

Modelo Destrutivo com Variável Terminal em Experimentos Quimiopreventivos de Tumores em Animais

Katherine Elizabeth Coaguila Zavaleta

Universidade Federal de São Carlos
Centro de Ciências Exatas e de Tecnologia
Departamento de Estatística

Modelo Destrutivo com Variável Terminal em Experimentos Quimiopreventivos de Tumores em Animais

Katherine Elizabeth Coaguila Zavaleta
Orientador: Prof. Dr. Josemar Rodrigues

Dissertação apresentada ao Departamento de Estatística da Universidade Federal de São Carlos - DEs/UFSCar como parte dos requisitos para obtenção do título de Mestre em Estatística.

UFSCar - São Carlos - SP
Abril de 2012

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária da UFSCar**

C652md

Coaguila Zavaleta, Katherine Elizabeth.

Modelo destrutivo com variável terminal em experimentos quimiopreventivos de tumores em animais / Katherine Elizabeth Coaguila Zavaleta. -- São Carlos : UFSCar, 2012. 50 f.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2012.

1. Estatística. 2. Modelo destrutivo. 3. Distribuição de Poisson. 4. Distribuição binomial negativa. 5. Modelos inflacionados de zeros. I. Título.

CDD: 519.5 (20^a)

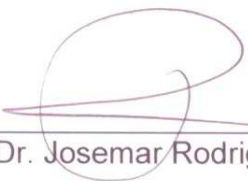
Katherine Elizabeth Coaguila Zavaleta

**MODELO DESTRUTIVO COM VARIÁVEL TERMINAL EM
EXPERIMENTOS QUIMIOPREVENTIVOS DE TUMORES EM ANIMAIS**

Dissertação apresentada à Universidade Federal de São Carlos, como parte dos requisitos para obtenção do título de Mestre em Estatística.

Aprovada em 12 de abril de 2012.

BANCA EXAMINADORA

Presidente 
Prof. Dr. Josemar Rodrigues (DEs-UFSCar/Orientador)

1º Examinador 
Profa. Dra. Juliana Cobre (ICMC-USP)

2º Examinador 
Prof. Dr. Vicente Garibay Gancho (ICMC-USP)

Agradecimentos

À Deus pela minha vida, por cada dia que amanheço viva, por minha saúde, por minha família, por meus estudos, por me iluminar durante os momentos difíceis e me dar forças para superar os obstáculos quando se apresentam.

Ao meu orientador de mestrado, Josemar Rodrigues, pelas orientações valiosas, pela disponibilidade de tempo, pelos esclarecimentos das minhas dúvidas, pela confiança, compreensão, paciência e incentivo na elaboração e condução deste trabalho. Também pelos conhecimentos que me transmitiu incentivaram-me a aprofundar os meus estudos e aspirar um alto nível de proficiência nesta área.

Ao professor Vicente Cancho, pela explanação de ideias, ajuda na elaboração dos programas computacionais e pela disposição em me ajudar, sempre que necessário, na realização deste trabalho. À professora Estela Pereira Bereta, pelo carinho, solidariedade e apoio prestado.

Aos professores membros da banca pelas críticas e sugestões para o melhoramento deste trabalho.

Aos professores do Departamento de Estatística da Universidade Federal de São Carlos, em especial Galvão Leite, Carlos Diniz, Vera Tomazella, Luis Milan, Josemar Rodrigues, Marcio Diniz, pelas aulas ministradas do mestrado. À Isabel uma excelente secretária, por todas as informações e serviços oferecidos. Em geral, a todas as pessoas que me abriram as portas, me oferecendo um ambiente acolhedor e sadio para que eu pudesse realizar o meu mestrado nesta Universidade.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo auxílio financeiro concedido durante o período de mestrado.

À todas as pessoas que colaboraram para minha vinda ao Brasil, que acreditaram em mim para que eu pudesse realizar este trabalho.

Às minhas amigas muito queridas, Patricia Tacuri que me acolheu quando cheguei ao Brasil me dando muito apoio e exemplo de estudo, Angelica Ribeiro, verdadeira companheira de pesquisa, sempre gentil, alegre e presente, Verónica Quispe, pelos ânimos, conselhos de vida e sincera amizade, Eveline Soares Costa e família pela amizade e contribuição incondicional para minha estadia aqui no Brasil.

À todos os meus amigos peruanos por compartilhar momentos como em família em São Carlos, fazendo nossas reuniões inesquecíveis e divertidas.

À todos os meus colegas do departamento de Estatística, Silvia Maria Prado, Katiane Conceição, pelas correções de redação em português, Rubiane Maria Pires, Vinicius cokin, Jhon Bernedo, Patrick Borges, Victor Fossaluza, Vitor Marchi, Hugo Kegler dos Santos, Adriano Kamimura, Paulo Henrique Ferreira, Thiago Arrabal, todos eles sempre estiveram carinhosamente presentes contribuindo com sugestões, tirando minhas dúvidas, além da amizade incondicional.

E com muito amor,

À toda minha família, que sempre me apoiaram em tudo, em particular aos meus pais José Luis e Carolin e a minha irmã Karla, por todo amor que me dedicaram, porque nunca mediram esforços para os sucesso das suas filhas, pelo incentivo, pelas orações desde Perú e pelas palavras de apoio neste período de trabalho. Este sucesso é também vosso.

Abstract

The chemical induction of carcinogens in chemopreventive animal experiments is becoming increasingly frequent in biological research. The purpose of these biological experiments is to evaluate the effect of a particular treatment on the rate of tumors incidence in animals. In this work, the number of promoted tumors per animal will be parametrically modeled following the suggestions given by Kokoska (1987) and Freedman *et al.* (1993). The study of these chemopreventive experiments will be presented in the context of the destructive model proposed by Rodrigues *et al.* (2010) with terminal variable that allows or censors the experiment at time of the animal death. Since the data analyzed in this field are subject to excess of zeros (Freedman *et al.* (1993)), we propose for the number of promoted tumors a negative binomial distribution (NB), a zero-inflated Poisson distribution (ZIP), and a zero-inflated Negative Binomial distribution (ZINB). The selection of these models will be made through the likelihood ratio test and the AIC, BIC criteria. The estimation of its parameters will be obtained by using the method of maximum likelihood, and further simulation studies will also be realized. As a future proposition to finalize this project, it is suggested the Bayesian methodology as an alternative to the method of maximum likelihood via the EM algorithm.

Keywords: Destructive Model, Poisson Model, Negative Binomial Model, Zero-Inflated Poisson Model, Zero-Inflated Negative Binomial Model, Chemopreventive Experiments.

Resumo

A indução química de substâncias cancerígenas em experimentos quimiopreventivos em animais é cada vez mais frequente em pesquisas biológicas. O objetivo destes experimentos biológicos é avaliar o efeito de um determinado tratamento na taxa de incidência de tumores em animais. Neste trabalho o número de tumores promovidos por animal será modelado parametricamente seguindo as sugestões dadas por Kokoska (1987) e por Freedman *et al.* (1993). O estudo desses experimentos quimiopreventivos será apresentado no contexto do modelo destrutivo proposto por Rodrigues *et al.* (2010) com variável terminal que condiciona ou censura o experimento no instante de morte do animal. Os dados analisados possuem uma grande quantidade de zeros, portanto será proposto para o número de tumores promovidos as seguintes distribuições: binomial negativa, a distribuição de Poisson com zeros inflacionados e a distribuição binomial negativa com zeros inflacionados. A seleção destes modelos será feita através do teste da razão de verossimilhança e os critérios AIC, BIC. As estimativas dos respectivos parâmetros serão obtidas utilizando o método de máxima verossimilhança e serão feitos estudos de simulação. Para continuar este projeto, a proposta futura é utilizar a metodologia Bayesiana como alternativa ao método de máxima verossimilhança via algoritmo EM.

Palavras-chave: Modelo destrutivo, Modelo de Poisson, Modelo binomial negativo, Modelo de Poisson zero inflacionada, Modelo binomial negativa zero inflacionada, Experimentos quimiopreventivos.

Sumário

Lista de Tabelas	v
1 Introdução	1
1.1 Modelo destrutivo (Rodrigues <i>et al.</i> (2010))	3
1.2 Motivação	6
1.3 Propostas do trabalho	6
1.4 Apresentação dos capítulos	6
2 Revisão de literatura	7
2.1 Distribuição de Poisson	7
2.2 Distribuição binomial negativa (BN)	8
2.3 Número excessivo de zeros nos dados	8
2.3.1 Distribuição inflacionada de zeros (DIZ)	9
2.3.2 Distribuição de Poisson com zeros inflacionados (PZI)	10
2.3.3 Distribuição binomial negativa com zeros inflacionados (BNZI)	11
2.4 Inferência	12
2.4.1 Estimação de máxima verossimilhança	12
2.4.2 Métodos numéricos	13
2.5 Teste de hipóteses	14
2.6 Intervalos de confiança	15
2.7 Critérios de seleção do modelo	15
3 Modelo destrutivo com variável terminal	17
3.1 Formulação do modelo	17
3.1.1 Observações	18
3.1.2 Distribuição de probabilidades com variável terminal	19
3.1.3 Função de verossimilhança	19
3.2 Casos especiais do modelo destrutivo com variável terminal	21
3.2.1 Modelo destrutivo de Poisson com variável terminal	21
3.2.2 Modelo destrutivo binomial negativo com variável terminal	22
3.2.3 Modelo destrutivo Poisson com zeros inflacionados e variável terminal	23
3.2.4 Modelo destrutivo binomial negativo com zeros inflacionados e variável terminal	27
3.2.5 Estudo de simulação para o modelo destrutivo PZI com variável terminal	31
4 Aplicação com Dados Reais	34
5 Considerações Finais	38
A	40
A.1 Câncer, Carcinogênese (INCA, 2011)	40
A.2 Esperança e Variância do modelo destrutivo com variável terminal	41
A.3 condições de 1ª e 2ª ordem do modelo destrutivo de Poisson com variável terminal	41
A.4 Distribuição da variável destrutiva BN dada a variável terminal	43

A.4.1	Condições de 1 ^a e 2 ^a ordem do modelo destrutivo BN com variável terminal .	43
A.5	Distribuição da variável destrutiva PZI dada a variável terminal	45
A.5.1	Função de verossimilhança ampliada do modelo destrutivo PZI com variável terminal	46
A.6	Distribuição da variável destrutiva BNZI com variável terminal	46
A.6.1	Função de verossimilhança ampliada do modelo destrutivo BNZI com variável terminal	47
Referências		49

Lista de Tabelas

1.1	Tabela de dados.	5
2.1	Tipos de zeros em dados.	9
3.1	Modelos Paramétricos.	20
3.2	Média, Viés, EQM das estimativas de máxima verossimilhança para 1000 réplicas com alta proporção de zeros aleatórios $\omega = 0,7$	32
3.3	Média, Viés, EQM das estimativas de máxima verossimilhança para 1000 réplicas com baixa proporção de zeros aleatórios $\omega = 0,4$	32
3.4	Média, Viés, EQM das estimativas de máxima verossimilhança para 1000 réplicas, estimando com o modelo destrutivo de Poisson para dados gerados com $\omega = 0,7$ (proporção de zeros aleatórios).	33
3.5	Média, Viés, EQM das estimativas de máxima verossimilhança para 1000 réplicas, estimando com o modelo destrutivo de Poisson para dados gerados com $\omega = 0,4$ (proporção de zeros aleatórios).	33
4.1	Resultados do modelo destrutivo de Poisson.	35
4.2	Resultados do modelo destrutivo binomial negativo.	35
4.3	Resultados do modelo destrutivo de Poisson com zeros inflacionados.	35
4.4	Resultados do modelo destrutivo binomial negativo com zeros inflacionados.	36
4.5	Seleção de modelo para o grupo controle.	36
4.6	Seleção de modelo para o grupo com agente A.	36
4.7	Frequências esperadas de cada modelo	37

Capítulo 1

Introdução

Os experimentos quimiopreventivos de câncer em animais têm como objetivo a avaliação do efeito de um determinado tratamento no combate de tumores. Neste sentido, a pesquisa para estudar melhores alternativas de tratamento dos tumores cancerígenos têm promovido um grande desenvolvimento na área biomédica. O uso de animais em experimentos de laboratório é uma poderosa ferramenta para as descobertas científicas com benefícios diretos na medicina. As contribuições significativas fornecidas pelos experimentos ajudam no desenvolvimento de diversos tratamentos, novas drogas, novos métodos cirúrgicos, vacinas, terapia gênica etc., que têm como finalidade curar a doença e melhorar a qualidade de vida do paciente.

O desenvolvimento destes experimentos gera prós e contras dos testes em animais (experimentação animal). Os cientistas fundamentam que sem os testes, as prateleiras das farmácias estariam vazias e não haveria como criar novos remédios sem experimentá-los antes em animais de laboratório. Os cientistas ressaltam também que seguem rígidos protocolos e códigos de ética que garantem o bem-estar e impedem o sofrimento dos animais. Os defensores dos direitos dos animais repudiam esses testes, afirmando que são cruéis e inúteis, mas os cientistas argumentam que, sem eles, os avanços da medicina seriam fortemente prejudicados.

A quimioprevenção é uma estratégia para lutar contra o câncer baseada na utilidade de agentes químicos, naturais ou sintéticos, para reverter, suprimir ou prevenir uma progressão carcinogênica. Essa estratégia é realizada em experimentos com animais que servem como cobaias e consiste em administrar um agente quimiopreventivo durante as etapas iniciais da carcinogênese, ou seja, antes do estabelecimento do estágio de progressão (ver Apêndice A.1), a fim de testar o efeito ou o potencial da substância quimiopreventiva. Camundongos e ratos, por terem um ciclo de vida curto e fácil reprodução, são utilizados nestes experimentos. Antes de tudo, estes animais são infectados com substâncias cancerígenas para produzir tumores, após aplica-se o agente quimiopreventivo para depois avaliar a sua eficácia e toxicidade. Finalizado o experimento segue um estudo histológico dos animais após serem sacrificados, dependendo dos resultados, o tratamento com agente quimiopreventivo é descartado. Se o teste for bem sucedido segue-se uma última fase, quando será testado o tratamento em pacientes.

Neste trabalho é realizado um estudo estatístico para testar o efeito de uma substância quimiopreventiva na taxa de incidência de tumores em animais.

Estudos feitos por [Thompson *et al.* \(1984\)](#), em que foi estudado a inibição de tumores mamários em ratos fêmeas, dois experimentos quimiopreventivos foram realizados: no primeiro experimento foi testado duas substâncias compostas quimiopreventivas, *selênio com sodium selenite (a)* e *selê-*

ção com selenometrionide (b) para eliminar o efeito do carcinógeno *1-Methyl-1-nitrosourea* (MNU) que foi inicialmente induzido nos animais. No segundo experimento foram utilizadas as substâncias quimiopreventivas anteriores contra um outro tipo de carcinógeno chamado *7-12-dimethylbenzanthraceno* (DMBA). A substância (a) foi um agente quimiopreventivo com mais eficácia no combate ao câncer, mostrando significativamente que houve uma redução no número de tumores por rato nos dois experimentos. No entanto, a substância (b) causou lesões graves e sérios danos vitais nos animais. Na análise estatística, os efeitos das substâncias selênio foram testadas via análise de regressão e a incidência do tumor foi testado usando o teste χ^2 .

No estudo de Kokoska (1987), o experimento foi primordialmente designado à pesquisa do efeito do tratamento quimiopreventivo sobre a taxa de incidência de tumores mamários em animais. Como variáveis respostas foram considerados o número de tumores e as respectivas taxas. O modelo matemático usado para caracterizar o experimento foi baseado nos dois primeiros estágios da carcinogênese, iniciação e promoção, que são resultados da exposição carcinogênica em que tumores malignos localizados nas glândulas mamárias são observados quando eles evoluem para um tamanho detectável por apalpação. Neste experimento os tempos de detecção dos tumores são censurados pelo tempo de laboratório pré-estabelecido. Muitos desses ratos morrem durante o experimento e são eliminados gerando conclusões viciadas (vide Kokoska (1988)). Foi utilizada a distribuição de Poisson para modelar o número de tumores, e a distribuição Gama para os tempos de detecção. Os parâmetros das distribuições descritas foram estimados através do método de máxima verossimilhança e foram construídos os seus respectivos intervalos de confiança. Para comparar os efeitos de dois tratamentos diferentes utilizaram o teste da razão de verossimilhança.

Boone *et al.* (1990) trabalhou na identificação de agentes quimiopreventivos em diferentes tipos de câncer: pele (papilomas), mama, bexiga, cólon e pulmão. Os cânceres de pele, mama, bexiga e cólon foram estudados em ratos e o câncer de pulmão em hamsters. Algumas avaliações de ensaios clínicos foram feitos em humanos. Boone *et al.* (1990) testou 16 agentes quimiopreventivos de 105 combinações de compostos. No artigo foi apresentada a substância DFMO (*difluorometilornitina*) como agente quimiopreventivo contra a substância carcinogênica DMBA (*7-12-dimetilbenzantraceno*) na avaliação de tumores de mama em ratos fêmeas. Foram realizadas três formas de testar o efeito do agente quimiopreventivo. O agente DFMO é aplicado aos animais durante a indução do carcinógeno DMBA; DFMO é aplicado somente depois do carcinógeno DMBA e DFMO é administrado durante e depois do DMBA. Destas três formas, o último caso resultou a forma mais eficiente de administrar um agente de quimioprevenção contra um carcinógeno, apresentando uma redução do número de mutações (células iniciadas). Por outro lado, a administração depois do DMBA, retardou o desenvolvimento desses tumores. No segundo caso, o agente DFMO atua como um agente de bloqueio da promoção dos tumores e no primeiro caso DMFO atua como um agente de bloqueio das mutações. Foi questionado se o agente quimiopreventivo DFMO possui ambos mecanismos de bloqueio, um só, ou nenhum deles. Esta questão é direcionada a uma análise estatística que possa distinguir entre estes mecanismos para avaliar o efeito de um determinado tratamento quimiopreventivo. Cabe ressaltar que o estudo está baseado no número de tumores desenvolvidos durante o experimento e no tempo de aparecimento de cada tumor em cada animal. Estes dados podem ser observados apenas em alguns tipos de câncer como mama e pele, entretanto em outros tipos de câncer as informações são obtidas somente no final do experimento quando o animal é sacrificado.

Freedman *et al.* (1993) estendeu o modelo de Kokoska (1987) correspondente ao número de

tumores e seus tempos de detecção. Eles descobriram que os resultados são sensíveis aos modelos paramétricos em particular a distribuição do número de tumores promovidos por animal. Para a escolha das distribuições utilizaram o teste χ^2 . Nesse trabalho é considerado que os tempos de detecção são independentes entre eles e são identicamente distribuídos dentro de um grupo de animais em tratamento, uma suposição que não é realista devido à variabilidade alta entre animais. As estimativas dos parâmetros foram obtidas pelo método de máxima verossimilhança.

Dado que os animais podem morrer antes de todos os tumores induzidos chegarem a um tamanho detectável, a separação desses efeitos do tratamento pode ser difícil. [Dunson & Dinse \(2000\)](#), interessados em distinguir estes efeitos em multiplicidade tumoral e taxa de crescimento dos tumores em experimentos de quimioprevenção, desenvolveram uma metodologia bayesiana, via algoritmo *Gibbs sampling*, para a estimativa dos parâmetros. Os métodos foram ilustrados através da aplicação de dados de um experimento de quimioprevenção do câncer de mama.

Trabalhos de [Kodell & Chen \(2001\)](#) e [Molefe et al. \(2005\)](#) também foram baseados no trabalho de [Kokoska \(1987\)](#), todos esses trabalhos dos autores mencionados têm em comum o estudo de modelos paramétricos para a análise dos dados. Foram consideradas as distribuições de Poisson e binomial negativa para o número de tumores promovidos e as distribuições Gama e Weibull para os tempos de detecção dos tumores. Cabe ressaltar que nesse tipo de experimentos é comum encontrar uma grande quantidade de zeros nos conjuntos de dados, pois existe uma grande quantidade de animais que não apresentam tumores após a indução e o tratamento quimiopreventivo, até o instante de morte. Portanto uma das propostas deste trabalho é considerar a presença excessiva de zeros, que não foi abordado pelos esses autores.

1.1 Modelo destrutivo ([Rodrigues et al. \(2010\)](#))

O trabalho de [Rodrigues et al. \(2010\)](#) foi uma extensão da proposta de [Yakovlev & Tsodikov \(1996\)](#), em que as células iniciadas ou contaminadas passam por um processo destrutivo, e apenas parte delas sobreviveu a uma intervenção. Portanto, o modelo destrutivo foi formulado para descrever matematicamente um processo carcinogênico ao nível de células. Existem fatores de risco que interatuam nas células de um indivíduo resultando uma quantidade de células "iniciadas" dispostas a virar células malignas (câncer). Estas células entram em um processo de destruição após um tratamento ou sistema de imunidade capaz de reparar as células. Como consequência, têm-se a formação ou não de lesões pré-cancerosas no genoma das células que são células danificadas ou malignas que sobreviveram ao tratamento contra o câncer.

Seja $M = m$ uma variável aleatória latente representando o número de células "iniciadas" e $X_j, j = 1, \dots, m$ variáveis aleatórias independentes, independentes de M , seguindo uma distribuição de Bernoulli com probabilidade de sucesso p , indicando se a j -ésima célula iniciada está ou não danificada (ou seja se foi ou não reparada). A variável J representa o número total entre as M células iniciadas que não foram eliminadas pelo sistema imunológico ou tratamento, definida como:

$$J = \begin{cases} X_1 + X_2 + \dots + X_M & , \text{ para } M > 0, \\ 0 & , \text{ para } M = 0. \end{cases} \quad (1.1)$$

Devido à destruição ou não reparo das células, temos que $J \leq M$. O processo destrutivo (mecanismo de destruição) é caracterizado pela distribuição condicional de $J = a$ dado $M = m$, dada

por:

$$P[J = a|M = m] = \binom{m}{a} p^a (1-p)^{m-a}, \quad (1.2)$$

em que $0 < p < 1$.

A distribuição do número total de células danificadas que não foram eliminadas pelo tratamento pode ser calculada por:

$$P[J = a] = \sum_{m=0}^{\infty} P[J = a|M = m]P[M = m], \quad (1.3)$$

em que $P[M = m]$ é a função de probabilidade da variável latente M .

A proposta deste trabalho será utilizar o modelo proposto Kokoska (1987) e Freedman *et al.* (1993), no contexto do modelo destrutivo formulado por Rodrigues *et al.* (2010), para experimentos quimiopreventivos em animais.

A principal diferença entre a nossa proposta de trabalho e a do modelo destrutivo formulado por Rodrigues *et al.* (2010) é que a modelagem será desenvolvida para variáveis observáveis e não com variáveis latentes. Como exemplo de aplicação será utilizado o banco de dados fornecido por Freedman *et al.* (1993) (ver Tabela 1.1), em que os dados são:

- i*) O número de tumores detectados em cada animal,
- ii*) o tempo de detecção do tumor gerado,
- iii*) o tempo de morte do animal, variável terminal.

Tabela 1.1: Tabela de dados.

Tratamento controle		Tratamento agente A		
Tempo dos tumores (dias)	t_i^*	Tempos dos tumores (dias)	t_i^*	t_i^*
0	92	0	135	
0	185	0	206	
0	191	0	217	
0	200	0	247	
0	203	0	260	
0	208	0	268	
0	214	0	272	
0	276	0	280	
0	279	0	280	
0	279	48	204	
0	279	70	186	
0	279	113	212	
0	279	120	145	
34	66	135	230	
58	232	176	280	
85	234	280 ^a	280	
92	279	280 ^a	280	
155	279	34	280 ^a	
279 ^a	279	98	280	
280 ^a	280	98	246	
280 ^a	280	148	231	
62	78	155	230 ^a	
85	279	225	255 ^a	
183	279 ^a	141	155	219 ^a
231	279 ^a	148	231	239
62	70	70	85	85
92	170 ^a	98	126	178 ^a
120	113	92	148	233
113	155	92	176	255 ^a
113	135	58	64	211
92	113	73	85	98
	120	92	85	126 ^a
	197	92	98	162
	203 ^a	92	113	155
			113	162
			135	190
			159	169
			162	231
			267 ^a	267 ^a
			267 ^a	267 ^a

^a indica quando o tumor foi descoberto na necropsia, e t_i^* é o tempo de morte do animal.

1.2 Motivação

A principal motivação para a realização deste trabalho é a reformulação do modelo destrutivo [Rodrigues *et al.* \(2010\)](#) para experimentos com animais censurados ou condicionados na variável terminal que é o instante de morte durante o experimento. Usualmente, os dados deste tipo de experimentos apresentam excesso de zeros e o modelo de Poisson não é apropriado. Portanto, será introduzido os modelos alternativos, conhecidos como modelo de Poisson com zeros inflacionados e binomial negativo com zeros inflacionados, para levar em consideração a dispersão e o excesso de zeros.

1.3 Propostas do trabalho

Os objetivos desta pesquisa são:

1. Formular o modelo proposto por [Kokoska \(1987\)](#) e [Freedman *et al.* \(1993\)](#) do ponto de vista do modelo destrutivo ([Rodrigues *et al.* \(2010\)](#)).
2. Estudar o comportamento das distribuições Poisson, binomial negativa, Poisson com zeros inflacionados (PZI) e binomial negativa com zeros inflacionados (BNZI) na análise de experimentos quimiopreventivos de câncer em animais.
3. Desenvolver um algoritmo computacional para simular dados de acordo com o modelo proposto.
4. Avaliar, utilizando dados simulados, as propriedades dos estimadores de máxima verossimilhança.
5. Aplicar a metodologia desenvolvida em um conjunto de dados reais.

Para atingir estes objetivos utilizaremos o método de máxima verossimilhança via algoritmo EM.

1.4 Apresentação dos capítulos

O Capítulo 2 traz uma revisão de literatura, envolvendo alguns tópicos necessários para o desenvolvimento do nosso trabalho. O Capítulo 3 apresenta o modelo destrutivo com variável terminal e alguns casos particulares. Também será mostrado um estudo de simulação do modelo proposto, em que o algoritmo computacional mostrou-se consistente. O Capítulo 4 apresenta os resultados de uma aplicação do modelo destrutivo com variável terminal com dados reais. Finalmente, no Capítulo 5 são discutidas as conclusões e algumas perspectivas futuras.

Capítulo 2

Revisão de literatura

Em muitas aplicações envolvendo dados de contagens são utilizados modelos discretos que foram largamente desenvolvidos na literatura. Os modelos mais utilizados são Poisson e binomial negativo (ver [Kokoska \(1987\)](#) e [Freedman *et al.* \(1993\)](#)). Estes modelos não consideram a possibilidade de excesso de zeros. Portanto, serão introduzidas as distribuições inflacionadas, Poisson com zeros inflacionados e binomial negativa com zeros inflacionados, que levam em consideração a dispersão e o excesso de zeros nos dados. Serão apresentados conceitos básicos de inferência clássica.

2.1 Distribuição de Poisson

O modelo de Poisson é amplamente utilizado na modelagem de dados discretos e foi introduzido pelo matemático francês Siméon Denis Poisson (1781–1840). Segundo [Johnson \(1969\)](#), uma variável aleatória M tem distribuição de Poisson com parâmetro λ se sua função de probabilidade é dada por:

$$P(M = m; \lambda) = \frac{e^{-\lambda} \lambda^m}{m!}, \quad m = 0, 1, 2, \dots, \quad (2.1)$$

em que $\lambda > 0$. A média e a variância da distribuição de Poisson são dadas por:

$$E(M) = Var(M) = \lambda. \quad (2.2)$$

A variável aleatória M é o número de ocorrências de um evento em um intervalo de tempo ou espaço específico, em que λ é o parâmetro que corresponde ao número médio de eventos ocorrendo nesse intervalo. Um exemplo típico de aplicação da distribuição de Poisson é o número de chamadas recebidas por uma central telefônica num intervalo de tempo fixo. Entretanto, existem diversas outras aplicações envolvendo dados de contagem, como o controle de qualidade (contagem de defeitos de um produto), demografia (número de filhos por casal), medicina (número de recorrências de um determinado tipo de doença), ecologia (número de ovos do mosquito), estudos sociais (número de parceiros de indivíduos homossexuais), etc.

A distribuição de Poisson apresenta as seguintes características:

1. O parâmetro λ é a média e variância da distribuição.
2. Em dados reais, muitas variáveis de contagens têm uma variância maior do que a média, o que é chamado de superdispersão. Neste caso a distribuição de Poisson não é recomendável.

3. Quando λ aumenta, a distribuição de Poisson aproxima-se de uma distribuição Normal.

2.2 Distribuição binomial negativa (BN)

A distribuição binomial negativa é conhecida na literatura como distribuição de Pascal ou distribuição de Polya. Por exemplo, considere um experimento estatístico com dois resultados possíveis, sucesso ou fracasso, em que o sucesso ocorre com probabilidade ν e o fracasso com probabilidade $1 - \nu$. Se o experimento é repetido indefinidamente e os ensaios são independentes, a variável aleatória M indica o número de tentativas necessárias para obter N sucessos, sendo a última tentativa um sucesso. A distribuição binomial negativa pode ser escrita de diversas maneiras dependendo da parametrização utilizada. Nesta seção a distribuição binomial negativa será utilizada na versão paramétrica de [Nelder & Wedderburn \(1972\)](#) dada por:

$$P(M = m; N, \nu) = \frac{\Gamma(N + m)}{\Gamma(N)m!} \nu^N (1 - \nu)^m. \quad (2.3)$$

A média e a variância são dadas pelas seguintes expressões:

$$\begin{aligned} E(M) &= N(1 - \nu)/\nu = \mu, \\ Var(M) &= \mu + \tau\mu^2, \end{aligned} \quad (2.4)$$

em que $\tau = 1/N$ e N é o parâmetro de dispersão, para $N > 0$, e $0 < \nu < 1$.

Observa-se que a variância da distribuição binomial negativa tem um termo adicional $\tau\mu^2$ quando comparado com a variância da Poisson, que será útil para ter um melhor ajuste quando existe superdispersão nos dados.

2.3 Número excessivo de zeros nos dados

[Silva \(2009\)](#) mostra que o número excessivo de zeros nos dados dificulta a elaboração de uma análise estatística para o problema, pois os modelos usuais não conseguem modelar a presença excessiva de zeros. Por exemplo, em uma linha de produção em que se aplica controle de qualidade, a contagem de defeitos de um produto apresenta-se cada vez menor, ou seja, há um grande número de zeros, isto ocorre devido a melhoria da qualidade e controle dos processos de fabricação. Logo os zeros correspondem a zeros verdadeiros ou chamados também de determinísticos. Na área de ecologia, os zeros podem ocorrer devido à espécie estando presente ou ausente na área amostrada e que não foi observada pelo pesquisador, logo, os zeros são aleatórios (falsos). Portanto, em um processo de contagem os zeros podem ser obtidos de forma determinística ou aleatória. Na grande maioria de situações práticas, não é possível separar estes dois tipos de zeros.

Os zeros inflacionados ocorrem de duas formas para cada tipo. O primeiro tipo de zero (determinístico) *surge de uma baixa frequência de ocorrência ou ser o resultado de um efeito que conduz para locais que não têm nenhum indivíduo presente*. O segundo tipo de zero (aleatório) é causado por: *apesar de existir, ocupar o local, não estava presente na hora da pesquisa*. ou *acontece quando a unidade de estudo ocupa o local e está presente na hora de amostrar; mas o pesquisador não o encontra*. ([Silva \(2009\)](#) apud [Martin et al. \(2005\)](#))

A Tabela 2.1 apresenta um resumo sobre os tipos de zeros e suas respectivas definições.

Tabela 2.1: *Tipos de zeros em dados.*

Tipo de zero	Descrição
Zero determinístico	<ul style="list-style-type: none"> • Baixa frequência de ocorrência. • Realmente não existe nenhum indivíduo presente.
Zero aleatório	<ul style="list-style-type: none"> • Apesar de existir e ocupar o local, não estava presente na hora da pesquisa. • Acontece quando ocupa o local e está presente na hora de amostrar, mas o pesquisador não o encontra.

Como exemplo, podemos supor que para os tumores cancerígenos em animais, os zeros determinísticos surgem devido à baixa ocorrência de tumores após um tratamento eficiente, ou os tumores estão temporariamente ausentes e não surgiram durante o tempo do experimento. Também podemos supor que se o tempo de período fosse grande a ausência dos tumores constituiria um zero aleatório.

Em um conjunto de dados é possível ocorrer diferentes tipos de zeros: as informações podem não conter nenhum zero inflacionado, ou ter zeros inflacionados devido a zeros verdadeiros ou devido a aleatórios; ou ainda, devido a excesso de ambos os casos, verdadeiros e aleatórios. Concluindo, existe uma incerteza sobre a fonte dos zeros inflacionados. A seguir apresentamos algumas sugestões:

1. Em um conjunto de dados, quando os zeros são determinísticos, é recomendável a utilização de uma média ponderada de duas distribuições, ou seja, usa uma distribuição degenerada no ponto zero e uma distribuição discreta adequada aos não zeros.
2. Se os zeros aleatórios estão presentes nos dados, a modelagem também pode ser feita através de misturas de modelos. Em muitas aplicações as distribuições zero inflacionadas são soluções possíveis.

2.3.1 Distribuição inflacionada de zeros (DIZ)

Muitos autores propuseram modelos diferentes, mas semelhantes para a análise de dados com zeros inflacionados, baseando-se no uso de mistura de distribuições. Estes modelos são capazes de incorporar o excesso de zeros nos dados. Os modelos inflacionados de zeros têm sido usados para modelagem de dados nas mais diversas áreas, tais como Epidemiologia, Sociologia, Medicina, Odontologia, Agricultura, Econometria dentre outras.

Lambert (1992), foi a pioneira no uso de modelos de regressão para contagens com distribuição inflacionada de zeros. A autora propôs uma mistura finita das distribuições Bernoulli e Poisson para modelar o excesso de dados de contagem envolvendo defeitos de fabricação de placas de circuito impresso. Essa metodologia, aplicada a dados de contagem, mostrou-se de grande utilidade devido ao fato de ser possível associar regressores que influenciam no parâmetro de mistura.

Um conjunto de dados, às vezes, apresenta variabilidade maior do que a esperada pelos modelos probabilísticos padrões. Esse fenômeno é conhecido como superdispersão, podendo ocorrer devido a

uma variabilidade da média, excessos de zeros, correlação entre indivíduos e/ou omissão de variáveis não observadas. Exemplos interessantes sobre esse tipo de dados com zeros inflacionados podem ser encontrados em Lambert (1992), Gupta *et al.* (1996), Ridout *et al.* (1998), Bohning *et al.* (2000) Kumara & Chin (2003), Borgatto (2004) entre outros. Rodrigues (2003) e Saito (2005) abordam o tema do ponto de vista bayesiano.

As motivações para usar os modelos com zeros inflacionados são a simplicidade para indicar alguma heterogeneidade não observada nos dados. A função de probabilidade de uma variável discreta com zeros inflacionados pode ser representada pela seguinte expressão (Saito (2005)):

$$P(M; \phi, \omega) = (1 - \omega)I_{\{0\}}(M) + \omega P(M; \phi) \quad M = 0, 1, 2, \dots, \quad (2.5)$$

em que

$$I_{\{0\}}(M = m) = \begin{cases} 1 & , m = 0 \\ 0 & , m \neq 0 \end{cases},$$

e $0 \leq \omega \leq 1$ e $P(M; \phi)$ é uma distribuição de probabilidade discreta, com vetor de parâmetros ϕ que teoricamente se adequaria aos dados caso não houvesse a presença excessiva de zeros nos dados. O parâmetro ω é a proporção de zeros aleatórios gerados pela distribuição $P(M; \phi)$ e $1 - \omega$ é a proporção de zeros de forma determinística. Na equação (2.5) apresenta uma mistura, da distribuição degenerada no zero e a distribuição da função de probabilidade $P(M; \phi)$.

2.3.2 Distribuição de Poisson com zeros inflacionados (PZI)

Seja M uma variável aleatória representando o número de tumores promovidos no animal. A distribuição de Poisson com zeros inflacionados é definida como

$$P(M = m; \omega, \lambda) = (1 - \omega)I_{\{0\}}(m) + \omega \frac{e^{-\lambda} \lambda^m}{m!} \quad m = 0, 1, 2, \dots, \quad (2.6)$$

em que

$$I_{\{0\}}(m) = \begin{cases} 1 & , m = 0 \\ 0 & , m \neq 0 \end{cases},$$

$0 \leq \omega \leq 1$, $\lambda > 0$ e $(1 - \omega)$ é a proporção de zeros não explicada pelo modelo de Poisson. A partir da função de probabilidade do modelo PZI temos que a média é dada por

$$\begin{aligned} E(M) &= \sum_{m=0}^{\infty} m P(M, \omega, \lambda) = \sum_{m=1}^{\infty} m \omega \frac{e^{-\lambda} \lambda^m}{m!}, \\ &= \omega e^{-\lambda} \sum_{m=1}^{\infty} m \frac{\lambda^m}{m!} = \omega e^{-\lambda} \lambda e^{\lambda} = \omega \lambda. \end{aligned} \quad (2.7)$$

Como,

$$\begin{aligned}
 E(M^2) &= \sum_{m=0}^{\infty} m^2 P(M, \omega, \lambda) = \sum_{m=1}^{\infty} m^2 \omega \frac{e^{-\lambda} \lambda^m}{m!}, \\
 &= \omega e^{\lambda} \sum_{m=1}^{\infty} m^2 \frac{\lambda^m}{m!} = \omega e^{-\lambda} \lambda e^{\lambda} (\lambda + 1), \\
 &= \omega (\lambda^2 + \lambda) = \omega \lambda (\lambda + 1),
 \end{aligned} \tag{2.8}$$

em que $0 < \omega < 1$, a variância da variável M é dada por:

$$\begin{aligned}
 Var(M) &= E(M^2) - (E(M))^2 = [\omega (\lambda^2 + \lambda)] - [\omega \lambda]^2, \\
 &= \omega \lambda^2 + \omega \lambda - \omega^2 \lambda^2, \\
 &= \omega \lambda [1 + \lambda(1 - \omega)].
 \end{aligned} \tag{2.9}$$

Quando $\omega = 1$, o modelo PZI corresponde ao modelo de Poisson com $E(M) = Var(M) = \lambda$.

2.3.3 Distribuição binomial negativa com zeros inflacionados (BNZI)

Seja M uma variável aleatória latente (não observável) representando o número de tumores promovidos no animal. A distribuição binomial negativa com zeros inflacionados é definida como:

$$P(M = m; \omega, N, \nu) = (1 - \omega) I_{\{0\}}(m) + \omega \frac{\Gamma(N + m)}{\Gamma(N) m!} \nu^N (1 - \nu)^m \quad m = 0, 1, 2, \dots, \tag{2.10}$$

em que

$$I_{\{0\}}(m) = \begin{cases} 1 & , m = 0 \\ 0 & , m \neq 0 \end{cases},$$

$0 \leq \omega \leq 1$, $0 < \nu < 1$ e $N > 0$ é o parâmetro de dispersão. A partir do modelo binomial negativo com zeros inflacionados temos que a média é dada por:

$$\begin{aligned}
 E(M) &= \sum_{m=0}^{\infty} m P(M; \omega, N, \nu) = \frac{\omega \nu^N}{\Gamma(N)} \sum_{m=1}^{\infty} m \frac{\Gamma(N + m)}{m!} (1 - \nu)^m, \\
 &= \frac{\omega \nu^N (1 - \nu)}{\Gamma(N)} \sum_{m=1}^{\infty} \frac{\Gamma(N + 1 + m - 1)}{(m - 1)!} (1 - \nu)^{m-1}, \\
 &= \frac{\omega \nu^N (1 - \nu) \Gamma(N + 1)}{\Gamma(N) \nu^{N+1}}, \\
 &= \omega \frac{N(1 - \nu)}{\nu} = \omega \mu.
 \end{aligned} \tag{2.11}$$

Como,

$$\begin{aligned}
E(M^2) &= \sum_{m=0}^{\infty} m^2 P(M; \omega, N, \nu), \\
&= \frac{\omega \nu^N (1 - \nu)}{\Gamma(N)} \sum_{m=1}^{\infty} m \frac{\Gamma(N + 1 + m - 1)}{(m - 1)!} (1 - \nu)^{m-1}, \\
&= \frac{\omega \nu^N (1 - \nu)}{\Gamma(N)} \frac{\Gamma(N + 1)(N - \nu N + 1)}{\nu^{N+2}}, \\
&= \frac{\omega N (1 - \nu)}{\nu} \left(\frac{N(1 - \nu)}{\nu} + \frac{1}{\nu} \right), \\
&= \omega \mu (1 + \mu + \mu \tau),
\end{aligned}$$

em que μ e τ são definidas em (2.4), a variância da variável M é dada por:

$$\begin{aligned}
Var(M) &= E(M^2) - (E(M))^2, \\
&= \omega \mu (1 + \mu + \mu \tau) - \omega^2 \mu^2, \\
&= \omega \mu (1 + \mu(1 - \omega) + \mu \tau).
\end{aligned} \tag{2.12}$$

Se $\omega = 1$, o modelo BNZI corresponde ao modelo binomial negativo com $E(M) = \mu$ e $Var(M) = \mu + \tau \mu^2$.

2.4 Inferência

2.4.1 Estimação de máxima verossimilhança

Pelo método de máxima verossimilhança (Bolfarine & Sandoval (2001)), as estimativas dos parâmetros serão obtidas a partir da maximização da função de verossimilhança.

Seja $L(\psi; D)$ a função de verossimilhança, em que $\psi = (\psi_1, \psi_2, \dots, \psi_k)$ é o vetor de parâmetros do modelo e D os dados. O estimador de máxima verossimilhança de ψ é o valor $\hat{\psi}$ que maximiza a função de verossimilhança $L(\psi; D)$. O logaritmo da função de verossimilhança de ψ é dada por:

$$l(\psi; D) = \log L(\psi; D). \tag{2.13}$$

O estimador de máxima verossimilhança pode ser encontrado seguindo as condições maximização de 1ª e 2ª ordem dadas por:

1. Condição de 1ª ordem

$$l'(\psi; D) = \frac{\partial l(\psi; D)}{\partial \psi} = 0, \tag{2.14}$$

sendo que a solução da equação de verossimilhança pode ser obtida explicitamente ou, em situações mais complicadas, a solução (estimativas) de (2.14), $\hat{\psi} = (\hat{\psi}_1, \hat{\psi}_2, \dots, \hat{\psi}_k)$, será em geral obtida por procedimentos numéricos.

2. Condição de 2ª ordem

A condição de 2ª ordem consiste em verificar se a matriz hessiana dada por:

$$H(\psi) = \begin{bmatrix} \frac{\partial^2 l}{\partial \psi_1^2} & \frac{\partial^2 l}{\partial \psi_1 \partial \psi_2} & \cdots & \frac{\partial^2 l}{\partial \psi_1 \partial \psi_k} \\ \frac{\partial^2 l}{\partial \psi_2 \partial \psi_1} & \frac{\partial^2 l}{\partial \psi_2^2} & \cdots & \frac{\partial^2 l}{\partial \psi_2 \partial \psi_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 l}{\partial \psi_k \partial \psi_1} & \frac{\partial^2 l}{\partial \psi_k \partial \psi_2} & \cdots & \frac{\partial^2 l}{\partial \psi_k^2} \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1k} \\ h_{21} & h_{22} & \cdots & h_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ h_{k1} & h_{k2} & \cdots & h_{kk} \end{bmatrix} \quad (2.15)$$

é negativa definida. Uma matriz simétrica é negativa definida se

$$|h_{11}| < 0, \begin{vmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{vmatrix} > 0, \dots, (-1)^n \begin{vmatrix} h_{11} & \cdots & h_{1k} \\ \vdots & \ddots & \vdots \\ h_{k1} & \cdots & h_{kk} \end{vmatrix} > 0,$$

em que $|A|$ denota o determinante da matriz A .

2.4.2 Métodos numéricos

• Método de Newton-Raphson

Vamos a denotar por $U(\psi)$ a função escore, ou seja,

$$U(\psi) = \frac{\partial l(\psi; D)}{\partial \psi}, \quad (2.16)$$

com $U(\hat{\psi}) = 0$. Expandindo $U(\hat{\psi}) = 0$ em série de Taylor em torno de um ponto ψ_0 , obtemos:

$$\hat{\psi} \cong \psi^0 - \frac{U(\psi^0)}{U'(\psi^0)}. \quad (2.17)$$

De (2.17), obtemos o procedimento iterativo de Newton-Raphson dado por

$$\psi^{K+1} = \psi^K - \frac{U(\psi^K)}{U'(\psi^K)}, \quad (2.18)$$

em que ψ^0 é o valor inicial de um novo valor ψ^1 obtido a partir de (2.18) e assim por diante, até que o processo se estabiliza, ou seja, para um dado ϵ pequeno, $|\psi^{K+1} - \psi^K| < \epsilon$. Nesse caso, o ponto $\hat{\psi}$ em que o processo se estabiliza é o estimador de máxima verossimilhança de ψ (vide [Bolfarine & Sandoval \(2001\)](#)).

• Algoritmo EM

A maximização da log-verossimilhança dos dados observados $l(\psi; D)$ pode ser obtida através do algoritmo *EM* (vide [Dempster et al. \(1977\)](#)). É um algoritmo iterativo, utilizado para maximizar probabilidades complexas e aplicado em problemas de dados incompletos. Os dados incompletos são definidos pela falta de dados, elementos desconhecidos, variáveis latentes, tipos de censura de observações, e assim por diante. Seja D_a os dados ampliados em que $l(\psi; D_a)$ é o logaritmo da função de máxima verossimilhança dos dados ampliados. A seguir apresentamos os passos do algoritmo *EM*.

Seja $\psi^{(0)}$ o valor inicial para $\psi = (\psi_1, \dots, \psi_\kappa)$, onde κ é o número de parâmetros a ser estimado. A obtenção da K -ésima iteração o algoritmo é formado pelos seguintes passos:

1. **Passo E.** Primeiro calcula-se a esperança condicional do logaritmo da função de máxima verossimilhança para dados ampliados

$$Q(\psi, \psi^{(K)}) = E_{\psi^{(K)}}[\log L(\psi; D_a)|D]$$

2. **Passo M.** Maximização de $Q(\psi, \psi^{(K)})$ para obter o valor $\psi^{(K+1)}$, i.e.,

$$Q(\psi^{(K+1)}, \psi^{(K)}) \geq Q(\psi, \psi^{(K)}) \quad \forall \psi$$

As estimativas de máxima verossimilhança ψ são então obtidas com a convergência do algoritmo, segundo o critério de parada estabelecido. Consideramos aqui como critério a condição

$$\left| \psi^{(K+1)} - \psi^{(K)} \right| < \epsilon,$$

para um $\epsilon > 0$.

2.5 Teste de hipóteses

Neste tipo de experimentos, geralmente são testadas diferentes substâncias quimiopreventivas por grupos de animais, gerando diferentes amostras independentes. A seguir formularemos primeiramente um teste de hipóteses que será utilizado para comparar a eficiência das substâncias quimiopreventivas de cada tratamento (vide Kokoska (1988)).

Seja $G \geq 2$ os tratamentos aplicados em diferentes grupos de animais, suponhamos que desejamos testar as seguintes hipóteses

$$\begin{cases} H_0 & : \text{o vetor de parâmetros de interesse, todos ou alguns, são iguais para todos os tratamentos;} \\ H_a & : \text{ao menos dois vetores de parâmetros são diferentes.} \end{cases} \quad (2.19)$$

Seja $\hat{\eta}_i, \hat{\theta}_i, 1 \leq i \leq G$ as estimativas de máxima verossimilhança de cada grupo de animais com diferentes tratamentos, em que $\hat{\eta}$ se refere a estimativa do vetor de parâmetros da distribuição do número de tumores, $\hat{\theta}$ é a estimativa do vetor de parâmetros da distribuição do tempo de detecção dos tumores e $\hat{\eta}_c, \hat{\theta}_c$ as estimativas de máxima verossimilhança dos dados combinados.

A função verossimilhança sob H_0 é dada por $L_{H_0} = L(\hat{\eta}_c, \hat{\theta}_c)$. A função verossimilhança sob H_a é dada por:

$$L_{H_a} = \prod_{i=1}^G L(\hat{\eta}_i, \hat{\theta}_i). \quad (2.20)$$

Teste da razão de verossimilhança (TRV)

O teste da razão de verossimilhança consiste em comparar os valores maximizados do logaritmo da função de verossimilhança sem restrição e sob a hipótese nula H_0 . O TRV é definido pela estatística:

$$TRV = \chi^2 = -2 \log(r), \quad \text{em que } r = L_{H_0}/L_{H_a}.$$

Se $\chi^2 \geq \chi_{k(G-1)}^2(\alpha')$, rejeita-se a hipótese nula a um nível de significância α' em que $k(G-1)$ é o número de graus de liberdade e k é o número de parâmetros sob a hipótese H_0 .

Se o teste resulta significativo, isto é, a hipótese nula será rejeitada então tem-se aqueles tratamentos que fazem efeito no experimento.

2.6 Intervalos de confiança

A ideia de estimação por intervalos é construir um intervalo em torno da estimativa pontual tal que tenha uma probabilidade conhecida de conter o verdadeiro valor do parâmetro. Neste trabalho utilizaremos as distribuições amostrais das estimativas dos parâmetros desconhecidos baseado na teoria assintótica. A seguir, introduzimos as ferramentas principais para o cálculo de intervalos de confiança. Seja D a variável que representa os dados observados com função de densidade $f(D; \psi)$, onde $\psi = (\psi_1, \dots, \psi_\kappa)$. A matriz de informação esperada de Fisher do vetor paramétrico ψ , através de D , é definida por:

$$I(\psi) = E \left[-\frac{\partial^2 \log f(\psi; D)}{\partial \psi \partial \psi'} \right]. \quad (2.21)$$

A definição acima é uma medida de informação global, enquanto na prática trabalha-se com uma medida de informação local, chamada de matriz de informação observada de Fisher, definida como:

$$J(\psi) = -H(\psi) = -\frac{\partial^2 \log f(\psi; D)}{\partial \psi \partial \psi'}, \quad (2.22)$$

onde $H(\psi)$ é uma matriz hessiana dada em (2.15). A inversa da matriz de informação observada é definida como:

$$J^{-1}(\psi) = \begin{bmatrix} Var(\hat{\psi}_1) & Cov(\hat{\psi}_1, \hat{\psi}_2) & \cdots & Cov(\hat{\psi}_1, \hat{\psi}_\kappa) \\ Cov(\hat{\psi}_2, \hat{\psi}_1) & Var(\hat{\psi}_2) & \cdots & Cov(\hat{\psi}_2, \hat{\psi}_\kappa) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\hat{\psi}_\kappa, \hat{\psi}_1) & Cov(\hat{\psi}_\kappa, \hat{\psi}_2) & \cdots & Var(\hat{\psi}_\kappa) \end{bmatrix}, \quad (2.23)$$

conhecida também como a matriz de variância-covariância. Usando este resultado podemos fazer inferência aproximada sobre cada parâmetro estimado e assim os intervalos de confiança aproximados de $100(1 - \alpha')\%$ são obtidos fazendo, $i = 1, \dots, \kappa$:

$$IC(\psi_i; 100(1 - \alpha')\%) = \hat{\psi}_i \pm z_{\alpha'/2} \sqrt{Var(\hat{\psi}_i)}. \quad (2.24)$$

no qual $z_{\alpha'/2}$ é o quantil da distribuição normal padrão e α' é o nível de confiança.

2.7 Critérios de seleção do modelo

Os critérios de seleção são regras que permitem comparar modelos através de medidas escalares, que medem a qualidade do ajuste do modelo. Neste trabalho serão utilizadas algumas medidas de ajuste, tais como:

1. DEV

A deviance é calculada através da expressão:

$$Dev = -2 \log L(\hat{\psi}; D) \quad (2.25)$$

2. AIC

O critério de informação de Akaike (1973) é conhecido como *AIC* é definido como:

$$AIC = -2 \log L(\hat{\psi}; D) + 2\kappa, \quad (2.26)$$

em que κ é o número de parâmetros no modelo em questão.

3. BIC

O critério de informação bayesiano *BIC* foi proposto por Schwarz (1978) e dado por:

$$BIC = -2 \log L(\hat{\psi}; D) + \kappa(\log n), \quad (2.27)$$

em que κ é o número de parâmetros e n é o número de observações, neste caso será o número de animais.

Os três critérios que foram apresentados, apesar de ser diferentemente definidos, utilizam o mesmo critério estatístico, o valor máximo da função de verossimilhança como medida do ajustamento, entretanto, definem valores críticos diferentes. Esta é a diferença fundamental entre esses critérios. Portanto, para modelos em avaliação com o mesmo conjunto de dados, será selecionado o melhor modelo, aquele que possui o menor valor de critério de seleção.

Capítulo 3

Modelo destrutivo com variável terminal

Neste capítulo apresentamos a modelagem matemática dos experimentos estudados e utilizados para avaliar o efeito de um determinado tratamento sobre a taxa de incidência de tumores. Para caracterizar os experimentos quimiopreventivos será realizado uma reformulação do modelo proposto por Kokoska (1987) e Freedman *et al.* (1993) no contexto do modelo de Rodrigues *et al.* (2010) que envolve dois estágios.

3.1 Formulação do modelo

O modelo destrutivo com variável terminal é caracterizado pelos seguintes estágios:

1. Estágio de iniciação

Neste estágio de iniciação o animal foi exposto aos agentes cancerígenos, para produzir tumores malignos, isto é, células normais sofrem o efeito dos agentes cancerígenos ou carcinógenos que provocam mutação na estrutura genética das células fazendo elas se transformarem em células geneticamente alteradas, células "iniciadas". Estas células encontram-se "preparadas", ou seja, "iniciadas", para transformar-se em células malignas gerando os tumores. Neste trabalho supomos que o desenvolvimento dos tumores é de forma independente um de outro.

Seja M_i a variável aleatória latente (não observável) que representa o número de tumores promovidos gerados pelas células malignas, no i -ésimo animal.

2. Estágio de destruição

No estágio de destruição o animal foi submetido ao tratamento quimiopreventivo, podendo ou não eliminar os tumores promovidos. Conseqüentemente, os tumores que sobreviveram ao tratamento podem ser detectados mediante a técnica de apalpação feita no corpo do animal. Esta técnica consiste na apalpação contínua para identificar a existência ou não de tumores com tamanho detectável. A observação desses tumores é possível somente durante a vida do animal sujeito a um tempo pré-estabelecido para o experimento.

Espera-se que o tratamento quimiopreventivo diminua o número de tumores malignos gerados pelas células malignas e/ou reduza a taxa de crescimento do tumor. Neste trabalho estamos interessados nos tumores que sobreviveram ao tratamento ou que não foram destruídos e que após o tratamento foram detectados.

Seja d a duração do experimento quimiopreventivo e t_i^* , $i = 1, 2, \dots, n$, uma variável aleatória, observável que chamaremos de variável terminal, referida ao instante de morte devido aos tumores do i -ésimo animal. Consideremos também X_{ij} como uma variável aleatória (independente de M_i), com distribuição de Bernoulli que indica se foi ou não detectado o j -ésimo tumor promovido no i -ésimo animal antes do tempo terminal t_i^* , isto é,

$$X_{ij} = \begin{cases} 1 & , \text{ tumor detectado no } i\text{-ésimo animal antes do tempo } t_i^*; \\ 0 & , \text{ caso contrario,} \end{cases} \quad (3.1)$$

com probabilidade de sucesso $F(t_i^*; \theta)$, em que θ é o parâmetro de interesse. É fundamental enfatizar que a identificação do tumor está condicionada ao instante de morte do animal.

Os tempos em que foram detectados os tumores serão representados pelas variáveis $T_{(ij)}$, $i = 1, 2, \dots, n$, $j = 1, \dots, M_i^*$ que são variáveis aleatórias observáveis, independentes das variáveis M_i , identicamente distribuídas e estatisticamente ordenadas geradas pela distribuição de probabilidade acumulada $F(t; \theta)$.

Seja M_i^* uma variável aleatória discreta que representa o número de tumores detectados entre os M_i tumores promovidos, que não foram destruídos pelo tratamento no i -ésimo animal condicionado ou censurado no instante de morte t_i^* , $i = 1, 2, \dots, n$. Portanto a variável M_i^* pode ser expressada como:

$$M_i^* = \begin{cases} \sum_{j=1}^{m_i} X_{ij} & , \text{ para } M_i > 0, \\ 0 & , \text{ para } M_i = 0. \end{cases} \quad (3.2)$$

Dado $M_i = m_i$ e t_i^* , $i = 1, 2, \dots, n$, a distribuição de probabilidade da variável aleatória discreta, M_i^* , é uma distribuição Binomial com parâmetros m_i e $F(t_i^*; \theta)$, neste caso conhecida como mecanismo de destruição dos tumores, condicionado ao instante de morte t_i^* (vide [Rodrigues et al. \(2010\)](#)). É importante enfatizar que a probabilidade de encontrar um tumor antes do instante t_i^* depende (probabilidade de sucesso) do tempo de vida do animal, diferentemente do modelo que foi formulado por [Rodrigues et al. \(2010\)](#), em que a probabilidade de sucesso é constante.

3.1.1 Observações

1. Se $t_i^* < d$, $i = 1, 2, \dots, n$, temos que o animal morreu antes de finalizar o experimento, isto é, o animal foi censurado aleatoriamente no instante t_i^* .
2. Se $t_i^* = d$, $i = 1, 2, \dots, n$, temos que o animal sobreviveu até finalizar o experimento, ou seja, foi sacrificado para permitir a confirmação histológica dos tumores observados após o tratamento quimiopreventivo.
3. As variáveis destrutivas M_i^* são observadas enquanto que no modelo destrutivo ([Rodrigues et al. \(2010\)](#)) são latentes.
4. Se $M_i^* < M_i$, temos que houve destruição dos tumores pelo agente quimiopreventivo.

5. O experimento em estudo, não será considerada a possibilidade de reparação das células iniciadas que são feitas pelo mecanismo de reparo do animal.
6. Se $M_i^* = M_i$, temos que não houve destruição, isto é, o tratamento quimiopreventivo não fez nenhum efeito no animal.

3.1.2 Distribuição de probabilidades com variável terminal

A distribuição de probabilidades da variável destrutiva com variável terminal, t_i^* , para o i -ésimo animal pode ser obtida a partir da expressão (3.2) e é dada por:

$$\begin{aligned}
 P[M_i^* = m_i^* | t_i^*; \eta, \theta] &= \sum_{m_i = m_i^*}^{\infty} P[M_i^* = m_i^* | M_i = m_i; \theta] P[M_i = m_i; \eta], \\
 &= \sum_{m_i = m_i^*}^{\infty} \overbrace{\binom{m_i}{m_i^*} F(t_i^*; \theta)^{m_i^*} (1 - F(t_i^*; \theta))^{m_i - m_i^*}}^{\text{Modelo destrutivo com variável terminal}} P[M_i = m_i; \eta],
 \end{aligned} \tag{3.3}$$

em que $P[M_i = m_i; \eta]$ é a distribuição de probabilidade do número de tumores promovidos com parâmetro η . Portanto, a variável M_i é atualizada pelo mecanismo de destruição anteriormente definido. A esperança e variância da variável destrutiva, M_i^* , dado a variável terminal, t_i^* , são dadas por (ver **Apêndice A.2**):

$$E[M_i^* | t_i^*] = E[M_i] F(t_i^*; \theta), \tag{3.4}$$

$$Var[M_i^* | t_i^*] = (Var[M_i] - E[M_i]) F(t_i^*; \theta)^2 + E[M_i] F(t_i^*; \theta). \tag{3.5}$$

3.1.3 Função de verossimilhança

A função de verossimilhança proposta por [Freedman et al. \(1993\)](#), para o i -ésimo animal será reformulada na seguinte forma.

$$L_i(\theta, \eta) = f(t_{(i1)}, \dots, t_{(im_i^*)} | t_i^*; \theta) P[M_i^* = m_i^* | t_i^*; \eta, \theta], \tag{3.6}$$

em que $t_{(ij)}$ é o tempo de detecção do j -ésimo tumor no i -ésimo animal, m_i^* é o número de tumores detectados no i -ésimo animal e t_i^* é o tempo de morte (variável terminal) do i -ésimo animal. A função densidade de probabilidade conjunta das observações ordenadas $t_{(ij)}$, dada a variável terminal, t_i^* , para $i = 1, 2, \dots, n$, $j = 1, \dots, m_i^*$, é dada por ([Casella & Berger \(2001\)](#)):

$$f(t_{(i1)}, \dots, t_{(im_i^*)} | t_i^*; \theta) = \begin{cases} m_i^*! f(t_{i1} | t_i^*; \theta) \cdots f(t_{im_i^*} | t_i^*; \theta) & , 0 < t_{i1} < \cdots < t_{im_i^*} < t_i^* < \infty; \\ 0 & , \text{ caso contrário.} \end{cases} \tag{3.7}$$

Logo, a distribuição conjunta para o i -ésimo animal é dada por:

$$L_i(\theta, \eta) = m_i^*! \prod_{j=1}^{m_i^*} f(t_{ij} | t_i^*; \theta) P[M_i^* = m_i^* | t_i^*; \eta, \theta], \tag{3.8}$$

em que $f(t_{ij}|t_i^*; \theta) = \frac{f(t_{ij}; \theta)}{F(t_i^*; \theta)}$ é a distribuição truncada no instante t_i^* e (θ, η) são os parâmetros envolvidos no modelo destrutivo com variável terminal. O parâmetro θ identifica a distribuição de probabilidade do tempo de detecção do tumor e η identifica a distribuição de probabilidade da variável M .

Portanto, a função de verossimilhança do modelo destrutivo com variável terminal, t_i^* , $i = 1, 2, \dots, n$, para n animais é dada por:

$$L(\theta, \eta) = \prod_{i=1}^n L_i(\theta, \eta) = \prod_{i=1}^n \left\{ m_i^*! \prod_{j=1}^{m_i^*} \frac{f(t_{ij}; \theta)}{F(t_i^*; \theta)} P[M_i^* = m_i^* | t_i^*; \eta, \theta] \right\}. \quad (3.9)$$

Como é usual encontrar excesso de zeros nos dados dos experimentos com animais, serão considerados os modelos Poisson, binomial negativo, Poisson com zeros inflacionados e binomial negativo com zeros inflacionados. A seguir apresentamos a Tabela 3.1 das distribuições a serem utilizadas neste trabalho.

Tabela 3.1: Modelos Paramétricos.

Distribuição de prob. de M	$P(M = m; \eta)$	Parâmetro
Poisson(λ)	$e^{-\lambda} \lambda^m / m!$	$\eta = \lambda$
NB(N, ν)	$\binom{N+m-1}{m} \nu^N (1-\nu)^m$	$\eta = (N, \nu)$
PZI(ω, λ)	$(1-\omega) I_{\{0\}}(m) + \omega \frac{e^{-\lambda} \lambda^m}{m!}$	$\eta = (\omega, \lambda)$
BNZI(ω, μ, τ)	$(1-\omega) I_{\{0\}}(m) + \omega \frac{\Gamma(N+m)}{\Gamma(N)m!} \nu^N (1-\nu)^m$	$\eta = (\omega, \mu, \tau)$
$I_{\{0\}}(m) = \begin{cases} 1 & , m = 0; \\ 0 & , m \neq 0, \end{cases}$		
Distribuição de prob. de T_{ij}	$f(t; \theta)$	Parâmetro
Gama(α, β)	$\frac{t^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} e^{-t/\beta}$	$\theta = (\alpha, \beta)$

3.2 Casos especiais do modelo destrutivo com variável terminal

Nesta seção, apresentamos alguns casos especiais do modelo destrutivo com variável terminal, t_i^* , proposto na seção anterior.

3.2.1 Modelo destrutivo de Poisson com variável terminal

A distribuição de probabilidade da variável destrutiva Poisson com variável terminal pode ser obtida da expressão (3.3), dada por:

$$\begin{aligned}
 P[M_i^* = m_i^* | t_i^*; \lambda, \theta] &= \sum_{m_i = m_i^*}^{\infty} P[M_i^* = m_i^* | M_i = m_i; \theta] P[M_i = m_i; \lambda], \\
 &= \sum_{m_i = m_i^*}^{\infty} \binom{m_i}{m_i^*} F(t_i^*; \theta)^{m_i^*} (1 - F(t_i^*; \theta))^{m_i - m_i^*} \frac{e^{-\lambda} \lambda^{m_i}}{m_i!}, \\
 &= \sum_{m_i = m_i^*}^{\infty} \frac{m_i!}{m_i^* (m_i - m_i^*)!} F(t_i^*; \theta)^{m_i^*} (1 - F(t_i^*; \theta))^{m_i - m_i^*} \frac{e^{-\lambda} \lambda^{m_i - m_i^* + m_i^*}}{m_i!}, \\
 &= \frac{[\lambda F(t_i^*; \theta)]^{m_i^*}}{m_i^*!} e^{-\lambda} \sum_{m_i = m_i^*}^{\infty} \frac{[\lambda(1 - F(t_i^*; \theta))]^{m_i - m_i^*}}{(m_i - m_i^*)!}, \\
 &= \frac{[\lambda F(t_i^*; \theta)]^{m_i^*}}{m_i^*!} e^{-\lambda} e^{\lambda(1 - F(t_i^*; \theta))}, \\
 &= e^{(-\lambda F(t_i^*; \theta))} \frac{[\lambda F(t_i^*; \theta)]^{m_i^*}}{m_i^*!}, \tag{3.10}
 \end{aligned}$$

em que $\lambda > 0$ e $\theta > 0$. Portanto, a variável destrutiva M_i^* , dada a variável terminal, t_i^* , tem uma distribuição de Poisson com parâmetro $\lambda F(t_i^*; \theta)$, $i = 1, \dots, n$. Assim

$$E[M_i^* | t_i^*] = \lambda F(t_i^*; \theta) = Var[M_i^* | t_i^*]. \tag{3.11}$$

De (3.9) a função verossimilhança com variável terminal, t_i^* , $i = 1, \dots, n$, para o modelo destrutivo de Poisson, em que $F(t; \theta)$ é a função distribuição acumulada de uma distribuição Gama com parâmetro $\theta = (\alpha, \beta)$, é dada por:

$$\begin{aligned}
 L(\lambda, \alpha, \beta) &= \prod_{i=1}^n \left\{ m_i^*! \prod_{j=1}^{m_i^*} \frac{f(t_{ij}; \alpha, \beta)}{F(t_i^*; \alpha, \beta)} e^{-\lambda F(t_i^*; \alpha, \beta)} \frac{(\lambda F(t_i^*; \alpha, \beta))^{m_i^*}}{m_i^*!} \right\}, \\
 &= \prod_{i=1}^n \prod_{j=1}^{m_i^*} f(t_{ij}; \alpha, \beta) e^{-\lambda \sum_{i=1}^n F(t_i^*; \alpha, \beta)} \prod_{i=1}^n \frac{\lambda^{m_i^*}}{m_i^*!}.
 \end{aligned}$$

O logaritmo da função de verossimilhança associada aos animais não necessariamente sacrificados, que são todos os animais do conjunto de dados, é dada por:

$$\log L(\lambda, \alpha, \beta) = \left[-\lambda \sum_{i=1}^n F(t_i^*; \alpha, \beta) + \sum_{i=1}^n m_i^* \log \lambda + (\alpha - 1) \sum_{i=1}^n \sum_{j=1}^{m_i^*} \log(t_{ij}) - \beta^{-1} \sum_{i=1}^n \sum_{j=1}^{m_i^*} t_{ij} \right] - \left[\alpha \sum_{i=1}^n m_i^* \log \beta + \sum_{i=1}^n m_i^* \log \Gamma(\alpha) \right],$$

em que $s_1 = \sum_{i=1}^n m_i^*$, $s_2 = \sum_{i=1}^n \sum_{j=1}^{m_i^*} \log t_{ij}$ e $s_3 = \sum_{i=1}^n \sum_{j=1}^{m_i^*} t_{ij}$.

Após algumas simplificações do logaritmo da função de verossimilhança obtemos:

$$\log[L(\lambda, \alpha, \beta)] = -\lambda \sum_{i=1}^n F(t_i^*; \alpha, \beta) + s_1(\log \lambda - \alpha \log \beta - \log \Gamma(\alpha)) + s_2(\alpha - 1) - \frac{s_3}{\beta}. \quad (3.12)$$

A seguir apresentamos as seguintes equações para a obtenção dos estimadores de máxima verossimilhança $\hat{\lambda}$, $\hat{\alpha}$ e $\hat{\beta}$ (obtidos no Apêndice A.3):

$$\hat{\lambda} = s_1 / \sum_{i=1}^n F(t_i^*; \hat{\alpha}, \hat{\beta}), \quad (3.13)$$

$$\Psi(\hat{\alpha}) - \log(\hat{\alpha}) + \log(s_3/s_1) - s_2/s_1 = 0, \quad (3.14)$$

$$\hat{\beta} = s_3 \hat{\alpha}^{-1} / s_1. \quad (3.15)$$

3.2.2 Modelo destrutivo binomial negativo com variável terminal

Supondo a variável M_i com distribuição binomial negativa, apresentamos a distribuição de probabilidade da variável destrutiva, M_i^* , com variável terminal, t_i^* , para o modelo binomial negativo dada por (Apêndice A.4):

$$\begin{aligned} P[M_i^* = m_i^* | t_i^*; N, \nu, \theta] &= \sum_{m_i = m_i^*}^{\infty} \binom{m_i}{m_i^*} F(t_i^*; \theta)^{m_i^*} (1 - F(t_i^*; \theta))^{m_i - m_i^*} \frac{\Gamma(N + m_i)}{\Gamma(N) m_i!} \nu^N (1 - \nu)^{m_i}, \\ &= \frac{\Gamma(N + m_i^*)}{\Gamma(N) m_i^*!} \left[\frac{\nu}{\nu + (1 - \nu) F(t_i^*; \theta)} \right]^N \left[1 - \frac{\nu}{\nu + (1 - \nu) F(t_i^*; \theta)} \right]^{m_i^*}, \end{aligned} \quad (3.16)$$

onde $0 < \nu < 1$, $N = 1, 2, \dots$ e $\theta > 0$.

Portanto, a distribuição de probabilidade da variável destrutiva M_i^* é uma atualização da distribuição binomial negativa com variável terminal, t_i^* , $i = 1, 2, \dots, n$, dada por:

$$M_i^* | t_i^* \sim \text{Binomial Neg.} \left(N, \frac{\nu}{\nu + (1 - \nu) F(t_i^*; \theta)} \right).$$

A esperança e variância podem ser obtidas de (2.4), isto é,

$$E[M_i^*|t_i^*] = \frac{N(1-\nu)F(t_i^*; \theta)}{\nu}, \quad (3.17)$$

$$Var[M_i^*|t_i^*] = N(1-\nu)F(t_i^*; \theta) \frac{\nu + (1-\nu)F(t_i^*; \theta)}{\nu^2}. \quad (3.18)$$

Para simplificar os algoritmos computacionais a serem desenvolvidos e evitar futuros problemas de convergência, vamos introduzir as seguintes reparametrizações: $\nu = 1/(1 + \mu\tau)$ e $N = 1/\tau$, em que μ é a média da distribuição binomial negativa. A função de verossimilhança com variável terminal t_i^* , $i = 1, \dots, n$, para o modelo destrutivo binomial negativo, em que $F(t; \theta)$ é uma distribuição gama com parâmetro $\theta = (\alpha, \beta)$, é dada por:

$$L(\mu, \tau, \alpha, \beta) = \prod_{i=1}^n \left\{ m_i^*! \prod_{j=1}^{m_i^*} \frac{f(t_{ij}; \alpha, \beta)}{F(t_i^*; \alpha, \beta)} \frac{\Gamma(1/\tau + m_i^*)}{\Gamma(1/\tau)m_i^*!} \left[\frac{1}{1 + \mu\tau F(t_i^*; \alpha, \beta)} \right]^{\frac{1}{\tau}} \left[\frac{\mu\tau F(t_i^*; \alpha, \beta)}{1 + \mu\tau F(t_i^*; \alpha, \beta)} \right]^{m_i^*} \right\}. \quad (3.19)$$

Fazendo as devidas simplificações em (3.19), temos que o logaritmo da função de verossimilhança é dado por:

$$\begin{aligned} \log L(\mu, \tau, \alpha, \beta) &= \sum_{i=1}^n \log \Gamma(1/\tau + m_i^*) - \sum_{i=1}^n [(1/\tau + m_i^*) \log(1 + \mu\tau F(t_i^*; \alpha, \beta))] - s_3/\beta \\ &\quad + s_2(\alpha - 1) + s_1(\log(\mu\tau) - \alpha \log \beta - \log \Gamma(\alpha)) - n \log \Gamma(1/\tau), \end{aligned}$$

sendo $s_1 = \sum_{i=1}^n m_i^*$, $s_2 = \sum_{i=1}^n \sum_{j=1}^{m_i^*} \log(t_{ij})$ e $s_3 = \sum_{i=1}^n \sum_{j=1}^{m_i^*} (t_{ij})$.

As condições de 1ª e 2ª ordem são apresentadas no Apêndice A.4.1.

3.2.3 Modelo destrutivo Poisson com zeros inflacionados e variável terminal

Supondo a variável M_i com distribuição de Poisson com zeros inflacionados, a distribuição de probabilidade da variável destrutiva, M_i^* , com variável terminal, t_i^* , $i = 1, \dots, n$, para o modelo Poisson com zeros inflacionados é dada por (ver Apêndice A.5):

$$P[M_i^* = m_i^* | t_i^*; \omega, \lambda, \theta] = (1 - \omega)I_{\{0\}}(m_i^*) + \omega \frac{e^{-\lambda F(t_i^*; \theta)} (\lambda F(t_i^*; \theta))^{m_i^*}}{m_i^*!}, \quad (3.20)$$

tal que

$$I_{\{0\}}(m_i^*) = \begin{cases} 1 & , m_i^* = 0; \\ 0 & , m_i^* \geq 1. \end{cases}$$

Portanto, a variável destrutiva, M_i^* , dado t_i^* , tem uma distribuição de Poisson com zeros inflacionados, com parâmetros $0 < \omega \leq 1$ e $\lambda F(t_i^*; \theta)$ para $\lambda > 0$ e $\theta > 0$. A média e a variância são obtidas

de (2.7) e (2.9) respectivamente. Logo temos,

$$E[M_i^* | t_i^*] = \omega \lambda F(t_i^*; \theta) \quad (3.21)$$

$$Var[M_i^* | t_i^*] = \omega \lambda F(t_i^*; \theta) [1 + \lambda F(t_i^*; \theta) (1 - \omega)] \quad (3.22)$$

Supondo que $F(t; \theta)$ é uma distribuição gama com parâmetro $\theta = (\alpha, \beta)$, a função de verossimilhança baseada em dados ampliados para o i -ésimo animal com variável terminal, t_i^* , $i = 1, \dots, n$ é dada por:

$$L_i(\omega, \lambda, \alpha, \beta | t_i^*) = m_i^*! \prod_{j=1}^{m_i^*} \frac{f(t_{ij}; \alpha, \beta)}{F(t_i^*; \alpha, \beta)} \left((1 - \omega) I_{\{0\}}(m_i^*) + \omega \frac{e^{-\lambda F(t_i^*; \alpha, \beta)} (\lambda F(t_i^*; \alpha, \beta))^{m_i^*}}{m_i^*!} \right), \quad (3.23)$$

em que m_i^* e t_i^* é o número de tumores observados e o tempo de vida do i -ésimo animal para $j = 1, \dots, m_i^*$ e $i = 1, \dots, n$, respectivamente. Logo, para os n animais a função de verossimilhança é expressa como:

$$\begin{aligned} L(\omega, \lambda, \alpha, \beta | t_i^*) &= \prod_{i=1}^{n_0} \left\{ 1 - \omega + \omega e^{-\lambda F(t_i^*; \alpha, \beta)} \right\} \prod_{i: M_i^* \notin A} \left\{ m_i^*! \prod_{j=1}^{m_i^*} \frac{f(t_{ij}; \alpha, \beta)}{F(t_i^*; \alpha, \beta)} \frac{\omega e^{-\lambda F(t_i^*; \alpha, \beta)} (\lambda F(t_i^*; \alpha, \beta))^{m_i^*}}{m_i^*!} \right\}, \\ &= \prod_{i=1}^{n_0} [1 - \omega + \omega e^{-\lambda F(t_i^*; \alpha, \beta)}] \omega^{n-n_0} \prod_{i: M_i^* \notin A} L_i(\lambda, \alpha, \beta), \end{aligned} \quad (3.24)$$

em que $A = \{M_i^* : M_i^* = 0\}$ e n_0 é o número de elementos do conjunto A . Quando $\omega = 1$, a função de verossimilhança em (3.24) reduz-se à função de verossimilhança correspondente ao modelo destrutivo de Poisson. Os elementos do conjunto A são gerados de duas formas diferentes: alguns elementos são gerados pela distribuição degenerada em zero (forma determinística) e outros pela distribuição de Poisson (forma aleatória). Como o modelo PZI é uma mistura de duas distribuições, a função de verossimilhança em (3.24) será simplificada utilizando o procedimento baseado em dados ampliados introduzido por Tanner (1996). Neste caso, vamos introduzir a seguinte variável latente:

$$Z_i = \begin{cases} 1, & \text{com probabilidade } \varphi_i; \\ 0, & \text{com probabilidade } 1 - \varphi_i, \end{cases} \quad (3.25)$$

no qual

$$\varphi_i = \frac{1 - \omega}{1 - \omega + \omega e^{-\lambda F(t_i^*; \alpha, \beta)}} \quad i = 1, \dots, n_0,$$

é a probabilidade de o zero ser determinístico. A função de verossimilhança baseada nos dados ampliados $D_a = \{D, Z_1, \dots, Z_{n_0}\}$, em que $D = \{M^*, T, t_i^*\}$, é dada por (Apêndice A.5.1):

$$L(\omega, \lambda, \alpha, \beta; D_a) = L(\omega, \lambda, \alpha, \beta) \prod_{i=1}^{n_0} [(\varphi_i)^{Z_i} (1 - \varphi_i)^{1-Z_i}], \quad (3.26)$$

$$\begin{aligned} &= \underbrace{\omega^{n-Z} (1 - \omega)^Z}_{\text{zeros inflacionados}} \underbrace{\exp \left\{ -\lambda \sum_{i=1}^{n_0} [(1 - Z_i) F(t_i^*; \alpha, \beta)] \right\} \prod_{i: M_i^* \notin A} L_i(\lambda, \alpha, \beta)}_{\text{dados do modelo original}}, \end{aligned} \quad (3.27)$$

em que $Z = \sum_{i=1}^{n_0} Z_i$. O logaritmo da função de verossimilhança ampliada em (3.27) é dado por:

$$\begin{aligned}
 \log L(\omega, \lambda, \alpha, \beta; D_a) &= (n - Z) \log \omega + Z \log(1 - \omega) - \lambda \sum_{i=1}^{n_0} (1 - Z_i) F(t_i^*; \alpha, \beta) + \sum_{i: M_i^* \notin A} \log L_i(\lambda, \alpha, \beta), \\
 &= (n - \sum_{i=1}^{n_0} Z_i) \log \omega + \log(1 - \omega) \sum_{i=1}^{n_0} Z_i + \lambda \sum_{i=1}^{n_0} Z_i F(t_i^*; \alpha, \beta) - \lambda \sum_{i=1}^{n_0} F(t_i^*; \alpha, \beta), \\
 &\quad + [\log \lambda - \alpha \log \beta - \log \Gamma(\alpha)] s_1 + (\alpha - 1) s_2 - \frac{s_3}{\beta} - \lambda \sum_{i: M_i^* \notin A} F(t_i^*; \alpha, \beta), \\
 &= (n - \sum_{i=1}^{n_0} Z_i) \log \omega + \log(1 - \omega) \sum_{i=1}^{n_0} Z_i + \lambda \sum_{i=1}^{n_0} Z_i F(t_i^*; \alpha, \beta) - \lambda \sum_{i=1}^n F(t_i^*; \alpha, \beta) \\
 &\quad + (\log \lambda - \alpha \log \beta - \log \Gamma(\alpha)) s_1 + (\alpha - 1) s_2 - s_3 / \beta,
 \end{aligned} \tag{3.28}$$

em que $s_1 = \sum_{i: M_i^* \notin A} m_i^*$, $s_2 = \sum_{i: M_i^* \notin A} \sum_{j=1}^{m_i^*} \log(t_{ij})$ e $s_3 = \sum_{i: M_i^* \notin A} \sum_{j=1}^{m_i^*} t_{ij}$.

Devido à presença da variável aleatória latente Z_i , os estimadores de máxima verossimilhança serão obtidos através da maximização de (3.28) utilizando o algoritmo EM (vide [Dempster et al. \(1977\)](#)) com os seguintes passos:

- **Passo E:** A esperança condicional do logaritmo da função de verossimilhança ampliada é dada por:

$$\begin{aligned}
 E[\log L(\omega, \lambda, \alpha, \beta; D_a) | D] &= (n - \sum_{i=1}^{n_0} E[Z_i | D_a]) \log \omega + \log(1 - \omega) \sum_{i=1}^{n_0} E[Z_i | D_a] \\
 &\quad + \lambda \sum_{i=1}^{n_0} E[Z_i | D_a] F(t_i^*; \alpha, \beta) - \lambda \sum_{i=1}^n F(t_i^*; \alpha, \beta) \\
 &\quad + (\log \lambda - \alpha \log \beta - \log \Gamma(\alpha)) s_1 + (\alpha - 1) s_2 - s_3 / \beta,
 \end{aligned}$$

no qual

$$\begin{aligned}
 E[Z_i | D_a] &= 0 \cdot (1 - \varphi_i) + 1 \cdot \varphi_i, \\
 &= \varphi_i, \\
 &= \frac{1 - \omega}{1 - \omega + \omega e^{-\lambda F(t_i^*; \alpha, \beta)}}.
 \end{aligned}$$

- **Passo M:** Maximização da esperança.

$$Q(\omega, \lambda, \alpha, \beta | \omega^{(K)}, \lambda^{(K)}, \alpha^{(K)}, \beta^{(K)}) = E[\log L(\omega^{(K)}, \lambda^{(K)}, \alpha^{(K)}, \beta^{(K)}; D_a) | D], \tag{3.29}$$

logo a esperança na K -ésima iteração é dada por:

$$\begin{aligned}
 Q(\omega, \lambda, \alpha, \beta | \omega^{(K)}, \lambda^{(K)}, \alpha^{(K)}, \beta^{(K)}) &= (n - \sum_{i=1}^{n_0} \varphi_i^{(K)}) \log \omega + \log(1 - \omega) \sum_{i=1}^{n_0} \varphi_i^{(K)} \\
 &+ \lambda \sum_{i=1}^{n_0} \varphi_i^{(K)} F(t_i^*; \alpha, \beta) - \lambda \sum_{i=1}^n F(t_i^*; \alpha, \beta) \\
 &+ (\log \lambda - \alpha \log \beta - \log \Gamma(\alpha)) s_1 + (\alpha - 1) s_2 - s_3 / \beta.
 \end{aligned} \tag{3.30}$$

As derivadas em relação aos parâmetros em (3.30), são apresentadas:

$$\frac{\partial}{\partial \omega} E[\log L(\psi^{(K)} | D_a)] = \frac{n - \sum_{i=1}^{n_0} \varphi_i^{(K)}}{\omega} - \frac{\sum_{i=1}^{n_0} \varphi_i^{(K)}}{1 - \omega} = 0, \tag{3.31}$$

$$\frac{\partial}{\partial \lambda} E[\log L(\psi^{(K)} | D_a)] = \sum_{i=1}^{n_0} \varphi_i^{(K)} F(t_i^*; \alpha, \beta) - \sum_{i=1}^n F(t_i^*; \alpha, \beta) + \frac{s_1}{\lambda} = 0, \tag{3.32}$$

$$\frac{\partial}{\partial \alpha} E[\log L(\psi^{(K)} | D_a)] = \lambda \sum_{i=1}^{n_0} \frac{\partial}{\partial \alpha} \varphi_i^{(K)} F(t_i^*; \alpha, \beta) - \lambda \sum_{i=1}^n \frac{\partial}{\partial \alpha} F(t_i^*; \alpha, \beta) - (\log(\beta) + \Psi(\alpha)) s_1 + s_2 = 0, \tag{3.33}$$

$$\frac{\partial}{\partial \beta} E[\log L(\psi^{(K)} | D_a)] = \lambda \sum_{i=1}^{n_0} \frac{\partial}{\partial \beta} \varphi_i^{(K)} F(t_i^*; \alpha, \beta) - \lambda \sum_{i=1}^n \frac{\partial}{\partial \beta} F(t_i^*; \alpha, \beta) - \alpha s_1 / \beta + s_3 / \beta^2 = 0, \tag{3.34}$$

em que $\Psi(\hat{\alpha}) = \frac{\partial}{\partial \alpha} \ln \Gamma(\hat{\alpha})$ é a função digama. A seguir apresentamos um procedimento para resolver o sistema de equações em (3.31):(3.34).

De (3.31) temos que:

$$\begin{aligned}
 (1 - \omega)(n - \sum_{i=1}^{n_0} \varphi_i^{(K)}) - \omega \sum_{i=1}^{n_0} \varphi_i^{(K)} &= 0, \\
 (n - \sum_{i=1}^{n_0} \varphi_i^{(K)}) - \omega(n - \sum_{i=1}^{n_0} \varphi_i^{(K)}) + \sum_{i=1}^{n_0} \varphi_i^{(K)} &= 0, \\
 \hat{\omega} &= 1 - \frac{\sum_{i=1}^{n_0} \varphi_i^{(K)}}{n}.
 \end{aligned}$$

De (3.32) temos que:

$$\begin{aligned}
 \frac{s_1}{\lambda} &= \sum_{i=1}^{n_0} \varphi_i^{(K)} F(t_i^*; \alpha, \beta) - \sum_{i=1}^n F(t_i^*; \alpha, \beta), \\
 \hat{\lambda} &= s_1 / (\sum_{i=1}^n F(t_i^*; \alpha, \beta) - \sum_{i=1}^{n_0} \varphi_i^{(K)} F(t_i^*; \alpha, \beta)).
 \end{aligned}$$

Derivando (3.32) em relação de α e β obtemos o seguinte:

$$\sum_{i=1}^{n_0} \frac{\partial}{\partial \alpha} \varphi_i^{(K)} F(t_i^*; \alpha, \beta) - \sum_{i=1}^n \frac{\partial}{\partial \alpha} F(t_i^*; \alpha, \beta) = 0,$$

$$\sum_{i=1}^{n_0} \frac{\partial}{\partial \beta} \varphi_i^{(K)} F(t_i^*; \alpha, \beta) - \sum_{i=1}^n \frac{\partial}{\partial \beta} F(t_i^*; \alpha, \beta) = 0.$$

Logo substituindo as expressões anteriores em (3.33) e (3.34), respectivamente, obtemos:

$$-(\log \beta + \Psi(\alpha))s_1 + s_2 = 0, \quad (3.35)$$

$$-\frac{\alpha}{\beta}s_1 + \frac{s_3}{\beta^2} = 0. \quad (3.36)$$

Resolvendo o sistema de equações (3.35) e (3.36), obtemos a seguinte expressão:

$$-(\log(s_3/s_1) - \log \alpha + \Psi(\alpha))s_1 + s_2 = 0. \quad (3.37)$$

Conseqüentemente, as estimativas de máxima verossimilhança são soluções das seguintes equações:

$$\hat{\lambda}^{K+1} = s_1 / \left(\sum_{i=1}^n F(t_i^*; \alpha, \beta) - \sum_{i=1}^{n_0} \varphi_i^{(K)} F(t_i^*; \alpha, \beta) \right), \quad (3.38)$$

$$\hat{\omega}^{K+1} = 1 - \sum_{i=1}^{n_0} \varphi_i^{(K)} / n, \quad (3.39)$$

$$\Psi(\hat{\alpha}) - \log \hat{\alpha} + \log(s_3/s_1) - s_2/s_1 = 0, \quad (3.40)$$

$$\hat{\beta} = \hat{\alpha}^{-1} s_3 / s_1. \quad (3.41)$$

As estimativas dos parâmetros foram obtidas utilizando rotinas de maximização do pacote *R*. As condições de 1ª e 2ª ordenes foram verificadas computacionalmente.

3.2.4 Modelo destrutivo binomial negativo com zeros inflacionados e variável terminal

Supondo a variável M_i com distribuição binomial negativa com zeros inflacionados, apresentamos a distribuição de probabilidade da variável destrutiva, M_i^* , com variável terminal, t_i^* , $i = 1, \dots, n$ para o modelo binomial negativo com zeros inflacionados é dada por (ver Apêndice A.6):

$$P[M_i^* = m_i^* | t_i^*; \omega, \mu, \tau, \theta] = (1 - \omega) I_{\{0\}}(m_i^*) + \omega \frac{\Gamma(\frac{1}{\tau} + m_i^*)}{\Gamma(\frac{1}{\tau}) m_i^*!} \left[\frac{1}{1 + \mu \tau F(t_i^*; \theta)} \right]^{\frac{1}{\tau}} \left[\frac{\mu \tau F(t_i^*; \theta)}{1 + \mu \tau F(t_i^*; \theta)} \right]^{m_i^*}, \quad (3.42)$$

em que

$$I_{\{0\}}(m_i^*) = \begin{cases} 1 & , m_i^* = 0; \\ 0 & , m_i^* \geq 1. \end{cases}$$

Portanto, a variável destrutiva, M_i^* , dado t_i^* , tem uma distribuição binomial negativa com zeros

inflacionados com parâmetros $0 < \omega \leq 1$, $\tau > 0$ e $\mu F(t_i^*; \theta)$ para $\mu > 0$ e $\theta > 0$. A média e a variância são obtidas de (2.11) e (2.12) respectivamente. Logo temos,

$$E[M_i^* | t_i^*] = \omega \mu F(t_i^*; \theta), \quad (3.43)$$

$$Var[M_i^* | t_i^*] = \omega \mu F(t_i^*; \theta) [\mu F(t_i^*; \theta) (1 - \omega) + 1 + \mu \tau F(t_i^*; \theta)]. \quad (3.44)$$

Supondo que $F(t_i^*; \theta)$ é uma distribuição gama com parâmetro $\theta = (\alpha, \beta)$, a função de verossimilhança baseada em dados ampliados sob o modelo destrutivo binomial negativo com zeros inflacionados e variável terminal, t_i^* , é dada por (ver Apêndice A.6.1):

$$L(\omega, \mu, \tau, \alpha, \beta) = \prod_{i=1}^{n_0} \left\{ 1 - \omega + \omega \left[\frac{1}{1 + \mu \tau F(t_i^*; \theta)} \right]^{\frac{1}{\tau}} \right\} \omega^{n-n_0} \prod_{i: M_i^* \notin A} L_i(\mu, \tau, \alpha, \beta), \quad (3.45)$$

onde $A = \{M_i^* : M_i^* = 0\}$ e n_0 é o número de elementos do conjunto A . Para $\omega = 1$, a função de verossimilhança em (3.45) reduz-se à função de verossimilhança correspondente ao modelo destrutivo binomial negativo com variável terminal dada em (3.19). O modelo destrutivo binomial negativo com zeros inflacionados e variável terminal, também é uma mistura de duas distribuições: distribuição degenerada em zero e a distribuição binomial negativa. O procedimento é similar, como apresentado na Seção (3.2.3). A função de verossimilhança em (3.45) será simplificada, utilizando o procedimento baseado em dados ampliados introduzido por Tanner (1996), com a seguinte variável latente.

$$Y_i = \begin{cases} 1 & , \text{ com probabilidade } \phi_i; \\ 0 & , \text{ com probabilidade } 1 - \phi_i, \end{cases} \quad (3.46)$$

no qual

$$\phi_i = \frac{1 - \omega}{1 - \omega + \omega [1 + \mu \tau F(t_i^*; \theta)]^{-\frac{1}{\tau}}} \quad i = 1, \dots, n_0,$$

é a probabilidade do zero ser determinístico para o modelo destrutivo BNZI com variável terminal.

A função de verossimilhança baseada nos dados ampliados $U_a = \{D, Y_1, \dots, Y_{n_0}\}$, em que $D = \{M^*, T, t_i^*\}$, é dada por (ver Apêndice A.6.1):

$$\begin{aligned} L(\omega, \mu, \tau, \alpha, \beta; U_a) &= L(\omega, \mu, \tau, \alpha, \beta) \prod_{i=1}^{n_0} [(\phi_i)^{Y_i} (1 - \phi_i)^{1-Y_i}], \\ &= \underbrace{\omega^{n-Y} (1 - \omega)^Y}_{\text{zeros inflacionados}} \underbrace{\prod_{i=1}^{n_0} (1 + \mu \tau F(t_i^*; \theta))^{\frac{Y_i-1}{\tau}} \prod_{i: M_i^* \notin A} L_i(\mu, \tau, \alpha, \beta)}_{\text{dados do modelo original}}, \end{aligned} \quad (3.47)$$

onde $Y = \sum_{i=1}^{n_0} Y_i$. O logaritmo da função de verossimilhança ampliada em (3.47) é:

$$l(\eta, \theta; U_a) = \log L(\omega, \mu, \tau, \alpha, \beta; U_a)$$

em que $\eta = (\omega, \mu, \tau)$ e $\theta = (\alpha, \beta)$, temos que:

$$\begin{aligned}
 l(\eta, \theta; U_a) &= (n - Y) \log \omega + Y \log(1 - \omega) + \sum_{i=1}^{n_0} \frac{(Y_i - 1)}{\tau} \log(1 + \mu\tau F(t_i^*; \alpha, \beta)) + \sum_{i: M_i^* \notin A} \log L_i(\mu, \tau, \alpha, \beta), \\
 &= (n - \sum_{i=1}^{n_0} Y_i) \log \omega + Y \log(1 - \omega) + \sum_{i=1}^{n_0} \frac{(Y_i - 1)}{\tau} \log(1 + \mu\tau F(t_i^*; \alpha, \beta)) \\
 &\quad + \sum_{i: M_i^* \notin A} \log \Gamma(1/\tau + m_i^*) - \sum_{i: M_i^* \notin A} [(\frac{1}{\tau} + m_i^*) \log(1 + \mu\tau F(t_i^*; \alpha, \beta))] - s_3/\beta \\
 &\quad + s_2(\alpha - 1) + s_1(\log(\mu\tau) - \alpha \log \beta - \log \Gamma(\alpha)) - (n - n_0) \log \Gamma(1/\tau),
 \end{aligned}$$

$$\text{onde } s_1 = \sum_{i: M_i^* \notin A} m_i^*, \quad s_2 = \sum_{i: M_i^* \notin A} \sum_{j=1}^{m_i^*} \log(t_{ij}) \quad \text{e} \quad s_3 = \sum_{i: M_i^* \notin A} \sum_{j=1}^{m_i^*} t_{ij}.$$

Devido à presença da variável aleatória latente Y_i os estimadores de máxima verossimilhança serão obtidos através da maximização de (3.48) utilizando o algoritmo EM (ver [Dempster et al. \(1977\)](#)) com os seguintes passos:

- **Passo E:** A esperança condicional do logaritmo da função de verossimilhança ampliada é dada por:

$$\begin{aligned}
 E[l(\eta, \theta; U_a)|D] &= (n - \sum_{i=1}^{n_0} E[Y_i|U_a]) \log \omega + \log(1 - \omega) \sum_{i=1}^{n_0} E[Y_i|U_a] \\
 &\quad + \sum_{i=1}^{n_0} (E[Y_i|U_a] - 1) \log(1 + \mu\tau F(t_i^*; \alpha, \beta))/\tau + \sum_{i: M_i^* \notin A} \log \Gamma(\frac{1}{\tau} + m_i^*) \\
 &\quad - \sum_{i: M_i^* \notin A} [(\frac{1}{\tau} + m_i^*) \log(1 + \mu\tau F(t_i^*; \alpha, \beta))] - s_3/\beta + s_2(\alpha - 1) \\
 &\quad + s_1(\log(\mu\tau) - \alpha \log \beta - \log \Gamma(\alpha)) - (n - n_0) \log \Gamma(1/\tau),
 \end{aligned}$$

onde

$$\begin{aligned}
 E[Y_i|U_a] &= 0 \cdot (1 - \phi_i) + 1 \cdot \phi_i, \\
 &= \phi_i, \\
 &= \frac{1 - \omega}{1 - \omega + \omega[1 + \mu\tau F(t_i^*; \theta)]^{-\frac{1}{\tau}}}.
 \end{aligned}$$

- **Passo M:** Maximização da esperança.

$$Q(\omega, \mu, \tau, \alpha, \beta | \omega^{(K)}, \mu^{(K)}, \tau^{(K)}, \alpha^{(K)}, \beta^{(K)}) = E[l(\omega^{(K)}, \mu^{(K)}, \tau^{(K)}, \alpha^{(K)}, \beta^{(K)}; U_a)|D], \quad (3.48)$$

logo a esperança na K -ésima iteração é dada por:

$$\begin{aligned}
 Q(\omega, \mu, \tau, \alpha, \beta | \omega^{(K)}, \mu^{(K)}, \tau^{(K)}, \alpha^{(K)}, \beta^{(K)}) &= (n - \sum_{i=1}^{n_0} \phi_i^{(K)}) \log \omega + \log(1 - \omega) \sum_{i=1}^{n_0} \phi_i^{(K)} \\
 &+ \sum_{i=1}^{n_0} (\phi_i^{(K)} - 1) \log(1 + \mu\tau F(t_i^*; \alpha, \beta)) / \tau + \sum_{i: M_i^* \notin A} \log \Gamma\left(\frac{1}{\tau} + m_i^*\right) \\
 &- \sum_{i: M_i^* \notin A} \left[\left(\frac{1}{\tau} + m_i^*\right) \log(1 + \mu\tau F(t_i^*; \alpha, \beta)) \right] - s_3 / \beta + s_2(\alpha - 1) \\
 &+ s_1(\log(\mu\tau) - \alpha \log \beta - \log \Gamma(\alpha)) - (n - n_0) \log \Gamma(1/\tau),
 \end{aligned} \tag{3.49}$$

As derivadas em relação aos parâmetros em (3.49), são apresentadas:

$$\frac{\partial}{\partial \omega} E[l(\psi^{(K)} | U_a)] = \frac{n - \sum_{i=1}^{n_0} \phi_i^{(K)}}{\omega} - \frac{\sum_{i=1}^{n_0} \phi_i^{(K)}}{1 - \omega} = 0, \tag{3.50}$$

$$\frac{\partial}{\partial \mu} E[l(\psi^{(K)} | U_a)] = \sum_{i=1}^{n_0} \frac{(\phi_i^{(K)} - 1) F(t_i^*; \alpha, \beta)}{1 + \mu\tau F(t_i^*; \alpha, \beta)} - \sum_{i: M_i^* \notin A} \frac{(1 + \tau m_i^*) F(t_i^*; \alpha, \beta)}{1 + \mu\tau F(t_i^*; \alpha, \beta)} + \frac{s_1}{\mu} = 0, \tag{3.51}$$

$$\begin{aligned}
 \frac{\partial}{\partial \tau} E[l(\psi^{(K)} | U_a)] &= \sum_{i=1}^{n_0} (\phi_i^{(K)} - 1) \left[\frac{\mu F(t_i^*; \alpha, \beta) / \tau}{1 + \mu\tau F(t_i^*; \alpha, \beta)} - \frac{\log(1 + \mu\tau F(t_i^*; \alpha, \beta))}{\tau^2} \right] + \sum_{i: M_i^* \notin A} \frac{\partial}{\partial \tau} \log \Gamma(1/\tau + m_i^*) \\
 &- \sum_{i: M_i^* \notin A} \frac{(1/\tau + m_i^*) \mu F(t_i^*; \alpha, \beta)}{1 + \mu\tau F(t_i^*; \alpha, \beta)} - \frac{\log(1 + \mu\tau F(t_i^*; \alpha, \beta))}{\tau^2} + \frac{s_1}{\tau} - (n - n_0) \frac{\partial \Gamma(1/\tau)}{\partial \tau} = 0
 \end{aligned} \tag{3.52}$$

$$\frac{\partial}{\partial \alpha} E[l(\psi^{(K)} | U_a)] = \sum_{i=1}^{n_0} \frac{\mu(\phi_i^{(K)} - 1)}{1 + \mu\tau F(t_i^*; \alpha, \beta)} \frac{\partial F(t_i^*; \alpha, \beta)}{\partial \alpha} - \sum_{i: M_i^* \notin A} \frac{\mu\tau(1/\tau + m_i^*)}{1 + \mu\tau F(t_i^*; \alpha, \beta)} \frac{\partial F(t_i^*; \alpha, \beta)}{\partial \alpha} + s_2 - s_1 \frac{\partial}{\partial \alpha} \log \Gamma(\alpha) = 0 \tag{3.53}$$

$$\frac{\partial}{\partial \beta} E[l(\psi^{(K)} | U_a)] = \sum_{i=1}^{n_0} \frac{\mu(\phi_i^{(K)} - 1)}{1 + \mu\tau F(t_i^*; \alpha, \beta)} \frac{\partial F(t_i^*; \alpha, \beta)}{\partial \beta} - \sum_{i: M_i^* \notin A} \frac{\mu\tau(1/\tau + m_i^*)}{1 + \mu\tau F(t_i^*; \alpha, \beta)} \frac{\partial F(t_i^*; \alpha, \beta)}{\partial \beta} + \frac{s_3}{\beta^2} - \frac{\alpha s_1}{\beta} \tag{3.54}$$

De (3.50) temos que:

$$\begin{aligned}
 (1 - \omega)(n - \sum_{i=1}^{n_0} \phi_i^{(K)}) - \omega \sum_{i=1}^{n_0} \phi_i^{(K)} &= 0, \\
 (n - \sum_{i=1}^{n_0} \phi_i^{(K)}) - \omega(n - \sum_{i=1}^{n_0} \phi_i^{(K)}) + \sum_{i=1}^{n_0} \phi_i^{(K)} &= 0, \\
 \hat{\omega} &= 1 - \frac{1}{n} \sum_{i=1}^{n_0} \phi_i^{(K)}.
 \end{aligned}$$

Das equações (3.51): (3.54), não foi possível expressar de maneira explícita os parâmetros μ, τ, α e β . Portanto as estimativas dos parâmetros serão obtidas utilizando rotinas de maximização do pacote R e as condições de 1^a e 2^a ordem serão verificadas computacionalmente.

3.2.5 Estudo de simulação para o modelo destrutivo PZI com variável terminal

Um estudo de simulação foi realizado com objetivo de estudar as propriedades dos estimadores de máxima verossimilhança do modelo destrutivo PZI com variável terminal (dada na equação (3.20)), com a distribuição Gama para os tempos de detecção.

A simulação foi realizada com a geração de 1000 réplicas dos dados da seguinte forma:

Passo 1 : Gerar a variável terminal t_i^* de uma distribuição exponencial com taxa de 0,002 sendo que os valores gerados maiores que $d = 280$ (duração do experimento) são censurados em $d = 280$. Nesta geração de dados está sendo considerada a censura no tempo d .

Passo 2 : Gerar o número dos tumores, M_i^* , da distribuição destrutiva Poisson com zeros inflacionados dada em (3.20) com parâmetros $\omega = 0,7$ (proporção de zeros aleatórios) e $\lambda F(t_i^*, \alpha, \beta)$, onde $F(\cdot)$ uma distribuição Gama de parâmetros $\alpha = 3$, $\beta = 30$ e $\lambda = 5$.

Passo 3 : Gerar os tempos de detecção $t_{(ij)}$, de uma distribuição Gama truncada, isto é $0 \leq t_{(ij)} \leq t_i^*$.

Geramos amostras de tamanho $n = 20, 30, 50, 100, 200$. Para cada amostra são estimados os parâmetros da função de verossimilhança através do algoritmo EM implementado no Sistema R. Foram replicadas 1000 vezes cada amostra, foram calculados a média, o viés e o erro médio quadrático das estimativas dos parâmetros ω, λ e α, β . Foi obtido o intervalo de confiança de 95% para cada parâmetro baseado na teoria assintótica e verificou-se a sua probabilidade de cobertura (PC).

Analogamente foram gerados os dados com uma proporção de zeros aleatórios mais baixa, isto é, para $\omega = 0,4; \lambda = 5$ e $\alpha = 3, \beta = 30$. Os resultados desta simulação são apresentados nas Tabelas 3.2 e 3.3 e mostram que o procedimento de estimação pontual e os intervalos de confiança são consistentes assintoticamente, isto é, a medida que aumentamos o tamanho da amostra as estimativas ficam cada vez mais próximas dos valores reais e a cobertura intervalar de confiança é mais próxima dos valores nominais.

Neste estudo de simulação para o modelo destrutivo PZI com variável terminal, surge uma questão interessante: Se os dados gerados apresentam excesso de zeros, qual seria o impacto em utilizar um modelo que ignora a proporção desses zeros?.

A solução desta pergunta começa na geração dos dados seguindo os passos do algoritmo de uma distribuição destrutiva PZI com $\omega = 0,7; \lambda = 5$ e $\alpha = 3, \beta = 30$, e no momento da estimação adota-se o modelo destrutivo de Poisson. Os resultados da simulação são apresentados nas Tabelas 3.4 e 3.5, que mostram claramente que, utilizando o modelo destrutivo de Poisson, o impacto sobre as estimativas e probabilidade de cobertura do parâmetro λ é muito grande. O mesmo não ocorre com os demais parâmetros.

Tabela 3.2: Média, Viés, EQM das estimativas de máxima verossimilhança para 1000 réplicas com alta proporção de zeros aleatórios $\omega = 0,7$.

n	parâmetros	Média	Viés	EQM	PC
20	$\hat{\omega}$	0,7045139	-0,0045139	0,0127141	0,920
	$\hat{\lambda}$	4,9949257	0,0050743	0,4838406	0,945
	$\hat{\alpha}$	3,1566124	-0,1566124	0,4081186	0,951
	$\hat{\beta}$	29,6606835	0,3393165	44,9315619	0,913
30	$\hat{\omega}$	0,7064255	-0,006426	0,0073830	0,937
	$\hat{\lambda}$	4,9712719	0,0287281	0,2959103	0,952
	$\hat{\alpha}$	3,0949310	-0,094931	0,2541409	0,948
	$\hat{\beta}$	29,8567982	0,1432018	32,3748227	0,925
50	$\hat{\omega}$	0,6989514	0,0010486	0,0044365	0,947
	$\hat{\lambda}$	5,0112842	-0,0112842	0,1880805	0,953
	$\hat{\alpha}$	3,0520857	-0,0520857	0,1370806	0,947
	$\hat{\beta}$	29,9313482	0,0686518	16,8934158	0,934
100	$\hat{\omega}$	0,700371	-0,0003709	0,0024832	0,945
	$\hat{\lambda}$	4,993700	0,0063000	0,0876725	0,953
	$\hat{\alpha}$	3,024720	-0,0247199	0,0659668	0,948
	$\hat{\beta}$	30,009251	-0,0092508	8,8252368	0,941
200	$\hat{\omega}$	0,7000351	-0,0000351	0,0010687	0,966
	$\hat{\lambda}$	4,9990479	0,0000952	0,0441467	0,958
	$\hat{\alpha}$	3,0132447	-0,0132448	0,0292417	0,954
	$\hat{\beta}$	29,9339954	-0,0660046	3,9832379	0,951

Tabela 3.3: Média, Viés, EQM das estimativas de máxima verossimilhança para 1000 réplicas com baixa proporção de zeros aleatórios $\omega = 0,4$.

n	parâmetros	Média	Viés	EQM	PC
20	$\hat{\omega}$	0,4017424	-0,0017424	0,0127812	0,931
	$\hat{\lambda}$	5,0207547	-0,0207547	1,254357	0,942
	$\hat{\alpha}$	3,3298874	-0,3298874	1,1758167	0,958
	$\hat{\beta}$	29,8280784	0,1719216	196,776984	0,873
30	$\hat{\omega}$	0,4049084	-0,0049083	0,0083819	0,936
	$\hat{\lambda}$	4,9815571	0,01844294	0,5545584	0,955
	$\hat{\alpha}$	3,1906962	-0,1906962	0,5214549	0,958
	$\hat{\beta}$	29,6977165	0,30228351	54,882946	0,914
50	$\hat{\omega}$	0,3995461	0,0004538	0,0052149	0,943
	$\hat{\lambda}$	5,0181623	-0,0181623	0,2980737	0,965
	$\hat{\alpha}$	3,1118545	-0,1118545	0,2798006	0,955
	$\hat{\beta}$	29,7798639	0,2201361	32,555758	0,921
100	$\hat{\omega}$	0,3991638	0,0008362	0,0026213	0,950
	$\hat{\lambda}$	4,9880804	0,0119196	0,1540574	0,955
	$\hat{\alpha}$	3,0457961	-0,045796	0,1179760	0,959
	$\hat{\beta}$	30,0250010	-0,025001	14,9516759	0,940
200	$\hat{\omega}$	0,3986471	0,0013529	0,0012638	0,960
	$\hat{\lambda}$	4,9991332	0,0008668	0,0854705	0,944
	$\hat{\alpha}$	3,0202925	-0,020293	0,0557506	0,954
	$\hat{\beta}$	30,0007716	-0,000772	7,3388651	0,960

Tabela 3.4: Média, Viés, EQM das estimativas de máxima verossimilhança para 1000 réplicas, estimando com o modelo destrutivo de Poisson para dados gerados com $\omega = 0.7$ (proporção de zeros aleatórios).

n	parâmetros	Média	Viés	EQM	PC
20	$\hat{\lambda}$	3,503274	1,4967258	2,725266	0,221
	$\hat{\alpha}$	3,156974	-0,156974	0,408142	0,952
	$\hat{\beta}$	29,624621	0,3753789	44,064460	0,914
30	$\hat{\lambda}$	3,506881	1,49311947	2,5324332	0,112
	$\hat{\alpha}$	3,096573	-0,0965731	0,2544574	0,948
	$\hat{\beta}$	29,830636	0,16936363	32,5697685	0,926
50	$\hat{\lambda}$	3,499384	1,50061620	2,4499523	0,030
	$\hat{\alpha}$	3,054785	-0,0547855	0,1386691	0,947
	$\hat{\beta}$	29,893469	0,10653051	17,0479867	0,928
100	$\hat{\lambda}$	3,496602	1,5033980	2,3609845	0,003
	$\hat{\alpha}$	3,028388	-0,0283881	0,0663408	0,951
	$\hat{\beta}$	29,960163	0,0398372	8,8678034	0,940
200	$\hat{\lambda}$	3,498318	1,5016820	2,3033004	0,000
	$\hat{\alpha}$	3,014575	-0,014575	0,0291785	0,957
	$\hat{\beta}$	29,913907	0,0860930	3,9977442	0,958

Tabela 3.5: Média, Viés, EQM das estimativas de máxima verossimilhança para 1000 réplicas, estimando com o modelo destrutivo de Poisson para dados gerados com $\omega = 0.4$ (proporção de zeros aleatórios).

n	parâmetros	Média	Viés	EQM	PC
20	$\hat{\lambda}$	2,006636	2,993364	9,407521	0,000
	$\hat{\alpha}$	3,335323	-0,33532	1,176453	0,958
	$\hat{\beta}$	29,408389	0,591611	93,51392	0,865
30	$\hat{\lambda}$	1,988324	3,011676	9,365486	0,000
	$\hat{\alpha}$	3,214872	-0,21487	0,570726	0,952
	$\hat{\beta}$	29,477966	0,522034	59,46436	0,895
50	$\hat{\lambda}$	1,998077	3,001923	9,178324	0,000
	$\hat{\alpha}$	3,115483	-0,11548	0,285386	0,955
	$\hat{\beta}$	29,731764	0,268236	33,479471	0,919
100	$\hat{\lambda}$	1,991451	3,008549	9,141515	0,000
	$\hat{\alpha}$	3,053069	-0,05307	0,119221	0,955
	$\hat{\beta}$	29,926103	0,073897	15,018572	0,931
200	$\hat{\lambda}$	1,992111	3,007889	9.091372	0,000
	$\hat{\alpha}$	3,025773	-0,02577	0.055429	0,955
	$\hat{\beta}$	29,927710	0,072290	7.276706	0,946

Capítulo 4

Aplicação com Dados Reais

Mesmo com o avanço do desenvolvimento de novas drogas preventivas sobre o câncer, o índice de mortalidade vem crescendo de forma assustadora. Nos últimos anos, tem havido um aumento no número experimentos que envolvem a indução química de tumores em animais, cujo objetivo é avaliar o efeito de tratamentos quimiopreventivos. O interesse em tais experimentos são duas variáveis respostas, número de tumores observados em cada animal e os tempos de detecção dos tumores, variáveis que dependem do tempo de vida do animal (variável terminal). Como ilustração iremos aplicar em animais os resultados obtidos no Capítulo 3, onde os dados reais podem ser encontrados em [Freedman *et al.* \(1993\)](#). A implementação de todos os programas computacionais necessários para obter estes resultados estão disponíveis em linguagem *R*. Neste capítulo, o modelo destrutivo com variável terminal será utilizado em um conjunto de dados reais sobre a reincidência de tumores em ratos após um tratamento quimiopreventivo.

O conjunto de dados utilizado foi obtido em um experimento realizado na Universidade do Alabama, Birmingham. Nestes dados o tratamento experimental de prevenção é conhecido como agente *A*. O experimento foi planejado da seguinte forma: primeiro foi aplicada uma substância química cancerígena em dois grupos de 30 ratos fêmeas para produzir tumores. Um grupo foi submetido no tratamento controle em que receberam uma dieta placebo e o outro no tratamento com agente *A*. Os animais receberam os tratamentos antes, durante e após a indução cancerígena. Para obter os dados desses experimentos, os animais foram apalpados semanalmente, registrando o tempo de promoção destes tumores. Os ratos que sobreviveram após 280 dias foram sacrificados e verificou-se a existência de tumores ou confirmaram-se os tumores identificados. Os tumores que foram encontrados na necropsia não foram considerados no experimento. Os dados são apresentados na Tabela 1.1.

Considerando os modelos de Poisson, binomial negativo, Poisson com zeros inflacionados e binomial negativo com zeros inflacionados apresentados na Tabela 3.1, os estimadores de máxima verossimilhança serão obtidos para cada grupo e um teste de hipótese via razão de verossimilhança será utilizado para selecionar um desses modelos e verificar se existe efeito do agente *A*. O teste de hipótese é formulado da seguinte forma:

$$\begin{cases} H_0 : & \text{os parâmetros correspondentes a } M \text{ são iguais para os dois tratamentos;} \\ H_a : & \text{os parâmetros são diferentes.} \end{cases} \quad (4.1)$$

Se H_0 for rejeitada, o tratamento com agente quimiopreventivo é diferente do tratamento controle.

Caso contrário, não existe diferença entre os efeitos dos dois tratamentos aplicados nos animais. Os resultados do ajuste dos modelos são apresentados nas Tabelas 4.1 a 4.4, que foram obtidos de seguir as metodologias dadas nas Seções 2.4, 2.5 e 2.6. Preferímo-nos às estimativas com sub-índice 1 como pertencente ao tratamento de controle e as outras com sub-índice 2 para o tratamento com agente A.

Tabela 4.1: Resultados do modelo destrutivo de Poisson.

$P(m)/f(t)$	Estimativa (Erro padrão)	Intervalos de confiança	$-2\log L_{H_a}$	TRV ($\lambda_1 = \lambda_2$)	P-valor ρ
Poisson/Gama	$\hat{\lambda}_1 = 1,265$ (0,279)	[0,718 1,812]	997,34	7,151	0,007
	$\hat{\lambda}_2 = 2,323$ (0,396)	[1,547 3,099]			
	$\hat{\alpha}_1 = 3,585$ (1,168)	[1,296 5,874]			
	$\hat{\alpha}_2 = 4,119$ (1,036)	[2,088 6,151]			
	$\hat{\beta}_1 = 44,108$ (20,147)	[4,621 83,595]			
	$\hat{\beta}_2 = 40,608$ (14,208)	[12,759 68,456]			

Tabela 4.2: Resultados do modelo destrutivo binomial negativo.

$P(m)/f(t)$	Estimativa (Erro padrão)	Intervalos de confiança	$-2\log L_{H_a}$	TRV ($\mu_1 = \mu_2$) ($\tau_1 = \tau_2$)	P-valor ρ
BN/Gama	$\hat{\mu}_1 = 1,556$ (0,582)	[0,414 2,695]	945,584	1,168	0,56
	$\hat{\mu}_2 = 2,417$ (0,679)	[1,087 3,747]			
	$\hat{\tau}_1 = 1,871$ (1,030)	[0,147 3,889]			
	$\hat{\tau}_2 = 1,413$ (0,579)	[0,277 2,55]			
	$\hat{\alpha}_1 = 3,721$ (1,214)	[1,342 6,099]			
	$\hat{\alpha}_2 = 4,239$ (1,055)	[2,172 6,308]			
$\hat{\beta}_1 = 44,701$ (21,353)	[2,847 86,555]				
$\hat{\beta}_2 = 39,593$ (13,665)	[12,809 66,376]				

Tabela 4.3: Resultados do modelo destrutivo de Poisson com zeros inflacionados.

$P(m)/f(t)$	Estimativa (Erro padrão)	Intervalos de confiança	$-2\log L_{H_a}$	TRV ($\omega_1 = \omega_2$) ($\lambda_1 = \lambda_2$)	P-valor ρ
PZI/Gama	$\hat{\omega}_1 = 0,548$ (0,113)	[0,327 0,769]	964,438	2,678	0,26
	$\hat{\omega}_2 = 0,676$ (0,095)	[0,489 0,861]			
	$\hat{\lambda}_1 = 2,448$ (0,624)	[1,225 3,672]			
	$\hat{\lambda}_2 = 3,656$ (0,735)	[2,216 5,095]			
	$\hat{\alpha}_1 = 3,706$ (1,192)	[1,369 6,044]			
	$\hat{\alpha}_2 = 3,956$ (1,029)	[1,938 5,974]			
$\hat{\beta}_1 = 42,664$ (19,186)	[5,06 80,268]				
$\hat{\beta}_2 = 44,135$ (16,824)	[11,161 77,109]				

Tabela 4.4: Resultados do modelo destrutivo binomial negativo com zeros inflacionados.

$P(m)/f(t)$	Estimativa (Erro padrão)	Intervalos de confiança	$-2\log L_{H_0}$	TRV ($\omega_1 = \omega_2$) ($\mu_1 = \mu_2$) ($\tau_1 = \tau_2$)	P-valor ρ
BNZI/Gama	$\hat{\omega}_1 = 0,976$ (1,164) $\hat{\omega}_2 = 1,000$ (0,332)	$[-1,305$ 3,256] $[0,3490$ 1,650]	945,5838	1,1692	0,76
	$\hat{\mu}_1 = 1,590$ (1,844) $\hat{\mu}_2 = 2,418$ (1,021)	$[-2,023$ 5,204] $[0,4160$ 4,420]			
	$\hat{\tau}_1 = 1,780$ (4,419) $\hat{\tau}_2 = 1,414$ (1,092)	$[-6,880$ 10,441] $[-0,727$ 3,554]			
	$\hat{\alpha}_1 = 3,719$ (1,214) $\hat{\alpha}_2 = 4,239$ (1,056)	$[1,3400$ 6,099] $[2,1700$ 6,309]			
	$\hat{\beta}_1 = 44,686$ (21,348) $\hat{\beta}_2 = 39,597$ (13,679)	$[2,8450$ 86,527] $[0,3490$ 66,407]			

Os resultados do teste da razão de verossimilhança apresentados nas Tabelas 4.2, 4.3 e 4.4 mostram que o modelo binomial negativo, modelo PZI e o modelo BNZI são preferidos em relação ao modelo de Poisson, quando $\rho = 0,56$, $\rho = 0,26$, e $\rho = 0,76$, respectivamente. Esses resultados confirmam que não houve significância, isto é, o tratamento com agente A não teve efeito no experimento. A seguir são apresentados os resultados dos critérios de seleção do modelo para cada grupo de tratamento.

Tabela 4.5: Seleção de modelo para o grupo controle.

Modelo destrutivo	N° de parâmetros no modelo	DEV	AIC	BIC
Poisson	3	365,578	371,578	375,782
BN	4	353,045	361,045	366,650
PZI	4	355,367	363,367	368,971
BNZI	5	353,045	363,045	370,051

Tabela 4.6: Seleção de modelo para o grupo com agente A.

Modelo destrutivo	N° de parâmetros no modelo	DEV	AIC	BIC
Poisson	3	624,611	630,611	634,815
BN	4	592,539	600,539	606,144
PZI	4	609,072	617,072	622,677
BNZI	5	592,539	602,539	609,545

As Tabelas 4.5 e 4.6 mostram que o modelo BN e o modelo BNZI são os mais indicados para ajustar os dados. A seguir apresentamos uma tabela de frequências esperadas onde podemos observar o ajuste em relação ao número de tumores.

Tabela 4.7: *Frequências esperadas de cada modelo*

M^*	O	<i>Poisson</i>	<i>BN</i>	<i>PZI</i>	<i>BNZI</i>
0	27	14,80	24,22	26,95	28,06
1	16	20,71	11,24	7,89	12,01
2	6	14,50	7,01	9,49	6,91
3-4	5	9,13	8,10	12,17	7,12
≥ 5	4	0,86	7,69	3,50	5,24

A Tabela 4.7 apresenta as frequências observadas, O , e as frequências esperadas segundo os modelos de Poisson, binomial negativo, PZI e BNZI. Para este conjunto de dados os modelos PZI e BNZI são superiores em distinguir zeros porque estas distribuições foram formuladas para justificar o excesso de zeros. O modelo destrutivo PZI é o modelo que mais se ajusta ao números de zeros, mais a um custo muito alto já que o ajuste para valores diferentes de zero não é satisfatório em relação aos resultados apresentados pelos modelos destrutivos BN e BNZI.

Capítulo 5

Considerações Finais

Neste trabalho foi apresentado o modelo destrutivo com variável terminal, o qual é uma adaptação do modelo proposto por [Rodrigues *et al.* \(2010\)](#) onde foram utilizados os modelos discretos de Poisson, binomial negativo, Poisson com zeros inflacionados e binomial negativo com zeros inflacionados.

No Capítulo 3 os procedimentos de estimação dos parâmetros de interesse, foram desenvolvidos utilizando o método de máxima verossimilhança via algoritmo EM os quais produziram resultados satisfatórios.

Foi apresentado um algoritmo e o programa computacional de simulação dos dados, para o modelo destrutivo de Poisson com zeros inflacionados e variável terminal. De acordo com os resultados observados no estudo de simulação, observou-se que o procedimento de estimação puntual e probabilidade de cobertura foram consistentes assintoticamente.

Na aplicação com dados reais os modelos destrutivos propostos Poisson com zeros inflacionados e binomial negativo com zeros inflacionados se mostraram capazes de incorporar o excesso de zeros nos dados.

O teste da razão de verossimilhança dos modelos destrutivos binomial negativo, Poisson com zeros inflacionados e binomial negativo com zeros inflacionados, mostraram que o efeito do agente A não foi estatisticamente significativo, enquanto, o modelo destrutivo de Poisson mostrou o contrário, ao nível de significância de 5%.

Entre os quatro modelos mencionados o modelo binomial negativo com zeros inflacionados foi o mais adequado para o ajuste dos dados, de acordo com o teste da razão de verossimilhança e os critérios AIC e BIC.

É importante mencionar que os quatro modelos envolvidos, não foram analisados na presença de covariáveis. Portanto, a inclusão de covariáveis na modelagem do modelo destrutivo com variável terminal faz parte de pesquisas futuras.

Considerar o modelo destrutivo com variável terminal, na análise de dados forneceu resultados mais próximos da realidade, pois considera na análise, o processo de destruição dos tumores promovidos pelo tratamento avaliado.

Como pesquisas futuras enfatizamos os seguintes tópicos:

- Uma das propostas futuras para finalizar este trabalho de mestrado é implementar a simulação do modelo destrutivo com variável terminal para o modelo binomial negativo com zeros inflacionados.

- Aplicar a metodologia do modelo destrutivo de Poisson com zeros inflacionados e variável terminal, para dados reais com covariáveis. Podemos estudar o trabalho recente de [Chin-Chang \(2011\)](#), no tema de identificabilidade que foi feito para o modelo de Poisson com zeros inflacionados.
- Formular um procedimento Bayesiano para o modelo destrutivo com variável terminal .
- Tentativas para modelar a carcinogênese foram feitas do ponto de vista da análise de sobrevivência para ensaios clínicos humanos. Entre os pesquisadores existe o interesse em propor modelos de sobrevivência para carcinogênese que incorporam a possibilidade de indivíduos não serem susceptíveis ao câncer, ou seja, existe uma proporção da população que devido a certa intervenção (tratamento e/ou defesas naturais do organismo) pode vir a não ser susceptível ao câncer. Nesta área, existem pesquisas na literatura tais como: [Berkson & Gage \(1952\)](#), [Farewell \(1986\)](#), [Yang & Chen \(1991\)](#), [Yakovlev & Tsodikov \(1996\)](#), [Chen & Ibrahim \(2001\)](#), assim como modelos mais recentes [Yin & Ibrahim \(2005\)](#), [Chen & Liu \(2006\)](#), [Cooner *et al.* \(2007\)](#), [Rodrigues *et al.* \(2009\)](#), [Rodrigues *et al.* \(2010\)](#), entre muitos outros. Uma proposta é abordar este problema do ponto de vista do modelo destrutivo proposto por [Rodrigues *et al.* \(2010\)](#) com variável terminal, via análise de sobrevivência de longa duração, onde o número de tumores e os respectivos tempos de detecção, por animal são observados.
- O trabalho de [Wang *et al.* \(2001\)](#) apresenta uma análise de eventos recorrentes de dados que possuem censura informativa para pacientes com AIDS. Outra proposta seria adaptar o modelo destrutivo com variável terminal no contexto do trabalho de [Wang *et al.* \(2001\)](#) em relação à recorrência dos tumores de animais, onde a variável terminal comporta-se como um mecanismo de censura informativo.

Apêndice A

A.1 Câncer, Carcinogênese (INCA, 2011)

Câncer é o nome dado a um conjunto de mais de 100 doenças que têm em comum o crescimento desordenado de células, que invadem tecidos e órgãos. Dividindo-se rapidamente, estas células tendem a ser incontroláveis, determinando a formação de tumores malignos, que podem espalhar-se para outras regiões do corpo. Pesquisas sobre o câncer indicam uma alta incidência de mortalidade no mundo todo. Fatores como cigarro, álcool, exposição excessiva ao sol, etc. atuam como agentes cancerígenos ou carcinógenos alterando a estrutura genética (DNA) das células chamado processo de carcinogênese. O surgimento do câncer depende da intensidade e duração da exposição das células aos agentes cancerígenos. Por exemplo, o risco de uma pessoa desenvolver câncer de pulmão é diretamente proporcional ao número de cigarros fumados por dia e ao número de anos que ela vem fumando. A formação do câncer ou processo de carcinogênese é um processo que passa por vários estágios antes de chegar ao tumor visível. Os estágios são:

1. **Estágio de iniciação:** O estágio de iniciação é o primeiro estágio da carcinogênese. Nele as células sofrem o efeito dos agentes cancerígenos ou carcinógenos que provocam mutação na estrutura genética das células. Nesta fase um número inicial de células se encontram geneticamente alteradas ou danificadas, porém ainda não é possível clinicamente detectar o tumor. Estas células encontram-se "preparadas", ou seja, "iniciadas", para transformar-se em células malignas gerando os tumores.
2. **Estágio de promoção:** O estágio de promoção é o estágio intermediário da carcinogênese. Nele, as células geneticamente alteradas, ou seja, "iniciadas", sofrem o efeito dos agentes cancerígenos classificados como oncopromotores. A célula iniciada é transformada em célula maligna, de forma lenta e gradual. Para que ocorra essa transformação, é necessário um longo e continuado contato com o agente cancerígeno promotor. Esse estágio é caracterizado pela reversibilidade, ao contrario da iniciação e progressão. Células malignas podem regredir com a descontinuação dos agentes cancerígenos, porém elas só podem reaparecer com a restituição do agente cancerígeno. A suspensão do contato com agentes promotores muitas vezes interrompe o processo nesse estágio. Alguns componentes da alimentação e a exposição excessiva e prolongada a hormônios são exemplos de fatores que promovem a transformação de células iniciadas em malignas.
3. **Estágio de progressão:** Este é o último estágio e se caracteriza pela multiplicação descontrolada e irreversível das células alteradas. Nesse estágio o câncer já está instalado, evoluindo até o surgimento das primeiras manifestações clínicas da doença.

A.2 Esperança e Variância do modelo destrutivo com variável terminal

Usando as propriedades de esperança condicional, a esperança da variável destrutiva, M_i^* , com variável terminal, t_i^* , é dada por:

$$\begin{aligned} E[M_i^*|t_i^*] &= E\{E[M_i^* = m_i^*|M_i = m_i; \theta]\}, \\ &= E\{m_i F(t_i^*; \theta)\}, \\ &= E[M_i]F(t_i^*; \theta), \end{aligned} \tag{A.1}$$

com as mesmas propriedades, a variância da variável destrutiva, M_i^* , com variável terminal, t_i^* , é dada por:

$$\begin{aligned} Var[M_i^*|t_i^*] &= E\{Var[M_i^* = m_i^*|M_i = m_i; \theta]\} + Var\{E[M_i^* = m_i^*|M_i = m_i; \theta]\}, \\ &= E\{m_i F(t_i^*; \theta)(1 - F(t_i^*; \theta))\} + Var\{m_i F(t_i^*; \theta)\}, \\ &= E[M_i]F(t_i^*; \theta) - F(t_i^*; \theta)^2 + F(t_i^*; \theta)^2 Var[M_i]. \\ &= (Var[M_i] - E[M_i])F(t_i^*; \theta)^2 + E[M_i]F(t_i^*; \theta). \end{aligned} \tag{A.2}$$

A.3 condições de 1ª e 2ª ordem do modelo destrutivo de Poisson com variável terminal

Condições de 1ª ordem.

$$\frac{\partial}{\partial \lambda} \log[L(\lambda, \alpha, \beta|t_i^*)] = -\sum_{i=1}^n F(t_i^*; \alpha, \beta) + \frac{s_1}{\lambda} = 0, \tag{A.3}$$

$$\frac{\partial}{\partial \alpha} \log[L(\lambda, \alpha, \beta|t_i^*)] = -\lambda \frac{\partial}{\partial \alpha} \sum_{i=1}^n F(t_i^*; \alpha, \beta) - s_1(\log \beta + \frac{\partial}{\partial \alpha} \log \Gamma(\alpha)) + s_2 = 0, \tag{A.4}$$

$$\frac{\partial}{\partial \beta} \log[L(\lambda, \alpha, \beta|t_i^*)] = -\lambda \frac{\partial}{\partial \beta} \sum_{i=1}^n F(t_i^*; \alpha, \beta) - \alpha s_1/\beta + s_3/\beta^2 = 0. \tag{A.5}$$

Os estimadores de máxima verossimilhança $\hat{\lambda}$, $\hat{\alpha}$ e $\hat{\beta}$ serão obtidos derivando a equação A.3 em relação a α e β , obtendo-se

$$\frac{\partial}{\partial \alpha} \sum_{i=1}^n F(t_i^*; \alpha, \beta) = \frac{\partial}{\partial \beta} \sum_{i=1}^n F(t_i^*; \alpha, \beta) = 0.$$

Substituindo esta última expressão nas equações A.4 e A.5 obtemos que

$$\Psi(\hat{\alpha}) + \log(\hat{\beta}) = s_2/s_1, \quad (\text{A.6})$$

$$\hat{\beta} = s_3\hat{\alpha}^{-1}/s_1, \quad (\text{A.7})$$

onde $\Psi(\hat{\alpha}) = \frac{\partial}{\partial \alpha} \log \Gamma(\hat{\alpha})$ é a função digama.

Logo, substituindo (A.6) em (A.7) obtemos as seguintes equações após algumas simplificações

$$\hat{\lambda} = s_1 / \sum_{i=1}^n F(t_i^*; \hat{\alpha}, \hat{\beta}), \quad (\text{A.8})$$

$$\Psi(\hat{\alpha}) - \log(\hat{\alpha}) + \log(s_3/s_1) - s_2/s_1 = 0, \quad (\text{A.9})$$

A equação (A.9) será resolvida via método Newton-Raphson. Assim, substituindo $\hat{\alpha}$ em (A.7) obtemos $\hat{\beta}$ e de (A.8) obtemos $\hat{\lambda}$. Como as estimativas não são obtidas explicitamente, verificaremos que o ponto estimado $(\hat{\lambda}, \hat{\alpha}, \hat{\beta})$ seja um ponto máximo da função de verossimilhança através das condições de 2ª ordem dada no Capítulo 2.

Condições de 2ª ordem.

$$\frac{\partial^2}{\partial \lambda^2} l(\hat{\lambda}, \hat{\alpha}, \hat{\beta}) = -s_1/\hat{\lambda}^2 \quad (\text{A.10})$$

$$\frac{\partial^2}{\partial \alpha^2} l(\hat{\lambda}, \hat{\alpha}, \hat{\beta}) = -\hat{\lambda} \sum_{i=1}^n \frac{\partial^2}{\partial \alpha^2} F(t_i^*, \hat{\alpha}, \hat{\beta}) - s_1 \frac{\partial^2}{\partial \alpha^2} \log(\Gamma(\hat{\alpha})) \quad (\text{A.11})$$

$$\frac{\partial^2}{\partial \beta^2} l(\hat{\lambda}, \hat{\alpha}, \hat{\beta}) = -\hat{\lambda} \sum_{i=1}^n \frac{\partial^2}{\partial \beta^2} F(t_i^*, \hat{\alpha}, \hat{\beta}) + \frac{s_1 \hat{\alpha}}{\hat{\beta}^2} + \frac{2s_3}{\hat{\beta}^3} \quad (\text{A.12})$$

as derivadas cruzadas em relação a λ , α e β são dadas por:

$$\frac{\partial^2}{\partial \alpha \partial \lambda} l(\hat{\lambda}, \hat{\alpha}, \hat{\beta}) = \frac{\partial^2}{\partial \lambda \partial \alpha} l(\hat{\lambda}, \hat{\alpha}, \hat{\beta}) = -\sum_{i=1}^n \frac{\partial}{\partial \alpha} F(t_i^*, \hat{\alpha}, \hat{\beta}) \quad (\text{A.13})$$

$$\frac{\partial^2}{\partial \alpha \partial \beta} l(\hat{\lambda}, \hat{\alpha}, \hat{\beta}) = \frac{\partial^2}{\partial \beta \partial \alpha} l(\hat{\lambda}, \hat{\alpha}, \hat{\beta}) = -\sum_{i=1}^n \frac{\partial}{\partial \beta} F(t_i^*, \hat{\alpha}, \hat{\beta}) \quad (\text{A.14})$$

$$\frac{\partial^2}{\partial \beta \partial \lambda} l(\hat{\lambda}, \hat{\alpha}, \hat{\beta}) = \frac{\partial^2}{\partial \lambda \partial \beta} l(\hat{\lambda}, \hat{\alpha}, \hat{\beta}) = -\hat{\lambda} \sum_{i=1}^n \frac{\partial^2}{\partial \alpha \partial \beta} F(t_i^*, \hat{\alpha}, \hat{\beta}) - \frac{s_1}{\hat{\beta}} \quad (\text{A.15})$$

Todas as derivadas parciais de segunda ordem são calculados computacionalmente para verificar que o ponto de estimativas é um ponto máximo, seguindo a definição dada em (2.15) definido no Capítulo 2.

A.4 Distribuição da variável destrutiva BN dada a variável terminal

$$\begin{aligned}
P[M_i^* = m_i^* | t_i^*; N, \nu, \theta] &= \sum_{m_i = m_i^*}^{\infty} P[M_i^* = m_i^* | M_i = m_i; \theta] P[M_i = m_i; N, \nu] \\
&= \sum_{m_i = m_i^*}^{\infty} \binom{m_i}{m_i^*} F(t_i^*; \theta)^{m_i^*} (1 - F(t_i^*; \theta))^{m_i - m_i^*} \binom{N + m_i - 1}{m_i} \nu^N (1 - \nu)^{m_i} \\
&= \sum_{m_i = m_i^*}^{\infty} \frac{m_i! F(t_i^*; \theta)^{m_i^*} (1 - F(t_i^*; \theta))^{m_i - m_i^*} \Gamma(N + m_i)}{m_i^*! (m_i - m_i^*)! \Gamma(N) m_i!} \nu^N (1 - \nu)^{m_i - m_i^* + m_i^*} \\
&= \frac{\nu^N [(1 - \nu) F(t_i^*; \theta)]^{m_i^*}}{\Gamma(N) m_i^*!} \sum_{m_i = m_i^*}^{\infty} \Gamma(N + m_i) \frac{[(1 - \nu)(1 - F(t_i^*; \theta))]^{m_i - m_i^*}}{(m_i - m_i^*)!} \\
&= \frac{\nu^N [(1 - \nu) F(t_i^*; \theta)]^{m_i^*}}{\Gamma(N) m_i^*!} \frac{\Gamma(N + m_i^*)}{[F(t_i^*; \theta) + \nu - \nu F(t_i^*; \theta)]^{N + m_i^*}} \\
&= \frac{\Gamma(N + m_i^*)}{\Gamma(N) m_i^*!} \left[\frac{\nu}{\nu + (1 - \nu) F(t_i^*; \theta)} \right]^N \left[\frac{(1 - \nu) F(t_i^*; \theta)}{\nu + (1 - \nu) F(t_i^*; \theta)} \right]^{m_i^*} \\
&= \binom{N + m_i^* - 1}{m_i^*} \left[\frac{\nu}{\nu + (1 - \nu) F(t_i^*; \theta)} \right]^N \left[1 - \frac{\nu}{\nu + (1 - \nu) F(t_i^*; \theta)} \right]^{m_i^*}.
\end{aligned}$$

A.4.1 Condições de 1ª e 2ª ordem do modelo destrutivo BN com variável terminal

Condições de 1ª ordem

$$\frac{\partial}{\partial \mu} \log[L(\lambda, \alpha, \beta)] = - \sum_{i=1}^n \left[\frac{(1/\tau + m_i^*) \tau F(t_i^*; \alpha, \beta)}{1 + \mu \tau F(t_i^*; \alpha, \beta)} \right] = 0 \quad (\text{A.16})$$

$$\begin{aligned}
\frac{\partial}{\partial \tau} \log[L(\lambda, \alpha, \beta)] &= \sum_{i=1}^n \left[\frac{1}{\Gamma(1/\tau + m_i^*)} \frac{\partial \Gamma(1/\tau + m_i^*)}{\partial \tau} \right] - \frac{n}{\Gamma(1/\tau)} \frac{\partial \Gamma(1/\tau)}{\partial \tau} + s_1/\tau \\
&\quad - \sum_{i=1}^n \left[\frac{(1/\tau + m_i^*) \mu F(t_i^*; \alpha, \beta)}{1 + \mu \tau F(t_i^*; \alpha, \beta)} - \frac{\log(1 + \mu \tau F(t_i^*; \alpha, \beta))}{\tau^2} \right] = 0 \quad (\text{A.17})
\end{aligned}$$

$$\frac{\partial}{\partial \alpha} \log[L(\lambda, \alpha, \beta)] = - \sum_{i=1}^n \left[\frac{\mu \tau (1/\tau + m_i^*)}{1 + \mu \tau F(t_i^*; \alpha, \beta)} \frac{\partial}{\partial \alpha} F(t_i^*; \alpha, \beta) \right] + s_2 - s_1 (\log \beta + \frac{\partial}{\partial \alpha} \log \Gamma(\alpha)) = 0 \quad (\text{A.18})$$

$$\frac{\partial}{\partial \beta} \log[L(\lambda, \alpha, \beta)] = - \sum_{i=1}^n \left[\frac{\mu \tau (1/\tau + m_i^*)}{1 + \mu \tau F(t_i^*; \alpha, \beta)} \frac{\partial}{\partial \beta} F(t_i^*; \alpha, \beta) \right] + \frac{s_3}{\beta^2} - \frac{\alpha s_1}{\beta} = 0 \quad (\text{A.19})$$

Condições de 2ª ordem

$$\frac{\partial^2}{\partial \mu^2} l(\hat{\mu}, \hat{\tau}, \hat{\alpha}, \hat{\beta}) = \sum_{i=1}^n \frac{(1/\hat{\tau} + m_i^*) \hat{\tau}^2 F(t_i^*; \hat{\alpha}, \hat{\beta})^2}{(1 + \hat{\mu} \hat{\tau} F(t_i^*; \hat{\alpha}, \hat{\beta}))^2} \quad (\text{A.20})$$

$$\begin{aligned} \frac{\partial^2}{\partial \tau^2} l(\hat{\mu}, \hat{\tau}, \hat{\alpha}, \hat{\beta}) &= \sum_{i=1}^n \frac{\partial^2}{\partial \tau^2} \log \Gamma(1/\hat{\tau} + m_i^*) - \sum_{i=1}^n \left[(1/\hat{\tau} + m_i^*) \frac{\partial^2}{\partial \tau^2} \log(1 + \hat{\mu} \hat{\tau} F(t_i^*; \hat{\alpha}, \hat{\beta})) \right. \\ &\quad - \frac{1}{\hat{\tau}} \frac{\partial}{\partial \tau} \log(1 + \hat{\mu} \hat{\tau} F(t_i^*; \hat{\alpha}, \hat{\beta})) - \frac{\mu F(t_i^*; \hat{\alpha}, \hat{\beta})}{\hat{\tau}^2 (1 + \hat{\mu} \hat{\tau} F(t_i^*; \hat{\alpha}, \hat{\beta}))} \\ &\quad \left. + \frac{2 \log(1 + \hat{\mu} \hat{\tau} F(t_i^*; \hat{\alpha}, \hat{\beta}))}{\hat{\tau}^3} \right] - n \frac{\partial^2}{\partial \tau^2} \log \Gamma(1/\hat{\tau}) \end{aligned} \quad (\text{A.21})$$

$$\frac{\partial^2}{\partial \alpha^2} l(\hat{\mu}, \hat{\tau}, \hat{\alpha}, \hat{\beta}) = - \sum_{i=1}^n \left[\frac{\hat{\mu} \hat{\tau} (1/\hat{\tau} + m_i^*)}{1 + \hat{\mu} \hat{\tau} F(t_i^*; \hat{\alpha}, \hat{\beta})} \frac{\partial^2}{\partial \alpha^2} F(t_i^*; \hat{\alpha}, \hat{\beta}) + \left(\frac{\partial}{\partial \alpha} F(t_i^*; \hat{\alpha}, \hat{\beta}) \right)^2 \right] - s_1 \frac{\partial^2}{\partial \alpha^2} \log \Gamma(\hat{\alpha}) \quad (\text{A.22})$$

$$\frac{\partial^2}{\partial \beta^2} l(\hat{\mu}, \hat{\tau}, \hat{\alpha}, \hat{\beta}) = - \sum_{i=1}^n \left[\frac{\hat{\mu} \hat{\tau} (1/\hat{\tau} + m_i^*)}{1 + \hat{\mu} \hat{\tau} F(t_i^*; \hat{\alpha}, \hat{\beta})} \frac{\partial^2}{\partial \beta^2} F(t_i^*; \hat{\alpha}, \hat{\beta}) + \left(\frac{\partial}{\partial \beta} F(t_i^*; \hat{\alpha}, \hat{\beta}) \right)^2 \right] - \frac{2s_3}{\hat{\beta}^3} + \frac{\hat{\alpha} s_1}{\hat{\beta}^2} \quad (\text{A.23})$$

as derivadas cruzadas em relação a μ, τ, α, β são dadas por:

$$\frac{\partial^2}{\partial \mu \partial \tau} l(\hat{\mu}, \hat{\tau}, \hat{\alpha}, \hat{\beta}) = \frac{\partial^2}{\partial \tau \partial \mu} l(\hat{\mu}, \hat{\tau}, \hat{\alpha}, \hat{\beta}) = - \sum_{i=1}^n \left[\frac{(m_i^* - \hat{\mu} F(t_i^*; \hat{\alpha}, \hat{\beta})) F(t_i^*; \hat{\alpha}, \hat{\beta})}{(1 + \hat{\mu} \hat{\tau} F(t_i^*; \hat{\alpha}, \hat{\beta}))^2} \right] \quad (\text{A.24})$$

$$\frac{\partial^2}{\partial \mu \partial \alpha} l(\hat{\mu}, \hat{\tau}, \hat{\alpha}, \hat{\beta}) = \frac{\partial^2}{\partial \alpha \partial \mu} l(\hat{\mu}, \hat{\tau}, \hat{\alpha}, \hat{\beta}) = - \sum_{i=1}^n \left[\frac{\tau (1/\hat{\tau} + m_i^*)}{(1 + \hat{\mu} \hat{\tau} F(t_i^*; \hat{\alpha}, \hat{\beta}))^2} \frac{\partial}{\partial \alpha} F(t_i^*; \hat{\alpha}, \hat{\beta}) \right] \quad (\text{A.25})$$

$$\frac{\partial^2}{\partial \mu \partial \beta} l(\hat{\mu}, \hat{\tau}, \hat{\alpha}, \hat{\beta}) = \frac{\partial^2}{\partial \beta \partial \mu} l(\hat{\mu}, \hat{\tau}, \hat{\alpha}, \hat{\beta}) = - \sum_{i=1}^n \left[\frac{\tau (1/\hat{\tau} + m_i^*)}{(1 + \hat{\mu} \hat{\tau} F(t_i^*; \hat{\alpha}, \hat{\beta}))^2} \frac{\partial}{\partial \beta} F(t_i^*; \hat{\alpha}, \hat{\beta}) \right] \quad (\text{A.26})$$

$$\frac{\partial^2}{\partial \tau \partial \alpha} l(\hat{\mu}, \hat{\tau}, \hat{\alpha}, \hat{\beta}) = \frac{\partial^2}{\partial \alpha \partial \tau} l(\hat{\mu}, \hat{\tau}, \hat{\alpha}, \hat{\beta}) = - \sum_{i=1}^n \left[\frac{\mu (m_i^* - \hat{\mu} F(t_i^*; \hat{\alpha}, \hat{\beta}))}{(1 + \hat{\mu} \hat{\tau} F(t_i^*; \hat{\alpha}, \hat{\beta}))^2} \frac{\partial}{\partial \alpha} F(t_i^*; \hat{\alpha}, \hat{\beta}) \right] \quad (\text{A.27})$$

$$\frac{\partial^2}{\partial \tau \partial \beta} l(\hat{\mu}, \hat{\tau}, \hat{\alpha}, \hat{\beta}) = \frac{\partial^2}{\partial \beta \partial \tau} l(\hat{\mu}, \hat{\tau}, \hat{\alpha}, \hat{\beta}) = - \sum_{i=1}^n \left[\frac{\mu (m_i^* - \hat{\mu} F(t_i^*; \hat{\alpha}, \hat{\beta}))}{(1 + \hat{\mu} \hat{\tau} F(t_i^*; \hat{\alpha}, \hat{\beta}))^2} \frac{\partial}{\partial \beta} F(t_i^*; \hat{\alpha}, \hat{\beta}) \right] \quad (\text{A.28})$$

$$\begin{aligned} \frac{\partial^2}{\partial \alpha \partial \beta} l(\hat{\mu}, \hat{\tau}, \hat{\alpha}, \hat{\beta}) &= \frac{\partial^2}{\partial \beta \partial \alpha} l(\hat{\mu}, \hat{\tau}, \hat{\alpha}, \hat{\beta}) = - \sum_{i=1}^n \left[\frac{-\hat{\mu}^2 \hat{\tau}^2 (1/\hat{\tau} + m_i^*)}{(1 + \hat{\mu} \hat{\tau} F(t_i^*; \hat{\alpha}, \hat{\beta}))^2} \frac{\partial}{\partial \alpha} F(t_i^*; \hat{\alpha}, \hat{\beta}) \frac{\partial}{\partial \beta} F(t_i^*; \hat{\alpha}, \hat{\beta}) \right. \\ &\quad \left. + \frac{\hat{\mu} \hat{\tau} (1/\hat{\tau} + m_i^*)}{1 + \hat{\mu} \hat{\tau} F(t_i^*; \hat{\alpha}, \hat{\beta})} \frac{\partial}{\partial \alpha \partial \beta} F(t_i^*; \hat{\alpha}, \hat{\beta}) - \frac{s_1}{\hat{\beta}} \right] \end{aligned} \quad (\text{A.29})$$

Todas as derivadas parciais de segunda ordem são calculados computacionalmente para verificar que o ponto de estimativas é um ponto máximo, seguindo a definição dada em (2.15).

A.5 Distribuição da variável destrutiva PZI dada a variável terminal

$$P[M_i^* = m_i^* | t_i^*; \omega, \lambda, \theta] = \sum_{m_i = m_i^*}^{\infty} \binom{m_i}{m_i^*} F(t_i^*; \theta)^{m_i^*} (1 - F(t_i^*; \theta))^{m_i - m_i^*} \left((1 - \omega) I_{\{0\}}(m_i) + \omega \frac{e^{-\lambda} \lambda^{m_i}}{m_i!} \right),$$

onde

$$I_{\{0\}}(m_i) = \begin{cases} 1 & , m_i = 0; \\ 0 & , m_i \neq 0. \end{cases}$$

1. $m_i^* = 0$

$$\begin{aligned} P[M_i^* = 0 | t_i^*; \omega, \lambda, \theta] &= \sum_{m_i=0}^{\infty} (1 - F(t_i^*; \theta))^{m_i} \left((1 - \omega) I_{\{0\}}(m_i) + \omega \frac{e^{-\lambda} \lambda^{m_i}}{m_i!} \right) \\ &= 1 - \omega + \omega e^{-\lambda} + \sum_{m_i=1}^{\infty} (1 - F(t_i^*; \theta))^{m_i} \omega \frac{e^{-\lambda} \lambda^{m_i}}{m_i!} \\ &= 1 - \omega + \omega e^{-\lambda} \left(\sum_{m_i=1}^{\infty} \frac{[\lambda(1 - F(t_i^*; \theta))]^{m_i}}{m_i!} + 1 \right) \\ &= 1 - \omega + \omega e^{-\lambda F(t_i^*; \theta)}. \end{aligned}$$

2. $m_i^* = 1, 2, 3, \dots$

$$\begin{aligned} P[M_i^* = m_i^* | t_i^*; \omega, \lambda, \theta] &= \sum_{m_i = m_i^*}^{\infty} \binom{m_i}{m_i^*} F(t_i^*; \theta)^{m_i^*} (1 - F(t_i^*; \theta))^{m_i - m_i^*} \omega \frac{e^{-\lambda} \lambda^{m_i - m_i^* + m_i^*}}{m_i!} \\ &= \omega e^{-\lambda} \frac{[\lambda F(t_i^*; \theta)]^{m_i^*}}{m_i^*!} \sum_{m_i = m_i^*}^{\infty} \frac{[\lambda(1 - F(t_i^*; \theta))]^{m_i - m_i^*}}{(m_i - m_i^*)!} \\ &= \omega \frac{e^{-\lambda F(t_i^*; \theta)} [\lambda F(t_i^*; \theta)]^{m_i^*}}{m_i^*!}. \end{aligned}$$

Logo, de 1 e 2 obtemos:

$$P[M_i^* = m_i^* | t_i^*; \omega, \lambda, \theta] = (1 - \omega) I_{\{0\}}(m_i^*) + \omega \frac{e^{-\lambda F(t_i^*; \theta)} (\lambda F(t_i^*; \theta))^{m_i^*}}{m_i^*!},$$

onde

$$I_{\{0\}}(m_i^*) = \begin{cases} 1 & , m_i^* = 0; \\ 0 & , m_i^* \geq 1. \end{cases}$$

A.5.1 Função de verossimilhança ampliada do modelo destrutivo PZI com variável terminal

$$\begin{aligned}
L(\omega, \lambda, \alpha, \beta; D) &= L(\omega, \lambda, \alpha, \beta) \prod_{i=1}^{n_0} [(\varphi_i)^{Z_i} (1 - \varphi_i)^{1-Z_i}] \\
&= \prod_{i=1}^{n_0} \left\{ 1 - \omega + \omega e^{-\lambda F(t_i^*; \alpha, \beta)} \right\} \omega^{n-n_0} \cdot \prod_{M_i^* \notin A} L_i(\lambda, \alpha, \beta) \cdot \prod_{i=1}^{n_0} [(\varphi_i)^{Z_i} (1 - \varphi_i)^{1-Z_i}], \\
&= \prod_{i=1}^{n_0} (1 - \omega + \omega e^{-\lambda F(t_i^*; \alpha, \beta)}) \omega^{n-n_0} \cdot \prod_{M_i^* \notin A} L_i(\lambda, \alpha, \beta) \cdot \prod_{i=1}^{n_0} \left[\frac{1 - \omega}{1 - \omega + \omega e^{-\lambda F(t_i^*; \alpha, \beta)}} \right]^{Z_i} \left[\frac{\omega e^{-\lambda F(t_i^*; \alpha, \beta)}}{1 - \omega + \omega e^{-\lambda F(t_i^*; \alpha, \beta)}} \right]^{1-Z_i}, \\
&= \prod_{i=1}^{n_0} (1 - \omega + \omega e^{-\lambda F(t_i^*; \alpha, \beta)}) \cdot \prod_{M_i^* \notin A} L_i(\lambda, \alpha, \beta) \omega^{n-n_0} \cdot \frac{(1 - \omega)^{\sum_{i=1}^{n_0} Z_i} \omega^{n_0 - \sum_{i=1}^{n_0} Z_i} e^{-\lambda \sum_{i=1}^{n_0} [(1 - Z_i) F(t_i^*; \alpha, \beta)]}}{\prod_{i=1}^{n_0} (1 - \omega + \omega e^{-\lambda F(t_i^*; \alpha, \beta)})}, \\
&= \omega^{n - \sum_{i=1}^{n_0} Z_i} (1 - \omega)^{\sum_{i=1}^{n_0} Z_i} e^{-\lambda \sum_{i=1}^{n_0} [(1 - Z_i) F(t_i^*; \alpha, \beta)]} \cdot \prod_{M_i^* \notin A} L_i(\lambda, \alpha, \beta), \\
&= \omega^{n-Z} (1 - \omega)^Z \exp \left\{ -\lambda \sum_{i=1}^{n_0} [(1 - Z_i) F(t_i^*)] \right\} \cdot \prod_{M_i^* \notin A} L_i(\lambda, \alpha, \beta),
\end{aligned}$$

onde $Z = \sum_{i=1}^{n_0} Z_i$.

A.6 Distribuição da variável destrutiva BNZI com variável terminal

$$P[M_i^* = m_i^* | t_i^*; \omega, \mu, \tau, \theta] = \sum_{m_i = m_i^*}^{\infty} \binom{m_i}{m_i^*} F(t_i^*; \theta)^{m_i^*} (1 - F(t_i^*; \theta))^{m_i - m_i^*} \left((1 - \omega) I_{\{0\}}(m_i) + \omega \frac{\Gamma(N + m_i)}{\Gamma(N) m_i!} \nu^N (1 - \nu)^{m_i} \right),$$

onde $\eta = (\omega, \mu, \tau)$ e $\theta = \alpha, \beta$ e

$$I_{\{0\}}(m_i^*) = \begin{cases} 1 & , m_i^* = 0; \\ 0 & , m_i^* \geq 1. \end{cases}$$

1. $m_i^* = 0$

$$\begin{aligned}
P[M_i^* = 0 | t_i^*; \eta, \theta] &= \sum_{m_i=0}^{\infty} (1 - F(t_i^*; \theta))^{m_i} \left((1 - \omega) I_{\{0\}}(m_i) + \omega \frac{\Gamma(N + m_i)}{\Gamma(N) m_i!} \nu^N (1 - \nu)^{m_i} \right) \\
&= 1 - \omega + \omega \theta^N + \sum_{m_i=1}^{\infty} (1 - F(t_i^*; \theta))^{m_i} \omega \frac{\Gamma(N + m_i)}{\Gamma(N) m_i!} \nu^N (1 - \nu)^{m_i} \\
&= 1 - \omega + \frac{\omega \nu^N}{\Gamma(N)} \left(\sum_{m_i=1}^{\infty} \Gamma(N + m_i) \frac{[(1 - F(t_i^*; \theta))(1 - \nu)]^{m_i}}{m_i!} + \Gamma(N) \right) \\
&= 1 - \omega + \frac{\omega \nu^N}{\Gamma(N)} \left(\sum_{m_i=0}^{\infty} \Gamma(N + m_i) \frac{[(1 - F(t_i^*; \theta))(1 - \nu)]^{m_i}}{m_i!} \right) \\
&= 1 - \omega + \frac{\omega \nu^N}{\Gamma(N)} \frac{\Gamma(N)}{\nu + (1 - \nu) F(t_i^*; \theta)} \quad N \\
&= 1 - \omega + \omega \left[\frac{\nu}{\nu + (1 - \nu) F(t_i^*; \theta)} \right]^N.
\end{aligned}$$

2. $m_i^* = 1, 2, 3, \dots$

$$\begin{aligned}
P[M_i^* = m_i^* | t_i^*; \eta, \theta] &= \sum_{m_i=m_i^*}^{\infty} \binom{m_i}{m_i^*} F(t_i^*; \theta)^{m_i^*} (1 - F(t_i^*; \theta))^{m_i - m_i^*} \omega \frac{\Gamma(N + m_i)}{\Gamma(N) m_i!} \nu^N (1 - \nu)^{m_i}, \\
&= \omega \frac{\nu^N [(1 - \nu) F(t_i^*; \theta)]^{m_i^*}}{\Gamma(N) m_i^*!} \sum_{m_i=m_i^*}^{\infty} \Gamma(N + m_i) \frac{[(1 - \nu)(1 - F(t_i^*; \theta))]^{m_i - m_i^*}}{(m_i - m_i^*)!}, \\
&= \omega \frac{\nu^N [(1 - \nu) F(t_i^*; \theta)]^{m_i^*}}{\Gamma(N) m_i^*!} \frac{\Gamma(N + m_i^*)}{[\nu + (1 - \nu) F(t_i^*; \theta)]^{N + m_i^*}}, \\
&= \omega \left[\frac{\nu}{\nu + (1 - \nu) F(t_i^*; \theta)} \right]^N \left[1 - \frac{\nu}{\nu + (1 - \nu) F(t_i^*; \theta)} \right]^{m_i^*},
\end{aligned}$$

Logo, de 1 e 2 obtemos:

$$P[M_i^* = m_i^* | t_i^*; \omega, \lambda, \theta] = (1 - \omega) I_{\{0\}}(m_i^*) + \omega \frac{\Gamma(N + m_i^*)}{\Gamma(N) m_i^*!} \left[\frac{\nu}{\nu + (1 - \nu) F(t_i^*; \theta)} \right]^N \left[1 - \frac{\nu}{\nu + (1 - \nu) F(t_i^*; \theta)} \right]^{m_i^*},$$

onde

$$I_{\{0\}}(m_i^*) = \begin{cases} 1 & , m_i^* = 0; \\ 0 & , m_i^* \geq 1. \end{cases}$$

A.6.1 Função de verossimilhança ampliada do modelo destrutivo BNZI com variável terminal

A função de verossimilhança baseada em dados ampliados para o i -ésimo animal do modelo destrutivo Binomial Negativo com zeros inflacionados e variável terminal, t_i^* , é dada por:

$$L_i(\omega, \mu, \tau, \alpha, \beta) = m_i^*! \prod_{j=1}^{m_i^*} \frac{f(t_{ij}; \alpha, \beta)}{F(t_i^*; \alpha, \beta)} \left((1 - \omega) I_{\{0\}}(m_i^*) + \frac{\omega \Gamma(\frac{1}{\tau} + m_i^*)}{\Gamma(\frac{1}{\tau}) m_i^*!} \left[\frac{1}{1 + \mu \tau F(t_i^*; \theta)} \right]^{\frac{1}{\tau}} \left[\frac{\mu \tau F(t_i^*; \theta)}{1 + \mu \tau F(t_i^*; \theta)} \right]^{m_i^*} \right).$$

Logo, para os n animais a função de verossimilhança é expressa como:

$$L(\omega, \mu, \tau, \alpha, \beta) = \prod_{i=1}^{n_0} L_i(\omega, \mu, \tau, \alpha, \beta) \prod_{M_i^* \notin A} L_i(\omega, \mu, \tau, \alpha, \beta)$$

segue

$$\begin{aligned} &= \prod_{i=1}^{n_0} \left\{ 1 - \omega + \omega \left[\frac{1}{1 + \mu\tau F(t_i^*; \theta)} \right]^{\frac{1}{\tau}} \right\} \prod_{M_i^* \notin A} \left\{ m_i^*! \prod_{j=1}^{m_i^*} \frac{f(t_{ij}; \alpha, \beta)}{F(t_i^*; \alpha, \beta)} \frac{\omega \Gamma(\frac{1}{\tau} + m_i^*)}{\Gamma(\frac{1}{\tau}) m_i^*!} \left[\frac{1}{1 + \mu\tau F(t_i^*; \theta)} \right]^{\frac{1}{\tau}} \left[\frac{\mu\tau F(t_i^*; \theta)}{1 + \mu\tau F(t_i^*; \theta)} \right]^{m_i^*} \right\}, \\ &= \prod_{i=1}^{n_0} \left\{ 1 - \omega + \omega \left[\frac{1}{1 + \mu\tau F(t_i^*; \theta)} \right]^{\frac{1}{\tau}} \right\} \cdot \omega^{n-n_0} \cdot \prod_{M_i^* \notin A} L_i(\mu, \tau, \alpha, \beta), \end{aligned}$$

Com uso da variável latente

$$Y_i = \begin{cases} 1 & , \text{ com probabilidade } \phi_i; \\ 0 & , \text{ com probabilidade } 1 - \phi_i, \end{cases}$$

os procedimentos da função de verossimilhança baseada em dados ampliados são mostrados a seguir:

$$\begin{aligned} L(\omega, \mu, \tau, \alpha, \beta; H_a) &= L(\omega, \mu, \tau, \alpha, \beta) \cdot \prod_{i=1}^{n_0} [(\phi_i)^{Y_i} (1 - \phi_i)^{1-Y_i}], \\ &= \prod_{i=1}^{n_0} \left\{ 1 - \omega + \omega [1 + \mu\tau F(t_i^*; \theta)]^{-\frac{1}{\tau}} \right\} \omega^{n-n_0} \cdot \prod_{M_i^* \notin A} L_i(\mu, \tau, \alpha, \beta) \\ &\quad \cdot \prod_{i=1}^{n_0} \left[\frac{1 - \omega}{1 - \omega + \omega [1 + \mu\tau F(t_i^*; \theta)]^{-\frac{1}{\tau}}} \right]^{Y_i} \left[\frac{\omega [1 + \mu\tau F(t_i^*; \theta)]^{-\frac{1}{\tau}}}{1 - \omega + \omega [1 + \mu\tau F(t_i^*; \theta)]^{-\frac{1}{\tau}}} \right]^{1-Y_i}, \\ &= \prod_{i=1}^{n_0} \left\{ 1 - \omega + \omega [1 + \mu\tau F(t_i^*; \theta)]^{-\frac{1}{\tau}} \right\} \omega^{n-n_0} \cdot \prod_{M_i^* \notin A} L_i(\mu, \tau, \alpha, \beta) \\ &\quad \cdot \frac{(1 - \omega)^{\sum_{i=1}^{n_0} Y_i} \omega^{n_0 - \sum_{i=1}^{n_0} Y_i} \prod_{i=1}^{n_0} (1 + \mu\tau F(t_i^*; \theta))^{\frac{Y_i-1}{\tau}}}{\prod_{i=1}^{n_0} [1 - \omega + \omega (1 + \mu\tau F(t_i^*; \theta))^{-\frac{1}{\tau}}]}, \\ &= \omega^{n - \sum_{i=1}^{n_0} Y_i} \cdot (1 - \omega)^{\sum_{i=1}^{n_0} Y_i} \cdot \prod_{i=1}^{n_0} (1 + \mu\tau F(t_i^*; \theta))^{\frac{Y_i-1}{\tau}} \cdot \prod_{M_i^* \notin A} L_i(\mu, \tau, \alpha, \beta), \\ &= \omega^{n-Y} \cdot (1 - \omega)^Y \cdot \prod_{i=1}^{n_0} (1 + \mu\tau F(t_i^*; \theta))^{\frac{Y_i-1}{\tau}} \cdot \prod_{M_i^* \notin A} L_i(\mu, \tau, \alpha, \beta), \end{aligned}$$

em que $Y = \sum_{i=1}^{n_0} Y_i$.

Referências

- Akaike, H. (1973). *Information theory and an extension of the maximum likelihood principle*, volume 1, pages 267–281. Akademiai Kiado. 16
- Berkson, J. & Gage, R. P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, **47**(259), 501–515. 39
- Bohning, D., Dietz, E., Schlattmann, P., Mendonça, L. & Kirchner, U. (2000). The zero-inflated poisson model and the decayed, missing and filled teeth index in dental epidemiology. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **163**(1), 121–122. 10
- Bolfarine, H. & Sandoval, M. (2001). *Introdução à inferência estatística*. SBM. 12, 13
- Boone, C. W., Kelloff, G. J. & Malone, W. E. (1990). Identification of candidate cancer chemopreventive agents and their evaluation in animal models and human clinical trials: A review. **50**(1), 2–9. 2
- Borgatto, A. F. (2004). *Modelos para proporções com superdispersão e excesso de zeros - Um procedimento bayesiano*. Tese (Doutorado em Agronomia). Universidade Federal de São Paulo - Piracicaba. 10
- Casella, G. & Berger, R. L. (2001). *Statistical Inference*. Duxbury/Thomson Learning, Pacific Grove, CA, second edition. 19
- Chen, M.-H. & Ibrahim, J. G. (2001). Maximum likelihood methods for cure rate models with missing covariates. *Biometrics*, **57**(1), 43–52. 39
- Chen, Y. H. J. & Liu, G. H. F. (2006). A note on the estimate of treatment effect from a Cox regression model when the proportionality assumption is violated. *Communications in Statistics – Theory and Methods*, **35**(1-3), 521–526. 39
- Chin-Chang (2011). Identifiability of zero-inflated poisson models. *Brazilian Journal of Probability and Statistics*. 39
- Cooner, F., Banerjee, S., Carlin, B. P. & Sinha, D. (2007). Flexible cure rate modeling under latent activation schemes. *Journal of the American Statistical Association*, **102**, 560–572. 39
- Dempster, A., Laird, N. & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39**, 1–38. 13, 25, 29
- Dunson, D. B. & Dinse, G. E. (2000). Distinguishing effects on tumor multiplicity and growth rate in chemoprevention experiments. *Biometrics*, **56**(4), 1068–1075. 3
- Farewell, V. (1986). Mixture models in survival analysis: Are they worth the risk. *Can. J. Statist.*, **14**, 257–262. 39
- Freedman, L., Midthune, D., Brown, C., Steele, V. & Kelloff, G. (1993). Statistical analysis of cancer chemoprevention experiments. *JSTOR: Biometrics*, **49**, 259–268. i, ii, 2, 4, 6, 7, 17, 19, 34
- Gupta, P. L., Ramesh C., G. & T., R. C. (1996). Analysis of zero-adjusted count data. *Computational Statistics & Data Analysis*, **23**(2), 207 – 218. 10
- Johnson, N. L.; Kotz, S. (1969). *Discrete Distributions: distribution in statistics* . John Wiley and Sons Inc.

- Kodell, R. L. & Chen, J. J. (2001). Inferring effects on tumor frequencies and times to observation in the analysis of tumor multiplicity data. *Biometrical Journal*, **43**(4), 447–460. [3](#)
- Kokoska, S. M. (1987). The analysis of cancer chemoprevention experiments. *Biometrics*, **43**, 525–534. [i](#), [ii](#), [2](#), [3](#), [4](#), [6](#), [7](#), [17](#)
- Kokoska, S. M. (1988). Including data from early deaths in the analysis of cancer chemoprevention experiments. *Applied Mathematics Letters*, **1**(2), 197 – 201. [2](#), [14](#)
- Kumara, S. S. P. & Chin, H. C. (2003). Modeling accident occurrence at signalized tee intersections with special emphasis on excess zeros. *Traffic Injury Prevention*, **4**(1), 53–57. [10](#)
- Lambert, D. (1992). Zero-inflated poisson regression with an application to defects in manufacturing. *Technometrics*, **34**(1), 1–14. [9](#), [10](#)
- Molefe, D. F., Chen, J. J., Howard, P. C., Miller, B. J., Sambuco, C. P., Forbes, P. D. & Kodell, R. L. (2005). Tests for effects on tumor frequency and latency in multiple dosing photocarcinogenicity experiments. *Journal of Statistical Planning and Inference*, **129**(1-2), 39 – 58. [3](#)
- Nelder, J. A. & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, **135**(3), 370–384. [8](#)
- Ridout, M., Demetrio, C. G. B. & Hinde, J. (1998). Models for Count Data with Many Zeros. pages 179–192. [10](#)
- Rodrigues, J. (2003). Bayesian analysis of zero-inflated distributions. *Communications in Statistics – Theory and Methods*, **32**(2), 281–289. [10](#)
- Rodrigues, J., Cancho, V. G., de Castro, M. & Louzada-Neto, F. (2009). On the unification of the long-term survival models. *Statistics and Probability Letters*, **79**, 753–759. [39](#)
- Rodrigues, J., de Castro, M., Balakrishnan, N. & Cancho, V. (2010). Destructive weighted Poisson cure rate models. *Lifetime Data Analysis*. [i](#), [ii](#), [3](#), [4](#), [6](#), [17](#), [18](#), [38](#), [39](#)
- Saito, M. (2005). *Inferencia Bayesiana para dados discretos com excesso de zeros*. Dissertação (Mestrado em Estatística). Universidade Federal de São Carlos. [10](#)
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464. [16](#)
- Silva, D. D. (2009). *Classe de distribuições serie de potencias inflacionadas com aplicações*. Dissertação (Mestrado em Estatística). Universidade Federal de São Carlos. [8](#)
- Tanner, M. A. (1996). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. Springer, New York, NY, third edition. [24](#), [28](#)
- Thompson, H. J., Meeker, L. D. & Kokoska, S. (1984). Effect of an inorganic and organic form of dietary selenium on the promotional stage of mammary carcinogenesis in the rat. *Cancer Research*, **44**(7), 2803–2806. [1](#)
- Wang, M.-C., Qin, J. & Chiang, C.-T. (2001). Analyzing recurrent event data with informative censoring. *Journal of the American Statistical Association*, **96**(455), pp. 1057–1065. [39](#)
- Yakovlev, A. Y. & Tsodikov, A. D. (1996). *Stochastic Models of Tumor Latency and Their Biostatistical Applications*. World Scientific, Singapore. [3](#), [39](#)
- Yang, G. L. & Chen, C. W. (1991). A stochastic two-stage carcinogenesis model: a new approach to computing the probability of observing tumor in animal bioassays. *Mathematical Biosciences*, **104**, 247–258. [39](#)
- Yin, G. & Ibrahim, J. G. (2005). Cure rate models: a unified approach. *The Canadian Journal of Statistics*, **33**(4), 559–570. [39](#)