

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

**Distribuição COM-Poisson na análise de dados de
experimentos de quimioprevenção do câncer em animais**

Angélica Maria Tortola Ribeiro
Orientador: Prof. Dr. Josemar Rodrigues

São Carlos, março de 2012

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

**Distribuição COM-Poisson na análise de dados de
experimentos de quimioprevenção do câncer em animais**

Angélica Maria Tortola Ribeiro
Orientador: Prof. Dr. Josemar Rodrigues

Dissertação apresentada ao
Departamento de Estatística da
Universidade Federal de São Carlos
para obtenção do Título de
Mestre em Estatística

São Carlos, março de 2012

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

R484dc

Ribeiro, Angélica Maria Tortola.

Distribuição COM-Poisson na análise de dados de experimentos de quimioprevenção do câncer em animais / Angélica Maria Tortola Ribeiro. -- São Carlos : UFSCar, 2012.
80 f.

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

1. Análise de sobrevivência. 2. Carcinogênese. 3. Distribuição COM-Poisson. 4. Modelo destrutivo. 5. Modelos de sobrevivência de longa-duração. I. Título.

CDD: 519.9 (20^a)

Angélica Maria Tortola Ribeiro

**DISTRIBUIÇÃO COM-POISSON NA ANÁLISE DE EXPERIMENTOS
DE QUIMIOPREVENÇÃO DO CÂNCER EM ANIMAIS**

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

Aprovada em 16 de março de 2012.

BANCA EXAMINADORA

Presidente



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

1º Examinador



Prof. Dr. Heleno Bolfarine (IME-USP)

2º Examinador



Prof. Dr. Mário de Castro Andrade Filho (ICMC-USP)

Agradecimentos

Primeiramente a Deus, por todas as graças que me concedeu pelas bênçãos e oportunidades com as quais me presenteou, por sempre iluminar-me e guiar-me em meus caminhos.

À minha maravilhosa família que sempre me apoiou e incentivou-me a buscar minha realização pessoal e profissional em especial à minha querida mãe Helena Maria e às minhas irmãs Jaqueline e Heloisa.

Ao Prof. Dr. Josemar Rodrigues, meu orientador, por me presentear com tantas oportunidades, sempre me incentivando a buscar continuamente o aprendizado, além da confiança depositada.

Aos membros da banca pela participação, crítica e sugestões.

Aos professores do Departamento de Estatística, por me auxiliarem em minhas dúvidas sempre prestativos e pacientes.

Aos meus colegas, pelo apoio incondicional, pela amizade que me foi tão importante, por todo conhecimento que pude adquirir com o auxílio deles.

Com muito carinho,
Angélica Maria Tortola Ribeiro.

Resumo

Experimentos que envolvem a indução química de substâncias cancerígenas em animais são comuns na área biológica. O interesse destes experimentos é, em geral, avaliar o efeito de uma substância quimiopreventiva na destruição das células danificadas. Neste tipo de estudo, duas variáveis de interesse são o número de tumores induzidos e seus tempos de desenvolvimento. Exploramos o uso do modelo estatístico proposto por Kokoska (1987) para a análise de dados de experimentos de quimioprevenção de câncer em animais. Flexibilizamos o modelo de Kokoska (1987), posteriormente utilizado por Freedman (1993), considerando para a variável número de tumores induzidos a distribuição Conway-Maxwell Poisson (COM-Poisson). Esta distribuição tem demonstrado eficiência devido à sua grande flexibilidade, quando comparada a outras distribuições discretas, para acomodar problemas relacionados à subdispersão e sobredispersão encontrados frequentemente em dados de contagem. A proposta deste trabalho consiste em adaptar a teoria de modelo destrutivo de longa duração (Rodrigues *et al.*, 2011) para experimentos de quimioprevenção do câncer em animais, com o propósito de avaliar a eficiência de tratamentos contra o câncer. Diferente da proposta de Rodrigues *et al.* (2011), formulamos um modelo para a variável número de tumores malignos detectados por animal, supondo que sua probabilidade de detecção não é mais constante, e sim dependente do instante de tempo. Esta é uma abordagem extremamente importante para experimentos quimiopreventivos de câncer, pois torna a análise mais realista e precisa. Realizamos um estudo de simulação com o propósito de avaliar a eficiência do modelo proposto e verificar as propriedades assintóticas dos estimadores de máxima verossimilhança. Analisamos também um conjunto de dados reais apresentado no artigo de Freedman (1993), visando demonstrar a eficiência do modelo COM-Poisson em relação aos resultados por ele obtidos com as distribuições Poisson e Binomial Negativa.

Palavras-chave: Iniciação e Promoção do Tumor; Carcinogênese; Distribuição COM-Poisson; Modelo Destrutivo; Modelos de Sobrevivência de Longa Duração.

Abstract

Experiments involving chemical induction of carcinogens in animals are common in the biological area. Interest in these experiments is, in general, evaluating the chemopreventive effect of a substance in the destruction of damaged cells. In this type of study, two variables of interest are the number of induced tumors and their development times. We explored the use of statistical model proposed by Kokoska (1987) for the analysis of experimental data of chemoprevention of cancer in animals. We flexibility the Kokoska's model, subsequently used by Freedman (1993), whereas for the variable number of tumors induced Conway-Maxwell Poisson (COM-Poisson) distribution. This distribution has demonstrated efficiency due to its great flexibility, when compared to other discrete distributions to accommodate problems related to sub-dispersion and super-dispersion often found in count data. The purpose of this paper is to adapt the theory of long-term destructive model (Rodrigues *et al.*, 2011) for experiments chemoprevention of cancer in animals, in order to evaluate the effectiveness of cancer treatments. Unlike the proposed Rodrigues *et al.* (2011), we formulate a model for the variable number of detected malignant tumors per animal, assuming that the probability of detection is no longer constant, but dependent on the time step. This is an extremely important approach to cancer chemoprevention experiments, because it makes the analysis more realistic and accurate. We conducted a simulation study, in order to evaluate the efficiency of the proposed model and to verify the asymptotic properties of maximum likelihood estimators. We also analyze a real data set presented in the article by Freedman (1993), to demonstrate the efficiency of the COM-Poisson model compared to results obtained by him with the Poisson and Negative Binomial distributions.

Keywords: Tumor Initiation and Promotion; Carcinogens; COM-Poisson Distribution; Destructive Model; Long-Term Survival Models.

Sumário

Lista de Figuras	vii
Lista de Tabelas	ix
1 Introdução	1
1.1 Análise de Sobrevivência de Longa Duração	4
1.2 O Modelo Destrutivo	6
1.3 Objetivos	8
1.3.1 Objetivos Específicos	8
1.4 Contribuições	8
1.5 Organização do Trabalho	9
2 Processo Destrutivo em Dois Estágios	11
2.1 Longa Duração em Dois Estágios	12
2.1.1 Função de Sobrevivência Mínima	13
2.2 Função de Probabilidade de $M^*(t)$	19
3 A Distribuição COM-Poisson	29
3.1 Função de Probabilidade de $M^*(t)$	32
3.2 Função de Sobrevivência Mínima	34
4 O Modelo Kokoska/Freedman	37
4.1 Função de Verossimilhança	38
4.2 Estudo de Simulação	40
4.3 Aplicação a Dados Reais	48
5 Conclusões	60
5.1 Considerações Finais	61
5.2 Propostas para Pesquisas Futuras	61
A Modelo Unificado de Sobrevivência	63

B Modelos de Sobrevivência com Fração de Cura	72
C Estatísticas de Ordem	75
Referências Bibliográficas	77

Lista de Figuras

1.1	Esquema de experimentos com ratos.	4
1.2	Estimativas de Kaplan-Meier para $S(t)$	5
2.1	Comportamento de $S_{min}(t; \lambda, \Psi)$ considerando a distribuição de Poisson com parâmetros $\lambda = 1$, $\lambda = 2$ e $\lambda = 3$, respectivamente, para M	27
2.2	Comportamento de $S_{min}(t; \mu, \tau, \Psi)$ considerando a distribuição Binomial Negativa com parâmetros $(\mu = 2.5, \tau = 0.5)$, $(\mu = 3.5, \tau = 0.5)$ e $(\mu = 6.5, \tau = 0.5)$, respectivamente, para M	28
3.1	Comportamento de $S_{min}(t; \eta, \phi, \Psi)$ utilizando a distribuição COM-Poisson com parâmetros $(\eta = 2.5, \phi = 1.5)$, $(\eta = 3.5, \phi = 1.5)$ e $(\eta = 6.5, \phi = 1.5)$, respectivamente, para a variável M	36
4.1	Estimativas de $S_{min}(t)$ para os dados simulados com $\phi = 0, 7$	45
4.2	Estimativas de $S_{min}(t)$ para os dados simulados com $\phi = 1, 0$	46
4.3	Estimativas de $S_{min}(t)$ para os dados simulados com $\phi = 1, 5$	47
4.4	Estimativas de $S_{min}(t)$ para os dados simulados considerando as porcentagens de censura 0%, 30%, 70% e 100%.	49
4.5	Estimativas de Kaplan-Meier para os grupos Controle e Tratamento.	51
4.6	Estimativas da média da variável $M^*(t)$ para os grupos controle e tratamento, considerando o modelo COM-Poisson/Weibull.	56
4.7	Estimativas da variância da variável $M^*(t)$ para os grupos controle e tratamento, considerando o modelo COM-Poisson/Weibull.	57
4.8	Distribuições ajustadas e Estimativas de Kaplan-Meier para os grupos Controle e Tratamento, respectivamente.	59
4.9	Estimativas das funções de cura para as distribuições Poisson, Binomial Negativa e COM-Poisson, avaliadas para os grupos controle e tratamento.	59
A.1	Comportamento de $f(t)$ para quatro grupos fictícios.	66
A.2	Comportamento de $S(t)$ para quatro grupos fictícios.	67

A.3 Comportamento de $h(t)$ para quatro grupos fictícios. 69

Lista de Tabelas

2.1	Representação de $S_{min}(t; \theta, \Psi)$, $p_0(t; \theta, \Psi)$ e p_0 para as distribuições Poisson e Binomial Negativa.	27
3.1	Representação das funções $S_{min}(t; \theta, \Psi)$, $p_0(t; \theta, \Psi)$ e p_0 para as distribuições de Poisson com parâmetro $\theta = \lambda$, Binomial Negativa com parâmetros $\theta = (\mu, \tau)$ e COM-Poisson com parâmetros $\theta = (\eta, \phi)$	35
4.1	Modelos Paramétricos	40
4.2	Média, viés, EQM e PC para os estimadores de máxima verossimilhança das 1000 amostras simuladas utilizando $\phi = 0,7$	43
4.3	Média, viés, EQM, e PC para os estimadores de máxima verossimilhança das 1000 amostras simuladas utilizando $\phi = 1,0$	44
4.4	Média, viés, EQM, e PC para os estimadores de máxima verossimilhança das 1000 amostras simuladas utilizando $\phi = 1,5$	44
4.5	Valores reais para as taxas de cura comparadas as taxas de cura estimadas para tamanhos de amostra $n = 30$, $n = 50$, $n = 100$ e $n = 300$, considerando valores de $\phi = 0,7$, $\phi = 1,0$ e $\phi = 1,5$	45
4.6	Média, viés, EQM, e PC para os estimadores de máxima verossimilhança das 1000 amostras simuladas de tamanho 100, considerando diferentes porcentagens de censuras nos dados.	48
4.7	Valores reais para as taxas de cura comparadas as taxas de cura estimadas para as porcentagens de censura 0%, 30%, 70%, e 100%, considerando $\phi = 0,7$ e $n = 100$	48
4.8	Conjunto de dados de tumores mamários em ratos experimentais	50
4.9	Resultados dos Modelos Ajustados usando a Distribuição de Poisson	52
4.10	Resultados dos Modelos Ajustados usando a Distribuição Binomial Negativa	52
4.11	Resultados dos Modelos Ajustados usando a Distribuição COM-Poisson	52
4.12	Teste da Razão de Verossimilhança (TRV) para a distribuição do número de tumores iniciados	53

4.13	Teste da Razão de Verossimilhança (TRV) para as distribuições dos tempos de detecção dos tumores	54
4.14	Estimativas da média e variância das variáveis M e T para os modelos apresentados.	54
4.15	Estimação da função de cura, $p_0(t; \theta, \Psi)$, e da taxa de cura, p_0 , em um processo destrutivo para o grupo controle.	56
4.16	Estimação da função de cura, $p_0(t; \theta, \Psi)$, e da taxa de cura, p_0 , em um processo destrutivo para o grupo tratamento.	57

Capítulo 1

Introdução

A palavra câncer é proveniente do latim e significa caranguejo. Esta denominação se refere às células afetadas no organismo de um indivíduo que se proliferam da mesma forma que as patas de um caranguejo se sobressaem de seu cefalotórax.

O termo câncer é utilizado para denominar doenças que se caracterizam pelo crescimento e divisão desordenados das células de um indivíduo (no caso do câncer maligno) se espalhando e destruindo tecidos e órgãos próximos, podendo atingir até outras regiões do corpo por um processo conhecido por metástase.

O desenvolvimento do câncer ocorre devido à anomalias no material genético do indivíduo. Estas anomalias podem surgir devido a diversos fatores como por exemplo, efeitos de radiação, agentes infecciosos, substâncias químicas, dentre outros fatores. Estes fatores responsáveis pelo agravamento ou sensibilização do organismo ao desenvolvimento de um câncer são chamados de carcinógenos ou agentes cancerígenos. Além destes fatores, as anomalias no material genético do indivíduo podem surgir também devido a erros na replicação do DNA (Ácido Desoxirribonucleico) ou podem ser decorrentes de herança genética, ou seja, o indivíduo já nasce com esta característica em suas células.

De acordo com INCA (2012), 80% a 90% dos cânceres estão associados a fatores ambientais (carcinógenos), como o cigarro, a exposição excessiva ao sol, etc. O tempo de exposição celular aos agentes cancerígenos e a intensidade com que esta exposição ocorre influenciam significativamente no aparecimento do câncer.

O processo de formação do câncer é denominado por carcinogênese e pode ser dividido em três fases, apresentadas a seguir (INCA, 2012).

- **1ª Fase - Iniciação**

Esta fase inicial do processo de carcinogênese caracteriza-se pela ação dos agentes cancerígenos nas células do indivíduo, que começam a modificar alguns dos genes, ou seja, as células passam a ficar geneticamente alteradas. Elas encontram-se iniciadas, isto

é, preparadas para receberem a ação dos próximos agentes. Nesta fase ainda não é possível detectar clinicamente um tumor.

- **2ª Fase - Promoção**

Nesta fase as células que já se encontram iniciadas passam a sofrer a ação dos agentes cancerígenos promotores transformando-as em malignas, de forma lenta e gradual. Para que esta transformação ocorra, a célula deve continuar em contato com o agente cancerígeno, do contrário o processo pode ser interrompido.

- **3ª Fase - Progressão**

Esta fase é caracterizada pela multiplicação descontrolada e irreversível das células que foram promovidas. Esta é a fase em que o câncer já se encontra instalado, e prossegue seu desenvolvimento até surgirem as primeiras manifestações clínicas da doença.

Considerando o aumento significativo no número de indivíduos que desenvolvem algum tipo de câncer, estudos relacionados a esta doença vêm crescendo nos últimos anos. Desta forma, experimentos são frequentemente realizados na prática médica cujo interesse consiste em identificar e testar agentes químicos que podem prevenir o câncer em humanos. Devido ao fato de nem sempre ser possível realizar testes em humanos, animais experimentais são amplamente utilizados com esta finalidade.

Os experimentos com animais são uma das bases científicas para desenvolver diversos tratamentos que possam ser utilizados futuramente em humanos, como por exemplo, na elaboração de novas drogas, novos métodos cirúrgicos e vacinas, entre outros. Entretanto, deve ser realizada uma pesquisa detalhada pra decidir qual espécie de animal é a mais indicada para determinada pesquisa, pois nem sempre é possível uma analogia entre os aspectos anatômicos e fisiológicos. Um problema muito grave que ocorreu no final dos anos 1950, decorrente desta falta de informação, foi numa pesquisa realizada para testar um medicamento conhecido por talidomida, usualmente utilizada como sedativo, anti-inflamatório e hipnótico. Os testes que haviam sido realizados somente em roedores não apresentaram problemas, pois estes metabolizavam a droga diferentemente dos humanos. Desta forma, o medicamento foi receitado para milhares de mulheres, inclusive grávidas, para combater enjôos matinais. Como consequência, muitas crianças nasceram com malformação congênita e o medicamento foi retirado do mercado para estudos mais detalhados. Posteriormente, novos testes foram realizados em outros animais, como primatas e coelhos, confirmando as malformações fetais dos membros.

Devido a este grande erro médico, atualmente existe um rigor muito maior na realização de testes com animais. Sendo assim, estes testes devem ser realizados primeiramente

em animais de pequeno porte, como roedores, passando posteriormente para animais de maior porte, como coelhos e cachorros, para então serem realizados em animais que são geneticamente mais parecidos com os homens, como os primatas.

Além disso, existem vários códigos de ética que devem ser rigidamente seguidos pelos pesquisadores, de modo a garantir o melhor tratamento possível aos animais, evitando que sejam maltratados e sintam dor. Para a realização de testes com animais no Brasil, é necessário que o projeto de pesquisa seja aprovado pelos comitês de ética. Estes comitês, que pertencem à própria instituição na qual a pesquisa é realizada, são os responsáveis pela aprovação e fiscalização dos projetos de pesquisa. Os resultados obtidos com testes em animais são utilizados com o propósito de produzir conhecimento científico útil aos seres humanos.

De todos os animais utilizados para testes laboratoriais, os ratos são os mais utilizados por terem uma fisiologia e um aspecto imunológico semelhantes aos do homem. Além disso, os ratos apresentam um período de gestação de apenas 21 dias, fazendo com que os resultados dos experimentos possam ser checados rapidamente. As fêmeas dão de três a seis ninhadas por ano, isso os torna extremamente vantajosos quando o desejo do cientista é testar o efeito dos medicamentos sobre os descendentes de quem irá consumi-los. Outro fator importante da utilização de ratos em experimentos laboratoriais é que eles podem ser manipulados mais facilmente, e os custos nestes estudos também são reduzidos, quando comparados aos estudos realizados com animais de maior porte.

São considerados neste trabalho os experimentos que consistem da indução química de tumores em animais experimentais com o propósito de verificar os efeitos quimiopreventivos de várias substâncias. Neste tipo de experimento, duas variáveis de interesse são o número de tumores induzidos e seus tempos de desenvolvimento. Espera-se que uma substância quimiopreventiva eficiente reduza o número de tumores e/ou desacelere sua taxa de crescimento. A Figura 1.1 exemplifica o funcionamento de testes experimentais realizados em ratos.

Considerar o processo de destruição das células afetadas pela exposição a um agente cancerígeno na modelagem deste tipo de problema é muito importante, pois fornece informações de extrema utilidade em experimentos de quimioprevenção de câncer, sendo uma ferramenta a mais na comparação de tratamentos. A teoria de Análise de Sobrevivência de Longa Duração é utilizada para obter estimativas da probabilidade de cura dos animais, auxiliando na verificação da eficiência dos tratamentos em estudo.

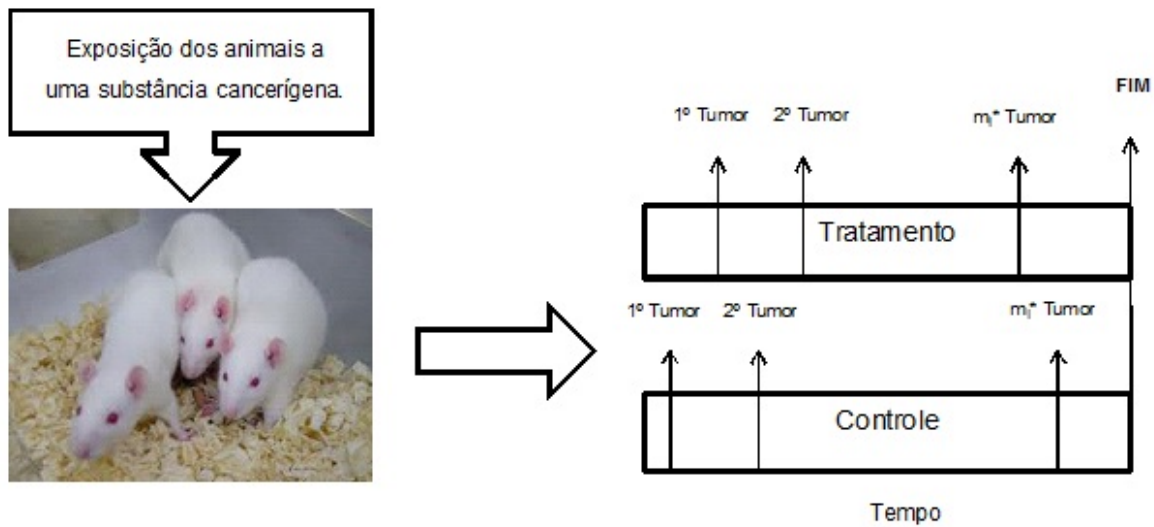


Figura 1.1: Esquema de experimentos com ratos.

1.1 Análise de Sobrevivência de Longa Duração

Estudos de tratamentos com substâncias químicas cujo objetivo é avaliar a destruição de células com câncer tem progredido muito nos últimos anos. O aperfeiçoamento destes tratamentos tornou-os cada vez mais eficientes no combate ao câncer e, desta forma, é de se esperar que uma parcela substancial dos indivíduos da população em estudo possa ser curada.

A teoria tradicional de Análise de Sobrevivência não contempla a possibilidade de cura dos indivíduos em estudo. Sendo assim, a abordagem de uma metodologia que leve em consideração a proporção de curados na população faz-se necessária. Na literatura, a área da estatística responsável pela modelagem de dados de sobrevivência com fração de cura é denominada por Análise de Sobrevivência de Longa Duração (ASLD) (Ibrahim et al., 2001) a qual vem sendo amplamente utilizada em estudos clínicos. Dentre os inúmeros trabalhos encontrados na literatura com modelos de fração de cura, ou modelos de longa duração, alguns resultados podem ser obtidos nos trabalhos de Maller & Zhou (1996), em Tsodikov *et al.* (2003), Cancho & Bolfarine (2001), Chen & Ibrahim (2001), Yakovlev & Tsodikov (1996), além da proposta da teoria unificada apresentada por Rodrigues *et al.* (2008).

A teoria de modelos de longa duração difere da teoria usual de sobrevivência pelo fato de incluir em sua modelagem a probabilidade de cura dos indivíduos. Neste caso, a função de sobrevivência não tende mais a zero, e sim para uma probabilidade p_0 , também conhecida por fração de cura. Um dos principais interesses desta modelagem é estimar esta fração de cura, a fim de avaliar com mais precisão a eficiência dos tratamentos em

estudo.

Com o intuito de ilustrar uma situação em que a probabilidade dos indivíduos sobreviverem além de um tempo t converge para um certo valor, p_0 , apresentaremos um exemplo de um ensaio clínico aleatorizado em que os dados se referem a 64 pacientes com anemia aplástica grave. Os pacientes em estudo foram atribuídos a dois tratamentos, um com ciclosporina e metotrexate (CSP+MTX), e outro somente com metotrexate (MTX) (vide Kalbfleisch & Prentice, 2002). As estimativas de Kaplan-Meier para as funções de sobrevivência dos dois grupos de tratamentos são apresentadas na Figura 1.2.

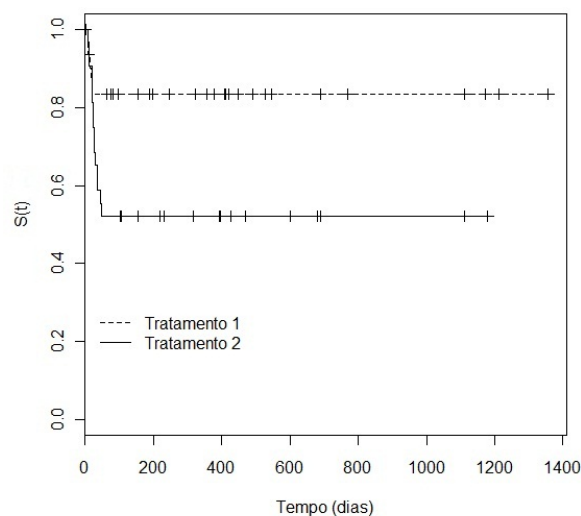


Figura 1.2: *Estimativas de Kaplan-Meier para $S(t)$.*

Realizando ligeiramente uma análise gráfica, pode-se observar que as estimativas das funções de sobrevivência dos tempos de ambos os tratamentos não convergem para zero, mas sim para uma proporção de indivíduos que não apresentou o evento de interesse, podendo ter morrido de outras causas que não a estudada, o que caracteriza uma possível cura destes indivíduos em relação à doença avaliada. Esta é uma situação em que a abordagem de modelos de longa duração faz-se necessária. Observa-se também que os indivíduos tratados por ciclosporina e metotrexate (Tratamento 1), apresentaram maior proporção de cura quando comparados aos indivíduos tratados somente com metotrexate (Tratamento 2), levando-nos a crer que a presença da substância ciclosporina aparentemente aumenta a chance de cura dos pacientes em tratamento.

No contexto de dados de sobrevivência, a variável de interesse é, geralmente, o tempo até a ocorrência de um evento de interesse (que pode ser a morte de um paciente, a recidiva de uma doença, a falha de um equipamento eletrônico, etc.). Entretanto, num cenário mais específico considera-se o processo de iniciação e promoção de tumores nas

células de indivíduos, devido a alguma exposição às substâncias cancerígenas, como alvo do estudo.

O interesse de estudos de sobrevivência é frequentemente avaliar a eficácia de tratamentos químicos na destruição de células com câncer. Com este objetivo, experimentos que envolvem a indução química de substâncias cancerígenas em indivíduos são cada vez mais frequentes em pesquisas biológicas.

Em geral o objetivo de tais experimentos é investigar o efeito de um determinado tratamento nas taxas de incidência de câncer. Estes experimentos são muito utilizados para testar os efeitos quimiopreventivos de várias substâncias, sendo que para comparar tratamentos são utilizadas como variáveis resposta, o número de tumores induzidos e o tempo que eles se desenvolvem. Na realização de experimentos deste tipo, espera-se que uma substância quimiopreventiva eficiente reduza o número de tumores e/ou desacelere sua taxa de crescimento. É comum neste tipo de estudo a presença de censuras, isto ocorre porque o experimento é finalizado antes que todos os tumores induzidos sejam detectados. Além disso, avaliar o processo de destruição das células afetadas pela exposição a um agente cancerígeno faz que o modelo utilizado seja uma alternativa mais real do que os modelos que não levam em conta esta suposição.

Na seção 1.2 é apresentado o conceito de modelo destrutivo, apresentando algumas definições que são utilizadas neste trabalho. Os objetivos propostos para este trabalho são apresentados na seção 1.3. As contribuições relevantes são apresentadas na seção 1.4, enquanto que o projeto de organização do trabalho é exposto na seção 1.5.

1.2 O Modelo Destrutivo

Os modelos destrutivos de sobrevivência com fração de cura foram propostos com o objetivo de representar, de uma maneira mais realista, dados de sobrevivência de longa duração. Muitos são os problemas da área de análise de sobrevivência que envolvem o estudo da eficiência de tratamentos no combate ao câncer e, desta forma, é de extrema relevância que este processo destrutivo (de destruição das células iniciadas devido a alguma exposição a substâncias cancerígenas) seja também considerado na análise dos dados.

Os modelos de sobrevivência que contemplam este processo destrutivo estão sendo cada vez mais utilizados, principalmente em estudos biomédicos, cujos problemas objetivam-se a analisar o processo de destruição ou reparação das células lesionadas após o indivíduo ser submetido à algum tratamento prolongado. Espera-se, caso o tratamento seja eficiente, que mais células cancerígenas sejam destruídas ao longo do tratamento.

Considere uma variável aleatória não observável M , a qual representa o número de causas associadas à ocorrência de um evento de interesse no início do experimento. A

função de probabilidade (f.p.) da variável M com parâmetro θ é então definida por $p(m; \theta) = P(M = m; \theta)$, para $m = 0, 1, 2, \dots$. Desta forma, dado um valor $M = m$, considere $X_j, j = 1, 2, \dots, m$ variáveis aleatórias independentes com distribuição Bernoulli e probabilidade de sucesso π , independentemente de M , ou seja,

$$X_j = \begin{cases} 1, & \text{se a } j\text{-ésima célula foi lesionada;} \\ 0, & \text{caso contrário,} \end{cases} \quad (1.1)$$

para $j = 1, \dots, m$.

Considerando então uma variável aleatória M^* que representa o número total de causas que não foram eliminadas (ou destruídas) durante o tratamento dentre as M causas iniciais, tem-se que esta variável M^* pode então ser definida como

$$M^* = \begin{cases} X_1 + X_2 + \dots + X_M, & \text{se } M > 0; \\ 0, & \text{se } M = 0. \end{cases} \quad (1.2)$$

Quando $M^* = M$ pode-se concluir que não houve destruição de células lesionadas, caracterizando uma ineficiência do tratamento avaliado. Portanto espera-se que $M^* < M$, para caracterizar o processo de destruição das células.

A distribuição conhecida por danificada ou destrutiva refere-se à distribuição condicional da variável M^* dado $M = m$. Assim, utilizando as propriedades de probabilidade condicional, tem-se que

$$P[M^* = m^*, M = m; \theta, \pi] = P[M^* = m^* | M = m; \pi] p(m; \theta). \quad (1.3)$$

A f.p. da variável M^* pode ser obtida utilizando-se a equação (1.3) e as propriedades de probabilidades condicionais, resultando na seguinte expressão:

$$\begin{aligned} P[M^* = m^*; \theta, \pi] &= \sum_{m=m^*}^{\infty} P[M^* = m^*, M = m; \theta, \pi] \\ &= \sum_{m=m^*}^{\infty} \underbrace{P[M^* = m^* | M = m; \pi]}_{\text{Distribuição destrutiva}} p(m; \theta). \end{aligned} \quad (1.4)$$

Utilizando uma distribuição binomial com parâmetros m e π para a distribuição destrutiva, $P[M^* = m^* | M = m; \pi]$, obtém-se a seguinte expressão para a distribuição de M^* :

$$P[M^* = m^*; \theta, \pi] = \sum_{m=m^*}^{\infty} \binom{m}{m^*} \pi^{m^*} (1 - \pi)^{m-m^*} p(m; \theta). \quad (1.5)$$

Esta distribuição está detalhada na seção 3.1 conjuntamente com a distribuição COM-Poisson para $p(m; \theta)$.

1.3 Objetivos

O principal objetivo deste trabalho consiste em avaliar o comportamento da distribuição COM-Poisson na análise de dados de experimentos de quimioprevenção do câncer em animais, comparando seu ajuste a outras distribuições tradicionais utilizadas frequentemente, como a Poisson e a Binomial Negativa, a fim de verificar sua eficiência. Os objetivos específicos são descritos a seguir.

1.3.1 Objetivos Específicos

- Formular os resultados de experimentos de quimioprevenção de câncer em animais no contexto de modelo destrutivo de longa duração, considerando que o número de tumores induzidos depende de um instante t , tornando a análise dinâmica. Além disso, verificamos que o número de tumores observados e seus tempos de detecção são variáveis observadas, diferente da proposta de Rodrigues *et al.* (2011), em que estas variáveis são latentes;
- Realizar um estudo de simulação de experimentos quimiopreventivos de câncer em animais, visando avaliar o comportamento da distribuição COM-Poisson no ajuste dos parâmetros do modelo proposto;
- Estimar os parâmetros dos modelos considerados na análise para um conjunto de dados reais retirado do artigo do Freedman (1993), considerando, além das distribuições Poisson e Binomial Negativa, a distribuição COM-Poisson;
- Comparar os resultados obtidos para os diferentes modelos paramétricos considerados, a fim de verificar qual deles melhor explica o número de tumores promovidos por animal;
- Estimar as probabilidades de cura dos tratamentos em estudo como funções paramétricas dos parâmetros dos modelos propostos.

1.4 Contribuições

O interesse em avaliar o desempenho de tratamentos quimioterápicos na destruição de células que sofreram alguma mutação devido às exposições químicas vem crescendo

significativamente nos últimos anos. A ciência avança cada vez mais em busca de novas descobertas de curas e tratamentos para enfermidades que vem assolando a população mundial, sendo o câncer uma delas, ganhando lugar de destaque na priorização da obtenção de formas de tratamentos cada vez mais eficazes. Desta forma, experimentos cujo interesse é avaliar a eficiência de tratamentos na destruição de células cancerígenas, são muito comuns na área médica. Além das vantagens relacionadas à saúde, outras contribuições importantes deste trabalho em estatística são as seguintes:

- Avaliar e desenvolver novas ferramentas estatísticas de análise de dados, adaptando resultados teóricos aos problemas frequentes da realidade;
- Apresentar o uso da distribuição COM-Poisson para experimentos com animais, como alternativa eficaz de análise, devido à sua grande flexibilidade para acomodar problemas como a subdispersão e a sobredispersão, problemas que são frequentemente encontrados em dados discretos.

1.5 Organização do Trabalho

O trabalho apresentado é composto por cinco capítulos. No capítulo 1 são apresentados alguns conceitos introdutórios, dando ênfase à metodologia de modelos destrutivos, suas principais definições e aplicações.

Os conceitos de processo destrutivo em dois estágios são apresentados no capítulo 2, no qual são introduzidos os conceitos básicos de iniciação e promoção de tumores em células que sofreram alguma exposição cancerígena, além de obter as estimativas das funções de sobrevivência mínimas de longa duração para as distribuições de Poisson e Binomial Negativa, considerando os máximos dos tempos de detecção dos tumores. As estimativas para as probabilidades de cura da população em estudo são também definidas neste capítulo.

O capítulo 3 refere-se à distribuição COM-Poisson, suas principais definições, aplicações e comparações com algumas distribuições muito difundidas na prática estatística de análise de dados discretos, como por exemplo, as distribuições Poisson e Binomial Negativa.

As aplicações da metodologia apresentada durante todo o trabalho dão-se no capítulo 4, no qual se enfatiza o modelo proposto por Kokoska (1987), observando-o do ponto de vista de modelo destrutivo. Neste capítulo também é realizado um estudo de simulação e analisa-se um conjunto de dados reais a fim de avaliar a eficiência do modelo proposto.

As principais conclusões e considerações finais retiradas deste estudo, além de algumas propostas para trabalhos posteriores são apresentadas no capítulo 5. Realiza-se uma breve

comparação entre o modelo proposto e os modelos tradicionais da literatura usados com frequência na análise deste tipo de dados.

Capítulo 2

Processo Destrutivo em Dois Estágios

Em problemas de experimentação com animais, duas variáveis de interesse são o número de tumores desenvolvidos e seus tempos de detecção até certo instante t . O interesse destes experimentos consiste, geralmente, em investigar o efeito de tratamentos na destruição de células que foram afetadas por alguma substância cancerígena, além da comparação de tratamentos, visando destacar o mais eficiente no combate da doença.

Na seção 1.2 foram apresentados alguns conceitos básicos de modelo destrutivo e como obter a função de probabilidade de uma variável aleatória M^* que representa este processo na análise de dados. A idéia de destruição utilizada neste trabalho refere-se ao fato de que após induzir um determinado número de tumores nos animais experimentais no início do estudo e designá-los aleatoriamente à seus respectivos tratamentos, o número de tumores que foram observados após a realização do experimento é caracterizado pela variável aleatória M^* . Ou seja, a variável M^* representa, neste caso de experimentos com animais, o número de tumores, dentre os tumores induzidos, que não foram eliminados ou destruídos pelo tratamento avaliado e se desenvolveram até atingirem um tamanho detectável.

Considera-se nesta abordagem a análise de experimentos com animais utilizando os conceitos de modelo destrutivo proposto por Rodrigues *et al.* (2011), apresentados na seção 1.2. Tal proposta torna o modelo mais dinâmico e realista quando comparado à proposta de Rodrigues *et al.* (2011), pois supõe que a variável número de tumores observados depende de um instante t . Dessa forma, a probabilidade de detectar um tumor até um determinado instante t não será considerada constante, como na proposta de Rodrigues, mas será considerada uma função que depende do tempo de observação.

Em experimentos com animais, a função de probabilidade da variável M^* é utilizada para a obtenção da função de verossimilhança, como foi apresentado por Kokoska (1987) e Freedman (1993). Entretanto, estes autores não observam estes experimentos do ponto de vista de um processo destrutivo, como será aqui abordado. Além disso, considera-se o uso da distribuição COM-Poisson para o número de tumores por animal, visando compará-la

com os resultados obtidos para as distribuições de Poisson e Binomial Negativa abordadas no artigo de Freedman, tornando a análise mais flexível.

Especificamente, quando o interesse é investigar o efeito de tratamentos nas taxas de incidência de câncer em seres humanos, é frequente a realização de testes em animais experimentais, os quais são submetidos à exposição de alguma radiação ou substância química e, então, são acompanhados por um tempo pré-determinado para avaliar o desenvolvimento de seus tumores.

Quando o interesse está na comparação de tratamentos, durante o acompanhamento, os animais são aleatorizados em grupos que receberão tratamentos distintos e, assim, verifica-se qual dos grupos de tratamentos aplacou com mais rigor os tumores iniciados pelo agente cancerígeno.

Estes experimentos são caracterizados por dois estágios principais que consistem em induzir quimicamente substâncias cancerígenas nos animais, iniciando células susceptíveis ao desenvolvimento de tumores e, posteriormente, acompanhá-los para verificar o desenvolvimento destes tumores, os quais são detectados por apalpação. O protocolo usual para estes experimentos consiste em examinar os tumores dos animais em intervalos frequentes, iniciando em duas a três semanas após a exposição ao agente cancerígeno. A dose de carcinogênese frequentemente utilizada é suficiente para induzir de cinco a sete tumores por animal (Kokoska, 1987).

Diferente da proposta de Rodrigues *et al.* (2008) da teoria unificada de longa duração, supõe-se neste trabalho que o número de tumores desenvolvidos por animal ao longo do experimento e seus tempos de detecção, são variáveis observáveis. A seguir são caracterizados os dois principais estágios de um processo de experimentação com animais.

2.1 Longa Duração em Dois Estágios

Pode-se estabelecer que num processo experimental de quimioprevenção de câncer em animais o evento de interesse pode ser caracterizado por dois principais estágios, definidos por:

- **1º Estágio: Iniciação**

Considera que o agente cancerígeno inicia M tumores no animal. Cada um destes tumores pode ou se desenvolver até um tamanho que seja detectável por apalpação (contribuindo para a ocorrência do evento de interesse), ou ser destruído pelo tratamento utilizado. Assume-se que esta variável aleatória M é representada por uma função de probabilidade com parâmetro θ , $p(m; \theta)$, ou seja,

$$p(m; \theta) = P(M = m; \theta), \quad m = 0, 1, 2, \dots$$

A situação em que $M = 0$ caracteriza que o agente cancerígeno não iniciou tumores no animal, representando a situação de que até o fim do estudo este animal não desenvolverá tumores. Para esta situação considera-se que o animal foi curado, e sua probabilidade será denotada por p_0 , também chamada de fração de cura, ou seja,

$$p_0 = p(0; \theta) = P(M = 0; \theta).$$

• 2º Estágio: Maturação

Considera que os tumores crescem independentemente (o que não representa estritamente uma verdade biológica, mas esta suposição é usualmente utilizada por simplicidade técnica). Dentre os M tumores iniciados em cada animal, o que se observa são $M^* = m^*$ tumores por animal, cujos tempos de detecção são dados por T_1, T_2, \dots, T_{m^*} , que são variáveis aleatórias independentes e identicamente distribuídas com função de distribuição acumulada com parâmetro Ψ , $F(t; \Psi)$. Após realizar o experimento, identifica-se a ordem em que os tumores foram detectados e, desta forma, os tempos de detecção dos tumores são observados como estatísticas de ordem, ou seja, são dados respectivamente por $t_{(1)}, t_{(2)}, \dots, t_{(m^*)}$, em que $t_{(1)} \leq t_{(2)} \leq \dots \leq t_{(m^*)}$.

Pode-se observar que o animal não apresenta o evento de interesse (morte) até o tempo do último tumor detectado, ou seja, ele passa a estar em situação de risco a partir do tempo $T_{(m^*)}$. Denotando o tempo de vida ou de censura dos animais por T^* , cujo tempo observado será representado por $t_i^*, i = 1, 2, \dots, n$ (sendo n o número de animais do grupo em estudo), observa-se que o tempo de vida dos animais é maior ou igual ao tempo de detecção do último tumor observado, ou seja, $T^* \geq T_{(m^*)}$.

A probabilidade de o animal sobreviver além do tempo $T_{(m^*)}$ será chamada de função de sobrevivência mínima e é apresentada mais detalhadamente a seguir.

2.1.1 Função de Sobrevivência Mínima

A função de sobrevivência mínima considera a situação em que o animal ainda estava vivo até o tempo de detecção do último tumor observado $T_{(m^*)}$ e, desta forma, o interesse consiste em calcular a probabilidade de o animal sobreviver além deste tempo. Entretanto, como este tempo $T_{(m^*)}$ não representa necessariamente o tempo de morte do animal, a probabilidade de o animal sobreviver além deste tempo não caracteriza uma função de sobrevivência de longa duração, pois todos os animais que não morreram durante o experimento são sacrificados no fim do estudo. Assim, denotamos a probabilidade de o animal sobreviver além do tempo $T_{(m^*)}$ como uma função de sobrevivência mínima, $S_{min}(t; \theta, \Psi)$, que utiliza como variável de interesse os tempos máximos de detecção dos tumores.

Dado $M = m$ tumores induzidos, sejam $X_j(t), j = 1, \dots, m$ variáveis aleatórias independentes seguindo uma distribuição de Bernoulli com probabilidade de sucesso $\pi(t; \Psi) = P[X_j(t) = 1; \Psi] = P[T_j \leq t; \Psi] = F(t; \Psi)$, indicando que o j -ésimo tumor induzido no início do experimento se desenvolveu até atingir um tamanho detectável por apalpamento no intervalo de tempo $(0, t]$, ou seja,

$$X_j(t) = \begin{cases} 1, & \text{se } T_j \leq t; \\ 0, & \text{se } T_j > t. \end{cases} \quad (2.1)$$

Desta forma, podemos definir a variável aleatória $M^*(t)$, a qual representa o número de tumores observados dentre os M tumores iniciados por um agente cancerígeno, que não foram destruídos pelo tratamento no intervalo de tempo $(0, t]$, da seguinte forma:

$$M^*(t) = \begin{cases} X_1(t) + X_2(t) + \dots + X_M(t), & \text{se } M > 0; \\ 0, & \text{se } M = 0. \end{cases} \quad (2.2)$$

A situação em que $M^*(t) \leq M$ caracteriza o processo de destruição dos tumores induzidos no início do experimento. Quando $M^*(t) = 0$ significa que não foram detectados tumores malignos no animal até o tempo t . Esta dependência das variáveis de Bernoulli $X_j(t)$ e, conseqüentemente, da variável $M^*(t)$ com o instante t , torna a análise dinâmica e realista, sendo extremamente vantajosa em relação à outras abordagens de modelos destrutivos, pois permite calcular a probabilidade de observar um determinado número de tumores para qualquer que seja o instante de tempo $t > 0$. Esta é uma das diferenças fundamentais entre as definições tradicionais de modelo destrutivo apresentados na seção 1.2 e a abordagem que estamos considerando para análise de experimentos com animais.

Por simplicidade de notação, a probabilidade de observar m^* tumores num animal no intervalo de tempo $(0, t]$ apresentada na equação 1.4, será denotada por $p(m^*; \theta, \Psi)$.

A distribuição de probabilidade da variável $M^*(t)$, $p(m^*; \theta, \Psi)$, será abordada mais detalhadamente na seção 2.2, considerando as distribuições Poisson e Binomial Negativa para a variável número de tumores induzidos por animal, M , e na seção 3.1, considerando a distribuição COM-Poisson para M .

Caso o experimento se prolongue por um tempo muito grande, ($t \rightarrow \infty$), a proporção de animais que não apresentaram tumores é conhecida como a proporção de animais que possivelmente foram curados pelo tratamento. Assim, a fração de cura pode ser calculada da seguinte forma:

$$\lim_{t \rightarrow \infty} P[M^*(t) = 0; \theta, \Psi] = \lim_{t \rightarrow \infty} p_0(t; \theta, \Psi) = P(M = 0; \theta) = p_0, \quad (2.3)$$

em que $p_0(t; \theta, \Psi) = P[M^*(t) = 0; \theta, \Psi]$. Ou seja, a longo prazo a probabilidade de não observar tumores malignos detectáveis por animal é chamada de probabilidade de cura ou fração de cura do tratamento, p_0 .

O resultado anterior segue diretamente das equações (2.1) e (2.2), pois quando $t \rightarrow \infty$ observa-se que $P[X_j(t) = 1; \Psi] = F(t; \Psi) = 1$, para $j = 1, \dots, M$, seguindo da equação (2.2) que $P[M^*(t) = m^*; \theta, \Psi] = P[M = m; \theta]$.

A função $p_0(t; \theta, \Psi)$ é uma novidade neste trabalho para análise de experimentos quimiopreventivos em animais, pois através dela pode-se calcular uma probabilidade de cura dinâmica, ou seja, para qualquer instante de tempo t é possível obter a probabilidade de o animal não desenvolver nenhum tumor detectável até este instante. Tomando t relativamente grande ($t \rightarrow \infty$), esta probabilidade se aproxima da fração de cura p_0 . Fazendo t próximo de zero, tem-se que esta probabilidade $p_0(t; \theta, \Psi)$ fica próxima de 1.

A variável de interesse utilizada na obtenção da função de sobrevivência mínima, representada por $T_{(M^*(t))}$, é dada por

$$T_{(M^*(t))} = \max\{T_1, T_2, \dots, T_{M^*(t)}\}, \quad (2.4)$$

para $M^*(t) \geq 1$ e $T_{(M^*(t))} = \infty$ se $M^*(t) = 0$, conduzindo a uma proporção p_0 dos indivíduos não suscetíveis à ocorrência do evento de interesse, ou seja, a proporção de indivíduos que foram curados.

O tempo de vida dos animais será representado pela variável T^* , em que $T^* \geq T_{(M^*(t))}$ e, portanto, para $t > 0$ pode-se relacionar a função de sobrevivência de longa duração, $S_p(t; \theta, \Psi)$, com a função de sobrevivência mínima, $S_{min}(t; \theta, \Psi)$, da seguinte forma:

$$S_p(t; \theta, \Psi) = P[T^* \geq t; \theta, \Psi] \geq P[T_{(M^*(t))} \geq t; \theta, \Psi] = S_{min}(t; \theta, \Psi). \quad (2.5)$$

Portanto, tem-se que $S_p(t; \theta, \Psi) \geq S_{min}(t; \theta, \Psi)$. Na situação em que o animal morre no instante de detecção do último tumor observado, ou seja, quando $T^* = T_{(M^*(t))}$, tem-se o caso particular em que $S_p(t; \theta, \Psi) = S_{min}(t; \theta, \Psi)$.

Com o propósito de obter a função de sobrevivência mínima, $S_{min}(t; \theta, \Psi)$, será definida a seguir a função geradora de probabilidades.

Definição 2.1.1: Seja $a = \{a_{m^*}\}$ uma sequência de números reais. Se

$$\begin{aligned} A_a(s) &= a_0 + a_1 s + a_2 s^2 + \dots \\ &= \sum_{m^*} s^{m^*} a_{m^*} \end{aligned} \quad (2.6)$$

converge para valores de s no intervalo $[0, 1]$, então $A_a(s)$ é chamada de função geradora

da sequência a (Feller, 1968).

Utilizando a Definição 2.1.1 observa-se que, se $a = \{a_{m^*}\} = \{p(m^*; \theta, \Psi)\}$, a função geradora de probabilidades (f.g.p.) pode ser expressa como um valor esperado, ou seja, $A_p(s) = \sum_{m^*} s^{m^*} p(m^*; \theta, \Psi) = E[s^{M^*(t)}]$. A partir destas definições pode-se enunciar a função de sobrevivência mínima no teorema a seguir.

Teorema 2.1.1. *Para a variável aleatória $T_{(M^*(t))}$ definida na equação (2.4) num modelo de longa duração em dois estágios, supondo independência entre o número de tumores induzidos, M , e as variáveis $X_j(t)$, a função de sobrevivência mínima é dada por*

$$\begin{aligned} S_{min}(t; \theta, \Psi) &= 1 + P[M^*(t) = 0; \theta, \Psi] - A_{M^*(t)}(F(t; \Psi)) \\ &= 1 + p_0(t; \theta, \Psi) - A_{M^*(t)}(F(t; \Psi)), \end{aligned} \quad (2.7)$$

em que $A_{M^*(t)}(F(t; \Psi))$ representa a função geradora de probabilidades da variável aleatória $M^*(t)$ calculada no valor $F(t; \Psi)$.

Prova:

Utilizando a variável $T_{(M^*)}$ apresentada na equação 2.4, a função de sobrevivência mínima é calculada por

$$\begin{aligned} S_{min}(t; \theta, \Psi) &= P[T_{(M^*(t))} \geq t; \theta, \Psi] = P[\max\{T_1, T_2, \dots, T_{M^*(t)}\} \geq t; \theta, \Psi] \\ &= 1 - P[\max\{T_1, T_2, \dots, T_{M^*(t)}\} < t; \theta, \Psi] \end{aligned}$$

Pela Regra da Probabilidade Total e pelas definições de Probabilidade Condicional, tem-se

$$\begin{aligned} S_{min}(t; \theta, \Psi) &= 1 - \sum_{m^*=1}^{\infty} P[\max\{T_1, T_2, \dots, T_{M^*(t)}\} < t, M^*(t) = m^*; \theta, \Psi] \\ &= 1 - \sum_{m^*=1}^{\infty} P[M^*(t) = m^*; \theta, \Psi] P[\max\{T_1, T_2, \dots, T_{M^*(t)}\} < t | M^*(t) = m^*; \theta, \Psi] \\ &= 1 - \sum_{m^*=1}^{\infty} p(m^*; \theta, \Psi) P[\max\{T_1, T_2, \dots, T_{m^*}\} < t; \theta, \Psi] \\ &= 1 - \sum_{m^*=1}^{\infty} p(m^*; \theta, \Psi) P[T_1 < t, T_2 < t, \dots, T_{m^*} < t; \theta, \Psi] + p_0(t; \theta, \Psi) - p_0(t; \theta, \Psi) \\ &= 1 + p_0(t; \theta, \Psi) - \left\{ p_0(t; \theta, \Psi) + \sum_{m^*=1}^{\infty} p(m^*; \theta, \Psi) F(t; \Psi)^{m^*} \right\} \\ &= 1 + p_0(t; \theta, \Psi) - A_{M^*(t)}(F(t; \Psi)), \end{aligned} \quad (2.8)$$

em que $p_0(t; \theta, \Psi) = P[M^*(t) = 0; \theta, \Psi]$ e $A_{M^*(t)}(F(t; \Psi))$ é a função geradora de probabilidades apresentada na Definição 2.1.1.

Para a variável aleatória M , pode-se supor qualquer distribuição conhecida, e desta forma sua f.g.p. é facilmente obtida. Entretanto, a variável $M^*(t)$ por ser composta pode apresentar um comportamento mais complexo, dificultando a obtenção de sua f.g.p.. Desta forma, com o intuito de facilitar os cálculos para a função de sobrevivência mínima e sabendo que $M^*(t)$ é uma soma de M variáveis independentes com distribuição de Bernoulli, pode-se utilizar o seguinte teorema (Feller, 1968):

Teorema 2.1.2. *Seja a variável $X_j(t)$ definida na equação (2.1) com f.g.p. $A_X(s)$ e considerando a variável M com função de probabilidade $p(m; \theta)$ e f.g.p. $A_M(s)$. Então, a função geradora de probabilidades da variável composta $M^*(t)$ definida na equação (2.2) é dada por*

$$\begin{aligned} A_{M^*(t)}(s) &= A_M(A_X(s)) \\ &= A_M[1 - F(t; \Psi)(1 - s)]. \end{aligned} \quad (2.9)$$

Prova:

Considere as variáveis aleatórias $X_j(t)$ definidas na equação (2.1), tem-se que sua função de probabilidade é dada por

$$P[X_j(t) = k; \Psi] = F(t; \Psi)^k [1 - F(t; \Psi)]^{1-k}, \quad k = 0, 1. \quad (2.10)$$

Assim, sua função geradora de probabilidades pode ser obtida por

$$\begin{aligned} A_X(s) &= E(s^{X_j(t)}) = \sum_k s^k P[X_j(t) = k; \Psi] \\ &= s^0 P[X_j(t) = 0; \Psi] + s^1 P[X_j(t) = 1; \Psi] \\ &= 1 - F(t; \Psi) + sF(t; \Psi) = 1 - F(t; \Psi)(1 - s). \end{aligned} \quad (2.11)$$

Considere também a variável M cuja função de probabilidade é $p(m; \theta)$ e a f.g.p é $A_M(s) = E(s^M) = \sum_m s^m p(m; \theta)$. Desta forma, utilizando as propriedades de probabilidade condicional, pode-se escrever a f.g.p da variável $M^*(t)$ como

$$\begin{aligned} A_{M^*(t)}(s) &= E(s^{M^*(t)}) \\ &= E\{E(s^{M^*(t)} | M)\}. \end{aligned} \quad (2.12)$$

Utilizando a distribuição destrutiva definida na equação (1.5) da seção 1.2, substi-

tuindo π por $\pi(t; \Psi) = P[X_j(t) = 1; \Psi] = F(t; \Psi)$ e, conseqüentemente, $1 - \pi$ por $1 - \pi(t; \Psi) = P[X_j(t) = 0; \Psi] = 1 - F(t; \Psi)$, tem-se que

$$\begin{aligned} E(s^{M^*(t)}|M = m) &= \sum_{m^*} s^{m^*} P(M^*(t) = m^*|M = m; \Psi) \\ &= \sum_{m^*} s^{m^*} \binom{m}{m^*} F(t; \Psi)^{m^*} (1 - F(t; \Psi))^{m-m^*} \\ &= \sum_{m^*} \binom{m}{m^*} [sF(t; \Psi)]^{m^*} [1 - F(t; \Psi)]^{m-m^*} \\ &= [1 - F(t; \Psi) + sF(t; \Psi)]^m \\ &= [1 - F(t; \Psi)(1 - s)]^m, \quad s \in [0, 1]. \end{aligned}$$

Desta forma, $E(s^{M^*(t)}|M) = [1 - F(t; \Psi)(1 - s)]^M$ e, portanto, a f.g.p da variável $M^*(t)$ fica

$$\begin{aligned} A_{M^*(t)}(s) &= E\{E(s^{M^*(t)}|M)\} \\ &= E\{[1 - F(t; \Psi)(1 - s)]^M\} \\ &= A_M[1 - F(t; \Psi)(1 - s)] = A_M(A_X(s)). \end{aligned}$$

Calculando a função $A_X(s)$ definida na equação (2.11), no ponto $s = F(t; \Psi)$, tem-se que

$$A_X(F(t; \Psi)) = 1 - F(t; \Psi)(1 - F(t; \Psi)) = 1 - F(t; \Psi)S(t; \Psi). \quad (2.13)$$

Portanto, utilizando o resultado apresentado na equação (2.9), tem-se que

$$\begin{aligned} A_{M^*(t)}(F(t; \Psi)) &= A_M(A_X(F(t; \Psi))) \\ &= A_M[1 - F(t; \Psi)S(t; \Psi)]. \end{aligned} \quad (2.14)$$

Assim, substituindo o resultado da equação (2.14) na equação (2.8), tem-se outra expressão para a função de sobrevivência mínima, dada por

$$S_{min}(t; \theta, \Psi) = 1 + p_0(t; \theta, \Psi) - A_M[1 - F(t; \Psi)S(t; \Psi)]. \quad (2.15)$$

Após calcular a função $S_{min}(t; \theta, \Psi)$ para um processo em dois estágios, as funções de densidade e de risco impróprias mínimas são facilmente obtidas. A função de densidade

mínima de longa duração é dada por

$$\begin{aligned} f_{min}(t; \theta, \Psi) &= -\frac{d}{dt}S_{min}(t; \theta, \Psi) = -\frac{d}{dt}[1 + p_0(t; \theta, \Psi) - A_M(1 - F(t; \Psi)S(t; \Psi))] \\ &= \frac{d}{dt}A_M(1 - F(t; \Psi)S(t; \Psi)) - \frac{d}{dt}p_0(t; \theta, \Psi). \end{aligned}$$

Conseqüentemente, a função de risco mínima é obtida calculando

$$h_{min}(t; \theta, \Psi) = \frac{f_{min}(t; \theta, \Psi)}{S_{min}(t; \theta, \Psi)}.$$

Note que $p_0(t; \theta, \Psi)$ é a probabilidade de o animal não apresentar tumores detectáveis até o instante t . Para calcular a fração de cura de longa duração na população, p_0 , basta tomar o limite de $S_{min}(t; \theta, \Psi)$, quando $t \rightarrow \infty$, ou seja,

$$\begin{aligned} \lim_{t \rightarrow \infty} S_{min}(t; \theta, \Psi) &= \lim_{t \rightarrow \infty} [1 + p_0(t; \theta, \Psi) - A_M(1 - F(t; \Psi)S(t; \Psi))] \\ &= \lim_{t \rightarrow \infty} [1 + p_0(t; \theta, \Psi) - A_M(1)] \\ &= \lim_{t \rightarrow \infty} p_0(t; \theta, \Psi) = p_0. \end{aligned}$$

Desta forma, observa-se que embora a função de sobrevivência de longa duração seja maior que a função de sobrevivência mínima, a longo prazo as duas funções de sobrevivência convergem para a mesma fração de cura, p_0 , ou seja,

$$\lim_{t \rightarrow \infty} S_p(t; \theta, \Psi) = \lim_{t \rightarrow \infty} S_{min}(t; \theta, \Psi) = p_0. \quad (2.16)$$

2.2 Função de Probabilidade de $M^*(t)$

Para o problema de experimentos com animais o interesse é considerar, na análise de dados, o processo de destruição dos tumores induzidos por algum tratamento aplicado aos animais durante o estudo. Desta forma, utilizando os resultados da seção 1.2, considerando que θ representa o vetor de parâmetros da distribuição da variável aleatória M , Ψ representa o vetor de parâmetros da distribuição da variável aleatória T e utilizando a probabilidade de sucesso $F(t; \Psi)$ para as variáveis de Bernoulli $X_j(t), j = 1, \dots, m$, a distribuição da variável $M^*(t)$ fica caracterizada por

$$P[M^*(t) = m^*; \theta, \Psi] = \sum_{m=m^*}^{\infty} P[M^*(t) = m^* | M = m; \Psi] p(m; \theta).$$

Supondo uma distribuição binomial com parâmetros m e $F(t; \Psi)$ para a distribuição

destrutiva, $P[M^*(t) = m^* | M = m; \Psi]$, obtém-se a seguinte expressão para a distribuição do número de tumores detectados até um instante t , dentre os M tumores iniciados:

$$P[M^*(t) = m^*; \theta, \Psi] = \sum_{m=m^*}^{\infty} \binom{m}{m^*} [F(t; \Psi)]^{m^*} [1 - F(t; \Psi)]^{m-m^*} p(m; \theta). \quad (2.17)$$

Do resultado acima segue que

$$M^*(t) | M = m; \Psi \sim \text{Binomial}(m, F(t; \Psi)), \quad (2.18)$$

Desta forma, a média e a variância de $M^*(t) | M = m; \Psi$ é dada por

$$E[M^*(t) | M; \Psi] = MF(t; \Psi) \quad e \quad \text{Var}[M^*(t) | M; \Psi] = MF(t; \Psi)(1 - F(t; \Psi)). \quad (2.19)$$

Portanto, pode-se obter também a média e a variância da variável $M^*(t)$, utilizando as propriedades de esperança e variância condicionais, ou seja,

$$E[M^*(t)] = E\{E[M^*(t) | M; \Psi]\} = E(M)F(t; \Psi), \quad (2.20)$$

e

$$\begin{aligned} \text{Var}[M^*(t)] &= E\{\text{Var}[M^*(t) | M; \Psi]\} + \text{Var}\{E[M^*(t) | M; \Psi]\} \\ &= E\{MF(t; \Psi)(1 - F(t; \Psi))\} + \text{Var}\{MF(t; \Psi)\} \\ &= E(M)F(t; \Psi) - E(M)F^2(t; \Psi) + \text{Var}(M)F^2(t; \Psi) \\ &= [\text{Var}(M) - E(M)]F^2(t; \Psi) + E(M)F(t; \Psi). \end{aligned} \quad (2.21)$$

Observa-se da equação (2.17) que a probabilidade de não detectar tumores observáveis até o instante t pode ser escrita como

$$\begin{aligned} p_0(t; \theta, \Psi) &= \sum_{m=0}^{\infty} \binom{m}{0} [F(t; \Psi)]^0 [1 - F(t; \Psi)]^m p(m; \theta) \\ &= \sum_{m=0}^{\infty} [1 - F(t; \Psi)]^m p(m; \theta) \\ &= A_M [1 - F(t; \Psi)]. \end{aligned} \quad (2.22)$$

Ou seja, a probabilidade de cura dinâmica $p_0(t; \theta, \Psi)$ é a função geradora de probabilidades da variável M aplicada no valor $1 - F(t; \Psi)$. Como já mostrado anteriormente,

tomando seu limite quando $t \rightarrow \infty$ tem-se

$$\begin{aligned}
 \lim_{t \rightarrow \infty} p_0(t; \theta, \Psi) &= \lim_{t \rightarrow \infty} A_M [1 - F(t; \Psi)] \\
 &= \lim_{t \rightarrow \infty} \sum_{m=0}^{\infty} [1 - F(t; \Psi)]^m p(m; \theta) \\
 &= \lim_{t \rightarrow \infty} \{ [1 - F(t; \Psi)]^0 p(0; \theta) + [1 - F(t; \Psi)]^1 p(1; \theta) + \dots \} \\
 &= p(0; \theta) = p_0,
 \end{aligned} \tag{2.23}$$

e fazendo $t \rightarrow 0$, tem-se

$$\begin{aligned}
 \lim_{t \rightarrow 0} p_0(t; \theta, \Psi) &= \lim_{t \rightarrow 0} A_M [1 - F(t; \Psi)] \\
 &= \lim_{t \rightarrow 0} \{ [1 - F(t; \Psi)]^0 p(0; \theta) + [1 - F(t; \Psi)]^1 p(1; \theta) + \dots \} \\
 &= \sum_{m=0}^{\infty} p(m; \theta) = 1.
 \end{aligned} \tag{2.24}$$

Utilizando uma distribuição de Poisson para M com parâmetro $\theta = \lambda$, ou seja, supondo que o agente cancerígeno inicia em média λ tumores por animal no início do experimento, tem-se que sua função de probabilidade é dada por

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

Assim, a função de probabilidade da variável $M^*(t)$ é dada por

$$\begin{aligned}
P[M^*(t) = m^*; \lambda, \Psi] &= \sum_{m=m^*}^{\infty} \binom{m}{m^*} [F(t; \Psi)]^{m^*} [1 - F(t; \Psi)]^{m-m^*} \frac{e^{-\lambda} \lambda^m}{m!} \\
&= \sum_{m=m^*}^{\infty} \frac{m!}{(m-m^*)! m^*!} [F(t; \Psi)]^{m^*} [1 - F(t; \Psi)]^{m-m^*} \frac{e^{-\lambda} \lambda^m}{m!} \\
&= \frac{e^{-\lambda} [F(t; \Psi)]^{m^*}}{m^*!} \sum_{m=m^*}^{\infty} \frac{\lambda^m [1 - F(t; \Psi)]^{m-m^*}}{(m-m^*)!} \times \frac{\lambda^{m^*}}{\lambda^{m^*}} \\
&= \frac{\lambda^{m^*} e^{-\lambda} [F(t; \Psi)]^{m^*}}{m^*!} \sum_{m=m^*}^{\infty} \frac{\lambda^{m-m^*} [1 - F(t; \Psi)]^{m-m^*}}{(m-m^*)!} \\
&= \frac{\lambda^{m^*} e^{-\lambda} [F(t; \Psi)]^{m^*}}{m^*!} \sum_{m=m^*}^{\infty} \frac{[\lambda(1 - F(t; \Psi))]^{m-m^*}}{(m-m^*)!} \\
&= \frac{\lambda^{m^*} e^{-\lambda} [F(t; \Psi)]^{m^*}}{m^*!} \sum_{m-m^*=0}^{\infty} \frac{[\lambda(1 - F(t; \Psi))]^{m-m^*}}{(m-m^*)!} \\
&= \frac{\lambda^{m^*} e^{-\lambda} [F(t; \Psi)]^{m^*}}{m^*!} e^{\lambda(1-F(t;\Psi))} \\
&= \frac{[\lambda F(t; \Psi)]^{m^*} e^{-\lambda F(t;\Psi)}}{m^*!} \sim Poisson(\lambda F(t; \Psi)). \tag{2.25}
\end{aligned}$$

Ou seja, supondo uma distribuição de Poisson com parâmetro λ para M , verifica-se que a variável $M^*(t)$ tem distribuição de Poisson com média e variância dadas por $\lambda F(t; \Psi)$.

A probabilidade de um animal não apresentar tumores detectáveis até o instante de tempo t é dada por

$$P[M^*(t) = 0; \lambda, \Psi] = \frac{[\lambda F(t; \Psi)]^0 e^{-\lambda F(t; \Psi)}}{0!} = e^{-\lambda F(t; \Psi)} = p_0(t; \lambda, \Psi). \tag{2.26}$$

Note que tomando o limite quando $t \rightarrow \infty$ na equação (2.26), obtém-se a fração de cura de longa duração p_0 , ou seja,

$$\lim_{t \rightarrow \infty} p_0(t; \lambda, \Psi) = P[M = 0; \theta] = e^{-\lambda} = p_0. \tag{2.27}$$

Supondo M uma variável seguindo uma distribuição Binomial Negativa com vetor de parâmetros $\theta = (N, p)$, tem-se que sua função de probabilidade é dada por

$$p(m; N, p) = \binom{N+m-1}{m} p^N (1-p)^m, \quad m = 0, 1, \dots$$

A média e a variância da distribuição Binomial Negativa, são dadas respectivamente

por

$$\mu = \frac{N(1-p)}{p} \quad e \quad \sigma^2 = \mu + \tau\mu^2,$$

em que $\tau = \frac{1}{N}$.

Com o propósito de facilitar as interpretações dos parâmetros do modelo e auxiliar na convergência dos procedimentos numéricos, utilizou-se uma reparametrização da distribuição Binomial Negativa. Desta forma, será reparametrizada a função de probabilidade de M em função dos parâmetros μ e τ para que facilite a estimação por máxima verossimilhança.

Assim, a função de probabilidade da variável M em função dos parâmetros μ e τ é dada por

$$p(m; \mu, \tau) = \binom{1/\tau + m - 1}{m} (1 + \tau\mu)^{-1/\tau} \left(\frac{\tau\mu}{1 + \tau\mu} \right)^m, \quad m = 0, 1, \dots \quad (2.28)$$

Desta forma, a distribuição de probabilidade da variável $M^*(t)$ fica

$$\begin{aligned} P[M^*(t) = m^*; \mu, \tau, \Psi] &= \sum_{m=m^*}^{\infty} \binom{m}{m^*} F(t; \Psi)^{m^*} (1 - F(t; \Psi))^{m-m^*} \left(\frac{1/\tau + m - 1}{m} \right) (1 + \tau\mu)^{-\frac{1}{\tau}} \left(\frac{\tau\mu}{1 + \tau\mu} \right)^m \\ &= \sum_{m=m^*}^{\infty} \frac{m! F(t; \Psi)^{m^*} (1 - F(t; \Psi))^{m-m^*} \left(\frac{1/\tau + m - 1}{m} \right)!}{(m - m^*)! m^*! \left(\frac{1/\tau - 1}{m} \right)!} (1 + \tau\mu)^{-\frac{1}{\tau}} \left(\frac{\tau\mu}{1 + \tau\mu} \right)^m \\ &= \frac{(1 + \tau\mu)^{-\frac{1}{\tau}} F(t; \Psi)^{m^*}}{m^*! \left(\frac{1/\tau - 1}{m^*} \right)!} \sum_{m=m^*}^{\infty} \left(\frac{\tau\mu}{1 + \tau\mu} \right)^m \frac{\left(\frac{1/\tau + m - 1}{m} \right)! [1 - F(t; \Psi)]^{m-m^*}}{(m - m^*)!} \times \frac{\left(\frac{\tau\mu}{1 + \tau\mu} \right)^{m^*}}{\left(\frac{\tau\mu}{1 + \tau\mu} \right)^{m^*}} \\ &= \frac{(1 + \tau\mu)^{-\frac{1}{\tau}}}{m^*! \left(\frac{1/\tau - 1}{m^*} \right)!} \left(\frac{F(t; \Psi)\tau\mu}{1 + \tau\mu} \right)^{m^*} \sum_{m=m^*}^{\infty} \frac{\left(\frac{1/\tau + m - 1}{m} \right)!}{(m - m^*)!} \left[\frac{(1 - F(t; \Psi))\tau\mu}{1 + \tau\mu} \right]^{m-m^*} \end{aligned}$$

Multiplicando e dividindo por $(1/\tau + m^* - 1)!$ tem-se

$$P[M^*(t) = m^*; \mu, \tau, \Psi] = \frac{\left(\frac{1/\tau + m^* - 1}{m^*} \right)! (1 + \tau\mu)^{-\frac{1}{\tau}}}{m^*! \left(\frac{1/\tau - 1}{m^*} \right)!} \left(\frac{F(t; \Psi)\tau\mu}{1 + \tau\mu} \right)^{m^*} \sum_{m=m^*}^{\infty} \binom{\frac{1/\tau + m - 1}{m}}{m - m^*} \left[\frac{(1 - F(t; \Psi))\tau\mu}{1 + \tau\mu} \right]^{m-m^*}. \quad (2.29)$$

Utilizando as propriedades de coeficientes binomiais, tem-se que o somatório apresentado na equação (2.29) pode ser representado pela expressão $\left[1 - \frac{(1 - F(t; \Psi))\tau\mu}{1 + \tau\mu} \right]^{-\left(\frac{1/\tau + m^*}{m^*} \right)}$.

Assim, tem-se

$$\begin{aligned}
P[M^*(t) = m^*; \mu, \tau, \Psi] &= \frac{\left(\frac{1}{\tau} + m^* - 1\right)! (1 + \tau\mu)^{-\frac{1}{\tau}}}{m^*! \left(\frac{1}{\tau} - 1\right)!} \left(\frac{F(t; \Psi)\tau\mu}{1 + \tau\mu}\right)^{m^*} \left[1 - \frac{(1 - F(t; \Psi))\tau\mu}{1 + \tau\mu}\right]^{-(1/\tau + m^*)} \\
&= \binom{\frac{1}{\tau} + m^* - 1}{m^*} \left(\frac{1}{1 + \tau\mu}\right)^{\frac{1}{\tau}} \left(\frac{F(t; \Psi)\tau\mu}{1 + \tau\mu}\right)^{m^*} \left[\frac{1 + \tau\mu}{1 + \tau\mu F(t; \Psi)}\right]^{\frac{1}{\tau} + m^*} \\
&= \binom{\frac{1}{\tau} + m^* - 1}{m^*} \left(\frac{1}{1 + \tau\mu F(t; \Psi)}\right)^{\frac{1}{\tau}} \left(\frac{\tau\mu F(t; \Psi)}{1 + \tau\mu F(t; \Psi)}\right)^{m^*}.
\end{aligned}$$

Com este resultado, observa-se que a variável $M^*(t)$ apresenta distribuição

$$M^*(t) \sim BN\left(\frac{1}{\tau}, \frac{1}{1 + \tau\mu F(t; \Psi)}\right). \quad (2.30)$$

Ou seja, supondo distribuição Binomial Negativa com parâmetros μ e τ para M , verifica-se que a variável $M^*(t)$ tem distribuição Binomial Negativa com média $\mu F(t; \Psi)$ e variância $\tau\mu^2 F^2(t; \Psi) + \mu F(t; \Psi)$.

Da mesma forma que no caso da distribuição de Poisson, a probabilidade do animal não apresentar tumores detectáveis até o instante t é calculada por

$$\begin{aligned}
P[M^*(t) = 0; \mu, \tau, \Psi] &= \binom{\frac{1}{\tau} + 0 - 1}{0} \left(\frac{1}{1 + \tau\mu F(t; \Psi)}\right)^{\frac{1}{\tau}} \left(\frac{\tau\mu F(t; \Psi)}{1 + \tau\mu F(t; \Psi)}\right)^0 \\
&= \left(\frac{1}{1 + \tau\mu F(t; \Psi)}\right)^{\frac{1}{\tau}} = p_0(t; \mu, \tau, \Psi).
\end{aligned} \quad (2.31)$$

A fração de cura de longa duração é obtida tomando-se o limite da equação (2.31) quando t tende a infinito, ou seja,

$$p_0 = \lim_{t \rightarrow \infty} p_0(t; \mu, \tau, \Psi) = \left(\frac{1}{1 + \tau\mu}\right)^{\frac{1}{\tau}} = (1 + \tau\mu)^{-1/\tau}. \quad (2.32)$$

Com o propósito de exemplificar os resultados desenvolvidos anteriormente, são calculadas as funções de sobrevivência mínima e a taxa de cura de longa duração para as distribuições de Poisson e Binomial Negativa. No próximo capítulo estes resultados são obtidos também fazendo uso da distribuição COM-Poisson (Shmueli et al., 2005) com o propósito de comparar os resultados obtidos para estas três distribuições. A seguir são apresentados os exemplos.

Exemplo 2.3: Distribuição de Poisson

Considerando que o número de tumores iniciados pelo agente cancerígeno, M , segue

uma distribuição de Poisson com parâmetro λ , tem-se que sua f.g.p. é dada por

$$\begin{aligned} A_M(s) &= E[s^M] = \sum_{m=0}^{\infty} s^m \frac{e^{-\lambda} [-\lambda]^m}{m!} \\ &= e^{-\lambda} \sum_{m=0}^{\infty} \frac{(-\lambda s)^m}{m!} \\ &= e^{-\lambda} e^{-\lambda s} = e^{-\lambda(1-s)}. \end{aligned} \quad (2.33)$$

Calculando a função apresentada na equação (2.33) no valor $s = 1 - F(t; \Psi)S(t; \Psi)$, tem-se

$$\begin{aligned} A_M[1 - F(t; \Psi)S(t; \Psi)] &= e^{-\lambda[1 - (1 - F(t; \Psi)S(t; \Psi))]} \\ &= e^{-\lambda F(t; \Psi)S(t; \Psi)}. \end{aligned} \quad (2.34)$$

Conseqüentemente, a função de sobrevivência imprópria mínima é dada por

$$\begin{aligned} S_{min}(t; \lambda, \Psi) &= 1 + p_0(t; \lambda, \Psi) - A_M[1 - F(t; \Psi)S(t; \Psi)] \\ &= 1 + e^{-\lambda F(t; \Psi)} - e^{-\lambda F(t; \Psi)S(t; \Psi)}. \end{aligned} \quad (2.35)$$

As funções de densidade de probabilidade e de risco impróprias são dadas, respectivamente, por:

$$\begin{aligned} f_{min}(t; \lambda, \Psi) &= -\frac{d}{dt} \{1 + e^{-\lambda F(t; \Psi)} - e^{-\lambda F(t; \Psi)S(t; \Psi)}\} \\ &= -\frac{d}{dt} e^{-\lambda F(t; \Psi)} + \frac{d}{dt} e^{-\lambda F(t; \Psi)S(t; \Psi)} \\ &= -\frac{d}{dt} [-\lambda F(t; \Psi)] e^{-\lambda F(t; \Psi)} + \frac{d}{dt} [-\lambda F(t; \Psi)S(t; \Psi)] e^{-\lambda F(t; \Psi)S(t; \Psi)} \\ &= \lambda f(t; \Psi) e^{-\lambda F(t; \Psi)} - \lambda [f(t; \Psi)S(t; \Psi) - F(t; \Psi)f(t; \Psi)] e^{-\lambda F(t; \Psi)S(t; \Psi)} \\ &= \lambda f(t; \Psi) e^{-\lambda F(t; \Psi)} - \lambda f(t; \Psi) [S(t; \Psi) - F(t; \Psi)] e^{-\lambda F(t; \Psi)S(t; \Psi)} \\ &= \lambda f(t; \Psi) [e^{-\lambda F(t; \Psi)} - e^{-\lambda F(t; \Psi)S(t; \Psi)} - 2F(t; \Psi)e^{-\lambda F(t; \Psi)S(t; \Psi)}] \end{aligned} \quad (2.36)$$

$$h_{min}(t; \lambda, \Psi) = \frac{\lambda f(t; \Psi) [e^{-\lambda F(t; \Psi)} - e^{-\lambda F(t; \Psi)S(t; \Psi)} - 2F(t; \Psi)e^{-\lambda F(t; \Psi)S(t; \Psi)}]}{1 + e^{-\lambda F(t; \Psi)} - e^{-\lambda F(t; \Psi)S(t; \Psi)}}$$

A fração de cura é obtida fazendo-se

$$\begin{aligned} p_0 &= \lim_{t \rightarrow \infty} p_0(t; \lambda, \Psi) = \lim_{t \rightarrow \infty} e^{-\lambda F(t; \Psi)} \\ &= e^{-\lambda}. \end{aligned} \quad (2.37)$$

Exemplo 2.4: Distribuição Binomial Negativa

Supondo que o número de tumores iniciados pelo agente cancerígeno, M , segue uma distribuição binomial negativa com parâmetros μ e τ , tem-se que sua função geradora de probabilidades é obtida por

$$\begin{aligned} A_M(s) &= E[s^M] = \sum_{m=0}^{\infty} s^m \binom{1/\tau + m - 1}{m} (1 + \tau\mu)^{-1/\tau} \left(\frac{\tau\mu}{1 + \tau\mu} \right)^m \\ &= (1 + \tau\mu)^{-1/\tau} \sum_{m=0}^{\infty} \binom{1/\tau + m - 1}{m} \left(\frac{s\tau\mu}{1 + \tau\mu} \right)^m \times \frac{\left(1 - \frac{s\tau\mu}{1 + \tau\mu} \right)^{1/\tau}}{\left(1 - \frac{s\tau\mu}{1 + \tau\mu} \right)^{1/\tau}} \\ &= (1 + \tau\mu - s\tau\mu)^{-1/\tau} \sum_{m=0}^{\infty} \binom{1/\tau + m - 1}{m} \left(\frac{s\tau\mu}{1 + \tau\mu} \right)^m \left(1 - \frac{s\tau\mu}{1 + \tau\mu} \right)^{1/\tau} \\ &= (1 + \tau\mu - s\tau\mu)^{-1/\tau}. \end{aligned} \quad (2.38)$$

Calculando a função apresentada na equação (2.38) no valor $s = 1 - F(t; \Psi)S(t; \Psi)$, tem-se

$$\begin{aligned} A_M[1 - F(t; \Psi)S(t; \Psi)] &= [1 + \tau\mu - \tau\mu(1 - F(t; \Psi)S(t; \Psi))]^{-1/\tau} \\ &= [1 + \tau\mu F(t; \Psi)S(t; \Psi)]^{-1/\tau}. \end{aligned} \quad (2.39)$$

Conseqüentemente, a função de sobrevivência mínima é dada por

$$\begin{aligned} S_{min}(t; \mu, \tau, \Psi) &= 1 + p_0(t; \mu, \tau, \Psi) - A_M[1 - F(t; \Psi)S(t; \Psi)] \\ &= 1 + [1 + \tau\mu F(t; \Psi)]^{-1/\tau} - [1 + \tau\mu F(t; \Psi)S(t; \Psi)]^{-1/\tau}. \end{aligned} \quad (2.40)$$

A fração de cura é obtida fazendo-se

$$\begin{aligned} p_0 &= \lim_{t \rightarrow \infty} S_{min}(t; \mu, \tau, \Psi) = \lim_{t \rightarrow \infty} p_0(t; \mu, \tau, \Psi) = \lim_{t \rightarrow \infty} [1 + \tau\mu F(t; \Psi)]^{-1/\tau} \\ &= (1 + \tau\mu)^{-1/\tau}. \end{aligned} \quad (2.41)$$

A seguir apresentamos uma tabela que resume as expressões obtidas nesta seção.

Dada uma distribuição para T com função de distribuição $F(t; \Psi)$, em que Ψ é o

Tabela 2.1: Representação de $S_{min}(t; \theta, \Psi)$, $p_0(t; \theta, \Psi)$ e p_0 para as distribuições Poisson e Binomial Negativa.

Modelo Destrutivo	$S_{min}(t; \theta, \Psi)$	$p_0(t; \theta, \Psi)$	p_0
Poisson	$1 + e^{-\lambda F(t; \Psi)} - e^{-\lambda F(t; \Psi)S(t; \Psi)}$	$e^{-\lambda F(t; \Psi)}$	$e^{-\lambda}$
BN	$1 + [1 + \tau \mu F(t; \Psi)]^{\frac{-1}{\tau}} - [1 + \tau \mu F(t; \Psi)S(t; \Psi)]^{\frac{-1}{\tau}}$	$[1 + \tau \mu F(t; \Psi)]^{\frac{-1}{\tau}}$	$(1 + \tau \mu)^{\frac{-1}{\tau}}$

vetor de parâmetros desconhecido, e fixando valores para o parâmetro θ da distribuição do número de tumores induzidos por animal, M , pode-se obter o comportamento da função de sobrevivência mínima para vários valores de t . Sua fração de cura também pode ser estimada utilizando os valores obtidos para os parâmetros da distribuição de M . A fração de cura de longa duração não depende da distribuição dos tempos de detecção dos tumores, mas a função $p_0(t; \theta, \Psi)$ depende.

A seguir são apresentados alguns comportamentos para as funções de sobrevivência mínimas, fixando uma distribuição de Weibull para a variável T , e variando-se os valores dos parâmetros das distribuições de Poisson e Binomial Negativa consideradas para a variável M . As frações de cura são apresentadas entre parênteses.

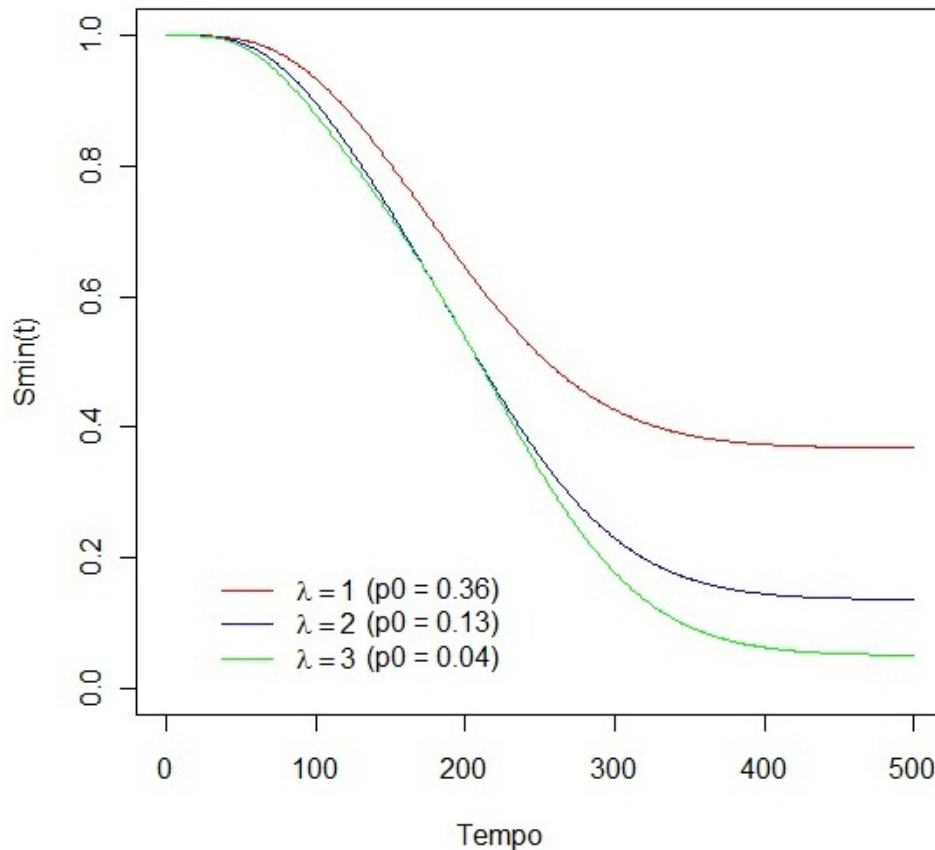


Figura 2.1: Comportamento de $S_{min}(t; \lambda, \Psi)$ considerando a distribuição de Poisson com parâmetros $\lambda = 1$, $\lambda = 2$ e $\lambda = 3$, respectivamente, para M .

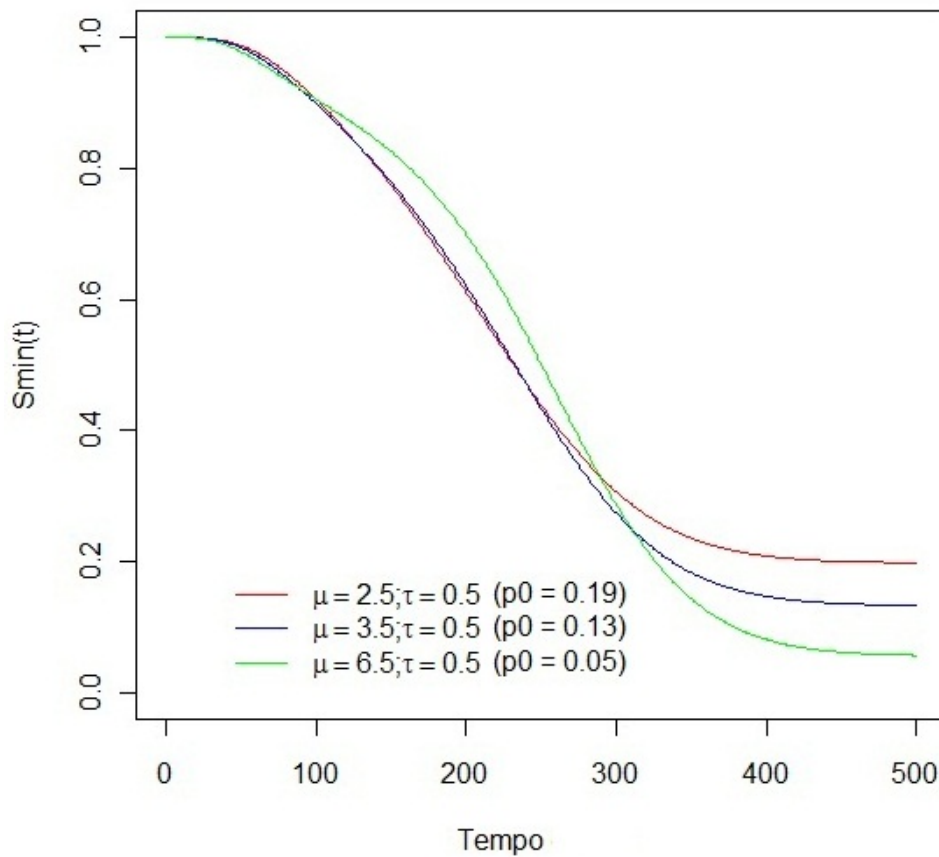


Figura 2.2: Comportamento de $S_{min}(t; \mu, \tau, \Psi)$ considerando a distribuição Binomial Negativa com parâmetros $(\mu = 2.5, \tau = 0.5)$, $(\mu = 3.5, \tau = 0.5)$ e $(\mu = 6.5, \tau = 0.5)$, respectivamente, para M .

Pode-se observar com base na Figura 2.1, que utiliza a distribuição de Poisson para a variável M , que conforme λ aumenta, a função de sobrevivência mínima decresce mais rapidamente para zero, reduzindo a probabilidade de cura de 0,36 para 0,04, conforme λ varia de 1 para 3. A interpretação biológica para isto é que a probabilidade de cura do animal reduz conforme aumenta-se o número de tumores induzidos por animal.

Já na Figura 2.2 quando se utiliza a distribuição binomial negativa para a variável M , observa-se uma redução menor na probabilidade de cura conforme aumenta-se o número médio de tumores induzidos, ou seja, a distribuição binomial negativa por apresentar um parâmetro a mais em sua modelagem, o parâmetro τ , apresenta um comportamento mais realista do que a distribuição de Poisson.

Capítulo 3

A Distribuição COM-Poisson

Em problemas de análise de sobrevivência de longa duração é comum o interesse na modelagem de variáveis discretas, como por exemplo, o número de tumores desenvolvidos em um indivíduo. Na modelagem desta variável, as distribuições mais comumente utilizadas são a Poisson e a Binomial Negativa. Entretanto, em alguns problemas mais complexos estas distribuições apresentam certas limitações, dificultando um bom ajuste ao conjunto de dados.

Uma distribuição reutilizada recentemente além de bastante flexível para o ajuste de variáveis discretas, como por exemplo para o número de tumores desenvolvidos em animais experimentais, é a Distribuição Conway-Maxwell Poisson (Conway e Maxwell, 1962), conhecida como COM-Poisson, que tem se tornado uma alternativa flexível para o ajuste de dados de sobrevivência que envolvem fração de cura. Uma grande vantagem deste modelo em relação aos modelos tradicionais para dados de contagem é o fato de ele conter um parâmetro de dispersão e assim acomodar com mais facilidade problemas de sobredispersão ou subdispersão, frequentemente encontrados em dados discretos. No contexto de experimentos com animais, ênfase será dada ao uso da distribuição COM-Poisson.

A distribuição Conway-Maxwell Poisson foi utilizada inicialmente pelos autores Conway e Maxwell (1962) e retomada recentemente por Shmueli et al. (2005). Ela flexibiliza a distribuição de Poisson através de seu parâmetro de dispersão ϕ , sendo adequada para problemas com sobredispersão bem como subdispersão. A função de probabilidade de uma variável aleatória M com distribuição COM-Poisson é dada por

$$P[M = m; \eta, \phi] = \frac{1}{Z(\eta, \phi)} \frac{\eta^m}{(m!)^\phi}, \quad m = 0, 1, 2, \dots, \quad (3.1)$$

em que $\phi \geq 0$, $\eta > 0$ e $Z(\eta, \phi) = \sum_{j=0}^{\infty} \frac{\eta^j}{(j!)^\phi}$ é uma constante de normalização.

Sua função geradora de probabilidades é obtida por meio da expressão

$$\begin{aligned} A_M(s) &= E(s^M) = \sum_{m=0}^{\infty} P[M = m; \eta, \phi] s^m \\ &= \sum_{m=0}^{\infty} s^m \frac{1}{Z(\eta, \phi)} \frac{\eta^m}{(m!)^\phi} = \frac{1}{Z(\eta, \phi)} \sum_{m=0}^{\infty} \frac{(s\eta)^m}{(m!)^\phi} = \frac{Z(s\eta, \phi)}{Z(\eta, \phi)}. \end{aligned} \quad (3.2)$$

Para a Distribuição COM-Poisson (Shmueli et al., 2005), os momentos populacionais podem ser obtidos por meio da expressão

$$E[M^{r+1}] = \begin{cases} \eta E[M + 1]^{1-\phi}, & \text{se } r = 0; \\ \eta \frac{d}{d\eta} E[M^r] + E[M] E[M^r], & \text{se } r > 0. \end{cases}$$

Desta forma, o valor médio e a variância para esta distribuição são dados respectivamente por

$$E[M] = \sum_{m=0}^{\infty} \frac{m\eta^m}{Z(\eta, \phi)(m!)^\phi} \quad e \quad Var[M] = \sum_{m=0}^{\infty} \frac{m^2\eta^m}{Z(\eta, \phi)(m!)^\phi} - \mu^2.$$

Quando o parâmetro ϕ da COM-Poisson assume valor um, a distribuição de Poisson torna-se um caso particular da distribuição COM-Poisson.

Além da distribuição de Poisson, a distribuição COM-Poisson apresenta como casos particulares outras distribuições muito conhecidas na literatura como a distribuição de Bernoulli e a Geométrica, as quais podem ser obtidas variando-se tanto o parâmetro de dispersão ϕ , como o valor de η . Para valores de $\phi > 1$ obtém-se situações correspondentes à subdispersão, enquanto que para $0 \leq \phi < 1$ obtém-se situações correspondentes à sobdispersão, relativa à Distribuição de Poisson. Quando $\phi \rightarrow \infty$, a Distribuição COM-Poisson se aproxima da Distribuição Bernoulli com parâmetro $\eta/(1 + \eta)$. Quando $\phi = 0$ a COM-Poisson se reduz à Distribuição Geométrica com parâmetro $1 - \eta$, com $0 < \eta < 1$.

Com o propósito de verificar se a distribuição COM-Poisson pertence à família exponencial de distribuições, realizar-se-ão algumas manipulações algébricas. Assim, aplicando-se o exponencial e o logaritmo na equação (3.1) simultaneamente, obtém-se a forma:

$$\begin{aligned} P[M = m; \eta, \phi] &= \exp\{m \log(\eta) + \log(m!)^{-\phi} - \log Z(\eta, \phi)\} \\ &= (m!)^{-\phi} \exp\{m \log(\eta) - \log Z(\eta, \phi)\}. \end{aligned}$$

Considerando $\omega = \log(\eta) \Rightarrow \eta = e^\omega$, a função acima assume a forma:

$$P[M = m; \omega, \phi] = (m!)^{-\phi} \exp\{m\omega - \log Z(e^\omega, \phi)\}. \quad (3.3)$$

Assim, com $\omega = \log(\eta)$, $K(\omega, \phi) = \log Z(e^\omega, \phi)$ e $\varphi(m; \phi) = (m!)^{-\phi}$, tem-se a distribuição COM-Poisson na forma da família exponencial natural,

$$p(m; \omega, \phi) = \varphi(m; \phi) \exp\{m\omega - K(\omega, \phi)\}. \quad (3.4)$$

onde $p(m; \omega, \phi) = P[M = m; \omega, \phi]$.

Conforme será mostrado no próximo teorema, a distribuição COM-Poisson também pode ser considerada um caso particular da distribuição Poisson Ponderada (Kokonendji et al., 2008) cuja densidade é dada por

$$p_w(m; \omega) = \frac{w(m; \phi)p(m; \omega)}{E_\omega[w(M; \phi)]} = P(M^w = m), \quad (3.5)$$

em que a variável aleatória M segue uma distribuição de Poisson com parâmetro $\omega > 0$ e função de probabilidade $p(m; \omega) = P(M = m)$, a variável M^w é a versão ponderada de M , com $w(m; \phi)$ sendo uma função não-negativa chamada de função peso que pode depender ou não do parâmetro ϕ e o valor $E_\omega(\cdot)$ é o valor médio que depende de ω .

O teorema a seguir (Kokonendji et al., 2008) estabelece a relação de uma distribuição na forma exponencial natural com a classe de distribuições conhecidas por funções ponderadas de Poisson.

Teorema 3.0.1. *Seja $\phi > 0$ e $\omega > 0$. Então a função densidade de probabilidade dada pela expressão*

$$p(m; \omega, \phi) = \varphi(m; \phi) \exp\{m\omega - K(\omega, \phi)\}, \quad m \in \mathfrak{N}, \quad (3.6)$$

em que \mathfrak{N} representa o conjunto dos números naturais, é uma distribuição de Poisson ponderada, em relação à distribuição de Poisson, com média $\eta = \exp(\omega)$. Neste caso, a função peso correspondente é,

$$w(m; \phi) = m!\varphi(m; \phi), \quad m \in \mathfrak{N}.$$

Prova:

Considerando que M segue uma distribuição de Poisson com parâmetro e^ω , tem-se:

$$\begin{aligned}
E_\omega[w(M; \phi)] &= \sum_{m \in \mathbb{N}} w(m; \phi) \frac{e^{-e^\omega} (e^\omega)^m}{m!} \\
&= \sum_{m \in \mathbb{N}} \frac{w(m; \phi)}{m!} \exp\{m\omega - e^\omega\} \\
&= \sum_{m \in \mathbb{N}} \varphi(m; \phi) \exp\{m\omega - e^\omega\} \\
&= \exp \left\{ \log \left[\sum_{m \in \mathbb{N}} \varphi(m; \phi) \exp(m\omega - e^\omega) \right] \right\} \\
&= \exp \left\{ -e^\omega + \log \left[\sum_{m \in \mathbb{N}} \varphi(m; \phi) \exp(m\omega) \right] \right\} \\
&= \exp\{-e^\omega + K(\omega; \phi)\}, \tag{3.7}
\end{aligned}$$

pode-se, então, escrever a função de probabilidade de M como,

$$\begin{aligned}
p(m; \omega, \phi) &= \varphi(m; \phi) \exp\{m\omega - K(\omega; \phi)\} \\
&= \varphi(m; \phi) \exp\{m\omega - K(\omega; \phi)\} \times \frac{m! \exp(-e^\omega)}{m! \exp(-e^\omega)} \\
&= \frac{m! \varphi(m; \phi)}{\exp\{-e^\omega + K(\omega; \phi)\}} \times \frac{\exp\{m\omega - e^\omega\}}{m!} \\
&= \frac{w(m; \phi)}{E_\omega[w(M; \phi)]} \times p(m; \omega) \\
&= p_w(m; \omega). \tag{3.8}
\end{aligned}$$

para todo $m \in \mathbb{N}$, completando a prova.

Este teorema estabelece que uma variável aleatória M com distribuição COM-Poisson com parâmetros $\eta = e^\omega$ e ϕ , pode ser escrita na forma da família exponencial natural apresentada na equação (3.4) e, portanto, pode ser vista como uma distribuição de Poisson Ponderada com função peso dada por

$$w(m; \phi) = m! \varphi(m; \phi) = (m!)^{1-\phi}.$$

3.1 Função de Probabilidade de $M^*(t)$

Considerando que o número de tumores induzidos, M , segue uma distribuição COM-Poisson com parâmetro $\theta = (\eta, \phi)$, tem-se que a distribuição da variável destrutiva $M^*(t)$

é dada por

$$\begin{aligned}
P[M^*(t) = m^*; \eta, \phi, \Psi] &= \sum_{m=m^*}^{\infty} \binom{m}{m^*} [F(t; \Psi)]^{m^*} [1 - F(t; \Psi)]^{m-m^*} \frac{\eta^m}{Z(\eta, \phi)(m!)^\phi} \\
&= \sum_{m=m^*}^{\infty} \frac{m!}{(m-m^*)!m^*!} [F(t; \Psi)]^{m^*} [1 - F(t; \Psi)]^{m-m^*} \frac{\eta^m}{Z(\eta, \phi)(m!)^\phi} \\
&= \frac{[F(t; \Psi)]^{m^*}}{m^*!Z(\eta, \phi)} \sum_{m=m^*}^{\infty} \frac{(m!)^{1-\phi} [1 - F(t; \Psi)]^{m-m^*} \eta^m}{(m-m^*)!} \times \frac{\eta^{m^*}}{\eta^{m^*}} \\
&= \frac{[\eta F(t; \Psi)]^{m^*}}{m^*!Z(\eta, \phi)} \sum_{m=m^*}^{\infty} \frac{(m!)^{1-\phi} [\eta(1 - F(t; \Psi))]^{m-m^*}}{(m-m^*)!}.
\end{aligned}$$

Fazendo $j = m - m^*$, tem-se que $m = m^* + j$. Desta forma, a expressão acima se torna

$$\begin{aligned}
p(m^*; \eta, \phi, \Psi) &= \frac{[\eta F(t; \Psi)]^{m^*}}{m^*!Z(\eta, \phi)} \sum_{j=0}^{\infty} \frac{[(m^* + j)!]^{1-\phi} [\eta(1 - F(t; \Psi))]^j}{j!} \\
&= \frac{[\eta F(t; \Psi)]^{m^*}}{m^*!Z(\eta, \phi)} \sum_{j=0}^{\infty} \frac{[(m^*!)(m^* + 1) \times \dots \times (m^* + j)]^{1-\phi} [\eta(1 - F(t; \Psi))]^j}{j!} \\
&= \frac{[\eta F(t; \Psi)]^{m^*}}{m^*!Z(\eta, \phi)} (m^*!)^{1-\phi} \sum_{j=0}^{\infty} \frac{[(m^* + 1)(m^* + 2) \times \dots \times (m^* + j)]^{1-\phi} [\eta(1 - F(t; \Psi))]^j}{j!} \\
&= \frac{[\eta F(t; \Psi)]^{m^*}}{(m^*!)^\phi Z(\eta, \phi)} {}_1F_\phi(m^* + 1; m^* + 1, \dots, m^* + 1; \eta(1 - F(t; \Psi))) \\
&= \frac{[\eta F(t; \Psi)]^{m^*}}{(m^*!)^\phi} \frac{{}_1F_\phi(m^* + 1; m^* + 1, \dots, m^* + 1; \eta(1 - F(t; \Psi)))}{{}_1F_\phi(1; 1, \dots, 1; \eta)},
\end{aligned}$$

em que $p(m^*; \eta, \phi, \Psi) = P[M^*(t) = m^*; \eta, \phi, \Psi]$. A função ${}_uF_v(a_1, \dots, a_u; b_1, \dots, b_v; x)$ é chamada de função hipergeométrica generalizada e é definida como

$${}_uF_v(a_1, \dots, a_u; b_1, \dots, b_v; x) = \sum_{m=0}^{\infty} \frac{(a_1)_m (a_2)_m \times \dots \times (a_u)_m x^m}{(b_1)_m (b_2)_m \times \dots \times (b_v)_m m!},$$

em que $(a)_m = a(a+1) \times \dots \times (a+m-1)$.

Da mesma forma que nos casos das distribuições de Poisson e Binomial Negativa, a probabilidade de um animal não apresentar tumores detectáveis até o instante de tempo

t considerando o modelo COM-Poisson, é calculada por

$$\begin{aligned}
 p_0(t; \eta, \phi, \Psi) &= \frac{[\eta F(t; \Psi)]^0}{(0!)^\phi} \frac{{}_1F_\phi(0+1; 0+1, \dots, 0+1; \eta(1-F(t)))}{{}_1F_\phi(1; 1, \dots, 1; \eta)} \\
 &= \frac{{}_1F_\phi(1; 1, \dots, 1; \eta(1-F(t; \Psi)))}{{}_1F_\phi(1; 1, \dots, 1; \eta)} \\
 &= \frac{Z(\eta(1-F(t; \Psi)); \phi)}{Z(\eta; \phi)}, \tag{3.9}
 \end{aligned}$$

sendo que $p_0(t; \eta, \phi, \Psi) = P[M^*(t) = 0; \eta, \phi, \Psi]$.

3.2 Função de Sobrevivência Mínima

Na equação (2.15) da seção 2.1 verificou-se que a função de sobrevivência utilizando a variável $M^*(t)$ na modelagem é dada por

$$S_{min}(t; \theta, \Psi) = 1 + p_0(t; \theta, \Psi) - A_M [1 - F(t; \Psi)S(t; \Psi)], \tag{3.10}$$

em que θ é o vetor de parâmetros da distribuição da variável M , Ψ é o vetor de parâmetros da distribuição da variável T , $F(t; \Psi)$ é a função de distribuição acumulada e $S(t; \Psi) = 1 - F(t; \Psi)$.

Será apresentada agora o comportamento da função de sobrevivência mínima supondo uma distribuição COM-Poisson para M . Conforme foi demonstrado anteriormente a função geradora de probabilidades da variável aleatória M com distribuição COM-Poisson com parâmetro $\theta = (\eta, \phi)$, é dada por

$$A_M(s) = \frac{Z(\eta s, \phi)}{Z(\eta, \phi)}, \tag{3.11}$$

Tomando na equação (3.11) o valor $s = 1 - F(t; \Psi)S(t; \Psi)$, tem-se

$$A_M[1 - F(t; \Psi)S(t; \Psi)] = \frac{Z[\eta(1 - F(t; \Psi)S(t; \Psi)), \phi]}{Z(\eta, \phi)}. \tag{3.12}$$

Desta forma, substituindo os resultados apresentados nas equações (3.9) e (3.12) na equação (3.10), tem-se que a função de sobrevivência mínima, considerando uma distribuição COM-Poisson para a variável M e um processo destrutivo representado pela

variável $M^*(t)$, é dada por

$$\begin{aligned} S_{min}(t; \eta, \phi, \Psi) &= 1 + p_0(t; \eta, \phi, \Psi) - A_M [1 - F(t; \Psi)S(t; \Psi)] \\ &= 1 + \frac{Z(\eta(1 - F(t; \Psi)); \phi)}{Z(\eta; \phi)} - \frac{Z[\eta(1 - F(t; \Psi)S(t; \Psi)), \phi]}{Z(\eta, \phi)}. \end{aligned} \quad (3.13)$$

A fração de cura, p_0 , para o modelo destrutivo de longa duração COM-Poisson, é obtida tomando-se o limite da equação (3.13) quando $t \rightarrow \infty$, ou seja,

$$p_0 = \lim_{t \rightarrow \infty} S_{min}(t; \eta, \phi, \Psi) = P[M = 0; \eta, \phi] = \frac{1}{Z(\eta; \phi)}. \quad (3.14)$$

A tabela a seguir apresenta os principais resultados obtidos para a distribuição COM-Poisson, juntamente com os resultados já expostos anteriormente para as distribuições de Poisson e Binomial Negativa, considerando um processo destrutivo na análise de experimentos com animais.

Tabela 3.1: Representação das funções $S_{min}(t; \theta, \Psi)$, $p_0(t; \theta, \Psi)$ e p_0 para as distribuições de Poisson com parâmetro $\theta = \lambda$, Binomial Negativa com parâmetros $\theta = (\mu, \tau)$ e COM-Poisson com parâmetros $\theta = (\eta, \phi)$.

Modelo Destrutivo	$S_{min}(t; \theta, \Psi)$	$p_0(t; \theta, \Psi)$	p_0
Poisson	$1 + e^{-\lambda F(t; \Psi)} - e^{-\lambda F(t; \Psi)S(t; \Psi)}$	$e^{-\lambda F(t; \Psi)}$	$e^{-\lambda}$
BN	$1 + [1 + \tau \mu F(t; \Psi)]^{-\frac{1}{\tau}} - [1 + \tau \mu F(t; \Psi)S(t; \Psi)]^{-\frac{1}{\tau}}$	$[1 + \tau \mu F(t; \Psi)]^{-\frac{1}{\tau}}$	$[1 + \tau \mu]^{-\frac{1}{\tau}}$
COM-Poisson	$1 + \frac{Z[\eta(1 - F(t; \Psi)); \phi]}{Z(\eta; \phi)} - \frac{Z[\eta(1 - F(t; \Psi)S(t; \Psi)); \phi]}{Z(\eta, \phi)}$	$\frac{Z[\eta(1 - F(t; \Psi)); \phi]}{Z(\eta; \phi)}$	$\frac{1}{Z(\eta, \phi)}$

Com o propósito de exemplificar o comportamento da função de sobrevivência mínima considerando o modelo destrutivo COM-Poisson, apresentou-se na Figura 3.1 o gráfico de $S_{min}(t; \eta, \phi, \Psi)$, empregando uma distribuição de Weibull para a variável T com vetor de parâmetros Ψ conhecido e fixo. Variou-se os valores do parâmetro η da distribuição COM-Poisson considerada para a variável M , fazendo $\eta = 2.5, 3.5, 6.5$. Fixou-se também o valor do parâmetro de dispersão $\phi = 1.5$, ou seja, considerou-se uma situação de sub-dispersão nos valores da variável M . As frações de cura são mostradas entre parênteses.

Pode-se observar na Figura 3.1 que em uma situação com subdispersão nos dados ($\phi = 1, 5$), as funções de sobrevivência mínimas apresentam um comportamento bem próximo até o tempo $t = 250$, aproximadamente, a partir deste tempo as curvas se distanciam sendo que o maior valor de η , $\eta = 6, 5$, é o que apresenta menor probabilidade de cura, $p_0 = 0, 01$, enquanto que para o maior valor de η , $\eta = 2, 5$, esta probabilidade sobe para $p_0 = 0, 13$.

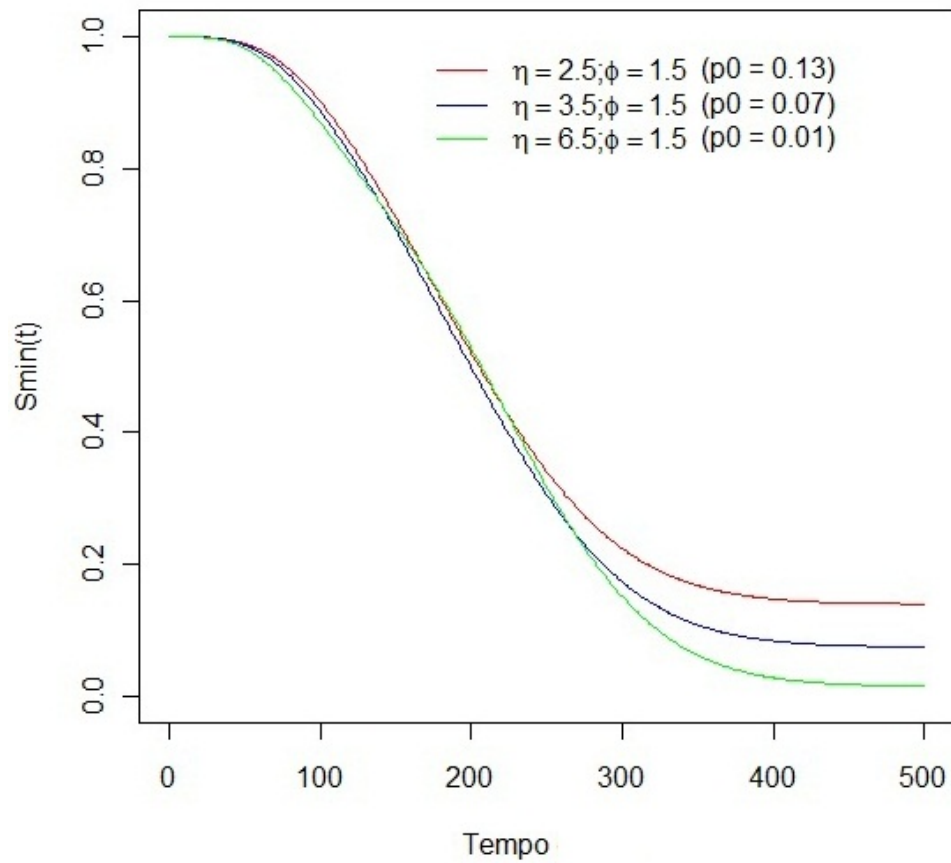


Figura 3.1: Comportamento de $S_{min}(t; \eta, \phi, \Psi)$ utilizando a distribuição COM-Poisson com parâmetros $(\eta = 2.5, \phi = 1.5)$, $(\eta = 3.5, \phi = 1.5)$ e $(\eta = 6.5, \phi = 1.5)$, respectivamente, para a variável M .

Capítulo 4

O Modelo Kokoska/Freedman

Análise de dados envolvendo experimentos com animais são comumente utilizados na ciência médica cujo objetivo é avaliar se um tratamento é eficiente ao combate de alguma doença. Estes experimentos consistem na indução química de tumores em animais experimentais, por meio de algum agente cancerígeno, cujo objetivo é avaliar a eficiência de tratamentos quimiopreventivos testados. As variáveis resposta comumente utilizadas para avaliar a eficiência destes tratamentos são o número de tumores induzidos e seus tempos de desenvolvimento.

Após a realização do experimento, o que se observa é o número de tumores detectados por animal e seus tempos de detecção. Considera-se nestes experimentos quimiopreventivos com animais, que a indução química dos tumores é realizada por uma única exposição ao agente cancerígeno, que pode-se dar direta ou indiretamente. Um tempo após essa exposição, os animais são designados aleatoriamente aos tratamentos considerados e, então, são acompanhados para avaliar o processo de desenvolvimento dos tumores. No caso de experimentos com ratos, durante o acompanhamento os animais são frequentemente avaliados por apalpação. O tumor e seu tempo de detecção são registrados quando o tumor atinge um tamanho palpável.

No caso de experimentos com animais, supõe-se que o agente cancerígeno inicia M tumores malignos por animal. Os tempos de detecção de todos os tumores induzidos são denotados por T_1, T_2, \dots, T_M . Considera-se como tempos de detecção observados, os tempos que satisfazem a relação $T_j \leq d$, em que d é o tempo de duração do experimento. Ao fim do experimento, todos os animais são sacrificados.

Assim, denotamos o número de tumores induzidos pela variável aleatória M , cuja média e variância são μ_M e σ_M^2 , respectivamente, e consideramos a distribuição acumulada $F(t; \Psi)$ para os tempos de desenvolvimento. A variável $M^*(t)$, já definida na equação (2.2), é o número de tumores detectáveis por animal no instante t , cuja média e variância são $\mu_M F(t; \Psi)$ e $(\sigma_M^2 - \mu_M) F^2(t; \Psi) + \mu_M F(t; \Psi)$, respectivamente.

Dentre os M tumores induzidos, o que se observa são $m^* \leq M$ tumores, cujos tempos de detecção são T_1, \dots, T_{m^*} . Seus valores observados são dados pelas estatísticas de ordem $t_{(1)}, \dots, t_{(m^*)}$.

A função de verossimilhança é apresentada a seguir.

4.1 Função de Verossimilhança

Utilizando a equação (2.17), que apresenta a distribuição da variável $M^*(t)$, aplicada no valor t_i^* , obtém-se a probabilidade de observar m_i^* tumores malignos no i -ésimo animal, $i = 1, \dots, n$, até o tempo de morte do animal t_i^* , ou seja,

$$P[M_i^*(t_i^*) = m_i^*; \theta, \Psi] = \sum_{m_i=m_i^*}^{\infty} \binom{m_i}{m_i^*} [F(t_i^*; \Psi)]^{m_i^*} [1 - F(t_i^*; \Psi)]^{m_i - m_i^*} p(m_i; \theta) \quad (4.1)$$

em que θ é o vetor de parâmetros da distribuição da variável número de tumores induzidos M e Ψ é o vetor de parâmetros da distribuição da variável tempos de detecção T .

Seja $t_{i(j)}$, $j = 1, \dots, m_i^*$, o j -ésimo menor tempo de detecção observado dentre os tempos de detecção de todos os tumores $t_{i1}, t_{i2}, \dots, t_{im_i^*}$ no i -ésimo animal, em que os valores ordenados são $t_{i(1)} \leq t_{i(2)} \leq \dots \leq t_{i(j)} \leq \dots \leq t_{i(m_i^*)}$. Considera-se que cada animal morre ou é censurado no tempo t_i^* , e que estes tempos são independentes do número de tumores e seus tempos de aparecimento. Esta é uma suposição utilizada nos artigos de Kokoska (1987) e Fredman (1993), que caracteriza censura não-informativa nos tempos de vida dos animais, o que não representa uma verdade biológica. Uma forma mais realista de analisar esta situação seria levar em conta esta dependência considerando assim uma censura do tipo informativa na análise, e utilizar modelos para eventos recorrentes (Lawless (1995) e Cheng Wang *et al.* (2001)) para solucionar este problema. Sugere-se esta abordagem para pesquisas futuras.

Assim, a função de verossimilhança para o i -ésimo animal é dada por

$$\begin{aligned} L_i(\theta, \Psi) &= P[M_i^*(t_i^*) = m_i^*; \theta, \Psi] \times f(t_{i(1)}|t_i^*; \Psi) \times f(t_{i(2)}|t_i^*; \Psi) \times \dots \times f(t_{i(m_i^*)}|t_i^*; \Psi) \\ &= P[M_i^*(t_i^*) = m_i^*; \theta, \Psi] \prod_{j=1}^{m_i^*} f(t_{i(j)}|t_i^*; \Psi) \\ &= P[M_i^*(t_i^*) = m_i^*; \theta, \Psi] \prod_{j=1}^{m_i^*} \frac{f(t_{i(j)}; \Psi)}{F(t_i^*; \Psi)}, \end{aligned} \quad (4.2)$$

em que $f(t_{i(j)}|t_i^*; \Psi) = f(t_{i(j)}; \Psi)/F(t_i^*; \Psi)$. Desta forma, tem-se

$$\begin{aligned}
L_i(\theta, \Psi) &= \frac{P[M_i^*(t_i^*) = m_i^*; \theta, \Psi]}{F(t_i^*; \Psi)^{m_i^*}} f_{T_{i(1)}, T_{i(2)}, \dots, T_{i(m_i^*)}}(t_{i1}, t_{i2}, \dots, t_{im_i^*}; \Psi) \\
&= \left[\sum_{m_i=m_i^*}^{\infty} \binom{m_i}{m_i^*} [1 - F(t_i^*; \Psi)]^{m_i - m_i^*} p(m_i; \theta) \right] f_{T_{i(1)}, T_{i(2)}, \dots, T_{i(m_i^*)}}(t_{i1}, t_{i2}, \dots, t_{im_i^*}; \Psi)
\end{aligned} \tag{4.3}$$

A distribuição conjunta das estatísticas de ordem $t_{i(j)}$ na equação 4.3 será obtida utilizando-se o Teorema C.0.1 apresentado no apêndice C retirado do livro do Ross, 1997. Assim, a função de verossimilhança do i -ésimo animal é dada por

$$\begin{aligned}
L_i(\theta, \Psi) &= \left[\sum_{m_i=m_i^*}^{\infty} \binom{m_i}{m_i^*} [1 - F(t_i^*; \Psi)]^{m_i - m_i^*} p(m_i; \theta) \right] \times m_i^*! \prod_{j=1}^{m_i^*} f(t_{ij}; \Psi) \\
&= m_i^*! \left[\left\{ \prod_{j=1}^{m_i^*} f(t_{ij}; \Psi) \right\} \sum_{m_i=m_i^*}^{\infty} \binom{m_i}{m_i^*} [1 - F(t_i^*; \Psi)]^{m_i - m_i^*} p(m_i; \theta) \right].
\end{aligned} \tag{4.4}$$

Supondo n animais no grupo de interesse, tem-se que a função de verossimilhança para a amostra completa é obtida fazendo-se o produto das verossimilhanças individuais, ou seja,

$$\begin{aligned}
L(\theta, \Psi) &= \prod_{i=1}^n L_i(\theta, \Psi) \\
&= \prod_{i=1}^n m_i^*! \left[\left\{ \prod_{j=1}^{m_i^*} f(t_{ij}; \Psi) \right\} \sum_{m_i=m_i^*}^{\infty} \binom{m_i}{m_i^*} [1 - F(t_i^*; \Psi)]^{m_i - m_i^*} p(m_i; \theta) \right],
\end{aligned} \tag{4.5}$$

Utilizou-se o método de máxima verossimilhança para estimar os parâmetros de interesse do modelo.

Seja $\vartheta = (\theta, \Psi)'$ o vetor de parâmetros de dimensão $p \times 1$ a ser estimado e $\hat{\vartheta} = (\hat{\theta}, \hat{\Psi})'$ o estimador de máxima verossimilhança de ϑ . Sob condições de regularidade, a distribuição assintótica de $\hat{\vartheta}$ é normal multivariada com vetor média ϑ e matriz de variâncias e covariâncias $\Sigma(\vartheta)$ de dimensão $p \times p$, dada por

$$\Sigma(\vartheta) = \left[\left\{ -E \left(\frac{\partial^2 \log L(\vartheta)}{\partial \vartheta_i \partial \vartheta_j} \right) \right\}_{i,j=1, \dots, p} \right]^{-1} \tag{4.6}$$

A matriz de variâncias e covariâncias para $\hat{\vartheta}$ depende dos valores dos parâmetros θ e Ψ que são quantidades desconhecidas. Desta forma, após estimar estes parâmetros pelo

método de máxima verossimilhança, pode-se utilizar estas estimativas na obtenção da matriz de variâncias e covariâncias. A estimativa da matriz de variâncias e covariâncias para $\hat{\vartheta}$ será denotada por $\hat{\Sigma}(\hat{\vartheta})$.

O interesse é estimar os parâmetros das distribuições $f(t; \Psi)$ e $p(m_i; \theta)$ e comparar os resultados obtidos ao longo de uma gama de modelos paramétricos. No caso em que $f(t; \Psi)$ tem distribuição Gama com vetor de parâmetros $\Psi = (\alpha, \beta)$ e $p(m_i; \theta)$ tem distribuição de Poisson com parâmetro $\theta = \lambda$, tem-se o caso particular do modelo de Kokoska (1987).

Na Tabela 1 são apresentadas duas distribuições para $f(t; \Psi)$ e três distribuições para $p(m; \theta)$. A fim de flexibilizar as distribuições para o número de tumores induzidos por animal, priorizou-se o uso da distribuição COM-Poisson, amplamente utilizada para dados discretos por acomodar com facilidade a dispersão contida nos dados.

Tabela 4.1: Modelos Paramétricos

Tempo de Aparecimento do Tumor, t		Número de Tumores, m	
Distribuição	$f(t; \Psi)$	Distribuição	$p(m; \theta)$
Gama(α, β)	$\frac{t^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} \exp\left[-\frac{t}{\beta}\right]$	Poisson(λ)	$\frac{e^{-\lambda} \lambda^m}{m!}$
Weibull(b, c)	$\frac{ct^{c-1}}{b^c} \exp\left[-\left(\frac{t}{b}\right)^c\right]$	BN (μ, τ)	$\binom{\frac{1}{\tau} + m - 1}{m} (1 + \tau\mu)^{-\frac{1}{\tau}} \left(\frac{\tau\mu}{1 + \tau\mu}\right)^m$
		COM-Poisson (η, ϕ)	$\frac{\eta^m}{Z(\eta, \phi)(m)^\phi}$

4.2 Estudo de Simulação

Para avaliar o comportamento do modelo COM-Poisson para diferentes valores dos parâmetros, tamanhos de amostra e porcentagens de censura, realizou-se um estudo de simulação com o propósito de verificar as propriedades dos estimadores de máxima verossimilhança. Para esta finalidade, calculou-se as medidas de viés, erro quadrático médio (EQM) e probabilidade de cobertura (PC), cujas definições são apresentadas a seguir. Utilizou-se nestas definições, ϑ_i como sendo o i -ésimo componente de ϑ , $i = 1, \dots, p$, em que $\vartheta = (\vartheta_1, \vartheta_2, \dots, \vartheta_i, \dots, \vartheta_p)'$. Seja $\hat{\vartheta}_i$ o estimador de máxima verossimilhança do parâmetro ϑ_i , e $E(\hat{\vartheta}_i)$ sendo o valor esperado de $\hat{\vartheta}_i$. Assim, pode-se definir,

- **Viés**

O viés do estimador $\hat{\vartheta}_i$ é a diferença entre o valor esperado de $\hat{\vartheta}_i$ e o parâmetro ϑ_i . Ou seja,

$$\text{Viés}(\hat{\vartheta}_i) = E(\hat{\vartheta}_i) - \vartheta_i, \quad (4.7)$$

para todo $i = 1, \dots, p$.

Um estimador cujo viés é zero é chamado de estimador não-viesado (ou não tendencioso), e satisfaz a condição

$$E(\hat{\vartheta}_i) = \vartheta_i, \quad \text{para todo } i = 1, \dots, p.$$

Para que a estimativa seja considerada boa, espera-se que seu valor médio seja próximo do valor que está sendo estimado. Ou seja, considerando o cálculo da estimativa para muitas amostras da população, espera-se que a média destas estimativas seja próxima do verdadeiro valor do parâmetro.

- **Erro Quadrático Médio (EQM)**

O erro quadrático médio (EQM) do estimador $\hat{\vartheta}_i$ do parâmetro ϑ_i , é definido por

$$EQM(\hat{\vartheta}_i) = E[(\hat{\vartheta}_i - \vartheta_i)^2]. \quad (4.8)$$

O EQM apresenta duas principais vantagens em relação às outras medidas, primeiramente por ser uma medida bastante tratável analiticamente e, segundo, que ele pode ser reescrito como

$$EQM(\hat{\vartheta}_i) = Var(\hat{\vartheta}_i) + [Viés(\hat{\vartheta}_i)]^2. \quad (4.9)$$

em que $Var(\hat{\vartheta}_i) = E(\hat{\vartheta}_i - E(\hat{\vartheta}_i))^2$ é a variância do estimador $\hat{\vartheta}_i$.

- **Probabilidade de Cobertura (PC)**

Para cada amostra simulada estimou-se o vetor de parâmetros $\hat{\boldsymbol{\vartheta}}$ do modelo pelo método de máxima verossimilhança e então, estimou-se a matriz de variâncias e covariâncias apresentada na equação (4.6). Seja $\hat{\sigma}_{ii}^2$ um estimador da variância do parâmetro ϑ_i , $i = 1, \dots, p$. Pode-se calcular desta forma o intervalo de confiança para o i -ésimo parâmetro, considerando a distribuição assintótica normal de $\hat{\vartheta}_i$, da seguinte forma:

$$IC(\vartheta_i, (1 - \alpha)\%) = \hat{\vartheta}_i \pm z_{\alpha/2} \hat{\sigma}_{ii}. \quad (4.10)$$

onde α é o nível de significância considerado, $z_{\alpha/2}$ é o valor tabelado da distribuição normal padrão e $\hat{\sigma}_{ii} = \sqrt{\hat{\sigma}_{ii}^2}$.

Assim para cada amostra simulada, calculou-se os intervalos de confiança para cada parâmetro do modelo. A proporção de intervalos de confiança que incluem o verdadeiro valor do parâmetro dentre todos os intervalos calculados é chamada de Probabilidade de Cobertura (PC).

Inicialmente foram simuladas 1000 amostras de tamanhos $n=30, 50, 100$ e 300 . Atribuiu-se para a variável número de tumores induzidos, M , uma distribuição COM-Poisson com parâmetros $\eta = 2, 0$ e $\phi = 0, 7$, ou seja, uma situação de dados com sobredispersão. Para a variável tempo de detecção dos tumores considerou-se uma distribuição Gama com parâmetros $\alpha = 4, 5$ e $\beta = 25$. Foi considerado que 30% dos animais foram censurados, ou seja, 30% das observações não apresentaram o evento de interesse (morte) até o fim do estudo (280 dias). Utilizou-se o método quase-Newton L-BFGS-B, por meio da função `optim` da linguagem R (R Development Core Team, 2009).

Considerando as amostras simuladas, calculou-se a média das estimativas, a variância, o viés, o EQM e obtive-se os intervalos de confiança de 95% para cada parâmetro em estudo com base na teoria assintótica normal, verificando a proporção de intervalos que continham o verdadeiro valor do parâmetro com o propósito de calcular a probabilidade de cobertura (PC) para cada parâmetro. As amostras para as quais não houve convergência foram descartadas. Os resultados são apresentados na Tabela 4.2.

Na Tabela 4.3 atribuiu-se para a variável M uma distribuição COM-Poisson com parâmetros $\eta = 2, 0$ e $\phi = 1, 0$, ou seja, uma situação de dados sem dispersão em que a COM-Poisson se reduz a distribuição de Poisson com $\eta = 2, 0$.

Na Tabela 4.4 considerou-se para a variável número de tumores induzidos, M , uma distribuição COM-Poisson com parâmetros $\eta = 2, 0$ e $\phi = 1, 5$, ou seja, uma situação de dados com sub-dispersão relativa à distribuição de Poisson.

Observa-se por meio dos resultados apresentados nas tabelas 4.2, 4.3 e 4.4 que conforme aumenta-se o tamanho da amostra, as estimativas tendem a ficar mais próximas dos verdadeiros valores dos parâmetros, fazendo com que o viés das estimativas fiquem cada vez mais próximos de zero e, desta forma, os valores do EQM diminuem, ficando mais próximos da variância dos estimadores, ou seja, $EQM \rightarrow Var(\hat{\theta})$ quando $n \rightarrow \infty$, indicando que os estimadores obtidos estão de acordo com as propriedades dos estimadores de máxima verossimilhança. Além disso, verifica-se que a variância das estimativas diminui conforme o tamanho da amostra cresce, tornando os intervalos de confiança para os parâmetros mais precisos.

Observando os valores obtidos da PC nas tabelas 4.2, 4.3 e 4.4, verifica-se que para $\phi = 0, 7$ os valores da PC vão se distanciando de 95% conforme a amostra o tamanho da amostra cresce. Essa diferença diminui conforme reduzimos a dispersão nos dados, ou seja, conforme aumenta-se o valor de ϕ , saindo de uma situação de sobre-dispersão para uma situação de sub-dispersão, os resultados para a probabilidades de cobertura tornam-se melhores, aproximando-se cada vez mais no nível de confiança de 95% que foi utilizado na obtenção dos intervalos de confiança para os parâmetros em estudo.

Pode-se observar também que a sobre-dispersão existente nos dados afeta principalmente o parâmetro de escala β da distribuição Gama. Isso nos leva a crer que as propriedades assintóticas dos estimadores de máxima verossimilhança estão sendo afetadas pelo parâmetro de dispersão ϕ da distribuição COM-Poisson, sendo que os resultados tendem a ficar melhores na situação em que se tem menos dispersão nos dados (sub-dispersão), ou seja, quando $\phi > 1$.

Devido a esta influência do parâmetro ϕ nas propriedades assintóticas dos estimadores de máxima verossimilhança dos parâmetros do modelo proposto, uma sugestão alternativa de análise que surge como provável solução para este problema é a abordagem bayesiana para análise de dados. Esta metodologia é sugerida como proposta de pesquisas futuras para a abordagem deste problema, visando solucionar com mais eficiência e maior precisão o problema aqui proposto de análise de dados em experimentos com animais.

Tabela 4.2: Média, viés, EQM e PC para os estimadores de máxima verossimilhança das 1000 amostras simuladas utilizando $\phi = 0,7$.

n	Parâmetro	Média	Variância	Viés	EQM	PC
30	$\alpha = 4,5$	4,980	0,509	0,480	0,740	0,952
	$\beta = 25,0$	22,885	10,151	-2,114	14,622	0,883
	$\eta = 2,0$	2,185	0,035	0,185	0,069	0,969
	$\phi = 0,7$	0,726	0,014	0,026	0,015	0,964
50	$\alpha = 4,5$	4,915	0,274	0,415	0,446	0,949
	$\beta = 25,0$	23,008	6,433	-1,991	10,398	0,866
	$\eta = 2,0$	2,164	0,019	0,164	0,046	0,952
	$\phi = 0,7$	0,711	0,007	0,011	0,008	0,977
100	$\alpha = 4,5$	4,865	0,129	0,365	0,262	0,910
	$\beta = 25,0$	23,130	3,105	-1,869	6,599	0,831
	$\eta = 2,0$	2,163	0,010	0,163	0,037	0,955
	$\phi = 0,7$	0,710	0,003	0,010	0,003	0,984
300	$\alpha = 4,5$	4,847	0,042	0,347	0,163	0,700
	$\beta = 25,0$	23,086	1,045	-1,913	4,705	0,603
	$\eta = 2,0$	2,154	0,003	0,154	0,027	0,965
	$\phi = 0,7$	0,702	0,001	0,002	0,001	0,949

Com o intuito de avaliar o comportamento das funções de sobrevivência mínimas estimadas, realizou-se uma representação gráfica considerando os valores simulados para $\phi = 0,7$, $\phi = 1,0$ e $\phi = 1,5$, respectivamente. Estas funções são apresentadas nas Figuras 4.1, 4.2 e 4.3.

Observou-se por meio das Figuras 4.1, 4.2 e 4.3 que as estimativas para todos os valores dos parâmetros mantiveram-se próximas para os diferentes tamanhos de amostra. Para valores menores de t as funções de sobrevivência mínimas estimadas ficaram mais próximas da função real, enquanto que para valores grandes de t as funções de sobrevivência

Tabela 4.3: Média, viés, EQM, e PC para os estimadores de máxima verossimilhança das 1000 amostras simuladas utilizando $\phi = 1, 0$.

n	Parâmetro	Média	Variância	Viés	EQM	PC
30	$\alpha = 4, 5$	4,997	0,720	0,497	0,968	0,971
	$\beta = 25, 0$	23,033	15,388	-1,966	19,254	0,897
	$\eta = 2, 0$	2,553	0,067	0,553	0,373	0,970
	$\phi = 1, 0$	1,119	0,041	0,119	0,055	0,980
50	$\alpha = 4, 5$	4,979	0,428	0,479	0,658	0,959
	$\beta = 25, 0$	22,838	9,182	-2,161	13,852	0,875
	$\eta = 2, 0$	2,543	0,037	0,543	0,333	0,979
	$\phi = 1, 0$	1,108	0,021	0,108	0,033	0,977
100	$\alpha = 4, 5$	4,882	0,191	0,382	0,338	0,942
	$\beta = 25, 0$	23,119	4,616	-1,880	8,151	0,860
	$\eta = 2, 0$	2,530	0,019	0,530	0,300	0,970
	$\phi = 1, 0$	1,100	0,011	0,100	0,021	0,971
300	$\alpha = 4, 5$	4,848	0,067	0,348	0,188	0,790
	$\beta = 25, 0$	23,201	1,697	-1,798	4,931	0,726
	$\eta = 2, 0$	2,512	0,006	0,512	0,268	0,965
	$\phi = 1, 0$	1,083	0,003	0,083	0,010	0,965

Tabela 4.4: Média, viés, EQM, e PC para os estimadores de máxima verossimilhança das 1000 amostras simuladas utilizando $\phi = 1, 5$.

n	Parâmetro	Média	Variância	Viés	EQM	PC
30	$\alpha = 4, 5$	5,108	1,161	0,608	1,532	0,971
	$\beta = 25, 0$	22,952	21,846	-2,047	26,037	0,898
	$\eta = 2, 0$	2,996	0,095	0,996	1,088	0,983
	$\phi = 1, 5$	1,756	0,114	0,256	0,179	0,984
50	$\alpha = 4, 5$	5,036	0,685	0,536	0,973	0,947
	$\beta = 25, 0$	22,938	13,625	-2,061	17,875	0,877
	$\eta = 2, 0$	2,966	0,056	0,966	0,990	0,985
	$\phi = 1, 5$	1,720	0,065	0,220	0,114	0,984
100	$\alpha = 4, 5$	4,953	0,280	0,453	0,486	0,944
	$\beta = 25, 0$	22,945	6,527	-2,054	10,749	0,871
	$\eta = 2, 0$	2,968	0,027	0,968	0,966	0,978
	$\phi = 1, 5$	1,714	0,031	0,214	0,077	0,977
300	$\alpha = 4, 5$	4,870	0,093	0,370	0,230	0,847
	$\beta = 25, 0$	23,204	2,270	-1,795	5,496	0,797
	$\eta = 2, 0$	2,955	0,009	0,955	0,923	0,967
	$\phi = 1, 5$	1,696	0,010	0,196	0,048	0,967

estimadas ficaram mais próximas da real quando $\phi = 0, 7$, sendo que para $\phi = 1$ ficaram mais distantes e quando $\phi = 1, 5$ distanciaram-se mais ainda dos verdadeiros valores da

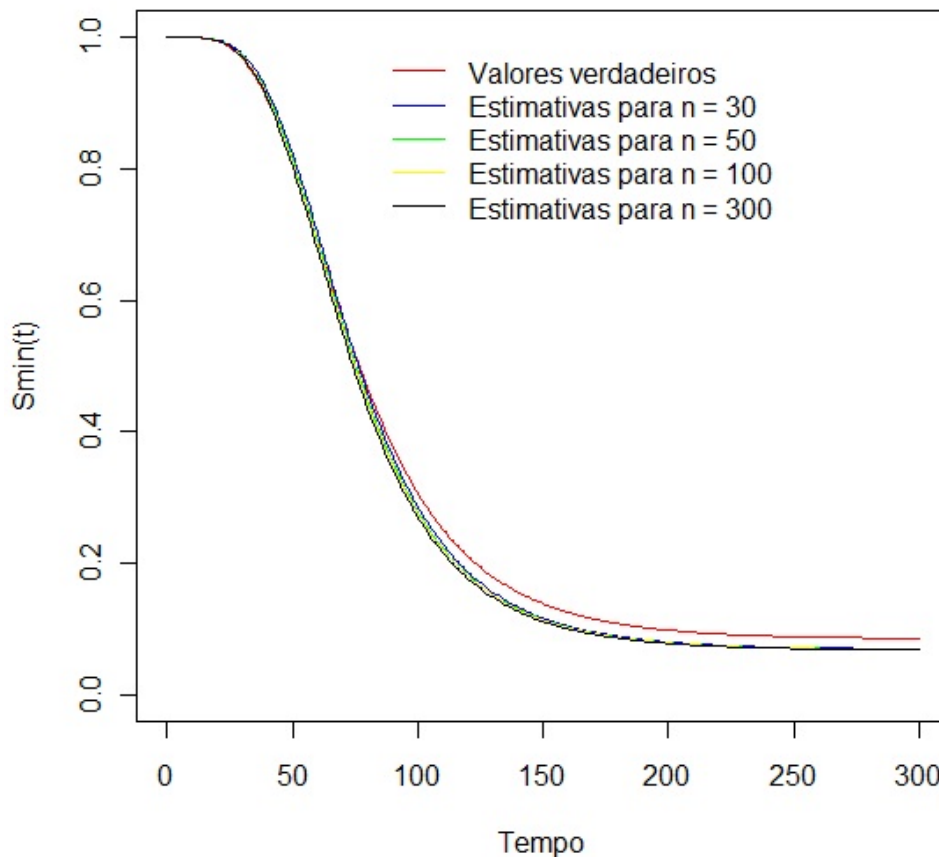


Figura 4.1: Estimativas de $S_{min}(t)$ para os dados simulados com $\phi = 0,7$.

função mínima de sobrevivência. A Tabela 4.5 apresenta as frações de cura calculadas para os verdadeiros valores dos parâmetros, considerando os diferentes tamanhos de amostras e os diferentes valores utilizados para ϕ .

Tabela 4.5: Valores reais para as taxas de cura comparadas as taxas de cura estimadas para tamanhos de amostra $n = 30$, $n = 50$, $n = 100$ e $n = 300$, considerando valores de $\phi = 0,7$, $\phi = 1,0$ e $\phi = 1,5$.

		$n = 30$	$n = 50$	$n = 100$	$n = 300$
ϕ	p_0	\hat{p}_0	\hat{p}_0	\hat{p}_0	\hat{p}_0
0,7	0,08461	0,0688	0,0682	0,0682	0,0676
1,0	0,1353	0,0929	0,0924	0,0924	0,0918
1,5	0,1952	0,1223	0,1194	0,1208	0,1203

Com o propósito de avaliar o comportamento das estimativas conforme a proporção de censuras existentes nos dados, realizou-se um estudo de simulação com valores dos parâmetros fixados em $\alpha = 4,5$ e $\beta = 25,0$, para a distribuição Gama da variável tempos de detecção dos tumores e $\eta = 2,0$ e $\phi = 0,7$, para a distribuição COM-Poisson da variável

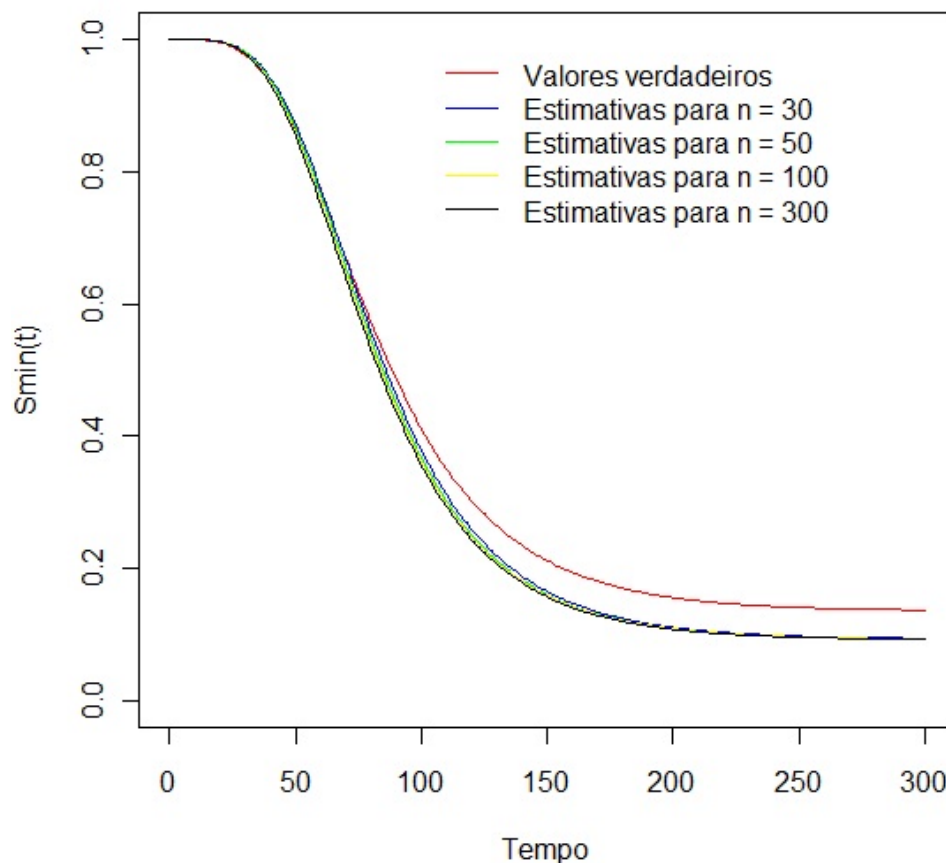


Figura 4.2: Estimativas de $S_{min}(t)$ para os dados simulados com $\phi = 1, 0$.

número de tumores iniciados, considerando um tamanho de amostra $n = 100$. Utilizou-se as seguintes porcentagens de censuras: 0%, 30%, 70% e 100%.

A censura aqui utilizada, é definida como o indivíduo que não apresentou o evento de interesse, neste caso a morte, até o fim do experimento, ou seja, o indivíduo que foi observado durante todo o tempo do experimento. O valor referente a 0% de censura representa a situação em que todos os indivíduos morreram antes que o experimento chegasse ao fim, sendo assim, todos os indivíduos considerados no estudo tiveram seus tempos de morte observados. O valor de censura correspondente a 100% representa a situação em que todos os indivíduos avaliados foram censurados, o que quer dizer que todos os indivíduos sobreviveram até o fim do experimento, sendo assim, não tiveram seus tempos de morte observados, o que se sabe é apenas que até o fim do experimento eles estavam vivos.

Os resultados obtidos para os estimadores dos parâmetros, a média, a variância, o viés, o EQM e a PC, para as quatro diferentes porcentagens de censura consideradas na análise, são dispostos na Tabela 4.6.

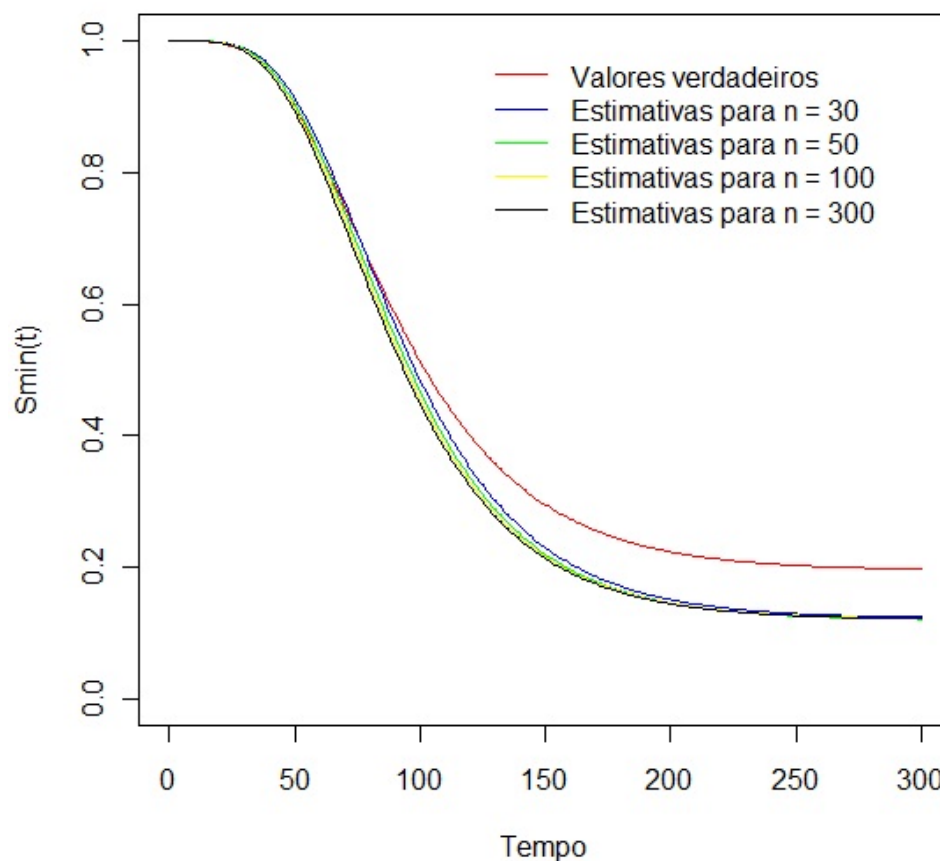


Figura 4.3: Estimativas de $S_{min}(t)$ para os dados simulados com $\phi = 1, 5$.

Pode-se observar por meio da Tabela 4.6 que conforme aumentamos a porcentagem de censura nos dados as estimativas tendem a ficar um pouco mais distantes dos verdadeiros valores dos parâmetros, aumentando ligeiramente as medidas de variância, viés e erro quadrático médio. Observa-se também que apesar de considerar uma situação de sobredispersão, a probabilidade de cobertura não foi tão afetada (quando comparada aos resultados anteriores) pelo parâmetro de dispersão ϕ , apresentando-se estável mesmo para altas porcentagens de censura. A Figura 4.4 apresenta o comportamento das funções de sobrevivência mínimas para as diferentes porcentagens de censura consideradas com $\phi = 0, 7$. A Tabela 4.7 apresenta os valores estimados das taxas de cura para as diferentes porcentagens de censura.

Observa-se por estes resultados que quanto maior a porcentagem de censura mais distante a função estimada da função de sobrevivência mínima real, verifica-se também que a porcentagem de censura afeta ligeiramente a probabilidade de cura estimada, que fica mais distante da verdadeira taxa de cura, conforme aumenta-se a proporção de observações censuradas no modelo.

Tabela 4.6: Média, viés, EQM, e PC para os estimadores de máxima verossimilhança das 1000 amostras simuladas de tamanho 100, considerando diferentes porcentagens de censuras nos dados.

Censura	Parâmetro	Média	Variância	Viés	EQM	PC
0%	$\alpha = 4,5$	4,873	0,120	0,373	0,260	0,899
	$\beta = 25,0$	23,131	2,985	-1,868	6,476	0,817
	$\eta = 2,0$	1,951	0,007	-0,048	0,010	0,963
	$\phi = 0,7$	0,625	0,003	-0,074	0,008	0,930
30%	$\alpha = 4,5$	4,865	0,129	0,365	0,262	0,910
	$\beta = 25,0$	23,130	3,105	-1,869	6,599	0,831
	$\eta = 2,0$	2,163	0,010	0,163	0,037	0,955
	$\phi = 0,7$	0,710	0,003	0,010	0,003	0,984
70%	$\alpha = 4,5$	4,881	0,124	0,381	0,270	0,904
	$\beta = 25,0$	22,963	2,829	-2,036	6,978	0,820
	$\eta = 2,0$	2,492	0,012	0,492	0,254	0,964
	$\phi = 0,7$	0,827	0,004	0,127	0,021	0,942
100%	$\alpha = 4,5$	4,882	0,133	0,381	0,279	0,886
	$\beta = 25,0$	22,936	2,990	-2,063	7,248	0,799
	$\eta = 2,0$	2,768	0,010	0,768	0,600	0,946
	$\phi = 0,7$	0,912	0,005	0,212	0,050	0,935

Todas estas informações obtidas do estudo de simulação realizado, nos dá indícios de que as estimativas dos parâmetros do modelo proposto não estão sendo significativamente afetadas pela censura presente nos dados, o que nos leva a concluir sobre a eficiência do modelo em acomodar com flexibilidade a censura inerente aos dados, estimando com precisão os parâmetros de interesse do modelo, mesmo quando consideradas altas porcentagens de censuras.

Tabela 4.7: Valores reais para as taxas de cura comparadas as taxas de cura estimadas para as porcentagens de censura 0%, 30%, 70%, e 100%, considerando $\phi = 0,7$ e $n = 100$.

	0%	30%	70%	100%
p_0	\hat{p}_0	\hat{p}_0	\hat{p}_0	\hat{p}_0
0.0846	0,0758	0,0682	0,0586	0,0519

4.3 Aplicação a Dados Reais

Visando exemplificar a utilização do modelo apresentado, realizou-se a análise de um conjunto de dados real retirado do artigo de Freedman (1993), o qual se refere a um experimento conduzido na Universidade do Alabama, Birmingham. Este experimento consistiu de dois grupos de 30 ratos fêmeas que foram alimentados com uma dieta padrão ou dieta

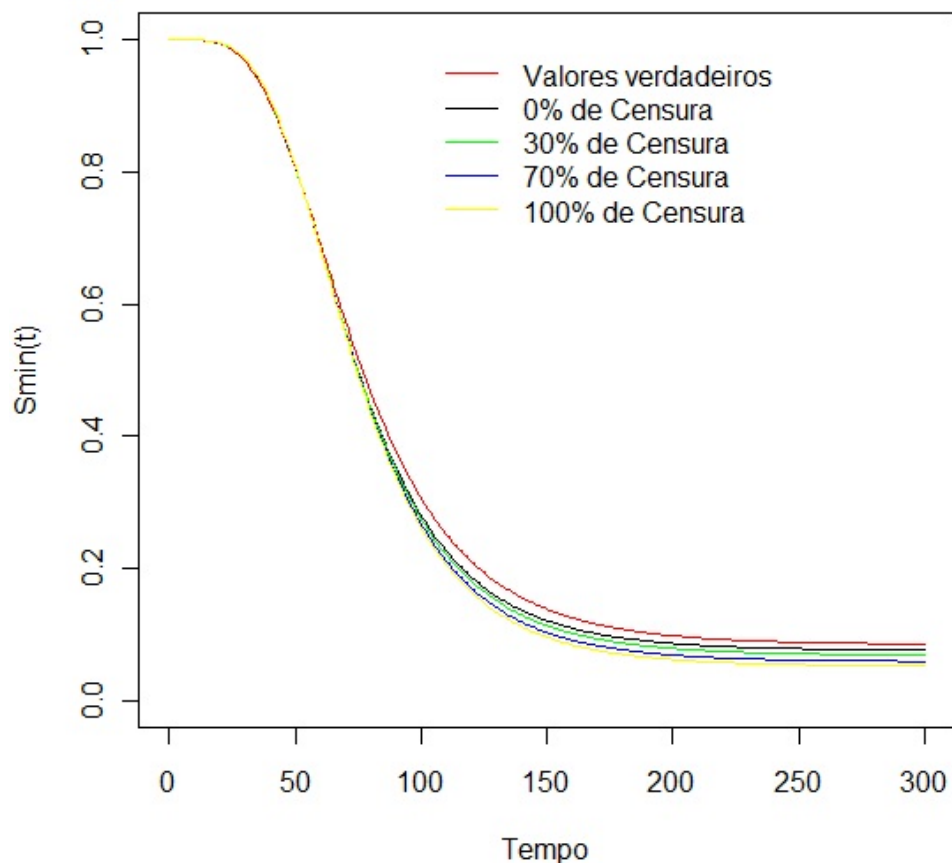


Figura 4.4: Estimativas de $S_{min}(t)$ para os dados simulados considerando as porcentagens de censura 0%, 30%, 70% e 100%.

acrescida com um agente A, antes, durante e após a administração da substância cancerígena. Os animais foram examinados semanalmente para a identificação dos tumores de mama palpável. Estes tumores foram posteriormente examinados histologicamente e foram considerados como um evento somente se fossem encontrados neles adenocarcinomas. Todos os animais que morreram ou que foram sacrificados ao final do experimento (280 dias) foram submetidos a autópsia para verificar adenocarcinomas não descobertos. Os dados são dispostos a seguir.

Tabela 4.8: Conjunto de dados de tumores mamários em ratos experimentais

Controle		Agente A	
Tempo do tumor (t_{ij})	Tempo de morte (t_i^*)	Tempo do tumor (t_{ij})	Tempo de morte (t_i^*)
—	92	—	135
—	185	—	206
—	191	—	217
—	200	—	247
—	203	—	260
—	208	—	268
—	214	—	272
—	276	—	280
—	279	—	280
—	279	48	204
—	279	70	186
—	279	113	212
—	279	120	145
—	66	135	230
34	232	176	280
58	234	280 ^a	280
85	279	280 ^a	280
92	279	34, 280 ^a	280
155	279	98, 246	280
279 ^a	279	148, 231	253
280 ^a	280	155, 230 ^a	237
280 ^a	280	225, 255 ^a	230
62, 62	78	141, 155, 219 ^a	255
85, 205	279	148, 231, 239	219
183, 279 ^a	279	70, 85, 85, 178 ^a	272
231, 279 ^a	279	98, 126, 148, 233, 255 ^a	178
62, 70, 107 ^a	107	92, 148, 176, 211, 211, 280 ^a	255
92, 113, 144, 191 ^a	191	58, 58, 64, 98, 98, 126 ^a , 126 ^a	280
120, 155, 225, 246	280	73, 85, 85, 98, 155, 162, 162, 190, 190, 205	126
113, 135, 169, 266, 279 ^a	279	92, 92, 113, 113, 135, 159, 162, 169, 169, 231, 267 ^a , 267 ^a , 267 ^a	280
92, 113, 120, 126, 197, 203 ^a	203		267

^aIndica que o tumor foi descoberto na autópsia.

Fonte: Dados retirados do artigo Freedman (1993).

Observa-se no conjunto de dados que o grupo de animais tratados com o agente A desenvolveu mais tumores do que o grupo controle. Ou seja, aparentemente o agente A está aumentando a produção de tumor, ao invés de preveni-lo. As questões de interesse a serem abordadas neste problema são: Para os diferentes modelos paramétricos considerados, o agente A pode aumentar a taxa de células lesionadas? Acelerar seus tempos de desenvolvimento? Ou ambos?

Para o conjunto de dados avaliado em Freedman (1993), o gráfico das estimativas de Kaplan-Meier para os grupos controle e tratamento considerando o máximo dos tempos de detecção (T^*) é apresentado na Figura 4.5.

Para obter as estimativas dos parâmetros das distribuições realizou-se a maximização da função de verossimilhança.

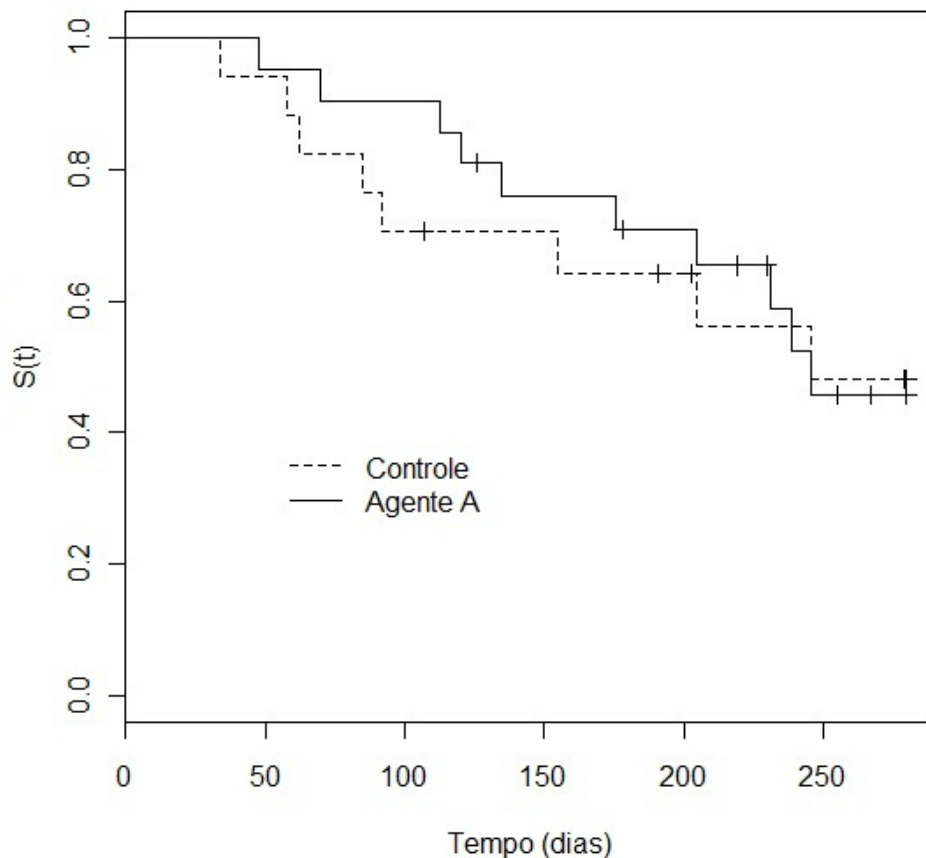


Figura 4.5: *Estimativas de Kaplan-Meier para os grupos Controle e Tratamento.*

As estimativas considerando os modelos de Poisson, Binomial Negativa e COM-Poisson para a variável número de tumores induzidos por animal e os modelos Gama e Weibull para os tempos de detecção são apresentadas nas Tabelas 4.9, 4.10 e 4.11. Os valores entre parênteses correspondem ao erro padrão das estimativas.

Utilizando os critérios de seleção de modelos AIC (Akaike information criterion) e

Tabela 4.9: Resultados dos Modelos Ajustados usando a Distribuição de Poisson

$p(m; \lambda)/f(t; \Psi)$	λ	Ψ	AIC	BIC
Poisson/Gamma	$\hat{\lambda}_1=1,2646$ (0,2)	$\hat{\alpha}_1=3,5895$ (1,1)	1002,18	1014,74
	$\hat{\lambda}_2=2,3207$ (0,3)	$\hat{\alpha}_2=4,1311$ (1,0)		
		$\hat{\beta}_1=44,025$ (20,0)		
		$\hat{\beta}_2=40,441$ (14,0)		
Poisson/Weibull	$\hat{\lambda}_1=1,2130$ (0,2)	$\hat{b}_1=168,175$ (22,8)	1004,80	1017,36
	$\hat{\lambda}_2=2,1714$ (0,3)	$\hat{b}_2=174,124$ (15,8)		
		$\hat{c}_1=2,1786$ (0,3)		
		$\hat{c}_2=2,4962$ (0,3)		

Tabela 4.10: Resultados dos Modelos Ajustados usando a Distribuição Binomial Negativa

$p(m; \mu, \tau)/f(t; \Psi)$	(μ, τ)	Ψ	AIC	BIC
BN/Gamma	$\hat{\mu}_1=1,5488$ (0,5)	$\hat{\alpha}_1=3,7286$ (1,2)	961,59	978,34
	$\hat{\mu}_2=2,4122$ (0,6)	$\hat{\alpha}_2=4,2574$ (1,0)		
	$\hat{\tau}_1=1,8623$ (1,0)	$\hat{\beta}_1=44,510$ (21,0)		
	$\hat{\tau}_2=1,4126$ (0,5)	$\hat{\beta}_2=39,346$ (13,4)		
BN/Weibull	$\hat{\mu}_1=1,4851$ (0,5)	$\hat{b}_1=177,77$ (27,7)	963,89	980,64
	$\hat{\mu}_2=2,2639$ (0,6)	$\hat{b}_2=174,34$ (15,0)		
	$\hat{\tau}_1=1,8573$ (1,0)	$\hat{c}_1=2,2594$ (0,4)		
	$\hat{\tau}_2=1,4279$ (0,5)	$\hat{c}_2=2,5692$ (0,3)		

Tabela 4.11: Resultados dos Modelos Ajustados usando a Distribuição COM-Poisson

$p(m; \eta, \phi)/f(t; \Psi)$	(η, ϕ)	Ψ	AIC	BIC
COM/Gamma	$\hat{\eta}_1=2,0208$ (1,7)	$\hat{\alpha}_1=4,4090$ (1,1)	796,76	813,51
	$\hat{\eta}_2=1,8389$ (1,4)	$\hat{\alpha}_2=5,4725$ (1,0)		
	$\hat{\phi}_1=0,8944$ (1,1)	$\hat{\beta}_1=30,409$ (8,5)		
	$\hat{\phi}_2=0,5252$ (0,6)	$\hat{\beta}_2=25,494$ (5,0)		
COM/Weibull	$\hat{\eta}_1=2,0246$ (1,7)	$\hat{b}_1=151,37$ (13,2)	796,16	812,91
	$\hat{\eta}_2=1,8568$ (1,4)	$\hat{b}_2=156,85$ (8,3)		
	$\hat{\phi}_1=0,8920$ (1,1)	$\hat{c}_1=2,3068$ (0,3)		
	$\hat{\phi}_2=0,5288$ (0,6)	$\hat{c}_2=2,7046$ (0,2)		

BIC (Bayesian information criterion), observou-se que o modelo COM-Poisson/Weibull apresentou menor estatística AIC e BIC. Assim, existem evidências de que este modelo foi o que melhor se ajustou ao conjunto de dados em estudo.

Considerando este modelo, observou-se que a média estimada para o número de tumores induzidos por animal é de 2,27 tumores para o grupo controle, enquanto para o

grupo tratamento é de 3,72 tumores. Já para a variável tempos de aparecimento dos tumores, observou-se um tempo médio de 134,10 dias para o grupo controle, e de 139,49 dias para o grupo tratamento. Ou seja, existem indícios de que o tratamento está estimulando o desenvolvimento dos tumores ao invés de preveni-lo.

Com o propósito de verificar se houve diferença significativa entre os grupos controle e tratamento, ou seja, de verificar se o tratamento utilizado para destruir as células afetadas surtiu efeito no número de tumores e seus tempos de detecção, realizou-se um teste da razão de verossimilhanças (TRV) entre os grupos avaliados. Este teste é utilizado quando o interesse é comparar o ajuste de dois modelos e, para isto, realiza-se a razão entre as funções de verossimilhança ou, equivalentemente, a razão entre o log das funções de verossimilhanças, dos modelos em estudo. A estatística teste é dada por:

$$TRV = -2\ln \left(\frac{\text{Verossimilhança do modelo nulo}}{\text{Verossimilhança do modelo alternativo}} \right) \quad (4.11)$$

Assintoticamente, a estatística TRV segue uma distribuição Qui-Quadrado com graus de liberdade iguais a diferença entre o número de parâmetros dos dois modelos. Os valores das estatísticas de teste para os modelos são apresentados nas Tabelas 4.12 e 4.13.

Tabela 4.12: *Teste da Razão de Verossimilhança (TRV) para a distribuição do número de tumores iniciados*

Modelo	$-2 \times LV$	TRV ($\eta_1 = \eta_2; \phi_1 = \phi_2$)	valor-p
COM-Poisson/Gamma	780,76	0,5504	0,7594
COM-Poisson/Weibull	780,16	0,5512	0,7591

Através do teste apresentado na Tabela 4.12 para comparar os parâmetros da distribuição COM-Poisson nos dois grupos em estudo, observou-se que não houve diferença significativa entre as distribuições COM-Poisson para o número de tumores induzidos por animal, ou seja, não existem evidências de que o tratamento avaliado teve efeito na destruição das células que foram iniciadas pela substância cancerígena, ao nível de 5% de significância. Através do teste apresentado observou-se que não houve diferença entre as dispersões dos dois grupos em estudo ($\phi_1 = \phi_2$). Esta é mais uma vantagem do modelo COM-Poisson em relação aos demais modelos apresentados, pois o modelo COM-Poisson leva em consideração a dispersão relativa a cada um dos grupos em estudo, tornando a análise mais próxima das situações reais encontradas, situação esta que não é considerada quando se utiliza o modelo de Poisson, por exemplo, que atribui a mesma variabilidade à ambos os grupos de animais.

Tabela 4.13: *Teste da Razão de Verossimilhança (TRV) para as distribuições dos tempos de detecção dos tumores*

Modelo	TRV ($\alpha_1 = \alpha_2; \beta_1 = \beta_2$)	valor-p
COM-Poisson/Gamma	0,6482	0,7231

Modelo	TRV ($b_1 = b_2; c_1 = c_2$)	p-valor
COM-Poisson/Weibull	0,8022	0,6695

Através dos testes apresentados na Tabela 4.13 para comparar os tempos de detecção dos grupos em estudo, verificou-se que não houve diferença significativa entre as distribuições dos tempos de aparecimento dos tumores nos dois grupos avaliados (controle e tratamento). Assim, pode-se dizer que existem evidências de que o tratamento não teve efeito na desaceleração dos tempos de aparecimento do tumor, ao nível de significância de 5%.

Outra forma de comparação entre os modelos que utilizamos é a estimação da média e da variância para cada distribuição em estudo. A Tabela 4.14 resume estas medidas descritivas para cada modelo, considerando os parâmetros estimados.

Tabela 4.14: *Estimativas da média e variância das variáveis M e T para os modelos apresentados.*

Modelos	Controle		Tratamento		Controle		Tratamento	
	$\hat{\mu}_m$	$\hat{\sigma}_m^2$	$\hat{\mu}_m$	$\hat{\sigma}_m^2$	$\hat{\mu}_t$	$\hat{\sigma}_t^2$	$\hat{\mu}_t$	$\hat{\sigma}_t^2$
Poisson/Gamma	1,26	1,26	2,32	2,32	158,02	6957,17	167,06	6756,47
Poisson/Weibull	1,21	1,21	2,17	2,17	148,93	5197,49	154,48	4382,00
BN/Gamma	1,54	6,01	2,41	10,63	165,95	7386,87	167,51	6590,91
BN/Weibull	1,48	5,58	2,26	9,58	157,46	5442,17	154,79	4178,25
COM/Gamma	2,26	2,44	3,69	6,01	134,07	4077,03	139,51	3556,81
COM/Weibull	2,27	2,46	3,72	6,03	134,10	3803,25	139,49	3094,80

Observa-se por meio da Tabela 4.14 que o grupo tratamento apresenta um número médio de tumores iniciados maior que o grupo controle para todos os modelos em estudo. Entretanto, as distribuições Poisson e Binomial Negativa não captam adequadamente a variabilidade em cada grupo. A distribuição COM-Poisson identifica a existência de uma variabilidade nos dados e que, além disso, tal variabilidade é verificada no grupo tratamento, cuja variância é aproximadamente o dobro da média. No grupo controle, a distribuição COM-Poisson identificou que a média está bem próxima da variância, ou

seja, não se observa dispersão significativa neste grupo. Isto pode ser verificado também por meio das estimativas dos parâmetros da distribuição COM-Poisson, em que para o grupo controle tem-se $\hat{\phi}$ próximo de 1 indicando que não há dispersão, e para o grupo tratamento tem-se $\hat{\phi}$ menor que 1, indicando sobredispersão nos dados para este grupo.

Como foi visto nas equações (2.20) e (2.21), pode-se calcular a média e a variância do número de tumores detectados $M^*(t)$, para qualquer t . Assim, considerando o modelo COM-Poisson/Weibull, tem-se que a média e a variância de $M^*(t)$ para o grupo controle, fica definida por:

$$\begin{aligned} E(M^*(t)) &= \mu_M F(t; \Psi) \\ &= 2,27F(t; b = 151, 37, c = 2, 30), \end{aligned} \quad (4.12)$$

e

$$\begin{aligned} Var(M^*(t)) &= [\sigma_M^2 - \mu_M]F^2(t; \Psi) + \mu_M F(t; \Psi) \\ &= (2,46 - 2,27)F^2(t; b = 151, 37, c = 2, 30) + 2,27F(t; b = 151, 37, c = 2, 30). \end{aligned} \quad (4.13)$$

Para o grupo tratamento, tem-se:

$$\begin{aligned} E(M^*(t)) &= \mu_M F(t; \Psi) \\ &= 3,72F(t; b = 156, 85, c = 2, 70), \end{aligned} \quad (4.14)$$

e

$$\begin{aligned} Var(M^*(t)) &= [\sigma_M^2 - \mu_M]F^2(t; \Psi) + \mu_M F(t; \Psi) \\ &= [6,03 - 3,72]F^2(t; b = 156, 85, c = 2, 70) + 3,72F(t; b = 156, 85, c = 2, 70). \end{aligned} \quad (4.15)$$

As Figuras 4.6 e 4.7 representam o comportamento da média e variância da variável $M^*(t)$, para os grupos controle e tratamento. Observa-se que tanto a média quanto a variância de $M^*(t)$ convergem para a média e variância de M , respectivamente, quando t cresce. Os grupos aparentam ter o mesmo número médio de tumores detectados no início do experimento, sendo que depois, aumentando o tempo, o grupo tratamento apresentou média de tumores detectados relativamente maior que o grupo controle.

Pode-se obter as funções de sobrevivência de longa duração para um processo destrutivo e suas respectivas proporções de curados estimadas. Utilizando as estimativas das distribuições de Poisson, Binomial negativa e COM-Poisson obtidas para a variável número

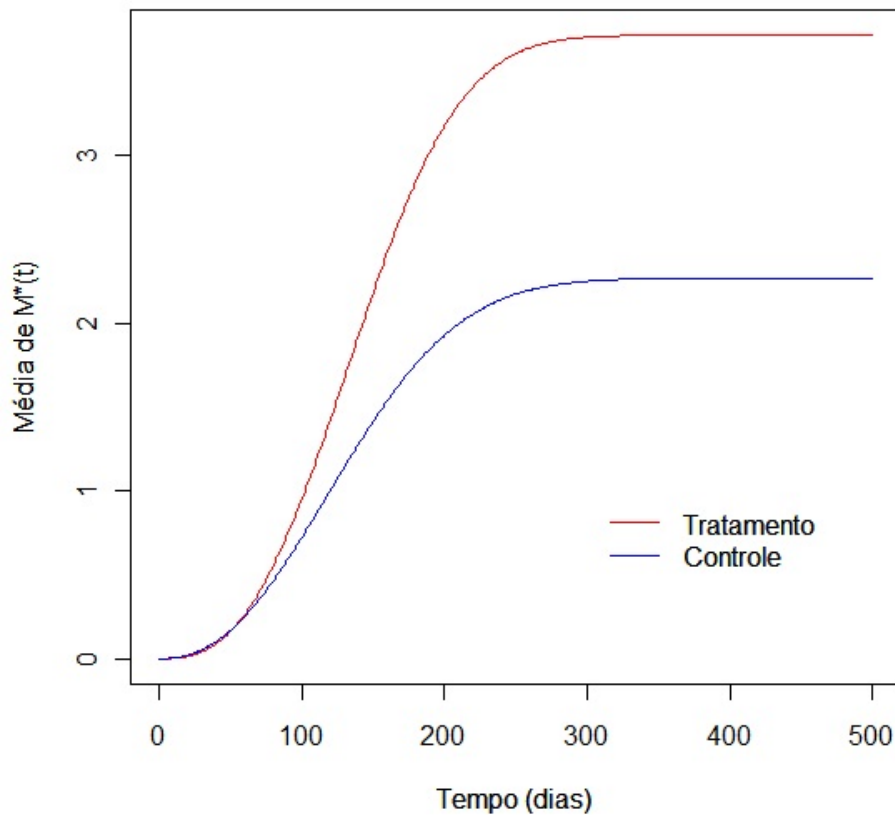


Figura 4.6: Estimativas da média da variável $M^*(t)$ para os grupos controle e tratamento, considerando o modelo COM-Poisson/Weibull.

de tumores induzidos, e considerando a distribuição Weibull para os tempos de detecção, as respectivas taxas de cura estimadas para os grupos controle e tratamento, são apresentadas nas Tabelas 4.15 e 4.16.

Tabela 4.15: Estimação da função de cura, $p_0(t; \theta, \Psi)$, e da taxa de cura, p_0 , em um processo destrutivo para o grupo controle.

Modelo Destrutivo	$p_0(t; \theta, \Psi)$	p_0
Poisson	$e^{-1,21F(t)}$	0,29
Binomial Negativa	$[1 + (1,85)(1,48)F(t)]^{-\frac{1}{1,85}}$	0,49
COM-Poisson	$\frac{Z[2,02(1-F(t));0,89]}{Z(2,02;0,89)}$	0,11

As comparações das funções de sobrevivência ajustadas mínimas com as estimativas de Kaplan-Meier são realizadas a seguir, para os grupos controle e tratamento, respectivamente. Observa-se por meio da Figura 4.8 que as curvas de sobrevivência mínimas

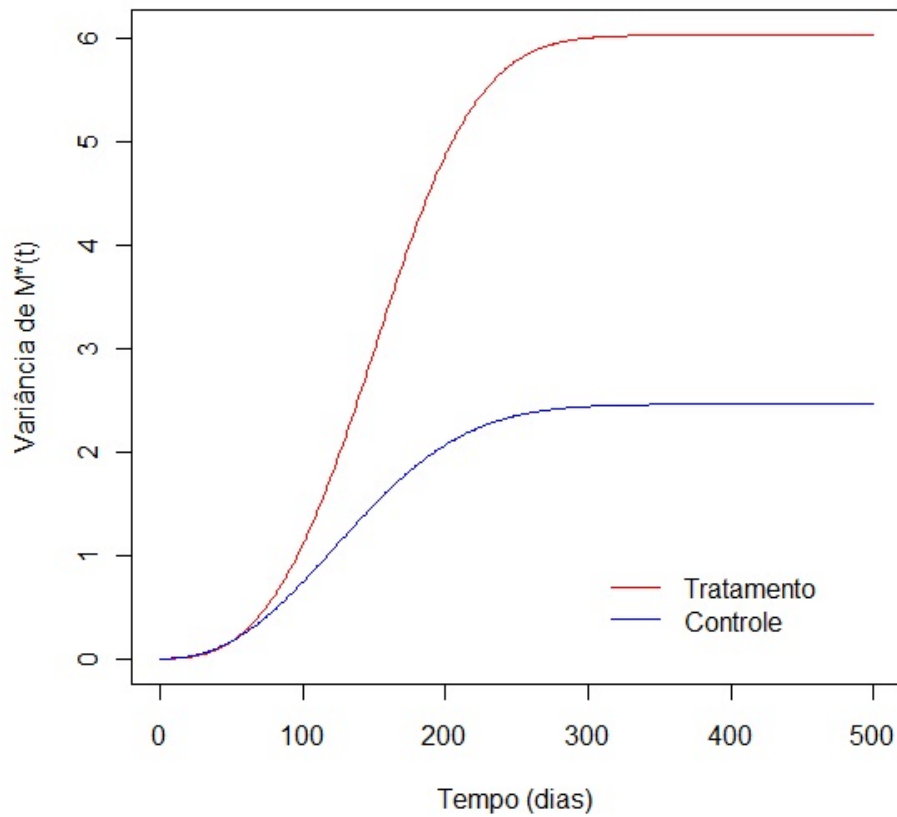


Figura 4.7: Estimativas da variância da variável $M^*(t)$ para os grupos controle e tratamento, considerando o modelo COM-Poisson/Weibull.

Tabela 4.16: Estimação da função de cura, $p_0(t; \theta, \Psi)$, e da taxa de cura, p_0 , em um processo destrutivo para o grupo tratamento.

Modelo Destrutivo	$p_0(t; \theta, \Psi)$	p_0
Poisson	$e^{-2,17F(t)}$	0,11
Binomial Negativa	$[1 + (1,42)(2,26)F(t)]^{-\frac{1}{1,42}}$	0,36
COM-Poisson	$\frac{Z[1,85(1-F(t));0,52]}{Z(1,85;0,52)}$	0,06

apresentam-se relativamente distantes das estimativas de KM. Isto decorre do fato de que o método não paramétrico de Kaplan-Meier não é o mais adequado para realizar comparações com as funções de sobrevivência mínimas, pois utiliza os tempos de vida dos animais enquanto que as funções de sobrevivência mínimas aqui apresentadas, utilizam os tempo máximos de detecção dos tumores, que não necessariamente representam o tempo de vida do animal. Outros métodos mais eficientes para avaliar a adequação de modelos em problemas de experimentos de quimioprevenção de câncer em animais são propostos

na literatura. Outra abordagem pra este problema seria analisá-lo utilizando modelos para eventos recorrentes. Esta metodologia que pode ser encontrada nos artigos de Lawless and Nadeau (1995) e Cheng Wang *et al.* (2001) considera situações mais realistas, podendo se ajustar com mais eficiência ao problema aqui trabalhado. A utilização da modelagem de eventos recorrentes para experimentos de quimioprevenção de câncer em animais é sugerida também como proposta para pesquisas futuras.

A fim de verificar o comportamento da função de curados no modelo, $p_0(t; \theta, \Psi)$, realizou-se uma representação gráfica que apresenta o comportamento da função $p_0(t; \theta, \Psi)$ ao longo do tempo.

Observa-se por meio da Figura 4.9 que conforme aumentamos o tempo, a função $p_0(t; \theta, \Psi)$ converge para a taxa de cura de longa duração, p_0 . Pode-se observar também que o grupo controle apresenta maior probabilidade de cura do que o grupo tratamento. O modelo COM-Poisson captou melhor a falta de efeito do tratamento (o que pode ser observado pela proximidade das curvas da função $p_0(t; \theta, \Psi)$ para os grupos controle e tratamento), sendo que sua eficiência quando comparado ao grupo controle, não foi significativa. Este modelo também detectou que o tratamento acelera o número de tumores desenvolvidos por animal e seus tempos de detecção, o que pode ser verificado com base na estimativa da taxa de cura próxima de zero. Assim, pode-se concluir que a flexibilidade do modelo COM-Poisson se adequou melhor ao conjunto de dados de experimentos com ratos tornando a análise mais próxima das situações reais encontradas, levando a obtenção de informações mais precisas e confiáveis.

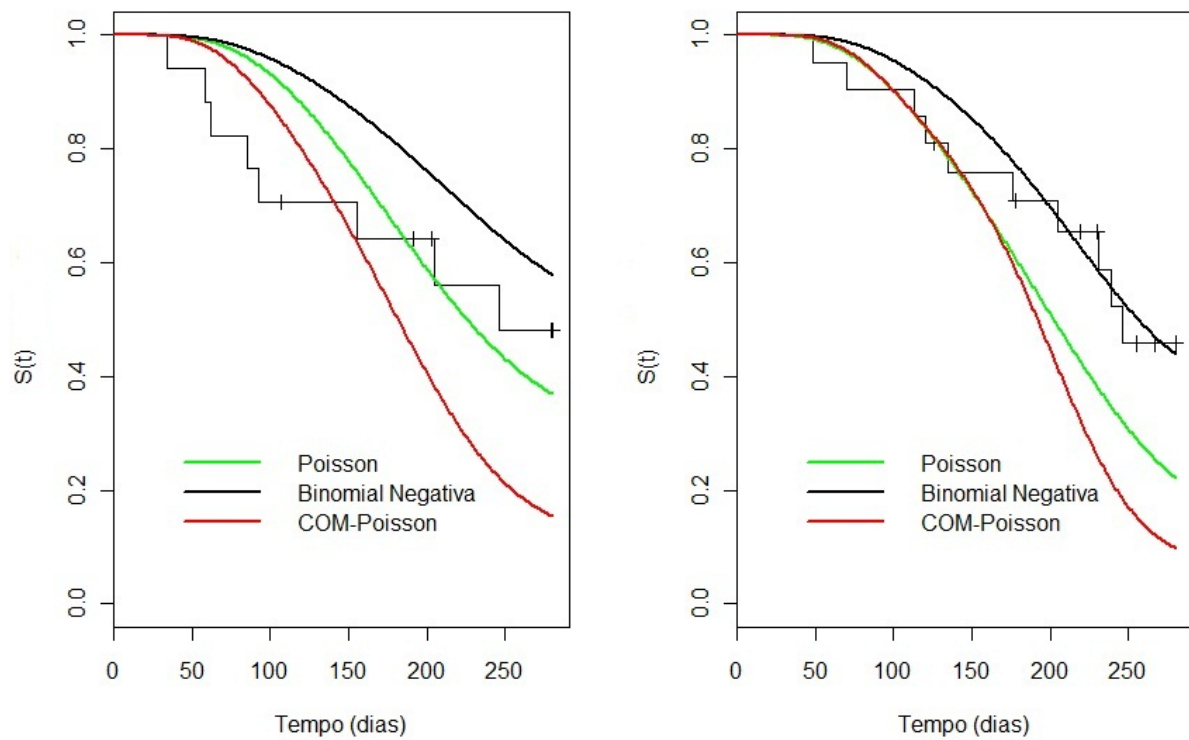


Figura 4.8: Distribuições ajustadas e Estimativas de Kaplan-Meier para os grupos Controle e Tratamento, respectivamente.

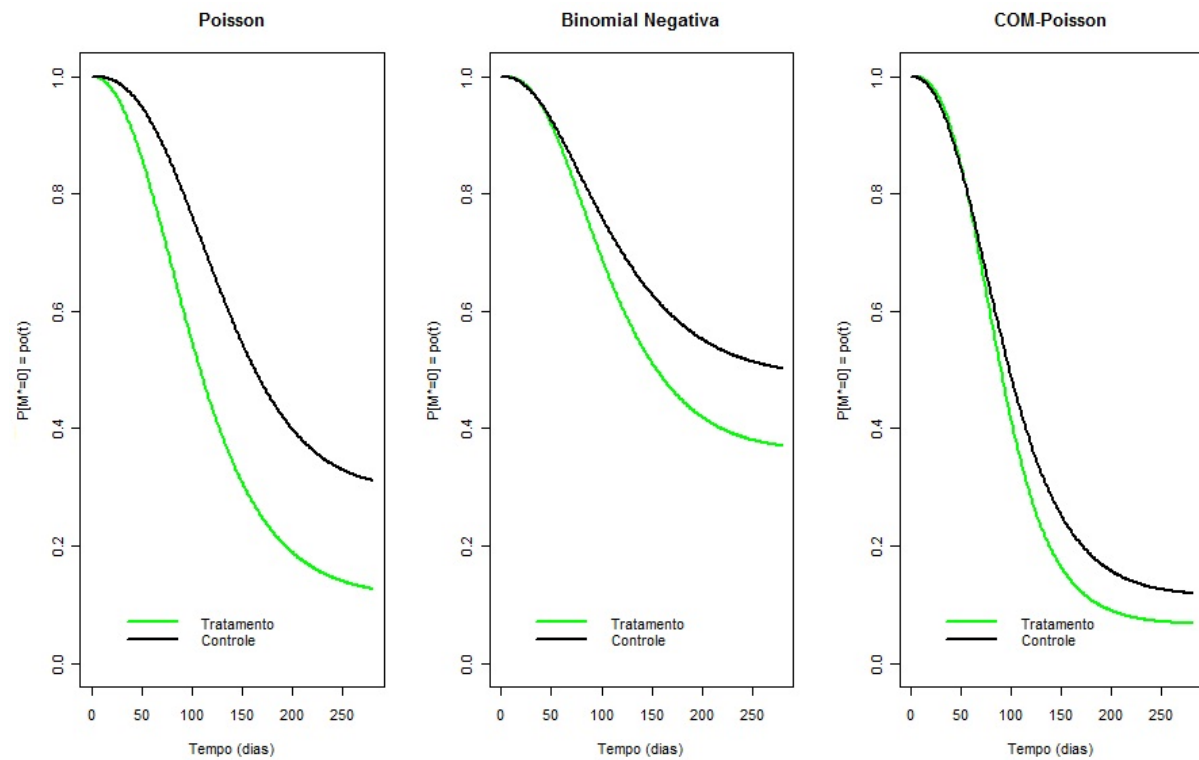


Figura 4.9: Estimativas das funções de cura para as distribuições Poisson, Binomial Negativa e COM-Poisson, avaliadas para os grupos controle e tratamento.

Capítulo 5

Conclusões

A análise apresentada é uma primeira tentativa de estudar criticamente o papel de uma avaliação mais específica com base na modelagem de agentes que são testados em experimentos de quimioprevenção de câncer em animais.

Neste estudo, explorou-se o uso do modelo estatístico proposto por Kokoska (1987) e generalizado posteriormente por Freedman (1993). Avaliou-se o modelo Kokoska/Freedman observando-o do ponto de vista de modelo destrutivo de longa duração. Os resultados obtidos por este método são sensíveis às formas paramétricas assumidas, especialmente para a distribuição do número de tumores induzidos por animal M . Além das distribuições usuais consideradas para esta variável, priorizou-se o uso da distribuição COM-Poisson, proposta por Conway e Maxwell em 1962 e reutilizada recentemente por Shumueli *et al.* (2005). Esta distribuição é uma generalização da distribuição de Poisson, pois contém um parâmetro adicional ϕ responsável pela modelagem de super-dispersão ou sub-dispersão, problemas que são frequentemente encontrados em dados discretos. Devido à sua ampla flexibilidade, a distribuição COM-Poisson vêm sendo frequentemente utilizada como ferramenta de análise de dados clínicos. As análises realizadas tanto no estudo de simulação, quanto no conjunto de dados reais, foram fundamentais para concluir sobre a eficiência da distribuição COM-Poisson em relação às demais para este tipo de dados.

Utilizou-se modelos de longa duração em experimentos de quimioprevenção de câncer em animais. Estimou-se a função de sobrevivência mínima e posteriormente a probabilidade de cura dos grupos controle e tratamento. Esta é mais uma ferramenta extremamente importante na análise deste tipo de dados, pois permite avaliar a probabilidade de cura de determinado tratamento, permitindo assim, verificar com mais precisão se o tratamento em questão é realmente eficiente ou não.

5.1 Considerações Finais

Nesta dissertação foi apresentada uma proposta para analisar dados de experimentos de quimioprevenção de câncer em animais da perspectiva de modelo destrutivo de longa duração. Este modelo foi apresentado em Rodrigues et al. (2011), considerando o mínimo dos tempos. Diferentemente da proposta de Rodrigues et al. (2011), esta dissertação se propôs a avaliar os máximos destes tempos, considerando que o animal atinge o evento de interesse quando o máximo dos tumores desenvolvidos são detectados. Outra diferença a se ressaltar é que aqui o número de tumores desenvolvidos por animal e seus tempos de detecção são observados durante o experimento, enquanto na proposta de Rodrigues et al. (2011), estas variáveis são consideradas latentes, ou seja, não observadas.

Priorizou-se o uso na distribuição COM-Poisson na análise, considerando sua ampla flexibilidade para ajustar problemas que envolvem sub ou super dispersão. Os resultados obtidos comprovaram sua eficiência para o conjunto de dados apresentado, em experimentos de quimioprevenção do câncer em animais.

5.2 Propostas para Pesquisas Futuras

Considerando a ampla aplicabilidade da modelagem proposta a problemas clínicos, muitas sugestões de pesquisas futuras podem ser avaliadas. Entre elas, podemos citar:

- Simular situações que considerem que os dados são provenientes de outras distribuições, como a Poisson e a Binomial Negativa, a fim de ajustar os modelos paramétricos apresentados e verificar o comportamento do modelo destrutivo COM-Poisson na adequação de diferentes situações que podem ocorrer na prática.
- Avaliar do ponto de vista bayesiano os resultados da análise de dados para experimentos de quimioprevenção do câncer em animais, considerando o modelo destrutivo de longa duração. Esta é uma abordagem extremamente interessante, visando o fato de que o parâmetro de dispersão ϕ afetou as propriedades assintóticas dos estimadores de máxima verossimilhança dos parâmetros do modelo proposto, principalmente na situação de sobredispersão, $\phi < 1$.
- Incluir covariáveis na estimação dos parâmetros do modelo destrutivo COM-Poisson, de modo a formular um modelo mais flexível que acomode tanto os efeitos de diferentes tratamentos, quanto os efeitos de animais no número de tumores induzidos e nos tempos de detecção destes tumores (Dunson & Dinse, 2000). Estes autores generalizam o modelo de Kokoska (1987) acomodando covariáveis e os efeitos de animais por meio de um modelo log-linear pra a média. Como proposta futura

sugere-se utilizar esta idéia de inclusão de covariáveis no contexto de modelo destrutivo COM-Poisson, visando tornar o modelo ainda mais próximo das situações reais encontradas.

- Utilizar a metodologia de análise de eventos recorrentes (Lawless (1995) e Cheng Wang *et al.* (2001)). Esta abordagem é uma forma alternativa de análise de dados de experimentos quimiopreventivos, que vem sendo muito utilizada e desenvolvida em muitos artigos científicos.

Apêndice A

Modelo Unificado de Sobrevivência

Dados de sobrevivência que envolvem a possibilidade de cura dos indivíduos são frequentemente encontrados em estudos médicos, principalmente em testes clínicos cujo interesse é avaliar a eficiência de novas drogas em pacientes com câncer. Para isto, experimentos são realizados com animais para investigar substâncias químicas efetivas na prevenção do câncer (Freedman et al., 1993).

Com o rápido avanço de técnicas e ferramentas computacionais a utilização da análise de sobrevivência cresceu significativamente nos últimos anos. "(...) Uma evidência qualitativa deste sucesso é o número de aplicações de análise de sobrevivência em medicina (...)"(Colosimo & Giolo, 2006), sendo que a inclusão da fração de cura no modelo vem sendo amplamente difundido em artigos recentes.

A variável de interesse em análise de sobrevivência é o "tempo de falha" das observações, que caracteriza o tempo até que ocorra o evento de interesse no estudo (este evento de interesse pode ser a morte de pacientes submetidos à um tratamento, a cura de uma doença, a falha de equipamentos eletrônicos, etc.).

O que difere a análise de sobrevivência das técnicas estatísticas clássicas utilizadas é a presença de censura nos dados (observações incompletas da resposta). A informação correspondente a um indivíduo que abandonou o tratamento é o tempo até seu último registro, por exemplo, este tempo é chamado de censura, pois a única informação que se sabe à respeito deste indivíduo é ele não havia apresentado o evento da falha até aquele instante. Os tempos de falha e de censuras são as informações que constituem a resposta em dados de sobrevivência. Geralmente, observa-se também um conjunto de covariáveis para as observações.

Existem diferentes mecanismos de censuras encontrados em dados de sobrevivência. Os principais tipos são apresentados a seguir.

- **Censura do Tipo I:**

É considerada nas situações em que se determina um prazo pré-estabelecido de tempo

para que o estudo seja finalizado. Assim, as observações que não falharam até o fim deste prazo são caracterizadas pelo tempo do fim do estudo, que é um tempo de censura. Se o experimento tem duração L , por exemplo, então o que se vai observar são os valores (t_i, δ_i) , onde,

$$\delta_i = \begin{cases} 1 & , \text{se } t_i \leq L \\ 0 & , \text{se } t_i > L \end{cases}$$

para $i = 1, \dots, n$.

- **Censura do Tipo II:**

É considerada nas situações em que se determina um número pré-estabelecido de falhas na amostra. Supondo, por exemplo, que o interesse é observar d falhas numa amostra de tamanho n , então os tempos observados dos $n - d$ indivíduos restantes são considerados tempos de censura do tipo II. Este tipo de censura costuma ser menos problemática para a análise estatística do que os outros tipos.

- **Censura Aleatória:**

É considerada nas situações em que o paciente deixa o estudo sem ter experimentado o evento de interesse. Esse tipo de censura pode ocorrer porque indivíduo abandona o estudo, morre por outra causa que não seja a que está sendo estudada ou permanece vivo até o fim do estudo. Desta forma, o instante da censura é aleatório, ou seja, difere para cada observação em estudo. Então o que se vai observar são os valores (t_i, δ_i) , onde,

$$\delta_i = \begin{cases} 1 & , \text{se } t_i \leq L_i \\ 0 & , \text{se } t_i > L_i \end{cases}$$

para $i = 1, \dots, n$.

- **Censura à Esquerda:**

É considerada nas situações em que o evento de interesse já aconteceu em um momento anterior ao início do estudo, ou seja, este mecanismo de censura é encontrado quando o tempo de registro do estudo é maior que o tempo de falha.

- **Censura Intervalar:**

É considerada nas situações em que os tempos observados dos indivíduos são medidos periodicamente, e portanto, quando o indivíduo apresenta o evento de interesse sabe-se apenas que a falha ocorreu em algum intervalo de tempo mas não se sabe seu tempo

exato. Como exemplo, pode-se citar a situação em que pacientes em um teste clínico ou estudo longitudinal têm avaliações periódicas e o tempo de falha do paciente só é conhecido dentro de algum intervalo de tempo. Os dados de sobrevivência caracterizados por este mecanismo de censura são também conhecidos, na literatura, por dados de censura intervalar.

Para representar os dados de sobrevivência, considera-se uma amostra de tamanho n , sendo que para cada observação $(i, i = 1, \dots, n)$ da amostra tem-se o par de informações (t_i, δ_i) , onde t_i representa os tempos observados para cada indivíduo e a variável δ_i indica se estes tempos são tempos de falha ou de censura, ou seja,

$$\delta_i = \begin{cases} 1 & , \text{ se } t_i \text{ representa tempo de falha da observação } i \\ 0 & , \text{ se } t_i \text{ representa tempo de censura da observação } i \end{cases}$$

Caso sejam consideradas covariáveis medidas para cada observação, o conjunto de dados de sobrevivência é caracterizado por (t_i, δ_i, x_i) , onde x_i representa uma ou mais covariáveis medidas para a i -ésima observação.

- **Modelo de Sobrevivência Usual**

A análise de dados cujo interesse é estudar o tempo até a ocorrência de um evento de interesse, ou tempo de falha, é tratada em uma área da estatística amplamente utilizada principalmente em estudos biomédicos, problemas de ciências sociais, de engenharia, conhecida por Análise de Sobrevivência. A necessidade de uma área específica para dados de sobrevivência vem do fato de que neste tipo de estudo é comum a presença de censuras, que são observações parciais da resposta.

Representação dos Tempos de Falha

Considere uma variável aleatória T que representa o tempo até a ocorrência de um evento de interesse, onde $T \geq 0$. A variável T pode ser expressa por várias funções matemáticas, dentre elas as mais utilizadas são:

- **Função Densidade de Probabilidade:**

Representa o limite da probabilidade de uma observação falhar num curto intervalo de tempo $[t, t + \Delta t]$ por unidade de tempo, e é definida como:

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P[t \leq T < t + \Delta t]}{\Delta t} \tag{A.1}$$

Pode-se definir também a função de distribuição acumulada de t , expressa por:

$$F(t) = P[T \leq t] = \int_0^t f(x)dx.$$

Na Figura A.1, são apresentadas algumas funções de densidades para uma variável t , não-negativa, considerando grupos de tratamentos simulados por diferentes funções de $f(t)$, com o propósito de ilustrar alguns possíveis comportamentos que esta função pode assumir.

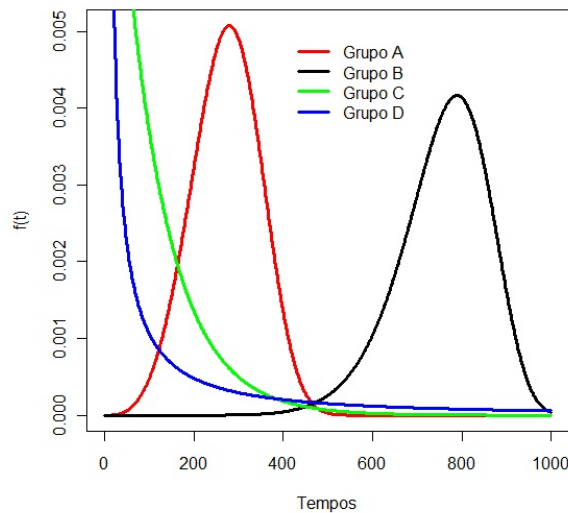


Figura A.1: Comportamento de $f(t)$ para quatro grupos fictícios.

A seguir são apresentadas algumas características importantes de $f(t)$:

1. $f(t) \geq 0$, para todo t ;
2. $\int_0^{\infty} f(t)dt = 1$;
3. $f(t) = \frac{d}{dt}F(t)$;

A variável aleatória T pode ser expressa também em termos de seus quantis, definidos probabilisticamente por $P(T \leq t_q) = q$, em que t_q representa a distribuição do q -ésimo quantil de T . Esta distribuição pode ser obtida por meio da expressão $t_q = F^{-1}(q)$. Utilizando a definição apresentada, pode-se obter facilmente a mediana (estatística muito utilizada em dados de sobrevivência) da distribuição de T , tomando o quantil $q = 0.5$ (Lawless, 2003).

- **Função de Sobrevivência:**

Utilizada amplamente para representar dados de sobrevivência, é caracterizada pela probabilidade de uma observação sobreviver além de um determinado tempo t , ou seja,

não apresentar o evento de interesse até t . Sua expressão probabilística é dada por:

$$S(t) = P[T \geq t] = \int_t^{\infty} f(x)dx = 1 - P[T < t] = 1 - F(t) \quad (\text{A.2})$$

Logo, $F(t)$ representa a probabilidade de uma observação não sobreviver até um certo tempo t . Pode-se observar na Figura A.2 o comportamento das funções de sobrevivência para as observações dos quatros grupos gerados (A, B, C e D). Caso estes grupos representem, por exemplo, tratamentos distintos para determinada doença, o tratamento representado pelo Grupo B apresenta tempos de vida superiores aos demais grupos na maior parte do acompanhamento. Outra observação importante é que para este grupo, praticamente todos os indivíduos ainda estão vivos após o tempo 400, enquanto que para os demais grupos, menos de 20 % dos indivíduos ainda estão vivos até este tempo.

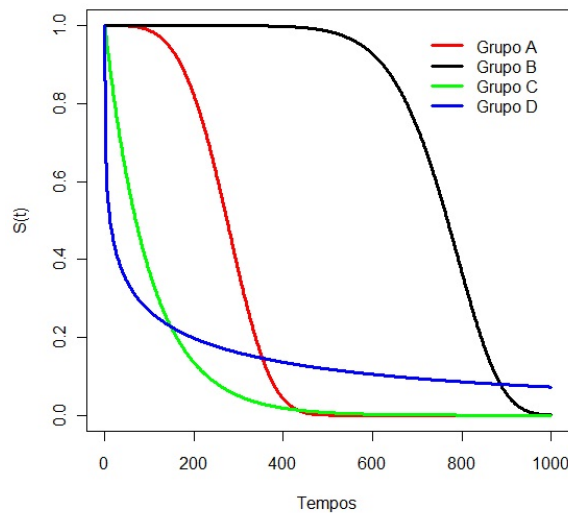


Figura A.2: *Comportamento de $S(t)$ para quatro grupos fictícios.*

A função de sobrevivência definida na Equação (A.2) possui algumas propriedades importantes, citadas a seguir:

1. $S(0) = 1$;
2. $S(t)$ é uma função contínua monótona não-crescente em t ;
3. $\lim_{t \rightarrow \infty} S(t) = 0$.

A terceira propriedade estabelece que ao longo do tempo a probabilidade de sobrevivência dos indivíduos tende a zero, ou seja, não considera que o tratamento forneça aos indivíduos uma possível cura, a fração de cura neste caso, é zero. Entretanto, muitos

estudos recentes (Yakovlev & Tsodikov, 1996) reconhecem a possibilidade da existência de indivíduos curados em tratamentos médicos ($S(\infty) > 0$) e, portanto, é de extrema importância que essa informação seja considerada na análise. Este conceito de possibilidade de cura na modelagem é discutido com mais detalhes na próxima seção.

- **Função de Risco:**

A função de risco, também conhecida por função de taxa de falha, é definida como a razão da probabilidade de uma observação falhar num intervalo de tempo $[t, t + \Delta t]$, dado que tenha sobrevivido até o tempo t , pelo comprimento do intervalo (Δt), ou seja,

$$h(t) = \frac{P[t \leq T < t + \Delta t | T \geq t]}{\Delta t} \quad (\text{A.3})$$

Fazendo $\lim_{\Delta t \rightarrow 0} h(t)$, a função $h(t)$ caracteriza o risco instantâneo da observação falhar no tempo t , dado que sobreviveu até aquele instante, e pode ser expressa por:

$$\begin{aligned} h(t) &= \lim_{\Delta t \rightarrow 0} \frac{P[t \leq T < t + \Delta t | T \geq t]}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{P[t \leq T < t + \Delta t, T \geq t]}{P[T \geq t] \Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{P[t \leq T < t + \Delta t]}{S(t) \Delta t} \\ &= \frac{1}{S(t)} \lim_{\Delta t \rightarrow 0} \frac{P[t \leq T < t + \Delta t]}{\Delta t} \\ &= \frac{f(t)}{S(t)} \end{aligned} \quad (\text{A.4})$$

As igualdades foram obtidas utilizando propriedades de probabilidade condicional e as definições das funções $f(t)$ e $S(t)$, apresentadas anteriormente.

Colosimo & Giolo (2006) referem-se à função de risco (ou taxa de falha) da seguinte forma:

"A função de taxa de falha é mais informativa do que a função de sobrevivência. Diferentes funções de sobrevivência podem ter formas semelhantes, enquanto que as respectivas funções de taxa de falha podem diferir drasticamente. Desta forma, a modelagem da função de taxa de falha é um importante método para dados de sobrevivência."

Desta forma, a função de risco vem sendo amplamente utilizada para representar dados de sobrevivência. Além disso, ela relata como o risco instantâneo das observações se comporta ao longo do tempo.

Na Figura A.3, são apresentados alguns comportamentos distintos para a função $h(t)$, para os dados fictícios gerados. Conforme a função de risco cresce, significa que a chance

da observação apresentar o evento da falha, aumenta ao longo do tempo. Quando a função decresce, significa que essa chance vai diminuindo com o passar do tempo. Quando a função permanece constante, significa que o risco não varia com o passar do tempo, sempre será o mesmo independente do tempo.

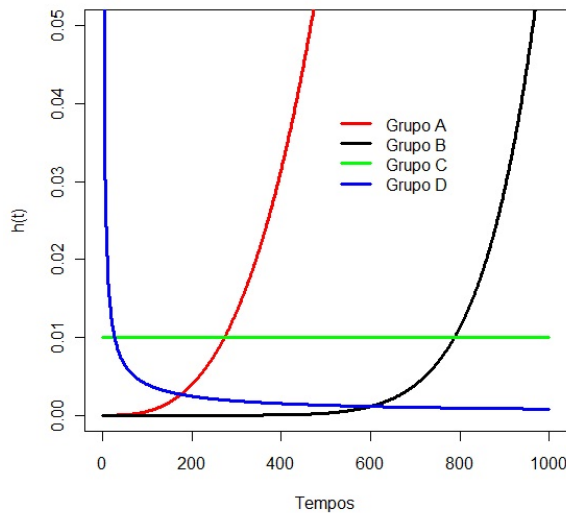


Figura A.3: Comportamento de $h(t)$ para quatro grupos fictícios.

A função de risco acumulada é denotada por:

$$H(t) = \int_0^t h(x)dx.$$

Conhecendo uma das funções apresentadas ($f(t)$, $S(t)$ ou $h(t)$), as outras podem ser obtidas facilmente por meio das seguintes expressões matemáticas:

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} (\log S(t)),$$

$$H(t) = \int_0^t h(x)dx = -\log S(t),$$

$$S(t) = \exp\{-H(t)\} = \exp\left\{-\int_0^t h(x)dx.\right\}.$$

Estimação Não-Paramétrica

Freqüentemente, na análise de dados de sobrevivência o interesse é obter informações para os tempos de falha de modo a comparar tratamentos distintos para uma certa doença. A primeira coisa a se fazer em qualquer análise estatística é resumir as informações contidas nos dados por meio de medidas descritivas. Entretanto, na presença de dados cen-

surados as técnicas usuais de estatística devem ser utilizadas com critério e tomando os cuidados necessários.

Na análise de dados de sobrevida, uma forma segura de realizar uma descrição dos dados é estimar a função de sobrevivência, e a partir dela obter as informações relevantes, como os tempos médios ou medianos de falha, os quantis, os percentis, etc.

Um estimador não-paramétrico para a função de sobrevivência muito utilizado na prática estatística de análise de dados com censuras é o popular Estimador de Kaplan-Meier, proposto por Kaplan e Meier em 1958. O número de intervalos de tempo que ele considera em sua obtenção, são tantos quanto o número de falhas distintas observadas nos dados.

Desta forma, considerando uma amostra de n observações onde k apresentaram o evento de interesse (falha) nos tempos $t_1 < t_2 < \dots < t_k$. Supondo que $d_j, j = 1, \dots, k$ representa o número de falhas no instante t_j , e que n_j representa o número de observações em risco no instante $t_j, j = 1, \dots, k$, o estimador de Kaplan-Meier para a função de sobrevivência $S(t)$, pode ser obtido por meio da seguinte expressão:

$$\hat{S}(t) = \prod_{j:t_j < t} \left(\frac{n_j - d_j}{n_j} \right) = \prod_{j:t_j < t} \left(1 - \frac{d_j}{n_j} \right) \quad (\text{A.5})$$

Algumas propriedades interessantes do estimador $\hat{S}(t)$, são:

1. Assintoticamente, ele é não viciado e converge para uma distribuição normal;
2. Ele é fracamente consistente;
3. Ele é o estimador de máxima verossimilhança de $S(t)$.

A demonstração do terceiro item pode ser observada no artigo original de Kaplan e Meier (1958), enquanto que a prova dos dois primeiros itens é verificada por Breslow e Crowley (1974), para certas condições de regularidade.

A variância assintótica do estimador $\hat{S}(t)$, necessária para a obtenção de intervalos de confiança e testes de hipóteses para a função de sobrevivência $S(t)$, pode ser definida pela seguinte expressão:

$$\widehat{Var}(\hat{S}(t)) = \left[\hat{S}(t) \right]^2 \sum_{j:t_j < t} \frac{d_j}{n_j(n_j - d_j)}. \quad (\text{A.6})$$

Desta forma, considerando distribuição normal assintótica para $\hat{S}(t)$, um intervalo de confiança aproximado, com $(1 - \alpha) \times 100\%$ de confiança para a função $S(t)$, é calculado

por meio da expressão:

$$IC[S(t), 100(1 - \alpha)\%] = \hat{S}(t) \pm z_{\alpha/2} \sqrt{\hat{Var}(\hat{S}(t))} \quad (\text{A.7})$$

No qual $\alpha/2$ representa o $\alpha/2$ -percentil da distribuição normal padrão.

Apêndice B

Modelos de Sobrevivência com Fração de Cura

Esta classe de modelos caracteriza uma metodologia que acrescenta à teoria usual de análise de sobrevivência a proporção de indivíduos que foram curados ao longo de um tratamento. Esta abordagem é conhecida também por Modelos de Sobrevivência de Longa Duração.

Avaliar esta "fração de cura" na modelagem de dados de sobrevivência é de extrema importância na análise estatística, pois esta situação representa com mais fidelidade as situações reais encontradas na prática médica, tornando assim, a análise mais confiável e precisa. A estimação desta quantidade também possibilita a comparação de tratamentos em teste, pois espera-se que tratamentos mais eficientes apresentem estimativas de proporção de curados maiores quando comparados a outros tratamentos. Diferentemente da teoria usual de análise de sobrevivência, esta metodologia considera que a "taxa de cura" de determinado tratamento assume um valor positivo, representado por $p_0 > 0$.

Considerar que a proporção de curados assume um valor positivo, faz com que haja mudanças nas funções apresentadas anteriormente para representar a variável do tempo de falha (função de sobrevivência, de densidade e de risco). Desta forma, dentro na metodologia de modelos de longa duração, essas funções serão chamadas de impróprias, e são especificadas a seguir.

• Função de Sobrevivência Imprópria

É muito utilizada para representar dados de sobrevivência que envolvem proporção de curados, e considera em sua obtenção uma flexibilização da função de densidade $f(t)$, que agora deve satisfazer a condição:

$$\int_0^{\infty} f(x)dx = \Gamma \leq 1.$$

Assim, a expressão probabilística da função de sobrevivência imprópria é dada por:

$$\begin{aligned}
 S_p(t) &= P[T \geq t] = \int_t^\infty f(x)dx = 1 - P[T < t] \\
 &= 1 - \left[\int_0^\infty f(x)dx - \int_t^\infty f(x)dx \right] \\
 &= 1 - \Gamma + \int_t^\infty f(x)dx
 \end{aligned} \tag{B.1}$$

onde $t \geq 0$.

A função de sobrevivência imprópria ou de longa duração $S_p(t)$ se restringe à função de sobrevivência tradicional $S(t)$ quando $\Gamma = 1$ e é dotada de algumas propriedades importantes, que são:

1. $S_p(0) = 1$;
2. $S_p(t)$ é uma função contínua monótona não-crescente em t ;
3. $\lim_{t \rightarrow \infty} S_p(t) = 1 - \Gamma$.

Pela terceira propriedade, observa-se que diferentemente de $S(t)$ a função de sobrevivência de longa duração não tende a zero conforme o tempo aumenta mas sim para uma proporção $1 - \Gamma$ que é considerada a fração de curados na amostra, ou seja, $p_0 = 1 - \Gamma$. Isto caracteriza a fração de indivíduos que ao fim do experimento não atingiram o evento de interesse (falha), podendo ter sobrevivido até o fim do estudo, ou terem morrido de outras causas que não as estudadas.

• **Função de Densidade Imprópria**

Esta função representa o comportamento da variável aleatória T , nas situações em que a fração de cura é considerada na análise de dados de sobrevivência e é também chamada de função de densidade de longa duração. Pode ser obtida por meio da seguinte relação:

$$\begin{aligned}
 f_p(t) &= -\frac{d}{dt} S_p(t) = -\frac{d}{dt} \left\{ 1 - \Gamma + \int_t^\infty f(x)dx \right\} \\
 &= -\frac{d}{dt} \int_t^\infty f(x)dx
 \end{aligned} \tag{B.2}$$

Quando se tem $\Gamma = 1$, a expressão para $f_p(t)$ é simplificada por:

$$f_p(t) = -\frac{d}{dt} \int_t^\infty f(x)dx = -\frac{d}{dt} (1 - F(t)) = f(t)$$

onde $f(t)$ é a função de densidade de probabilidade própria de T .

Características de $f_p(t)$:

1. $f_p(t) \geq 0$, para todo t ;
2. $\int_0^\infty f_p(t)dt = \int_0^t f_p(x)dx + \int_t^\infty f_p(x)dx \leq 1$;
3. $f_p(t) = -\frac{d}{dt}S_p(t)$.

• **Função de Risco Imprópria**

Baseando-se na definição de função de risco para o caso em que a fração de curados na amostra é zero ($h(t)$), como definido anteriormente, é possível estabelecer a função de risco num cenário em que a fração de cura é incluída no modelo. Assim, a função de risco imprópria, $h_p(t)$, é calculada por:

$$\begin{aligned} h_p(t) &= -\frac{d}{dt}\log(S_p(t)) = -\frac{1}{S_p(t)} \times \frac{d}{dt}S_p(t) \\ &= -\frac{1}{S_p(t)}(-f_p(t)) = \frac{f_p(t)}{S_p(t)} \end{aligned} \tag{B.3}$$

A função de risco imprópria é tão importante para a análise de modelos de longa duração quanto a função de risco própria para modelos de sobrevivência tradicionais e, portanto, sua aplicabilidade é vasta para essa abordagem.

Apêndice C

Estatísticas de Ordem

Teorema C.0.1. *Sejam $T_{i1}, T_{i2}, \dots, T_{im_i^*}$, m_i^* variáveis aleatórias independentes e identicamente distribuídas, com função densidade de probabilidade comum f e função de distribuição acumulada F . Considere os valores ordenados destas variáveis denotados por $T_{i(1)}, T_{i(2)}, \dots, T_{i(m_i^*)}$, onde $T_{i(1)} \leq T_{i(2)} \leq \dots \leq T_{i(m_i^*)}$. Então a função densidade de probabilidade conjunta das estatísticas de ordem é dada por:*

$$\begin{aligned} f_{T_{i(1)}, T_{i(2)}, \dots, T_{i(m_i^*)}}(t_{i1}, t_{i2}, \dots, t_{im_i^*}) &= m_i^*! f(t_{i1}) f(t_{i2}) \dots f(t_{im_i^*}) \\ &= m_i^*! \prod_{j=1}^{m_i^*} f(t_{ij}), \quad t_{i1} < t_{i2} < \dots < t_{im_i^*}. \end{aligned} \quad (\text{C.1})$$

Prova:

Sejam $T_{i1}, T_{i2}, \dots, T_{im_i^*}$, m_i^* variáveis aleatórias independentes e identicamente distribuídas, com função densidade de probabilidade comum f e função de distribuição acumulada F . Considere,

$$\begin{aligned} T_{i(1)} &= \text{menor valor dos } T_{i1}, T_{i2}, \dots, T_{im_i^*} \\ T_{i(2)} &= \text{segundo menor valor dos } T_{i1}, T_{i2}, \dots, T_{im_i^*} \\ &\vdots \\ T_{i(j)} &= \text{j-ésimo menor valor dos } T_{i1}, T_{i2}, \dots, T_{im_i^*} \\ &\vdots \\ T_{i(m_i^*)} &= \text{maior valor dos } T_{i1}, T_{i2}, \dots, T_{im_i^*} \end{aligned}$$

Os valores ordenados $T_{i(1)} \leq T_{i(2)} \leq \dots \leq T_{i(m_i^*)}$ são conhecidos como as *estatísticas de ordem* correspondentes às variáveis aleatórias $T_{i1}, T_{i2}, \dots, T_{im_i^*}$. Em outras palavras, $T_{i(1)}, T_{i(2)}, \dots, T_{i(m_i^*)}$ são os valores ordenados de $T_{i1}, T_{i2}, \dots, T_{im_i^*}$.

A função densidade conjunta das estatísticas de ordem é obtida observando-se que as estatísticas de ordem $T_{i(1)}, T_{i(2)}, \dots, T_{i(m_i^*)}$ assumirão os valores $t_{i1} \leq t_{i2} \leq \dots \leq t_{im_i^*}$ se, e somente se, para alguma permutação $(a_{i1}, a_{i2}, \dots, a_{im_i^*})$ de $(i1, i2, \dots, im_i^*)$

$$T_{i1} = t_{a_{i1}}, T_{i2} = t_{a_{i2}}, \dots, T_{im_i^*} = t_{a_{im_i^*}}.$$

Então para alguma permutação $(a_{i1}, a_{i2}, \dots, a_{im_i^*})$ de $(i1, i2, \dots, im_i^*)$, tem-se que,

$$\begin{aligned} P \left\{ t_{a_{i1}} - \frac{\epsilon}{2} < T_{i1} < t_{a_{i1}} + \frac{\epsilon}{2}, \dots, t_{a_{im_i^*}} - \frac{\epsilon}{2} < T_{im_i^*} < t_{a_{im_i^*}} + \frac{\epsilon}{2} \right\} &\approx \epsilon^{m_i^*} f_{T_{i1}, \dots, T_{im_i^*}}(t_{a_{i1}}, \dots, t_{a_{im_i^*}}) \\ &= \epsilon^{m_i^*} f(t_{a_{i1}}) \dots f(t_{a_{im_i^*}}) \\ &= \epsilon^{m_i^*} f(t_{i1}) \dots f(t_{im_i^*}) \end{aligned}$$

Vemos que para $t_{i1} < t_{i2} < \dots < t_{im_i^*}$, temos:

$$P \left\{ t_{i1} - \frac{\epsilon}{2} < T_{i(1)} < t_{i1} + \frac{\epsilon}{2}, \dots, t_{im_i^*} - \frac{\epsilon}{2} < T_{i(m_i^*)} < t_{im_i^*} + \frac{\epsilon}{2} \right\} \approx m_i^*! \epsilon^{m_i^*} f(t_{i1}) \dots f(t_{im_i^*})$$

Dividindo por $\epsilon^{m_i^*}$ e fazendo $\epsilon \rightarrow 0$, temos que:

$$\begin{aligned} f_{T_{i(1)}, T_{i(2)}, \dots, T_{i(m_i^*)}}(t_{i1}, t_{i2}, \dots, t_{im_i^*}) &= m_i^*! f(t_{i1}) f(t_{i2}) \dots f(t_{im_i^*}) \\ &= m_i^*! \prod_{j=1}^{m_i^*} f(t_{ij}), \quad t_{i1} < t_{i2} < \dots < t_{im_i^*}. \end{aligned} \quad (\text{C.2})$$

A equação C.2 pode ser interpretada da seguinte forma: para que o vetor $(T_{i(1)}, T_{i(2)}, \dots, T_{i(m_i^*)})$ assumira o valor $(t_{i1}, t_{i2}, \dots, t_{im_i^*})$, é necessário e suficiente que o vetor $(T_{i1}, T_{i2}, \dots, T_{im_i^*})$ seja igual a uma das $m_i^*!$ permutações de $(t_{i1}, t_{i2}, \dots, t_{im_i^*})$. Segundo Ross (1997), como a probabilidade de que $(T_{i1}, T_{i2}, \dots, T_{im_i^*})$ seja igual a alguma dada permutação de $(t_{i1}, t_{i2}, \dots, t_{im_i^*})$ é exatamente $f(t_{i1}) f(t_{i2}) \dots f(t_{im_i^*})$, obtemos o resultado apresentado na equação C.2.

Referências Bibliográficas

Berkson, J. & Gage, R. P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, 47, 501-515.

Bolfarine, H., Rodrigues, J. & Achcar, J. A. (1991). *Análise de Sobrevida*. II Escola de Modelos de Regressão, Associação Brasileira de Estatística, Universidade Federal do Rio de Janeiro-RJ.

Breslow, N. & Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *The Annals of Statistics*, 2, 437-453.

Cancho, V. G. & Bolfarine, H. (2001). Modeling the presence of immunes by using the exponentiated-Weibull model. *Journal of Applied Statistics*, 28, 659-671.

Cancho, V. G., de Castro, M. & Rodrigues, J. (2010). A Bayesian analysis of the Conway-Maxwell-Poisson cure rate model. *Statistical Papers*, 53, 165-176.

Casella, G. & Berger, R. L. (2001). *Statistical Inference*. Duxbury/Thompson Learning, Pacific Grove, CA, second edition.

Chen, M.-H., Harrington, D. P. & Ibrahim, J. G. (2002). Bayesian Cure Rate Models for Malignant Melanoma: A Case-Study of Eastern Cooperative Oncology Group Trial E1690. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 51, 135-150.

Chen, M.-H. & Ibrahim, J. G. (2001). Maximum likelihood methods for cure rate models with missing covariates. *Biometrics*, 57, 43-52.

Chen, M.-H. & Ibrahim, J. G. & Sinha, D. (1999). A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, 94, 909-919.

Chen, M.-H. & Ibrahim, J. G. & Sinha, D. (2002). Bayesian inference for multivariate survival data with a cure fraction. *Journal of Multivariate Analysis*, 80, 101-126.

- Cheng W. M., Qin J. & Chiang C. T. (2001). Analyzing Recurrent Event Data With Informative Censoring. *Journal of the American Statistical Association*, vol. 96, nº 455.
- Colosimo, E. A. & Giolo, S. R. (2006). *Análise de Sobrevivência Aplicada*. Edgard Blücher, São Paulo-SP.
- Conway, R. W. & Maxwell, W. L. (1962). A queuing model with state dependent services rates. *The Journal of Industrial Engineering* XII, 132136.
- Dunson, D. B. & Dinse, G. E. (2000). Distinguishing Effects on Tumor Multiplicity and Growth Rate in Chemoprevention Experiments. *Biometrics*, 56, 1068-1075.
- Feller, W. (1968). *An Introduction to Probability Theory and its Applications*, volume I. Wiley, New York, NY, third edition.
- Freedman, L. S., Midthume, D. N., Brown, C. C., Steele, V. & Kelloff, G. J. (1993). Statistical Analysis of Animal Cancer Chemoprevention Experiments. *Biometrics*, vol. 49, 259-268.
- Gu, Y., Sinha, D., Banerjee, S. (2010). Analysis of cure rate survival data under proportional odds model. *Lifetime Data Analysis*, 17, 123-134.
- Ibrahim, J. G., Chen, M.-H. & Sinha, D. (2001). *Bayesian Survival Analysis*. Springer, New York, NY.
- INCA (2012). Instituto Nacional do Câncer. Homepage: www.inca.gov.br.
- Kalbfleisch, J. D. & Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. Wiley, New York, NY, second edition.
- Kaplan, E. L. & Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, 53, 457-481.
- Kokonendji, C. C., Mizère, D. & Balakrishnan, N. (2008). Connections of the Poisson weight function to overdispersion and underdispersion. *Journal of Statistical Planning and Inference*, 138, 1287-1296.
- Kokoska, S. M. (1987). The analysis of cancer chemoprevention experiments. *Biometrics*, 43, 525-534.

- Lawless, J. F. & Nadeau, C. (1995). Some Simple Robust Methods for the Analysis of Recurrent Events. *Technometrics*, 37, 158-168.
- Louzada-Neto, F., Mazucheli, J. & Achcar, J. A. (2002). *Introdução à Análise de Sobre-vivência e Confiabilidade*. III Jornada Regional de Estatística e II Semana de Estatística, Universidade Estadual de Maringá, Maringá-PR.
- Maller, R. A. & Zhou, S. (1995). Testing for the presence of immune or cured individuals in censored survival data. *Biometrics*, 51, 1197-1205.
- Maller, R. A. & Zhou, S. (1996). *Survival Analysis with Long-Term Survivors*. Wiley, New York, NY.
- Migon, H. & Gamerman, D. (1999). *Statistical Inference: An Integrated Approach*. Arnold, London.
- R Development Core Team. *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2009.
- Rodrigues, J., Cancho, V. & de Castro, M. (2008). *Teoria Unificada de análise de sobre-vivência*. Associação Brasileira de Estatística, São Paulo-SP.
- Rodrigues, J., de Castro, M., Cancho, V. G. & Balakrishnan, N. (2009). COM-Poisson cure rate survival models and an application to a cutaneous melanoma data. *Journal of Statistical Planning and Inference*, 139, 3605-3611.
- Rodrigues, J., Cancho, V., de Castro, M. & Balakrishnan, N. (2010). A Bayesian destructive weighted Poisson cure rate model and an application to a cutaneous melanoma data. *Statistical Methods in Medical Research*, John Wiley & Sons, Ltd.
- Rodrigues, J., de Castro, M., Balakrishnan, N. & Cancho, V. (2011). Destructive weighted Poisson cure rate models. *Lifetime Data Analysis*, 17, 333-346.
- Rodrigues, J., Cancho, V., de Castro, M. & Louzada-Neto, F. (2008). On the unification of long-term survival models. *Statistics & Probability Letters*, 39, 753-759.
- Ross, G. J. S. & Preece, D. A. (1985). The negative binomial distribution. *The Statistician*, 34, 323-336.
- Ross, S. M. (1997). *A First Course in Probability*. Prentice Hall College, fifth edition.

Shmueli, G., Minka, T. P., Kadane, J. B., Borle, S. & Boatwright, P. (2005). A useful distribution for fitting discrete data: Revival of the Conway-Maxwell-Poisson distribution, *Journal of the Royal Statistical Society Ser. C*, 54, 127-142.

Tsodikov, A. D., Ibrahim, J. G. & Yakovlev, A. Y. (2003). Estimating cure rates from survival data: an alternative to two-component mixture models. *Journal of the American Statistical Association*, 98, 1063-1078.

Wang, M.C., Qin, J. & Chiang, C.T. (2001). Analyzing recurrent event data with informative censoring. *Journal of the American Statistical Association*, 96, 1057-1065.

Yakovlev, A. Y. & Tsodikov, A. D. (1996). *Stochastic Models of Tumor Latency and Their Biostatistical Applications*. World Scientific, Singapore.