

**UNIVERSIDADE DE SÃO PAULO**  
Instituto de Ciências Matemáticas e de Computação

**Modelos de sobrevivência induzidos por fragilidade discreta  
Série de Potência Zero-Modificada**

**Katy Rocio Cruz Molina**

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



SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

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

**Katy Rocio Cruz Molina**

## Modelos de sobrevivência induzidos por fragilidade discreta Série de Potência Zero-Modificada

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

Orientadora: Profa. Dra. Vera Lucia Damasceno Tomazella

Coorientador: Dr. Vinicius Fernando Calsavara

**USP – São Carlos**  
**Abril de 2020**

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

C955m Cruz Molina, Katy Rocio  
Modelos de sobrevivência induzidos por  
fragilidade discreta Série de Potência Zero-  
Modificada / Katy Rocio Cruz Molina; orientadora  
Vera Lucia Damasceno Tomazella; coorientador  
Vinicius Fernando Calsavara. -- São Carlos, 2020.  
81 p.

Dissertação (Mestrado - Programa  
Interinstitucional de Pós-graduação em Estatística) --  
Instituto de Ciências Matemáticas e de Computação,  
Universidade de São Paulo, 2020.

1. Modelos de fragilidade discreto. 2.  
Distribuição Série de Potência Zero-Modificada. 3.  
Fragilidade zero. I. Damasceno Tomazella, Vera  
Lucia, orient. II. Calsavara, Vinicius Fernando,  
coorient. III. Título.

**Katy Rocio Cruz Molina**

**Survival models induced by Zero-Modified Power Series  
discrete frailty**

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

Concentration Area: Statistics

Advisor: Prof. Dra. Vera Lucia  
Damasceno Tomazella

Co-advisor: Dr. Vinicius Fernando Calsavara

**USP – São Carlos**  
**April 2020**





---

**Folha de Aprovação**

---

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado da candidata Katy Rocio Cruz Molina, realizada em 13/03/2020:

---

Profa. Dra. Vera Lucia Damasceno Tomazella  
UFSCar

---

Prof. Dr. Pedro Luiz Ramos

---

Prof. Dr. Eder Angelo Milani  
UFG

*Dedico com muito amor a meus pais e irmão por ser meu apoio incondicional na minha vida.  
Amo vocês.*



# AGRADECIMENTOS

---

---

Agradeço em primeiro lugar a Deus pela vida e saúde, por me abençoar, não me deixar sozinha neste caminho, permitir terminar com sucesso o mestrado e cuidar da minha família.

Aos meus pais, Victoria e Emilio, pelo seu amor, conselhos, apoio incondicional para alcançar meus objetivos e ser minha inspiração desde criança que todo esforço e sacrifício é recompensado grandemente. A meu irmão George, meu segundo pai nesta vida, por seu carinho e apoio. Mesmo estando longe deles, sempre estarão presentes nas minhas orações e no meu coração.

À minha orientadora Professora Doutora Vera Tomazella e meu coorientador Doutor Vinicius Calsavara pelos ensinamentos, disponibilidade, dedicação, carinho e principalmente pela orientação deste trabalho.

Aos membros da banca da qualificação e a defesa, por todas suas sugestões e conselhos que são de enriquecimento do trabalho, especialmente à Professora Doutora Katiane Conceição pela ajuda.

Ao Jonathan, pelo amor, companherismo, risadas e apoios.

Aos meus amigos e colegas feitos em todo este tempo, por sua sincera amizade e vontade de ajudar me nos momentos que estava atrapalhando. Obrigada por todos os momentos compartilhados com vocês.

À Fundação Oncocentro de São Paulo (FOSP) pelos dados fornecidos do estudo de melanoma.

Agradeço a CAPES, pois este trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.



*“Antes que nada, la preparación es llave del éxito.”  
(Alexander Graham Bell)*



# RESUMO

MOLINA, K. R. C. **Modelos de sobrevivência induzidos por fragilidade discreta Série de Potência Zero-Modificada**. 2020. 81 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, 2020.

Modelos de sobrevivência com um termo de fragilidade são apresentados como uma extensão do modelo de risco proporcional de Cox (COX, 1972), em que, um efeito aleatório também denominado fragilidade, é introduzido na função de risco de forma multiplicativa com o objetivo de modelar a heterogeneidade não observada das unidades em estudo. A distribuição para a variável de fragilidade, é assumida contínua e não negativa. Entretanto, há algumas situações nas quais é apropriado considerar a fragilidade distribuída discretamente, por exemplo, quando a heterogeneidade dos tempos de sobrevivência surge por causa da presença de um número aleatório de falhas por unidade ou pela causa da exposição a danos em um número aleatório de ocasiões. Neste trabalho, desenvolvemos diferentes modelos de fragilidade usando algumas distribuições pertencentes à família Série de Potência Zero-Modificada (SPZM). Neste contexto, com o uso da distribuição SPZM para a fragilidade, podemos observar a possibilidade de indivíduos com fragilidade zero, que corresponde a um modelo de falha limitado que contém uma proporção de unidades que nunca falham (sobreviventes de longa duração) ou modelo de fração de cura. O modelo proposto é aplicado a um conjunto de dados reais de melanoma.

**Palavras-chave:** Modelos de fragilidade discreto, distribuição Série de Potência Zero-Modificada, fragilidade zero.



# ABSTRACT

MOLINA, K. R. C. **Survival models induced by Zero-Modified Power Series discrete frailty**. 2020. 81 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, 2020.

Survival models with a frailty term are presented as an extension of Cox's proportional risk model (COX, 1972), in which a random effect, called frailty, is introduced in the risk function in a multiplicative way with the aim of modeling the unobserved heterogeneity from the units under study. The distribution for the frailty variable is assumed to be continuous and not negative. However, there are some situations it is appropriate to consider discretely-distributed frailty, for example, when heterogeneity in lifetimes arises because of the presence of a random number of flaws in a unit or because of exposure to damage on a random number of occasions. In this work, we developed different frailty models applying some distributions belonging to Zero-Modified Power Series (ZMPS) family. In this context, with the use of the ZMPS distribution for frailty, we can notice the possibility of individuals with zero frailty, which corresponds to a limited failure model that contains a proportion of units that never fail (long-term survivors) or cure fraction model. The proposed model is applied to a set of real melanoma data.

**Keywords:** Discrete frailty models, Zero-Modified Power Series distributions, zero frailty.





# LISTA DE ILUSTRAÇÕES

---

---

Figura 1 – Espaço paramétrico de $p$ em função do parâmetro $\mu$ para algumas distribuições SPZM. . . . .	35
Figura 2 – Média dos Estimadores de Máxima Verossimilhança (MEMV), Raiz dos Erros Quadráticos Médios (REQM), Desvio Padrão (DP), Probabilidades de Cobertura (PC) de 90% e 95% dos parâmetros versus o tamanho da amostra para dados simulados do Modelo de Fragilidade PZM com covariáveis (Zero-Inflação). . . . .	52
Figura 3 – Comportamento assintótico dos parâmetros envolvidos do Modelo de Fragilidade PZM com covariáveis para o primeiro estudo de simulação (Zero-Inflação). . . . .	53
Figura 4 – Média dos Estimadores de Máxima Verossimilhança (MEMV), Raiz dos Erros Quadráticos Médios (REQM), Desvio Padrão (DP), Probabilidade de Cobertura (PC) de 90% e 95% dos parâmetros versus o tamanho da amostra para dados simulados do Modelo de Fragilidade PZM com covariáveis (Zero-Deflação). . . . .	56
Figura 5 – Comportamento assintótico dos parâmetros envolvidos do Modelo de Fragilidade PZM com covariáveis para o segundo estudo de simulação (Zero-Deflação). . . . .	57
Figura 6 – Estimativa de Kaplan-Meier para os dados de melanoma. . . . .	60
Figura 7 – Estimativa de Kaplan-Meier e a curva de sobrevivência do Modelo de Fragilidade PZM para os dados de melanoma. . . . .	63
Figura 8 – Estimativa de Kaplan-Meier e a curva de sobrevivência do Modelo de Fragilidade GZM para os dados de melanoma. . . . .	64
Figura 9 – Estimativas de Kaplan-Meier e curvas de sobrevivência do Modelo de Fragilidade PZM estratificado por (a) sexo e (b) estadiamento clínico do câncer para os dados de melanoma. . . . .	67
Figura 10 – Estimativas de Kaplan-Meier e curvas de sobrevivência do Modelo de Fragilidade PZM estratificado por (a) radioterapia e (b) quimioterapia para os dados de melanoma. . . . .	68
Figura 11 – Estimativas de Kaplan-Meier e curvas de sobrevivência do Modelo de Fragilidade GZM estratificado por (a) sexo e (b) estadiamento clínico do câncer para os dados de melanoma. . . . .	70

Figura 12 – Estimativas de Kaplan-Meier e curvas de sobrevivência do Modelo de Fragilidade GZM estratificado por (a) radioterapia e (b) quimioterapia para os dados de melanoma. . . . . 71

# LISTA DE TABELAS

---

---

Tabela 1	– Algumas distribuições pertencentes à família SP. . . . .	28
Tabela 2	– Média e variância de algumas distribuições pertencentes à família SP. . . . .	28
Tabela 3	– Média dos Estimadores de Máxima Verossimilhança (MEMV), Raiz do Erro Quadrático Médio (REQM), Desvio Padrão (DP) e Probabilidade de Cobertura (PC) do Modelo de Fragilidade PZM com covariáveis, sendo, $p = 0,5$ , $\lambda = 0,2$ , $\beta_0 = 0,4$ e $\beta_1 = 1$ (Zero-Inflação). . . . .	51
Tabela 4	– Média dos Estimadores de Máxima Verossimilhança (MEMV), Raiz do Erro Quadrático Médio (REQM), Desvio Padrão (DP) e Probabilidade de Cobertura (PC) da fração de cura e heterogeneidade não observada captada pelo Modelo de Fragilidade PZM com covariáveis sendo $q_1 = 0,612$ , $q_2 = 0,509$ , $\theta_1 = 1,302$ e $\theta_2 = 6,139$ (Zero-Inflação). . . . .	54
Tabela 5	– Média dos Estimadores de Máxima Verossimilhança (MEMV), Raiz do Erro Quadrático Médio (REQM), Desvio Padrão (DP) e Probabilidade de Cobertura (PC) do Modelo de Fragilidade PZM com covariáveis sendo $p = 1,3$ , $\lambda = 0,4$ , $\beta_0 = -1$ e $\beta_1 = 0,5$ (Zero-Deflação). . . . .	55
Tabela 6	– Média dos Estimadores de Máxima Verossimilhança (MEMV), Raiz do Erro Quadrático Médio (REQM), Desvio Padrão (DP) e Probabilidade de Cobertura (PC) da fração de cura e heterogeneidade não observada captada pelo Modelo de Fragilidade PZM com covariáveis, sendo $q_1 = 0,600$ , $q_2 = 0,409$ , $\theta_1 = 0,425$ e $\theta_2 = 0,645$ (Zero-Deflação). . . . .	58
Tabela 7	– Descrição das covariáveis pertencentes aos dados de melanoma. . . . .	60
Tabela 8	– Estimativa de Máxima Verossimilhança (EMV), Erro Padrão (EP), Intervalo de Confiança (IC) de 95%, fração de cura ( $q_0$ ), valores AIC e BIC para o Modelo de Fragilidade PZM. . . . .	61
Tabela 9	– Estimativa de Máxima Verossimilhança (EMV), Erro Padrão (EP), Intervalo de Confiança (IC) de 95%, fração de cura ( $q_0$ ), valores AIC e BIC para o Modelo de Fragilidade GZM. . . . .	62
Tabela 10	– Heterogeneidade não observada (Variância) e Intervalo de Confiança (IC) de 95% atingida pelos Modelos de Fragilidade PZM e GZM. . . . .	62
Tabela 11	– Estimativa de Máxima Verossimilhança (EMV), Erro Padrão (EP), Intervalo de Confiança (IC) de 95%, valores AIC e BIC do Modelo de Fragilidade PZM classificado por sexo, estadiamento clínico do câncer, radioterapia e quimioterapia para os dados de melanoma. . . . .	66

Tabela 12 – Estimativa de Máxima Verossimilhança (EMV), Erro Padrão (EP), Intervalo de Confiança (IC) de 95%, valores AIC e BIC do Modelo de Fragilidade GZM classificado por sexo, estadiamento clínico do câncer, radioterapia e quimioterapia para os dados de melanoma. . . . .	69
Tabela 13 – Estimativa de Máxima Verossimilhança (EMV), Erro Padrão (EP), Intervalo de Confiança (IC) de 95%, valores AIC e BIC do Modelo de Fragilidade PZM para os dados de melanoma considerando todas as covariáveis. . . . .	73
Tabela 14 – Estimativa de Máxima Verossimilhança (EMV), Erro Padrão (EP), Intervalo de Confiança (IC) de 95%, valores AIC e BIC do Modelo de Fragilidade GZM para os dados de melanoma considerando todas as covariáveis. . . . .	73
Tabela 15 – Fração de cura univariada ( $q_0$ ), Erro Padrão (EP), e Intervalo de Confiança (IC) de 95% do Modelo de Fragilidade GZM analisada para sexo, estadiamento clínico, radioterapia e quimioterapia para as idades 30 e 60 anos. . . . .	75

# LISTA DE ABREVIATURAS E SIGLAS

---

---

B	Binomial
BN	Binomial Negativa
BZD	Binomial Zero-Deflacionada
BZI	Binomial Zero-Inflacionado
BZM	Binomial Zero-Modificada
BZT	Binomial Zero-Truncada
FOSP	Fundação Oncocentro São Paulo
FVP	Funções de Variação de Potência
G	Geométrica
GZD	Geométrica Zero-Deflacionada
GZI	Geométrica Zero-Inflacionado
GZM	Geométrica Zero-Modificada
GZT	Geométrica Zero-Truncada
P	Poisson
PG	Poisson Generalizada
PPCD	Processo de Poisson Composto Discreto
PZD	Poisson Zero-Deflacionada
PZI	Poisson Zero-Inflacionado
PZM	Poisson Zero-Modificada
PZT	Poisson Zero-Truncada
SP	Série de Potência
SPZD	Série de Potência Zero-Deflacionada
SPZI	Série de Potência Zero-Inflacionada
SPZM	Série de Potência Zero-Modificadas
SPZT	Série de Potência Zero-Truncada
ZMPS	Zero-Modified Power Series



# SUMÁRIO

---

---

1	INTRODUÇÃO . . . . .	23
1.1	Objetivo . . . . .	25
1.2	Organização dos Capítulos . . . . .	25
2	DISTRIBUIÇÕES SÉRIE DE POTÊNCIA ZERO-MODIFICADAS (SPZM) . . . . .	27
2.1	Distribuições Série de Potência . . . . .	27
2.2	Distribuições Série de Potência Zero-Modificadas . . . . .	29
2.3	Casos Particulares da Distribuição SPZM . . . . .	30
2.4	Algumas Distribuições da família SPZM . . . . .	31
2.4.1	<i>Distribuição Poisson Zero-Modificada (PZM)</i> . . . . .	31
2.4.2	<i>Distribuição Geométrica Zero-Modificada (GZM)</i> . . . . .	32
2.4.3	<i>Distribuição Binomial Zero-Modificada (BZM)</i> . . . . .	33
2.5	Distribuição Série de Potência Zero-Modificada na versão hurdle . . . . .	36
2.5.1	<i>Casos Particulares</i> . . . . .	36
2.6	Síntese . . . . .	37
3	MODELOS DE SOBREVIVÊNCIA INDUZIDOS POR FRAGILIDADE DISCRETA SPZM . . . . .	39
3.1	Modelo de Fragilidade . . . . .	39
3.2	Modelos de Fragilidade Discreta . . . . .	41
3.2.1	<i>Modelo de Fragilidade PZM</i> . . . . .	42
3.2.2	<i>Modelo de Fragilidade GZM</i> . . . . .	44
3.2.3	<i>Modelo de Fragilidade BZM</i> . . . . .	46
3.3	Modelos de Fragilidade SPZM com covariáveis . . . . .	48
3.4	Estudo de Simulação . . . . .	49
3.5	Síntese . . . . .	58
4	APLICAÇÃO EM DADOS DE MELANOMA . . . . .	59
5	CONCLUSÕES E PROPOSTAS FUTURAS . . . . .	77
	REFERÊNCIAS . . . . .	79





---

## INTRODUÇÃO

---

Uma das áreas da Estatística que tem apresentado um crescimento considerável, nas últimas décadas, é a Análise de Sobrevivência e confiabilidade. Esta consiste de procedimentos estatísticos para analisar dados relacionados ao tempo até a ocorrência de algum evento de interesse específico (morte, cura, recidiva de uma doença, falha de um componente eletrônico, queima de uma lâmpada, tempo até a germinação de uma semente, etc.), que em geral, pode ser modelado por uma distribuição de probabilidade adequada. Os dados de sobrevivência são descritos por uma variável  $T$  que indica o tempo até a ocorrência de um evento de interesse e  $\delta$  uma variável indicadora de censura, que são observações incompletas ou parciais em que esta assume o valor 1 se o tempo de sobrevivência for completamente observado e 0 caso contrário. Além disso, os dados incorporam tanto os tempos de sobrevivência quanto um conjunto de variáveis observáveis que podem estar relacionadas com seus respectivos tempos. Essas variáveis também são conhecidas como variáveis explicativas, covariáveis ou covariadas. Uma forma de verificar a influência dessas covariáveis nos tempos de sobrevivência é através de Modelos de Regressão.

Em aplicações de análise de sobrevivência, geralmente apenas algumas covariáveis como idade, sexo, gravidade da doença ou dados laboratoriais, são conhecidas e medidas. Mas existem muitos outros fatores que podem influenciar a sobrevivência, por exemplo estado de saúde, estilo de vida, ocupação, risco genético, fatores ambientais, entre outros e esses fatores são desconhecidos ou não são incluídos na análise. Nesse contexto, [Vaupel, Manton e Stallard \(1979\)](#) considerou o Modelo de Fragilidade que é caracterizado pela utilização de um efeito aleatório, ou seja, de uma variável aleatória não observável que representa as informações que não podem ou que não foram observadas por algum motivo. Assim, a idéia desses modelos é que os indivíduos têm diferentes fragilidades e os indivíduos mais frágeis ou propensos morrerão mais cedo que os menos frágeis ([ANDREAS, 2010](#)). Além disso, frequentemente tem-se a suposição de que todos os indivíduos são suscetíveis a experimentar o evento de interesse.

Diferentes estudos sobre Modelos de Fragilidade geralmente consideram possíveis candidatas à distribuição da fragilidade como sendo contínuas e não negativas, tais como a distribuição Gama (VAUPEL; MANTON; STALLARD, 1979; HOUGAARD, 1984), Log-Normal (MCGILCHRIST; AISBETT, 1991; SANTOS; DAVIES; FRANCIS, 1995), Gaussiana Inversa (HOUGAARD, 1984), Estável Positiva (HOUGAARD, 1986) e Gama Generalizada (BALAKRISHNAN; PENG, 2006; CHEN; ZHANG; ZHANG, 2013). Em Tomazella (2003), as distribuições Gama e Gaussiana Inversa são apresentadas em um estudo comparativo para modelos de regressão em processos de Poisson. Do ponto de vista computacional, essas distribuições são convenientes para derivar funções de sobrevivência incondicionais em forma fechada, funções de densidade e taxas de risco usando a transformada de Laplace. Contudo, uma distribuição contínua da fragilidade não permite a existência de risco zero.

Com o rápido desenvolvimento dos tratamentos médicos, em geral os dados revelam a existência de uma fração de indivíduos os quais não se espera que vivenciem o evento de interesse. Assim, segundo Maller e Zhou (1996) e Ibrahim, Chen e Sinha (2001) as unidades experimentais em que o evento de interesse não aconteceu mesmo após um período longo de observação podem ser chamados curados ou imunes. Neste sentido, pode ser apropriado considerar distribuições discretas para a fragilidade pois a partir da função de sobrevivência pode ser obtido a fragilidade zero ou também chamado fração de cura que indica a existência de um subgrupo de indivíduos não suscetíveis onde o evento de interesse não ocorreu mesmo após um longo período de observação. Diferentes autores estudaram o uso de distribuições discretas para a fragilidade, por exemplo Caroni, Crowder e Kimber (2010) propuseram os modelos de fragilidade discreta utilizando as distribuições Binomial Negativa, Poisson e Geométrica, Ata e Özel (2013) desenvolveram o Modelo de Fragilidade baseados no Processo de Poisson Composto Discreto (PPCD), Zavaleta *et al.* (2016), Cancho *et al.* (2019) consideraram as distribuições Série de Potência Zero-Inflacionadas para o Modelo de Fragilidade discreta e Souza *et al.* (2017) propuseram a distribuição estendida de Conway-Maxwell e a distribuição hiper Poisson (hP) como caso particular para a modelagem da sobrevivência induzida por uma variável discreta de fragilidade.

Considere, por exemplo, um cenário em que uma paciente foi diagnosticada com câncer de mama, o termo câncer de mama refere-se a um tumor maligno que se desenvolveu a partir de células da mama. Geralmente, o câncer de mama se origina nas células dos lóbulos, que são as glândulas produtoras de leite, ou nos ductos, que são as vias que transportam o leite dos lóbulos para o mamilo (BREASTCANCER, 2020). Nesta situação, não observamos diretamente o comportamento das células cancerígenas, mas temos conhecimento que a paciente tem câncer pela presença de um número aleatório de células malignas. Se a paciente não faz nenhum tratamento, ela vai continuar em risco de morrer. No momento em que a paciente tenha o câncer sob controle e não esteja mais em risco podemos inferir que a paciente estará curada. Desta forma, relacionamos o termo de fragilidade com a contagem de células malignas.

Assim, a análise de dados de contagem ocupa um importante lugar dentro da estatística aplicada, pois muitos problemas reais são expressos em termos de enumerações. Geralmente, a distribuição de Poisson é utilizada para analisar esse tipo de dados e a presença de zeros em um conjunto de dados pode ser maior ou menor do que o esperado, tornando assim errado o uso de uma distribuição de contagem padrão. Neste sentido, é chamado zero-inflacionado a alta ocorrência de zeros nas observações e zero-deflacionado a baixa presença de zeros. [Dietz e Böhning \(2000\)](#) apresentaram uma distribuição que pode ser utilizado sem o conhecimento prévio da inflação ou deflação de zero na amostra e que o parâmetro oriundo da distribuição de Poisson é estimado sem fazer o uso das observações zero. Segundo estes autores, a ocorrência de muitos ou poucos zeros pode ocorrer das seguintes situações:

- a) Nem todos os membros da população de estudo são afetados pelo processo, de modo que, a inflação de zero ocorre devido a resposta zero dos membros não afetados.
- b) Certos problemas inevitáveis ao processo de amostragem levam a um crescimento ou decrescimento da chance de um indivíduo da população com contagem zero de ser amostrado.
- c) Não existe chance de ocorrer uma observação zero dentro da amostra (distribuição positiva, truncada em zero).
- d) Uma combinação das situações a) e c), sendo que uma parte da população vem de um processo positivo e a outra parte, a qual não é afetada, fornecem as observações zero.

## 1.1 Objetivo

Este trabalho, tem por objetivo desenvolver Modelos de Sobrevivência induzidos por fragilidade para dados de tempo de vida que permitem o uso de distribuições discretas para a variável de fragilidade. Propomos o uso de uma nova família de distribuições chamada Série de Potência Zero-Modificadas (SPZM) para a variável de fragilidade, sendo esta família estudada amplamente por [Conceição \(2013\)](#), [Conceição et al. \(2017\)](#). A vantagem da utilização desta família de distribuições é que são flexíveis e capazes de capturar a presença inflacionada, deflacionada ou truncada de zero na população. Além disso a modelagem proposta permite ter indivíduos com fragilidade zero ou risco zero que corresponde a um modelo com fração de cura.

## 1.2 Organização dos Capítulos

Este trabalho está organizado como segue: no [Capítulo 2](#) apresentamos a família de distribuições Série de Potência Zero-Modificada, suas propriedades e particularidades. No [Capítulo 3](#) apresentamos os Modelos de Sobrevivência induzidos por fragilidade discreta SPZM, em que desenvolvemos os Modelos de Fragilidade SPZM sem e com a presença de covariáveis. Ademais

mostramos o respectivo estudo de simulação. No [Capítulo 4](#) dissertamos a metodologia proposta no capítulo anterior, numa aplicação em dados reais de melanoma. Finalmente indicamos as conclusões e propostas futuras deste trabalho no [Capítulo 5](#).

Para o desenvolvimento deste trabalho, utilizamos o Software [R Core Team \(2019\)](#) para a obtenção dos resultados e as simulações.

# DISTRIBUIÇÕES SÉRIE DE POTÊNCIA ZERO-MODIFICADAS (SPZM)

---



---

Neste Capítulo, apresentamos a família de distribuições Série de Potência Zero-Modificadas (SPZM) e suas propriedades, proposta por [Conceição \(2013\)](#). Esta família de distribuições, cujo suporte se inicia no zero, tem várias distribuições pertencente a essa classe, tais como: Poisson Zero-Modificada (PZM), Geométrica Zero-Modificada (GZM), Binomial Zero-Modificada (BZM), entre outras.

Para a formulação das distribuições SPZM, primeiro é necessário definir a distribuição Série de Potência (SP), pois a partir desta, é feita uma modificação na probabilidade da presença de zeros.

## 2.1 Distribuições Série de Potência

Uma grande classe de variáveis aleatórias com distribuição discreta de probabilidade podem ser derivadas de certas série de potências. Assim, [Noack \(1950\)](#) considerou uma classe de variável aleatória com distribuição discreta, onde definiu distribuições de Série de Potência e mostrou que muitas distribuições discretas importantes pertencem a essa classe. No entanto, [Consul \(1990\)](#) desenvolveu e estudou a reparametrização na distribuição SP em termos da média.

Seja uma variável aleatória  $Y$  definida sobre os inteiros não negativos, que segue uma distribuição SP,  $Y \sim SP(\mu, \phi)$ , com parâmetro de média  $\mu > 0$  e parâmetro de dispersão  $\phi \geq 0$ . A função de probabilidade é definida por:

$$\pi_{SP}(y; \mu, \phi) = \frac{\alpha(y, \phi) (g(\mu, \phi))^y}{f(\mu, \phi)}, \quad y \in A_s, \quad (2.1)$$

em que  $A_s$  é o suporte do subconjunto dos inteiros  $\{s, s+1, \dots\}$ , com  $s \geq 0$ ,  $\alpha(y, \phi)$  é uma função positiva,  $g(\mu, \phi)$  e  $f(\mu, \phi)$  são funções positivas, finitas e duas vezes diferenciáveis,

sendo que  $f(\mu, \phi) = \sum_{y \in A_s} \alpha(y, \phi) g(\mu, \phi)^y$  (CORDEIRO; ANDRADE; CASTRO, 2009).

A família SP, definida em (2.1) contém várias distribuições conhecidas, como Poisson, Poisson Generalizado, Geométrica, Binomial, Binomial Negativa e entre várias outras distribuições. Na Tabela 1 exibimos as respectivas funções  $\alpha(y, \phi)$ ,  $g(\mu, \phi)$  e  $f(\mu, \phi)$  para algumas distribuições pertencentes à família SP, sendo que o suporte destas distribuições iniciam no zero.

Tabela 1 – Algumas distribuições pertencentes à família SP.

Distribuição	$\alpha(y, \phi)$	$g(\mu, \phi)$	$f(\mu, \phi)$	$A_s$
Poisson (P)	$\frac{1}{y!}$	$\mu$	$e^\mu$	$\{0, 1, \dots\}$
Poisson Generalizada (PG)	$\frac{(1+\phi y)^{y-1}}{y!}$	$\frac{\mu e^{-\mu\phi(1+\mu\phi)^{-1}}}{1+\mu\phi}$	$e^{\mu(1+\mu\phi)^{-1}}$	$\{0, 1, \dots\}$
Geométrica (G)	1	$\frac{\mu}{1+\mu}$	$1 + \mu$	$\{0, 1, \dots\}$
Binomial (B)	$\binom{m}{y}$	$\frac{\mu}{m-\mu}$	$\left(\frac{m}{m-\mu}\right)^m$	$\{0, 1, \dots, m\}$
Binomial Negativa (BN)	$\frac{\Gamma(\phi+y)}{y!\Gamma(\phi)}$	$\frac{\mu}{\mu+\phi}$	$\left(\frac{\phi}{\mu+\phi}\right)^{-\phi}$	$\{0, 1, \dots\}$

Fonte: Adaptada de Conceição (2013).

A média e variância de uma variável aleatória  $Y$  com distribuição SP são, respectivamente:

$$\mathbb{E}(Y) = \mu_{SP} = \frac{f'(\mu, \phi) g(\mu, \phi)}{f(\mu, \phi) g'(\mu, \phi)} \quad \text{e} \quad \text{Var}(Y) = \sigma_{SP}^2 = \frac{g(\mu, \phi)}{g'(\mu, \phi)},$$

em que  $f'(\mu, \phi)$  e  $g'(\mu, \phi)$  denotam a derivada da função com relação a  $\mu$  (GUPTA, 1974).

Na Tabela 2 mostramos as respectivas média e variância para algumas distribuições pertencentes à família SP.

Tabela 2 – Média e variância de algumas distribuições pertencentes à família SP.

Distribuição	Média	Variância
Poisson (P)	$\mu$	$\mu$
Poisson Generalizada(PG)	$\mu$	$\mu(1 + \phi\mu)^2$
Geométrica(G)	$\mu$	$\mu(1 + \mu)$
Binomial (B)	$\mu$	$\mu\left(1 - \frac{\mu}{m}\right)$
Binomial Negativa (BN)	$\frac{\mu(1 + \phi)}{(1 + \mu)}$	$\mu\left(1 + \frac{\mu}{\phi}\right)$

Fonte: Elaborada pelo autor.

Uma modificação na probabilidade da observação zero na distribuição SP foi proposta e estudadas por [Conceição \(2013\)](#), [Conceição et al. \(2017\)](#). A família de distribuições SPZM são estudadas na seguinte seção.

## 2.2 Distribuições Série de Potência Zero-Modificadas

De acordo com [Conceição \(2013\)](#), a família de distribuições SPZM é adequada para conjunto de dados que tem presença de zeros diferentes das esperadas pelas distribuições SP usuais. Desta forma, a família de distribuições SPZM são construídas a partir da sua distribuição SP.

Seja uma variável aleatória  $Y$  definida sobre os inteiros não negativos, que segue uma distribuição SPZM,  $Y \sim \text{SPZM}(\mu, \phi, p)$ . A função de probabilidade é definida por:

$$\pi_{\text{SPZM}}(y; \mu, \phi, p) = (1 - p)I(y) + p\pi_{\text{SP}}(y; \mu, \phi), \quad y \in A_y, \quad (2.2)$$

em que  $A_y$  é o suporte formado pelo subconjunto dos inteiros  $\{0, 1, \dots\}$ ,  $\pi_{\text{SP}}(y; \mu, \phi)$  é a distribuição SP definida em [\(2.1\)](#),  $I(y)$  é uma função indicadora tal que:

$$I(y) = \begin{cases} 1, & \text{se } y = 0, \\ 0, & \text{se } y > 0, \end{cases}$$

e  $p$  é o parâmetro responsável pela modificação das probabilidades das distribuições SP, obedecendo a condição:

$$0 \leq p \leq \frac{1}{1 - \pi_{\text{SP}}(0; \mu, \phi)},$$

sendo  $\pi_{\text{SP}}(0; \mu, \phi)$  a distribuição SP no ponto zero.

A média e variância da variável aleatória  $Y$  com distribuição SPZM são, respectivamente:

$$\mathbb{E}(Y) = \mu_{\text{SPZM}} = p\mu \quad \text{e} \quad \text{Var}(Y) = \sigma_{\text{SPZM}}^2 = p[\sigma_{\text{SP}}^2 + \mu^2(1 - p)].$$

A função geradora de probabilidade da variável aleatória  $Y$  com distribuição SPZM é dada por:

$$G_Y(x) = 1 - p[1 - G(x)],$$

em que  $G(x)$  é a função geradora de probabilidade da distribuição SP.

A distribuição SPZM tem diferentes distribuições que são casos particulares desta família e serão estudadas na seguinte seção.

## 2.3 Casos Particulares da Distribuição SPZM

Da diferença no ponto zero da distribuição SPZM e a distribuição SP, temos.

$$\begin{aligned}\pi_{SPZM}(0; \mu, \phi, p) - \pi_{SP}(0; \mu, \phi) &= 1 - p + p\pi_{SP}(0; \mu, \phi) - \pi_{SP}(0; \mu, \phi) \\ &= 1 - p - \pi_{SP}(0; \mu, \phi)(1 - p) \\ &= (1 - p)[1 - \pi_{SP}(0; \mu, \phi)].\end{aligned}\quad (2.3)$$

Portanto, diferentes valores de  $p$  em (2.3), levam a diferentes distribuições SPZM, considerando a presença de zeros adicionais ou ausentes.

- i) Se  $p = 0$  em (2.3), temos que  $\pi_{SPZM}(0; \mu, \phi, p) = 1$ , então  $\pi_{SPZM}(y; \mu, \phi, p)$  é uma distribuição degenerada com toda massa no ponto zero.
- ii) Se  $0 < p < 1$  em (2.3), temos que  $(1 - p)(1 - \pi_{SP}(0; \mu)) > 0$ , portanto obtemos  $\pi_{SPZM}(0; \mu, \phi, p) > \pi_{SP}(0; \mu, \phi)$ . Então  $\pi_{SPZM}(y; \mu, \phi, p)$  é a distribuição Série de Potência Zero-Inflacionada (SPZI).
- iii) Se  $p = 1$  em (2.3), temos que  $\pi_{SPZM}(0; \mu, \phi, p) - \pi_{SP}(0; \mu, \phi) = 0$ , portanto obtemos  $\pi_{SPZM}(0; \mu, \phi, p) = \pi_{SP}(0; \mu, \phi)$ . Então  $\pi_{SPZM}(0; \mu, \phi, p)$  é a distribuição Série de Potência usual.
- iv) Se  $1 < p < \frac{1}{1 - \pi_{SP}(0; \mu, \phi)}$  em (2.3), temos que  $\pi_{SPZM}(0; \mu, \phi, p) - \pi_{SP}(0; \mu, \phi) < 0$ , portanto  $\pi_{SPZM}(0; \mu, \phi, p) < \pi_{SP}(0; \mu, \phi)$ . Então  $\pi_{SPZM}(y; \mu, \phi, p)$  é a distribuição Série de Potência Zero-Deflacionada (SPZD).
- v) Se  $p = \frac{1}{1 - \pi_{SP}(0; \mu, \phi)}$  em (2.3), temos que  $\pi_{SPZM}(0; \mu, \phi, p) = 0$ . Então  $\pi_{SPZM}(y; \mu, \phi, p)$  é uma distribuição Série de Potência Zero-Truncada (SPZT).

A distribuição SPZT associada à distribuição da variável aleatória  $Y$  é dada por:

$$\pi_{SPZT}(y; \mu, \phi) = \left[ \frac{\pi_{SP}(y; \mu, \phi)}{1 - \pi_{SP}(0; \mu, \phi)} \right] I_{A_1}(y), \quad (2.4)$$

em que  $A_1$  é o suporte formado pelo subconjunto dos inteiros  $\{1, 2, \dots\}$ ,  $\pi_{SP}(y; \mu, \phi)$  é a distribuição SP associada e  $I_{A_1}(y)$  é uma função indicadora, tal que:

$$I_{A_1}(y) = \begin{cases} 1, & \text{se } y \in A_1, \\ 0, & \text{se } y \notin A_1. \end{cases}$$

A média e a variância da variável aleatória  $Y$  com distribuição SPZT são, respectivamente:

$$\mathbb{E}(Y) = \mu_{SPZT} \frac{\mu}{1 - \pi_{SP}(0; \mu, \phi)} \quad \text{e} \quad \text{Var}(Y) = \sigma_{SPZT}^2 \frac{\sigma_{SP}^2}{f(\mu, \phi) - 1} \left[ f(\mu, \phi) - \mu \frac{f'(\mu, \phi)}{f(\mu, \phi) - 1} \right].$$



A família de distribuições SPZM tem como casos particulares as distribuições SPZI, SP, SPZD e SPZT. O parâmetro  $p$  na distribuição SPZM desempenha o papel de aumentar ( $0 < p < 1$ ) ou diminuir  $\left(1 < p < \frac{1}{1 - \pi_{SP}(0; \mu, \phi)}\right)$  a probabilidade de zero com relação a distribuição SP usual (CONCEIÇÃO, 2013).

Portanto, uma das vantagens em utilizar uma distribuição SPZM é a flexibilidade de poder ser ajustada a qualquer conjunto de dados de contagem, dispensando qualquer conhecimento prévio da caracterização dos dados (zero-inflacionados, zero-deflacionados ou zero-truncado) ocasionada pela frequência de zero.

## 2.4 Algumas Distribuições da família SPZM

Nesta Seção, apresentamos uma breve descrição de algumas distribuições pertencentes à família SPZM, por exemplo a distribuição PZM, GZM e BZM.

### 2.4.1 Distribuição Poisson Zero-Modificada (PZM)

A distribuição Poisson foi descoberta e publicada por Poisson (1837) conjuntamente com a sua teoria da probabilidade. Seja  $Y$  uma variável aleatória com distribuição Poisson,  $Y \sim P(\mu)$ , com parâmetro de média  $\mu > 0$ . A função de probabilidade da variável aleatória  $Y$  é dada por:

$$\pi_P(y; \mu) = \frac{e^{-\mu} \mu^y}{y!}, \quad y = 0, 1, \dots \quad (2.5)$$

A distribuição PZM é construída a partir de (2.5). Desta forma, a função de probabilidade da distribuição PZM é definida por:

$$\pi_{PZM}(y; \mu, p) = (1 - p)I(y) + p \frac{e^{-\mu} \mu^y}{y!}, \quad \mu > 0, y = 0, 1, \dots, \quad (2.6)$$

em que  $0 \leq p \leq \frac{1}{1 - \pi_P(0; \mu)}$  é a restrição para o parâmetro  $p$ , sendo que  $\pi_P(0; \mu)$  é a distribuição Poisson calculada no ponto zero.

A partir da diferença no ponto zero da distribuição PZM e Poisson temos:

$$\begin{aligned} \pi_{PZM}(0; \mu, p) - \pi_P(0; \mu) &= 1 - p + pe^{-\mu} - e^{-\mu} \\ &= 1 - p - e^{-\mu}(1 - p) \\ &= (1 - p)(1 - e^{-\mu}). \end{aligned} \quad (2.7)$$

Desta forma, diversos valores de  $p$  levam a diferentes modificações na frequência do zero na distribuição PZM.

- i) Se  $p = 0$  em (2.7), temos que  $\pi_{PZM}(0; \mu, p) = 1$ , portanto temos que  $\pi_{PZM}(y; \mu, p)$  é uma distribuição degenerada com toda massa no ponto zero.

- ii) Se  $0 < p < 1$  em (2.7), temos que  $\pi_{PZM}(0; \mu, p) - \pi_P(0; \mu) > 0$ . Portanto  $\pi_{PZM}(y; \mu, p)$  é a distribuição Poisson Zero-Inflacionado (PZI).
- iii) Se  $p = 1$  em (2.7), temos que  $\pi_{PZM}(0; \mu, p) - \pi_P(0; \mu) = 0$ . Portanto  $\pi_{PZM}(y; \mu, p)$  é a distribuição Poisson (P).
- iv) Se  $1 < p < \frac{1}{1 - e^{-\mu}}$  em (2.7), temos que  $\pi_{PZM}(0; \mu, p) - \pi_P(0; \mu) < 0$ . Portanto  $\pi_{PZM}(y; \mu, p)$  é a distribuição Poisson Zero-Deflacionada (PZD).
- v) Se  $p = \frac{1}{1 - e^{-\mu}}$  em (2.7), temos que  $\pi_{PZM}(0; \mu, p) = 0$ . Portanto  $\pi_{PZM}(y; \mu, p)$  é uma distribuição Poisson Zero-Truncada (PZT).

A média e a variância de uma variável aleatória  $Y$  com distribuição PZM são, respectivamente:

$$\mathbb{E}(Y) = \mu_{PZM} = p\mu \quad \text{e} \quad \text{Var}(Y) = \sigma_{PZM}^2 = p\mu + p(1-p)\mu^2. \quad (2.8)$$

A respectiva função geradora de probabilidade de uma variável aleatória  $Y$  com distribuição PZM é definida por:

$$G_Y(x) = 1 - p \left[ 1 - e^{-\mu(1-x)} \right]. \quad (2.9)$$

#### 2.4.2 Distribuição Geométrica Zero-Modificada (GZM)

Seja  $Y$  uma variável aleatória com distribuição Geométrica,  $Y \sim G(\mu)$ , com parâmetro de média  $\mu > 0$ . A função de probabilidade da variável aleatória  $Y$  é dada por:

$$\pi_G(y; \mu) = \frac{\left( \frac{\mu}{1+\mu} \right)^y}{1+\mu}, \quad y = 0, 1, \dots \quad (2.10)$$

A distribuição GZM é construída a partir de (2.10). Desta forma, a função de probabilidade da distribuição GZM é dada por:

$$\pi_{GZM}(y; \mu, p) = (1-p)I(y) + p \frac{\left( \frac{\mu}{1+\mu} \right)^y}{1+\mu}, \quad \mu > 0, y = 0, 1, \dots, \quad (2.11)$$

em que  $0 \leq p \leq \frac{1}{1 - \pi_G(0; \mu)}$  é a restrição para o parâmetro  $p$ , sendo que  $\pi_G(0; \mu)$  é a distribuição Geométrica calculada no ponto zero.

Da diferença no ponto zero da distribuição GZM e Geométrica temos:

$$\begin{aligned} \pi_{GZM}(0; \mu, p) - \pi_G(0; \mu) &= 1 - p + p \frac{1}{1+\mu} - \frac{1}{1+\mu} \\ &= 1 - p - \frac{1}{1+\mu} (1-p) \\ &= (1-p) \left( 1 - \frac{1}{1+\mu} \right). \end{aligned} \quad (2.12)$$

Desta forma, diferentes valores de  $p$  levam a diferentes modificações na frequência do zero na distribuição GZM.

- i) Se  $p = 0$  em (2.12), temos que  $\pi_{GZM}(0; \mu, p) = 1$ , portanto  $\pi_{GZM}(y; \mu, p)$  é uma distribuição degenerada com toda massa no ponto zero.
- ii) Se  $0 < p < 1$  em (2.12), temos que  $\pi_{GZM}(0; \mu, p) - \pi_G(0; \mu) > 0$ . Portanto  $\pi_{GZM}(y; \mu, p)$  é a distribuição Geométrica Zero-Inflacionado (GZI).
- iii) Se  $p = 1$  em (2.12), temos que  $\pi_{GZM}(0; \mu, p) - \pi_G(0; \mu) = 0$ . Portanto  $\pi_{GZM}(y; \mu, p)$  é a distribuição Geométrica (G).
- iv) Se  $1 < p < \frac{1+\mu}{\mu}$  em (2.12), temos que  $\pi_{GZM}(0; \mu, p) - \pi_G(0; \mu) < 0$ . Portanto  $\pi_{GZM}(y; \mu, p)$  é a distribuição Geométrica Zero-Deflacionada (GZD).
- v) Se  $p = \frac{1+\mu}{\mu}$  em (2.12), temos que  $\pi_{GZM}(0; \mu, p) = 0$ . Portanto  $\pi_{GZM}(y; \mu, p)$  é uma distribuição Geométrica Zero-Truncada (GZT).

A média e a variância de uma variável aleatória  $Y$  com distribuição GZM são, respectivamente:

$$\mathbb{E}(Y) = \mu_{GZM} = p\mu \quad \text{e} \quad \text{Var}(Y) = \sigma_{GZM}^2 = p\mu(1+\mu) + p(1-p)\mu^2.$$

A correspondente função geradora de probabilidade de uma variável aleatória  $Y$  com distribuição GZM é definida por:

$$G_Y(x) = 1 - p \left[ 1 - \frac{1}{1 + \mu - \mu x} \right]. \quad (2.13)$$

### 2.4.3 Distribuição Binomial Zero-Modificada (BZM)

Seja uma variável aleatória  $Y$  com distribuição Binomial,  $Y \sim B(\mu)$ , com parâmetro de média  $\mu > 0$ . A função de probabilidade da variável aleatória  $Y$  é dada por:

$$\pi_B(y; \mu) = \frac{\binom{m}{y} \left( \frac{\mu}{m-\mu} \right)^y}{\left( \frac{m}{m-\mu} \right)^m}, \quad y = 0, 1, \dots \quad (2.14)$$

A distribuição BZM é construída a partir de (2.14). Desta forma, a função de probabilidade da distribuição BZM é definida por:

$$\pi_{BZM}(y; \mu, p) = (1-p)I(y) + p \frac{\binom{m}{y} \left( \frac{\mu}{m-\mu} \right)^y}{\left( \frac{m}{m-\mu} \right)^m}, \quad 0 < \mu < m, y = 0, 1, \dots, \quad (2.15)$$

em que  $0 \leq p \leq \frac{1}{1 - \pi_B(0; \mu)}$  é a restrição para o parâmetro  $p$ , sendo que  $\pi_B(0; \mu)$  é a distribuição Binomial calculada no ponto zero.

Da diferença no ponto zero da distribuição BZM e Binomial temos:

$$\begin{aligned} \pi_{BZM}(0; \mu, p) - \pi_B(0; \mu) &= 1 - p + p \left( \frac{m - \mu}{m} \right)^m - \left( \frac{m - \mu}{m} \right)^m \\ &= 1 - p - \left( \frac{m - \mu}{m} \right)^m (1 - p) \\ &= (1 - p) \left[ 1 - \left( \frac{m - \mu}{m} \right)^m \right]. \end{aligned} \quad (2.16)$$

Desta forma, distintos valores de  $p$  levam a diferentes modificações na frequência do zero na distribuição de BZM.

- i) Se  $p = 0$  em (2.16), temos que  $\pi_{BZM}(0; \mu, p) = 1$ , portanto  $\pi_{BZM}(y; \mu, p)$  é uma distribuição degenerada com toda massa no ponto zero.
- ii) Se  $0 < p < 1$  em (2.16), temos que  $\pi_{BZM}(0; \mu, p) - \pi_B(0; \mu) > 0$ . Portanto  $\pi_{BZM}(y; \mu, p)$  é a distribuição Binomial Zero-Inflacionado (BZI).
- iii) Se  $p = 1$  em (2.16), temos que  $\pi_{BZM}(0; \mu, p) - \pi_B(0; \mu) = 0$ . Portanto  $\pi_{BZM}(y; \mu, p)$  é a distribuição Binomial (B).
- iv) Se  $1 < p < \frac{m^m}{m^m - (m - \mu)^m}$  em (2.16), temos que  $\pi_{BZM}(0; \mu, p) - \pi_B(0; \mu) < 0$ . Portanto,  $\pi_{BZM}(y; \mu, p)$  é a distribuição Binomial Zero-Deflacionada (BZD).
- v) Se  $p = \frac{m^m}{m^m - (m - \mu)^m}$  em (2.16), temos que  $\pi_{BZM}(0; \mu, p) = 0$ . Portanto  $\pi_{BZM}(y; \mu, p)$  é uma distribuição Binomial Zero-Truncada (BZT).

A média e a variância de uma variável aleatória  $Y$  com distribuição BZM são, respectivamente:

$$\mathbb{E}(Y) = \mu_{BZM} = p\mu \quad \text{e} \quad \text{Var}(Y) = \sigma_{BZM}^2 = p\mu \left( 1 - \frac{\mu}{m} \right) + p(1 - p)\mu^2.$$

A função geradora de probabilidade de uma variável aleatória  $Y$  com distribuição BZM é definida por:

$$G_Y(x) = 1 - p \left[ 1 - \left( \frac{m - \mu + \mu x}{m} \right)^m \right]. \quad (2.17)$$

O espaço paramétrico para  $p$  possui um limitante superior que depende da probabilidade de ocorrência de zero da distribuição SP. A Figura 1 apresenta graficamente o limite de variação do comportamento do parâmetro  $p$  para diferentes valores de  $\mu$  considerando algumas

distribuições pertencentes à família SPZM. Podemos observar que o valor do limite superior do parâmetro  $p$  diminui conforme aumenta o valor do parâmetro  $\mu$ . As regiões indicadas por \ \ \ \ \ indicam valores de  $p$  em que se tem a distribuição SPZI e as regiões em \ \ \ \ \ indicam valores de  $p$  em que se tem a distribuição SPZD.

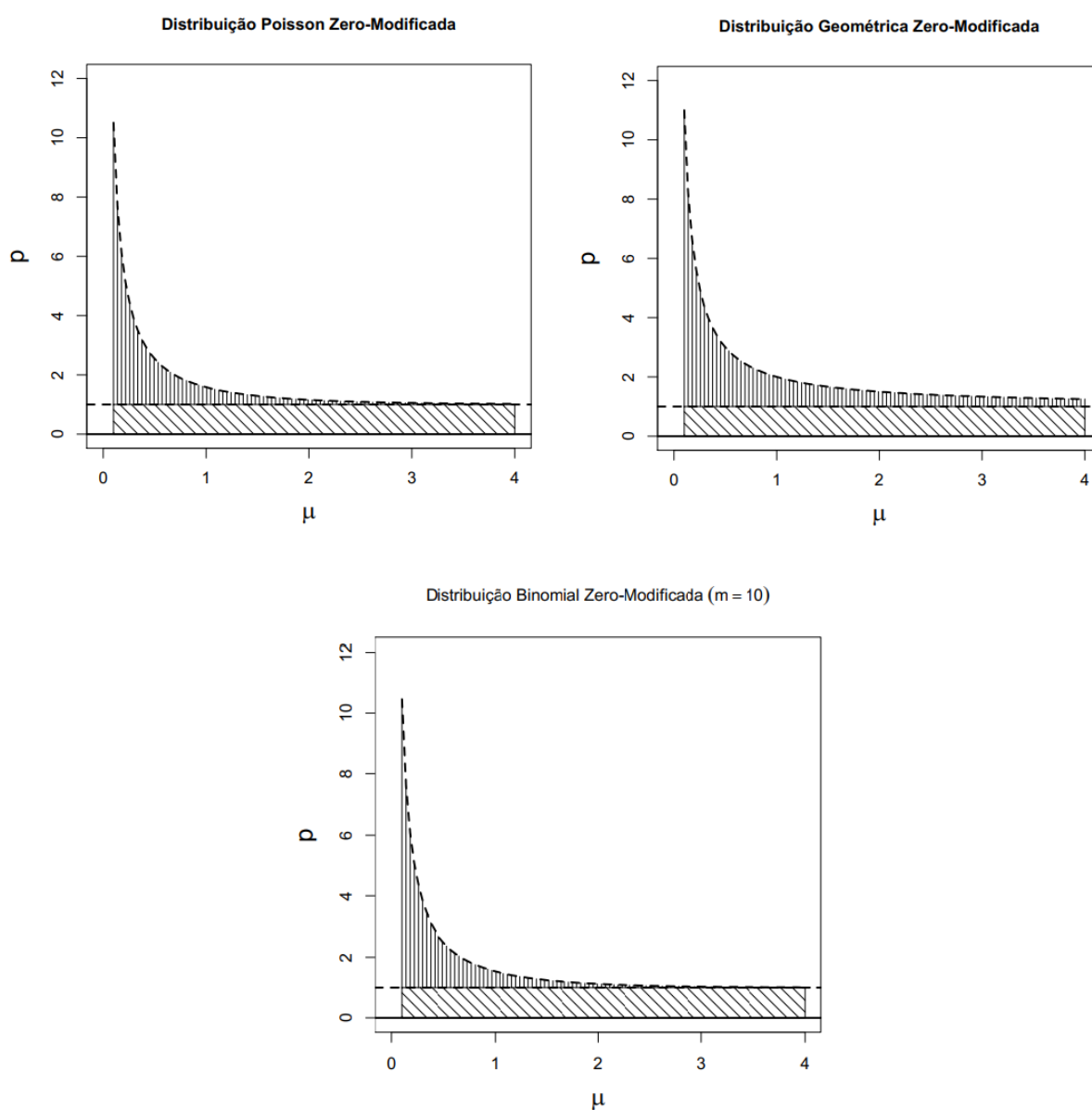


Figura 1 – Espaço paramétrico de  $p$  em função do parâmetro  $\mu$  para algumas distribuições SPZM.

Fonte: Adaptada de [Conceição \(2013\)](#).

## 2.5 Distribuição Série de Potência Zero-Modificada na versão hurdle

A reparametrização em  $p$ , foi considerado como sendo:

$$p = \frac{\omega}{1 - \pi_{SP}(0; \mu, \phi)}, \quad 0 \leq \omega \leq 1,$$

em que  $\pi_{SP}(0; \mu, \phi)$  é a distribuição SP no ponto zero. Logo, temos que:  $\omega = p(1 - \pi_{SP}(0; \mu, \phi))$ . Para maior detalhe, ver [Conceição \(2013\)](#), [Conceição et al. \(2017\)](#).

A partir da definição da distribuição SPZM em (2.2), podemos reescrever a distribuição SPZM como:

$$\pi_{SPZM}(y; \mu, \phi, \omega) = (1 - \omega)I(y) + \omega\pi_{SPZT}(y; \mu, \phi), \quad y \in A_t, \quad (2.18)$$

em que  $A_t$  é o suporte formado pelo subconjunto dos inteiros  $\{0, 1, \dots\}$  e  $\pi_{SPZT}(y; \mu, \phi)$  é a distribuição SPZT definida em (2.4).

Podemos interpretar  $\pi_{SPZM}(y; \mu, \phi, \omega)$  como uma distribuição mistura, na qual a probabilidade de ocorrência de zero ( $y = 0$ ) é  $1 - \omega$  e a probabilidade de ocorrência de uma observação ( $y > 0$ ) é  $\pi_{SPZT}(y; \mu, \phi)$ .

A média e a variância da variável aleatória  $Y$  com distribuição SPZM na versão hurdle são, respectivamente:

$$\mathbb{E}(Y) = \mu_{SPZM} = \frac{\omega\mu_{SP}}{1 - \pi_{SP}(0; \mu, \phi)} \quad \text{e} \quad \text{Var}(Y) = \sigma_{SPZM}^2 = \frac{\omega(\sigma_{SP}^2 + \mu_{SP}^2)}{1 - \pi_{SP}(0; \mu, \phi)} - (\omega\mu_{SPZT})^2,$$

em que  $\pi_{SP}(0; \mu, \phi)$  é a distribuição SP calculada no ponto zero.

A partir da definição em (2.3), podemos reescrever a respectiva função geradora de probabilidade para a distribuição SPZM na versão hurdle, como sendo:

$$G_Y(x) = 1 - \frac{\omega}{1 - \pi_{SP}(0; \mu, \phi)} [1 - G(x)], \quad (2.19)$$

em que  $G(x)$  é a função geradora de probabilidade da distribuição SP associada.

### 2.5.1 Casos Particulares

A partir da diferença no ponto zero da distribuição SPZM na versão hurdle e a distribuição SP, temos:

$$\pi_{SPZM}(0; \mu, \phi, \omega) - \pi_{SP}(0; \mu, \phi) = 1 - \omega - \pi_{SP}(0; \mu, \phi). \quad (2.20)$$

Assim, diferentes valores de  $\omega$  em (2.20) levam a diferentes modificações na frequência de zeros da distribuição SPZM na versão hurdle.

- i) Se  $\omega = 0$  em (2.20), temos que  $\pi_{SPZM}(0; \mu, \phi, \omega) = 1$ , então  $\pi_{SPZM}(y; \mu, \phi, \omega)$  é uma distribuição degenerada com toda massa no ponto zero.
- ii) Se  $0 < \omega < 1 - \pi_{SP}(0; \mu, \phi)$  em (2.20), temos que  $\pi_{SPZM}(0; \mu, \phi, \omega) - \pi_{SP}(0; \mu, \phi) > 0$ . Então  $\pi_{SPZM}(y; \mu, \phi, \omega)$  é a distribuição SPZI.
- iii) Se  $\omega = 1 - \pi_{SP}(0; \mu, \phi)$  em (2.20), temos que  $\pi_{SPZM}(0; \mu, \phi, \omega) - \pi_{SP}(0; \mu, \phi) = 0$ . Então  $\pi_{SPZM}(y; \mu, \phi, \omega)$  é a distribuição SP.
- iv) Se  $1 - \pi_{SP}(0; \mu, \phi) < \omega < 1$  em (2.20), temos que  $\pi_{SPZM}(0; \mu, \phi, \omega) - \pi_{SP}(0; \mu, \phi) < 0$ . Então  $\pi_{SPZM}(y; \mu, \phi, \omega)$  é a distribuição SPZD.
- v) Se  $\omega = 1$  em (2.20), temos que  $\pi_{SPZM}(0; \mu, \phi, \omega) = 0$ . Então  $\pi_{SPZM}(y; \mu, \phi, \omega)$  é uma distribuição SPZT.

Esta parametrização, tem a vantagem de que o  $\omega$  e o  $\mu$  são ortogonais, permitindo assim, estimar  $\omega$  independentemente de  $\mu$ . No entanto, a parametrização dada em (2.2), permite fazer inferências sobre o parâmetro  $p$ , o que permite identificar diretamente o tipo de modificação zero (inflacionada, deflacionada ou truncada).

## 2.6 Síntese

Neste Capítulo, apresentamos a família de distribuições SPZM pois são uma extensão da distribuição SP e tem como casos particulares as distribuições como SPZI, SP, SPZD e SPZT. Mostramos a flexibilidade de algumas distribuições SPZM para modelar dados de contagem, sem ter conhecimento prévio do comportamento da frequência dos zeros. Finalmente, foi considerado uma reparametrização em  $p$  e a distribuição resultante pode ser interpretada como uma distribuição de mistura.





---

# MODELOS DE SOBREVIVÊNCIA INDUZIDOS POR FRAGILIDADE DISCRETA SPZM

---

Neste Capítulo, serão desenvolvidos os Modelos de Sobrevivência induzidos por fragilidade discreta, utilizando as distribuições pertencentes à família SPZM, que resulta de nossa formulação. Particularmente, são estudados as situações em que a variável de fragilidade segue uma distribuição PZM, GZM e BZM. O Modelo de Fragilidade oferece uma forma conveniente de introduzir a heterogeneidade não observada das unidades em estudo.

## 3.1 Modelo de Fragilidade

O Modelo de Fragilidade é caracterizado pela inclusão de um efeito aleatório, quer dizer, de uma variável aleatória não observável, que representa as informações que não podem ou não foram observadas como fatores ambientais, genéticos ou informações que por algum motivo não foram considerados no planejamento do estudo. O termo de fragilidade foi introduzido por [Vaupel, Manton e Stallard \(1979\)](#). Uma forma para incorporar a variável de fragilidade, é introduzi-la na modelagem da função de risco, com o objetivo de controlar a heterogeneidade não observável das unidades em estudo, ou seja, uma mistura de indivíduos com diferentes riscos.

Assim, temos que a função de risco condicionado à fragilidade é definida por:

$$h(t|Z) = Zh_0(t), \quad (3.1)$$

em que  $Z$  representa a variável de fragilidade e  $h_0(t)$  é a função de risco comum a todos os indivíduos no estudo. O Modelo de Fragilidade em (3.1), assume estrutura de risco proporcional condicionado ao efeito aleatório (fragilidade).

Considere que  $z_i$  representa um valor da variável aleatória não observável  $Z$ , o risco individual cresce quando  $z_i > 1$ , decresce se  $z_i < 1$ , e para  $z_i = 1$  o Modelo de Fragilidade reduz-se ao Modelo de Risco Proporcional de Cox.

O fato da variável de fragilidade atuar de forma multiplicativa na função de risco, implica que quanto maior for o valor da variável de fragilidade, maior será a chance do indivíduo de apresentar o evento de interesse, dessa forma, quanto maior for  $z_i$ , mais frágil o indivíduo  $i$  está para falhar. Portanto, é esperado que o evento de interesse ocorra para os indivíduos mais frágeis (ANDREAS, 2010).

A função de sobrevivência definida como a probabilidade de um indivíduo sobreviver ao tempo  $t$ , condicionado a fragilidade é dado por:

$$S(t|Z) = P(T > t|Z) = e^{\{-Z \int_0^t h_0(s) ds\}} = S_0(t)^Z, \quad (3.2)$$

em que  $S_0(t)$  é a função de sobrevivência base.

Integrando  $S(t|Z)$  sobre  $Z$ , obtemos a função de sobrevivência marginal  $S(t)$ , e é dada por:

$$S(t) = \int_0^\infty S(t|Z)g(Z)dZ,$$

em que  $g(Z)$  é a função de densidade de probabilidade da variável de fragilidade.

Em geral, assumimos que as fragilidades são independentes e identicamente distribuídas para cada unidade, supondo que este termo de fragilidade segue uma distribuição de probabilidade, e a escolha de tal distribuição é um ponto importante a ser abordado. Devido ao fato desse termo de fragilidade atuar de maneira multiplicativa na função de risco, as candidatas à distribuição de fragilidade são não negativas, usualmente contínuas e não dependentes do tempo, tais como a distribuição Gama, Log-Normal, Gaussiana Inversa, Estável Positiva e Gama Generalizada.

Existe uma família generalizada de distribuições de fragilidade que inclui as distribuições Gama, Gaussiana Inversa e Estável Positiva que é a família de distribuições de Funções de Variação de Potência (FVP) sugeridas por Tweedie (1984) e derivadas independentemente por Hougaard (1986). Na literatura a distribuição Gama é a mais usada para modelar a fragilidade, esta preferência está relacionada ao fato de que a Gama é uma distribuição de tratamento algébrico conveniente.

Por outro lado, o uso de distribuições contínuas para a fragilidade não permitem ter fragilidade zero ou também conhecido como fração de cura. Na seguinte Seção, estudaremos o uso de distribuições discretas para a fragilidade.

## 3.2 Modelos de Fragilidade Discreta

De acordo com [Caroni, Crowder e Kimber \(2010\)](#), existem algumas situações nas quais é apropriado considerar a fragilidade distribuída discretamente, por exemplo, quando a heterogeneidade dos tempos de sobrevivência surge por causa da presença de um número aleatório de falhas por unidade ou pela causa da exposição a danos em um número aleatório de ocasiões.

Neste contexto, supomos que a variável de fragilidade  $Z$ , definida em (3.2), pode assumir valores inteiros não negativos  $\{0, 1, \dots\}$ . Seja a distribuição de probabilidade de  $Z$  especificada por  $P(Z = z) = q_z$ , em que  $z = 0, 1, \dots$ . A função de sobrevivência marginal é dada por:

$$\begin{aligned} S(t) &= \mathbb{E} \left\{ [S_0(t)]^Z \right\} \\ &= \sum_{z=0}^{\infty} P(Z = z) [S_0(t)]^z \\ &= \sum_{z=0}^{\infty} q_z [S_0(t)]^z \\ &= G_Z [S_0(t)], \end{aligned} \tag{3.3}$$

em que  $G_Z(\cdot)$  é a função geradora de probabilidade da variável aleatória  $Z$  e  $S_0(t)$  é a função de sobrevivência base ([RODRIGUES et al., 2009](#)).

O uso de uma distribuição discreta para a variável de fragilidade tem a característica de atribuir fragilidade zero a indivíduos que não estão em risco, o que corresponde a um modelo que contém uma proporção de indivíduos livres do evento de interesse, ou seja, risco zero, fragilidade zero ou fração de cura.

Desta forma, a função de sobrevivência não condicional definida em (3.3), pode ser reescrita como:

$$S(t) = q_0 + \sum_{z=1}^{\infty} q_z [S_0(t)]^z.$$

A fragilidade zero (fração de cura) é obtida quando:

$$\lim_{t \rightarrow \infty} S(t) = q_0.$$

Distribuições de fragilidade que permitem  $q_0 > 0$ , podem gerar unidades com fragilidade zero. Para tais unidades, o modelo de riscos proporcionais implica risco zero, isto é,  $S_0(t)^0 = 1, \forall t$ . Isso pode ser usado para descrever sobreviventes de longo prazo ou unidades que nunca falharão. No contexto médico, tais indivíduos são imunes ou curados da doença em questão ([ANDREAS, 2010](#)).

A distribuição da fragilidade permite relacionar a heterogeneidade não observada ao longo do tempo com o coeficiente de variação da distribuição da fragilidade ([ROCHA, 1995](#)).

Nesse sentido, é razoável avaliar o grau de heterogeneidade não observada em populações heterogêneas a partir das medidas de dispersão da distribuição de fragilidade. Um menor grau implica que as unidades pertencem a uma população mais homogênea.

Neste trabalho, desenvolvemos o uso de algumas distribuições pertencentes à família SPZM como possíveis candidatas para a fragilidade que resulta de nossa formulação. Particularmente, são estudados as situações em que a variável de fragilidade  $Z$  segue uma distribuição PZM, GZM e BZM.

### 3.2.1 Modelo de Fragilidade PZM

Vamos assumir que a variável de fragilidade tem distribuição PZM,  $Z \sim \text{PZM}(p, \mu)$ . A partir da definição em (3.3) e considerando a função geradora de probabilidade da distribuição PZM em (2.9), definimos a função de sobrevivência do Modelo de Fragilidade PZM como:

$$S(t) = 1 - p \left\{ 1 - e^{-\mu[1-S_0(t)]} \right\}. \quad (3.4)$$

A função densidade e função de risco são dadas respectivamente por

$$f(t) = p\mu f_0(t) e^{-\mu[1-S_0(t)]},$$

e

$$h(t) = \frac{p\mu f_0(t) e^{-\mu[1-S_0(t)]}}{1 - p [1 - e^{-\mu[1-S_0(t)]}]} \quad (3.5)$$

A fragilidade zero (fração de cura) é dada por

$$q_0 = 1 - p + pe^{-\mu}, \quad (3.6)$$

sendo que,  $f_0(t)$  é a função de densidade base e  $S_0(t)$  é a função de sobrevivência base.

Em Análise de Sobrevivência, observamos para cada indivíduo em estudo, o tempo  $t_i$  em que ocorre o evento de interesse e  $\delta_i$  é uma variável indicadora, tal que:

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

O mecanismo de censura pode ser informativa e não informativa. Expressamos que a censura é não informativa, se a ocorrência desta é independente do mecanismo que ocasiona a falha. A censura informativa ocorre quando observações censuradas são mais (ou menos) propensas a experimentar um evento específico das observações restantes no futuro (COLLETT, 2015). Neste trabalho consideramos que o mecanismo de censura é não informativa.

A função de verossimilhança com dados censurados, segundo Lawless (2003), assume a seguinte forma:

$$\mathcal{L}(\boldsymbol{\theta}; t_i, \delta_i) = \prod_{i=1}^n [h(t_i, \boldsymbol{\theta})]^{\delta_i} S(t_i, \boldsymbol{\theta}),$$

em que  $\boldsymbol{\theta}$  é o vetor de parâmetros a serem estimados.

Assim, a construção da função de máxima verossimilhança do Modelo de Fragilidade PZM é a partir da função de sobrevivência e risco, definidas em (3.4) e (3.5), respectivamente. Então a função de máxima verossimilhança é dada por:

$$\mathcal{L}(\boldsymbol{\mu}, p; t_i, \delta_i) = \prod_{i=1}^n \left\{ \frac{p\boldsymbol{\mu}f_0(t) e^{-\boldsymbol{\mu}[1-S_0(t)]}}{1-p[1-e^{-\boldsymbol{\mu}[1-S_0(t)]}]} \right\}^{\delta_i} \times \left\{ 1-p[1-e^{-\boldsymbol{\mu}[1-S_0(t)]}] \right\},$$

e a correspondente log-verossimilhança é:

$$\begin{aligned} l(\boldsymbol{\mu}, p; t_i, \delta_i) &= \log[\mathcal{L}(\boldsymbol{\mu}, p; t_i, \delta_i)] \\ &= \sum_{i=1}^n \delta_i \left[ \log \left( p\boldsymbol{\mu}f_0(t) e^{-\boldsymbol{\mu}[1-S_0(t)]} \right) - \log \left( 1-p+pe^{-\boldsymbol{\mu}[1-S_0(t)]} \right) \right] \\ &\quad + \sum_{i=1}^n \log \left( 1-p+pe^{-\boldsymbol{\mu}[1-S_0(t)]} \right). \end{aligned}$$

Note que, quando  $S_0(t)$  e  $f_0(t)$  são especificados, ou seja, definimos uma distribuição base para defini-los, com isso o número de parâmetros a serem estimados aumentará.

Considerando a reparametrização de  $p$ , vista em (2.18) e assumindo que a variável de fragilidade tem distribuição PZM,  $Z \sim \text{PZM}(\boldsymbol{\mu}, \boldsymbol{\omega})$ . A função de sobrevivência em (3.4) pode ser reescrita como

$$S(t) = 1 - \frac{\boldsymbol{\omega}}{1-e^{-\boldsymbol{\mu}}} \left\{ 1 - e^{-\boldsymbol{\mu}[1-S_0(t)]} \right\}, \quad (3.7)$$

A função densidade e risco são dadas respectivamente por

$$f(t) = \frac{\boldsymbol{\omega}\boldsymbol{\mu}f_0(t) e^{-\boldsymbol{\mu}[1-S_0(t)]}}{1-e^{-\boldsymbol{\mu}}},$$

e

$$h(t) = \frac{\boldsymbol{\omega}\boldsymbol{\mu}f_0(t) e^{-\boldsymbol{\mu}[1-S_0(t)]}}{1-e^{-\boldsymbol{\mu}} - \boldsymbol{\omega} + \boldsymbol{\omega}e^{-\boldsymbol{\mu}[1-S_0(t)]}}, \quad (3.8)$$

e a fragilidade zero (fração de cura) é:

$$q_0 = 1 - \boldsymbol{\omega},$$

em que,  $f_0(t)$  e  $S_0(t)$  são as funções de densidade e sobrevivência base, respectivamente.

A função de máxima verossimilhança do Modelo de Fragilidade PZM( $\mu, \omega$ ), é construída a partir da função de sobrevivência e risco, vistas em (3.7) e (3.8), respectivamente. Então a função de máxima verossimilhança é:

$$\mathcal{L}(\mu, \omega; t_i, \delta_i) = \prod_{i=1}^n \left\{ \frac{\omega \mu f_0(t) e^{-\mu[1-S_0(t)]}}{1 - e^{-\mu} - \omega + \omega e^{-\mu[1-S_0(t)]}} \right\}^{\delta_i} \times \left\{ 1 - \frac{\omega}{1 - e^{-\mu}} [1 - e^{-\mu[1-S_0(t)]}] \right\},$$

e a correspondente log-verossimilhança é:

$$\begin{aligned} l(\mu, \omega; t_i, \delta_i) &= \log[\mathcal{L}(\mu, \omega; t_i, \delta_i)] \\ &= \sum_{i=1}^n \delta_i \left\{ \log[\omega \mu f_0(t) e^{-\mu[1-S_0(t)]}] - \log[1 - e^{-\mu} - \omega + \omega e^{-\mu[1-S_0(t)]}] \right\} \\ &\quad + \sum_{i=1}^n \left\{ \log[1 - e^{-\mu} - \omega + \omega e^{-\mu[1-S_0(t)]}] - \log[1 - e^{-\mu}] \right\}. \end{aligned}$$

### 3.2.2 Modelo de Fragilidade GZM

Vamos assumir que a variável de fragilidade tem distribuição GZM,  $Z \sim \text{GZM}(p, \mu)$ . A partir da definição em (3.3) e considerando a função geradora de probabilidade da distribuição GZM em (2.13), definimos a função de sobrevivência do Modelo de Fragilidade GZM como:

$$S(t) = 1 - p \left[ 1 - \frac{1}{1 + \mu - \mu S_0(t)} \right]. \quad (3.9)$$

A função densidade e risco são dadas por:

$$f(t) = \frac{p\mu f_0(t)}{[1 + \mu - \mu S_0(t)]^2},$$

e

$$h(t) = \frac{p\mu f_0(t)}{[1 + \mu - \mu S_0(t)] \{(1-p)[1 + \mu - \mu S_0(t)] + p\}}, \quad (3.10)$$

e a fragilidade zero (fração de cura) é:

$$q_0 = 1 - p + \frac{p}{1 + \mu},$$

sendo que,  $f_0(t)$  é a função de densidade base e  $S_0(t)$  é a função de sobrevivência base.

Para a construção da função de máxima verossimilhança do Modelo de Fragilidade GZM, precisamos da função de sobrevivência e risco, dadas em (3.9) e (3.10), respectivamente. Então a função de verossimilhança é definida por:

$$\begin{aligned} \mathcal{L}(\mu, p; t_i, \delta_i) &= \prod_{i=1}^n \left\{ \frac{p\mu f_0(t)}{[1 + \mu - \mu S_0(t)] [(1-p)(1 + \mu - \mu S_0(t)) + p]} \right\}^{\delta_i} \\ &\quad \times \left\{ 1 - p \left[ 1 - \frac{1}{1 + \mu - \mu S_0(t)} \right] \right\}, \end{aligned}$$

e o respectivo log-verossimilhança é:

$$\begin{aligned}
l(\mu, p; t_i, \delta_i) &= \log [\mathcal{L}(\mu, p; t_i, \delta_i)] \\
&= \sum_{i=1}^n \delta_i \{ \log [p\mu f_0(t)] - \log [1 + \mu - \mu S_0(t)] \} \\
&\quad - \sum_{i=1}^n \delta_i \{ \log [(1-p)(1 + \mu - \mu S_0(t)) + p] \} \\
&\quad + \sum_{i=1}^n \{ \log [(1-p)(1 + \mu - \mu S_0(t)) + p] - \log [1 + \mu - \mu S_0(t)] \}.
\end{aligned}$$

Considerando a reparametrização de  $p$ , vista em (2.18) e assumindo que a variável de fragilidade tem distribuição GZM,  $Z \sim \text{GZM}(\mu, \omega)$ . A função de sobrevivência em (3.9) pode ser reescrita como:

$$S(t) = 1 - \frac{\omega(1+\mu)}{\mu} \left[ 1 - \frac{1}{1 + \mu - \mu S_0(t)} \right]. \quad (3.11)$$

A função de densidade e risco é dada por

$$f(t) = \frac{\omega(1+\mu)f_0(t)}{[1 + \mu - \mu S_0(t)]^2},$$

e

$$h(t) = \frac{\mu\omega(1+\mu)f_0(t)}{[1 + \mu - \mu S_0(t)] \{ [\mu - \omega(1+\mu)][1 + \mu - \mu S_0(t)] + \omega(1+\mu) \}}, \quad (3.12)$$

e a fragilidade zero (fração de cura) é:

$$q_0 = 1 - \omega,$$

em que,  $f_0(t)$  e  $S_0(t)$  são a função de densidade base e função de sobrevivência base, respectivamente.

A função de máxima verossimilhança do Modelo de Fragilidade GZM( $\mu, \omega$ ), é construída a partir da função de sobrevivência e risco, vistas em (3.11) e (3.12), respectivamente. A função de máxima verossimilhança é:

$$\begin{aligned}
\mathcal{L}(\mu, \omega; t_i, \delta_i) &= \prod_{i=1}^n \left\{ \frac{\mu\omega(1+\mu)f_0(t)}{[1 + \mu - \mu S_0(t)] \{ [\mu - \omega(1+\mu)][1 + \mu - \mu S_0(t)] + \omega(1+\mu) \}} \right\}^{\delta_i} \\
&\quad \times \left\{ 1 - \frac{\omega(1+\mu)}{\mu} \left[ 1 - \frac{1}{1 + \mu - \mu S_0(t)} \right] \right\},
\end{aligned}$$

e a correspondente log-verossimilhança é:

$$\begin{aligned}
l(\mu, \omega; t_i, \delta_i) &= \log [\mathcal{L}(\mu, \omega; t_i, \delta_i)] \\
&= \sum_{i=1}^n \delta_i \{ \log [\mu \omega (1 + \mu) f_0(t)] - \log [1 + \mu - \mu S_0(t)] \} \\
&\quad - \sum_{i=1}^n \delta_i \{ \log [(\mu - \omega(1 + \mu))(1 + \mu - \mu S_0(t)) + \omega(1 + \mu)] \} \\
&\quad + \sum_{i=1}^n \log [(\mu - \omega(1 + \mu))(1 + \mu - \mu S_0(t)) + \omega(1 + \mu)] \\
&\quad - \sum_{i=1}^n \log [\mu] - \log [1 + \mu - \mu S_0(t)].
\end{aligned}$$

### 3.2.3 Modelo de Fragilidade BZM

Vamos assumir que a variável de fragilidade tem distribuição BZM,  $Z \sim \text{BZM}(p, \mu)$ . A partir da definição em (3.3) e considerando a função geradora de probabilidade da distribuição BZM em (2.17), definimos a função de sobrevivência do Modelo de Fragilidade BZM como:

$$S(t) = 1 - p \left[ 1 - \left( \frac{m - \mu + \mu S_0(t)}{m} \right)^m \right]. \quad (3.13)$$

A função de densidade e risco são dadas por

$$f(t) = \frac{p\mu f_0(t) [m - \mu + \mu S_0(t)]^{m-1}}{m^m},$$

e

$$h(t) = \frac{p\mu f_0(t) [m - \mu + \mu S_0(t)]^{m-1}}{(1-p)m^m + p[m - \mu + \mu S_0(t)]^m}. \quad (3.14)$$

A fragilidade zero (fração de cura) é dada por

$$q_0 = 1 - p + p \left( \frac{m - \mu}{m} \right)^m,$$

sendo que,  $f_0(t)$  é a função de densidade base e  $S_0(t)$  é a função de sobrevivência base.

Para a construção da função de máxima verossimilhança do Modelo de Fragilidade BZM, precisamos da função de sobrevivência e risco, dadas em (3.13) e (3.14) respectivamente. Então a função de verossimilhança é definida por:

$$\mathcal{L}(\mu, p; t_i, \delta_i) = \prod_{i=1}^n \left\{ \frac{p\mu f_0(t) [m - \mu + \mu S_0(t)]^{m-1}}{(1-p)m^m + p[m - \mu + \mu S_0(t)]^m} \right\}^{\delta_i} \times \left\{ 1 - p \left[ 1 - \left( \frac{m - \mu + \mu S_0(t)}{m} \right)^m \right] \right\},$$



e o respectivo log-verossimilhança é:

$$\begin{aligned}
l(\mu, p; t_i, \delta_i) &= \log [\mathcal{L}(\mu, p; t_i, \delta_i)] \\
&= \sum_{i=1}^n \delta_i \{ (m-1) \log [p\mu f_0(t) (m-\mu + \mu S_0(t))] \} \\
&\quad - \sum_{i=1}^n \delta_i \{ \log [(1-p)m^m + p(m-\mu + \mu S_0(t))^m] \} \\
&\quad + \sum_{i=1}^n \log [(1-p)m^m + p(m-\mu + \mu S_0(t))^m] - m \log [m].
\end{aligned}$$

Considerando a reparametrização de  $p$ , vista em (2.18) e assumindo que a variável de fragilidade tem distribuição BZM,  $Z \sim \text{BZM}(\mu, \omega)$ . A função de sobrevivência em (3.13) pode ser reescrita como:

$$S(t) = 1 - \frac{\omega m^m}{m^m - (m-\mu)^m} \left[ 1 - \left( \frac{m-\mu + \mu S_0(t)}{m} \right)^m \right]. \quad (3.15)$$

A função de densidade e risco respectivamente são dadas por:

$$f(t) = \frac{\omega \mu f_0(t) [m-\mu + \mu S_0(t)]^{m-1}}{m^m - (m-\mu)^m},$$

e

$$h(t) = \frac{\omega \mu f_0(t) [m-\mu + \mu S_0(t)]^{m-1}}{m^m (1-\omega) - (m-\mu)^m + \omega [m-\mu + \mu S_0(t)]^m}. \quad (3.16)$$

A fragilidade zero (fração de cura) é obtida como

$$q_0 = 1 - \omega,$$

em que,  $f_0(t)$  e  $S_0(t)$  são a função densidade base e a função de sobrevivência base, respectivamente.

A função de máxima verossimilhança do Modelo de Fragilidade BZM( $\mu, \omega$ ), é construída a partir da função de sobrevivência e risco, vistas em (3.15) e (3.16), respectivamente. Então a função de máxima verossimilhança é:

$$\begin{aligned}
\mathcal{L}(\mu, \omega; t_i, \delta_i) &= \prod_{i=1}^n \left\{ \frac{\omega \mu f_0(t) [m-\mu + \mu S_0(t)]^{m-1}}{m^m (1-\omega) - (m-\mu)^m + \omega [m-\mu + \mu S_0(t)]^m} \right\}^{\delta_i} \\
&\quad \times \left\{ 1 - \frac{\omega m^m}{m^m - (m-\mu)^m} \left[ 1 - \left( \frac{m-\mu + \mu S_0(t)}{m} \right)^m \right] \right\},
\end{aligned}$$

e a correspondente log-verossimilhança é:

$$\begin{aligned}
l(\boldsymbol{\mu}, \boldsymbol{\omega}; t_i, \boldsymbol{\delta}_i) &= \log [\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\omega}; t_i, \boldsymbol{\delta}_i)] \\
&= \sum_{i=1}^n \delta_i \{ (m-1) \log [\omega m \mu f_0(t) (m - \mu + \mu S_0(t))] \} \\
&\quad - \sum_{i=1}^n \delta_i \{ \log [m^m (1 - \omega) - (m - \mu)^m + \omega (m - \mu + \mu S_0(t))^m] \} \\
&\quad + \sum_{i=1}^n \{ \log [m^m (1 - \omega) - (m - \mu)^m + \omega (m - \mu + \mu S_0(t))^m] - \log [m^m - (m - \mu)^m] \}.
\end{aligned}$$

Até agora, desenvolvemos as diferentes funções de sobrevivência, densidade e risco para os diferentes Modelos de Fragilidade SPZM. Este desenvolvimento foi feito sem a inclusão de covariáveis mas em uma aplicação real são estudadas e medidas diferentes covariáveis que influenciam ao estudo. Na seguinte Seção, serão estendidos os diferentes Modelos de Fragilidade com a presença de covariáveis.

### 3.3 Modelos de Fragilidade SPZM com covariáveis

A inclusão de covariáveis em Modelos de Fragilidade discreto, geralmente, é realizada por meio de uma função de ligação, que vai associar as covariáveis estudadas com um dos parâmetros de interesse, sendo que, a função de ligação dependerá do espaço paramétrico do parâmetro. As ligações usadas podem ser: logarítmica, logística, identidade, probit, entre outras.

Neste trabalho, mostramos dois cenários para a família de distribuições SPZM, o primeiro quando  $Z \sim \text{SPZM}(\mu, p)$ , vista em (2.2). Faremos a ligação com o parâmetro  $\mu$  e a função de ligação usada é a logarítmica, desta forma, temos:

$$\mu_i = e^{\mathbf{x}_i \boldsymbol{\beta}}, \quad i = 1, \dots, n, \quad (3.17)$$

o segundo cenário é  $Z \sim \text{SPZMR}(\mu, \omega)$ , vista em (2.18). A inclusão das covariáveis será feita com o parâmetro  $\omega$  por meio da ligação logito. Assim, temos:

$$\omega_i = \frac{e^{\mathbf{x}_i \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i \boldsymbol{\beta}}}, \quad i = 1, \dots, n, \quad (3.18)$$

sendo que,  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$  é o vetor de covariáveis estudadas para o  $i$ -ésimo indivíduo e  $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)'$  é o vetor de parâmetros a serem estimados.

No segundo cenário, fizemos a ligação com o parâmetro  $\omega$  pois as respectivas frações de cura dos Modelos de Fragilidade PZM, GZM e BZM na versão hurdle, estão em função deste parâmetro.

Analogamente, para a obtenção das respectivas funções de sobrevivência, densidade, risco e fração de cura com a presença de covariáveis, só basta substituir adequadamente (3.17) e (3.18) nas expressões correspondentes dos Modelos de Fragilidade PZM, GZM e BZM.

Para se obter as estimativas de todos os parâmetros envolvidos, será necessário maximizar a função de verossimilhança. As vezes torna-se impossível encontrar seu ponto de máximo analiticamente devido a complexidade, mas com a utilização de métodos numéricos podemos obter tais valores. Para tal finalidade, neste trabalho, foi usado o método “BFGS” por meio da rotina “optim” do software (R Core Team, 2019).

### 3.4 Estudo de Simulação

Nesta seção, apresentamos dois estudos de simulação a partir de conjuntos de dados gerados para analisar o comportamento dos EMV dos parâmetros envolvidos no Modelo de Fragilidade PZM com covariáveis, vista na (Seção 3.3). O objetivo é avaliar o desempenho do método de máxima verossimilhança na estimação dos parâmetros e suas propriedades assintóticas.

Analizamos dois estudos, cada um considerando diferentes situações na frequência de zero. O primeiro estudo é caracterizado como zero-inflação pois ( $0 < p < 1$ ) e o segundo, identificado como zero-deflação pois ( $1 < p < \frac{1}{1-e^{-\mu}}$ ). Para cada situação consideramos 1000 repetições de diferentes tamanhos  $n = 50, 100, 300, 500$  e  $1000$ , respectivamente.

Para o primeiro estudo, os dados dos tempos de falha foram simulados com  $p = 0,5$ , introduzimos dois parâmetros de regressão em  $\mu_i = e^{\beta_0 + \beta_1 x_i}$ , sendo que  $\beta_0 = 0,4$  é o intercepto e  $\beta_1 = 1$  é o parâmetro associado à covariável. Consideramos a distribuição Exponencial para a função de risco base com parâmetro  $\lambda = 0,2$ .

O segundo estudo, correspondente à deflação, os dados dos tempos foram simulados com  $p = 1,3$ , inserimos dois parâmetros de regressão a partir da função de ligação logarítmica em  $\mu_i$  e fixamos  $\beta_0 = -1$  e  $\beta_1 = 0,5$ . Da mesma forma, consideramos a distribuição Exponencial como a função de risco base com  $\lambda = 0,4$ .

É importante resaltar que a censura gerada é maior que a fração de cura.

O seguinte algoritmo foi usado para gerar dados do Modelo de Fragilidade PZM com covariáveis, nos dois estudos:

1. Gerar  $\mathbf{x}_i \sim \text{Binomial}(1, 0,5)$  e  $\eta_i \sim \text{Uniforme}(0, 1)$ ;
2. Gerar  $q_{0i} = 1 - p + pe^{-\mu_i}$ , sendo que  $\mu_i = e^{\beta_0 + \beta_1 x_i}$ ;
3. Se  $\eta_i < q_{0i}$ ,  $t_i = \infty$ , caso contrario:

$$t_i = \frac{1}{\lambda} \left\{ -\log \left[ 1 + \log \left( 1 - \frac{\eta_i}{p} \right) \times \frac{1}{\mu_i} \right] \right\};$$

4. Gerar  $c_i \sim \text{Exp}(v)$ ,  $v = \frac{e^{0,1}(q_{0i} - q_{1i})}{1 - (q_{0i} - q_{1i})}$ , em que  $q_{1i} = q_{0i} + 1$ ;
5. Fazer  $\omega_i = \min \{t_i, c_i\}$ ;

6. Se  $t_i < c_i$ , então  $\delta_i = 1$ , caso contrario  $\delta_i = 0$ , para  $i = 1, \dots, n$ .

Para realizar os respectivos estudos de simulação usamos o método “CG” por meio da rotina “mle” do software (R Core Team, 2019).

As propriedades do método de estimação proposto são avaliadas em termos da Média dos Estimadores de Máxima Verossimilhança (MEMV), Raiz do Erro Quadrático Médio (REQM), Desvio Padrão (DP) e Probabilidades de Cobertura (PC) dos intervalos de confiança de 90% e 95% para os parâmetros no modelo. Os resultados das simulações para cada estudo, Zero-Inflação e Zero-Deflação, estão apresentados na Tabela 3 e 5, respectivamente.

Na Tabela 3 exibimos os resultados do primeiro estudo de simulação para os quatro parâmetros envolvidos. As respectivas MEMV destes parâmetros estão próximas dos valores fixados a medida que o tamanho da amostra aumenta. Também observamos que REQM e DP diminuem e tendem para zero quando aumenta o tamanho da amostra. A PC(90%) e PC(95%) para os parâmetros envolvidos, em algumas situações os resultados obtidos não estão ficando próximos do nível nominal fixado. A melhoria da MEMV, REQM, DP e PC acontece a partir da amostra de tamanho 300. Na Figura 2 observamos graficamente os resultados obtidos do primeiro estudo da simulação e os resultados da PC(90%) e PC(95%) apresentam que não todos os resultados estão dentro do intervalo de confiança correspondente. Portanto, depois de fazer o primeiro estudo de simulação (Zero-Inflação) concluímos que o comportamento dos parâmetros em estudo, tendem a ser assintoticamente normais e na Figura 3 mostramos gráficamente este comportamento assintótico normal para o tamanho da amostra  $n = 1000$ .

Para este primeiro estudo fixamos os parâmetros como sendo  $p = 0,5$ ,  $\beta_0 = 0,4$  e  $\beta_1 = 1$  e consideramos uma covariável dicotômica no estudo de simulação. A partir de esta informação, podemos obter as respectivas frações de cura e heterogeneidade não observada (variância) para cada grupo. Assim, temos que para os pacientes pertencentes ao grupo 1 ( $\mathbf{x}_1 = 0$ ), a respectiva fração de cura é  $q_1 = 0,612$  e a heterogeneidade não observada é  $\theta_1 = 1,302$ . Para os pacientes referentes ao grupo 2 ( $\mathbf{x}_1 = 1$ ), a fração de cura obtida é  $q_2 = 0,509$  e a heterogeneidade não observada é  $\theta_2 = 6,139$ . As respectivas frações de cura e heterogeneidade não observada foram calculadas por meio de (3.6) e (2.8), respectivamente.

Na Tabela 4 apresentamos os resultados para as respectivas frações de cura e heterogeneidade não observada (variância) captadas para cada grupo. Vemos que a medida que o tamanho da amostra aumenta a MEMV estão próximas dos valores calculados e observamos que REQM e DP vão diminuindo. A PC(90%) e PC(95%) em algumas situações se mostra que os resultados da simulação não estão ficando próximos do nível nominal fixado.

Apresentamos os respectivos resultados do segundo estudo de simulação na Tabela 5. Apreciamos que a medida que o tamanho da amostra aumenta as MEMV dos parâmetros estão próximas dos valores fixados. Vemos que REQM e DP diminuem e tendem para zero ao aumentar o tamanho da amostra. Os resultados para a PC(90%) e PC(95%) mostram que não estão

ficando próximos do nível nominal fixado para todos os parâmetros abrangidos. A melhoria das MEMV, REQM e DP acontece a partir da amostra de tamanho 300. Na Figura 4 contemplamos graficamente os respectivos resultados do segundo estudo de simulação. Consequentemente, logo de executar o segundo estudo de simulação (Zero-Deflação) induzimos que os parâmetros envolvidos no estudo tendem a ter um comportamento assintoticamente normal. Na Figura 5 observamos este comportamento dos parâmetros envolvidos para o tamanho de amostra  $n = 1000$ .

No segundo estudo, fixamos os valores dos parâmetros da seguinte forma:  $p = 1,3$ ,  $\beta_0 = -1$ ,  $\beta_1 = 0,5$ . Com esta informação, para cada grupo determinamos as frações de cura e a respectiva heterogeneidade não observada (variância). Desta forma, a fração de cura para os pacientes do grupo 1 ( $\mathbf{x}_1 = 0$ ) é  $q_1 = 0,600$  e a correspondente heterogeneidade não observada é  $\theta_1 = 0,425$ . A fração de cura para os pacientes do grupo 2 ( $\mathbf{x}_1 = 1$ ) é  $q_2 = 0,409$  e a heterogeneidade não observada obtida é  $\theta_2 = 0,645$ .

Na Tabela 6 mostramos as frações de cura e heterogeneidade não observada (variância) percebidas para os dois grupos. Observamos que as MEMV estão próximas dos respectivos valores obtidos quando o tamanho da amostra aumenta. A REQM e DP tendem para zero a medida que o tamanho da amostra aumenta. A PC(90%) e PC(95%) mostra que para alguns casos os resultados da simulação não estão próximos do nível nominal fixado.

Tabela 3 – Média dos Estimadores de Máxima Verossimilhança (MEMV), Raiz do Erro Quadrático Médio (REQM), Desvio Padrão (DP) e Probabilidade de Cobertura (PC) do Modelo de Fragilidade PZM com covariáveis, sendo,  $p = 0,5$ ,  $\lambda = 0,2$ ,  $\beta_0 = 0,4$  e  $\beta_1 = 1$  (Zero-Inflação).

n	Parâmetros	MEMV	REQM	DP	PC (90%)	PC (95%)
50	$p$	0,718	0,514	0,466	0,924	0,956
	$\lambda$	0,254	0,165	0,156	0,888	0,936
	$\beta_0$	0,037	0,965	0,894	0,935	0,969
	$\beta_1$	1,063	0,627	0,624	0,854	0,923
100	$p$	0,564	0,213	0,203	0,937	0,971
	$\lambda$	0,218	0,104	0,102	0,898	0,954
	$\beta_0$	0,285	0,612	0,601	0,923	0,962
	$\beta_1$	1,035	0,453	0,451	0,884	0,931
300	$p$	0,510	0,053	0,052	0,910	0,950
	$\lambda$	0,204	0,050	0,049	0,928	0,971
	$\beta_0$	0,380	0,313	0,312	0,925	0,961
	$\beta_1$	1,011	0,235	0,235	0,925	0,961
500	$p$	0,503	0,037	0,037	0,893	0,942
	$\lambda$	0,201	0,037	0,037	0,931	0,968
	$\beta_0$	0,410	0,252	0,251	0,915	0,956
	$\beta_1$	0,993	0,194	0,194	0,914	0,955
1000	$p$	0,502	0,024	0,024	0,917	0,959
	$\lambda$	0,201	0,026	0,026	0,900	0,959
	$\beta_0$	0,405	0,174	0,174	0,933	0,967
	$\beta_1$	0,998	0,130	0,130	0,925	0,967

Fonte: Elaborada pelo autor.

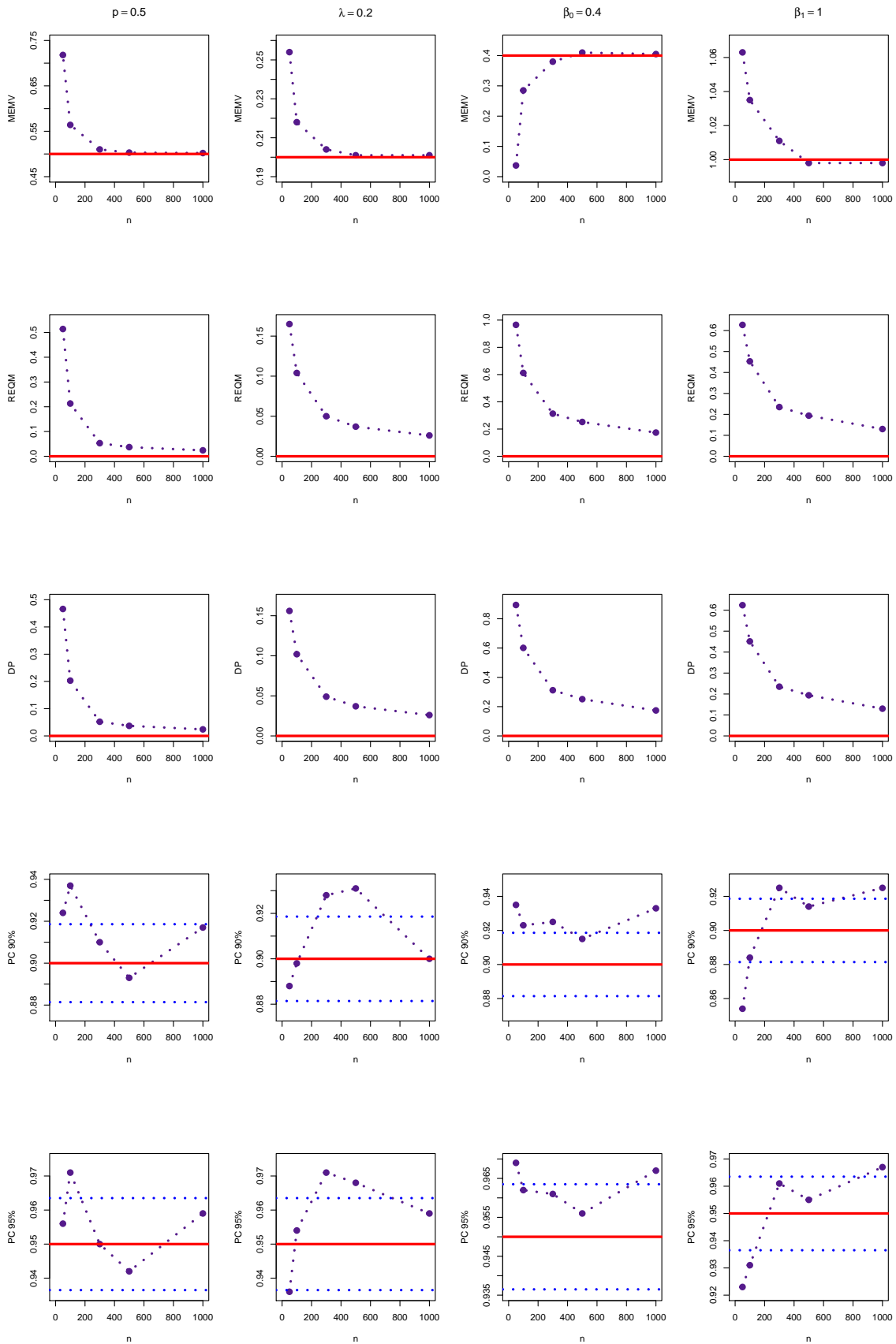


Figura 2 – Média dos Estimadores de Máxima Verossimilhança (MEMV), Raiz dos Erros Quadráticos Médios (REQM), Desvio Padrão (DP), Probabilidades de Cobertura (PC) de 90% e 95% dos parâmetros versus o tamanho da amostra para dados simulados do Modelo de Fragilidade PZM com covariáveis (Zero-Inflação).

Fonte: Elaborada pelo autor.

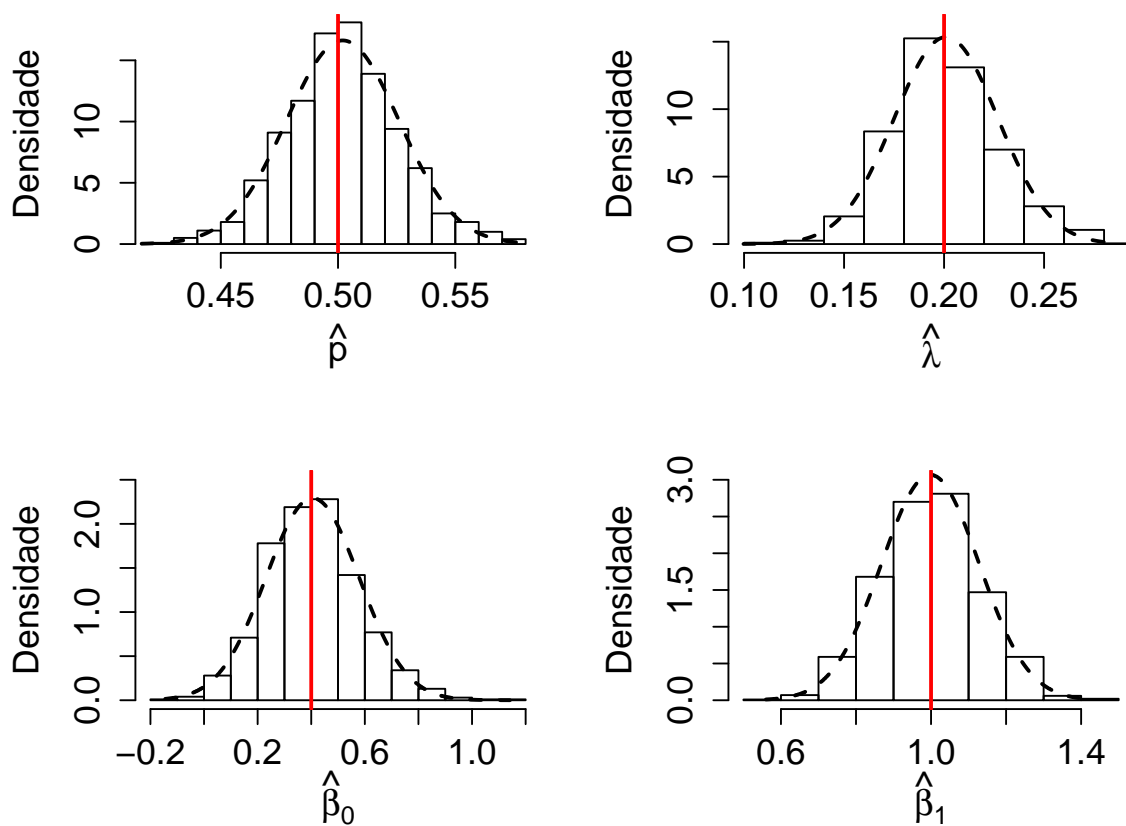


Figura 3 – Comportamento assintótico dos parâmetros envolvidos do Modelo de Fragilidade PZM com covariáveis para o primeiro estudo de simulação (Zero-Inflação).

Fonte: Elaborada pelo autor.

Tabela 4 – Média dos Estimadores de Máxima Verossimilhança (MEMV), Raiz do Erro Quadrático Médio (REQM), Desvio Padrão (DP) e Probabilidade de Cobertura (PC) da fração de cura e heterogeneidade não observada captada pelo Modelo de Fragilidade PZM com covariáveis sendo  $q_1 = 0,612$ ,  $q_2 = 0,509$ ,  $\theta_1 = 1,302$  e  $\theta_2 = 6,139$  (Zero-Inflação).

n	Parâmetros	MEMV	REQM	DP	PC (90%)	PC (95%)
50	$q_1$	0,627	2,786	0,091	0,882	0,928
	$q_2$	0,489	2,859	0,098	0,868	0,928
	$\theta_1$	1,688	3,282	2,371	0,707	0,737
	$\theta_2$	11,893	24,259	22,247	0,743	0,767
100	$q_1$	0,620	2,788	0,064	0,903	0,945
	$q_2$	0,502	2,853	0,065	0,906	0,954
	$\theta_1$	1,637	2,831	1,585	0,804	0,846
	$\theta_2$	9,070	12,092	9,717	0,857	0,876
300	$q_1$	0,616	2,791	0,036	0,916	0,963
	$q_2$	0,505	2,852	0,040	0,895	0,946
	$\theta_1$	1,422	2,526	0,776	0,857	0,891
	$\theta_2$	6,859	6,316	3,478	0,886	0,927
500	$q_1$	0,613	2,790	0,029	0,916	0,961
	$q_2$	0,509	2,849	0,032	0,892	0,941
	$\theta_1$	1,416	2,513	0,525	0,877	0,908
	$\theta_2$	6,611	5,640	2,466	0,909	0,945
1000	$q_1$	0,612	2,791	0,021	0,895	0,947
	$q_2$	0,508	2,850	0,021	0,919	0,957
	$\theta_1$	1,357	2,486	0,341	0,916	0,942
	$\theta_2$	6,380	5,094	1,644	0,904	0,939

Fonte: Elaborada pelo autor.



Tabela 5 – Média dos Estimadores de Máxima Verossimilhança (MEMV), Raíz do Erro Quadrático Médio (REQM), Desvio Padrão (DP) e Probabilidade de Cobertura (PC) do Modelo de Fragilidade PZM com covariáveis sendo  $p = 1, 3, \lambda = 0, 4, \beta_0 = -1$  e  $\beta_1 = 0, 5$  (Zero-Deflação).

n	Parâmetros	MEMV	REQM	DP	PC (90%)	PC (95%)
50	$p$	1,662	0,875	0,797	0,815	0,824
	$\lambda$	0,401	0,123	0,123	0,916	0,948
	$\beta_0$	-1,130	0,784	0,773	0,926	0,966
	$\beta_1$	0,640	0,499	0,480	0,899	0,957
100	$p$	1,522	0,704	0,669	0,815	0,826
	$\lambda$	0,393	0,088	0,088	0,921	0,953
	$\beta_0$	-1,031	0,661	0,660	0,902	0,934
	$\beta_1$	0,565	0,356	0,351	0,896	0,948
300	$p$	1,422	0,510	0,495	0,863	0,885
	$\lambda$	0,395	0,053	0,053	0,937	0,964
	$\beta_0$	-1,013	0,471	0,471	0,914	0,932
	$\beta_1$	0,530	0,181	0,178	0,928	0,970
500	$p$	1,347	0,395	0,392	0,863	0,889
	$\lambda$	0,395	0,042	0,041	0,939	0,969
	$\beta_0$	-0,972	0,392	0,392	0,913	0,930
	$\beta_1$	0,517	0,144	0,143	0,916	0,953
1000	$p$	1,310	0,215	0,215	0,958	0,965
	$\lambda$	0,398	0,026	0,026	0,965	0,980
	$\beta_0$	-0,988	0,217	0,217	0,969	0,976
	$\beta_1$	0,510	0,091	0,090	0,951	0,978

Fonte: Elaborada pelo autor.

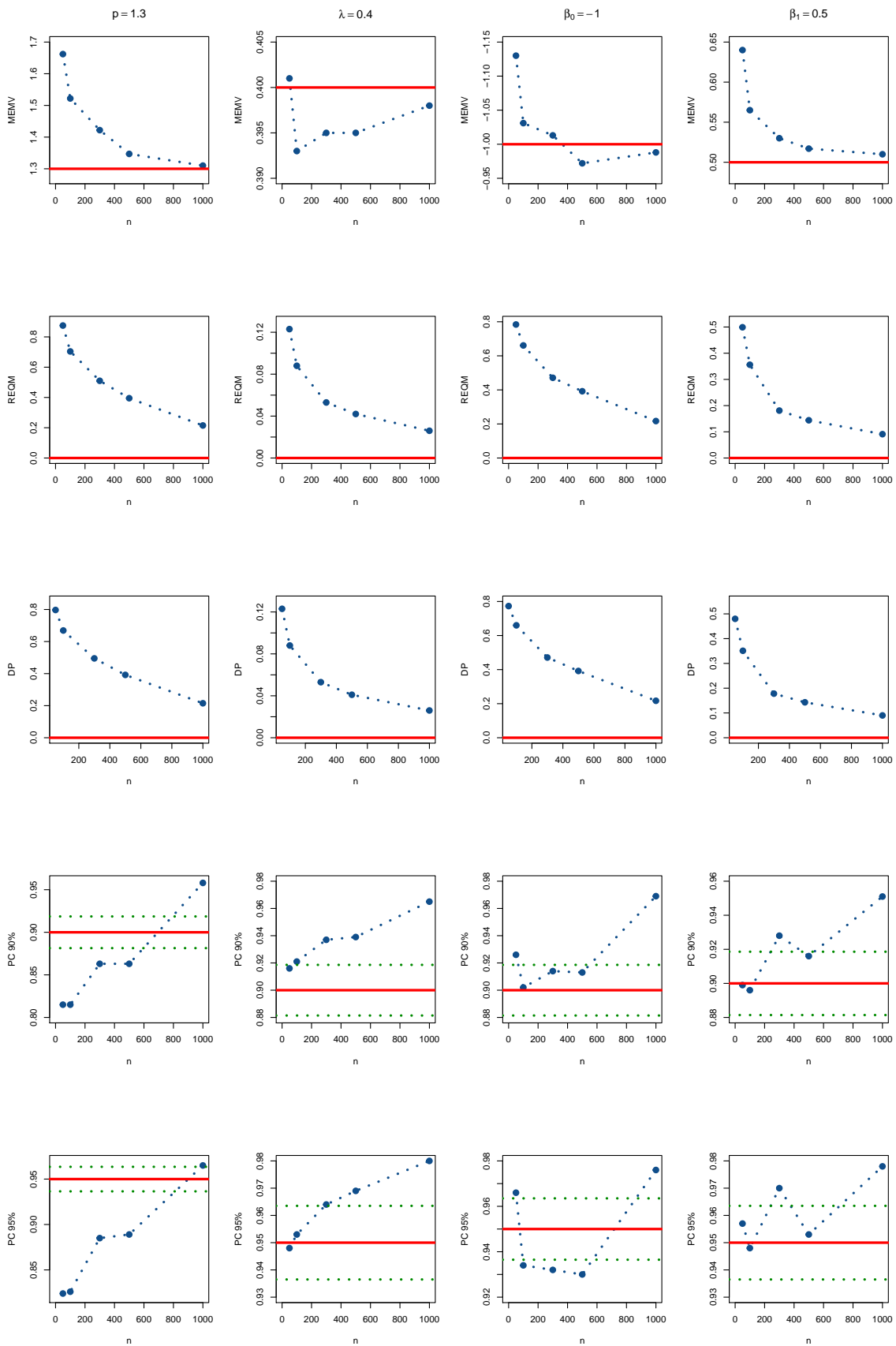


Figura 4 – Média dos Estimadores de Máxima Verossimilhança (MEMV), Raiz dos Erros Quadráticos Médios (REQM), Desvio Padrão (DP), Probabilidade de Cobertura (PC) de 90% e 95% dos parâmetros versus o tamanho da amostra para dados simulados do Modelo de Fragilidade PZM com covariáveis (Zero-Deflação).

Fonte: Elaborada pelo autor.

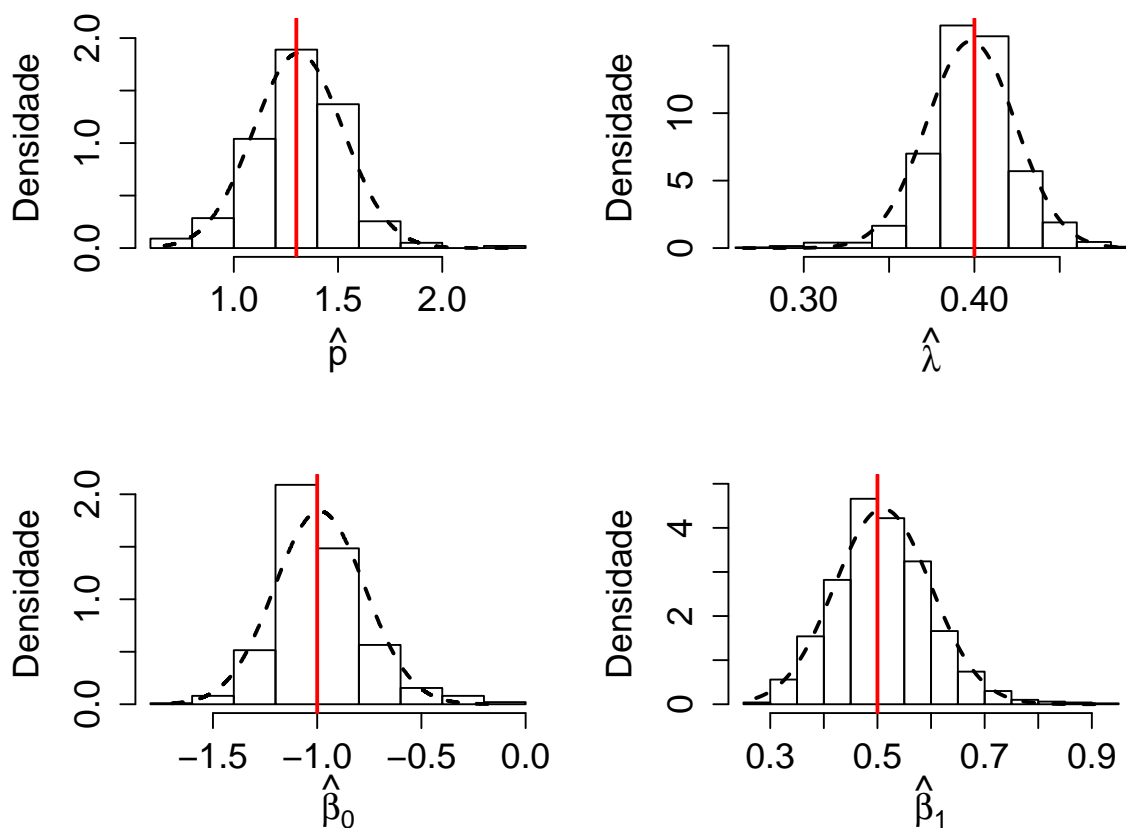


Figura 5 – Comportamento assintótico dos parâmetros envolvidos do Modelo de Fragilidade PZM com covariáveis para o segundo estudo de simulação (Zero-Deflação).

Fonte: Elaborada pelo autor.

Tabela 6 – Média dos Estimadores de Máxima Verossimilhança (MEMV), Raiz do Erro Quadrático Médio (REQM), Desvio Padrão (DP) e Probabilidade de Cobertura (PC) da fração de cura e heterogeneidade não observada captada pelo Modelo de Fragilidade PZM com covariáveis, sendo  $q_1 = 0,600$ ,  $q_2 = 0,409$ ,  $\theta_1 = 0,425$  e  $\theta_2 = 0,645$  (Zero-Deflação).

n	Parâmetros	MEMV	REQM	DP	PC (90%)	PC (95%)
50	$q_1$	0,604	0,167	0,100	0,884	0,935
	$q_2$	0,399	0,188	0,100	0,877	0,919
	$\theta_1$	0,523	0,602	0,598	0,919	0,946
	$\theta_2$	1,315	2,167	2,014	0,977	0,989
100	$q_1$	0,600	0,152	0,071	0,878	0,931
	$q_2$	0,411	0,165	0,071	0,890	0,946
	$\theta_1$	0,523	0,630	0,626	0,957	0,975
	$\theta_2$	1,061	1,363	1,250	0,982	0,996
300	$q_1$	0,599	0,136	0,039	0,914	0,958
	$q_2$	0,408	0,155	0,041	0,885	0,944
	$\theta_1$	0,458	0,196	0,153	0,970	0,987
	$\theta_2$	0,797	0,612	0,534	0,978	0,996
500	$q_1$	0,598	0,135	0,031	0,897	0,950
	$q_2$	0,410	0,152	0,031	0,906	0,953
	$\theta_1$	0,456	0,171	0,119	0,969	0,987
	$\theta_2$	0,758	0,450	0,365	0,975	0,998
1000	$q_1$	0,600	0,134	0,021	0,907	0,953
	$q_2$	0,408	0,153	0,022	0,906	0,960
	$\theta_1$	0,435	0,144	0,060	0,983	0,989
	$\theta_2$	0,683	0,269	0,189	0,981	0,993

Fonte: Elaborada pelo autor.

### 3.5 Síntese

Neste Capítulo, desenvolvemos os Modelos de Sobrevivência induzidos por fragilidade discreta SPZM. Especificamente foram estudados os Modelos de Fragilidade PZM, GZM e BZM. Apresentamos em detalhes as respectivas funções de sobrevivência, densidade, risco e fração de cura destes modelos com e sem a presença de covariáveis quando temos a parametrização  $p$  e  $\mu$  e a versão hurdle. Finalmente fizemos o respectivo estudo de simulação para o Modelo de Fragilidade PZM com a presença de uma variável dicotômica, em que analisamos dois cenários com diferentes frequências de zero (Zero-Inflação e Zero-Deflação). Desta forma, verificamos que os parâmetros envolvidos nos estudos de simulações tendem a ser assintoticamente normais. No processo de efetuar o estudo de simulação para os dois cenários, tivemos algumas dificuldades nos resultados da probabilidade da cobertura para alguns parâmetros envolvidos, mudamos os shutes iniciais e mesmo assim continuou esse problema.

---

## APLICAÇÃO EM DADOS DE MELANOMA

---

Neste Capítulo, apresentamos uma aplicação em um conjunto de dados reais usando os Modelos de Fragilidade PZM e GZM. A escolha do modelo conveniente para o conjunto de dados será feito por meio de critérios de seleção.

A Fundação Oncocentro São Paulo (FOSP) é uma instituição pública vinculada à Secretaria Estadual de Saúde, que auxilia no estudo e na implementação de políticas públicas na área de Oncologia. Os registros foram fornecidos pela FOSP, responsável pela coordenação do Registro de Câncer Hospitalar do Estado de São Paulo. O conjunto de dados deste estudo pertencem a pacientes diagnosticados com melanoma no estado de São Paulo, Brasil. O estudo incluiu pacientes entre 2000 e 2014, com acompanhamento realizado até 2018. O evento de interesse foi o óbito atribuído apenas ao câncer de melanoma. Da amostra original foram removidos 592 pacientes, pois não tinham informações sobre as covariáveis observadas, restando  $n = 7823$  pacientes no estudo. Das observações 73,4% foram censuradas, ou seja, pacientes que não apresentaram o evento de interesse no tempo estudado.

Foram pesquisadas as covariáveis: idade medida em anos, sexo, estadiamento clínico do câncer (que é o processo para determinar a extensão do câncer presente no corpo de uma pessoa e onde está localizado), radioterapia e quimioterapia. Na [Tabela 7](#) apresentamos em detalhe a descrição das covariáveis estudadas. Na amostra observada, 49,41% foram sexo masculino, 31,85% foram pacientes pertencentes ao estadiamento clínico II, só 8,51% receberam radioterapia e 84,10% não fizeram quimioterapia.

Tabela 7 – Descrição das covariáveis pertencentes aos dados de melanoma.

Covariável	Categoria	Descrição	n	%
$X_1$ : idade	-	medida em anos $\mu = 58,42$ $\sigma = 16,36$	7823	-
$X_2$ : sexo	0	Masculino	3865	49,91%
	1	Feminino	3958	50,59%
$X_3$ : estadiamento clínico do câncer	0	Estadiamento I	5331	68,15%
	1	Estadiamento II	2492	31,85%
$X_4$ : radioterapia	0	Não receberam	7157	91,49%
	1	Receberam	666	8,51%
$X_5$ : quimioterapia	0	Não receberam	6579	84,10%
	1	Receberam	1244	15,90%

Fonte: Dados da pesquisa.

Na [Figura 6](#) apresentamos a estimativa de Kaplan-Meier da função de sobrevivência para os dados de melanoma. Observamos que a curva de Kaplan-Meier aproximadamente após os 16 anos se estabiliza, sugerindo a existência de uma fração de indivíduos com risco zero, pelo menos no que concerne ao intervalo de tempo abrangido pelo estudo. Assim, a probabilidade de um paciente com melanoma sobreviver 16 anos é aproximadamente 0,55.

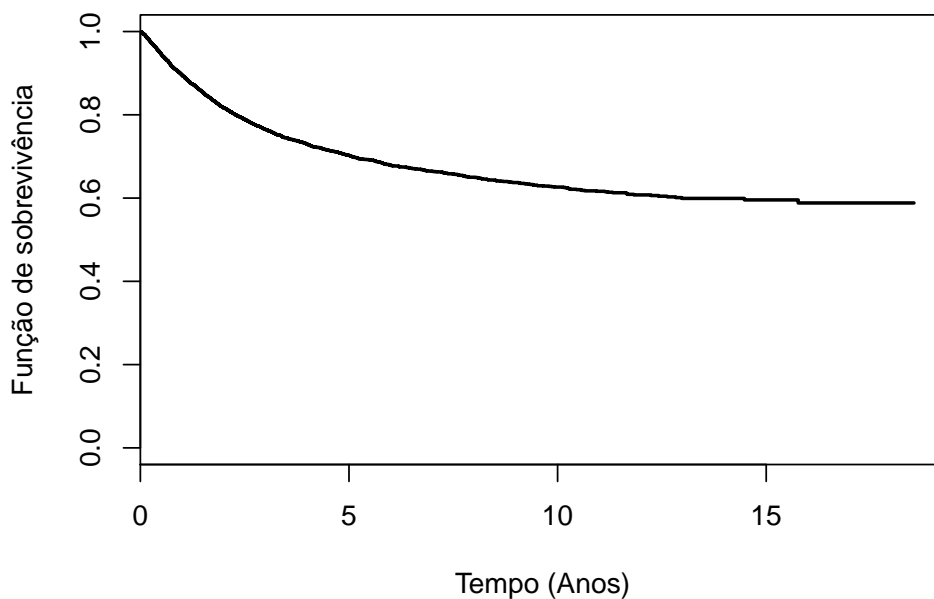


Figura 6 – Estimativa de Kaplan-Meier para os dados de melanoma.

Fonte: Elaborada pelo autor.

Em relação à distribuição base dos tempos de melanoma, consideramos a distribuição Weibull, cuja função densidade e sobrevivência são, respectivamente:

$$f_0(t) = \gamma \lambda^\gamma t^{\gamma-1} e^{-(\lambda t)^\gamma}, \quad \gamma, \lambda > 0, t \geq 0,$$

$$S_0(t) = e^{-(\lambda t)^\gamma},$$

estas duas funções serão substituídas devidamente nas respectivas funções de sobrevivência e risco dos modelos propostos.

Fizemos uma comparação entre a estimativa de Kaplan-Meier para os dados de melanoma e as respectivas curvas de sobrevivência dos Modelos de Fragilidade PZM e GZM, respectivamente. Para fazer esta comparação, obtemos as respectivas estimativas dos parâmetros para cada modelo e plotamos as correspondentes curvas de sobrevivência, desta forma, verificamos se os modelos propostos explicam adequadamente os dados de melanoma. É importante resaltar que esta análise foi feita sem a inclusão de covariáveis.

Na [Tabela 8](#) e [9](#) exibimos as respectivas estimativas dos parâmetros para os Modelos de Fragilidade PZM e GZM. Para o Modelo de Fragilidade PZM, na [Tabela 8](#) observamos que o parâmetro  $p = 0,471$  e em relação ao Modelo de Fragilidade GZM, vemos na [Tabela 9](#) que  $p = 0,610$ . Em ambos modelos observamos ( $0 < p < 1$ ), o que revela a presença inflacionada de zeros no estudo, quer dizer, existe mais pacientes com risco individual zero do que o esperado. Os parâmetros de forma e escala  $\gamma$  e  $\lambda$ , respectivamente, provém da distribuição de base Weibull para cada um dos modelos propostos.

Tabela 8 – Estimativa de Máxima Verossimilhança (EMV), Erro Padrão (EP), Intervalo de Confiança (IC) de 95%, fração de cura ( $q_0$ ), valores AIC e BIC para o Modelo de Fragilidade PZM.

Parâmetro	EMV	EP	IC 95%	
			Inferior	Superior
$p$	0,471	0,045	0,383	0,558
$\mu$	2,048	0,661	0,752	3,344
$\gamma$	1,065	0,024	1,017	1,112
$\lambda$	0,144	0,033	0,080	0,208
$q_0$	0,590	0,011	0,569	0,612
$\max l(\cdot)$		-7494,663		
AIC		14997,33		
BIC		15025,19		

Fonte: Elaborada pelo autor.

Tabela 9 – Estimativa de Máxima Verossimilhança (EMV), Erro Padrão (EP), Intervalo de Confiança (IC) de 95%, fração de cura ( $q_0$ ), valores AIC e BIC para o Modelo de Fragilidade GZM.

Parâmetro	EMV	EP	IC 95%	
			Inferior	Superior
$p$	0,610	0,037	0,537	0,683
$\mu$	2,129	0,398	1,349	2,908
$\gamma$	1,135	0,029	1,080	1,191
$\lambda$	0,130	0,015	0,100	0,158
$q_0$	0,585	0,011	0,563	0,608
$\max l(\cdot)$		-7489,96		
AIC		14987,92		
BIC		15015,78		

Fonte: Elaborada pelo autor.

Agora, é relevante obter a respectiva restrição do parâmetro  $p$  para os Modelos de Fragilidade PZM e GZM. A restrição do parâmetro  $p$  para o Modelo de Fragilidade PZM é dado por

$$0 \leq p \leq \frac{1}{1 - e^{-\mu}} \implies 0 \leq p \leq 1,148;$$

e a restrição para o Modelo de Fragilidade GZM é

$$0 \leq p \leq \frac{1 + \mu}{\mu} \implies 0 \leq p \leq 1,470;$$

desta forma, corroboramos que as estimativas obtidas para o parâmetro  $p$  atendem à restrição dos Modelos de Fragilidade PZM e GZM, respectivamente.

O Modelo de Fragilidade também quantifica a heterogeneidade não observada e é dada pela variância da distribuição usada. Na [Tabela 10](#) apresentamos as respectivas variâncias captadas por cada um dos modelos. De acordo com os Modelos de Fragilidade PZM e GZM, existem fatores não observados ou que não foram considerados no estudo que afetam os tempos de sobrevivência dos pacientes de melanoma.

Tabela 10 – Heterogeneidade não observada (Variância) e Intervalo de Confiança (IC) de 95% atingida pelos Modelos de Fragilidade PZM e GZM.

Modelo	Variância	IC 95%	
		Inferior	Superior
PZM	2,010	0,269	3,780
GZM	5,139	2,213	8,064

Fonte: Elaborada pelo autor.



Na [Figura 7](#) e [8](#) mostramos a comparação entre a estimativa de Kaplan-Meier e as curvas de sobrevivência dos Modelos de Fragilidade PZM e GZM, respectivamente. Observamos o acompanhamento das curvas de sobrevivência destes modelos à estimativa de Kaplan-Meier no transcurso do estudo, o que indica um bom ajuste destes modelos aos dados de melanoma. O Modelo de Fragilidade PZM fornece a estimativa de que 0,590 dos pacientes não morreram devido ao câncer enquanto o Modelo de Fragilidade GZM revela que 0,585 dos pacientes não morreram devido ao câncer no período estudado.

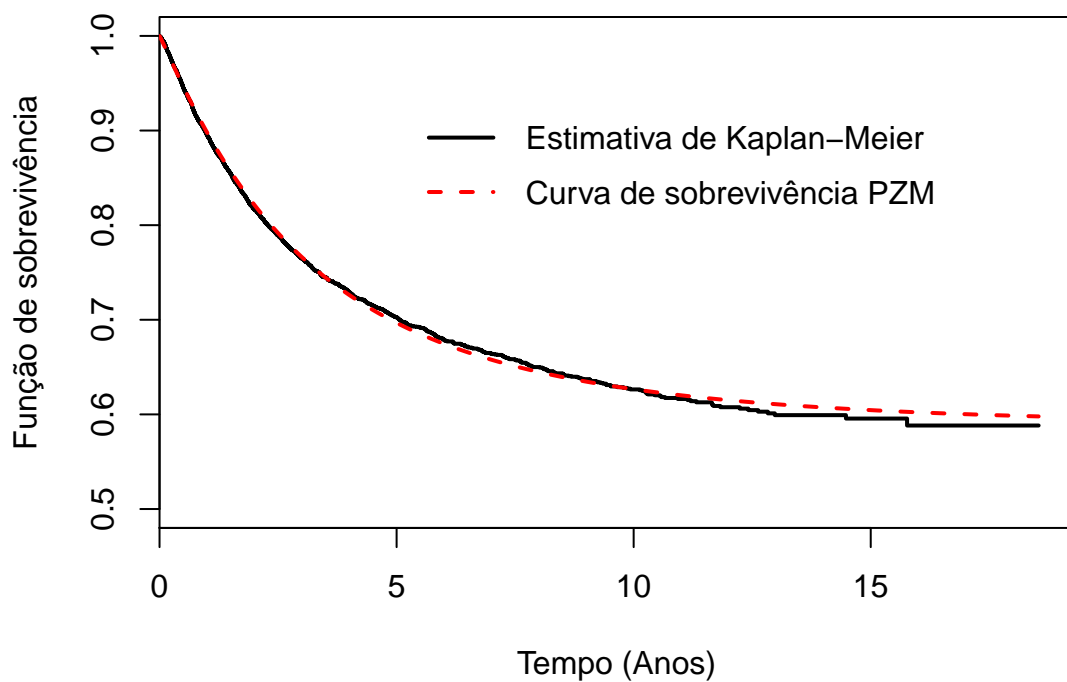


Figura 7 – Estimativa de Kaplan-Meier e a curva de sobrevivência do Modelo de Fragilidade PZM para os dados de melanoma.

Fonte: Elaborada pelo autor.

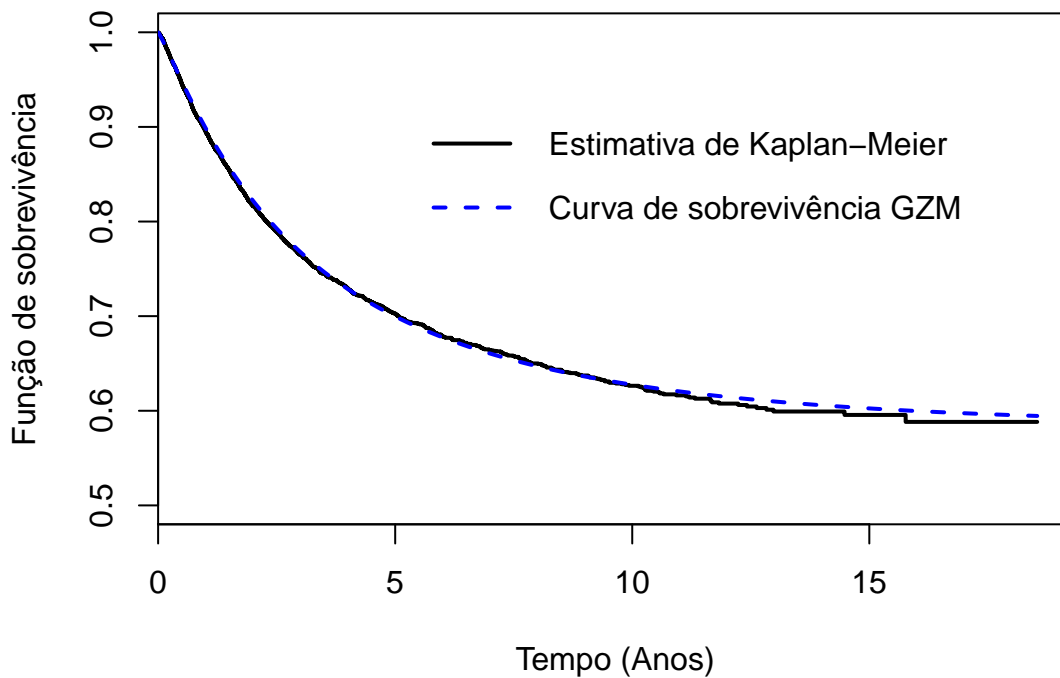


Figura 8 – Estimativa de Kaplan-Meier e a curva de sobrevivência do Modelo de Fragilidade GZM para os dados de melanoma.

Fonte: Elaborada pelo autor.

Agora, levando em consideração as informações medidas nos pacientes, ajustaremos os Modelos de Fragilidade PZM e GZM com covariáveis, respectivamente, ver [Seção 3.3](#) deste trabalho.

Primeiro ostentamos quatro cenários em que concebemos apenas o efeito de cada covariável por vez para cada Modelo de Fragilidade proposto. O objetivo é analisar a influência separada de cada uma das covariáveis estudadas nos tempos de óbito devido ao melanoma dos pacientes. Na [Tabela 11](#) e [12](#) apresentamos as correspondentes estimativas dos parâmetros para cada cenário proposto dos Modelos de Fragilidade. Desta forma, temos que:

- O modelo 1 sugere, independentemente do Modelo de Fragilidade usado, um efeito significativo do sexo dos pacientes, pois só basta observar o intervalo de confiança de 95% para  $\beta_1$ . Na [Figura 9a](#) e [11a](#), apreciamos que as mulheres têm maior chance de não morrer devido ao câncer em relação aos homens.
- O intervalo de confiança para o parâmetro  $\beta_2$ , no modelo 2, demonstra que o estadiamento clínico nos pacientes tem uma considerável influência nos tempos de óbito, independentemente do Modelo de Fragilidade empregado. Na [Figura 9b](#) e [11b](#), se vê refletido esta

influência nas frações de cura. Quanto maior o nível do tumor menor a proporção de cura, pois nesse caso a doença estaria num estágio mais avançado.

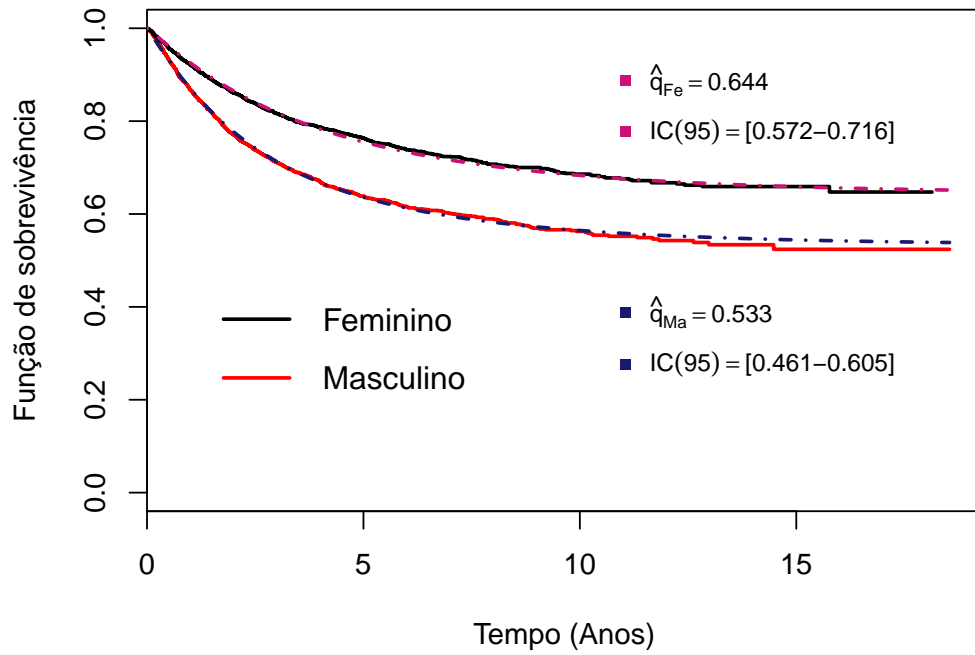
- Independentemente do Modelo de Fragilidade utilizado, ao observar o intervalo de confiança de 95% para o parâmetro  $\beta_3$ , no modelo 3, se evidencia que a recepção da radioterapia influencia nos tempos de óbito dos pacientes. Da [Figura 10a](#) e [12a](#), inferimos que os pacientes que não receberam radioterapia têm maior chance de não morrer devido ao câncer em relação aos pacientes que receberam.
- A partir do intervalo de confiança de 95% para o parâmetro  $\beta_4$ , no modelo 4, inferimos que a recepção da quimioterapia tem efeito significativo nos tempos de óbito dos pacientes, independentemente do Modelo de Fragilidade empregado. A [Figura 10b](#) e [12b](#), manifestam que os pacientes que receberam quimioterapia têm maior probabilidade de morrer devido ao câncer em comparação aos pacientes que não tiveram quimioterapia.
- O parâmetro  $p$  quando é analisado a covariável sexo e estadiamento clínico II, autônomo do Modelo de Fragilidade usado, sugere que existe inflação de zeros na população, quer dizer, que existem mais pacientes do que o esperado com risco individual de zero. No momento em que, só estudamos a influência da quimioterapia nos pacientes, o intervalo de confiança para  $p$  do Modelo de Fragilidade PZM, indica que provavelmente a distribuição Poisson também poderia ser usada para fazer a análise, enquanto o  $p$  do Modelo de Fragilidade GZM sugere que existe deflação de zeros. Finalmente, na ocasião em que, pesquisamos a ação da covariável quimioterapia nos tempos de óbito dos pacientes, independente do Modelo de Fragilidade empregado,  $p$  demonstra que existe deflação de zeros, isto é, que tem menos pacientes do que o esperado com risco individual zero .

Na [Figura 9](#), [10](#), [11](#) e [12](#) apresentamos as respectivas estimativas de Kaplan-Meier considerando os Modelos de Regressão para Fragilidade PZM e GZM dividido pelas covariáveis sexo, estadiamento do câncer, radioterapia e quimioterapia. Observamos em cada um dos gráficos, independente do Modelo de Fragilidade usado, a existência de uma forte evidência de fração de cura na população e o acompanhamento das respectivas curvas de sobrevivência dos Modelos de Fragilidade (linhas tracejadas) à estimativa de Kaplan-Meier (linha contínua). Assim, inferimos que os dois Modelos de Fragilidade indicam um bom ajuste aos dados de melanoma.

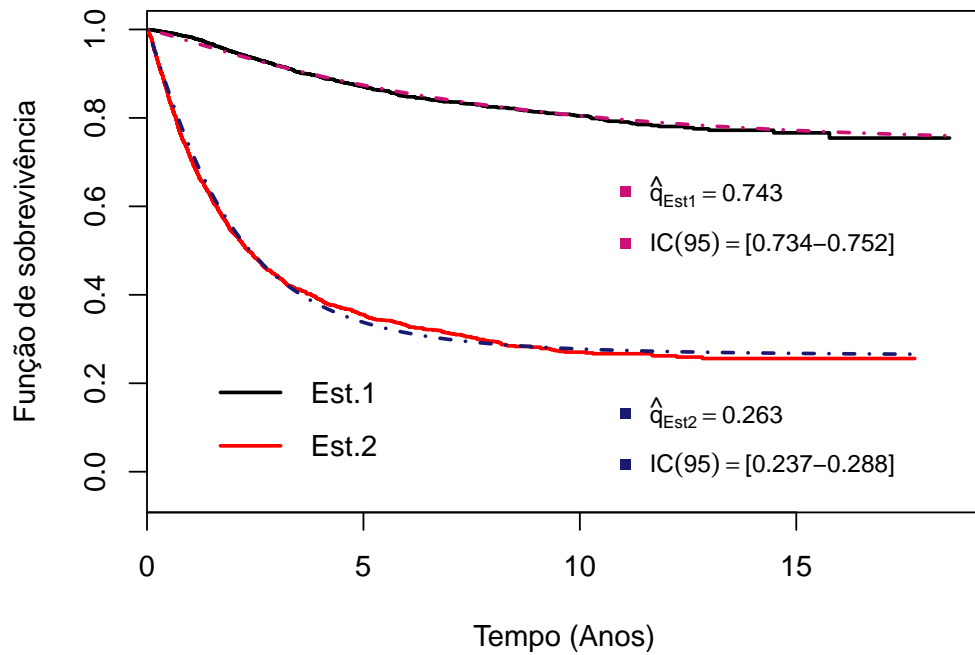
Tabela 11 – Estimativa de Máxima Verossimilhança (EMV), Erro Padrão (EP), Intervalo de Confiança (IC) de 95%, valores AIC e BIC do Modelo de Fragilidade PZM classificado por sexo, estadiamento clínico do câncer, radioterapia e quimioterapia para os dados de melanoma.

Parâmetro	EMV	EP	IC 95%	
			Inferior	Superior
$p$	0,541	0,043	0,458	0,625
$\gamma$	1,065	0,024	1,018	1,113
$\lambda$	0,167	0,021	0,127	0,208
$\beta_0$	1,301	0,244	0,823	1,779
$\beta_1$ (Sexo: feminino)	-0,617	0,060	-0,735	-0,500
$\max l(\cdot)$		-7433,41		
AIC		14876,82		
BIC		14911,65		
$p$	0,741	0,013	0,715	0,767
$\gamma$	1,138	0,025	1,089	1,187
$\lambda$	0,123	0,014	0,095	0,152
$\beta_0$	-3,343	0,115	-3,568	-3,118
$\beta_2$ (Estadiamento II)	2,487	0,059	2,372	2,602
$\max l(\cdot)$		-6424,11		
AIC		12858,22		
BIC		12893,04		
$p$	0,976	0,065	0,848	1,104
$\gamma$	1,044	0,022	1,000	1,088
$\lambda$	0,227	0,013	0,201	0,253
$\beta_0$	-0,771	0,100	-0,968	-0,574
$\beta_3$ (Radioterapia: sim)	1,402	0,072	1,261	1,543
$\max l(\cdot)$		-7254,15		
AIC		14518,30		
BIC		14553,12		
$p$	1,002	0,060	0,885	1,120
$\gamma$	1,056	0,022	1,013	1,100
$\lambda$	0,227	0,014	0,200	0,254
$\beta_0$	-0,994	0,095	-1,180	-0,808
$\beta_4$ (Quimioterapia: sim)	1,536	0,062	1,414	1,657
$\max l(\cdot)$		-7017,49		
AIC		14044,98		
BIC		14079,80		

Fonte: Elaborada pelo autor.



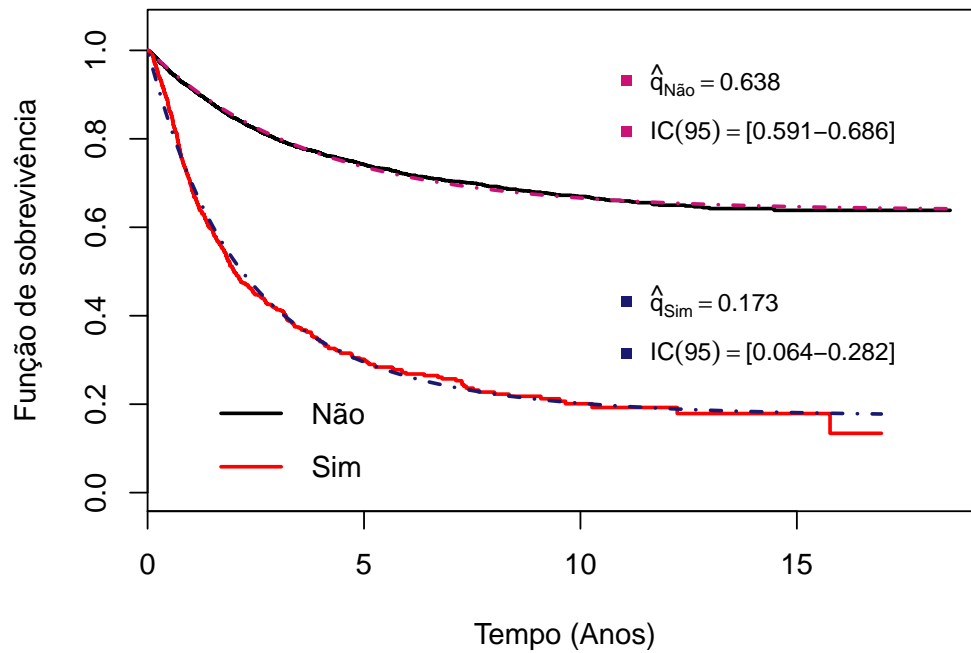
(a) Sexo.



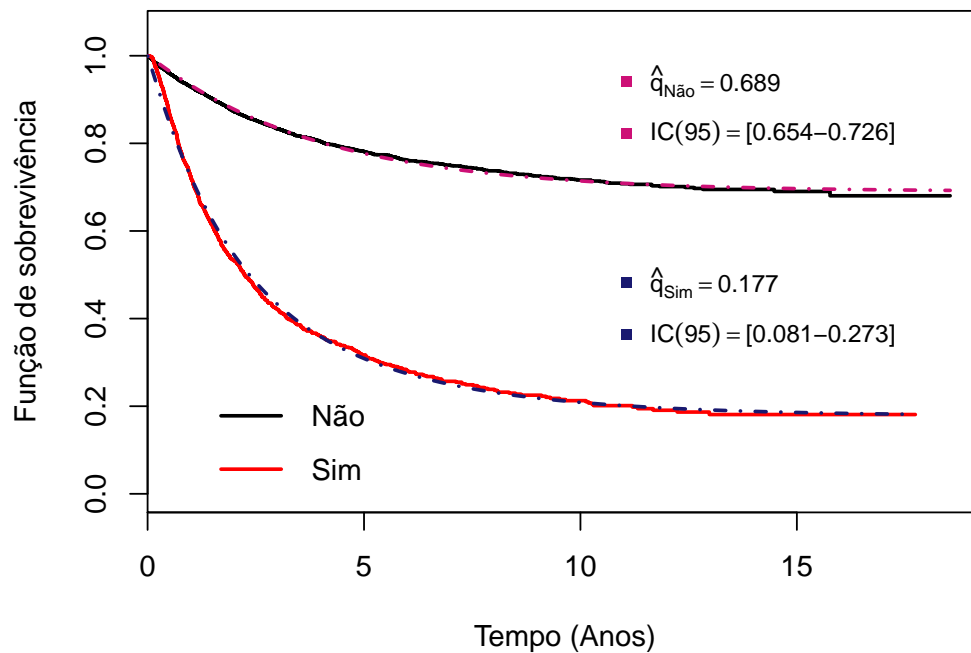
(b) Estadiamento clínico do câncer.

Figura 9 – Estimativas de Kaplan-Meier e curvas de sobrevivência do Modelo de Fragilidade PZM estratificado por (a) sexo e (b) estadiamento clínico do câncer para os dados de melanoma.

Fonte: Elaborada pelo autor.



(a) Radioterapia.



(b) Quimioterapia.

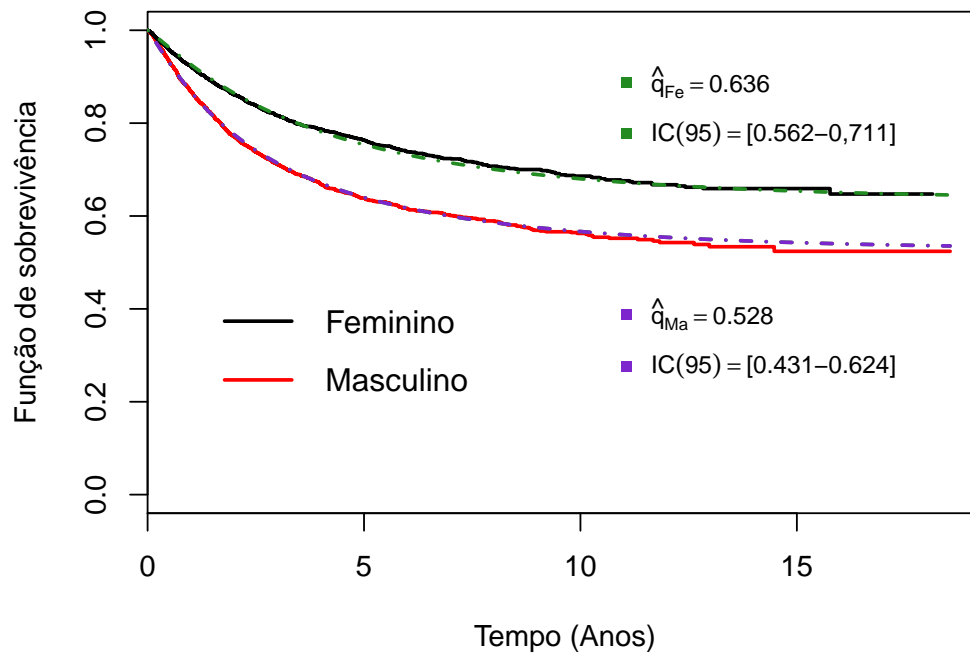
Figura 10 – Estimativas de Kaplan-Meier e curvas de sobrevivência do Modelo de Fragilidade PZM estratificado por (a) radioterapia e (b) quimioterapia para os dados de melanoma.

Fonte: Elaborada pelo autor.

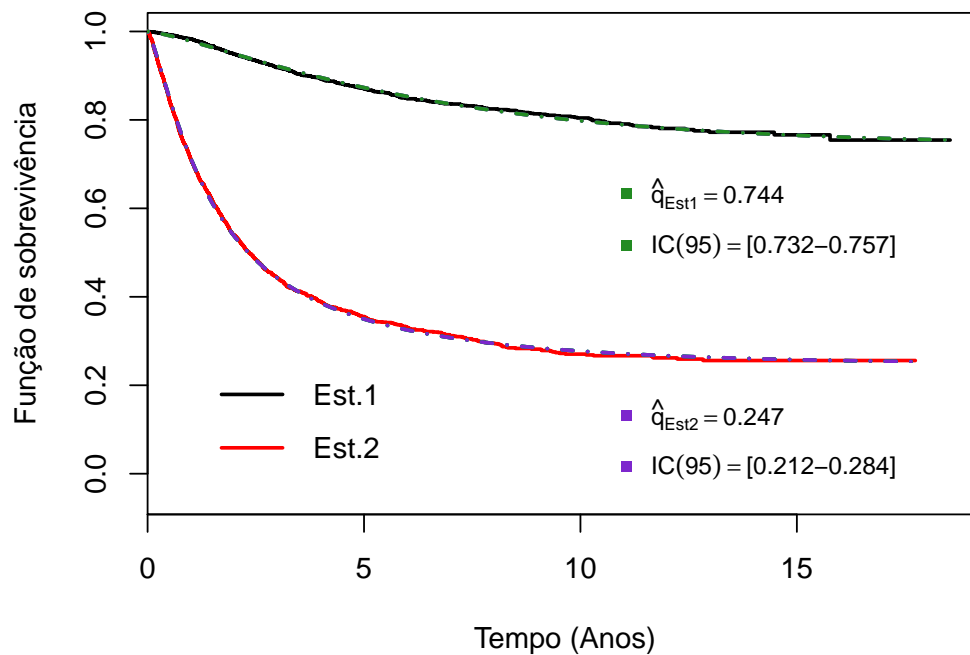
Tabela 12 – Estimativa de Máxima Verossimilhança (EMV), Erro Padrão (EP), Intervalo de Confiança (IC) de 95%, valores AIC e BIC do Modelo de Fragilidade GZM classificado por sexo, estadiamento clínico do câncer, radioterapia e quimioterapia para os dados de melanoma.

Parâmetro	EMV	EP	IC 95%	
			Inferior	Superior
$p$	0,678	0,071	0,539	0,816
$\gamma$	1,122	0,030	1,063	1,180
$\lambda$	0,142	0,023	0,095	0,188
$\beta_0$	1,518	0,367	0,799	2,238
$\beta_1$ (Sexo: feminino)	-0,685	0,070	-0,824	-0,547
$\max l(\cdot)$		-7430,388		
AIC		14870,78		
BIC		14905,60		
$p$	0,843	0,021	0,803	0,884
$\gamma$	1,300	0,030	1,242	1,359
$\lambda$	0,121	0,012	0,097	0,146
$\beta_0$	-3,778	0,125	-4,022	-3,533
$\beta_2$ (Estadiamento II)	2,946	0,083	2,782	3,109
$\max l(\cdot)$		-6403,653		
AIC		12817,31		
BIC		12852,13		
$p$	1,390	0,137	1,121	1,659
$\gamma$	1,064	0,024	1,017	1,111
$\lambda$	0,219	0,014	0,192	0,247
$\beta_0$	-1,048	0,148	-1,339	-0,757
$\beta_3$ (Radioterapia: sim)	1,459	0,087	1,289	1,629
$\max l(\cdot)$		-7254,082		
AIC		14518,16		
BIC		14552,99		
$p$	1,476	0,134	1,213	1,738
$\gamma$	1,078	0,024	1,031	1,125
$\lambda$	0,225	0,014	0,197	0,252
$\beta_0$	-1,333	0,136	-1,599	-1,067
$\beta_4$ (Quimioterapia: sim)	1,580	0,075	1,433	1,726
$\max l(\cdot)$		-7018,712		
AIC		14047,42		
BIC		14082,25		

Fonte: Elaborada pelo autor.



(a) Sexo.

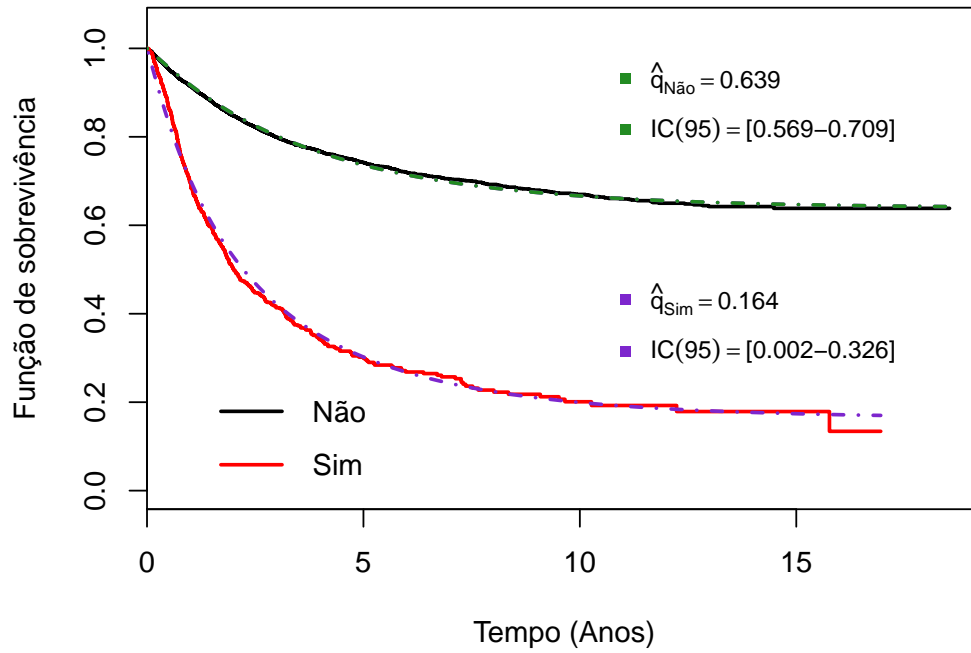


(b) Estadiamento clínico do câncer.

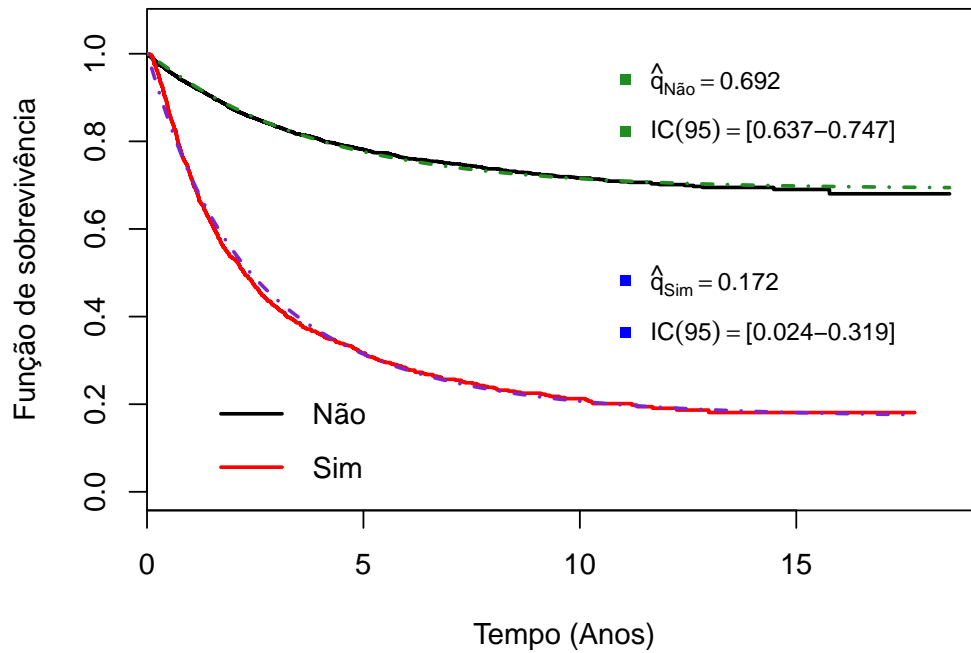
Figura 11 – Estimativas de Kaplan-Meier e curvas de sobrevivência do Modelo de Fragilidade GZM estratificado por (a) sexo e (b) estadiamento clínico do câncer para os dados de melanoma.

Fonte: Elaborada pelo autor.





(a) Radioterapia.



(b) Quimioterapia.

Figura 12 – Estimativas de Kaplan-Meier e curvas de sobrevivência do Modelo de Fragilidade GZM estratificado por (a) radioterapia e (b) quimioterapia para os dados de melanoma.

Fonte: Elaborada pelo autor.

Já foi realizado a análise dos Modelos de Fragilidade PZM e GZM considerando uma covariável por vez. A finalidade foi conhecer a influência separada das covariáveis nos tempos de óbito devido ao câncer dos pacientes. Agora, faremos a inclusão de todas as covariáveis ao mesmo tempo, no parâmetro  $\mu$  para os Modelos de Fragilidade propostos.

As correspondentes estimativas dos parâmetros dos Modelos de Fragilidade PZM e GZM com a presença de todas as covariáveis estão apresentadas na [Tabela 13](#) e [14](#), respectivamente. Deste modo, temos que:

- Dos resultados pertencentes ao Modelo de Fragilidade PZM, inferimos que o tempo de sobrevida de uma mulher é maior comparado ao de um homem, também se observa que para os pacientes que receberam radioterapia o chance de não morrer é menor em comparação aos pacientes que não receberam.
- Em relação ao Modelo de Fragilidade GZM, concluímos que para os pacientes diagnosticados com estadiamento II, a sobrevida é menor, pois quanto maior o nível do tumor menor a possibilidade de ser curado. Além disso, apreciamos que a possibilidade de morrer devido ao câncer dos pacientes que receberam quimioterapia é maior em comparação aos pacientes que não receberam.
- Se observamos os Intervalos de Confiança de 95%, independente do Modelo de Fragilidade usado, concluímos que as covariáveis: idade, sexo, estadiamento clínico do câncer, radioterapia e quimioterapia têm um efeito significativo no tempo de óbito devido ao câncer dos pacientes.
- O Intervalo de Confiança de 95% para o parâmetro  $p$  do Modelo de Fragilidade PZM, ainda com a presença de todas as covariáveis, sugere a existência inflacionada de zeros enquanto para o Modelo de Fragilidade GZM, o Intervalo de Confiança de 95% propõe que a distribuição Geométrica também poderia ser empregada para fazer a análise.

Tabela 13 – Estimativa de Máxima Verossimilhança (EMV), Erro Padrão (EP), Intervalo de Confiança (IC) de 95%, valores AIC e BIC do Modelo de Fragilidade PZM para os dados de melanoma considerando todas as covariáveis.

Parâmetro	EMV	EP	IC 95%	
			Inferior	Superior
$p$	0,877	0,022	0,834	0,919
$\gamma$	1,130	0,024	1,084	1,177
$\lambda$	0,165	0,013	0,142	0,19
$\beta_0$	-3,332	0,154	-3,634	-3,031
$\beta_1$ (Idade)	0,013	0,002	0,010	0,016
$\beta_2$ (Sexo: feminino)	-0,408	0,050	-0,505	-0,310
$\beta_3$ (Estadiamento II)	1,881	0,066	1,752	2,010
$\beta_5$ (Radioterapia: sim)	0,409	0,067	0,277	0,542
$\beta_6$ (Quimioterapia: sim)	0,598	0,060	0,480	0,716
$\max l(\cdot)$		-6300,21		
AIC		12618,42		
BIC		12681,10		

Fonte: Elaborada pelo autor.

Tabela 14 – Estimativa de Máxima Verossimilhança (EMV), Erro Padrão (EP), Intervalo de Confiança (IC) de 95%, valores AIC e BIC do Modelo de Fragilidade GZM para os dados de melanoma considerando todas as covariáveis.

Parâmetro	EMV	EP	IC 95%	
			Inferior	Superior
$p$	0,995	0,029	0,939	1,052
$\gamma$	1,258	0,028	1,202	1,313
$\lambda$	0,147	0,012	0,124	0,170
$\beta_0$	-3,704	0,176	-4,048	-3,359
$\beta_1$ (Idade)	0,015	0,002	0,012	0,019
$\beta_2$ (Sexo: feminino)	-0,457	0,058	-0,571	-0,342
$\beta_3$ (Estadiamento II)	2,230	0,082	2,069	2,391
$\beta_5$ (Radioterapia: sim)	0,563	0,083	0,400	0,727
$\beta_6$ (Quimioterapia: sim)	0,706	0,071	0,568	0,845
$\max l(\cdot)$		-6271,448		
AIC		12560,90		
BIC		12623,58		

Fonte: Elaborada pelo autor.

Em todas as análises feitas, vimos que os Modelos de Fragilidade PZM e GZM têm um bom ajuste aos dados de melanoma. Para a escolha do melhor modelo foi feito o critério AIC e BIC. Desta forma, temos que:

- A primeira análise efetuada, foi sem a incorporação das covariáveis para os Modelos de Fragilidade PZM e GZM. Na [Tabela 8](#) e [9](#) temos os critérios AIC e BIC, ao compará-los revelam que o melhor modelo para os dados de melanoma é o Modelo de Fragilidade GZM.
- Seguidamente, consideramos quatro cenários considerando apenas o efeito de cada covariável por vez para cada Modelo de Fragilidade proposto. Analisamos os respectivos valores AIC e BIC, na [Tabela 11](#) e [12](#). Estes critérios indicam que o modelo razoável para os dados de melanoma quando consideramos separadamente as covariáveis sexo, estadiamento clínico e radioterapia é o Modelo de Fragilidade GZM. Entretanto, o Modelo de Fragilidade PZM é favorável quando fazemos a análise só com a covariável quimioterapia.
- Finalmente, inserimos todas as covariáveis estudadas ao mesmo tempo no parâmetro  $\mu$ . Contrastamos os correspondentes critérios AIC e BIC destes Modelos de Fragilidade, ubicados na [Tabela 13](#) e [14](#). O Modelo de Fragilidade GZM é selecionado como o modelo adequado para o conjunto de dados de melanoma.

O Modelo de Fragilidade GZM com covariáveis foi escolhido como um modelo favorável para o conjunto de dados de melanoma, segundo os criterios AIC e BIC. Agora fixaremos situações específicas pois nosso interesse é obter a respectiva fração de cura para um determinado paciente. Fizemos as diferentes combinações com todas as covariáveis fixando as idades de 30 e 60 anos para realizar as respectivas comparações. Na [Tabela 15](#) mostramos as respectivas frações de cura univariada para todas as combinações possíveis das covariáveis. Apreciamos que:

- Existe diferença na fração de cura para uma mulher de 30 anos (Estadiamento Clínico: II com a presença de radioterapia e quimioterapia) em relação a um homem de 30 anos com as mesmas condições em um nível de significância de 5%.
- Vemos que existe uma diferença na fração de cura individual para um homem de 60 anos (Estadiamento Clínico: II, sem radioterapia e com quimioterapia) em comparação a uma mulher de 60 anos sobre as mesmas características.
- A fração de cura para as mulheres é maior em comparação dos homens, isto acontece porque uma proporção das mulheres descobriram que tinham câncer na primeira fase e isso é vantajoso, pois dessa forma o câncer pode ser removido ou curado enquanto os homens quando foram para o médico acharam que eles possuíam melanoma na fase final, nessa situação é mais difícil ser curado pois a doença está numa etapa mais avançada.

Tabela 15 – Fração de cura univariada ( $q_0$ ), Erro Padrão (EP), e Intervalo de Confiança (IC) de 95% do Modelo de Fragilidade GZM analisada para sexo, estadiamento clínico, radioterapia e quimioterapia para as idades 30 e 60 anos.

Idade	Sexo	Estadiamento Clínico	Radioterapia	Quimioterapia	$q_0$	EP	IC 95%	
							Inferior	Superior
30	M	I	Não	Não	0,962	0,006	0,951	0,973
				Sim	0,924	0,012	0,901	0,948
			Não	0,934	0,011	0,913	0,956	
		II	Não	Não	0,871	0,022	0,828	0,914
				Sim	0,697	0,032	0,634	0,760
			Sim	0,482	0,048	0,388	0,576	
30	F	I	Não	Não	0,531	0,050	0,433	0,629
				Sim	0,279	0,053	0,176	0,382
			Não	0,976	0,003	0,970	0,982	
		II	Não	Sim	0,951	0,007	0,938	0,965
				Não	0,958	0,006	0,945	0,970
			Sim	0,916	0,013	0,890	0,942	
60	M	I	Não	Não	0,796	0,018	0,761	0,830
				Sim	0,630	0,032	0,566	0,693
			Não	0,670	0,033	0,604	0,735	
		II	Não	Sim	0,444	0,046	0,355	0,533
				Não	0,940	0,008	0,925	0,956
			Sim	0,883	0,018	0,848	0,918	
60	F	I	Não	Não	0,898	0,016	0,866	0,929
				Sim	0,804	0,031	0,742	0,865
			Não	0,565	0,038	0,492	0,639	
		II	Não	Sim	0,316	0,047	0,224	0,408
				Sim	0,368	0,051	0,268	0,468
			Não	0,134	0,038	0,059	0,209	
60	F	I	Não	Não	0,962	0,004	0,954	0,970
				Sim	0,924	0,010	0,904	0,944
			Não	0,934	0,009	0,916	0,952	
		II	Não	Sim	0,871	0,019	0,833	0,909
				Sim	0,697	0,021	0,656	0,737
			Sim	0,481	0,036	0,410	0,552	
II	Sim	Não	0,530	0,038	0,455	0,605		
		Sim	0,278	0,042	0,195	0,361		

Fonte: Elaborada pelo autor.



---

## CONCLUSÕES E PROPOSTAS FUTURAS

---

Neste trabalho, foi considerado uma família de distribuições discretas para a modelagem da variável de fragilidade, como uma nova alternativa aos modelos já existentes. A família de distribuições considerada foi a SPZM que é uma extensão da família de distribuição SP. A vantagem do uso desta distribuição é a flexibilidade, pois pode ser ajustada a qualquer conjunto de dados sem ter conhecimento prévio da frequência de zero. Foram estudadas as características e propriedades das distribuições SPZM. Também foi apresentado a versão hurdle desta família. O benefício do uso de uma distribuição discreta para a fragilidade é que permite a existência de indivíduos que não vão apresentar o evento de interesse mesmo após um longo período de tempo.

Desenvolvemos Modelos de Sobrevivência induzidas por fragilidade discreta usando as distribuições PZM, GZM e BZM. Também estudamos os Modelos de Fragilidade SPZM com a presença de covariáveis e apresentamos as respectivas funções de sobrevivência, risco, densidade e fragilidade zero (fração de cura) para cada modelo, considerando a parametrização  $p$  e  $\mu$  e a versão hurdle. Foi realizado o devido estudo de simulação, em que foram desenvolvidos procedimentos de inferências baseados na teoria de máxima verossimilhança e verificamos as propriedades frequentistas dos EMV dos parâmetros envolvidos no Modelo de Fragilidade PZM com a presença de uma covariável dicotômica. Este estudo foi executado para dois cenários, o primeiro referente a Zero-Inflação e o segundo relacionado ao Zero-Deflação. Apresentamos alguns impasses na parte da simulação, sugerimos utilizar algum outro método para resolver o comportamento da Probabilidade de Cobertura, mesmo assim, os parâmetros implicados têm um comportamento assintótico normal.

A metodologia proposta foi aplicada a um conjunto de dados reais de melanoma fornecidos pela FOSP. Efetuamos a respectiva análise dos dados, com e sem a presença das covariáveis estudadas, obtendo as respectivas frações de cura e heterogeneidade não observada (variância). As estimativas dos parâmetros envolvidos foram determinadas mediante a abordagem de máxima

verossimilhança. O propósito da aplicação foi comparar os Modelos de Fragilidade PZM e GZM, e a seleção do modelo apropriado para os dados foi realizado mediante um critério de seleção, sendo eleito o Modelo de Fragilidade GZM.

A principal contribuição deste trabalho foi abordar uma família de distribuições discretas flexíveis para a fragilidade, esta família tem como casos particulares a diferentes distribuições conhecidas como SPZI, SPZD, SPZT e SP. Portanto, seu uso é vantajoso pois dependendo do comportamento do parâmetro  $p$  vai indicar em que situação está a presença de zero (inflacionada, deflacionada ou truncada).

A continuidade deste trabalho inclui usar outras distribuições pertencentes à família SPZM, como por exemplo a BN, PG, entre outras, em que o parâmetro de dispersão  $\phi$  esteja presente nas análises. Desta forma, medir a dispersão dos dados se houver. Também pretendemos estudar em maior detalhe o Modelo de Fragilidade BZM e assim poder fixar um valor razoável para  $m$ . Além disso, explorar com maior detalhe a interpretação do parâmetro  $p$  nos Modelos de Fragilidade e considerar outra reparametrização deste. Ademais, planejamos considerar a versão hurdle para ser usado em métodos bayesianos e assim estimar os parâmetros do modelo proposto, realizando as respectivas análises de simulação considerando este enfoque.



## REFERÊNCIAS

---

ANDREAS, W. **Frailty models in survival analysis**. [S.l.]: Chapman and Hall/CRC, 2010. Citado nas páginas 23, 40 e 41.

ATA, N.; ÖZEL, G. Funções de sobrevivência para os modelos de fragilidade baseados no processo de poisson composto discreto. **Jornal de Computação Estatística e Simulação**, v. 83, n. 11, p. 2105–2116, 2013. Citado na página 24.

BALAKRISHNAN, N.; PENG, Y. Generalized gamma frailty model. **Statistics in medicine**, Wiley Online Library, v. 25, n. 16, p. 2797–2816, 2006. Citado na página 24.

BREASTCANCER. **Câncer de mama**. 2020. Disponível em: <[https://www.breastcancer.org/es/sintomas/cancer\\_de\\_mama/que\\_es\\_cancer\\_mama](https://www.breastcancer.org/es/sintomas/cancer_de_mama/que_es_cancer_mama)>. Citado na página 24.

CANCHO, V. G.; MACERA, M. A.; SUZUKI, A. K.; LOUZADA, F.; ZAVALETA, K. E. A new long-term survival model with dispersion induced by discrete frailty. **Lifetime data analysis**, Springer, p. 1–24, 2019. Citado na página 24.

CARONI, C.; CROWDER, M.; KIMBER, A. Proportional hazards models with discrete frailty. **Lifetime data analysis**, Springer, v. 16, n. 3, p. 374–384, 2010. Citado nas páginas 24 e 41.

CHEN, P.; ZHANG, J.; ZHANG, R. Estimation of the accelerated failure time frailty model under generalized gamma frailty. **Computational Statistics & Data Analysis**, Elsevier, v. 62, p. 171–180, 2013. Citado na página 24.

COLLETT, D. **Modelling survival data in medical research**. [S.l.]: CRC press, 2015. Citado na página 42.

CONCEIÇÃO, K. S. Modelos serie de potência zero-modificados. 2013. Citado nas páginas 25, 27, 28, 29, 31, 35 e 36.

CONCEIÇÃO, K. S.; LOUZADA, F.; ANDRADE, M.; HELOU, E. Zero-modified power series distribution and its hurdle distribution version. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 87, n. 9, p. 1842–1862, 2017. Citado nas páginas 25, 29 e 36.

CONSUL, P. New class of location-parameter discrete probability distributions and their characterizations. **Communications in Statistics-Theory and Methods**, Taylor & Francis, v. 19, n. 12, p. 4653–4666, 1990. Citado na página 27.

CORDEIRO, G. M.; ANDRADE, M. G.; CASTRO, M. de. Power series generalized nonlinear models. **Computational Statistics & Data Analysis**, Elsevier, v. 53, n. 4, p. 1155–1166, 2009. Citado na página 28.

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 nas páginas 11 e 13.

- DIETZ, E.; BÖHNING, D. On estimation of the poisson parameter in zero-modified poisson models. **Computational Statistics & Data Analysis**, Elsevier, v. 34, n. 4, p. 441–459, 2000. Citado na página 25.
- GUPTA, R. C. Modified power series distribution and some of its applications. **Sankhyā: The Indian Journal of Statistics, Series B**, JSTOR, p. 288–298, 1974. Citado na página 28.
- HOUGAARD, P. Life table methods for heterogeneous populations: distributions describing the heterogeneity. **Biometrika**, Oxford University Press, v. 71, n. 1, p. 75–83, 1984. Citado na página 24.
- \_\_\_\_\_. Survival models for heterogeneous populations derived from stable distributions. **Biometrika**, Oxford University Press, v. 73, n. 2, p. 387–396, 1986. Citado nas páginas 24 e 40.
- IBRAHIM, J. G.; CHEN, M.-H.; SINHA, D. **Bayesian Survival Analysis**. [S.l.]: Springer, 2001. Citado na página 24.
- LAWLESS, J. **Statistical Models and Methods for Lifetime Data**. [S.l.]: John Wiley & Sons, Inc., 2003. Citado na página 43.
- MALLER, R. A.; ZHOU, X. **Survival analysis with long-term survivors**. [S.l.]: John Wiley & Sons, 1996. Citado na página 24.
- MCGILCHRIST, C.; AISBETT, C. Regression with frailty in survival analysis. **Biometrics**, v. 47, n. 2, p. 461–466, 1991. Citado na página 24.
- NOACK, A. A class of random variables with discrete distributions. **The Annals of Mathematical Statistics**, JSTOR, v. 21, n. 1, p. 127–132, 1950. Citado na página 27.
- POISSON, S. D. **Recherches sur la probabilité des jugements en matière criminelle et en matière civile précédées des règles générales du calcul des probabilités par sd poisson**. [S.l.]: Bachelier, 1837. Citado na página 31.
- R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2019. Disponível em: <<https://www.R-project.org/>>. Citado nas páginas 26, 49 e 50.
- ROCHA, C. M. T. S. **Modelos com fragilidade em análise de sobrevivência**. [S.l.: s.n.], 1995. Citado na página 41.
- RODRIGUES, J.; CASTRO, M. de; CANCHO, V. G.; BALAKRISHNAN, N. Com–poisson cure rate survival models and an application to a cutaneous melanoma data. **Journal of Statistical Planning and Inference**, Elsevier, v. 139, n. 10, p. 3605–3611, 2009. Citado na página 41.
- SANTOS, D. M. dos; DAVIES, R. B.; FRANCIS, B. Nonparametric hazard versus nonparametric frailty distribution in modelling recurrence of breast cancer. **Journal of Statistical Planning and Inference**, Elsevier, v. 47, n. 1-2, p. 111–127, 1995. Citado na página 24.
- 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 na página 24.
- TOMAZELLA, V. L. D. **Modelagem de dados de eventos recorrentes via processo de Poisson com termo de fragilidade**. Tese (Doutorado) — Universidade de São Paulo, 2003. Citado na página 24.

TWEEDIE, M. C. An index which distinguishes between some important exponential families. In: **Statistics: Applications and new directions: Proc. Indian statistical institute golden Jubilee International conference**. [S.l.: s.n.], 1984. v. 579, p. 579–604. Citado na página 40.

VAUPEL, J. W.; MANTON, K. G.; STALLARD, E. The impact of heterogeneity in individual frailty on the dynamics of mortality. **Demography**, Springer, v. 16, n. 3, p. 439–454, 1979. Citado nas páginas 23, 24 e 39.

ZAVALETA, K. E. C. *et al.* Modelos série de potência com excesso de zeros observáveis e latentes. Universidade Federal de São Carlos, 2016. Citado na página 24.

