

**ANÁLISE DE SOBREVIVÊNCIA NA
PRESENÇA DE CENSURA
INFORMATIVA: UMA ABORDAGEM
BAYESIANA**

Renata Camila de Souza

Departamento de Estatística - ICEX - UFMG

Junho de 2015

ANÁLISE DE SOBREVIVÊNCIA NA PRESENÇA DE CENSURA INFORMATIVA: UMA ABORDAGEM BAYESIANA

Renata Camila de Souza

Orientador: Prof. Dr. Vinícius Diniz Mayrink

Coorientador: Prof. Dr. Fábio Nogueira Demarqui

Dissertação submetida ao Programa de Pós-Graduação em Estatística da UFMG,
como parte dos requisitos necessários à obtenção do grau de Mestre em Estatística.

Departamento de Estatística
Instituto de Ciências Exatas
Universidade Federal de Minas Gerais

Belo Horizonte, MG - Brasil

Junho de 2015

Aos meus pais.

“O sucesso nasce do querer, da determinação e persistência em se chegar a um objetivo. Mesmo não atingindo o alvo, quem busca e vence obstáculos, no mínimo fará coisas admiráveis” .

José de Alencar

Agradecimentos

Agradeço a Deus por renovar minhas forças todos os dias e iluminar a minha mente para possa superar os obstáculos.

Aos meus pais, pelo amor, apoio e orgulho que demonstram diariamente por mim e por meus irmãos.

Ao meu amor, Normandes, pelo apoio incondicional, por me ajudar a persistir, por ser paciente com minha ausência e por estar presente sempre que foi preciso.

A minha querida irmã Elaine por ser a voz da razão e por acreditar em mim, muitas vezes mais que eu mesma. Ao meu irmão Marcelo, pelos bons ouvidos e por estar sempre disponível para me socorrer. Ao meu amigo Roberto, pelo incentivo e pelo auxílio com os programas. Sua ajuda foi essencial para apresentação desse trabalho. A Viviane que entrou em nossas vidas a pouco mas já tem compartilhado conosco sua generosidade.

Aos meus orientadores Prof. Dr. Vinícius D. Mayrink e Prof. Dr. Fábio N. Demarqui, por partilharem comigo seu conhecimento, serem pacientes ao longo da jornada, pelas críticas e diretrizes e todo apoio que me proporcionaram neste período.

Aos meus colegas de mestrado, Marcela, Rachel, Larissa, Lívia, Maurício, Juliane e Luiza, por dividirem frustrações e alegria ao longo desses dois anos. Vocês tornaram a jornada mais fácil. Em especial a minha querida amiga Luiza, que me incentivou a inscrever-me no mestrado e tem estado bem presente em minha vida, muito obrigada.

Aos professores membros da banca pelas críticas e sugestões que irão aprimorar o trabalho.

A Fundação de Amparo à Pesquisa do Estado de Minas Gerais – FAPEMIG - pelo apoio financeiro.

Resumo

Grande parte dos procedimentos existentes na literatura para a modelagem de dados de sobrevivência baseiam-se na suposição de que o mecanismo de censura é não informativo, ou seja, que os tempos de falha e censura são independentes. Apesar da independência ser uma suposição da modelagem, testar essa condição requer informações adicionais sobre os dados que normalmente não estão disponíveis. Em diversas situações reais, tal suposição não é válida e o mecanismo gerador de censura é dito informativo. Uma alternativa para modelagem de dados sob o mecanismo de censura informativo baseia-se na inclusão de um efeito aleatório, conhecido como fragilidade, que afeta tanto os tempos de falha quanto de censura. Neste trabalho, propomos dois modelos paramétricos: em um modelo o efeito da fragilidade é relacionado aos tempos de censura por meio de um parâmetro e o outro considera uma mistura para modelar o efeito da fragilidade. Estes modelos permitem que o tipo de correlação e, conseqüentemente, o mecanismo de censura do conjunto de observações seja apontado pelos dados. Um estudo simulado, com foco na distribuição Weibull, considerando diferentes proporções de falha, tamanho amostral e o tipo de correlação entre falha e censura foi realizado para comparação do desempenho de algumas especificações de modelo: (i) sem fragilidade, (ii) com fragilidade apenas no componente de falha, (iii) fragilidade em ambos os componentes e afetada por um parâmetro na censura, (iv) fragilidade em ambos componentes mas modelada por uma mistura na censura. Finalmente, realiza-se uma aplicação destes modelos a dados reais.

Palavras-chave: Estatística Bayesiana, Análise de Sobrevivência, Censura Informativa, Modelo de Mistura, Fragilidade.

Abstract

Most procedures found in the literature to model survival data are based on the assumption of non-informative censoring mechanism, meaning that failure and censored times are independent. Although independence is a model assumption, testing this condition requires additional data which are often unavailable. In several real situations this assumption is not valid and the censoring mechanism is thus called informative. An alternative for data modeling under censoring mechanism is based on the inclusion of a random effect known as frailty, which influences both failure and censoring times. In this study, two Bayesian models are proposed: in one of them, the frailty effect is related to the censoring times through a parameter, the other one considers a mixture distribution to model the frailty effect. These models can indicate the type of association and, consequently, the censoring mechanism of the data. A simulated study, focused on the Weibull distribution, considering different failure proportions, sample sizes and correlation type between failure and censoring was developed to compare the performances of some model specifications: (i) without frailty, (ii) with frailty only on the failure component, (iii) frailty in both components and affected by a censoring parameter, (iv) with frailty in both components but affecting the censoring via mixture. Finally, the models are applied to real data.

Keywords: Bayesian Statistics, Survival Analysis, Informative Censoring, Mixture Model, Frailty.

Sumário

1	Introdução	1
2	Conceitos Básicos em Análise de Sobrevida	6
2.1	Modelo de Taxas de Falha Proporcionais	11
2.1.1	Modelo de Regressão Weibull	13
3	Modelos de Fragilidade	15
3.1	Função de Verossimilhança Marginal	17
3.2	Censura Informativa	19
3.2.1	Modelo de Huang e Wolfe	20
4	Modelagem com Mistura	24
4.1	Modelo de Mistura com Três Componentes	25
5	Estudo com Dados Simulados	29
5.1	Especificação de Modelos	32
5.1.1	Modelo M_0	32
5.1.2	Modelo M_1	34
5.1.3	Modelo M_2	35
5.1.4	Modelo M_3	37
5.2	Resultados	38
5.2.1	Resultados para amostras de tamanho 50.	40
5.2.2	Resultados para amostra de tamanho 250.	56

6	Aplicação: Mieloma Múltiplo	68
7	Conclusões	74
7.1	Trabalhos futuros	76
	Apêndice	78

Capítulo 1

Introdução

Em análise de sobrevivência a variável resposta representa o tempo até a ocorrência de um evento de interesse, sendo denominado tempo de falha, como por exemplo, o tempo decorrido entre o surgimento dos primeiros sintomas e o diagnóstico de uma determinada doença, o tempo de vida útil de um equipamento eletrônico ou o tempo até a cura de um paciente.

Dados de sobrevivência, em geral, são caracterizados por tempos com distribuição assimétrica. A maior parte dos tempos observados são pequenos e poucos elementos são acompanhados por um longo período, ou seja, um grande volume de dados concentra-se próximo ao tempo inicial enquanto valores distantes de zero são mais escassos.

Outra característica peculiar de dados de sobrevivência é a presença de observações censuradas, que correspondem à informação parcial sobre o tempo até a ocorrência do evento de interesse. Observações censuradas podem ser causadas por limitação do tempo de acompanhamento ou de recursos, perda do acompanhamento de determinado elemento (indivíduo ou componente), maneira como o experimento foi planejado, entre outros.

A censura pode ser classificada como (Colosimo e Giolo, 2006):

- (a) censura à direita: o tempo até a ocorrência do evento de interesse está à direita do tempo observado;
 - ◇ Ex.: Em um estudo sobre o tempo entre o diagnóstico até a cura de determinada doença por um tempo determinado, alguns indivíduos podem chegar ao

fim do estudo sem vivenciar o evento de interesse.

(b) censura à esquerda: o tempo até a ocorrência do evento está à esquerda do tempo observado;

◇ Ex.: Ao se realizar uma pesquisa sobre o tempo de alfabetização de crianças, é possível que na amostra selecionada já existam algumas crianças alfabetizadas no início do estudo, não sendo possível determinar quando isso ocorreu;

(c) censura intervalar: a falha ocorre em um intervalo de tempo, mas não é possível especificar o momento exato de sua ocorrência.

◇ Ex.: Em alguns casos, o evento de interesse pode ser o tempo entre os primeiros sintomas e o diagnóstico de determinada doença. Entretanto, o diagnóstico é dado de acordo com um resultado de exame feito mensalmente, assim não se conhece o tempo exato da ocorrência da doença, apenas o intervalo entre os resultados de exame que indicaram a existência da doença.

Nesta dissertação iremos focar na análise de dados com censura à direita, visto que esse tipo de censura é bastante comum em situações reais.

O mecanismo gerador de censura refere-se à relação entre o tempo até o evento de interesse e o tempo até a observação de uma censura. Ele tem papel importante na formulação de modelos em análise de sobrevivência. A suposição largamente usada é de que o mecanismo de censura é não informativo, ou seja, que as observações de falha e censura são independentes. Segundo Staplin (2012), isto significa que observações censuradas – no caso de censura à direita – irão fornecer somente a informação de que o tempo de sobrevivência verdadeiro excede o tempo observado e que não há qualquer outra informação sobre o tempo de sobrevivência subsequente. Esta suposição é bastante conveniente no sentido de que simplifica a função de verossimilhança do modelo. Entretanto, ela é adequada apenas para alguns casos, por exemplo, quando os indivíduos deixam de ser acompanhados devido ao término do estudo ou em casos em que eles são

retirados do estudo por experimentarem eventos não relacionados ao evento de interesse, tais como mudança de localidade, acidente de carro, etc.

Em algumas situações reais, é possível perceber que existe algum tipo de dependência entre os tempos de falha e censura, caracterizando, de acordo com Staplin (2012), o mecanismo de censura informativo. Segundo Link (1989), pode-se supor que o mecanismo de censura é informativo, por exemplo, quando um indivíduo tem o fim de seu acompanhamento ao abandonar o estudo por razões relacionadas ao tipo de terapia. Outro exemplo, são situações em que a falha é causada por eventos secundários e assim registrada como observação censurada.

A preocupação com a modelagem de dados de sobrevivência na presença de censura informativa não é recente. A literatura que trata desse assunto tem aumentado em volume e evoluído ao longo do tempo. Link (1989) propôs uma modelagem para dados sob o mecanismo informativo através de uma modificação no estimador de Kaplan Meier. Leung et al. (1997) apontam alguns problemas de modelagem caso a suposição de independência seja violada. Scharfstein e Robins (2002) propõem o uso de uma função associada ao modelo para determinar a dependência da censura, fazendo algumas suposições sobre ela. Um modelo de cópula para dados de censura informativa é proposto por Jiang et al. (2005). Lu e Zhang (2012) propõem um modelo de taxas de falha proporcionais que compartilha parâmetros de regressão na distribuição de falha e censura para modelagem de dados sob o mecanismo de censura informativo.

Entre as diversas propostas usadas na literatura para a modelagem de dados com alguma correlação, pode-se citar o uso de modelos de fragilidade, como apresentado por Link (1989), Kosorok et al. (2002), entre outros. O termo fragilidade foi introduzido em análise de sobrevivência por Vaupel et al. (1979) para explicar como indivíduos com mesmas características observáveis e conhecidas podiam possuir riscos diferentes de experimentar o evento de interesse. Desde então, seu uso tem crescido e se diversificado. Segundo Hanagal (2011) a fragilidade representa um efeito aleatório não observado que é compartilhado por indivíduos que possuem características comuns não mensuráveis. Sua principal aplicação acontece em situações nas quais se deseja fazer inferência sobre tempos de sobrevivência com algum tipo de relação, por exemplo, quando se deseja analisar in-

divíduos da mesma família, ou grupos de indivíduos em diferentes centros de tratamento. Colosimo e Giolo (2006) e Collett (2003) apresentam um estudo introdutório sobre o uso da fragilidade em análise de sobrevivência, considerando diversas aplicações. Uma discussão mais detalhada sobre modelos de fragilidade pode ser encontrada em Hanagal (2011).

Segundo Siannis (2004) mensurar a dependência entre falha e censura é uma tarefa difícil. Dessa forma, é necessário inserir na modelagem elementos que possam captar fontes da variação não mensuráveis, como é o caso da fragilidade. Huang e Wolfe (2002), por exemplo, propõem um modelo de fragilidade semiparamétrico que avalia o grau da censura informativa e ajusta seu modelo a dados relacionados espacialmente por meio do algoritmo EM. Santos (2012) utiliza uma ideia semelhante ajustando um modelo paramétrico para dados de ocorrência de infecção em pacientes que realizaram transplante de medula óssea em um único centro, utilizando a distribuição *Gama* para a modelagem da fragilidade, o que minimiza as dificuldades algébricas e torna o algoritmo EM desnecessário. Em ambos os trabalhos, os tempos de falha e censura são considerados independentes quando condicionados ao termo de fragilidade.

Nesta dissertação, modelos de fragilidade também serão utilizados para captar a dependência entre falha e censura em duas propostas. Na primeira, o efeito aleatório será adicionado ao modelo através de um parâmetro associado aos tempos de censura, de forma captar a correlação. Na segunda, a fragilidade será incorporada à estrutura do modelo por meio de uma mistura com três componentes degenerados que caracterizam a relação existente entre os tempos de falha e censura. A estrutura hierárquica é construída de forma que a mistura identifique o mecanismo de censura presente nos dados, tornando a suposição inicial de independência entre falha e censura dispensável. Este tipo de modelagem envolvendo misturas originou-se no contexto Bayesiano de seleção de variáveis em regressão e tem sido explorado em diversas pesquisas; ver George e McCulloch (1997) e Geweke (1996).

A inferência sobre os modelos é realizada através de uma abordagem Bayesiana. A análise de dados de sobrevivência sob o mecanismo de censura não informativo com a inclusão de uma fragilidade é apresentada por Ibrahim et al. (2010), porém trabalhos que

utilizam essa abordagem para tratamento de dados de censura informativa são escassos quando comparados ao mecanismo não informativo. Uma das vantagens do uso da inferência Bayesiana está na inclusão do componente de fragilidade de forma mais natural na modelagem, visto que a incerteza sobre qualquer objeto (variáveis aleatórias, variáveis latentes e parâmetros) é descrita através de uma distribuição de probabilidade.

O foco deste trabalho é desenvolver uma abordagem Bayesiana para a análise de dados de sobrevivência na presença de censura informativa, utilizando a suposição de que os tempos de falha e censura são condicionalmente independentes dado uma fragilidade. Utilizaremos a distribuição Weibull para modelar os tempos observados. Esta é a distribuição mais utilizada para este contexto devido à grande flexibilidade de sua função de taxa de falha (Lawless, 2003). Apresentaremos um estudo considerando o comportamento de diferentes modelos ajustados a dados simulados com três proporções de falha (pequena, média e alta). Iremos avaliar cenários com dados sob o mecanismo não informativo (independência entre os tempos de falha e censura) e cenários sob o mecanismo informativo (um deles com correlação positiva entre falha e censura, outro com correlação negativa).

Esta dissertação está organizada da seguinte maneira: o Capítulo 2 descreve os fundamentos de análise de sobrevivência e alguns elementos específicos como modelos de taxas de falha proporcionais. O Capítulo 3 trata da fragilidade e apresenta o modelo de Huang e Wolfe (2002) para modelagem de dados de sobrevivência sob mecanismo informativo. Um proposta para a modelagem de misturas é apresentada no Capítulo 4, sendo a mistura de três componentes tratada como caso particular. O Capítulo 5 descreve o estudo dos cenários simulados e analisa o ajuste dos diferentes modelos. Uma aplicação real é desenvolvida no Capítulo 6 explorando todos os modelos considerados no Capítulo 5, enquanto o Capítulo 7 trará um resumo de toda discussão abordada nesta dissertação; destaque será dado às principais conclusões tiradas nos estudos. O encerramento é feito com algumas propostas de trabalhos futuros.

Capítulo 2

Conceitos Básicos em Análise de Sobrevivência

Segundo Collett (2003), o termo análise de sobrevivência é usado para descrever dados representando períodos de tempo transcorridos entre um tempo inicial e a ocorrência de um particular evento. Hanagal (2011) define dados de sobrevivência como a medida de tempo até a ocorrência de um certo evento. De maneira geral, podemos dizer que a análise de sobrevivência é o acompanhamento de elementos (indivíduos, pacientes, equipamentos, etc) por um período, tendo como principal objetivo analisar o tempo até a ocorrência de um evento de interesse, usualmente denominado tempo de falha. Este ramo da Estatística pode ser aplicado, por exemplo, em situações na quais se deseja avaliar a reincidência de uma doença, tempo até a ocorrência de um crime em uma determinada região, vida útil de equipamentos eletrônicos, entre outros.

O tempo até a falha pode ser descrito como uma variável aleatória T , não negativa, em geral com distribuição assimétrica, que representa o tempo de sobrevivência de indivíduos em alguma população. Essa variável é usualmente especificada pela função de sobrevivência ou pela função de taxa de falha, sendo estas relacionadas entre si. A função de sobrevivência $S(t)$ descreve a probabilidade de um indivíduo não experimentar o evento de interesse até o instante t , sendo dada por:

$$S(t) = 1 - F(t) = P(T > t) = \int_t^{\infty} f(u)du, \quad (2.1)$$

em que $F(t)$ e $f(t)$ representam a função de distribuição acumulada e a função densidade de probabilidade, respectivamente. Nota-se que $S(t)$ é uma função monótona não crescente com $S(0) = 1$ e $S(\infty) = \lim_{t \rightarrow \infty} S(t) = 0$.

A função de taxa de falha é um importante conceito em análise de tempos de vida. Segundo Collett (2003), esta função é amplamente usada para expressar o risco do indivíduo experimentar o evento de interesse em determinado instante t , sendo obtida a partir da probabilidade da falha ocorrer no tempo t dado que o indivíduo sobreviveu até este tempo. Sua forma indica o comportamento do evento de interesse em indivíduos da população, ou seja, se ao longo do tempo a probabilidade do indivíduo sofrer a falha aumenta, diminui ou é constante. A função de taxa de falha é dada por:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}. \quad (2.2)$$

As funções $h(t)$, $f(t)$ e $S(t)$ são matematicamente equivalentes. Podemos determinar a expressão de $f(t)$ a partir de $S(t)$ e vice-versa.

$$f(t) = \frac{d}{dt}F(t) = \frac{d}{dt}[1 - S(t)] = -\frac{d}{dt}S(t).$$

Pode-se também substituir a expressão acima em (2.2), obtendo:

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

Ao integrarmos $h(t)$, apresentado acima, no intervalo $[0, t]$, teremos a seguinte igualdade:

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

Finalmente, é possível concluir que:

$$f(t) = h(t) \exp\{-H(t)\}.$$

Uma revisão sobre essas funções e suas relações pode ser encontrada em Collett (2003), Lawless (2003) e Colosimo e Giolo (2006), entre outros.

Dados de sobrevivência possuem uma importante característica: a presença de observações censuradas e, portanto, devem receber tratamento especial. Uma observação censurada é uma observação parcial do evento de interesse. Dessa forma, a censura não deve ser ignorada durante a análise dos dados, sob risco de levar a estimativas viesadas dos parâmetros, (Colosimo e Giolo, 2006). Observações censuradas costumam ocorrer pela saída de alguns indivíduos do estudo, perda de contato durante o acompanhamento, término do estudo sem ocorrência do evento de interesse, entre outros motivos.

Lawless (2003) afirma que para aplicar procedimentos estatísticos baseados nos dados censurados é necessário considerar no processo todos os tempos observados, sejam eles falha ou censura. Uma representação simples e amplamente utilizada para o ajuste de dados com presença de censura é composta por um par de variáveis, sendo: Y_i representando o tempo observado para o indivíduo i e δ_i uma indicadora de falha ou censura daquela observação. Considere T uma variável aleatória contínua e não negativa que representa o tempo de falha de um indivíduo. De forma análoga, podemos definir C como outra variável aleatória que representa o tempo de censura associado a um indivíduo. A notação geral, usada por Lawless (2003) e Colosimo e Giolo (2006), é:

$$Y_i = \min\{T_i, C_i\}$$

$$\delta_i = \begin{cases} 1, & T_i \leq C_i \\ 0, & T_i > C_i. \end{cases} \quad (2.3)$$

A variável aleatória Y é caracterizada, em análise de sobrevivência, pela função de sobrevivência $S(y)$, apresentada em (2.1), ou pela função de taxa de falha $h(y)$ descrita em (2.2). Essas funções podem ser modeladas de forma paramétrica ou não paramétrica. Neste trabalho, optou-se pelo uso de técnicas paramétricas em que devemos assumir uma distribuição de probabilidade ao conjunto de observações. Esta distribuição deve ter características que se adequam ao conjunto de dados. Entre as distribuições am-

plamente usadas são: exponencial, Weibull, log-normal, logística; detalhes em Collett (2003). Neste texto, o foco será a distribuição Weibull, amplamente usada na literatura, cujas características serão discutidas com mais detalhes na Subseção 3.1.1.

De acordo com Collett (2003) o uso de modelos paramétricos pode levar a inferências mais precisas, desde que as suposições do modelo sejam satisfeitas. Os modelos paramétricos destacam-se pela necessidade de estimação de quantidades desconhecidas (parâmetros) que determinam a forma e o comportamento das distribuições. A estimação dos parâmetros pode ser realizada por meio de duas abordagens distintas: Frequentista ou Bayesiana. Em ambas é fundamental a construção da função de verossimilhança, que contém as informações dos dados para a estimação dos parâmetros.

Para a construção da função de verossimilhança é necessário apresentar primeiramente a distribuição conjunta para as variáveis δ_i e Y_i . Esta distribuição descreve a relação da variável “tempo de falha” com a variável “tempo de censura”. O modelo geralmente proposto para descrever a distribuição de Y_i para as observações de censura ($\delta_i = 0$) é dado pela seguinte expressão, Lawless (2003):

$$P(y_i, \delta_i = 0) = P(C \in (y_i, y_i + \Delta y_i], T > y_i). \quad (2.4)$$

Similarmente, para as observações de falha temos:

$$P(y_i, \delta_i = 1) = P(T \in (y_i, y_i + \Delta y_i], C > y_i). \quad (2.5)$$

As distribuições dadas em (2.4) e (2.5) são usadas para descrever a função de verossimilhança para uma amostra de n observações, em que r são falhas e $(n - r)$ são censuras. Denote $D = \{(y_i, \delta_i), i = 1, \dots, n\}$ e seja $\Theta = \{\Theta_T, \Theta_C\}$, em que Θ_T e Θ_C são vetores de parâmetros associados à distribuição dos tempos de falha e censura, respectivamente. Então a função de verossimilhança é dada por:

$$\begin{aligned} L(\Theta|D) &= \prod_{i=1}^r \{P(T \in (y_i, y_i + \Delta y_i], C > y_i|\Theta)\} \prod_{i=r+1}^n \{P(C \in (y_i, y_i + \Delta y_i], T > y_i|\Theta)\} \\ &= \prod_{i=1}^n \{P(T \in (y_i, y_i + \Delta y_i], C > y_i|\Theta)\}^{\delta_i} \{P(C \in (y_i, y_i + \Delta y_i], T > y_i|\Theta)\}^{1-\delta_i}. \end{aligned}$$

Assumindo independência entre T e C , a função de verossimilhança pode ser reescrita como produto simplificando as operações. Se definirmos $f^{(T)}(y_i|\Theta_T)$ e $f^{(C)}(y_i|\Theta_C)$ como funções densidade, $S^{(T)}(y_i|\Theta_T)$ e $S^{(C)}(y_i|\Theta_C)$ como função de sobrevivência, $h^{(T)}(y_i|\Theta_T)$ e $h^{(C)}(y_i|\Theta_C)$ como função de taxa de falha para as observações de falha e censura, respectivamente, a função de verossimilhança passa a ser escrita como:

$$\begin{aligned}
L(\Theta|D) &= \prod_{i=1}^n P(T \in (y_i, y_i + \Delta y_i]|\Theta_T)^{\delta_i} P(C > y_i|\Theta_T)^{\delta_i} \\
&\times P(C \in (y_i, y_i + \Delta y_i]|\Theta_C)^{1-\delta_i} P(T > y_i|\Theta_C)^{1-\delta_i} \\
&= \prod_{i=1}^n [f^{(T)}(y_i|\Theta_T)^{\delta_i} S^{(T)}(y_i|\Theta_T)^{1-\delta_i}] [f^{(C)}(y_i|\Theta_C)^{1-\delta_i} S^{(C)}(y_i|\Theta_C)^{\delta_i}] \\
&= \prod_{i=1}^n [h^{(T)}(y_i|\Theta_T)^{\delta_i} S^{(T)}(y_i|\Theta_T)] [h^{(C)}(y_i|\Theta_C)^{1-\delta_i} S^{(C)}(y_i|\Theta_C)].
\end{aligned}$$

Quando existe independência entre os tempos de falha e censura (T e C) dizemos que a censura é não informativa, ou seja, ela não traz informação sobre os parâmetros da distribuição de falha (Θ_T) e pode ser suprimida da equação de forma que:

$$L(\Theta_T|D) \propto \prod_{i=1}^n h^{(T)}(y_i|\Theta_T)^{\delta_i} S^{(T)}(y_i|\Theta_T).$$

A suposição de independência desempenha papel importante na análise de dados de sobrevivência, visto que é muito difícil determinar a distribuição conjunta dos tempos de falha e censura. Além disso, a independência entre os tempos traz simplificação algébrica que auxilia na inferência sobre o modelo. Entretanto, Leung et al. (1997) afirmam que não é possível testar a suposição de independência sem fazer restrições adicionais, como por exemplo restrições sobre a distribuição conjunta de falha e censura. Assim, na maioria das vezes, a suposição de independência é adotada sem que qualquer teste seja realizado.

Segundo Wu e Carroll (1988) a modelagem de dados de censura informativa por modelos que utilizam a suposição de independência pode causar um pequeno viés nas estimativas. Staplin (2012), também alerta sobre o de desconsiderar a relação entre falha e censura levando à superestimação da função de sobrevivência caso a correlação seja

positiva e, por outro lado, uma correlação negativa entre tempo até o evento e o tempo de censura poderia levar a uma subestimação da função de sobrevivência.

Estudos relacionados a dados de sobrevivência, muitas vezes, envolvem covariáveis que podem estar relacionadas com o tempo de sobrevivência dos indivíduos. Uma forma adequada para incluir o efeito dessas covariáveis na modelagem é utilizar o modelo de taxas de falha proporcionais.

2.1 Modelo de Taxas de Falha Proporcionais

Na área médica, por exemplo, pode-se imaginar que alguns fatores (covariáveis) influenciam no surgimento de uma doença (idade, sexo, peso) ou que fazem parte de um diagnóstico (exame laboratorial, exame físico). Uma maneira extensivamente usada para incluir covariáveis no modelo é proposta por Cox (1972). Segundo Colosimo e Giolo (2006), o modelo de Cox permite a análise de dados provenientes de estudos de tempo de vida considerando covariáveis. Um estudo sobre este modelo e suas propriedades assintóticas é desenvolvido em Tsiatis (1981).

O modelo representa uma forma de inflação/deflação do risco do indivíduo. Neste modelo as covariáveis são acrescentadas a função de taxa de falha de forma multiplicativa. Considere k covariáveis de modo que $\mathbf{x}_i = (x_{i1}, \dots, x_{ik})$. A expressão do modelo de regressão de Cox é dada por:

$$h^{(T)}(y_i) = h_0(y_i)^{(T)} \exp\{\mathbf{x}_i \boldsymbol{\beta}_T\}. \quad (2.6)$$

Na expressão acima, o termo $h_0^{(T)}$ representa a função de taxa de falha base, e $\boldsymbol{\beta} = \{\beta_{T,1}, \dots, \beta_{T,k}\}$ é o vetor de coeficientes de regressão. Observe que tanto $\exp\{\mathbf{x}_i \boldsymbol{\beta}_T\}$ quanto $h_0^{(T)}$ são funções não negativas. Além disso, a função de taxa de falha base $h_0^{(T)}$ pode ser modelada parametricamente ou não. O "modelo de Cox" é usualmente empregado quando modelamos $h_0^{(T)}$ semi-parametricamente.

O modelo de taxas de falha proporcionais tem a propriedade de garantir a proporcionalidade do risco entre dois indivíduos como constante ao longo do tempo. Isso pode ser facilmente observado na seguinte equação:

$$\frac{h^{(T)}(y|\mathbf{x}_i)}{h^{(T)}(y|\mathbf{x}_j)} = \frac{h_0^{(T)}(y) \exp\{\mathbf{x}_i\boldsymbol{\beta}_T\}}{h_0^{(T)}(y) \exp\{\mathbf{x}_j\boldsymbol{\beta}_T\}} = \exp\{\mathbf{x}_i\boldsymbol{\beta}_T - \mathbf{x}_j\boldsymbol{\beta}_T\} = \exp\{\boldsymbol{\beta}_T(\mathbf{x}_i - \mathbf{x}_j)\}.$$

Por exemplo, se o risco de um indivíduo ser diagnosticado com determinada doença é três vezes maior que o risco de outro indivíduo, essa razão permanecerá a mesma durante todo o período estudado. A suposição básica para o uso do modelo de taxas de falha proporcionais, portanto, é que o risco entre indivíduos seja constante ao longo do tempo. Esse modelo é caracterizado pelos coeficientes β 's, que medem os efeitos das covariáveis sob a função de taxa de falha. Para escrever a verossimilhança deste modelo (para o mecanismo de censura não informativo), faremos uso da identidade:

$$S(t) = \exp\{-H(t)\}, \quad (2.7)$$

considerando a nova função de taxa de falha apresentada acima. Neste caso, $D = \{(y_i, \delta_i, x_i), i = 1, \dots, n\}$ representa os dados observados, $\boldsymbol{\Theta}_T = \{\theta_T, \boldsymbol{\beta}_T\}$ e a função de verossimilhança passa a ser escrita da seguinte forma:

$$L(\boldsymbol{\Theta}_T|D) \propto \prod_{i=1}^n h_0^{(T)}(y_i|\boldsymbol{\theta}_T) \exp\{\mathbf{x}_i\boldsymbol{\beta}_T\}^{\delta_i} \exp\{H_0^{(T)}(y_i|\boldsymbol{\theta}_T) \exp\{\mathbf{x}_i\boldsymbol{\beta}_T\}\}. \quad (2.8)$$

Uma das vantagens em utilizar o modelo de taxas de falha proporcionais é que as propriedades de log-concavidade da distribuição são mantidas, sendo útil para amostrar as componentes do vetor $\boldsymbol{\beta}_T$. Segundo Dellaportas e Smith (1993), na análise Bayesiana, se a distribuição *a priori* e a função de verossimilhança são log-côncavas então a distribuição condicional completa *a posteriori* também será. Essa propriedade permite o uso de técnicas de simulação do tipo envelope como o algoritmo *Adaptive Rejection Sampling* (ARS) – ver Gamerman e Lopes (2006). Para mais informações sobre a log-concavidade ver Dellaportas e Smith (1993) e Gilks e Wild (1992).

As covariáveis tem o efeito de acelerar ou desacelerar o tempo até a falha. Entretanto, podem existir fatores não mensuráveis ou desconhecidos que também tenham o poder de influenciar o tempo de sobrevivência de indivíduos em estudo. Estes casos são normal-

mente modelados com a inclusão de um componente de fragilidade, situação abordada no Capítulo 4.

2.1.1 Modelo de Regressão Weibull

A utilização do modelo de Cox para a modelagem de dados de sobrevivência não exige que alguma distribuição de probabilidade em particular seja assumida. Como resultado, a função de taxa de falha não é restrita a uma forma específica, tornando o modelo de Cox bastante flexível. Por outro lado, se as suposições sobre determinada distribuição de probabilidade para os dados é válida, inferências baseadas em tais suposições podem ser mais precisas. Em particular, estimativas de quantidades relativas a médias e medianas dos tempos de sobrevivência tendem a ter menores erros padrão [Collett (2003)]. Modelos em que uma distribuição de probabilidade é assumida para os tempos de sobrevivência são conhecidos como modelos paramétricos.

Segundo Lawless (2003) a distribuição Weibull é a distribuição mais utilizada para a modelagem de tempos de vida. Sua grande aplicação a situações reais deve-se, principalmente, á grande variedade de formas que sua função de taxa de falha pode apresentar. Independente da forma assumida, uma característica básica desta distribuição é possuir uma função de taxa de falha monótona, sendo crescente para valores de $\alpha > 1$, decrescente para $\alpha < 1$ e constante quando temos $\alpha = 1$. Algumas formas da função de taxa de falha são apresentadas na Figura 2.1.

Para uma variável aleatória T com distribuição Weibull, tem-se que a função densidade de probabilidade é dada por:

$$f(t|\alpha, \lambda) = \alpha \lambda t^{\alpha-1} \exp\{-\lambda t^\alpha\}$$

em que $\alpha > 0$ é o parâmetro de forma e $\lambda > 0$ é o parâmetro de escala. Para o caso particular em que $\alpha = 1$ temos a distribuição exponencial.

As função de sobrevivência $S(t)$ e de taxa de falha $h(t)$ podem ser facilmente obtidas por meio das identidades matemáticas apresentadas anteriormente.

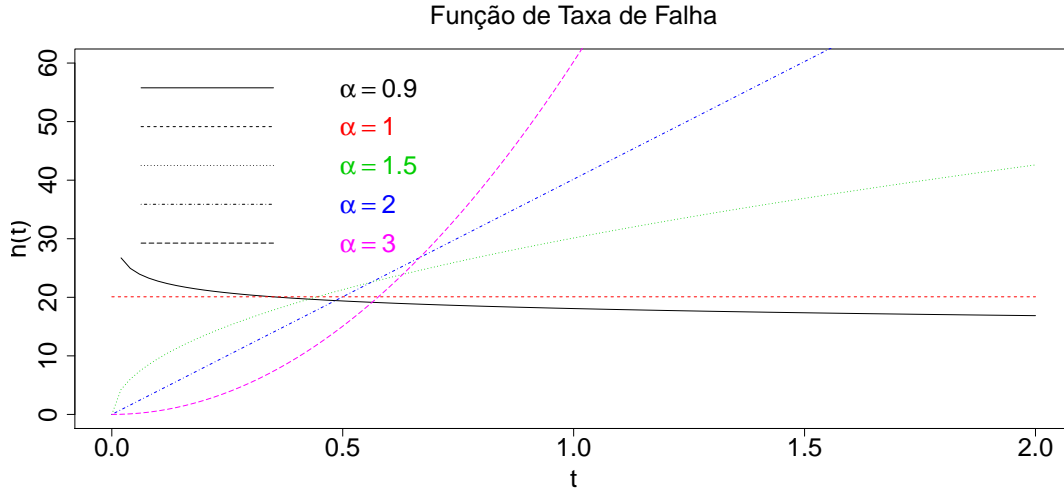


Figura 2.1: *Diferentes formas da distribuição Weibull*

O modelo de taxas de falha proporcionais consiste em definir a função de taxa de falha base $h_0(t)$ como uma função modelada por esta distribuição. Assim, as funções de taxa de falha e sobrevivência passam a ser descritas da seguinte forma:

$$h(t|\alpha, \lambda) = \alpha \lambda t^{\alpha-1} \exp\{\mathbf{x}_i \boldsymbol{\beta}\}$$

$$S(t|\alpha, \lambda) = \exp\{-\lambda t^\alpha \exp\{\mathbf{x}_i \boldsymbol{\beta}\}\}$$

Note que o tempo de sobrevivência do i -ésimo indivíduo tem distribuição Weibull com o parâmetro de escala $\lambda \exp\{\mathbf{x}_i \boldsymbol{\beta}\}$ e o parâmetro de forma α . Este resultado mostra que o efeito das variáveis explicativas alteram apenas o parâmetro de escala, enquanto o parâmetro de forma permanece constante. Uma representação comum é aplicar a transformação $\lambda = \exp\{\beta_0\}$ e definir a distribuição Weibull(α, λ_i) em que $\lambda_i = \exp\{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k\}$, em que k é o número de covariáveis conhecidas da amostra. Essa estrutura mantém a propriedade de riscos proporcionais do modelo de Cox e garante a identificabilidade dos parâmetros do modelo, (Collett, 2003).

Capítulo 3

Modelos de Fragilidade

Segundo Rocha (2011), uma abordagem comum na análise de dados de sobrevivência é assumir uma população homogênea de indivíduos com a mesma estrutura de covariáveis. No entanto, é evidente que indivíduos idênticos em muitos aspectos, tais como a idade, sexo e mesmo tratamento, diferem devido a fatores desconhecidos ou não mensuráveis. A necessidade de incluir na modelagem fontes de heterogeneidade não observadas levou ao surgimento dos chamados modelos de fragilidade. O termo fragilidade, introduzido por Vaupel et al. (1979), é usado para representar um efeito aleatório não observável inserido na modelagem para captar fatores de riscos (não mensuráveis) para um indivíduo.

Pode-se dizer que um indivíduo é frágil quando ele está mais (ou menos) suscetível a experimentar o evento de interesse, em relação aos outros. Hanagal (2011) afirma que modelos de fragilidade são utilizados para tentar explicar a existência de atributos não mensuráveis, como o genótipo ou outras características individuais, que acabam por introduzir heterogeneidade em uma população de estudo.

Modelos de fragilidade também possuem grande aplicação a dados de sobrevivência multivariados. Para Hanagal (2011), esses modelos são úteis para captar a correlação na modelagem de dados de sobrevivência multivariados (subgrupos e eventos recorrentes), quando alguma característica é compartilhada por mais de um indivíduo ou quando temos vários tempos medidos para um mesmo indivíduo (tempos do mesmo indivíduo formam um grupo). Nestes casos, o efeito das características não mensuráveis, compartilhado pelos elementos de um determinado grupo pode ser incluído na modelagem através da

adição do componente de fragilidade à função de taxa de falha:

$$h^{(T)}(y_{i,j}|x_{i,j}, U_j) = h_0^{(T)}(y_{i,j})U_j \exp\{\mathbf{x}_{i,j}\boldsymbol{\beta}_T\};$$

em que $h_0^{(T)}(y_{i,j})$ é a função de taxa de falha base, $\mathbf{x}_{i,j}$ é o vetor de covariáveis associado ao i -ésimo indivíduo do j -ésimo grupo, $\boldsymbol{\beta}$ são os coeficientes de regressão e U_j é a fragilidade (não observada) associada ao j -ésimo grupo.

O modelo de fragilidade compartilhada pode ter vários níveis dependendo do número de subgrupos da amostra, como por exemplo num estudo composto por grupos de pacientes de diferentes famílias ou diferentes centros de tratamento ou, ainda, pesquisas com casais; ver Collett (2003), Ibrahim et al. (2010) e Hanagal (2011), entre outros.

Neste estudo, nosso objetivo é modelar fatores individuais não observáveis que justificam o fato de que indivíduos com características conhecidas e iguais apresentam tempos de sobrevivência diferentes. Observe que, neste caso, a fragilidade deve ser incluída para cada indivíduo de forma que cada observação é considerada um grupo diferente. Hanagal (2011) apresenta o seguinte modelo:

$$\begin{aligned} h^{(T)}(y_i) &= h_0^{(T)}(y_i) \exp\{\mathbf{x}_i\boldsymbol{\beta}_T\}U_i \\ &= h_0^{(T)}(y_i) \exp\{\mathbf{x}_i\boldsymbol{\beta}_T + w_i\}, \end{aligned} \quad (3.1)$$

em que $w_i = \log(U_i)$. Para garantir a identificabilidade do modelo, temos que $E(U_i) = 1$ ou $E(w_i) = 0$ (Rocha, 2011).

A introdução do componente de fragilidade no contexto univariado busca modelar a heterogeneidade dos indivíduos, bem como avaliar o efeito de covariáveis ignoradas na modelagem de forma que, se alguma covariável importante não for acrescentada, a fragilidade do modelo cresce afetando as inferências feitas sobre as demais. Logo, podemos dizer que a fragilidade tem também o papel de modelar a heterogeneidade não medida dos dados, (Hanagal, 2011). Entretanto, neste contexto, o modelo perde a interpretação de riscos proporcionais apresentada em (2.7), ver Colosimo e Giolo (2006) e Rocha (2011). De fato, o risco relativo depende agora tanto da diferença entre as covariáveis quanto da fragilidade, tornando impossível identificar qual fator influencia o risco dos indivíduos, como apresentado abaixo:

$$\frac{h^{(T)}(y|\mathbf{x}_i)}{h^{(T)}(y|\mathbf{x}_j)} = \frac{h_0^{(T)}(y) \exp\{\mathbf{x}_i\boldsymbol{\beta}_T\}U_i}{h_0^{(T)}(y) \exp\{\mathbf{x}_j\boldsymbol{\beta}_T\}U_j} = \frac{U_i}{U_j} \exp\{(\mathbf{x}_i - \mathbf{x}_j)\boldsymbol{\beta}_T\}.$$

Seja D_{obs} o conjunto de dados observados, U o conjunto de dados não observados e $D_{comp.} = \{D_{obs}, U\}$ o conjunto de dados completos. A verossimilhança (condicionada à fragilidade) sob a suposição de falha e censura independentes é dada por:

$$\begin{aligned} L(\boldsymbol{\Theta}_T|D_{comp.}) &\propto \prod_{i=1}^n (h_0^{(T)}(y_i|\boldsymbol{\theta}_T) \exp\{\mathbf{x}_i\boldsymbol{\beta}_T\}U_i)^{\delta_i} \\ &\times \exp\{H_0^{(T)}(y_i|\boldsymbol{\theta}_T) \exp\{\mathbf{x}_i\boldsymbol{\beta}_T\}U_i\}. \end{aligned} \quad (3.2)$$

Existem diversas distribuições na literatura indicadas para modelar a fragilidade. Hanagal (2011) apresenta algumas sugestões, como a gama, log-normal e a normal inversa. Dentre estas, a gama é amplamente usada devido à sua conveniência algébrica.

3.1 Função de Verossimilhança Marginal

Sob a abordagem Frequentista, o cálculo da função de verossimilhança requer que as variáveis aleatórias não observáveis sejam integradas para obtenção da função de verossimilhança marginal na qual se baseiam as análises dos dados. A função de verossimilhança marginal condicionada é dada por:

$$L(\boldsymbol{\Theta}|D_{obs.}) \propto \prod_{i=1}^n \int_U h^{(T)}(y_i|\boldsymbol{\theta}_T, U)^{\delta_i} S^{(T)}(y_i|\boldsymbol{\theta}_T, U) g(U|\psi) dU, \quad (3.3)$$

em que $D_{obs.} = \{y_i, \delta_i, \mathbf{x}_i\}$ é o vetor de dados observados, $\boldsymbol{\Theta} = \{\boldsymbol{\Theta}_T, \psi\}$, $\boldsymbol{\Theta}_T = \{\boldsymbol{\theta}_T, \boldsymbol{\beta}_T\}$ e ψ é um vetor de parâmetros associado a distribuição que modela a fragilidade.

A solução desta integral, entretanto, pode ser bastante complicada, uma vez que, em geral, não é possível obter uma expressão fechada. Neste sentido, a distribuição gama tem

vido bastante utilizada, pois, segundo Hanagal (2011), é a única distribuição que permite a obtenção de formas fechadas das funções de taxa de falha e sobrevivência marginais.

Assim, supondo que $U \sim Ga(\kappa, \vartheta)$, cuja densidade é dada por $f(U) = (\vartheta^\kappa U^{\kappa-1} e^{-\vartheta U})/\Gamma(\kappa)$, a função de verossimilhança marginal condicionada apenas nos dados observados pode ser obtida. Note que teremos uma forma fechada sempre que as funções $h_0(y)$ e $H_0(y)$ (para tempos de falha e censura) também apresentarem forma fechada. A verossimilhança obtida é:

$$\begin{aligned} L(\Theta|D_{obs.}) &= \prod_{i=1}^n \int_U \left[h_0^{(T)}(y_i|\Theta_T) e^{\mathbf{x}_i \boldsymbol{\beta}_T U_i} \right]^{\delta_i} \exp \left\{ -H_0^{(T)}(y_i|\Theta_T) e^{\mathbf{x}_i \boldsymbol{\beta}_T U_i} \right\} \\ &\times \vartheta^\kappa U_i^{\kappa-1} \exp \{-\vartheta U_i\} / \Gamma(\kappa) dU_i \\ &= \prod_{i=1}^n h_0^{(T)}(y_i|\Theta_T)^{\delta_i} e^{\delta_i \mathbf{x}_i \boldsymbol{\beta}_T} \vartheta^\kappa / \Gamma(\kappa) \\ &\times \int_U U_i^{\delta_i + \kappa - 1} \exp \left\{ -U_i \left(H_0^{(T)}(y_i|\Theta_T) e^{\mathbf{x}_i \boldsymbol{\beta}_T} + \vartheta \right) \right\} dU_i. \end{aligned}$$

O núcleo dentro da integral acima é de uma distribuição gama $[(\delta_i + \kappa), (\vartheta + H_0^{(T)}(y_i|\Theta_T) e^{\mathbf{x}_i \boldsymbol{\beta}_T})]$.

Assim, a função de verossimilhança marginal tem a seguinte forma:

$$L(\Theta|D_{obs.}) = \prod_{i=1}^n \left\{ \frac{\left(h_0^{(T)}(y_i|\Theta_T) e^{\mathbf{x}_i \boldsymbol{\beta}_T} \right)^{\delta_i} \vartheta^\kappa \Gamma(\delta_i + \kappa)}{\Gamma(\kappa) \left(\vartheta + H_0^{(T)}(y_i|\Theta_T) e^{\mathbf{x}_i \boldsymbol{\beta}_T} \right)^{(\delta_i + \kappa)}} \right\}.$$

A log-normal é outra distribuição bastante utilizada para a modelagem da fragilidade. Para isso deve-se considerar a transformação indicada em (3.1), e definir $w \sim N(0, \tau)$, sendo τ a precisão. Banerjee et al. (2003) afirmam que, para o modelo de fragilidade log-normal ser identificável, o termo do efeito aleatório deve ter média zero.

O parâmetro $\tau = 1/\sigma^2$ tem a propriedade de captar a variabilidade não explicada no modelo; valores pequenos de τ indicam grande variabilidade não explicada. Entretanto, diferente da distribuição gama, não é possível determinar uma forma fechada para a função de verossimilhança marginal. Nesse sentido, sob a abordagem frequentista, a utilização de técnicas numéricas como o algoritmo EM são necessárias. Outra alternativa é o uso da classe de métodos MCMC, na abordagem Bayesiana, conforme descrito no Capítulo 2. A escolha destas técnicas torna desnecessária a solução da integral apresen-

tada em (3.3). Neste estudo, a distribuição log-nomal será utilizada na modelagem da fragilidade.

3.2 Censura Informativa

Segundo Collett (2003), o mecanismo de censura é não informativo quando não existe correlação entre os tempos de falha e censura, ou seja, T e C são independentes. Entretanto, não é difícil encontrar situações nas quais é possível perceber associação entre falha e censura, tornando a suposição de independência inadequada. Quando existe alguma relação entre os tempos de falha e censura dizemos que o mecanismo de censura é informativo. Para Staplin (2012) uma possível causa da censura informativa está relacionada ao fato de que fatores que controlam o tempo de censura estarem também associados ao tempo de falha.

Link (1989) cita algumas situações em que a suposição de independência pode ser violada, como por exemplo em estudos clínicos nos quais alguns pacientes deixam o tratamento por razões relacionadas ao tipo de terapia interferindo na possibilidade de sofrer o evento de interesse, ou quando pacientes deixam de ser acompanhados por experimentarem algum evento crítico específico relacionado ao evento de interesse (ex: alguma doença secundária). Nesse casos, é possível supor que esta observação censurada pode trazer alguma informação sobre o tempo de falha.

Entretanto, uma limitação na análise de dados de sobrevivência é testar o mecanismo de censura como não informativo, tendo como hipótese alternativa o caso informativo, (Staplin, 2012). Esta limitação ocorre devido ao teste requerer outras suposições sobre o modelo (Leung et al., 1997), como por exemplo o conhecimento da distribuição conjunta das variáveis T e C de forma que a independência entre estas variáveis seja um caso particular. Porém Staplin (2012) afirma que determinar uma distribuição conjunta para uma particular aplicação é muito difícil.

Observe que a suposição de independência desempenha um papel importante para a modelagem dos dados, conforme discutido no Capítulo 3. Considerando a dificuldade para se testar a suposição de independência, alguns autores têm proposto modelos que

ajustem dados sob ambos os mecanismos de censura de forma que a suposição de independência seja desnecessária. Emoto e Matthews (1990), por exemplo, assumiram o modelo Weibull bivariado para os tempos de falha e de censura e mostraram que os estimadores de máxima verossimilhança para a distribuição conjunta são consistentes. Uma modificação do estimador de Kaplan Meier é proposto por Kaciroti et al. (2012) para a análise Bayesiana de dados sob o mecanismo de censura informativa e dados faltantes. Siannis (2004) utiliza um modelo que contém um parâmetro para a associação entre os tempos de falha e censura e uma função viés na modelagem das observações de falha.

Modelos de fragilidade também são úteis para análise de dados de sobrevivência quando existe dependência entre os tempos de falha e censura, ou seja, dados com censura informativa. Link (1989) propõe um modelo para censura informativa utilizando um estimador de Kaplan–Meier modificado com um componente de fragilidade. A principal proposta deste trabalho é condicionar a função de verossimilhança aos dados observados e não observados (efeito aleatório) e assim tratar tempo de falha e censura como variáveis aleatórias independentes.

Considerar os tempos de falha e censura independentes dado a fragilidade é uma alternativa que simplifica a modelagem dos dados, tornando desnecessário conhecer sua distribuição conjunta. Um modelo para análise de dados sob o mecanismo informativo é proposto por Huang e Wolfe (2002).

3.2.1 Modelo de Huang e Wolfe

Uma das principais dificuldades no tratamento de dados de análise de sobrevivência é avaliar se a observação censurada traz alguma informação sobre o tempo até o evento de interesse e, conseqüentemente, identificar o mecanismo gerador dos dados. Diversos autores, como Leung et al. (1997) e Staplin (2012), discutem o uso indiscriminado da suposição de que os tempos de falha e censura são independentes devido, principalmente, a dificuldades de se testar essa premissa. Dessa forma, uma solução para a modelagem é permitir que o mecanismo de censura seja determinado pela informação contida no conjunto de dados.

Uma proposta de modelagem, que avalia ambos os mecanismos e utiliza um compo-

nente de fragilidade, foi desenvolvida por Huang e Wolfe (2002). Estes autores utilizam um modelo semiparamétrico que indica, através de um parâmetro, o mecanismo de censura sob o qual os dados foram gerados. O trabalho considera que os tempos de falha e censura são independentes quando condicionados ao componente de fragilidade e realizam uma análise de dados multivariados baseada em duas suposições: a primeira é que os indivíduos do mesmo grupo compartilham uma fragilidade comum; a segunda é que, dentro de cada grupo, a censura é independente da falha. Portanto, a associação entre falha e censura foi considerada a nível do grupo, permitindo a modelagem da censura informativa.

Sejam m grupos com n_i observações pertencentes a cada um deles, de forma que a amostra tenha tamanho $N = \sum_{i=1}^m n_i$. Para a construção do modelo assume-se que a fragilidade compartilhada representa também a correlação entre os indivíduos do mesmo grupo. Assume-se ainda que, condicional nas covariáveis e fragilidade, todas as N observações da amostras são independentes. A função de taxa de falha para os tempos de falha e censura é apresentada abaixo:

$$\begin{aligned} h^{(T)}(y_{i,j}|\delta_i, \mathbf{x}_{i,j}, w_i,) &= h_0^{(T)}(y_{i,j}) \exp\{\mathbf{x}_{i,j}\boldsymbol{\beta}_T + w_i\}; \\ h^{(C)}(y_{i,j}|\delta_i, \mathbf{x}_{i,j}, w_i,) &= h_0^{(C)}(y_{i,j}) \exp\{\mathbf{x}_{i,j}\boldsymbol{\beta}_C + \phi w_i\}, \end{aligned} \quad (3.4)$$

em que $i = 1, \dots, n_j$ e $j = 1, \dots, m$. A fragilidade w_i segue uma distribuição Normal com média zero e precisão τ .

As funções de taxa de falha base $h_0^{(T)}(\cdot)$ e $h_0^{(C)}(\cdot)$ são modeladas de forma não paramétrica. As variáveis aleatórias T_i e C_i são condicionalmente independentes dadas as covariáveis e a fragilidade, sendo que a dependência marginal ocorre devido à variável não observável afetar tanto a falha quanto a censura. Além disso, é possível observar que o risco de que a censura ocorra para determinado paciente é inflado pelo fator ϕ , em $(e^{w_i})^\phi$, de forma que se o paciente apresentar fragilidade $e^0 = 1$, este fator não terá influência na modelagem. Entretanto, quando $\phi > 0$, observa-se que a censura ocorrerá mais rapidamente para o indivíduo se a fragilidade apresentar um grande valor pois o

risco aumenta. Por outro lado, para $\phi < 0$, um grande valor de fragilidade permite que o indivíduo permaneça no estudo por mais tempo, pois o risco diminui com o aumento da fragilidade. Perceba que, para qualquer valor de $\phi \neq 0$, a fragilidade está presente na modelagem dos tempos de censura, indicando dependência entre T e C . Huang e Wolfe (2002) propõem que $\phi \in [-1.5, 1.5]$, sendo este parâmetro responsável por descrever a associação entre falha e censura, de forma que quanto maior a correlação, maior o valor em módulo de ϕ ; quando a correlação é positiva temos $\phi > 0$ e quando a correlação é negativa teremos $\phi < 0$. Segundo os autores, para valores de ϕ fora deste intervalo não há diferença significativa nos valores da correlação entre T e C .

É importante observar que, quando $\phi = 0$, a fragilidade não afetaria a censura, porém, ela continua presente na modelagem das observações de falha sendo responsável por captar a dependência dentro de cada grupo. Neste caso, retornamos ao modelo apresentado em (3.1) para o qual não é necessário estimar os parâmetros relacionados à censura, pois estes não são de interesse em nossa inferência.

No modelo proposto por Huang e Wolfe (2002), os tempos de falha e censura são independentes dentro de cada grupo, mas dependentes quando consideramos o conjunto da população. Isso significa, por exemplo, que dentro de cada grupo, a saída de alguns pacientes acontece de forma aleatória. Entretanto, alguns grupos podem possuir um número maior de saídas em detrimento de outros grupos, indicando que aquele grupo pode ter alguma característica que tenha influência no tempo de falha. O trabalho considera também um tipo de censura chamado de administrativa. Este caso inclui situações em que o indivíduo é censurado aleatoriamente por um motivo fora de controle do estudo; por exemplo, o paciente morre em um acidente de carro.

Como discutido na Seção anterior, a escolha da distribuição Normal para modelagem da fragilidade tem, como consequência, o fato de que as funções $h^{(T)}(\cdot)$ e $h^{(C)}(\cdot)$ não possuam formas fechadas. Assim, Huang e Wolfe (2002) optaram pelo ajuste usando o algoritmo EM Monte Carlo. Nesta dissertação, utilizaremos técnicas Bayesianas que tornam desnecessário a obtenção de funções de taxas de falha marginais para T e C .

Esta modelagem possui a vantagem de avaliar a existência de associação entre falha e censura e permite a identificação do mecanismo informativo ($\phi \neq 0$) ou não informativo

($\phi = 0$). Entretanto, o modelo é pouco flexível no sentido que caracteriza toda a amostra apenas em um tipo de mecanismo. É razoável pensar que existem situações nas quais o conjunto de dados possui alguns elementos com censura administrativa (independência entre falha e censura) e outros com dependência entre falha e censura. Se houver poucas censuras administrativas, seus efeitos serão oprimidos e praticamente ignorados.

Capítulo 4

Modelagem com Mistura

Os modelos de mistura são relevantes em Estatística devido a sua grande aplicabilidade em diversas áreas do conhecimento, tais como: Astronomia, Biologia, Genética, Medicina, entre outros. A importância destes modelos deve-se ao fato dos mesmos serem adequados para modelagem de populações formadas por vários grupos cujas proporções de indivíduos pertencentes a cada subpopulação é desconhecida.

A flexibilidade dos modelos de misturas tem sido aproveitada em diversas situações de análise de sobrevivência, como modelos de fração de cura, modelagem de dados heterogêneos, entre outros; algumas aplicações podem ser encontradas em Lawless (2003) e Ibrahim et al. (2010).

Gelman et al. (2014) apresenta uma modelagem geral para misturas. Defina uma amostra de n observações, para a qual existem M subpopulações, sendo M conhecido e definido. A distribuição conjunta desta população, modelada por M distribuições independentes, é dada por:

$$f(y_i|q) = q_1 f_1(y_i) + q_2 f_2(y_i) + \dots + q_M f_M(y_i) = \sum_{j=1}^M q_j f_j(y_i)$$

em que $f_j(y_i)$ é a função densidade de probabilidade da j -ésima componente e q_j é a probabilidade da observação i pertencer ao subgrupo j , satisfazendo $q_j \geq 0$ e $\sum_{j=1}^M q_j = 1$.

Uma forma simples e amplamente utilizada para modelagem de misturas é interpretá-la em termos de variáveis latentes (não observáveis) que indicam a componente da mistura à qual pertence cada elemento da amostra. Para o i -ésimo indivíduo da amostra temos um vetor $Z_i = (z_{i_1}, \dots, z_{i_M})$, no qual apenas uma componente $Z_j = 1$ sendo as demais iguais a zero. Assim $z_{i,j}$ é definido como:

$$z_{i,j} = \begin{cases} 1, & \text{se } y_i \text{ é proveniente da componente } j, \\ 0, & \text{caso contrário.} \end{cases}$$

em que $j = 1, \dots, M$ indica o subgrupo da população ao qual pertence aquela observação.

Gelman et al. (2014) acrescentam que, dado q_j , cada vetor Z_i pode ser modelado por uma distribuição Multinomial($1, q_1, \dots, q_M$), em que q_j representa a probabilidade de uma observação pertencer ao subgrupo. Neste caso, o modelo de mistura pode ser associado ao modelo hierárquico Bayesiano, em que $q = (q_1, \dots, q_j, \dots, q_M)$ pode ser pensado como um hiperparâmetro que determina a distribuição de Z_i . Logo, a distribuição conjunta da variável observada Y e da variável latente Z , condicionada nos parâmetros do modelo, pode ser escrita como:

$$f(y, Z|q) = f(Z|q)f(y|Z) = \prod_{i=1}^n \prod_{j=1}^M [q_j f_j(y_i)]^{z_{i,j}},$$

em que $f(y, Z|q)$ é função densidade da distribuição conjunta de Y e Z , com exatamente um $z_{i,j} = 1$ para cada i .

4.1 Modelo de Mistura com Três Componentes

O modelo de mistura, assim como na proposta de Huang e Wolfe (2002), corresponde a uma alternativa à modelagem que utiliza a informação contida nos dados para determinar o mecanismo de censura gerador da amostra. Este modelo também assume os tempos de falha e censura condicionalmente independentes dada a fragilidade, o que permite fatorar a função de verossimilhança em expressões que modelam a falha e a censura separadamente. Dessa forma, a dependência entre T e C ocorre apenas devido ao componente de

fragilidade.

Inicialmente pensou-se na utilização de uma mistura contendo dois componentes, sendo uma degenerada em zero e uma distribuição definida para a fragilidade das observações de falha. Entretanto, esta configuração não permitia a identificação de uma associação negativa entre falha e censura. Nesse sentido, optou-se pelo uso de uma distribuição multinomial que classificaria a fragilidade de acordo com três diferentes tipos de correlação: positiva, negativa ou independente. Essa multinomial também pode ser interpretada como uma mistura de degeneradas em -1 , 0 e 1 , como mostra a expressão (4.1):

$$\begin{aligned} h^{(T)}(y_i|\mathbf{x}_i, w_i) &= h_0^{(T)}(y_i) \exp\{\boldsymbol{\beta}_T \mathbf{x}_i + w_i\}; \\ h^{(C)}(y_i|\mathbf{x}_i, w_i, Z_i) &= h_0^{(C)}(y_i) \exp\{\boldsymbol{\beta}_C \mathbf{x}_i + (Z_i - 2)w_i\}. \end{aligned} \quad (4.1)$$

$$\begin{aligned} (Z_i) &\sim \text{Multinomial}(1, \mathbf{q}), \text{ sendo } Z_i = 1, 2, 3; \\ (\mathbf{q}) &\sim \text{Dirichlet}(\boldsymbol{\eta}), \end{aligned}$$

em que $\mathbf{q} = (q_1, q_2, q_3)$, representa a proporção de indivíduos pertencentes a cada grupo, com $q_1 + q_2 + q_3 = 1$ e $\boldsymbol{\eta} = (\eta_1, \eta_2, \eta_3)$.

Seja $Z_i^* = Z_i - 2$, indicada em (4.1). Esta componente pode ser representada por uma mistura de componentes degeneradas, de acordo com a notação a seguir:

$$f(Z_i^*) = q_1 d_{(-1)} + q_2 d_{(0)} + q_3 d_{(1)},$$

em que $d_{(\cdot)}$ é uma distribuição degenerada nos pontos -1 , 0 ou 1 , indicando, respectivamente, a existência de correlação negativa, independência e correlação positiva entre os tempos de falha e censura. Dessa forma os tempos de falha e censura serão considerados independentes (mecanismo não informativo) se a maior parte das observações estiverem associadas a $Z_i = 2$, ou seja, quando $q_2 > q_1$ e $q_2 > q_3$. Por outro lado, os tempos serão

considerados dependentes (mecanismo informativo) se a maior parte das observações estiverem associadas a $Z_i = 1$ ou $Z_i = 3$, sendo ainda possível identificar a correlação como negativa ($q_1 > q_2$ e $q_1 > q_3$) ou positiva ($q_3 > q_1$ e $q_3 > q_2$).

É possível observar que a modelagem de Huang e Wolfe (2002) e o modelo de misturas proposto possuem ideias semelhantes para o tratamento de dados de sobrevivência. Entretanto, existem algumas diferenças importantes. O modelo de Huang e Wolfe (2002) utiliza um parâmetro ϕ para avaliar o tipo e o impacto da fragilidade na modelagem da correlação entre os tempos de falha e censura, como mostrado em (3.4), enquanto o modelo de misturas agrupa os dados de acordo com três tipos de associação (positiva, negativa e nula). Veja que o modelo de mistura não possui um parâmetro ϕ para avaliar o impacto da fragilidade, entretanto, este impacto pode ser, de certa forma, também estudado através da variância da distribuição indicada para a fragilidade. Uma variância grande sugere fragilidades bem diferentes de zero e, portanto, maior impacto. Se consideramos uma distribuição Normal($0, \tau$), em que τ é a precisão, para a modelagem da fragilidade, como indicado no artigo de Huang e Wolfe (2002) haverá uma troca de informações entre ϕ e τ podendo acarretar dificuldades para estimar estes parâmetros. Isso não ocorrerá no modelo de misturas.

A grande vantagem no uso do modelo de misturas está em sua flexibilidade para tratar conjuntos de dados com censura parcialmente informativa, caso não contemplado pela proposta de Huang e Wolfe (2002). O modelo de mistura permite que cada elemento da amostra seja classificado de acordo com a relação entre seu tempo de falha e censura. Isso permite que alguns elementos da amostra sejam classificados como informativos, enquanto outros sejam considerados independentes, o que é razoável supor que aconteça em situações reais nas quais alguns elementos são censurados por causas não relacionadas ao tempo de falha.

Observe que o modelo proposto é bastante flexível, pois utiliza os pesos q_1 , q_2 e q_3 para classificar o conjunto de dados, de acordo com a relação entre o tempo de falha e censura para cada observação. Essa proposta substitui o parâmetro ϕ usado por Huang e Wolfe (2002), que tem uma função semelhante, porém dificulta a estimação de τ . Além disso, no modelo de Huang e Wolfe, parte da variabilidade que deveria ser captada por τ na

fragilidade é expressa em ϕ ; isso não ocorre no modelo de mistura visto que a fragilidade é multiplicada por valores fixos ($-1, 0$ ou 1). O parâmetro \mathbf{q} possui importante interpretação visto que indica o tipo de correlação predominando na amostra. Se avaliarmos Z_i para cada observação, teremos uma ideia sobre quais observações contribuem para a dependência e quais seriam censuras administrativas ($Z_i = 2$ ou $Z_i^* = 0$).

A principal desvantagem no uso do modelo de misturas é o grande número parâmetros. Entretanto, com o uso de métodos Bayesianos, distribuições a *priori* mais informativas podem melhorar o desempenho do modelo e conseqüentemente levar a melhores estimativas para os parâmetros mais sensíveis.

Capítulo 5

Estudo com Dados Simulados

Para a realização do estudo com dados simulados geramos três conjuntos de dados com diferentes características. O intuito é analisar e comparar o comportamento dos modelos discutidos ao longo do texto em diferentes cenários de forma a avaliar o ajuste do modelo a situações para os quais é indicado, bem como de seu comportamento em cenários para os quais sua escolha não é indicada. Foram construídos os seguintes cenários:

- $C_{ind.}$: observações geradas sem correlação entre os tempos de falha e censura (mecanismo de censura não informativo). Neste cenário, o efeito aleatório é incluído apenas na equação dos tempos de falha - expressão (3.1) - para a geração dos dados.
- $C_{pos.}$: observações geradas com correlação positiva entre os tempos de falha e censura (mecanismo informativo). O efeito aleatório compõe os componentes de falha e censura - expressão (3.4). Isto significa que observações que apresentarem grandes valores de fragilidade permanecerão por pouco tempo no estudo visto que sofrem a censura num curto período de tempo.
- $C_{neg.}$: observações geradas com correlação negativa entre os tempos de falha e censura (mecanismo informativo). A fragilidade está presente nas equações de falha e censura - expressão (3.4) - porém na equação de censura a fragilidade é multiplicada pelo fator -1 . Isto significa que observações que apresentarem grandes valores de fragilidade permanecerão mais tempo sob estudo.

Os dados foram gerados usando uma adaptação da técnica de inversão proposta por Bender et al. (2005), de forma que a fragilidade entre os tempos estivesse inclusa. Para compor o modelo, definiu-se o uso de uma covariável x_{i1} vinda de uma distribuição Normal com média zero e variância 1, e uma covariável dicotômica x_{i2} gerada de uma distribuição *Bernoulli* com probabilidade 0.5. O passo-a-passo a seguir mostra a obtenção do conjunto de observações referente ao cenário $C_{neg.}$. Assim temos:

1. Gere o valor das covariáveis $x_{1i} \sim \text{Normal}(0, 1)$ e $x_{2i} \sim \text{Bernoulli}(0.5)$;
2. Gere a fragilidade w_i de uma $\text{Normal}(0, 1)$;
3. Gere os valores u e v pertencentes ao intervalo $[0, 1]$;
4. Gere os tempos de falha $T_i \sim \text{Weibull}(\alpha_T, \lambda_{T,i})$ em que $\lambda_{T,i} = \exp\{\mathbf{x}_i\boldsymbol{\beta}_T + w_i\}$ através de $t_i = (-\log(u)/\lambda_{T,i})^{1/\alpha_T}$. Essa é uma adaptação da técnica de inversão do modelo de Cox (Bender et al., 2005);
5. Para gerar os tempos de censura, precisamos definir a correlação. No cenário de correlação negativa a fragilidade w é multiplicada pelo fator -1 . Assim os tempos são gerados por $C_i \sim \text{Weibull}(\alpha_C, \lambda_{C,i})$ em que $\lambda_{C,i} = \exp\{\mathbf{x}_i\boldsymbol{\beta}_C - 1w_i\}$, utilizando novamente a técnica de inversão do modelo de Cox: $c_i = (-\log(v)/\lambda_{C,i})^{1/\alpha_C}$;
6. Faça $Y_i = \min\{T_i, C_i\}$. Considere $\delta_i = 1$ se $Y_i = T_i$ e $\delta_i = 0$ caso contrário.

A construção dos cenários de correlação positiva e independência é semelhante. A principal diferença ocorre na geração dos tempos de censura (passo 5). Quando temos o cenário de correlação positiva multiplicamos o termo de fragilidade por $+1$ e no caso de independência este termo é multiplicado por zero, sendo que o restante dos passos permanecem os mesmos.

A proporção de falha também tem grande relevância no ajuste dos modelos, visto que, como o interesse reside em estimar os parâmetros associados à distribuição dos tempos de falha, um grande número de observações censuradas pode influenciar a estimação dos parâmetros. Dessa forma, nossa análise considerou proporções de falha pequena (aproximadamente 30%), media (aproximadamente 50%) e alta (aproximadamente 80%),

que foram definidas alternando-se apenas o parâmetro de forma. Além disso, considerou-se dois tamanhos amostrais (50 e 250).

Os valores simulados dos coeficientes das covariáveis são os mesmos utilizados por Huang e Wolfe (2002) em seu artigo. Os demais valores foram definidos de forma a produzir a proporção de falha desejada e garantir a correlação. São eles:

Tabela 5.1: Valores simulados.

%falha	$\alpha^{(T)}$	$\beta_0^{(T)}$	$\beta_1^{(T)}$	$\beta_2^{(T)}$	$\alpha^{(C)}$	$\beta_0^{(C)}$	$\beta_1^{(C)}$	$\beta_2^{(C)}$	τ
baixa	1.5	log(0.2)	0.1	-1.4	2.5	log(0.04)	0.2	1.2	0.5
média	1.5	log(0.2)	0.1	-1.4	1.5	log(0.04)	0.2	1.2	0.5
alta	2.5	log(0.2)	0.1	-1.4	1.5	log(0.04)	0.2	1.2	0.5

A Figura 5.1 traz os *boxplots* da correlação entre os tempos de falha e censura calculada para 50 amostras compostas de 50 tempos cada uma. Observe que os dados apresentam correlação adequada, conforme cenário em que estão alocados.

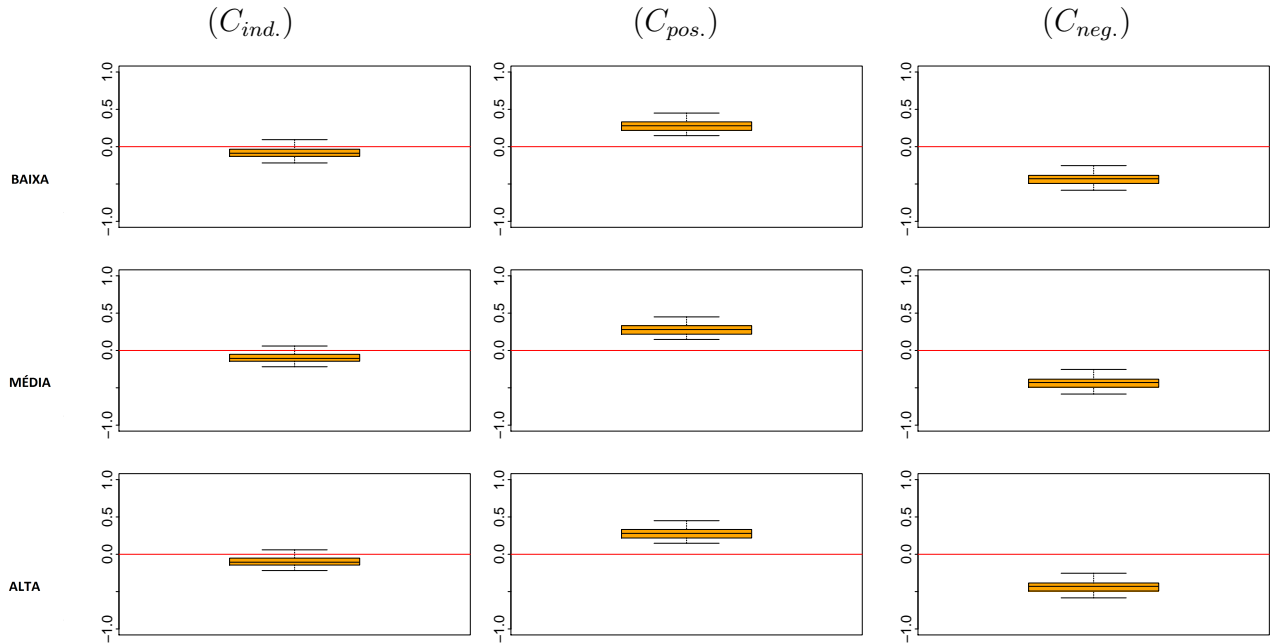


Figura 5.1: Correlação entre os tempos de falha e censura registrada nas 50 amostras geradas com 50 tempos.

Os dados referentes a cada cenário foram ajustados por quatro diferentes modelos, discutidos em 6.1. Todas as estimativas foram obtidas por meio de métodos MCMC,

aplicados através do *software JAGS* (Plummer, 2003). Para geração dos dados e análise das amostras *a posteriori*, utilizou-se o *software R* (R Core Team, 2014), juntamente com o pacote *coda* (Plummer et al., 2006) e *R2jags* (Su e Yajima, 2014). A simulação via MCMC foi realizada com a seguinte configuração: uma cadeia com 30 mil iterações, sendo as 20 mil primeiras referentes ao período de aquecimento. A amostra *a posteriori* é de tamanho 1000, visto que utilizamos *lag* igual a 10 após o *burn-in*. Este MCMC será executado para 50 amostras diferentes de tempos de sobrevivência geradas com a mesma configuração de parâmetros. Esta estratégia será usada para evitar que trabalhem com uma única amostra que pode ser eventualmente problemática (contendo *outliers* por exemplo). A análise dos resultados baseou-se na estatística descritiva das 1000 amostras *a posteriori* obtidas para cada parâmetro em todos os modelos e cenários. A descrição e especificação dos modelos é apresentada a seguir.

5.1 Especificação de Modelos

Para melhor organização do trabalho, nesta seção apresentaremos a forma hierárquica, distribuições *a priori* e especificação dos hiperparâmetros adotada para quatro modelos de sobrevivência mencionados ao longo do texto, sendo dois modelos indicados para cenários de mecanismo não informativo e dois modelos para o mecanismo informativo. Omitiremos no corpo do texto os cálculos para obtenção das distribuições *a posteriori* que serão apresentados no Apêndice.

5.1.1 Modelo M_0

O modelo M_0 é a modelagem mais simples para tratamento de dados de sobrevivência na presença de covariáveis e sua aplicação está bem consolidada na literatura. Esta modelagem possui como principal suposição a independência entre observações de falha e censura, o que simplifica a estrutura do modelo permitindo a análise apenas de parâmetros associados à distribuição da falha. A função de taxa de falha do modelo M_0 é dada por:

$$h^{(T)}(y_i) = \alpha_T y_i^{\alpha_T - 1} \exp\{\mathbf{x}_i \boldsymbol{\beta}_T\}.$$

Note que o número reduzido de parâmetros é uma grande vantagem de $M0$ e esperamos que ele tenha melhor desempenho quando a suposição de independência for válida. Quando não houver independência entre falha e censura, as estimativas dos parâmetros podem ser viesadas, sendo esta a principal desvantagem do modelo. Além disso, esse modelo mais simples não considera fontes de variabilidade extra que interferem nas estimativas. Defina $D = \{(y_i, \delta_i, x_i), i = 1, \dots, n\}$. O modelo hierárquico completo para $M0$ é dado por:

$$\begin{aligned} T_i | D, \Theta_T &\sim \text{Weibull}(\alpha_T, \lambda_{T,i}), \\ \alpha_T &\sim \text{Gama}(a_T, b_T), \\ \beta_{T,j} &\sim \text{Normal}(0, V_{T,j}), \quad j = 0, 1, 2. \end{aligned} \tag{5.1}$$

Os hiperparâmetros $V_{T,j}$, a_T e b_T são definidos pelo pesquisador, $\lambda_{T,i} = \exp\{\mathbf{x}_i \boldsymbol{\beta}_T\}$, em que $\boldsymbol{\beta}_T$ é o vetor de coeficientes de regressão e $\mathbf{x}_i = (1, x_{1i}, \dots, x_{ki})$ é um vetor com k covariáveis. A parametrização utilizada neste trabalho para a distribuição Normal considera $V_{T,j}$ como a precisão ao invés da variância, de acordo com a parametrização adotada no *software JAGS*.

Os valores dos hiperparâmetros foram escolhidos de forma a estabelecer distribuições *a priori* pouco informativas. Para a Gama especificamos $a_T = 0.6$ e $b_T = 0.3$ que garantem uma distribuição com média 2, mediana 1 e variância igual a 7. Esta escolha concentra grande parte da massa de probabilidade no intervalo $[0, 3]$ e ainda apresenta uma cauda que se estende além de 10. Para a Normal, assumimos uma precisão $V_{T,j} = 0.01$ (variância 100) para $j = 1, 2$, o que indica variabilidade grande no suporte da distribuição dos $\beta_{T,j}$. Para o parâmetro $\beta_{T,0}$, uma variância um pouco menor foi definida, com $V_{T,0} = 0.1$. Esta escolha foi estabelecida para que haja um maior controle sobre o valor do intercepto que tem forte influência sobre o modelo. Uma distribuição *a priori* muito vaga levaria a uma maior incerteza *a posteriori* permitindo maior variação de $\beta_{T,0}$, o que poderia prejudicar a estimação dos demais parâmetros.

O uso da distribuição Weibull para a modelagem dos dados não permite a obtenção de distribuições *a posteriori* que possuem forma fechada, sendo necessário o uso de métodos MCMC para amostragem indireta. O núcleo das distribuições *a posteriori* condicionais completas está disponível no Apêndice A.

5.1.2 Modelo $M1$

Para a construção do modelo $M1$, um efeito aleatório (fragilidade) é incluído apenas na distribuição dos tempos de falha, tornando-se uma extensão do modelo simples $M0$. O modelo $M1$ ainda tem como principal suposição a independência entre os tempos de falha e censura e, portanto, pode ser escrito levando em conta apenas a modelagem de falha. A inclusão da componente de fragilidade neste modelo traz maior flexibilidade permitindo que qualquer variabilidade extra não explicada pelas covariáveis observadas possa ser incorporada de forma a melhorar o ajuste, sendo esta sua principal vantagem. Entretanto, se não houverem fontes extras de variação o efeito aleatório pode interferir na estimação dos demais parâmetros. A função de taxa de falha deste modelo é dada por:

$$h^{(T)}(y_i) = \alpha_T y_i^{\alpha_T - 1} \exp\{\mathbf{x}_i \boldsymbol{\beta}_T + w_i\}.$$

A hierarquia deste modelo difere pouco da utilizada para o modelo $M0$, sendo que incluiremos a modelagem do efeito aleatório da fragilidade. Considerando $D = \{(y_i, \delta_i, \mathbf{x}_i, w_i), i = 1, \dots, n\}$, o modelo hierárquico completo para $M1$ é dado por:

$$\begin{aligned} T_i | D, \boldsymbol{\Theta}_T &\sim \text{Weibull}(\alpha_T, \lambda_{T,i}), & (5.2) \\ \alpha_T &\sim \text{Gama}(a_T, b_T), \\ \beta_{T,j} &\sim \text{Normal}(0, V_{T,j}), \quad j = 0, 1, 2, \\ w | \tau &\sim \text{Normal}(0, \tau), \\ \tau &\sim \text{Gama}(m_1, m_2). \end{aligned}$$

Novamente, os hiperparâmetros a_T , b_T , precisão $V_{T,j}$, m_1 e m_2 , são definidos pelo pesquisador, $\lambda_{T,i} = \exp\{\mathbf{x}_i\boldsymbol{\beta}_T + w_i\}$. Para a distribuição da fragilidade optou-se pela Normal com média zero e precisão τ , comumente utilizada para descrever este efeito. Para a modelagem do parâmetro τ utilizou-se a distribuição Gama, por ser conjugada com a distribuição Normal.

Procurando manter as mesmas características em todos os modelos, as especificações *a priori* utilizadas em $M0$ foram mantidas, sendo: $a_T = 0.6$, $b_T = 0.3$, $V_{T,j} = 0.01$ para $j = 1, 2$ e $V_{T,0} = 0.1$. A especificação *a priori* de τ foi $m_1 = m_2 = 0.1$, que possui média 1 e variância 10, colocando grande massa em valores próximos de zero; veja que τ é o parâmetro de precisão, então concentrar maior massa de probabilidade próximo de zero significa indicar que a variância do efeito aleatório pode ser grande.

5.1.3 Modelo $M2$

Desenvolvido por Huang e Wolfe (2002), o modelo $M2$ propõe uma generalização de modelos de sobrevivência no qual a informação dos dados é utilizada para identificar o mecanismo de censura gerador da amostra. Este modelo é também adequado para a modelagem de conjunto de dados sob o mecanismo de censura informativa, permitindo avaliar o grau de correlação entre os tempos (medida pelo parâmetro ϕ). Reveja a discussão sobre esta proposta apresentada na Subseção 4.2.1. As funções de taxa de falha para os tempos de falha e censura no caso em que cada indivíduo que compõe a amostra constitui seu próprio *cluster*, são:

$$\begin{aligned} h^{(T)}(y_i) &= \alpha_T y_i^{\alpha_T - 1} \exp\{\mathbf{x}_i\boldsymbol{\beta}_T + w_i\}; \\ h^{(C)}(y_i) &= \alpha_C y_i^{\alpha_C - 1} \exp\{\mathbf{x}_i\boldsymbol{\beta}_C + \phi w_i\}. \end{aligned}$$

Originalmente, Huang e Wolfe (2002) propõem uma abordagem frequentista para modelagem. Nesta dissertação apresentamos uma versão Bayesiana do modelo proposto por Huang e Wolfe (2002). Seja $D = \{(y_i, \delta_i, \mathbf{x}_i, w_i), i = 1, \dots, n\}$, a descrição hierárquica do modelo é dada por:

$$\begin{aligned}
T_i|D, \Theta_T &\sim \text{Weibull}(\alpha_T, \lambda_{T,i}), \\
C_i|D, \Theta_C &\sim \text{Weibull}(\alpha_C, \lambda_{C,i}), \\
\alpha_T &\sim \text{Gama}(a_T, b_T), \\
\alpha_C &\sim \text{Gama}(a_C, b_C), \\
\beta_{T,j} &\sim \text{Normal}(0, V_{T,j}), \quad j = 0, 1, 2, \\
\beta_{C,j} &\sim \text{Normal}(0, V_{C,j}), \quad j = 0, 1, 2, \\
\phi &\sim \text{Uniforme}(l_1, l_2), \\
w|\tau &\sim \text{Normal}(0, \tau), \\
\tau &\sim \text{Gama}(m_1, m_2).
\end{aligned} \tag{5.3}$$

Veja acima que para a composição deste modelo optou-se por manter as distribuições *a priori* para a falha já definidas em (6.2). Visto que os parâmetros de falha e censura possuem o mesmo suporte, utilizaremos especificações *a priori* semelhantes para os parâmetros de censura. A fragilidade foi modelada pela Normal $(0, \tau)$ e o hiperparâmetro τ pela distribuição gama para garantir a conjugação. A distribuição Uniforme foi escolhida para ϕ de acordo com as características desse parâmetro (intervalo real e centrado em zero) dadas por Huang e Wolfe (2002).

A especificação dos hiperparâmetros é igual para as distribuições de falha e censura, sendo: $a_T = a_C = 0.6$, $b_T = b_C = 0.3$, $m_1 = m_2 = 0.1$, precisão $V_{T,j} = V_{C,j} = 0.01$ para $j = 1, 2$ e $V_{T,0} = V_{C,0} = 0.1$, novamente privilegiando distribuições pouco informativas. Os hiperparâmetros para a distribuição de ϕ são $l_1 = -50$ e $l_2 = 50$. É importante observar que este intervalo é bem mais abrangente do que o proposto por Huang e Wolfe (2002) que fixam ϕ no intervalo $[-1.5, 1.5]$. Estamos optando por introduzir maior incerteza *a priori* sobre esse parâmetro.

Considerando independência condicional entre T e C , os núcleos das distribuições condicionais completas *a posteriori* para os parâmetros de falha Θ_T e τ deste modelo serão iguais às obtidas para $M1$.

5.1.4 Modelo $M3$

O modelo $M3$ é uma alternativa de modelagem que também permite que os dados determinem o tipo de relação entre falha e censura. Este modelo ajusta amostras sob ambos os mecanismos de censura (informativo e não informativo). Ele permite também avaliar o tipo de associação (correlação positiva ou negativa) entre T_i e C_i . Lembramos ao leitor que a Seção 5.1 descreveu elementos básicos sobre modelagem de misturas com três componentes que serve como base para construção do modelo $M3$. A função taxa de falha será dada por:

$$\begin{aligned} h^{(T)}(y_i) &= \alpha_T y_i^{\alpha_T - 1} \exp\{\mathbf{x}_i \boldsymbol{\beta}_T + w_i\}; \\ h^{(C)}(y_i) &= \alpha_C y_i^{\alpha_C - 1} \exp\{\mathbf{x}_i \boldsymbol{\beta}_C + (Z_i - 2)w_i\}. \end{aligned}$$

A fragilidade na modelagem das observações censuradas é multiplicada por $(Z_i - 2)$ que determina o tipo de relação existente entre falha e censura. Defina $D = \{(y_i, \delta_i, \mathbf{x}_i, w_i, Z_i), i = 1, \dots, n\}$. A descrição hierárquica do modelo é dada por:

$$\begin{aligned} T_i | D, \boldsymbol{\Theta}_T &\sim \text{Weibull}(\alpha_T, \lambda_{T,i}), \\ C_i | D, \boldsymbol{\Theta}_C &\sim \text{Weibull}(\alpha_C, \lambda_{C,i}), \\ \alpha_T &\sim \text{Gama}(a_T, b_T), \\ \alpha_C &\sim \text{Gama}(a_C, b_C), \\ \beta_{T,j} &\sim \text{Normal}(0, V_{T,j}), \quad j = 0, 1, 2, \\ \beta_{C,j} &\sim \text{Normal}(0, V_{C,j}), \quad j = 0, 1, 2, \\ Z_i &\sim \text{Multinomial}(1, \mathbf{q}), \quad \mathbf{q} = (q_1, q_2, q_3) \text{ e } (q_1 + q_2 + q_3) = 1, \\ w | \tau &\sim \text{Normal}(0, \tau), \\ \tau &\sim \text{Gama}(m_1, m_2), \\ \mathbf{q} | Z &\sim \text{Dirichlet}(\boldsymbol{\eta}), \quad \boldsymbol{\eta} = (\eta_1, \eta_2, \eta_3). \end{aligned} \tag{5.4}$$

Como ocorreu nos modelos anteriores, as distribuições *a priori* foram escolhidas de forma a respeitar o domínio dos parâmetros; mantemos as mesmas distribuições para os parâmetros de falha e censura. Para a variável latente Z_i utilizou-se uma Multinomial($1, \mathbf{q}$), de forma a distinguir o tipo de correlação existente, sendo que cada observação é classificada em um único grupo, conforme apresentado na Seção 5.1. O hiperparâmetro \mathbf{q} , que desempenha importante função neste modelo indicando a probabilidade do tipo de associação, foi modelado por uma distribuição Dirichlet, que apresenta conjugação com a Multinomial.

As especificações *a priori* utilizadas nas demais modelagens foram mantidas, sendo: $a_T = a_C = 0.6$, $b_T = b_C = 0.3$, $V_{T,j} = V_{C,j} = 0.01$, $V_{T,0} = V_{C,0} = 0.1$ e $m_1 = m_2 = 0.1$. A escolha para $\boldsymbol{\eta} = (1, 1, 1)$ corresponde a uma especificação Uniforme no espaço $[0, 1] \times [0, 1]$ para q_j , $j = 1, 2, 3$. Veja que não estamos privilegiando qualquer valor de q_j .

Novamente, a independência condicional entre T e C é considerada e, consequentemente, o núcleo das distribuições condicionais completas *a posteriori* são os mesmos obtidos para $M1$ em relação aos parâmetros de falha Θ_T . As demais distribuições condicionais completas *a posteriori* estão disponíveis no Apêndice A.

5.2 Resultados

Adotando as configurações mencionadas neste capítulo para o MCMC e as especificações *a priori* de cada modelo apresentado nas Subseções 6.1.1 a 6.1.4, os resultados a serem analisados neste estudo simulado são obtidos de acordo com os seguintes passos:

1. Gerar um conjunto de dados com uma das configurações ($C_{ind.}$, $C_{pos.}$ e $C_{neg.}$) apresentadas no início desta Seção;
2. Ajustar os quatro modelos: $M0$, $M1$, $M2$, $M3$; obtendo assim, via MCMC por meio do pacote *R2jags* (Su e Yajima, 2014) e *snowfall* (Knaus, 2013) disponíveis no R (R Core Team, 2014), as cadeias da distribuição *a posteriori* de cada parâmetro;
3. Assumindo a função de perda quadrática estimaremos os parâmetros tomando a média da amostras *a posteriori*. Os intervalos de credibilidade HPD também serão

calculados;

4. Visto que os dados são simulados, os valores verdadeiros são conhecidos (Tabela 5.1), podemos calcular o vício relativo, definido por $VR(\theta) = (\hat{\theta} - \theta)/|\theta|$; sendo θ o valor verdadeiro e $\hat{\theta}$ a média a *posteriori*
5. Seja $VR_{M_i}(\cdot)$ o vício relativo do modelo M_i , para $i = 0, 1, 2, 3$. Calcularemos a razão de vícios para comparar os modelos. Neste cálculo, o valor absoluto da razão de vícios é usado para evitar que o resultado oscile entre valores negativos. A fórmula é dada por:

$$RV = \left| \frac{VR_{M_i}(\cdot)}{VR_{M_j}(\cdot)} \right|, \quad (5.5)$$

para $i = 0, 1, 2, 3$ e $j = 2, 3$ com $i \neq j$ para cada parâmetro de interesse

Note que a razão de vícios considera no denominador os modelos M_2 e M_3 . Uma razão próxima de 1 indica desempenho similar dos modelos sob comparação, razão abaixo de 1 indica melhor desempenho do modelo no numerador e razão acima de 1 sugere melhor performance do modelo no denominador.

6. Retorne ao item 1 por uma quantidade determinada de vezes (neste trabalho consideramos 50 repetições). Destacamos novamente que esta estratégia de replicação visa evitar que o MCMC seja executado para uma única amostra que poderia conter *outliers* levando assim a interpretações incorretas sobre o desempenho de um modelo.

Veja que ao final deste procedimento teremos uma amostra (tamanho 50) de vícios relativos, razões de vícios, médias a *posteriori* e intervalos de credibilidade HPD de 95%.

Todos os gráficos analisados nesta seção foram construídos obedecendo à seguinte configuração: as colunas representam os cenários de geração de dados, sendo a Coluna 1 representando o cenário $C_{ind.}$, a Coluna 2 representando $C_{pos.}$ e Coluna 3 relacionado ao cenário $C_{neg.}$. As linhas representam os percentuais de falha sendo a Linha 1 referente à proporção de falha pequena, a Linha 2 representando a proporção média e a Linha 3

referente à proporção grande. Dessa forma, o primeiro gráfico de cada quadro refere-se ao cenário independente com pequena proporção de falha, enquanto o último refere-se ao cenário de correlação negativa com grande proporção de falha. Esses gráficos exibem resultados relativos aos parâmetros α_T , $\beta_{T,0}$, $\beta_{T,1}$ e $\beta_{T,2}$ que modelam os tempos de falha.

5.2.1 Resultados para amostras de tamanho 50.

Tabela 5.2: Estimativas para amostra tamanho 50

cenário	% falha	modelo	α			β_0			β_1			β_2		
			VR	EP	PC	VR	EP	PC	VR	EP	PC	VR	EP	PC
$C_{ind.}$	baixa	M0	-0.26	0.21	0.46	0.13	0.34	0.86	-0.63	0.25	0.96	0.29	0.59	0.78
		M1	2.05	2.37	0.94	-1.07	1.67	0.96	-0.75	1.41	0.98	-2.71	3.71	0.96
		M2	0.77	1.57	1.00	-0.49	1.28	1.00	-0.41	0.86	1.00	-1.02	2.42	0.96
		M3	0.54	1.08	1.00	-0.46	1.05	0.98	-0.62	0.67	0.98	-0.69	1.72	0.96
	média	M0	-0.32	0.15	0.20	0.14	0.33	0.92	-0.42	0.21	0.94	0.46	0.45	0.64
		M1	2.08	2.21	0.84	-1.05	1.60	0.96	0.51	1.36	1.00	-2.52	3.38	0.92
		M2	0.87	1.61	1.00	-0.54	1.25	0.98	0.68	0.84	0.98	-0.91	2.40	0.96
		M3	0.78	1.21	1.00	-0.58	1.06	0.98	-0.34	0.73	0.98	-0.89	1.86	0.94
	alta	M0	-0.37	0.20	0.02	0.17	0.31	0.86	-0.88	0.19	0.88	0.51	0.38	0.46
		M1	1.51	2.94	0.86	-0.81	1.36	0.98	-0.70	0.97	0.96	-1.68	2.60	0.96
		M2	0.55	2.01	1.00	-0.34	1.02	0.98	-0.37	0.64	0.98	-0.53	1.69	0.98
		M3	0.81	1.90	0.92	-0.56	1.04	0.92	-0.66	0.68	0.94	-0.84	1.68	0.92
$C_{pos.}$	baixa	M0	-0.41	0.16	0.18	0.14	0.34	0.88	-0.82	0.25	0.92	-0.04	0.70	0.96
		M1	1.54	1.86	0.94	-1.15	1.62	0.90	-0.90	1.35	0.98	-3.37	3.88	0.86
		M2	0.71	1.43	1.00	-0.35	1.06	0.96	-0.07	0.72	0.96	-0.91	2.05	0.96
		M3	0.58	1.04	0.98	-0.52	0.96	0.98	-0.16	0.71	1.00	-1.28	1.92	0.92
	média	M0	-0.41	0.13	0.06	0.17	0.32	0.86	-1.23	0.21	0.84	0.11	0.48	0.92
		M1	1.70	1.91	0.86	-1.07	1.57	0.90	-1.82	1.23	0.98	-3.21	3.59	0.84
		M2	0.65	1.30	0.98	-0.29	1.03	1.00	-0.57	0.68	0.96	-0.75	1.70	1.00
		M3	0.90	1.27	0.94	-0.59	1.04	0.94	-1.07	0.79	0.96	-1.54	2.15	0.86
	alta	M0	-0.39	0.19	0.06	0.15	0.30	0.88	-1.28	0.19	0.84	0.25	0.38	0.74
		M1	1.20	2.57	0.94	-0.81	1.31	0.92	-1.54	0.89	0.96	-1.94	2.56	0.94
		M2	0.33	1.56	0.96	-0.14	0.84	0.98	-0.79	0.52	0.96	-0.43	1.34	0.98
		M3	1.06	2.04	0.78	-0.65	1.08	0.90	-1.37	0.76	0.98	-1.34	1.89	0.94
$C_{neg.}$	baixa	M0	-0.10	0.24	0.86	0.17	0.34	0.86	-0.44	0.24	0.98	0.39	0.52	0.72
		M1	2.42	2.74	0.90	-0.80	1.61	0.94	0.88	1.33	1.00	-2.15	3.40	0.94
		M2	0.68	1.24	1.00	-0.51	1.17	0.98	-0.77	0.75	1.00	-0.76	2.01	0.92
		M3	0.93	1.21	0.94	-0.60	1.18	0.98	-0.22	0.76	0.98	-0.84	1.87	0.90
	média	M0	-0.15	0.20	0.62	0.15	0.33	0.78	0.06	0.21	0.94	0.52	0.43	0.54
		M1	2.37	2.63	0.84	-0.75	1.48	0.94	2.30	1.27	0.98	-1.83	3.02	0.92
		M2	0.58	1.09	0.96	-0.37	1.03	0.92	-0.07	0.64	1.00	-0.44	1.66	0.88
		M3	1.22	1.36	0.84	-0.70	1.19	0.94	0.50	0.84	0.98	-0.94	1.91	0.94
	alta	M0	-0.20	0.27	0.42	0.14	0.32	0.78	0.001	0.20	0.92	0.59	0.38	0.34
		M1	1.70	3.50	0.94	-0.57	1.28	0.96	0.86	1.02	0.98	-1.12	2.44	0.96
		M2	0.32	1.42	0.96	-0.29	0.90	0.94	-0.11	0.49	0.98	-0.02	1.26	0.90
		M3	1.24	2.24	0.82	-0.58	1.15	0.96	0.75	0.80	0.98	-0.69	1.77	0.94

VR = vício relativo - EP = erro padrão - PC = probabilidade de cobertura

Os resultados exibidos aqui são obtidos para amostras contendo 50 tempos (incluindo falha e censura). Esta amostra é bem menor em relação àquela que será avaliada na subseção seguinte (250 tempos). Nosso objetivo aqui é avaliar o comportamento dos

modelos diante de um pequeno conjunto de dados. As médias da probabilidade de cobertura, vício relativo e erro padrão obtidas para as 50 replicações MCMC são mostradas na Tabela 5.2 para os parâmetros da distribuição dos tempos de falha.

A probabilidade de cobertura representa a proporção de vezes que o intervalo HPD incluiu o valor real do parâmetro. Observe que o modelo $M2$ tem bom ajuste, e apresenta para todos os cenários probabilidade de cobertura próxima a 95% para a maioria dos parâmetros. Os modelos $M1$ e $M3$ também apresentam desempenhos satisfatórios e similares entre si para a maioria dos cenários. Para estes modelos, a probabilidade de cobertura está, em geral, acima de 90%. As menores probabilidades de cobertura podem ser observadas para o modelo $M0$ (destaque na tabela), principalmente em relação aos parâmetros α e β_2 em todos os cenários. O modelo $M0$ não possui estrutura para explicar a variabilidade extra dos dados, desta forma a estimação dos parâmetros fica prejudicada na tentativa de acomodar essa informação extra. Para os demais parâmetros (β_0 e β_1) porém, $M0$ exibe probabilidade de cobertura próxima à nominal indicando uma adequada estimação.

O aumento da proporção de falha e mudança do tipo de correlação alteram levemente a qualidade do ajuste do modelo $M0$, sendo possível observar um decréscimo dos valores médios de probabilidade de cobertura à medida que aumentamos o número de falha. Para os demais modelos não identificamos padrão na variação dos níveis de probabilidade de cobertura relacionados aos diferentes percentuais de falha ou tipos de correlação.

Ao gerarmos os tempos de falha e censura no cenário de correlação negativa, estamos introduzindo o mesmo valor de fragilidade w_i , porém com sinal contrário na determinação do tempo de censura. Isto terá influência sobre a proporção de falha controlada através das escolhas dos valores reais de α_T e α_C . Note que essa troca de sinal determina: um aumento em w_i causa aumento em $\lambda_{T,i}$ e aumento em T_i , porém teremos diminuição em $\lambda_{C,i}$ e diminuição em C_i . Visto que trabalhamos com censura à direita e definimos $Y_i = \min(T_i, C_i)$, diminuir C_i e aumentar T_i tem impacto sobre a proporção nominal de falha. Neste cenário negativo, a proporção de falha estará mais próxima de 50%. Isso faz com que, para o tipo de correlação negativa, $M0$ apresenta probabilidade de convergência superior ao observado para os demais cenários em relação ao parâmetro α .

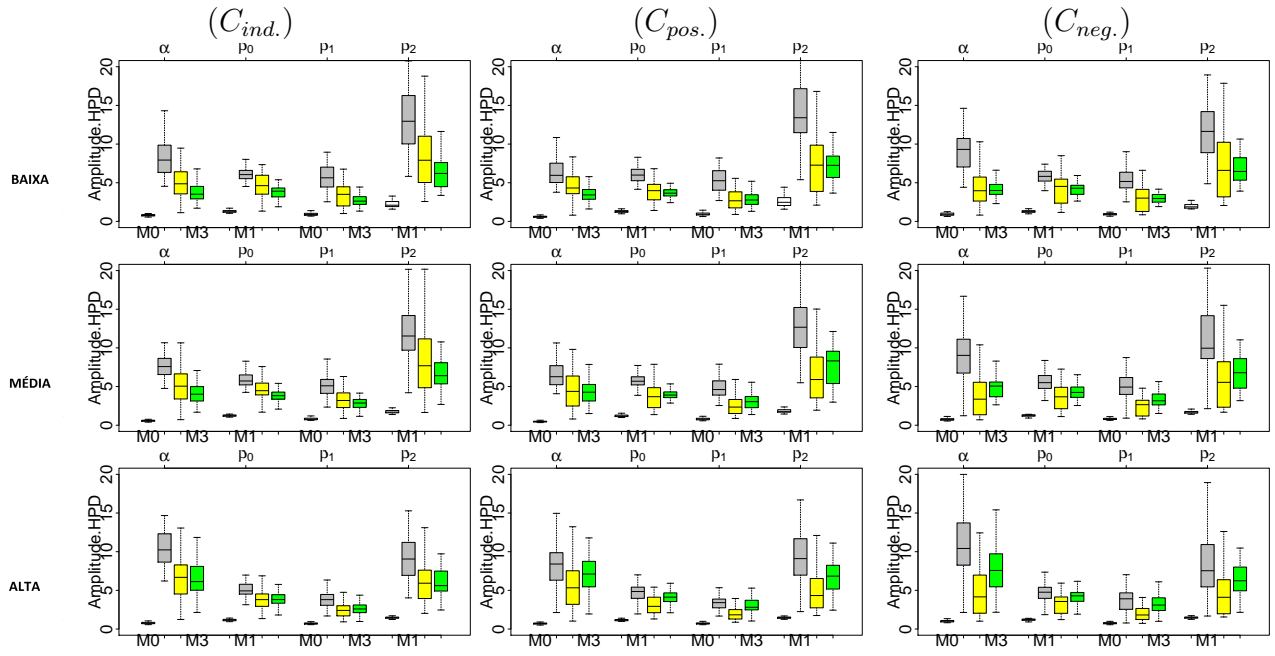


Figura 5.2: Amplitude do intervalo HPD registrada nas 50 replicações do MCMC para o ajuste de uma amostra de 50 tempos.

O erro padrão também auxilia na análise do intervalo HPD. Note que os menores erros padrões são obtidos para o modelo $M0$ o que implica menores intervalos de credibilidade. Ocorre que, com intervalos muito pequenos, o intervalo HPD para $M0$ não inclui o valor real do parâmetro prejudicando as inferências desses modelos. O modelo $M1$, entretanto, apresenta o maior erro padrão indicando grande variabilidade das estimativas *a posteriori*. Esses resultados também podem ser observados na Figura 5.2 que representa a amplitude do intervalo HPD para os parâmetros de falha dos modelos $M0$, $M1$, $M2$ e $M3$. Na parte superior do gráfico temos indicado os parâmetros avaliados, enquanto que na parte inferior é possível identificar o modelo.

A amplitude do HPD é um bom indicativo da qualidade do ajuste do modelo, visto que intervalos muito amplos sugerem grande incerteza *a posteriori*. O modelo $M0$ produz os intervalos com menores amplitudes indicando pouca variabilidade *a posteriori*. O modelo $M1$ apresenta os intervalos de maior amplitude. Para os modelos $M2$ e $M3$, observamos que existe um comportamento similar entre eles, principalmente para pequenas

proporções de falha, entretanto o modelo $M2$ tem maior variabilidade de amplitudes nas 50 réplicas, seus *boxplots* são um pouco mais amplos. Em quase todos os cenários, a amplitude do intervalo HPD é maior e varia mais para o parâmetro β_2 , indicando incerteza a *posteriori*. Lembre-se que β_2 é o parâmetro associado a uma variável binária e, em geral, o coeficiente desse tipo de covariável é mais difícil de estimar. Um comportamento similar ao de β_2 também pode ser observado em relação ao parâmetro α em alguns cenários. Este é parâmetro está no expoente, sendo influenciado pelas estimativas de β_0 . Observe que os gráficos 5.3 e 5.3 são espelhados, indicando que a superestimação de um parâmetros e compensado na subestimativa do outro.

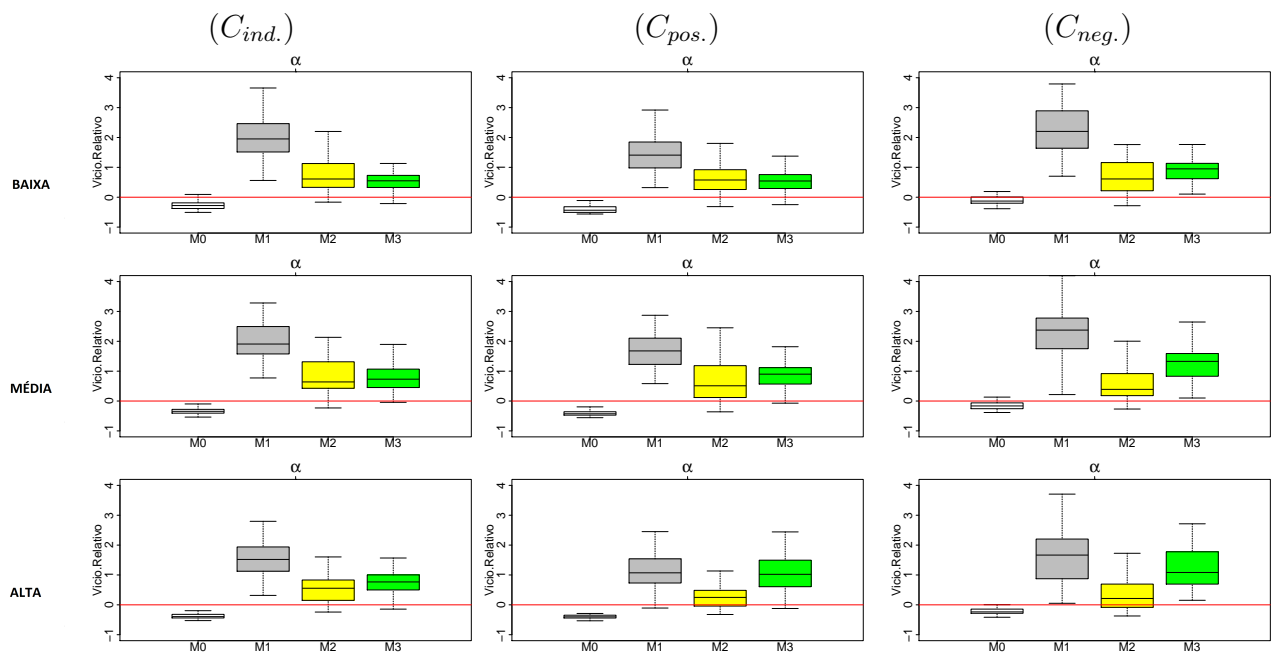


Figura 5.3: *Boxplots* dos vícios relativos para α obtidos em cada replicação do MCMC para ajustes considerando amostras de 50 tempos.

As Figuras 5.3, 5.4, 5.5 e 5.6 apresentam os *boxplots* do vício relativo para os parâmetros de falha de cada modelo. Nestes gráficos, um traço horizontal no patamar zero é marcado para auxiliar visualmente a análise.

Para o parâmetro α , nota-se o maior vício do Modelo $M1$ em comparação aos demais modelos. O modelo $M3$ também apresenta vício relativo em torno de 1 com exceção do cenário de correlação negativa cujo vício observado para este modelo é um pouco

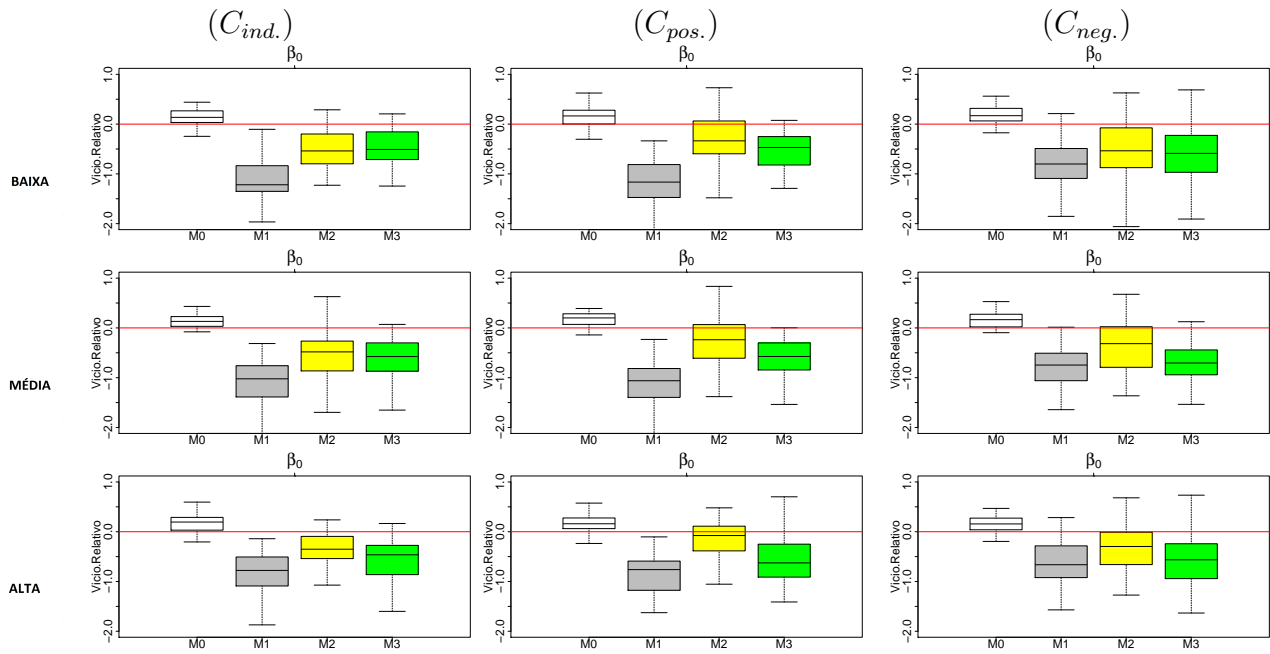


Figura 5.4: *Boxplots dos vícios relativos para β_0 obtidos em cada replicação do MCMC para ajustes considerando amostras de 50 tempos.*

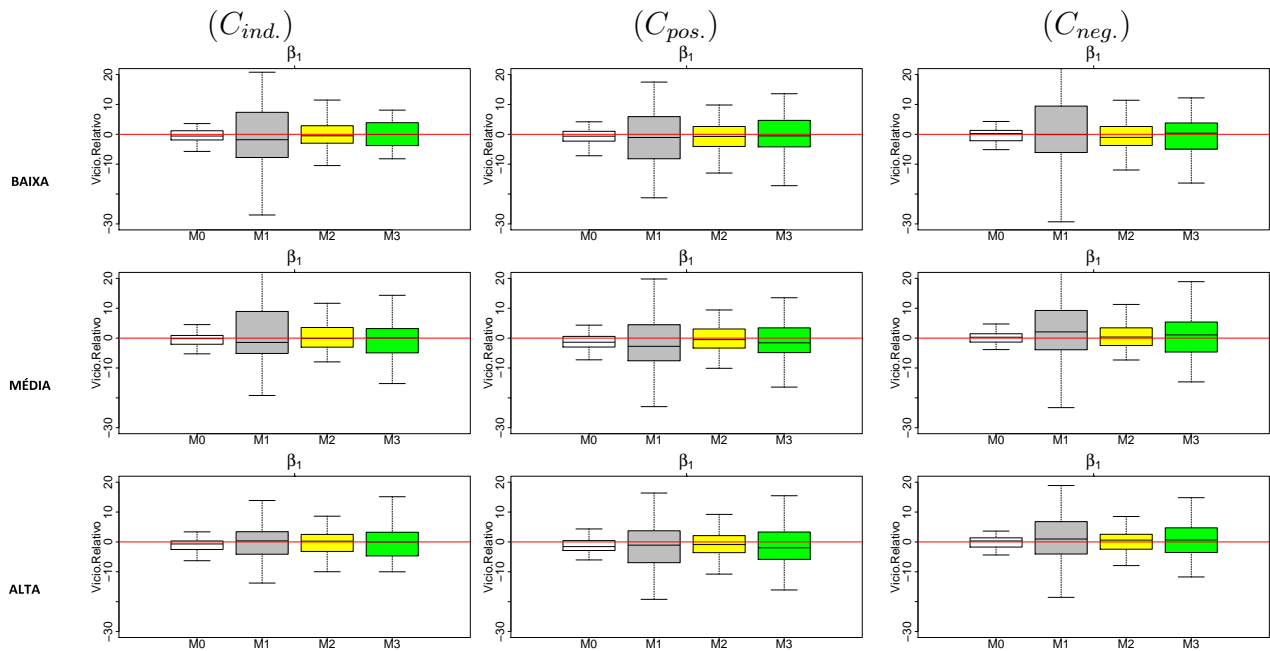


Figura 5.5: *Boxplots dos vícios relativos para β_1 obtidos em cada replicação do MCMC para ajustes considerando amostras de 50 tempos.*

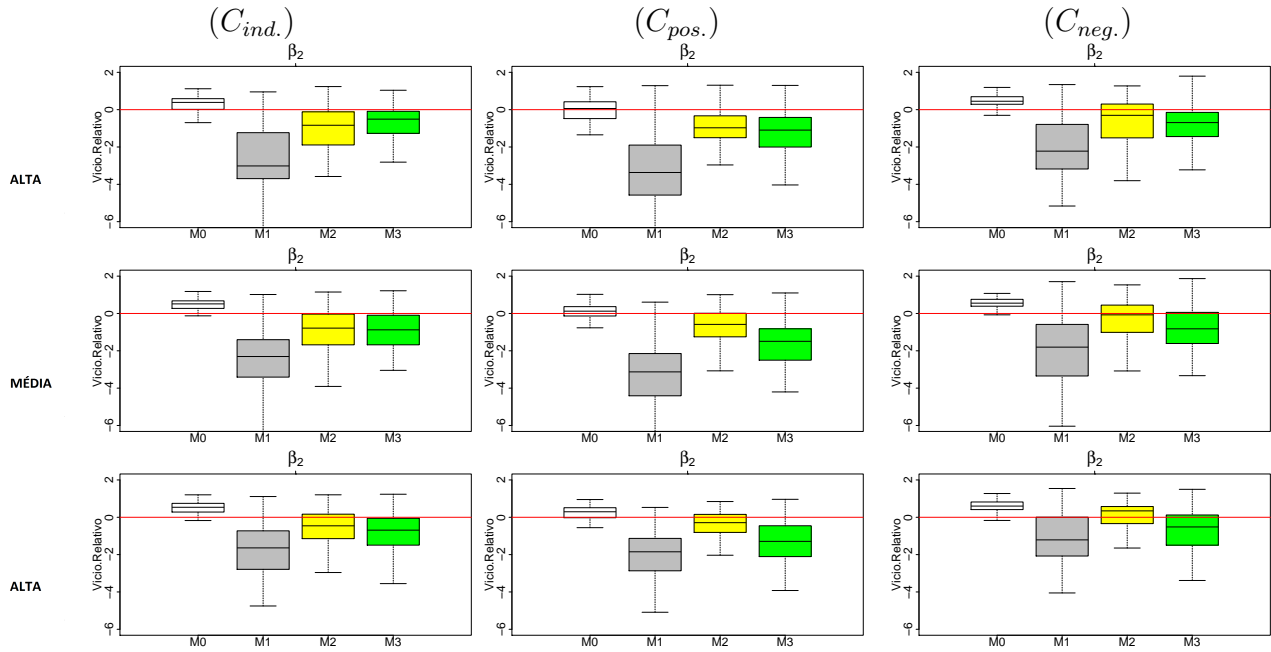


Figura 5.6: *Boxplots* dos vícios relativos para β_2 obtidos em cada replicação do MCMC para ajustes considerando amostras de 50 tempos.

maior. Os modelos que tem melhores estimativas $M0$ e $M2$. Estes modelos apesar de apresentarem um pequeno vício relativo, mostraram *boxplots* mais próximos do patamar em zero, sendo que, em alguns cenários o *boxplot* do modelo $M2$ continha o valor zero. O modelo $M0$ subestima o valor real do parâmetro α nos cenários analisados, porém o vício é bem pequeno (em geral menor que 0.5). Destacamos também que apesar de possuir, geralmente, menor vício, a amplitude dos intervalos HPD de $M0$ são as menores (Figura 5.2) e, conseqüentemente, a cobertura destes intervalos é insuficiente deixando o valor real fora da região de 95 % de probabilidade. Os demais modelos mostram superestimação na maioria dos cenários. Apesar do modelo $M2$ apresentar nas estimativas do cenário independente um vício relativo pouco maior que o observado nos demais cenários, a alteração do tipo de correlação e do percentual de falha não parecem influenciar no desempenho do modelo.

Em relação ao parâmetro β_0 , novamente é evidente o pior desempenho nas estimativas produzidas pelo modelo $M1$ que apresenta o maior vício relativo observado. O modelo $M3$ também demonstra grande vício relativo, subestimando o valor real de β_0 em todos

os cenários e proporções de falha. De forma inversa ao que ocorreu para α , o modelo $M0$, para este tamanho de amostra, tem estimativas pouco maiores que o valor real de β_0 , mostrando nos *boxplots* vícios relativos bem pequenos. A exceção do cenário independente, o *boxplot* do modelo $M2$ indica que suas estimativas são muito similares as obtidas por $M0$, tendo maior variabilidade - a amplitude do *boxplot* para o modelo $M2$ é maior que a observada em $M0$ - em relação as estimativas para β_0 . Em alguns cenários, $M2$ apresenta melhores resultados que $M0$. Seu melhor desempenho é observado em relação ao parâmetro β_0 no cenário de correlação positiva com alta proporção de falha.

O parâmetro β_1 tem as melhores estimativas, sendo que todos os modelos estimam adequadamente o parâmetro sem indicação de vício. Para todos os modelos a mediana do vício relativo sobrepõe ou está muito próxima do valor zero, indicando qualidade das estimativas. Além disso, é possível perceber maior variabilidade nos *boxplots*. Este é o parâmetro para qual se obtêm as melhores estimativas em todos os modelos e cenários.

Considerando β_2 , novamente identificamos estimativas viesadas, principalmente em relação a $M1$ e $M3$. O modelo $M1$ subestima o valor real do parâmetro apresentando mediana de aproximadamente -3 . Já o modelo $M3$ apresenta estimativas viesadas com mediana de aproximadamente -1 , sendo que os *boxplots* para esse modelo estão muito próximos ou incluem o valor zero. O modelo $M0$, que apresenta *boxplots* com menor amplitude, também indica viés nas estimativas superestimando β_2 . Este modelo apresenta melhor desempenho para o caso de correlação positiva e proporções de falha baixa e média. O modelo $M2$ tem o melhor desempenho em relação as estimativas para β_2 estimando corretamente no cenário de correlação negativa em todas as proporções de falha e nos cenários independente e correlação positiva para grandes proporções de falha.

A média do vício relativo, disponível na Tabela 5.2, confirma que $M1$ apresenta maior vício para todos os parâmetros e cenários analisados, sendo $M0$ o modelo com menor vício relativo. O modelo $M2$, para este tamanho de amostra, também apresenta vício relativamente pequeno tendo desempenho próximo ao observado para modelo $M0$.

Os gráficos de vício relativo oferecem uma visão geral da qualidade da estimação dos modelos em relação a cada parâmetro. No entanto, os *boxplots* dos vícios relativos obtidos nas replicações do MCMC ignoram a comparação de vícios dos modelos para uma mesma

amostra gerada em cada replicação. Esta informação é perdida na representação gráfica. Para corrigir este problema, optamos por avaliar o valor absoluto da razão de vícios de dois modelos (RV) para uma mesma amostra gerada em uma replicação do MCMC. Esse calculo é realizado pela fórmula dada em (5.5).

A análise é realizada de forma que, quando o modelo do denominador tiver desempenho semelhante ao obtido pelo modelo no numerador, a razão entre os vícios será próxima de 1. Por outro lado, se o modelo do denominador apresentar melhor desempenho então a razão entre os vícios relativos será maior do que 1; quando a razão entre vícios for menor do que 1, o vício relativo do modelo no denominador é maior que do numerador.

Tabela 5.3: Razão de Vícios Relativos

RV	%	parâmetro	$C_{ind.}$			$C_{pos.}$			$C_{neg.}$		
			$M0$	$M1$	$M3$	$M0$	$M1$	$M3$	$M0$	$M1$	$M3$
VR_{M_i}/VR_{M_2}	baixa	α	0.22	1.00	0.38	0.34	0.92	0.44	0.10	0.96	0.64
		β_0	0.18	0.88	0.44	0.26	0.96	0.58	0.22	0.72	0.50
		β_1	0.20	0.94	0.54	0.18	0.82	0.66	0.22	0.80	0.56
		β_2	0.30	0.90	0.32	0.18	0.94	0.58	0.40	0.74	0.46
	média	α	0.18	0.94	0.50	0.46	0.94	0.64	0.30	0.92	0.84
		β_0	0.20	0.86	0.50	0.42	0.90	0.72	0.32	0.70	0.68
		β_1	0.08	0.86	0.58	0.26	0.82	0.66	0.16	0.86	0.68
		β_2	0.32	0.86	0.48	0.26	0.92	0.72	0.56	0.74	0.64
	alta	α	0.36	0.96	0.62	0.64	0.92	0.84	0.40	0.94	0.92
		β_0	0.34	0.94	0.72	0.42	0.92	0.82	0.36	0.68	0.68
		β_1	0.22	0.72	0.52	0.26	0.78	0.70	0.16	0.76	0.74
		β_2	0.48	0.78	0.60	0.46	0.82	0.78	0.70	0.72	0.60
VR_{M_i}/VR_{M_3}	baixa	α	0.20	0.98	0.62	0.42	0.94	0.56	0.06	0.96	0.36
		β_0	0.32	0.94	0.56	0.22	0.94	0.42	0.32	0.74	0.50
		β_1	0.16	0.90	0.46	0.16	0.76	0.34	0.14	0.76	0.44
		β_2	0.42	0.90	0.68	0.24	0.94	0.42	0.38	0.74	0.54
	média	α	0.16	0.98	0.50	0.20	0.88	0.36	0.02	0.80	0.16
		β_0	0.22	0.86	0.50	0.20	0.90	0.28	0.14	0.58	0.32
		β_1	0.06	0.78	0.42	0.16	0.78	0.34	0.10	0.60	0.32
		β_2	0.38	0.92	0.52	0.10	0.86	0.28	0.40	0.64	0.36
	alta	α	0.20	0.88	0.38	0.12	0.58	0.16	0.02	0.60	0.08
		β_0	0.20	0.82	0.28	0.20	0.68	0.18	0.22	0.52	0.32
		β_1	0.22	0.74	0.48	0.22	0.58	0.30	0.10	0.54	0.26
		β_2	0.42	0.76	0.40	0.26	0.64	0.22	0.44	0.68	0.40

Para facilitar as análises, calculamos a proporção de vezes que a razão de vícios relativos foi maior que 1 nas comparações entre demais modelos com $M2$ e $M3$. Os resultados são mostrados na Tabela 5.3. Nas comparações entre $M0$ e $M3$, observamos que a razão entre vícios produz poucos valores acima de 1, o que sugere melhor desempenho de $M0$ em relação a $M3$, ou seja, o vício relativo nas estimativas apresentadas por $M3$ é maior que o vício relativo das estimativas obtidas com o modelo $M0$.

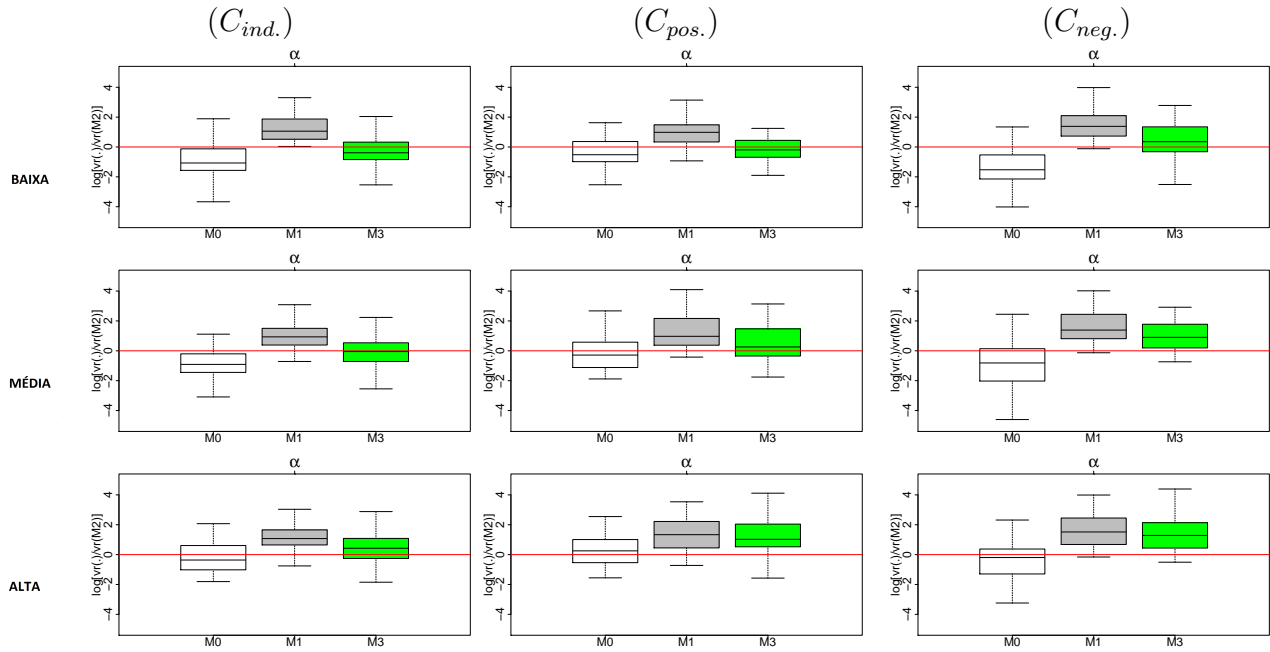


Figura 5.7: *Boxplots dos valores do log do valor absoluto da razão de vícios relativos de α obtidos nas 50 replicações do MCMC para ajustes considerando amostras de 50 tempos. O modelo de referência no denominador da razão é o M2.*

Na proporção de falha pequena no cenário de correlação positiva e negativa e na proporção de falha média no cenário independente, note que, em relação a α o modelo $M2$ apresenta resultados semelhantes aos obtidos via $M3$, sendo que a proporção de vício relativo entre esses modelos é próxima ou pouco inferior a 0.5. Em relação aos β s, $M3$ é competitivo no cenário independente, entretanto o desempenho de $M2$ é um pouco superior. O melhor desempenho de $M3$ em relação a $M2$ pode ser observado para baixa proporção de falha nos parâmetros α e β_2 para os quais a razão entre os vícios foi maior que 1 para aproximadamente 60% das amostras. Para os demais cenários e proporções de falha, $M2$ tem melhor desempenho. Os resultados obtidos pelo modelo $M3$ são melhores apenas em relação ao modelo $M1$ para todos os parâmetros e proporções de falha, visto que mais de 50% da razão de vício entre esse modelos mostraram resultados superiores a 1.

Na comparação entre $M2$ com $M1$, verificamos, também melhor desempenho de $M2$ que apresenta proporção de vício sempre superiores a 0.7. Entretanto, na comparação

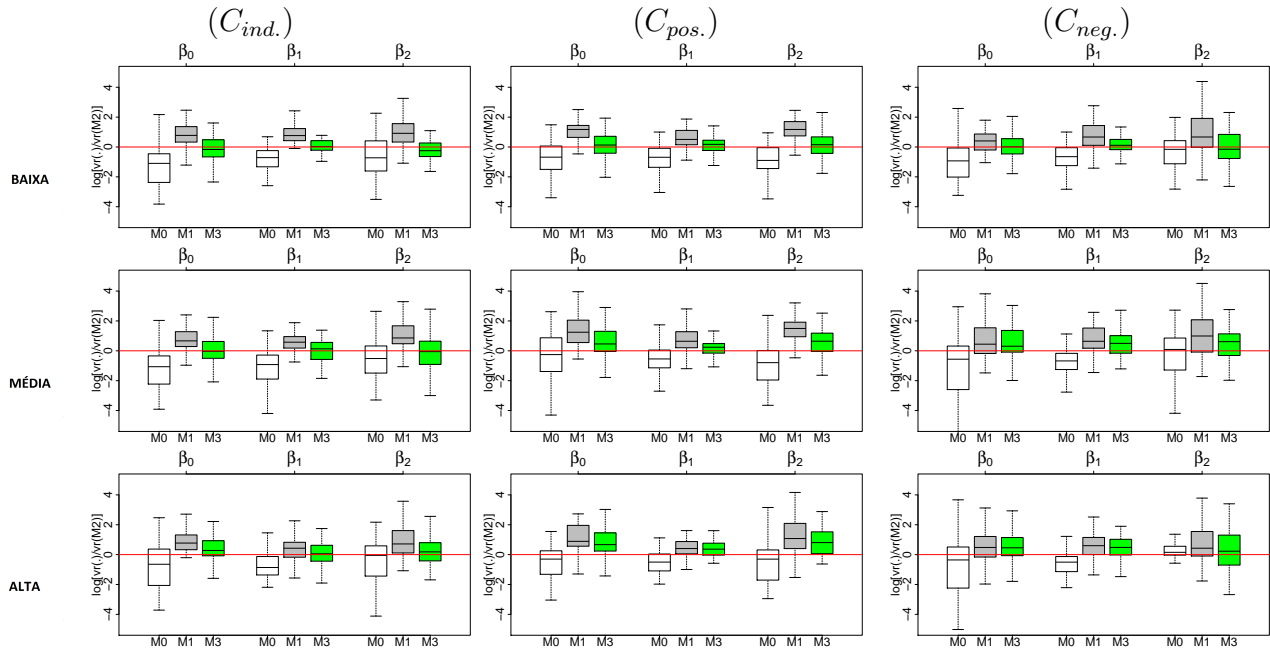


Figura 5.8: *Boxplots dos valores do log do valor absoluto da razão de vícios relativos de β obtidos nas 50 replicações do MCMC para ajustes considerando amostras de 50 tempos. O modelo de referência no denominador da razão é o M2.*

com $M0$, para a maioria dos cenários o modelo $M0$ mostra melhor desempenho. Isto pode ser verificado nas Figuras 5.7 e 5.8 que mostram o log da razão entre os vícios relativos dos demais modelos em comparação a $M2$. Neste gráfico, um traço horizontal demarcando o ponto $\log(1) = 0$ foi adicionado para orientar as análises.

Em ambas as figuras, é evidente o melhor desempenho de $M2$ em relação aos resultados obtidos por $M1$. Comparando-se os modelos $M2$ e $M3$, observa-se que os *boxplots* estão acima do traço em 0 na maioria dos cenários. Isso indica que $M2$ produz amostras menos viesadas. Para proporção de falha pequena os dois modelos ($M2$ e $M3$) demonstram ter estimativas semelhantes visto que os *boxplots*, além da pequena amplitude, incluem o valor 0. Considerando a comparação entre $M0$ e $M2$, pode-se dizer que $M0$ é competitivo em relação a $M2$ para os cenários de correlação positiva e negativa, apresentando sempre desempenho igual ou pouco superior. No cenário independente, o modelo $M0$ tende a apresentar melhores resultados, a exceção da proporção de falha grande para α , na qual não há indício de diferença nos resultados produzidos pelos dois modelos.

Outra exceção é β_2 , para o qual o resultado produzido por cada modelo é equivalente.

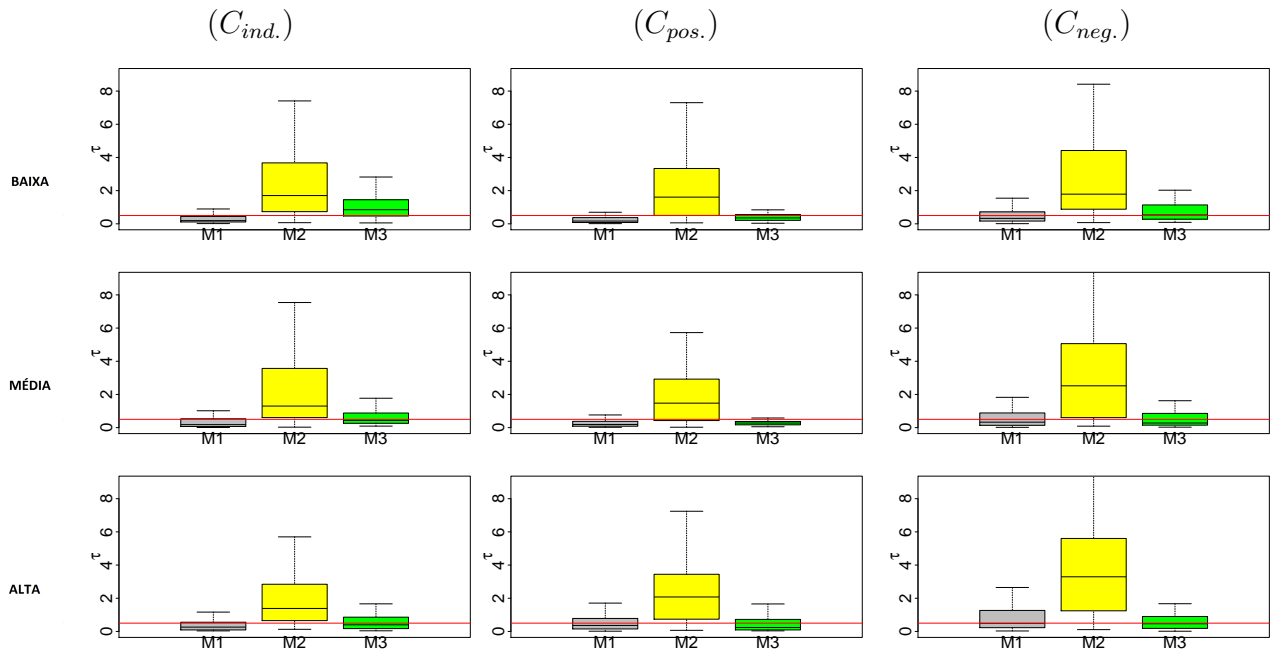


Figura 5.9: *Boxplots* representando as médias *a posteriori* de τ obtidas em cada replicação do MCMC e considerando amostras de 50 tempos. Tração sobre o valor real $\tau = 0.2$.

Observe que em todas as comparações realizadas acima, o modelo $M0$ se mostrou competitivo em relação aos modelos $M2$ e $M3$, mesmo com a existência de correlação entre os tempos de falha e censura. Vale lembrar que os modelos $M2$ e $M3$ são mais complexos e possuem maior número de parâmetros que $M0$. Além disso, $M2$ e $M3$ tem a vantagem de classificar a amostra de acordo com o tipo de correlação.

A Figura 5.9 mostra os *boxplots* das médias amostrais *a posteriori* para τ fornecida por cada modelo que inclui a fragilidade ($M1$, $M2$ e $M3$). É importante lembrar que o parâmetro τ capta a variabilidade extra dos dados, de forma que quanto menor seu valor, maior a variabilidade extra existente. Note que as estimativas para τ obtidas por $M1$ e $M3$ são muito similares, sendo que $M3$ produz estimativas um pouco maiores que $M1$ para alguns cenários. Entretanto os dois modelos parecem captar bem o valor real do parâmetro.

Ainda analisando os gráficos da Figura 5.9, veja que as estimativas obtidas para $M2$ merecem destaque, visto que a estimativa do parâmetro τ produzida por esse modelo é

sempre muito superior à produzida por $M1$ e $M3$.

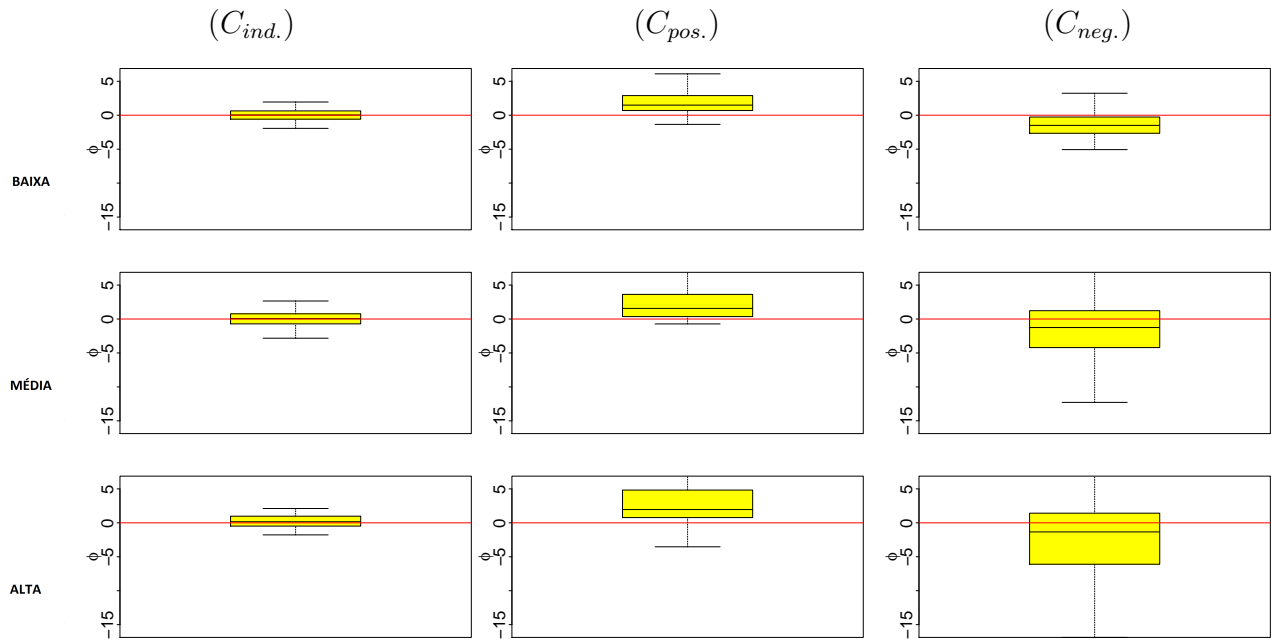


Figura 5.10: *Boxplots* representando as médias a posteriori de ϕ obtidas em cada replicação do MCMC e considerando amostras de 50 tempos.

Salientamos que, para o modelo $M2$, na distribuição dos tempos de censura, a fragilidade é multiplicada por ϕ . Este produto tem relação direta na estimação de τ uma vez que, $\phi N(0, 1/\tau)$ implica em $N(0, \phi^2/\tau)$ e pode, portanto, interferir na estimação.

A Figura 5.10 descreve as estimativas obtidas para ϕ através de $M2$. Este parâmetro tem o objetivo de classificar o mecanismo de censura de acordo com o tipo de correlação entre os tempos de falha e censura. Uma linha horizontal é traçada sobre o ponto zero para orientar o leitor quanto à posição dos *boxplots*. Os gráficos mostram que ϕ indica adequadamente o tipo de correlação existente entre os tempos de falha e censura para os cenários de independência e correlação positiva. Observe que em $C_{pos.}$ as estimativas estão claramente acima de zero. No cenário $C_{ind.}$, as estimativas estão concentradas em torno de zero, indicando a independência como adequado ao cenário.

Alguns problemas surgem na avaliação do $C_{neg.}$ para as proporções de falha média e grande. Nesses casos, o intervalo interquartil do *boxplot* contem o valor zero: isto sugere incerteza sobre a classificação, entretanto note que a mediana está sempre abaixo

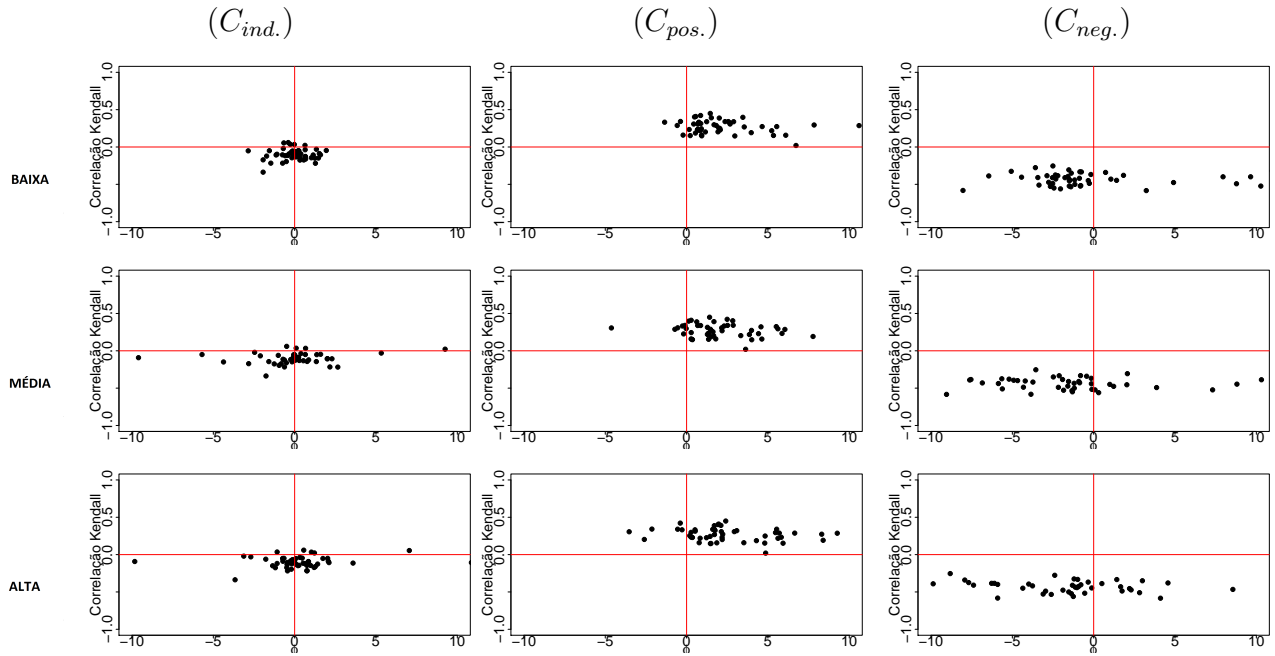


Figura 5.11: Gráfico de dispersão de ϕ contra a correlação Kendall obtidas em cada replicação do MCMC e considerando amostras de 50 tempos.

de zero mostrando que a correlação negativa é percebida pelo modelo.

Para uma comparação entre a correlação existente nos dados e ϕ apresentamos a Figura 5.11. No eixo vertical está a correlação, enquanto no eixo horizontal temos a variação de ϕ . Linhas horizontais e verticais marcam os quadrantes em cada gráfico. Note que, no cenário independente existe uma concentração de pontos na região central do gráfico, indicando que ϕ captou adequadamente a independência da amostra. Para o cenário de correlação positiva, os pontos estão concentrados no primeiro quadrante, ou seja, existe uma correlação positiva e que, para maioria das amostras foi corretamente identificada por ϕ . No cenário de correlação negativa, observamos que apesar da maioria das amostras mostrarem sua correlação indicada adequadamente alguns valores de ϕ estão situadas no quarto quadrante, onde ϕ é positivo. Esses valores acabam por influenciar a estimação do parâmetro gerando a incerteza mostrada nos *boxplots* da Figura 5.10.

No modelo $M3$ o hiperparâmetro $\mathbf{q} = (q_1, q_2, q_3)$ é responsável por indicar o tipo de correlação existente entre os tempos de falha e censura. A estimativa desse parâmetro é apresentada na Figura 5.12. Uma linha sobre o ponto 0.5 foi traçada de forma a

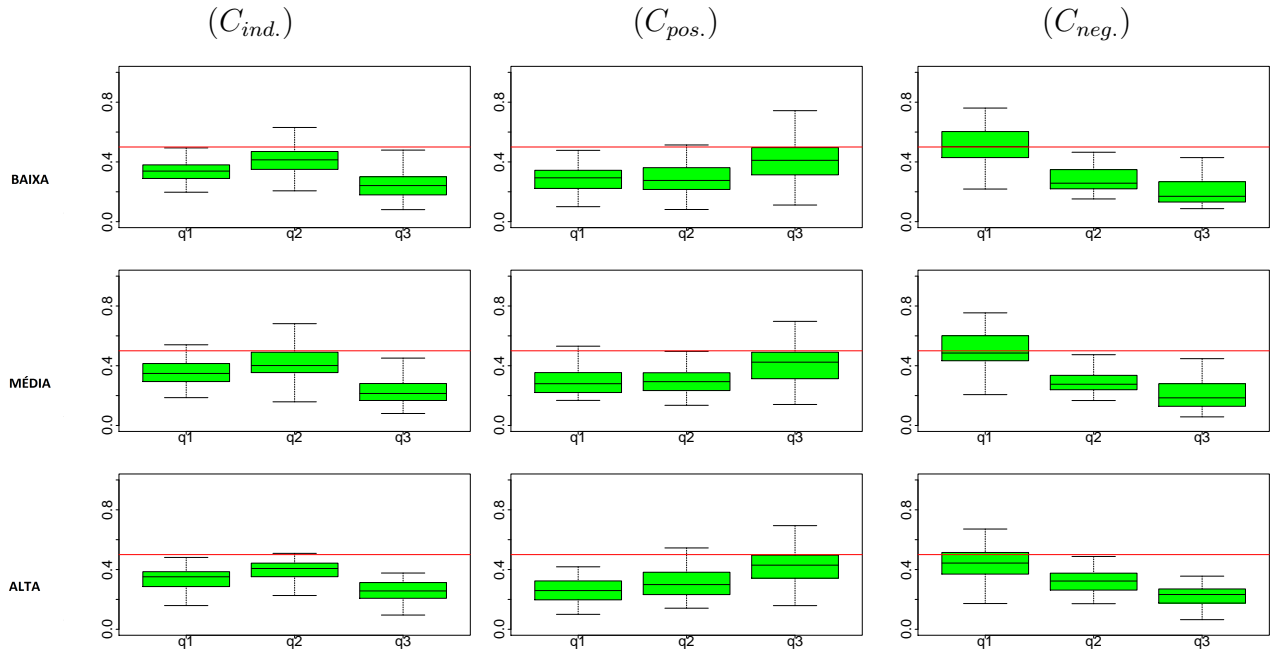


Figura 5.12: *Boxplots* representando as médias a posteriori de \mathbf{q} obtidas em cada replicação do MCMC e considerando amostras de 50 tempos.

orientar as análises. Os detalhes sobre a modelagem estão descritas nas Seções 5.1 e 6.1.3, porém vale lembrar que as componentes q_1 , q_2 e q_3 identificam os cenários de correlação negativa, independência e correlação positiva, respectivamente. Para este tamanho de amostra, o modelo $M3$ se mostrou mais eficiente na classificação dos dados quanto ao tipo de correlação existente. Para todos os gráficos é possível perceber que o maior componente q_j é justamente aquele que representa o cenário real dos dados; em outras palavras, $M3$ indica corretamente o cenário de associação. Este parâmetro parece não sofrer influência com o aumento da proporção de falha. Esse modelo também possui a vantagem de classificar a amostra sem influenciar as estimativas de τ .

Finalizando as análises, apresentaremos os resultados de dois critérios de comparação de modelos. A Figura 5.13 mostra o critério DIC; o cálculo desta medida considerou, para todos os modelos, apenas os parâmetros relacionados à distribuição de falha, descartando os parâmetros relacionados à censura. O modelo que apresentar o menor DIC seria, comparativamente, melhor que os demais. Ressaltamos que, apesar de descartar a parte da verossimilhança associada aos parâmetros de censura, o modelo $M0$ ainda permanece

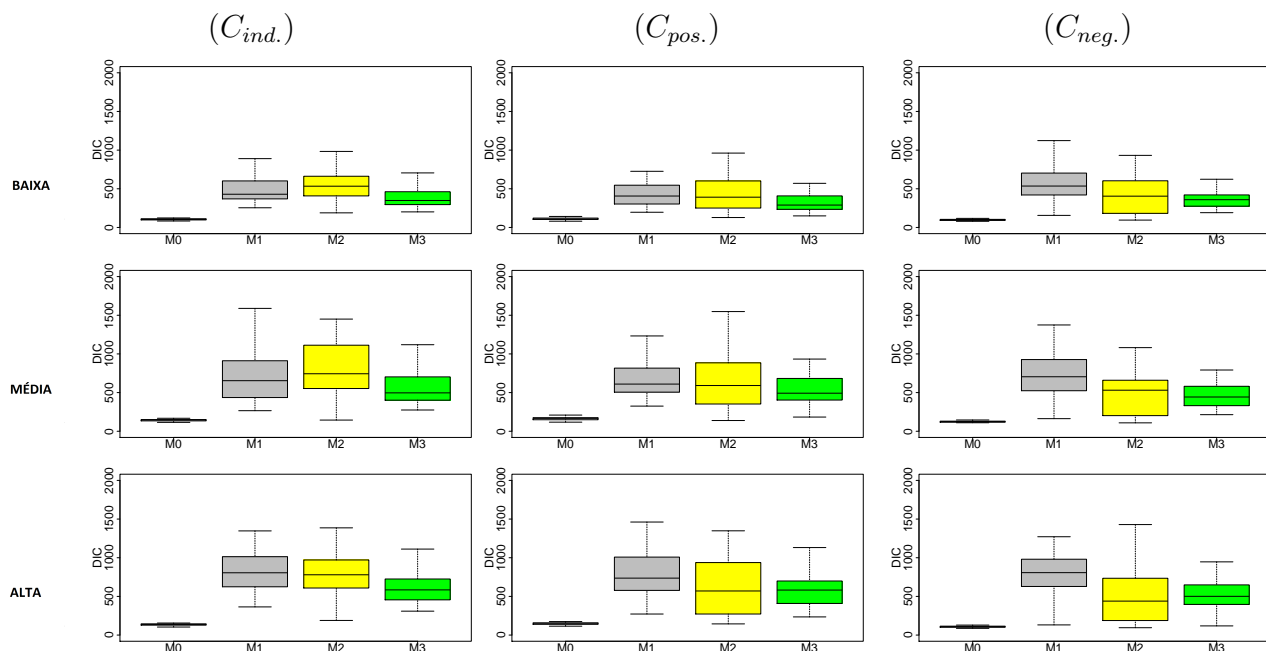


Figura 5.13: *Boxplots* representando os valores do DIC obtidos em cada replicação do MCMC e considerando amostras de 50 tempos.

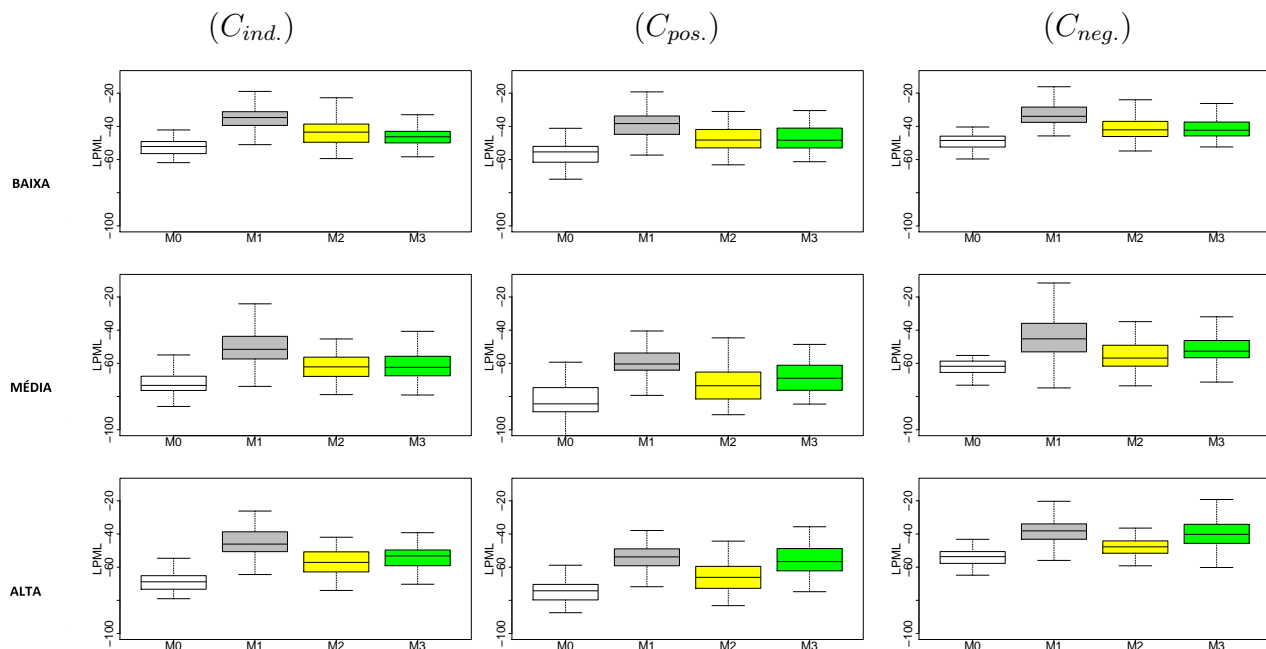


Figura 5.14: *Boxplots* representando os valores do LPML obtidos em cada replicação do MCMC e considerando amostras de 50 tempos.

com a grande vantagem de ser parcimonioso visto que não inclui a fragilidade. A presença de efeitos aleatórios distorce a comparação de $M0$ com os demais modelos e isso será observado em nossas análises a seguir.

O bom desempenho do modelo $M0$ nas estimativas dos parâmetros da distribuição dos tempos de falha é evidenciado no critério DIC, onde os *boxplots* para esse modelo são menores em todos os cenários. Comparando o desempenho dos modelos $M1$, $M2$ e $M3$, percebemos que para o cenário independente, $M3$ mostra o melhor resultado, com *boxplots* menores que os dois outros modelos. Para os cenários de correlação positiva e negativa, observamos que $M2$ tem maior amplitude nos *boxplots* e atinge valores menores. Entretanto sua mediana está muito próxima do modelo $M3$.

Considerando o critério LPML, representado na Figura 5.14, temos a indicação de $M1$ como melhor modelo. Entretanto, como apresentado ao longo desta seção, este modelo mostrou estimativas viesadas e piores para todos os parâmetros. Observe que, para este critério, o modelo $M0$, antes classificado como o melhor, agora se mostra como a pior qualificação. Em relação aos modelos $M2$ e $M3$ observamos que, em geral, o modelo $M3$ apresenta desempenho melhor ou similar nos cenários de correlação positiva e negativa. No cenário independente, $M2$ apresenta melhor desempenho quando a proporção de falha é baixa. Para as demais proporções de falha, $M2$ e $M3$ possuem performances semelhantes.

Tabela 5.4: DIC e LPML.

%	modelo	$C_{ind.}$		$C_{pos.}$		$C_{neg.}$	
		DIC	LPML	DIC	LPML	DIC	LPML
baixa	M0	103.45	-52.13	109.84	-55.40	95.48	-48.49
	M1	429.37	-34.69	406.26	-38.32	535.33	-33.95
	M2	533.37	-43.48	391.43	-48.25	404.54	-42.09
	M3	348.69	-46.29	289.39	-48.35	358.79	-42.32
média	M0	145.11	-73.28	166.48	-84.41	123.04	-61.75
	M1	654.02	-51.55	609.93	-60.38	704.74	-45.23
	M2	743.59	-62.12	591.36	-73.50	530.79	-56.82
	M3	496.47	-62.40	492.13	-68.95	442.92	-52.67
alta	M0	135.68	-68.83	146.11	-74.25	106.27	-53.65
	M1	805.51	-46.10	736.28	-53.83	807.15	-38.13
	M2	779.67	-57.15	570.18	-66.13	439.04	-47.75
	M3	584.31	-53.22	581.80	-56.65	500.87	-40.22

A Tabela 5.4 traz as medianas dos DIC e LPML para 50 replicações MCMC. É possível perceber que a diferença entre as medianas obtidas pelo critério LPML para $M2$ e $M3$ é

bem pequena. Verificamos que o melhor modelo, segundo esse critério é $M3$ pois possui menor valor de LPML na maioria dos cenários. Entretanto, considerando o vício relativo e a razão entre vício, vimos que $M2$ produz estimativas menos viesadas, sendo o modelo preferível para este tamanho de amostra.

5.2.2 Resultados para amostra de tamanho 250.

Tabela 5.5: Estimativas para amostra tamanho 250

cenário	% falha	modelo	α			β_0			β_1			β_2		
			VR	EP	PC	VR	EP	PC	VR	EP	PC	VR	EP	PC
$C_{ind.}$	Pequeno	M0	-0.27	0.09	0.02	0.22	0.14	0.32	-0.09	0.10	0.96	0.32	0.23	0.42
		M1	1.70	1.68	0.70	-1.11	1.25	0.70	2.07	0.48	0.96	-1.90	2.05	0.80
		M2	0.58	0.98	0.94	-0.43	0.83	0.94	0.75	0.29	0.98	-0.69	1.25	0.96
		M3	0.10	0.44	0.90	-0.10	0.43	0.90	0.25	0.18	0.98	-0.14	0.62	0.96
	Média	M0	-0.33	0.07	0.00	0.22	0.14	0.24	-0.09	0.09	0.92	0.40	0.19	0.14
		M1	1.58	1.45	0.58	-0.97	1.04	0.68	1.89	0.43	0.96	-1.75	1.74	0.64
		M2	0.66	1.00	0.94	-0.45	0.79	0.96	0.86	0.29	0.98	-0.76	1.24	0.96
		M3	0.22	0.49	0.86	-0.20	0.45	0.92	0.35	0.19	0.96	-0.28	0.68	0.96
	Grande	M0	-0.38	0.09	0.00	0.24	0.13	0.16	-0.18	0.08	0.88	0.43	0.16	0.08
		M1	1.15	1.88	0.74	-0.70	0.82	0.74	1.29	0.34	0.98	-1.22	1.30	0.70
		M2	0.48	1.26	0.90	-0.33	0.61	0.96	0.66	0.23	0.98	-0.49	0.87	0.92
		M3	0.28	0.84	0.90	-0.24	0.45	0.92	0.46	0.19	0.98	-0.31	0.63	0.96
$C_{pos.}$	Pequeno	M0	-0.42	0.07	0.00	0.22	0.14	0.24	-0.34	0.10	0.94	0.03	0.27	0.94
		M1	1.39	1.39	0.74	-1.31	1.28	0.58	1.16	0.52	0.94	-2.92	2.35	0.48
		M2	0.33	0.66	0.84	-0.18	0.50	0.84	0.26	0.21	0.98	-0.27	0.66	0.98
		M3	0.17	0.43	0.96	-0.20	0.39	0.92	0.11	0.19	0.98	-0.39	0.55	0.88
	Média	M0	-0.43	0.05	0.00	0.25	0.14	0.14	-0.38	0.09	0.90	0.07	0.20	0.90
		M1	1.33	1.30	0.66	-1.01	1.05	0.64	1.31	0.44	0.96	-2.65	2.10	0.42
		M2	0.51	0.70	0.74	-0.27	0.52	0.80	0.67	0.22	0.96	-0.35	0.67	0.90
		M3	0.32	0.49	0.86	-0.28	0.44	0.90	0.20	0.20	0.98	-0.45	0.58	0.88
	Grande	M0	-0.42	0.08	0.00	0.24	0.13	0.10	-0.37	0.08	0.94	0.20	0.16	0.52
		M1	0.84	1.47	0.74	-0.64	0.72	0.76	0.89	0.28	0.98	-1.51	1.26	0.64
		M2	0.29	0.95	0.84	-0.14	0.43	0.90	0.59	0.19	0.98	-0.26	0.60	0.90
		M3	0.33	0.79	0.88	-0.26	0.42	0.92	0.61	0.19	0.98	-0.43	0.57	0.92
$C_{neg.}$	Pequeno	M0	-0.11	0.11	0.54	0.24	0.14	0.22	0.05	0.10	0.92	0.43	0.21	0.18
		M1	0.83	0.98	0.90	-0.26	0.61	0.86	0.57	0.27	0.96	-0.43	0.99	0.90
		M2	0.43	0.68	0.88	-0.33	0.66	0.86	0.47	0.24	0.96	-0.44	0.93	0.88
		M3	0.24	0.40	0.92	-0.14	0.48	0.98	0.39	0.19	0.96	-0.20	0.63	0.92
	Média	M0	-0.18	0.08	0.06	0.23	0.14	0.20	0.16	0.09	0.94	0.48	0.18	0.08
		M1	0.96	1.07	0.86	-0.34	0.67	0.80	0.85	0.28	0.96	-0.58	1.00	0.76
		M2	0.49	0.76	0.90	-0.34	0.67	0.90	0.68	0.24	0.98	-0.53	1.00	0.82
		M3	0.22	0.43	0.94	-0.14	0.48	0.96	0.54	0.18	0.96	-0.14	0.65	0.94
	Grande	M0	-0.25	0.11	0.00	0.23	0.13	0.18	0.11	0.08	0.96	0.51	0.17	0.02
		M1	0.66	1.46	0.86	-0.20	0.53	0.86	0.85	0.22	0.96	-0.30	0.84	0.82
		M2	0.31	1.05	0.90	-0.26	0.58	0.94	0.61	0.20	0.98	-0.30	0.81	0.94
		M3	0.29	0.83	0.88	-0.17	0.50	0.94	0.58	0.18	0.98	-0.17	0.66	0.86

VR = vício relativo - EP = erro padrão - PC = probabilidade de cobertura

Nesta subseção, iremos explorar os mesmos tipos de resultados investigados anteriormente para amostras de 50 tempos. A diferença é que agora aumentamos o tamanho amostral para 250. Visto que o conjunto dados é maior, espera-se que a identificação do

tipo de associação e as estimativas sejam mais precisas. Ressaltamos que nesta análise continuaremos a considerar 50 replicações do MCMC.

A média da probabilidade de cobertura, erro padrão e vício relativo são apresentadas na Tabela 5.5. Para a amostra de tamanho 250, observamos que os modelos com probabilidade de cobertura próximas de 95% são $M2$ e $M3$. O modelo $M1$ também apresenta alguns bons resultados, principalmente em relação ao parâmetro β_1 . O modelo $M0$, entretanto, mostra os piores resultados, sendo que para alguns cenários o valor verdadeiro de α não é coberto pelo intervalo de credibilidade. Este modelo também não apresenta bons resultados para os β s, sendo exceção β_1 para o qual $M0$ mostra probabilidades de cobertura próximas ao nominal. Ao compararmos os resultados para as amostras de tamanho 50 e 250, verificamos que a probabilidade de cobertura produzida por todos os modelos sofre uma pequena redução com o aumento da amostra, sendo mais acentuada para os modelos $M0$ e $M1$. Isso se deve ao fato de termos uma amostra maior e consequentemente mais informação disponível.

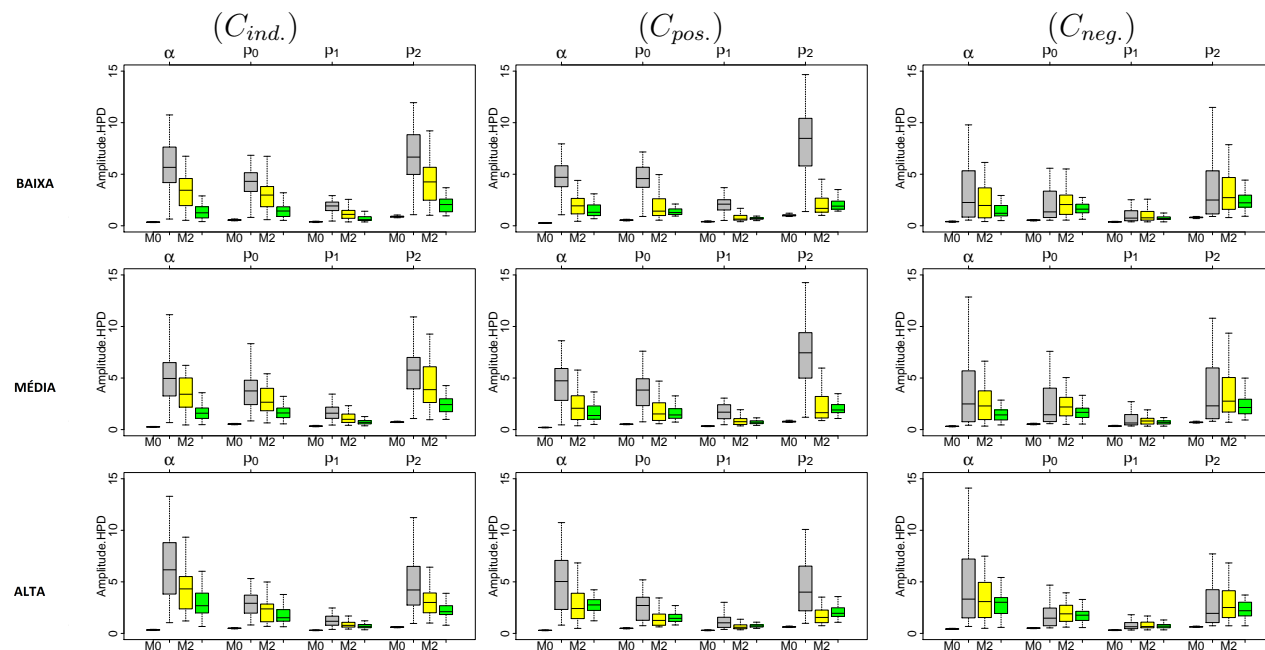


Figura 5.15: Amplitude do intervalo HPD registrada nas 50 replicações do MCMC para o ajuste de uma amostra de 250 tempos.

A Figura 5.15 apresenta os boxplots das amplitudes dos intervalos HPD de 95% ob-

Tabela 5.6: Razão de Vícios Relativos - amostra de 250 tempos

RV	%	parâmetro	$C_{ind.}$			$C_{pos.}$			$C_{neg.}$		
			M0	M1	M3	M0	M1	M3	M0	M1	M3
VR_{M_i}/VR_{M_2}	baixa	α	0.42	0.84	0.28	0.74	0.86	0.30	0.32	0.56	0.42
		β_0	0.54	0.80	0.36	0.38	0.98	0.44	0.54	0.60	0.30
		β_1	1.00	1.00	0.60	1.00	1.00	0.32	1.00	1.00	0.28
		β_2	0.00	0.02	0.74	0.00	0.04	0.48	0.00	0.02	0.14
	média	α	0.34	0.80	0.20	0.66	0.74	0.26	0.40	0.56	0.40
		β_0	0.54	0.70	0.24	0.38	0.88	0.44	0.70	0.50	0.30
		β_1	1.00	1.00	0.80	1.00	1.00	0.36	1.00	1.00	0.28
		β_2	0.00	0.02	0.88	0.04	0.10	0.58	0.00	0.00	0.18
	alta	α	0.58	0.76	0.38	0.72	0.66	0.58	0.56	0.48	0.44
		β_0	0.74	0.70	0.40	0.70	0.74	0.46	0.82	0.70	0.50
		β_1	1.00	1.00	0.68	1.00	1.00	0.30	1.00	1.00	0.30
		β_2	0.00	0.00	0.86	0.00	0.02	0.60	0.00	0.02	0.18
RV	%	parâmetro	M0	M1	M2	M0	M1	M2	M0	M1	M2
VR_{M_i}/VR_{M_3}	baixa	α	0.72	0.88	0.72	0.80	0.98	0.70	0.42	0.56	0.58
		β_0	0.88	0.90	0.64	0.46	1.00	0.56	0.86	0.76	0.70
		β_1	1.00	1.00	0.40	1.00	1.00	0.68	1.00	1.00	0.72
		β_2	0.00	0.02	0.26	0.00	0.02	0.52	0.00	0.02	0.86
	média	α	0.68	0.88	0.80	0.76	0.84	0.74	0.50	0.62	0.60
		β_0	0.84	0.84	0.76	0.54	0.94	0.56	0.90	0.70	0.70
		β_1	1.00	1.00	0.20	1.00	1.00	0.64	1.00	1.00	0.72
		β_2	0.00	0.02	0.12	0.02	0.06	0.42	0.00	0.02	0.82
	alta	α	0.68	0.78	0.62	0.70	0.72	0.42	0.52	0.56	0.56
		β_0	0.80	0.82	0.60	0.58	0.78	0.54	0.88	0.66	0.50
		β_1	1.00	1.00	0.32	1.00	1.00	0.70	1.00	1.00	0.70
		β_2	0.00	0.00	0.14	0.00	0.02	0.40	0.02	0.04	0.82

tidos nas 50 replicações do MCMC. Novamente, observamos que o modelo $M0$ tem os intervalos com menor amplitude, enquanto o modelo $M1$ produz os maiores intervalos. Essas informações também são verificadas na Tabela 5.5 em que $M0$ e $M1$ têm, respectivamente, o menor e o maior erro padrão. Os intervalos de menor amplitude podem ser observados para o parâmetro β_1 , sendo que este parâmetro apresentou grande probabilidades de cobertura, indicando qualidade das estimativas em todos os modelos. O aumento no tamanho da amostra proporcionou uma redução na amplitude dos intervalos.

O vício relativo dos parâmetros da distribuição de falha para uma amostra de 250 tempos apresenta resultados similares aos ocorridos para amostra de tamanho 50. O vício calculado para as estimativas do modelo $M1$ é maior que o obtido pelos outros modelos (ver Tabela 5.5), sendo que para este tamanho de amostra a média de vício relativo é menor que a média obtida para amostras de tamanho 50, indicando melhora nas estimativas. Os resultados em geral permanecem com as mesmas características apontadas para amostra de tamanho 50.

Para α , os *boxplots* da amostra de tamanho 250 (gráficos disponíveis no Apêndice

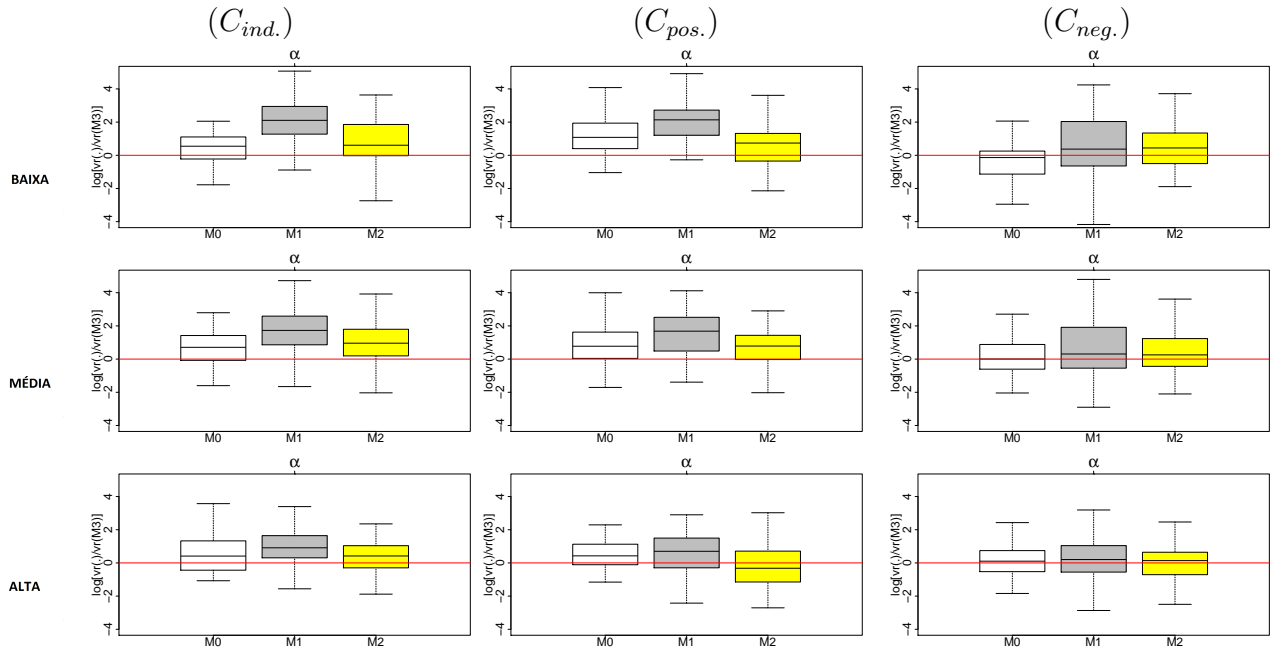


Figura 5.16: *Boxplots dos valores do log do valor absoluto da razão de vícios relativos de α obtidos nas 50 replicações do MCMC para ajustes considerando amostras de 250 tempos. O modelo de referência no denominador da razão é o M3.*

B) tem, em geral, comportamento semelhante aos observados para amostras de tamanho 50. Entretanto, notam-se pequenas diferenças como o aumento do vício relativo para o modelo $M0$, e redução do vício para os modelos $M1$, $M2$ e $M3$. Apesar da melhora nas estimativas, o modelo $M1$ continua apresentando vício em todos os cenários, sendo seus melhores resultados registrados no cenário de correlação negativa. O modelo $M3$ é mais eficiente no cenário independente, enquanto no cenário de correlação positiva e negativa, $M2$ e $M3$ tem desempenho semelhante.

Em relação ao parâmetro β_0 , temos a obtenção de estimativas um pouco menos viciadas para todos os modelos. Para grandes proporções de falha, os modelos $M2$ e $M3$ tem resultados bem semelhantes, indicando que ambos os modelos são competitivos. Resultado peculiar é observado no cenário de correlação negativa em que o modelo $M1$ obtém melhores resultados que o modelo $M2$ (até então $M2$ apresentava melhores resultados em cenários com existência de correlação) e demonstra obter estimativas semelhante a obtidas pelo modelo $M3$, apenas com maior variabilidade nos resultados.

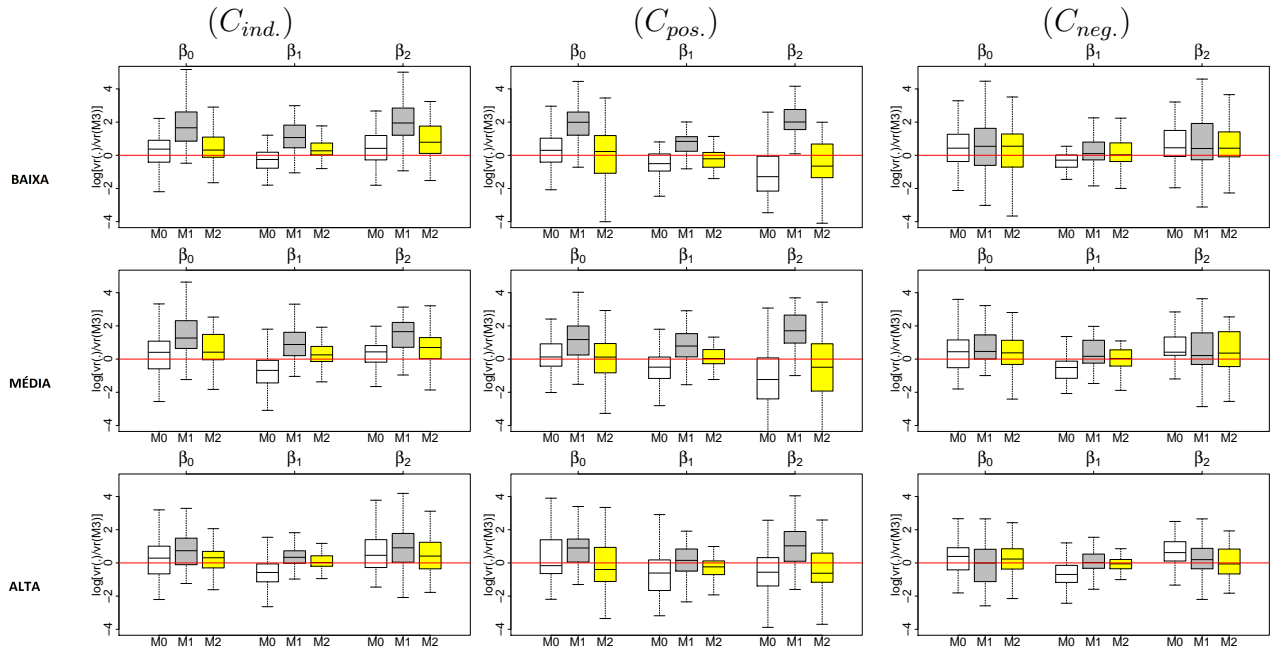


Figura 5.17: *Boxplots* dos valores do \log do valor absoluto da razão de vícios relativos de β obtidos nas 50 replicações do MCMC para ajustes considerando amostras de 250 tempos. O modelo de referência no denominador da razão é o $M3$.

Os resultados observados para β_1 e β_2 também tem similaridade com aqueles obtidos para amostras de tamanho 50. O principal ganho foi a redução da variabilidade do vício relativo. Para o cenário independente com proporção de falha média β_1 se destaca entre os demais, por apresentar estimativas menores que o valor real nos modelos $M1$ e $M2$. Os modelos $M2$ e $M3$ apresentaram os melhores resultados em todos os cenário.

A Tabela 5.6 mostra a proporção de razão entre vícios relativos (RV) que são maiores que 1 nas comparações realizadas entre os demais modelos com $M2$ e $M3$. Para este tamanho de amostra é possível perceber uma melhora no desempenho dos modelos $M2$ e $M3$, visto que para mais parâmetros e cenários, a proporção de razões de vício relativo foi superior a 0.6, como destacado na Tabela 5.6. Observe também que em relação ao parâmetro β_1 , para todos as proporções de falha, os modelos $M2$ e $M3$ apresentaram melhores estimativas na comparação com $M0$ e $M1$, com $RV = 1$. Nas comparações realizadas entre $M0$ e $M3$ percebemos que o desempenho de $M3$ foi superior para maioria dos parâmetros. O modelo $M2$ confirmou o melhor desempenho em relação a $M1$. Neste

cenário, o $M3$ mostrou-se um pouco melhor que o $M2$, principalmente nas proporções de falha pequena e média.

O log das razões entre os vícios relativos de cada modelo em relação a $M3$ são apresentados nas Figuras 5.16 e 5.17. Observe que o modelo $M3$ apresenta estimativas menos viesadas para α , β_0 e β_2 no cenário independente visto que maior parte do intervalo interquartil está acima do traço em zero. Os melhores resultados podem ser vistos na comparação entre $M1$ e $M3$, já que para este modelo os gráficos apresentam maior amplitude. Em comparação com os modelos $M0$ e $M2$, observamos que $M3$, em geral, tem resultados semelhantes ou pouco melhores. Se a proporção de falha é alta, o 'impacto' da censura informativa, isto é, da dependência entre falha e censura é menor tornando os resultados obtidos por todos os modelos mais semelhantes.

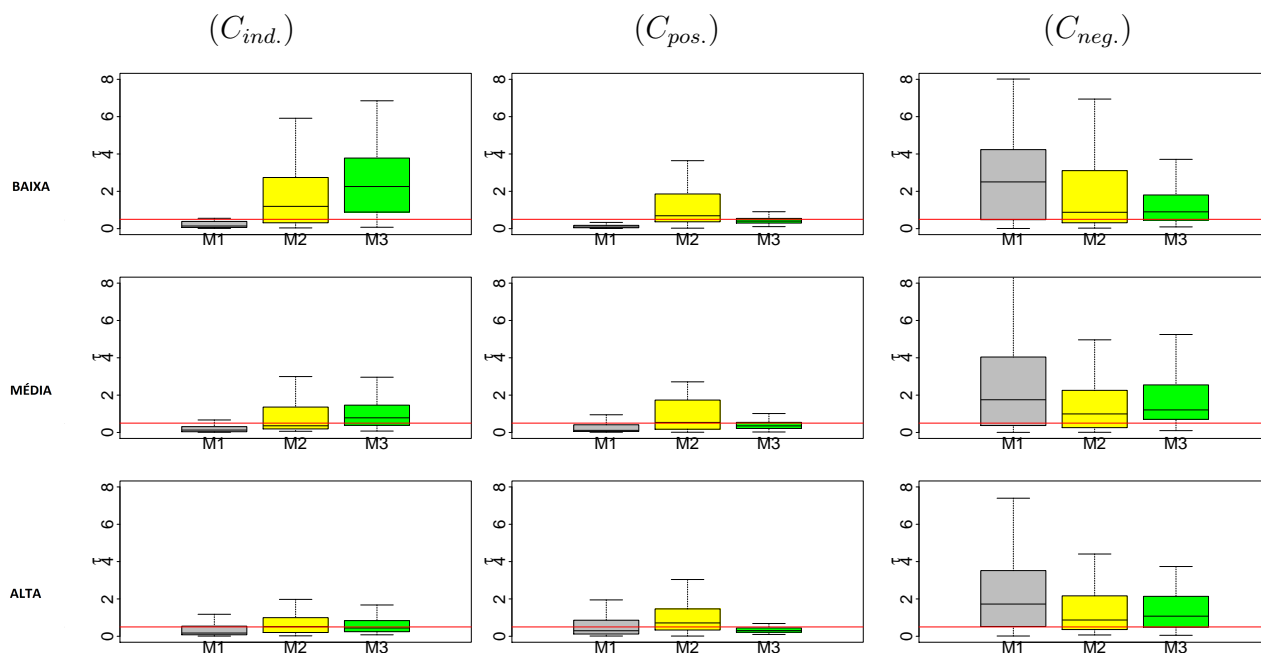


Figura 5.18: *Boxlots representando as médias a posteriori de τ obtidas em cada replicação do MCMC e considerando amostras de 250 tempos. Tração sobre o valor real $\tau = 0.2$.*

No cenário de correlação positiva, $M3$ produz melhores estimativas em relação ao parâmetro α , entretanto para os parâmetros β_0 e β_2 , seus resultados são similares ou pouco piores que os resultados obtidos por $M0$ e $M2$. O modelo $M3$ mostra melhores resultados que $M1$ mesmo para grandes proporções de falha.

No cenário de correlação negativa, a comparação de $M3$ com os demais modelos, para todos os parâmetros indica desempenho igual ou superior do modelo $M3$, com os *boxplots* acima do traço em zero. A exceção é o modelo $M0$ que mostra ter melhores resultados em relação ao parâmetro β_1 . No cenário independente, fica evidenciado que o modelo $M3$ tem melhores estimativas. O aumento da proporção de falha tornam todos os modelos mais competitivos em relação a $M3$. Quanto maior a proporção de falha, meno peso e contribuição terá a parte da modelagem dedicada às observações de censura. Desta forma relação com a censura não contribuirá muito para um melhor desempenho de $M3$ ou $M2$.

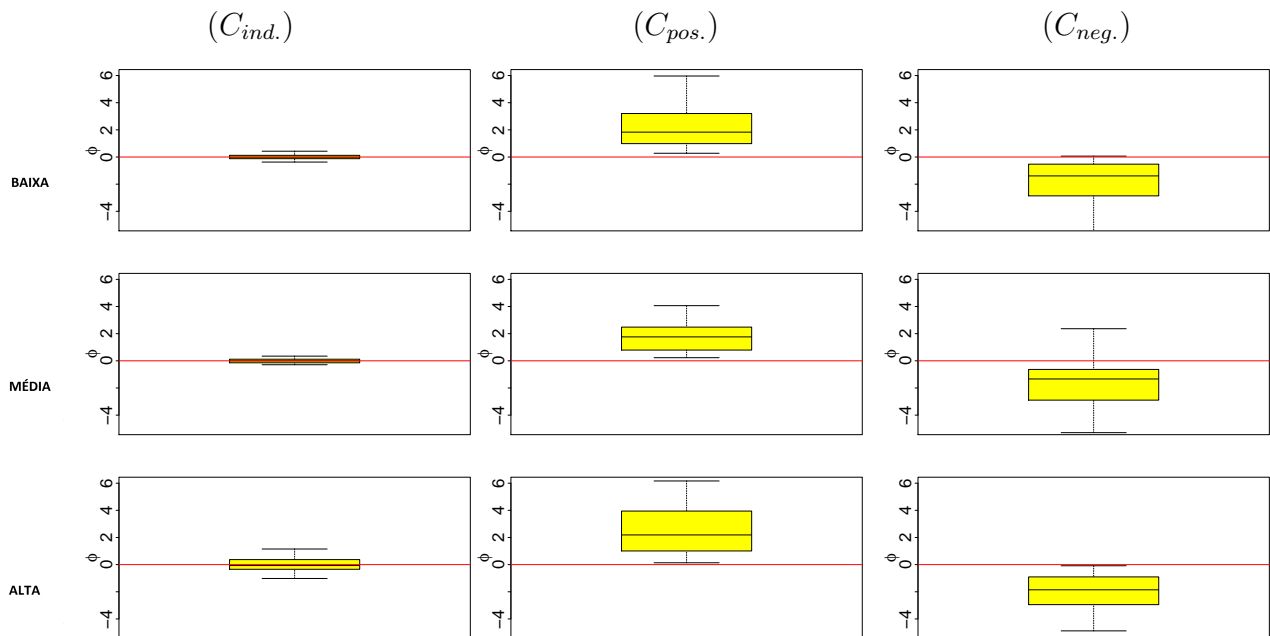


Figura 5.19: *Boxplots* representando as médias a posteriori de ϕ obtidas em cada replicação do MCMC e considerando amostras de 250 tempos.

As estimativas *a posteriori* de τ estão apresentadas na Figura 5.18. Para este tamanho de amostra, observamos alguma variação nas estimativas do parâmetro, em comparação ao observado para o tamanho 50, principalmente relacionado aos modelos $M1$ e $M3$. Lembrando que as estimativas de τ sofrem influência de ϕ no modelo $M2$, como anteriormente comentado. Note que para cada cenário a estimativa produzidas por um dos modelo são maiores que as produzidas pelos demais, sendo $M3$ no cenário independente,

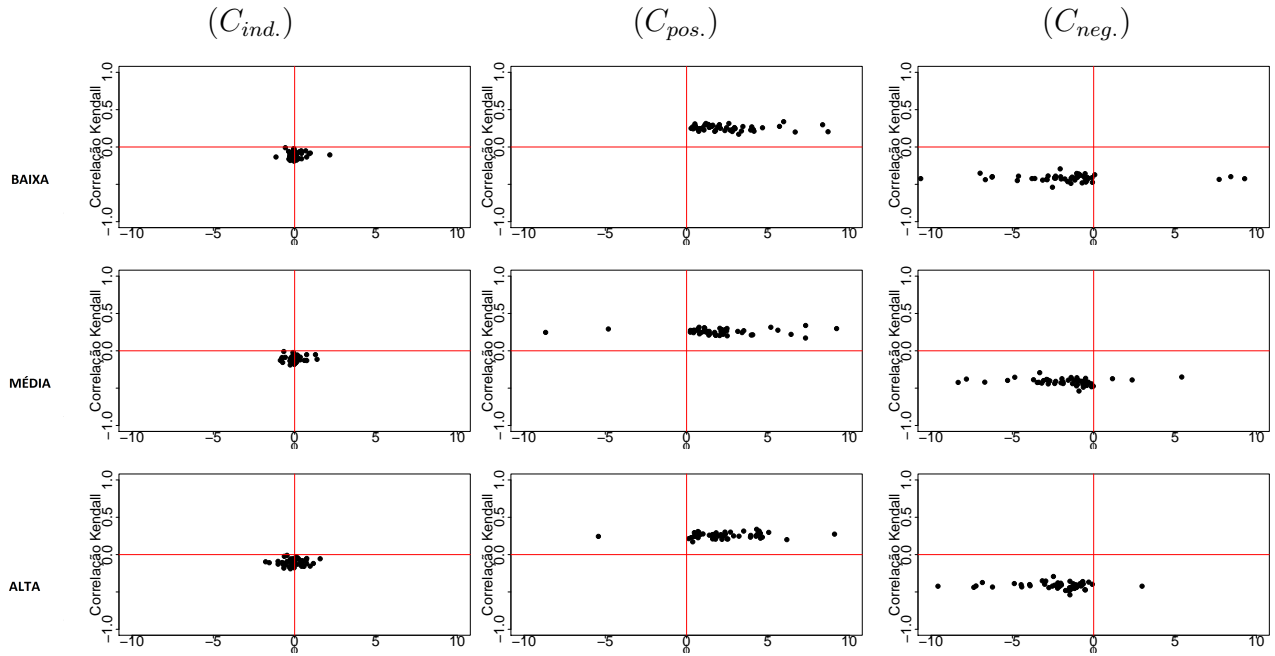


Figura 5.20: Gráfico de dispersão de ϕ contra a correlação Kendall obtidas em cada replicação do MCMC e considerando amostras de 250 tempos.

$M2$ no cenário com correlação positiva e $M1$ no cenário de correlação negativa. Para proporções de falha pequena, observamos que o valor real de τ não é corretamente estimado nos cenários independente e de correlação negativa, visto que todos os modelos apresentam algum viés. Entretanto a qualidade das estimativas melhora a medida que aumenta o percentual de falha. O modelo $M3$ mostra bons resultados na estimação de τ no cenário de correlação positiva para todos as proporções de falha. Novamente $M1$ apresenta maior viés, enquanto $M2$, tem os melhores resultados apesar da maior variabilidade de seus *boxplots* em geral.

O bom desempenho do parâmetro ϕ para amostra de tamanho 50 também pode ser observado na Figura 5.19. Para todos os cenários e proporções de falha, ϕ identifica corretamente a correlação existente entre os tempos de falha e censura. Observe que os *boxplots* estão concentrados na região que identifica corretamente o cenário. Por exemplo, no cenário independente há um pequena variação em torno de zero em todas as proporções de falha, enquanto no cenário de correlação positiva todo gráfico está na região acima de zero. Para o cenário de correlação negativa com proporção de falha média vemos

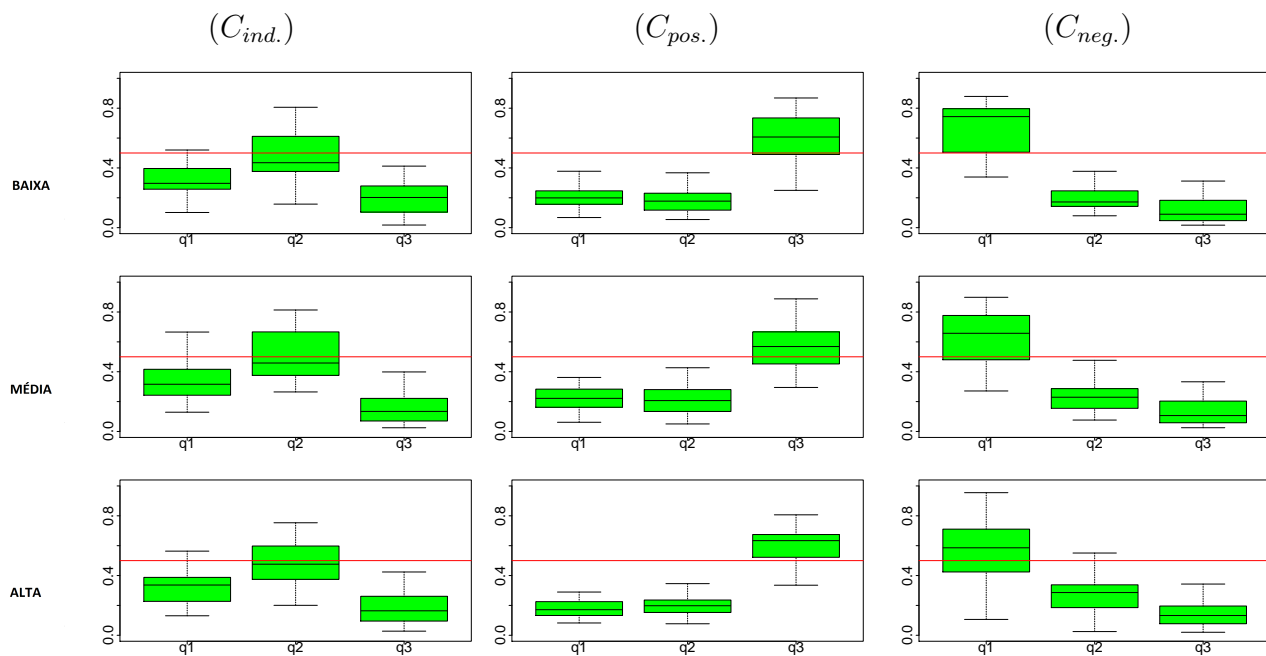


Figura 5.21: *Boxplots* representando as médias a posteriori de q obtidas em cada replicação do MCMC e considerando amostras de 250 tempos.

que parte do *boxplot* está na região positiva do gráfico, entretanto a maior parte das observações está alocada na parte negativa. Nas demais proporções de falha o gráfico encontra-se totalmente abaixo do traço em zero, indicando a correta escolha do tipo de correlação. O aumento do tamanho da amostra trouxe maior definição em relação ao tipo de correlação existente entre os tempos de falha e censura.

A comparação de ϕ com o correlação existente em cada cenário é apresentado na Figura 5.20. Observe que, com o aumento do tamanho da amostra os pontos ficaram concentrados na região adequada ao cenário. No cenário independente, os pontos estão concentrados na região central. Enquanto no cenário de correlação positiva a maioria dos pontos esta situada no quadrante superior direito, indicando que ϕ apresentou estimativas maiores que zero, sendo adequadas ao cenário. Para o cenário de correlação negativa, o aumento do tamanho da amostra, melhorou as estimações de ϕ . Note que, para esse cenário, apenas poucos valores de ϕ obtiveram as estimativas positivas o que reduziu a incerteza na estimação desse parâmetro.

Para o modelo $M3$, a definição do tipo de correlação na amostra é dada pelo parâmetro

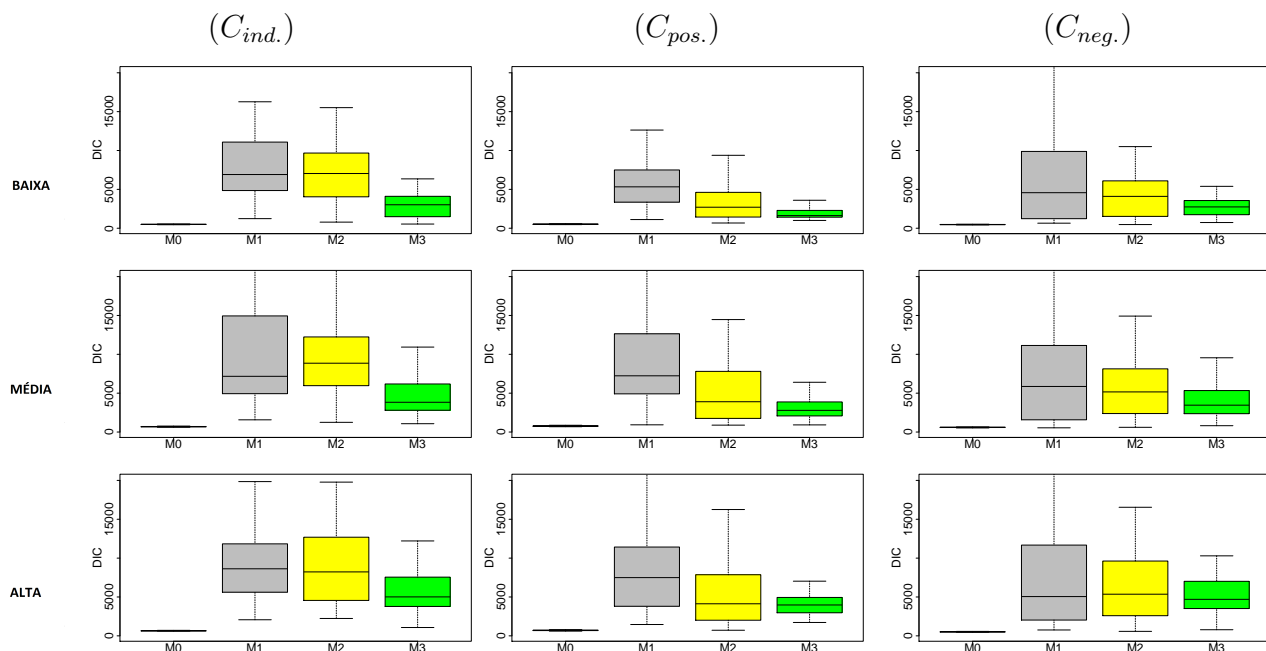


Figura 5.22: *Boxplots* representando os valores do DIC obtidos em cada replicação do MCMC e considerando amostras de 250 tempos.

\mathbf{q} , representado na Figura 5.21. Um traço horizontal sobre o patamar 0.5 auxilia as análises. Novamente, a maior componente de probabilidade é justamente aquela que indica o cenário correto dos dados. As estimativas da componente correta q_j apresenta valores iguais ou maiores a 0.5.

Tabela 5.7: DIC e LPML.

%	modelo	$C_{ind.}$		$C_{pos.}$		$C_{neg.}$	
		DIC	LPML	DIC	LPML	DIC	LPML
baixa	M0	500.49	-250.31	514.72	-257.50	469.70	-234.94
	M1	6909.13	-182.62	5324.78	-191.07	4571.97	-222.43
	M2	7039.27	-220.41	2703.04	-239.84	4098.62	-214.35
	M3	3018.76	-235.93	1626.43	-240.83	2742.09	-221.23
média	M0	687.62	-343.87	768.34	-385.21	654.02	-327.32
	M1	7168.48	-269.94	7235.07	-308.44	8627.69	-252.14
	M2	8855.22	-300.58	3898.63	-356.51	8227.19	-281.82
	M3	3828.54	-318.75	2787.98	-361.44	5011.73	-285.42
alta	M0	654.02	-327.32	708.86	-355.78	518.05	-259.26
	M1	8627.69	-252.14	7492.50	-279.91	5053.36	-237.58
	M2	8227.19	-281.82	4125.47	-328.97	5357.50	-230.60
	M3	5011.73	-285.42	3972.34	-308.49	4696.36	-233.02

Encerrando a análise, apresentamos a Figura 5.22 que mostra o critério DIC de comparação de modelos. Novamente, $M0$ apresenta melhor desempenho em comparação

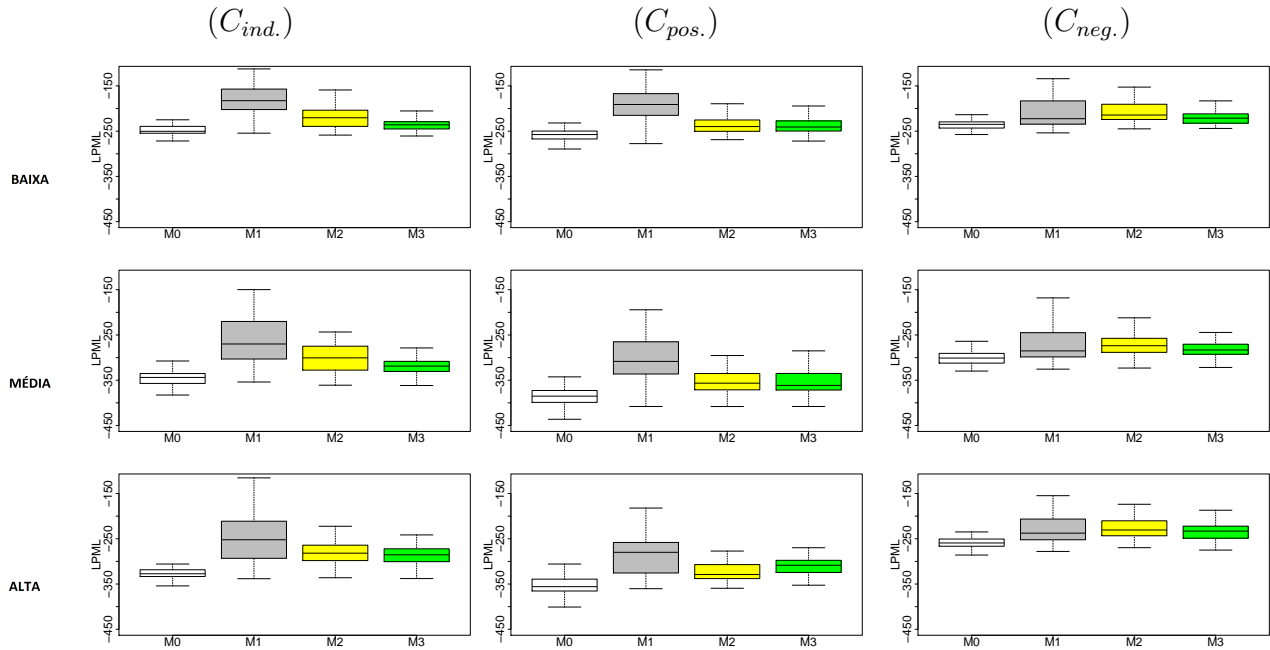


Figura 5.23: *Boxplots* representando os valores do LPML obtidos em cada replicação do MCMC e considerando amostras de 250 tempos.

com os demais modelos. Isso ocorre principalmente devido ao seu número reduzido de parâmetros, sendo o único que não possui efeito aleatório na modelagem. Considerando $M1$, $M2$ e $M3$, o critério DIC indica $M3$ como o melhor modelo para o cenário independente, visto ter os menores *boxplots*. Para pequenas proporções de falha, $M3$ tem claramente o menor DIC. Entretanto, para as proporções de falha média e grande, apesar da maior amplitude dos *boxplots* do modelo $M2$, as medianas de $M2$ e $M3$ são bem semelhantes.

Comparando estes resultados com os obtidos para amostras de tamanho 50, verificamos que o aumento do conjunto de dados confirma a posição de $M1$ como modelo menos indicado ao ajuste, os valores do DIC para $M1$ eram similares aos de $M2$ e agora são um pouco maiores. Além disso, os valores do DIC para um conjunto de dados com 50 observações é muito menor que os DICs calculados para uma amostra de 250 observações, como é possível observar no eixo horizontal.

O critério LPML é apresentado na Figura 5.23. Observe que para todos os cenários, o modelo que apresenta os maiores resultados para LPML é o $M1$, principalmente nos

cenários independente e de correlação positiva. No cenário de correlação negativa o *boxplot* de $M1$ está mais próximo do *boxplot* de $M2$, sendo $M1$ um pouco melhor. Veja que $M0$ aparece com o pior resultado dentre os modelos, enquanto $M2$ e $M3$ mostram resultados semelhantes para a maioria dos cenários.

A mediana do DIC e LPML, calculada para as 50 réplicas MCMC, estão registradas na Tabela 5.7. As informações da tabela também apontam $M0$ e $M1$ como melhores modelos de acordo com os critérios DIC e LPML, respectivamente. Os modelos $M2$ e $M3$ tem, em geral, classificação intermediária. No cenário de correlação negativa, para proporções de falha média e grande, o LPML indica os modelos $M2$ e $M3$ como possuindo melhor desempenho. A diferença entre os resultados obtidos por $M2$ e $M3$ em relação a esses critérios é maior que a observada para amostra de tamanho 50 (antes apresentavam valores muito próximos).

É interessante observar que os dois critérios de seleção de modelos produzem resultados completamente opostos, sendo DIC indicando $M0$ e $M1$ como melhor e pior modelo, respectivamente, e LPML indicando $M0$ como pior modelo e $M1$ como melhor. Estes resultados conflitantes também podem ser observados para a amostra de tamanho 50. Assim, para definição dos maiores modelos observamos os intervalos de credibilidade e o modelo com menor vício. Considerando os resultados apresentados, podemos dizer que para amostras de tamanho 50, $M2$ mostrou-se como melhor modelo, enquanto para amostras de tamanho 250, $M3$ foi melhor. O melhor resultado de $M3$ para amostras maiores esta relacionado ao seu número de parâmetros que é melhor estimado com o maior números de observações.

Capítulo 6

Aplicação: Mieloma Múltiplo

O mieloma múltiplo é uma doença que afeta a medula óssea e se caracteriza pelo aumento do plasmócito (tipo de célula que produz imunoglobulina), sendo relacionada ao nosso sistema de defesa. O mieloma múltiplo acontece quando um grupo de plasmócitos se multiplica de forma desgovernada e compromete o funcionamento da medula óssea. Os problemas de saúde causados pelo mieloma múltiplo podem afetar os ossos, sistema imunológico, rins e contagem de células vermelhas do sangue (Maiolino e Magalhães, 2015). A doença tem maior incidência em pessoas idosas, sendo rara em indivíduos com menos de 35 anos. Ainda não estão elucidadas as causas que levam a doença, entretanto, alguns fatores genéticos e ambientais podem aumentar a predisposição do indivíduo como: exposição a radiação, profissões na agricultura e indústria de materiais pesados e idade. O tratamento é realizado principalmente por meio de transplante de medula óssea, radioterapia e quimioterapia.

Os tempos de sobrevivência de 65 pacientes com mieloma múltiplo, que foram diagnosticados e tratados com agentes alquilantes no centro médico da Universidade de West Virginia, EUA, foram disponibilizados por Krall et al. (1975) num estudo que propõe uma técnica para seleção de covariáveis. Dos pacientes em estudo 48 morreram (74%) durante o período de acompanhamento, enquanto 17 (26%) permaneceram vivos e foram considerados observações censuradas. O evento de interesse era o tempo entre o diagnóstico e óbito dos pacientes.

O estudo também disponibiliza 16 covariáveis medidas no momento do diagnóstico.

Entre essas covariáveis, o artigo indica que apenas três são úteis ao diagnóstico, sendo elas: o nível de nitrogênio ureico no sangue (X_1) (*blood urea nitrogen*- BUN), cálcio (X_2) e hemoglobina (X_3), ver Krall et al. (1975). Segundo os autores, outros estudos reforçam que estas covariáveis estão fortemente ligadas ao tempo de sobrevivência de pacientes e, portanto consideramos apenas essas covariáveis no estudo.

Dois pacientes da amostra apresentaram falha em tempo inferior a 15 dias após o diagnóstico, sendo atribuído um tempo de sobrevivência de 0.25 mês. Krall et al. (1975) propõem uma transformação nos dados sendo tempo de falha igual ao tempo observado mais 1 mês. Essa transformação será mantida aqui. O log da covariável nitrogênio ureico no sangue (BUN) foi tomado, sendo que as demais permaneceram na mesma escala.

Este mesmo conjunto de dados foi utilizado por Siannis et al. (2005) como exemplo para uma análise de sensibilidade de um modelo para censura informativa proposto por Siannis (2004). Os autores, inicialmente, discutem o banco de dados que, aparentemente, não possui associação entre os tempos de falha e censura, dado que as observações censuradas são essencialmente interrompidas com o fim do estudo (censura administrativa). Entretanto, esses dados tratam de um típico exemplo de análise de sobrevivência para o qual a censura ocorre por razões não controladas inicialmente, por exemplo, os pacientes são selecionados de acordo com o surgimento da doença e não de forma aleatória. Segundo Siannis et al. (2005), nesses casos a suposição de independência deve ser vista com cautela. De fato, o artigo mostra que há indícios de pequena correlação entre os tempos de falha e censura que, ainda segundo os autores, pode ser suficiente para ter efeito sobre as análises.

Os modelos M_0 , M_1 , M_2 e M_3 foram ajustados a esses dados. As distribuições *a priori*, discutidas nas Seções 6.1.1 a 6.1.4, foram mantidas assim como os hiperparâmetros, com exceção dos parâmetros α_T e α_C , para os quais optamos por uma Gama(2, 1) que é uma distribuição um pouco mais informativa cuja média é aproximadamente 2 e a mediana é igual a 1.5. Considerando as três covariáveis na modelagem, β_{Tj} e β_{Cj} passam a ser definidas para $j = 0, 1, 2, 3$, sem alteração nas especificações *a priori*. Para geração das amostras *a posteriori* os *softwares* R (R Core Team, 2014) e *JAGS* (Plummer, 2003) foram utilizados. Para o ajuste de cada modelo, uma única cadeia com 30 mil iterações

foi gerada, para o qual as primeiras 20 mil iterações foram descartadas (*burn-in*). Consideramos um *lag* de tamanho 10 e uma amostra *a posteriori* de tamanho 1000 para cada parâmetro. A Figura 6.1 mostra o traço da cadeia e a densidade *a posteriori* para os parâmetros da distribuição de falha considerando o ajuste de M_0 .

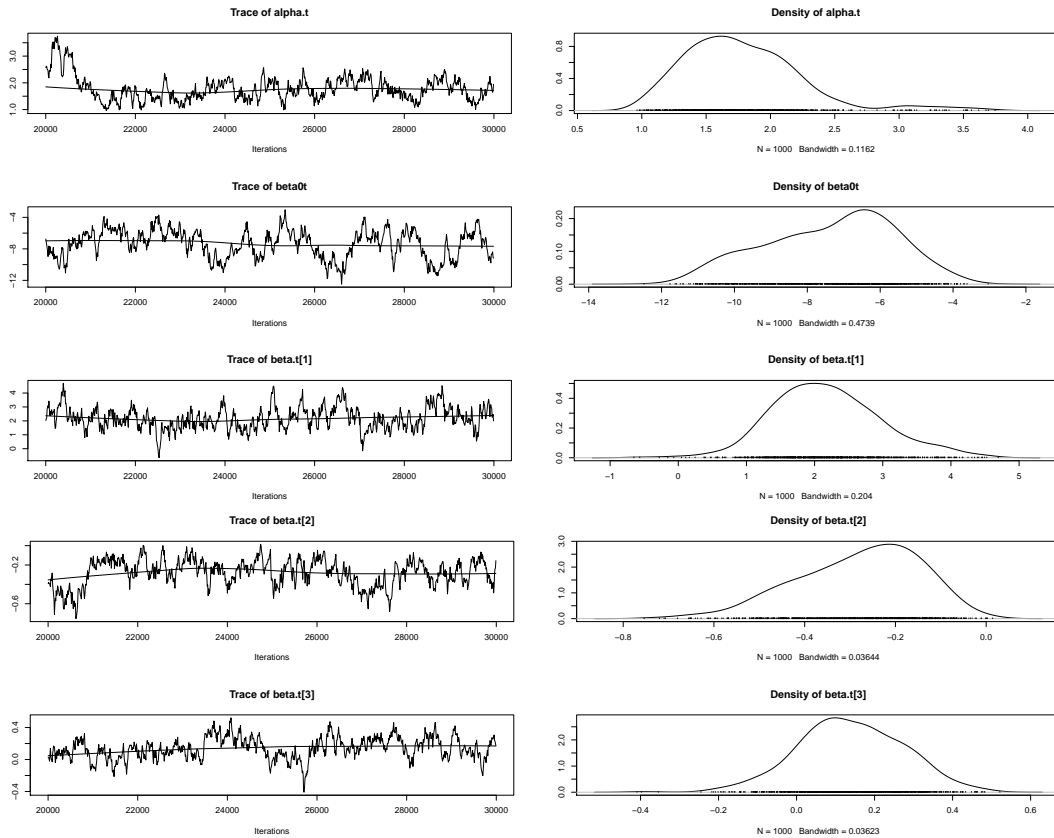


Figura 6.1: Traço da cadeia indicando convergência e densidade *a posteriori* para os parâmetros da distribuição de falha relativos ao modelo M_3 .

As estimativas dos parâmetros de falha obtidos com o ajuste dos modelos aos dados são apresentados na Tabela 6.1. A tabela está dividida em três partes, de forma que cada uma apresenta a média da amostra *a posteriori* e o intervalo de credibilidade HPD para cada parâmetro.

As médias *a posteriori* indicam que $\alpha_T > 1$ para todos os modelos ajustados sugerindo que os dados seguem uma distribuição Weibull, cuja função de taxa de falha tem comportamento crescente. O modelo M_0 apresenta a menor estimativa e um pequeno intervalo de credibilidade. Lembre-se que para simulações realizadas com amostras de

Tabela 6.1: Estimativa dos parâmetros de falha

		α			τ		
		HPD		HPD			
modelo	média	Li	Ls	média	Li	Ls	
M0	1.12	0.89	1.34	-	-	-	
M1	1.79	0.91	2.86	1.77	0.04	7.71	
M2	1.66	1.01	2.31	0.51	0.06	1.90	
M3	2.20	1.21	3.72	0.36	0.02	1.07	

		β_0			β_1		
		HPD		HPD			
modelo	média	Li	Ls	média	Li	Ls	
M0	-5.27	-7.82	-2.80	1.49	0.44	2.62	
M1	-6.78	-11.84	-3.38	1.89	-0.01	3.81	
M2	-7.86	-11.20	-4.93	1.76	0.09	3.47	
M3	-7.52	-11.13	-3.13	1.88	-0.01	4.04	

		β_2			β_3		
		HPD		HPD			
modelo	média	Li	Ls	média	Li	Ls	
M0	-0.13	-0.23	-0.02	0.07	-0.14	0.22	
M1	-0.27	-0.55	-0.05	0.11	-0.20	0.45	
M2	-0.30	-0.56	-0.04	0.26	-0.02	0.60	
M3	-0.37	-0.84	-0.04	0.16	-0.28	0.47	

Li = limite inferior - Ls = limite superior

tamanho 50 o modelo $M2$ apresentava estimativas com menor viés para este parâmetro, enquanto $M0$ apresentava estimativas menores que o valor real. Para estes dados, $M0$ produz intervalos HPD de menor amplitude, seguindo o comportamento apresentado no estudo simulado. O modelo $M2$ também produz um intervalo relativamente pequeno, enquanto $M1$ e $M3$ têm as maiores amplitude de intervalo HPD.

Para o parâmetro τ (Tabela 6.1) observamos que $M2$ e $M3$ tem estimativas médias bem próximas, assim como são similares os limites do intervalo de credibilidade, sendo a amplitude do intervalo para $M2$ um pouco maior. Para $M1$, observamos estimativas para τ um pouco maiores que as dos demais modelos e uma grande variação no intervalo de credibilidade.

Para o intercepto β_0 , as médias *a posteriori* e o intervalo de credibilidade são muito parecidos para $M2$ e $M3$. Nos modelos $M0$ e $M1$ temos estimativas diferentes de $M2$ e $M3$, porém a média é também negativa e com intervalo de credibilidade HPD de 95%

não incluindo o zero.

Para o vetor de parâmetros β_T note que existe coerência nas estimativas obtidas em todos os modelos, sendo razoavelmente próximas umas das outras, à exceção do modelo $M0$ que apresenta valores estimados mais próximos de zero em todos os parâmetros. Os intervalos de confiança incluem o valor zero para β_3 , indicando que esta covariável pode não ser significativa.

Em seu artigo, Krall et al. (1975) usa uma distribuição exponencial, cuja densidade é dada por $f_i(y) = \theta_i \exp\{-\theta_i y\}$ com $1/\theta_i = \mathbf{x}_i \beta_i$ e $i = 0, 1, 2, 3$, para modelagem dos dados, obtendo os seguintes parâmetros: $1/\theta_i = 82.72 - 37.46x_{1,i} + 2.399x_{2,i} - 2.085x_{3,i}$. Essa modelagem dificulta as comparações com nossa proposta visto que utilizamos uma função exponencial para incluir as covariáveis. Contudo, os dados de mieloma múltiplo parecem ser adequadamente ajustados à distribuição Weibull com $\alpha > 1$. O intercepto β_0 tem estimativas próximas a -7 , o que indica um risco inicialmente pequeno do paciente experimentar o evento de interesse. O coeficiente de regressão β_1 foi estimado em aproximadamente 1.8, indicando um aumento do risco do indivíduo para grandes valores de X_1 . Este resultado é coerente com o diagnóstico, visto que altos níveis de BUN indicam prejuízo da função renal (sintoma comum em paciente com mieloma múltiplo). O coeficiente β_2 está relacionado ao nível de cálcio, sendo estimado em aproximadamente -0.3 . Neste caso, quanto maior o valor da covariável, menor será o risco do indivíduo. Finalmente, o coeficiente β_3 , estimado em 0.1. Esta covariável indica os níveis da hemoglobina, sendo que altos valores absolutos aumentam o risco.

Ambos modelos, $M2$ e $M3$ identificaram a existência de uma correlação negativa entre os tempos de falha e censura, como mostra a Tabela 6.2. Observe que a média de ϕ e seu intervalo de credibilidade está todo compreendido na região abaixo de zero. O modelo $M3$ também indica correlação negativa através do componente q_1 estimado em 0.49 e maior que q_2 e q_3 .

Avaliando os critérios de seleção de modelos DIC e LPML apresentados na Tabela 6.3, verificamos que devido ao reduzido número de parâmetros, o critério DIC privilegia o modelo $M0$, ficando $M2$ com a segunda colocação. Lembrando que $M0$ ignora a existência de correlação entre os dados e não modela a variabilidade extra que parece

Tabela 6.2: Estimativa dos parâmetros ϕ e \mathbf{q} .

modelo	ϕ			\mathbf{q}		
	média	HPD		q_1	q_2	q_3
		Li	Ls			
M2	-4.85	-10.20	-1.10	-	-	-
M3	-	-	-	0.49	0.23	0.28

ser grande, visto as estimativas pequenas obtidas para a precisão τ (Tabela 6.1). Já o critério LPML indica $M3$ como melhor modelo e $M2$ como o segundo melhor.

Tabela 6.3: Estimativas para DIC e LPML.

modelo	DIC	LPML
M0	426.19	-213.76
M1	911.47	-207.14
M2	622.31	-205.56
M3	886.69	-202.13

Capítulo 7

Conclusões

Neste trabalho foram feitos estudos de modelos de sobrevivência para dados com censura informativa. Foram considerados modelos de taxas de falha proporcionais sem efeito aleatório e com efeito aleatório apenas na componente de falha (modelos indicados ao ajuste de censura não informativa), além de modelos com efeitos aleatórios em ambos os componentes - falha e censura - sendo o efeito aleatório da componente de censura afetada por um parâmetro real ϕ em um dos modelos e controlado por uma mistura em outro. Estes modelos permitem a identificação do mecanismo de censura e, em caso de censura informativa, determinam o tipo de correlação entre falha e censura.

Uma estudo simulado, focado na distribuição Weibull, ajustou estes modelos considerando cenários de censura não informativa e informativa com correlação positiva ou negativa entre os tempos de falha e censura. Considerou-se tamanhos amostrais de 50 e 250 observações e para estes dois tamanhos de amostra foi possível observar viés nas estimativas de algum parâmetro para todos os cenários e modelos. Não foi possível identificar um modelo perfeito estimando muito bem todos os parâmetros.

O modelo $M0$ apresenta, para a maioria dos cenários, algum viés nas estimativas se aproximando do valor real para alguns parâmetros. O modelo $M1$, de maneira geral, apresentou os piores resultados em todos os cenários, inclusive no cenário independente para o qual seria o mais adequado. As estimativas produzidas por esse modelo estavam bastante viesadas e com grande variabilidade. Modelar a falha e a censura (como é feito em $M2$ e $M3$) parece ser vantajoso, visto que estaremos utilizando mais informação (todos

os tempos) para estimar a variância τ da fragilidade. O modelo $M1$ modela a fragilidade (estima τ) levando em conta apenas os tempos de falha. Se estivermos no cenário com proporção de falha baixa, o desempenho ruim de $M1$ será mais visível.

Para $n = 50$, o modelo $M2$ apresenta pequeno viés para alguns parâmetros. Esse modelo é, aparentemente, melhor ou competitivo em relação a $M3$ em todos os cenários. Para amostras de tamanho 250, $M2$ apresentou resultados bem semelhantes, entretanto para alguns cenários o modelo $M3$ forneceu estimativas menos viesadas. Levando em conta as estimativas de τ produzidas por $M2$, a forma como o modelo é definido faz com que o parâmetro ϕ interfira nas estimativas de τ .

O modelo $M3$ mostrou boa adaptação ao cenário independente para ambos tamanhos amostrais, porém existiu algum viés nas estimativas dos parâmetros para os cenários com correlação. Seus resultados apresentaram melhora com o aumento do tamanho da amostra tornando-o mais competitivo em comparação com $M2$ para os cenários de correlação positiva e negativa. Tanto $M2$, através do valor de ϕ , quanto $M3$, pelas estimativas de \mathbf{q} , mostraram-se eficientes na indicação correta do tipo de correlação existente entre os tempos de falha e censura.

O critério de comparação DIC apresenta dificuldades para definição do melhor modelo, principalmente devido à inclusão de um efeito aleatório em três dos modelos o que traz uma vantagem em termos de DIC para a modelagem mais simples $M0$ (sem fragilidade). Observe que para o LPML o modelo $M1$ foi indicado como o melhor, sendo exatamente aquele que apresentou maior viés nas estimativas.

De maneira geral, os coeficientes de regressão β_{T_j} , com $j = 1, 2$ apresentaram estimativas mais precisas dentre todos os parâmetros, provavelmente por estarem associados a covariáveis, o que torna a estimação mais estável. O principal ganho com o aumento do tamanho da amostra é a redução da amplitude dos *boxplots* indicando menor variabilidade do vício para todos os modelos. Os modelos $M2$ e $M3$ tem um maior número de parâmetros o que torna os modelos mais complexos, entretanto existe o ganho da obtenção de melhores estimativas no cenário com correlação entre os tempos de falha e censura, além de determinar o tipo de correlação existente. Essas características permitem ao pesquisador obter mais informações sobre o conjunto de dados. Os modelos

$M2$ e $M3$ mostraram resultados razoáveis nas estimativas dos parâmetros em todos os cenários. O modelo $M2$ mostrou melhor desempenho para amostras menores, enquanto $M3$ foi melhor para a maior parte dos cenários quando a amostra de tempos era maior.

Os modelos estudados aqui também foram ajustados a dados reais com três covariáveis, disponibilizados em Krall et al. (1975), referente a pacientes com diagnóstico de Mieloma Múltiplo tratados com agentes alquilantes no centro médico da Universidade de West Virginia, EUA. Para este ajuste consideramos uma *priori* mais informativa para α do que aquela utilizada no estudo simulado. Os resultados encontrados pelos modelos $M2$ e $M3$ indicam existência de correlação negativa entre os tempos de falha e censura.

As estimativas obtidas pelos demais parâmetros indicam coerência nas estimativas dos modelos, entretanto, a covariável X_3 foi considerada não significativa, visto que os intervalos de credibilidade de 95% do parâmetro β_3 incluem o valor zero em todos os ajustes. Em relação aos critérios de seleção de modelos, o DIC indicou $M0$ como o melhor modelo, enquanto $M3$ foi apontado como o melhor segundo o critério LPML. Para os dois critérios $M2$ ocupou a segunda posição.

7.1 Trabalhos futuros

Entre as propostas para continuidade deste trabalho podemos citar:

- Adaptar o modelo $M2$ para que ele considere dados com censura parcialmente informativa. Uma maneira de fazer isso é adotar a mistura *a priori* $(\phi|Z_i) \sim (1 - Z_i)d_0 + Z_i\phi$ em que $\phi \sim \text{Uniforme}[-a, a]$ com $a > 0$ e $Z_i \sim \text{Bernoulli}(q)$. Isto permite uma classificação de existência de efeito de fragilidade para cada i -ésimo tempo de sobrevivência dada através de Z_i .

- Incluir na geração de dados observações referentes à censura administrativa. Podemos considerar um limite superior para o tempo máximo de estudo e também a ocorrência de censuras devido a fatores não ligados ao estudo como a morte de um paciente por acidente de carro.

- Gerar dados utilizando estratégias alternativas. Uma delas é considerar cópulas para determinar o tipo de correlação entre falha e censura; como por exemplo as cópulas *Arqui-*

medianas, uma classe que abrange uma grande variedade de estruturas de dependência e são frequentemente utilizadas para realizar conexão com a noção de fragilidade (Suzuki, 2012). A subclasse conhecida como cópula de *Frank* permite gerar dados com correlação positiva e negativa por meio da escolha de um parâmetro θ . Neste caso, um efeito aleatório seria introduzido apenas para caracterizar a existência de uma variabilidade extra.

- Reduzir a correlação das cadeias aumentando o espaçamento (*lag*) da cadeia *a posteriori*;

- Utilizar outros métodos de seleção de modelos para melhor comparação, como por exemplo o *widely applicable information criterion* (WAIC), proposto por Watanabe (2010).

- Escolher outras distribuições, diferentes da Weibull, para a modelar a função de taxa de falha. Para cada modelo paramétrico adotado, incluindo a Weibull, seria interessante também desenvolver uma análise de sensibilidade avaliando o impacto de várias escolhas de especificações *a priori*.

Apêndice

Apêndice A: Conceitos Básicos de Inferência Bayesiana

Segundo Robert (2007), o principal objetivo da teoria estatística é obter, a partir de observações de um fenômeno aleatório, uma inferência sobre a distribuição de probabilidade que descreva este acontecimento de forma a fornecer tanto uma análise desse fato quanto uma previsão sobre um futuro evento de natureza semelhante. Esta distribuição é indexada por uma quantidade de interesse (parâmetro) que representa características que se deseja conhecer para ter uma visão mais completa do fenômeno.

Em inferência Bayesiana, não apenas o evento de interesse é representado por uma distribuição de probabilidade, como todas as quantidades não observáveis (parâmetros e variáveis latentes relacionadas a ele) sobre as quais possuímos algum grau de incerteza. Segundo Gamerman e Lopes (2006), a principal diferença entre o método Bayesiano e o Frequentista está no fato de que o primeiro incorpora conhecimento prévio a respeito de quantidades de interesse (parâmetros), enquanto o segundo não admite essa forma de incorporação da informação subjetiva por não ser observável e, portanto, não ser passível de comprovação. O uso da abordagem Bayesiana possui vantagens como a especificação natural e intuitiva de modelos, a incorporação sequencial de informação, maior robustez e interpretabilidade de resultados. Gamerman e Lopes (2006), Gelman et al. (2014), Robert (2007) são algumas referências que tratam da Inferência Bayesiana.

A regra de Bayes é o teorema que fundamenta a abordagem Bayesiana. Ela trata de probabilidades condicionais sendo utilizada para atualizar a informação inicial disponível sobre a quantidade desconhecida de interesse antes de observarmos os dados. A proposta é resumir toda a informação prévia sobre um vetor de parâmetros desconhecidos θ através de uma distribuição de probabilidade $p(\theta)$, denominada distribuição *a priori*. Esta distribuição expressa nossa incerteza sobre o parâmetro antes que qualquer dado seja observado. A informação dos dados observados é introduzida através da função de verossimilhança $L(\theta|D)$, assim a distribuição *a priori* será atualizada de maneira a reduzir a incerteza inicial. Essa distribuição atualizada é denominada distribuição *a posteriori*, e denotada por $p(\theta|D)$, em que D representa o conjunto de dados observados. A

distribuição *a posteriori* é obtida via Teorema de Bayes como segue:

$$p(\boldsymbol{\theta}|D) = \frac{p(\boldsymbol{\theta}, D)}{p(D)} = \frac{L(\boldsymbol{\theta}|D)p(\boldsymbol{\theta})}{\int L(\boldsymbol{\theta}|D)p(\boldsymbol{\theta})d\boldsymbol{\theta}} \quad (0.1)$$

Note que o termo do denominador não depende de $\boldsymbol{\theta}$, sendo apenas uma constante normalizadora. Dessa forma a regra de Bayes pode ser escrita como:

$$p(\boldsymbol{\theta}|D) \propto L(\boldsymbol{\theta}|D)p(\boldsymbol{\theta})$$

Uma característica da distribuição *a posteriori* é a redução da influência da distribuição *a priori* na medida em que há o aumento do número de observações, ou seja, quando é possível retirar mais informação dos dados.

A escolha da distribuição *a priori* é uma etapa importante do estudo; uma escolha inadequada pode dificultar a análise. A distribuição selecionada deve expressar toda a informação prévia sobre a quantidade de interesse desconhecida. Entretanto, na maioria das vezes, não se dispõe de qualquer informação sobre essa quantidade. Nestes casos, as distribuições *a priori* são chamadas não informativas. Em algumas situações a especificação escolhida *a priori* não pode ser definida como distribuição de probabilidade por não respeitar propriedades básicas de uma distribuição; este caso ocorre, por exemplo, quando utilizamos distribuições *a priori* de Jeffreys (Gamerman e Lopes, 2006). Isto, entretanto, não configura um problema desde que a expressão *a posteriori* resultante seja uma distribuição de probabilidade. Em outros casos, dada a escolha da distribuição *a priori*, temos que a distribuição *a posteriori* resultante pertence a mesma classe da distribuição *a priori*. Nesse caso, dizemos que a distribuição *a priori* é conjugada. A escolha de distribuições conjugadas minimiza o uso de técnicas computacionais e simplifica a estimação dos parâmetros, pois a passagem da distribuição *a priori* para *a posteriori* envolve agora uma simples atualização dos parâmetros. Gamerman e Lopes (2006) alertam para o cuidado de se utilizar apenas distribuições *a priori* conjugadas, visto que essa escolha nem sempre representa adequadamente a incerteza sobre o parâmetro.

A especificação das distribuições *a priori* é, normalmente, realizada pelo pesquisador que deve considerar as características já conhecidas do conjunto de dados e representar

a incerteza sobre a quantidade de interesse. Durante a especificação, o pesquisador deve definir os valores dos parâmetros da distribuição *a priori*, denominados hiperparâmetros. Em alguns casos, esses hiperparâmetros também podem ser associados a uma distribuição de probabilidade. É importante ressaltar que os conceitos de distribuição *a priori* e *a posteriori* são sempre relativos à observação considerada no momento.

Quando $\boldsymbol{\theta}$ é multivariado, distribuições marginais e condicionais podem ser obtidas para cada componente θ_i a partir da densidade conjunta, como apresentado por Gamerman e Lopes (2006). A densidade marginal *a posteriori* de θ_i , com $i = 1, \dots, k$ é dada por:

$$P(\theta_i|D) = \int P(\boldsymbol{\theta}|D)d\boldsymbol{\theta}_{-i}, \quad (0.2)$$

em que $\boldsymbol{\theta}_{-i}$ representa o vetor de parâmetros $\boldsymbol{\theta}$ à exceção do i -ésimo componente. Denotaremos por $P(\theta_i|D, \boldsymbol{\theta}_{-i})$ a distribuição condicional completa do parâmetro θ_i , ou seja, a distribuição do parâmetro dado que todas as demais quantidades são conhecidas.

A distribuição *a posteriori* de um parâmetro deve resumir toda informação probabilística atualizada a respeito do mesmo. Uma vez obtida, a ideia é sumarizar a informação nela contida em estimativas pontuais ou intervalares. No entanto, resumir a informação da distribuição *a posteriori* em um único elemento tem um custo equivalente a uma perda de informação. Gamerman e Lopes (2006) e Bolfarine e Sandoval (2002) indicam o uso de medidas de tendência central como média, mediana e moda, e medidas de dispersão como variância, desvio-padrão e precisão (inverso da variância), para avaliar o parâmetro de interesse.

O estimador $\hat{\theta}$ que minimiza o custo de utilizar um único valor para caracterizar uma distribuição é chamado estimador de Bayes. Nesse sentido, é importante avaliar o quão precisa é essa informação por meio das medidas de dispersão. A variabilidade da medida é controlada pelo intervalo de confiança Bayesiano, também chamado de intervalo de credibilidade. Um intervalo de credibilidade Cr com $100(1 - \alpha)\%$ nível de probabilidade para θ é apresentado por Bolfarine e Sandoval (2002) como $P(\theta \in Cr) \geq 1 - \alpha$. Essa definição permite a construção de inúmeros intervalos para θ , sendo natural o interesse

em intervalos com menor comprimento, que corresponde ao intervalo onde o parâmetro θ toma valores com maior densidade *a posteriori*. O intervalo de máxima densidade *a posteriori* (*highest posterior density* - HPD) será $Cr = \{\theta \in \Theta : p(\theta|x) \geq k(\alpha)\}$, em que $k(\alpha)$ é a maior constante tal que $P(\theta \in Cr) \geq 1 - \alpha$. Para mais detalhes consulte Gelman et al. (2014) e Robert (2007).

Em modelos mais complexos, com vários parâmetros e variáveis latentes, a especificação de um modelo Bayesiano completo deve ser feita de forma cuidadosa, respeitando as relações entre as várias quantidades desconhecidas. Essa estruturação pode ser feita mais naturalmente através de um modelo hierárquico. Um modelo hierárquico divide a especificação da distribuição *a priori* em estágios definindo uma sequência de distribuições condicionais que reflete uma hierarquia de dependência entre as variáveis. Este tipo de estrutura também incorpora com facilidade a distribuição dos hiperparâmetros e qualquer outra informação que seja relevante para o modelo. Gelman et al. (2014) ressalta que modelos hierárquicos podem ter parâmetros suficientes de forma a modelar melhor os dados ao usar a distribuição da população para estruturar a dependência entre os parâmetros.

Apesar das vantagens de especificação e estruturação de modelos que a inferência Bayesiana permite, a complexidade da forma que a distribuição *a posteriori* pode assumir ainda é um grande obstáculo. Em muitos casos, conhecer a constante normalizadora da distribuição, como mostrado em (0.1) é impossível; não podemos calcular de forma exata a integral mostrada no denominador. Neste caso, desconhecemos a forma completa da distribuição *a posteriori* que permitiria, entre outras coisas, a comparação de modelos e o cálculo de estimativas pontuais e intervalares. A fim de superar esse problema, a inferência Bayesiana faz uso intenso de técnicas computacionais. A ideia é realizar a inferência por meio de simulação utilizando amostras da distribuição *a posteriori* para extrair informações sobre os parâmetros. Alguns dos principais métodos serão resumidos na próxima Seção.

Técnicas de Simulação

A inferência baseada em técnicas de simulação utiliza amostras da distribuição *a*

posteriori para extrair e sumarizar informações a respeito da quantidade de interesse. Segundo Gamerman e Lopes (2006), como a amostra é sempre um substituto parcial da informação contida em uma distribuição, métodos baseados em simulação são aproximações que devem ser utilizadas apenas quando constatada a impossibilidade de extração analítica das informações. Apesar disso, o avanço tecnológico tem ampliado o uso de métodos baseados em simulações.

Obter uma amostra da distribuição *a posteriori* não é uma tarefa fácil. Entretanto, mesmo quando a distribuição *a posteriori* não apresenta forma analítica, ainda é possível conhecer o núcleo desta distribuição (expressão sem a constante normalizadora). A partir desse núcleo várias técnicas para obtenção de uma amostra da distribuição *a posteriori* podem ser usadas. Para um vetor de parâmetros θ , Gamerman e Lopes (2006) descrevem como se procede a inferência para se obter uma amostra da componente θ_i , bastando tomar as outras componentes (θ_{-i}) como valores conhecidos que serão gerados de uma determinada distribuição. Este passo torna desnecessária a operação de integração para obtenção analítica da distribuição marginal mostrada em (0.2).

Existem hoje vários métodos que podem ser usados para se obter uma amostra da distribuição *a posteriori*, como por exemplo *rejection sampling* e *importance sampling*. Neste texto, apresentaremos apenas dois deles dentro de uma classe de métodos conhecidos como *Markov Chain Monte Carlo* (MCMC). Segundo Gelman et al. (2014), os métodos MCMC formam uma classe mais geral de métodos de simulação que extraem valores de distribuições aproximadas para θ e em seguida corrigem esses valores de forma a caminhar em direção à distribuição alvo. Estes métodos objetivam gerar uma amostra da distribuição *a posteriori* conhecendo somente o núcleo dessa distribuição, utilizando para isto, a informação da verossimilhança e da distribuição *a priori*. Dessa forma, o método busca, por meio de simulações iterativas, observações amostrais que tenham as características da distribuição alvo.

As amostras são obtidas sequencialmente, com a distribuição geradora dependente do último valor obtido numa amostragem anterior, assim como acontece em um cadeia de Markov. Para Gelman et al. (2014), uma cadeia de Markov pode ser definida como uma sequência de variáveis aleatória $\theta^{(1)}, \theta^{(2)}, \dots$, tal que a distribuição da variável θ

no instante atual (s), depende apenas da última realização conhecida dessa variável, ou seja $P(\theta^{(s)} \in A | \theta^{(0)}, \dots, \theta^{(s-1)}) = P(\theta^{(s)} \in A | \theta^{(s-1)})$, em que A é um subconjunto do espaço paramétrico, sendo essa distribuição chamada de transição da cadeia. Além disso, as cadeias possuem duas importantes propriedades que garantem sua eficiência em simulações, sendo elas: (a) homogeneidade: a probabilidade de uma nova realização da variável não altera com o tempo; (b) irreducibilidade: existe uma probabilidade positiva de que a cadeia visite todo espaço amostral. Segundo Gamerman e Lopes (2006) após várias realizações da cadeia, a distribuição de transição se aproxima da distribuição alvo de forma que os valores finais gerados da cadeia terão aproximadamente a distribuição *a posteriori*.

Para a composição da amostra *a posteriori*, devemos descartar os primeiros valores da cadeia que formam o chamado período de aquecimento (*burn-in*). Estas observações iniciais fazem parte de um período em que a cadeia ainda está se adaptando antes de atingir a convergência; ver Gamerman e Lopes (2006) para mais detalhes. Após a convergência, todos os valores gerados da cadeia serão provenientes da distribuição *a posteriori* formando assim nossa amostra final para inferência. Entretanto, como os valores gerados são resultado de uma cadeia de Markov, pode haver uma dependência (autocorrelação) muito forte entre eles, o que pode influenciar a estimação. Para superar este obstáculo, uma estratégia é selecionar para análise os valores gerados a cada k iterações (*lag*); esta amostra selecionada terá maior grau de independência.

Existem várias técnicas para amostrar da distribuição *a posteriori*, sendo descritas aqui apenas as duas principais pertencentes a classe de métodos MCMC: *Gibbs Sampling* e *Metropolis-Hastings*.

Gibbs Sampling

O *Gibbs Sampling* é um método MCMC bastante popular para a solução de problemas multidimensionais, sendo um caso particular do algoritmo *Metropolis-Hastings*. Essencialmente, trata-se de um esquema iterativo de amostragem de uma cadeia de Markov para o qual é necessário conhecer as distribuições condicionais completas *a posteriori* de cada parâmetro. O método foi usado pela primeira vez por Geman e Geman (1984)

em um artigo que utiliza distribuições condicionais completas aplicadas por um algoritmo iterativo definindo uma cadeia de Markov em um contexto de processamento de imagens. Mas foram Gelfand e Smith (1990) que compararam a proposta a outros esquemas de simulação tornando este método bastante conhecido na área da Estatística.

O uso dessa técnica fornece uma alternativa baseada em gerações sucessivas das distribuições condicionais completas, visto que gerar uma amostra diretamente da distribuição conjunta de interesse é complicado, tem alto custo e, muitas vezes, é impossível, pois não podemos identificar, no núcleo obtido via Teorema de Bayes, uma distribuição de probabilidade conhecida. A sequência de passos a seguir, apresentada por Gamerman e Lopes (2006), exemplifica o funcionamento do *Gibbs Sampling*. Considere o vetor $\boldsymbol{\theta}$ composto por k parâmetros, isto é, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$. Seja $P(\theta_i | \boldsymbol{\theta}_{-i}, D)$ a distribuição condicional completa de cada elemento que compõem $\boldsymbol{\theta}$. Um conjunto de valores iniciais $\{\theta_1^{(0)}, \dots, \theta_k^{(0)}\}$ deve ser sugerido. Para $s = 1, 2, \dots, S$, faça:

1. Gere $\theta_1^{(s)}$ a partir de $p(\theta_1 | \boldsymbol{\theta}_{-1}^{(s-1)}, D)$
2. Gere $\theta_2^{(s)}$ a partir de $p(\theta_2 | \theta_1^{(s)}, \theta_3^{(s-1)}, \dots, \theta_k^{(s-1)}, D)$
- ⋮
- k. Gere $\theta_k^{(s)}$ a partir de $p(\theta_k | \theta_1^{(s)}, \theta_2^{(s)}, \dots, \theta_{k-1}^{(s)}, D)$

Após o período de *burn-in* e a seleção final baseada no espaçamento *lag*, teremos uma amostra da distribuição conjunta alvo. A definição do *burn-in* pode ser feita de maneira informal, olhando para o gráfico da cadeia e localizando visualmente onde ocorreu a convergência, ou de maneira mais formal por meio de testes de convergência propostos na literatura. Um teste apresentado por Geweke (1992) consiste em dividir as observações geradas na cadeia em dois grandes blocos, calculando estimadores de média e variância para cada bloco que serão utilizados em uma estatística de teste de forma que se a diferença entre as médias de cada bloco for grande, há indicação de não convergência da cadeia. Outros testes são discutidos em Gamerman e Lopes (2006).

Metropolis-Hastings

Outro importante algoritmo é o *Metropolis-Hastings*, bastante usado, por exemplo, dentro do *Gibbs Sampling* quando a forma das distribuições condicionais completas também são desconhecidas. O algoritmo *Metropolis-Hastings* usa a mesma ideia de métodos de rejeição não pertencentes à classe MCMC, em que um valor é gerado de uma distribuição auxiliar e aceito com uma determinada probabilidade. Essa validação do valor gerado garante a convergência da distribuição de transição para a distribuição *a posteriori* desejada.

O primeiro passo do algoritmo consiste em escolher a distribuição auxiliar $g(\cdot)$ da qual os valores de θ serão gerados. Inicialmente a única exigência para escolha dessa distribuição é a necessidade de garantir as propriedades de irredutibilidade e homogeneidade da cadeia já citadas. No entanto, a eficiência do algoritmo é determinada por essa distribuição, que deve ser o mais próxima possível da distribuição de interesse. A sequência de passos para gerar uma amostra de acordo com o método é descrita por Gamerman e Lopes (2006) e Gelman et al. (2014), sendo apresentada a seguir:

1. Escolha um valor inicial para θ ;
2. Inicie o contador de iterações $s = 0$;
3. Gere $u^{(s+1)} \sim U(0, 1)$;
4. Gere θ^* da distribuição auxiliar escolhida $g(\cdot|\theta^{(s)})$;
5. Calcule a probabilidade de aceitação $r = \min \left\{ \frac{\pi(\theta^*)g(\theta^{(s)}|\theta^*)}{\pi(\theta^{(s)})g(\theta^*|\theta^{(s)})}, 1 \right\}$, em que $\pi(\cdot)$ representa o núcleo conhecido da distribuição *a posteriori*;
6. Se $u^{(s+1)} < r$ então $\theta^{(s+1)} = \theta^*$, caso contrário $\theta^{(s+1)} = \theta^{(s)}$;
7. Faça $s = s + 1$ e retorne ao terceiro passo.

Observe que caso o valor proposto θ^* não seja aceito, a cadeia não evolui. Quando há um grande número de rejeições recomenda-se alterar a distribuição auxiliar de forma a melhorar a taxa de aceitação r . Segundo Gamerman e Lopes (2006), o sucesso do método depende de taxas de aceitação não muito baixas e de distribuições auxiliares fáceis de

simular, de forma a substituir uma geração difícil da distribuição de interesse por várias outras gerações propostas da distribuição de transição. Novamente aqui, o período de aquecimento da cadeia e o *lag* devem ser respeitados.

Na maioria dos casos, a amostra *a posteriori* é gerada de uma combinação das duas técnicas apresentadas aqui, entretanto, outras técnicas de amostragem, como por exemplo *Adaptive Rejection Sampling* (ARS), *Adaptive Rejection Metropolis Sampling* (ARMS) e o *Slice Sampling*, também podem ser utilizadas. Para detalhes consulte Gamerman e Lopes (2006), Gelman et al. (2014) e Neal (2003).

Seleção de Modelos

Existem diversas metodologias que se propõem a avaliar se um modelo está adequado a um certo conjunto de dados, além de selecionar a melhor opção dentro de uma coleção de modelos. Para este trabalho, priorizamos alguns critérios de seleção, frequentemente usados em inferência Bayesiana, em que as amostras *a posteriori* para cada parâmetro são obtidas usando métodos MCMC.

O critério DIC (*Deviance Information Criterion*), inicialmente proposto por Spiegelhalter et al. (2002) como uma generalização do critério *Akaike* (AIC) (Akaike, 1974), é baseado na média *a posteriori* do desvio (\bar{D}) que pode ser aproximado por (Gelman et al., 2014):

$$\begin{aligned}\bar{D} &= \frac{1}{R} \sum_{r=1}^R D(\boldsymbol{\theta}_r), \\ D(\boldsymbol{\theta}_r) &= -2 \sum_{i=1}^n \log f(y_i | \boldsymbol{\theta}),\end{aligned}$$

sendo r o índice que indica o r -ésimo elemento da amostra *a posteriori* de tamanho R e $f(y_i | \boldsymbol{\theta})$ a função densidade de probabilidade correspondente ao modelo.

O DIC pode ser estimado utilizando $\widehat{DIC} = \bar{D} + \widehat{p}_D$, em que \widehat{p}_D é uma medida do número de parâmetros do modelo, $\widehat{p}_D = E[D(\boldsymbol{\theta})] - D[E(\boldsymbol{\theta})] = \bar{D} - \widehat{D}$. Gelman et al. (2014) sugere a seguinte estimativa para \widehat{D} :

$$\widehat{D} = D \left(\frac{1}{R} \sum_{r=1}^R \theta_r \right).$$

Dessa forma temos que $\widehat{DIC} = \bar{D} + \widehat{D}/2$, sendo que, no contexto de comparação, o modelo escolhido será aquele que apresentar o menor DIC.

Outro critério bastante utilizado, no contexto Bayesiano, é a densidade preditiva condicional ordenada (CPO). Segundo Gelman et al. (2014), o CPO pode ser interpretado como o peso da densidade marginal para a observação y_i , de forma que altos valores de CPO implicam em um melhor ajuste do modelo.

Seja D os dados completos e $D_{(-i)}$ os dados excluindo a i -ésima observação. A distribuição *a posteriori* do vetor de parâmetros $\boldsymbol{\theta}$ dado $D_{(-i)}$ é denotada por $p(\boldsymbol{\theta}|D_{(-i)})$. Para a i -ésima observação, o CPO_i pode ser escrito como (Gelman et al., 2014):

$$CPO_i = \int_{\Theta} f(y_i|\boldsymbol{\theta})p(\boldsymbol{\theta}|D_{(-i)})d\boldsymbol{\theta} = \int_{\Theta} \frac{p(\boldsymbol{\theta}|D)}{f(y_i|\boldsymbol{\theta})}d\boldsymbol{\theta}, \quad i = 1, \dots, n.$$

em que D_{-i} é o vetor de dados observados y sem a i -ésima observação.

Considerando que a modelagem proposta não permite obter uma expressão fechada para a solução desta integral, uma estimativa baseada na amostra da distribuição *a posteriori* resultante do MCMC pode ser obtida. Seja $\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \dots, \boldsymbol{\theta}^{(R)}$ uma amostra de tamanho R da distribuição *a posteriori* $p(\boldsymbol{\theta}|D)$. O CPO_i pode ser calculado pela seguinte aproximação (Chen et al., 2000):

$$\widehat{CPO}_i = \left[\frac{1}{R} \sum_{r=1}^R \frac{1}{f(y_i|\boldsymbol{\theta}^{(r)})} \right]^{-1}.$$

Pode-se resumir a informação do \widehat{CPO}_i utilizando a estatística chamada logaritmo pseudo marginal da verossimilhança (*Logarithm of the Pseudomarginal Likelihood*) $LPML = \sum_{i=1}^n \log \widehat{CPO}_i$. De forma análoga, temos que os maiores valores de LPML indicam o melhor modelo no ajuste dos dados.

Apêndice B: Distribuições *a posteriori*.

- Modelo $M0$.

Seja $D = \{(y_i, \delta_i, x_i), i = 1, \dots, n\}$, então:

$$P(\alpha_T | D, \Theta_T) \propto \alpha_T^{\sum_{i=1}^n \delta_i + a_T - 1} \prod_{i=1}^n \left[y_i^{\delta_i(\alpha_T - 1)} \right] \exp \left\{ - \left[b_T \alpha_T + \sum_{i=1}^n y_i^{\alpha_T} e^{x_i} \beta_T \right] \right\};$$

$$P(\beta_{T,j} | D, \Theta_T) \propto \exp \left\{ \frac{-V_{T,j}}{2} \left[\beta_{T,j}^2 - 2\beta_{T,j} \left(\frac{\sum_{i=1}^n \delta_i x_{ji}}{V_{T,j}} - \frac{e^{\sum_{i=1}^n \delta_i x_i} \beta_T}{\beta_{T,j} V_{T,j}} \right) \right] \right\},$$

em que $j = 0, 1, 2$.

- Modelo $M1$.

Seja $D = \{(y_i, \delta_i, \mathbf{x}_i, w_i), i = 1, \dots, n\}$, então:

$$P(\alpha_T | D, \Theta_T, \psi) \propto \alpha_T^{\sum_{i=1}^n \delta_i + a_T - 1} \prod_{i=1}^n \left[y_i^{\delta_i(\alpha_T - 1)} \right] \exp \left\{ - \left[b_T \alpha_T + \sum_{i=1}^n y_i^{\alpha_T} e^{x_i} \beta_{T+w_i} \right] \right\};$$

$$P(\beta_{T,j} | D, \Theta_T, \psi) \propto \exp \left\{ \frac{-V_{T,j}}{2} \left[\beta_{T,j}^2 - 2\beta_{T,j} \left(\frac{\sum_{i=1}^n \delta_i x_{ji}}{V_{T,j}} - \frac{e^{\sum_{i=1}^n \delta_i x_i} \beta_{T+w_i}}{\beta_{T,j} V_{T,j}} \right) \right] \right\};$$

$$P(\tau | D, \Theta_T) \propto \tau^{(n+m_1)-1} \exp \left\{ - \left(\frac{\sum_{i=1}^n w_i^2}{2} + m_2 \right) \tau \right\},$$

em que $j = 0, 1, 2$.

A distribuição condicional completa de τ é a Gama $\left[(n + m_1), \left(\frac{\sum_{i=1}^n w_i^2}{2} + m_2 \right) \right]$. Entretanto para os demais parâmetros a distribuição *a posteriori* não possui forma conhecida.

- Modelo $M2$.

Seja $D = \{(y_i, \delta_i, \mathbf{x}_i, w_i), i = 1, \dots, n\}$, então:

$$\begin{aligned}
P(\alpha_C | \Theta, D) &\propto \alpha_C^{(n + \sum_{i=1}^n \delta_i + a_C - 1)} \prod_{i=1}^n \left[y_i^{(1-\delta_i)(\alpha_C - 1)} \right] \exp \left\{ - \left[b_C \alpha_C + \sum_{i=1}^n y_i^{\alpha_C} e^{x_i \beta_C + \phi w_i} \right] \right\}; \\
P(\beta_{C,j} | \Theta, D) &\propto \exp \left\{ \frac{-V_{C,j}}{2} \left[\beta_{C,j}^2 - 2\beta_{C,j} \left(\frac{\sum_{i=1}^n x_{ij}}{V_{C,j}} - \frac{\sum_{i=1}^n \delta_i x_{ij}}{V_{C,j}} - \frac{\sum_{i=1}^n \delta_i x_i \beta_C + \phi w_i}{\beta_{C,j} V_{C,j}} \right) \right] \right\}; \\
P(\phi | \Theta, D) &\propto \exp \left\{ \phi \sum_{i=1}^n w_i - \phi \sum_{i=1}^n \delta_i w_i - \sum_{i=1}^n e^{x_i \beta_C + \phi w_i} y_i^{\alpha_C} \right\}.
\end{aligned}$$

- Modelo M3.

Seja $D = \{(y_i, \delta_i, \mathbf{x}_i, w_i, z_i), i = 1, \dots, n\}$, então:

$$\begin{aligned}
P(\alpha_C | \Theta, D) &\propto \alpha_C^{(n + \sum_{i=1}^n \delta_i + a_C - 1)} \prod_{i=1}^n \left[y_i^{(1-\delta_i)(\alpha_C - 1)} \right] \exp \left\{ - \left[b_C \alpha_C + \sum_{i=1}^n y_i^{\alpha_C} e^{x_i \beta_C + (Z_i - 2)w_i} \right] \right\}; \\
P(\beta_{C,j} | \Theta, D) &\propto \exp \left\{ \frac{-V_{C,j}}{2} \left[\beta_{C,j}^2 - 2\beta_{C,j} \left(\frac{\sum_{i=1}^n x_{ij}}{V_{C,j}} - \frac{\sum_{i=1}^n \delta_i x_{ij}}{V_{C,j}} - \frac{\sum_{i=1}^n \delta_i x_i \beta_C + (Z_i - 2)w_i}{\beta_{C,j} V_{C,j}} \right) \right] \right\}; \\
P(\mathbf{q} | Z) &\propto \prod_{j=1}^3 q_j^{I_{(Z_i=j)} + \eta_j - 1}.
\end{aligned}$$

Observe que a distribuição *a posteriori* condicional completa de \mathbf{q} tem núcleo conhecido, sendo uma Dirichlet($\boldsymbol{\eta}^*$), em que $\eta_j^* = I_{(Z_i=j)} + \eta_j$, para $j = 1, 2, 3$ e $I_{(Z_i=j)} = 1$ se $Z_i = j$ (0 caso contrário).

Apêndice C: Convergência da amostra *a posteriori*.

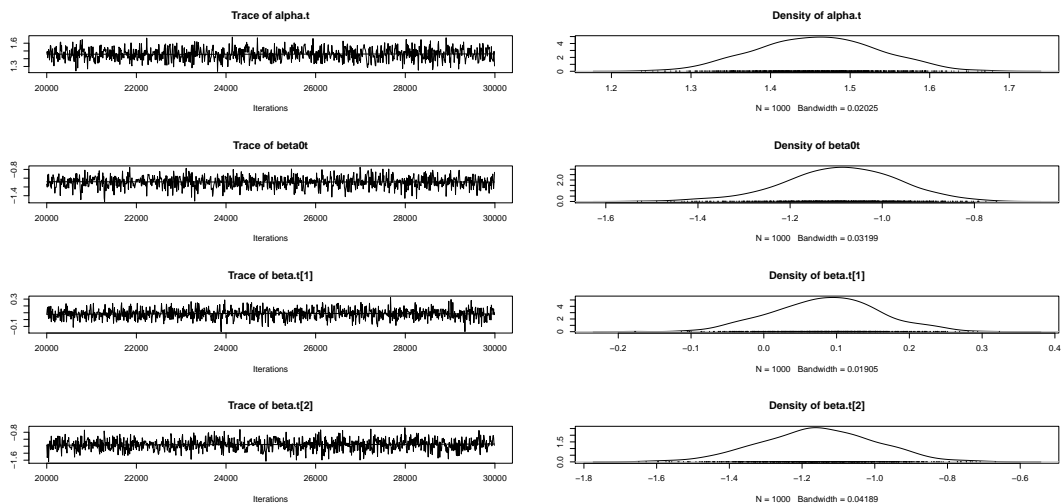


Figura C.1: Traço da cadeia indicando convergência e densidade *a posteriori* para os parâmetros da distribuição de falha relativos ao modelo M0 considerando amostras de 250 tempos no cenário independente com grande proporção de falha.

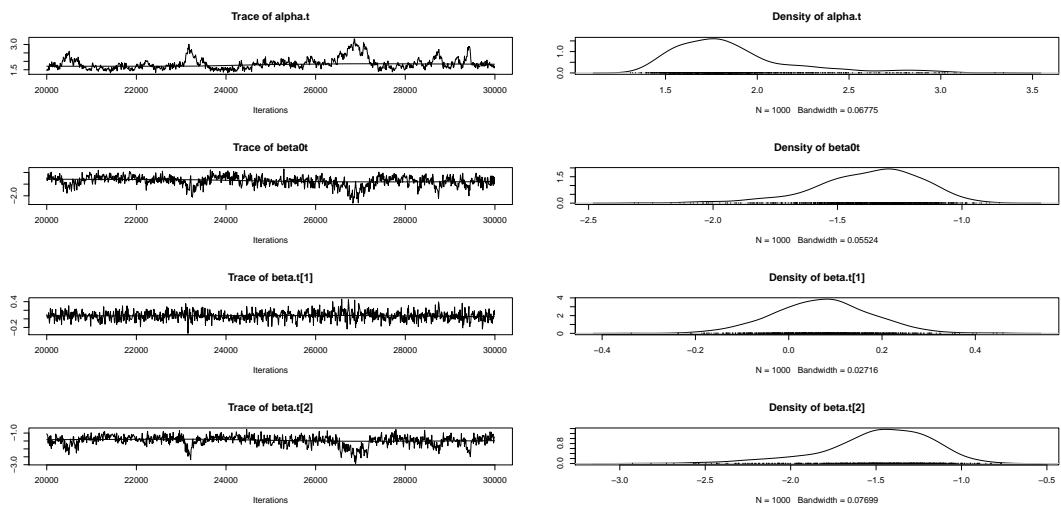


Figura C.2: Traço da cadeia indicando convergência e densidade *a posteriori* para os parâmetros da distribuição de falha relativos ao modelo M1 considerando amostras de 250 tempos no cenário independente com grande proporção de falha.

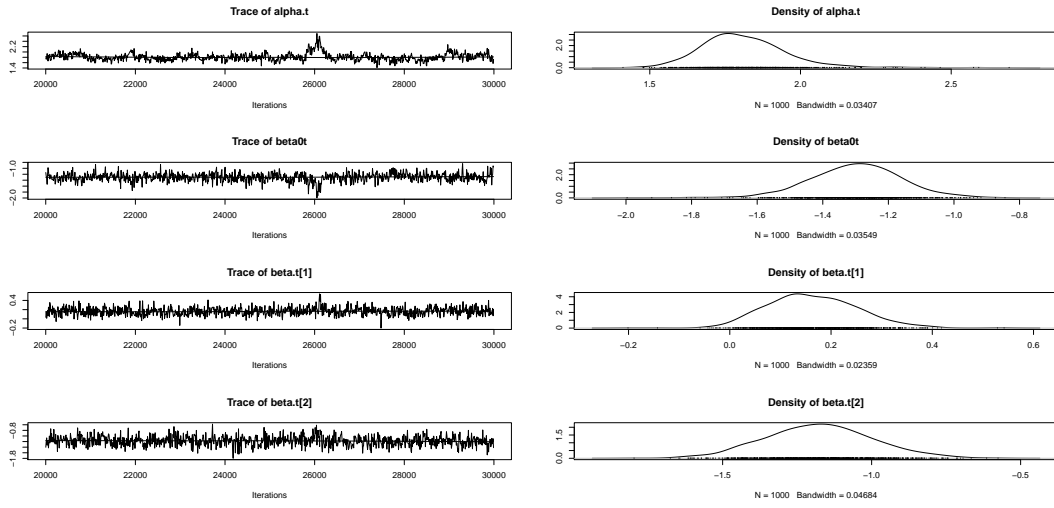


Figura C.3: *Traço da cadeia indicando convergência e densidade a posteriori para os parâmetros da distribuição de falha relativos ao modelo M2 considerando amostras de 250 tempos no cenário independente com grande proporção de falha.*

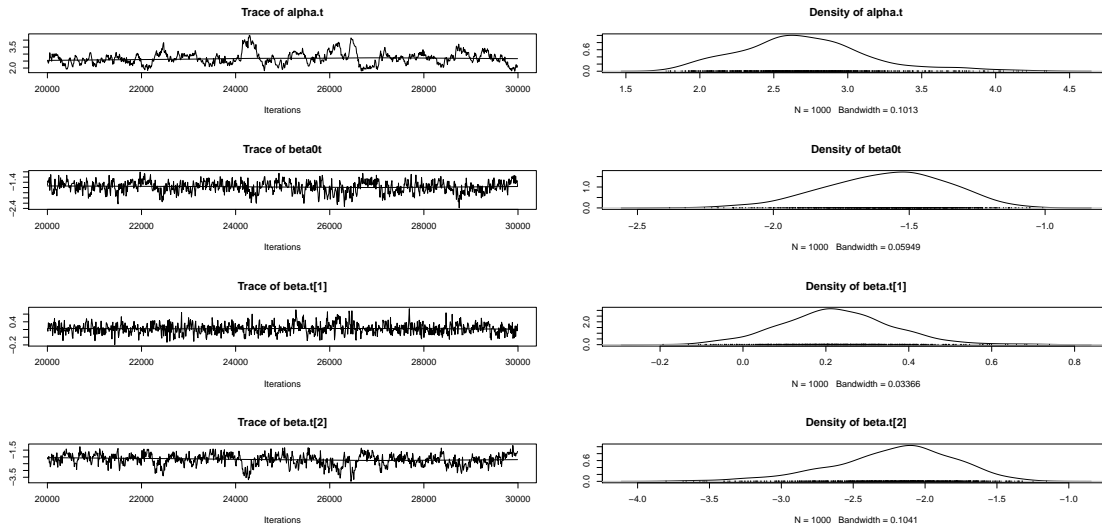


Figura C.4: *Traço da cadeia indicando convergência e densidade a posteriori para os parâmetros da distribuição de falha relativos ao modelo M3 considerando amostras de 250 tempos no cenário independente com grande proporção de falha.*

Apêndice D: Gráficos adicionais para o estudo simulado.

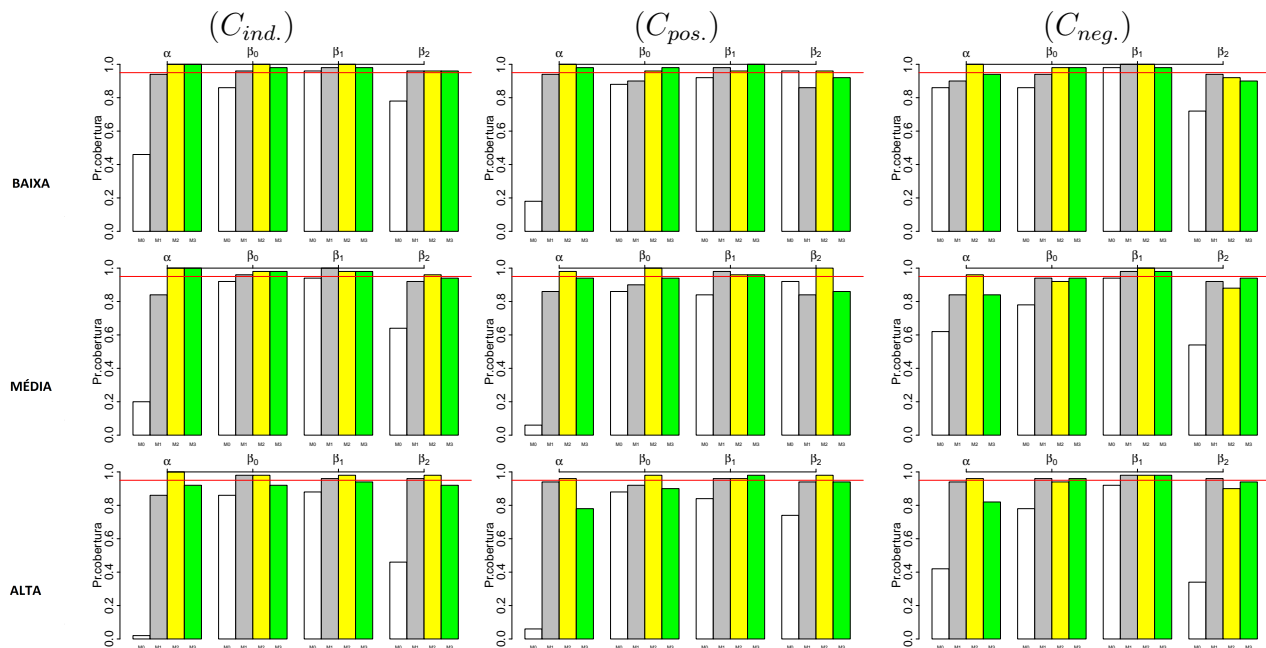


Figura D.1: Probabilidade de cobertura registrada nas 50 replicações do MCMC para o ajuste de uma amostra de 50 tempos.

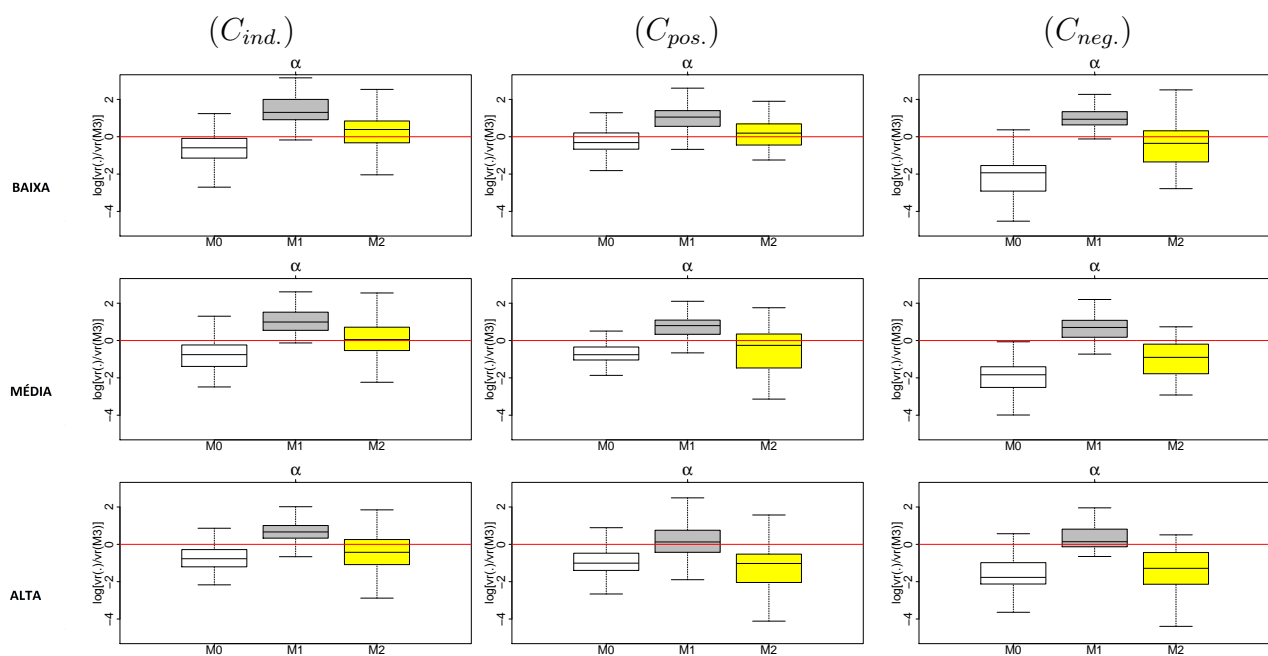


Figura D.2: *Boxplots dos valores do log do valor absoluto da razão de vícios relativos de α obtidos nas 50 replicações do MCMC para ajustes considerando amostras de 50 tempos. O modelo de referência no denominador da razão é o M3.*

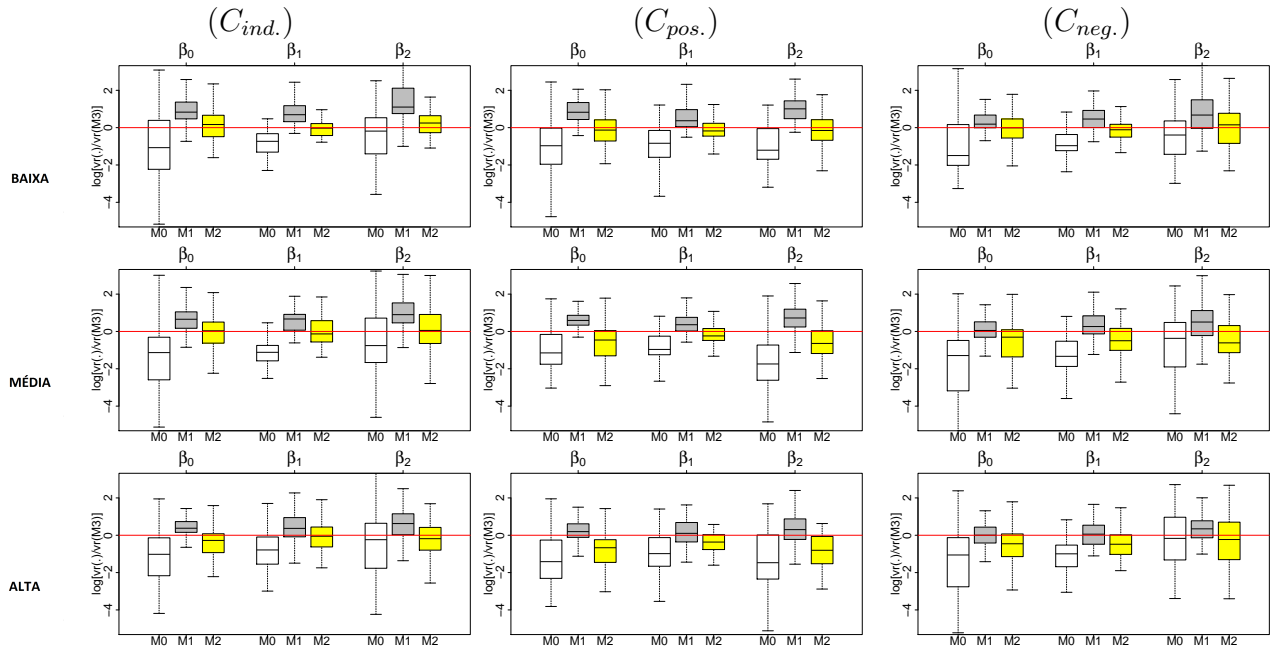


Figura D.3: *Boxplots dos valores do log do valor absoluto da razão de vícios relativos de β obtidos nas 50 replicações do MCMC para ajustes considerando amostras de 50 tempos. O modelo de referência no denominador da razão é o M3.*

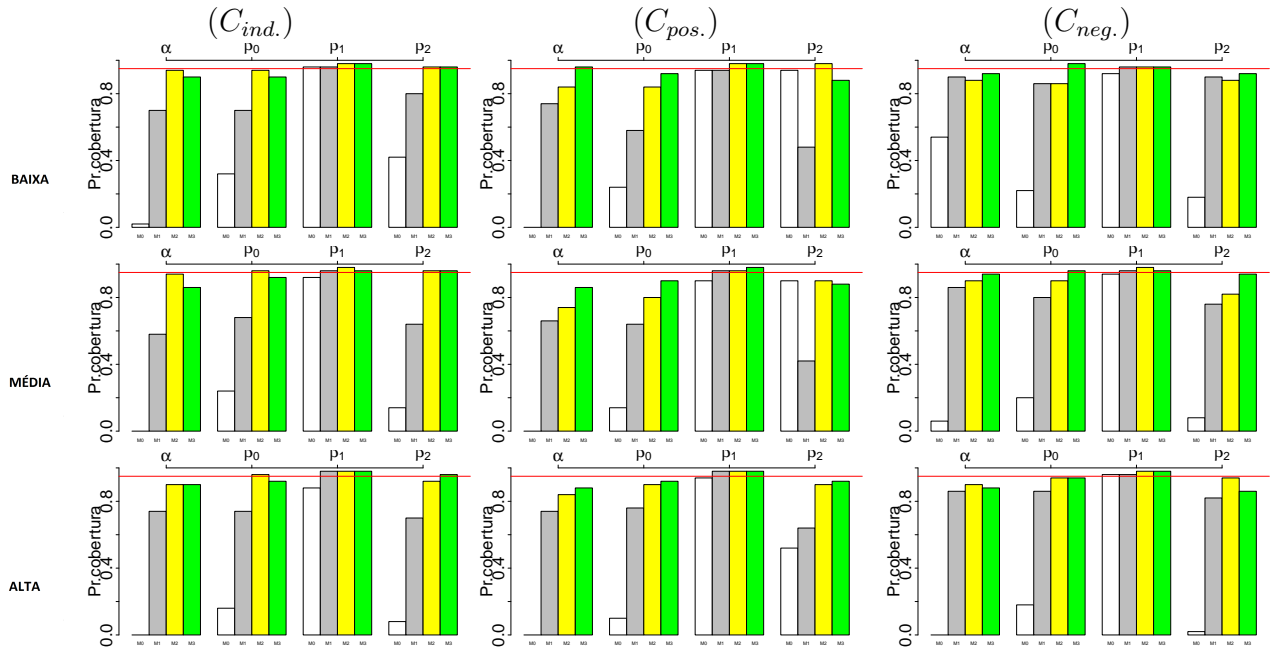


Figura D.4: *Probabilidade de cobertura registrada nas 50 replicações do MCMC para o ajuste de uma amostra de 250 tempos.*

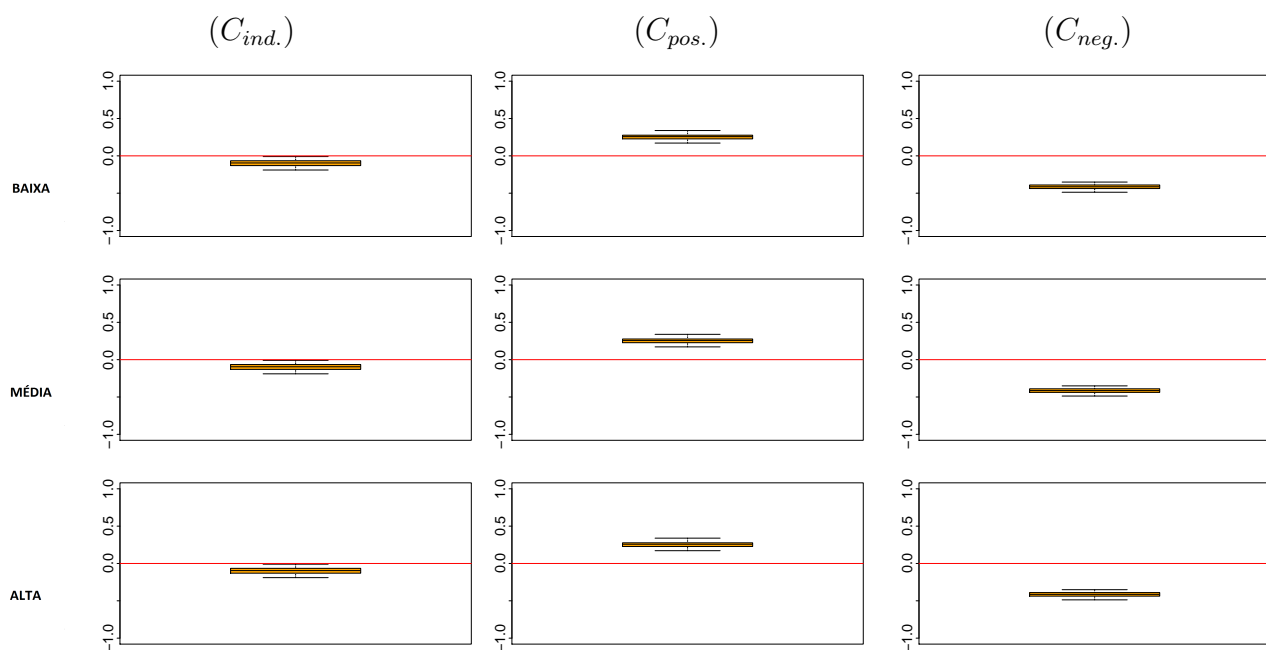


Figura D.5: Correlação entre os tempos de falha e censura registrada nas 50 replicações do MCMC para o ajuste de uma amostra de 250 tempos.

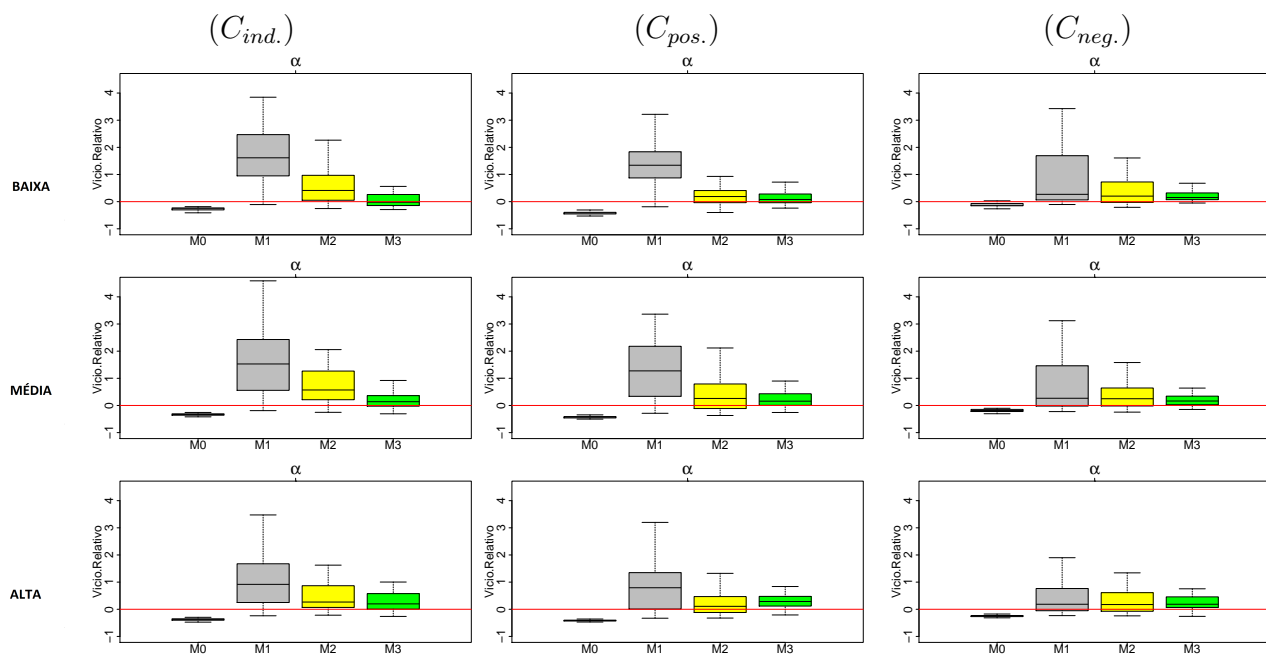


Figura D.6: Boxplots dos vícios relativos para α obtidos em cada replicação do MCMC para ajustes considerando amostras de 250 tempos.

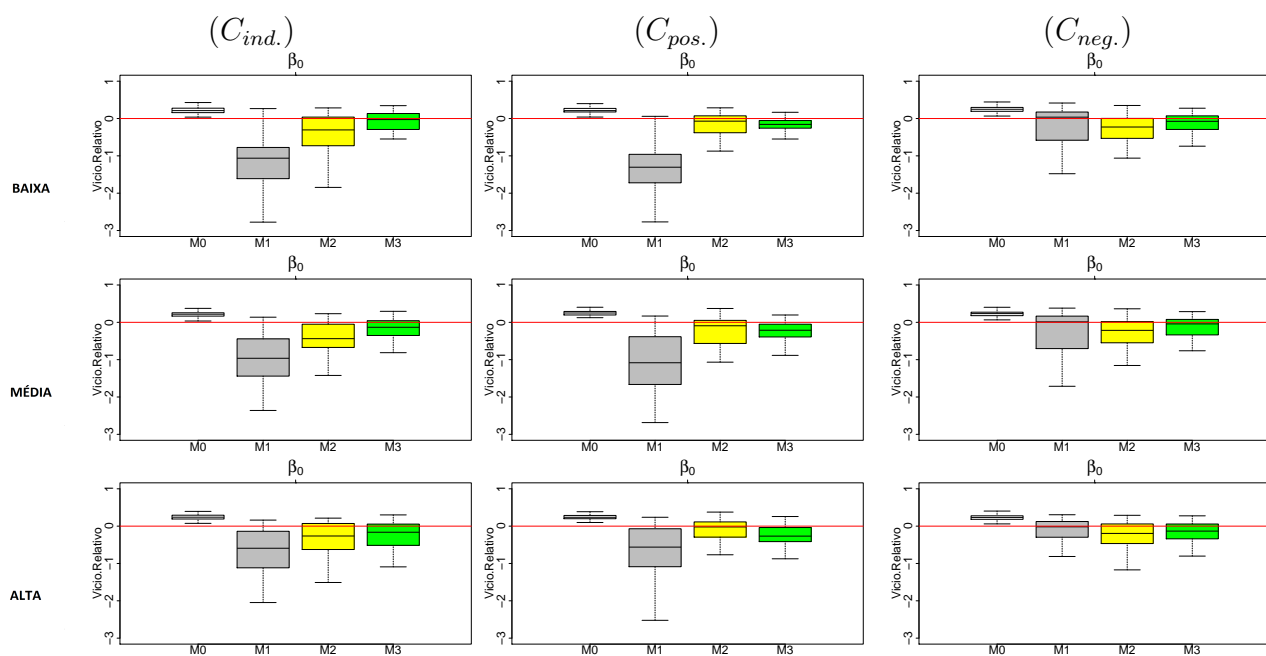


Figura D.7: *Boxplots dos vícios relativos para β_0 obtidos em cada replicação do MCMC para ajustes considerando amostras de 250 tempos.*

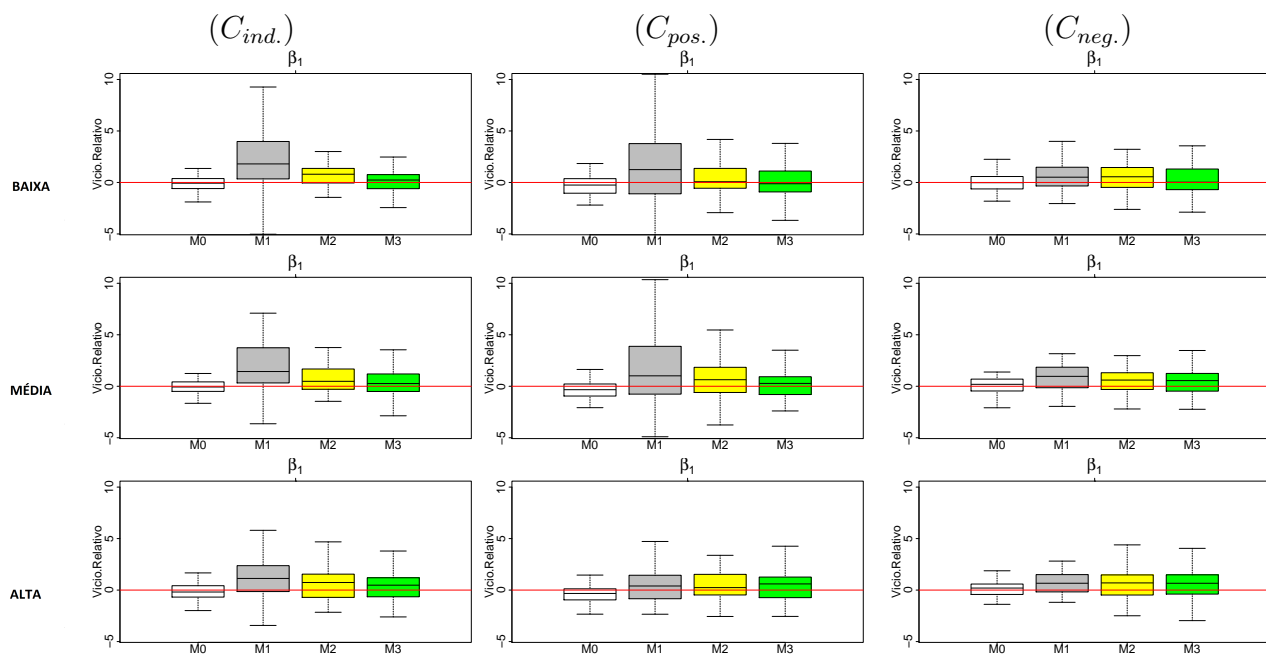


Figura D.8: *Boxplots dos vícios relativos para β_1 obtidos em cada replicação do MCMC para ajustes considerando amostras de 250 tempos.*

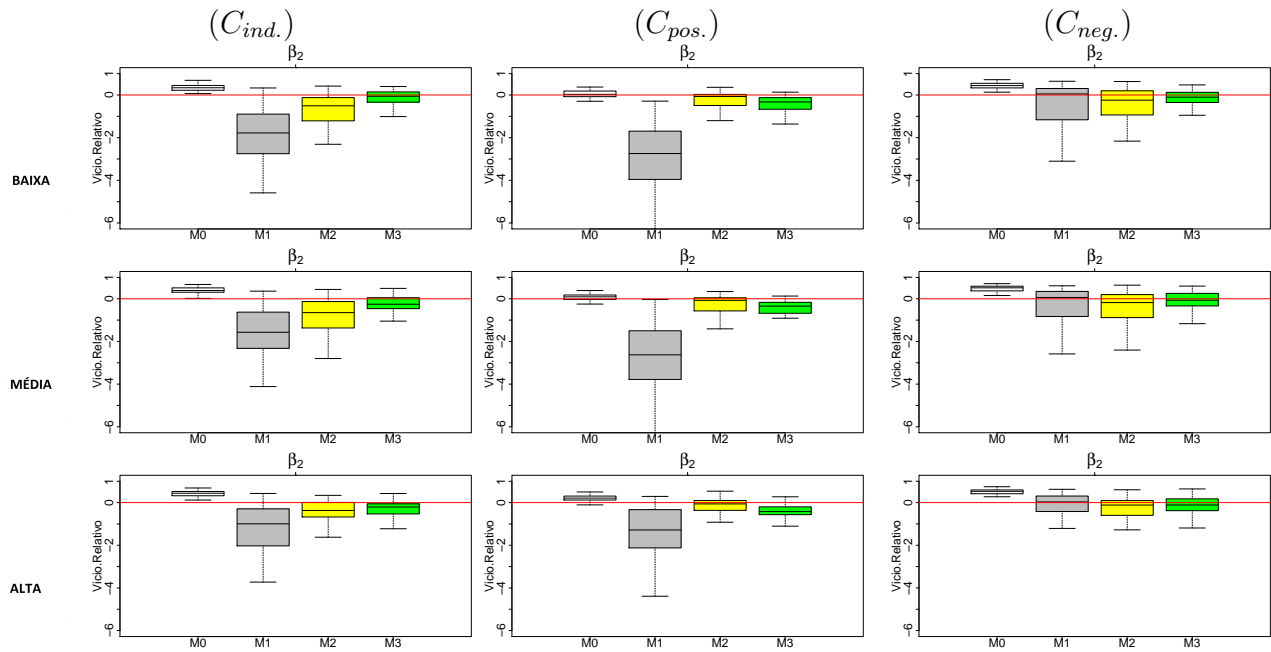


Figura D.9: *Boxplots dos vícios relativos para β_2 obtidos em cada replicação do MCMC para ajustes considerando amostras de 250 tempos.*

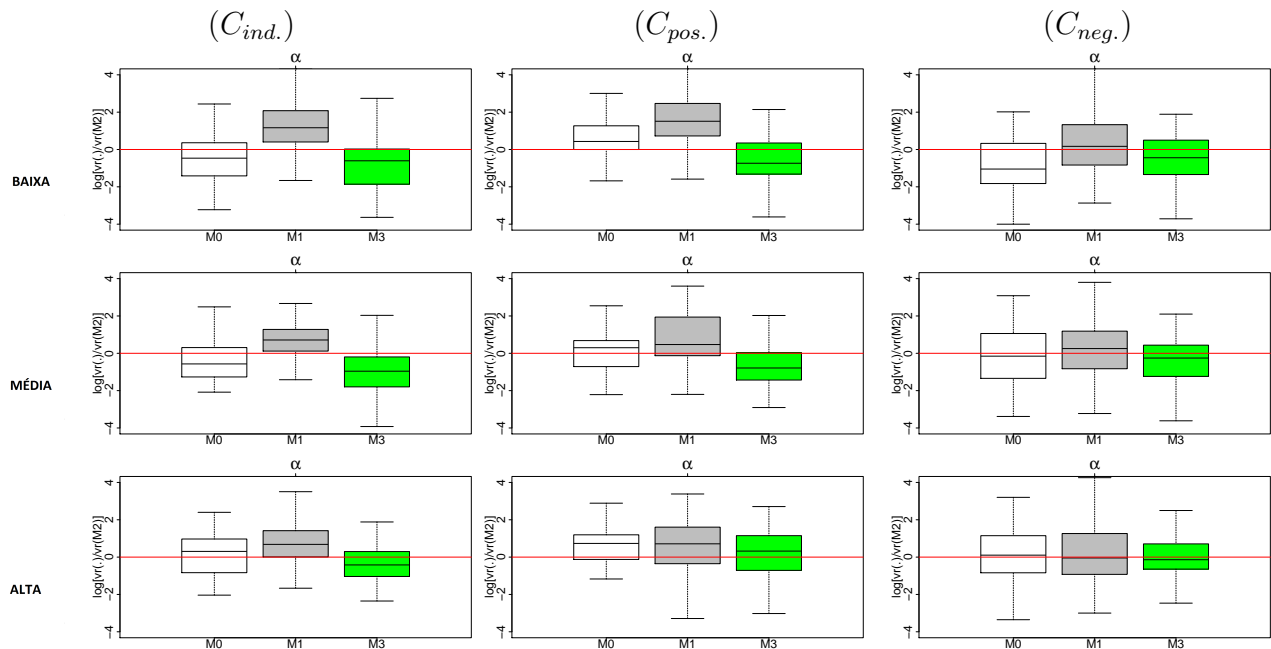


Figura D.10: *Boxplots dos valores do log do valor absoluto da razão de vícios relativos de α obtidos nas 50 replicações do MCMC para ajustes considerando amostras de 250 tempos. O modelo de referência no denominador da razão é o M2.*

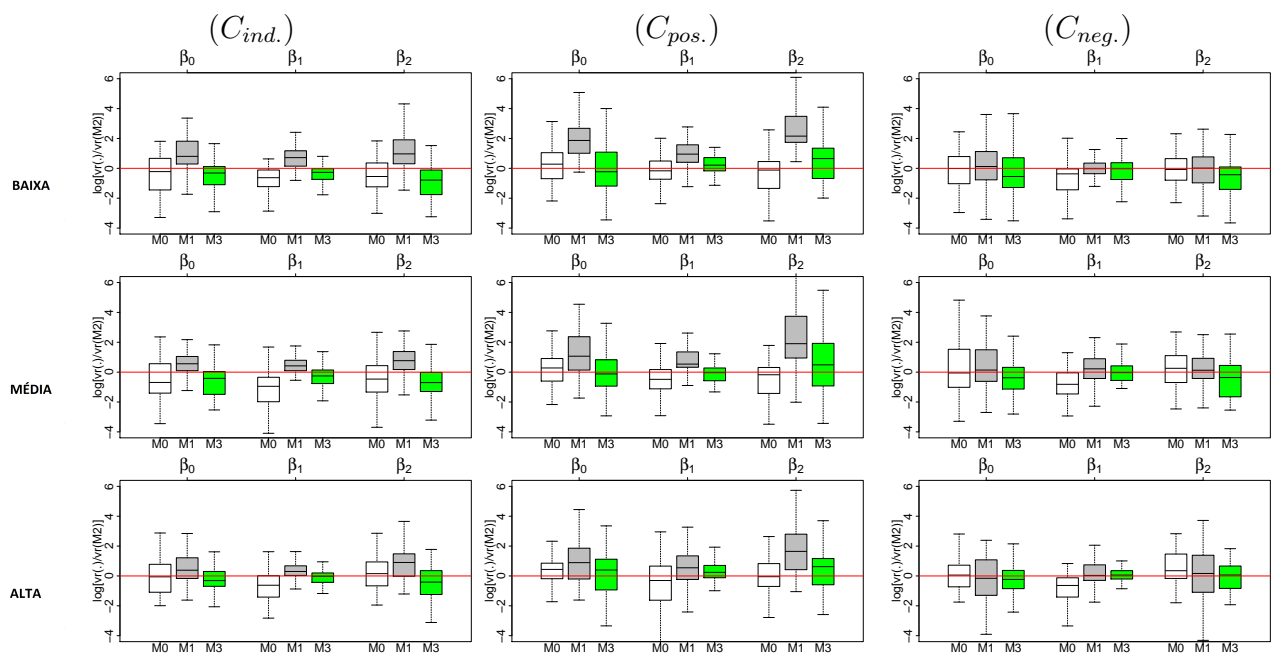


Figura D.11: *Boxplots dos valores do log do valor absoluto da razão de vícios relativos de β obtidos nas 50 replicações do MCMC para ajustes considerando amostras de 250 tempos. O modelo de referência no denominador da razão é o M2.*

Referências Bibliográficas

- Akaike, H. (1974), “A new look at the statistical model identification,” *I.E.E.E Transactions on Automatic Control*, 19, 716–723.
- Banerjee, S., Wall, M. M., e Carlin, B. P. (2003), “Frailty Modeling for Spatially Correlated Survival Data, with Application to Infant Mortality in Minnesota,” *Biostatistics*, 4, 123–142.
- Bender, R., Augustin, T., e Blettner, M. (2005), “Generating Survival Times to Simulate Cox Proportional Hazards Models,” *Statistics in Medicine*, 24, 1713–1723.
- Bolfarine, H. e Sandoval, M. (2002), *Introdução a Inferência Estatística*, SBM.
- Chen, M. H., Shao, Q. M., e Ibrahim, J. (2000), *Monte Carlo Methods in Bayesian Computation*, Springer, New York.
- Collett, D. (2003), *Modelling Survival Data in Medical Research*, Chapman & Hall, Boca Raton, 2^a edn.
- Colosimo, E. e Giolo, S. (2006), *Análise de Sobrevivência Aplicada*, Edgar Blücher, São Paulo, 1^a edn.
- Cox, D. R. (1972), “Regression Models and Life-Tables,” *Journal of the Royal Statistical Society*, 34, 187–220, Series B (Methodological).
- Dellaportas, P. e Smith, A. F. M. (1993), “Bayesian Inference for Generalized Linear and Proportional Hazards Models via Gibbs Sampling,” *Journal of Royal Statistical Society*, 42, 443–459, Series C (Applied Statistics).

- Emoto, S. E. e Matthews, P. C. (1990), “A Weibull Model for Dependent Censoring,” *The Annals of Statistics*, 18, 1556–1577.
- Gamerman, D. e Lopes, H. F. (2006), *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, vol. 68, Chapman and Hall/CRC, London, 2 edn.
- Gelfand, A. E. e Smith, A. F. M. (1990), “Sampling-Based Approaches to Calculating Marginal Densities,” *Journal of the American Statistical Association*, 85, 398–409.
- Gelman, A., Carlin, J. B., Stern, A. S., e Rubin, D. B. (2014), *Bayesian Data Analysis*, Chapman & Hall, Boca Raton, 3^a edn.
- Geman, S. e Geman, D. (1984), “Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.
- George, E. I. e McCulloch, R. E. (1997), “Approaches for Bayesian Variable Selection,” *Statistica Sinica*, 7, 339–373.
- Geweke, J. (1992), “Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments,” *Bayesian Statistics*, 4, 625–631, J.M. Bernardo, J. Berger, A.P. Dawid and A.F.M. Smith.
- Geweke, J. (1996), “Variable selection and model comparison in regression,” *Bayesian Statistics*, 5, 609–620.
- Gilks, W. R. e Wild, P. (1992), “Adaptive Rejection Sampling for Gibbs Sampling,” *Journal of the Royal Statistical Society*, 41, 337–348, Series C (Applied Statistics).
- Hanagal, D. D. (2011), *Modeling Survival Data Using Frailty Models*, Chapman & Hall, Boca Raton.
- Huang, X. e Wolfe, R. A. (2002), “A Frailty Model for Informative Censoring,” *Biometrics*, 58, 510–520.
- Ibrahim, J. G., Chen, M.-H., e Sinha, D. (2010), *Bayesian Survival Analysis*, Springer, New York, USA.

- Jiang, H., Fine, J. P., Kosorok, M. R., e Chappell, R. (2005), “Pseudo Self - Consistent Estimation of Copula Model with Informative Censoring,” *Scandinavian Journal of Statistics*, 32, 1–20.
- Kaciroti, N. A., Raghunathan, T. E., e Taylor, J. M. G. (2012), “A Bayesian model for Time-to-event Data with Informative Censoring,” *Biostatistics*, 13, 341–354.
- Knaus, J. (2013), *snowfall: Easier cluster computing (based on snow)*., R package version 1.84-6.
- Kosorok, M. R., Fine, J. P., Jiang, H., e Chappell, R. (2002), “Asymptotic Theory for the Gama Frailty Model with Dependent Censoring,” *Annals of the Institute of Statistical Mathematics*, 54, 476–499.
- Krall, J. M., Uthoff, V. A., e Harley, J. B. (1975), “A Step-up Procedure for Selecting Variables Associated with Survival,” *Biometrics*, 31, 49–57.
- Lawless, J. F. (2003), *Statistical Models an Methods for Lifetime Data*, John Wiley & Sons, New Jersey, 2^a edn.
- Leung, K., Elashoff, R. M., e Afifi, A. A. (1997), “Censoring Issues in Survival Analysis,” *Annual Review of Public Health*, 18, 83–104.
- Link, W. A. (1989), “A Model for Informative Censoring,” *Journal of the American Statistical Association*, 84, 749–752.
- Lu, Z. e Zhang, W. (2012), “Semiparametric Likelihood Estimation in Survival Models with Informative Censoring,” *Journal of Multivariate Analysis*, 106, 187–211.
- Maiolino, A. e Magalhães, R. J. R. (2015), *Mieloma Múltiplo*, ABRALÉ, São Paulo - SP.
- Neal, R. M. (2003), “Slice Sampling,” *The Annals of Statistics*, 31, 705–741.
- Plummer, M. (2003), “JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling,” .

- Plummer, M., Best, N., Cowles, K., e Vines, K. (2006), “CODA: Convergence Diagnosis and Output Analysis for MCMC,” *R News*, 6, 7–11.
- R Core Team (2014), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
- Robert, C. P. (2007), *The Bayesian Choice*, Springer, New York, 2º edn.
- Rocha, C. S. (2011), “Modelos com Fragilidade: Aplicação à Modelagem da Heterogeneidade não Observada,” *Boletim da Sociedade Portuguesa de Estatística*, 4, 26–34.
- Santos, P. C. J. (2012), “Análise de Sobrevivência na Presença de Censura Informativa,” Dissertação de Mestrado, UFMG - Departamento de Estatística - ICEX.
- Scharfstein, D. O. e Robins, J. M. (2002), “Estimation of the failure time distribution in the presence of informative censoring,” *Biometrika*, 89, 617–634.
- Siannis, F. (2004), “Applications of a Parametric Model for Informative Censoring,” *Biometrics*, 60, 704–714.
- Siannis, F., Copas, J., e Lu, G. (2005), “Sensitivity Analysis for Informative Censoring in Parametric Survival Models,” *Biostatistics*, 6, 77–91.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., e Lin, A. V. D. (2002), “Bayesian measures of model complexity and fit,” *Journal of the Royal Statistical Society*, 64, 583–639, Series B (Statistical Methodology).
- Staplin, N. (2012), “Informative Censoring in Transplantation Statistics,” Tese de Doutorado, University of Southampton - Faculty of Social and Human Sciences - Mathematics.
- Su, Y.-S. e Yajima, M. (2014), *R2jags: A Package for Running JAGS from R*, R package version 0.04-03.
- Suzuki, A. K. (2012), “Modelos de Sobrevivência Bivariados Baseados na Cópula FGM: Uma abordagem Bayesiana,” Tese de Doutorado, UFSCar - Departamento de Estatística, São Paulo.

- Tsiatis, A. A. (1981), “A Large Sample Study of Cox’s Regression Model,” *The Annals of Statistics*, 9, 93–108.
- Vaupel, J. W., Manton, K. G., e Stallard, E. (1979), “The impact of heterogeneity in individual frailty on dynamics of mortality,” *Demography*, 16, 439–454.
- Watanabe, S. (2010), “Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory,” *Journal of Machine Learning Research*, 11, 3571–3594.
- Wu, M. C. e Carroll, R. J. (1988), “Estimation and Comparison of Changes in the Presence of Informative Right Censoring by Modeling the Censored Process,” *Biometrics*, 44, 175–188.