescente. Quando a curvatura do gráfico é apenas convexa, a função de risco é monótona e decrescente.

Para obtermos o gráfico TTT pela proposta do MON, seguimos as definições de Sun e Kececloglu (1999). Assim, com um total de n observações, tempo vida em que r são falhas observadas e $(n - r)$ observações censuradas. Definindo TF_j como o tempo de falha da j -ésima falha ($j = 1, 2, \dots, r$) e TC_{ij} o tempo da i -ésima observação censurada entre $(j - 1)$ -ésima e j -ésima falha, com $i = 1, 2, \dots, m_j; j = 1, 2, \dots, r + 1$. temos que os tempos de falha ou de censura são arranjados em ordem crescente da seguinte forma (SUN; KECECLOGLU, 1999):

$$\begin{aligned}
0 < TC_{11} \leq TC_{12} \leq \dots \leq TC_{1m_1} < TF_1 \\
< TC_{21} \leq TC_{22} \leq \dots \leq TC_{2m_2} < TF_2 \\
< \dots < TF_r < TC_{r+1,1} \leq TC_{r+1,2} \leq \dots \leq TC_{r+1,m_{r+1}},
\end{aligned}$$

observamos que $TF_0 = 0$ é o início do teste, enquanto que a $(r+1)$ -ésima falha, $TF_{r+1} = T_{FIM}$ é o final do teste.

Para a presença de censuras entre as falhas, o número de ordem de falha se torna incerto e o MON é estimado como:

$$MON_j = MON_{j-1} + I_j, j = 1, 2, \dots, r, \quad (B.1)$$

em que MON_{j-1} o número de ordem médio da $(j-1)$ -ésima falha; MON_j o número de ordem médio da j -ésima falha; I_j o incremento do número de ordem médio da j -ésima falha devido à(s) censura(s) ocorrida(s) entre a $(j-1)$ -ésima e a j -ésima falha.

O incremento do número de ordem médio da j -ésima falha é calculado como:

$$I_j = \frac{(n+1) - MON_{j-1}}{1 + NIBPSS}, \quad (B.2)$$

em que NIBPSS é o número de observações além do atual conjunto de censuras.

Considerando a [Equação B.1](#) e a [Equação B.2](#) temos que: se o primeiro tempo for uma falha então $MON_1 = 1$ e se o segundo tempo for uma falha então $MON_2 = MON_1 + 1 = 2$ e assim sucessivamente sendo valido até ser encontrada uma censura; se o primeiro tempo for uma censura, então $MON_0 = 0$; se não houver observações censuradas entre $(j-1)$ -ésima e a j -ésima falha $I_j = I_{j-1}$.

Determinando o MON de cada falha observada, temos que a função de probabilidade acumulada da falha e a função de sobrevivência podem ser estimadas com as [Equação B.3](#) e [Equação B.4](#) a seguir:

$$F^{MON}(T) = \begin{cases} 0, & 0 \leq T < TF_1 \\ \frac{MON_1}{n}, & TF_1 \leq T < TF_2 \\ \dots & \\ \frac{MON_r}{n}, & TF_r \leq T < T_{FIM} \\ 1, & T \geq T_{FIM} \end{cases} \quad (B.3)$$

e

$$S^{\text{MON}}(T) = \begin{cases} 1, & 0 \leq T < TF_1 \\ 1 - \frac{\text{MON}_1}{n}, & TF_1 \leq T < TF_2 \\ \dots & \\ 1 - \frac{\text{MON}_r}{n}, & TF_r \leq T < TF_{\text{FIM}} \\ 0, & T \geq TF_{\text{FIM}} \end{cases}, \quad (\text{B.4})$$

em que $TF_0 = 0$; $\text{MON}_0 = 0$; $TF_{r+1} = TF_{\text{FIM}}$ uma pseudo falha.

Especificamente para $T = 0, TF_1, TF_2, \dots, TF_r$ e TF_{FIM} temos:

$$\begin{cases} \text{Para } T = 0, & F_0 = 0 \text{ e } R_0 = 1 \\ \text{Para } T = TF_1, & F_1 = \frac{\text{MON}_1}{n} \text{ e } R_1 = 1 - \frac{\text{MON}_1}{n}, \\ \text{Para } T = TF_2, & F_2 = \frac{\text{MON}_2}{n} \text{ e } R_2 = 1 - \frac{\text{MON}_2}{n}, \\ \dots & \\ \text{Para } T = TF_r, & F_r = \frac{\text{MON}_r}{n} \text{ e } R_r = 1 - \frac{\text{MON}_r}{n}, \\ \text{Para } T = TF_{\text{FIM}}, & F_{\text{FIM}} = 1 \text{ e } R_{\text{FIM}} = 0 \end{cases}. \quad (\text{B.5})$$

O TTT até cada falha observada TF_j , $j = 1, 2, \dots, r+1$ é dado pela [Equação B.6](#) a seguir:

$$TTT_j = n \int_0^T F_j R^{\text{MON}}(T) dT \quad (\text{B.6})$$

Temos que a [Equação B.5](#) pode ser escrita de forma mais explícita, como é mostrado pela [Equação B.7](#) a seguir:

$$\begin{cases} \text{Para } T = 0, & TTT_0 = 0 \\ \text{Para } T = TF_1, & TTT_1 = TTT_0 + n * R_0(TF_1 - 0), \\ \text{Para } T = TF_2, & TTT_2 = TTT_1 + n * R_1(TF_2 - TF_1), \\ \dots & \\ \text{Para } T = TF_r, & TTT_r = TTT_{r-1} + n * R_{r-1}(TF_r - TF_{r-1}), \\ \text{Para } T = TF_{\text{FIM}}, & TTT_{\text{FIM}} = TTT_r + n * R_r(TF_{\text{FIM}} - TF_r) \end{cases}. \quad (\text{B.7})$$

Dessa forma, temos a estatística TTT escalada é dada por:

$$\left\{ \begin{array}{ll} ParaT = 0, & \varphi_0 = 0 \\ ParaT = T_{ON_{1:n}}, & \varphi_1 = TTT_1/TTT_{FIM}, \\ ParaT = T_{ON_{2:n}}, & \varphi_2 = TTT_2/TTT_{FIM}, \\ \dots & \\ ParaT = T_{ON_{r:n}}, & \varphi_r = TTT_r/TTT_{FIM}, \\ ParaT = T_{ON_{1:n}}, & \varphi_{FIM} = 1 \end{array} \right. . \quad (B.8)$$

Plotando $(F_0, \varphi_0), (F_1, \varphi_1), \dots, (F_r, \varphi_r)$ e (F_{FIM}, φ_{FIM}) obtidas pela [Equação B.7](#) e [Equação B.8](#) e, conectando eles com os segmentos de linha pelo TTT escalonado para a amostra com observações censuradas temos o gráfico TTT utilizado.

ANEXOS LIG

A.1 Demonstração da **Proposição 1**

Proposição 1 A FGP da distribuição logarítmica começando em zero pode ser calculada da seguinte forma:

$$\begin{aligned}
 G_z(s) &= E[s^z] \\
 &= \sum_{z=0}^{\infty} s^z \left(-\frac{1}{\ln(1-\theta)} \frac{\theta^{z+1}}{(z+1)} \right) \\
 &= \frac{\theta \ln(1-s\theta)}{\ln(1-\theta)} \sum_{z=0}^{\infty} -\frac{(s\theta)^z}{\ln(1-s\theta)(z+1)}.
 \end{aligned}$$

Fazendo a mudança de variável $z = y - 1$, temos que:

$$\begin{aligned}
 G_z(s) &= \frac{\theta \ln(1-s\theta)}{\ln(1-\theta)} \sum_{y=1}^{\infty} -\frac{(s\theta)^{(y-1)}}{\ln(1-s\theta)y} \\
 &= \frac{\ln(1-s\theta)}{\ln(1-\theta)s} \sum_{y=1}^{\infty} -\frac{(s\theta)^y}{\ln(1-s\theta)y} \\
 &= \frac{\ln(1-\theta s)}{s \ln(1-\theta)},
 \end{aligned}$$

para $|s| < 1$.

A.2 FGM da distribuição Logarítmica começando em zero

A FGM da distribuição Logarítmica começando em zero é calculada como:

$$\begin{aligned}
M_z(a) &= \sum_{z=0}^{\infty} e^{zt} \left(-\frac{1}{\ln(1-\theta)} \frac{\theta^{(z+1)}}{(z+1)} \right) \\
&= \frac{\theta \ln(1-e^a \theta)}{\ln(1-\theta)} \sum_{z=0}^{\infty} \left(-\frac{(e^a \theta)^z}{\ln(1-e^a \theta)(z+1)} \right).
\end{aligned}$$

Fazendo a mudança de variável $z = y - 1$, temos que:

$$\begin{aligned}
M_z(a) &= \frac{\theta \ln(1-e^a \theta)}{\ln(1-\theta)} \sum_{y=1}^{\infty} \left(-\frac{(e^a \theta)^{y-1}}{\ln(1-e^a \theta)(y)} \right) \\
&= \frac{\ln(1-e^a \theta)}{\ln(1-\theta)e^a} \sum_{y=1}^{\infty} \left(-\frac{(e^a \theta)^{y-1}}{\ln(1-e^a \theta)(y)} \right) \\
&= \frac{\ln(1-\theta e^a)}{e^a \ln(1-\theta)},
\end{aligned}$$

em que $0 < \theta < 1$.

A.3 Esperança e Variância da distribuição Logarítmica começando em zero

Com a FGM da distribuição é possível encontrar a esperança de Z ($E[Z] = M'_z(0)$) e a variância de Z ($Var[Z] = M''_z(0) - (M'_z(0))^2$).

Fazendo a primeira e segunda derivada da FGM da variável aleatória Z temos [Equação A.1](#) e [Equação A.2](#):

$$\begin{aligned}
M'_z(a) &= \frac{d}{da} \left(\frac{\ln(1-\theta e^a)}{e^a \ln(1-\theta)} \right) \\
&= \frac{[(-\theta e^a) e^a \ln(1-\theta)] / (1-\theta e^a) - [\ln(1-\theta e^a) \ln(1-\theta) e^a]}{(e^a \ln(1-\theta))^2} \\
&= -\frac{\theta}{(1-\theta e^a) \ln(1-\theta)} - \frac{e^{-a} \ln(1-\theta e^a)}{(1-\theta e^a) \ln(1-\theta)} + \frac{\theta \ln(1-\theta e^a)}{(1-\theta e^a) \ln(1-\theta)} \\
&= -\frac{\theta}{(1-\theta e^a) \ln(1-\theta)} - \frac{-e^{-a} \ln(1-\theta e^a)}{\ln(1-\theta)} \tag{A.1}
\end{aligned}$$

e,

$$\begin{aligned}
M''_z(a) &= \frac{d}{da} \left(-\frac{\theta}{(1-\theta e^a) \ln(1-\theta)} - \frac{-e^{-a} \ln(1-\theta e^a)}{\ln(1-\theta)} \right) \\
&= \frac{\theta(1-\theta e^a)^{-2}(-\theta e^a)}{\ln(1-\theta)} + \frac{e^{-a} \ln(1-\theta e^a)}{\ln(1-\theta)} + \frac{\theta}{(1-\theta e^a) \ln(1-\theta)}. \tag{A.2}
\end{aligned}$$

Logo, a esperança da distribuição Z é:

$$\begin{aligned}
 E[Z] &= M'_z(0) \\
 &= -\frac{\theta}{(1-\theta e^0)\ln(1-\theta)} - \frac{e^{-0}\ln(1-\theta e^0)}{\ln(1-\theta)} \\
 &= -\frac{\theta}{(1-\theta)\ln(1-\theta)} - 1.
 \end{aligned}$$

E, a variância da distribuição Z é dada por:

$$\begin{aligned}
 Var[Z] &= M''_z(0) - (M'_z(0))^2 \\
 &= 1 + \frac{\theta(1-\theta) - \theta^2}{(1-\theta)^2\ln(1-\theta)} - \left(-\frac{\theta}{(1-\theta)\ln(1-\theta)} - 1\right)^2 \\
 &= \frac{\theta(1-\theta)\ln(1-\theta) - \theta^2\ln(1-\theta) - 2\theta(1-\theta)\ln(1-\theta)}{(1-\theta)^2(\ln(1-\theta))^2} - \frac{\theta^2}{(1-\theta)^2(\ln(1-\theta))^2} \\
 &= -\frac{\theta}{(1-\theta)^2\ln(1-\theta)} - \frac{\theta}{(1-\theta)^2(\ln(1-\theta))^2} \\
 &= \frac{-\theta}{(1-\theta)^2\ln(1-\theta)} \left(1 + \frac{\theta}{\ln(1-\theta)}\right).
 \end{aligned}$$

ANEXOS GEPGW

B.1 FGP da distribuição Geométrica começando em zero

Podemos obter a FGP da distribuição Geométrica começando em zero como é mostrado na [Equação 4.2](#):

Demonstração.

$$\begin{aligned}
 G_z(s) &= E[s^z] \\
 &= \sum_{z=0}^{\infty} s^z (1-\omega)^z \omega \\
 &= \omega \sum_{z=0}^{\infty} (s(1-\omega))^z \\
 &= \omega \frac{1}{1 - (s(1-\omega))},
 \end{aligned}$$

em que $|s(1-\omega)| < 1$.

□

B.2 A FGM da distribuição Geométrica começando em zero

A FGM da distribuição Geométrica começando em zero é dada pela equação a seguir:

Demonstração.

$$\begin{aligned}
 M_z(a) &= \sum_{z=0}^{\infty} e^{az}(1-\omega)^z \omega \\
 &= \omega \sum_{z=0}^{\infty} (e^a)^z (1-\omega)^z \\
 &= \omega \sum_{z=0}^{\infty} [e^a(1-\omega)]^z \\
 &= \omega \frac{1}{1 - [e^a(1-\omega)]}.
 \end{aligned}$$

□

B.3 Esperança e Variância da distribuição Geométrica começando em zero

Com a [Equação 4.3](#) podemos encontrar o primeiro e o segundo momento da distribuição Z , $E[Z] = M'_Z(0)$ e $E[Z^2] = M''_Z(0)$ respectivamente. A primeira derivada da FGM de Z é dada pela [Equação B.1](#), já a segunda derivada é dada pela [Equação B.1](#):

$$\begin{aligned}
 M'_z(a) &= \frac{d}{da} \left(\frac{\omega}{1 - e^a(1-\omega)} \right) \\
 &= \omega(-1)[1 - e^a(1-\omega)]^{-2} \left(\frac{d}{da} (1 - e^a(1-\omega)) \right) \\
 &= \frac{-\omega}{[1 - e^a(1-\omega)]^2} (-e^a(1-\omega)) \\
 &= \frac{\omega e^a(1-\omega)}{[1 + e^a(\omega-1)]^2}
 \end{aligned}$$

e,

$$\begin{aligned}
 M''_z(a) &= \frac{d}{da} \left(\frac{\omega e^a(1-\omega)}{[1 + e^a(\omega-1)]^2} \right) \\
 &= \frac{\omega(1-\omega)e^a(1 + e^a(\omega-1))^2 - 2\omega(1-\omega)e^a(1 + e^a(\omega-1))(e^a(\omega-1))}{(1 + e^a(\omega-1))^4} \\
 &= \frac{[\omega e^a(1-\omega)(1 + e^a(\omega-1)) - 2\omega e^a(1-\omega)(\omega-1)e^a](1 + e^a(\omega-1))}{(1 + e^a(\omega-1))^4} \\
 &= \frac{\omega e^a(1-\omega) + \omega e^{2a}(\omega-1)(1-\omega) - 2\omega e^{2a}(1-\omega)(\omega-1)}{(1 + e^a(\omega-1))^3} \\
 &= \frac{-\omega e^a(\omega-1) + \omega e^{2a}(\omega-1)(\omega-1)}{(1 + e^a(\omega-1))^3} \\
 &= \frac{\omega e^a(\omega-1)[(\omega-1)\omega e^a - 1]}{(1 + e^a(\omega-1))^3}.
 \end{aligned} \tag{B.1}$$

A esperança da distribuição Z é:

$$\begin{aligned} E[Z] &= M'_z(0) \\ &= \frac{\omega e^0(1-\omega)}{[1+e^0(\omega-1)]^2} \\ &= \frac{1-\omega}{\omega}. \end{aligned}$$

A variância da distribuição Z é dada abaixo:

$$\begin{aligned} Var[Z] &= M''_z(0) - (M'_z(0))^2 \\ &= \frac{\omega e^0(\omega-1)[(\omega-1)\omega e^0 - 1]}{(1+e^0(\omega-1))^3} - \left(\frac{1-\omega}{\omega}\right)^2 \\ &= \frac{\omega(\omega-1)((\omega-1)-1)}{(1+\omega-1)^3} - \left(\frac{1-2\omega+\omega^2}{\omega^2}\right) \\ &= \frac{(\omega-1)(\omega-2)}{\omega^2} + \frac{-1+2\omega-\omega^2}{\omega^2} \\ &= \frac{1-\omega}{\omega^2}. \end{aligned}$$

