
Modelos de sobrevivência para estimação do
período de latência do Câncer

Bárbara Beltrame Bettim

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA
UFSCar-USP

BÁRBARA BELTRAME BETTIM

**MODELOS DE SOBREVIVÊNCIA PARA ESTIMAÇÃO DO
PERÍODO DE LATÊNCIA DO CÂNCER**

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

Orientador: Prof. Dr. Francisco Louzada Neto

**São Carlos
Dezembro de 2016**

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA
UFSCar-USP

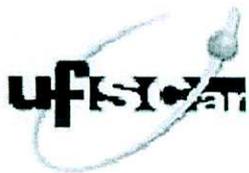
BÁRBARA BELTRAME BETTIM

**SURVIVAL MODELS TO ESTIMATE THE LATENCY PERIOD OF
CANCER**

Master dissertation submitted to the Departamento de Estatística – DEs-UFSCar and to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP, in partial fulfillment of the requirements for the degree of the Master joint Graduate Program in Statistics.

Advisor: Prof. Dr. Francisco Louzada Neto

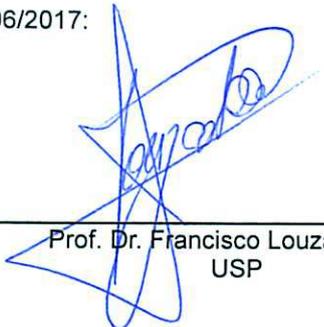
**São Carlos
December 2016**



UNIVERSIDADE FEDERAL DE SÃO CARLOS
Centro de Ciências Exatas e de Tecnologia
Programa Interinstitucional de Pós-Graduação em Estatística

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a defesa de dissertação de mestrado da candidata Bárbara Beltrame Bettim realizada em 29/06/2017:



Prof. Dr. Francisco Louzada Neto
USP



Profa. Dr. Maria Aparecida de Paiva Franco
USP

Prof. Dr. Paulo Henrique Ferreira da Silva
UFBA

Certifico que a sessão de defesa foi realizada com a participação à distância do membro Prof. Dr. Paulo Henrique Ferreira da Silva e, depois das arguições e deliberações realizadas, o participante à distância está de acordo com o conteúdo do parecer da comissão examinadora redigido no relatório de defesa do(a) aluno(a) Bárbara Beltrame Bettim.



Prof. Dr. Francisco Louzada Neto
Presidente da Comissão Examinadora
USP

Este trabalho é dedicado à minha família.

AGRADECIMENTOS

Agradeço, primeiramente, à minha família pelo apoio em todas as decisões da minha vida, por entenderem minha falta enquanto eu morava fora e por me darem amor incondicional.

Em especial aos meus pais Elisabete e Edilson, e aos meus pais de coração Paulo e Luciana.

Aos meus avôs e avós, tios e primos pelo amor, incentivo e apoio.

À minha irmã Letícia por ser minha companheira de vida, companheira de quarto e meu maior amor.

Ao Alexandre por todo o carinho, ajuda e amor neste anos e por ser meu exemplo de estatístico e maior companheiro. Ele é responsável por grande parte dessa conquista e sem ele não sei se teria conseguido.

Aos meus companheiros desde a graduação até o fim do mestrado: Nicholas e Karine. Só nós sabemos tudo o que passamos juntos e eu fico muito feliz por tudo que conquistamos. Vocês moram no meu coração.

Às minhas melhores amigas: Larissa, Milena, Gabriela, Ana Carolina, Bruna e Denise que estão sempre comigo independente do tempo e da distância.

Aos meus companheiros de trabalho da AEGEA pelo apoio e motivação.

Aos membros das bancas de Qualificação e Defesa deste trabalho: Dra. Maria Aparecida Paiva Franco, Dr. Domingos Alves e Dr. Paulo Henrique Ferreira pelas sugestões.

Ao meu orientador Prof. Dr. Francisco Louzada Neto pela orientação e ensinamentos.

À Prof. Dra. Gleici Perdoná, que além de professora e coorientadora foi uma grande amiga. Obrigada por todos os ensinamentos.

A todos os professores do DEs-UFSCar e do ICMC-USP com quem tive a oportunidade de aprimorar meus conhecimentos em Estatística.

A todos que contribuíram e torceram por mim para que mais essa conquista em minha vida fosse concretizada.

Muito obrigada!

*Another turning point, a fork stuck in the road
Time grabs you by the wrist, directs you where to go
So make the best of this test, and don't ask why
It's not a question, but a lesson learned in time...*

RESUMO

BETTIM, B. B. **Modelos de sobrevivência para estimação do período de latência do Câncer**. 2016. 67 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2016.

O câncer é responsável por aproximadamente 13% de todas as mortes no mundo, sendo que elas ocorrem principalmente em pessoas que são diagnosticadas tardiamente e em estágios avançados. Devido às suas características devastadoras e à prevalência cada vez maior da doença, é inquestionável a necessidade de investigações e pesquisas constantes na área, no sentido de aprimorar a detecção precoce e auxiliar em sua prevenção e tratamento. Dentre as diversas abordagens existentes, uma alternativa é a criação de técnicas para estimar o período de crescimento "silencioso" do câncer, que significa conhecer o momento do início do processo cancerígeno, também chamado de período de latência. A partir da revisão da literatura realizada, foi verificada uma escassez de modelos que estimam a latência do câncer, indicando a necessidade de estudo sobre o tema. Nesse contexto, métodos de análise de sobrevivência surgem como uma ferramenta útil para a construção desses modelos. No presente trabalho, é apresentada uma revisão de um modelo já existente, bem como sua formulação e métodos de estimação. Além disso, apresenta-se uma aplicação em um conjunto de dados reais e uma discussão dos resultados obtidos. Foi identificada a necessidade da formulação de um novo modelo, visto que o método estudado apresenta algumas limitações. Com isso são apresentadas 3 alternativas de modelos que solucionam os pontos apresentados na discussão, com respectivas aplicações.

Palavras-chave: Período de Latência, Análise de Sobrevivência, Câncer.

ABSTRACT

BETTIM, B. B. **Survival models to estimate the latency period of cancer**. 2016. 67 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2016.

Cancer is responsible for about 13% of all deaths in the world occurring mainly in people who are late diagnosed and in advanced stages. Due to its devastating characteristics and the growing prevalence of the disease, it is unquestionable the need of constant investigation and research in this area, in order to improve the early detection and to help in its prevention and treatment. Among the existing approaches, one alternative is the creation of techniques to estimate the "silent" growth period of cancer, which means to know the beginning moment of the carcinogen period, also known as latency period. In a literature review, it was found an shortage of models that estimate the latency of cancer, indicating the need of study about this theme. In this context, survival analysis methods appear as an useful tool to build these models. In this study, a review of an existing model is presented, as well as its formulation and estimation methods. Furthermore, an application on real data and a discussion of the obtained results are made. As a result, it was identified the need to formulate a new model, because of the limitations of the studied one. We present 3 alternative models that solve the points presented in the discussion, with applications.

Keywords: Latency period, Survival Analysis, Cancer.

LISTA DE ILUSTRAÇÕES

Figura 1 – Exemplo de curva de sobrevivência indicando os períodos observados e não observados	24
Figura 2 – Variações da distribuição Weibull	28
Figura 3 – Sobrevivência empírica obtida pelo método de Kaplan Meier e bandas de confiança	38
Figura 4 – Estimação do período aproximado do câncer de fígado	39
Figura 5 – Sobrevivência empírica obtida pelo método de Kaplan Meier e bandas de confiança considerando censura	46
Figura 6 – Estimação do período aproximado do câncer de fígado	47
Figura 7 – Boxplots das estimativas dos parâmetros para diferentes valores de β iniciais	49
Figura 8 – Verossimilhança para o parâmetro λ	49
Figura 9 – Verossimilhança para o parâmetro β	50
Figura 10 – Verossimilhança para o parâmetro u	50
Figura 11 – Kaplan Meier para Okuda	57

LISTA DE TABELAS

Tabela 1 – Resultados do ajuste do modelo de regressão para λ e β descrito em subseção 2.2.1	38
Tabela 2 – Resultados do modelo de regressão para a e b descrito em subseção 2.2.1	38
Tabela 3 – Estimativas dos parâmetros pelo Método Exato	40
Tabela 4 – Resultados do modelo de regressão para λ e β descrito em subseção 2.2.1 com censura	45
Tabela 5 – Resultados do modelo de regressão para a e b descrito em subseção 2.2.1 com censura	45
Tabela 6 – Estimativas dos parâmetros pelo Método Exato e com censura	47
Tabela 7 – Comparação das estimativas do parâmetro de latência pelos Métodos Aproximado e Exato com e sem censura	48
Tabela 8 – Tabela de classificação de Okuda	56
Tabela 9 – Tabela de frequências para Okuda	56
Tabela 10 – Testes de log-rank e Wilcoxon para Okuda com censura	56
Tabela 11 – Estimativas dos parâmetros	58
Tabela 12 – Estimativas das fragilidades para os estadiamentos	58
Tabela 13 – Estimativas dos parâmetros	61
Tabela 14 – Estimativas das latências (em anos) para os estadiamentos	61
Tabela 15 – AIC dos modelos	64

SUMÁRIO

1	INTRODUÇÃO	23
1.1	Análise de Sobrevivência	25
1.2	Revisão Bibliográfica	28
1.3	Objetivos	31
1.3.1	<i>Gerais</i>	31
1.3.2	<i>Específicos</i>	31
1.4	Considerações Finais	31
2	MODELO DE LATÊNCIA	33
2.1	O Modelo	33
2.2	Estimação dos Parâmetros	34
2.2.1	<i>Método Aproximado</i>	35
2.2.2	<i>Método Exato</i>	36
2.3	Ilustração dos Métodos	37
2.3.1	<i>Método Aproximado</i>	37
2.3.2	<i>Método Exato</i>	39
2.4	Discussão	40
2.5	Considerações Finais	41
3	MODELO 1 - MODELO DE LATÊNCIA CONSIDERANDO CEN- SURA	43
3.1	O Modelo	43
3.2	Estimação dos Parâmetros	43
3.3	Ilustração dos Métodos com Censura	45
3.3.1	<i>Método Aproximado</i>	45
3.3.2	<i>Método Exato</i>	46
3.4	Estudo de Simulação	47
3.5	Considerações Finais	51
4	MODELO 2 - MODELO COM FRAGILIDADE COMPARTILHADA	53
4.1	O Modelo	53
4.2	Formulação do Modelo	54
4.3	Estimação dos Parâmetros	54
4.4	Aplicação do Modelo	55

4.5	Considerações Finais	58
5	MODELO 3 - MODELO DE LATÊNCIA COM COVARIÁVEL	59
5.1	O Modelo	59
5.2	Estimação dos Parâmetros	59
5.3	Aplicação do Modelo	60
5.4	Considerações Finais	61
6	CONSIDERAÇÕES FINAIS	63
6.1	Propostas Futuras	64
	REFERÊNCIAS	65

INTRODUÇÃO

O câncer é uma doença complexa que acontece quando há o crescimento desordenado de células que podem invadir tecidos e órgãos do corpo ([Inca \(2014\)](#)). Devido à esse crescimento, há a formação de tumores, que são o acúmulo de células cancerígenas. Essa doença pode afetar pessoas de todas as idades, porém o risco de desenvolvimento cresce conforme a idade aumenta ([Cancer-Research-UK \(2014\)](#)).

Suas causas variam de acordo com o tipo de câncer, podendo ser causas externas ou do próprio organismo. As causas externas correspondem à maioria dos casos e estão relacionadas ao meio-ambiente ou à hábitos sociais como cigarro, exposição excessiva ao sol, alimentos, álcool, entre outros. Por outro lado, as causas do próprio organismo são, geralmente, ligadas à capacidade do organismo de se defender e são geneticamente pré-determinadas. Essas causas são conhecidas como fatores de risco e o surgimento do câncer depende da intensidade e duração da exposição das células aos agentes causadores ([Inca \(2014\)](#)).

Em geral, a taxa de mortalidade associada ao câncer está relacionada com seu tipo e com o estágio em que a doença se encontra na fase de diagnóstico. Essa doença é responsável por aproximadamente 13% de todas as mortes no mundo, sendo que os tipos que mais matam são o câncer de pulmão, estômago, fígado, cólon e mama no caso das mulheres. Ainda, essas mortes ocorrem principalmente em pessoas que são diagnosticadas tardiamente e em estágios avançados ([World-Health-Statistics \(2014\)](#)).

Devido às suas características devastadoras e à prevalência cada vez maior da doença, é inquestionável a necessidade de investigações e pesquisas constantes na área, no sentido de aprimorar a detecção precoce e auxiliar em sua prevenção e tratamento. No geral, o tempo real em que as células cancerígenas se desenvolveram antes do diagnóstico é desconhecido, visto que os sintomas característicos da doença podem demorar meses e até anos para aparecer. Além disso, fatores de risco podem agir para iniciar a carcinogênese e dez ou mais anos podem passar entre a exposição à fatores externos e a detecção do câncer. Dessa forma, a detecção precoce

é uma das armas mais promissoras para evitar o avanço do câncer, enquanto o mesmo ainda é curável e está restrito a apenas um local, evitando não apenas a mortalidade mas também reduzindo os custos de tratamento (Nadler e Zurbenko (2012)).

Nesse contexto, o desenvolvimento de técnicas para estimar o período de crescimento “silencioso” do câncer, que significa conhecer o momento do início do processo cancerígeno (também chamado de período de latência), é de grande importância.

Para contextualizar o problema, a Figura 1 apresenta um diagrama do que é o tempo latente. O primeiro evento a ocorrer é a inicialização da doença, ou seja, o momento em que ela se manifestou, mesmo que ainda não diagnosticada, enquanto que o segundo apresenta quando se inicializa o registro, isto é, o diagnóstico da doença. Nesse caso, o período de latência é o tempo entre o início da doença e o diagnóstico.

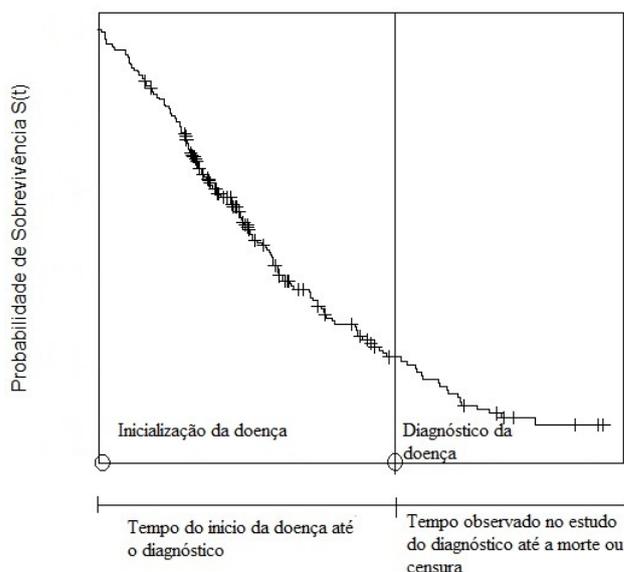


Figura 1 – Exemplo de curva de sobrevivência indicando os períodos observados e não observados

Fonte: Elaborada pelo autor.

No entanto, o início da doença nem sempre é bem definido ou seu fator de risco não é bem conhecido ou determinado. Também, a definição do que seria o período latente pode ser diferente dependendo do tipo de câncer. Nos casos em que há um fator de risco bem definido, o período de latência é dado pelo tempo entre a exposição ao fator de risco e o diagnóstico da doença. Por outro lado, quando esse fator de risco não existe ou não é bem definido, o período de latência pode ser o tempo entre o aparecimento da primeira célula cancerígena e o diagnóstico, ou o tempo entre o aparecimento do tumor e o diagnóstico, entre outros. Há ainda os casos em que o fator de risco pode durar um longo período. Como exemplo, pode-se citar o cigarro, que pode desencadear o aparecimento de um câncer, mas o período de latência exato não pode ser calculado. Como citado por Howard (2013), a avaliação dos períodos de latência para os tipos de

câncer é direta quando as exposições ocorrem em um único ponto no tempo ou regularmente. No entanto, a maioria das exposições humanas a cancerígenos varia significativamente ao longo do tempo, tornando difícil a determinação dos períodos mínimos de latência. Os autores ainda afirmam que é importante compreender que a literatura científica que avalia períodos mínimos de latência para tipos específicos de câncer é escassa.

A partir da realização de uma revisão da literatura, observou-se a escassez de modelos que estimam o período de latência, indicando a necessidade de estudo desse tema devido à sua grande importância. Dentre os modelos existentes, grande parte deles já tem um período de latência definido com base em um fator de risco anterior. Por outro lado, dentre os modelos que estimam o período de latência, destacamos os trabalhos de [Nadler e Zurbenko \(2012\)](#), [Nadler e Zurbenko \(2014\)](#), que designam um parâmetro específico para a latência e fazem sua estimação. O trabalho desses autores será estudado em um primeiro momento para reunir conteúdo visando a proposição de um novo modelo ou até mesmo a modificação desse já existente.

Nesse contexto, métodos de análise de sobrevivência surgem como uma ferramenta útil para a construção de modelos que visam a estimação do período de latência, uma vez que essa abordagem vem sendo muito utilizada em análise de dados de câncer que envolvem o tempo até algum evento de interesse.

Esse trabalho está organizado da seguinte forma: o Capítulo 1 apresenta uma introdução ao problema da latência, à análise de sobrevivência, uma revisão da literatura e os objetivos do trabalho. O estudo do modelo de Nadler e Zurbenko é apresentado no Capítulo 2, contendo a formulação do modelo, uma ilustração em dados reais e a discussão dos resultados. No Capítulo 3 é feita a inserção da censura no modelo dos autores, uma ilustração e um estudo de simulação. O Capítulo 4 apresenta uma outra proposta de modelo considerando a abordagem de Fragilidade Compartilhada, com respectiva aplicação em dados reais. No Capítulo 5 é apresentada uma proposta de modelo de latência com covariável, também contendo uma aplicação. Por fim, o Capítulo 6 apresenta as considerações finais e propostas futuras.

1.1 Análise de Sobrevivência

Modelos de análise de sobrevivência consideram que a variável de interesse é o tempo até que um determinado evento ocorra, também conhecido como tempo de falha ([Saude \(2014\)](#)). Esse tempo pode ser o tempo entre o diagnóstico do câncer e o óbito ou entre o diagnóstico e a reincidência da doença. Além disso, outra característica dessa metodologia é a de incorporar informações de tempos incompletos (tempos censurados), que são os tempos dos indivíduos para os quais não se observou o evento de interesse. Para esses casos, toda a informação referente à resposta se resume ao conhecimento de que o tempo de falha é superior àquele observado.

Nesse contexto, podemos caracterizar os três tipos de censura da seguinte maneira ([Louzada-Neto, Mazucheli e Achcar \(2001\)](#)):

1. Censura tipo I: A censura acontece nos indivíduos que ao término de um tempo pré-estabelecido para o estudo não tiveram a ocorrência do evento de interesse.
2. Censura tipo II: O estudo termina quando ocorreu o evento de interesse em um número pré estabelecido de pacientes.
3. Censura tipo III: Quando o processo de censura acontece ao acaso (ex: paciente sai do estudo sem ter ocorrido a falha ou morre por uma razão diferente daquela em estudo).

Os dados de sobrevivência para um indivíduo i ($i = 1, \dots, n$) são geralmente representados pelo par (t_i, δ_i) , sendo t_i o tempo de falha ou de censura e δ_i a variável indicadora de censura, que assume valor 1 se t_i é um tempo de falha e 0 se t_i é um tempo censurado.

Seja T o tempo de vida de um paciente com câncer uma variável aleatória contínua e não negativa cuja função de distribuição acumulada é dada por:

$$F(t) = Pr(T \leq t),$$

que é definida como a probabilidade de um paciente não sobreviver ao tempo t .

A função de sobrevivência $S(t)$ é o complemento da função de distribuição acumulada e definida como (Louzada-Neto, Mazucheli e Achcar (2001)):

$$S(t) = 1 - F(t) = Pr(T > t)$$

sendo que essa função é monótona decrescente, com $S(0) = 1$ e $S(\infty) = 0$.

A função de risco, denotada por $h(t)$, é dada por:

$$h(t) = \lim_{\Delta_t \rightarrow 0} \frac{Pr(t \leq T < t + \Delta_t | T \geq t)}{\Delta_t}.$$

Pela expressão acima conclui-se que, para Δ_t pequeno em valor absoluto, $h(t)$ representa aproximadamente a probabilidade do indivíduo morrer no intervalo $(t, t + \Delta_t]$ condicional ao fato de haver sobrevivido ao tempo t .

Essa função é também uma aproximação da taxa de falha por unidade de tempo instantânea e pode assumir diversas formas. As duas formas básicas são: crescente (taxa de risco cresce conforme aumenta o tempo) e decrescente (taxa de risco decresce conforme aumenta o tempo). No entanto, ela também pode assumir forma de “banheira” (começa decrescente, se estabiliza e volta a crescer) ou forma unimodal. Além disso, a função de risco pode ser constante, ou seja, não varia no tempo.

A função de taxa de falha acumulada, denotada por $H(t)$, fornece a taxa de falha acumulada do indivíduo e é definida por:

$$H(t) = \int_0^t h(u) du. \tag{1.1}$$

Dadas as funções definidas anteriormente, temos algumas relações matemáticas importantes, tais como:

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt}(\log S(t)), \quad (1.2)$$

$$H(t) = \int_0^t h(u)du = -\log S(t), \quad (1.3)$$

$$S(t) = \exp\{-H(t)\} = \exp\left\{-\int_0^t h(u)du\right\}, \quad (1.4)$$

sendo que essas relações mostram que o conhecimento de uma função, por exemplo $S(t)$, implica no conhecimento das demais, ou seja, $F(t)$, $f(t)$, $h(t)$ e $H(t)$.

A estimação da função de sobrevivência pode acontecer de forma paramétrica ou não paramétrica. Os métodos de estimação não paramétricos são utilizados quando não é feita nenhuma suposição da distribuição de probabilidade do tempo de sobrevivência, ou seja, quando não é imposto nenhum modelo teórico para as falhas e/ou censuras. O método não paramétrico mais comum é o estimador de Kaplan-Meier, que é definido como (Colosimo e Giolo (2006)):

$$\hat{S}(t) = \prod_{j:t_j < t} \left(\frac{n_j - d_j}{n_j} \right), \quad (1.5)$$

em que:

- $t_1 < t_2 < \dots < t_k$ são os k tempos de falha ordenados;
- d_j é o número de falhas em t_j , $j = 1 \dots, k$;
- n_j é o número de indivíduos em risco em t_j .

Prova-se que esse método é um estimador de máxima verossimilhança não-paramétrico da função de sobrevivência (Colosimo e Giolo (2006)).

Já na abordagem paramétrica considera-se que o tempo segue uma distribuição de probabilidade conhecida como, por exemplo, distribuição Exponencial, Weibull, Log-Normal, entre outras. Para esse caso, a escolha do modelo probabilístico adequado deve ser cuidadosa, uma vez que cada distribuição pode gerar estimadores diferentes para a mesma quantidade desconhecida (Colosimo e Giolo (2006)).

Dentre essas distribuições, a Weibull é um modelo muito popular e bastante utilizado na área médica por sua função de risco ser bastante flexível (Nadler e Zurbenko (2012)). Alguns exemplos de estudos que utilizaram a distribuição Weibull e suas variações para dados que envolvem o tempo até o aparecimento de um câncer ou a morte são os trabalhos de Cobre *et al.* (2013), Pike (1966), Peto e Lee (1973), Martinez *et al.* (2013) e Ortega *et al.* (2015).

Nesse caso, seja T uma variável aleatória representando o tempo de falha e seguindo uma distribuição Weibull. Sua função densidade de probabilidade é dada por (Nadler e Zurbenko (2012)):

$$f(t; \lambda, \beta) = \beta \lambda^\beta t^{\beta-1} \exp\{-(t\lambda)^\beta\}, \quad t \geq 0, \quad (1.6)$$

em que β é o parâmetro de forma e λ o de escala, sendo que ambos são positivos. As respectivas funções de sobrevivência e de risco para essa distribuição são definidas como:

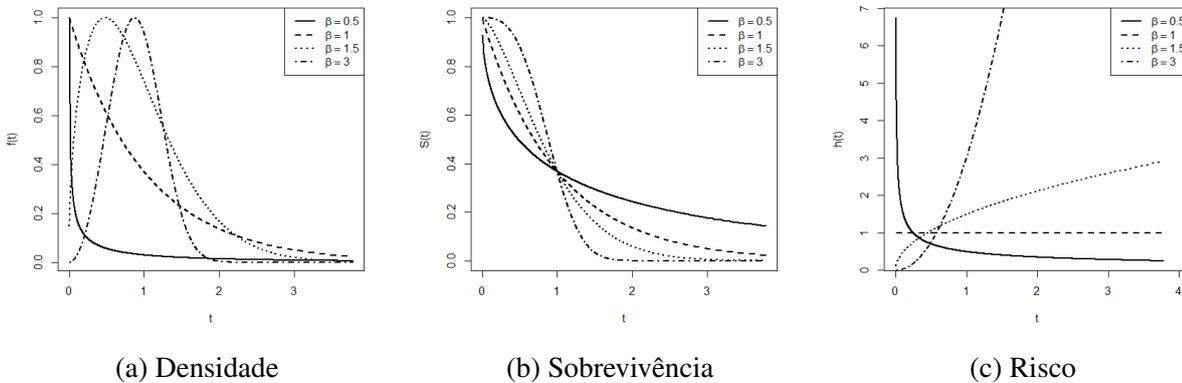
$$S(t) = \exp\{-(\lambda t)^\beta\}, \quad (1.7)$$

$$h(t) = \beta \lambda^\beta t^{\beta-1}, \quad (1.8)$$

sendo que diferentes valores de β definem o comportamento dessa distribuição. Se $\beta = 1$ a distribuição se reduz a uma Exponencial, que tem risco constante no tempo. Por outro lado, se $\beta < 1$ a distribuição assume um risco decrescente e se $\beta > 1$ um risco crescente.

A Figura 2 mostra diferentes formas das funções de densidade, sobrevivência e risco para diferentes valores de β de acordo com essa distribuição.

Figura 2 – Variações da distribuição Weibull



Fonte: Elaborada pelo autor.

1.2 Revisão Bibliográfica

Diversos trabalhos foram dedicados à modelagem do período de latência em diferentes contextos. No geral, os autores já têm um período de latência definido com base em algum fator de risco anterior, porém, novos estudos tentam calcular o período de latência com base em outros critérios.

Uma das primeiras contribuições para o entendimento do período de latência do câncer foi feita por Blum, Grady e Kirby-Smith (1942a), Blum, Grady e Kirby-Smith (1942b) e Armenian

(1987), em que os autores estudaram o câncer de pele em ratos que sofriam radiação ultravioleta e descobriram que o período de incubação era normalmente distribuído. Além disso, verificaram que essa distribuição não varia sistematicamente com a dosagem de radiação, intervalos entre exposições, intensidade da radiação e idade. Tais descobertas foram uma das primeiras afirmações sobre a possível robustez do período de latência em relação à dosagem de exposição (Nadler e Zurbenko (2014)).

O trabalho de Cobb, Miller e Wald (1959) faz a estimativa do período de incubação da leucemia após exposição à radiação resultante da bomba atômica de Hiroshima. Esse estudo avaliou a relação entre o período de incubação e a distância da explosão, o tempo após a explosão, exposição ao tratamento com raio-x de espondilite e experiências passadas. Ainda, os autores enfatizaram a utilidade do período de incubação em identificar o período de tempo na vida do paciente que deve ser pesquisado para a exposição etiológica (Armenian (1987)).

No trabalho de Lui *et al.* (1986), o tempo de latência é definido como o tempo entre a transfusão de sangue e o diagnóstico da AIDS. Eles sugerem que o período de incubação pode assumir qualquer distribuição de probabilidade apropriada, no entanto, eles utilizam a distribuição Weibull e o método de máxima verossimilhança para estimação dos parâmetros. Como resultado, eles notaram que o tempo de incubação é progressivamente menor conforme o ano de transfusão é mais recente.

O modelo proposto por Bentzen, Thames e Overgaard (1989) consiste em um modelo de misturas que incorpora tanto o efeito das covariáveis quanto o efeito do tempo de latência até aparição de uma fibrose subcutânea resultante do recebimento de doses de um tratamento para câncer de mama. Um dos resultados encontrados sobre o efeito da latência é que a frequência de complicações tardias cresce consideravelmente com uma observação prolongada.

O objetivo do estudo de Hessol *et al.* (1990) concentrou-se em verificar o efeito do período de latência da AIDS (tempo entre a soroconversão do HIV e o diagnóstico) no período de sobrevivência da doença (tempo entre o diagnóstico e a morte do paciente) para pacientes homossexuais e bissexuais. Para isso, os autores utilizaram um modelo de Cox com riscos proporcionais e o uso do estimador de Kaplan Meier para estimar o período médio e mediano de latência e de sobrevivência.

No trabalho desenvolvido por Amaral, Pereira e Paixão (2005), o objetivo era determinar a incidência da AIDS e a prevalência do HIV. Para tanto, foi utilizado um método chamado *Back Calculation*, sendo que esse método permite a estimação de taxas de infecção passadas através de dados de incidência da AIDS. Além disso, ele é útil em obter projeções a curto prazo, permitindo a estimação no número de pessoas infectadas por HIV em diferentes estágios da doença. O *Back Calculation* usa o número de casos de HIV diagnosticados e uma estimação confiável da distribuição do período de incubação (tempo entre a infecção pelo HIV e o diagnóstico da AIDS). A ideia básica é usar os dados de incidência da AIDS juntamente com uma estimação da distribuição do período de incubação para estabelecer o número de indivíduos que foram

previamente infectados a fim de ajustar o padrão da incidência de AIDS observada. A abordagem usada é baseada em uma estimação por verossimilhança condicional da distribuição de atraso.

O tempo de latência foi definido como o tempo entre a exposição de amianto e o diagnóstico de *Malignant mesothelioma* no estudos de [Marinaccio et al. \(2007\)](#). Os autores realizaram uma análise descritiva do tempo de latência e também uma ANOVA de 1 fator para verificar se existiam diferenças entre o ano de diagnóstico. Além disso, foi feito um modelo linear generalizado para o tempo de latência relacionado a diversas covariáveis.

No artigo de [Torres et al. \(2013\)](#), os autores estudam o tempo até o aparecimento de angiosarcomas associados à radiação como, por exemplo, um câncer de mama resultante de tratamentos de radioterapia de tumores anteriores. Foram feitas curvas de Kaplan Meier, calculados tempos medianos de latência e comparação das curvas para covariáveis de interesse, além de um modelo de Cox para identificar fatores associados ao tempo de sobrevivência.

[Ornstein et al. \(2014\)](#) estudaram o período de latência da leucemia mielóide aguda relacionado à terapia, que é definido como o intervalo entre o diagnóstico da desordem antecedente e o diagnóstico da leucemia. Os autores utilizaram o método de Kaplan Meier para resumir o tempo até o evento e análises univariadas paramétricas e não paramétricas foram aplicadas para comparar grupos, como teste de Fisher, teste qui-quadrado, teste de Wilcoxon, entre outros. Também foi utilizado um modelo de riscos proporcionais para verificar o efeito simultâneo do efeito de várias covariáveis no tempo de sobrevivência.

Verifica-se que, em todos os artigos descritos acima, conhecia-se os tempos de latência e/ou quando houve exposição determinante a um fator de risco.

[Nadler e Zurbenko \(2012\)](#) propuseram uma extensão do modelo Weibull para estimar o tempo de latência do câncer utilizando um modelo Weibull de dois parâmetros, em que o tamanho do período de latência é designado por um parâmetro específico. O estimador de Kaplan Meier foi utilizado para estimar as probabilidades de sobrevivência que foram utilizadas como resposta em um modelo de regressão utilizado para estimar os parâmetros do modelo. Os mesmos autores ([Nadler e Zurbenko \(2014\)](#)) fizeram uma extensão do modelo anterior em que os parâmetros foram estimados pelo método de máxima verossimilhança. Ainda, eles fizeram uma comparação dos tempos de latência obtidos pelo método dos dois trabalhos. Os métodos propostos por esses autores serão estudados com mais detalhes no decorrer do presente trabalho.

Tem-se então duas situações: quando o tempo de latência é conhecido ou pré-determinado e quando o tempo de latência é desconhecido e precisa ser estimado ou calculado.

Como pode-se notar, poucos trabalhos são específicos para a estimação do período de latência. A maioria trabalha com valores já conhecidos e apenas [Nadler e Zurbenko \(2012\)](#), [Nadler e Zurbenko \(2014\)](#) apresentam métodos de estimação. Devido à este fato, estudaremos aqui esse modelo em particular.

1.3 Objetivos

1.3.1 Gerais

O objetivo do presente trabalho é a descrição e análise crítica, e posterior construção de modelos mais eficazes, de modelos de sobrevivência capazes de estimar o período de latência relacionando covariáveis de interesse e fornecendo ferramentas que possam auxiliar na detecção precoce do câncer, auxiliando profissionais da saúde na prevenção e tratamento da doença.

1.3.2 Específicos

- Estudo e revisão crítica da literatura referente à modelagem do tempo de latência;
- Avaliação de modelos para o tempo de latência;
- Propor modelo de sobrevivência para o tempo de latência;
- Aplicação em dados de câncer.

1.4 Considerações Finais

Nesse Capítulo foi apresentada uma introdução ao problema do período de latência e sua relação com o câncer. Foi também apresentada uma introdução à Análise de Sobrevivência, metodologia que se apresentou útil em problemas desse tipo, uma revisão bibliográfica do assunto e os objetivos do trabalho.

MODELO DE LATÊNCIA

2.1 O Modelo

O modelo proposto por [Nadler e Zurbenko \(2012\)](#), [Nadler e Zurbenko \(2014\)](#) se baseia na distribuição Weibull para estimar a latência do câncer. Como dito anteriormente, essa distribuição é muito utilizada por sua função de risco assumir diferentes formas.

Os autores utilizaram a estrutura de um modelo Weibull e desenvolveram um modelo de sobrevivência condicional com dois parâmetros que assume que os indivíduos sobreviveram ao tempo de diagnóstico. Introduzindo esse parâmetro adicional e utilizando a propriedade de memória da Weibull foi estabelecido o que os autores acreditam ser o tempo entre o início do câncer e o diagnóstico, ou seja, o período de latência.

Para aplicação do modelo, eles utilizaram um banco de dados proveniente do programa SEER (*Surveillance, Epidemiology and End Results*), vinculado ao Instituto Nacional do Câncer - EUA. Foram escolhidos os cânceres com altas taxas de mortalidade e disponibilidade limitada de opções de tratamento efetivos, permitindo que a doença seguisse seu curso natural e evitando possíveis vícios. No total, 44 tipos de cânceres foram selecionados e analisados com uma amostra total de 1.608.484 pacientes.

De acordo com dados de registros de câncer disponíveis, os autores observaram que o parâmetro de forma da Weibull, β , é menor que 1 para todos os cânceres analisados, indicando função de risco decrescente no tempo e que a distribuição tem uma memória forte de tempo de sobrevivência passados, o que é um fator crucial nessa análise.

Para encontrar a probabilidade de sobrevivência condicional, eles partiram da propriedade de probabilidade condicional $P(A|B) = P(A \cap B)/P(B)$. Sabendo que a função de sobrevivência da Weibull é $S(t) = \exp\{-(\lambda t)^\beta\}$ e que $S(t) = P(T > t)$, então a probabilidade condicional de sobreviver no tempo $(u + t)$ dado que o paciente sobreviveu até a data de diagnóstico pode ser

encontrada da forma:

$$\begin{aligned} S(u+t|u) &= \frac{P((T > u+t) \cap (T > u))}{P(T > u)} = \frac{P(T > u+t)}{P(T > u)} = \frac{S(u+t)}{S(u)} \\ &= \frac{\exp\{-[\lambda(u+t)]^\beta\}}{\exp\{-[\lambda u]^\beta\}} = \exp\{-\lambda^\beta [(u+t)^\beta - u^\beta]\}, t > 0. \end{aligned} \quad (2.1)$$

Nesse modelo, u é o parâmetro designado ao período de latência, o β é o parâmetro de forma e λ o de escala, sendo que, para cada indivíduo, t_i representa o tempo entre o diagnóstico e a morte ou o fim do estudo.

2.2 Estimação dos Parâmetros

Os autores propõem a estimação dos parâmetros por dois métodos, que eles chamam de método exato e método aproximado. O método exato é baseado na estimação por máxima verossimilhança, enquanto que o método aproximado é obtido por meio de técnicas de regressão linear.

A função de verossimilhança é obtida por meio da função de sobrevivência condicional apresentada em [Equação 2.1](#) e, com isso, a função de distribuição acumulada condicional é definida como:

$$F(u+t|u) = 1 - \exp\{-\lambda^\beta [(u+t)^\beta - u^\beta]\}. \quad (2.2)$$

Derivando a [Equação 2.2](#) com respeito a t , obtém-se a seguinte função de densidade condicional:

$$f(u+t|u) = \beta \lambda^\beta (u+t)^{\beta-1} \exp\{-[(u+t)\lambda]^\beta + (\lambda u)^\beta\}. \quad (2.3)$$

Além disso, sabemos que:

$$h(u+t|u) = \beta \lambda^\beta (u+t)^{\beta-1}, \quad (2.4)$$

$$H(u+t|u) = \lambda^\beta [(u+t)^\beta - u^\beta]. \quad (2.5)$$

Por fim, é possível obter as funções de verossimilhança e log-verossimilhança correspondente a uma amostra aleatória de tamanho n de T aplicando o logaritmo no produto das funções $f(u+t_i|u)$ para $i = 1, \dots, n$, ou seja:

$$L(u, \beta, \lambda; t) = \prod_{i=1}^n f(u+t_i|u) \quad (2.6)$$

$$\ell(u, \beta, \lambda; t) = n\beta \ln(\lambda) + n\ln(\beta) + (\beta - 1) \sum_{i=1}^n \ln(u+t_i) - \lambda^\beta \sum_{i=1}^n (u+t_i)^\beta + n(\lambda u)^\beta. \quad (2.7)$$

A seguir são apresentados os detalhes dos dois métodos propostos.

2.2.1 Mtodo Aproximado

Para estimaco dos parâmetros no mtodo aproximado, os autores utilizaram mtodos de anlise de regresso linear devido à sua simplicidade computacional e fcil interpretao. O mtodo de Kaplan Meier foi utilizado para encontrar as probabilidades de sobrevivncia que foram utilizadas como varivel resposta no modelo.

Sabe-se que a distribuio Weibull tem a propriedade de que $\ln(-\ln(S(t)))$ é linear com relao a $\ln(t)$, sendo $-\beta \ln(\lambda)$ o intercepto e β o coeficiente angular da reta. Logo, para encontrar os valores das estimativas dos parâmetros β e λ , é preciso determinar o intercepto e coeficiente angular da equao:

$$\ln(-\ln(\hat{S}(u+t|u))) = -\beta \ln(\lambda) + \beta \ln(u+t). \quad (2.8)$$

Segundo os autores, para obter \hat{u} é necessrio encontrar o valor de u que maximiza a funo de log-verossimilhana, igualando a derivada parcial com respeito à u a zero.

Derivando a funo de log-verossimilhana apresentada em [Equao 2.7](#) com relao à u , obtm-se que:

$$\begin{aligned} \frac{\partial \ell(u, \beta, \lambda; t)}{\partial u} &= \sum_{i=1}^n \frac{(\beta-1)}{(u+t_i)} - \beta \lambda^\beta \sum_{i=1}^n (u+t_i)^{\beta-1} + n \beta \lambda^\beta u^{\beta-1} \\ &= \sum_{i=1}^n \left\{ \frac{(\beta-1)}{(u+t_i)} - \beta \lambda^\beta \left[(u+t_i)^{\beta-1} - u^{\beta-1} \right] \right\} \\ &= \sum_{i=1}^n \left\{ \frac{(\beta-1)}{(u+t_i)} - \beta \lambda^\beta \left[\frac{(u+t_i)^\beta}{(u+t_i)} - \frac{u^\beta}{u} \right] \right\}. \end{aligned} \quad (2.9)$$

E igualando o resultado a 0 temos que:

$$\begin{aligned} \frac{\partial \ell(u+t|u; \beta, \lambda)}{\partial u} = 0 &= \sum_{i=1}^n \left\{ \frac{(\beta-1)}{(u+t_i)} - \beta \lambda^\beta \left[\frac{(u+t_i)^\beta}{(u+t_i)} - \frac{u^\beta}{u} \right] \right\} \\ &\Rightarrow \frac{-(\beta-1)}{\beta} = a + b(u+t_i). \end{aligned}$$

Encontra-se ento uma estimativa aproximada do tamanho do perodo de latncia no tempo de diagnstico, em que t_i é igual a 0, da forma:

$$\hat{u} = \frac{[-(\beta-1)/\beta] - a}{b}. \quad (2.10)$$

Sendo assim, essa funo representa o tempo em que a regresso de $\ln(S(t))$ com o tempo t é igual ao fator de correo $[-(\beta-1)/\beta]$.

Logo, para determinar $\hat{\lambda}$ e $\hat{\beta}$, os autores sugerem a regresso:

$$\ln(-\ln(\hat{S}(u+t|u))) = -\beta \ln(\lambda) + \beta \ln(u+t) \quad (2.11)$$

e para encontrar \hat{a} e \hat{b} , a regressão:

$$\ln(\hat{S}(u+t|u)) = a + b(u+t_i). \quad (2.12)$$

As demonstrações dos resultados apresentados podem ser vistas nos artigos originais de [Nadler e Zurbenko \(2014\)](#).

2.2.2 Método Exato

Para a estimação dos parâmetros no método exato, [Nadler e Zurbenko \(2014\)](#) utilizaram a estimação por máxima verossimilhança, um dos métodos mais utilizados e versáteis da literatura. Nesse caso, a função de log-verossimilhança é da forma definida na [Equação 2.7](#) e para sua maximização é preciso encontrar o valor dos parâmetros em que a derivada parcial em relação a cada um deles é igual a 0, obtendo assim os parâmetros estimados $\hat{\lambda}$, $\hat{\beta}$ e \hat{u} . Considerando o modelo proposto, a derivada parcial com respeito a cada um dos parâmetros é dada como se segue:

$$\frac{\partial \ell(u, \beta, \lambda; t)}{\partial \lambda} = \frac{n\beta}{\lambda} - \beta \lambda^{\beta-1} \left[\sum_{i=1}^n (u+t_i)^\beta - nu^\beta \right], \quad (2.13)$$

$$\begin{aligned} \frac{\partial \ell(u, \beta, \lambda; t)}{\partial \beta} &= n \ln(\lambda) + \frac{n}{\beta} + \sum_{i=1}^n \ln(u+t_i) \\ &\quad - \lambda^\beta \sum_{i=1}^n (u+t_i)^\beta \left\{ \ln(\lambda) + \ln \left[\sum_{i=1}^n (u+t_i) \right] \right\} \\ &\quad + n\lambda^\beta u^\beta [\ln(\lambda) + \ln(u)], \end{aligned} \quad (2.14)$$

$$\frac{\partial \ell(u, \beta, \lambda; t)}{\partial u} = \frac{(\beta-1)}{\sum_{i=1}^n (u+t_i)} - \beta \lambda^\beta \sum_{i=1}^n (u+t_i)^{\beta-1} + n\beta \lambda^\beta u^{\beta-1}. \quad (2.15)$$

No entanto, as equações apresentadas em [Equação 2.13](#), [Equação 2.14](#) e [Equação 2.15](#) igualadas a zero não podem ser resolvidas diretamente por se tratarem de um sistema de equações não-lineares e a sua solução não apresentar uma forma fechada. Nesse caso, o método iterativo de Newton-Raphson é utilizado para calcular as estimativas de máxima verossimilhança.

Esse método é uma técnica muito poderosa para encontrar a solução de sistemas de equações não-lineares e pode ser usado para maximizar funções de log-verossimilhança multiparamétricas. A j -ésima interação do algoritmo de Newton-Raphson é dada por ([Lawless \(2011\)](#)):

$$\theta_j = \theta_{j-1} - H(\theta_{j-1})^{-1} U(\theta_{j-1}), \quad j = 1, 2, \dots \quad (2.16)$$

O processo começa com um chute inicial θ_0 e na [Equação 2.16](#) o valor de θ_{j-1} corresponde ao valor "atual" do parâmetro e θ_j ao novo valor do parâmetro. Ainda, o vetor $U(\theta)$ é

o vetor de primeiras derivadas, $U(\theta) = \partial\ell(\theta)/\partial\theta$, e $H(\theta)$ é a matriz de segundas derivadas, $H(\theta) = \partial^2\ell(\theta)/\partial\theta\partial\theta'$.

Como valores iniciais para λ_0 , β_0 e u_0 , os autores utilizaram os valores obtidos no método aproximado, descrito anteriormente na [subseção 2.2.1](#).

2.3 Ilustração dos Métodos

Com a finalidade de ilustrar o funcionamento dos métodos apresentados, foi realizada uma aplicação em dados reais de pacientes com um tipo de câncer de fígado, o carcinoma hepatocelular (CHC).

O CHC é uma das complicações mais comuns em doenças de fígado, sendo o quinto câncer mais comum do mundo. Além disso, tem-se que 80% dos pacientes o desenvolvem através da cirrose ([Llovet \(2005\)](#)). Sua incidência anual varia entre 5,5 e 14,9 pessoas a cada 100 mil, resultando em 600 mil a 1 milhão de mortes anuais ([Mazzanti, Gramantieri e Bolondi \(2008\)](#)). Apesar de não ser o câncer sólido mais comum, é a terceira ou quarta causa de morte por câncer do mundo ([Mazzanti, Gramantieri e Bolondi \(2008\)](#)), mas sua incidência tende a subir, atingindo o topo entre 2015 e 2020 ([Llovet \(2005\)](#)).

Os dados provêm de uma população de pacientes do Ambulatório de Lesões Focais da Divisão de Gastroenterologia do Departamento de Clínica Médica, do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo. Um total de 99 pacientes com suspeita de CHC, acompanhados no período de janeiro de 2000 a dezembro de 2004, foram avaliados. Destes, 80 pacientes eram elegíveis para o estudo, sendo que dezenove pacientes foram excluídos devido à não confirmação do CHC ou porque apresentavam outra neoplasia.

2.3.1 Método Aproximado

Como descrito na [subseção 2.2.1](#), esse método considera métodos de análise de regressão linear para estimar os parâmetros de interesse. O método de Kaplan Meier foi utilizado para encontrar as probabilidades de sobrevivência, que foram utilizadas como variável resposta no modelo e o gráfico com suas estimativas e intervalos de 95% de confiança é apresentado na [Figura 3](#).

Como visto anteriormente, para encontrar os valores $\hat{\beta}$ e $\hat{\lambda}$, é preciso determinar o intercepto e coeficiente angular da equação:

$$\ln(-\ln(\hat{S}(u+t|u))) = -\beta \ln(\lambda) + \beta \ln(u+t).$$

Por meio dos resultados dispostos na [Tabela 1](#) tem-se que $\hat{\beta} = 0,9203$ e $\hat{\lambda} = \exp\{-0,2055/\hat{\beta}\} = \exp\{-0,2055/0,9203\} = 0,7998$. O valor 0,92 para o parâmetro β indica uma função de risco

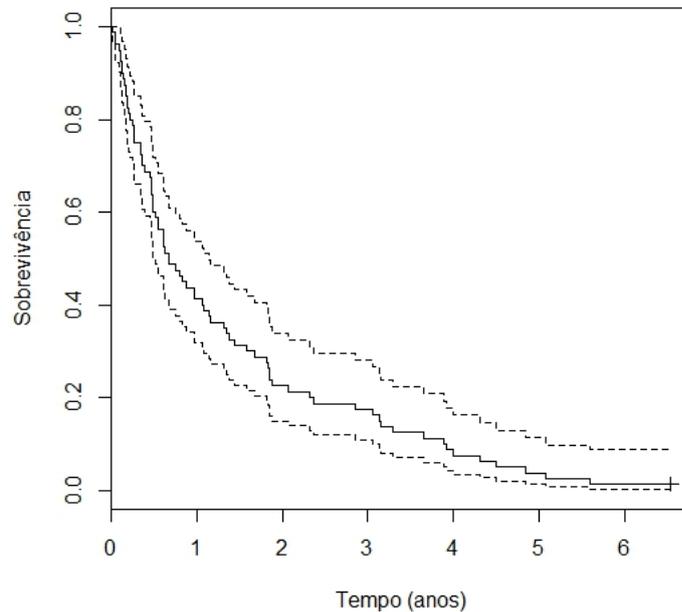


Figura 3 – Sobrevivência empírica obtida pelo método de Kaplan Meier e bandas de confiança

decrecente para os dados, o que significa que o risco de falha diminui conforme o tempo passa. Ainda, o parâmetro de escala, que define a compressão ou expansão da curva, teve valor de 0,8.

Tabela 1 – Resultados do ajuste do modelo de regressão para λ e β descrito em [subseção 2.2.1](#)

	Estimativa	Erro Padrão	Estatística T	P-Valor
Intercepto	-0,2055	0,0206	-9,95	<0,0001
ln(Tempo)	0,9203	0,0153	60,07	<0,0001

Para encontrar \hat{a} e \hat{b} , é preciso determinar o intercepto e coeficiente angular da equação:

$$\ln(\hat{S}(u + t_i|u)) = a + b(u + t_i).$$

A partir dos resultados do modelo, dispostos na [Tabela 2](#), tem-se que $\hat{a} = -0,1305$ e $\hat{b} = -0,6456$.

Tabela 2 – Resultados do modelo de regressão para a e b descrito em [subseção 2.2.1](#)

	Estimativa	Erro Padrão	Estatística T	P-Valor
Intercepto	-0,1305	0,0233	-5,60	<0,0001
Tempo	-0,6456	0,0119	-54,18	<0,0001

Com isso, encontra-se então uma estimativa aproximada do tamanho do período de latência no tempo de diagnóstico, em que t_i é igual a 0, da seguinte maneira:

$$\hat{u} = \frac{(-(\hat{\beta} - 1)/\hat{\beta}) - \hat{a}}{\hat{b}} = \frac{[-(0,9203 - 1)/0,9203] - (-0,1305)}{-0,6456} = -0,3363. \quad (2.17)$$

A [Figura 4](#) apresenta uma representação gráfica do método em questão. Nota-se que o parâmetro β estimado foi de 0,92, o que resulta em um fator de correção igual a 0,08. Estendendo a reta da regressão até o fator de correção no eixo y, é possível estimar o período de latência do câncer de fígado em 0,33 anos.

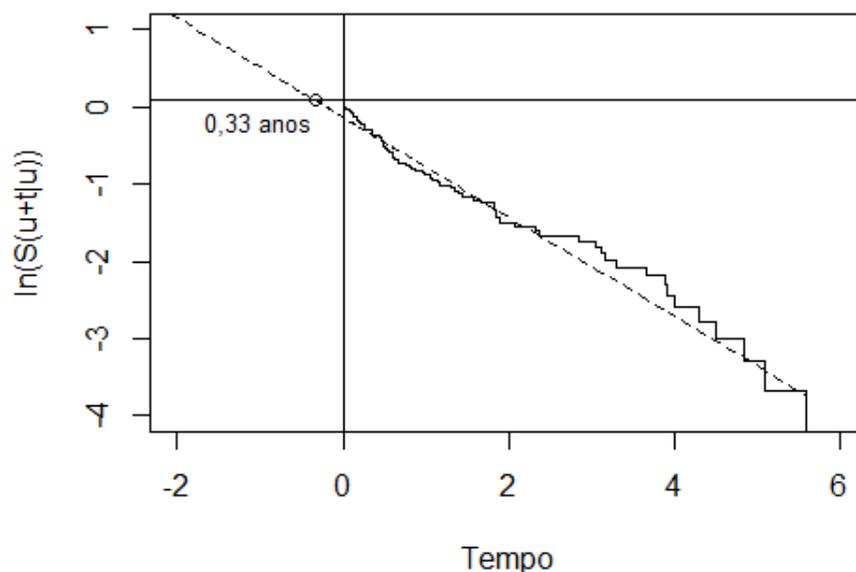


Figura 4 – Estimação do período aproximado do câncer de fígado

Com isso, o tempo de latência do CHC estimado pelo método aproximado sem a inclusão de censuras é de 0,33 anos, o que é equivalente a aproximadamente 4 meses. Isso nos traz uma informação de que provavelmente o início da manifestação da doença se deu 4 meses antes do diagnóstico.

2.3.2 Método Exato

Para a estimação dos parâmetros nesse método, foi utilizada a estimação por máxima verossimilhança considerando a função definida em [Equação 2.7](#). Para maximizar essa função, é preciso encontrar os valores dos parâmetros em que a derivada parcial em relação a cada um deles é igual a 0. Como valores iniciais no método foram utilizadas as estimativas obtidas pelo método aproximado.

Os resultados dos parâmetros estimados com respectivos erros padrão e intervalos de 95% de confiança estão dispostos na [Tabela 3](#). Por meio dos valores obtidos tem-se que o tempo de latência do CHC estimado é de 0,41 anos, o que é equivalente à aproximadamente 5 meses. Esse resultado indica que, pelo método exato, provavelmente o início da manifestação da doença se deu 5 meses antes do diagnóstico.

Tabela 3 – Estimativas dos parâmetros pelo Método Exato

Parâmetros	Estimativa	Erro Padrão	IC (95%)
λ	1,0462	0,3037	{0,45;1,64}
β	0,7631	0,0470	{0,67;0,85}
u	0,4116	0,3732	{-0,31;1,14}

2.4 Discussão

Durante a aplicação dos modelos propostos por [Nadler e Zurbenko \(2012\)](#), [Nadler e Zurbenko \(2014\)](#), identificamos algumas inconsistências e limitações.

Primeiramente, o modelo proposto só funciona para β estritamente menor que 1, o que não acontece na prática, já que é possível ter risco crescente para alguns tipos de câncer, como observado nos trabalhos de [Wahed, Luong e Jeong \(2009\)](#) e [Baker \(1998\)](#). Os autores colocam essa restrição pois, para $\beta < 1$, a função de risco decresce no tempo, o que faz a distribuição ter uma forte “memória” dos tempos de sobrevivência passados, visto que esse fato é crucial na análise proposta por eles. Além disso, aplicando $\beta = 1$ na [Equação 2.1](#), temos:

$$S(u+t|u) = \exp\{-\lambda[(u+t) - u]\} = \exp\{-\lambda t\}. \quad (2.18)$$

Como podemos observar, o parâmetro correspondente à latência (u) é anulado na equação, fazendo com que o modelo em questão se reduza a um modelo Exponencial, que tem a característica da falta de memória. Esse fato impossibilita a análise do tempo de latência quando o risco é constante no tempo.

Segundamente, os autores utilizam a relação apresentada em [Equação 2.8](#) para estimar os parâmetros β e λ . No entanto, se calcularmos $\ln(-\ln(\hat{S}(u+t|u)))$ para a função de sobrevivência condicional proposta por eles, temos:

$$\begin{aligned} S(u+t|u) &= \exp\{-\lambda^\beta[(u+t)^\beta - u^\beta]\} & (2.19) \\ \ln(S(u+t|u)) &= -\lambda^\beta[(u+t)^\beta - u^\beta] \\ -\ln(S(u+t|u)) &= \lambda^\beta(u+t)^\beta - \lambda^\beta u^\beta \\ \ln(-\ln(S(u+t|u))) &= \beta \ln(\lambda) + \ln[(u+t)^\beta - u^\beta]. \end{aligned}$$

Observa-se que a equação resultante é diferente daquilo que os autores apresentam, impossibilitando a aplicação do método aproximado, que se baseia em uma regressão linear, para a estimação dos parâmetros do modelo. Outra relação que os autores utilizam para a estimação de a e b , apresentada em [Equação 2.12](#), não respeita a linearização apresentada, pois:

$$\ln(\hat{S}(u+t_i|u)) = -\lambda^\beta[(u+t_i)^\beta - u^\beta] = \lambda^\beta u^\beta - \lambda^\beta (u+t_i)^\beta = a + b(u+t_i)^\beta, \quad (2.20)$$

diferentemente daquilo que os autores apresentam em seus trabalhos.

Em terceiro lugar, os autores consideram a latência como um parâmetro que assume um valor geral para todos os indivíduos. Isso não acontece na prática, já que vários fatores podem influenciar no período de latência do câncer individualmente.

Por fim, os modelos propostos pelos autores não incorporam informações censuradas, o que resulta em uma perda de informação.

2.5 Considerações Finais

No presente capítulo foi apresentado o estudo de um modelo para o período de latência já existente, contendo uma introdução ao modelo, estimação dos parâmetros, uma ilustração do método em um conjunto de dados reais. Além disso, incluímos uma discussão com as inconsistências e limitações encontradas no modelo. Nos próximos capítulos serão apresentadas novas propostas de modelo para o período de latência baseadas no modelo já apresentado, que visam solucionar as limitações citadas na discussão.

MODELO 1 - MODELO DE LATÊNCIA CONSIDERANDO CENSURA

3.1 O Modelo

Se tratando de dados de tempo, é muito comum a observação de tempos incompletos (tempos censurados), que são os tempos dos indivíduos para os quais não se observou o evento de interesse por diversos motivos. A consideração dessas informações é muito importante para não haver perda de informação na análise dos dados.

Para esses casos, toda a informação referente à resposta se resume ao conhecimento de que o tempo de falha é superior àquele observado.

Nesse caso, os dados de sobrevivência para um indivíduo i ($i = 1, \dots, n$) são geralmente representados pelo par (t_i, δ_i) , sendo t_i o tempo de falha ou de censura e δ_i a variável indicadora de censura, que assume valor 1 se t_i é um tempo de falha e 0 se t_i é um tempo censurado.

As informações de censura não foram consideradas pelos autores na construção dos modelos propostos por eles e apresentados anteriormente no [Capítulo 2](#). Com isso, desenvolvemos um modelo baseado naquele proposto por [Nadler e Zurbenko \(2012\)](#), [Nadler e Zurbenko \(2014\)](#), porém considerando a presença de censura na estimação dos parâmetros de interesse.

3.2 Estimação dos Parâmetros

Os autores utilizaram o método de máxima verossimilhança para estimação dos parâmetros supondo que todas as observações t_i são não censuradas. Assim, supondo que a população seja caracterizada por uma função de densidade $f(t)$, a função de verossimilhança para um

parâmetro θ dessa população é expressa por (Colosimo e Giolo (2006)):

$$L(\theta) = \prod_{i=1}^n f(t_i; \theta). \quad (3.1)$$

O método se baseia em encontrar o valor de θ que maximize a função $L(\theta)$. Nesse caso, a função de verossimilhança mostra que a contribuição de cada observação não censurada é a sua função de densidade (Colosimo e Giolo (2006)).

Porém, quando a observação é censurada, a contribuição da observação em $L(\theta)$ é a função de sobrevivência $S(t)$, uma vez que essas observações informam que o tempo de falha é maior que o tempo de censura observado. Com isso, a função de verossimilhança, quando existem dados censurados à direita na amostra, é dada por (Louzada-Neto, Mazucheli e Achcar (2001)):

$$L(\theta) = \prod_{i=1}^n [f(t_i, \theta)]^{\delta_i} [S(t_i, \theta)]^{1-\delta_i}, \quad (3.2)$$

em que δ_i é a variável indicadora de censura definida anteriormente.

Como já foi descrito, para encontrar a probabilidade de sobrevivência condicional, os autores Nadler e Zurbenko (2012) partiram da propriedade de probabilidade condicional e encontraram que a probabilidade condicional de sobreviver no tempo $(u+t)$, dado que o paciente sobreviveu até a data de diagnóstico, é dada por:

$$S(u+t|u) = \exp\{-\lambda^\beta [(u+t)^\beta - u^\beta]\}, \quad t > 0. \quad (3.3)$$

Nesse modelo, u é o parâmetro designado ao período de latência, β é o parâmetro de forma e λ o de escala, sendo que, para cada indivíduo, t_i representa o tempo entre o diagnóstico e a morte ou o fim do estudo.

A função de distribuição acumulada condicional é obtida por meio da função de sobrevivência condicional apresentada em Equação 3.3 e definida como:

$$F(u+t|u) = 1 - \exp\{-\lambda^\beta [(u+t)^\beta - u^\beta]\}. \quad (3.4)$$

Derivando a Equação 3.4 com respeito a t , obtém-se a seguinte função de densidade condicional:

$$f(u+t|u) = \beta \lambda^\beta (u+t)^{\beta-1} \exp\{-((u+t)\lambda)^\beta + (\lambda u)^\beta\}. \quad (3.5)$$

Sendo assim, a função de verossimilhança para o modelo descrito é dada por:

$$\begin{aligned} L(u, \beta, \lambda; t) &= \prod_{i=1}^n [f(t_i, \theta)]^{\delta_i} [S(t_i, \theta)]^{1-\delta_i} \\ &= \prod_{i=1}^n \left[\beta \lambda^\beta (u+t)^{\beta-1} \exp\{-((u+t)\lambda)^\beta + (\lambda u)^\beta\} \right]^{\delta_i} \left[\exp\{-\lambda^\beta [(u+t)^\beta - u^\beta]\} \right]^{1-\delta_i} \end{aligned} \quad (3.6)$$

e a respectiva função de log-verossimilhança é dada por:

$$\begin{aligned} \ell(u, \beta, \lambda; t) &= \sum_{i=1}^n \delta_i \ln \left[\beta \lambda^\beta (u + t_i)^{\beta-1} \exp\{ -[(u + t_i)\lambda]^\beta + (\lambda u)^\beta \} \right] \\ &+ \sum_{i=1}^n (1 - \delta_i) \{ -\lambda^\beta [(u + t)^\beta - u^\beta] \}. \end{aligned} \quad (3.7)$$

Com a função de log-verossimilhança, a estimação dos parâmetros pode ser feita por meio dos mesmos procedimentos descritos em [subseção 2.2.1](#) e [subseção 2.2.2](#), em que, para a estimação dos parâmetros no método aproximado, os autores utilizaram métodos de análise de regressão linear e, no método exato, a estimação é feita por máxima verossimilhança, em que, para que haja a maximização da função é preciso encontrar os valores dos parâmetros em que a derivada parcial em relação a cada um deles é igual a 0.

3.3 Ilustração dos Métodos com Censura

3.3.1 Método Aproximado

Os mesmos dados sobre câncer de fígado foram utilizados, porém agora considerando a informação de censura. Um total de 78% dos dados são censurados com censura tipo I. O gráfico exibido na [Figura 5](#) apresenta as estimativas de Kaplan Meier utilizadas como variável resposta no modelo quando considerada a censura com os intervalos de 95% de confiança.

Os resultados do modelo para λ e β estão dispostos na [Tabela 4](#). Por meio desses resultados, temos que $\hat{\beta} = 0,8978$ e $\hat{\lambda} = \exp\{-0,3473/\hat{\beta}\} = \exp\{-0,3473/0,8978\} = 0,6792$. Ainda, o valor da estimativa do parâmetro de forma β (menor que 1) novamente indica que a função de risco é decrescente.

Tabela 4 – Resultados do modelo de regressão para λ e β descrito em [subseção 2.2.1](#) com censura

	Estimativa	Erro Padrão	Estatística T	P-Valor
Intercepto	-0,3473	0,022	-15,74	<0,0001
log(Tempo)	0,8978	0,0161	55,62	<0,0001

O mesmo procedimento feito anteriormente foi realizado para encontrar os valores dos parâmetros a e b e os resultados do modelo estão dispostos na [Tabela 5](#). Considerando esses resultados, tem-se que $\hat{a} = -0,1611$ e $\hat{b} = -0,502$

Tabela 5 – Resultados do modelo de regressão para a e b descrito em [subseção 2.2.1](#) com censura

	Estimativa	Erro Padrão	Estatística T	P-Valor
Intercepto	-0,1611	0,0203	-7,92	<0,0001
Tempo	-0,502	0,0123	-40,75	<0,0001

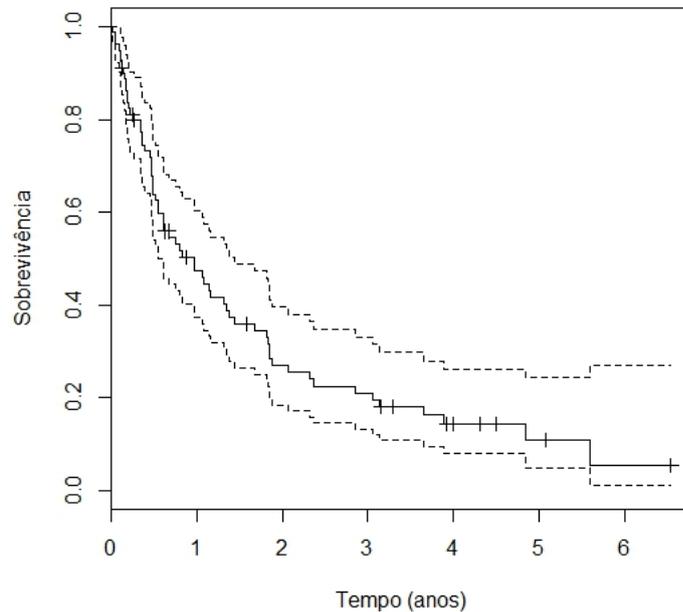


Figura 5 – Sobrevivência empírica obtida pelo método de Kaplan Meier e bandas de confiança considerando censura

Com isso, uma estimativa aproximada do tamanho do período de latência no tempo de diagnóstico, em que t_i é igual a 0, quando considera a censura, é dada por:

$$\hat{u} = \frac{[-(\hat{\beta} - 1)/\hat{\beta}] - \hat{a}}{\hat{b}} = \frac{-(0,8978 - 1)/0,8978 - (-0,1611)}{-0,502} = -0,5476. \quad (3.8)$$

A Figura 6 apresenta uma representação gráfica do método considerando que o parâmetro β estimado foi de 0,9, resultando em um fator de correção igual a 0,11. Novamente, estendendo a reta da regressão até o fator de correção no eixo y é possível estimar o período de latência do câncer de fígado em 0,54 anos.

Portanto, o tempo de latência do CHC estimado pelo método aproximado com a inclusão de censuras é de 0,54 anos, o que equivale a aproximadamente 6 meses e meio. Esse resultado indica que provavelmente o início da manifestação da doença se deu 6 meses e meio antes do diagnóstico.

3.3.2 Método Exato

O mesmo procedimento realizado anteriormente, em que as censuras não foram consideradas, é aplicado no método quando as censuras são consideradas. Os resultados dos parâmetros estimados com respectivos erros padrão e intervalos de 95% de confiança estão dispostos na Tabela 6.

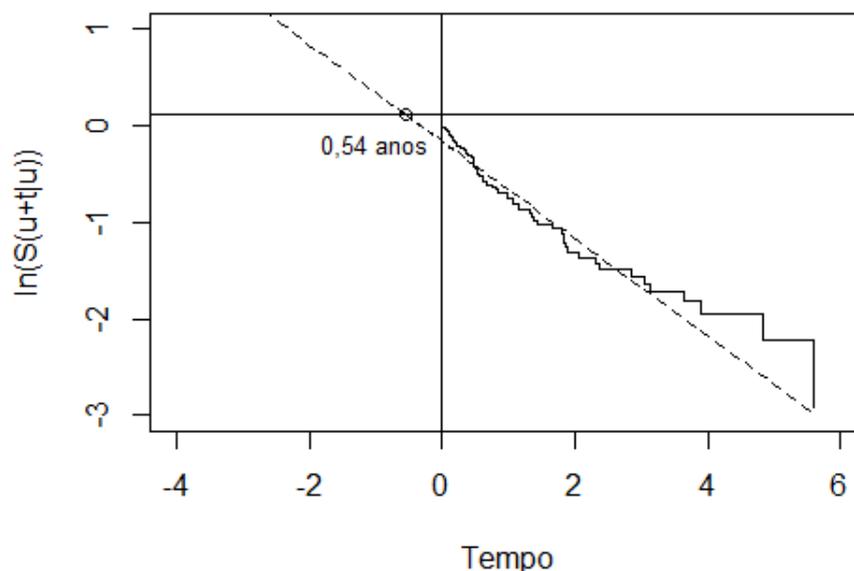


Figura 6 – Estimação do período aproximado do câncer de fígado

Dessa forma, o tempo de latência do CHC estimado pelo método exato com a inclusão de censuras é de 0,67 anos, o que equivale a aproximadamente 8 meses.

Tabela 6 – Estimativas dos parâmetros pelo Método Exato e com censura

Parâmetros	Estimativa	Erro Padrão	IC (95%)
$\hat{\lambda}$	1,2971	0,2366	{0,83;1,73}
$\hat{\beta}$	0,6325	0,0340	{0,56;0,69}
\hat{u}	0,6781	0,3112	{0,06;1,28}

A partir dos resultados apresentados, observamos no [Capítulo 2](#) que, sem censuras, o método aproximado estima a latência em aproximadamente 4 meses, enquanto que o método exato estima em aproximadamente 5 meses. Por outro lado, com a presença de censuras, o método aproximado estima uma latência aproximada de 6 meses e meio para o CHC, enquanto que o método exato estima em aproximadamente 8 meses. Dessa forma, existe uma diferença entre as duas abordagens, mostrando que a inclusão da censura impacta na estimativa do tempo latente para ambos os métodos ([Tabela 7](#)). Diante desse fato, apresentamos um estudo de simulação para verificar o desempenho do método exato com censura.

3.4 Estudo de Simulação

Um estudo de simulação foi realizado com a finalidade de avaliar o método de estimação exato considerando censuras, o comportamento das estimativas e sua variação, uma vez que esse

Tabela 7 – Comparação das estimativas do parâmetro de latência pelos Métodos Aproximado e Exato com e sem censura

	Método Aproximado	Método Exato
Sem censura	0,3363	0,4116
Com censura	0,5476	0,6781

estudo não foi feito pelos autores [Nadler e Zurbenko \(2014\)](#).

Para a geração dos dados de tempos, foram encontradas as raízes de t na equação:

$$\begin{aligned}
 S(u+t|u) &= \exp\{-\lambda^\beta[(u+t)^\beta - u^\beta]\} & (3.9) \\
 \ln(S(u+t|u)) &= -\lambda^\beta[(u+t)^\beta - u^\beta] \\
 -\ln(S(u+t|u)) &= \lambda^\beta(u+t)^\beta - \lambda^\beta u^\beta \\
 \ln(-\ln(S(u+t|u))) &= \beta \ln(\lambda) + \ln[(u+t)^\beta - u^\beta] \\
 0 &= -\ln(-\ln(S(u+t|u))) + \beta \ln(\lambda) + \ln[(u+t)^\beta - u^\beta].
 \end{aligned}$$

O parâmetro u foi fixado em 10, o parâmetro λ em 2 e o parâmetro β foi variado entre 0,1 e 0,9, respeitando a restrição imposta pelos autores. Seguindo essas condições, foram geradas através do *Software R*, 500 amostras de tamanho 300 e construídos os boxplots para as estimativas de cada parâmetro. Para alguns casos o algoritmo não convergiu, resultando em observações mal estimadas, ou seja, que apresentaram valores muito diferentes dos valores utilizados na geração. Esses casos foram omitidos na apresentação dos boxplots.

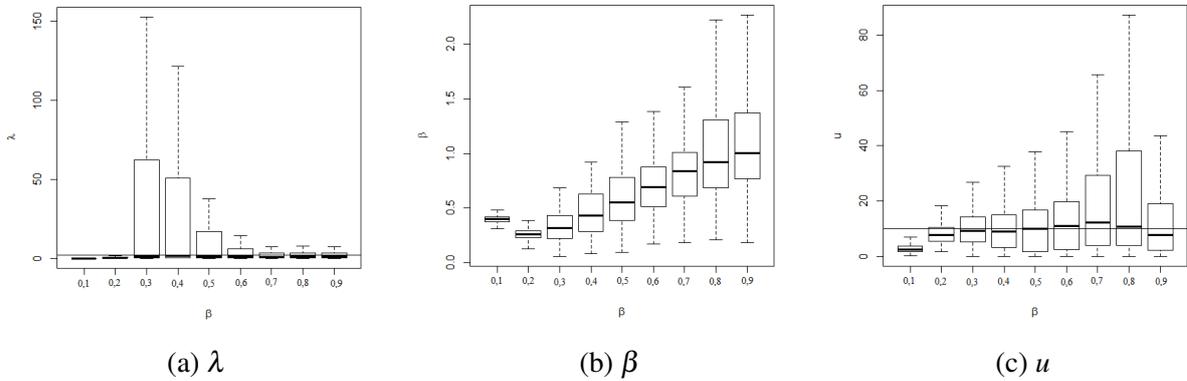
Analisando a [Figura 7](#) nota-se que as estimativas de variância aumentam conforme o parâmetro β utilizado na geração se aproxima do valor de restrição $\beta = 1$. Ainda é possível notar que, em relação ao parâmetro λ , o método apresenta dificuldade na estimação, demonstrada por sua ampla variabilidade, principalmente para $\beta \in [0,3; 0,6]$.

Além disso, como o método estudado é baseado na função de log-verossimilhança, a precisão da estimativa de um parâmetro - por exemplo $\hat{\theta}$ - depende da curvatura da função de log-verossimilhança em torno de seu verdadeiro valor. Se a log-verossimilhança é bastante curvada ou "íngreme" em torno de $\hat{\theta}$, então θ é estimado precisamente e, com isso, temos bastante informação sobre θ ([Zivot \(2009\)](#)). Por outro lado, se a função de log-verossimilhança é não curvada ou "plana (*flat*)" perto de $\hat{\theta}$, então θ não é precisamente estimado. Nesse caso, é dito que não se tem muita informação sobre θ .

Nesse contexto, as [Figura 8](#), [Figura 9](#) e [Figura 10](#) apresentam as funções de log-verossimilhança para os parâmetros λ , β e u , respectivamente. Verifica-se comportamento instável para as amostras geradas e diversos pontos de máximo locais, o que impossibilita a determinação do ponto de máximo para as funções pelo método utilizado. Observa-se ainda

que os gráficos de algumas funções são planas, resultando numa falta de informação para o parâmetro, o que explica os resultados encontrados na simulação.

Figura 7 – Boxplots das estimativas dos parâmetros para diferentes valores de β iniciais



Fonte: Elaborada pelo autor.

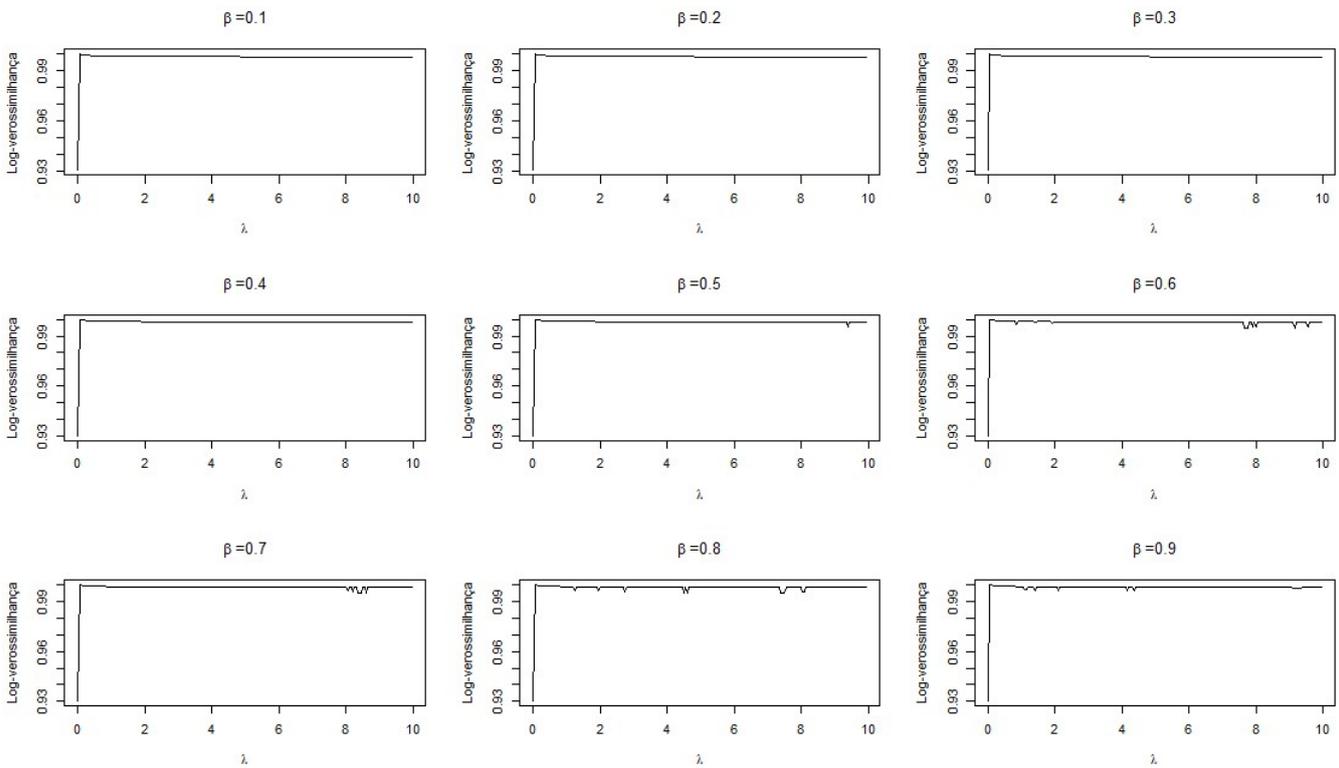
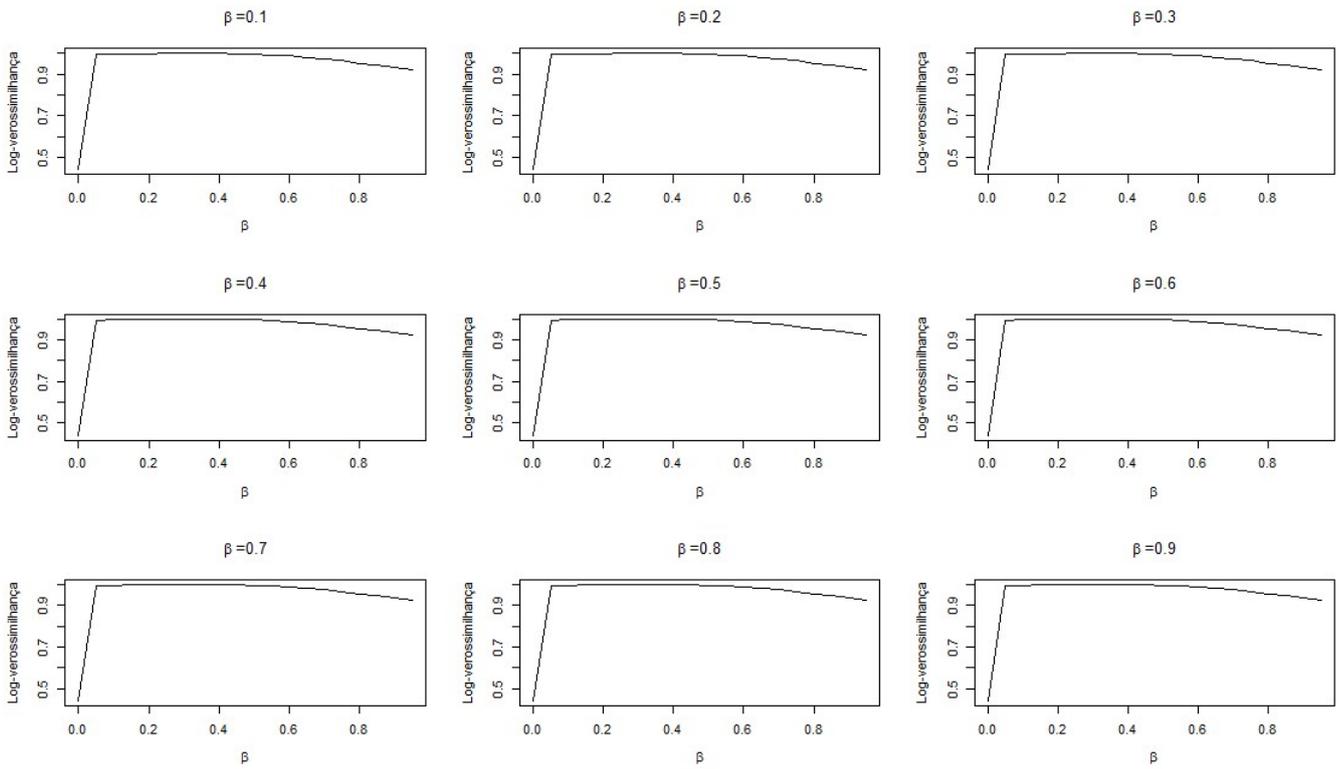
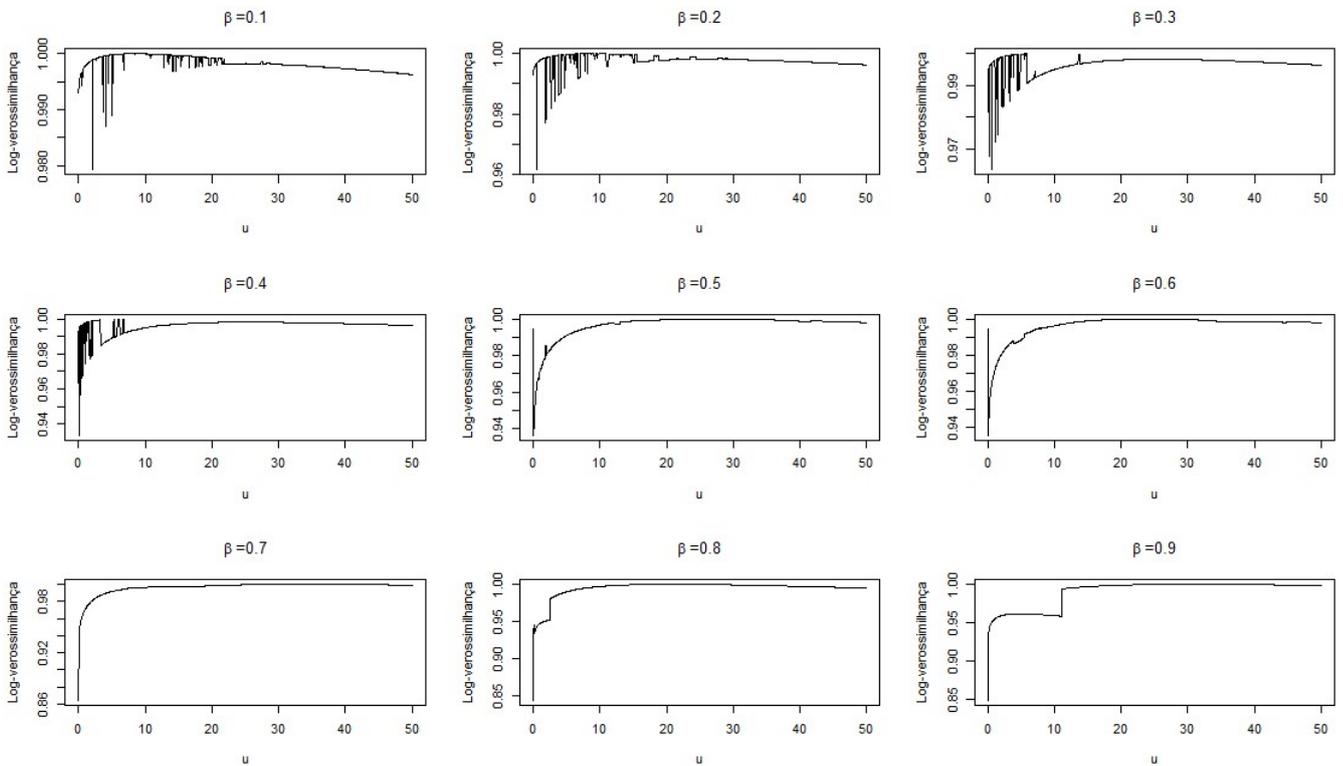


Figura 8 – Verossimilhança para o parâmetro λ

Figura 9 – Verossimilhança para o parâmetro β Figura 10 – Verossimilhança para o parâmetro u

3.5 Considerações Finais

Nesse capítulo foi apresentada a inclusão da censura no modelo anterior, pois esse foi um dos pontos acusados na discussão sobre as limitações do modelo inicial. Verificamos que a inclusão da censura impacta na estimativa do tempo latente para ambos os métodos de estimação apresentados. Também foi realizada uma simulação desse modelo com a finalidade de estudar o método de estimação exato considerado, na qual observamos alguns problemas de estimação.

MODELO 2 - MODELO COM FRAGILIDADE COMPARTILHADA

4.1 O Modelo

Como apontado na discussão, o modelo apresentado pelos autores [Nadler e Zurbenko \(2014\)](#) considera uma única latência para todos os indivíduos no estudo, o que não acontece na prática, já que vários fatores podem influenciar no período de latência do câncer. Uma outra alternativa que apresentamos são os modelos de fragilidade compartilhada. Ele é caracterizado pela inclusão de um efeito aleatório não observável na função de risco, com o objetivo de controlar a heterogeneidade não observável das unidades sob estudo. Apesar desse modelo não resultar em um valor de latência para cada indivíduo ou para cada grupo específico, temos um efeito aleatório incorporado na função de risco de forma multiplicativa, sendo que, quanto maior o valor do efeito aleatório, maior será o risco de uma falha ocorrer. Essa fragilidade pode explicar, além de outros fatores, a heterogeneidade dos indivíduos em estudo ([Colosimo e Giolo \(2006\)](#)).

Um caso particular dos modelos de fragilidade são os modelos de fragilidade compartilhada, que são utilizados quando a população está dividida em "*clusters*", e os indivíduos do mesmo *cluster* são mais parecidos do que aqueles que estão em outros. Os modelos de fragilidade compartilhada são assim chamados porque os sujeitos do mesmo *cluster* compartilham do mesmo fator de fragilidade ([Hougaard \(2012\)](#)).

Nesse caso, a fragilidade representa um efeito aleatório que descreve o risco comum, isto é, a fragilidade é compartilhada por indivíduos dentro de um mesmo grupo. Isso significa que os grupos apresentam fragilidades diferentes e os grupos com maior fragilidade deverão experimentar o evento de interesse em menor tempo do que aqueles com menores valores ([Colosimo e Giolo \(2006\)](#)).

No contexto do presente trabalho, os indivíduos que estão no mesmo estadiamento do

câncer (mesmo *cluster*) vão compartilhar da mesma fragilidade.

4.2 Formulação do Modelo

Seja j os indivíduos do estudo, $j = 1, \dots, n$, e os clusters i , $i = 1, \dots, s$. Observamos y_{ij} , sendo o mínimo entre os tempos de censura c_{ij} e os tempos de evento t_{ij} e δ_{ij} a variável indicadora de censura, como definido anteriormente. O número de eventos observados no i -ésimo *cluster* é $d_i = \sum_{j=1}^{n_i} \delta_{ij}$. Com isso, o modelo de fragilidade compartilhada é definido como

$$h_{ij}(t) = h_0(t) \exp(x_{ij}^t \alpha + z_i), \quad (4.1)$$

em que $h_{ij}(t)$ é a função de risco condicional do j -ésimo indivíduo do i -ésimo *cluster* (condicional a z_i), $h_0(t)$ é a função de risco basal, α é o vetor de parâmetros, x_{ij} é o vetor de covariáveis e z_i é o efeito aleatório do i -ésimo *cluster*. Muitas possibilidades para a função de risco base são consideradas na literatura, como Weibull, Exponencial, Log-Normal, entre outras.

A função de risco base que iremos utilizar será aquela proposta por [Nadler e Zurbenko \(2012\)](#), que contém o parâmetro da latência.

O modelo apresentado em [Equação 4.1](#) pode ser reescrito como:

$$h_{ij}(t) = h_0(t) w_i \exp(x_{ij}^t \alpha) \quad (4.2)$$

em que $w_i = \exp(z_i)$ é a fragilidade do i -ésimo *cluster*.

Os valores w_i 's, $i = 1, \dots, s$, são valores de uma amostra de uma densidade f_W . Como w_i multiplica a função de risco, W deve ser não negativa. Uma escolha matemática conveniente, e usualmente utilizada para a distribuição dos w_i 's é a distribuição Gama, cuja função densidade de probabilidade é dada por:

$$f_W(w) = \frac{w^{1/\theta-1} \exp\{-w/\theta\}}{\theta^{1/\theta} \Gamma(1/\theta)}, \quad (4.3)$$

sendo Γ a função matemática gama. Nesse caso, $E(W) = 1$ e $Var(W) = \theta$, o que significa que indivíduos de um grupo i com $w_i > 1$ ($w_i < 1$) têm maior risco (menor risco). Já o parâmetro θ diz respeito à variabilidade dos indivíduos nos *clusters*.

4.3 Estimação dos Parâmetros

Numa abordagem frequentista, as estimativas dos parâmetros de interesse são obtidas através da maximização da função de log-verossimilhança.

Através da equação apresentada em [Equação 4.2](#), podemos escrever a função de verossimilhança condicional para o i -ésimo *cluster* como:

$$L_i(\xi, \alpha | w_i) = \prod_{j=1}^{n_i} [h_0(y_{ij}) w_i \exp\{x_{ij}^t \alpha\}]^{\delta_{ij}} \exp(-H_0(y_{ij}) w_i \exp\{x_{ij}^t \alpha\}) \quad (4.4)$$

sendo que ξ contém os parâmetros da função de risco base.

Com isso, a função de verossimilhança marginal $L_{\text{marg},i}(\xi)$, com o vetor de parâmetros $\xi = (\xi, \theta, \alpha)$, para o i -ésimo *cluster* é:

$$L_{\text{marg},i}(\xi) = \int_0^\infty \prod_{j=1}^{n_i} [h_0(y_{ij}) w \exp\{x_{ij}^t \alpha\}]^{\delta_{ij}} \exp(-H_0(y_{ij}) w \exp\{x_{ij}^t \alpha\}) \times \frac{w^{1/\theta-1} \exp\{-w/\theta\}}{\theta^{1/\theta} \Gamma(1/\theta)} dw. \quad (4.5)$$

Considerando $z = 1/\theta + \sum_{j=1}^{n_i} H_0(y_{ij}) \exp\{x_{ij}^t \alpha\}$, a equação apresentada em [Equação 4.5](#) pode ser reescrita como:

$$L_{\text{marg},i}(\xi) = \frac{\prod_{j=1}^{n_i} [h_0(y_{ij}) \exp\{x_{ij}^t \alpha\}]^{\delta_{ij}}}{z^{1/\theta+d_i} \theta^{1/\theta} \Gamma(1/\theta)} \int_0^\infty (zw)^{1/\theta+d_i-1} \exp\{-zw\} d(zw). \quad (4.6)$$

e trabalhando na integral, temos:

$$L_{\text{marg},i}(\xi) = \frac{\Gamma(d_i + 1/\theta) \prod_{j=1}^{n_i} [h_0(y_{ij}) \exp\{x_{ij}^t \alpha\}]^{\delta_{ij}}}{\left[1/\theta + \sum_{j=1}^{n_i} H_0(y_{ij}) \exp\{x_{ij}^t \alpha\}\right]^{1/\theta+d_i} \theta^{1/\theta} \Gamma(1/\theta)} \quad (4.7)$$

Aplicando o logaritmo em [Equação 4.7](#) e somando para os s *clusters*, obtemos a função de log-verossimilhança marginal $l_{\text{marg}}(\xi)$:

$$l_{\text{marg}}(\xi) = \sum_{i=1}^s [d_i \log(\theta) - \log(\Gamma(1/\theta)) + \log(\Gamma(1/\theta + d_i)) - (1/\theta + d_i) \log\left(1 + \theta \sum_{j=1}^{n_i} H_{ij,c}(y_{ij})\right) + \sum_{j=1}^{n_i} \delta_{ij} (x_{ij}^t \alpha + \log(h_0(y_{ij})))], \quad (4.8)$$

em que $H_{ij,c}(y_{ij}) = H_0(y_{ij}) \exp\{x_{ij}^t \alpha\}$. As estimativas de máxima verossimilhança para ξ , θ e α podem ser encontradas maximizando essa log-verossimilhança.

Além disso, temos que, no modelo de fragilidade Gama compartilhada com covariáveis, as fragilidades não-observáveis de cada cluster podem ser estimadas como ([Nielsen et al. \(1992\)](#)):

$$\hat{w}_i = \frac{1/\hat{\theta} + d_i}{1/\hat{\theta} + \sum_{j=1}^{n_i} h_0(y_{ij}) \exp\{x_{ij}^t \hat{\alpha}\}} \quad (4.9)$$

4.4 Aplicação do Modelo

Os dados utilizados nessa aplicação são os mesmos utilizados anteriormente com dados reais de pacientes com carcinoma hepato celular (CHC). Para os 80 pacientes elegíveis para o estudo, além de observar os tempos e censuras de cada um deles, também consideramos uma classificação prognóstica (estadiamento) chamado OKUDA, que foi desenvolvido por Okuda *et al.* em 1985 em um estudo com 850 pacientes com CHC que foram analisados quanto à sobrevida e tratamento.

Esse foi o primeiro sistema de estadiamento a incluir variáveis de função hepática, como nível de bilirrubina ($<$ ou \geq 3 mg/dl), nível de albumina (\geq ou $<$ 3 g/dl) e presença de ascite, associado a fatores tumorais ($>$ ou $<$ 50% de envolvimento hepático). Para essa classificação foram propostos três estágios (I-III), conforme a [Tabela 8](#):

Tabela 8 – Tabela de classificação de Okuda

	(-)	(+)
TAMANHO TUMORAL	< 50% DO FÍGADO'	> 50% DO FÍGADO
ASCITE	AUSENTE	PRESENTE
ALBUMINA SÉRICA	\geq 3 g/dl	$<$ 3 g/dl
BILIRRUBINA	$<$ 3 mg/dl	\geq 3 mg/dl

OKUDA I: Nenhum dos fatores positivos;

OKUDA II: 1 ou 2 fatores positivos;

OKUDA III: 3 ou 4 fatores positivos.

Classificando os pacientes segundo o sistema de estadiamento de Okuda *et al.* (1985), observa-se a existência de pacientes nos três estágios estudados, com menor concentração deles no estágio 3, conforme a [Tabela 9](#).

Tabela 9 – Tabela de frequências para Okuda

	N (%)
Est. 1	29 (36,25)
Est. 2	31 (38,75)
Est. 3	20 (25,00)

As curvas de sobrevivência estimadas por meio do estimador de Kaplan-Meier para os três estágios são mostradas na [Figura 11](#), em que visualmente é possível perceber a existência de diferença de sobrevivência para os diferentes estágios, pois no estágio III o paciente sobrevive menos tempo do que nos demais estágios, sobrevivendo este por mais tempo estando no estágio I.

Os valores das estatísticas log-rank e Wilcoxon para os estadiamentos ([Tabela 10](#)), que sob a hipótese de igualdade das curvas de sobrevivência, possuem distribuição qui-quadrado com dois graus de liberdade, resultam em $T = 25,7$ e $T = 23,6$ com p -valores menores que 0,0001, nos dando evidências de que existe diferença entre os estágios.

Tabela 10 – Testes de log-rank e Wilcoxon para Okuda com censura

	Teste (valor p)	
	Log-rank	Wilcoxon
Okuda	25,7 ($<$ 0,0001)	23,6 ($<$ 0,0001)

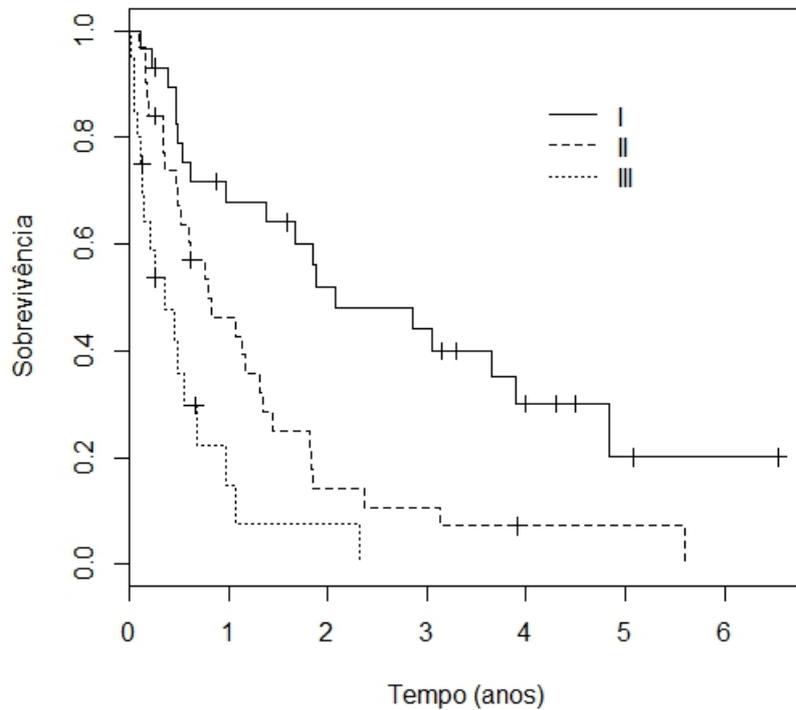


Figura 11 – Kaplan Meier para Okuda

Para encontrar as estimativas dos parâmetros precisamos encontrar os valores que maximizam a função de verossimilhança, dada por:

$$l_{marg}(\zeta) = \sum_{i=1}^s [d_i \log(\theta) - \log(\Gamma(1/\theta)) + \log(\Gamma(1/\theta + d_i))] \quad (4.10)$$

$$- (1/\theta + d_i) \log \left(1 + \theta \sum_{j=1}^{n_i} H_{ij,c}(y_{ij}) \right) + \sum_{j=1}^{n_i} \delta_{ij} (x'_{ij} \alpha + \log(h_0(y_{ij}))).$$

Nesse caso temos que:

$$h_0(t_{ij}) = \beta \lambda^\beta (u + t_{ij})^{\beta-1}, \quad (4.11)$$

$$H_0(t_{ij}) = \lambda^\beta [(u + t_{ij})^{\beta-1} - u^{\beta-1}]. \quad (4.12)$$

No entanto, ao derivar essa equação com relação a cada um dos parâmetros, chegamos em um sistema de equações não-lineares que não apresenta uma forma fechada. Nesse caso, métodos numéricos são utilizados para calcular as estimativas de máxima verossimilhança.

Os resultados das estimativas com respectivos erros padrão e intervalos de 95% de confiança são apresentados na [Tabela 11](#). Temos um valor de β estimado de 2,58, indicando

função de risco decrescente. Já o parâmetro λ , que define a compressão ou expansão da curva, teve valor de 1,44. O parâmetro θ nos dá informação sobre a variabilidade dos indivíduos dos *clusters* e o valor de α diz respeito à covariável idade. Por fim, temos um valor de latência u estimado de 1,08, o que nos traz uma informação de que provavelmente o início da manifestação da doença se deu 1.08 anos antes do diagnóstico.

De acordo com Howard (2013), o período de latência para cânceres sólidos é de 4 anos. No entanto, como já citado anteriormente, a literatura científica que avalia períodos mínimos de latência para tipos específicos de câncer é escassa e a maioria das exposições humanas a cancerígenos varia significativamente ao longo do tempo, tornando difícil a determinação dos períodos mínimos de latência, além dessa ser uma estimativa geral para os cânceres sólidos, e não específica para o câncer de fígado.

Na Tabela 12, temos os valores de fragilidade estimados para cada estadiamento, conforme descrito na Equação 4.9. Sabemos que, quanto maior o valor da fragilidade, maior será o risco de uma falha ocorrer. Observamos que o estadiamento 1 contém menor valor de fragilidade e que o estadiamento 2 tem maior valor, porém sem variar muito com relação ao estadiamento 3. Isso significa que os indivíduos do estadiamento 1 vão demorar mais a experimentar o evento de interesse do que os dos demais estadiamentos.

Tabela 11 – Estimativas dos parâmetros

Parâmetro	Estimativa	Erro Padrão	IC
θ	1,6002	0,7564	{0,11;3,08}
β	2,5792	0,4803	{1,63;3,52}
λ	1,4386	0,4399	{0,57;2,30}
α	1,0162	0,0127	{0,99;1,04}
u	1,0875	0,5178	{0,07;2,10}

Tabela 12 – Estimativas das fragilidades para os estadiamentos

w_1	w_2	w_3
0,7341	1,0079	0,9026

4.5 Considerações Finais

No presente capítulo foi apresentada uma abordagem para o modelo de latência considerando fragilidade compartilhada. Apesar dessa abordagem não resultar em um valor de latência para cada grupo específico, ela nos permitiu analisar a fragilidade de cada estadiamento estudado, no qual pudemos observar que, quanto maior o estadiamento do paciente, maior sua fragilidade e, conseqüentemente, maior o risco da falha ocorrer.

MODELO 3 - MODELO DE LATÊNCIA COM COVARIÁVEL

5.1 O Modelo

Uma outra opção de modelo que podemos utilizar é considerando o estadiamento do paciente como uma covariável dentro do parâmetro de latência. Essa abordagem é interessante pois, como dito anteriormente, o modelo proposto por [Nadler e Zurbenko \(2012\)](#), [Nadler e Zurbenko \(2014\)](#), apresentado no [Capítulo 2](#), considera apenas um período de latência geral para todos os indivíduos no estudo, o que não acontece na prática, já que vários fatores podem influenciar na estimativa desse período.

Com esse modelo conseguimos incorporar a informação de covariáveis que podem influenciar no período de latência em sua estimação, aumentando assim a precisão dessas estimativas. Apesar de não estimarmos um período de latência por paciente, é possível estimar uma latência para cada grupo de indivíduos de acordo com o seu estadiamento.

Isso é feito através da inserção da covariável que mede o estadiamento do câncer no paciente dentro do parâmetro de latência u , obtendo assim um parâmetro estimado para cada estadiamento e , conseqüentemente, um período de latência para cada estadiamento.

5.2 Estimação dos Parâmetros

Como visto anteriormente, os autores [Nadler e Zurbenko \(2014\)](#) definem a probabilidade condicional de sobreviver no tempo $(u + t)$ dado que o paciente sobreviveu até a data de diagnóstico da forma:

$$S(u + t|u) = \exp\{-\lambda^\beta [(u + t)^\beta - u^\beta]\}, t > 0, \quad (5.1)$$

em que u é o parâmetro designado ao período de latência, o β é o parâmetro de forma e λ o de escala, sendo que, para cada indivíduo, t_i representa o tempo entre o diagnóstico e a morte ou o fim do estudo.

A função de verossimilhança é obtida por meio da função de sobrevivência condicional apresentada em [Equação 5.1](#) e, com isso, a função de distribuição acumulada condicional é definida como:

$$F(u+t|u) = 1 - \exp\{-\lambda^\beta[(u+t)^\beta - u^\beta]\}. \quad (5.2)$$

Derivando a [Equação 5.2](#) com respeito a t , obtém-se a seguinte função de densidade condicional:

$$f(u+t|u) = \beta\lambda^\beta(u+t)^{\beta-1} \exp\{-((u+t)\lambda)^\beta + (\lambda u)^\beta\}. \quad (5.3)$$

E então temos que as funções de verossimilhança e log-verossimilhança correspondentes a uma amostra aleatória de tamanho n de T são dadas, respectivamente, por:

$$\begin{aligned} L(u, \beta, \lambda; t) &= \prod_{i=1}^n [f(t_i, \theta)]^{\delta_i} [S(t_i, \theta)]^{1-\delta_i} \\ &= \prod_{i=1}^n \left[\beta\lambda^\beta(u+t)^{\beta-1} \exp\{-((u+t)\lambda)^\beta + (\lambda u)^\beta\} \right]^{\delta_i} \left[\exp\{-\lambda^\beta[(u+t)^\beta - u^\beta]\} \right]^{1-\delta_i} \end{aligned} \quad (5.4)$$

e a respectiva função de log-verossimilhança é dada por:

$$\begin{aligned} \ell(u, \beta, \lambda; t) &= \sum_{i=1}^n \delta_i \ln \left[\beta\lambda^\beta(u+t_i)^{\beta-1} \exp\{-[(u+t_i)\lambda]^\beta + (\lambda u)^\beta\} \right] \\ &\quad + \sum_{i=1}^n (1-\delta_i) \{-\lambda^\beta[(u+t_i)^\beta - u^\beta]\}. \end{aligned} \quad (5.5)$$

Nesse caso, vamos definir que $u = \exp\{x^t \boldsymbol{\eta}\}$, em que x é a matriz de planejamento para o estadiamento e $\boldsymbol{\eta}$ é o vetor de parâmetros para o estadiamento a ser estimado. Dessa forma, teremos um valor $\boldsymbol{\eta}$ estimado para cada estadiamento e, com isso, uma latência estimada para cada estadiamento.

As estimativas de máxima verossimilhança para β , λ , u e $\boldsymbol{\eta}$ podem ser encontradas maximizando essa log-verossimilhança. No entanto, novamente ao derivar a função de log-verossimilhança com relação a cada um dos parâmetros de interesse, chegamos em um sistema de equações não-lineares que não apresenta uma forma fechada. Nesse caso, métodos numéricos são utilizados para calcular as estimativas de máxima verossimilhança.

5.3 Aplicação do Modelo

Os resultados das estimativas com respectivos erros padrão e intervalos de 95% de confiança são apresentados na [Tabela 13](#). Temos um valor de β estimado de 2,24, indicando

função de risco crescente. Já o parâmetro λ , que define a compressão ou expansão da curva, teve valor de 0,04. Os valores de $\boldsymbol{\eta}$ dizem respeito ao estadiamento e são utilizados para calcular a latência através da relação $u = \exp\{x^t \boldsymbol{\eta}\}$.

Na Tabela 14, temos os valores de latência estimados para cada estadiamento. Observamos que o estadiamento 1 obteve uma latência estimada de 1,47, o estadiamento 2 de 1,59 e o estadiamento 3 de 1,70 anos. Isso nos mostra que quanto maior o estadiamento do câncer, maior o tempo de latência, o que significa que provavelmente o início da manifestação da doença se deu antes comparado a um paciente com menor estadiamento.

Tabela 13 – Estimativas dos parâmetros

Parâmetro	Estimativa	Erro Padrão	IC
β	2,2427	0,2826	{1,688;2,796}
λ	0,0390	0,0183	{0,002;0,075}
η_1	4,3848	0,8721	{2,675;6,094}
η_2	0,5282	0,2394	{0,058;0,997}
η_3	1,1134	0,3175	{0,491;1,735}

Tabela 14 – Estimativas das latências (em anos) para os estadiamentos

u_1	u_2	u_3
1,47	1,59	1,70

5.4 Considerações Finais

Neste capítulo estudamos o modelo de latência com o estadiamento inserido na covariável. Nessa abordagem pudemos estimar um período de latência para cada grupo de indivíduos de acordo com seu estadiamento. Como resultado, observamos que quanto maior o estadiamento do câncer, maior o tempo de latência, o que significa que provavelmente o início da manifestação da doença se deu antes comparado a um paciente com menor estadiamento.

CONSIDERAÇÕES FINAIS

Nesse trabalho vimos que é inquestionável a necessidade de investigações e pesquisas constantes na estimação do período de latência do câncer no sentido de aprimorar a detecção precoce e auxiliar em sua prevenção e tratamento. Essa é uma das armas mais promissoras para evitar seu avanço enquanto o mesmo ainda é curável e está localizado em apenas um local, evitando não apenas a mortalidade mas também reduzindo os custos de tratamento ([Nadler e Zurbenko \(2012\)](#)).

Dessa forma, o desenvolvimento de técnicas para estimar o período de crescimento “silencioso” do câncer, que significa conhecer o momento do início do processo cancerígeno (também chamado de período de latência), é de grande importância. A partir da revisão da literatura observamos uma escassez de modelos que estimam a latência do câncer, indicando a necessidade do estudo desse tema. Dentre os modelos existentes, grande parte já tem um período de latência definido com base em um fator de risco anterior. Por outro lado, dentre os modelos que estimam o período de latência, podemos considerar os trabalhos de [Nadler e Zurbenko \(2012\)](#), [Nadler e Zurbenko \(2014\)](#), que designam um parâmetro específico para a latência e fazem sua estimação. No entanto, como observado nesse trabalho, foram identificadas limitações e inconsistências neste modelo.

Portanto, devido à necessidade de modelos para determinação do tempo de latência do câncer e a identificação de limitações no modelo citado, verificamos a necessidade de melhorias no modelo apresentado.

A primeira delas foi a inserção da censura na estimação dos parâmetros do modelo. Esse primeiro estudo foi importante pois detectamos uma diferença entre as estimativas com e sem a censura, mostrando o impacto que ela tem no cálculo das estimativas.

Após a inserção das censuras, vimos a necessidade de uma abordagem que diferenciasse os indivíduos entre si para o cálculo do período de latência. Usamos como alternativa a inserção do estadiamento do paciente de duas formas: através de um modelo de fragilidade e inserindo

essa informação como covariável dentro do parâmetro.

A primeira forma foi útil para mostrar que a fragilidade dos pacientes é diferente dependendo do estadiamento que ele se encontra e, conseqüentemente, o risco de uma falha ocorrer aumenta quanto maior o valor da fragilidade.

Já a segunda forma nos permitiu calcular um período de latência para cada estadiamento, o que não era possível em nenhum dos outros modelos. Através dele observamos que quanto maior o estadiamento em que o paciente se encontrava, maior era o seu período de latência.

Para comparar os modelos utilizamos o critério Akaike (AIC) (Akaike (1974)), que busca o modelo com o melhor ajuste, sendo que o melhor modelo é aquele com menor AIC. Na Tabela 15 podemos comparar o AIC dos modelos apresentados, incluindo o modelo original proposto pelos autores. Observamos que o modelo original foi o que apresentou maior AIC, enquanto que o último modelo, cuja abordagem era o estadiamento como covariável no parâmetro de latência, foi o que apresentou menor AIC, sendo assim considerado o modelo com melhor ajuste.

Tabela 15 – AIC dos modelos

Modelo	AIC
Modelo Original	214,16
Modelo 1	196,96
Modelo 2	185,19
Modelo 3	179,41

6.1 Propostas Futuras

Como proposta futuras podemos avaliar a viabilidade de novas abordagens para a detecção do período de latência do câncer. Dentre elas temos o estudo de modelos aditivos com eventos recorrentes, com dois eventos de interesse: o início da doença e o diagnóstico. Uma outra opção é a abordagem bayesiana, em que uma distribuição a priori pode ser assumida para um dos parâmetros de interesse. Além disso, podemos considerar distribuições truncadas à esquerda para obter um período de latência estimado para cada indivíduo.

REFERÊNCIAS

AKAIKE, H. A new look at the statistical model identification. **IEEE transactions on automatic control**, Ieee, v. 19, n. 6, p. 716–723, 1974. Citado na página 64.

AMARAL, J.; PEREIRA, E.; PAIXÃO, M. Data and projections of hiv/aids cases in portugal: an unstoppable epidemic? **Journal of Applied Statistics**, Taylor & Francis, v. 32, n. 2, p. 127–140, 2005. Citado na página 29.

ARMENIAN, H. K. Incubation periods of cancer: old and new. **Journal of chronic diseases**, Elsevier, v. 40, p. 9S–15S, 1987. Citado na página 29.

BAKER, R. D. Use of a mathematical model to evaluate breast cancer screening policy. **Health Care Management Science**, Springer, v. 1, n. 2, p. 103–113, 1998. Citado na página 40.

BENTZEN, S. M.; THAMES, H. D.; OVERGAARD, M. Latent-time estimation for late cutaneous and subcutaneous radiation reactions in a single-follow-up clinical study. **Radiotherapy and Oncology**, Elsevier, v. 15, n. 3, p. 267–274, 1989. Citado na página 29.

BLUM, H. F.; GRADY, H. G.; KIRBY-SMITH, J. S. Limits of accuracy in experimental carcinogenesis as exemplified by tumor induction with ultraviolet radiation. **Journal of the National Cancer Institute**, Oxford University Press, v. 3, n. 1, p. 83–89, 1942. Citado na página 28.

_____. Relationships between dosage and rate of tumor induction by ultraviolet radiation. **Journal of the National Cancer Institute**, Oxford University Press, v. 3, n. 1, p. 91–97, 1942. Citado na página 28.

CANCER-RESEARCH-UK. **Incidence, survival and mortality - Cancer Resarch UK**. [S.l.], 2014. Acessado em: 24 de Outubro de 2014. Disponível em: <<http://www.cancerresearchuk.org/about-cancer/cancers-in-general/what-is-cancer/statistics/incidence-survival-and-mortality>>. Citado na página 23.

COBB, S.; MILLER, M.; WALD, N. On the estimation of the incubation period in malignant disease: the brief exposure case, leukemia. **Journal of chronic diseases**, Elsevier, v. 9, n. 4, p. 385–393, 1959. Citado na página 29.

COBRE, J.; PERDONÁ, G. S. C.; PERIA, F. M.; LOUZADA, F. A mechanistic breast cancer survival modelling through the axillary lymph node chain. **Statistics in medicine**, Wiley Online Library, v. 32, n. 9, p. 1536–1546, 2013. Citado na página 27.

COLLETTE, S.; BONNETAIN, F.; PAOLETTI, X.; DOFFOEL, M.; BOUCHE, O.; RAOUL, J.; ROUGIER, P.; MASSKOURI, F.; BEDENNE, L.; BARBARE, J. Prognosis of advanced hepatocellular carcinoma: comparison of three staging systems in two french clinical trials. **Annals of oncology**, Eur Soc Med Oncology, v. 19, n. 6, p. 1117–1126, 2008. Nenhuma citação no texto.

COLOSIMO, E. A.; GIOLO, S. R. Análise de sobrevivência aplicada. In: **ABE-Projeto Fisher**. [S.l.]: Edgard Blücher, 2006. Citado nas páginas 27, 44 e 53.

- HESSOL, N. A.; BYERS, R. H.; LIFSON, A. R.; O'MALLEY, P. M.; CANNON, L.; BARNHART, J. L.; HARRISON, J. S.; RUTHERFORD, G. W. Relationship between aids latency period and aids survival time in homosexual and bisexual men. **JAIDS Journal of Acquired Immune Deficiency Syndromes**, LWW, v. 3, n. 11, p. 1078–1085, 1990. Citado na página 29.
- HOUGAARD, P. **Analysis of multivariate survival data**. [S.l.]: Springer Science & Business Media, 2012. Citado na página 53.
- HOWARD, J. Minimum latency & types or categories of cancer. **Replaces administrators white paper on minimum latency & types of cancer. Centers for disease control and prevention**, p. 128–147, 2013. Citado nas páginas 24 e 58.
- INCA. **O que é o cancer?** [S.l.], 2014. Acessado em: 24 de Outubro de 2014. Disponível em: <http://www.inca.gov.br/conteudo/_view.asp?id=322>. Citado na página 23.
- LAWLESS, J. F. **Statistical models and methods for lifetime data**. [S.l.]: John Wiley & Sons, 2011. v. 362. Citado na página 36.
- LLOVET, J. M. Updated treatment approach to hepatocellular carcinoma. **Journal of gastroenterology**, Springer, v. 40, n. 3, p. 225–235, 2005. Citado na página 37.
- LOUZADA-NETO, F.; MAZUCHELI, J.; ACHCAR, J. A. **Uma introducao à análise de sobrevivencia e confiabilidade**. [S.l.]: Sociedad Chilena de Estadística, 2001. Citado nas páginas 25, 26 e 44.
- LUI, K.-J.; LAWRENCE, D. N.; MORGAN, W. M.; PETERMAN, T. A.; HAVERKOS, H. W.; BREGMAN, D. J. A model-based approach for estimating the mean incubation period of transfusion-associated acquired immunodeficiency syndrome. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 83, n. 10, p. 3051–3055, 1986. Citado na página 29.
- MARINACCIO, A.; BINAZZI, A.; CAUZILLO, G.; CAVONE, D.; ZOTTI, R. D.; FERRANTE, P.; GENNARO, V.; GORINI, G.; MENEGOZZO, M.; MENSI, C. *et al.* Analysis of latency time and its determinants in asbestos related malignant mesothelioma cases of the italian register. **European journal of cancer**, Elsevier, v. 43, n. 18, p. 2722–2728, 2007. Citado na página 30.
- MARTINEZ, E. Z.; ACHCAR, J. A.; JÁCOME, A. A.; SANTOS, J. S. Mixture and non-mixture cure fraction models based on the generalized modified weibull distribution with an application to gastric cancer data. **Computer methods and programs in biomedicine**, Elsevier, v. 112, n. 3, p. 343–355, 2013. Citado na página 27.
- MAZZANTI, R.; GRAMANTIERI, L.; BOLONDI, L. Hepatocellular carcinoma: epidemiology and clinical aspects. **Molecular aspects of medicine**, Elsevier, v. 29, n. 1, p. 130–143, 2008. Citado na página 37.
- NADLER, D. L.; ZURBENKO, I. G. Developing a weibull model extension to estimate cancer latency. **ISRN Epidemiology**, Hindawi Publishing Corporation, v. 2013, 2012. Citado nas páginas 24, 25, 27, 28, 30, 33, 40, 43, 44, 54, 59 e 63.
- _____. Estimating cancer latency times using a weibull model. **Advances in Epidemiology**, Hindawi Publishing Corporation, v. 2014, 2014. Citado nas páginas 25, 29, 30, 33, 36, 40, 43, 48, 53, 59 e 63.

- NIELSEN, G. G.; GILL, R. D.; ANDERSEN, P. K.; SØRENSEN, T. I. A counting process approach to maximum likelihood estimation in frailty models. **Scandinavian journal of Statistics**, JSTOR, p. 25–43, 1992. Citado na página 55.
- ORNSTEIN, M. C.; MUKHERJEE, S.; MOHAN, S.; ELSON, P.; TIU, R. V.; SAUNTHARARAJAH, Y.; KENDEIGH, C.; ADVANI, A.; KALAYCIO, M.; MACIEJEWSKI, J. P. *et al.* Predictive factors for latency period and a prognostic model for survival in patients with therapy-related acute myeloid leukemia. **American journal of hematology**, Wiley Online Library, v. 89, n. 2, p. 168–173, 2014. Citado na página 30.
- ORTEGA, E. M.; CORDEIRO, G. M.; CAMPELO, A. K.; KATTAN, M. W.; CANCHO, V. G. A power series beta weibull regression model for predicting breast carcinoma. **Statistics in medicine**, Wiley Online Library, v. 34, n. 8, p. 1366–1388, 2015. Citado na página 27.
- PETO, R.; LEE, P. Weibull distributions for continuous-carcinogenesis experiments. **Biometrics**, JSTOR, p. 457–470, 1973. Citado na página 27.
- PIKE, M. A method of analysis of a certain class of experiments in carcinogenesis. **Biometrics**, JSTOR, p. 142–161, 1966. Citado na página 27.
- SAUDE, a. Analise de Sobrevida-Teoria e Aplicacoes em. **Analise de Sobrevida - Teoria e Aplicacoes em Saude**. [S.l.], 2014. Acessado em: 4 de Outubro de 2014. Disponível em: <<http://sobrevida.fiocruz.br/material.html>>. Citado na página 25.
- TORRES, K. E.; RAVI, V.; KIN, K.; YI, M.; GUADAGNOLO, B. A.; MAY, C. D.; ARUN, B. K.; HUNT, K. K.; LAM, R.; LAHAT, G. *et al.* Long-term outcomes in patients with radiation-associated angiosarcomas of the breast following surgery and radiotherapy for breast cancer. **Annals of surgical oncology**, Springer, v. 20, n. 4, p. 1267–1274, 2013. Citado na página 30.
- WAHED, A. S.; LUONG, T. M.; JEONG, J.-H. A new generalization of weibull distribution with application to a breast cancer data set. **Statistics in medicine**, Wiley Online Library, v. 28, n. 16, p. 2077–2094, 2009. Citado na página 40.
- WORLD-HEALTH-STATISTICS. **World Health Statistics**. [S.l.], 2014. Acessado em: 24 de Outubro de 2014. Disponível em: <http://www.who.int/gho/publications/world_health_statistics/en/>. Citado na página 23.
- ZIVOT, E. **Maximum Likelihood Estimation**. [S.l.], 2009. Acessado em: 27 de Dezembro de 2015. Disponível em: <<http://faculty.washington.edu/ezivot/econ583/mleLectures.pdf>>. Citado na página 48.