

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
DEPARTAMENTO DE ESTATÍSTICA

**ABORDAGEM BAYESIANA PARA MODELOS
PARAMÉTRICOS NA ANÁLISE DE SOBREVIVÊNCIA**

Sabrina Letícia Couto da Silva

Orientadora: Profa. Dra. Patrícia Klarmann Ziegelmann

Monografia apresentada para a obtenção
do título de Bacharel em Estatística.

Porto Alegre, Dezembro de 2005.

AGRADECIMENTOS

A conclusão de um curso de graduação é apenas o primeiro passo no desenvolvimento pessoal e criativo. Aumentar a nossa criatividade significa abrir novas oportunidades o que é fundamental para enfrentar os desafios de nosso tempo.

Agradeço a todos os professores do curso de Estatística, que foram os principais responsáveis pela minha formação.

A meus familiares que sempre me apoiaram, meu pai Clóvis e minha mãe Gelsa pela educação primorosa, amor e dedicação; minhas duas irmãs, Michele e Jaqueline, obrigada pela paciência nas horas difíceis e pelo constante incentivo.

A todos os amigos que souberam entender a minha ausência devido aos estudos e aos colegas do curso de Estatística por terem me ensinado muito nas matérias: convivência, respeito e coleguismo.

À amiga, professora e orientadora Patrícia Klarmann Ziegelmann pela sua grande dedicação e conhecimento não só na realização desta monografia como também em disciplinas ao longo do curso, agradeço muito a ela pelo respeito e atenção com que sempre me tratou. Muito da minha formação está solidificado devido ao conhecimento e lições que me passastes. Obrigada por tudo!

A professora Jandyra Fachel pela orientação na bolsa de Iniciação Científica, por seu conhecimento e o aprendizado que fora de grande importância para a minha formação acadêmica e ainda, por sua dedicação e contribuição à ciência e à profissão.

Ao professor Daniel Furtado Ferreira, que mesmo à distância colaborou com informações para a confecção deste trabalho.

Ao Fernando pela sua dedicação e amor durante quase todo este curso, obrigado por estar sempre ao meu lado.

Agradeço a todos que, de alguma forma, colaboraram para que este momento chegasse.

E, acima de tudo, gostaria de agradecer a Deus por estar comigo ao longo de toda a minha vida, sempre me capacitando a enfrentar as barreiras e dificuldades encontradas ao longo do caminho.

SUMÁRIO

1 – INTRODUÇÃO.....	01
2 – ELEMENTOS DE ANÁLISE DE SOBREVIVÊNCIA.....	03
2.1 – Exemplos de Aplicações	04
2.2 – Definições Básicas	04
2.2.1 – Tempo de Sobrevivência	04
2.2.2 – Censura	06
2.2.3 – Função de Sobrevivência	08
2.2.4 – Função Risco	09
2.3 – Análise dos Dados de Sobrevivência	10
2.3.1 – Métodos Não-Paramétricos	11
2.3.2 – Modelo de Regressão de Cox	15
2.3.3 – Modelos Paramétricos	17
2.4 – Modelo Exponencial	17
2.5 – Modelo Weibull	19
3 – ABORDAGEM CLÁSSICA PARA MODELOS PARAMÉTRICOS.....	23
3.1 – Método de Máxima Verossimilhança	24
3.2 – Modelo Exponencial	25
3.2.1 – Modelo sem covariáveis	26
3.2.2 – Modelo com covariáveis	27
3.3 – Modelo Weibull	32
3.3.1 – Modelo sem covariáveis	32
3.3.2 – Modelo com covariáveis	33
4 – ABORDAGEM BAYESIANA PARA MODELOS PARAMÉTRICOS.....	37
4.1 – Definições Básicas de Inferência Bayesiana	38
4.1.1 – Distribuição a Posteriori	39
4.1.2 – Distribuição a Priori	40
4.1.3 – Priori Conjugada	41
4.1.4 – Priori Não-Informativa	41
4.1.5 – Priori Imprópria	42
4.1.6 – Preditiva a Posteriori	42
4.1.7 – Resultados a Posteriori	43
4.1.8 – Métodos MCMC	45

4.2 – Modelo Exponencial	47
4.2.1 – Modelo Exponencial sem covariáveis	47
4.2.2 – Modelo Exponencial com covariáveis	49
4.3 – Modelo Weibull	51
4.3.1 – Modelo Weibull sem covariáveis	52
4.3.2 – Modelo Weibull com covariáveis	53
5 – EXEMPLO DE APLICAÇÃO.....	54
5.1 – Descrição do Exemplo	54
5.2 – Modelo Exponencial	54
5.2.1 – Análise Clássica	55
5.2.2 – Análise Bayesiana	63
5.3 – Modelo Weibull	77
5.3.1 – Análise Bayesiana	78
6 – CONCLUSÕES.....	81
7 – BIBLIOGRAFIA.....	83

1. INTRODUÇÃO

A análise de sobrevivência é uma das áreas da Estatística que mais tem crescido nos últimos 20 anos. Este crescimento pode ser observado tanto no desenvolvimento e aprimoramento de métodos estatísticos, quanto no número de aplicações. Atualmente, a sua importância na área médica se revela cada vez mais, porém abrangendo mais áreas do conhecimento, tais como Engenharia (Confiabilidade), Ciências Sociais (Demografia), etc.

Em análise de sobrevivência a variável resposta é, geralmente, o tempo até ocorrência de um evento de interesse. Este tempo é denominado como o tempo de falha, podendo ser o tempo até a morte de um paciente, tempo de vida de um componente elétrico, estudos de coortes específicas, etc. A principal característica de dados de sobrevivência é a presença de censura, que é a observação parcial da variável resposta. Por exemplo, por alguma razão o acompanhamento do paciente pode ser interrompido antes do término do estudo ou o estudo é finalizado antes da morte do paciente. Usualmente, sem a presença de censura, as técnicas estatísticas clássicas, como Análise de Regressão e Planejamento de Experimentos, por exemplo, poderiam ser utilizadas na análise desse tipo de dados.

O objetivo desse trabalho é apresentar e ilustrar a análise de dados de sobrevivência sob o enfoque Bayesiano e estabelecer comparação com a abordagem clássica, extensamente utilizada e descrita na literatura.

A inferência Bayesiana paramétrica é o processo de ajuste de um modelo probabilístico para um conjunto de dados resumindo os resultados através de uma distribuição de probabilidade para os parâmetros do modelo, dado que os parâmetros são vistos como variáveis aleatórias. Juntando-se a informação a priori a respeito do parâmetro com a função de verossimilhança, deriva-se a função de probabilidade a posteriori para os parâmetros desconhecidos a qual resume toda a informação atual (após observar os dados) a respeito dos parâmetros. Com a distribuição a posteriori é possível obter inferências para os parâmetros desconhecidos do modelo com base na informação prévia e nos dados.

Assim como a análise clássica de dados de sobrevivência, há a abordagem Bayesiana de dados de sobrevivência, que tem recebido atenção especial devido aos avanços das técnicas computacionais e de modelagem. Diversos trabalhos têm sido desenvolvidos nesta área onde podemos citar o livro de Klein e Moeschberger (1997). O assunto é muito extenso e por este motivo serão focados neste trabalho apenas alguns modelos paramétricos e seus métodos de análise baseado em Chen *et al* (2001).

A abordagem clássica para modelos paramétricos também será apresentada com o objetivo de comparação com a Bayesiana e teve como base Hosmer e Lemeshow (1999), Colosimo (2000) e Lee e Wang (2003).

Neste trabalho serão apresentados os modelos paramétricos Exponencial e Weibull no contexto da análise de dados de sobrevivência. As abordagens clássica e Bayesiana serão comparadas empiricamente através de um exemplo de dados reais.

No Capítulo 2 serão introduzidas algumas definições de análise de sobrevivência clássica e uma apresentação dos modelos paramétricos que serão abordados no trabalho: Exponencial e Weibull. O Capítulo 3 dedica-se à abordagem clássica e nele serão mostrados os estimadores de máxima verossimilhança para os parâmetros dos modelos, considerando tanto os modelos sem incluir covariáveis quanto para modelo com inclusão de covariáveis. O Capítulo 4 refere-se à análise Bayesiana onde serão apresentadas algumas idéias e definições básicas de Inferência Bayesiana, bem como para cada um dos dois modelos, a análise considerando inclusão e não inclusão de covariáveis no modelo. Para cada modelo, serão apresentadas as derivações das distribuições a posteriori para os respectivos parâmetros. No Capítulo 5 será mostrado um exemplo de aplicação das abordagens clássica e Bayesiana para os modelos Exponencial e Weibull. Será utilizado um banco de dados reais para os dois modelos a título de ilustração dos mesmos.

Finalmente, o Capítulo 6 apresentará algumas considerações finais sobre o trabalho.

2. ELEMENTOS DE ANÁLISE DE SOBREVIVÊNCIA

A análise de sobrevivência é uma técnica estatística que visa analisar dados onde a variável de interesse é o tempo até a ocorrência de um evento. Esse tempo é denominado tempo de sobrevivência, podendo ser, por exemplo, o tempo até a morte de um paciente, bem como até a cura ou reincidência de uma doença. Em geral, estudos de dados de sobrevivência têm por objetivo estimar médias de tempo de sobrevivência e comparar curvas de sobrevivência estimadas para diferentes grupos e identificar fatores prognósticos relacionados com a resposta. Ainda, se analisarmos de outro ponto de vista, a distribuição do tempo de sobrevivência usualmente é assimétrica, logo seria mais adequado estimar o tempo mediano de sobrevivência. Desta forma, entende-se por dados de sobrevivência não só dados da variável resposta (tempo de sobrevivência) como também dados de covariáveis associadas tais como tratamento, idade, sexo, etc.

A principal característica de dados de sobrevivência é a presença de censura, que é a observação parcial da variável resposta. Como análise preliminar de dados de sobrevivência pode-se citar o estimador não-paramétrico de Kaplan-Meier para a função de sobrevivência e o teste log-rank para comparar funções de sobrevivência. Quando o estudo envolve covariáveis, uma forma eficiente de incluí-las nas análises é utilizando um modelo de regressão. Nessa linha, destaca-se o modelo de regressão de Cox que é um modelo semi-paramétrico e os modelos paramétricos.

Nesse capítulo, inicialmente, na Seção 2.1 serão apresentados alguns tipos de situações onde a técnica de análise de sobrevivência se aplica. A Seção 2.2 segue com algumas definições de elementos básicos constituintes da análise de dados de sobrevivência. Na Seção 2.3 serão citados os métodos não-paramétricos de análise de dados de sobrevivência e uma breve introdução sobre o Modelo de Regressão de Cox. Nas Seções 2.4 e 2.5 serão apresentados, respectivamente, os modelos Exponencial e Weibull.

2.1. EXEMPLOS DE APLICAÇÃO

Existem diversos exemplos de aplicação das técnicas de análise de sobrevivência nas mais diversas áreas do conhecimento.

Na área médica, são bastante utilizadas na identificação de fatores de prognóstico para doenças, bem como na comparação de tratamentos. Em oncologia, por exemplo, qualquer nova droga ou tratamento para o controle do câncer requer um estudo, no qual a variável resposta de interesse é, na maioria dos casos, o tempo de sobrevivência dos pacientes, chamado de sobrevivência global pelos especialistas dessa área. Estudos epidemiológicos da AIDS são outros exemplos, nos quais as técnicas de análise de sobrevivência vêm sendo aplicadas com frequência.

Na área da sociologia, criminalistas estudam o tempo entre a liberação de presos e a ocorrência de crimes; estudos na área trabalhista podem se concentrar em mudança de empregos, desempregos e aposentadorias; os demógrafos podem querer estudar nascimentos, mortes, migrações.

Outra área de bastante aplicação é a da Engenharia. Nesta área, estudos de sobrevivência são chamados de estudos de Confiabilidade, onde o tempo envolvido é o tempo de falha de equipamentos eletrônicos, componentes, etc.

2.2. DEFINIÇÕES BÁSICAS

Nesta seção serão apresentados e definidos alguns elementos e terminologias básicas do contexto de Análise de Sobrevivência.

2.2.1. TEMPO DE SOBREVIVÊNCIA

Refere-se ao tempo observado desde o início do estudo até a ocorrência de uma resposta de interesse no estudo. É a variável resposta de um estudo de sobrevivência. Podemos observar três elementos que constituem o tempo de sobrevivência:

a) Tempo Inicial

É o tempo de início do estudo. Deve-se ter cuidado para não confundir o tempo inicial com a data que o indivíduo entra no estudo, isto é, durante o período pré-determinado para o estudo pode-se observar diferentes indivíduos com diferentes datas de inclusão no estudo.

Outro aspecto também importante é que, a menos de conhecidas diferenças nas variáveis medidas, todos os indivíduos deverão ser comparados na origem do estudo, ou seja, deve-se ter cuidado para não incluir na análise variáveis que dependam do tempo de ingresso do indivíduo no estudo. Entretanto, existem algumas variáveis com essa característica: variáveis tempo-dependentes.

b) Escala de Medida

Refere-se ao tempo real de observação, isto é, pode-se definir estudos onde o tempo observado é o número de anos que o indivíduo levou até a ocorrência do evento de interesse, ou ainda o número de semestres, número de meses, etc. Em testes de engenharia, podem surgir outras escalas de medida como, por exemplo, o número de ciclos, a quilometragem de um automóvel ou qualquer medida de carga.

c) Evento de Interesse ou “Falha”

Refere-se ao evento de interesse que, na maioria dos casos, são indesejáveis, como a morte de um paciente, a queima de um equipamento eletrônico, a evasão de um aluno da escola, etc. Porém, em alguns estudos o evento de interesse é algo desejável como, por exemplo, a cura de um paciente. Independente de ser algo indesejado ou desejado, o acontecimento do evento de interesse é, muitas vezes, chamado de falha. Esse evento deve ser bem definido antes de iniciar o estudo a fim de não haver dúvidas sobre o que considerar como falha.

O enfoque apresentado neste trabalho será na aplicação da análise de sobrevivência na área médica, por esse motivo será sempre utilizada a expressão tempo de sobrevivência.

Após a definição dos elementos citados anteriormente, é possível determinar o tempo de sobrevivência também chamado de tempo de falha ou o tempo de morte.

2.2.2. CENSURA

Partiremos sempre da hipótese de que as observações disponíveis para os n indivíduos são independentes. Ao se analisar dados de sobrevivência, muitas vezes ocorrem algumas informações incompletas sobre o tempo de sobrevivência, caracterizando assim, a presença de dados censurados. Isso se deve ao fato de que os estudos clínicos, por exemplo, os quais envolvem uma resposta temporal, são na sua maioria prospectivos e de longa duração. Porém, apesar de longos, esses estudos costumam terminar antes que todos os indivíduos envolvidos no estudo venham a falhar.

A censura é definida como sendo a não-observação do tempo exato de sobrevivência. O que se tem de informação para os indivíduos censurados é uma observação imprecisa do tempo de sobrevivência chamada de tempo de censura. Note que, se considerarmos essas unidades como informações completas (tempos exatos de sobrevivência) podemos obter resultados viciados para o tempo médio de sobrevivência. Ou ainda, se não considerarmos essas observações censuradas nas análises a serem realizadas, estaremos produzindo resultados menos eficientes pois o tamanho da amostra estará reduzido.

São geralmente três razões pelas quais a censura pode ocorrer:

- i) O indivíduo não apresenta o evento de interesse antes do fim do estudo;
- ii) O indivíduo é “perdido” (muda de endereço, mudança de hospital, perde o interesse no estudo, etc.) durante o período de estudo;
- iii) O indivíduo retirou-se do estudo porque morreu (isso se a morte não for o evento de interesse, ou se morrer de outra razão diferente da estudada, etc.) ou alguma outra razão.

A Figura 2.1 mostra os diferentes tipos de mecanismos de censura existentes.

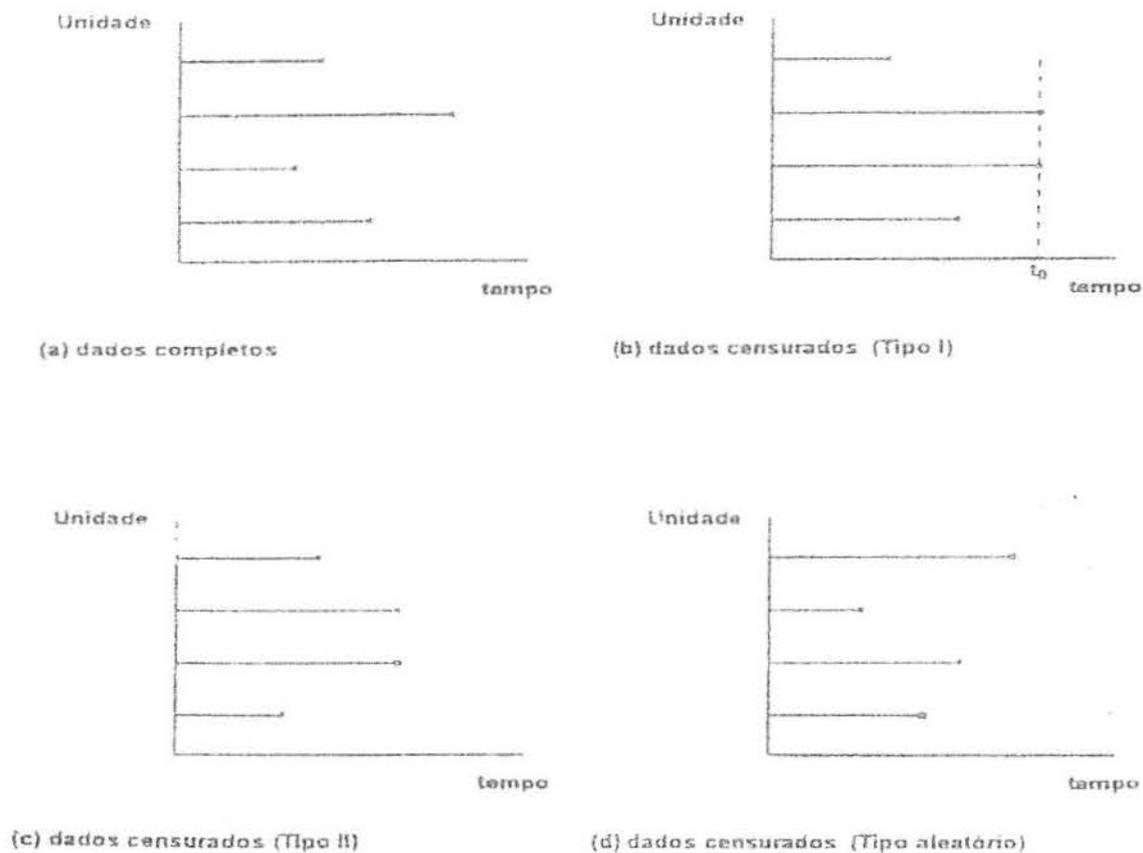


Figura 2.1 - Gráfico dos diferentes mecanismos de censura

Os mecanismos de censura apresentados na Figura 2.1 são conhecidos por censura à direita, pois o evento de interesse está à direita do tempo registrado. Essa situação é muito encontrada em estudos envolvendo dados de sobrevivência. Apesar disso, outros tipos de censura podem ocorrer: censura à esquerda e censura intervalar.

A censura à esquerda acontece quando o tempo registrado é maior do que o tempo de sobrevivência, isto é, o evento de interesse já ocorreu quando o indivíduo foi observado.

A censura intervalar é um tipo mais geral de censura que ocorre quando se sabe apenas que o tempo de sobrevivência ocorreu num determinado intervalo.

Enfim, todos os dados de um estudo de sobrevivência, até mesmo os censurados, devem ser incluídos na análise estatística pois mesmo sendo

informação incompleta, os dados censurados fornecem informações a respeito do tempo de sobrevivência dos pacientes. Porém, deve-se ter o cuidado de considerar um modelo onde os dados de sobrevivência sejam considerados de maneira diferente do que os dados de censura.

2.2.3. FUNÇÃO DE SOBREVIVÊNCIA

É uma função probabilística utilizada para descrever os estudos de sobrevivência. A função de sobrevivência é fundamental para a análise de sobrevivência, pois obtém probabilidades de sobrevivência para diferentes valores do tempo.

Seja T_i o tempo de sobrevivência para o indivíduo i , e seja C_i o tempo de censura para o indivíduo i . É claro que somente o menor de tais tempos, ou seja, $\min(T_i, C_i)$, é observado. Outra hipótese a considerar é que os tempos T_i e C_i são independentes. O tempo de sobrevivência T é uma variável aleatória contínua não-negativa.

Seja $f(t)$ a função densidade de probabilidade de T e $F(t)$ a sua respectiva função de distribuição acumulada.

A função de sobrevivência $S(t)$ é definida como sendo a probabilidade de um indivíduo sobreviver até o tempo t , ou ainda, é a probabilidade de um indivíduo não falhar até o tempo t . A função de sobrevivência é dada por:

$$S(t) = 1 - F(t) = P(T \geq t), \quad (2.1)$$

e possui as seguintes características:

- i) $S(t)$ é uma função decrescente;
- ii) No tempo $t = 0$, $S(t) = 1$, isto é, no início do estudo, desde que não possa ter ocorrido um evento ainda, a probabilidade de sobreviver é igual a 1.
- iii) $\lim_{t \rightarrow \infty} S(t) = 0$.

A função de sobrevivência $S(t)$ é utilizada para estimar características dos tempos de sobrevivência tais como a média, mediana e outros percentis. Ela também é utilizada para efetuar a comparação entre os tempos de sobrevivência de dois ou mais grupos.

2.2.4. FUNÇÃO RISCO

A probabilidade de um indivíduo falhar em um intervalo de tempo $[t_1; t_2)$ é dada pela diferença entre as funções de sobrevivência para cada tempo do intervalo, ou seja,

$$S(t_1) - S(t_2).$$

A taxa de risco no intervalo de tempo $[t_1; t_2)$ é dada pela probabilidade de que a falha ocorra neste intervalo de tempo, dado que não ocorreu antes de t_1 , dividindo-se pelo comprimento do intervalo. Dessa forma, a taxa de falha em $[t_1; t_2)$ é dada por:

$$\frac{S(t_1) - S(t_2)}{(t_2 - t_1)} \times \frac{1}{S(t_1)}$$

De modo geral, definindo o intervalo como $[t; t + \Delta t)$, a expressão acima definida fica da seguinte forma:

$$\frac{S(t) - S(t + \Delta t)}{\Delta t S(t)}$$

Ao definir Δt como um valor muito pequeno, $h(t)$ representará a função risco instantâneo ao tempo t , condicionada à sobrevivência até o tempo t . Ou seja,

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

A função risco ou função "hazard" em Inglês é também conhecida como função taxa de falha.

Existe uma relação matemática entre a função risco $h(t)$ e a função de sobrevivência $S(t)$ que é dada por:

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt}[-\log S(t)] \quad (2.2)$$

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

Uma outra relação bastante utilizada em Análise de Sobrevivência é a que relaciona a função de sobrevivência com a função risco acumulada $H(t)$, isto é,

$$H(t) = \int_0^t h(u).du = -\log(S(t))$$

2.3. ANÁLISE DE DADOS DE SOBREVIVÊNCIA

Por mais complexo que seja o problema, a resposta às perguntas de interesse será dada a partir de um conjunto de dados de sobrevivência. O primeiro passo para uma análise estatística desses dados é realizar uma análise descritiva. Para isso, existem técnicas não-paramétricas que são utilizadas para descrever os dados de sobrevivência, sendo de fácil aplicação pois não envolvem nenhuma estrutura paramétrica. Dentre as técnicas não-paramétricas destacam-se o estimador de Kaplan-Meier para a função de sobrevivência e o teste log-rank para comparar funções de sobrevivência. Entretanto, essas análises não permitem analisar modelos mais complexos nos quais incluem-se covariáveis. A forma mais apropriada para analisar o efeito de covariáveis é utilizar um modelo de regressão adequado para dados censurados. Os modelos de regressão para dados de sobrevivência dividem-se em modelos paramétricos e modelos semi-paramétricos. Na linha de modelos semi-paramétricos, o modelo de regressão de Cox é o mais utilizado em estudos clínicos por sua versatilidade. Esse modelo permite incorporar facilmente covariáveis, e uma breve introdução sobre ele será apresentada nessa seção. Entre os modelos

paramétricos citamos o Exponencial e o Weibull, cuja abordagem clássica será vista no Capítulo 3 e a abordagem Bayesiana no Capítulo 4.

2.3.1. MÉTODOS NÃO-PARAMÉTRICOS

Os objetivos de uma análise estatística envolvendo dados de sobrevivência geralmente estão relacionados na área médica com a identificação de fatores prognósticos para doenças ou a comparação de tratamentos em um estudo clínico com o controle de outros fatores. Para realizar uma análise desse tipo, primeiramente precisamos realizar uma descrição dos dados. As técnicas estatísticas não-paramétricas comumente utilizadas em análise de sobrevivência para esse fim são o estimador de Kaplan-Meier e o teste log-rank.

Estimador de Kaplan-Meier

O Kaplan-Meier é um estimador não-paramétrico para a função de sobrevivência. Ele foi proposto por Kaplan e Meier (1958) e é também chamado de estimador limite-produto. Trata-se de uma adaptação da função de sobrevivência empírica que, na ausência de censuras é definida como sendo:

$$\hat{S}(t) = \frac{\text{n}^\circ \text{ de observações que não falharam até o tempo } t}{\text{n}^\circ \text{ de observações no estudo}} \quad (2.3)$$

onde $\hat{S}(t)$ é uma função escada com degraus nos tempos observados de sobrevivência de tamanho $1/n$, onde n é o tamanho da amostra. Se existirem empates em um certo tempo t , o tamanho do degrau fica multiplicado pelo número de empates. O estimador de Kaplan-Meier considera tantos intervalos de tempo quantos forem o número de falhas distintas. Os limites dos intervalos de tempo são os tempos de sobrevivência da amostra. A função de sobrevivência $S(t)$ pode ser escrita em termos de probabilidades condicionais. Suponha que existem n pacientes no estudo e k ($k \leq n$) falhas distintas nos tempos $t_1 < t_2 < \dots < t_k$. Então,

$$S(t) = (1 - q_1)(1 - q_2) \dots (1 - q_i), \quad t \in [t_{i-1}; t_i) \quad (2.4)$$

onde q_i é a probabilidade de um indivíduo falhar no intervalo $[t_{i-1}; t_i)$ sabendo que não falhou até t_{i-1} e $t_0 = 0$. Ou seja, podemos escrever t_i como:

$$q_i = P(T \in [t_{i-1}, t_i) | T \geq t_{i-1}). \quad (2.5)$$

Desta forma escrevemos a expressão geral de $S(t)$ em termos de probabilidades condicionais. O estimador de Kaplan-Meier se reduz então, a estimar q_i adaptado da expressão (2.3) é dado por:

$$\hat{q}_i = \frac{n^\circ \text{ de falhas em } [t_{i-1}; t_i)}{n^\circ \text{ sob risco em } t_{i-1}} \quad (2.6)$$

para $i = 1, 2, \dots, k$.

A expressão geral para o estimador de Kaplan-Meier pode ser apresentada após essas considerações preliminares. Vamos introduzir a notação:

- d_i = número de falhas no tempo t_i ;
- n_i = número de observações sob risco (não falhou e não é censurado) até o tempo t_i (exclusive).

Assim, o estimador de Kaplan-Meier de $S(t)$ é então definido como:

$$\hat{S}(t) = \prod_{i|t_i < t} \left(\frac{n_i - d_i}{n_i} \right) = \prod_{i|t_i < t} \left(1 - \frac{d_i}{n_i} \right). \quad (2.7)$$

Uma justificativa para a expressão (2.7) vem da decomposição de $S(t)$ em termos dos q_i 's apresentada em (2.4). O estimador de Kaplan-Meier é obtido a partir de (2.5) se estimarmos os q_i 's por d_i/n_i que foi expresso em palavras em (2.6). No artigo original, Kaplan e Meier justificam a expressão (2.7) mostrando que ela é o estimador de máxima verossimilhança de $S(t)$.

Teste Log-Rank

Considere, inicialmente, o teste de duas funções de sobrevivência $S_1(t)$ e $S_2(t)$. Sejam $t_1 < t_2 < \dots < t_k$ os tempos de sobrevivência distintos da amostra formada pela combinação das duas amostras individuais.

As hipóteses do teste são:

H_0 : $S_1(t) = S_2(t)$ (os tratamentos são igualmente efetivos)

H_1 : $S_1(t) > S_2(t)$ (o tratamento 1 é mais efetivo que o 2).

Suponha que no tempo t_j acontecem d_j mortes e n_j indivíduos estão sob risco em um tempo imediatamente inferior a t_j na amostra combinada e, respectivamente, d_{ij} e n_{ij} na amostra i ; $i = 1, 2$ e $j = 1, 2, \dots, k$. Em cada tempo de sobrevivência, os dados podem ser dispostos em uma tabela de contingência 2×2 com d_{ij} mortes e $n_{ij} - d_{ij}$ sobreviventes na coluna i . Isso é mostrado na Tabela 2.1. Condicional à experiência de morte e censura até o tempo t_j (fixando as marginais de coluna) e ao número de mortes no tempo t_j (fixando as marginais de linha), a distribuição de d_{2j} é dada por:

$$\frac{\binom{n_{1j}}{d_{1j}} \binom{n_{2j}}{d_{2j}}}{\binom{n_j}{d_j}} \quad (2.8)$$

onde a expressão (2.8) é de uma distribuição Hipergeométrica.

A média de d_{2j} é dada por:

$$w_{2j} = n_{2j} d_j n_j^{-1},$$

o que equivale dizer que, se não houver diferença entre as duas populações no tempo t_j , o número total de mortes (d_j) pode ser dividido entre as duas amostras de acordo com a razão entre o número de indivíduos sob risco em cada amostra e o número total sob risco. A variância de d_{2j} é dada por:

$$(V_j)_2 = n_{2j} (n_j - n_{2j}) d_j (n_j - d_j) n_j^{-2} (n_j - 1)^{-1}.$$

Assim, a estatística $d_{2j} - w_{2j}$ tem média zero e variância $(V_j)_2$. Se as k tabelas de contingência forem independentes, um teste aproximado para a igualdade das duas funções de sobrevivência é dado por:

$$T = \frac{\left[\sum_{j=1}^k (d_{2j} - w_{2j}) \right]^2}{\sum_{j=1}^k (V_j)_2},$$

onde T tem uma distribuição qui-quadrado com 1 grau de liberdade para grandes amostras.

Tabela 2.1 – Tabela de Contingência gerada no tempo t_j .

	Grupos		t_j
	1	2	
Morre	d_{1j}	d_{2j}	d_j
Não morre	$n_{1j} - d_{1j}$	$n_{2j} - d_{2j}$	$n_j - d_j$
	n_{1j}	n_{2j}	n_j

A generalização do teste log-rank para a igualdade de $r > 2$ funções de sobrevivência $S_1(t), \dots, S_r(t)$ não é complicada. Considere a mesma notação anterior, com o índice i agora variando entre 1 e r . Assim, os dados são arranjados em tabela de contingência $2 \times r$ com d_{ij} mortes e $n_{ij} - d_{ij}$ sobreviventes na coluna i . Isto é, a Tabela 2.1 passa a ter r colunas. Condicional à experiência de morte e censura até o tempo t_j e ao número de mortes no tempo t_j , a distribuição conjunta de $d_{1j}, d_{2j}, \dots, d_{rj}$ será uma distribuição Hipergeométrica multivariada dada pela expressão:

$$\frac{\prod_{i=1}^r \binom{n_{1j}}{d_{1j}}}{\binom{n_j}{d_j}}$$

A média de d_{ij} é:

$$w_{ij} = n_{ij} d_j n_j^{-1},$$

A variância de d_{ij} e a covariância de d_{ij} são, respectivamente:

$$(V_j)_{ii} = n_{ij}(n_j - n_{ij})d_j(n_j - d_j)n_j^{-2}(n_j - 1)^{-1}$$

e

$$(V_j)_{ii} = -n_{ij}n_{ij}d_j(n_j - d_j)n_j^{-2}(n_j - 1)^{-1}.$$

Então, a estatística $v'_j = (d_{1j} - w_{1j}, \dots, d_{rj} - w_{rj})$ tem média zero e matriz de variância-covariância V_j .

Assim, podemos formar a estatística v , somando sobre todos os tempos de morte,

$$v = \sum_j^k v_j,$$

onde v é um vetor de dimensão $r \times 1$ cujos elementos são as diferenças entre os totais observados e esperados de morte.

Considerando novamente a suposição de que as k tabelas de contingência são independentes, a variância da estatística v será $V = V_1 + \dots + V_k$. Um teste aproximado para a igualdade de r funções de sobrevivência é dado por:

$$T = v'Vv^{-1},$$

onde T tem uma distribuição qui-quadrado com $r - 1$ graus de liberdade para grandes amostras.

2.3.2. MODELO DE REGRESSÃO DE COX

O modelo de regressão de Cox permite a análise de dados provenientes de estudos de tempo de vida em que a variável resposta é o tempo até a ocorrência do evento de interesse, ajustando por covariáveis.

De uma forma genérica, vamos considerar p covariáveis e assim, \mathbf{x} será um vetor com os componentes (x_1, x_2, \dots, x_p) . A expressão geral do modelo de regressão de Cox considera,

$$h(t) = h_0(t)g(\mathbf{x}'\boldsymbol{\beta}) \quad (2.9)$$

onde g é uma função que deve ser especificada, tal que $g(0) = 1$. Este modelo é composto pelo produto de dois componentes, um não-paramétrico e outro paramétrico. O componente não-paramétrico, $h_0(t)$, não é especificado e é uma função não-negativa do tempo. Ele é usualmente chamado de função de base pois $h(t) = h_0(t)$ quando $\mathbf{x} = 0$. O componente paramétrico é freqüentemente usado na seguinte forma multiplicativa:

$$g(\mathbf{x}'\boldsymbol{\beta}) = \exp(\mathbf{x}'\boldsymbol{\beta}) = \exp(\beta_1 x_1 + \dots + \beta_p x_p), \quad (2.10)$$

onde $\boldsymbol{\beta}$ é o vetor de parâmetros associado às covariáveis. Essa forma garante que $h(t)$ será positiva.

Este modelo é chamado de modelo de riscos proporcionais, pois a razão das taxas de riscos de dois indivíduos diferentes é constante ao longo do tempo. Isto é, a razão das funções risco para dois indivíduos i e j é dada por:

$$\begin{aligned} \frac{h_i(t)}{h_j(t)} &= \frac{h_0(t)\exp(\mathbf{x}_i'\boldsymbol{\beta})}{h_0(t)\exp(\mathbf{x}_j'\boldsymbol{\beta})} \\ \frac{h_i(t)}{h_j(t)} &= \exp(\mathbf{x}_i'\boldsymbol{\beta} - \mathbf{x}_j'\boldsymbol{\beta}) \end{aligned}$$

que não depende do tempo. Por exemplo, se um indivíduo no início do estudo tem um risco de morte igual a duas vezes o risco de um segundo indivíduo, então essa razão de risco será a mesma para todo o período de acompanhamento.

O modelo de regressão de Cox é utilizado extensivamente em estudos médicos. A principal razão para a sua popularidade é a presença do componente não-paramétrico que torna o modelo bastante flexível. Apesar de sua popularidade, o modelo de Cox não será explorado com mais detalhes por não ser o objetivo desta

monografia. Mais detalhes sobre a abordagem clássica para modelo de Cox podem ser vistos em Hosmer e Lemeshow (1999) e Cox e Oakes (1984). Em relação a abordagem Bayesiana citamos Kalbfleisch, J.D. *et al* (1973) e Chen *et al* (2001).

2.3.3. MODELOS PARAMÉTRICOS

Modelos paramétricos caracterizam-se pela atribuição de um modelo probabilístico para a variável resposta.

A análise de dados de sobrevivência pode ser realizada seguindo a modelagem paramétrica em casos onde a distribuição do tempo de sobrevivência dos indivíduos possa ser supostamente conhecida.

Apesar de existirem diversos modelos probabilísticos que são utilizados na análise de dados de sobrevivência, alguns são mais destacados devido à sua comprovada adequação a diversas situações práticas tais como: Exponencial, Weibull, Log-Normal, Gama Generalizada e Log-Logístico. Nesta monografia apresentaremos o Exponencial e o Weibull. Nas seções 2.4 e 2.5 serão apresentadas as características de cada um destes modelos, tais como, suas funções densidade de probabilidade, funções de sobrevivência, função risco, etc.

Técnicas de inferência para esses modelos serão apresentadas nas próximas seções.

2.4. MODELO EXPONENCIAL

Considere que T seja a variável tempo de sobrevivência de um indivíduo. Assumindo-se que T tenha distribuição Exponencial de parâmetro λ , a sua função densidade de probabilidade é dada por:

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & , t \geq 0, \lambda > 0 \\ 0 & , t < 0 \end{cases} \quad (2.11)$$

A distribuição Exponencial é caracterizada por um risco constante λ , sendo seu único parâmetro. Um valor alto para λ indica um alto risco e curta sobrevivência, já um valor pequeno para λ indica um risco pequeno e longa sobrevivência. A média e a variância do tempo de sobrevivência T são, respectivamente, dadas por:

$$E(T) = \frac{1}{\lambda} \quad e \quad Var(T) = \frac{1}{\lambda^2}.$$

A função distribuição acumulada de T é dada por,

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

e a função de sobrevivência é dada por,

$$S(t) = 1 - F(t) = e^{-\lambda t}, \quad t \geq 0$$

Dessa forma, usando (2.2) temos que, para o modelo Exponencial, a função risco é dada por,

$$h(t) = \lambda, \quad t \geq 0,$$

uma função constante, independente do valor de t .

A Figura 2.2 mostra respectivamente: (a) a função de sobrevivência para o modelo Exponencial, (b) a função distribuição de probabilidade $f(t)$ do modelo Exponencial e (c) a função risco do modelo Exponencial.

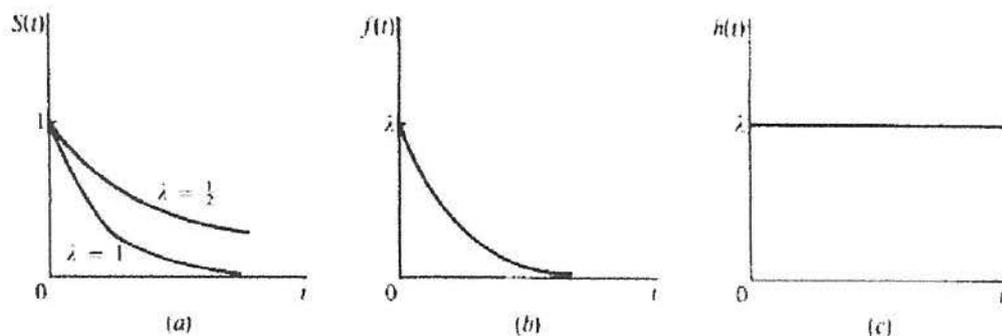


Figura 2.2 - Gráficos da função sobrevivência, densidade e risco do modelo Exponencial.

Quando $\lambda = 1$, a distribuição é comumente chamada de Distribuição Exponencial Unitária.

A hipótese de risco constante pode significar que, depois que uma peça estiver em uso, por exemplo, a sua probabilidade de falhar não se altera. De modo menos rigoroso, não existe efeito de “desgaste” quando o modelo Exponencial é estipulado. Nesse sentido, pode-se dizer que a distribuição Exponencial admite que a probabilidade de falhar seja independente do que se tenha passado. Isto é, enquanto a peça estiver funcionando, ela será “tão boa quanto nova”.

O modelo Exponencial não parece muito adequado para tempo de vida de pessoas, pois não é muito plausível considerar um risco constante para tempo de vida de pessoas.

2.5. MODELO WEIBULL

A distribuição de Weibull foi proposta por W. Weibull (1951) em estudos relacionados ao tempo de falha devido à fadiga de metais. É freqüentemente utilizada para descrever o tempo de vida de produtos industriais (área de confiabilidade). Sua grande aplicação em situações práticas deve-se ao fato de apresentar uma grande variedade de formas, todas com uma propriedade básica: a função taxa de falha é monótona, ou seja, ou ela é crescente ou decrescente ou constante. Descreve, de forma adequada, a vida de componentes eletrônicos, cerâmicas, capacitores, etc.

A distribuição de Weibull pode ser considerada uma generalização da distribuição Exponencial. Entretanto, ao contrário da distribuição Exponencial, não assume risco constante ao longo do tempo t .

Caracteriza-se por dois parâmetros, γ e λ . O valor de γ determina a forma da curva de distribuição e o valor de λ determina a escala. Conseqüentemente, γ e λ são ditos parâmetros de forma e escala, respectivamente.

Quando o tempo de sobrevivência T segue a distribuição de Weibull com parâmetros λ e γ , a função densidade de probabilidade de T é dada por:

$$f(t) = \lambda \gamma (\lambda t)^{\gamma-1} e^{-(\lambda t)^\gamma}, \quad t \geq 0, \gamma, \lambda > 0, \quad (2.12)$$

com função distribuição acumulada dada por:

$$F(t) = 1 - e^{-(\lambda t)^\gamma}.$$

A Figura 2.3 mostra algumas curvas da função densidade Weibull com parâmetro de escala $\lambda = 1$ variando os valores para o parâmetro de forma γ .

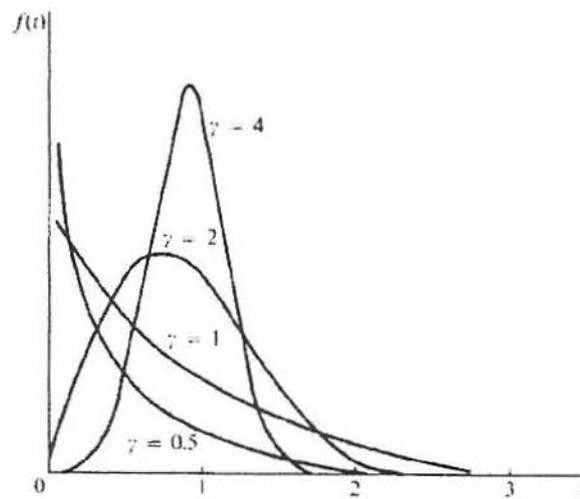


Figura 2.3 - Gráficos da densidade Weibull para diversos valores de γ .

A média da distribuição Weibull de parâmetros λ e γ é dada por,

$$\mu = \frac{\Gamma(1+1/\gamma)}{\lambda}$$

com variância,

$$\sigma^2 = \frac{1}{\lambda^2} [\Gamma(1+2/\gamma) - \Gamma^2(1+1/\gamma)]$$

onde $\Gamma(\beta)$ é a função matemática Gama definida como:

$$\Gamma(\beta) = \int_0^{\infty} x^{\beta-1} e^{-x} dx = (\beta-1)! \quad , \beta > 0$$

A função de sobrevivência de T será dada por:

$$S(t) = e^{-(\lambda t)^\gamma}.$$

Lembrando que, a função risco é dada pelo quociente entre a função densidade $f(t)$ e a função de sobrevivência $S(t)$, temos para o modelo Weibull que $h(t)$

$$h(t) = \lambda\gamma(\lambda t)^{\gamma-1}.$$

é dada por:

A relação entre os valores de γ e o tempo de sobrevivência pode ser vista na Figura 2.4 que mostra a taxa de falha da distribuição Weibull com $\gamma = 0.5$, $\gamma = 1$, $\gamma = 2$ e $\gamma = 4$. Quando $\gamma = 1$ a taxa de falha permanece constante ao longo de t ; este é o caso da distribuição Exponencial. A taxa de falha cresce quando $\gamma > 1$ e decresce quando $\gamma < 1$ à medida que t aumenta. Então, a distribuição Weibull pode ser usada para modelar a distribuição de sobrevivência de uma população com risco crescente, risco decrescente ou risco constante.

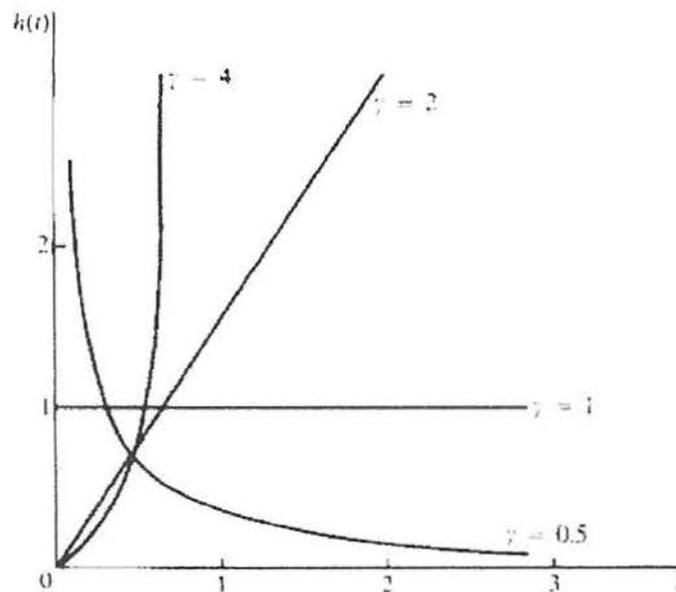


Figura 2.4 - Gráficos da função Risco para valores diversos de γ .

Uma maneira usual de representar a função de sobrevivência é aplicando a função logaritmo. Ou seja,

$$\log S(t) = -(\lambda t)^\gamma.$$

A Figura 2.5 apresenta gráficos do logaritmo da função de sobrevivência de T para alguns valores de γ . Pode-se visualizar que: quando $\gamma = 1$, a função $\log S(t)$ é uma linha reta com inclinação negativa; quando $\gamma < 1$ a função $\log S(t)$ decresce lentamente a partir do valor zero e então se aproxima de um valor constante. Quando $\gamma > 1$ a função $\log S(t)$ decresce de forma mais severa a partir do valor zero à medida que t cresce.

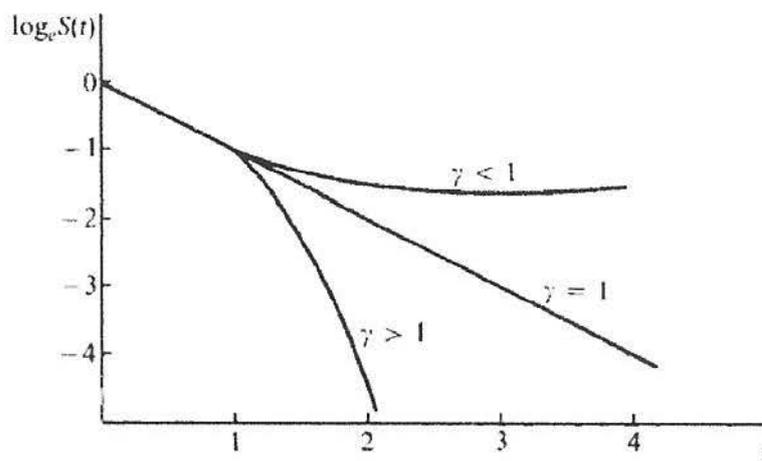


Figura 2.5 - Gráfico do logaritmo da função de sobrevivência de T

3. ABORDAGEM CLÁSSICA PARA MODELOS PARAMÉTRICOS

O que caracteriza um modelo paramétrico é a suposição de uma distribuição de probabilidade para a variável tempo de sobrevivência T . Nas seções 2.4 e 2.5 foram apresentados elementos básicos dos modelos Exponencial e Weibull. Neste capítulo apresentaremos a abordagem clássica utilizando esses modelos para análise de dados de sobrevivência.

Existem diversos métodos de estimação dos parâmetros descritos na literatura, tais como: método dos momentos, método de mínimos quadrados, método de máxima verossimilhança, entre outros. Talvez o mais conhecido deles seja o método de mínimos quadrados que é utilizado no contexto de regressão linear. Entretanto, esse método não é o mais apropriado para estudos de tempo de sobrevivência pois ele é incapaz de incorporar censuras no seu processo de estimação. Para dados censurados, o método de máxima verossimilhança surge como uma opção adequada já que incorpora as censuras e possui propriedades ótimas para grandes amostras.

Na seção 3.1 será definido o método de máxima verossimilhança e o modo como incorpora os dados censurados em seu processo de estimação dos parâmetros. Na seção 3.2 serão apresentados os estimadores de máxima verossimilhança para o parâmetro λ do modelo Exponencial, considerando tanto o caso de não incluir covariáveis no modelo quanto para o caso de incluir covariáveis. Na seção 3.3 serão mostrados os estimadores para os parâmetros γ e λ do modelo Weibull também considerando a inclusão ou não de covariáveis.

3.1. MÉTODO DE MÁXIMA VEROSSIMILHANÇA

O método de máxima verossimilhança trata o problema da estimação da seguinte maneira: baseado nos resultados obtidos na amostra, qual é a distribuição, entre todas definidas pelos possíveis valores dos seus parâmetros, com maior probabilidade de ter gerado tal amostra? Por exemplo, se a distribuição do tempo de sobrevivência é Weibull com parâmetros γ e λ , para cada combinação diferente de γ e λ tem-se diferentes distribuições de Weibull. O estimador de máxima verossimilhança escolhe aquele par $(\gamma; \lambda)$ que melhor explique a amostra observada.

Suponha uma amostra de observações independentes t_1, \dots, t_n de uma certa população de interesse com função distribuição de probabilidade dada por $f(t; \theta)$. Considere, de início, que todas observações da amostra são não-censuradas.

Neste caso, a função de verossimilhança para o parâmetro desta população é dada por:

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

Na expressão (3.1), θ pode estar representando um único parâmetro ou um vetor de parâmetros. A tradução em termos matemáticos para a expressão “a distribuição que melhor explique a amostra observada” é achar o valor de θ que maximize a função $L(\theta)$. Isto é, achar o valor de θ que maximiza a probabilidade da amostra observada ocorrer.

A função de verossimilhança $L(\theta)$ mostra que a contribuição de cada observação não-censurada é sua função densidade de probabilidade. A contribuição de cada observação censurada não deve ser sua função densidade, pois essas observações somente informam que o tempo de sobrevivência é maior que o tempo de censura observado. Portanto, a sua contribuição para a função de verossimilhança é a sua função de sobrevivência $S(t)$.

As observações são divididas em dois grupos, sendo as r primeiras as observações não-censuradas ($1, 2, \dots, r$) e as $n - r$ seguintes são as observações censuradas ($r+1, r+2, \dots, n$). Desta maneira, a função de verossimilhança assume a seguinte forma,

$$L(\theta) = \prod_{i=1}^r f(t_i; \theta) \cdot \prod_{i=r+1}^n S(t_i; \theta) \quad (3.2)$$

A expressão (3.2) é válida para os mecanismos de censura do tipo I e II e sob a suposição de que o mecanismo de censura é não-informativo (não revela informações sobre os parâmetros). Vale também para mecanismos do tipo aleatório.

É sempre mais conveniente, em termos de cálculos, trabalhar com o logaritmo da função de verossimilhança. Então, os estimadores de máxima verossimilhança são os valores de θ que maximizam $L(\theta)$ ou $\log(L(\theta))$, sendo derivado através da solução do sistema de equações abaixo:

$$U(\theta) = \frac{\partial \log L(\theta)}{\partial \theta} \Big|_{\theta=\hat{\theta}} = 0.$$

3.2. MODELO EXPONENCIAL

Seja T o tempo de sobrevivência e suponha que T segue o modelo Exponencial com único parâmetro desconhecido λ . Desta forma, $f(t)$ é dado por (2.12). Precisamos estimar λ para poder obter uma estimativa para o tempo médio de sobrevivência, para o tempo mediano de sobrevivência, enfim, para poder efetuar as análises relevantes para o tempo de sobrevivência. Nessa seção serão apresentados os estimadores de máxima verossimilhança para λ considerando cada um dos modelos: o modelo sem incluir covariáveis e o modelo incluindo covariáveis.

3.2.1. MODELO SEM COVARIÁVEIS

A função de verossimilhança para o parâmetro λ da distribuição Exponencial é dada por:

$$L(\lambda) = \left(\prod_{i=1}^r (1/\lambda) \exp(-t_i/\lambda) \right) \cdot \left(\prod_{i=r+1}^n \exp(-t_i/\lambda) \right)$$

sendo o logaritmo de $L(\lambda)$ e sua derivada em relação à λ dados por:

$$\log(L(\lambda)) = -r \log \lambda - (1/\lambda) \sum_{i=1}^n t_i$$

e

$$\frac{\partial \log L(\lambda)}{\partial \lambda} = -\frac{r}{\lambda} + \left(\frac{1}{\lambda^2} \right) \sum_{i=1}^n t_i$$

Igualando a derivada da expressão acima a zero, tem-se a expressão do estimador de máxima verossimilhança para λ :

$$\hat{\lambda} = \frac{\sum_{i=1}^n t_i}{r} \quad (3.3)$$

Na expressão (3.3), o termo $\sum_{i=1}^n t_i$ é denominado “tempo real sob teste”.

Observe que, se todas as informações são não-censuradas, $\hat{\lambda} = \bar{t}$, onde \bar{t} é a média amostral.

Colosimo (2000) mostra que um intervalo com aproximadamente $(1 - \alpha)\%$ de confiança para λ pode ser estimado através de:

$$\hat{\lambda} \pm Z_{\alpha/2} \sqrt{\frac{\hat{\lambda}^2}{r}} \quad (3.4)$$

onde r é o número de observações não-censuradas.

3.2.2. MODELO COM COVARIÁVEIS

Nesta seção serão apresentadas as idéias para a inclusão de apenas uma covariável no modelo. A inclusão de duas ou mais covariáveis segue analogamente e a restrição a uma covariável foi feita aqui somente para simplificar a notação.

Um modo conveniente e plausível para descrever o tempo de sobrevivência em um modelo de regressão com a inclusão de uma covariável pode ser dado pela equação:

$$T = e^{\beta_0 + \beta_1 x} \varepsilon. \quad (3.5)$$

O tempo de sobrevivência T está sendo expresso em (3.5) como o produto entre a componente sistemática do modelo $e^{\beta_0 + \beta_1 x}$ e a componente de erro ε .

Este modelo pode ser colocado na forma de um modelo de regressão linear se aplicarmos o logaritmo natural de ambos os lados da equação. Dessa maneira, ficamos com,

$$\ln T = \beta_0 + \beta_1 x + \varepsilon^*, \quad (3.6)$$

onde $\varepsilon^* = \ln(\varepsilon)$ e a componente de erro ε^* segue distribuição valor extremo mínimo, denotada por $G(0; \sigma)$.

Nota-se que, conforme esse modelo, o tempo de sobrevivência seguirá uma distribuição Exponencial quando $\sigma = 1$ e a distribuição Weibull quando $\sigma \neq 1$.

Os modelos de tempo de sobrevivência que podem ser linearizados através da aplicação da função logaritmo natural são ditos Modelos de Tempos de Sobrevivência Acelerados. A razão para essa nomenclatura se deve ao fato de que o efeito da covariável será multiplicativo na escala do tempo. Isto é, o efeito das covariáveis pode acelerar ou desacelerar o tempo de sobrevivência T .

Os modelos paramétricos de regressão para tempo de sobrevivência acelerado fornecem resultados e análises de fácil interpretação com a presença de dados censurados para tempo de sobrevivência.

Consideremos o modelo com uma covariável dado em (3.6), onde a distribuição do termo de erro será dada por uma distribuição Log- exponencial, isto é, segue uma distribuição Valor Extremo Mínimo representada por $G(0,1)$. Os detalhes da distribuição Valor Extremo Mínimo estão presentes em Lawless (1982). Neste caso T terá distribuição exponencial com média $E(T) = e^{\beta_0 + \beta_1 x}$. Ou seja,

$$f(t) = \frac{1}{e^{\beta_0 + \beta_1 x}} e^{\left(-\frac{1}{e^{\beta_0 + \beta_1 x}} t\right)} \quad (3.7)$$

Note que, se olharmos a expressão (3.7) como a distribuição exponencial de parâmetro λ definida em (2.12) teremos que $\lambda = 1/e^{\beta_0 + \beta_1 x}$.

Conseqüentemente, a função de sobrevivência para o modelo exponencial com covariáveis será dada por,

$$S(t, x, \beta) = \exp(-t / e^{\beta_0 + \beta_1 x}). \quad (3.8)$$

Para obter a mediana do tempo de sobrevivência, iguala-se o termo à direita de (3.8) a 0,5 e resolve-se a equação resultante, de onde fica:

$$t_{50}(x, \beta) = -e^{\beta_0 + \beta_1 x} \ln(0.5).$$

Quando a covariável em (3.6) é dicotômica, codificada como 0 ou 1, a razão do tempo de sobrevivência mediano do grupo com $x = 1$ pelo tempo de sobrevivência mediano do grupo com $x = 0$, denotada por $TR(x = 1, x = 0)$, é dada por:

$$TR(x = 1; x = 0) = \frac{t_{50}(x = 1, \beta)}{t_{50}(x = 0, \beta)} = \frac{-e^{\beta_0 + \beta_1 x} \cdot \ln(0.5)}{-e^{\beta_0} \cdot \ln(0.5)} = e^{\beta_1}.$$

Por exemplo, se $e^{\beta_1} = 2$, o tempo de sobrevivência mediano no grupo com $x = 1$ é igual a duas vezes o tempo de sobrevivência mediano do grupo com $x = 0$. Do mesmo modo que, se $e^{\beta_1} = 0,5$, então o tempo de sobrevivência mediano no grupo com $x = 1$ é igual a 0,5 vezes o tempo de sobrevivência mediano do grupo com $x = 0$.

No caso geral teremos:

$$TR(x; x^*) = \frac{t_{50}(x, \beta)}{t_{50}(x^*, \beta)} = e^{\beta_1(x-x^*)}. \quad (3.9)$$

Note que, se $x = x^* + 1$ a expressão (3.9) fica igual a e^{β_1} representando a razão do tempo de sobrevivência mediano para o aumento de uma unidade em x .

A quantidade e^{β_1} é comumente referida como o fator de aceleração, muito embora seu efeito possa ou acelerar ou desacelerar o tempo de sobrevivência.

Uma alternativa para apresentar o efeito multiplicativo do fator de aceleração e^{β_1} é via função de sobrevivência.

Segue de (3.8) que a relação entre as funções de sobrevivência de dois grupos é dada por:

$$S(t, x, \beta) = S(te^{-\beta_1}, x = 0, \beta). \quad (3.10)$$

A interpretação do resultado em (3.10) é que o valor da função sobrevivência ao tempo t para o grupo com $x = 1$ pode ser obtido através da estimação da função de sobrevivência para o grupo com $x = 0$ ao tempo $te^{-\beta_1}$. A mudança no sinal do coeficiente é devida ao fato de que os percentis de tempo e as probabilidades de sobrevivência são operações inversas uma da outra.

A função risco para o modelo em (3.8) é,

$$h(t, x, \beta) = e^{-(\beta_0 + \beta_1 x)}$$

e a razão de risco:

$$HR(x, x^*) = e^{-\beta_1(x-x^*)}.$$

Então, podemos ver que o modelo de regressão Exponencial é um exemplo de um modelo de tempo de falha acelerado que tem riscos proporcionais. Cox e Oakes (1984) mostraram que os únicos modelos de tempo de falha acelerado que tinham riscos proporcionais eram os modelos de regressão Exponencial e Weibull.

Se um modelo de regressão Exponencial ajusta-se aos dados, pode-se expressar o efeito de quaisquer covariáveis como uma razão de tempo ou como uma razão de risco. A inclusão de covariáveis categóricas não-binárias é feita através de variáveis dummy, analogamente como o que é feito no modelo de regressão múltipla.

Estimação dos Parâmetros

A estimação dos parâmetros do modelo com covariáveis também é feita utilizando o método de máxima verossimilhança. Os parâmetros do modelo com p covariáveis serão representados por $\beta' = (\beta_0, \beta_1, \dots, \beta_p)$. Assumiremos que as observações são censuradas apenas à direita, mas a análise pode ser estendida para outros tipos de censura.

Suponha que T tenha distribuição exponencial com função densidade de probabilidade dada por (2.12). Considere n observações independentes de T , p covariáveis e um indicador de dado censurado denotado por v_i , $i = 1, 2, 3, \dots, n$.

Então, para cada indivíduo i , tem-se a tripla (t, x_i, v_i) onde $x_i' = (1, x_{i1}, x_{i2}, \dots, x_{ip})$. Desta forma, o logaritmo da função de verossimilhança de β é dado por,

$$l(\beta) = \sum_{i=1}^n v_i z_i - e^{z_i}, \quad (3.11)$$

onde $z_i = y_i - x_i' \beta$ e $y_i = \ln(t_i)$.

As equações de verossimilhança são obtidas pela diferenciação do logaritmo da função de verossimilhança com respeito aos parâmetros

desconhecidos e igualando as expressões a zero. Esse processo fornece a equação para o termo constante,

$$\sum_{i=1}^n (v_i - e^{z_i}) = 0 \quad (3.12)$$

e equações para as covariáveis não-constantes,

$$\sum_{i=1}^n x_{ij} (v_i - e^{z_i}) = 0 \quad (3.13)$$

para $j = 1, 2, \dots, p$. Se denotarmos as soluções numéricas para (3.12) e (3.13) como $\hat{\beta}' = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)$, o modelo baseado nos valores de tempo preditos ou ajustados será computado como $\hat{t}_i = \exp(x_i' \hat{\beta})$ para $i = 1, 2, \dots, n$.

No modelo de regressão linear, a inclusão do termo constante tem como consequência que a média dos valores observados e dos valores ajustados sejam iguais. Entretanto, no modelo de regressão exponencial, esse efeito é no sentido de fazer com que a soma de todas as razões dos valores observados pelos valores ajustados seja igual ao número de observações não-censuradas, isto é:

$$\sum_{i=1}^n v_i = \sum_{i=1}^n \frac{t_i}{\hat{t}_i} = m. \quad (3.14)$$

A expressão (3.14) mostra que o verdadeiro valor de $\hat{\beta}_0$ dependerá em grande parte de m , o número de observações não-censuradas. Assim, os valores ajustados são predições do tempo de sobrevivência de um sujeito com covariável de valor x_i .

Obtêm-se os estimadores da variância e covariância do estimador dos coeficientes β através do método de estimação por máxima verossimilhança tomando-se a segunda derivada parcial da função log-verossimilhança. A forma geral da segunda derivada parcial da expressão (3.11) é dada por,

$$\frac{\partial^2 L(\beta)}{\partial \beta_j \partial \beta_k} = - \sum_{i=1}^n x_{ij} x_{ik} e^{z_i}, \quad j, k = 0, 1, \dots, p. \quad (3.15)$$

Note que não existem soluções analíticas para a derivada em (3.15). Desta forma, a estimação dos coeficientes β é feita através de métodos numéricos.

Os estimadores de variância e covariância são baseados na matriz de informação observada, denotada por $I(\hat{\beta})$ que é uma matriz com elementos dados pelo sinal contrário (negativo) de (3.15), validada para o estimador dos coeficientes. Eles são dados por:

$$\text{Vâr}(\hat{\beta}) = I(\hat{\beta})^{-1}.$$

3.3. MODELO WEIBULL

Seja T o tempo de sobrevivência e suponha que T segue o modelo Weibull com parâmetros desconhecidos γ e λ . Precisamos estimar γ e λ para poder obter a estimativa para o tempo médio ou mediano de sobrevivência, tempo mediano de sobrevivência, enfim, para poder efetuar as análises relevantes para o tempo de sobrevivência. Nessa seção serão apresentados os estimadores de máxima verossimilhança para γ e λ considerando cada um dos modelos: o modelo sem incluir covariáveis e o modelo incluindo covariáveis.

3.3.1. MODELO SEM COVARIÁVEIS

A função de log-verossimilhança para uma amostra de dados de tempos de sobrevivência provenientes de uma distribuição Weibull de parâmetros γ e λ é dada por:

$$\log(L(\lambda, \gamma)) = r \log(\gamma) - \gamma r \log(\lambda) + (\gamma - 1) \sum_{i=1}^r \log(t_i) - \lambda^{-\gamma} \sum_{i=1}^n t_i^{\gamma}. \quad (3.16)$$

De forma alternativa, fazendo $y_i = \log(t_i)$ e utilizando a distribuição do valor extremo, tem-se que,

$$\log(L(\mu, \sigma)) = -r \log(\sigma) + \sum_{i=1}^r \left(\frac{y_i}{\sigma} \right) - \frac{r\mu}{\sigma} - \sum_{i=1}^n \exp \left[\frac{(y_i - \mu)}{\sigma} \right],$$

que é mais simples que a log-verossimilhança obtida em (3.16).

Derivando a log-verossimilhança $\log(L(\mu; \sigma))$ e igualando a zero, obtém-se as equações de verossimilhança,

$$\frac{\partial \log(L(\mu, \sigma))}{\partial \mu} = \frac{1}{\sigma} \left\{ -r + \sum_{i=1}^n \exp\left[\frac{(y_i - \mu)}{\sigma}\right] \right\} = 0 \quad (3.17)$$

e

$$\frac{\partial \log(L(\mu, \sigma))}{\partial \sigma} = \frac{1}{\sigma^2} \left\{ -r\sigma - \sum_{i=1}^r y_i + r\mu + \sum_{i=1}^n \exp\left[\frac{(y_i - \mu)}{\sigma}\right] (y_i - \mu) \right\} = 0. \quad (3.18)$$

Os estimadores de máxima verossimilhança são os valores de μ e σ que satisfazem às equações (3.17) e (3.18). A solução desse sistema de equações para um conjunto de dados particular deve ser obtida, de forma aproximada, via métodos numéricos. Um método muito utilizado para esse fim é o método de Newton-Raphson que não será apresentado nesta monografia. Maiores detalhes podem ser encontrados em Colosimo (2000).

3.3.2. MODELO COM COVARIÁVEIS

A forma mais comum de um modelo de regressão para o tempo de sobrevivência acelerado foi mostrada em (3.6). Além de haver mais um parâmetro desconhecido a ser estimado, a principal diferença entre o modelo de regressão Weibull e o modelo de regressão Exponencial está no parâmetro σ da distribuição da componente de erro do modelo em (3.6) que agora pode ser diferente do valor 1. Neste caso, T segue distribuição Weibull com função densidade de probabilidade dada por (2.12). A inclusão de mais um parâmetro no modelo nos leva a uma função risco um tanto mais complicada e um modelo de regressão mais complexo. Esta seção irá considerar a inclusão de uma covariável no modelo assim como foi apresentado no modelo Exponencial.

A função risco para o modelo de regressão Weibull foi vista no capítulo anterior e é dada por:

$$h(t) = \lambda^\gamma (\lambda t)^{\gamma-1} = \lambda^\gamma \gamma t^{\gamma-1}.$$

Por conveniência trocaremos na expressão anterior λ por γ e vice-versa. Desta forma,

$$h(t) = \gamma^\lambda \lambda t^{\lambda-1}.$$

A inclusão de covariáveis será feita através do parâmetro γ , ou seja, $\gamma = \frac{1}{e^{\beta_0 + \beta_1 x}}$. Então:

$$h(t, x, \beta, \lambda) = \frac{\lambda t^{\lambda-1}}{(e^{\beta_0 + \beta_1 x})^\lambda}, \quad (3.19)$$

onde, por conveniência, utilizou-se $\lambda = 1/\sigma$. Essa função risco pode ser expressa novamente em termos de riscos proporcionais ou na forma de tempo de sobrevivência acelerado. Ou seja,

$$\begin{aligned} h(t, x, \beta, \lambda) &= \lambda t^{\lambda-1} e^{-\lambda(\beta_0 + \beta_1 x)} \\ &= \lambda t^{\lambda-1} e^{-\lambda\beta_0} e^{-\lambda\beta_1 x} \\ &= \lambda \gamma t^{\lambda-1} e^{-\lambda\beta_1 x} \\ &= h_0(t) e^{\theta_1 x}, \end{aligned} \quad (3.20)$$

onde $\gamma = \exp(-\beta_0 / \sigma) = \exp(\theta_0)$, $\theta_1 = -\beta_1 / \sigma$. $h_0(t)$ é a função risco de base.

A expressão (3.20) leva à interpretação da razão de risco através do parâmetro θ_1 .

Outra forma de apresentar a função risco é dada re-arranjando (3.19) como segue:

$$\begin{aligned} h(t, x, \beta, \lambda) &= \lambda t^{\lambda-1} e^{-\lambda(\beta_0 + \beta_1 x)} \\ &= \lambda \gamma (te^{-\beta_1 x})^{\lambda-1} e^{-\beta_1 x}. \end{aligned} \quad (3.21)$$

A função risco dada pela fórmula (3.21) é muito utilizada nos aplicativos, por isso a ilustração dela aqui.

A função de sobrevivência correspondente à forma da função risco para o tempo de sobrevivência acelerado em (3.21) é dada por:

$$S(t, x, \beta, \lambda) = \exp\left\{-t^\lambda \exp\left[(-1/\sigma)(\beta_0 + \beta_1 x)\right]\right\}. \quad (3.22)$$

Podemos obter a equação para o tempo de sobrevivência mediano igualando a expressão (3.22) a 0,5, que assim fica:

$$t_{50}(x, \beta, \sigma) = [-\ln(0,5)]^\sigma e^{\beta_0 + \beta_1 x}.$$

Por exemplo, se a covariável for dicotômica, codificada em 0 ou 1, a razão dos tempos de sobrevivência medianos é:

$$TR(x=1; x=0) = \frac{t_{50}(x=1, \beta, \sigma)}{t_{50}(x=0, \beta, \sigma)} = \frac{[-\ln(0,5)]^\sigma e^{\beta_0 + \beta_1 x}}{[-\ln(0,5)]^\sigma e^{\beta_0}} = e^{\beta_1}.$$

Estimação dos Parâmetros

A estimação dos parâmetros do modelo com covariáveis também é feita utilizando o método de máxima verossimilhança. Os parâmetros do modelo com p covariáveis serão representados por $\beta' = (\beta_0, \beta_1, \dots, \beta_p)$. Assumiremos que as observações são censuradas apenas à direita, mas a análise pode ser estendida para outros tipos de censura.

Suponha que T tenha distribuição weibull denotada por $(\alpha; \gamma)$. Utilizando a parametrização $\lambda = \log(\gamma)$ e $\alpha = 1/\sigma$, a função densidade de probabilidade de T será dada por:

$$f(t) = \frac{1}{\sigma} t^{\frac{1}{\sigma}-1} \exp(\beta_0 + \beta_1 x - \exp(\beta_0 + \beta_1 x)t^{\frac{1}{\sigma}}).$$

Considere n observações independentes de T , p covariáveis e um indicador de dado censurado denotado por v_i , $i = 1, 2, 3, \dots, n$.

Então, para cada indivíduo i tem-se a tripla (t, x_i, v_i) onde $x_i' = (1, x_{i1}, x_{i2}, \dots, x_{ip})$. Desta forma, o logaritmo da função de verossimilhança de β é dado por,

$$l(\beta, \sigma) = \sum_{i=1}^n v_i (-\ln(\sigma) + z_i) - e^{z_i}. \quad (3.23)$$

Os valores para a equação para o j -ésimo coeficiente de regressão é obtida tomando a derivada de (3.23) com respeito a β_j e igualando a zero. Assim, fica:

$$\frac{\partial l(\beta, \sigma)}{\partial \beta_j} = \sum_{i=1}^n \frac{-x_{ij}}{\sigma} (v_i - e^{z_i}) = 0, \quad j = 0, 1, 2, \dots, p.$$

O valor da equação para o parâmetro de forma σ é dado por:

$$\frac{\partial l(\beta, \sigma)}{\partial \sigma} = \frac{-m}{\sigma} + \sum_{i=1}^n \frac{-z_i}{\sigma} (v_i - e^{z_i}) = 0, \quad j = 0, 1, 2, \dots, p.$$

Note que, assim como no modelo Exponencial, a estimação de β e σ é feita através de métodos numéricos.

Ao solucionar as equações acima, a matriz de informação pode ser expressa como:

$$I(\hat{\beta}; \hat{\sigma}) = \frac{1}{\hat{\sigma}^2} \begin{bmatrix} X' \hat{V} X & X' \hat{V} \hat{z} \\ \hat{z}' \hat{V} X & \hat{z}' \hat{V} \hat{z} + m \end{bmatrix},$$

onde X é uma matriz $n \times (p+1)$ contendo os valores das covariáveis, $\hat{V} = \text{diag}(e^{\hat{z}_i})$ uma matriz $n \times n$ diagonal e $\hat{z}' = (\hat{z}_1, \hat{z}_2, \dots, \hat{z}_n)$ com z_i definido como sendo:

$$z_i = \frac{y_i - x_i' \beta}{\sigma}.$$

O estimador da matriz de covariância dos estimadores dos parâmetros é dado por:

$$\text{Var}(\hat{\beta}; \hat{\sigma}) = [I(\hat{\beta}; \hat{\sigma})]^{-1}.$$

4. ABORDAGEM BAYESIANA PARA MODELOS PARAMÉTRICOS

Dentro do contexto Bayesiano, a meta principal de qualquer análise é a busca pela distribuição a posteriori para os parâmetros de interesse. Em relação aos modelos Exponencial e Weibull, apenas o Exponencial sem covariáveis é que apresenta forma fechada para a posteriori de interesse. Para as análises do modelo Exponencial com covariáveis e o modelo Weibull são necessários métodos de simulação. Um método de simulação estocástica que resolve este e muitos outros problemas de inferência Bayesiana é o *Gibbs Sampling*. O Gibbs Sampling é um tipo de simulação Monte Carlo usando Cadeias de Markov (MCMC).

Os modelos paramétricos possuem importante papel na Análise de Sobrevivência Bayesiana, já que diversas análises desse tipo, na prática, são realizadas utilizando modelos paramétricos. Essa modelagem oferece menores dificuldades e técnicas de análise menos complexas.

Este capítulo inicia com algumas definições básicas de Inferência Bayesiana na seção 4.1. Nas seções 4.2 e 4.3, serão apresentados os modelos paramétricos exponencial e Weibull utilizados na análise Bayesiana para dados de sobrevivência univariados e censurados à direita, descrevendo as características de cada um tanto para modelagem sem covariável quanto para a modelagem considerando a presença da covariável.

4.1. DEFINIÇÕES BÁSICAS DE INFERÊNCIA BAYESIANA

A Inferência Estatística Bayesiana é o processo de diminuição de incerteza que se baseia em dados estatísticos e evidências pessoais.

Na Inferência Estatística Bayesiana Paramétrica, a forma da distribuição de probabilidade de um vetor de variáveis aleatórias \mathbf{X} é suposta conhecida e dependente de um parâmetro θ (θ pode ser um vetor de parâmetros mas, por simplicidade de exposição, usaremos θ como escalar. A generalização para θ vetor é direta. Os parâmetros θ dessa distribuição são desconhecidos e por isso são considerados como variáveis aleatórias.

Por serem considerados variáveis aleatórias, toda a informação que se tenha sobre θ antes de observar os dados pode ser incluída na análise através da distribuição a priori para θ . As inferências a respeito dos parâmetros θ são baseadas na distribuição a posteriori de θ . A idéia central da Inferência Bayesiana é, após observar os dados, combiná-los com a informação a priori de θ , derivando assim a distribuição a posteriori para θ , notada por $\pi(\theta | \mathbf{x})$.

Além disso, existe a distribuição preditiva a posteriori para X^* , onde X^* representa uma observação futura da variável X no experimento em questão a qual estamos querendo prever. Isto é, em Inferência Estatística Bayesiana, as previsões para valores futuros de X podem ser feitas diretamente através da distribuição preditiva a posteriori para X^* . Denota-se a distribuição preditiva a posteriori por $\pi(x^* | \mathbf{x})$.

4.1.1. DISTRIBUIÇÃO A POSTERIORI

A distribuição a posteriori combina o conhecimento que se tenha a priori de θ com a informação sobre o parâmetro θ contida nos dados observados (x), com o objetivo de gerar um conhecimento final a respeito do desconhecido parâmetro θ .

Através da posteriori, podemos obter todas as decisões e inferências a respeito do desconhecido parâmetro θ .

A distribuição a posteriori de θ é condicionada no valor amostral observado x e é denotada por $\pi(\theta|x)$. A notação indica que a posteriori é uma distribuição condicionada à observação amostral x . Denota-se que θ e X têm distribuição densidade conjunta dada por:

$$f(\theta, x) = \pi(\theta)f(x|\theta), \quad (4.1)$$

onde $\pi(\theta)$ é a distribuição a priori para θ e $f(x|\theta)$ é a distribuição ou função densidade de probabilidade de X .

A densidade marginal de X , também chamada de preditiva a priori de X , será dada por,

$$m(x) = \int_{\theta} f(x|\theta)dF(\theta) = \begin{cases} \int_{\theta} f(x|\theta)\pi(\theta)d\theta, & \text{se } \theta \text{ é contínuo} \\ \sum_{\theta} f(x|\theta)\pi(\theta), & \text{se } \theta \text{ é discreto} \end{cases}$$

A distribuição a posteriori de θ é encontrada através da aplicação da regra de Bayes. Ou seja,

$$\pi(\theta/x) = \frac{f(\theta, x)}{m(x)} = \frac{\pi(\theta)f(x|\theta)}{m(x)}. \quad (4.2)$$

Depois da amostra observada, $f(x|\theta)$ é a verossimilhança $L(\theta|x)$ e $m(x)$ será uma constante que não depende de θ . Sendo assim, uma forma equivalente para a expressão (4.2) é obtida omitindo-se o fator $m(x)$, ficando,

$$\pi(\theta|x) \propto \pi(\theta)L(\theta|x) \quad . \quad (4.3)$$

A expressão (4.3) mostra o "jargão" da Inferência Estatística Bayesiana que diz: "A Posteriori é proporcional ao produto da priori pela verossimilhança".

Essas expressões acima descrevem a idéia fundamental da Inferência Bayesiana Paramétrica: o primeiro passo para qualquer aplicação da técnica é assumir um modelo probabilístico $f(\theta ; x)$ envolvendo θ e x . Depois, determina-se uma priori para θ e, através de métodos analíticos ou computacionais apropriados, deriva-se a distribuição a posteriori $\pi(\theta|x)$.

Note que a distribuição a priori de θ reflete o conhecimento sobre θ antes da observação da amostra (antes do experimento), já a distribuição a posteriori de θ reflete o conhecimento atualizado sobre θ após a observação da amostra x .

4.1.2. DISTRIBUIÇÃO A PRIORI

É uma distribuição de probabilidade que expressa o conhecimento sobre o parâmetro desconhecido θ que se tenha antes da observação da amostra. Este conhecimento é baseado, muitas vezes, no entendimento do pesquisador, experiência, estudo anteriores, etc. Como θ é visto como uma variável aleatória, o valor de θ é pensado como uma realização da distribuição a priori, assim como um valor de x é pensado como sendo uma realização da distribuição $f(x)$.

A distribuição a priori é a maneira de incorporar toda a informação e conhecimento que se tenha sobre θ que não seja proveniente dos dados amostrais, já que um pesquisador possui, por menor que seja, algum conhecimento sobre θ .

A partir do conhecimento que se tenha sobre θ , pode-se definir uma família paramétrica de densidades. Neste caso, a distribuição a priori é representada por uma forma funcional, cujos parâmetros devem ser especificados de acordo com

esse conhecimento. Os parâmetros indexadores da família de distribuição a priori são chamados de *hiperparâmetros* para diferenciar do parâmetro de interesse θ .

Para a derivação da distribuição a posteriori, precisamos da especificação completa da distribuição a priori para θ . Ou seja, é preciso especificar uma forma para essa distribuição e valores para os seus parâmetros. Esta especificação deve espelhar a informação a priori que se tenha.

A seguir serão apresentadas algumas das diferentes formas de especificação da distribuição a priori para θ .

4.1.3. PRIORI CONJUGADA

O uso de priors conjugadas, em geral, facilita as análises pois envolverá apenas uma mudança nos hiperparâmetros. A idéia é que quando as distribuições a priori e a posteriori pertencem a uma mesma classe de distribuições dizemos que a priori é conjugada ao modelo. Por exemplo, considere o modelo Exponencial de parâmetro λ . A priori conjugada para λ é dada por uma distribuição Gama de parâmetros α e β .

4.1.4. PRIORI NÃO-INFORMATIVA

Existem situações nas quais o conhecimento a respeito de θ não existe ou é algo de pouca convicção. Apesar da inexistência desse conhecimento, precisamos da distribuição a priori para θ para poder derivar a distribuição a posteriori de θ .

Precisamos então determinar uma priori que represente a falta de conhecimento sobre θ recebendo o nome de Priori Não-Informativa.

Ao utilizar uma Priori Não-Informativa para θ , a distribuição a posteriori de θ será um retrato dos dados observados.

O trabalho original de Bayes (1763) considerava o uso de Priori Uniforme, isto é, atribui igual probabilidade para todo $\theta \in \Theta$ (espaço paramétrico). Ou seja, é o conhecido postulado de Bayes que diz: "Se nada sabemos sobre θ devemos assumir que a priori de θ é uma distribuição de probabilidade constante". Uma priori não-informativa muito utilizada é $\pi(\theta) \propto c$, onde c é uma constante qualquer.

Uma outra forma muito utilizada para definir uma Priori Não-Informativa para θ é a Priori de Jeffreys, definida por,

$$\pi(\theta) \propto [I(\theta)]^{1/2},$$

onde $I(\theta)$ é a Informação de Fisher para θ .

A Priori de Jeffreys possui a propriedade de invariância em relação à parametrização. Ou seja, para qualquer parametrização utilizada para θ , a priori será a mesma, a menos de algum parâmetro de escala.

Prioris conjugadas também podem ser prioris não-informativas dependendo do valor especificado para os seus parâmetros. Por exemplo, uma priori dada por uma Gama (0.0001; 0.0001). Essa priori, devido aos valores de seus parâmetros, será não-informativa.

4.1.5. PRIORI IMPRÓPRIA

São distribuições que não integram um no espaço paramétrico. Um exemplo de priori imprópria é a priori não-informativa $\pi(\theta) \propto c$. Em geral, prioris impróprias só são utilizadas quando não se tem nenhuma informação sobre o parâmetro. O uso de prioris impróprias deve ser cuidadoso pois nem sempre uma priori imprópria conduz a uma posteriori própria (ou seja, que integre um no seu espaço paramétrico).

4.1.6. PREDITIVA A POSTERIORI

A realização de inferências a respeito de uma quantidade observável mas ainda desconhecida, freqüentemente chamada de inferência preditiva, segue uma lógica similar a da distribuição a posteriori para o parâmetro de interesse. Antes dos dados amostrais serem considerados, a distribuição dessa quantidade desconhecida porém observável x , é dada por,

$$m(x) = \int f(x, \theta) d\theta = \int f(x / \theta) \cdot \pi(\theta) \cdot d\theta \quad (4.4)$$

A expressão acima é geralmente chamada de distribuição marginal de x , mas uma nomenclatura mais informativa é a de distribuição preditiva a priori: priori porque não é condicional a uma observação prévia, e preditiva porque é a distribuição para uma quantidade observável. Após os dados terem sido observados, podemos prever um valor observável desconhecido x^* do mesmo processo.

Por exemplo, seja $\mathbf{x} = (x_1, x_2, \dots, x_n)$ um vetor de observações de pesos de um objeto que foi observado em n tempos distintos em uma escala de medida e seja $\theta = (\mu; \sigma^2)$ o vetor de parâmetros, onde μ é o valor verdadeiro, porém desconhecido, para o peso médio do objeto e σ^2 a medida de variância da escala, e ainda, x^* uma futura observação de peso do objeto.

A distribuição de x^* é chamada distribuição preditiva a posteriori: posteriori porque é condicional aos dados observados \mathbf{x} e preditiva por tratar-se de uma previsão para um valor futuro, x^* . Assim, a distribuição preditiva a posteriori para x^* será dada por:

$$\begin{aligned} \pi(x^* | \mathbf{x}) &= \int \pi(x^*, \theta | \mathbf{x}) d\theta \\ &= \int \pi(x^* | \theta, \mathbf{x}) \pi(\theta | \mathbf{x}) d\theta \\ &= \int \pi(x^* | \theta) \pi(\theta | \mathbf{x}) d\theta, \end{aligned} \tag{4.5}$$

onde $\pi(\theta | \mathbf{x})$ é a posteriori para θ .

A última igualdade na expressão acima se deve ao fato de x^* ser independente de X_1, \dots, X_n .

4.1.7. RESULTADOS A POSTERIORI

O principal objetivo da inferência Bayesiana é obter a distribuição a posteriori, pois é através dela que se pode obter estimativas para os parâmetros de interesse. A distribuição a posteriori apresenta o atual conhecimento sobre o parâmetro θ expresso em probabilidade. Porém podemos calcular estatísticas descritivas da distribuição a posteriori, tais como a média, a moda, a mediana, etc, e utilizá-las como estimativas pontuais para θ .

Para podermos comparar os resultados calculados pela abordagem Bayesiana com os obtidos com a abordagem clássica é mais adequado se utilizar a moda da posteriori já que a mesma baseia-se na idéia de valor mais freqüente, mais provável, idéia similar a dos estimadores de máxima verossimilhança. Quando a distribuição a posteriori possui forma simétrica ou aproximadamente simétrica sabe-se que a média, a moda e a mediana são iguais ou aproximadamente iguais. Assim, não há problemas em se utilizar a média a posteriori para comparar os resultados clássicos e Bayesianos.

Sob a abordagem Bayesiana também é possível obter estimativas por intervalo para o parâmetro θ . Na Inferência Bayesiana calculam-se Intervalos de Credibilidade (também chamados de Intervalos de Probabilidade a Posteriori) para o parâmetro θ . Na inferência clássica calculam-se Intervalos de Confiança para o parâmetro θ . A construção desses intervalos é similar nas duas abordagens, porém a sua interpretação é diferente. A diferença nas interpretações deve-se ao fato de que, na inferência Bayesiana o parâmetro θ é considerado aleatório, e na clássica ele é considerado fixo. Desta forma, o intervalo de confiança clássico é um intervalo que, antes da amostra ser observada, tem uma probabilidade $(1 - \alpha)$ de incluir o parâmetro desconhecido. O intervalo de credibilidade Bayesiano é aquele que, após a observação da amostra, o parâmetro θ possui probabilidade $(1-\alpha)$ tirar os espaços para $1-\alpha$ ficar na mesma linha de pertencer a ele.

Podemos distinguir dois tipos de intervalos de credibilidade Bayesianos: o Intervalo de Credibilidade Central e o HDR (*High Density Region*).

Seja $\pi(\theta | x)$ a posteriori de θ . O intervalo $(c_1; c_2)$, onde c_1 e c_2 são tais que:

$$\int_{-\infty}^{c_1} \pi(\theta/x) d\theta = \int_{c_2}^{+\infty} \pi(\theta/x) d\theta = \alpha/2,$$

é chamado de Intervalo de Credibilidade Central para θ com probabilidade a posteriori igual a $(1 - \alpha)$.

O Intervalo de Credibilidade HDR (*High Density Region*) para θ com probabilidade $(1 - \alpha)$ é definido como sendo a região da posteriori que:

- i) A densidade da posteriori em qualquer ponto pertencente a essa região é igual ou maior a densidade da posteriori em qualquer ponto fora da região.
- ii) A amplitude da região é a menor possível.
- iii) A probabilidade da posteriori nessa região é igual a $(1 - \alpha)$.

Observe que, se a distribuição a posteriori é uma distribuição unimodal e simétrica, então o Intervalo de Credibilidade Central é também o Intervalo de Credibilidade HDR.

O estimador Bayesiano por intervalo mais intuitivo é o Intervalo de Credibilidade HDR, pois é a região que contém os valores mais prováveis para o parâmetro desconhecido. Qualquer que seja a credibilidade $(1 - \alpha)$, todos os pontos que pertencem à região de credibilidade têm maior densidade de probabilidade que qualquer outro ponto que não pertença a ela. O problema do HDR é que, dependendo da posteriori, sua determinação pode não ser possível. Este é o caso quando a posteriori não é derivável analiticamente sendo utilizado métodos MCMC para simular dela. Nestes casos, utilizam-se os intervalos de credibilidade central.

4.1.8. MÉTODO MCMC

Um dos grandes empecilhos da inferência Bayesiana não ter se desenvolvido no passado é o fato de que a posteriori para os parâmetros desconhecidos de muitos modelos não apresenta forma fechada. Sem conhecer ou poder amostrar da posteriori de interesse não é possível fazer análise Bayesiana. Uma maneira de superar este problema é via utilização dos métodos MCMC.

Essa metodologia é, essencialmente, integração via método Monte Carlo utilizando cadeias de Markov. A integração Monte Carlo extrai amostras de uma distribuição necessária e usa as médias amostrais como aproximações da esperança desta distribuição. O MCMC extrai dessas amostras para construir uma

cadeia de Markov ao longo do tempo, cuja distribuição estacionária será a distribuição a posteriori de interesse.

Um dos métodos MCMC é o Gibbs Sampling. Ele é um caso particular do algoritmo de Metropolis-Hasting (1970). Ele é adequado quando o número de parâmetros desconhecidos envolvidos é maior do que 1. O Gibbs Sampling se caracteriza pelo fato de utilizar as distribuições condicionais completas como distribuições propostas. Digamos que queremos simular uma amostra da posteriori $\pi(\theta | x)$ onde $\theta = (\theta_1, \dots, \theta_d)$. O algoritmo *Gibbs Sampling* é descrito da seguinte maneira:

1. Inicialize o contador de iteração da cadeia em $j=1$ e fixe o conjunto de valores iniciais $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_d^{(0)})'$.
2. Obtenha o próximo valor $\theta^{(j)} = (\theta_1^{(j)}, \dots, \theta_d^{(j)})'$ de $\theta^{(j-1)}$ através da sucessiva geração de valores:

$$\begin{aligned} \theta_1^{(j)} &\sim \pi(\theta_1 | \theta_2^{(j-1)}, \dots, \theta_d^{(j-1)}, \text{dados}) \\ \theta_2^{(j)} &\sim \pi(\theta_2 | \theta_1^{(j)}, \theta_3^{(j-1)}, \dots, \theta_d^{(j-1)}, \text{dados}) \\ &\vdots \\ \theta_d^{(j)} &\sim \pi(\theta_d | \theta_1^{(j)}, \dots, \theta_{d-1}^{(j)}, \text{dados}), \end{aligned}$$

onde, por exemplo $\pi(\theta_1 | \theta_2^{(j-1)}, \dots, \theta_d^{(j-1)})$ é a distribuição condicional completa de θ_1 , que nada mais é do que a distribuição a posteriori para θ_1 condicionada em $(\theta_2, \dots, \theta_d, \text{dados})$.

3. Mude o contador de j para $j+1$ e retorne ao passo 2 até que a convergência seja alcançada. Quando a convergência for alcançada, o valor resultante $\theta^{(j)}$ será uma extração da posteriori de interesse.

Desta forma, para obter uma amostra de tamanho n da posteriori $\pi(\theta_1, \theta_2, \dots, \theta_d | \text{dados})$ basta aplicar o algoritmo de Gibbs Sampling, descartar m primeiros

valores simulados (período antes da convergência chamado de burn-in) e tomar os n valores seguintes. Mais detalhes sobre MCMC veja, por exemplo, Gilks *et al* (1997).

4.2. MODELO EXPONENCIAL

Suponha que os tempos de sobrevivência $\mathbf{t}' = (t_1, t_2, \dots, t_n)$ são independentes e identicamente distribuídos (i.i.d.) e cada um possui distribuição exponencial com parâmetro λ , denotada por $\mathcal{E}(\lambda)$. Seja o indicador de censura denotado por $\mathbf{v} = (v_1, v_2, \dots, v_n)'$, onde $v_i = 0$ se t_i é censurado à direita e $v_i = 1$ se t_i é um tempo de sobrevivência. Neste caso, a função densidade de probabilidade do tempo de sobrevivência T é dada por,

$$f(t_i | \lambda) = \lambda \exp(-\lambda t_i),$$

a função de sobrevivência no modelo Exponencial é dada por,

$$S(t_i | \lambda) = \exp(-\lambda t_i).$$

A função de verossimilhança de λ pode ser escrita como:

$$\begin{aligned} L(\lambda / t, \mathbf{v}) &= \prod_{i=1}^n f(t_i / \lambda)^{v_i} S(t_i / \lambda)^{(1-v_i)} \\ &= \lambda^d \exp(-\lambda \sum_{i=1}^n t_i) \end{aligned}$$

onde $d = \sum_{i=1}^n v_i$.

4.2.1. MODELO EXPONENCIAL SEM COVARIÁVEIS

No modelo exponencial sem a inclusão de covariáveis, o único parâmetro desconhecido é λ .

A priori conjugada para λ para esse modelo é dada por uma distribuição Gama ($\alpha_0; \lambda_0$) com função densidade de probabilidade dada por:

$$\pi(\lambda / \alpha_0; \lambda_0) = \frac{\lambda_0^{\alpha_0}}{\Gamma(\alpha_0)} \lambda^{\alpha_0-1} \exp(-\lambda_0 \lambda), \quad \text{para } \lambda > 0.$$

Então, tomando-se uma distribuição Gama ($\alpha_0 ; \lambda_0$) como a priori para λ , a distribuição a posteriori de λ será dada por:

$$\begin{aligned}\pi(\lambda/t, \nu) &= \frac{L(\lambda/t, \nu)\pi(\lambda/\alpha_0; \lambda_0)}{\int_0^\infty L(\lambda/t, \nu)\pi(\lambda/\alpha_0; \lambda_0)d\lambda} \\ &= \frac{(\lambda_0 + \sum_{i=1}^n t_i)^{\alpha_0+d}}{\Gamma(\alpha_0+d)} \lambda^{\alpha_0+d-1} \exp\left\{-\lambda(\lambda_0 + \sum_{i=1}^n t_i)\right\} \text{ para } \lambda > 0.\end{aligned}$$

Ou seja, a posteriori de λ é uma distribuição Gama ($\alpha_0 + d ; \lambda_0 + \sum_{i=1}^n t_i$).

Assim, a média a posteriori de λ e a variância a posteriori de λ serão dadas, respectivamente, por:

$$E(\lambda/t, \nu) = \frac{\alpha_0 + d}{\lambda_0 + \sum_{i=1}^n t_i} \quad e \quad Var(\lambda/t, \nu) = \frac{\alpha_0 + d}{(\lambda_0 + \sum_{i=1}^n t_i)^2}.$$

A distribuição preditiva a posteriori de um futuro tempo T_f será dada por:

$$\begin{aligned}\pi(t_f/t, \nu) &= \int_0^\infty f(t_f/\lambda)\pi(\lambda/t, \nu)d\lambda \\ &= \int_0^\infty \lambda \exp(-\lambda t_f) \frac{(\lambda_0 + \sum_{i=1}^n t_i)^{\alpha_0+d}}{\Gamma(\alpha_0+d)} \lambda^{\alpha_0+d-1} \exp\left\{-\lambda(\lambda_0 + \sum_{i=1}^n t_i)\right\} d\lambda \\ &= \begin{cases} \frac{(\alpha_0+d)(\lambda_0 + \sum_{i=1}^n t_i)^{\alpha_0+d}}{(\lambda_0 + \sum_{i=1}^n t_i + t_f)^{\alpha_0+d+1}}, & \text{se } t_f > 0 \\ 0, & \text{c.c.} \end{cases} \quad (4.6)\end{aligned}$$

Note que a expressão (4.6) é conhecida como distribuição Inversa Beta.

4.2.2. MODELO EXPONENCIAL COM COVARIÁVEIS

Seja $x_i' = (1, x_{i1}, x_{i2}, \dots, x_{ip})$ o vetor de p -covariáveis para o indivíduo i . A inclusão de x_i no modelo será feita utilizando um modelo de regressão para λ . Usaremos que $\lambda_i = \exp(x_i' \beta) = \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip})$. O parâmetro desconhecido é o vetor β .

Nesse caso, a função de verossimilhança para o modelo com a inclusão de covariáveis através de λ (verossimilhança de β) fica:

$$\begin{aligned} L(\beta/t, \nu) &= \prod_{i=1}^n f(t_i/\lambda)^{\nu_i} S(t_i/\lambda)^{(1-\nu_i)} \\ &= \prod_{i=1}^n [\exp(x_i' \beta) \exp(-t_i \exp(x_i' \beta))]^{\nu_i} [\exp(-t_i \exp(x_i' \beta))]^{1-\nu_i} \\ &= \exp\left\{ \sum_{i=1}^n \nu_i x_i' \beta \right\} \exp\left\{ - \sum_{i=1}^n t_i \exp(x_i' \beta) \right\}. \end{aligned}$$

É usual especificar uma priori conjugada para o parâmetro de interesse. Porém, neste caso não existe priori conjugada para β e uma alternativa razoável é especificar uma priori Normal p -dimensional para β , denotada por $N_p(\mu_0; \Sigma_0)$, onde μ_0 é a média a priori para β e Σ_0 a matriz de covariância a priori para β . Apesar da distribuição Normal não ser uma priori conjugada para este modelo, ela facilita as contas para a posteriori e é bastante flexível para representar a informação a priori podendo até se tornar não-informativa (especificando uma variância grande). Sendo assim, a distribuição a posteriori de β será dada por,

$$\pi(\beta/t, \nu) = \frac{L(\beta/t, \nu) \pi(\beta/\mu_0; \Sigma_0)}{\int \dots \int L(\beta/t, \nu) \pi(\beta/\mu_0; \Sigma_0) d\beta_0 d\beta_1 \dots d\beta_p} \quad (4.7)$$

$$\pi(\beta/t, \nu) \propto L(\beta/t, \nu) \pi(\beta/\mu_0; \Sigma_0). \quad (4.8)$$

onde $\pi(\beta/\mu_0; \Sigma_0)$ é uma função densidade Normal Multivariada com média μ_0 e matriz de covariância Σ_0 .

A integral do denominador na expressão (4.7) não apresenta uma solução analítica e o produto que aparece na expressão (4.8) não revela nenhum núcleo

de uma distribuição que seja conhecida. Com isso, verifica-se que a distribuição a posteriori para β não possui uma forma fechada. Como alternativa, os métodos de simulação MCMC podem ser utilizados para amostrar da distribuição a posteriori para β .

Para realizar o Gibbs Sampling para o modelo anterior, onde a posteriori para β não possui forma fechada, precisamos conhecer as distribuições condicionais completas para cada β_j , componente do vetor β .

Seja β_j o j - ésimos componente de β e seja $\beta^{(-j)}$ o vetor sem este componente. Então, a j - ésima distribuição condicional completa pode ser escrita como,

$$\pi(\beta_j / y, n, \beta^{(-j)}) \propto L(\beta_j, \beta^{(-j)} / y, n) \pi(\beta_j, \beta^{(-j)} / \mu_0; \Sigma_0) \quad (4.9)$$

para $j = 0, 1, 2, \dots, p$. Note que cada uma das $p+1$ condicionais completas em (4.9) são proporcionais a função de verossimilhança multiplicada pela priori, onde em cada caso, a variável aleatória é β_j e $\beta^{(-j)}$ é mantido fixo.

É possível verificar que todas as $p+1$ condicionais completas deste modelo não apresentam forma fechada. Porém, isto não implica que não seja possível amostrar delas. Para aplicar o algoritmo de Gibbs Sampling não precisamos, necessariamente, conhecer as distribuições condicionais completas, mas sim saber como amostrar delas. Gilks e Wild (1992) descrevem o método denominado Adaptive Rejection que é um algoritmo de simulação usando rejeição. Para poder utilizar esse algoritmo, precisamos que cada uma das distribuições condicionais completas a serem amostradas seja uma função log-côncava. Chen *et al* (2001) mostram que todas as condicionais completas do modelo Exponencial com covariáveis são log-côncavas.

Um resultado a posteriori de interesse freqüente é a distribuição a posteriori da função de sobrevivência em um particular valor t . Dessa forma ao gerar as amostras através da simulação da distribuição a posteriori de β , as estimativas dos parâmetros serão calculadas utilizando os momentos desta amostra. Assim,

será possível gerar os resultados a posteriori para $S(t)$, tempo médio ou mediano de sobrevivência, razão de sobrevivência, etc.

4.3. MODELO WEIBULL

Suponha que os tempos de sobrevivência $\mathbf{t}' = (t_1, t_2, \dots, t_n)$ são independentes e identicamente distribuídos (i.i.d.) e cada um possui distribuição Weibull de parâmetros α e γ , denotada por $W(\alpha; \gamma)$. Seja o indicador de censura denotado por $\mathbf{v} = (v_1, v_2, \dots, v_n)'$, onde $v_i = 0$ se t_i é censurado à direita e $v_i = 1$ se t_i é um tempo de sobrevivência.

Com a finalidade de facilitar os cálculos, utilizaremos a parametrização dada por $\lambda = \log(\gamma)$, então, a distribuição de um tempo de sobrevivência T será denotada por $W(\alpha; \lambda)$. Nesse caso, a função densidade de probabilidade do tempo de sobrevivência T será dada por,

$$f(t/\alpha; \lambda) = \alpha t^{\alpha-1} \exp(\lambda - \exp(\lambda t^\alpha)). \quad (4.10)$$

e a função de sobrevivência no modelo Weibull será dada por,

$$S(t/\alpha; \lambda) = \exp(-\exp(\lambda t^\alpha)).$$

Note que a notação utilizada para a distribuição Weibull neste capítulo não é a mesma da utilizada no Capítulo 3.

A função de verossimilhança de $(\alpha; \lambda)$ pode ser escrita como:

$$\begin{aligned} L(\alpha, \lambda / t, \mathbf{v}) &= \prod_{i=1}^n f(t_i / \alpha, \lambda)^{v_i} S(t_i / \alpha, \lambda)^{1-v_i} \\ &= \alpha^d \exp\left\{d\lambda + \sum_{i=1}^n (v_i(\alpha - 1) \log(t_i) - \exp(\lambda t_i^\alpha))\right\}, \end{aligned}$$

onde, $d = \sum_{i=1}^n v_i$.

4.3.1. MODELO WEIBULL SEM COVARIÁVEIS

No modelo Weibull sem a inclusão de covariáveis, os únicos parâmetros desconhecidos são α e λ . Neste caso não há nenhuma priori conjugada possível para $(\alpha; \lambda)$. Porém, uma suposição plausível é considerar que, a priori, α e λ são independentes. Desta forma, a priori conjunta $\pi(\alpha; \lambda)$ pode ser fatorada como:

$$\pi(\alpha; \lambda) = \pi(\alpha)\pi(\lambda).$$

Considere $G(\alpha_0; k_0)$ a priori Gama para α , e $N(\mu_0; \sigma_0^2)$ a priori Normal para λ . Neste caso, a distribuição a posteriori conjunta de $(\alpha; \lambda)$ será dada por:

$$\begin{aligned} \pi(\alpha, \lambda | t, v) &\propto L(\alpha, \lambda | t, v) \pi(\alpha | \alpha_0; k_0) \pi(\lambda | \mu_0; \sigma_0^2) \\ &\propto \prod_{i=1}^n f(t_i | \alpha, \lambda)^{v_i} S(t_i | \alpha, \lambda)^{1-v_i} \\ &= \alpha^{\alpha_0 + d - 1} \exp \left\{ d\lambda + \sum_{i=1}^n (v_i(\alpha - 1) \log(t_i) - \exp(\lambda t_i^\alpha)) - \right. \\ &\quad \left. - k_0 \alpha - \frac{1}{2\sigma_0^2} (\lambda - \mu_0)^2 \right\} \end{aligned}$$

A distribuição a posteriori conjunta de $(\alpha; \lambda)$ não possui uma forma fechada, o que nos leva a pensar em aplicar o *Gibbs Sampling* para poder amostrar das condicionais completas. Porém, as distribuições condicionais completas $\pi(\alpha | \lambda, t, v)$ e $\pi(\lambda | \alpha, t, v)$ também não possuem forma fechada o que implica que não podemos amostrar diretamente delas. Entretanto, o algoritmo *Adaptive Rejection* proposto por Gilks e Wild (1992) também pode ser utilizado para esse modelo uma vez que cada uma das distribuições a posteriori condicionais $\pi(\alpha | \lambda, t, v)$ e $\pi(\lambda | \alpha, t, v)$ são log-côncavas. Sendo assim, esse algoritmo é uma ferramenta útil que será utilizada para amostrar das condicionais completas.

4.3.2. MODELO WEIBULL COM COVARIÁVEIS

Seja $\mathbf{x}_i' = (1, x_{i1}, x_{i2}, \dots, x_{ip})$ o vetor de p-covariáveis para o indivíduo i. A inclusão de x_i no modelo será feita utilizando um modelo de regressão para λ . Usaremos que $\lambda_i = \mathbf{x}_i' \boldsymbol{\beta} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}$.

Os parâmetros desconhecidos desse modelo são α e $\boldsymbol{\beta}$. Considere que α e $\boldsymbol{\beta}$ são independentes a priori, isto é:

$$\pi(\alpha; \boldsymbol{\beta}) = \pi(\alpha)\pi(\boldsymbol{\beta}).$$

Assumindo uma distribuição $N_p(\boldsymbol{\mu}_0; \boldsymbol{\Sigma}_0)$ como distribuição a priori para $\boldsymbol{\beta}$ e uma distribuição Gama como distribuição a priori para α , teremos como distribuição a posteriori conjunta para $(\boldsymbol{\beta}, \alpha)$ a expressão abaixo:

$$\begin{aligned} \pi(\alpha; \boldsymbol{\beta} | t, \nu) \propto \alpha^{\alpha_0 + d - 1} \exp\left\{ \sum_{i=1}^n \nu_i x_i' \boldsymbol{\beta} + \nu_i (\alpha - 1) \log(t_i) - \right. \\ \left. - t_i^\alpha \exp(x_i' \boldsymbol{\beta}) - k_0 \alpha - \frac{1}{2} (\boldsymbol{\beta} - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}_0) \right\}. \end{aligned} \quad (4.11)$$

Como no modelo anterior sem covariável, a expressão (4.11) não possui uma forma fechada, levando ao uso do *Gibbs Sampling* para amostrar das condicionais completas. Para o modelo de regressão Weibull as posteriores condicionais completas $\pi(\alpha | \boldsymbol{\beta}, t, \nu)$ e $\pi(\boldsymbol{\beta} | \alpha, t, \nu)$ não apresentam forma fechada. Porém, é mostrado na literatura que as distribuições a posteriori condicionais completas são log-côncavas, logo, o uso do algoritmo *Adaptive Rejection* poderá ser considerado uma forma plausível para amostrar das condicionais completas para esse modelo.

5. EXEMPLO DE APLICAÇÃO

Esse capítulo apresentará um exemplo utilizado para ilustrar os modelos paramétricos vistos anteriormente, apresentando os resultados obtidos segundo os dois enfoques para a análise de dados de sobrevivência: clássico e Bayesiano. As análises sob o enfoque clássico foram realizadas com o auxílio do *software* STATA e para as análises sob o enfoque Bayesiano foi utilizado o WinBUGS.

5.1. DESCRIÇÃO DO EXEMPLO

Os dados são provenientes de um ensaio clínico duplo-cego realizado em pacientes de um grupo de estudos oncológicos (ECOG) com um determinado tipo de câncer de pele (Melanoma) os quais foram submetidos à quimioterapia pós-operatória. Esse exemplo foi retirado de Chen et al (2001).

O banco de dados utilizado para fins desse exemplo foi retirado de <http://merlot.stat.uconn.edu/~mhchen/survbook/> em Setembro de 2005 e possui o nome de E1684.

Esse ensaio foi realizado para comparar um grupo de pacientes submetidos a doses elevadas de Interferon (IFN) com um grupo controle sem medicação pois recentemente diversos tratamentos quimioterápicos pós-operatórios têm sido propostos para esse tipo de pacientes. O tamanho da amostra foi de 255 pacientes estudados entre os anos de 1984 e 1990. A variável resposta é sobrevivência livre da reincidência, que pode ser censurada à direita.

5.2. MODELO EXPONENCIAL

A distribuição do tempo de sobrevivência T é suposta ser Exponencial com função densidade de probabilidade dada por (2.11). Esse modelo apresenta um único parâmetro desconhecido λ , onde $1/\lambda$ representa o tempo médio de sobrevivência.

A análise sem inclusão de covariáveis se resume em estimar λ considerando cada grupo de indivíduos em separado. Ou seja, estima um λ para o grupo de indivíduos tratados e outro λ para o grupo não tratado. Para ilustrar o modelo sem covariável usaremos apenas os dados dos indivíduos tratados. Para ilustrar o modelo com covariáveis, incluiremos a variável tratamento (TRAT) através da regressão $\lambda = 1 / \exp(\beta_0 + \beta_1 * \text{TRAT})$, onde $\text{TRAT} = 0$ se o indivíduo é não tratado e $\text{TRAT} = 1$ se o indivíduo é tratado com Interferon.

Logo, há dois parâmetros a serem estimados: β_0 e β_1 , onde β_0 representa o efeito de não utilizar o Interferon enquanto $\beta_0 + \beta_1$ representa o efeito do uso do Interferon.

Apesar de o exemplo considerar a inclusão de uma única covariável, o modelo é viável para a inclusão de mais covariáveis de modo simples e direto.

5.2.1. ANÁLISE CLÁSSICA

Essas análises foram realizadas no software STATA. Assim, obtemos:

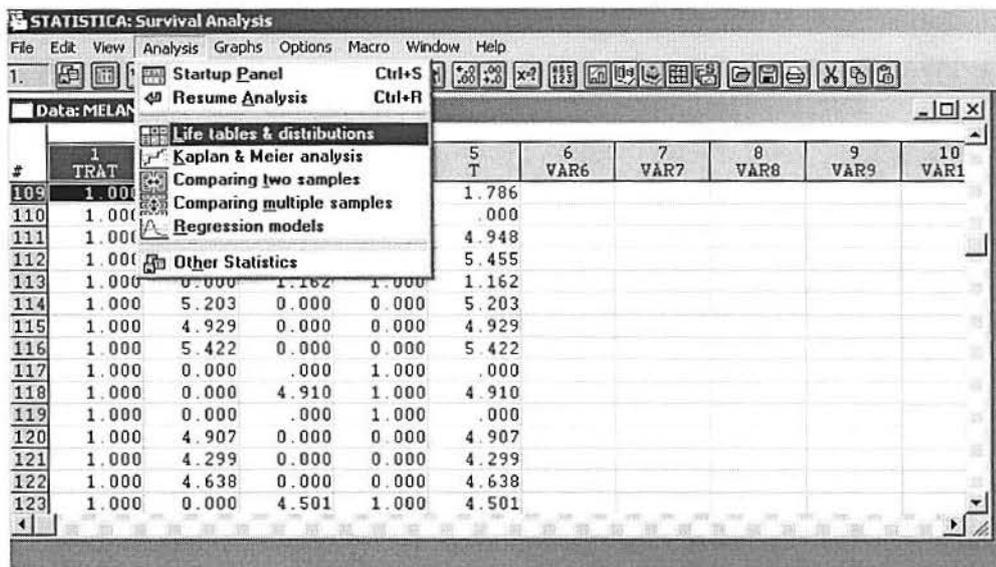
MODELO SEM COVARIÁVEIS:

Quando não há covariáveis no modelo, o interesse é estimar o tempo de sobrevivência diretamente. Serão analisados os dados do grupo de pacientes tratados com Interferon (IFN) a fim de verificar qual o tempo médio de sobrevivência desse grupo. Ou seja, o banco de dados a ser utilizado nesta análise compreende apenas os dados referentes aos indivíduos com valor 1 para a variável tratamento (TRAT).

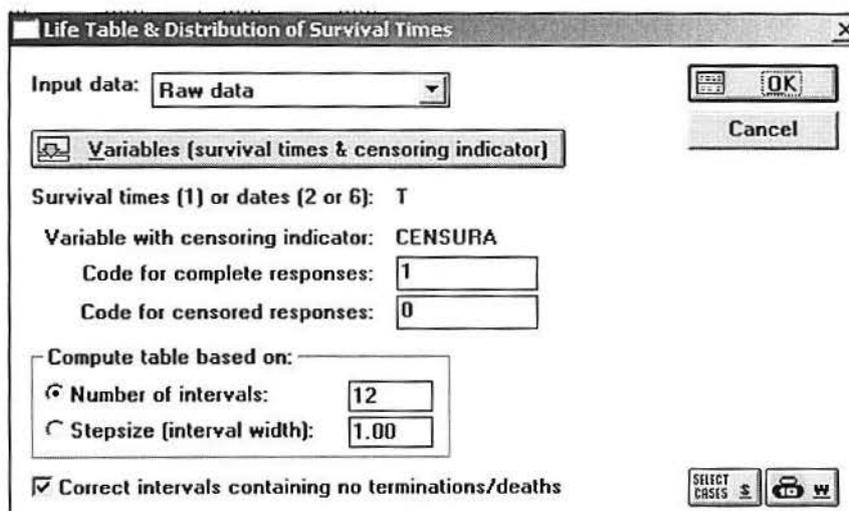
Nas páginas a seguir, apresentaremos os passos que devem ser realizados no *software* STATA.

1º - Abra o banco de dados. Este poderá estar em formato Excel, formato de planilha.

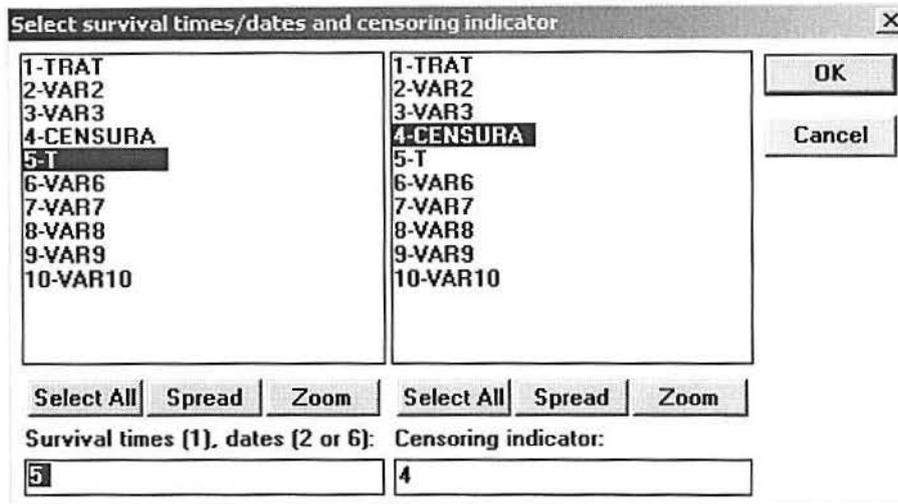
2º - Ir em: **ANALYSIS – LIFE TABLES AND DISTRIBUTIONS**



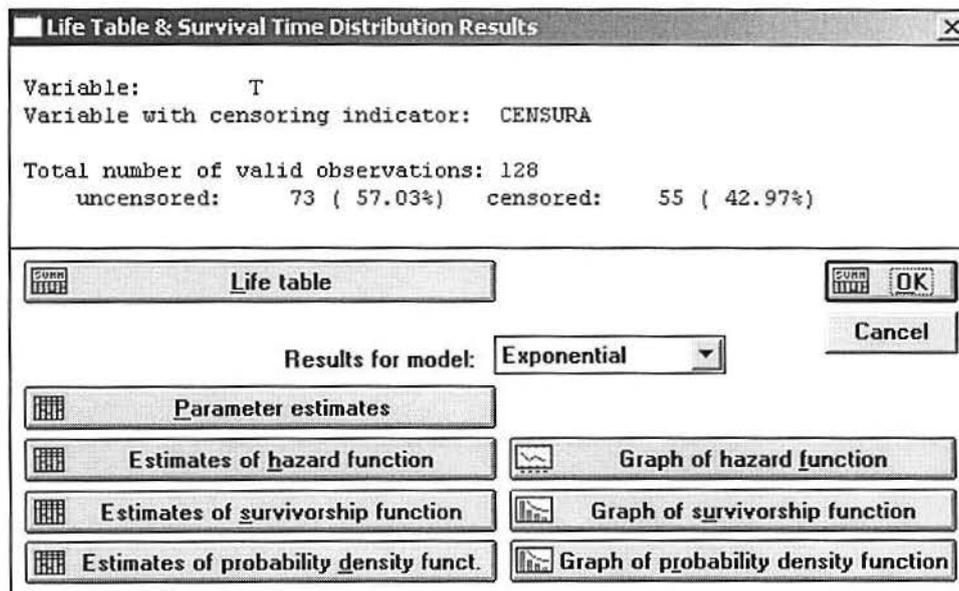
3º - Na janela de **LIFE TABLE**, aparecerá o menu *Input Data*. Selecione *Raw Data* e após clique em variables para definir a variável tempo de sobrevivência e variável indicadora de censura:



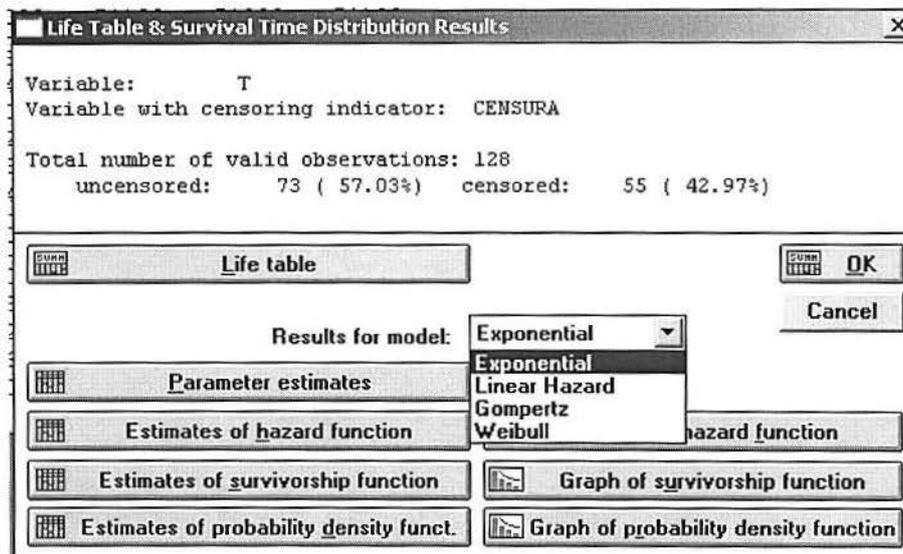
4º - Selecione as variáveis: tempo de sobrevivência e o indicador de censura.
Após, clique em OK.



5º – Na janela abaixo, defina o código de observação censurada e não censurada e clique em OK:



6º - Na janela de resultados, escolha o modelo que se quer para o tempo de sobrevivência. No caso do presente exemplo, escolha Exponencial como a seguir:



7º - Agora é só clicar em PARAMETER ESTIMATES para obter a estimativa para o parâmetro λ da Exponencial. O aplicativo fornece três valores para essa estimativa, baseados em 3 métodos de estimação diferentes. O primeiro deles estima o(s) parâmetro(s) por mínimos quadrados não ponderados, o segundo estima ponderando pelo inverso da variância da estimativa do risco e o terceiro método estima pelo produto do tamanho do intervalo pelo número de indivíduos expostos ao risco no mesmo intervalo. Neste trabalho será utilizado o primeiro método (*Weight*). Vejamos:

Parameter Estimates, Model: Exponential (melanoma.sta)							
Note: Weights: 1=1., 2=1./V, 3=N(I)*H(I)							
Estimate Method	Lambda	Variance Lambda	Std. Err. Lambda	Log-Likelihood	Chi-Sqr.	df	p
Weight 1	.119033	.000905	.030087	-225.878	33.23970	10	.000249
Weight 2	.085303	.000167	.012917	-233.134	47.75188	10	.000001
Weight 3	.149682	.000298	.017270	-224.656	30.79576	10	.000637

A estimativa obtida para o parâmetro λ é 0,119033.

Sabemos que a média para o tempo de sobrevivência dos pacientes submetidos ao tratamento com Interferon (IFN) é dada por $1/\lambda$. Assim, com a estimativa obtida, o tempo médio de sobrevivência estimado dos pacientes tratados é de, aproximadamente, 9,1 unidades de tempo. Usando a expressão em (3.4), construímos o IC-95% aproximado para λ . Ou seja, com 95% de confiança o tempo médio de sobrevivência dos pacientes tratados é um valor entre [6,64; 11,42]. Como a mediana de uma variável com distribuição Exponencial é dada por $(1/\lambda) \ln(2)$, então o tempo de sobrevivência mediano estimado dos pacientes submetidos ao tratamento com Interferon (IFN) é de 5,82 unidades de tempo, aproximadamente.

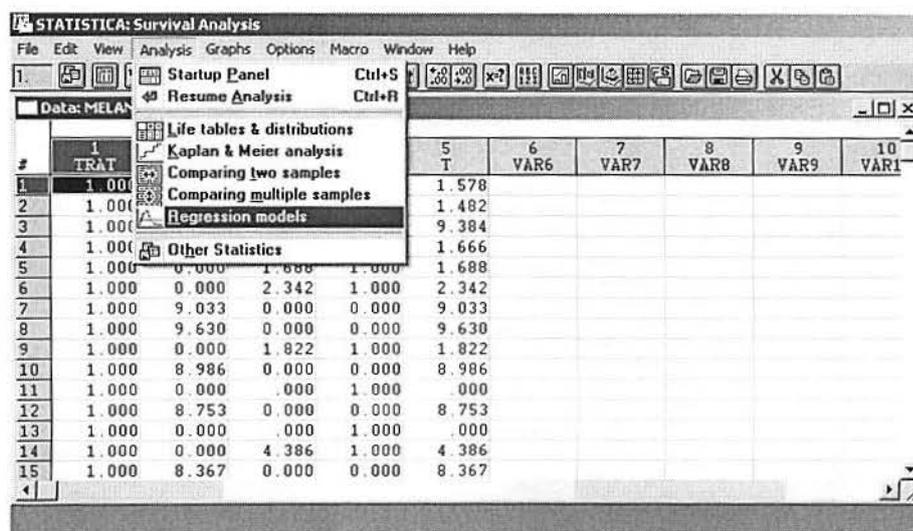
MODELO COM COVARIÁVEIS:

Para essa análise, o banco de dados utilizado foi o completo, considerando tanto os indivíduos tratados como também os não tratados.

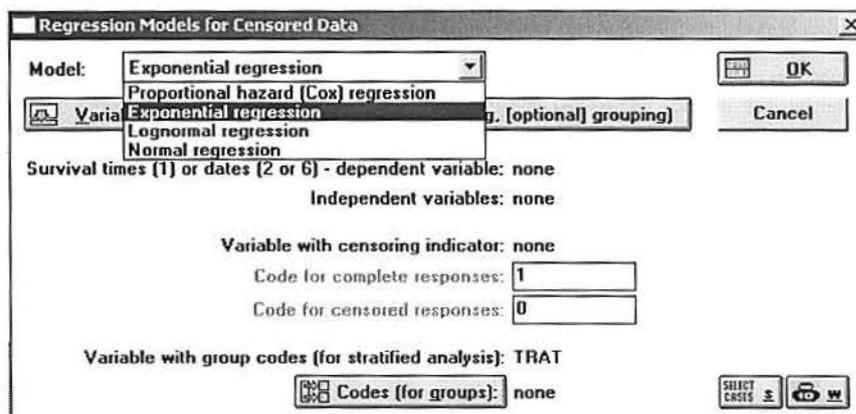
1º - Abra o arquivo de dados.

FILE – OPEN DATA... Ache o diretório onde estão salvos os dados, ou copie os mesmos de uma planilha Excel.

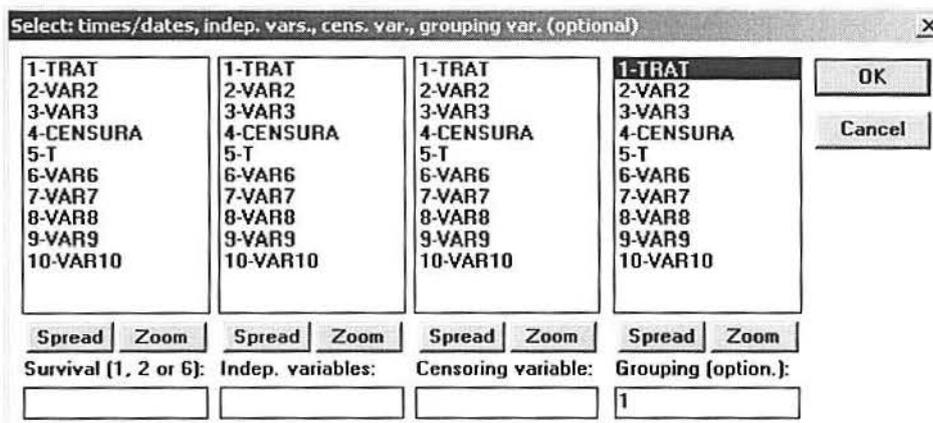
2º - Ir em: *ANALYSIS – REGRESSION MODELS*.



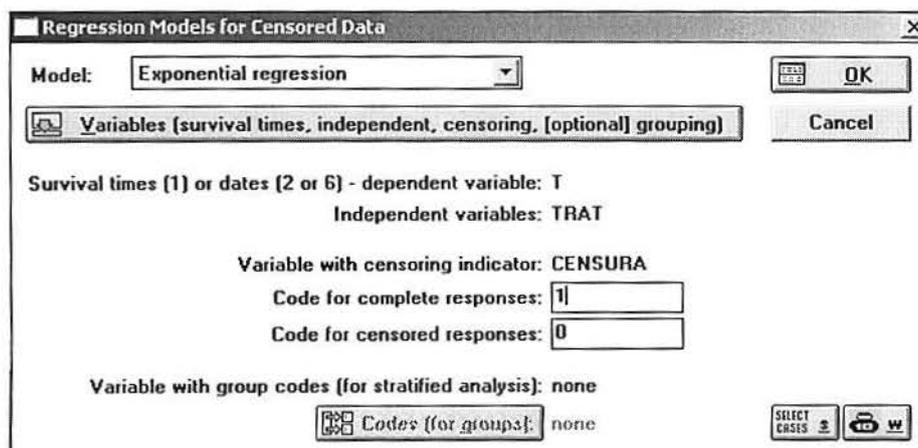
3º - Em modelos de regressão, escolha o modelo adequado, no caso do exemplo, modelo Exponencial:



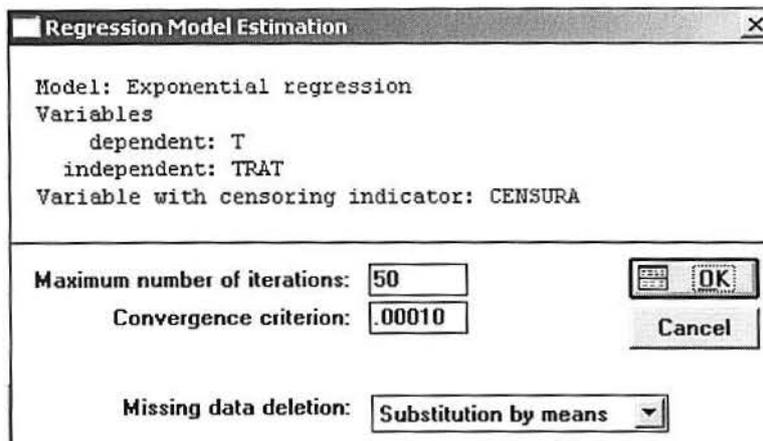
4º - Na janela anterior, clicar em *VARIABLES* para definir a variável de tempo de sobrevivência, censura, grupos, etc. Após definir as variáveis, clicar OK.



5º - Após a escolha das variáveis, colocar o código para observações censuradas e não-censuradas. Depois de definir a codificação de censura, clicar OK.



6º – Aparece a janela a seguir, com informações sobre o modelo, variáveis dependentes e independentes, variável de censura, número de interações, convergência e método de substituição de *missing data*:



Regression Model Estimation

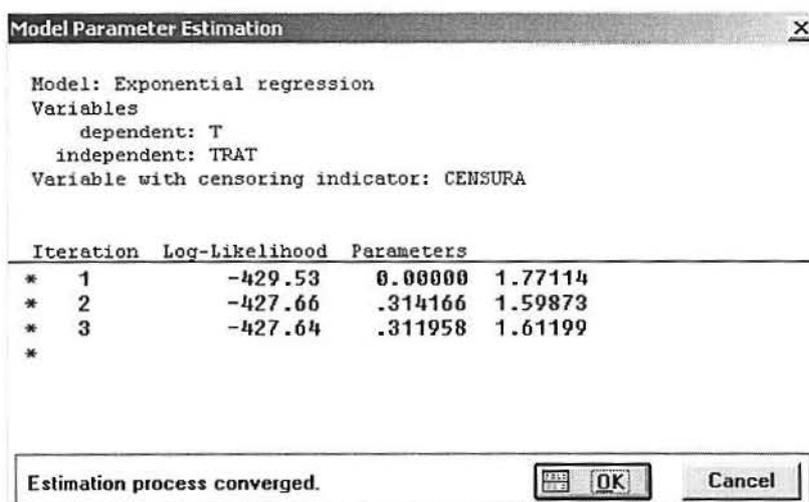
Model: Exponential regression
Variables
dependent: T
independent: TRAT
Variable with censoring indicator: CENSURA

Maximum number of iterations: 50
Convergence criterion: .00010

Missing data deletion: Substitution by means

OK Cancel

7º - Clique OK, surge a janela de Estimação dos parâmetros do modelo:



Model Parameter Estimation

Model: Exponential regression
Variables
dependent: T
independent: TRAT
Variable with censoring indicator: CENSURA

Iteration	Log-Likelihood	Parameters
* 1	-429.53	0.00000 1.77114
* 2	-427.66	.314166 1.59873
* 3	-427.64	.311958 1.61199
*		

Estimation process converged.

OK Cancel

8º - Na janela anterior, clique em OK:

Regression Results

Model: Exponential regression
 Variables
 dependent: T
 independent: TRAT
 Variable with censoring indicator: CENSURA

Total number of valid observations: 255
 uncensored: 155 (60.78%) censored: 100 (39.22%)

Log-Likelihood of final solution: -427.644
 Log-Likelihood of Null model (all β 's=0): -429.526
 Chi-Square (Null model - final solution): 3.765176 df= 1 p= .05234

Parameter estimates

Parameter variance/covariance matrix Graph residuals vs. alpha

Stratified analysis Graph residuals vs. theta

Means & Standard Deviations Graph residuals vs. log(t)

9º - Clicando em *PARAMETER ESTIMATES*, tem-se as estimativas para os parâmetros do modelo que estamos ajustando:

Dependent Variable: T (melanoma.sta)			
<input type="button" value="Continue..."/>	Censoring var.: CENSURA Chi ² = 3.76518 df = 1 p = .05234		
N=255	Beta	Standard Error	t-value
TRAT	.311958	.160909	1.93872
Constant	1.611993	.110427	14.59786

O coeficiente β_0 estimado é de 1,611993 e o coeficiente β_1 estimado é de 0,311958. Ou seja, o tempo médio de sobrevivência estimado do grupo de pacientes tratados com Interferon é dado por $\exp(1,611993 + 0,311958) = 6,8479$ unidades de tempo. Já o tempo médio de sobrevivência estimado dos pacientes não tratados é de $\exp(1,611993) = 5,0128$ unidades de tempo.

A razão dos tempos de sobrevivência medianos dos grupos tratado em relação aos não tratados é dada por: $\exp(\beta_1) = \exp(0,311958) \approx 1,37$. Ou seja, estima-se que o tempo de sobrevivência mediano do grupo tratado é 1,37 vezes o tempo de sobrevivência mediano do grupo não tratado. A razão de risco entre os grupos tratado e não tratado é dada por: $\exp(-\beta_1) = \exp(-0,311958) = 0,73$. Ou

seja, estima-se que o risco do grupo tratado é igual a 0,73 vezes o risco do grupo não tratado.

5.2.2. ANÁLISE BAYESIANA

Essa análise será feita com auxílio do Excel para o caso do modelo sem covariável. Já as análises com inclusão de uma covariável serão feitas no software WinBUGS. O software WinBUGS é de uso livre e pode ser obtido em: <http://www.mrc-bsu.cam.ac.uk/bugs>. O aplicativo serve para auxiliar nas análises Bayesianas de modelos estatísticos complexos utilizando-se das técnicas de MCMC.

MODELO SEM COVARIÁVEIS:

Dos 255 pacientes estudados, 128 recebem o tratamento (possuem valor 1 na variável tratamento) e, dentre esses, 55 são tempos de sobrevivência e os demais são tempos censurados. Assim, estabelecendo, primeiramente, uma priori não-informativa para o parâmetro λ da Exponencial e aplicando os resultados obtidos na Sessão 4.2.1, temos:

PRIORI PARA λ : $\pi(\lambda) \sim \text{Gama}(0,001; 0,001)$.

POSTERIORI PARA λ : $\pi(\lambda / x) \sim \text{Gama}(55,001; 499,933)$

Média a posteriori para λ : $E(\lambda / x) = (55,001 / 499,933) = 0,11$.

O valor esperado para o tempo médio de sobrevivência é dado pelo inverso da média da distribuição a posteriori que, neste exemplo, é de $1/0,11 \approx 9,1$ unidades de tempo. Analogamente, o valor mais provável para o tempo médio de sobrevivência a posteriori é dado pelo inverso da moda da distribuição a posteriori que, neste exemplo, é de $1/0,10802 = 9,26$ unidades de tempo.

A estimativa por Intervalo de Credibilidade HDR 95% para λ é: [0,0817; 0,1395]. Com estes valores podemos estimar que com 0,95 de probabilidade a posteriori, o tempo médio de sobrevivência dos pacientes tratados com Interferon é um valor entre [7,17; 12,24].

A comparação dos resultados da análise clássica com a análise Bayesiana deve ser feita utilizando as estimativas baseadas na moda da posteriori. Note que os resultados obtidos são análogos ao resultado clássico. Isso era esperado, pois não há nenhum conhecimento a priori sobre o parâmetro λ , sendo assim, ao utilizar priori não-informativa, a posteriori será o retrato dos dados observados.

Com o objetivo de ilustrar a influência da informação a priori para λ na respectiva posteriori, vamos estabelecer prioris informativas para o parâmetro λ da Exponencial. Suponha que, a priori, o pesquisador tem grande convicção de que, em média, o tempo médio de sobrevivência é algo em torno de 1 unidade de tempo, ou seja que $\lambda = 1$. Essa opinião é hipotética e escolhida para ser bem extrema em relação à informação dos dados. Uma priori que representa essa opinião é a Gama (64; 64) que apresenta média igual a 1 e variância igual a 0,0156. Os valores dos parâmetros da priori Gama foram escolhidos de modo a representar a opinião do pesquisador.

A Figura 5.1 apresenta a priori, a verossimilhança padronizada e a posteriori em um mesmo gráfico. Este gráfico tem por objetivo mostrar a influência do uso de priori informativa sobre os dados. Analisando o gráfico é possível observar que a média da priori é algo em torno de 1, que a média da verossimilhança padronizada é algo em torno de 0,1 e que a média da posteriori é algo um pouco maior que 0,2. Observa-se também que a variabilidade da verossimilhança é bem menor do que a variabilidade da priori. Com isto, a posteriori acaba sofrendo maior influência da verossimilhança e fica deslocada para seu lado.

Agora, vamos supor que um outro pesquisador tenha conhecimento de que, em média, o tempo médio de sobrevivência é algo em torno de 0,5 unidade de tempo, ou seja, que $\lambda = 2$. Uma priori que representa essa opinião é a Gama (256,410; 128,205) que apresenta média igual a 2 e variância igual a 0,0156. Os valores dos parâmetros da priori Gama foram escolhidos de modo a representar a opinião do pesquisador.

A Figura 5.2 apresenta a priori, a verossimilhança padronizada e a posteriori em um mesmo gráfico. Este gráfico tem por objetivo mostrar a influência

do uso de priori informativa sobre os dados. Analisando o gráfico é possível observar que a média da priori é algo em torno de 2, que a média da verossimilhança padronizada é algo em torno de 0,1 e que a média da posteriori é algo em torno de 0,5. Observa-se ainda que a variabilidade da verossimilhança é menor do que a variabilidade da priori. Assim, a posteriori acaba sofrendo maior influência da verossimilhança e fica deslocada para seu lado, porém está mais próxima da priori, se compararmos com a suposição do primeiro pesquisador.

Notamos que, considerando a mesma variabilidade, a segunda priori utilizada influencia mais a informação da amostra deslocando mais o gráfico da posteriori, uma vez que a informação a priori no segundo gráfico é mais discrepante em relação aos dados amostrais.

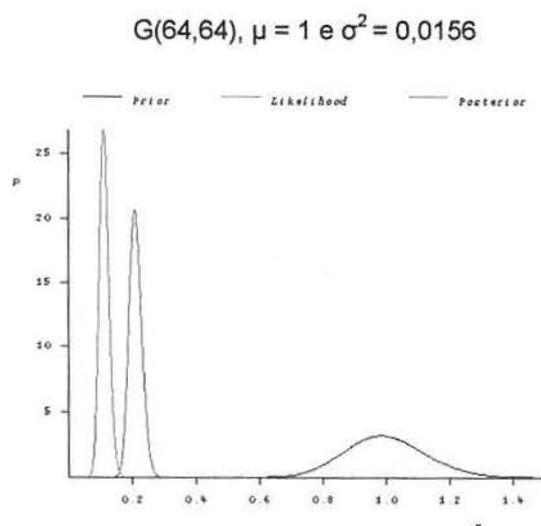


Figura 5.1

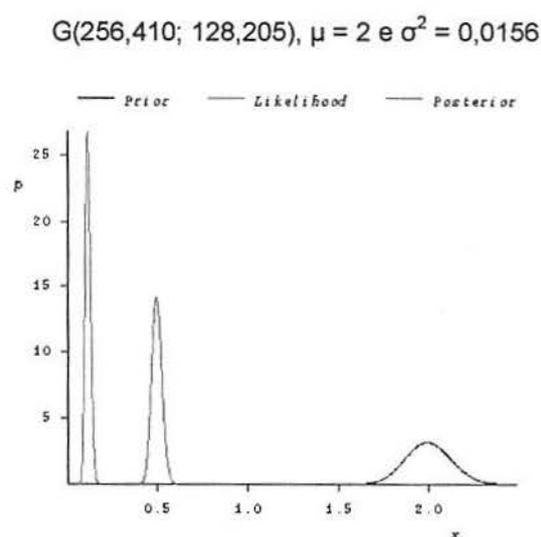


Figura 5.2

Ao utilizar uma priori informativa, estamos agregando informação aos dados obtidos na amostra e a partir da combinação dessa informação com a informação amostral as inferências para o parâmetro de interesse serão feitas.

Por exemplo, para a opinião do primeiro pesquisador temos que:

PRIORI PARA λ : $\pi(\lambda) \sim \text{Gama}(64; 64)$.

POSTERIORI PARA λ : $\pi(\lambda / x) \sim \text{Gama}(119; 563,932)$

Média a posteriori para λ : $E(\lambda / x) = (119 / 563,932) = 0,2110$.

A posteriori, o valor esperado para o tempo médio de sobrevivência dos pacientes tratados é de, aproximadamente, $1/0,2110 \approx 4,7$ unidades de tempo.

Note que, a priori especificada influenciou bastante a informação da amostra, uma vez que com ausência de informação a priori, a estimativa do tempo médio de sobrevivência a posteriori era de 9,1 unidades de tempo e passa para 4,7 unidades de tempo com o uso dessa informação convicta do pesquisador.

MODELO COM COVARIÁVEL:

Para os coeficientes β_0 e β_1 assumiram-se distribuições a priori Normais independentes, com média zero e precisão 10^{-4} , denotadas por $\beta \sim N_2(\mathbf{0}; 10^{-4} I_2)$, onde I_2 representa a matriz identidade de dimensão 2. Note que a precisão de uma Normal é igual ao inverso da variância. Aqui estamos utilizando precisão pois é a maneira utilizada pelo software WinBUGS. Uma precisão de 10^{-4} é igual a uma variância de 10^4 . Este valor para a variância torna essa priori não-informativa visto que uma variância de 10^4 pode ser considerada grande. Neste caso, isto pode ser verificado comparando resultados da posteriori para β_0 e β_1 com esse valor de variância e com um valor maior ainda. Resultados parecidos significam que a priori não está influenciando na posteriori, ou seja, que ela é não-informativa.

Através do pacote WinBUGS foram geradas 4000 Gibbs Samples com um burn-in de 1000, resultando em uma amostra de 3000 valores.

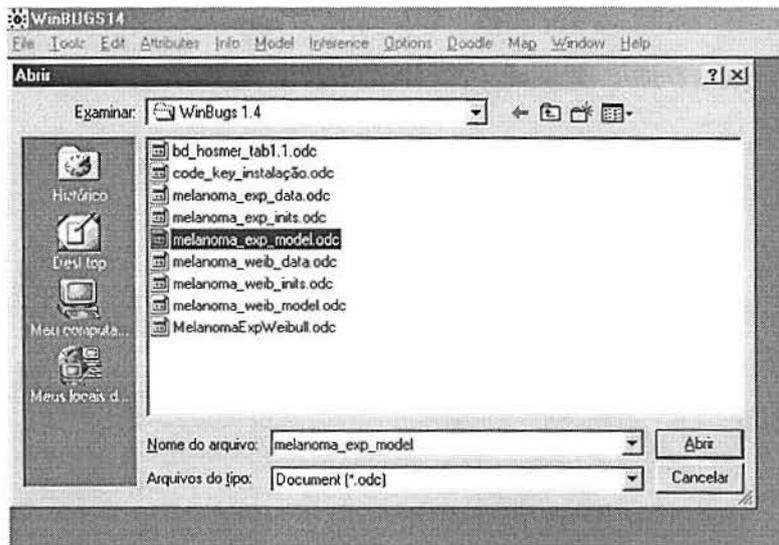
Com esses dados simulados, obtiveram-se os resultados a posteriori para β e as estimativas pela abordagem bayesiana. A seguir, os resultados bem como um passo-a-passo de como realizar uma análise no software WinBUGS:

1º - Abra o programa clicando no ícone

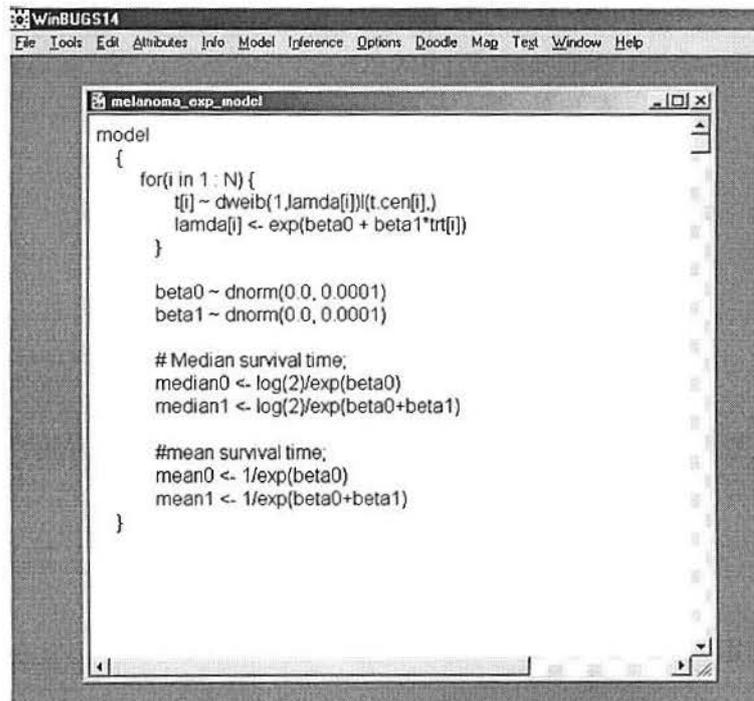


3º - Abra o arquivo do programa (modelo) em nova janela. Há diversos modelos prontos no software, porém o usuário tem total liberdade para programar seus modelos e salvar em um formato txt.

FILE – OPEN – ABRIR

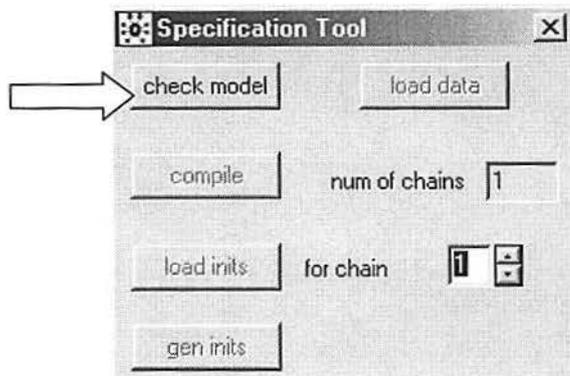


A janela da linguagem de programação estará aberta:

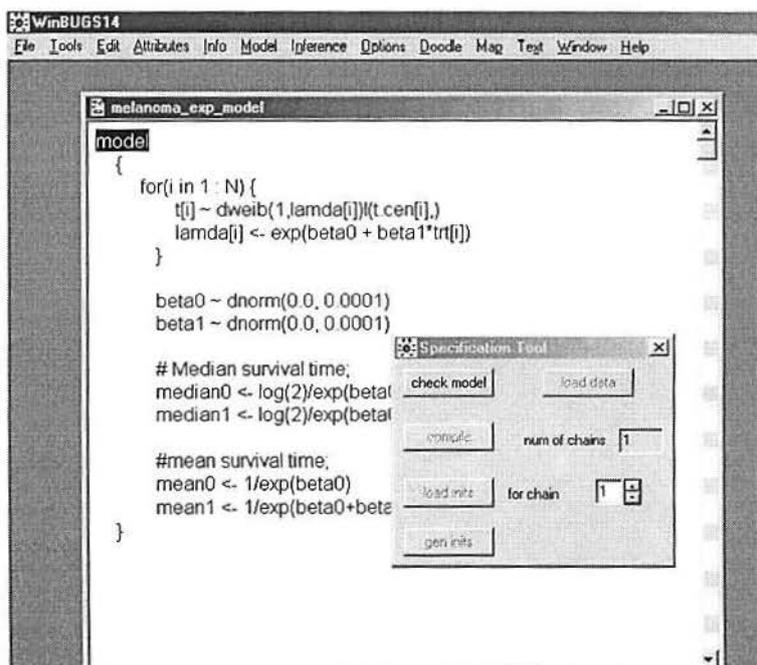


4º - Com o arquivo do programa aberto, podemos começar a analisar os dados. Para isso, começamos pela verificação do modelo:

MODEL – SPECIFICATION... surgirá uma janela como a abaixo (*Specification Tool*):

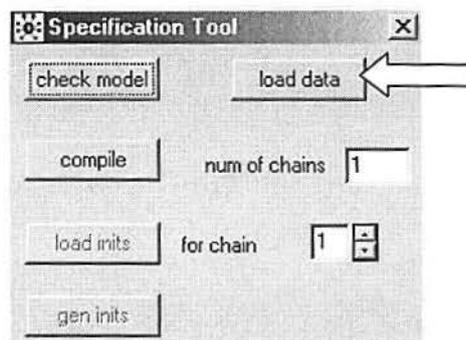


5º - Nessa janela aparece um botão "*check model*" que é através dele que iremos verificar se o modelo está com a sintaxe correta. Na janela onde está a linguagem de programação, selecionar a palavra *MODEL* com o mouse e deixar selecionada.

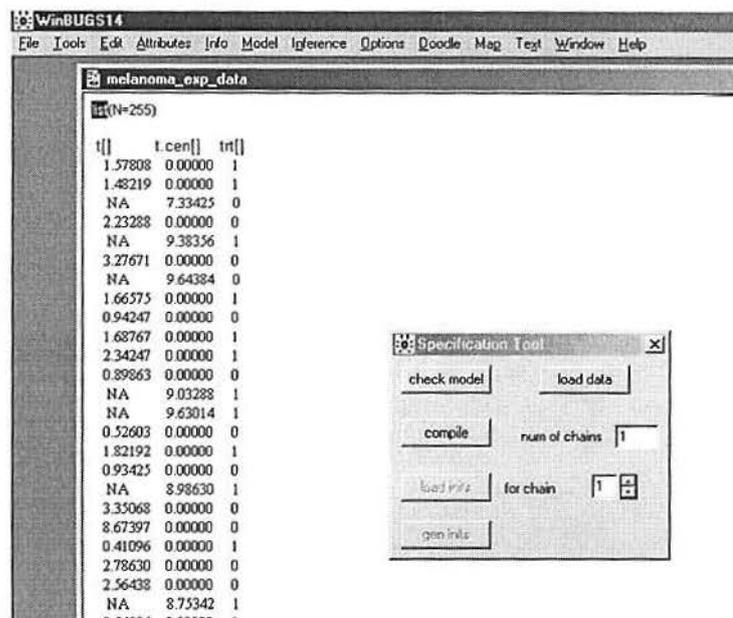


6º - Com a palavra selecionada, voltar à janela *Specification Tool* e clicar em “*check model*”. Se o modelo estiver correto, deve aparecer no canto esquerdo da janela do software a mensagem: “*model is syntactically correct*”.

E ainda, após isso feito, a palavra *model* da linguagem de programação ficará dentro de um retângulo, então se abre um novo botão em *Specification Tool*: “*load data*”.



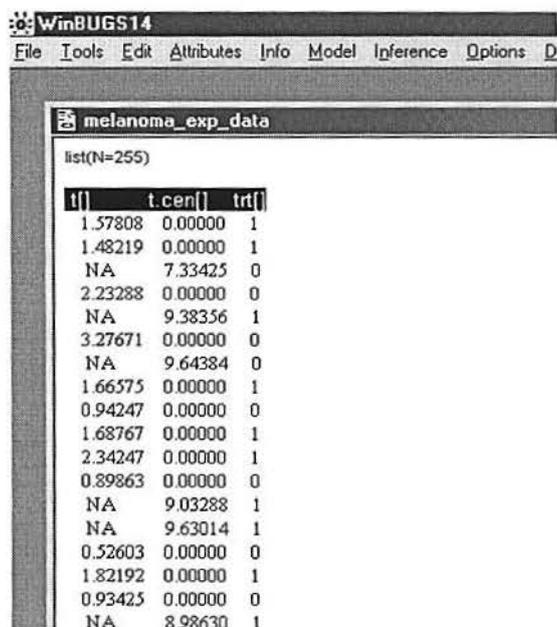
7º - Ir para a janela dos dados que está aberta (passo1). Nessa janela, selecionar primeiro a palavra *list* que deve estar no topo da janela. Veja:



8º - Estando a palavra *list* selecionada, clique em “*load data*”. Deverá surgir a mensagem: “*Data Loaded*” no canto inferior esquerdo do aplicativo.

9º - Após, na janela dos dados selecione os nomes das variáveis, logo abaixo de *list*. Com o nome das variáveis selecionado como na figura anterior, clicar

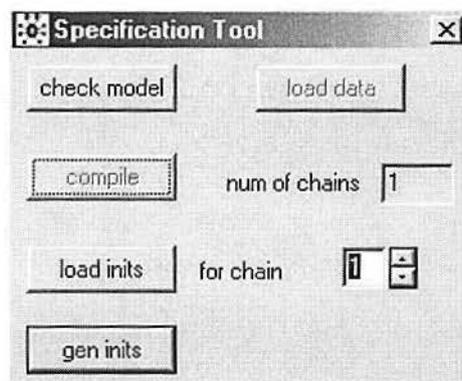
novamente em "load data". Irá surgir no canto inferior esquerdo da tela do aplicativo a mensagem: *Data Loaded*.



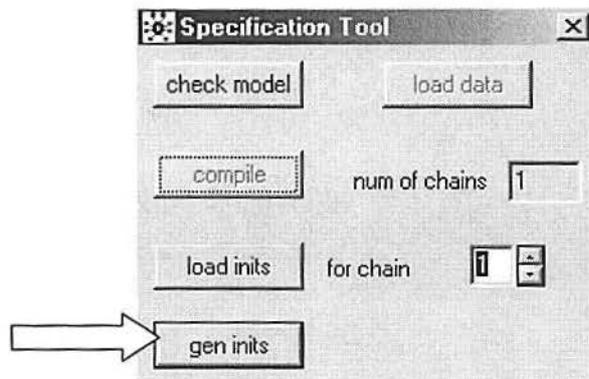
list(N=255)

t[]	t.cen[]	tri[]
1.57808	0.00000	1
1.48219	0.00000	1
NA	7.33425	0
2.23288	0.00000	0
NA	9.38356	1
3.27671	0.00000	0
NA	9.64384	0
1.66575	0.00000	1
0.94247	0.00000	0
1.68767	0.00000	1
2.34247	0.00000	1
0.89863	0.00000	0
NA	9.03288	1
NA	9.63014	1
0.52603	0.00000	0
1.82192	0.00000	1
0.93425	0.00000	0
NA	8.98630	1

10º - A partir de agora, com os dados carregados, devemos compilar o modelo, clicando em *COMPILE MODEL*. Aparecerá no canto inferior esquerdo do aplicativo a mensagem: "*model compiled*". A janela *Specification Tool* ficará assim:

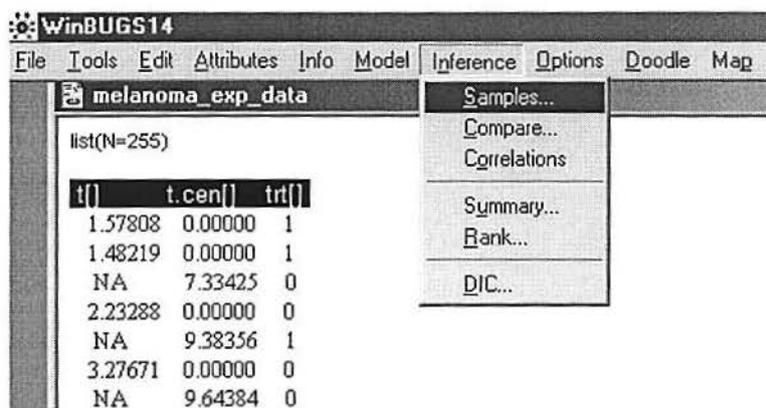


11º - Agora, precisamos gerar os valores iniciais da simulação. Isso pode ser feito clicando no botão *gen inits*.



12º - Clicando-se em *gen inits*, o processo de simulação irá gerar valores iniciais e o modelo poderá ser iniciado. A mensagem: *initial values generated, model initialized* deve surgir no canto inferior esquerdo do aplicativo.

13º - Fecha-se a janela *Specification Tool* e clica-se em:
INFERENCE – SAMPLES



14º - Na janela a seguir, entrar com os nomes dos parâmetros a serem estimados no modelo. Esses nomes devem estar em acordo com o que foi escrito na linguagem de programação. Onde aparece *NODE* escreve-se o nome do

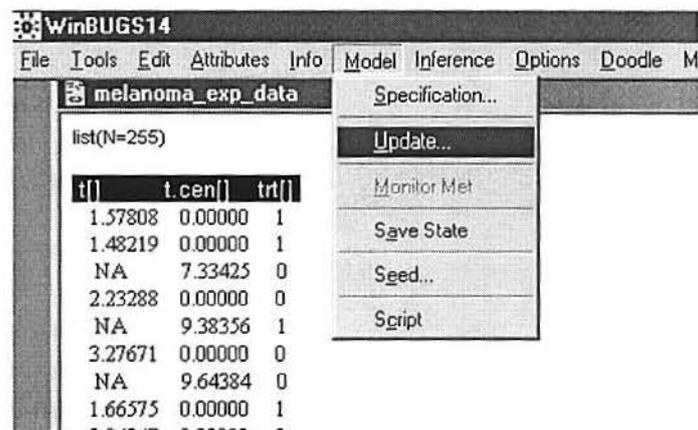
parâmetro e após clica-se em SET. Repete-se até que todos os parâmetros que desejamos estimar tenham sido escritos e incluídos no conjunto de parâmetros.

Essa janela seria um monitor dos parâmetros, devemos especificar aqui para quais parâmetros iremos querer visualizar os resultados.

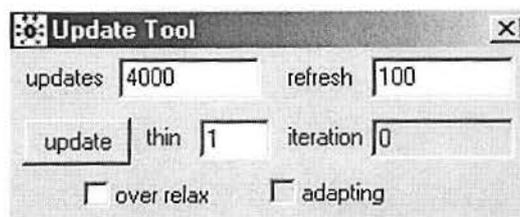


15º - Fecha a janela.

16º - Ir em: *MODEL – UPDATE*



17º - Aqui, iremos escolher o tamanho da cadeia a ser simulada. Escolhe-se o número de iterações (no exemplo são 4000) e de quanto em quanto a simulação irá acontecer, nesse caso, de 100 em 100.

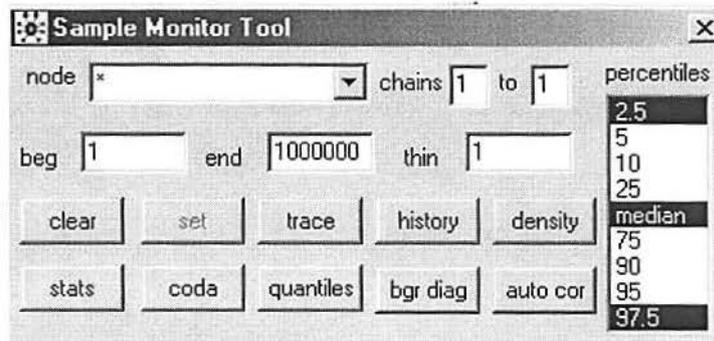


18° - Aguarda-se o *update* do modelo. Quando acabar, aparece no canto inferior da tela: *model is updating*. Ao término da simulação, aparecerá *updated took ...s*, dando o tempo total de duração da simulação.

19° - Fecha-se essa janela.

20° - Volta-se em INFERENCE – SAMPLES.

Agora, podemos visualizar os resultados a posteriori para os parâmetros do modelo estimado. Basta digitar um asterisco (*) na caixa NODE que teremos os resultados para todos os parâmetros que foram definidos antes



21° - Clicando em *history* podemos obter os gráficos com as trajetórias dos valores simulados para termos para β_0 e β_1 .

22° - Clicando em *density* podemos obter gráficos com estimativas não paramétricas das distribuições a posteriori. Estas estimativas são feitas utilizando o método de Kernel e representam o núcleo das respectivas distribuições a posteriori.

A Figura 5.3 apresenta as trajetórias dos valores simulados para β_0 e β_1 . Através delas podemos visualizar que a partir do valor 1000 da simulação parece haver convergência. Para verificar se este período de burn-in igual a 1000 é suficiente, é importante olhar para o gráfico de trajetórias sem as primeiras 1000 simulações. Estes gráficos estão apresentados na Figura 5.4. Através dos gráficos da Figura 5.4 podemos visualizar que a convergência da cadeia parece estar razoável. A análise da convergência foi feita apenas através de inspeção visual.

Descrição de técnicas mais formais para analisar convergência podem ser encontradas em, por exemplo, Gilks *et al* (1997), mas não serão abordadas aqui.

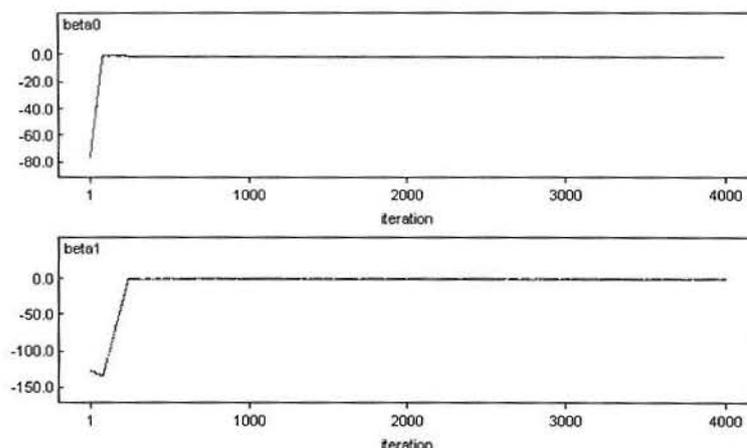


Figura 5.3 - Gráfico das trajetórias da simulação considerando o Burn-in

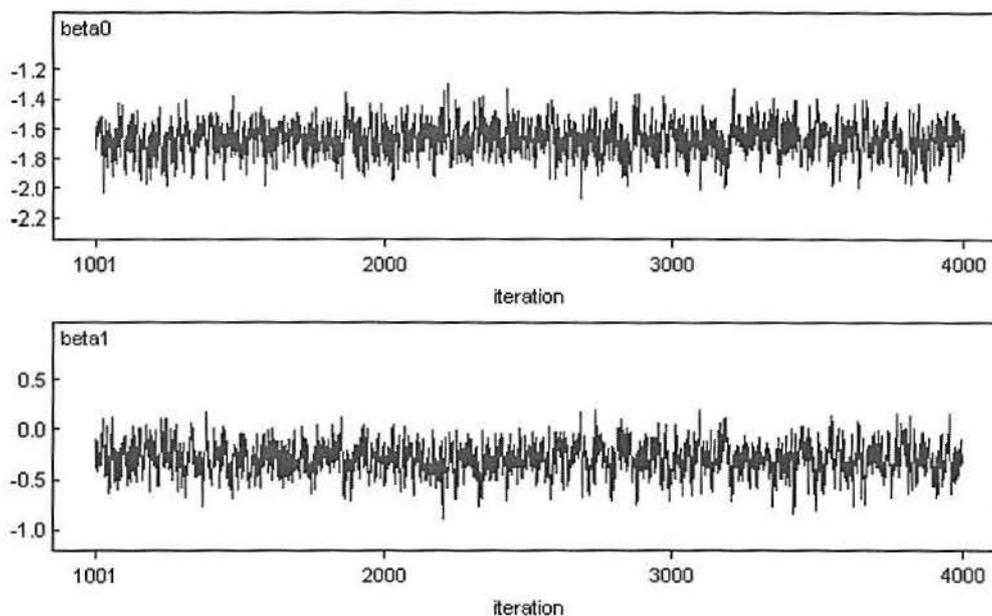


Figura 5.4 - Gráfico das trajetórias da simulação desconsiderando o Burn-in de 1000 amostras

A Figura 5.5 apresenta estimativas baseadas no método de Kernel para as distribuições a posteriori marginais de β_0 e β_1 . Ela revela uma certa simetria nas distribuições a posteriori dos parâmetros. Com isso, a média a posteriori dos parâmetros β_0 e β_1 serão utilizadas para estimação do tempo médio de

sobrevivência já que, em distribuições simétricas, a moda, a mediana e a média são as mesmas.

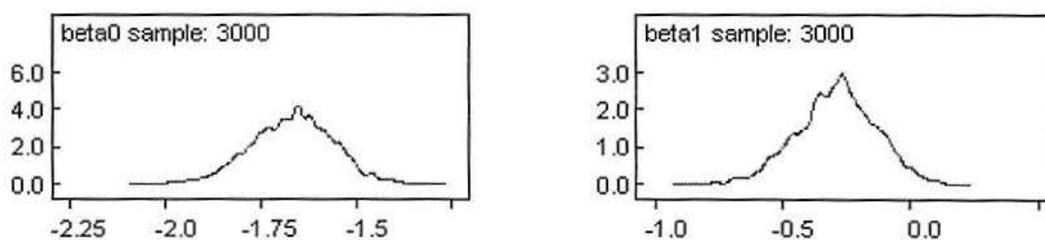


Figura 5.5 - Gráficos dos núcleos das densidades a posteriori para β_0 e β_1

A Tabela 5.1 apresenta os resultados a posteriori para β . Esses resultados serão utilizados para estimar o tempo médio de sobrevivência já que o modelo de regressão foi introduzido através do parâmetro λ da distribuição Exponencial. Assim, ficamos com:

Tempo médio de sobrevivência estimado para o grupo não tratado ($x = 0$):

$$E(T) = 1/\lambda = 1/\exp(-1,6690) = 1/0,1884 \approx 5,31 \text{ unidades de tempo.}$$

Intervalo de Credibilidade HDR 95% para o tempo médio de sobrevivência do grupo não tratado ($x = 0$):

$$\text{HDR } -95\% = (1/\exp(-1,4570); 1/\exp(-1,8850)) = (4,29; 6,59)$$

Tempo médio de sobrevivência estimado para o grupo tratado ($x = 1$):

$$E(T) = 1/\lambda = 1/\exp(-1,6690 - 0,2893) = 1/0,1411 \approx 7,09 \text{ unidades de tempo.}$$

Intervalo de Credibilidade HDR 95% para o tempo médio de sobrevivência do grupo tratado ($x = 0$):

$$\text{HDR } -95\% = (1/\exp(-1,4570 + 0,0079); 1/\exp(-1,8850 - 0,6083)) = (4,25; 12,10).$$

A estimativa pontual para o tempo médio de sobrevivência mostra superioridade para o grupo tratado. Porém, através do intervalo de credibilidade HDR podemos dizer que, com 0,95 de probabilidade a posteriori, o tempo médio de sobrevivência não deve diferir entre o grupo tratado e o grupo não tratado.

Tabela 5.1. Resultados a posteriori para os parâmetros β_0 e β_1

Parâmetro	Média	D.P.	2,5%	97,5%	Mediana
β_0	- 1,6690	0,1083	- 1,8850	- 1,4570	- 1,6640
β_1	- 0,2893	0,1575	- 0,6083	0,0079	- 0,2832

A razão dos tempos medianos estimados do grupo tratado em relação ao não tratado é igual a $1/\exp(-0,2893) = 1,33$, ou seja, estima-se que o tempo de sobrevivência mediano do grupo tratado é igual a 1,33 vezes o tempo de sobrevivência mediano do grupo não tratado. O risco relativo estimado dos pacientes tratados em relação aos não tratados é igual a 0,7534, aproximadamente. Ou seja, estima-se que o risco de morte no grupo tratado com Interferon seja igual a 0,7534 vezes o risco no grupo não tratado.

5.3. MODELO WEIBULL

A fim de ilustração, foram realizadas as análises para as duas abordagens no modelo Exponencial. Agora, serão feitas as análises apenas sob o enfoque Bayesiano para o modelo Weibull com covariáveis.

A distribuição do tempo de sobrevivência T é suposta ser Weibull com função densidade de probabilidade dada por (4.10). Esse modelo apresenta dois parâmetros desconhecidos: α e λ .

Para ilustrar o modelo com covariáveis, incluiremos a variável tratamento (TRAT) através da regressão $\lambda = \exp(\beta_0 + \beta_1 \cdot \text{TRAT})$, onde $\text{TRAT} = 0$ se o indivíduo é não tratado e $\text{TRAT} = 1$ se o indivíduo é tratado com Interferon. Logo há três parâmetros a serem estimados: α , β_0 e β_1 , onde β_0 representa o efeito de não utilizar o Interferon enquanto $\beta_0 + \beta_1$ representa o efeito do uso do Interferon. Apesar de o exemplo considerar a inclusão de uma única covariável, o modelo é viável para a inclusão de mais covariáveis de modo simples e direto.

5.3.1. ANÁLISE BAYESIANA

A análise será feita com auxílio do software WinBUGS.

MODELO COM COVARIÁVEIS:

Seja $x_i' = (1, x_{i1}, x_{i2}, \dots, x_{ip})$ o vetor de p-covariáveis para o indivíduo i. A inclusão de x_i no modelo será feita utilizando um modelo de regressão para λ . Usaremos que $\lambda_i = x_i' \beta = \beta_0 + \beta_1 \cdot \text{TRAT}$, onde TRAT é a covariável tratamento.

Para este exemplo foi utilizado o algoritmo *Gibbs Sampling* com o método *Adaptive Rejection* para simular das condicionais completas. Utilizou-se uma cadeia de 4000 valores com burn-in de 1000. Ou seja, para as estimativas a posteriori foi considerado uma amostra de 3000 valores.

A Figura 5.6 apresenta as trajetórias dos valores simulados para α , β_0 e β_1 desconsiderando o período burn-in. Através deles podemos visualizar que a convergência da simulação para as distribuições a posteriori marginais de α , β_0 e β_1 parece razoável.

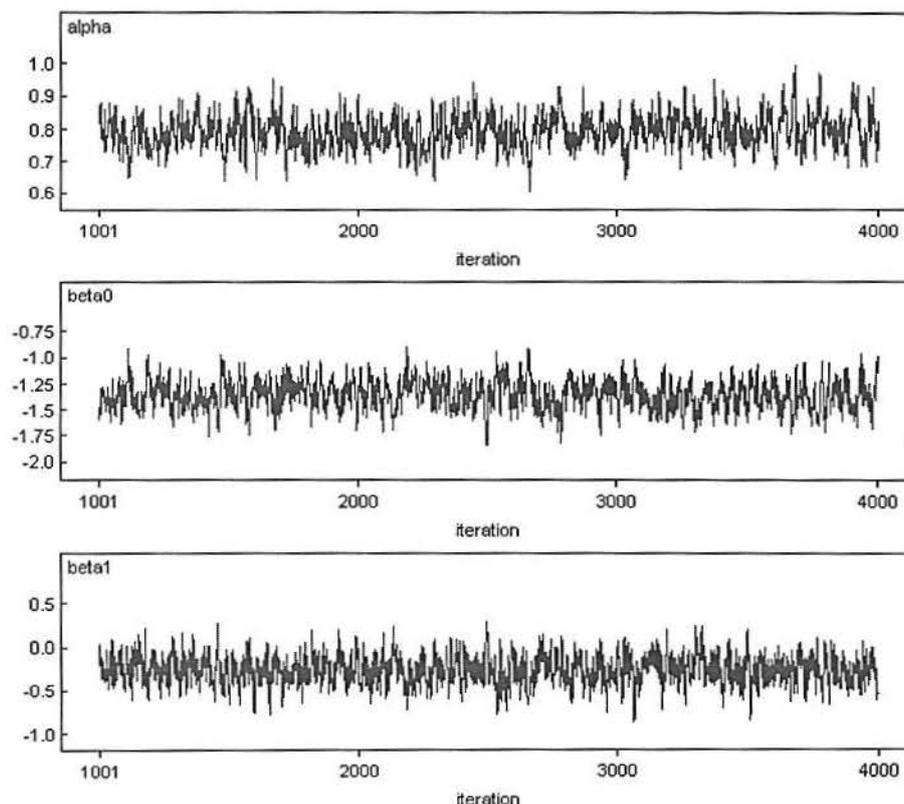


Figura 5.6 - Gráfico do histórico da simulação considerando um Burn-in de 1000 amostras

A Figura 5.7 apresenta estimativas pelo método de Kernel para as densidades a posteriori dos parâmetros. Note que, as distribuições a posteriori de todos os parâmetros parecem ser simétricas. Por este motivo, podemos utilizar a média das posteriores dos parâmetros β_0 e β_1 para estimar as quantidades de interesse. A Tabela 5.2 apresenta estas médias bem como Intervalos de Credibilidade Centrais para cada parâmetro.

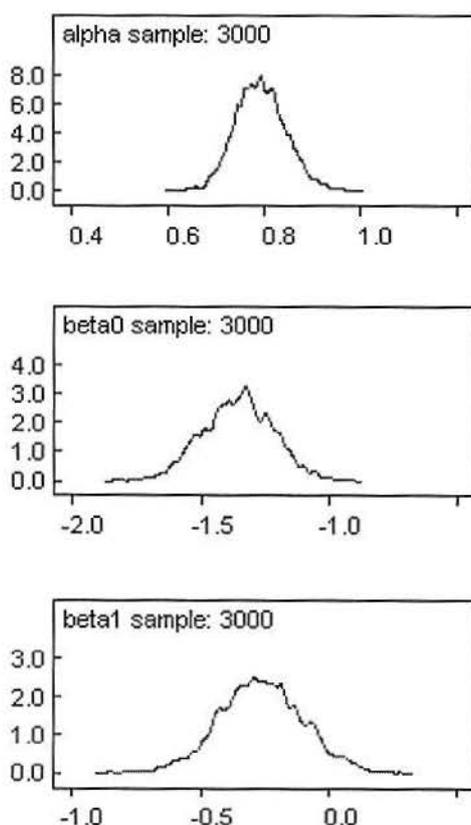


Figura 5.7 - Gráficos dos núcleos das densidades a posteriori para α , β_0 e β_1 .

Como exemplo de estimativas citamos a razão de tempos medianos de sobrevivência entre os grupos tratado e não tratado. Ela é dada por $1/e^{\beta_1}$. Neste caso teremos $1/\exp(-0,2695) = 1,31$. Ou seja, o tempo mediano de sobrevivência do grupo tratado com Interferon é igual a 1,31 vezes o tempo de sobrevivência mediano do grupo não tratado. O Intervalo de Credibilidade Central é construído substituindo β_1 pelas suas respectivas estimativas. Desta forma, teremos:

$$(1/e^{0,0741}; 1/e^{-0,6027}) = (0,93; 1,83).$$

Tabela 5.2. Resultados a posteriori para os parâmetros da Weibull

Parâmetro	Média	D.P.	2,5%	97,5%	Mediana
α	0,7962	0,0525	0,6981	0,9001	0,7945
β_0	- 1,3620	0,1373	- 1,6340	- 1,0920	- 1,3610
β_1	- 0,2662	0,1708	- 0,6027	0,0741	- 0,2695

6. CONCLUSÕES

Este trabalho apresenta a abordagem clássica e a abordagem Bayesiana para analisar dados de sobrevivência sob a suposição de que o tempo de sobrevivência tenha distribuição Exponencial ou distribuição Weibull. O principal objetivo foi estabelecer um comparativo e ilustrar as duas abordagens. Para este fim foi utilizado um banco de dados reais envolvendo a doença Melanoma. Através dos resultados do exemplo foi possível ilustrar a idéia de que a inferência clássica é o caso particular da inferência Bayesiana quando as prioris utilizadas são não-informativas. Segundo Chen *et al* (2001), a análise Bayesiana permite uma maior facilidade dos cálculos e ajuste de modelos de sobrevivência especialmente na presença de esquemas complexos de censura. Além disso, com o uso do Gibbs Sampling e outras técnicas de MCMC o ajuste de modelos de sobrevivência mais complexos se torna mais fácil. Somando-se a isso, os métodos MCMC permitem a realização de inferências exatas para qualquer tamanho de amostra sem restrições dos cálculos assintóticos. Ainda, com o enfoque Bayesiano, podemos incorporar a informação a priori que se tenha sobre os parâmetros do modelo enquanto que na clássica isso não é possível de ser feito.

O uso do software STATA foi útil na análise sob o enfoque clássico. O seu uso é bastante simples, não exigindo conhecimento de qualquer linguagem de programação para efetuar as análises necessárias, apenas supor o modelo ao qual os dados se ajustam melhor.

Sobre o software WinBUGS, esse é bastante prático nas análises bayesianas, porém o usuário deve ter algum conhecimento de linguagem de programação do pacote, pois a modelagem é toda ela realizada via programação do modelo. Porém, os algoritmos dos métodos MCMC estão implementados no pacote. Assim, de posse da linguagem de programação para o modelo, o software realiza as simulações estocásticas necessárias para obter as estimativas a posteriori dos parâmetros.

Este trabalho também serviu para ilustrar que, mesmo modelos simples como o Weibull sem incluir covariáveis, não há forma fechada para a distribuição a

posteriori, sendo necessários métodos alternativos para as análises. Neste sentido, os métodos MCMC, especialmente o Gibbs Sampling, mostraram-se bastante eficientes para derivação dos resultados a posteriori.

Vários assuntos poderiam ter sido explorados mas por razões de tempo não puderam ser vistos nessa monografia. Para citar alguns, a exploração de métodos de escolha das covariáveis a serem incluídas nos modelos bem como estudar os diversos métodos de adequação para modelos, idéias de previsão para futuros tempos de sobrevivência de indivíduos, etc. Ainda, podem ser explorados outros modelos paramétricos além dos vistos neste trabalho. Um assunto bastante interessante de ser explorado seria o modelo de regressão de Cox sob o contexto Bayesiano a fim de estabelecer uma análise comparativa com os métodos clássicos.

Esse trabalho pode ser uma fonte inicial de consulta sobre o assunto de análise de sobrevivência Bayesiana, uma vez que foram apresentados os métodos de estimação dos parâmetros para os modelos Exponencial e Weibull. Para ilustração dos modelos foi apresentado um exemplo com dados reais utilizando os softwares STATA e WinBUGS para as análises, mostrando as etapas dos aplicativos para que os usuários possam consultar e ter um material de auxílio para o uso dos mesmos.

7. BIBLIOGRAFIA

Berger, J.O. (1980). *Statistical Decision Theory and Bayesian Analysis*, second edition. New York: Springer-Verlag.

Chen, M. H.; Ibrahim, J.G.; Sinha, D. (2001). *Bayesian Survival Analysis*, . New York: Springer-Verlag.

Colosimo, E. A. (2000). *Análise de Sobrevivência Aplicada – Notas de aula*.

Cox, D.R.; Oakes, D. (1984). *Analysis of Survival Data*. London: Chapman & Hall.

Gamerman, D. (1997). *Markov Chain Monte Carlo: Stochastic simulation for Bayesian Inference*, first edition. London: Chapman & Hall.

Gelman, A.; Carlin, J.B.; Stern, H.S. and Rubin, D.B. (1997). *Bayesian Data Analysis*, first edition. USA: Chapman & Hall/ CRC.

Gilks, W.R. ; Richardson, S. and Spiegelhalter, D.J. (1997). *Markov Chain Monte Carlo in Practice*. Great Britain: Chapman & Hall.

Gilks, W.R. ; Wild, P. (1992). *Adaptive Rejection sampling for Gibbs Sampling*. Applied Statistics 41, 337-348.

Hastings, W.K. (1970). *Monte Carlo sampling methods using Markov chains and their applications*. Biometrika 57, 97-109.

Hosmer, D.W.; Lemeshow, S. (1999). *Applied Survival Analysis: regression modeling of time to event data*. New York: Wiley & Sons.

Kaplan, E.L.; Meier, P. (1958). *Nonparametric estimation from incomplete observations*. Journal of the American Statistical Association 53, 457-481.

Kalbfleisch, J.D.; Prentice, R.L. (1973). *Marginal likelihood's based on Cox's regression and life model*. Biometrika 60, 267-278.

Klein, J.P.; Moeschberger, M.L. (1997). *Survival Analysis*. New York: Springer-Verlag.

Lawless, J.F. (1982). *Statistical Models and Methods for Lifetime Data*. New York: Wiley and Sons.

Lee, E.T.; Wang, J.W. (2003). *Statistical Methods for Survival Data Analysis*, third edition. USA: Wiley & Sons.

Meyer, P.L. (1975). *Probabilidade: Aplicações à Estatística*. Rio de Janeiro: Livros Técnicos e Científicos S/A .

Storer, B.E.; Wacholder, S.; Breslow, N.E. (1983). *Maximum likelihood fitting of general risk models to stratified data*. Applied Statistics.

Weibull, W. (1951). *A Statistical Distribution of Wide Applicability*. Journal of Applied Mathematics 18, 293-297.