



Universidade Federal de Minas Gerais  
Instituto de Ciências Exatas e da Terra  
Programa de Pós-Graduação em Estatística

**Anna Rafaella da Silva Marinho**

**Modelos Bayesianos com Fração de Cura e Erro de Medida com  
Distribuição na Família Mistura de Escala da Normal**

Belo Horizonte, 30 de Agosto de 2018

Anna Rafaella da Silva Marinho

## **Modelos Bayesianos com Fração de Cura e Erro de Medida com Distribuição na Família Mistura de Escala da Normal**

Tese de doutorado<sup>a</sup> apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Minas Gerais, em cumprimento com as exigências legais para obtenção do título de Doutor.

---

<sup>a</sup>Pesquisa parcialmente financiada pela CAPES.

Orientadora:  
Prof<sup>a</sup>. Dr<sup>a</sup>. Rosângela H. Loschi

Belo Horizonte, 30 de Agosto de 2018

# Dedicatória

À minha mãe, Graça,  
meus irmãos  
e a Alyson, com amor.

# Agradecimentos

Agradeço, primeiramente a Deus por te me concedido a graça de chegar até aqui.

À minha mãe Graça, por todo esforço em prol da minha educação, pelo amor e apoio despendido. À toda a minha família, irmãos, D. Terezinha e S. Expedito; aos amigos queridos, pelo constante apoio emocional e toda a torcida pela conclusão deste trabalho.

Agradeço especialmente ao meu noivo, Alyson, pela compreensão, amor e companheirismo em todas as horas e principalmente naquelas mais difíceis de dúvidas e estresse. Sem você não sei se teria chegado até aqui. Eu e Rosângela, lhe somos gratas por todo o suporte computacional que permitiu realizar todas as simulações necessárias em tempo hábil.

À minha orientadora, Rosângela, pela parceria, confiança, conselhos, ensinamentos e amizade. Lhe sou eternamente grata por me guiar na construção dessa tese.

À minha orientadora do mestrado, Débora, por me aconselhar e apoiar a seguir em frente.

Aos professores das disciplinas cursadas durante o doutorado: Duczmal, Marco, Rosângela, Flávio e Roger, seus ensinamentos foram fundamentais para o desenvolvimento deste trabalho e para o meu crescimento acadêmico.

Aos professores Reinaldo e Vinícius que estiveram em minha banca de qualificação e de defesa de tese, à Lourdes e Josemar que completaram a banca da defesa, pelas valiosas sugestões à melhoria deste trabalho e conselhos para trabalhos futuros.

À todos os professores que participaram da minha vida escolar e acadêmica. Cada um tem alguma contribuição à formação da profissional que sou hoje.

Aos amigos de pós-graduação: Juliana, Juliane, Lívia, Rachel, Marcela, Mariana, Larissa, Nívea, Paulo, Zayda, Cristiano, Guilherme, José Luiz, Luís Gustavo, Jussiane, Douglas, Fádua, Raquel, Maurício, Fernanda, Uriel, Ana Gabriela e a todos os outros colegas da pós, pelas horas de estudo compartilhado, conselhos e amizade. Que felicidade foi compartilhar tantos momentos com vocês!

Agradeço a todos os funcionários do Depto de Estatística da UFMG pelos serviços prestados e a Maísa pelo café fresquinho e chá de toda tarde.

À capes pelo auxílio financeiro.

”Todo caminho leva a algum lugar, mas não leva a todos os lugares. Trilhos e trilhas sempre tem um destino certo, e eventualmente, chega-se no fim da linha, aquele lugar que interrompe a trajetória das vigas de aço, com uma parede de madeira, que é abrupta e inconveniente, como são todas as cruciais mudanças da vida. E se a existência passa a ser um projeto de constantes deslocamentos, não seria ingenuidade acreditar no alcance do ”paraíso“? “

**Henrique Haddefinir**

# Resumo

Os avanços médicos no tratamento do câncer e o desenvolvimento de técnicas de diagnóstico mais eficientes nos últimos anos, contribuíram para o aumento do número de pacientes curados. Com isso também houve um crescimento no interesse de se desenvolver modelos estatísticos capazes de incorporar uma fração de cura em sua estrutura. É sabido que algumas covariáveis relevantes ao tempo de vida dos pacientes podem ser medidas com erro. Neste trabalho, propomos modelos Bayesianos de fração de cura que acomodam covariáveis desta natureza. Nossa proposta estende modelos já discutidos na literatura por: (i) assumir que o erro de medida pertence a classe de distribuições mistura de escala da Normal (família NI); (ii) por estimar a variância do erro de medida e, também, por (iii) considerar abordagens Bayesianas para modelos já propostos. Sugerimos três especificações *a priori* diferentes para fazermos inferência. Em todas estas, as distribuições *a posteriori* não têm forma fechada. Por essa razão, lançamos mão do amostrador de *Gibbs* com passos de *Metropolis* adaptativo para obter amostras das distribuições *a posteriori* dos parâmetros. Para avaliar os modelos propostos, apresentamos um extenso estudo de simulação e também uma aplicação na área médica, analisando dados de um ensaio clínico de pacientes com melanoma cutâneo já investigado na literatura.

**Palavras-chave:** Modelo de Tempo de Promoção. Modelo Estrutural. Covariáveis Medidas com Erro. Inferência Bayesiana. Métodos MCMC.

# Abstract

The medical advances in cancer treatment and the development of efficient diagnosis techniques in the recent years have contributed to the increase in the fraction of cured patients. Because of this, it can be noted an increasing the interest to develop statistical models to appropriately deal with lifetime data in the presence of cure fraction. It is well known that some covariates that may influence the patient lifetime can be mismeasured. In this work, we propose a Bayesian cure rate model that accommodates covariates of this nature. Our proposal extends models that have already been discussed in the literature by: (i) assuming that the measurement error belongs to the scale mixture of normal distributions class; (ii) by estimating the variance of the measurement error and also by (iii) considering Bayesian approaches to the models already proposed. Three different prior specifications are proposed to model this parameter. In all models, the expressions of the posterior distributions does not have a closed form. For this reason we used the Gibbs Sampler with Adaptive Metropolis to obtain samples from the posterior distributions. A Monte Carlo simulation study is presented and also an analysis of a melanoma clinical trial data that has already been explored in the literature.

**Keywords:** Promotion time cure rate model. Structural Model. Mismeasured Covariates. Bayesian Inference. MCMC Methods.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>20</b>
1.1	Considerações Iniciais . . . . .	20
1.2	Modelo Linear com Erro nas Covariáveis . . . . .	23
1.3	Modelos de Fração de Cura . . . . .	25
1.3.1	Modelo de Tempo de Promoção (MTP) . . . . .	25
1.3.2	Modelo de Fração de Cura com Erro nas Covariáveis de Mizoi <i>et al.</i> (2007) . . . . .	30
1.3.3	Modelo de Fração de Cura com Erro nas Covariáveis de Ma e Yin (2008) . . . . .	31
<b>2</b>	<b>Modelo Bayesiano de Fração de Cura (MBFC) com Distribuição Normal para o Erro de Medida</b>	<b>36</b>
2.1	Inferência Bayesiana no MBFC . . . . .	38
2.1.1	MBFC1: MBFC com $\sigma_\epsilon^2$ fixado . . . . .	41
2.1.2	MBFC2: $\sigma_\epsilon^2$ é estimado . . . . .	43
2.1.3	MBFC3: $\sigma_\epsilon^2 = k\sigma_x^2$ é estimado . . . . .	44
2.2	Análise Bayesiana do Modelo de Ma e Yin (2008) . . . . .	45
2.3	Estudo de Simulação de Monte Carlo para Avaliar o MBFC . . . . .	50
2.3.1	Estudo 1: O Efeito de Supor a Variância do Erro Conhecida . . . . .	51
2.3.2	Estudo 2: O Efeito de Ignorar o Erro de Medida Presente nos Dados . . . . .	57
2.3.3	Estudo 3: O Efeito de Pressupor Erro de Medida Quando Não Existe . . . . .	66
2.3.4	Estudo 4: Avaliando a Suposição de Normalidade para $\mathbf{X}_\epsilon$ . . . . .	69
2.3.5	Estudo 5: Avaliando a Escolha da Distribuição <i>a priori</i> da constante $k$ no MBFC3 . . . . .	74
2.3.6	Comentários Gerais . . . . .	79
2.4	Aplicação . . . . .	81
2.4.1	Análise Assumindo Distribuição Weibull para os Tempos de Falha . . . . .	83
2.4.2	Análise Assumindo Distribuição Exponencial para os Tempos de Falha . . . . .	96



<b>3</b>	<b>Modelo Bayesiano de Fração de Cura com Erro de Medida Normal Independente (MEFC-NI)</b>	<b>108</b>
3.1	MBFC com Erro de Medida Normal Independente . . . . .	109
3.2	Inferência Bayesiana no MBFC-NI . . . . .	111
3.2.1	MBFCNI1: MBFC-NI com $\sigma_\epsilon^2$ conhecido . . . . .	112
3.2.2	MBFC-NI2: MBFC-NI com $\sigma_\epsilon^2$ desconhecido . . . . .	115
3.2.3	MBFC-NI3: MBFC-NI com $\sigma_\epsilon^2 = k\sigma_x^2$ . . . . .	116
3.3	Estudo de Simulação de Monte Carlo para Avaliar o MBFC-NI . . . . .	117
3.3.1	Estudo 1: O Efeito de Supor $\nu$ Conhecido . . . . .	119
3.3.2	Estudo 2: O Efeito de Pressupor Erro Quando Não Existe . . . . .	128
3.3.3	Estudo 3: Avaliando a Robustez do Modelo . . . . .	135
3.4	Aplicação . . . . .	146
<b>4</b>	<b>Considerações Finais</b>	<b>156</b>
<b>A</b>	<b>Família de Distribuições Normal Independente</b>	<b>165</b>
<b>B</b>	<b>Distribuição <i>a posteriori</i> dos Parâmetros do MBFC-NI</b>	<b>170</b>
<b>C</b>	<b>Alguns Critérios de Seleção de Modelos</b>	<b>172</b>
<b>D</b>	<b>Código para Gerar Dados com Erro de Medida Normal</b>	<b>175</b>
<b>E</b>	<b>Código para Gerar Dados com Erro de Medida <i>t-Student</i></b>	<b>178</b>
<b>F</b>	<b>Programa Implementado para Simular o MBFC2</b>	<b>181</b>
<b>G</b>	<b><i>Box-plots</i> das Estimativas da Fração de Cura Obtidas nos Estudos de Monte Carlo da Seção 2.3</b>	<b>197</b>
<b>H</b>	<b>Gráficos Diagnósticos de Convergência</b>	<b>210</b>

# Lista de Figuras

2.1	Distribuições <i>a priori</i> de $\sigma_x^2$ , $GI(102, 1010)$ (linha sólida), $GI(100, 100)$ (linha tracejada) e $GI(0.1, 0.1)$ (linha pontilhada). . . . .	52
2.2	Médias do valor esperado <i>a posteriori</i> para a fração de cura dos indivíduos do Cenário 1, ajustando-se o MBFC1 nos casos em que fixamos $\sigma_\epsilon^2 = 1$ (linha com $\Delta$ ), $\sigma_\epsilon^2 = 0.5$ (linha com $+$ ) e $\sigma_\epsilon^2 = 10$ (linha com $*$ ) e supondo $\sigma_x^2 \sim GI(0.1, 0.1)$ (a), $\sigma_x^2 \sim GI(102, 1010)$ (b) e $\sigma_x^2 \sim GI(100, 100)$ (c). A linha com $\circ$ é referente aos valores reais de fração de cura para cada indivíduo. . . . .	55
2.3	Histograma das frações de cura reais superposto pelas distribuições empíricas das médias dos valores esperados <i>a posteriori</i> da fração de cura, obtidas ajustando-se o MBFC1, supondo $\sigma_x^2 \sim GI(0.1, 0.1)$ (Gráfico a), $\sigma_x^2 \sim GI(102, 1010)$ (Gráfico b) e $\sigma_x^2 \sim GI(100, 100)$ (Gráfico c), fixando $\sigma_\epsilon^2 = 1$ (linha sólida), $0.1$ (linha pontilhada), $10$ (linha traço-ponto). . . . .	56
2.4	$X_\epsilon$ (estimado) $\times X_2$ . Nos gráficos (b), (c) e (d) temos as estimativas obtidas ajustando-se o MBFC1 quando $\sigma_\epsilon^2 = 1$ , MBFC2 e MBFC3, respectivamente. No gráfico (a), correspondente ao MTP, temos os valores observados $(X_\epsilon + \epsilon) \times X_2$ . . . . .	60
2.5	$X_\epsilon$ (estimado) $\times X_2$ . Nos gráficos (b), (c) e (d) temos as estimativas obtidas ajustando-se o MBFC1 quando $\sigma_\epsilon^2 = 5$ , MBFC2 e MBFC3, respectivamente. No gráfico (a), correspondente ao MTP, temos os valores observados $(X_\epsilon + \epsilon) \times X_2$ . . . . .	60
2.6	Histograma das frações de cura reais superpostos pelas distribuições empíricas das médias dos valores esperados <i>a posteriori</i> da fração de cura, ajustando-se todos os modelos, considerando o Cenário 1 (Gráfico a) e o Cenário 2 (Gráfico b). . . . .	61
2.7	Estimativas médias do valor esperado <i>a posteriori</i> para fração de cura dos indivíduos dos Cenário 1 e 2 Fig.a e Fig.b, respectivamente, ajustando-se os modelos: MTP (linha com $\times$ ), MBFC1 ( $\sigma_\epsilon^2 = 1$ – linha com $\circ$ ), MBFC2 (linha com $\diamond$ ) e MBFC3 (linha com $\nabla$ ). $\square$ representa os valores reais da fração de cura. . . . .	62

2.8	Viés do valor esperado <i>a posteriori</i> para a fração de cura dos indivíduos dos Cenários 1 e 2, Fig.a e Fig.b, respectivamente, ajustando-se os modelos: MTP (linha com ×), MBFC1 ( $\sigma_\epsilon^2 = 1$ – linha com o, $\sigma_\epsilon^2 = 0.5$ – linha com $\triangle$ , $\sigma_\epsilon^2 = 5$ – linha com +), MBFC2 (linha com $\diamond$ ) e MBFC3 (linha com $\nabla$ ). $\square$ representa os valores reais da fração de cura. . . . .	63
2.9	Histograma do erro de medida em 9 das 250 réplicas de Monte Carlo. As curvas sobre cada histograma representam as densidades de distribuições Normais centradas em zero e cuja a variância foi fixada em 1, no caso do MBFC1 e as densidades empíricas de amostras do erro de medida estimadas aos ajustarmos o MBFC2 e o MBFC3, considerando o Cenário 1. . . . .	64
2.10	Histograma do erro de medida em 9 das 250 réplicas de Monte Carlo. As curvas sobre cada histograma representam as densidades de distribuições Normais centradas em zero e cuja a variância foi fixada em 1, 0.5 e 5, no caso do MBFC1 e as densidades empíricas de amostras do erro de medida estimadas aos ajustarmos o MBFC2 e o MBFC3, considerando o Cenário 2. . . . .	65
2.11	$X_\epsilon$ no eixo das abscissas representa as estimativas obtidas para $X_2$ ajustando-se o MBFC1 quando $\sigma_\epsilon^2 = 1$ (Gráfico a), o MBFC2 (Gráfico b) e o MBFC3 (Gráfico c). . . . .	67
2.12	Médias (a) e viés (b) do valor esperado <i>a posteriori</i> para a fração de cura de cada um dos 100 indivíduos do Estudo 3, ajustando-se os modelos: MTP, MBFC1, MBFC2 e MBFC3. . . . .	68
2.13	Histograma das frações de cura reais superposto pelas distribuições empíricas das estimativas médias dos valores esperados <i>a posteriori</i> para a fração de cura, obtidas ajustando-se os modelos MTP e MBFC1 a MBFC3. . . . .	68
2.14	$X_\epsilon$ no eixo das abscissas representa as estimativas obtidas para $X_2$ ajustando-se o MBFC2 quando a distribuição <i>a priori</i> de $\sigma_x^2$ é $GI(0.1, 0.1)$ (Gráfico a), $GI(3, 20)$ (Gráfico b), $GI(12, 110)$ (Gráfico c) e $GI(102, 1010)$ (Gráfico d). . . . .	71
2.15	$X_\epsilon$ no eixo das abscissas representa as estimativas obtidas para $X_2$ ajustando-se o MBFC3 quando a distribuição <i>a priori</i> de $\sigma_x^2$ é $GI(0.1, 0.1)$ (Gráfico a), $GI(3, 20)$ (Gráfico b), $GI(12, 110)$ (Gráfico c) e $GI(102, 1010)$ (Gráfico d). . . . .	71
2.16	Estimativas médias do valor esperado <i>a posteriori</i> para a fração de cura dos indivíduos do Cenário 4, obtidas ajustando-se os modelos MBFC2 (Fig.a) e MBFC3 (Fig.b), assumindo que <i>a priori</i> $\sigma_x^2$ tem distribuição: $GI(0.1, 0.1)$ (linha com ×), $GI(3, 20)$ (linha com +), $GI(12, 20)$ (linha com $\nabla$ ) e $GI(102, 1010)$ (linha com o). O símbolo $\square$ representa os valores reais da fração de cura. . . . .	72
2.17	Viés das estimativas médias do valor esperado <i>a posteriori</i> para a fração de cura dos indivíduos do Cenário 4, obtidas ajustando-se os modelos MBFC2 (Gráfico a) e MBFC3 (Gráfico b), assumindo que <i>a priori</i> $\sigma_x^2$ tem distribuição: $GI(0.1, 0.1)$ (linha com ×), $GI(3, 20)$ (linha com +), $GI(12, 20)$ (linha com $\nabla$ ) e $GI(102, 1010)$ (linha com o). O símbolo $\square$ representa os valores reais da fração de cura. . . . .	73
2.18	Histograma das frações de cura reais superposto pelas distribuições empíricas das estimativas médias dos valores esperados <i>a posteriori</i> da fração de cura, obtidas ajustando-se os modelos MEFC2 (Gráfico a) e MEFC3 (Gráfico b). . . . .	73

2.19	$X_\epsilon$ no eixo das abscissas representa as estimativas obtidas para $X_2$ ajustando-se o MBFC3 aos dados do Cenário 1, supondo como distribuições <i>a priori</i> para $k$ : Gama(0.1, 1) (Gráfico a), Gama(0.5, 0.001) (Gráfico b), $U(0, 0.5)$ (Gráfico c) e $U(0, 1)$ (Gráfico d). . . . .	76
2.20	$X_\epsilon$ no eixo das abscissas representa as estimativas obtidas para $X_2$ ajustando-se o MBFC3 aos dados do Cenário 2 supondo como distribuição <i>a priori</i> para $k$ : Gama(0.5, 0.001) (Gráfico a), Gama(0.5, 5) (Gráfico b), Gama(0.5, 50) (Gráfico c) e $U(0, 1)$ (Gráfico d). . . . .	76
2.21	Estimativas médias do valor esperado <i>a posteriori</i> para a fração de cura dos indivíduos dos os Cenários 1 (Gráfico a) e 2 (Gráfico b), ajustando-se o MBFC3 e assumindo diferentes distribuições <i>a priori</i> para $k$ : Gama(0.1, 1) (linha com $\circ$ ), Gama(0.5, 0.001) (linha com $\times$ ), $U(0, 0.5)$ (linha com $+$ ) e $U(0, 1)$ (linha com $\nabla$ ). O símbolo $\square$ representa os valores reais da fração de cura. . . . .	77
2.22	Viés das estimativas médias do valor esperado <i>a posteriori</i> para a fração de cura dos indivíduos do Cenário 1 (Gráfico a) e 2 (Gráfico b), ajustando-se o MBFC3 e assumindo diferentes distribuições <i>a priori</i> para $k$ : Gama(0.1, 1) (linha com $\circ$ ), Gama(0.5, 0.001) (linha com $\times$ ), $U(0, 0.5)$ (linha com $+$ ) e $U(0, 1)$ (linha com $\nabla$ ). . . . .	78
2.23	Histograma das frações de cura reais superposto pelas distribuições empíricas das médias dos valores esperados <i>a posteriori</i> da fração de cura, obtidas ajustando-se o MBFC3 considerando os Cenários 1 (Gráfico a) e 2 (Gráfico b). . . . .	79
2.24	Curvas de sobrevivência para os 234 pacientes do estudo E1684. . . . .	82
2.25	Fração de cura calculada para pacientes com idade $X_1 = 15, 35, 55, 75$ anos, do sexo feminino ( $X_2 = 1$ ), submetidos ao tratamento ( $X_3 = 1$ ), com profundidade de Breslow $X_4 = 1, \dots, 10$ e ativo $X_5 = 1$ . Considerando as estimativas para os coeficientes de regressão obtidas ajustando-se o MTP (Gráfico a), o MBFC1 (Gráfico b), o MBFC2 (Gráfico c) e o MBFC3 (Gráfico d). . . . .	85
2.26	HPD's de 95% para a variável latente $X_\epsilon$ dos primeiros 100 indivíduos, obtidos ajustando-se o MBFC1. Os pontos destacados são os valores observados da covariável $X_4$ para cada um dos indivíduos. . . . .	86
2.27	HPD's de 95% para a variável latente $X_\epsilon$ dos primeiros 100 indivíduos, obtidos ajustando-se o MBFC2. Os pontos destacados são os valores observados da covariável $X_4$ para cada um dos indivíduos. . . . .	86
2.28	HPD's de 95% para a variável latente $X_\epsilon$ dos primeiros 100 indivíduos, obtidos ajustando-se o MBFC3. Os pontos destacados são os valores observados da covariável $X_4$ para cada um dos indivíduos. . . . .	87
2.29	<i>Box-plots</i> das estimativas para a média <i>a posteriori</i> da fração de cura de todos os indivíduos observados, considerando os ajustes dos modelos MBFC1, 2 e 3 e MTP, respectivamente. . . . .	88
2.30	<i>Box-plots</i> das estimativas para a média <i>a posteriori</i> da fração de cura de todos os indivíduos observados que receberam doses de Interferon ( $X_3 = 1$ ) ou que ficaram apenas em observação ( $X_3 = 0$ ), considerando os ajustes dos modelos MBFC1, 2, 3 e MTP, respectivamente. . . . .	88
2.31	<i>Box-plots</i> das estimativas para a média <i>a posteriori</i> da fração de cura de todos os indivíduos observados para cada faixa etária considerada, sob o ajuste dos modelos MBFC1 (Gráfico a), 2 (Gráfico b), 3 (Gráfico c) e MTP (Gráfico d). . . . .	89

2.32	<i>Box-plots</i> das estimativas para a média <i>a posteriori</i> da fração de cura de todos os indivíduos observados para cada faixa etária considerada divididas entre os indivíduos que receberam Interferon ( $X_3 = 1$ ) e os que permaneceram apenas em observação ( $X_3 = 0$ ), sob o ajuste dos modelos MBFC1 (Gráfico a), 2 (Gráfico b), 3 (Gráfico c) e MTP (Gráfico d). . . . .	91
2.33	<i>Box-plots</i> das estimativas para a média <i>a posteriori</i> da fração de cura de todos os indivíduos observados que receberam doses de Interferon (Gráfico a) ou que ficaram somente em observação (Gráfico b) e são do sexo feminino (F) ou do sexo masculino (M), considerando cada um dos modelos ajustados. . . . .	92
2.34	<i>Box-plots</i> das estimativas para a média <i>a posteriori</i> da fração de cura de todos os indivíduos observados que receberam doses de Interferon e possuem nódulo do Tipo I, II, III e IV, considerando os ajustes dos modelos MBFC1 (Gráfico a), 2 (Gráfico b), 3 (Gráfico c) e MTP (Gráfico d). . . . .	93
2.35	Divergência K-L. Os gráficos (a), (b), (c) e (d) são relativos aos ajustes dos modelos MTP, MBFC1, 2 e 3, respectivamente. . . . .	94
2.36	Distribuições <i>a posteriori</i> para a fração de cura. As linhas mista, sólida, tracejada e pontilhada são das distribuições obtidas ajustando-se os modelos MTP, MBFC1, 2 e 3, respectivamente. . . . .	95
2.37	HPD's de 95% para a variável latente $X_\epsilon$ dos primeiros 100 indivíduos, obtidos ajustando-se o MBFC1. Os pontos destacados são os valores observados da covariável $X_4$ para cada um dos indivíduos. . . . .	97
2.38	HPD's de 95% para a variável latente $X_\epsilon$ dos primeiros 100 indivíduos, obtidos ajustando-se o MBFC2. Os pontos destacados são os valores observados da covariável $X_4$ para cada um dos indivíduos. . . . .	98
2.39	HPD's de 95% para a variável latente $X_\epsilon$ dos primeiros 100 indivíduos, obtidos ajustando-se o MBFC3. Os pontos destacados são os valores observados da covariável $X_4$ para cada um dos indivíduos. . . . .	98
2.40	HPD's de 95% para a variável latente $X_\epsilon$ dos primeiros 100 indivíduos, obtidos ajustando-se o MBFC-Ma. Os pontos destacados são os valores observados da covariável $X_4$ para cada um dos indivíduos. . . . .	99
2.41	<i>Box-plots</i> das estimativas para a média <i>a posteriori</i> da fração de cura de todos os indivíduos observados, considerando os ajustes dos modelos MBFC1, 2 e 3, MTP e MBFC-Ma, respectivamente. . . . .	99
2.42	<i>Box-plots</i> das estimativas para a média <i>a posteriori</i> da fração de cura de todos os indivíduos observados que receberam doses de Interferon ( $X_3 = 1$ ) ou que ficaram apenas em observação ( $X_3 = 0$ ), considerando os ajustes dos modelos MBFC1, 2 e 3, MTP e MBFC-Ma, respectivamente. . . . .	100
2.43	<i>Box-plots</i> das estimativas para a média <i>a posteriori</i> da fração de cura de todos os indivíduos observados que receberam doses de Interferon e são do sexo feminino (F) ou do sexo masculino (M), considerando cada um dos modelos ajustados. . . . .	101
2.44	<i>Box-plots</i> das estimativas para a média <i>a posteriori</i> da fração de cura de todos os indivíduos observados que permaneceram apenas em observação e são do sexo feminino (F) ou do sexo masculino (M), considerando cada um dos modelos ajustados. . . . .	101
2.45	<i>Box-plots</i> das estimativas para a média <i>a posteriori</i> da fração de cura de todos os indivíduos observados que receberam doses de Interferon e possuem nódulo do Tipo I, II, III e IV, considerando os ajustes dos modelos MBFC1 (Gráfico a), 2 (Gráfico b) e 3 (Gráfico c), MTP (Gráfico d) e MBFC-Ma (Gráfico e). . . . .	102

2.46	<i>Box-plots</i> das estimativas para a média <i>a posteriori</i> da fração de cura de todos os indivíduos observados para cada faixa etária considerada, sob o ajuste dos modelos MEFC1 (Gráfico a), 2 (Gráfico b) e 3 (Gráfico c), MTP (Gráfico d) e MEFC-Ma (Gráfico e). . . . .	103
2.47	<i>Box-plots</i> das estimativas para a média <i>a posteriori</i> da fração de cura de todos os indivíduos observados para cada faixa etária considerada divididas entre os indivíduos que receberam Interferon ( $X_3 = 1$ ) e os que permaneceram apenas em observação ( $X_3 = 0$ ), sob o ajuste dos modelos MBFC1 (Gráfico a), 2 (Gráfico b) e 3 (Gráfico c), MTP (Gráfico d) e MEFC-Ma (Gráfico e). . . . .	105
2.48	Divergência K-L. Os gráficos (a), (b), (c), (d) e (e) são relativos aos ajustes dos modelos MTP, MBFC1, MBFC2, MBFC3 e MBFC-Ma, respectivamente. . . . .	106
2.49	Distribuições <i>a posteriori</i> para a fração de cura. As linhas mista, sólida, traços pequenos, pontilhada e traços grandes são das distribuições obtidas ajustando-se os modelos MTP, MBFC1, MBFC2, MBFC3 e MBFC-Ma, respectivamente. . . . .	107
3.1	Comparação entre as f.d.p da distribuição <i>t</i> -Student considerando $\nu = 2.5, 3, 5, 10$ e $30$ . . . . .	119
3.2	<i>Box-plots</i> das estimativas médias para $\sigma_\epsilon^2$ ajustando-se o modelo MBFC-NI2. Os gráficos (a), (b), (c), (d) e (e) são relativos aos ajustes supondo $\nu = 2.5, 3, 5, 10$ e $30$ , respectivamente. . . . .	122
3.3	<i>Box-plots</i> das estimativas médias para $\sigma_\epsilon^2$ ajustando-se o modelo MBFC-NI3. Os gráficos (a), (b), (c), (d) e (e) são relativos aos ajustes supondo $\nu = 2.5, 3, 5, 10$ e $30$ , respectivamente. . . . .	122
3.4	No eixo das abscissas temos as estimativas para $\mathbf{X}_\epsilon$ ajustando-se o MBFC-NI1 considerando o Cenário 1, supondo $\nu = 2.5, 3, 5, 10$ e $30$ nos gráficos (a), (b), (c), (d) e (e), respectivamente. . . . .	123
3.5	No eixo das abscissas temos as estimativas para $\mathbf{X}_\epsilon$ ajustando-se o MBFC-NI2 considerando o Cenário 1, supondo $\nu = 2.5, 3, 5, 10$ e $30$ nos gráficos (a), (b), (c), (d) e (e), respectivamente. . . . .	123
3.6	No eixo das abscissas temos as estimativas para $\mathbf{X}_\epsilon$ ajustando-se o MBFC-NI3 considerando o Cenário 1, supondo $\nu = 2.5, 3, 5, 10$ e $30$ nos gráficos (a), (b), (c), (d) e (e), respectivamente. . . . .	124
3.7	Estimativas médias do valor esperado <i>a posteriori</i> para a fração de cura de cada um dos 100 indivíduos, ajustando-se o MBFC-NI1 (a), o MBFC-NI2 (b) e o MBFC-NI3 (c) considerando o Cenário 1, assumindo $\nu = 2.5, 3, 5, 10$ e $30$ . . . . .	125
3.8	Viés das estimativas do valor esperado <i>a posteriori</i> para a fração de cura de cada um dos 100 indivíduos, ajustando-se o MBFC-NI1 (a), o MBFC-NI2 (b) e o MBFC-NI3 (c) considerando o Cenário 1, assumindo $\nu = 2.5, 3, 5, 10$ e $30$ . . . . .	126
3.9	Histograma das frações de cura reais superposto pelas distribuições empíricas das médias dos valores esperados <i>a posteriori</i> da fração de cura, considerando o Cenário 1 e ajustando-se os modelos MBFC-NI1, MBFC-NI2 e MBFC-NI3, nos gráficos (a), (b) e (c), respectivamente, para os diferentes valores de $\nu$ . . . . .	127
3.10	No eixo das abscissas temos as estimativas para $\mathbf{X}_\epsilon$ ajustando-se o MBFC-NI1, 2 e 3, considerando o Cenário 2, supondo $\nu = 2.5$ nos gráficos (a), (b) e (c), respectivamente. . . . .	130

3.11	No eixo das abscissas temos as estimativas para $\mathbf{X}_\epsilon$ ajustando-se o MBFC-NI1, 2 e 3 considerando o Cenário 2, supondo $\nu = 30$ nos gráficos (a), (b) e (c), respectivamente. . . . .	130
3.12	<i>Box-plots</i> das estimativas médias obtidas para $\sigma_\epsilon^2$ ajustando-se os modelos MEFC-NI2 (Fig.a e b) e MEFC-NI3 (Fig.c e d), assumindo $\nu = 2.5$ nas figuras (a) e (c), e $\nu = 30$ nas figuras (b) e (d). . . . .	131
3.13	Estimativas médias do valor esperado <i>a posteriori</i> para a fração de cura de cada um dos 100 indivíduos, ajustando-se os modelos MBFC-NI1, 2, 3 e MTP supondo $\nu = 2.5$ (a) e $\nu = 30$ (b). . . . .	132
3.14	Viés médio das estimativas do valor esperado <i>a posteriori</i> para a fração de cura de cada um dos 100 indivíduos, ajustando-se os modelos MBFC-NI1, 2, 3 e MTP supondo $\nu = 2.5$ (a) e $\nu = 30$ (b). . . . .	133
3.15	Histograma das frações de cura reais superposto pelas distribuições empíricas das médias dos valores esperados <i>a posteriori</i> da fração de cura, considerando o Cenário 2 e ajustando-se os modelos MBFC-NI1, MBFC-NI2, MBFC-NI3 e MTP supondo $\nu = 2.5$ (a) e $\nu = 30$ (b). . . . .	134
3.16	<i>Box-plots</i> das estimativas médias para $\sigma_\epsilon^2$ ajustando-se os modelos MBFC-NI2(a), MBFC-NI3(b), MBFC2(c) e MBFC3(d), considerando os dados do Cenário 1. . . . .	137
3.17	No eixo das abscissas temos as estimativas para $\mathbf{X}_\epsilon$ ajustando-se os modelos MBFC-NI1, 2, 3, MBFC1, 2 e 3, considerando o Cenário 1, nos gráficos (a), (b), (c), (d), (e) e (f), respectivamente. . . . .	137
3.18	Estimativas médias do valor esperado <i>a posteriori</i> para a fração de cura de cada um dos 100 indivíduos, ajustando-se os modelos MBFC1, 2, 3, MBFC-NI1, 2 e 3. . . . .	139
3.19	Viés médio das estimativas do valor esperado <i>a posteriori</i> para a fração de cura de cada um dos 100 indivíduos, ajustando-se os modelos MBFC1, 2, 3, MBFC-NI1, 2 e 3. . . . .	139
3.20	Histograma das frações de cura reais superposto pelas distribuições empíricas das médias dos valores esperados <i>a posteriori</i> da fração de cura, obtidas considerando os Cenário 1 e ajustando-se os modelos MBFC1, MBFC2, MEFC3, MBFC-NI1, MBFC-NI2 e MBFC-NI3. . . . .	140
3.21	<i>Box-plots</i> das estimativas médias para $\sigma_\epsilon^2$ ajustando-se os modelos MBFC-NI2(a), MBFC-NI3(b), MBFC2(c) e MBFC3(d), considerando os dados do Cenário 3. . . . .	142
3.22	No eixo das abscissas temos as estimativas para $\mathbf{X}_\epsilon$ ajustando-se os modelos MBFC-NI1, 2, 3, MBFC1, 2 e 3, considerando o Cenário 3, nos gráficos (a), (b), (c), (d), (e) e (f), respectivamente. . . . .	143
3.23	Estimativas médias do valor esperado <i>a posteriori</i> para a fração de cura de cada um dos 100 indivíduos, ajustando-se os modelos MBFC1, 2, 3, MBFC-NI1, 2 e 3, considerando o Cenário 3. . . . .	143
3.24	Viés médio das estimativas do valor esperado <i>a posteriori</i> para a fração de cura de cada um dos 100 indivíduos, ajustando-se os modelos MBFC1, 2, 3, MBFC-NI1, 2 e 3. . . . .	144
3.25	Histograma das frações de cura reais superposto pelas distribuições empíricas das médias dos valores esperados <i>a posteriori</i> da fração de cura, considerando o Cenário 3 e ajustando-se os modelos MBFC1, MBFC2, MBFC3, MBFC-NI1, MBFC-NI2 e MBFC-NI3. . . . .	144
3.26	HPD's de 95% para a variável latente $X_\epsilon$ dos primeiros 100 indivíduos, obtidos ajustando-se o MBFC-NI1. Os pontos destacados são os valores observados da covariável $X_4$ para cada indivíduo. . . . .	149

3.27	HPD's de 95% para a variável latente $X_\epsilon$ dos primeiros 100 indivíduos, obtidos ajustando-se o MBFC-NI2. Os pontos destacados são os valores observados da covariável $X_4$ para cada indivíduo. . . . .	150
3.28	HPD's de 95% para a variável latente $X_\epsilon$ dos primeiros 100 indivíduos, obtidos ajustando-se o MBFC-NI3. Os pontos destacados são os valores observados da covariável $X_4$ para cada indivíduo. . . . .	150
3.29	<i>Box-plots</i> das estimativas do valor esperado da distribuição <i>a posteriori</i> da fração de cura dos indivíduos observados que receberam doses de Interferon ( $X_3 = 1$ ) ou que ficaram apenas em observação ( $X_3 = 0$ ), considerando os ajustes dos modelos MBFC-NI1, 2 e 3, respectivamente. . . . .	151
3.30	<i>Box-plots</i> das estimativas do valor esperado da distribuição <i>a posteriori</i> da fração de cura dos indivíduos observados, considerando as diferentes faixa etárias, sob o ajuste dos modelos MBFC-NI1 (a), 2 (b) e 3 (c). . . . .	152
3.31	<i>Box-plots</i> das estimativas do valor esperado da distribuição <i>a posteriori</i> da fração de cura dos indivíduos observados para cada faixa etária considerada, divididas entre os indivíduos que receberam Interferon ( $X_3 = 1$ ) e aqueles que permaneceram apenas em observação ( $X_3 = 0$ ), sob o ajuste dos modelos MBFC-NI1 (a), 2 (b) e 3 (c). . . . .	153
3.32	Divergência K-L nos modelos MBFC-NI1 (a), MBFC-NI2 (b) e MBFC-NI3 (c). . . . .	154
A.1	Comparação entre as f.d.p. das distribuições normal (curva sólida), t-Student (curva tracejada), slash (curva pontilhada) e normal contaminada (curva mista – pontilhada e tracejada). . . . .	167
A.2	Na Figura (a), temos a distribuição t-Student para diferentes valores de $\nu$ . Na Figura(b), temos a distribuição slash para diferentes valores de $\nu$ e nas Figuras (c) e (d) temos a distribuição normal contaminada para diferentes valores de $\xi$ e $\varsigma$ . . . . .	168
G.1	<i>Box-plots</i> dos valores esperados <i>a posteriori</i> da fração de cura para cada um dos 50 primeiros indivíduos da amostra ajustando o MBFC1 e assumindo $\sigma_x^2 \sim GI(102, 1010)$ , fixando $\sigma_\epsilon^2 = 1$ (Fig.a), 0.5 (Fig.b) e 10 (Fig.c). O * representa o valor real para a fração de cura de cada indivíduo. . . . .	198
G.2	<i>Box-plots</i> dos valores esperados <i>a posteriori</i> da fração de cura para cada um dos 50 últimos indivíduos da amostra ajustando o MBFC1 e assumindo $\sigma_x^2 \sim GI(102, 1010)$ , fixando $\sigma_\epsilon^2 = 1$ (Fig.a), 0.5 (Fig.b) e 10 (Fig.c). O * representa o valor real para a fração de cura de cada indivíduo. . . . .	199
G.3	<i>Box-plots</i> dos valores esperados <i>a posteriori</i> da fração de cura para cada um dos 50 primeiros indivíduos da amostra ajustando o MBFC1 e assumindo $\sigma_x^2 \sim GI(0.1, 0.1)$ , fixando $\sigma_\epsilon^2 = 1$ (Fig.a), 0.5 (Fig.b) e 10 (Fig.c). O * representa o valor real para a fração de cura de cada indivíduo. . . . .	200
G.4	<i>Box-plots</i> dos valores esperados <i>a posteriori</i> da fração de cura para cada um dos 50 últimos indivíduos da amostra ajustando o MBFC1 e assumindo $\sigma_x^2 \sim GI(0.1, 0.1)$ , fixando $\sigma_\epsilon^2 = 1$ (Fig.a), 0.5 (Fig.b) e 10 (Fig.c). O * representa o valor real para a fração de cura de cada indivíduo. . . . .	201
G.5	<i>Box-plots</i> dos valores esperados <i>a posteriori</i> da fração de cura para cada um dos 50 primeiros indivíduos da amostra ajustando o MBFC1 e assumindo $\sigma_x^2 \sim GI(100, 100)$ , fixando $\sigma_\epsilon^2 = 1$ (Fig.a), 0.5 (Fig.b) e 10 (Fig.c). O * representa o valor real para a fração de cura de cada indivíduo. . . . .	202



G.6	<i>Box-plots</i> dos valores esperados <i>a posteriori</i> da fração de cura para cada um dos 50 últimos indivíduos da amostra ajustando o MBFC1 e assumindo $\sigma_x^2 \sim GI(100, 100)$ , fixando $\sigma_\epsilon^2 = 1$ (Fig.a), 0.5 (Fig.b) e 10 (Fig.c). O * representa o valor real para a fração de cura de cada indivíduo. . . . .	203
G.7	<i>Box-plots</i> dos valores esperados <i>a posteriori</i> da fração de cura para cada um dos 50 primeiros indivíduos da amostra ajustando os modelos MTP (Fig.a), MBFC1 (fixando $\sigma_\epsilon^2 = 1$ , Fig.b), MBFC2 (Fig.c) e MBFC3 (Fig.d), assumindo $\sigma_x^2 \sim GI(102, 1010)$ . O * representa o valor real para a fração de cura de cada indivíduo. . . . .	204
G.8	<i>Box-plots</i> dos valores esperados <i>a posteriori</i> da fração de cura para cada um dos 50 últimos indivíduos da amostra ajustando os modelos MTP (Fig.a), MBFC1 (fixando $\sigma_\epsilon^2 = 1$ , Fig.b), MBFC2 (Fig.c) e MBFC3 (Fig.d), assumindo $\sigma_x^2 \sim GI(102, 1010)$ . O * representa o valor real para a fração de cura de cada indivíduo.	205
G.9	<i>Box-plots</i> dos valores esperados <i>a posteriori</i> da fração de cura para cada um dos 50 primeiros indivíduos da amostra ajustando os modelos MTP (Fig.a), MBFC1 (fixando $\sigma_\epsilon^2 = 5$ , Fig.b), MBFC2 (Fig.c) e MBFC3 (Fig.d), assumindo $\sigma_x^2 \sim GI(102, 1010)$ . O * representa o valor real para a fração de cura de cada indivíduo. . . . .	206
G.10	<i>Box-plots</i> dos valores esperados <i>a posteriori</i> da fração de cura para cada um dos 50 últimos indivíduos da amostra ajustando os modelos MTP (Fig.a), MBFC1 (fixando $\sigma_\epsilon^2 = 5$ , Fig.b), MBFC2 (Fig.c) e MBFC3 (Fig.d), assumindo $\sigma_x^2 \sim GI(102, 1010)$ . O * representa o valor real para a fração de cura de cada indivíduo.	207
G.11	<i>Box-plots</i> dos valores esperados <i>a posteriori</i> da fração de cura para cada um dos 50 primeiros indivíduos da amostra ajustando os modelos MTP (Fig.a), MBFC1 (fixando $\sigma_\epsilon^2 = 1$ , Fig.b), MBFC2 (Fig.c) e MBFC3 (Fig.d), assumindo $\sigma_x^2 \sim GI(102, 1010)$ . O * representa o valor real para a fração de cura de cada indivíduo. . . . .	208
G.12	<i>Box-plots</i> dos valores esperados <i>a posteriori</i> da fração de cura para cada um dos 50 últimos indivíduos da amostra ajustando os modelos MTP (Fig.a), MBFC1 (fixando $\sigma_\epsilon^2 = 1$ , Fig.b), MBFC2 (Fig.c) e MBFC3 (Fig.d), assumindo $\sigma_x^2 \sim GI(102, 1010)$ . O * representa o valor real para a fração de cura de cada indivíduo.	209
H.1	Média Ergódica e autocorrelação dos parâmetros do MBFC2. . . . .	211
H.2	Gráficos da estatística Z de Geweke para as cadeias dos parâmetros $\beta_1$ (a), $\beta_2$ (b), $\alpha$ (c), $\lambda$ (d), $\mu_x$ (e), $\sigma_x^2$ (f) e $\sigma_\epsilon^2$ (g) do MBFC2 . . . . .	212
H.3	Média Ergódica e autocorrelação dos parâmetros Gráficos da estatística Z de Geweke para as cadeias dos parâmetros do MBFC-NI2. . . . .	213
H.4	Gráficos da estatística Z de Geweke para as cadeias dos parâmetros Gráficos da estatística Z de Geweke para as cadeias dos parâmetros $\beta_1$ (a), $\beta_2$ (b), $\alpha$ (c), $\lambda$ (d), $\mu_x$ (e), $\sigma_x^2$ (f) e $\sigma_\epsilon^2$ (g) do MBFC-NI2 . . . . .	214

# Lista de Tabelas

2.1	Modelos Bayesianos de fração de cura com distribuição Normal para o erro de medida. . . . .	37
2.2	Média, Viés e EQM das estimativas <i>a posteriori</i> obtidas ajustando-se o MBFC1, considerando $\sigma_\epsilon^2 = 1, 0.5$ e $10$ , e diferentes distribuições a priori para $\sigma_x^2$ . . . . .	53
2.3	Percentual médio de cobertura dos intervalos HPD's com 95% de credibilidade obtidos para a fração de cura, ajustando-se o MBFC1 para todas as especificações <i>a priori</i> para $\sigma_\epsilon^2$ e $\sigma_x^2$ . . . . .	56
2.4	Média do valor esperado <i>a posteriori</i> , Viés e Erro Quadrático Médio (EQM) das estimativas obtidas ajustando-se os modelos MTP, MBFC2, MBFC3 e o MBFC1, no qual consideramos $\sigma_\epsilon^2 = 1, 0.5, 10$ . O valor real de $\sigma_\epsilon^2$ é $1$ . . . . .	58
2.5	Média do valor esperado <i>a posteriori</i> , Viés e Erro Quadrático Médio (EQM) das estimativas obtidas ajustando-se os modelos MTP, MBFC2, MBFC3 e o MBFC1, no qual consideramos $\sigma_\epsilon^2 = 1, 0.5, 5$ . O valor real de $\sigma_\epsilon^2$ é $5$ . . . . .	59
2.6	Percentual médio de cobertura dos Intervalos HPD's com 95% de credibilidade obtidos para a fração de cura, ajustando-se o MTP e o MBFC sob todas as especificações de $\sigma_\epsilon^2$ e $\sigma_x^2 \sim GI(102, 1010)$ . . . . .	61
2.7	Média do valor esperado <i>a posteriori</i> , Viés e Erro Quadrático Médio (EQM) das estimativas obtidas ajustando-se os modelos MTP, MBFC1, no qual consideramos $\sigma_\epsilon^2 = 1$ , MBFC2 e MBFC3. . . . .	66
2.8	Percentual médio de cobertura dos intervalos de 95% de credibilidade para a fração de cura, ajustando-se o MTP e os MBFC's. . . . .	69
2.9	Média do valor esperado <i>a posteriori</i> , Viés e Erro Quadrático Médio (EQM) das estimativas obtidas ajustando-se os modelos MBFC2 e MBFC3. . . . .	70
2.10	Percentual médio de cobertura dos intervalos de 95% de credibilidade para a fração de cura, ajustando-se o MBFC2 e o MBFC3 para as diferentes escolhas <i>a priori</i> de $\sigma_x^2$ . . . . .	74
2.11	Média do valor esperado <i>a posteriori</i> , Viés e Erro Quadrático Médio (EQM) das estimativas obtidas ajustando-se o modelo MBFC3 aos dados do Cenário 1, supondo diferentes distribuições <i>a priori</i> para $k$ . . . . .	75

2.12	Média do valor esperado <i>a posteriori</i> , Viés e Erro Quadrático Médio (EQM) das estimativas obtidas ajustando-se o modelo MBFC3 aos dados do Cenário 2, supondo diferentes distribuições <i>a priori</i> para $k$ . . . . .	75
2.13	Percentual médio de cobertura dos intervalos de 95% de credibilidade da fração de cura, ajustando-se o MBFC3 aos dados dos Cenário 1 e 2, considerando as diferentes escolhas <i>a priori</i> para $k$ . . . . .	79
2.14	Resumos das distribuições <i>a posteriori</i> dos parâmetros dos modelos MTP, MBFC1, 2 e 3 para os dados do estudo E1684, assumindo-se distribuição Weibull para os tempos de falha. . . . .	84
2.15	Comparação dos Modelos ajustados aos dados do estudo E1684. . . . .	92
2.16	Resumos das distribuições <i>a posteriori</i> dos parâmetros dos modelos MTP, MBFC1, 2, 3 e do MBFC-Ma para os dados do estudo E1684, assumindo-se distribuição Exponencial para os tempos de falha. . . . .	96
2.17	Comparação dos Modelos ajustados aos dados do estudo E1684. . . . .	104
3.1	Média, Viés e EQM das estimativas <i>a posteriori</i> obtidas ajustando-se os modelos MBFC-NI1, considerando $\sigma_\epsilon^2 = 1$ , MBFC-NI2 e MBFC-NI3, supondo $\nu = 2.5, 3, 5, 10$ e $30$ , com os dados do Cenário 1 (em que $\nu = 3$ ). . . . .	120
3.2	Percentual de cobertura dos intervalos de 95% de credibilidade para a fração de cura, considerando o Cenário 1 e ajustando-se os modelos MBFC-NI1, MBFC-NI2 e MBFC-NI3 para os diferentes valores de $\nu$ . . . . .	127
3.3	Média, Viés e EQM das estimativas <i>a posteriori</i> obtidas ajustando-se os modelos MTP, MBFC-NI1, considerando $\sigma_\epsilon^2 = 1$ , MBFC-NI2 e MBFC-NI3, supondo $\nu = 2.5$ e $30$ , com os dados do Cenário 2. . . . .	129
3.4	Percentual de cobertura dos intervalos de 95% de credibilidade para a fração de cura, considerando o Cenário 2 e ajustando-se os modelos MBFC-NI1, MBFC-NI2, MBFC-NI3 e MTP para os diferentes valores de $\nu$ . . . . .	134
3.5	Média, Viés e EQM das estimativas <i>a posteriori</i> obtidas ajustando-se os modelos MBFC1 e MBFC-NI1, considerando $\sigma_\epsilon^2 = 1$ , MBFC2, MBFC-NI2, MBFC3 e MBFC-NI3, supondo $\nu = 3$ , aos dados do Cenário 1. . . . .	136
3.6	Percentual de cobertura dos intervalos de 95% de credibilidade para a fração de cura, considerando o Cenário 1 e ajustando-se os modelos MBFC1 a 3 e MBFC-NI1 a 3. . . . .	140
3.7	Média, Viés e EQM das estimativas <i>a posteriori</i> obtidas ajustando-se os modelos MBFC1 e MBFC-NI1, considerando $\sigma_\epsilon^2 = 1$ , MBFC2, MBFC-NI2, MBFC3 e MBFC-NI3, analisando os dados do Cenário 3. . . . .	141
3.8	Percentual de cobertura dos intervalos de 95% de credibilidade para a fração de cura, considerando o Cenário 3 e ajustando-se os modelos MBFC1 a 3 e MBFC-NI1 a 3. . . . .	145

3.9	Resumos das distribuições <i>a posteriori</i> dos parâmetros dos modelos MBFC-NI1, MBFC-NI2 e MBFC-NI3 para os dados do estudo E1684, assumindo-se distribuição Weibull para os tempos de falha. . . . .	148
3.10	Comparação dos Modelos ajustados aos dados do estudo E1684. . . . .	153

# Capítulo 1

## Introdução

### 1.1 Considerações Iniciais

Em Análise de Sobrevivência, o desenvolvimento de modelos nos quais uma fração de cura é incorporada em sua estrutura têm sido de grande interesse. Isso se deve ao fato de que, na prática, é possível que um certo percentual da população estudada não experimente o evento de interesse e “sobreviva” além do tempo de observação. A este percentual dá-se o nome de fração de cura e os modelos que a incluem são denominados como modelos de fração de cura.

Um exemplo clássico, onde esse tipo de modelo é considerado, são os estudos onde o objeto de interesse é o tempo até a morte dos indivíduos por um determinado tipo de doença, em que após serem submetidos a um tratamento, alguns dos indivíduos observados são curados.

Um modelo de fração de cura bastante utilizado é o modelo de mistura padrão proposto por Berkson e Gage (1952), o qual tem como grande vantagem a simplicidade na inclusão de covariáveis em sua estrutura. No entanto, assume que apenas uma causa pode provocar a falha, o que o torna irrealístico do ponto de vista prático. Além disso, na presença de covariáveis a estrutura de riscos proporcionais, que é uma propriedade bastante desejada em modelos de sobrevivência, nem sempre é observada. Outras desvantagens são apontadas por Chen *et al.* (1999).

Uma alternativa, que engloba a fração de curados e não curados em uma mesma expressão, é o modelo de tempo de promoção proposto por Yakovlev e Tsodikov (1996), que generaliza o modelo de mistura padrão, com a suposição de que várias causas competem para a ocorrência do evento de interesse, e possui propriedades bastante desejáveis, como destacam Zeng *et al.* (2006). Entre tais vantagens está a facilidade em incorporar covariáveis através de uma função de ligação (ver definição na Subseção 1.3.1). De acordo como explicam Rodrigues *et al.* (2008), o modelo de mistura padrão pode ser visto a partir do modelo de tempo de promoção ao considerarmos

que a falha ocorrerá em consequência de apenas uma causa.

O modelo de tempo de promoção que será o foco neste trabalho, também foi considerado por Demarqui *et al.* (2011), que propõem a distribuição exponencial por partes com grade aleatória para descrever o comportamento do tempo de promoção. Lopes e Bolfarine (2012) incluem efeitos aleatórios à modelagem para avaliar o efeito das clínicas na fração de cura, considerando apenas o intercepto como aleatório. Os modelos de mistura padrão e tempo de promoção são bastante utilizados por fornecerem bons resultados em várias situações (ver Farewell (1982); Tsodikov (1998); Chen *et al.* (1999); Chen e Ibrahim (2001); Chen *et al.* (2002a,b,c)). No entanto, na prática, algumas covariáveis contínuas são frequentemente passíveis de erro de medida, o que não é considerado por tais modelos. Tanto quanto sabemos, pouco tem sido feito nesta direção. Visando acomodar suposições mais realísticas e captar o máximo de informação que se tenha disponível, surgem os primeiros modelos de fração de cura com erro de medida nas covariáveis propostos por Mizoi (2004), Mizoi *et al.* (2007) e Ma e Yin (2008). Ambos assumem que algumas covariáveis (ou todas) podem ter sido indiretamente observadas (variáveis latentes) ou que foram coletadas com algum erro de medida. Em situações como esta, não considerarmos os erros de medida em variáveis relevantes para o estudo podem produzir estimativas muito imprecisas para os parâmetros. Neste sentido, a proposta de Mizoi (2004) e Mizoi *et al.* (2007) é incluir variáveis medidas com erro ao modelo de tempo de promoção. Tais autores fixam a variância do erro e utilizam o método do escore corrigido para obter estimativas dos parâmetros. Ma e Yin (2008) propõem um modelo diferente, utilizando uma abordagem semiparamétrica para fazer a inferência. As estimativas dos parâmetros são então obtidas via maximização da função de verossimilhança, aplicando o método dos multiplicadores de Lagrange. Apesar de incluir a suposição de erro nas covariáveis, as estimativas obtidas por Ma e Yin (2008) são ainda bastante enviesadas. Bertrand *et al.* (2016, 2017) propõem um método para a correção do viés nas estimativas dos parâmetros do modelo de Ma e Yin (2008), obtendo uma expressão para o viés dos estimadores dos parâmetros de regressão. O efeito do erro de medida no viés dos estimadores na proposta de Ma e Yin (2008) e Chen *et al.* (1999), são estudados por Bertrand *et al.* (2017), que utilizam o método do escore corrigido (Nakamura, 1990) para fazer inferência no modelo. Essa correção, é feita considerando-se uma adaptação do método SIMEX, um algoritmo de simulação e extrapolação (Apanasovich *et al.*, 2009), em que impõe-se que variância do erro de medida pertença a um conjunto finito de valores. Esta suposição sobre a variância é um pouco menos restritiva do que a estrutura considerada por Ma e Yin (2008) e Mizoi *et al.* (2007), que fixam tal variância. Em todos esses modelos a inferência é realizada sob o ponto de vista clássico e as covariáveis medidas com erro recebem um tratamento funcional. Na modelagem funcional, as variáveis latentes são consideradas constantes desconhecidas e, portanto, não há uma preocupação em estimar seu verdadeiro valor.

Uma construção ainda não explorada em modelos de fração de cura com erro nas covariáveis é aquela denotada por Fuller (2009) como estrutural (ver também Bolfarine e Arellano-Valle (1998)). No contexto de estatística clássica, na modelagem estrutural as variáveis latentes são tratadas como aleatórias, atribuindo-se a estas uma medida de probabilidade. No âmbito Bayesiano todo modelo com erro de medida é visto como estrutural, visto que uma distribuição de probabilidade é utilizada para descrever a tudo que é desconhecido. Com essa motivação, neste trabalho introduzimos modelos Bayesianos de fração de cura em que as covariáveis medidas com erro são tratadas como variáveis aleatórias latentes. Nesta estrutura, além dos parâmetros do modelo de sobrevivência, estimamos a média e a variância das variáveis latentes visando, com isto, considerarmos a incerteza sobre os valores das variáveis medidas com erro nas estimativas da fração de cura. Diferentemente do que é feito em Mizoi (2004), Mizoi *et al.* (2007), Ma e Yin (2008) e Bertrand *et al.* (2016) e Bertrand *et al.* (2017), a variância do erro de medida é estimada. Para as variáveis medidas com erro, consideramos um modelo com estrutura aditiva na qual supomos para uma distribuição na família mistura de escala da *Normal* para o erro de medida. Até onde sabemos, este é o primeiro trabalho a considerar tais suposições.

Nosso objetivo é propor modelos mais flexíveis, capazes de acomodar mais suposições práticas do que aquelas já consideradas na literatura. Todos os modelos de fração de cura com erro supracitados carregam a imposição de variância do erro conhecida, o que pode ser bastante restritivo na prática. Apresentamos aqui, três maneiras distintas para modelar a incerteza sobre a variância do erro de medida. Na primeira delas (tal como já explorado na literatura), consideramos a variância do erro como uma constante fixa conhecida; uma segunda modelagem é concebida ao considerarmos tal variância desconhecida, eliciando uma distribuição *a priori Gama Inversa* para tal parâmetro; e por último, admitimos que a razão entre a variância das variáveis latentes e a variância do erro de medida é uma constante  $0 < k \leq 1$ , atribuindo uma medida de probabilidade para  $k$  a fim de inferir a respeito da variância do erro a partir da estimação de tal constante. Essa última suposição é inspirada na proposta de modelagem de Fuller (2009) para resolver problemas de falta de identificabilidade em modelos estruturais com erro de medida nas covariáveis. Um outro diferencial entre os modelos propostos aqui e aqueles disponíveis na literatura, supracitados, é a presença do intercepto. Considerar tal parâmetro em nossos modelos resulta em problemas de falta de identificabilidade, que não são resolvidos seguindo a recomendação de Fuller (2009) de estabelecermos que a razão entre a variância das variáveis latentes e a variância do erro de medida é uma constante conhecida.

Este trabalho está organizado da maneira que segue. Na Seção 1.2 definimos modelo linear com erro nas covariáveis, diferenciando a abordagem funcional da estrutural. Na Seção 1.3, descrevemos brevemente alguns dos modelos supracitados, a saber: modelo de tempo de promoção (Chen *et al.*, 1999), e modelos de fração de cura com erro de medida (Mizoi, 2004;

Mizoi *et al.*, 2007; Ma e Yin, 2008), para efeitos de comparação e de melhor entendimento acerca dos modelos Bayesianos propostos ao longo desse texto. No Capítulo 2, definimos três modelos Bayesianos com erro de medida normal aditivo. Apresentamos a função de verossimilhança e calculamos as distribuições condicionais completas dos parâmetros, que serão utilizadas para obtermos amostras da distribuição *a posteriori* via métodos MCMC, nas três modelagens propostas. Além disso, na Seção 2.2, sugerimos uma versão estrutural, Bayesianas e paramétrica do modelo proposto por Ma e Yin (2008), com o objetivo de fazermos uma comparação entre tal modelo e aqueles desenvolvidos neste trabalho. Na Seção 2.3, usando simulação de Monte Carlo, avaliamos o desempenho das três modelagens apresentadas neste capítulo e fazemos uma comparação entre o desempenho de tais modelos e o modelo de tempo de promoção proposto por Chen *et al.* (1999). No capítulo 3, os modelos Bayesianos de fração cura com erro de medida são estendidos ao assumirmos que as covariáveis medidas com erro são modeladas por distribuições na família mistura de escala da *Normal* (Andrews e Mallows (1974); Santos (2015); Lange e Sinsheimer (1993); Rogers e Tukey (1972), entre outros). Os novos modelos são mais flexíveis e abrangem aqueles propostos no Capítulo 2. Na Seção 3.3, focamos nossa atenção no modelo Bayesiano de fração de cura com erro de medida com distribuição *t*-Student, a qual é membro da família mistura de escala da *Normal*. Uma das dificuldades encontradas neste contexto é a estimação dos graus de liberdade da distribuição do erro de medida. Aqui consideramos tal parâmetro como uma constante conhecida fixada. Comparamos o desempenho deste modelo com o desempenho daqueles nos quais consideramos erro normalmente distribuído e também com o modelo de tempo de promoção. Nas Seções 2.4 e 3.4, apresentamos uma aplicação considerando dados reais de um ensaio clínico pós cirúrgico de pacientes com melanoma cutâneo, o E1684. Nestas seções lançamos mão de alguns critérios de seleção de modelos para avaliar a qualidade dos ajustes apresentados. No Capítulo 4 apresentamos uma breve discussão e conclusões, bem como ideias para pesquisas futuras relacionadas a extensão dos modelos Bayesianos de fração de cura com erro nas covariáveis propostos ao longo deste trabalho.

## 1.2 Modelo Linear com Erro nas Covariáveis

Seja  $Y_i$  variável reposta de interesse associada ao indivíduo  $i$ . Assuma que  $Y_i$  é linearmente dependente de um vetor de covariáveis  $\mathbf{X}_i = (X_{i1}, X_{i2}, \dots, X_{ip})$ , de ordem  $1 \times p$ , tal que

$$Y_i = \mathbf{X}_i \boldsymbol{\beta} + e_i, \quad i = 1, \dots, n \quad (1.1)$$

em que  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^\top$  é um vetor de efeitos fixos de ordem  $p \times 1$  e  $e_i$  denota o erro do modelo. Usualmente, assume-se que os erros  $e_i$ 's são independentes e identicamente distribuídos



(i.i.d), com distribuição centrada em zero e com variância comum  $\sigma_e^2$ .

Em muitos contextos é razoável assumir que algumas covariáveis, ou até mesmo todas, podem ter sido registradas ou mensuradas com erro e que não seja possível acessar seus verdadeiros valores. Como exemplo, em estudos de sobrevivência, podemos citar os casos em que uma das covariáveis no estudo é o tamanho do tumor. Em indivíduos com melanoma cutâneo, por exemplo, a medição dessa variável é obtida a partir de imagens e classificada segundo o índice de Breslow, sendo, portanto, passível de erro. O índice de Breslow é o principal indicador de prognóstico em casos de câncer e relaciona a espessura vertical do tumor, em milímetros, com a gravidade do caso e a probabilidade de haver metástases (Wainstein e Belfort, 2004).

Supondo que a variável resposta é linearmente modelada como na Equação (1.1) e admitindo, sem perda de generalidade que a covariável  $\mathbf{X}_1 = (X_{11}, X_{21}, \dots, X_{n1})$  é medida com erro, de modo que a matriz de covariáveis observadas é  $\mathbf{W} = (\mathbf{X}_1^\top + \boldsymbol{\epsilon}, \mathbf{X}_2^\top, \dots, \mathbf{X}_p^\top)$ , de ordem  $n \times p$ , em que  $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^\top$  é o vetor  $n \times 1$  de erros de medida da covariável  $\mathbf{X}_1$ . Bolfarine *et al.* (1992) mostra que esta nova estrutura pode ser representada como:

$$\begin{aligned} Y_i &= \mathbf{W}_i \boldsymbol{\beta} + e_i, \quad i = 1, \dots, n, \\ \mathbf{W}_i &= (\mathbf{X}_{i1} + \epsilon_i, \mathbf{X}_{i2}, \dots, \mathbf{X}_{ip}), \end{aligned} \tag{1.2}$$

em que  $\epsilon_i$  é o erro cometido ao mensurarmos a variável  $X_{i1}$ . Os erros  $\epsilon_i$  são assumidos como sendo independentes e identicamente distribuídos (i.i.d), com distribuição centrada em zero e com variância comum  $\sigma_e^2$ . Supomos  $\epsilon_i$  independente da variável latente  $\mathbf{X}_{i1}$ ,  $i = 1, 2, \dots, n$ . O modelo dado na Equação (1.2), é conhecido como modelo com erro de medida nas covariáveis. Mais detalhes sobre modelos com erro de medida nas covariáveis podem ser encontrados em Fuller (2009) e Bolfarine *et al.* (1992).

Do ponto de vista clássico, os modelos com erro nas covariáveis podem ser classificados como **modelo funcional** ou **modelo estrutural**, os quais são determinados pelo caráter assumido para as covariáveis medidas com erro.

Em um modelo funcional os valores que a covariável  $\mathbf{X}_i$  pode assumir na Expressão (1.2) são considerados constantes fixas e desconhecidas. Por essa razão, dentro da óptica clássica, cada  $X_{ij}$  é visto como um parâmetro do modelo (Bolfarine *et al.*, 1992). Em um modelo estrutural, a covariável  $\mathbf{X}_i$  medida com erro também é desconhecida, mas é considerada uma variável aleatória para a qual se atribui uma distribuição de probabilidade. Uma suposição habitualmente feita é que cada  $X_{ij} \sim N(\mu_x, \sigma_x^2)$ . É importante salientar, no entanto, que a suposição de normalidade não é necessária para que o modelo seja considerado estrutural. Modelos mais flexíveis podem ser obtidos assumindo-se outras distribuições.

Destacamos que a diferença entre os modelos funcional e estrutural está apenas na natureza atribuída às variáveis latentes  $\mathbf{X}$ . Assim, do ponto de vista Bayesiano onde a tudo que é desconhecido atribuímos uma distribuição *a priori* a fim de se fazer inferência, modelos com erro de medida nas covariáveis serão sempre abordados como modelos estruturais.

Na seção a seguir, apresentamos uma breve revisão de modelos de fração de cura que já foram discutidos na literatura e que são citados no decorrer deste trabalho.

### 1.3 Modelos de Fração de Cura

Nesta seção, definiremos o modelo de tempo de promoção proposto por Yakovlev e Tsodikov (1996) e os modelos de fração de cura com erro de medida nas covariáveis propostos por Mizoi *et al.* (2007) e Ma e Yin (2008). O primeiro modelo proposto por Yakovlev e Tsodikov (1996) foi bastante utilizado e discutido na literatura por descrever bem o comportamento do tempo de falha no contexto de reincidência de câncer, onde os modelos de fração de cura são frequentemente empregados. Apresentamos a versão do modelo proposta por Chen *et al.* (1999) que inclui as covariáveis através da fração de curados, o que conduz a obtenção de distribuições *a posteriori* próprias para muitas famílias de distribuições *a priori* impróprias. Esta propriedade para a distribuição *a posteriori* não é conseguida ao utilizarmos os modelos de mistura padrão (Berkson e Gage, 1952) de tempo de promoção (Yakovlev e Tsodikov, 1996) quando utilizamos, por exemplo, a inclusão de covariáveis via modelo de regressão binomial padrão (Rodrigues *et al.*, 2008). O modelo proposto por Mizoi *et al.* (2007) é mais recente e foi pouco explorado até o momento. Tal modelo é um pouco mais geral que o modelo de tempo de promoção, pois, comporta a possibilidade de alguma covariável ter sido observada ou mensurada com erro, suposição muito razoável na prática. Por fim, definiremos o modelo proposto por Ma e Yin (2008), o qual, além de considerar que covariáveis possam ser observadas com erro, usando uma abordagem funcional, inova na maneira de incluir a fração de curados no modelo.

#### 1.3.1 Modelo de Tempo de Promoção (MTP)

Suponhamos que para um indivíduo qualquer na população de interesse,  $N$  denote o número de células carcinogênicas ativas após o tratamento inicial. Assumimos que  $N \sim \text{Poisson}(\theta)$ . Denotemos por  $R_i$  uma variável aleatória (v.a.) que caracteriza o instante de tempo em que a  $i$ -ésima célula carcinogênica produziu uma massa cancerígena detectável. Assuma que, dado  $N = n$ , as v.a's  $R_i$ , com  $i = 1, 2, \dots, n$ , são i.i.d., com função de distribuição acumulada (f.d.a) comum  $F(t)$  e que, também, são independentes de  $N$ . Denotemos por  $S(t)$  a função de sobrevivência tal que  $F(t) = 1 - S(t)$ . O tempo de reincidência do câncer é definido como sendo a v.a.

$T = \min \{R_i; 0 \leq i \leq N\}$ . Assumimos também que se  $N = 0$ , temos  $\mathbb{P}(R_0 = \infty) = 1$ , isto é, se o indivíduo não possui célula carcinogênica ativa, a probabilidade do indivíduo estar curado é 1. Com isto, a função de sobrevivência para a população é dada por

$$\begin{aligned}
 S_p(t) &= \mathbb{P}(\text{não ter câncer até o tempo } t) \\
 &= \mathbb{P}(T > t) \\
 &= \mathbb{P}(T > t, N = 0) + \mathbb{P}(T > t, N \geq 1) \\
 &= \mathbb{P}(T > t | N = 0)\mathbb{P}(N = 0) + \sum_{k=1}^{\infty} \mathbb{P}(T > t | N = k)\mathbb{P}(N = k) \\
 &= \mathbb{P}(N = 0) + \mathbb{P}(\min \{R_i, i = 1, \dots, N\} > t, N \geq 1) \\
 &= \exp\{-\theta\} + \sum_{k=1}^{\infty} \mathbb{P}(R_1 > t, \dots, R_N > t | N = k)\mathbb{P}(N = k) \\
 &= \exp\{-\theta\} + \sum_{k=1}^{\infty} (S(t))^k \mathbb{P}(N = k) \\
 &= \exp\{-\theta\} + \sum_{k=1}^{\infty} (1 - F(t))^k \frac{\theta^k \exp\{-\theta\}}{k!} \\
 &= \exp\{-\theta\} \sum_{k=0}^{\infty} \frac{[\theta(1 - F(t))]^k}{k!} \\
 &= \exp\{-\theta F(t)\}.
 \end{aligned} \tag{1.3}$$

Neste modelo, a fração de curados é dada por  $S_p(\infty) = \exp\{-\theta\}$ . Essa quantidade é estritamente positiva e é interpretada como  $S_p(\infty) = \mathbb{P}(N = 0)$ . Além disso, sob as suposições supracitadas, a função de densidade populacional é dada por:

$$f_p(t) = -\frac{dS_p(t)}{dt} = \theta f(t) \exp\{-\theta F(t)\}, \tag{1.4}$$

e, conseqüentemente, a função de risco populacional é,

$$h_p(t) = \frac{f_p(t)}{S_p(t)} = \theta f(t). \tag{1.5}$$

A função  $h_p(t)$  dada em (1.5) não pode ser considerada uma função de risco correspondente a uma distribuição de probabilidade, visto que  $S_p(t)$  não é uma função de sobrevivência propriamente dita. No entanto, a medida que  $t$  se aproxima de infinito,  $h_p(t)$  tende a zero. Para mais detalhes acerca dessa teoria, Chen *et al.* (1999), Giolo e Colosimo (2006), Rodrigues *et al.* (2008) e Rodrigues *et al.* (2009) podem ser consultados.

No modelo de tempo de promoção, Chen *et al.* (1999) relacionam o número médio de causas do evento de interesse à covariáveis para o  $i$ -ésimo indivíduo fazendo  $\mathbb{E}(N_i) = \theta_i = \exp\{\mathbf{X}_i\boldsymbol{\beta}\}$ , em que  $\mathbf{X}_i = (\mathbf{X}_{i1}, \dots, \mathbf{X}_{ip})$  o vetor de observações para o  $i$ -ésimo indivíduo de ordem  $1 \times p$  e  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$  é o vetor de coeficientes de regressão de ordem  $p \times 1$ . A função de risco  $h_p$ , assim definida, possui a propriedade de riscos proporcionais, isto é, se para dois indivíduos distintos na população,  $i \neq j$ , a razão  $\frac{h_{p_i}(t)}{h_{p_j}(t)}$  não depende de  $t$ . A propriedade de riscos proporcionais é desejável, pois, torna a implementação dos métodos computacionais relativamente simples, o que facilita a estimação sob o ponto de vista Bayesiano.

Assumindo  $\boldsymbol{\theta} = \exp\{\mathbf{X}\boldsymbol{\beta}\}$ , em que  $\mathbf{X}$  denota a matriz de delineamento de experimento de ordem  $n \times p$ , a fração de curados é dada por  $\exp\{-\exp\{\mathbf{X}\boldsymbol{\beta}\}\}$ . Isto permite uma fácil interpretação do papel dos coeficientes de regressão, uma vez que coeficientes negativos levam a uma fração de cura maior do que coeficientes positivos. Além disso, para o grupo de não curados, os coeficientes de regressão afetam bastante a função de risco, isto é, um coeficiente de regressão negativo leva a um risco muito maior do que um positivo.

### 1.3.1.1 Função de verossimilhança

Suponhamos que  $n$  indivíduos foram independentemente observados e denote por  $N_i$  o número de células carcinogênicas do  $i$ -ésimo indivíduo. Assuma que  $N_i \sim \text{Poisson}(\theta_i)$  com  $i = 1, \dots, n$ . As variáveis  $N_i$  não são observáveis e são vistas como variáveis latentes na formulação do modelo. Sejam  $R_{i1}, R_{i2}, \dots, R_{iN_i}$  os tempos em que as  $N_i$  células carcinogênicas produzem massa detectável. Denotemos por  $\boldsymbol{\gamma}$  o vetor de parâmetros relacionados aos tempos de incubação. Assim,  $F(\cdot|\boldsymbol{\gamma})$  e  $S(\cdot|\boldsymbol{\gamma})$  representam as funções de distribuição e sobrevivência, respectivamente, associadas a cada  $R_{ij}$ . Seja  $T_i$  o tempo de falha para o indivíduo  $i$  e assumamos que  $T_i$  pode ser censurado à direita. Denotemos por  $C_i$  o tempo de censura. Dessa maneira, o que se observa é  $y_i = \min\{T_i, C_i\}$ . Seja  $\delta_i$  o indicador de falha para o indivíduo  $i$ , isto é,

$$\delta_i = \begin{cases} 1, & \text{se } T_i \leq C_i, \\ 0, & \text{se } T_i > C_i. \end{cases}$$

Assuma que o mecanismo de falha e censura são independentes, isto é,  $T_i$  é independente de  $C_i$  para todo  $i$ .

Representamos o vetor de dados completos por  $\mathbf{D} = (n, \mathbf{N}, \mathbf{Y}, \boldsymbol{\delta}, \mathbf{X})$ , em que  $\mathbf{N}$  é um vetor cuja  $i$ -ésima entrada representa o número de células carcinogênicas ativas para o indivíduo  $i$  e  $\mathbf{X}_i = (\mathbf{X}_{i1}, \mathbf{X}_{i2}, \dots, \mathbf{X}_{ip})$  é o vetor de covariáveis associadas a este indivíduo. Aqui não

incluímos o intercepto para um melhor efeito comparativo entre o MTP e os modelos com erro nas covariáveis propostos por nós. A função de verossimilhança do modelo de tempo de promoção baseada nos dados completos, é dada por:

$$L(\gamma, \theta | \mathbf{D}) \propto \prod_{i=1}^n [f_p(y_i | \gamma, \theta_i)]^{\delta_i} [S_p(y_i | \gamma, \theta_i)]^{N_i - \delta_i} \times \exp \left\{ \sum_{i=1}^n [N_i \log(\theta_i) - \log(N_i!) - n\theta] \right\}. \quad (1.6)$$

em que  $\theta_i = \exp\{-\exp\{\mathbf{X}_i \boldsymbol{\beta}\}\}$ .

Denotemos o vetor de dados observados por  $\mathbf{D}_{\text{obs}} = (n, \mathbf{Y}, \boldsymbol{\delta}, \mathbf{X})$ . Podemos fazer inferência sobre os parâmetros  $(\gamma, \theta)$ , utilizando a abordagem Bayesiana, baseados apenas na função de verossimilhança com os dados observados como mostram Rodrigues *et al.* (2008) e Rodrigues *et al.* (2009), em seu trabalho sobre a unificação de modelos de longa duração. A função de verossimilhança marginal baseada nos dados observados, é dada por:

$$L(\gamma, \theta | \mathbf{D}_{\text{obs}}) \propto \prod_{i=1}^n [f_p(y_i | \gamma, \theta_i)]^{\delta_i} [S_p(y_i | \gamma, \theta_i)]^{1 - \delta_i} \propto \prod_{i=1}^n [\theta_i f(y_i | \gamma) \exp\{-\theta_i F(y_i | \gamma)\}]^{\delta_i} [\exp\{-\theta_i F(y_i | \gamma)\}]^{1 - \delta_i}. \quad (1.7)$$

### 1.3.1.2 Inferência Bayesiana no Modelo de Tempo de Promoção

Suponhamos que os tempos em que as  $N_i$  células carcinogênicas do indivíduo  $i$  produzem uma massa cancerígena detectável, sejam variáveis aleatórias i.i.d. com distribuição Weibull com parâmetro  $\gamma = (\alpha, \lambda)$ . Denotamos  $R_{ij} \sim \text{Weibull}(\alpha, \lambda)$ ,  $j = 1, 2, \dots, N_i$  e  $i = 1, \dots, n$ , e sua função densidade de probabilidade (f.d.p.) é dada por

$$f(z | \alpha, \lambda) = \exp\{\lambda\} \alpha z^{\alpha-1} \exp\{-\exp\{\lambda\} z^\alpha\}. \quad (1.8)$$

Suponhamos, *a priori*, que  $\boldsymbol{\beta}$ ,  $\alpha$  e  $\lambda$  são independentes. Além disso, consideremos  $\pi(\boldsymbol{\beta}) \propto 1$  de tal forma que a distribuição *a priori* conjunta destes parâmetros seja pouco informativa e denotada  $\pi(\boldsymbol{\beta}, \alpha, \lambda) = \pi(\alpha, \lambda)$ . Assumamos que  $\alpha \sim \text{Gama}(a, b)$  e  $\lambda \sim \text{Normal}(m, v)$ . Sob tais supostos, a distribuição *a posteriori* do vetor de parâmetros  $(\alpha, \lambda, \boldsymbol{\beta})$ , baseada nas observações  $\mathbf{D}_{\text{obs}} = (n, \mathbf{Y}, \boldsymbol{\delta}, \mathbf{X})$ , é dada por:

$$\begin{aligned}
\pi(\alpha, \lambda, \boldsymbol{\beta} | \mathbf{D}_{\text{obs}}) &\propto L(\alpha, \lambda, \boldsymbol{\beta} | \mathbf{D}_{\text{obs}}) \pi(\alpha | a, b) \pi(\lambda | m, v) \\
&= \left[ \prod_{i=1}^n \left( \exp \{ \mathbf{X}_i \boldsymbol{\beta}^\top \} f(y_i | \alpha, \lambda) \right)^{\delta_i} \exp \{ - \exp \{ \mathbf{X}_i \boldsymbol{\beta}^\top \} F(y_i | \alpha, \lambda) \} \right] \\
&\quad \times \pi(\alpha | a, b) \pi(\lambda | m, v) \\
&= \alpha^{\sum_{i=1}^n \delta_i} \left[ \prod_{i=1}^n y_i^{(\alpha-1)\delta_i} \right] \exp \left\{ \sum_{i=1}^n \delta_i (\mathbf{X}_i \boldsymbol{\beta}^\top + \lambda) - \exp \{ \lambda \} \sum_{i=1}^n \delta_i y_i^\alpha \right\} \\
&\quad \times \exp \left\{ - \sum_{i=1}^n \exp \{ \mathbf{X}_i \boldsymbol{\beta}^\top \} (1 - \exp \{ - \exp \{ \lambda \} y_i^\alpha \}) \right\}. \tag{1.9}
\end{aligned}$$

O resultado apresentado na Equação (1.9) pode ser visto com mais detalhes em (Rodrigues *et al.*, 2008; Chen *et al.*, 1999). As condições para que tal distribuição *a posteriori* seja própria, são dadas no Teorema 2 de Chen *et al.* (1999). Essa distribuição *a posteriori* não tem forma fechada conhecida. Por essa razão, a estimação dos parâmetros é feita obtendo-se aproximações da distribuição *a posteriori* via métodos MCMC. As distribuições condicionais completas dos parâmetros  $(\boldsymbol{\beta}, \alpha, \lambda)$  do modelo, necessárias para obter amostras das distribuições *a posteriori* via métodos MCMC são dadas a seguir.

i) Distribuição condicional completa para  $\boldsymbol{\beta}$ :

$$\pi(\boldsymbol{\beta} | \alpha, \lambda, \mathbf{D}_{\text{obs}}) \propto \exp \left\{ \sum_{i=1}^n \delta_i \mathbf{X}_i \boldsymbol{\beta}^\top - \exp \{ \mathbf{X}_i \boldsymbol{\beta}^\top (1 - \exp \{ - \exp \{ \lambda \} y_i^\alpha \}) \} \right\}. \tag{1.10}$$

ii) Distribuição condicional completa para  $\alpha$ :

$$\begin{aligned}
\pi(\alpha | \boldsymbol{\beta}, \lambda, \mathbf{D}_{\text{obs}}) &\propto \alpha^{\sum_{i=1}^n \delta_i} \exp \left\{ - \exp \{ \lambda \} \sum_{i=1}^n \delta_i y_i^\alpha + \sum_{i=1}^n [\exp \{ \mathbf{X}_i \boldsymbol{\beta}^\top - \exp \{ \lambda \} y_i^\alpha \}] \right\} \\
&\quad \times \exp \left\{ \lambda \sum_{i=1}^n \delta_i \right\} \pi(\alpha | a, b). \tag{1.11}
\end{aligned}$$

iii) Distribuição condicional completa para  $\lambda$ :

$$\begin{aligned}
\pi(\lambda | \boldsymbol{\beta}, \alpha, \mathbf{D}_{\text{obs}}) &\propto \exp \left\{ \sum_{i=1}^n [\delta_i (\lambda - \exp \{ \lambda \} y_i^\alpha) + \exp \{ \mathbf{X}_i \boldsymbol{\beta}^\top - \exp \{ \lambda \} y_i^\alpha \}] \right\} \\
&\quad \times \pi(\lambda | m, v). \tag{1.12}
\end{aligned}$$

Após obtermos o núcleo das condicionais completas de todos os parâmetros do modelo, podemos realizar amostragem indireta através de métodos MCMC. Um método a ser considerado aqui é o Amostrador de *Gibbs* (Gelfand *et al.*, 1990; Gelfand, 2000; Migon *et al.*, 2015) com passos de *Metropolis* adaptativo (Haario *et al.*, 2005, 2001), pelo fato das condicionais completas não possuírem forma fechada conhecida.

### 1.3.2 Modelo de Fração de Cura com Erro nas Covariáveis de Mizoi *et al.* (2007)

O modelo apresentado a seguir, é o de fração de cura com erro de medida nas covariáveis introduzido por Mizoi (2004) e discutido também por Mizoi *et al.* (2007) e Lima (2008). Este modelo é funcional e considera que apenas uma das covariáveis foi medida com erro. Na modelagem funcional as variáveis medidas com erro são consideradas constantes desconhecidas e fixas (ver Seção 3.2). A estimação dos parâmetros deste modelo é feita sob o ponto de vista clássico, através do método do score corrigido; Detalhes sobre essa metodologia de estimação podem ser obtidas em Nakamura (1992).

Suponhamos que no modelo de tempo de promoção, descrito na Subseção 1.3.1, a covariável  $\mathbf{X}_2$  seja medida com erro para cada indivíduo  $i = 1, 2, \dots, n$ . Admita que este erro seja aditivo de tal forma que o valor observável desta variável, seja tal que

$$W_{i2} = X_{i2} + \epsilon_i, \quad i = 1, 2, \dots, n$$

em que  $\epsilon_i$  é o erro cometido ao mensurar  $W_{i2}$ . Assuma que os erros  $\epsilon_i \sim \text{Normal}(0, \sigma_\epsilon^2)$  e são independentes de todas as outras variáveis do modelo de sobrevivência definido na Seção 1.3.1. Nesse caso, a função de sobrevivência para o modelo funcional de fração de cura com erro de medida nas covariáveis (MFFC) é dada por:

$$S(y|\boldsymbol{\gamma}, \boldsymbol{\beta}) = \exp \left\{ - \exp \left\{ \mathbf{W}_1\beta_1 + (\mathbf{X}_2 + \boldsymbol{\epsilon})\beta_2 + \mathbf{W}_3\beta_3 + \dots + \mathbf{W}_p\beta_p \right\} F(y|\boldsymbol{\gamma}) \right\}, \quad (1.13)$$

em que  $\mathbf{X}_2$  é uma constante desconhecida que representa o valor real e não observado da covariável.

A função de verossimilhança é análoga à do modelo de tempo de promoção, Equação (1.7). Substituindo na Equação (1.7) a função de ligação  $\exp \{ \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \mathbf{X}_3\beta_3 + \dots + \mathbf{X}_p\beta_p \}$ , por  $\exp \{ \mathbf{W}_1\beta_1 + (\mathbf{X}_2 + \boldsymbol{\epsilon})\beta_2 + \mathbf{W}_3\beta_3 + \dots + \mathbf{W}_p\beta_p \}$ , a função de verossimilhança para o MFFC

torna-se

$$\begin{aligned}
 L(\gamma, \theta | \mathbf{D}_{\text{obs}}) &\propto \prod_{i=1}^n [f_p(y_i | \gamma, \theta)]^{\delta_i} [S_p(y_i | \gamma, \theta)]^{1-\delta_i} \\
 &\propto \prod_{i=1}^n [\exp \{ \mathbf{W}_1 \beta_1 + (\mathbf{X}_2 + \boldsymbol{\epsilon}) \beta_2 + \mathbf{W}_3 \beta_3 + \cdots + \mathbf{W}_p \beta_p \} f(y_i | \gamma) \times \\
 &\quad \exp \{ - \exp \{ \mathbf{W}_1 \beta_1 + (\mathbf{X}_2 + \boldsymbol{\epsilon}) \beta_2 + \mathbf{W}_3 \beta_3 + \cdots + \mathbf{W}_p \beta_p \} F(y_i | \gamma) \}]^{\delta_i} \times \\
 &\quad [\exp \{ - \exp \{ \mathbf{W}_1 \beta_1 + (\mathbf{X}_2 + \boldsymbol{\epsilon}) \beta_2 + \mathbf{W}_3 \beta_3 + \cdots + \mathbf{W}_p \beta_p \} F(y_i | \gamma) \}]^{1-\delta_i}.
 \end{aligned} \tag{1.14}$$

Para estimar os parâmetros do modelo, os quais também incluem  $\mathbf{X}_2$ , Mizoi (2004) recorreu ao método do escore corrigido, comumente utilizado em modelos com erro de medida (Nakamura, 1992), resolvendo a equação

$$\mathbb{E} [l^*(\boldsymbol{\beta}, \boldsymbol{\gamma} | \mathbf{D}_{\text{obs}}, \mathbf{X}_2)] = l(\gamma, \theta | \mathbf{D}_{\text{obs}}),$$

em que  $l^*$  é o logaritmo natural da função de verossimilhança do MFFC corrigida a partir da log-verossimilhança não observada dada a seguir:

$$\begin{aligned}
 l(\gamma, \theta | \mathbf{D}_{\text{obs}}) &\propto \ln \left[ \prod_{i=1}^n [f_p(y_i | \gamma, \theta)]^{\delta_i} [S_p(y_i | \gamma, \theta)]^{1-\delta_i} \right] \\
 &\propto \ln \left[ \prod_{i=1}^n [\exp \{ \mathbf{W}_1 \beta_1 + \mathbf{X}_2 \beta_2 + \mathbf{W}_3 \beta_3 + \cdots + \mathbf{W}_p \beta_p \} f(y_i | \gamma) \times \right. \\
 &\quad \exp \{ - \exp \{ \mathbf{W}_1 \beta_1 + \mathbf{X}_2 \beta_2 + \mathbf{W}_3 \beta_3 + \cdots + \mathbf{W}_p \beta_p \} F(y_i | \gamma) \}]^{\delta_i} \times \\
 &\quad \left. [\exp \{ - \exp \{ \mathbf{W}_1 \beta_1 + \mathbf{X}_2 \beta_2 + \mathbf{W}_3 \beta_3 + \cdots + \mathbf{W}_p \beta_p \} F(y_i | \gamma) \}]^{1-\delta_i} \right].
 \end{aligned} \tag{1.15}$$

Para mais detalhes sobre como é feita a inferência no MFFC e aplicações deste modelo, ver Mizoi *et al.* (2007) e Mizoi (2004).

### 1.3.3 Modelo de Fração de Cura com Erro nas Covariáveis de Ma e Yin (2008)

Nesta seção, o modelo de fração de cura com erro de medida proposto por Ma e Yin (2008) é apresentado. As covariáveis com erro de medida também recebem uma modelagem funcional com erro normal aditivo. Esse modelo é baseado no modelo de riscos proporcionais de Cox (1972), que utiliza o mesmo preditor linear do MTP (Chen *et al.*, 1999) para introduzir



as covariáveis. Entretanto, a proposta de Ma e Yin (2008) não leva em consideração os riscos competitivos que é uma característica singular do MTP (Chen *et al.*, 1999; Ibrahim *et al.*, 2005; Rodrigues *et al.*, 2008). Além disso, Ma e Yin (2008) assumem uma abordagem semi-paramétrica frequentista de inferência, em que a maximização da função verossimilhança é feita empregando o método dos multiplicadores de Lagrange e uma aproximação para o escore corrigido.

Considere o modelo proposto por Yakovlev e Tsodikov (1996):

$$S(t|\mathbf{X}) = \exp\{-\theta(\mathbf{X})F(t)\}, \quad (1.16)$$

em que  $\theta(\mathbf{X})$  é uma função de ligação e  $F(t)$  é uma função de distribuição acumulada basal desconhecida. Neste modelo, a fração de cura é definida como sendo  $S(\infty|\mathbf{X}) = \exp\{-\theta(\mathbf{X})\}$  e a função de risco é  $h(t|\mathbf{X}) = \theta(\mathbf{X})f(t)$ , em que  $f(t) = \frac{dF(t)}{dt}$ .

Vamos admitir  $\theta(\mathbf{X}) = \exp\{\mathbf{X}\boldsymbol{\beta}\}$ , em que  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^\top$ , é um vetor  $1 \times p + 1$  de coeficientes de regressão e  $\mathbf{X}$  é a matriz de delineamento de ordem  $n \times (p + 1)$  cuja a linha  $\mathbf{X}_i = (1, X_{i1}, \dots, X_{ip+1})$  contém a informação das covariáveis associadas ao  $i$ -ésimo indivíduo,  $i = 1, \dots, n$ .

Seja  $\mathbf{W}$  a matriz de covariáveis observada com erro, isto é,  $\mathbf{W} = \mathbf{X} + \mathbf{E}$  sendo  $\mathbf{E}$  o erro de medição e  $\mathbf{X}$  a matriz que contém os verdadeiros valores não observados das covariáveis.

Suponhamos  $T_i$  e  $C_i$  os tempos de falha e censura, respectivamente, relacionados ao indivíduo  $i = 1, \dots, n$ . Defina  $Y_i = \min\{T_i, C_i\}$  e  $\delta_i = 1_{\{T_i \leq C_i\}}$  como sendo o indicador de falha para o indivíduo  $i$ . Denotemos por  $\mathbf{X}_i = (X_{i0}, X_{i1}, \dots, X_{ip})$ , em que  $X_{i0} = 1$  para todo  $i$ , o vetor linha de ordem  $p + 1$  composto pelas covariáveis relacionadas ao  $i$ -ésimo indivíduo, assim a matriz de delineamento de experimento é dada por  $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)^\top$ . Consideremos  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^\top$  o vetor de efeitos fixos de ordem  $(p + 1) \times 1$ . A informação disponível para o indivíduo  $i$  é representada por  $(Y_i, \delta_i, \mathbf{W}_i)$ . Assuma que os mecanismos de falha e censura são independentes, isto é,  $T_i$  é independente de  $C_i$ . Além disto, suponha que dado  $\mathbf{X}_i$  as informações para cada indivíduo são independentes, ou seja, dado o valor real da covariável  $\mathbf{X}$ , a covariável observada  $\mathbf{W}$  não contém nenhuma informação adicional sobre  $\mathbf{Y}$ .

Considerando que alguns indivíduos nunca irão experimentar o evento de interesse, ou seja, “ $Y_i = \infty$ ”, para alguns indivíduos, a fração de curados é definida da seguinte maneira. Considere  $L$  um limiar, dizemos que um indivíduo  $i$  é curado se  $Y_i > L$ . Ma e Yin (2008) sugerem que  $L$  seja o maior tempo de falha observado. No entanto, outros critérios podem se considerados para definir  $L$ , como por exemplo, o conhecimento de especialistas na área.

Supondo a estrutura aditiva para o erro nas covariáveis, a função de sobrevivência do

modelo introduzido por Ma e Yin (2008) é dada por

$$S(t|\mathbf{X}) = \exp\{-F(t) \exp\{\mathbf{W}\boldsymbol{\beta}\}\}, \quad \mathbf{W} = \mathbf{X} + \mathbf{E}; \quad (1.17)$$

em que  $\mathbf{E} = (\boldsymbol{\epsilon}_o^\top, \boldsymbol{\epsilon}_1^\top, \dots, \boldsymbol{\epsilon}_p^\top)$ . Suponhamos  $\boldsymbol{\epsilon}_p \sim N_n(\mathbf{0}, \mathbf{V}_p)$ . Para acomodar o caso em que algumas variáveis sejam medidas sem erro sejam  $\mathbf{X}_o$  e  $\mathbf{X}_E$  as variáveis medidas sem erro e com erro, respectivamente, em que  $\mathbf{X}_E$  é um vetor  $k$  dimensional e não observado. Desse modo,  $\mathbf{X} = (\mathbf{X}_o, \mathbf{X}_E)$  e, portanto,  $\mathbf{W} = (\mathbf{X}_o, \mathbf{W}_E)$  em que  $\mathbf{W}_E = \mathbf{X}_E + \mathbf{E}$ . O interesse consiste em modelar  $(\mathbf{Y}, \boldsymbol{\delta}, \mathbf{W})$ . Uma vez que,  $\mathbf{Y}$  e  $\mathbf{W}$  são independentes, condicional em  $\mathbf{X}$ , segue que:

$$\begin{aligned} f(\mathbf{Y}, \boldsymbol{\delta}, \mathbf{W} | \mathbf{X}) &= f(\mathbf{Y}, \boldsymbol{\delta}, \mathbf{W} | \mathbf{X}_o, \mathbf{X}_E) \\ &= f(\mathbf{Y}, \boldsymbol{\delta} | \mathbf{W}, \mathbf{X}_o, \mathbf{X}_E) f_{\mathbf{W}}(\mathbf{W} | \mathbf{X}_o, \mathbf{X}_E) \\ &= f(\mathbf{Y}, \boldsymbol{\delta} | \mathbf{X}_o, \mathbf{X}_E) f_{\mathbf{W}_E}(\mathbf{W}_E | \mathbf{X}_E). \end{aligned} \quad (1.18)$$

Um modelo com essa propriedade é chamado de modelo com erro não diferencial; ver Bolfarine e Arellano-Valle (1998).

A seguir, temos a definição da fração de cura do MFFC proposto por Ma e Yin (2008). A forma com que foi introduzida a fração de curados nesse caso, é bem diferenciada dentre os modelos de fração de cura presentes na literatura de análise de sobrevivência.

Defina a função de sobrevivência e a função densidade de probabilidade para o evento de interesse (tempo de falha) como sendo, respectivamente,

$$S_e(Y | \mathbf{X}) = \exp\{-F(Y) \exp\{\mathbf{X}\boldsymbol{\beta}\}\}, \quad (1.19)$$

$$f_e(Y | \mathbf{X}) = \exp\{-F(Y) \exp\{\mathbf{X}\boldsymbol{\beta}\}\} f(Y) \exp\{\mathbf{X}\boldsymbol{\beta}\}, \quad (1.20)$$

em que  $F(y) = \int_0^y f(t) dt$ . Semelhantemente, admita que tais funções para os tempos de censura são, respectivamente, dadas por

$$S_c(Y | \mathbf{X}) = \mathbb{P}(C > Y | \mathbf{X}), \quad (1.21)$$

$$f_c(Y | \mathbf{X}) = \frac{\partial}{\partial Y} S_c(Y | \mathbf{X}). \quad (1.22)$$

É natural esperar que um certo número de indivíduos nunca experimentarão a falha ou a censura à direita. Para esses indivíduos, “ $Y = \infty$ ” com probabilidade 1. Desse modo, a fração de curados pode ser definida da seguinte maneira: Seja  $L \in \mathbb{R}^+ \cup \{0\}$  um limiar de cura. Se  $Y_i > L$ , então o indivíduo é curado. Assuma que  $Y_i = \infty$ . Assim, a fração de curados para ambos os indivíduos que experimentaram o evento de interesse e os indivíduos que foram censurados,

são, respectivamente, dadas por:

$$S_e(\infty | \mathbf{X}) = \lim_{Y \rightarrow \infty} \exp\{-\exp\{\theta(\mathbf{X})F(Y)\}\} \quad (1.23)$$

e

$$S_c(\infty | \mathbf{X}) = \lim_{Y \rightarrow \infty} S_c(Y | \mathbf{X}). \quad (1.24)$$

Sabemos que quando  $Y \rightarrow \infty$ ,  $F(Y) \rightarrow 1$ , visto que  $F$  é uma função de distribuição acumulada, logo  $\lim_{Y \rightarrow \infty} \exp\{-\exp\{\theta(\mathbf{X})F(Y)\}\} = \exp\{-\exp\{-\theta(\mathbf{X})\}\}$ . Portanto, a fração de cura para os indivíduos que experimentaram o evento de interesse é dada pela função de sobrevivência na Equação (1.23), como segue

$$S_e(\infty | \mathbf{X}) = \exp\{-\exp\{-\theta(\mathbf{X})\}\}, \quad (1.25)$$

e para os indivíduos censurados, podemos notar que a fração de cura na Equação (1.24) é estritamente positiva,  $S_c(\infty | \mathbf{X}) > 0$ .

### 1.3.3.1 Função de Verossimilhança

Admitindo a formulação na Expressão (1.17) e assumindo uma amostra i.i.d de indivíduos, se  $\mathbf{X}$  é observado, a função de verossimilhança do vetor de parâmetros do modelo é

$$\begin{aligned} L(\psi | \mathbf{Y}, \delta, \mathbf{W}, \mathbf{X}) &= \left[ \prod_{i=1}^n f(y_i, \delta_i | \mathbf{X}_i) \right] f_{\mathbf{W}_E}(\mathbf{W}_E | X_E) \\ &= \left[ \prod_{i=1}^n f(y_i, \delta_i | \mathbf{X}_{i_o}, \mathbf{X}_{E_i}) \right] f_{\mathbf{W}_E}(\mathbf{W}_E | X_E) \\ &= \left[ \prod_{i=1}^n \left[ (f_e(y_i | \mathbf{X}_{o_i}, \mathbf{X}_{E_i}) S_c(y_i | \mathbf{X}_{o_i}, \mathbf{X}_{E_i}))^{\delta_i} (f_c(y_i | \mathbf{X}_{i_o}, \mathbf{X}_{i_e}) \right. \right. \\ &\quad \left. \left. \times S_e(y_i | \mathbf{X}_{o_i}, \mathbf{X}_{E_i})^{1-\delta_i} \right]^{\mathbb{1}_{\{y_i < \infty\}}} [S_e(\infty | \mathbf{X}_{o_i}, \mathbf{X}_{E_i}) S_c(\infty | \mathbf{X}_{o_i}, \mathbf{X}_{E_i})]^{\mathbb{1}_{\{y_i = \infty\}}} \right] \\ &\quad \times f_{\mathbf{W}_E}(\mathbf{W}_E | \mathbf{X}_E) \\ &= \left[ \prod_{i=1}^n \left[ (\exp\{-F(y_i) \exp\{\mathbf{X}_i \boldsymbol{\beta}\}\} f(y_i) \exp\{\mathbf{X}_i \boldsymbol{\beta}\} \right. \right. \\ &\quad \times S_c(y_i | \mathbf{X}_i))^{\delta_i} (f_c(y_i | \mathbf{X}_i) \exp\{-F(y_i) \exp\{\mathbf{X}_i \boldsymbol{\beta}\}\})^{1-\delta_i} \left. \right]^{\mathbb{1}_{\{y_i < \infty\}}} \\ &\quad \left. \times (\exp\{-\exp\{\mathbf{X}_i \boldsymbol{\beta}\}\} S_c(\infty | \mathbf{X}_i))^{\mathbb{1}_{\{y_i < \infty\}}} \right] f_{\mathbf{W}_E}(\mathbf{W}_E | X_E), \quad (1.26) \end{aligned}$$

em que  $\mathbb{1}_{\{A\}}$  denota a função indicadora assumindo 1 se  $A$  ocorre e 0 caso contrário, e  $\exp\{\mathbf{X}_i \boldsymbol{\beta}\} = \exp\{\mathbf{X}_{o_i} \boldsymbol{\beta}_o + \mathbf{X}_{E_i} \boldsymbol{\beta}_E\}$ . Sob a suposição de censura não informativa, a Equação

(1.26) torna-se:

$$L(\boldsymbol{\psi}, \mathbf{X} | \mathbf{Y}, \boldsymbol{\delta}, \mathbf{W}) \propto \left[ \prod_{i=1}^n \left[ (f(y_i) \exp\{\mathbf{X}_i \boldsymbol{\beta}\})^{\delta_i} \exp\{-F(y_i) \exp\{\mathbf{X}_i \boldsymbol{\beta}\}\} \right]^{\mathbb{1}_{\{y_i < \infty\}}} \right. \\ \left. (\exp\{-\exp\{\mathbf{X}_i \boldsymbol{\beta}\}\})^{\mathbb{1}_{\{y_i = \infty\}}} \right] f_{\mathbf{W}_E}(\mathbf{W}_E | X_E). \quad (1.27)$$

Para inferir sobre o modelo, Ma e Yin (2008) propõem maximizar o logaritmo natural da função de verossimilhança através do método dos Multiplicadores de Lagrange e usar uma aproximação para o escore corrigido, baseado no estimador de máxima verossimilhança não paramétrico. Além disso, propõem uma transformação nas covariáveis contaminadas, para modelar as situações em que a suposição de normalidade não é razoável. Mais detalhes sobre como estimativas para os parâmetros deste modelo são obtidas utilizando a metodologia empregada pelos autores, sugere-se consultar Ma e Yin (2008). Uma outra maneira de fazer inferência neste modelo foi considerada em Bertrand *et al.* (2016) e Bertrand *et al.* (2017), que propõem um método SIMEX para obter estimadores para os parâmetros do modelo.

No Capítulo 2 a seguir, propomos um novo modelo de fração de cura com erro de medida nas covariáveis no qual a modelagem das covariáveis medidas com erro é vista como estrutural no âmbito de estatística clássica. Consideramos o erro normal aditivo e modelamos a incerteza sobre as variáveis latentes atribuindo a elas uma distribuição *a priori*. Além disso, uma reformulação Bayesiana dos modelos propostos por Mizoi *et al.* (2007) e Ma e Yin (2008) é apresentada, desse modo tais modelos passam a receber um tratamento estrutural para as covariáveis medidas com erro. Na modelagem proposta, um modelo sem intercepto para descrever os tempos de falha é considerado. A abordagem Bayesiana será utilizada para fazer inferência sobre os parâmetros. Um estudo de simulação para avaliar o desempenho dos modelos propostos é realizado.

## Capítulo 2

# Modelo Bayesiano de Fração de Cura (MBFC) com Distribuição Normal para o Erro de Medida

Neste capítulo, apresentamos um novo modelo Bayesiano de fração de cura com erro de medida nas covariáveis (MBFC), com estrutura normal aditiva. As covariáveis medidas com erro são tratadas como variáveis latentes e, por essa razão, do ponto de vista Bayesiano, serão estimadas juntamente com os parâmetros do modelo. Diferentemente do que tem sido assumido na literatura, a variância do erro de medida  $\sigma_\epsilon^2$  não será assumida necessariamente conhecida. Para inferir acerca de tal parâmetro propomos duas distribuições *a priori* para modelar seu comportamento. No primeiro caso, atribuímos uma distribuição *a priori* Gama Invertida para  $\sigma_\epsilon^2$ , seguindo o padrão vigente na literatura. Como segunda proposta, nos inspirados em resultados estabelecidos na literatura de modelos com erro de medida (Fuller, 2009) para resolver problemas de falta de identificabilidade, modelamos a incerteza a respeito da razão  $k$  entre as variâncias do modelo  $\sigma_\epsilon^2$  e  $\sigma_x^2$  (variância da variável latente). Além destas duas abordagens, para fins de comparação, também fixamos  $\sigma_\epsilon^2$  em alguns valores, seguindo o que já foi proposto na literatura de modelos de fração de cura com erro de medida nas covariáveis. Denotando essas três modelagens por MBFC2, MBFC3 e MEBC1, respectivamente. As especificações consideradas são dadas na Tabela 2.1.

Consideremos o modelo de tempo de promoção proposto por Chen *et al.* (1999). Suponha que, de um modo mais geral, para um indivíduo  $i$  na população,  $N_i$  denote o número de causas do evento de interesse após o tratamento inicial. Assuma que  $N_i \sim \text{Poisson}(\theta_i)$ . Denotemos por  $R_{ij}$ ,  $j = 1, \dots, N_i$ , variáveis aleatórias (v.a.) que caracterizam o tempo de ocorrência do evento de interesse para as  $N_i$  causas. Considere que, dado  $N_i = N$ , as v.a.'s  $R_{ij}$  são i.i.d. com

Tabela 2.1: Modelos Bayesianos de fração de cura com distribuição Normal para o erro de medida.

Modelo	Suposição <i>a priori</i> sobre $\sigma_\epsilon^2$
MBFC1	$\sigma_\epsilon^2$ é uma constante fixa conhecida.
MBFC2	$\sigma_\epsilon^2 \sim GI(a, b)$ .
MBFC3	$\sigma_\epsilon^2 = k\sigma_x^2$ , $k \sim \pi(k)$ , em que $0 < k \leq 1$ .

f.d.a comum  $F(t) = 1 - S(t)$ . Admita também que  $R_{i1}, \dots, R_{iN}$  são independentes de  $N$ . O tempo de reincidência do câncer é definido como sendo a v.a.  $T = \min \{R_{ij}; 0 \leq j \leq N\}$ , em que  $\mathbb{P}(R_{i0} = \infty) = 1$ , isto é, dado que o número de células carcinogênicas ativas após o tratamento inicial é  $N_i = 0$ , a probabilidade de que haja reincidência da doença em um tempo finito é 0. A função de sobrevivência para a população é dada por

$$S_p(t) = \mathbb{P}(\text{não ter câncer até o tempo } t) = \exp\{-\theta_i F(t)\}. \quad (2.1)$$

No modelo em (2.1) a fração de curados é dada por  $S_p(\infty) = \exp\{-\theta_i\}$ , a qual é estritamente positiva e representa a probabilidade do indivíduo não desenvolver células cancerígenas até o tempo  $t$ .

Para  $n$  indivíduos da amostra, defina  $\boldsymbol{\theta} = \exp\{\mathbf{X}\boldsymbol{\beta}\}$ . Com esta suposição a função de sobrevivência populacional do modelo de tempo de promoção torna-se:

$$S_p(t | \mathbf{X}) = \exp\{-\exp\{\mathbf{X}\boldsymbol{\beta}\} F(t)\}, \quad (2.2)$$

em que  $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)^\top$  denota a matriz de delineamento de experimento de ordem  $n \times p$ ,  $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})$  é um vetor de covariáveis explicativas de ordem  $1 \times p$  relacionadas ao indivíduo  $i$  e  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$  é o vetor de coeficientes de regressão de ordem  $p \times 1$ , comum a todos os indivíduos. Definimos um modelo sem intercepto com o objetivo de resolver problemas de falta de identificabilidade no modelo.

Seguindo a proposta de Chen *et al.* (1999) para o modelo de tempo de promoção, relacionamos o número esperado de células carcinogênicas ativas ou causas do evento de interesse à covariáveis fazendo  $\theta_i = \mathbb{E}(N_i) = \exp\{\mathbf{X}_i\boldsymbol{\beta}\}$ . Ao fazermos isto, o modelo adquire a propriedade de riscos proporcionais, como destacamos ao definirmos o modelo de tempo de promoção na Seção 1.3. Essa propriedade é bastante desejável, pois, entre outras coisas, torna a implementação dos métodos MCMC relativamente mais simples, o que facilita a estimação do modelo sob o ponto de vista Bayesiano. Além disso, ao assumirmos que  $\boldsymbol{\theta} = \exp\{\mathbf{X}\boldsymbol{\beta}\}$ , o entendimento do papel dos coeficientes de regressão na fração de curados se torna mais simples: coeficientes negativos levam a uma fração de cura maior do que coeficientes positivos. Para o grupo de não

curados, os coeficientes de regressão afetam bastante a função de risco, isto é, um coeficiente de regressão negativo leva a um risco muito maior do que um positivo.

Na Equação (2.2) suponha que as covariáveis  $\mathbf{X}$  foram observadas com erro, isto é, observamos

$$\mathbf{W} = \mathbf{X} + \mathbf{E},$$

em que  $\mathbf{E} = (\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n)^\top$  é a matriz  $n \times p$  de erros de medida, tal que  $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{ip})$  em que  $\epsilon_{ip} \sim N(0, \sigma_p^2)$  é o erro associado a a covariável  $p$  para todo indivíduo  $i$ . Considere que  $\text{cov}(\boldsymbol{\epsilon}_j, \boldsymbol{\epsilon}_l | \sigma_j^2, \sigma_l^2) = 0$  para todo  $j \neq l$ ,  $j, l = 1, \dots, p$  e que  $\mathbf{X}_i$  é independente de  $\boldsymbol{\epsilon}_i$ , dado  $(\sigma_1^2, \dots, \sigma_p^2)$ . Para tornar o modelo mais realista, assuma que  $m$  entre as  $p$  covariáveis tenham sido medidas com erro. Neste contexto,  $\mathbf{W} = (\mathbf{X}_E + \mathbf{E}, \mathbf{X}_o)$  denota as covariáveis observadas, em que  $\mathbf{X}_E$  denota a matriz latente de variáveis medidas com erro e  $\mathbf{X}_o$  é a matriz de variáveis observadas. Assim, supondo que  $0 \leq m \leq p$ ,  $\mathbf{X}_E$  e  $\mathbf{E}$  têm dimensão  $n \times m$  e  $\mathbf{X}_o$  tem dimensão  $n \times (p - m)$ .

Podemos escrever a função de sobrevivência, baseada nos dados observados, da seguinte forma:

$$S(t|\boldsymbol{\beta}, \boldsymbol{\gamma}) = \exp \{- \exp \{ \mathbf{W}\boldsymbol{\beta} \} F(t|\boldsymbol{\gamma}) \}.$$

A função de sobrevivência baseada nos valores reais  $(\mathbf{X}_E, \mathbf{X}_o)$ , é dada por

$$S(t|\boldsymbol{\beta}, \boldsymbol{\gamma}) = \exp \{- \exp \{ \mathbf{X}_E\boldsymbol{\beta}_E + \mathbf{X}_o\boldsymbol{\beta}_o \} F(t|\boldsymbol{\gamma}) \}, \quad (2.3)$$

em que  $\boldsymbol{\beta}_E$  denota o vetor dos coeficientes das variáveis latentes e  $\boldsymbol{\beta}_o$  é o vetor dos coeficientes das variáveis observadas. Tais vetores são, respectivamente, de ordem  $m \times 1$  e  $(p - m) \times 1$ . A função de sobrevivência em (2.3) é aleatória, pois, depende das variáveis latentes  $\mathbf{X}_E$ .

## 2.1 Inferência Bayesiana no MBFC

Considerando a estrutura do modelo apresentado na seção anterior, nesta seção construiremos a função de verossimilhança para o MEFC e completamos as especificações do modelo estabelecendo distribuições *a priori* para os parâmetros. Três modelos distintos serão propostos a partir de diferentes especificações *a priori* para a variância do erro de medida.

Suponhamos que  $n$  indivíduos foram observados independentemente e que  $N_i$  é o número de células carcinogênicas do  $i$ -ésimo indivíduo observado. Assumamos que  $N_i | \theta_i \sim \text{Poisson}(\theta_i)$ , em que  $\theta_i = \exp \{ \mathbf{X}_{\boldsymbol{\epsilon}_i}\boldsymbol{\beta}_E + \mathbf{X}_{o_i}\boldsymbol{\beta}_o \}$  e  $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{im})$ , para todo  $i = 1, 2, \dots, n$ . As variáveis  $N_i$  são não observáveis e, portanto, são variáveis latentes na formulação do modelo. Relembramos que  $R_{i1}, R_{i2}, \dots, R_{iN_i}$  são os tempos de ocorrência das  $N_i$  causas do evento de interesse e

admitimos que a função de distribuição acumulada de cada  $R_{ij}$  é denotada por  $F(\cdot | \gamma)$ , em que  $\gamma$  é o vetor de parâmetros associados aos tempos de incubação. Nesse caso, temos que  $S(\cdot | \gamma)$  é a função de sobrevivência de cada tempo de ocorrência.

Seja  $T_i$  o tempo de falha do indivíduo  $i$  e considere que  $T_i$  pode ser censurado à direita. Denote por  $C_i$  o tempo de censura. Dessa maneira, observamos  $y_i = \min \{T_i, C_i\}$  e definimos o indicador de falha,  $\delta_i$ , como segue:

$$\delta_i = \begin{cases} 1, & \text{se } T_i \leq C_i, \\ 0, & \text{se } T_i > C_i. \end{cases}$$

Assuma que os mecanismos de falha e censura são independentes, isto é,  $T_i$  é independente de  $C_i$  para todo  $i$ .

Tal como foi considerado no modelo de tempo de promoção (Seção 1.3), a inferência no MBFC é baseada no vetor de dados observados, representado por  $\mathbf{D}_{\text{obs}} = (n, \mathbf{Y}, \boldsymbol{\delta}, \mathbf{W})$ , em que  $\mathbf{Y} = (y_1, y_2, \dots, y_n)$ ,  $\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_n)$  e  $\mathbf{W}$  é a matriz de delineamento de experimento de ordem  $n \times p$ . A função de verossimilhança relacionada ao MBFC, considerando um indivíduo  $i$  cujo o tempo de falha foi observado, é dada por

$$L(\gamma, \theta_i | y_i, \delta_i, \mathbf{W}_i) = f(y_i, \delta_i, \mathbf{W}_i | \gamma, \theta_i). \quad (2.4)$$

Condicionalmente em  $\mathbf{X}_{\epsilon_i}$  e  $\mathbf{X}_{o_i}$ , temos que  $y_i$  é independente de  $\mathbf{W}_i$  e a função de verossimilhança em (2.4) pode ser reescrita como segue.

$$\begin{aligned} L(\gamma, \boldsymbol{\beta} | y_i, \delta_i, \mathbf{W}_i, \mathbf{X}_{\epsilon_i}, \mathbf{X}_{o_i}) &= f(y_i, \delta_i, | \mathbf{W}_i, \mathbf{X}_{\epsilon_i}, \mathbf{X}_{o_i}, \gamma, \boldsymbol{\beta}, \Sigma) f_{\mathbf{W}_i}(\mathbf{w}_i | \mathbf{X}_{\epsilon_i}, \mathbf{X}_{o_i}, \Sigma) \\ &= f(y_i, \delta_i, | \mathbf{X}_{\epsilon_i}, \mathbf{X}_{o_i}, \gamma, \theta_i) f_{\mathbf{W}_i}(\mathbf{w}_i | \mathbf{X}_{\epsilon_i}, \mathbf{X}_{o_i}, \Sigma), \end{aligned} \quad (2.5)$$

em que  $\Sigma = (\sigma_1^2, \dots, \sigma_p^2)$ .

Para uma amostra independente de  $n$  indivíduos, considerando o resultado na expressão (2.5), temos que a função de verossimilhança é

$$L(\gamma, \theta | \mathbf{Y}, \boldsymbol{\delta}, \mathbf{X}_E, \mathbf{X}_O, \Sigma) \propto \underbrace{\left[ \prod_{i=1}^n f(y_i, \delta_i, \mathbf{X}_{\epsilon_i}, \mathbf{X}_{o_i} | \gamma, \theta_i) \right]}_{\text{I}} \underbrace{f_{\mathbf{W}}(\mathbf{w} | \mathbf{X}_E, \mathbf{X}_O, \Sigma)}_{\text{II}}. \quad (2.6)$$

A parte (I) da Equação(2.6) corresponde à função de verossimilhança do modelo de tempo de promoção sem erro de medida, proposto por Chen *et al.* (1999). A parte (II) descreve a incerteza sobre as variáveis medidas com erro, as quais são aqui assumidas como aleatórias.

Admita que parte dos indivíduos da amostra foram censurados. Sob este suposto, a função



de verossimilhança em (2.6) torna-se

$$L(\gamma, \theta | \mathbf{Y}, \delta, \mathbf{X}_E, \mathbf{X}_o, \Sigma) \propto \underbrace{\left[ \prod_{i=1}^n [f_p(y_i | \mathbf{X}_{\epsilon_i}, \mathbf{X}_{o_i}, \gamma, \boldsymbol{\beta})]^{1-\delta_i} [S_p(y_i | \mathbf{X}_{\epsilon_i}, \mathbf{X}_{o_i}, \gamma, \boldsymbol{\beta})]^{\delta_i} \right]}_I \times \underbrace{f_{\mathbf{W}}(\mathbf{w} | \mathbf{X}_E, \mathbf{X}_o, \Sigma)}_II. \quad (2.7)$$

em que  $f_p$  e  $S_p$  são as funções densidade de probabilidade e de sobrevivência, respectivamente, considerando todos os indivíduos da população.

Suponha que os tempos de falha  $y_i$  são i.i.d. com distribuição Weibull( $\alpha, \lambda$ ), cuja função densidade de probabilidade é dada por

$$f(y | \alpha, \lambda) = \exp\{\lambda\} \alpha y^{\alpha-1} \exp\{-\exp\{\lambda\} y^\alpha\}. \quad (2.8)$$

Sob tais supostos e assumindo que para o indivíduo  $i$  temos  $\theta_i = \exp\{-\exp\{\mathbf{X}_{\epsilon_i} \boldsymbol{\beta}_E + \mathbf{X}_{o_i} \boldsymbol{\beta}_o\}\}$ , em que  $\mathbf{X}_{\epsilon_i}$  e  $\mathbf{X}_{o_i}$  denotam os vetores de variáveis medidas com e sem erro, respectivamente, para tal indivíduo, segue da Equação (2.7) a função de verossimilhança para o MEFC, dada por

$$\begin{aligned} L(\alpha, \lambda, \boldsymbol{\beta}, \Sigma | \mathbf{D}_{\text{obs}}, \mathbf{X}_E) &\propto \left[ \prod_{i=1}^n \left( \exp\{\mathbf{X}_{\epsilon_i} \boldsymbol{\beta}_E + \mathbf{X}_{o_i} \boldsymbol{\beta}_o\} \exp\{\lambda\} y_i^{\alpha-1} \exp\{-\exp\{\lambda\} y_i^\alpha\} \right. \right. \\ &\quad \left. \left. \times \exp\{-\exp\{\mathbf{X}_{\epsilon_i} \boldsymbol{\beta}_E + \mathbf{X}_{o_i} \boldsymbol{\beta}_o\} [1 - \exp\{-\exp\{\lambda\} y_i^\alpha\}]\} \right)^{\delta_i} \right. \\ &\quad \left. \times \left( \exp\{-\exp\{\mathbf{X}_{\epsilon_i} \boldsymbol{\beta}_E + \mathbf{X}_{o_i} \boldsymbol{\beta}_o\} [1 - \exp\{-\exp\{\lambda\} y_i^\alpha\}]\} \right)^{1-\delta_i} \right] \\ &\quad \times f_{\mathbf{W}}(\mathbf{w} | \mathbf{X}_E, \mathbf{X}_o, \Sigma) \\ &= \left[ \prod_{i=1}^n \left( \exp\{\mathbf{X}_{\epsilon_i} \boldsymbol{\beta}_E + \mathbf{X}_{o_i} \boldsymbol{\beta}_o\} \exp\{\lambda\} y_i^{\alpha-1} \exp\{-\exp\{\lambda\} y_i^\alpha\} \right)^{\delta_i} \right. \\ &\quad \left. \times \exp\{-\exp\{\mathbf{X}_{\epsilon_i} \boldsymbol{\beta}_E + \mathbf{X}_{o_i} \boldsymbol{\beta}_o\} [1 - \exp\{-\exp\{\lambda\} y_i^\alpha\}]\} \right] \\ &\quad \times f_{\mathbf{W}}(\mathbf{w} | \mathbf{X}_E, \mathbf{X}_o, \Sigma). \quad (2.9) \end{aligned}$$

Para que o modelo fique completamente especificado, devemos estabelecer as distribuições *a priori* para as variáveis latentes  $\mathbf{X}_{\epsilon_i}$  e para os parâmetros envolvidos no modelo. Por razões práticas, nos estudos que vem a seguir supomos que apenas uma covariável foi medida com

erro. Denotemos por  $\mathbf{X}_\epsilon$  a variável latente no modelo. Desse modo a matriz de delineamento de experimento é dada por

$$\mathbf{W} = (\mathbf{X}_\epsilon + \boldsymbol{\epsilon}, \mathbf{X}_o),$$

em que  $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^\top$  é o vetor  $n \times 1$  de erros de medida associado a covariável  $\mathbf{W}_\epsilon = \mathbf{X}_\epsilon + \boldsymbol{\epsilon}$  e é tal que  $\epsilon_i \stackrel{i.i.d.}{\sim} N(0, \sigma_\epsilon^2)$ . Assumimos que  $\mathbf{X}_\epsilon = (X_{\epsilon_1}, \dots, X_{\epsilon_n})$ . Além disso, neste capítulo admitimos que as variáveis latentes  $X_{\epsilon_i} \mid \mu_x, \sigma_x^2 \stackrel{i.i.d.}{\sim} N(\mu_x, \sigma_x^2)$ , para todo  $i$ , em todas as modelagens propostas.

A seguir, introduzimos alguns modelos que diferem pela forma como a incerteza sobre a variância do erro de medida  $\sigma_\epsilon^2$  é quantificada, como expusemos na Tabela 2.1. Além disso, também calculamos as distribuições condicionais completas dos parâmetros nos três modelos, necessárias na implementação dos algoritmos MCMC utilizados para obter amostras das distribuições *a posteriori*. Para ver o cálculo da distribuição *a posteriori* conjunta dos parâmetros do MBFC, assim como a demonstração de que trata-se de uma distribuição de probabilidade própria, basta consultar o Apêndice B deste trabalho.

### 2.1.1 MBFC1: MBFC com $\sigma_\epsilon^2$ fixado

O modelo aqui proposto pode ser visto como uma versão Bayesiana e, sob o ponto de vista clássico, estrutural do modelo proposto por Mizoi *et al.* (2007) quando o intercepto é desconsiderado.

Denotemos por  $\boldsymbol{\eta} = (\boldsymbol{\beta}, \alpha, \lambda, \sigma_\epsilon^2, \mathbf{X}_\epsilon, \mu_x, \sigma_x^2)$  o parâmetro a ser estimado. Para completar a especificação do modelo assumamos *a priori* que

$$\begin{aligned} X_{\epsilon_i} &\stackrel{i.i.d.}{\sim} N(\mu_x, \sigma_x^2), \\ \pi(\boldsymbol{\beta}) &\propto 1, \\ \alpha &\sim G(a_1, b_1) \\ \lambda &\sim N(m_1, v_1) \\ \mu_x &\sim N(m_2, v_2) \\ \sigma_x^2 &\sim GI(a_2, b_2). \end{aligned} \tag{2.10}$$

Estas especificações *a priori* também serão usadas nos modelos apresentados nas próximas subseções. No modelo definido aqui, nomeado MBFC1, assumimos que  $\sigma_\epsilon^2$  é uma constante conhecida, isto é, admitimos uma distribuição *a priori* degenerada para este parâmetro. Sob tais especificações, as distribuições *a posteriori* não possuem forma fechada. As distribuições condicionais completas *a posteriori* (d.c.c) dos parâmetros, bem como, das variáveis latentes são então cal-

culadas, fornecendo os resultados apresentados a seguir.

Denotemos por  $\boldsymbol{\eta}_{(-\eta_i)}$  o vetor  $\boldsymbol{\eta}$  sem a componente  $\eta_i$ . Desse modo, temos que as d.c.c para os parâmetros do modelo são:

i) para cada regressor  $\beta_j$ ,  $j = 1, \dots, p$

$$\begin{aligned} \pi(\beta_j | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}_{(-\beta_j)}) &\propto \exp \left\{ \sum_{i=1}^n \delta_i (X_{ij} \beta_j) - \sum_{i=1}^n \exp \{ \mathbf{X}_{ei} \boldsymbol{\beta}_\epsilon + \mathbf{X}_{oi} \boldsymbol{\beta}_o \} \right. \\ &\quad \left. \times (1 - \exp \{ - \exp \{ \lambda \} y_i^\alpha \}) \right\}. \end{aligned} \quad (2.11)$$

ii) para o parâmetros  $\lambda$  associado à distribuição dos tempos de falha,

$$\begin{aligned} \pi(\lambda | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}_{(-\lambda)}) &\propto \exp \left\{ \lambda \sum_{i=1}^n \delta_i - \exp \{ \lambda \} \sum_{i=1}^n \delta y_i^\alpha + \sum_{i=1}^n \exp \{ \mathbf{X}_{ei} \boldsymbol{\beta}_\epsilon + \mathbf{X}_{oi} \boldsymbol{\beta}_o \} \right. \\ &\quad \left. - \exp \{ \lambda \} y_i^\alpha \right\} \exp \left\{ \frac{-(\lambda - m_1)^2}{2v_1} \right\}, \end{aligned} \quad (2.12)$$

iii) para o parâmetro  $\alpha$  associado à distribuição dos tempos de falha,

$$\begin{aligned} \pi(\alpha | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}_{(-\alpha)}) &\propto \exp \left\{ \sum_{i=1}^n [- \exp \{ \lambda \} \delta_i y_i^\alpha] + \exp \{ \mathbf{X}_{ei} \boldsymbol{\beta}_\epsilon + \mathbf{X}_{oi} \boldsymbol{\beta}_o - \exp \{ \lambda \} y_i^\alpha \} \right\} \times \\ &\quad \left[ \prod_{i=1}^n y_i^{(\alpha-1)\delta_i} \right] \alpha^{a_1-1+\sum_{i=1}^n \delta_i} \exp \{ -b_1 \alpha \}. \end{aligned} \quad (2.13)$$

iv) para a variável latente  $\mathbf{X}_\epsilon$  temos que

$$\begin{aligned} \pi(\mathbf{x}_\epsilon | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}_{(-\mathbf{X}_\epsilon)}) &\propto \exp \left\{ \sum_{i=1}^n \left[ \delta_i (\mathbf{X}_{ei} \boldsymbol{\beta}_\epsilon + \mathbf{X}_{oi} \boldsymbol{\beta}_o) - \exp \{ \mathbf{X}_{ei} \boldsymbol{\beta}_\epsilon + \mathbf{X}_{oi} \boldsymbol{\beta}_o \} \right] \right\} \times \\ &\quad \exp \left\{ \sum_{i=1}^n \exp \{ \mathbf{X}_{ei} \boldsymbol{\beta}_\epsilon + \mathbf{X}_{oi} \boldsymbol{\beta}_o \} \exp \{ - \exp \{ \lambda \} y_i^\alpha \} \right\} \times \\ &\quad f_{\mathbf{W}_\epsilon}(\mathbf{w}_\epsilon | \mathbf{x}_\epsilon, \sigma_\epsilon^2) \pi(\mathbf{x}_\epsilon | \mu_x, \sigma_x^2), \end{aligned} \quad (2.14)$$

em que  $f_{\mathbf{W}_\epsilon}(\mathbf{w}_\epsilon | \mathbf{x}_\epsilon, \sigma_\epsilon^2) = \phi_n(\mathbf{w} | \mathbf{x}_\epsilon, \sigma_\epsilon^2)$  é a f.d.p da Normal  $n$ -variada com vetor de médias  $\mathbf{x}_\epsilon$  e matriz de covariância  $\sigma_\epsilon^2 \mathcal{I}_n$ .

Para o  $i$ -ésimo indivíduo, a d.c.c de  $X_{\epsilon_i}$  é dada por

$$\begin{aligned} \pi(x_{\epsilon_i} | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}_{(-\mathbf{x}_{\epsilon})}) &\propto \exp \left\{ \left[ \delta_i (x_{\epsilon_i} \beta_{\epsilon} + \mathbf{X}_{oi} \boldsymbol{\beta}_o) - \exp \{ x_{\epsilon_i} \beta_{\epsilon} + \mathbf{X}_{oi} \boldsymbol{\beta}_o \} \right] \right\} \times \\ &\exp \{ \exp \{ x_{\epsilon_i} \beta_{\epsilon} + \mathbf{X}_{oi} \boldsymbol{\beta}_o \} \exp \{ - \exp \{ \lambda \} y_i^{\alpha} \} \} \times \\ &\phi \left( x_{\epsilon_i} \mid \frac{\sigma_x^2 w_{\epsilon_i} + \sigma_{\epsilon}^2 \mu_x}{\sigma_x^2 + \sigma_{\epsilon}^2}, \frac{\sigma_x^2 \sigma_{\epsilon}^2}{\sigma_x^2 + \sigma_{\epsilon}^2} \right). \end{aligned} \quad (2.15)$$

v) Para a média  $\mu_x$  e variância  $\sigma_x^2$  das variáveis latentes, respectivamente, temos:

$$\pi(\mu_x | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}_{(-\mu_x)}) \propto \exp \left\{ - \frac{\mathbf{v}_2 \mathcal{I}_n \mathbf{x}_{\epsilon} + \sigma_x^2 m_2}{2 \sigma_x^2 v_2} \left[ \mu_x - \frac{\mathbf{v}_2 \mathcal{I}_n \mathbf{x}_{\epsilon} + \sigma_x^2 m_2}{\sigma_x^2 + v_2} \right]^2 \right\}, \quad (2.16)$$

em que  $\mathbf{v}_2 = v_2 \mathbf{1}$ , tal que  $\mathbf{1}$  é um vetor de dimensão  $n \times 1$  de uns, e

$$\pi(\sigma_x^2 | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}_{(-\sigma_x^2)}) \propto (\sigma_x^2)^{-(a_2 + \frac{n}{2} - 1)} \exp \left\{ - \frac{[\frac{1}{2} (\mathbf{x}_{\epsilon} - \boldsymbol{\mu}_x)^{\top} \mathcal{I}_n (\mathbf{x}_{\epsilon} - \boldsymbol{\mu}_x) + b_2]}{\sigma_x^2} \right\}, \quad (2.17)$$

Como pode ser visto nas Equações (2.16) e (2.17), as d.c.c de  $\mu_x$  e  $\sigma_x^2$  são distribuições conhecidas que denotamos, respectivamente, por

$$\mu_x | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}_{(-\mu_x)} \sim N \left( \frac{\mathbf{v}_2 \mathcal{I}_n \mathbf{x}_{\epsilon} + \sigma_x^2 m_2}{\sigma_x^2 + v_2}, \frac{\sigma_x^2 v_2}{\sigma_x^2 + v_2} \right)$$

e

$$\sigma_x^2 | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}_{(-\sigma_x^2)} \sim GI \left( a_2 + \frac{n}{2}, \frac{1}{2} (\mathbf{x}_{\epsilon} - \boldsymbol{\mu}_x)^{\top} \mathcal{I}_n (\mathbf{x}_{\epsilon} - \boldsymbol{\mu}_x) + b_2 \right).$$

No MBFC1 decorre que a maioria das distribuições condicionais completas *a posteriori* não possuem formas fechadas conhecidas. Em razão disso, para amostrarmos destas distribuições *a posteriori* utilizamos um algoritmo *Gibbs Sampling* com passos de *Metropolis Adaptativo* (Haario *et al.*, 2001, 2005).

Uma grande limitação deste modelo é o requisito de termos informação *a priori* precisa sobre  $\sigma_{\epsilon}^2$ , o que nem sempre é viável. Exploraremos o efeito de fixarmos  $\sigma_{\epsilon}^2$  em valores errados no estudo de simulação apresentado na Seção 2.3.

### 2.1.2 MBFC2: $\sigma_{\epsilon}^2$ é estimado

No segundo MBFC proposto, o qual denotamos por MBFC2, consideramos que a variância do erro  $\sigma_{\epsilon}^2$  é desconhecida e deve ser estimada, ao contrário do que supomos no MBFC1. Este

modelo difere daqueles já discutidos na literatura por estimar tanto as variáveis latentes quanto a variância do erro de medida. Comparado ao MBFC1, este novo modelo traz mais flexibilidade à análise, uma vez que não pressupõe que conheçamos perfeitamente  $\sigma_\epsilon^2$ . No MBFC2 seguimos o padrão estabelecido na literatura para modelar a incerteza sobre a variância, assumindo que *a priori*,  $\sigma_\epsilon^2 \sim GI(a_3, b_3)$ . Consideramos as distribuições *a priori* para  $\beta$ ,  $\alpha$ ,  $\lambda$ ,  $X'_{\epsilon_i} s$ ,  $\mu_x$  e  $\sigma_x^2$  dadas na Expressão (2.10), também utilizadas no MBFC1. Por causa da estrutura de dependência existente entre os parâmetros, necessitamos apenas calcular a distribuição condicional completa de  $\sigma_\epsilon^2$ , a qual é dada por:

$$\pi(\sigma_\epsilon^2 | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}_{(-\sigma_\epsilon^2)}) \propto (\sigma_\epsilon^2)^{-(a_3 + \frac{n}{2} - 1)} \exp \left\{ -\frac{[\frac{1}{2}(\mathbf{w}_\epsilon - \mathbf{x}_\epsilon)^\top \mathcal{I}_n(\mathbf{w}_\epsilon - \mathbf{x}_\epsilon) + b_2]}{\sigma_\epsilon^2} \right\}, \quad (2.18)$$

ou seja,  $\sigma_\epsilon^2 | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}_{(-\sigma_\epsilon^2)} \sim GI\left(a_3 + \frac{n}{2}, \frac{1}{2}(\mathbf{w}_\epsilon - \mathbf{x}_\epsilon)^\top \mathcal{I}_n(\mathbf{w}_\epsilon - \mathbf{x}_\epsilon) + b_2\right)$ . Analogamente ao que observamos no MEFC1, as distribuições condicionais completas de alguns parâmetros desse modelo, não possuem forma fechada, o que nos sugere novamente o uso do algoritmo *Gibbs Sampling* com alguns passos de *Metropolis*.

### 2.1.3 MBFC3: $\sigma_\epsilon^2 = k\sigma_x^2$ é estimado

Na terceira especificação do MBFC assumimos que as variâncias do erro  $\sigma_\epsilon^2$  e da variável aleatória  $\mathbf{X}_\epsilon$  estão relacionadas. Fazemos isto supondo que  $\sigma_\epsilon^2 = k\sigma_x^2$ , em que  $0 < k \leq 1$ . Esta relação, assumindo  $k$  conhecido é indicada na literatura (Fuller, 2009; Rocha *et al.*, 2016) para resolver problemas de falta de identificabilidade em modelos estruturais. Aqui nos inspiramos nesta alternativa como uma nova maneira de modelarmos a incerteza acerca da variância do erro de medida. Como já mencionamos, a falta de identificabilidade no modelo com erro proposto não é resolvido utilizando tal opção, mas, através de uma modelagem sem intercepto. Buscando flexibilizar o modelo, supomos que  $k$  é desconhecido e *a priori* assumimos que  $k \sim \pi(k)$ . Para realizarmos inferência no modelo consideramos a função de verossimilhança descrita pela Equação (2.9), reescrita substituindo-se  $\sigma_\epsilon^2$  por  $k\sigma_x^2$ . Supondo as mesmas distribuições *a priori* para  $\beta$ ,  $\lambda$ ,  $\alpha$ ,  $\mu_x$  e  $\sigma_x^2$  que consideramos nos MBFC1 e MBFC2, dadas em (2.10), temos que as d.c.c desses parâmetros, com exceção de  $\sigma_x^2$ , permanecem inalteradas e são dadas na Subseção 2.1.1. Além de  $\sigma_x^2$ , temos mudanças também na d.c.c de  $X_{\epsilon_i}$ , a qual terá a mesma forma que na Equação (2.14) apenas substituindo  $\sigma_\epsilon^2$  por  $k\sigma_x^2$ , e no lugar de uma distribuição condicional completa para  $\sigma_\epsilon^2$ , temos uma distribuição condicional completa para  $k$ .

Assumindo que *a priori*  $k \sim U(a, b)$ , em que  $0 < a < b$ , a d.c.c para  $k$  é dada por

$$\begin{aligned} \pi(k | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}_{(-\sigma_\epsilon^2)}) &\propto f_{\mathbf{W}_\epsilon}(\mathbf{w}_\epsilon | \mathbf{x}_\epsilon, \mathbf{x}_o, \sigma_x^2, k) \frac{1}{b-a} \mathbb{1}_{(a,b)}(k) \\ &\propto \frac{1}{k^{\frac{n}{2}}(b-a)} \exp \left\{ -\frac{1}{k\sigma_x^2} (\mathbf{w}_\epsilon - \mathbf{x}_\epsilon)^\top \mathcal{I}_n (\mathbf{w}_\epsilon - \mathbf{x}_\epsilon) \right\} \mathbb{1}_{(a,b)}(k). \end{aligned} \quad (2.19)$$

E a d.c.c de  $\sigma_x^2$  no MEFC3 é dada por

$$\begin{aligned} \pi(\sigma_x^2 | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}_{(-\sigma_\epsilon^2)}, k) &\propto f_{\mathbf{W}_\epsilon}(\mathbf{w}_\epsilon | \mathbf{x}_\epsilon, \mathbf{x}_o, \sigma_x^2, k) \pi(\mathbf{x}_\epsilon | \mathbf{D}_{\text{obs}}, \mu_x, \sigma_x^2) \pi(\sigma_x^2 | a_2, b_2) \\ &\propto \left( \frac{1}{k\sigma_x^2} \right)^{\frac{n}{2}} |\mathcal{I}_n| \exp \left\{ -\frac{1}{2k\sigma_x^2} (\mathbf{w}_\epsilon - \mathbf{x}_\epsilon)^\top \mathcal{I}_n (\mathbf{w}_\epsilon - \mathbf{x}_\epsilon) \right\} \\ &\quad \times \left( \frac{1}{\sigma_x^2} \right)^{\frac{n}{2}} |\mathcal{I}_n| \exp \left\{ -\frac{1}{2\sigma_x^2} (\mathbf{x}_\epsilon - \mu_x)^\top \mathcal{I}_n (\mathbf{x}_\epsilon - \mu_x) \right\} \\ &\quad \times \left( \frac{1}{\sigma_x^2} \right)^{a_2+1} \exp \left\{ -\frac{b_2}{\sigma_x^2} \right\} \\ &\propto \left( \frac{1}{\sigma_x^2} \right)^{a_2+n+1} \exp \left\{ -\frac{1}{\sigma_x^2} \left[ \frac{1}{2k} (\mathbf{w}_\epsilon - \mathbf{x}_\epsilon)^\top \mathcal{I}_n (\mathbf{w}_\epsilon - \mathbf{x}_\epsilon) \right. \right. \\ &\quad \left. \left. + \frac{1}{2} (\mathbf{x}_\epsilon - \mu_x)^\top \mathcal{I}_n (\mathbf{x}_\epsilon - \mu_x) + b_2 \right] \right\} \end{aligned} \quad (2.20)$$

ou seja,

$$\sigma_x^2 | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}_{(-\sigma_\epsilon^2)}, k \sim GI \left( a_2 + n, \frac{1}{2k} (\mathbf{w}_\epsilon - \mathbf{x}_\epsilon)^\top \mathcal{I}_n (\mathbf{w}_\epsilon - \mathbf{x}_\epsilon) + \frac{1}{2} (\mathbf{x}_\epsilon - \mu_x)^\top \mathcal{I}_n (\mathbf{x}_\epsilon - \mu_x) + b_2 \right).$$

Tal como nos modelos MBFC1 e 2, as condicionais completas dos parâmetros desse modelo, não possuem forma fechada, por essa razão optamos novamente por utilizar o algoritmo de *Gibbs Sampling* com passos de *Metropolis Adaptativo* (Haario *et al.*, 2001, 2005).

## 2.2 Análise Bayesiana do Modelo de Ma e Yin (2008)

Nesta seção, propomos uma nova abordagem ao modelo em Ma e Yin (2008). Temos o objetivo de compará-lo aos modelos Bayesianos de fração de cura com erro de medida que construímos neste capítulo. A proposta definida por (Ma e Yin, 2008) é semi-paramétrica e utiliza o método dos multiplicadores de Lagrange para obter estimadores de máxima verossimilhança para os parâmetros. Nesse modelo, a variância do erro de medida é assumida conhecida. Tal suposição pode determinar viés para as estimativas. Consideramos uma interpretação paramétrica

sem intercepto deste modelo, em que a função de distribuição basal, presente na definição da função de sobrevivência em (1.16), é conhecida. Além disso, discutiremos uma versão Bayesiana para esta opção, a qual denotaremos por MBFC-Ma. Sob a abordagem Bayesiana, o modelo de Ma e Yin (2008) pode ser visto como estrutural pensando pelo ponto de vista clássico. Nessa perspectiva, estimamos a variância do erro de medida e também as variáveis latentes.

Consideremos o modelo dado na Equação (1.16), cuja a função de sobrevivência é

$$S(t|\mathbf{X}) = \exp\{-\theta(\mathbf{X})F(t)\}. \quad (2.21)$$

Assuma, diferentemente do que é considerado por Ma e Yin (2008), que  $F(t)$  é uma função de distribuição basal conhecida. Tal como no MBFC, defina  $\theta(\mathbf{X}) = \exp\{\mathbf{X}\boldsymbol{\beta}\}$  em que  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$  é o vetor de coeficientes de regressão de ordem  $p \times 1$  e  $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)^\top$  denota a matriz de delineamento de experimento de ordem  $n \times p$ , com  $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})$  correspondendo as covariáveis relacionadas ao indivíduo  $i = 1, \dots, n$ .

Neste modelo, a fração de cura para os indivíduos que experimentaram o evento de interesse, também é definida como sendo  $S(\infty|\mathbf{X}) = \exp\{-\theta(\mathbf{X})\}$  e a função de risco é dada por  $h(t|\mathbf{X}) = \theta(\mathbf{X})f(t)$ , em que  $f(t) = \frac{dF(t)}{dt}$ .

Admita que  $\mathbf{X}$  é medida com erro, de modo que observamos a matriz

$$\mathbf{W} = \mathbf{X} + \mathbf{E},$$

em que  $\mathbf{E} = (\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n)^\top$  é uma matriz  $n \times p$ , tal que  $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{ip})$ , para todo  $i = 1, \dots, n$  e  $\epsilon_{ip}$  é o erro associado a  $p$ -ésima covariável do indivíduo  $i$ . Assumimos que  $\boldsymbol{\epsilon}_{ip} \sim N(0, \sigma_p^2)$  e, dado  $\sigma_1^2, \dots, \sigma_p^2$ ,  $\boldsymbol{\epsilon}_i$  é independente de  $\mathbf{X}_i$ . Assim como no MBFC proposto na Seção 2.1,  $\mathbf{X}_i$ 's são v.a. latentes. Diferentemente do que é considerado em Ma e Yin (2008) aqui assumimos  $\sigma_j^2$  desconhecido, para todo  $j = 1, \dots, p$ .

Seja  $T_i$  e  $C_i$  os tempos de falha e censura, respectivamente, relacionados ao indivíduo  $i = 1, \dots, n$ . Consideremos  $Y_i = \min\{T_i, C_i\}$  e defina  $\delta_i = \mathbb{1}_{\{T_i \leq C_i\}}$  como sendo o indicador de falha para o indivíduo  $i$ . A informação disponível para o indivíduo  $i$  é representada por  $(y_i, \delta_i, \mathbf{X}_i)$ . Assumimos que os mecanismos de falha e censura são independentes, isto é,  $T_i$  é independente de  $C_i$ . Dado  $\mathbf{X}_i$ , as informações para cada indivíduo são independentes. Consequentemente, dado o valor real da covariável  $\mathbf{X}$ , assumimos que a covariável observada  $\mathbf{W}$  não contém qualquer informação adicional sobre  $\mathbf{Y}$ .

Seguindo Ma e Yin (2008), assumo que alguns indivíduos nunca experimentarão o evento de interesse, ou seja,  $y_i = \infty$  para estes indivíduos. Definimos a fração de curados da seguinte maneira: considere  $L$  um limiar, desse modo, diremos que um indivíduo  $i$  foi curado sempre que

$y_i > L$ . A sugestão é considerarmos  $L$  como sendo o maior tempo de falha observado.

Considere a estrutura de erro normal aditivo. Para comportar os casos em que não haja erro na observação de nenhuma das variáveis, Ma e Yin (2008), também supõem que  $m$  variáveis foram observadas com erro, tal que  $0 \leq m \leq p$ . Denotemos por  $\mathbf{X}_o$  e  $\mathbf{X}_E$  as variáveis medidas sem erro e com erro, respectivamente, em que  $\mathbf{X}_E$  é não observada. Desse modo temos  $\mathbf{X} = (\mathbf{X}_o, \mathbf{X}_E)$ , em que  $\mathbf{X}$ ,  $\mathbf{X}_o$  e  $\mathbf{X}_E$ , são, respectivamente, uma matrizes de dimensão  $n \times p$ ,  $n \times (p - m)$  e  $n \times m$ . Dessa maneira,  $\mathbf{W} = (\mathbf{X}_o, \mathbf{W}_E)$ , em que  $\mathbf{W}_E = \mathbf{X}_E + \mathbf{E}$ . Nosso interesse é, portanto, modelar o comportamento conjunto de  $(\mathbf{Y}, \delta, \mathbf{W})$ . Condicional em  $\mathbf{X}$  temos que  $\mathbf{Y}$  e  $\mathbf{W}$  são independentes e admitindo que o mecanismo de censura é independente, temos,

$$\begin{aligned} f(\mathbf{Y}, \delta, \mathbf{W} | \mathbf{X}) &= f(\mathbf{Y}, \delta, \mathbf{W} | \mathbf{X}_o, \mathbf{X}_E) \\ &= f(\mathbf{Y}, \delta | \mathbf{W}, \mathbf{X}_o, \mathbf{X}_E) f_{\mathbf{W}}(\mathbf{W} | \mathbf{X}_o, \mathbf{X}_E) \\ &= f(\mathbf{Y}, \delta | \mathbf{X}_o, \mathbf{X}_E) f_{\mathbf{W}_E}(\mathbf{W}_E | \mathbf{X}_E), \end{aligned} \quad (2.22)$$

tal como no MBFC, definido na seção anterior.

A grande diferença entre o modelo proposto por Ma e Yin (2008) e o MBFC introduzido na Seção 2.1 deste capítulo, está na forma como a fração de cura é definida nas Expressões (1.23) a (1.25) e no fato de o modelo não supor uma estrutura de riscos competitivos. A definição da função de sobrevivência para a nova abordagem do modelo de Ma e Yin (2008) que estamos propondo, é análoga ao que explicitamos nas Expressões (1.19) e (1.21). Assim como consideramos no MBFC, é natural esperarmos que um certo número de indivíduos nunca experimentarão falha ou censura à direita. Para esses indivíduos, temos que “ $Y = \infty$ ” com probabilidade 1. É sob esse raciocínio que a função de sobrevivência para a fração de curados é definida nesse modelo. Continuamos com a mesma proposta de Ma e Yin (2008) dadas nas Expressões (1.23) e (1.24).

Denotemos por  $\boldsymbol{\xi}$  o vetor de parâmetros do modelo e, como na Equação (1.26), assumamos  $\exp\{\mathbf{X}_i \boldsymbol{\beta}\} = \exp\{\mathbf{X}_{o_i} \boldsymbol{\beta}_o + \mathbf{X}_{E_i} \boldsymbol{\beta}_E\}$ . Sob tais supostos, temos que a função de verossimilhança



do modelo é:

$$\begin{aligned}
 L(\xi, \mathbf{X} | \mathbf{Y}, \delta, \mathbf{W}) &= \left[ \prod_{i=1}^n f(y_i, \delta_i | \mathbf{X}_i) \right] f_{\mathbf{W}_E}(\mathbf{W}_E | X_E) \\
 &= \left[ \prod_{i=1}^n f(y_i, \delta_i | \mathbf{X}_{o_i}, \mathbf{X}_{E_i}) \right] f_{\mathbf{W}_\epsilon}(\mathbf{W}_E | X_E) \\
 &= \left[ \prod_{i=1}^n \left[ (f_e(y_i | \mathbf{X}_i) S_c(y_i | \mathbf{X}_i))^{\mathbb{1}\{\delta_i=1\}} (f_c(y_i | \mathbf{X}_i) S_e(y_i | \mathbf{X}_i))^{1-\mathbb{1}\{\delta_i=1\}} \right]^{\mathbb{1}\{y_i < \infty\}} \right. \\
 &\quad \left. \times [S_e(\infty | \mathbf{X}_i) S_c(\infty | \mathbf{X}_i)]^{\mathbb{1}\{y_i = \infty\}} \right] f_{\mathbf{W}_E}(\mathbf{W}_E | X_E) \\
 &= \left[ \prod_{i=1}^n \left[ (\exp\{-H(y_i) \exp\{\mathbf{X}_i \boldsymbol{\beta}\}\} h(y_i) \exp\{\mathbf{X}_i \boldsymbol{\beta}\} \right. \right. \\
 &\quad \times S_c(y_i | \mathbf{X}_i))^{\mathbb{1}\{\delta_i=1\}} (f_c(y_i | \mathbf{X}_i) \exp\{-H(y_i) \exp\{\mathbf{X}_i \boldsymbol{\beta}\}\})^{1-\mathbb{1}\{\delta_i=1\}} \left. \right]^{\mathbb{1}\{y_i < \infty\}} \\
 &\quad \left. \times (\exp\{-\exp\{\mathbf{X}_i \boldsymbol{\beta}\}\} S_c(\infty | \mathbf{X}_i))^{\mathbb{1}\{y_i = \infty\}} \right] f_{\mathbf{W}_E}(\mathbf{W}_E | X_E). \tag{2.23}
 \end{aligned}$$

Suponha ainda que a censura é não informativa. Nesse caso a expressão dada na Equação (2.23) é simplificada e assume a seguinte forma:

$$\begin{aligned}
 L(\xi, \mathbf{X} | \mathbf{Y}, \delta, \mathbf{W}) &\propto \left[ \prod_{i=1}^n \left[ (h(y_i) \exp\{\mathbf{X}_i \boldsymbol{\beta}\})^{\mathbb{1}\{\delta_i=1\}} \exp\{-H(y_i) \exp\{\mathbf{X}_i \boldsymbol{\beta}\}\} \right]^{\mathbb{1}\{y_i < \infty\}} \right. \\
 &\quad \left. (\exp\{-\exp\{\mathbf{X}_i \boldsymbol{\beta}\}\})^{\mathbb{1}\{y_i = \infty\}} \right] f_{\mathbf{W}_E}(\mathbf{W}_E | X_E). \tag{2.24}
 \end{aligned}$$

Para os cálculos a seguir, suponha, sem perda de generalidade que apenas uma covariável foi observada com erro. Denotemos tal covariável por  $\mathbf{W}_\epsilon = \mathbf{X}_\epsilon + \boldsymbol{\epsilon}$ , a qual é um vetor  $n \times 1$ . Consideremos,  $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)$ , em que  $\epsilon_i \stackrel{i.i.d}{\sim} N(0, \sigma_\epsilon^2)$ . Desse modo, dado  $\mathbf{X}_\epsilon$  e  $\sigma_\epsilon^2$ , temos que  $\mathbf{W}_\epsilon | \mathbf{X}_\epsilon \sigma_\epsilon^2 \sim N_n(\mathbf{X}_\epsilon, \sigma_\epsilon^2 \mathcal{I}_n)$ , em que  $\mathcal{I}_n$  denota a matriz identidade de ordem  $n$ . Como no estudo de simulação apresentado em Ma e Yin (2008), admita que  $Y \sim \text{Exp}(\lambda)$ . Sob tais supostos, o vetor de parâmetros do modelo é  $\boldsymbol{\xi}^* = (\boldsymbol{\beta}_o, \boldsymbol{\beta}_\epsilon, \lambda, \sigma_\epsilon^2, \mathbf{X}_\epsilon)$  e a função de verossimilhança torna-se

$$\begin{aligned}
 L(\boldsymbol{\xi}^* | \mathbf{Y}, \delta, \mathbf{W}) &\propto \left[ \prod_{i=1}^n \left[ (\lambda \exp\{\theta(\mathbf{X}_i)\})^{\mathbb{1}\{\delta_i=1\}} \exp\{-\lambda y_i \theta(\mathbf{X}_i)\} \right]^{\mathbb{1}\{y_i < \infty\}} \right. \\
 &\quad \left. (\exp\{-\exp\{\theta(\mathbf{X}_i)\}\})^{\mathbb{1}\{y_i = \infty\}} \right] |\sigma_\epsilon^2 \mathcal{I}_n|^{-\frac{1}{2}} \exp\{-A\},
 \end{aligned}$$

em que  $\boldsymbol{\theta}(\mathbf{X}_i) = \exp\{\mathbf{X}_i\boldsymbol{\beta}\}$  e  $A = \frac{1}{2}(\mathbf{w}_\epsilon - \mathbf{x}_\epsilon)^\top (\sigma_\epsilon^2 \mathcal{I}_n)^{-1} (\mathbf{w}_\epsilon - \mathbf{x}_\epsilon)$ . Assuma que  $\sigma_\epsilon^2, \boldsymbol{\beta}, \lambda$  e  $\mathbf{X}_\epsilon$  são independentes. Consequentemente, as d.c.c dos parâmetros, são dadas por:

i) Seja  $\pi(\lambda)$  a função densidade de probabilidade *a priori* de  $\lambda$ , então,

$$\pi(\lambda | \xi_{(-\lambda)}^*, \mathbf{Y}, \boldsymbol{\delta}, \mathbf{W}) \propto \left[ \prod_{i=1}^n \left[ \lambda^{\mathbb{1}\{\delta_i=1\}} \exp\{-\lambda y_i \exp\{\theta(\mathbf{X}_i)\}\} \right]^{\mathbb{1}\{y_i < \infty\}} \right. \\ \left. \left[ \exp\{-\theta(\mathbf{X}_i)\}\} \right]^{\mathbb{1}\{y_i = \infty\}} \right] \pi(\lambda).$$

ii) Seja  $\pi(\beta_j); j = 1, \dots, p$  a função densidade de probabilidade *a priori* de  $\beta_j$ , então,

$$\pi(\beta_j | \xi_{(-\beta_j)}^*, \mathbf{Y}, \boldsymbol{\delta}, \mathbf{W}) \propto \left[ \prod_{i=1}^n \left[ (\exp\{X_{ij}\beta_j\})^{\mathbb{1}\{\delta_i=1\}} \exp\{-\lambda y_i \theta(\mathbf{X}_i)\} \right]^{\mathbb{1}\{y_i < \infty\}} \right. \\ \left. \left[ \exp\{-\theta(\mathbf{X}_i)\}\} \right]^{\mathbb{1}\{y_i = \infty\}} \right] \pi(\beta_j).$$

iii) Se a distribuição *a priori* de  $\mathbf{X}_\epsilon$  é  $\pi(\mathbf{X}_\epsilon)$ , então,

$$\pi(\mathbf{X}_\epsilon | \xi_{(-\mathbf{X}_\epsilon)}^*, \mathbf{Y}, \boldsymbol{\delta}, \mathbf{W}) \propto \pi(\mathbf{X}_\epsilon) \left[ \prod_{i=1}^n \left[ (\exp\{\mathbf{X}_{i\epsilon}\boldsymbol{\beta}_\epsilon\})^{\mathbb{1}\{\delta_i=1\}} \exp\{-\lambda y_i \theta(\mathbf{X}_i)\} \right]^{\mathbb{1}\{y_i < \infty\}} \right. \\ \left. \left[ \exp\{-\theta(\mathbf{X}_i)\}\} \right]^{\mathbb{1}\{y_i = \infty\}} \right] \exp \left\{ -\frac{1}{2} (\mathbf{w}_\epsilon - \mathbf{x}_\epsilon)^\top (\sigma_\epsilon^2 \mathcal{I}_n)^{-1} (\mathbf{w}_\epsilon - \mathbf{x}_\epsilon) \right\}.$$

iv) Se  $\sigma_\epsilon^2 \sim \pi(\sigma_\epsilon^2)$ , então,

$$\pi(\sigma_\epsilon^2 | \beta, \lambda, \mathbf{X}_\epsilon, \mathbf{Y}, \boldsymbol{\delta}, \mathbf{W}) \propto |\sigma_\epsilon^2 \mathcal{I}_n|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{w}_\epsilon - \mathbf{x}_\epsilon)^\top (\sigma_\epsilon^2 \mathcal{I}_n)^{-1} (\mathbf{w}_\epsilon - \mathbf{x}_\epsilon) \right\} \pi(\sigma_\epsilon^2).$$

As d.c.c de  $\mu_x, \sigma_x^2$  e  $\sigma_\epsilon^2$  são idênticas as obtidas para o MBFC nas Equações (2.16), (2.17) e (2.18), respectivamente, quando supomos as mesmas distribuições *a priori* da Expressão (2.10).

Tal como no caso dos modelos anteriormente definidos, a versão reformulada do modelo de Ma e Yin (2008) possui uma função de verossimilhança com forma bastante complexa, o que resulta em distribuições condicionais completas para os parâmetros do modelo, sem forma fechada conhecida. Novamente, será necessário utilizar métodos MCMC para obtermos amostras das distribuições *a posteriori* dos parâmetros.

## 2.3 Estudo de Simulação de Monte Carlo para Avaliar o MBFC

Nesta seção, apresentamos um estudo de simulação com a finalidade de validarmos o modelo proposto e de analisarmos o efeito, na inferência *a posteriori*, das diferentes distribuições *a priori* assumidas para a variância do erro de medida e variáveis latentes, visto que a estimação destes elementos são a principal contribuição deste capítulo. Também analisamos o impacto que ignorar o erro de medida no modelo provoca nas estimativas dos parâmetros e, por consequência, à fração de cura. É importante ressaltarmos que, pelo que sabemos, esse é o primeiro modelo de fração de cura com erro de medida nas covariáveis em que se leva em consideração a inferência sobre variância do erro de medida. Apenas o modelo de regressão sem intercepto para os tempos de falha foi ajustado. Esta escolha foi motivada pela dificuldade surgida na estimação de alguns parâmetros na sua presença. Na análise de todos os bancos de dados ajustamos os modelos MBFC1 a MBFC3. Para fins de comparação o modelo de tempo de promoção sem erro nas covariáveis também será ajustado.

Para os estudos de simulação apresentados neste capítulo, consideramos quatro cenários distintos para avaliar o desempenho dos modelos propostos. Nos Cenários 1 e 2, analisamos a capacidade de ajuste do modelo, quando a variável latente é normalmente distribuída. Além disso, analisamos o efeito de ajustarmos um modelo ingênuo quando existe erro de medida nos dados. No contexto deste trabalho, chamamos de modelo ingênuo àqueles que não supõem erro de medida nas covariáveis. Nesse texto, o modelo ingênuo considerado é o MTP proposto por Chen *et al.* (1999) e definido na Seção 1.3.1. No Cenário 3, avaliamos o que acontece quando supomos normalidade para a variável latente, quando na verdade esta vem de outra distribuição. No Cenário 4, a proposta é estudarmos o que acontece quando ajustamos dados que não possuem erro de medida, através de um modelo que carrega em sua estrutura a suposição de que alguma covariável foi observada com erro. Em todos os estudos, simulamos 250 amostras com 100 indivíduos em cada uma delas e em todas as análises dos dados assumiremos que a variável latente  $\mathbf{X}_2 \sim N(\mu_x, \sigma_x^2)$ . Os cenários que serão analisados são descritos a seguir.

**Cenário 1:** No primeiro estudo, que denotamos Cenário 1, os dados foram gerados de um modelo de fração de cura com erro nas covariáveis obedecendo as especificações a seguir: os tempos de falha  $\mathbf{Y} = (y_1, \dots, y_n)$ , são gerados da distribuição Weibull( $\alpha, \lambda$ ), em que  $\alpha = 1$  e  $\lambda = -1.2$  (na Equação(2.4.1) podemos ver a parametrização considerada); consideramos 30% de censura. A matriz de delineamento de experimento é dada por  $\mathbf{W} = (\mathbf{X}_1, \mathbf{X}_2 + \boldsymbol{\epsilon})$ , em que  $X_1 \sim \text{Bernoulli}(0.4)$ ,  $X_2 \sim \text{Normal}(\mu_x, \sigma_x^2)$  com  $\mu_x = 3.9$  e  $\sigma_x^2 = 10$ . Temos também  $\epsilon_i \stackrel{i.i.d}{\sim} N(0, \sigma_\epsilon^2)$  com  $\sigma_\epsilon^2 = 1$ . Os coeficientes de regressão são  $\beta_1 = 1.8$  e  $\beta_2 = -0.15$ .

O *script* do programa utilizado para gerar os bancos de dados deste cenário, encontra-se disponível no Apêndice D deste trabalho.

**Cenário 2:** Os dados são gerados usando a mesma estrutura do Cenário 1, agora assumindo  $\sigma_\epsilon^2 = 5$ .

**Cenário 3:** Neste cenário, supomos que as variáveis latentes do modelo foram geradas de uma distribuição Gama( $\psi_1, \psi_2$ ). O objetivo de definir um cenário em que essas variáveis não são normalmente distribuídas é verificar, como a suposição de normalidade para as variáveis latentes afetará as estimativas dos parâmetros do modelo.

Como nos casos anteriores, nos bancos de dados gerados para este cenário, assumimos que os tempos de falha  $\mathbf{Y} = (y_1, \dots, y_n)$  foram gerados da distribuição Weibull( $\alpha, \lambda$ ), em que  $\alpha = 1$  e  $\lambda = -1.2$ . Consideramos 30% de censura e que os coeficientes de regressão são  $\boldsymbol{\beta} = (1.8, -0.15)^\top$ . Diferentemente do que foi assumido nos outros cenários temos na matriz de delineamento de experimento  $\mathbf{W} = (\mathbf{X}_1, \mathbf{X}_2 + \boldsymbol{\epsilon})$  que  $X_1 \sim \text{Bernoulli}(0.4)$  e  $X_2$  tem distribuição Gama( $\psi_1, \psi_2$ ) com média  $\frac{\psi_1}{\psi_2} = \mu_x = 3.9$  e variância  $\frac{\psi_1}{\psi_2^2} = \sigma_x^2 = 10$ . Para cada indivíduo  $i$  temos  $\epsilon_i \stackrel{i.i.d}{\sim} N(0, \sigma_\epsilon^2)$  com  $\sigma_\epsilon^2 = 1$ .

**Cenário 4:** Neste cenário, consideramos que os dados foram completamente observados, isto é, não há nenhuma covariável medida com erro. Este cenário foi pensado para avaliar o que acontece quando supomos erro de medida, porém, este erro não existe. Nesse cenário, geramos os dados assumindo que os tempos de falha  $\mathbf{Y} = (y_1, \dots, y_n)$  foram gerados da distribuição Weibull( $\alpha, \lambda$ ), supondo  $\alpha = 1$ ,  $\lambda = -1.2$  e considerando 30% de censura; a matriz de delineamento de experimento é dada desta vez por  $\mathbf{W} = \mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ , em que, novamente  $X_1 \sim \text{Bernoulli}(0.4)$ ,  $X_2 \sim N(3.9, 10)$ ; e  $\boldsymbol{\beta} = (1.8, -0.15)^\top$  é o vetor de coeficientes de regressão.

Realizamos algumas simulações partindo de valores iniciais aleatórios no MCMC, verificando a convergência dos algoritmos de cada modelo proposto. Ao fazermos isto, observamos que a autocorrelação após a convergência é muito pequena. No Apêndice H temos alguns gráficos de diagnóstico de convergência obtidos. Sabendo que os algoritmos convergem, para reduzir o tempo computacional ao realizarmos os estudos de Monte Carlo com 250 réplicas, não utilizamos *lag* e *burn-in* no MCMC, considerando como ponto inicial da cadeia os valores reais dos parâmetros. Em cada réplica, fizemos um total de 2000 iterações, selecionando todas estas como amostra da distribuição *a posteriori* de cada parâmetro, a fim de calcularmos as estimativas *a posteriori* para cada um deles. Um dos programas implementados para obtermos as distribuições *a posteriori* dos parâmetros, executados no software livre R (R Core Team, 2016) tem o código disponibilizado no Apêndice F deste trabalho.

### 2.3.1 Estudo 1: O Efeito de Supor a Variância do Erro Conhecida

Neste estudo ajustamos o modelo MBFC1, o qual assume  $\sigma_\epsilon^2$  fixo e conhecido, para analisarmos dados do cenário 1. O objetivo é avaliarmos o impacto de errarmos a especificação de  $\sigma_\epsilon^2$

no MBFC1, nas estimativas *a posteriori*, e fazermos uma análise de sensibilidade considerando diferentes escolhas para a distribuição *a priori* de  $\sigma_x^2$ .

Comparamos três diferentes ajustes do modelo, em um deles fixamos  $\sigma_\epsilon^2$  em seu valor real que é 1, e nos outros dois supomos este valor fixo em 0.5 e 10. Três escolhas de distribuições *a priori* para  $\sigma_x^2$  são feitas. Uma pouco informativa, em que  $\sigma_x^2 \sim GI(0.1, 0.1)$ . Outra muito informativa em que centramos a distribuição no verdadeiro de  $\sigma_x^2 = 10$  e assumimos uma variância pequena, eliciando  $\sigma_x^2 \sim GI(102, 1010)$ . Por último, escolhemos uma distribuição que traz informação razoavelmente equivocada sobre  $\sigma_x^2$ , a qual centramos em um valor consideravelmente distante do real, considerando  $\sigma_x^2 \sim GI(100, 100)$ . Na Figura 2.1, podemos visualizar as f.d.p. destas três distribuições e percebemos que a probabilidade de observar valores em torno de dez é alta quando  $\sigma_x^2$  tem distribuição  $GI(102, 1010)$ . Além disso, quando a distribuição é  $GI(100, 100)$ , a probabilidade de observar valores longe de 1 tende a zero. Note que a curva da distribuição  $GI(0.1, 0.1)$  tem um formato bastante suavizado, indicando que a massa probabilística de intervalos  $(a, b)$  tal que  $1 \leq a < b$ , de mesmo comprimento é aproximadamente igual.

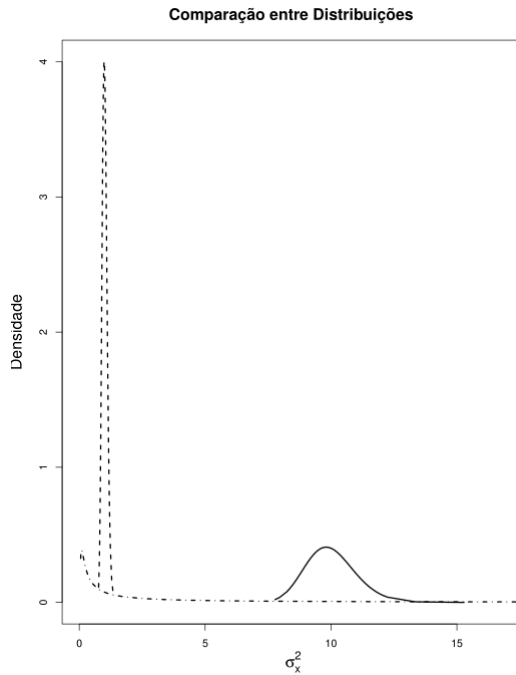


Figura 2.1: Distribuições *a priori* de  $\sigma_x^2$ ,  $GI(102, 1010)$  (linha sólida),  $GI(100, 100)$  (linha tracejada) e  $GI(0.1, 0.1)$  (linha pontilhada).

Também consideramos que, *a priori*, os coeficientes de regressão  $\beta_1$  e  $\beta_2$  tem distribuição

imprópria  $\pi(\beta_1, \beta_2) \propto 1$  e que os parâmetros da distribuição dos tempos de falha são tais que  $\lambda \sim \text{Normal}(0, 100)$  e  $\alpha \sim \text{Gama}(0.5, 0.001)$ . Para a variável latente  $\mathbf{X}_2$ , assumimos que  $\mu_x \sim \text{Normal}(5, 10)$ .

### Efeito nas Estimativas dos Parâmetros

Na Tabela 2.2, apresentamos a média, viés e erro quadrático médio (EQM) das estimativas dos parâmetros fornecidas pela média *a posteriori*, para todas as especificações *a priori* de  $\sigma_x^2$  e  $\sigma_\epsilon^2$ . Tais estimativas foram calculadas fazendo uma média de Monte Carlo de todas as médias, vieses e EQM's resultantes de cada réplica. Do ponto de vista Bayesiano, a avaliação do viés médio e EQM nos permite analisar se a inferência de diferentes especialistas que possuem a mesma informação *a priori* pode ser afetada por diferentes informações amostrais.

Tabela 2.2: Média, Viés e EQM das estimativas *a posteriori* obtidas ajustando-se o MBFC1, considerando  $\sigma_\epsilon^2 = 1, 0.5$  e  $10$ , e diferentes distribuições *a priori* para  $\sigma_x^2$ .

$\pi(\sigma_x^2)$	$\eta$	Real	$\sigma_\epsilon^2 = 1$			Modelo			$\sigma_\epsilon^2 = 10$		
			Média	Viés	EQM	$\sigma_\epsilon^2 = 0.5$	Média	Viés	EQM	Média	Viés
$GI(102, 1010)$	$\beta_1$	1.8	1.820	0.020	0.352	1.805	0.005	0.322	2.656	0.856	3.779
$GI(0.1, 0.1)$			1.817	0.017	0.346	1.810	0.010	0.343	2.061	0.261	1.189
$GI(100, 100)$			1.836	0.036	0.334	1.816	0.016	0.341	1.841	0.041	0.431
$GI(102, 1010)$	$\beta_2$	-0.15	-0.145	0.005	0.007	-0.135	0.015	0.007	-0.436	-0.286	0.311
$GI(0.1, 0.1)$			-0.144	0.006	0.008	-0.135	0.015	0.007	-0.127	0.023	0.227
$GI(100, 100)$			-0.172	-0.022	0.011	-0.146	0.004	0.008	-0.069	0.081	0.126
$GI(102, 1010)$	$\alpha$	1	1.078	0.078	0.052	1.070	0.070	0.050	1.344	0.344	0.414
$GI(0.1, 0.1)$			1.078	0.078	0.052	1.072	0.072	0.052	1.154	0.154	0.120
$GI(100, 100)$			1.089	0.089	0.055	1.075	0.075	0.051	1.101	0.101	0.059
$GI(102, 1010)$	$\lambda$	-1.2	-1.253	-0.053	0.411	-1.281	-0.081	0.386	-0.944	0.256	0.605
$GI(0.1, 0.1)$			-1.254	-0.054	0.417	-1.285	-0.085	0.419	-1.533	-0.333	2.565
$GI(100, 100)$			-1.134	0.066	0.433	-1.239	-0.039	0.409	-1.528	-0.328	2.214
$GI(102, 1010)$	$\mu_x$	3.9	3.921	0.021	0.010	3.920	0.020	0.010	3.876	-0.024	0.016
$GI(0.1, 0.1)$			3.920	0.020	0.010	3.921	0.021	0.011	3.873	-0.027	0.018
$GI(100, 100)$			3.912	0.012	0.010	3.913	0.013	0.010	3.892	-0.008	0.013
$GI(102, 1010)$	$\sigma_x^2$	10	10.028	0.028	0.032	10.190	0.191	0.073	9.122	-0.878	0.778
$GI(0.1, 0.1)$			10.297	0.297	0.480	10.820	0.823	1.071	1.858	-8.142	67.216
$GI(100, 100)$			3.005	-6.995	48.965	3.694	-6.306	39.806	1.036	-8.964	80.361

Ao analisarmos os resultados na Tabela 2.2, percebemos que fixar a variância do erro em um valor muito discrepante do real, causa um enorme impacto negativo na estimação da variância das variáveis latentes. Isso está evidenciado no viés e EQM das estimativas, que aumentam significativamente quando fixamos  $\sigma_\epsilon^2 = 10$ , que é um valor muito maior do que o valor real  $\sigma_\epsilon^2 = 1$ . Este estudo mostra que supor que a variância do erro é conhecida é um ponto bastante inconveniente e restritivo na prática. Se o valor fixado estiver muito longe do verdadeiro, o ajuste fornecido tende a não ser bom. Este efeito também pode ser observado nas

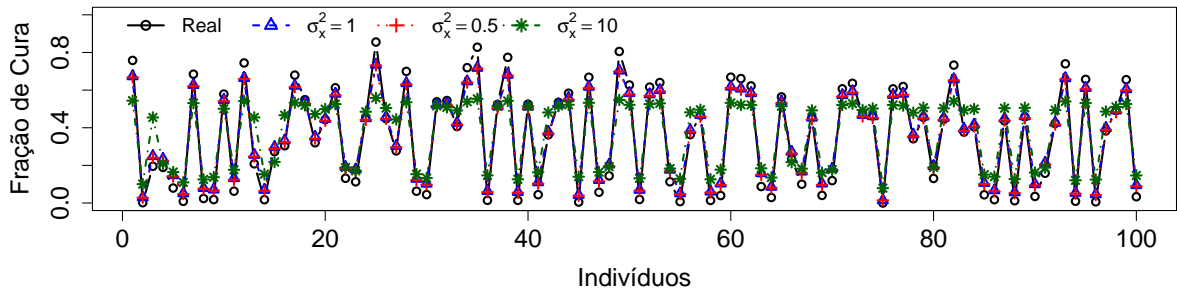
Figuras 2.2 e 2.3, em que exibimos médias dos valores esperados *a posteriori* da fração de cura (ver também Figuras G.1 à G.6 no Apêndice G deste texto). Em todos os gráficos fica evidente que o pior ajuste obtido foi através do MBFC1 no qual fixamos  $\sigma_x^2 = 10$ . Quando fixamos  $\sigma_\epsilon^2$  em um valor um pouco menor que o valor real ( $\sigma_\epsilon^2 = 0.5$ ), obtivemos estimativas comparáveis com aquelas obtidas fixando  $\sigma_\epsilon^2 = 1$ , com valor médio, viés e EQM muito próximos.

Quanto à escolha da distribuição *a priori* para  $\sigma_x^2$ , notamos, analisando a Tabela 2.2, assim como nos gráficos exibidos nesta seção, o quão importante é incluir alguma informação confiável que tenhamos *a priori* a respeito da variância das variáveis latentes. Embora as estimativas para os demais parâmetros sejam pouco influenciadas pela especificação *a priori* de  $\sigma_x^2$ , estimativas menos enviesadas são obtidas quando optamos por uma distribuição *a priori* muito informativa, centrada no valor real. O viés nas estimativas diminui consideravelmente, se comparado ao viés observado para as estimativas obtidas quando eliciamos uma distribuição para  $\sigma_x^2$  centrada muito abaixo do valor real. Na prática é razoável esperar que consigamos alguma informação a respeito das variáveis latentes com um especialista como, por exemplo, sobre a média da distribuição de  $\sigma_x^2$ . A opinião de especialistas acerca da média da distribuição *a priori*, ou qualquer outra medida de resumo é importante para que determinemos uma distribuição bem informativa. Caso não tenhamos essa opção, outra alternativa é especificar os hiperparâmetros de  $\sigma_x^2$  com base na média da amostra observada no estudo. Isto é, a distribuição de  $\sigma_x^2$  pode ser centrada na média (ou em um valor próximo a esta) da variância dos valores observados da variável medida com erro na amostra analisada. A incerteza sobre tal informação pode ser controlada de acordo com a variância escolhida para tal distribuição. Nesse estudo, por exemplo, a média das variâncias amostrais é 8.89, enquanto que o valor real desta variância é 10. Resultados observados em estudos anteriores também podem ser considerados com informação *a priori*.

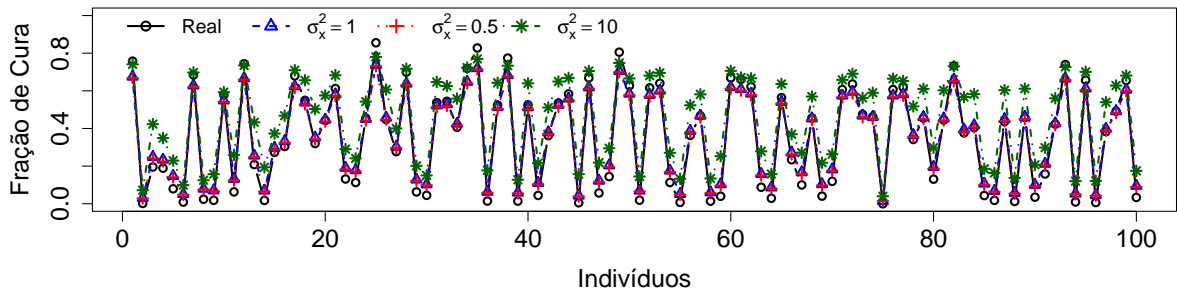
### Efeito nas estimativas da fração de cura

Na Figura 2.2, apresentamos as médias dos valores esperados *a posteriori* para a fração de cura obtidas em cada réplica de Monte Carlo, para para cada indivíduo  $i$ , quando ajustamos o MBFC1, nos casos em que fixamos  $\sigma_\epsilon^2$  em 1, 0.5 e 10. Além disso, nas Figuras 2.2a, 2.2b e 2.2c, podemos comparar os resultados que são obtidos quando as distribuições *a priori* de  $\sigma_x^2$  são  $GI(0.1, 01)$ ,  $GI(102, 1010)$  e  $GI(100, 100)$ , respectivamente. Nessa figura, obtemos o viés das estimativas da fração de cura. Podemos observar nos gráficos que a fração de cura é subestimada para alguns indivíduos, entretanto, o viés nas estimativas está perto de zero.

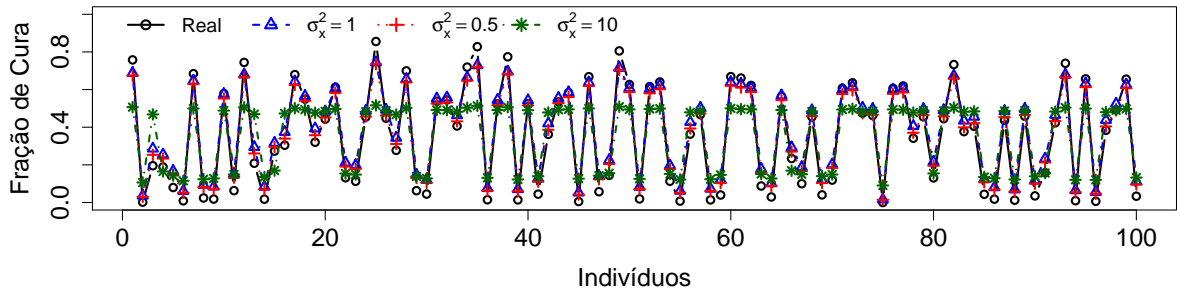
Na Figura 2.3, apresentamos um histograma dos valores que simulamos para a fração de cura (fração de cura real) e sobre ele estão as curvas da distribuição empírica das estimativas da fração de cura de todos os 100 indivíduos, obtidas a partir da média dos valores esperados



(a)



(b)



(c)

Figura 2.2: Médias do valor esperado *a posteriori* para a fração de cura dos indivíduos do Cenário 1, ajustando-se o MBFC1 nos casos em que fixamos  $\sigma_\epsilon^2 = 1$  (linha com  $\Delta$ ),  $\sigma_\epsilon^2 = 0.5$  (linha com  $+$ ) e  $\sigma_\epsilon^2 = 10$  (linha com  $*$ ) e supondo  $\sigma_x^2 \sim GI(0.1, 0.1)$  (a),  $\sigma_x^2 \sim GI(102, 1010)$  (b) e  $\sigma_x^2 \sim GI(100, 100)$  (c). A linha com  $\circ$  é referente aos valores reais de fração de cura para cada indivíduo.



*a posteriori*, ajustando-se os dados através de todos os modelos estudados nessa seção. Com o objetivo de entender melhor o comportamento das estimativas médias do valor esperado *a posteriori* para a fração de cura de cada indivíduo, no Apêndice G exibimos *box-plots* de tais estimativas para os 100 indivíduos nas 250 réplicas de Monte Carlo, nas Figuras G.1 a G.6. Na Tabela 2.3, apresentamos o percentual de intervalos HPD 95% que contém o valor real para a fração de cura de cada indivíduo, obtidos a partir das 250 réplicas de cada indivíduo no estudo de Monte Carlo.

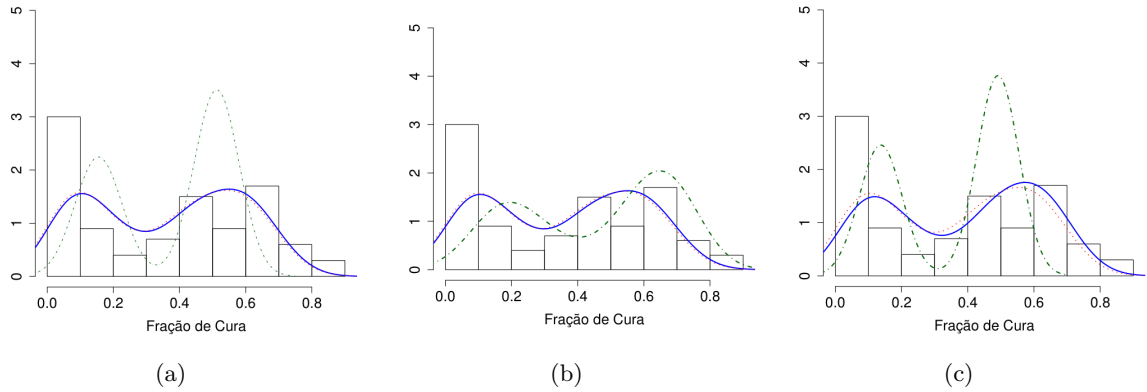


Figura 2.3: Histograma das frações de cura reais superposto pelas distribuições empíricas das médias dos valores esperados *a posteriori* da fração de cura, obtidas ajustando-se o MBFC1, supondo  $\sigma_x^2 \sim GI(0.1, 0.1)$  (Gráfico a),  $\sigma_x^2 \sim GI(102, 1010)$  (Gráfico b) e  $\sigma_x^2 \sim GI(100, 100)$  (Gráfico c), fixando  $\sigma_\epsilon^2 = 1$  (linha sólida), 0.1 (linha pontilhada), 10 (linha traço-ponto).

A Figura 2.3 mostra que a distribuição empírica da média das estimativas *a posteriori* da fração de cura se assemelha à distribuição da fração de cura real, exceto para valores até 10% e quando fixamos  $\sigma_x^2 = 10$ .

Tabela 2.3: Percentual médio de cobertura dos intervalos HPD's com 95% de credibilidade obtidos para a fração de cura, ajustando-se o MBFC1 para todas as especificações *a priori* para  $\sigma_\epsilon^2$  e  $\sigma_x^2$ .

Modelo	MBFC1		
	$\sigma_\epsilon^2 = 1$	$\sigma_\epsilon^2 = 0.5$	$\sigma_\epsilon^2 = 10$
$GI(0.1, 0.1)$	91.88%	89.99%	88.56%
$GI(102, 1010)$	92.13%	89.93%	95.18%
$GI(100, 100)$	91.90%	91.89%	86.05%

Analisando os resultados expostos na Tabela 2.3, vemos que o percentual dos intervalos com 95% de credibilidade que contém o valor verdadeiro da fração de cura de cada indivíduo é bastante alto (acima de 86% em todos os casos). O modelo que apresenta um percentual maior de cobertura da fração de cura real e, em geral, fornece melhores estimativas para todos os

parâmetros, é o modelo no qual fixamos  $\sigma_\epsilon^2 = 1$  (valor real), o que já é esperado. Vale salientar que quando fixamos  $\sigma_\epsilon^2 = 0.5$ , que é um valor bem perto do real, os resultados são muito similares aqueles obtidos quando assumimos o valor real no modelo. Dessa maneira percebemos que, ao optarmos por ajustar o MBFC1 no qual temos a suposição de  $\sigma_\epsilon^2$  conhecido, é importante admitirmos um valor bem perto do real para tal parâmetro, caso contrário poderemos obter estimativas bastante imprecisas para os parâmetros do modelo.

### 2.3.2 Estudo 2: O Efeito de Ignorar o Erro de Medida Presente nos Dados

Nesta seção, temos como objetivo avaliar o efeito nas estimativas dos parâmetros e da fração de cura ao utilizamos um modelo sem erro nas covariáveis quando alguma delas tenha sido observada com erro. Para isso, os dados dos Cenários 1 e 2 serão analisados usando o modelo de tempo de promoção (MTP), definido na Seção 1.3.1 e os modelos MBFC1 (considerando  $\sigma_\epsilon^2 = 1, 0.5, 10$ , no Cenário 1, e  $\sigma_\epsilon^2 = 1, 0.5, 5$ , no cenário 2), MBFC2 e MBFC3. Médias, viés e EQM das médias *a posteriori* de todos os parâmetros são dados nas Tabelas 2.4 e 2.5, a seguir.

As distribuições *a priori* utilizadas para os parâmetros que são comuns a todos as especificações do MBFC,  $\beta, \alpha, \lambda$  e  $\mu_x$ , são as mesmas escolhidas no Estudo 1. No MBFC2, escolhemos a distribuição  $GI(1, 1)$  para  $\sigma_\epsilon^2$ , e no MBFC3, assumimos que  $k \sim \text{Gama}(0.1, 1)$  ao analisarmos os dados do Cenário 1 e que  $k \sim U(0, 1)$  considerando o Cenário 2. As distribuições atribuídas *a priori* para  $\sigma_x^2$ , são as mesmas consideradas no Estudo 1.

#### Efeito nas Estimativas dos Parâmetros

Nas Tabelas 2.4 e 2.5, apresentamos as médias, viés e EQM das estimativas de todos os parâmetros dos modelos ajustados. As estimativas são fornecidas pela média *a posteriori*.

Os resultados dados na Tabela 2.4, indicam que quando o erro de medida possui uma variância consideravelmente pequena, o modelo ingênuo não fornece estimativas muito viesadas para os parâmetros  $\beta_1, \alpha$  e  $\lambda$ . Por outro lado, na Tabela 2.5, vemos que quando  $\sigma_\epsilon^2$  é um pouco maior temos um aumento no viés. Os parâmetros que são comuns aos modelos MTP e MBFC, com exceção do coeficiente de regressão  $\beta_2$  da covariável medida com erro, não diferem substancialmente, indicando que não são afetados por perturbações naqueles das distribuições da variável latente e do erro de medida,  $\mu_x, \sigma_x^2$  e  $\sigma_\epsilon^2$ . Isto parece acontecer devido a configuração da função de verossimilhança dos modelos com erro de medida, a qual foi construída supondo que, condicional nas variáveis latentes, os tempos de falha são independentes das observações, resultando em modelos com erro não diferencial (Bolfarine e Arellano-Valle, 1998). Nestes modelos a função de verossimilhança pode ser escrita em duas partes, sendo a parte (I) referente apenas a função de sobrevivência comum ao MTP e aos MBFC's e a (II) é a contribuição do erro de medida para a verossimilhança; ver Equação (2.6). Percebemos que os modelos MBFC2 e 3

Tabela 2.4: Média do valor esperado *a posteriori*, Viés e Erro Quadrático Médio (EQM) das estimativas obtidas ajustando-se os modelos MTP, MBFC2, MBFC3 e o MBFC1, no qual consideramos  $\sigma_\epsilon^2 = 1, 0.5, 10$ . O valor real de  $\sigma_\epsilon^2$  é 1.

$\pi(\sigma_x^2)$		Modelo										
		MTP			MBFC2			MBFC3				
		$\eta$	Real	Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM
$GI(102, 1010)$			1.798	-0.002	0.365	1.835	0.035	0.381	1.826	0.026	0.369	
$GI(0.1, 0.1)$	$\beta_1$	1.8	1.798	-0.002	0.365	1.905	0.105	0.796	1.820	0.020	0.326	
$GI(100, 100)$			1.798	-0.002	0.365	1.859	0.059	0.490	1.833	0.033	0.359	
$GI(102, 1010)$			-0.128	0.022	0.006	-0.152	-0.002	0.013	-0.144	0.006	0.007	
$GI(0.1, 0.1)$	$\beta_2$	-0.15	-0.128	0.022	0.006	-0.178	-0.028	0.047	-0.144	0.006	0.008	
$GI(100, 100)$			-0.128	0.022	0.006	-0.081	0.069	0.128	-0.161	-0.011	0.009	
$GI(102, 1010)$			1.067	0.067	0.048	1.085	0.085	0.055	1.077	0.077	0.053	
$GI(0.1, 0.1)$	$\alpha$	1	1.067	0.067	0.048	1.110	0.110	0.083	1.079	0.079	0.054	
$GI(100, 100)$			1.067	0.067	0.048	1.103	0.103	0.059	1.083	0.083	0.053	
$GI(102, 1010)$			-1.308	-0.108	0.436	-1.237	-0.037	0.381	-1.259	-0.059	0.439	
$GI(0.1, 0.1)$	$\lambda$	-1.2	-1.308	-0.108	0.436	-1.190	0.010	0.528	-1.255	-0.055	0.400	
$GI(100, 100)$			-1.308	-0.108	0.436	-1.499	-0.299	2.250	-1.184	0.016	0.428	
$GI(102, 1010)$			-	-	-	3.921	0.021	0.010	3.920	0.020	0.010	
$GI(0.1, 0.1)$	$\mu_x$	3.9	-	-	-	3.915	0.015	0.011	3.921	0.021	0.010	
$GI(100, 100)$			-	-	-	3.889	-0.011	0.016	3.913	0.013	0.010	
$GI(102, 1010)$			-	-	-	9.990	-0.010	0.024	10.140	0.141	0.059	
$GI(0.1, 0.1)$	$\sigma_x^2$	10	-	-	-	9.251	-0.749	3.340	10.780	0.778	1.044	
$GI(100, 100)$			-	-	-	1.050	-8.950	80.100	3.719	-6.281	39.496	
$GI(102, 1010)$			-	-	-	1.241	0.241	0.211	1.011	0.011	0.000	
$GI(0.1, 0.1)$	$\sigma_\epsilon^2$	1	-	-	-	2.081	1.081	3.647	1.011	0.011	0.000	
$GI(100, 100)$			-	-	-	9.911	8.911	79.807	0.891	-0.109	0.012	
		MBFC1										
		$\sigma_\epsilon^2$	1			0.5			10			
		$\eta$	Real	Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM
$GI(102, 1010)$			1.816	0.016	0.343	1.810	0.010	0.334	2.398	0.598	1.964	
$GI(0.1, 0.1)$	$\beta_1$	1.8	1.817	0.017	0.323	1.807	0.007	0.309	1.981	0.181	1.045	
$GI(100, 100)$			1.832	0.032	0.332	1.814	0.014	0.337	1.833	0.033	0.450	
$GI(102, 1010)$			-0.143	0.007	0.007	-0.134	0.016	0.006	-0.364	-0.214	0.167	
$GI(0.1, 0.1)$	$\beta_2$	-0.15	-0.142	0.008	0.008	-0.134	0.016	0.007	-0.096	0.054	0.190	
$GI(100, 100)$			-0.171	-0.021	0.011	-0.145	0.005	0.007	-0.097	0.053	0.093	
$GI(102, 1010)$			1.056	0.056	0.043	1.050	0.050	0.041	1.238	0.238	0.190	
$GI(0.1, 0.1)$	$\alpha$	1	1.056	0.056	0.044	1.051	0.051	0.043	1.107	0.107	0.085	
$GI(100, 100)$			1.067	0.067	0.045	1.054	0.054	0.043	1.077	0.077	0.045	
$GI(102, 1010)$			-1.272	-0.072	0.394	-1.307	-0.107	0.402	-0.968	0.231	0.621	
$GI(0.1, 0.1)$	$\lambda$	-1.2	-1.274	-0.074	0.388	-1.301	-0.101	0.371	-1.582	-0.382	2.184	
$GI(100, 100)$			-1.150	0.050	0.421	-1.256	-0.056	0.402	-1.419	-0.219	1.628	
$GI(102, 1010)$			3.920	0.020	0.010	3.920	0.020	0.010	3.888	-0.012	0.013	
$GI(0.1, 0.1)$	$\mu_x$	3.9	3.921	0.021	0.010	3.921	0.021	0.010	3.878	-0.022	0.016	
$GI(100, 100)$			3.912	0.012	0.010	3.913	0.013	0.010	3.892	-0.008	0.013	
$GI(102, 1010)$			10.029	0.029	0.032	10.190	0.190	0.072	9.114	-0.886	0.792	
$GI(0.1, 0.1)$	$\sigma_x^2$	10	10.297	0.297	0.479	10.820	0.824	1.072	1.764	-8.236	68.567	
$GI(100, 100)$			3.005	-6.995	48.961	3.694	-6.306	39.808	1.036	-8.964	80.359	

Tabela 2.5: Média do valor esperado *a posteriori*, Viés e Erro Quadrático Médio (EQM) das estimativas obtidas ajustando-se os modelos MTP, MBFC2, MBFC3 e o MBFC1, no qual consideramos  $\sigma_\epsilon^2 = 1, 0.5, 5$ . O valor real de  $\sigma_\epsilon^2$  é 5.

$\eta$	Modelo									
	MTP				MBFC2			MBFC3		
	Real	Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM
$\beta_1$	1.8	1.759	-0.041	0.330	1.826	0.140	0.782	2.126	0.326	1.391
$\beta_2$	-0.15	-0.094	0.056	0.007	-0.175	-0.025	0.050	-0.259	-0.109	0.090
$\alpha$	1	1.049	0.049	0.045	1.123	0.123	0.106	1.192	0.192	0.173
$\lambda$	-1.2	-1.433	-0.233	0.412	-1.269	-0.069	0.414	-1.080	0.120	0.502
$\mu_x$	3.9	-	-	-	3.925	0.025	0.048	3.914	0.014	0.050
$\sigma_x^2$	10	-	-	-	10.293	0.293	0.135	7.780	-2.220	5.594
$\sigma_\epsilon^2$	5	-	-	-	4.063	-0.937	0.978	7.588	2.588	7.333
MBFC1										
$\sigma_\epsilon^2$	1				0.5			5		
$\eta$	Real	Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM
$\beta_1$	1.8	1.772	-0.028	0.339	1.768	-0.032	0.341	1.882	0.082	0.451
$\beta_2$	-0.15	-0.105	0.044	0.007	-0.101	0.049	0.007	-0.165	-0.015	0.017
$\alpha$	1	1.056	0.056	0.046	1.053	0.053	0.048	1.104	0.104	0.063
$\lambda$	-1.2	-1.398	-0.198	0.423	-1.417	-0.217	0.429	-1.239	-0.039	0.467
$\mu_x$	3.9	3.927	0.027	0.048	3.927	0.027	0.049	3.928	0.028	0.048
$\sigma_x^2$	10	11.161	1.161	1.548	11.394	1.394	2.164	9.965	-0.035	0.073

forneem estimativas menos viesadas para tais parâmetros. Além disso, nas Figuras 2.4 e 2.5, vemos como as estimativas obtidas para as variáveis latentes ajustando-se os modelos MBFC1, 2 e 3, são próximas do valor real destas variáveis. Em tais figuras, temos diagramas de dispersão nos quais os valores médios estimados para a variável latente  $X_\epsilon$  são relacionados com o seu valor real. Nos dois cenários, as estimativas utilizadas são as médias dos valores esperados *a posteriori* da covariável  $X_\epsilon$  de cada indivíduo. Tais estimativas foram obtidas ajustando-se os modelos MBFC1, 2 e 3. Nestas figuras, também comparamos a covariável observada  $\mathbf{W}_2 = \mathbf{X}_\epsilon + \epsilon$  com os valores reais  $\mathbf{X}_\epsilon$ , uma vez que no MTP supõe-se que essa variável não possui erro de medida e queremos analisar como os valores observados com erro afetam na estimação dos parâmetros e da fração de cura.

Destacamos que os resultados apresentados na Tabela 2.5 na qual consideramos o Cenário 2, apontam que quando o erro de medida tem uma maior variabilidade,  $\sigma_\epsilon^2 = 5$ , o MTP forneceu estimativas mais viesadas para os parâmetros do que aquelas obtidas sob o Cenário 1, no qual  $\sigma_\epsilon^2 = 1$ . Por outro lado, as três versões do MBFC se comportaram de forma análoga nos dois cenários.

### Efeito nas Estimativas Fração de Cura

Nas Figuras 2.6 a 2.8 podemos visualizar as estimativas obtidas para a fração de cura ao ajustamos os quatro modelos. Além disso, *box-plots* para tais estimativas podem ser vistos

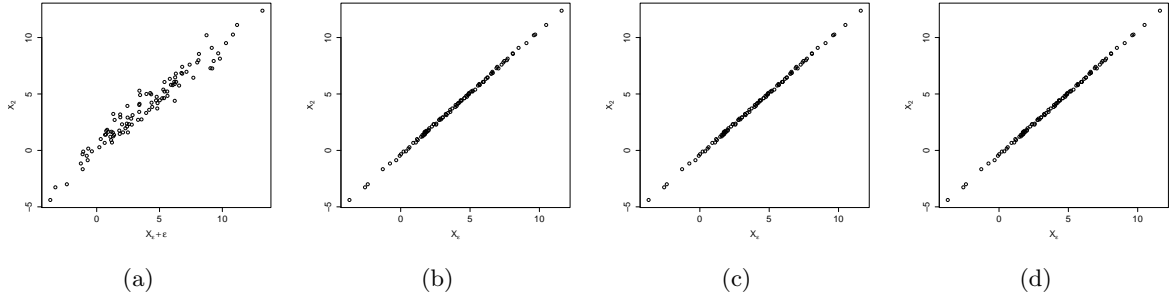


Figura 2.4:  $X_\epsilon$  (estimado)  $\times X_2$ . Nos gráficos (b), (c) e (d) temos as estimativas obtidas ajustando-se o MBFC1 quando  $\sigma_\epsilon^2 = 1$ , MBFC2 e MBFC3, respectivamente. No gráfico (a), correspondente ao MTP, temos os valores observados  $(X_\epsilon + \epsilon) \times X_2$

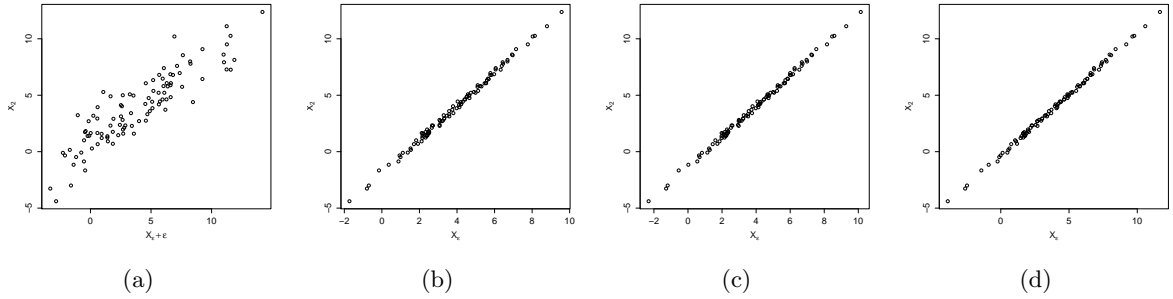


Figura 2.5:  $X_\epsilon$  (estimado)  $\times X_2$ . Nos gráficos (b), (c) e (d) temos as estimativas obtidas ajustando-se o MBFC1 quando  $\sigma_\epsilon^2 = 5$ , MBFC2 e MBFC3, respectivamente. No gráfico (a), correspondente ao MTP, temos os valores observados  $(X_\epsilon + \epsilon) \times X_2$

no Apêndice G (Figuras G.7 a G.10). É importante também incluímos nesta análise o que foi observado a respeito da estimação das variáveis latentes nas Figuras 2.4 e 2.5. Esse tipo de consideração é importante, pois, as estimativas para a fração de cura dependem, além dos parâmetros do modelo, das observações. Nesse caso, queremos avaliar se o viés nas estimativas da fração de cura é reduzido quando fornecemos boas estimativas para as variáveis latentes.

Através da análise dos gráficos nas Figuras 2.4 e 2.5, nos quais comparamos as estimativas obtidas para a variável latente com seu valor real, e além disso, as variáveis observadas também são comparadas aos valores reais, podemos observar o quão boas parecem ser as estimativas médias das variáveis latentes. Essa parece ser a razão de, apesar das estimativas médias da fração de cura terem o comportamento semelhante em todos os modelos nas Figuras 2.6 a 2.8, encontrarmos bastante diferença entre as estimativas intervalares obtidas sob cada ajuste, ao analisarmos os percentuais de HPDs que contém o verdadeiro valor da fração de cura. os quais são dados na Tabela 2.6 a seguir.

Tabela 2.6: Percentual médio de cobertura dos Intervalos HPD's com 95% de credibilidade obtidos para a fração de cura, ajustando-se o MTP e o MBFC sob todas as especificações de  $\sigma_\epsilon^2$  e  $\sigma_x^2 \sim GI(102, 1010)$ .

Cenário	Modelo			
	MTP	MBFC1	MBFC2	MBFC3
1	81.98%	92.14%	92.82%	92.34%
2	71.74%	92.00%	90.80%	93.08%

Na Tabela 2.6, podemos ver que o MTP é o modelo cujo ajuste fornece o menor número intervalos de credibilidade contendo o valor verdadeiro da fração de cura, apenas 81,9% no Cenário 1 e 71,7% no Cenário 2, enquanto que entre as versões do MBFC esse percentual varia de 90% à 93%. O modelo que apresenta uma melhor cobertura para o verdadeiro valor do parâmetro é o MBFC2 no caso em que  $\sigma_\epsilon^2 = 1$  e o MBFC3 quando  $\sigma_\epsilon^2 = 5$ .

Nos gráficos dados nas Figuras 2.6 a 2.8, nos quais comparamos as estimativas médias obtidas para a fração de cura de cada indivíduo, podemos analisar as estimativas obtidas para a fração de cura pontualmente, sob as condições dos dois cenários analisados. Na Figura 2.6 temos o histograma da fração de cura real e sobre ele estão as curvas das densidades empíricas das médias de Monte Carlo do valor esperado *a posteriori* de todos os 100 indivíduos, calculadas a partir das 250 réplicas.

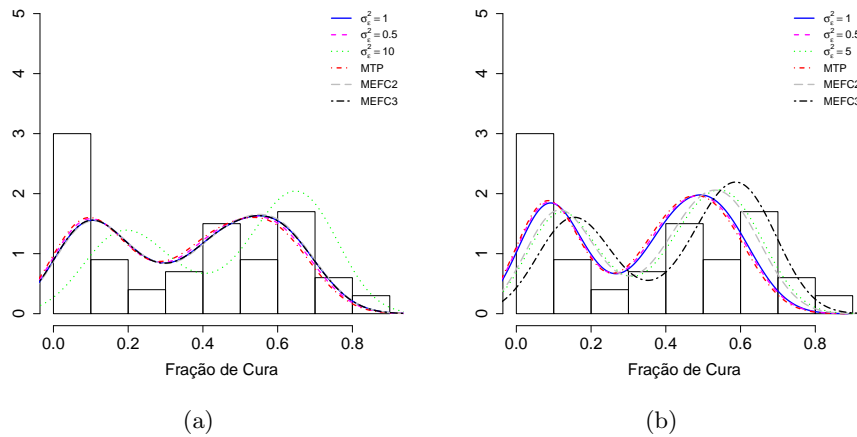
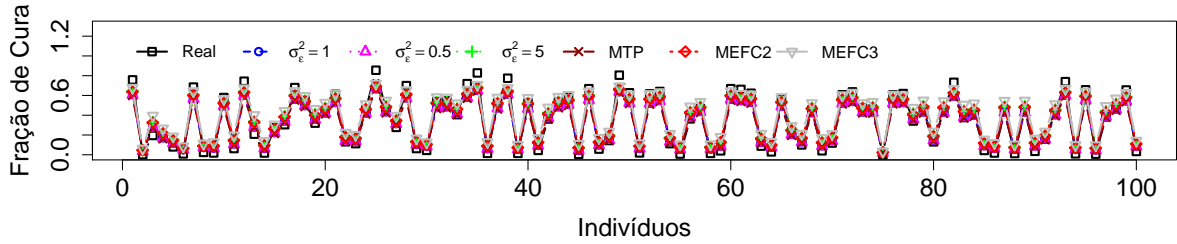


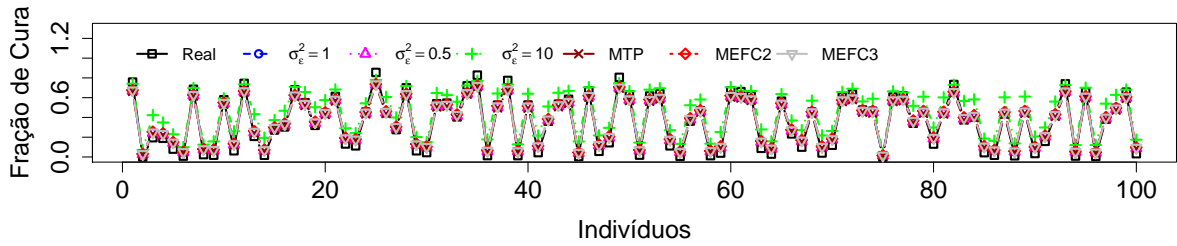
Figura 2.6: Histograma das frações de cura reais superpostos pelas distribuições empíricas das médias dos valores esperados *a posteriori* da fração de cura, ajustando-se todos os modelos, considerando o Cenário 1 (Gráfico a) e o Cenário 2 (Gráfico b).

Analisando as Figuras 2.7 e 2.8, percebemos que o viés das estimativas médias para o valor esperado *a posteriori* da fração de cura possuem comportamento similar em ambos os modelos,

de modo que quando comparamos estas estimativas graficamente não somos capazes de fazer distinção entre as curvas, o que acontece nos dois cenários analisados. Entretanto, na Figura 2.6, é possível enxergar um pouco melhor a diferença entre as estimativas obtidas, comparando-se o histograma da fração de cura real com as curvas de densidade estimadas para o valor esperado *a posteriori* da fração de cura sob cada um dos modelos.



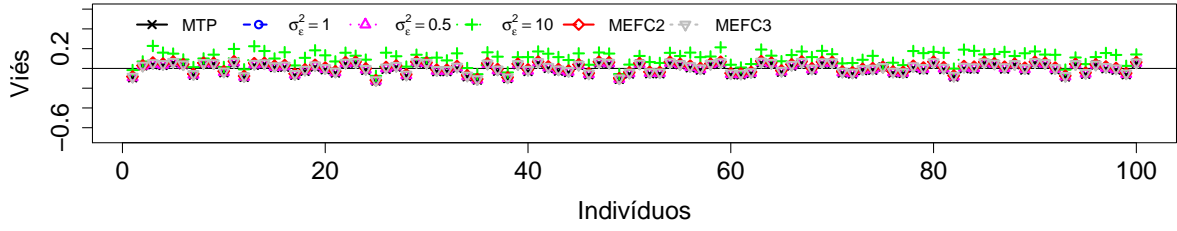
(a)



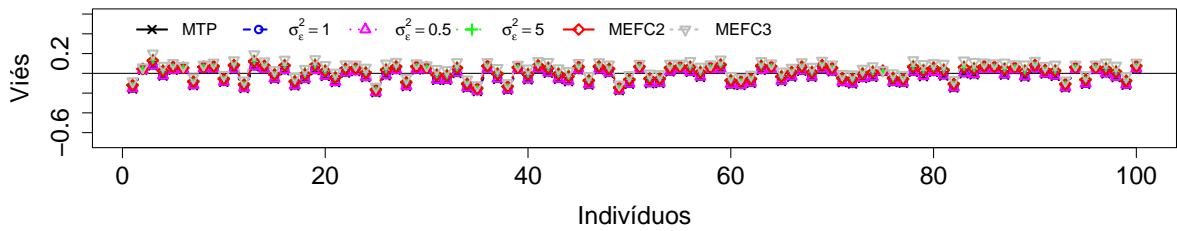
(b)

Figura 2.7: Estimativas médias do valor esperado *a posteriori* para fração de cura dos indivíduos dos Cenário 1 e 2 Fig.a e Fig.b, respectivamente, ajustando-se os modelos: MTP (linha com  $\times$ ), MBFC1 ( $\sigma_\epsilon^2 = 1$  – linha com  $\circ$ ), MBFC2 (linha com  $\diamond$ ) e MBFC3 (linha com  $\nabla$ ).  $\square$  representa os valores reais da fração de cura.

Na Figuras 2.9 e 2.10, exibimos histogramas do erro de medida em 9 das 250 réplicas do estudo de Monte Carlo que realizamos, considerando os Cenários 1 e 2, respectivamente. Sob cada histograma podemos observar as curvas da densidade empírica de amostras de tamanho 100 para o erro de medida, obtidas utilizando as estimativas médias para o valor esperado *a posteriori* de  $\sigma_\epsilon^2$ , ajustando-se o MBFC2 e o MBFC3. Para avaliar a escolha que fizemos para  $\sigma_\epsilon^2$  no MBFC1, no qual supomos esse valor conhecido, também consideramos as curvas das densidades empíricas do erro de medida obtido quando fixamos  $\sigma_\epsilon^2 = 1, 0.5, 5$ , em que 5 é o valor real de  $\sigma_\epsilon^2$ . Em resumo, geramos amostras da distribuição  $N(0, \hat{\sigma}_\epsilon^2)$ , em que  $\hat{\sigma}_\epsilon^2$  é a estimativa média fornecida para  $\sigma_\epsilon^2$  pelo ajuste de cada modelo.



(a)



(b)

Figura 2.8: Viés do valor esperado *a posteriori* para a fração de cura dos indivíduos dos Cenários 1 e 2, Fig.a e Fig.b, respectivamente, ajustando-se os modelos: MTP (linha com  $\times$ ), MBFC1 ( $\sigma_\epsilon^2 = 1$  – linha com  $\circ$ ,  $\sigma_\epsilon^2 = 0.5$  – linha com  $\triangle$ ,  $\sigma_\epsilon^2 = 5$  – linha com  $+$ ), MBFC2 (linha com  $\diamond$ ) e MBFC3 (linha com  $\nabla$ ).  $\square$  representa os valores reais da fração de cura.

Observando as Figuras 2.9 e 2.10, podemos analisar a capacidade que cada modelo tem em captar a forma como o erro de medida está distribuído. Em geral, notamos que o MBFC1, com  $\sigma_\epsilon^2$  fixado em seu valor real, parece fornecer a melhor aproximação para a distribuição verdadeira do erro de medida, o que é natural de se esperar, visto que esse é o modelo real dos dados. Por outro lado, quando fixamos esse valor em um valor longe do real, o MBFC1 nos fornece a distribuição mais distante da real. Entre os modelos MBFC2 e MBFC3, que supõem  $\sigma_\epsilon^2$  desconhecido, o MBFC2, em geral, fornece a melhor aproximação para a verdadeira distribuição do erro de medida.

Analisando os dois cenários estudados nessa seção, com base nos resultados expostos nas tabelas e gráficos apresentados, podemos concluir que quanto maior a variabilidade do erro de medida, mais importante se torna considerarmos um modelo supondo erro de medida nas covariáveis, pois, o MTP forneceu estimativas mais viesadas, quando temos um erro com maior variância. Enquanto que entre os MBFC, os modelos que forneceram um melhor ajuste foram o MBFC2 quando  $\sigma_\epsilon^2 = 1$  e o MBFC3 quando  $\sigma_\epsilon^2 = 5$ .



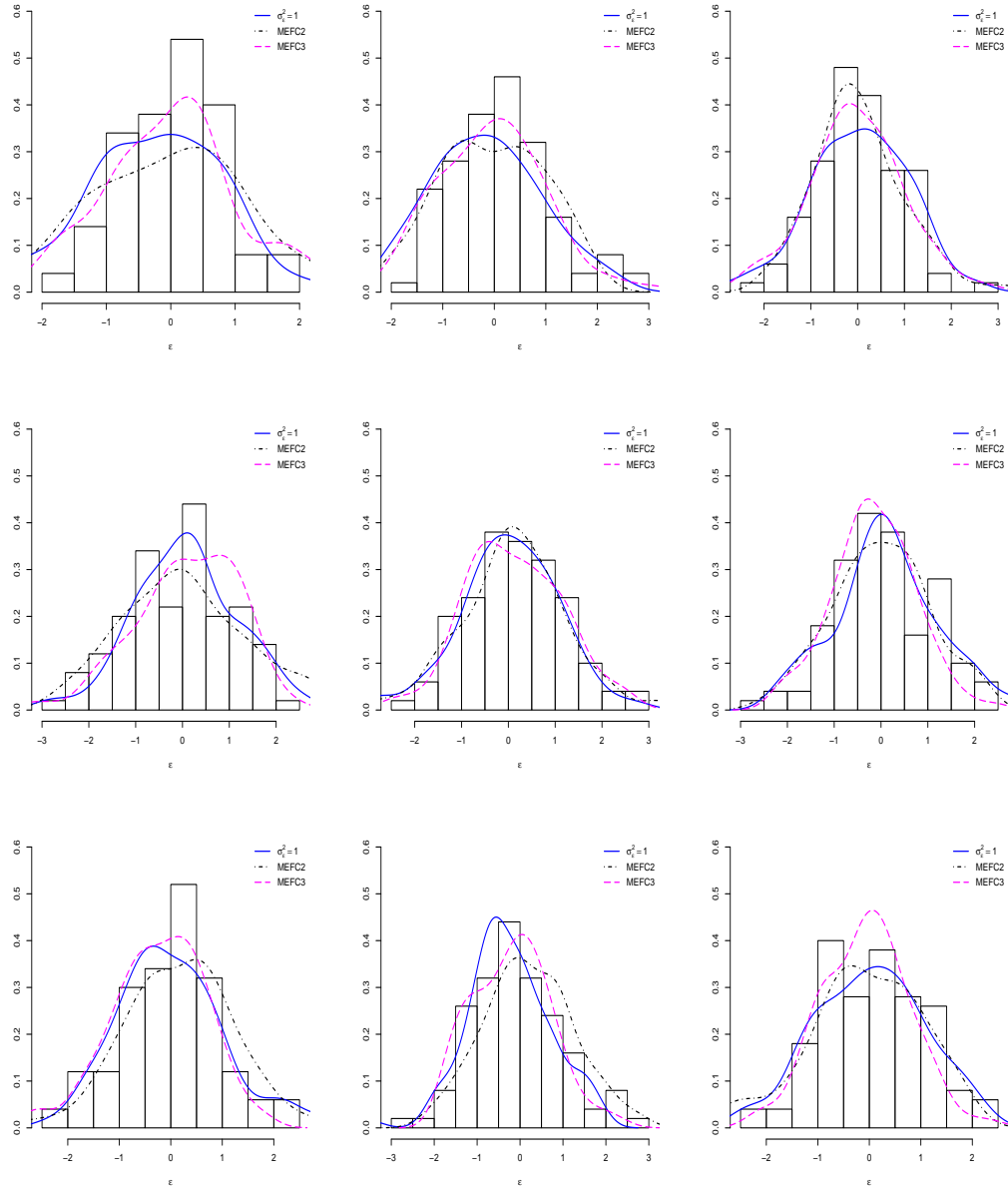


Figura 2.9: Histograma do erro de medida em 9 das 250 réplicas de Monte Carlo. As curvas sobre cada histograma representam as densidades de distribuições Normais centradas em zero e cuja a variância foi fixada em 1, no caso do MBFC1 e as densidades empíricas de amostras do erro de medida estimadas aos ajustarmos o MBFC2 e o MBFC3, considerando o Cenário 1.

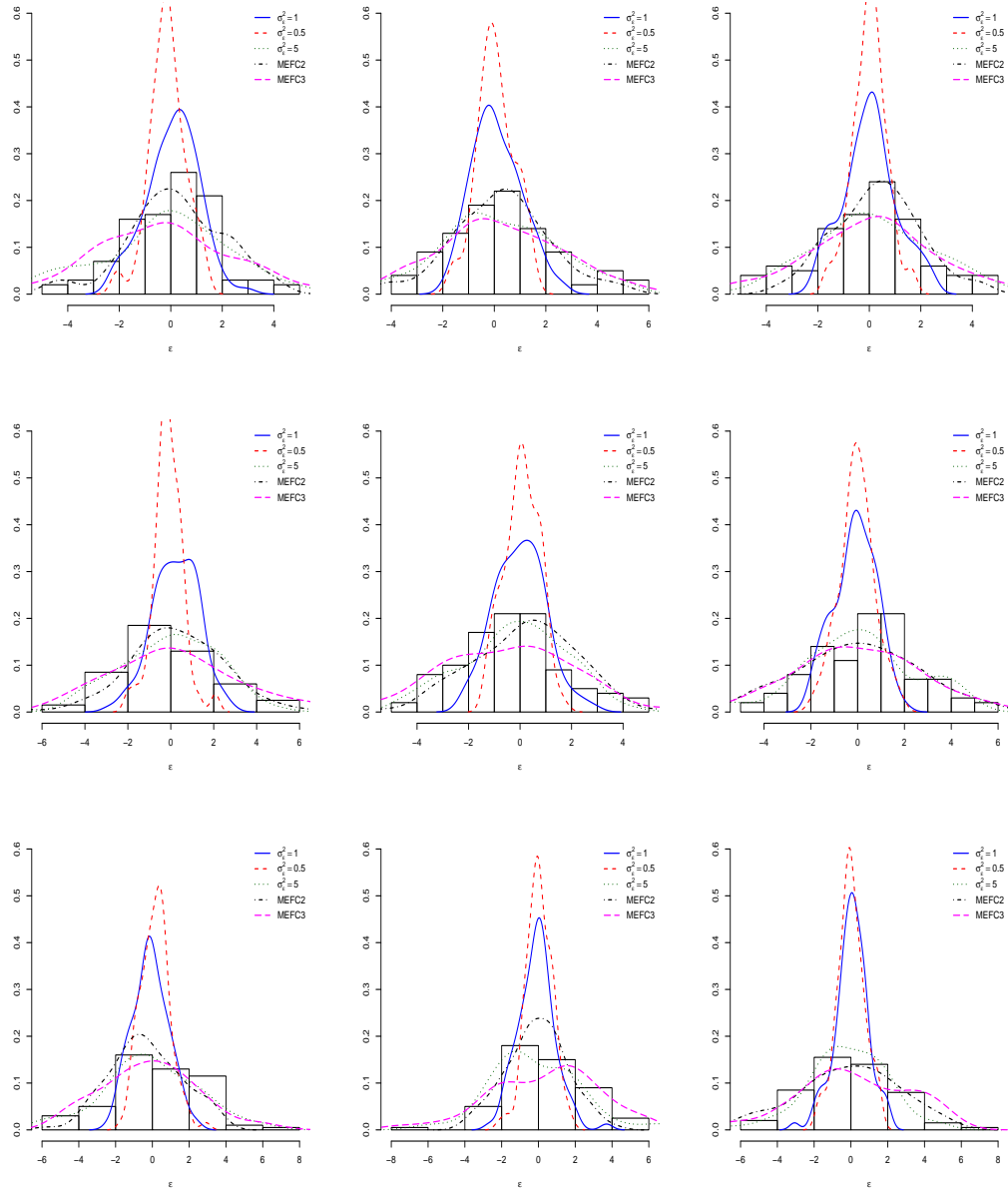


Figura 2.10: Histograma do erro de medida em 9 das 250 réplicas de Monte Carlo. As curvas sobre cada histograma representam as densidades de distribuições Normais centradas em zero e cuja a variância foi fixada em 1, 0.5 e 5, no caso do MBFC1 e as densidades empíricas de amostras do erro de medida estimadas aos ajustarmos o MBFC2 e o MBFC3, considerando o Cenário 2.

### 2.3.3 Estudo 3: O Efeito de Pressupor Erro de Medida Quando Não Existe

Nesta seção, apresentamos um estudo de simulação no qual consideramos o Cenário 4. Nesse cenário, geramos os dados sem acrescentar erro de medida em qualquer covariável. O objetivo desse estudo é avaliar o ajuste do MBFC para dados com essa natureza. Esperamos que os modelos propostos se ajustem bem a dados que não apresentam erro de medida, pois, na especificação do modelo o erro é definido com distribuição centrada em zero, e portanto, o que esperamos é que esses modelos estimem a variância do erro de medida bem próxima a zero.

Na Tabela 2.7, apresentamos as estimativas médias para os valores esperados *a posteriori*, viés e EQM, obtidas em 250 réplicas de Monte Carlo, ajustando-se os dados através dos modelos MTP, MBFC1, 2 e 3. Neste estudo supomos *a priori*  $\sigma_x^2 \sim GI(102, 1010)$ . Para os parâmetros  $\beta_1, \beta_2, \alpha, \lambda, \mu_x, \sigma_\epsilon^2$  e  $k$ , as distribuições *a priori* são as mesmas utilizadas nos Estudos 1 e 2.

Tabela 2.7: Média do valor esperado *a posteriori*, Viés e Erro Quadrático Médio (EQM) das estimativas obtidas ajustando-se os modelos MTP, MBFC1, no qual consideramos  $\sigma_\epsilon^2 = 1$ , MBFC2 e MBFC3.

$\eta$	Real	Modelo											
		MTP		MBFC1		MBFC2		MBFC3					
		Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM
$\beta_1$	1.8	1.798	-0.002	0.306	1.830	0.030	0.314	1.821	0.021	0.308	1.817	0.017	0.298
$\beta_2$	-0.15	-0.128	0.022	0.006	-0.144	0.006	0.007	-0.144	0.006	0.007	-0.145	0.005	0.007
$\alpha$	1	1.065	0.065	0.044	1.076	0.076	0.047	1.049	0.079	0.049	1.078	0.078	0.047
$\lambda$	-1.2	-1.316	-0.116	0.358	-1.272	-0.072	0.368	-1.262	-0.062	0.365	-1.251	-0.051	0.357
$\mu_x$	3.9	-	-	-	3.915	0.015	0.000	3.915	0.015	0.000	3.915	0.015	0.000
$\sigma_x^2$	10	-	-	-	9.748	-0.252	0.064	9.801	-0.199	0.043	9.828	-0.172	0.031
$\sigma_\epsilon^2$	-	-	-	-	-	-	-	0.869	-	-	1.029	-	-

#### Efeito nas Estimativas dos Parâmetros

Na Tabela 2.7 podemos observar que, em geral, todas as versões do MBFC fornecem boas estimativas para todos os parâmetros. Pontualmente, quando comparamos os quatro modelos percebemos que  $\beta_1$  e  $\alpha$  são melhor estimados pelo MTP. As estimativas médias fornecidas por este modelo apresentam menor viés e EQM do que aquelas obtidas através das três versões do MBFC. Dentre as três especificações que propomos para o MBFC, a que fornece uma estimativa média menos enviesada para  $\beta_1$  é o MBFC3 e para  $\alpha$  é o MBFC2. Os parâmetros  $\beta_2$  e  $\lambda$  foram melhor ajustados pelo MBFC3 (menor viés entre os modelos). O MTP é o modelo que apresenta maior viés e EQM para as estimativas de tais parâmetros. Para a média e variância da variável  $\mathbf{X}_2$  ( $\mu_x$  e  $\sigma_x^2$ ), suposta medida com erro pelos modelos MBFC1, 2 e 3, o modelo MBFC3 forneceu estimativas menos enviesadas. Em relação a variância do erro de medida, entre o MBFC2 e o MBFC3, que são os modelos que estimam tal parâmetro, o MBFC2 forneceu uma estimativa média mais próxima de zero. Como os dados foram gerados sem erro de medida, consideramos

que o valor verdadeiro da variância do erro é zero, logo, podemos considerar que o MBFC2 produziu uma estimativa menos enviesada para tal parâmetro.

Nas três especificações do MBFC estimamos a variável  $\mathbf{X}_2$ , suposta medida com erros nesses modelos. Na Figura 2.11 apresentamos um diagrama de dispersão relacionando os valores médios estimados,  $\mathbf{X}_\epsilon$ , para a variável  $\mathbf{X}_2$  com os valores dessa variável.

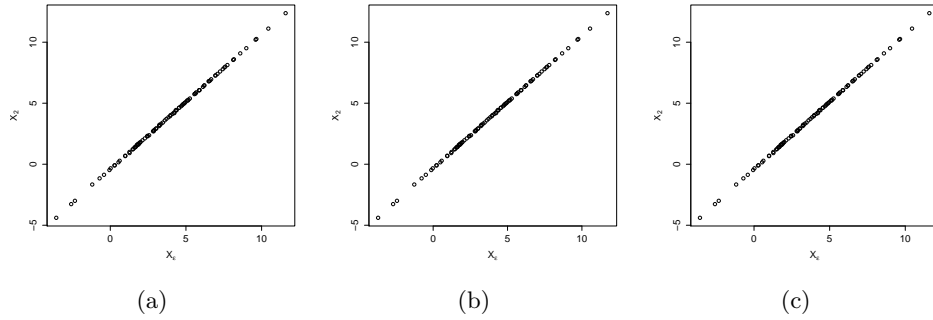


Figura 2.11:  $\mathbf{X}_\epsilon$  no eixo das abscissas representa as estimativas obtidas para  $\mathbf{X}_2$  ajustando-se o MBFC1 quando  $\sigma_\epsilon^2 = 1$  (Gráfico a), o MBFC2 (Gráfico b) e o MBFC3 (Gráfico c).

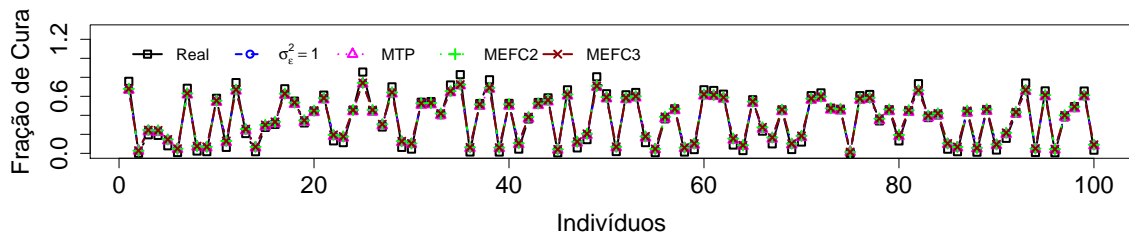
Avaliando os diagramas apresentados na Figura 2.11, percebemos a boa qualidade do ajuste fornecido para a variável  $\mathbf{X}_2$  pelos modelos MBFC1 a MBFC3. As estimativas consideradas são as médias dos valores esperados *a posteriori*. Esses gráficos mostram que há pouca ou nenhuma dispersão, indicando uma forte relação entre as estimativas médias fornecidas e o valores reais de  $\mathbf{X}_2$ , o que no contexto analisado dá uma ideia de viés muito próximo de zero.

### Efeito na Fração de cura

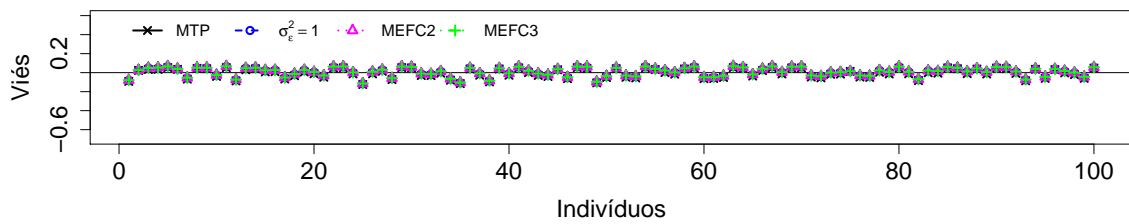
Nas Figuras 2.12 e 2.13, podemos analisar as estimativas obtidas para a fração de cura, quando ajustamos os dados através dos quatro modelos. Na Figura 2.12 comparamos tais estimativas pontualmente, além disso, o viés médio também pode ser observado. Na Figura 2.13 apresentamos um histograma da fração de cura real e sobreposto a ele estão as curvas das densidades empíricas dos valores esperados *a posteriori* para a fração de cura dos 100 indivíduos replicados no estudo de Monte Carlo. Podemos visualizar as estimativas médias obtidas por indivíduo para a fração de cura nos *box-plots* exibidos no Apêndice G.

Analisando os gráficos nas Figuras 2.12 e 2.13 vemos que o comportamento das estimativas médias fornecidas para a fração de cura, ajustando-se ambos os modelos, é bastante similar, de modo que não somos capazes fazer distinção entre o viés médio das estimativas fornecidas por cada modelo. Na Figura 2.13, vemos bem claramente que as distribuições empíricas dessas estimativas são aproximadamente iguais.

Para avaliar melhor as estimativas médias fornecidas para a fração de cura, na Tabela 2.8,



(a)



(b)

Figura 2.12: Médias (a) e viés (b) do valor esperado *a posteriori* para a fração de cura de cada um dos 100 indivíduos do Estudo 3, ajustando-se os modelos: MTP, MBFC1, MBFC2 e MBFC3.

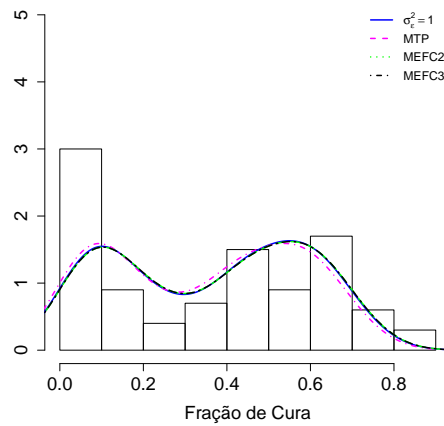


Figura 2.13: Histograma das frações de cura reais superposto pelas distribuições empíricas das estimativas médias dos valores esperados *a posteriori* para a fração de cura, obtidas ajustando-se os modelos MTP e MBFC1 a MBFC3.

apresentamos o percentual de intervalos de credibilidade que contém seu valor real. Nesta tabela, vemos que o MTP é o modelo que apresenta um menor percentual de intervalos de credibilidade que contém o verdadeiro valor da fração de cura, em torno de 92%. O percentual de cobertura fornecido por todas as especificações do MBFC é aproximadamente igual, em torno de 94%.

Tabela 2.8: Percentual médio de cobertura dos intervalos de 95% de credibilidade para a fração de cura, ajustando-se o MTP e os MBFC's.

Modelo			
MTP	MBFC1	MBFC2	MBFC3
91.74%	94.72%	94.05%	94.35%

Os resultados apresentados nessa seção sugerem que, mesmo quando não há erro de medida nos dados, os modelos que supõem erro de medida propostos ajustam-se bem, fornecendo em média estimativas com viés e EQM pequenos. Se comparados ao MTP, os modelos MBFC1 (quando fixado uma variância razoavelmente pequena), MBFC2 e MBFC3, apresentam uma melhor qualidade de ajuste para os dados de um modo geral, e, portanto, podem ser utilizados mesmo quando não há erro de medição em qualquer covariável do estudo.

#### 2.3.4 Estudo 4: Avaliando a Suposição de Normalidade para $\mathbf{X}_\epsilon$

Nesta subseção temos como objetivo avaliar a qualidade do ajuste fornecido pelos modelos MBFC2 e 3 quando supomos normalidade para a variável latente e esta variável na verdade não é normalmente distribuída. Comparamos as estimativas obtidas para os parâmetros supondo diferentes escolhas *a priori* para a variância  $\sigma_x^2$  da variável latente, visto que os Estudos 1 e 2, apresentados nesta seção, sugerem que a estimação desse parâmetro é bastante sensível a tal escolha.

Neste estudo, consideramos os dados do Cenário 3 e supomos as mesmas distribuições *a priori* eliciadas para  $\beta_1, \beta_2, \alpha, \lambda, \mu_x, \sigma_x^2$  e  $k$  no Estudo 3. Para a variância  $\sigma_x^2$  da variável latente  $\mathbf{X}_\epsilon$ , supomos três distribuições *a priori* Gama Inversa informativas, centradas em 10 que é o verdadeiro valor de  $\sigma_x^2$ , com três diferentes variâncias (1, 10 e 100); quanto maior é a variância suposta, menos informativa é a distribuição. Além destas, supomos a distribuição *a priori* não informativa  $\sigma_x^2 \sim GI(0.1, 0.1)$ .

Na Tabela 2.9, apresentamos as estimativas obtidas para todos os parâmetros do MBFC2 e do MBFC3 quando *a priori* temos  $\mathbf{X}_\epsilon \sim N(\mu_x, \sigma_x^2)$ , mas na verdade tais variáveis têm distribuição Gama com média  $\mu_x$  e variância  $\sigma_x^2$ . Além disso, consideramos as diferentes distribuições *a priori* para  $\sigma_x^2$ .

Tabela 2.9: Média do valor esperado *a posteriori*, Viés e Erro Quadrático Médio (EQM) das estimativas obtidas ajustando-se os modelos MBFC2 e MBFC3.

Modelo	$\eta$	Real	$\pi(\sigma_x^2)$											
			$GI(0.1, 0.1)$			$GI(3, 20)$			$GI(12, 110)$			$GI(102, 1010)$		
		Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM	
MBFC2	$\beta_1$	1.8	2.339	0.539	1.040	2.350	0.550	1.162	2.259	0.459	0.755	2.229	0.429	0.787
MBFC3			2.259	0.459	1.063	2.223	0.423	0.672	2.226	0.426	0.733	2.220	0.420	0.712
MBFC2	$\beta_2$	-0.15	-0.291	-0.141	0.062	-0.285	-0.135	0.051	-0.248	-0.098	0.027	-0.237	-0.087	0.046
MBFC3			-0.229	-0.079	0.018	-0.230	-0.080	0.018	-0.230	-0.080	0.018	-0.228	-0.078	0.018
MBFC2	$\alpha$	1	1.133	0.133	0.071	1.131	0.131	0.068	1.103	0.103	0.052	1.094	0.094	0.051
MBFC3			1.087	0.087	0.046	1.088	0.088	0.047	1.090	0.090	0.048	1.087	0.087	0.046
MBFC2	$\lambda$	-1.2	-1.168	0.032	0.540	-1.206	-0.006	0.659	-1.277	-0.077	0.542	-1.310	-0.110	0.521
MBFC3			-1.354	-0.154	0.933	-1.314	-0.114	0.562	-1.318	-0.118	0.640	-1.322	-0.122	0.622
MBFC2	$\mu_x$	3.9	3.741	-0.159	0.037	3.745	-0.155	0.035	3.747	-0.153	0.035	3.748	-0.152	0.034
MBFC3			3.748	-0.152	0.034	3.749	-0.151	0.034	3.748	-0.152	0.034	3.749	-0.151	0.034
MBFC2	$\sigma_x^2$	10	7.750	-2.250	7.388	8.006	-1.994	4.927	8.781	-1.219	1.766	9.637	-0.363	0.163
MBFC3			9.278	-0.722	1.010	9.127	-0.873	1.192	9.264	-0.736	0.836	9.699	-0.301	0.133
MBFC2	$\sigma_\epsilon^2$	1	2.147	1.147	3.104	1.837	0.837	1.461	1.301	0.301	0.266	0.998	-0.002	0.087
MBFC3			1.010	0.010	0.000	1.009	0.009	0.000	1.011	0.011	0.000	1.015	0.015	0.000

### Efeito nas Estimativas dos Parâmetros

Analisando a Tabela 2.9, vemos que as estimativas obtidas para os parâmetros, supondo que as variáveis latentes são normalmente distribuídas quando na verdade não são, apresentam viés e EQM maior do que aquelas estimativas obtidas quando a suposição de normalidade era válida (ver resultados nos Estudos de 1 a 3). Podemos observar também que o modelo MBFC3 em geral se ajusta melhor aos dados do que o MBFC2, produzindo estimativas menos enviesadas para a maioria dos parâmetros. Isso deve acontecer porque no MBFC2 utilizamos distribuições *a priori* não informativas para todos os parâmetros do modelo, salvo  $\sigma_x^2$  cuja a escolha da distribuição *a priori* está sendo analisada. Enquanto que no MBFC3, eliciamos uma distribuição *a priori* muito informativa para  $k$ , consequentemente, agregando informação sobre a variância do erro de medida  $\sigma_\epsilon^2$  ao modelo (no Estudo 5 fazemos uma análise de sensibilidade à escolha da distribuição *a priori* de  $k$  do MBFC3). Em relação a eliciação da distribuição *a priori* de  $\sigma_x^2$ , vemos que considerar uma distribuição muito informativa resulta em estimativas com viés e EQM menores do que aquelas obtidas quando supomos uma pouco informativa. Por exemplo, supondo  $\sigma_x^2 \sim GI(0.1, 0.1)$  o viés da estimativa média obtido ajustando-se o MBFC3 é  $-0.722$  ( $-2.250$  no MBFC2) e quando  $\sigma_x^2 \sim GI(102, 1010)$  o viés médio é  $-0.301$  (e  $-0.363$  no MBFC2).

Nas Figuras 2.14 e 2.15 a seguir, podemos avaliar as estimativas médias obtidas para o valor esperado *a posteriori* da variável latente de cada indivíduo através de um diagrama de dispersão, no qual relacionamos cada estimativa obtida com o valor verdadeiro  $X_2$ .

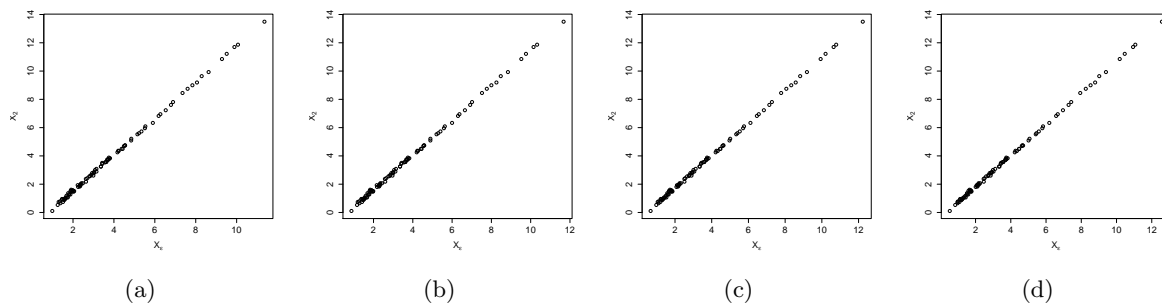


Figura 2.14:  $X_\epsilon$  no eixo das abscissas representa as estimativas obtidas para  $X_2$  ajustando-se o MBFC2 quando a distribuição *a priori* de  $\sigma_x^2$  é  $GI(0.1, 0.1)$  (Gráfico a),  $GI(3, 20)$  (Gráfico b),  $GI(12, 110)$  (Gráfico c) e  $GI(102, 1010)$  (Gráfico d).

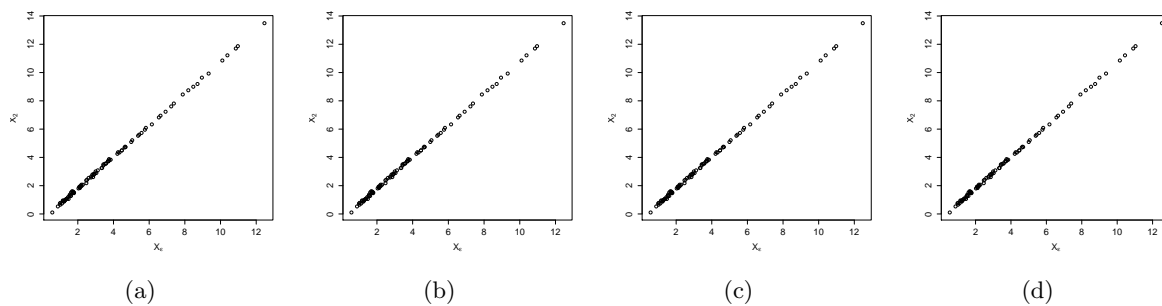


Figura 2.15:  $X_\epsilon$  no eixo das abscissas representa as estimativas obtidas para  $X_2$  ajustando-se o MBFC3 quando a distribuição *a priori* de  $\sigma_x^2$  é  $GI(0.1, 0.1)$  (Gráfico a),  $GI(3, 20)$  (Gráfico b),  $GI(12, 110)$  (Gráfico c) e  $GI(102, 1010)$  (Gráfico d).

Os diagramas de dispersão apresentados nas Figuras 2.14 e 2.15, indicam que as estimativas obtidas para as variáveis latentes são bem próximas ao valor real destas variáveis, o que acontece em todos os modelos ajustados. Os gráficos exibidos são muito parecidos e visualmente não é possível notar diferença entre eles. Analisando as estimativas pontualmente, notamos que elas diferem uma da outra a partir da terceira ou quarta casa decimal, indicando que tais estimativas não são sensíveis a escolha da distribuição *a priori* da variância  $\sigma_x^2$  dessas variáveis.

### Efeito na Fração de Cura

Para analisarmos as estimativas médias obtidas para o valor esperado *a posteriori* da fração de cura apresentamos, nas Figuras 2.16 a 2.18, uma comparação entre as estimativas obtidas e o valor real da fração de cura para cada indivíduo. Na Figura 2.17, exibimos o viés médio para as estimativas obtidas ajustando-se os modelos MBFC2 e MBFC3 supondo as diferentes distribuições *a priori* para  $\sigma_x^2$ .



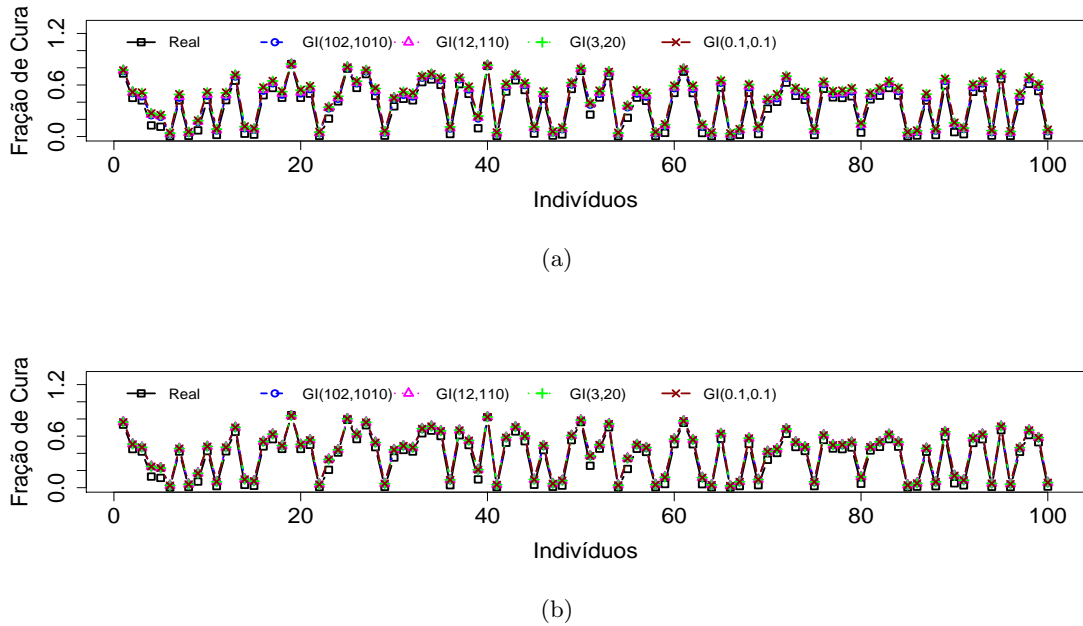


Figura 2.16: Estimativas médias do valor esperado *a posteriori* para a fração de cura dos indivíduos do Cenário 4, obtidas ajustando-se os modelos MBFC2 (Fig.a) e MBFC3 (Fig.b), assumindo que *a priori*  $\sigma_x^2$  tem distribuição:  $GI(0.1, 0.1)$  (linha com  $\times$ ),  $GI(3, 20)$  (linha com  $+$ ),  $GI(12, 20)$  (linha com  $\nabla$ ) e  $GI(102, 1010)$  (linha com  $\circ$ ). O símbolo  $\square$  representa os valores reais da fração de cura.

Na Figura 2.16 vemos o quanto as estimativas aproximam bem o valor real de maneira bastante similar em todos os ajustes fornecidos. Visualmente não é fácil fazermos distinção entre as curvas resultantes através de cada diferente ajuste. Essa conclusão fica mais clara ao observarmos a Figura 2.17, na qual podemos notar que o viés médio obtido, ajustando-se todos os modelos, se aproxima bastante de zero.

Em relação a sensibilidade da eliciação da distribuição *a priori* para  $\sigma_x^2$ , a Figura 2.18 nos mostra que o modelo MBFC2 é um pouco mais sensível a essa escolha, pois, a curva empírica que melhor aproxima o histograma das frações de cura reais é a curva das estimativas que foram obtidas supondo  $\sigma_x^2 \sim GI(102, 1010)$ . Enquanto que as estimativas fornecidas pelo MBFC3 não foram sensíveis a tal escolha *a priori*. Nesta Figura também vemos que a fração de cura é subestimada para os indivíduos cujos os valores estão nos intervalos (0, 10%] e [40%, 50%).

Na Tabela 2.10, apresentamos o percentual de intervalos de 95% de credibilidade para a fração de cura, que contém seus verdadeiros valores, obtidos ajustando-se os modelos MBFC2 e MBFC3 supondo diferentes distribuições *a priori* para  $\sigma_x^2$ .

Observando os percentuais exibidos na Tabela 2.10, podemos concluir que, no que diz

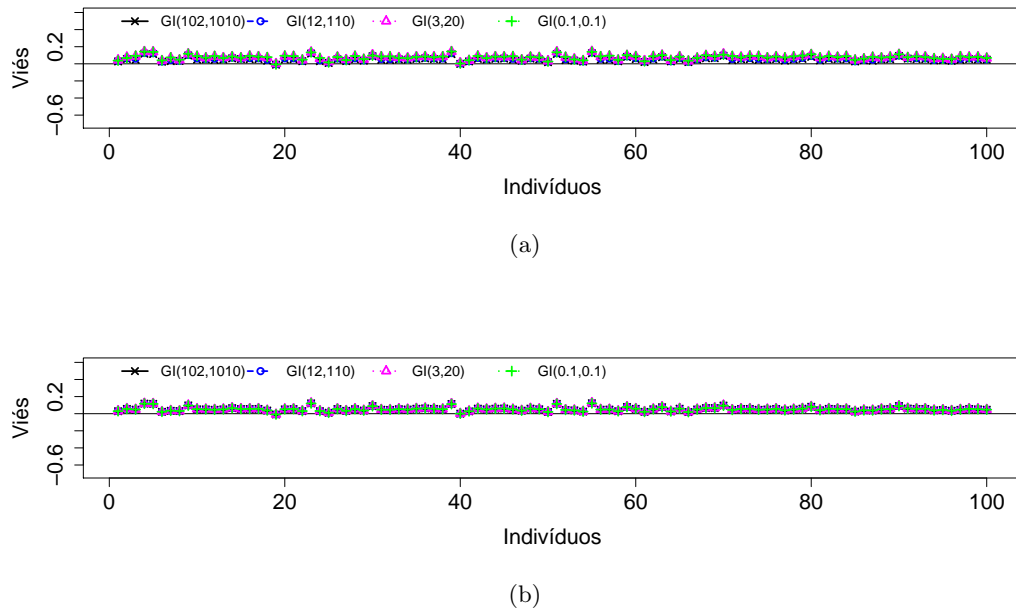


Figura 2.17: Viés das estimativas médias do valor esperado *a posteriori* para a fração de cura dos indivíduos do Cenário 4, obtidas ajustando-se os modelos MBFC2 (Gráfico a) e MBFC3 (Gráfico b), assumindo que *a priori*  $\sigma_x^2$  tem distribuição:  $GI(0.1, 0.1)$  (linha com  $\times$ ),  $GI(3, 20)$  (linha com  $+$ ),  $GI(12, 20)$  (linha com  $\nabla$ ) e  $GI(102, 1010)$  (linha com  $\circ$ ). O símbolo  $\square$  representa os valores reais da fração de cura.

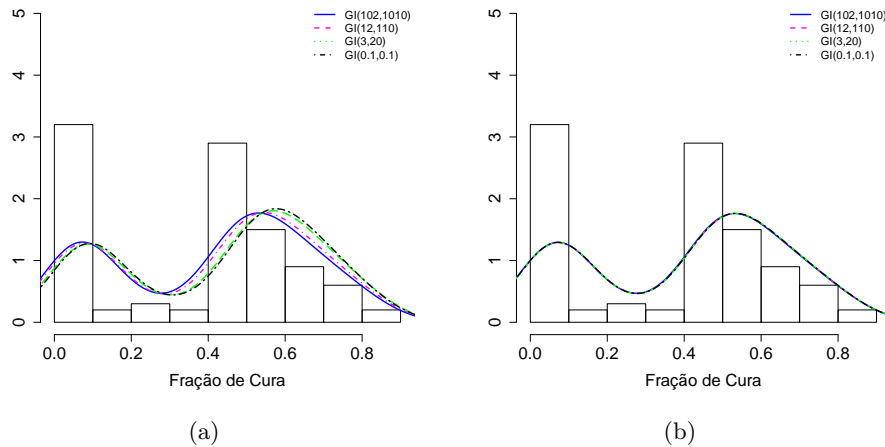


Figura 2.18: Histograma das frações de cura reais superposto pelas distribuições empíricas das estimativas médias dos valores esperados *a posteriori* da fração de cura, obtidas ajustando-se os modelos MEFC2 (Gráfico a) e MEFC3 (Gráfico b).

Tabela 2.10: Percentual médio de cobertura dos intervalos de 95% de credibilidade para a fração de cura, ajustando-se o MBFC2 e o MBFC3 para as diferentes escolhas *a priori* de  $\sigma_x^2$ .

Modelo	$GI(0.1, 0.1)$	$GI(3, 20)$	$GI(12, 110)$	$GI(102, 1010)$
MBFC2	90.26%	90.34%	90.32%	90.28%
MBFC3	90.28%	90.18%	90.06%	90.27%

respeito a estimação da fração de cura, todos os modelos analisados se comportam de maneira muito similar. Para todas as escolhas de distribuição *a priori* para  $\sigma_x^2$ , obtivemos um percentual de cobertura bem próximo a 90%, ao ajustarmos ambos os modelos. Esses resultados, confirmam o que nos mostram as Figuras 2.16 à 2.18. Portanto, apesar de concluirmos que existe uma certa sensibilidade em relação a escolha da distribuição *a priori* de  $\sigma_x^2$  nas estimativas de tal parâmetro, vemos que essa eliciação não influi significativamente na estimação da fração de cura.

### 2.3.5 Estudo 5: Avaliando a Escolha da Distribuição *a priori* da constante $k$ no MBFC3

No estudo de simulação apresentado nesta subseção, consideramos os Cenários 1 e 2. Nosso objetivo é analisar a escolha da distribuição *a priori* da razão  $k$  entre as variâncias do erro de medida e das variáveis latentes. Os valores reais de  $k$  nesse estudo são 0.1 e 0.5 para os dados dos Cenários 1 e 2, respectivamente.

Para o vetor de parâmetros  $\eta$  do MBFC3, permanecemos com as mesmas escolhas de distribuição *a priori* do Estudo 2, supondo  $\sigma_x^2 \sim GI(102, 1010)$ . Ao construirmos tal modelo, partimos do pressuposto que a variância do erro de medida não deve ser maior do que a variância das variáveis latentes, pois, se assim fosse, estaríamos assumindo um erro de medida grande, o que acreditamos não ser uma suposição razoável na prática, pois, indicaria que o instrumento utilizado na medição seria pouco ou nada preciso. Dessa maneira, focamos o estudo na escolha da distribuição *a priori* de  $k$ , com a finalidade de fazermos uma análise de sensibilidade. Assumindo que  $0 < \frac{\sigma_\epsilon^2}{\sigma_x^2} \leq 1$ , construímos distribuições *a priori* para  $k$ . Consideramos algumas distribuições pouco ou muito informativas obedecendo esse pressuposto. Para ajustar os dados do Cenário 1, escolhemos *a priori* as distribuições  $G(0.1, 1)$ ,  $G(0.5, 0.001)$ ,  $U(0, 0.5)$  e  $U(0, 1)$ , e para ajustar os dados do Cenário 2 eliciamos as distribuições  $G(0.5, 5)$ ,  $G(0.5, 50)$ ,  $G(0.5, 0.001)$  e  $U(0, 1)$ .

Nas Tabelas 2.11 e 2.12 a seguir, apresentamos as estimativas médias, viés e EQM para o valor esperado *a posteriori* de todos os parâmetros do MBFC3, ajustando-se os dados dos Cenários 1 e 2, supondo as diferentes distribuições *a priori*, supracitadas para  $k$ . Para ajustarmos os dados do Cenário 2, consideramos no MCMC um *lag* de tamanho 30, pois, mesmo após a convergência, os parâmetros apresentavam uma autocorrelação muito alta. Para esse estudo,

fizemos 60000 iterações selecionando uma amostra de tamanho 2000 da distribuição *a posteriori* dos parâmetros em cada uma das 250 réplicas de Monte Carlo. Sob o Cenário 1, no MCMC permanecemos com a mesma configuração dos demais estudos de simulação apresentados nesta seção.

Tabela 2.11: Média do valor esperado *a posteriori*, Viés e Erro Quadrático Médio (EQM) das estimativas obtidas ajustando-se o modelo MBFC3 aos dados do Cenário 1, supondo diferentes distribuições *a priori* para  $k$ .

$\eta$	Real	$\pi(k)$											
		Gama(0.1, 1)			Gama(0.5, 0.001)			U(0, 0.5)			U(0, 1)		
		Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM
$\beta_1$	1.8	1.826	0.026	0.369	1.830	0.030	0.366	1.988	0.188	0.571	2.534	0.734	3.329
$\beta_2$	-0.15	-0.144	0.006	0.007	-0.144	0.006	0.007	-0.220	-0.070	0.030	-0.413	-0.263	0.360
$\alpha$	1	1.077	0.077	0.053	1.077	0.077	0.053	1.139	0.139	0.080	1.307	0.307	0.371
$\lambda$	-1.2	-1.259	-0.059	0.439	-1.264	-0.064	0.431	-1.084	0.116	0.504	-0.937	0.263	0.573
$\mu_x$	3.9	3.920	0.020	0.010	3.920	0.020	0.010	3.919	0.019	0.011	3.891	-0.009	0.013
$\sigma_x^2$	10	10.141	0.141	0.059	10.140	0.140	0.058	9.171	-0.829	0.708	8.600	-1.400	1.978
$\sigma_\epsilon^2$	1	1.011	0.011	0.000	1.016	0.016	0.000	4.454	3.454	11.936	8.266	7.266	52.819

Tabela 2.12: Média do valor esperado *a posteriori*, Viés e Erro Quadrático Médio (EQM) das estimativas obtidas ajustando-se o modelo MBFC3 aos dados do Cenário 2, supondo diferentes distribuições *a priori* para  $k$ .

$\eta$	Real	$\pi(k)$											
		Gama(0.5, 0.001)			Gama(0.5, 5)			Gama(0.5, 50)			U(0, 1)		
		Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM
$\beta_1$	1.8	1.775	-0.025	0.341	1.774	-0.026	0.339	1.775	-0.025	0.342	2.126	0.326	1.391
$\beta_2$	-0.15	-0.105	0.045	0.008	-0.105	0.045	0.008	-0.105	0.045	0.008	-0.259	-0.109	0.090
$\alpha$	1	1.058	0.058	0.048	1.058	0.058	0.048	1.057	0.057	0.048	1.192	0.192	0.173
$\lambda$	-1.2	-1.400	-0.200	0.427	-1.399	-0.199	0.423	-1.403	-0.203	0.428	-1.080	0.120	0.502
$\mu_x$	3.9	3.928	0.028	0.051	3.928	0.028	0.051	3.928	0.028	0.051	3.914	0.014	0.050
$\sigma_x^2$	10	11.381	1.381	2.138	11.383	1.383	2.143	11.390	1.390	2.163	7.780	-2.220	5.594
$\sigma_\epsilon^2$	5	1.006	-3.994	15.954	1.001	-3.999	15.990	0.962	-4.038	16.308	7.588	2.588	7.333

### Efeito nas Estimativas dos Parâmetros

Os resultados apresentados na Tabela 2.11 sugerem que quando escolhemos uma distribuição Uniforme para  $k$ , as estimativas fornecidas para todos os parâmetros do MBFC3 são mais enviesadas do que aquelas obtidas quando a distribuição *a priori* eliciada é Gama. O mesmo ocorre no estudo apresentado na Tabela 2.12. No entanto, as estimativas obtidas para  $\sigma_\epsilon^2 = k\sigma_x^2$ , possuem um viés e EQM muito grande nesse segundo estudo, indicando que quando supomos distribuições *a priori* Gama para  $k$  as estimativas desse parâmetro se concentram em torno de 0.1, independente do seu valor real, mesmo quando a distribuição é centrada no verdadeiro valor de  $k$ , como são a Gama(0.5, 5) e a Gama(0.5, 50) ou quando é não informativa, como a

Gama(0.5, 0.001). Por outro lado, quando escolhemos as distribuições *a priori*  $U(a, b)$ , as estimativas do valor esperado *a posteriori* de  $k$  tendem a ser mais próximas ao hiperparâmetro  $b$ . Portanto, baseados nos resultados observados nas Tabelas 2.11 e 2.12 percebemos que a eliciação *a priori* de  $k$  é um desafio, visto que as estimativas obtidas para  $\sigma_\epsilon^2$  são bastante influenciadas por essa escolha. No entanto, em relação aos demais parâmetros, o ajuste fornecido quando escolhemos para  $k$  as distribuições *a priori* Gama é bastante satisfatório, visto que o viés e EQM das estimativas é relativamente pequeno.

Nas Figuras 2.19 e 2.20 mostramos um diagrama de dispersão no qual relacionamos as estimativas obtidas para as variáveis latentes com seus verdadeiros valores. Observando esses gráficos podemos avaliar a qualidade das estimativas fornecidas para as variáveis latentes quando assumimos diferentes distribuições *a priori* para  $k$  no MBFC3.

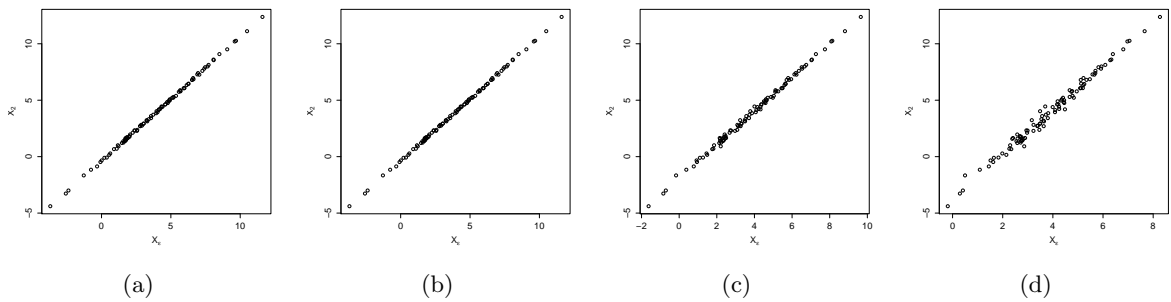


Figura 2.19:  $X_\epsilon$  no eixo das abscissas representa as estimativas obtidas para  $X_2$  ajustando-se o MBFC3 aos dados do Cenário 1, supondo como distribuições *a priori* para  $k$ : Gama(0.1, 1) (Gráfico a), Gama(0.5, 0.001) (Gráfico b),  $U(0, 0.5)$  (Gráfico c) e  $U(0, 1)$  (Gráfico d).

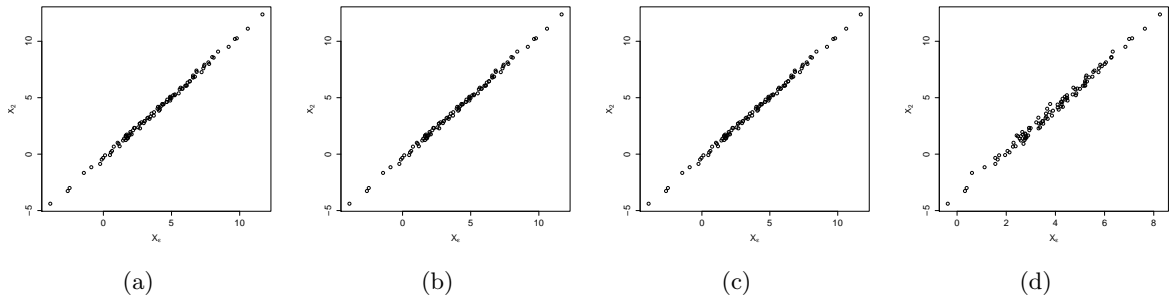


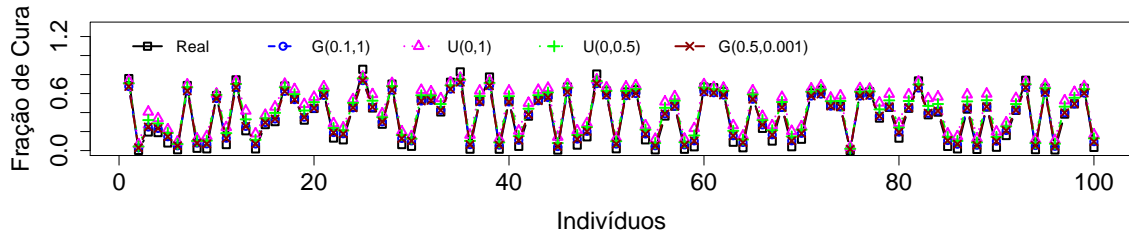
Figura 2.20:  $X_\epsilon$  no eixo das abscissas representa as estimativas obtidas para  $X_2$  ajustando-se o MBFC3 aos dados do Cenário 2 supondo como distribuição *a priori* para  $k$ : Gama(0.5, 0.001) (Gráfico a), Gama(0.5, 5) (Gráfico b), Gama(0.5, 50) (Gráfico c) e  $U(0, 1)$  (Gráfico d).

As Figuras 2.19 e 2.20 mostram que nos dois cenários analisados, o comportamento das

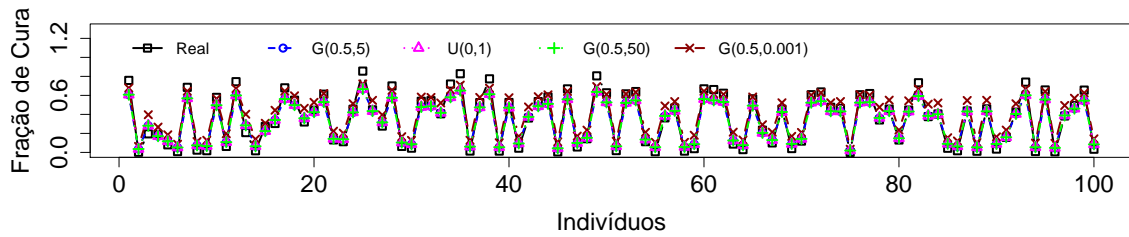
variáveis latentes é similar ao que observamos para os demais parâmetros do modelo. Isto é, nos dois cenários, quando supomos uma distribuição *a priori* Gama para  $k$  (Fig.2.19a, 2.19b, 2.20a, 2.20b e 2.20c), as estimativas fornecidas para as variáveis latentes apresentam uma forte relação com os valores reais de  $\mathbf{X}_2$ , indicando que são menos enviesadas do que as estimativas obtidas quando supomos uma distribuição Uniforme para  $k$  (Fig.2.19c, 2.19d e 2.20d). Esses resultados estão, possivelmente, relacionados com a variância destas distribuições.

### Efeito nas Estimativas da Fração de Cura

Para analisar as estimativas obtidas para o valor esperado *a posteriori* da fração de cura, mostramos nas Figuras 2.21 a 2.23 uma comparação entre as estimativas fornecidas ajustando-se o MBFC3, considerando os Cenários 1 e 2, assumindo as diferentes distribuições *a priori* para  $k$ . O viés de tais estimativas pode ser visualizado na Figura 2.22.



(a)



(b)

Figura 2.21: Estimativas médias do valor esperado *a posteriori* para a fração de cura dos indivíduos dos os Cenários 1 (Gráfico a) e 2 (Gráfico b), ajustando-se o MBFC3 e assumindo diferentes distribuições *a priori* para  $k$ : Gama(0.1,1) (linha com  $\circ$ ), Gama(0.5,0.001) (linha com  $\times$ ),  $U(0,0.5)$  (linha com  $+$ ) e  $U(0,1)$  (linha com  $\nabla$ ). O símbolo  $\square$  representa os valores reais da fração de cura.

A Figura 2.21 mostra que o comportamento das estimativas obtidas para a fração de cura é semelhante sob todas as suposições *a priori* de  $k$  nos dois cenários analisados. Entretanto, vemos que as estimativas fornecidas pelos modelos em que assumimos distribuição Gama(0.1,1)

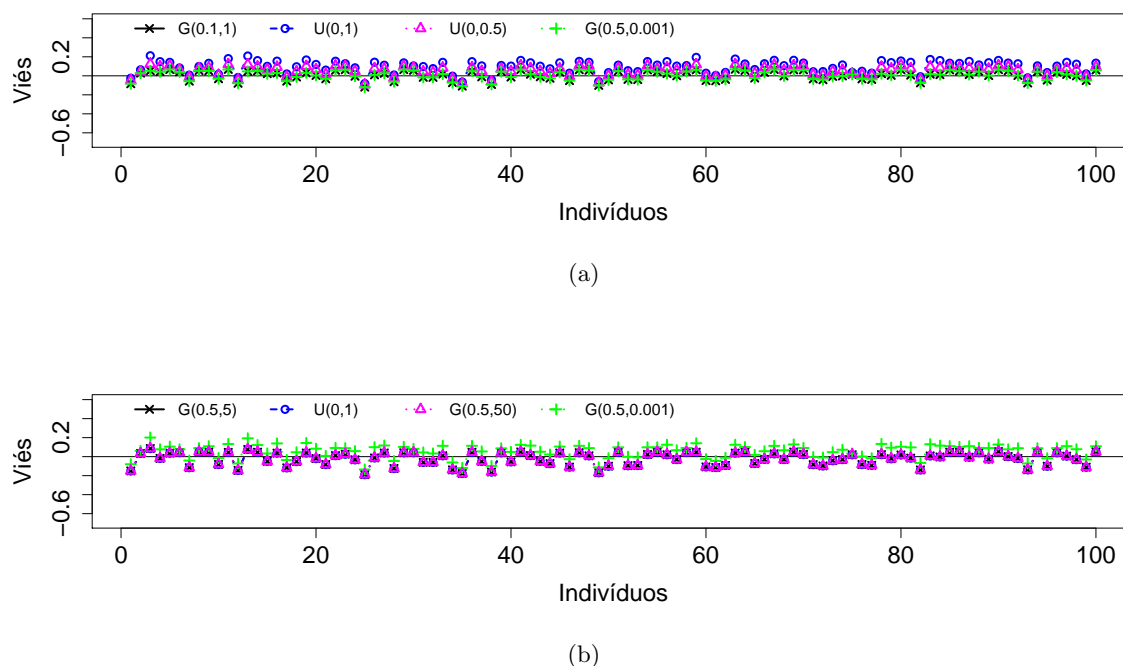


Figura 2.22: Viés das estimativas médias do valor esperado *a posteriori* para a fração de cura dos indivíduos do Cenário 1 (Gráfico a) e 2 (Gráfico b), ajustando-se o MBFC3 e assumindo diferentes distribuições *a priori* para  $k$ : Gama(0.1, 1) (linha com  $\circ$ ), Gama(0.5, 0.001) (linha com  $\times$ ),  $U(0, 0.5)$  (linha com  $+$ ) e  $U(0, 1)$  (linha com  $\nabla$ ).

(Cenário 1),  $U(0, 1)$  (Cenário 2) e Gama(0.5, 0.001) (nos dois cenários), aproximam melhor os valores reais. Isto fica mais evidenciado na Figura 2.22, na qual observamos que o viés das estimativas obtidas supondo-se tais distribuições é menor para a maioria dos indivíduos, visto que as linhas estão mais próximas de zero do que quando supomos as distribuições  $U(0, 0.5)$ ,  $U(0, 1)$  (Cenário 1), Gama(0.5, 5) e Gama(0.5, 50) (Cenário 2). Além disso, vemos na Figura 2.23, o histograma das frações de cura reais superposto pelas distribuições empíricas das médias dos valores esperados *a posteriori* da fração de cura, mostrando que os indivíduos com fração de cura de até 10% tem este valor subestimado em todos os ajustes fornecidos. Para os demais indivíduos, observamos que os modelos nos quais supomos distribuição *a priori* Uniforme para  $k$  captam melhor o formato da distribuição empírica da fração de cura e o mesmo acontece para os indivíduos do Cenário 2.

Na Tabela 2.13, exibimos o percentual de cobertura dos intervalos de 95% de credibilidade da fração de cura, obtidos ao ajustarmos o MBFC3, considerando os dados dos Cenários 1 e 2, assumindo as diferentes distribuições *a priori* para  $k$ .

Observando os percentuais apresentados a Tabela 2.13, percebemos que maiores coberturas

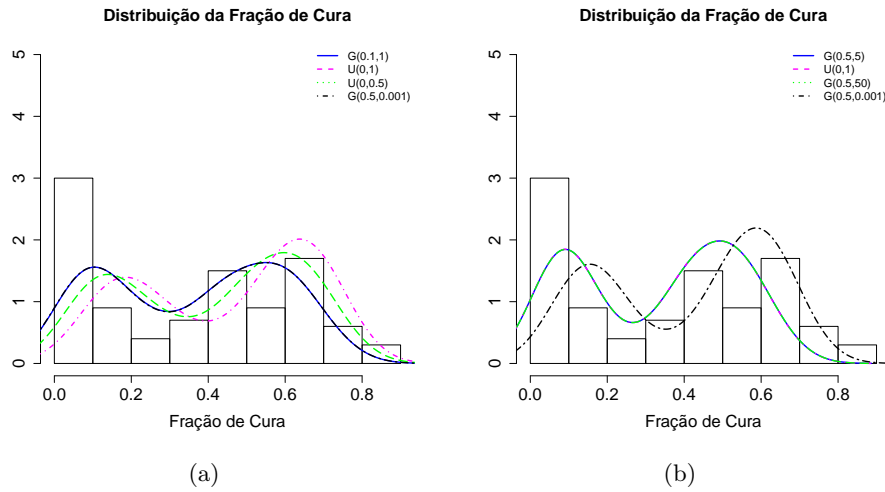


Figura 2.23: Histograma das frações de cura reais superposto pelas distribuições empíricas das médias dos valores esperados *a posteriori* da fração de cura, obtidas ajustando-se o MBFC3 considerando os Cenários 1 (Gráfico a) e 2 (Gráfico b).

Tabela 2.13: Percentual médio de cobertura dos intervalos de 95% de credibilidade da fração de cura, ajustando-se o MBFC3 aos dados dos Cenário 1 e 2, considerando as diferentes escolhas *a priori* para  $k$ .

Cenário 1				Cenário 2			
Gama(0.1, 1)	Gama(0.5, 0.001)	$U(0, 0.5)$	$U(0, 1)$	Gama(0.5, 0.001)	Gama(0.5, 5)	Gama(0.5, 50)	$U(0, 1)$
92.34%	92.31%	95.07%	94.34%	81.62%	81.71%	81.42%	93.08%

são alcançadas sob os modelos nos quais assumimos distribuição *a priori* Uniforme para  $k$ . Isto coincide com o que foi observado na Figura 2.23, na qual notamos que tais modelos fornecem uma melhor aproximação para a distribuição empírica da fração de cura real. Para os indivíduos do Cenário 1, o percentual de cobertura dos *HPD*'s para a fração de cura foi em torno de 92% ao ajustarmos o MBFC3 supondo distribuição *a priori* Gama para  $k$ , enquanto que este percentual foi de 94% e 95% quando supomos distribuição *a priori*  $U(0, 1)$  e  $U(0, 0.5)$ , respectivamente. Para os indivíduos do Cenário 2, a maior cobertura foi observada quando assumimos  $k \sim U(0, 1)$  *a priori*.

### 2.3.6 Comentários Gerais

Em geral, tivemos alguns problemas para fazer inferência nos modelos propostos. Inicialmente, tínhamos considerado uma modelagem com intercepto no preditor linear  $\theta$ , isto é, considerávamos  $\beta = (\beta_0, \beta_1, \dots, \beta_p)^\top$ . Porém, todos os modelos apresentavam problemas de falta de identificabilidade. O intercepto se confundia com o parâmetro de escala  $\lambda$  da distribuição



*Weibull* e para sanar tal dificuldade tentamos diversas alternativas. Primeiramente, consideramos algumas blocagens diferentes para os parâmetros no MCMC, de modo que amostrássemos de distribuições *a posteriori* conjuntas para alguns parâmetros. Utilizamos os blocos:  $(\beta_0, \lambda)$ ,  $(\beta_0, \beta_i)$  em que  $i = 1, 2$  e  $(\beta_0, \beta_1, \beta_2)$ . Porém, isso não fez mais do que diminuir a autocorrelação. Uma outra maneira que pensamos para resolver o problema de falta de identificabilidade, foi utilizar a técnica de aumento de dados (detalhes sobre o uso dessas técnicas podem ser vistos em Chib (1992); Tanner e Wong (1987)), porém, isso também não resolveu o problema. Tentamos ainda, aplicar transformações nos parâmetros de modo a modificar o suporte de cada um deles, por exemplo, o parâmetro  $\beta_0 \in (-\infty, +\infty)$ , então, aplicamos a transformação  $\beta_0 = \log\left(\frac{a}{1-a}\right)$  em que  $a \in (0, 1)$ . Tal reparametrização também não tornou o modelo identificável. Outra maneira pensada para resolver nosso problema foi considerar a razão entre variâncias  $k$  conhecida. Tal alternativa foi proposta por Fuller (2009) para resolver problemas de falta de identificabilidade em modelos com erros nas covariáveis. Porém, no nosso caso, não funcionou para este fim, mas, nos inspirou a propor o MBFC3. Essa nova modelagem proporcionou mais velocidade aos algoritmos MCMC e em alguns casos, se ajustou melhor aos dados, fornecendo estimativas menos enviesadas para os parâmetros (como observamos no estudos de simulação apresentados nesta seção). O problema de falta de identificabilidade foi solucionado utilizando uma alternativa bastante comum em modelos de sobrevivência que é considerar uma modelagem sem intercepto (Cox, 1972).

Na modelagem proposta, as variáveis latentes são muito bem aproximadas pelas estimativas fornecidas, o que pôde ser observado nos diagramas de dispersão apresentados nos Estudos 2 a 5, mesmo no caso em que tais variáveis não eram normalmente distribuídas (Estudo 4). Isso parece contribuir para a obtenção de estimativas menos enviesadas para a fração de cura, visto os resultados de tais estudos. Entretanto, a suposição de normalidade para tais variáveis, pode levar à estimativas muito enviesadas para as variâncias do erro e das variáveis latentes. Nessa perspectiva, uma proposta para trabalhos futuros é considerar uma classe mais ampla de distribuições para ajustar tais variáveis, com o objetivo de fornecer estimativas ainda mais precisas para as variâncias do modelo.

Os modelos MBFC2 e MBFC3 propostos, em geral, apresentam resultados muito satisfatórios em todos cenários analisados, visto que o viés e o EQM das estimativas obtidas para todos os parâmetros do modelo, fração de cura e variáveis latentes é bem próximo a zero. Tais modelos tem como vantagens algumas suposições que os tornam bastante realísticos do ponto de vista prático, e portanto, surgem como uma excelente alternativa para ajustar dados de sobrevivência com fração de cura, mesmo no contexto em que as variáveis não tenham sido medidas com erro (ver Estudo 3). Entretanto, o MBFC3 tem uma certa limitação em relação a elicitação da razão

de variâncias  $k$ , visto que, para obter estimativas menos enviesadas para a variância do erro de medida, é necessário uma distribuição *a priori* muito informativa.

Para tornar os modelos propostos ainda mais realísticos, sugerimos que outros tipos de modelagem sejam consideradas futuramente. Por exemplo, pode-se supor outras distribuições *a priori* para os parâmetros do modelo, com a finalidade de produzir estimativas mais próximas dos valores reais. Nesse sentido, aparece como um desafio a eliciação de distribuições *a priori* para  $k$ , com visto no Estudo 5. Outra proposta, é considerar um modelo semiparamétrico para os tempos de falha, como por exemplo, o modelo exponencial por partes (Demarqui *et al.*, 2011).

Concluimos, ao analisarmos o Cenário 4, que quando a variância do erro de medida é significativamente grande, torna-se necessário utilizar distribuições *a priori* mais informativas para a variância  $\sigma_x^2$  das variáveis latentes. Além disso, percebemos que um modelo ingênuo não é adequado nesse tipo de situação.

## 2.4 Aplicação

Consideremos o banco de dados relacionado ao ensaio clínico pós cirúrgico de pacientes com melanoma cutâneo, coletado no Eastern Cooperative Oncology Group (ECOG). O estudo, denotado por E1684, é bastante conhecido na literatura de modelos de fração de cura e foi previamente analisado por Chen *et al.* (1999), Chen *et al.* (2002d) e Ibrahim *et al.* (2005). Uma descrição detalhada dele pode ser vista em Kirkwood *et al.* (1996). O objetivo dessa triagem clínica foi avaliar o desempenho de um tratamento administrado nos pacientes após serem submetidos a cirurgia para a extração do melanoma. O tratamento consistiu em aplicar nos pacientes altas doses de Interferon alfa-2b, a fim de evitar a reincidência da doença. Dos 287 pacientes incluídos no estudo, sete foram cancelados antes do tratamento e 28 eram inelegíveis, restando um total de 252. Entre estes, consideramos aqui 234 indivíduos após retirarmos os pacientes que apresentavam informações faltantes. Somente uma parcela recebeu o tratamento e os demais permaneceram apenas em observação. Optamos pela análise desse banco de dados, pois, uma das covariáveis observadas é o índice de Breslow, que avalia a profundidade do tumor e é dada em milímetros. Como a medição é feita com base em imagens geradas em tomografia, a interpretação da profundidade do tumor é passível de erro. Isto sugere que utilizarmos um modelo que acomode erro nas covariáveis pode ser mais apropriado para analisar tais dados.

Além de considerarmos que o índice de Breslow pode ter sido observado ou registrado com erro, supomos que os dados apresentam uma fração de curados. Observando a Figura 2.24, que mostra o estimador de Kaplan-Meier para esses dados, percebe-se que a curva de sobrevivência parece estabilizar-se após um certo tempo. Esse comportamento é bastante comum em se tratando de estudos envolvendo dados de melanoma, e indica a presença de indivíduos

curados, inspirando-nos a usar um modelo que também acomode fração de cura.

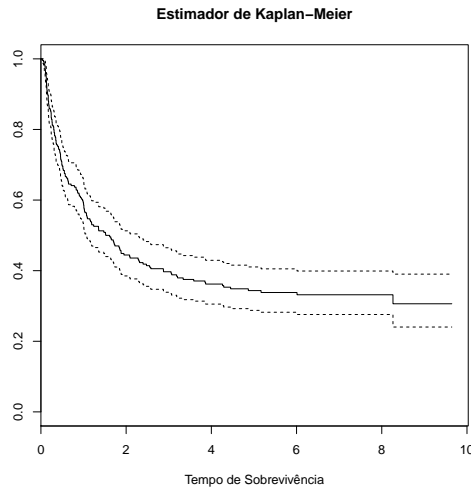


Figura 2.24: Curvas de sobrevivência para os 234 pacientes do estudo E1684.

No Estudo E1684, para cada indivíduo observado foi registrado: o tempo de falha em anos ( $Y$ ), o indicador de falha ou censura ( $\delta = 1$  caso o indivíduo tenha falhado e  $\delta = 0$  caso contrário), a idade em anos ( $X_1$ ), o sexo ( $X_2 = 0$  se o indivíduo for do sexo masculino e 1 se for do sexo feminino), tratamento ( $X_3 = 0$ , se o indivíduo ficou somente em observação e  $X_3 = 1$  se recebeu doses de Interferon alpha-2b), a profundidade de Breslow ( $X_4$ , medida em milímetros), e a performance física do paciente ( $X_5 = 1$  se o indivíduo pratica alguma atividade física e  $X_5 = 0$  caso seja sedentário). Consideramos em nossas análises que a covariável  $X_4$  é passível de ser mensurada ou registrada com erro, desse modo, supomos  $X_4 = X_\epsilon + \epsilon$ . Por razões computacionais, no MCMC consideramos o logaritmo natural das covariáveis  $X_1$  e  $X_4$ .

Consideramos duas distribuições para o tempo de falha: Weibull e Exponencial. Nas duas análises, ajustamos os modelos MBFC1, 2 e 3 introduzidos neste capítulo e o MTP (Chen *et al.*, 1999), que não supõe erro de medida nas covariáveis. Na segunda análise também ajustamos a versão estrutural Bayesiana do modelo proposto por Ma e Yin (2008), desenvolvido na Seção 2.2. Nas Tabelas 2.14 e 2.16 mostramos, respectivamente, os resultados obtidos assumindo distribuição Weibull e Exponencial para os tempos de falha.

No MCMC consideramos *lag* e *burn-in* de tamanhos 50 e 100000, respectivamente, resultando em um total de 200000 iterações e uma amostra de tamanho 2000 da distribuição *a posteriori* de cada parâmetro. Para os parâmetros  $\beta$ ,  $\alpha$  e  $\lambda$  que são comuns a todos os modelos, assumimos que a *a priori*  $\pi(\beta) \propto \mathbf{1}$ ,  $\alpha \sim \text{Gama}(1, 1)$  e  $\lambda \sim \text{Normal}(-0.63, 10)$ . Nos três MBFC, supomos *a priori* que  $X_\epsilon \sim \text{Normal}(\mu_x, \sigma_x^2)$ ,  $\mu_x \sim \text{Normal}(0, 10)$  e  $\sigma_x^2 \sim \text{GI}(100, 100)$ . Nos mo-

delos MBFC2 e MBFC3 assumimos, respectivamente, *a priori* que  $\sigma_\epsilon^2 \sim GI(1, 1)$  e  $k \sim U(0, 0.5)$ . Escolhas de distribuições *a priori* mais informativas foram feitas diante do cálculo de medidas de resumo dos dados, tal como sugerimos na Seção 2.3.

### 2.4.1 Análise Assumindo Distribuição Weibull para os Tempos de Falha

Na análise apresentada nesta subseção ajustamos os modelos MTP, MBFC1, MBFC2 e MBFC3, assumindo que os tempos de falha possuem distribuição Weibull( $\alpha, \lambda$ ). Resumos das distribuições *a posteriori* de todos os parâmetros dos modelos ajustados são dados na Tabela 2.14. Os modelos MBFC1, 2 e 3 supõem erro de medida nas covariáveis, desse modo também exibimos estimativas para os hiperparâmetros da variável latente  $X_\epsilon$ :  $\mu_x$  e  $\sigma_x^2$ ; e a variância do erro  $\sigma_\epsilon^2$ . No MBFC1,  $\sigma_\epsilon^2$  é suposto conhecido e assumimos tal parâmetro fixado em 1.

Observando a Tabela 2.14, vemos que as estimativas obtidas para a variância do erro de medida  $\sigma_\epsilon^2$  ajustando-se os modelos MBFC2 e MBFC3 são, respectivamente, 0.094 e 0.311. A estimativa de  $\sigma_\epsilon^2$  via MBFC2 sugere que o erro de medida tende a ser razoavelmente pequeno. A média e variância amostral da covariável  $X_4$ , que supomos ter sido medida com erro, são 0.913 e 0.789, respectivamente. Comparando esses valores com as estimativas obtidas pra a média  $\mu_x$  e a variância  $\sigma_x^2$  das variáveis latentes, vemos que as estimativas obtidas para tais parâmetros ajustando-se o MBFC1 e supondo  $\sigma_\epsilon^2 = 1$ , são as que mais se aproximam desses valores. Por outro lado, as estimativas obtidas ajustando-se o MBFC2 e o MBFC3 diferem um pouco desses valores, indicando a presença de erro de medida. Para os parâmetros  $\beta$ ,  $\alpha$  e  $\lambda$ , as estimativas pontuais e intervalares fornecidas pelos modelos MTP e MBFC2 são muito próximas, de um modo geral. Para alguns parâmetros o MTP fornece intervalos menores e para outros isto acontece via MBFC2.

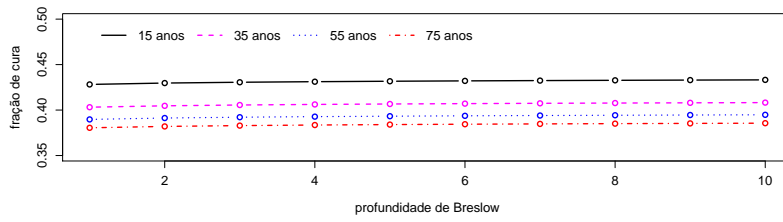
Nos modelos analisados, a interpretação dos coeficientes de regressão  $\beta$  é: coeficientes negativos levam a uma fração de cura maior do que aqueles positivos. Neste sentido, observando a Tabela 2.14 vemos que, sob todos os modelos, as estimativas pontuais de  $\beta_2$  a  $\beta_5$  são negativas, indicando que as covariáveis sexo, tratamento, profundidade do tumor e *status* físico devem levar a uma fração de cura maior. Por outro lado,  $\beta_1$  é positivo, indicando que idades maiores podem levar a frações de cura menores. Para entender melhor tal interpretação, consideremos a situação ilustrada na Figura 2.25 a seguir, na qual calculamos a fração de cura de pacientes com  $X_1 = 15, 35, 55$  e  $75$  anos de idade, do sexo feminino  $X_2 = 1$ , em tratamento por Interferon  $X_3 = 1$ , variando a profundidade do tumor supondo  $X_4 = 1, \dots, 10$  e que praticam atividades físicas  $X_5 = 1$ . Observando os gráficos dados na Figura 2.25, vemos que o único modelo cujos resultados não seguem tal interpretação é o MBFC3, cujo comportamento da fração de cura é tal que uma mulher com 15 anos apresenta uma fração de cura menor do que aquelas com 35 e

Tabela 2.14: Resumos das distribuições *a posteriori* dos parâmetros dos modelos MTP, MBFC1, 2 e 3 para os dados do estudo E1684, assumindo-se distribuição Weibull para os tempos de falha.

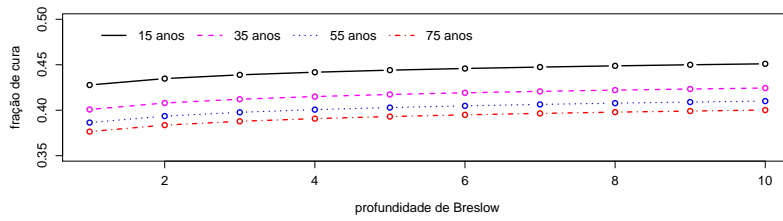
		Modelo			
$\eta$	Resumo	MTP	MBFC1	MBFC2	MBFC3
$\beta_1$	Média	0.081	0.087	0.080	0.083
	DP	0.042	0.074	0.045	0.049
	HPD 95%	(-0.003, 0.161)	(-0.051, 0.238)	(-0.012, 0.166)	(-0.014, 0.173)
$\beta_2$	Média	-0.017	-0.013	-0.015	-0.023
	DP	0.159	0.168	0.197	0.271
	HPD 95%	(-0.325, 0.294)	(-0.337, 0.316)	(-0.388, 0.382)	(-0.561, 0.504)
$\beta_3$	Média	-0.332	-0.336	-0.325	-0.334
	DP	0.163	0.161	0.157	0.162
	HPD 95%	(-0.636, -0.005)	(-0.656, -0.035)	(-0.630, -0.024)	(-0.646, -0.004)
$\beta_4$	Média	-0.006	-0.028	-0.003	-0.009
	DP	0.090	0.262	0.092	0.094
	HPD 95%	(-0.177, 0.171)	(-0.541, 0.483)	(-0.183, 0.168)	(-0.187, 0.179)
$\beta_5$	Média	-0.035	-0.050	-0.038	-0.041
	DP	0.266	0.278	0.269	0.269
	HPD 95%	(-0.537, 0.512)	(-0.579, 0.494)	(-0.558, 0.484)	(-0.588, 0.491)
$\alpha$	Média	1.003	1.009	1.006	1.006
	DP	0.067	0.065	0.065	0.066
	HPD 95%	(0.864, 1.124)	(0.884, 1.137)	(0.885, 1.135)	(0.879, 1.137)
$\lambda$	Média	-0.501	-0.516	-0.505	-0.507
	DP	0.114	0.113	0.114	0.117
	HPD 95%	(-0.722, -0.273)	(-0.745, -0.301)	(-0.730, -0.291)	(-0.749, -0.293)
$\mu_x$	Média	-	0.912	0.400	0.400
	DP	-	0.087	0.053	0.062
	HPD 95%	-	(0.736, 1.080)	(0.298, 0.502)	(0.285, 0.521)
$\sigma_x^2$	Média	-	0.781	0.606	0.635
	DP	-	0.062	0.043	0.048
	HPD 95%	-	(0.664, 0.904)	(0.523, 0.687)	(0.550, 0.731)
$\sigma_\epsilon^2$	Média	-	-	0.094	0.311
	DP	-	-	0.021	0.023
	HPD 95%	-	-	(0.056, 0.135)	(0.263, 0.353)

55 anos. Tal comportamento indica que as estimativas obtidas para os coeficientes de regressão do MBFC3 podem não seguir como o esperado por serem mais imprecisas.

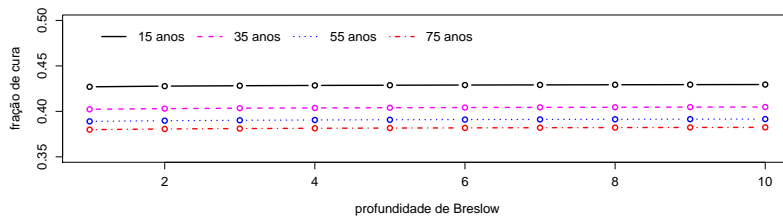
Ao compararmos os modelos com erro nas covariáveis vemos que os modelos MBFC1 e MBFC2 fornecem estimativas aproximadamente iguais apenas para os parâmetros  $\beta_2$  e  $\alpha$ . Por outro lado, os modelos MBFC2 e MBFC3 produzem estimativas aproximadamente iguais para os parâmetros  $\beta_1$ ,  $\beta_5$ ,  $\alpha$ ,  $\lambda$  e  $\mu_x$ . Embora perceba-se, em geral, que menores desvios padrão e intervalos de credibilidade são obtidos ajustando-se o MBFC2. Para compararmos melhor tais modelos, podemos analisar as estimativas obtidas para as variáveis latentes nas Figuras 2.26 à 2.28 a seguir, nas quais exibimos os intervalos de credibilidade de 95% para a variável latente dos 100 primeiros indivíduos, obtidos ajustando-se os modelos MBFC.



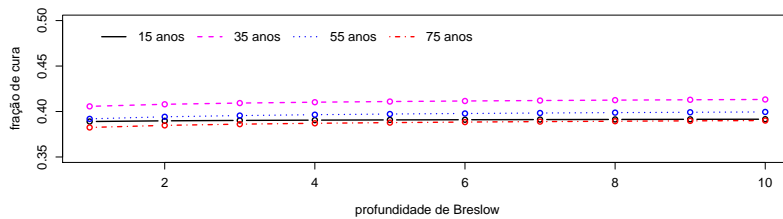
(a)



(b)



(c)



(d)

Figura 2.25: Fração de cura calculada para pacientes com idade  $X_1 = 15, 35, 55, 75$  anos, do sexo feminino ( $X_2 = 1$ ), submetidos ao tratamento ( $X_3 = 1$ ), com profundidade de Breslow  $X_4 = 1, \dots, 10$  e ativo  $X_5 = 1$ . Considerando as estimativas para os coeficientes de regressão obtidas ajustando-se o MTP (Gráfico a), o MBFC1 (Gráfico b), o MBFC2 (Gráfico c) e o MBFC3 (Gráfico d).

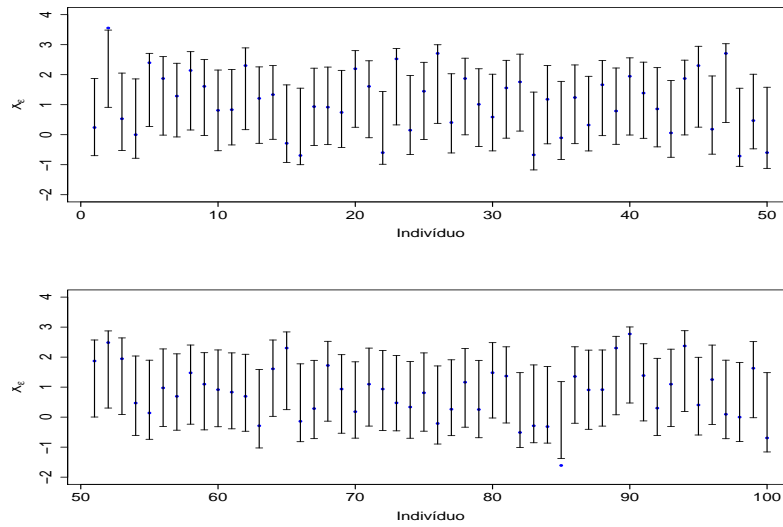


Figura 2.26: HPD's de 95% para a variável latente  $X_\epsilon$  dos primeiros 100 indivíduos, obtidos ajustando-se o MBFC1. Os pontos destacados são os valores observados da covariável  $X_4$  para cada um dos indivíduos.

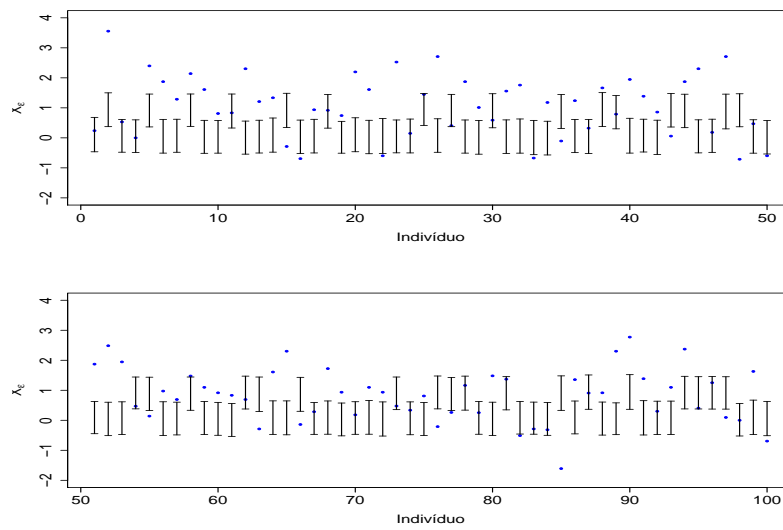


Figura 2.27: HPD's de 95% para a variável latente  $X_\epsilon$  dos primeiros 100 indivíduos, obtidos ajustando-se o MBFC2. Os pontos destacados são os valores observados da covariável  $X_4$  para cada um dos indivíduos.

Observando tais figuras podemos notar o comportamento das estimativas da variável latente para cada indivíduo baseado nos intervalos de credibilidade. Vemos que grande parte dos intervalos de credibilidade obtidos ao ajustarmos o MBFC1, contém o valor observado para a

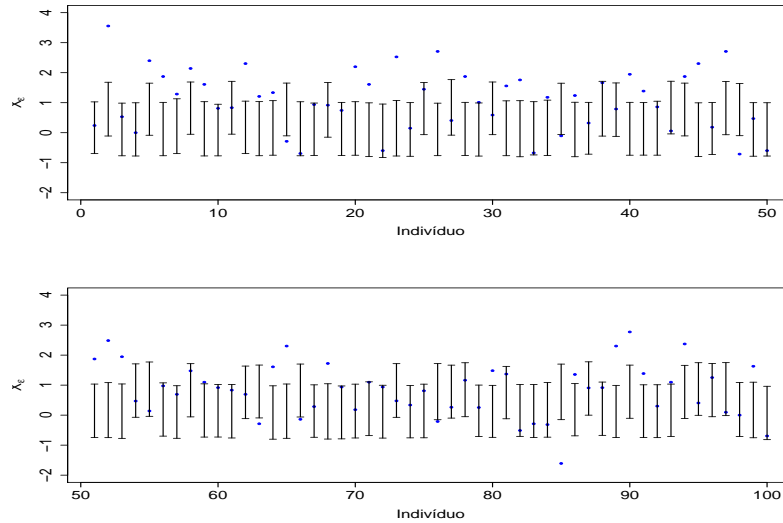


Figura 2.28: HPD's de 95% para a variável latente  $X_\epsilon$  dos primeiros 100 indivíduos, obtidos ajustando-se o MBFC3. Os pontos destacados são os valores observados da covariável  $X_4$  para cada um dos indivíduos.

covariável  $X_4$ , enquanto que ajustando-se o MBFC2 e o MBFC3 a maioria dos valores observados dessa covariável não pertencem aos HPD's 95% de  $X_\epsilon$ . Essa conclusão vai de encontro ao que concluímos ao analisarmos as estimativas obtidas para  $\mu_x$  e  $\sigma_x^2$ . O comprimento médio dos intervalos de credibilidade, considerando todos os indivíduos observados, é 2.585, 1.114 e 1.778, sob os modelos MBFC1, MBFC2 e MBFC3, respectivamente.

### Analisando o Efeito das Covariáveis na Fração de Cura

Ao analisarmos a Tabela 2.14, notamos que os fatores que devem ser mais relevantes no prognóstico do melanoma cutâneo nos pacientes do estudo E1684 são a idade e o tratamento por Interferon alpha-2b, representados pelas covariáveis  $X_1$  e  $X_3$ , respectivamente. O intervalo de credibilidade do parâmetro  $\beta_1$  apesar de conter o 0, possui limite inferior muito próximo desse valor, por esse motivo investigamos a relevância da covariável  $X_1$  na fração de cura dos indivíduos. A seguir, com o objetivo de avaliar como tais covariáveis afetam a probabilidade de cura dos indivíduos, exibimos nas Figuras 2.29 a 2.32 *box-plots* das estimativas para a média a *posteriori* da fração de cura dos indivíduos, considerando os quatro modelos ajustados.

Primeiramente, podemos observar na Figura 2.29 que as estimativas para a fração de cura dos indivíduos sob os modelos MBFC1, 2, 3 e o MTP se distribuem de modo bastante semelhante. As estimativas obtidas com os modelos MBFC2 e MTP são as que apresentam uma maior similaridade de comportamento.

Na Figura 2.30 a seguir, exibimos *box-plots* das estimativas para a fração de cura obtidas



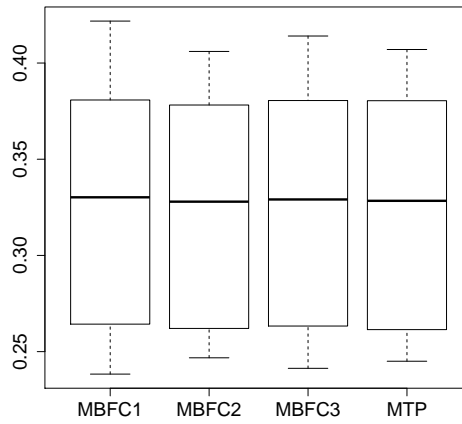


Figura 2.29: *Box-plots* das estimativas para a média *a posteriori* da fração de cura de todos os indivíduos observados, considerando os ajustes dos modelos MBFC1, 2 e 3 e MTP, respectivamente.

ajustando-se cada um dos quatro modelos analisados, considerando separadamente os indivíduos que receberam tratamento por Interferon alpha-2b, para os quais a covariável  $X_3$  é igual a 1, daqueles que ficaram apenas em observação ( $X_3 = 0$ ).

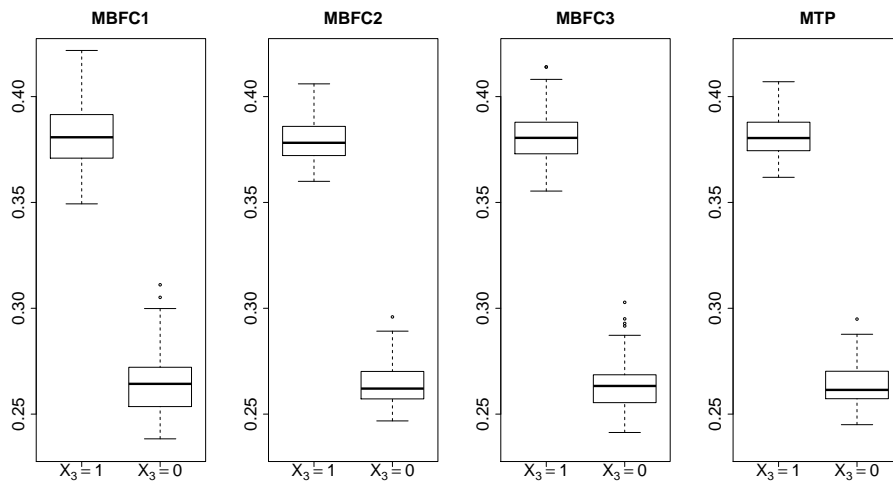


Figura 2.30: *Box-plots* das estimativas para a média *a posteriori* da fração de cura de todos os indivíduos observados que receberam doses de Interferon ( $X_3 = 1$ ) ou que ficaram apenas em observação ( $X_3 = 0$ ), considerando os ajustes dos modelos MBFC1, 2, 3 e MTP, respectivamente.

Ao observarmos os *box-plots* na Figura 2.30, vemos que o tratamento por Interferon teve

um efeito positivo nas estimativas da probabilidade de cura dos pacientes. Os indivíduos em tratamento apresentaram uma estimativa de fração de cura de no mínimo 35%, enquanto que as estimativas entre os indivíduos que permaneceram apenas em observação é no máximo 32%, aproximadamente. Os resultados sugerem que o tratamento com o uso da droga pode ser eficiente no aumento da probabilidade de cura.

Para verificarmos se a covariável  $X_1$ , a qual representa a idade dos indivíduos, é de fato significativa para o tempo de vida dos pacientes, dividimos a amostra observada em cinco faixas etárias e analisamos as estimativas obtidas para a fração de cura em cada uma delas. As faixas de idade, em anos, consideradas são: de 18 a 30.2, de 30.2 a 42.4, de 42.4 a 54.6, de 54.6 a 66.8 e de 66.8 a 79 anos.

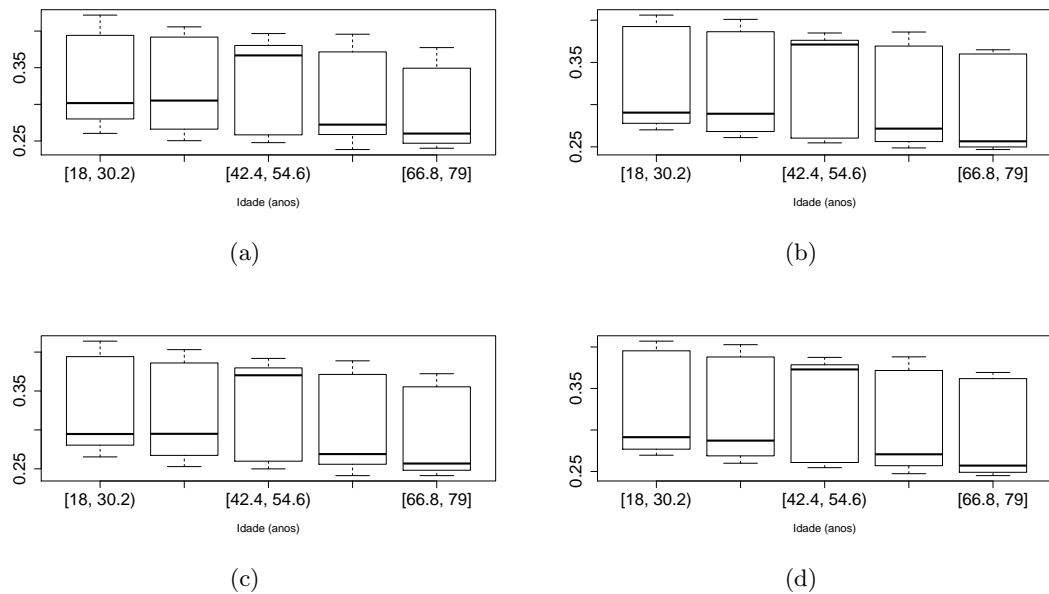


Figura 2.31: *Box-plots* das estimativas para a média *a posteriori* da fração de cura de todos os indivíduos observados para cada faixa etária considerada, sob o ajuste dos modelos MBFC1 (Gráfico a), 2 (Gráfico b), 3 (Gráfico c) e MTP (Gráfico d).

Observando os *box-plots* na Figura 2.31, percebemos que parece haver uma certa redução na probabilidade de cura a medida que a idade dos indivíduos aumenta, embora essa redução não ultrapasse no máximo 4%. Vamos analisar também, como se comportam as estimativas para fração de cura dos indivíduos classificados por faixa etária, separados entre os que fizeram tratamento com Interferon e aqueles que não receberam o tratamento. Vejamos os *box-plots* na Figura 2.32 a seguir. Constatamos que, para todos os modelos ajustados, não existe uma diferença considerável entre o comportamento das estimativas da fração de cura observadas por

faixa etária nesse caso. Indicando que pode não haver interação entre as covariáveis idade e tratamento.

Também estudamos como se comportam as estimativas obtidas para a fração de cura sob cada modelo, dentre os indivíduos que receberam doses de Interferon e aqueles que estiveram somente em observação, examinando-os de acordo com o sexo (Feminino – F e Masculino – M) e também em relação ao tipo do melanoma segundo o índice de Breslow que classifica a doença como: tipo *I* se  $X_4 < 1mm$ , tipo *II* se  $1mm \leq X_4 < 2mm$ , do tipo *III* se  $2mm \leq X_4 < 4mm$  e tipo *IV* se  $X_4 \geq 4$ .

Na Figura 2.33, podemos notar que embora em média (e mediana) os indivíduos do sexo masculino apresentem estimativas para a fração de cura menores do que aqueles do sexo feminino, tanto entre os indivíduos que fizeram o tratamento com a droga quanto entre aqueles que não fizeram, não percebemos uma diferença muito substancial. Esta conclusão, segue em concordância com o que é indicado pelas estimativas intervalares apresentadas na Tabela 2.14. No que diz respeito as estimativas obtidas para a fração de cura dos indivíduos classificados segundo o tipo do melanoma, observemos os *box-plots* apresentados na Figura 2.34 a seguir. Em tais gráficos, notamos que o tipo do melanoma, não parece representar uma influência significativa na fração de cura, visto que o comportamento das estimativas obtidas, sob todos os modelos ajustados, é aproximadamente igual.

Para analisar qual dos modelos se ajusta melhor aos dados, lançamos mão de algumas das medidas de seleção de modelos Bayesianos mais utilizadas em aplicações na literatura: a *Conditional Predictive Ordinate* (CPO), o *Deviance Information Criterion* (DIC) e o *Expected Bayesian Information Criterion* (EBIC). Na Tabela 2.15, exibimos o valor dessas medidas para cada modelo, na qual o LPML representa o resumo do logaritmo natural da CPO. Além disso, na Figura 2.15 podemos analisar a robustez dos modelos propostos através do diagnóstico de pontos influentes via divergência de Kullbach-Leibler. Detalhes a respeito dessas medidas podem ser vistas em Cancho *et al.* (2011); Gelman *et al.* (2013); Oliveira (2015); Spiegelhalter *et al.* (2002) e no Apêndice C desse texto. Em resumo, temos que quanto maior o LPML melhor é considerado o ajuste, enquanto que sob o DIC e EBIC valores menores indicam maior qualidade. Em relação a divergência de Kullbach-Leibler, analisamos como observações não influentes aquelas cujas linhas no gráfico não ultrapassam 0.5 no eixo das ordenadas.

Na Tabela 2.15 podemos observar que todas as medidas de qualidade de ajuste avaliam o MTP como sendo aquele que melhor ajusta-se aos dados. Entre os modelos com erro de medida, o MBFC2 é apontado por todos os critérios de comparação, como sendo o mais adequado. Na Figura 2.35, exibimos o gráfico da divergência K-L para todos os modelos. Ao observarmos os gráficos em tal figura percebemos que o modelo que apresenta uma maior quantidade de observações potencialmente influentes é o MBFC2. Esses gráficos indicam que para o banco

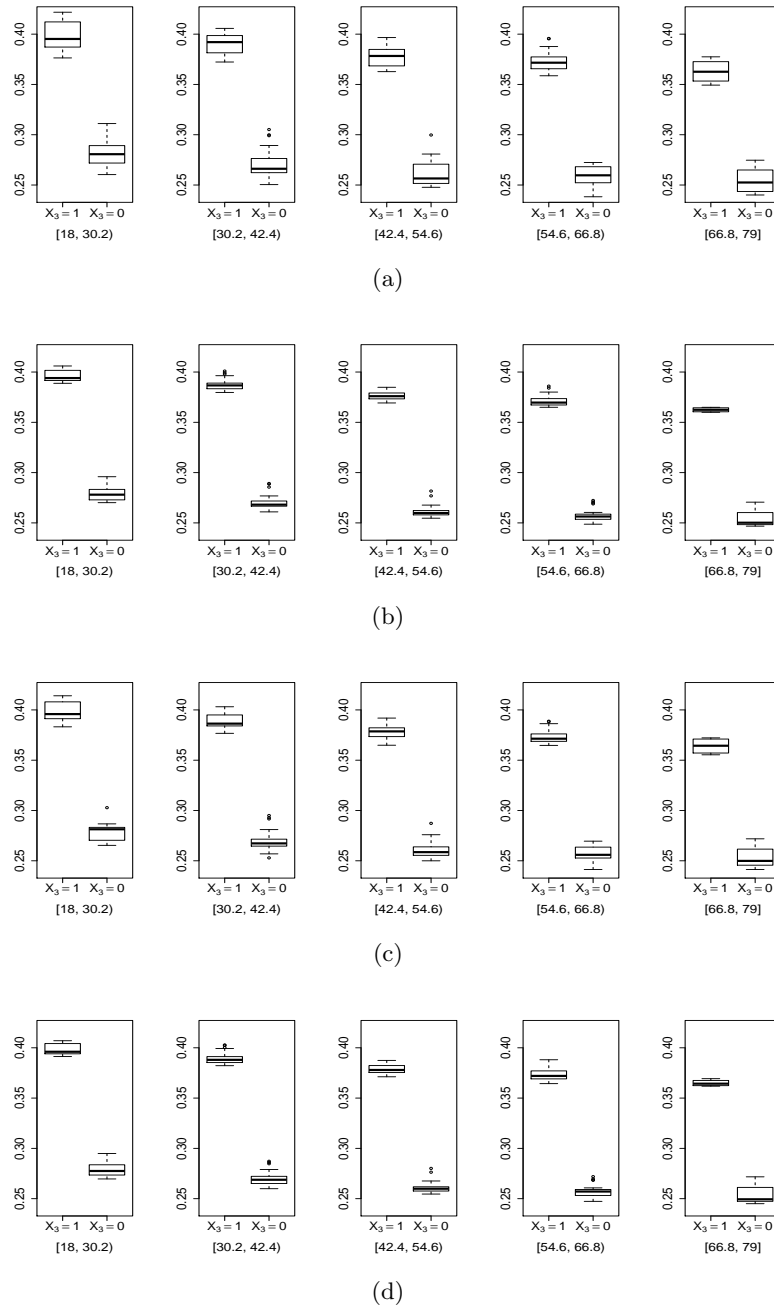


Figura 2.32: *Box-plots* das estimativas para a média *a posteriori* da fração de cura de todos os indivíduos observados para cada faixa etária considerada divididas entre os indivíduos que receberam Interferon ( $X_3 = 1$ ) e os que permaneceram apenas em observação ( $X_3 = 0$ ), sob o ajuste dos modelos MBFC1 (Gráfico a), 2 (Gráfico b), 3 (Gráfico c) e MTP (Gráfico d).

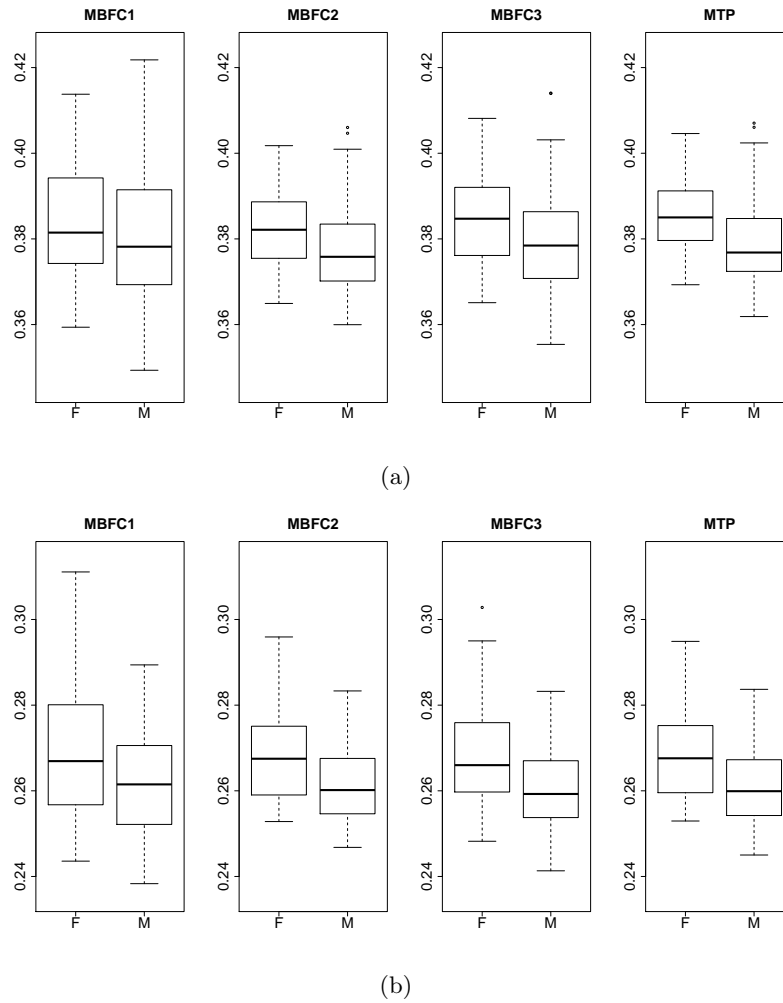
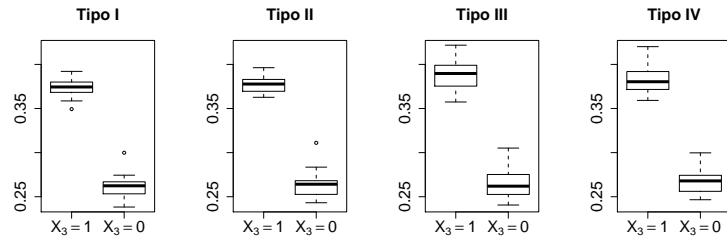


Figura 2.33: *Box-plots* das estimativas para a média *a posteriori* da fração de cura de todos os indivíduos observados que receberam doses de Interferon (Gráfico a) ou que ficaram somente em observação (Gráfico b) e são do sexo feminino (F) ou do sexo masculino (M), considerando cada um dos modelos ajustados.

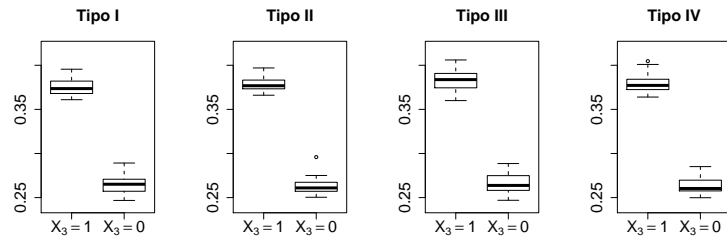
Tabela 2.15: Comparação dos Modelos ajustados aos dados do estudo E1684.

Modelo	Critério		
	LPML	DIC	EBIC
MTP	-75.155	141.663	180.243
MBFC1	-505.235	939.913	2162.769
MBFC2	-235.612	409.352	1544.395
MBFC3	-303.542	615.447	1797.354

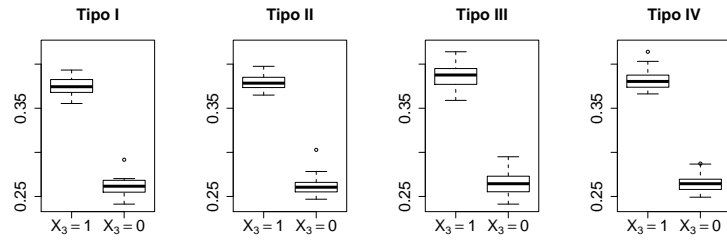
de dados analisado o modelo mais adequado é o MTP, para o qual não observamos a presença de pontos influentes. Entre os modelos com erro nas covariáveis, a opção cuja a presença de



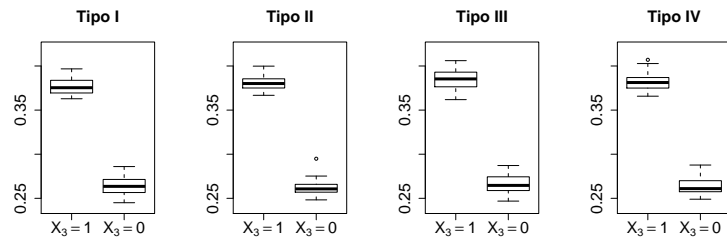
(a)



(b)



(c)



(d)

Figura 2.34: *Box-plots* das estimativas para a média *a posteriori* da fração de cura de todos os indivíduos observados que receberam doses de Interferon e possuem nódulo do Tipo I, II, III e IV, considerando os ajustes dos modelos MBFC1 (Gráfico a), 2 (Gráfico b), 3 (Gráfico c) e MTP (Gráfico d).

pontos potencialmente influentes é menor é o MBFC3. Portanto, tal modelo aparece como o mais apropriado via divergência K-L. Vale salientarmos que em toda a análise feita até aqui, os modelos MBFC2 e MTP ao serem ajustados forneceram resultados muito similares.

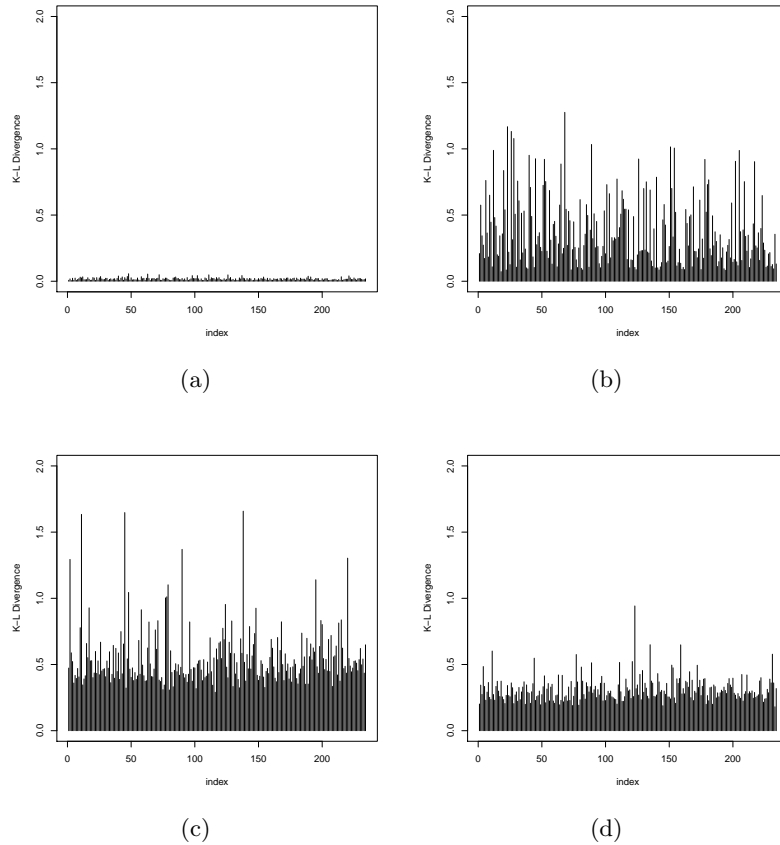


Figura 2.35: Divergência K-L. Os gráficos (a), (b), (c) e (d) são relativos aos ajustes dos modelos MTP, MBFC1, 2 e 3, respectivamente.

Para analisarmos a fração de cura estimada por indivíduo, exibimos na Figura 2.36 as curvas das distribuições *a posteriori*, considerando os primeiros 30 observados. Em cada gráfico temos as curvas que foram obtidas ajustando-se os modelos MTP, MBFC1, MBFC2 e MBFC3. Nestes gráficos vemos que os modelos MTP e MBFC2 produzem distribuições aproximadamente iguais para a fração de cura, e o modelo que em geral fornece uma distribuição mais distante das demais é o MBFC1.

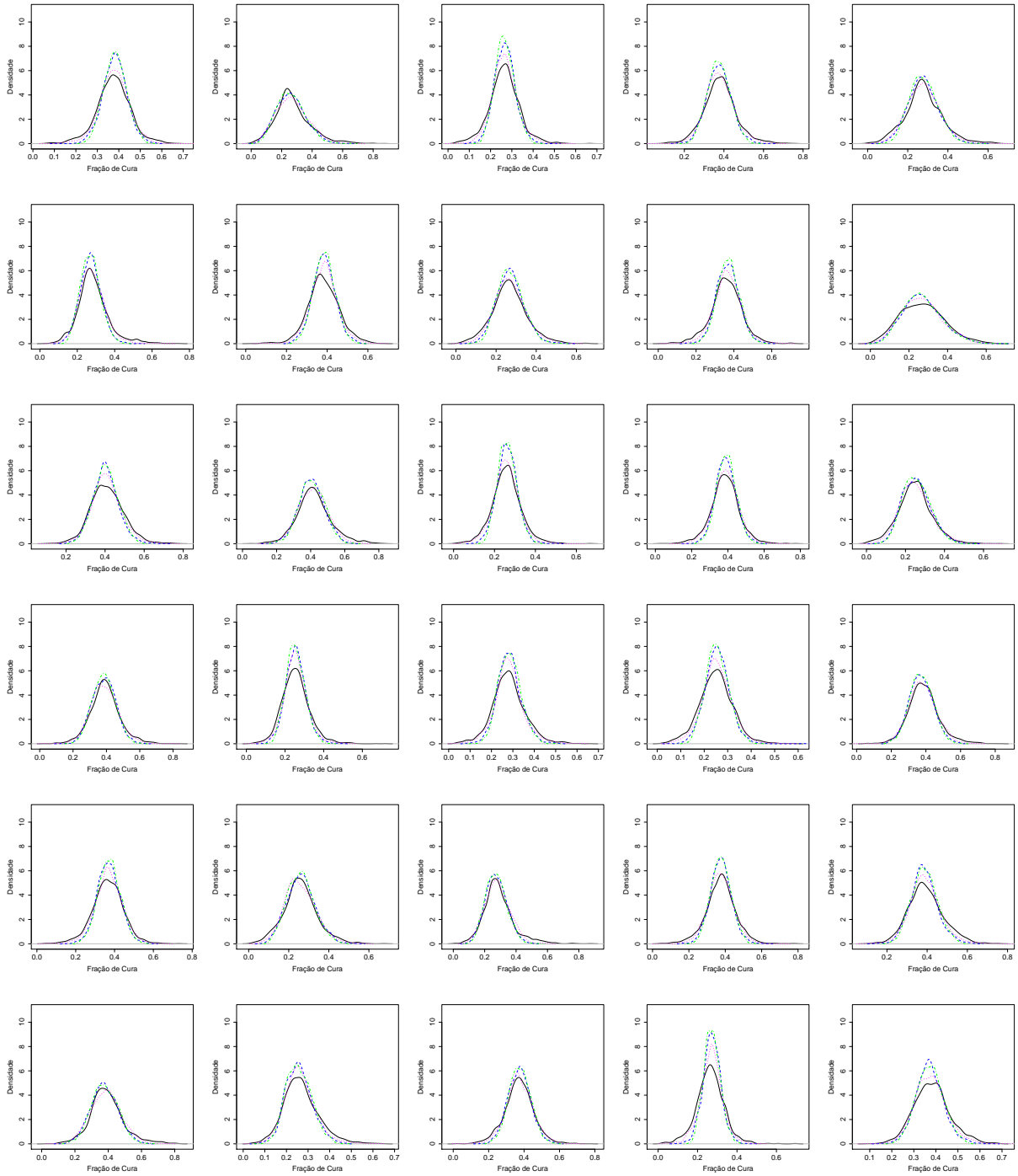


Figura 2.36: Distribuições *a posteriori* para a fração de cura. As linhas mista, sólida, tracejada e pontilhada são das distribuições obtidas ajustando-se os modelos MTP, MBFC1, 2 e 3, respectivamente.



### 2.4.2 Análise Assumindo Distribuição Exponencial para os Tempos de Falha

Nesta subseção, apresentamos uma análise dos dados do estudo E1684, ajustando os modelos MTP, MBFC1, MBFC2, MBFC3 e a versão Bayesiana d proposta de Ma e Yin (2008), a qual apresentamos na Seção 2.2. Em todas os modelos ajustados supomos que os tempos de falha tem distribuição Weibull( $1, \lambda$ ), que é equivalente a supormos uma distribuição Exponencial( $\lambda$ ). Novamente fixamos  $\sigma_\epsilon^2 = 1$  no MBFC1. Na Tabela 2.16, exibimos resumos das distribuições *a posteriori* dos parâmetros dos modelos ajustados.

Tabela 2.16: Resumos das distribuições *a posteriori* dos parâmetros dos modelos MTP, MBFC1, 2, 3 e do MBFC-Ma para os dados do estudo E1684, assumindo-se distribuição Exponencial para os tempos de falha.

$\eta$	Resumo	Modelo				
		MTP	MBFC1	MBFC2	MBFC3	MBFC-MA
$\beta_1$	Média	0.079	0.087	0.080	0.084	0.080
	DP	0.043	0.072	0.044	0.049	0.044
	HPD 95%	(-0.006, 0.161)	(-0.059, 0.217)	(-0.009, 0.164)	(-0.018, 0.173)	(-0.002, 0.171)
$\beta_2$	Média	-0.007	-0.014	-0.009	-0.033	-0.010
	DP	0.168	0.169	0.193	0.278	0.192
	HPD 95%	(-0.364, 0.289)	(-0.328, 0.329)	(-0.386, 0.361)	(-0.530, 0.553)	(-0.372, 0.385)
$\beta_3$	Média	-0.329	-0.338	-0.329	-0.333	-0.328
	DP	0.163	0.166	0.163	0.163	0.162
	HPD 95%	(-0.644, -0.012)	(-0.663, -0.021)	(-0.668, -0.023)	(-0.655, -0.018)	(-0.663, -0.025)
$\beta_4$	Média	-0.007	-0.027	-0.004	-0.008	-0.005
	DP	0.092	(-0.518, 0.454)	0.090	0.091	0.090
	HPD 95%	(-0.175, 0.174)	0.249	(-0.183, 0.172)	(-0.180, 0.172)	(-0.176, 0.180)
$\beta_5$	Média	-0.045	-0.049	-0.044	-0.042	-0.045
	DP	0.267	0.278	0.272	0.272	0.276
	HPD 95%	(-0.567, 0.439)	(-0.592, 0.507)	(-0.569, 0.498)	(-0.559, 0.469)	(-0.595, 0.481)
$\lambda$	Média	-0.493	-0.508	-0.499	-0.504	-0.502
	DP	0.112	0.114	0.113	0.113	0.110
	HPD 95%	(-0.714, -0.287)	(-0.750, -0.314)	(-0.727, -0.282)	(-0.734, -0.300)	(-0.719, -0.283)
$\mu_x$	Média	-	0.913	0.397	0.396	0.397
	DP	-	0.087	0.055	0.066	0.056
	HPD 95%	-	(0.742, 1.079)	(0.289, 0.504)	(0.262, 0.516)	(0.291, 0.508)
$\sigma_x^2$	Média	-	0.778	0.605	0.634	0.605
	DP	-	0.062	0.042	0.045	0.043
	HPD 95%	-	(0.657, 0.896)	(0.522, 0.688)	(0.551, 0.726)	(0.524, 0.692)
$\sigma_\epsilon^2$	Média	-	-	0.095	0.311	0.095
	DP	-	-	0.021	0.024	0.022
	HPD 95%	-	-	(0.056, 0.136)	(0.263, 0.355)	(0.058, 0.141)

Ao observarmos a Tabela 2.16, notamos que os modelos com erro de medida nas covariáveis fornecem estimativas semelhantes aquelas obtidas na subseção anterior para a variância do erro  $\sigma_\epsilon^2$ , de modo que é possível concluirmos que se de fato houve erro de medição para a covariável  $X_4$ , esse erro tende a ser pequeno, pois, a estimativa obtida via MBFC2 para a média *a posteriori* de

$\sigma_\epsilon^2$  é aproximadamente 0.1. Sob ao MBFC2 e o MBFC-Ma temos estimativas aproximadamente iguais para todos os parâmetros. Para  $\beta$  e  $\lambda$  que são comuns a todos os modelos ajustados, o MTP, MBFC2 e MBFC-Ma produziram estimativas bem similares. Os modelos MBFC1 e MBFC3 são os que apresentam estimativas com maior diferença, principalmente, para os parâmetros  $\beta_1$  e  $\beta_2$ .

Em relação a inferência sobre as variáveis latentes, os modelos com erro nas covariáveis, com exceção do MBFC1, forneceram estimativas aproximadamente iguais para a média  $\mu_x$  e a variância  $\sigma_x^2$ . O MBFC1, no qual fixamos variância do erro em 1, que é um valor significativamente maior do que aquele estimado para tal variância pelos demais modelos com erro nas covariáveis, produziu estimativas para  $\mu_x$  e  $\sigma_x^2$  aproximadamente iguais a média e variância amostral de  $X_4$ . Isso pode sugerir que não há erro nas covariáveis, contrariando o que a suposição do modelo.

Para analisarmos o comportamento da distribuição da distribuição *a posteriori* das variáveis latentes, exibimos nas Figuras 2.37 a 2.40 os intervalos de credibilidade obtidos ajustando-se os MBFCs, para os 100 primeiros indivíduos observados. Nestas figuras, destacamos o valor observado para a covariável  $X_4$  e podemos notar que os a maioria dos intervalos de credibilidades obtidos para o MBFC1, exibidos na Figura 2.37, contém o valor de  $X_4$  observado para cada indivíduo, corroborando às estimativas obtidas para a média e variância das variáveis latentes. Opostamente ao comportamento observado para o MEFC1, os modelos MBFC2, MBFC3 e MBFC-Ma fornecem intervalos de credibilidade para  $X_\epsilon$  que não contém o valor observado  $X_4$  para a maioria dos indivíduos (ver Figuras 2.38 à 2.40), apontando a presença de erro de medida na covariável  $X_4$ , tal como podemos conjecturar ao analisar as estimativas produzidas para  $\mu_x$  e  $\sigma_x^2$ .

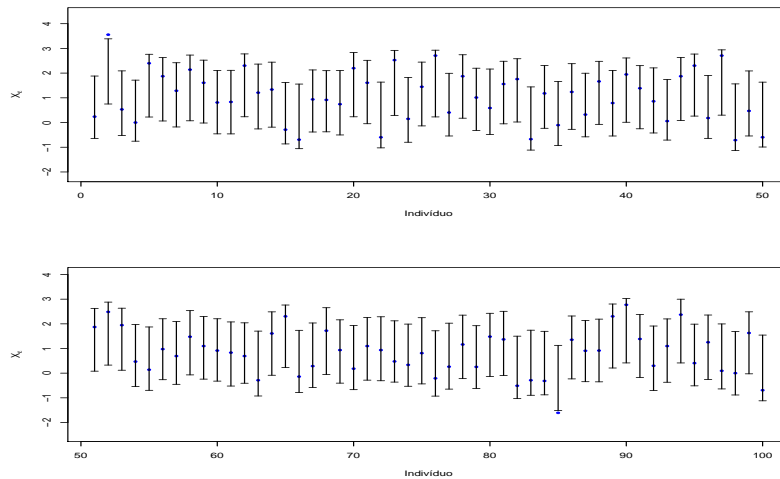


Figura 2.37: HPD's de 95% para a variável latente  $X_\epsilon$  dos primeiros 100 indivíduos, obtidos ajustando-se o MBFC1. Os pontos destacados são os valores observados da covariável  $X_4$  para cada um dos indivíduos.

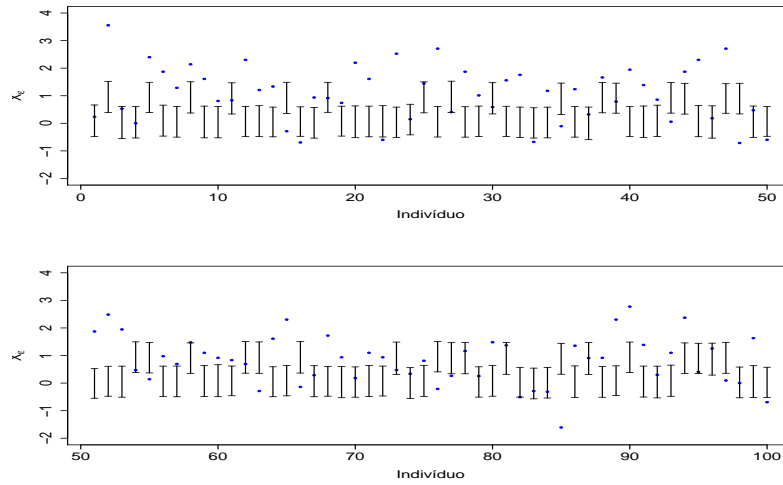


Figura 2.38: HPD's de 95% para a variável latente  $X_\epsilon$  dos primeiros 100 indivíduos, obtidos ajustando-se o MBFC2. Os pontos destacados são os valores observados da covariável  $X_4$  para cada um dos indivíduos.

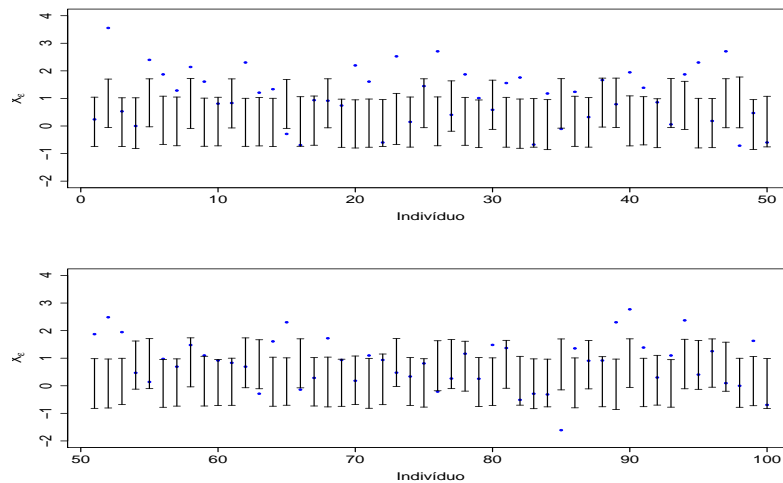


Figura 2.39: HPD's de 95% para a variável latente  $X_\epsilon$  dos primeiros 100 indivíduos, obtidos ajustando-se o MBFC3. Os pontos destacados são os valores observados da covariável  $X_4$  para cada um dos indivíduos.

### Analisando o Efeito das Covariáveis na Fração de Cura

Ao analisarmos as estimativas apresentadas na Tabela 2.16 percebemos que considerando a distribuição Exponencial, sob todos os modelos, o efeito da covariável  $X_3$  é significativo. Tal covariável representa se o indivíduo recebeu ou não o tratamento por Interferon. Tendo isso em vista, nas Figuras 2.42 a 2.45 estudamos o efeito dessa covariável no percentual de cura

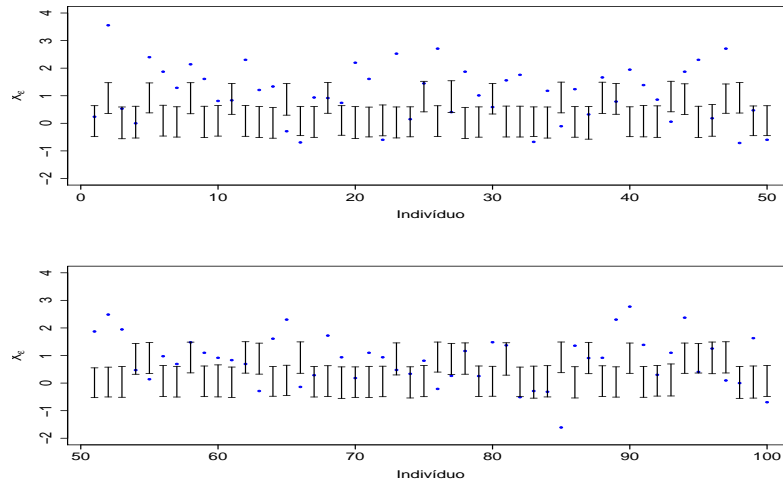


Figura 2.40: HPD's de 95% para a variável latente  $X_e$  dos primeiros 100 indivíduos, obtidos ajustando-se o MBFC-Ma. Os pontos destacados são os valores observados da covariável  $X_4$  para cada um dos indivíduos.

estimado para os indivíduos em geral, fazendo distinção entre os sexos, considerando o tipo do tumor segundo o índice de Breslow que classifica os nódulos, como especificamos na Subseção 2.4.1, e também por faixa etária. Para efeitos comparativos, na Figura 2.41 podemos observar o comportamento das estimativas da fração de cura ajustando-se os cinco modelos analisados, para todos os indivíduos irrestritamente.

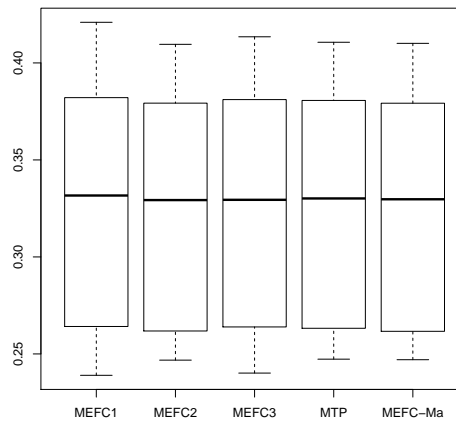


Figura 2.41: *Box-plots* das estimativas para a média *a posteriori* da fração de cura de todos os indivíduos observados, considerando os ajustes dos modelos MBFC1, 2 e 3, MTP e MBFC-Ma, respectivamente.

Na Figura 2.42 a seguir, vemos que sob todos os modelos ajustados, os indivíduos que receberam o tratamento por Interferon alpha-2b apresentaram uma fração de cura maior do que aqueles que permaneceram apenas sendo observados, mostrando que a covariável  $X_3$  influencia a fração de cura dos indivíduos estudados e, portanto, o uso do tratamento parece ser recomendado.

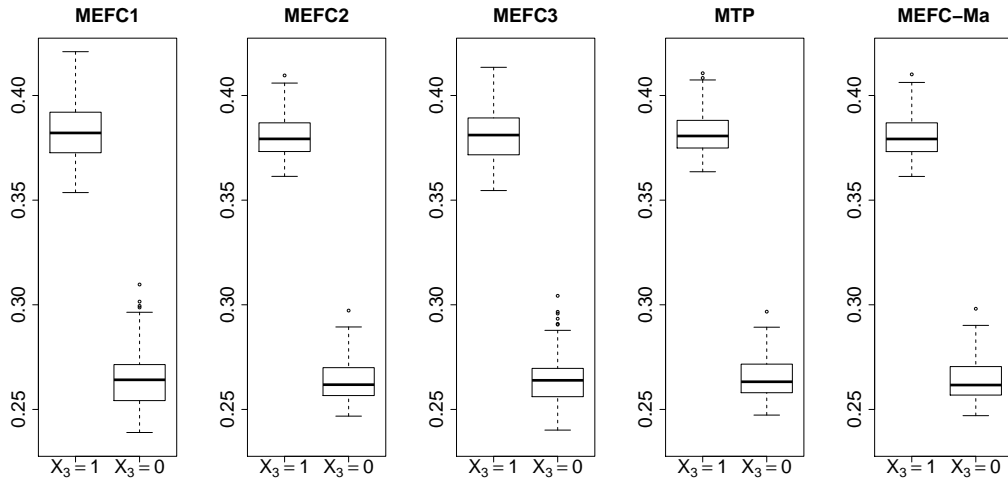


Figura 2.42: *Box-plots* das estimativas para a média *a posteriori* da fração de cura de todos os indivíduos observados que receberam doses de Interferon ( $X_3 = 1$ ) ou que ficaram apenas em observação ( $X_3 = 0$ ), considerando os ajustes dos modelos MBFC1, 2 e 3, MTP e MBFC-Ma, respectivamente.

Nos *box-plots* exibidos na Figuras 2.43 e 2.44 podemos observar, para cada modelo ajustado, como as estimativas da fração de cura dos indivíduos que receberam doses de Interferon (Fig.2.43) ou não (Fig.2.44) se comportam segundo o sexo, o qual é indicado pela covariável  $X_2$ . Ao observarmos tais *box-plots* concluímos que não há uma diferença substancial na fração de cura quando avaliamos os indivíduos classificados por sexo. Para todos os modelos ajustados as estimativas obtidas para indivíduos dos sexos feminino e masculino, olhados separadamente, não apresentam um comportamento consideravelmente diferente daquele na Figura 2.42 para os indivíduos sem distinção de sexo. Entretanto, sob todos os modelos ajustados e para os indivíduos em tratamento ou apenas em observação, aqueles do sexo masculino apresentam uma fração de cura em média (e mediana) um pouco mais baixa que os indivíduos do sexo feminino.

Para avaliar se o tipo do nódulo influencia o prognóstico do melanoma, nos gráficos exibidos na Figura 2.45, podemos avaliar o comportamento das estimativas da fração de cura dos indivíduos classificados em tratamento por Interferon ( $X_3 = 1$ ) ou apenas em observação ( $X_3 = 0$ ) para cada tipo de nódulo, sob cada modelo ajustado.

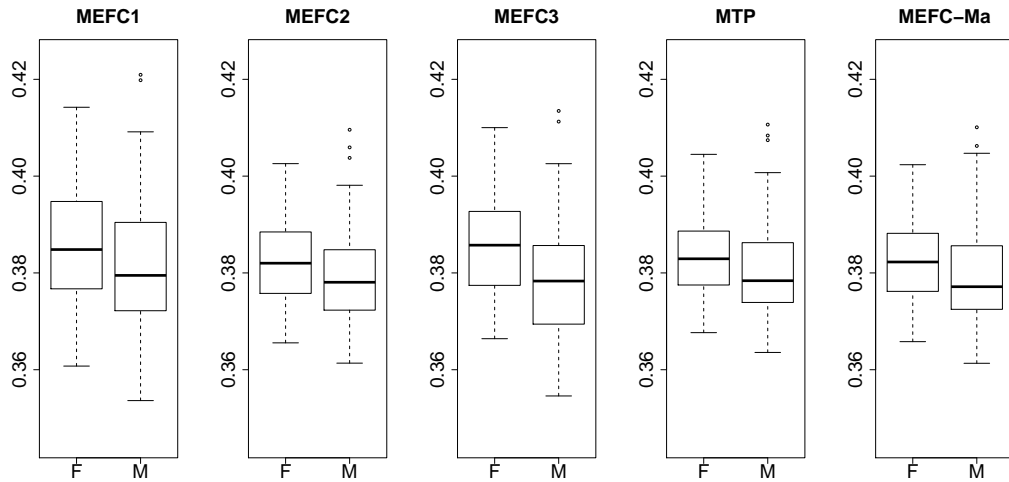


Figura 2.43: *Box-plots* das estimativas para a média *a posteriori* da fração de cura de todos os indivíduos observados que receberam doses de Interferon e são do sexo feminino (F) ou do sexo masculino (M), considerando cada um dos modelos ajustados.

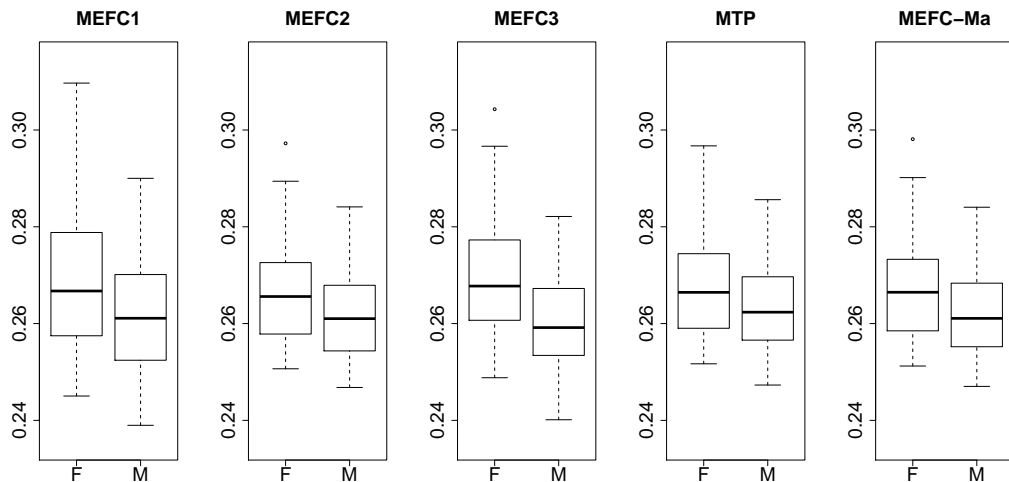
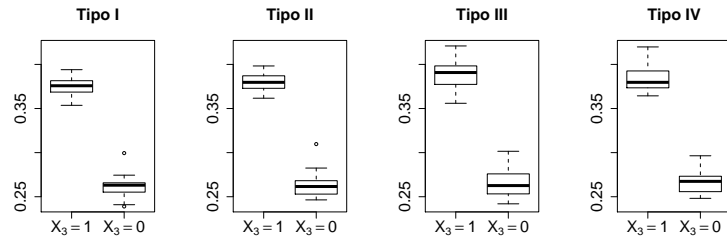
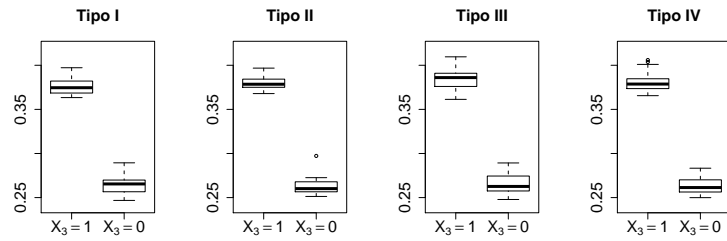


Figura 2.44: *Box-plots* das estimativas para a média *a posteriori* da fração de cura de todos os indivíduos observados que permaneceram apenas em observação e são do sexo feminino (F) ou do sexo masculino (M), considerando cada um dos modelos ajustados.

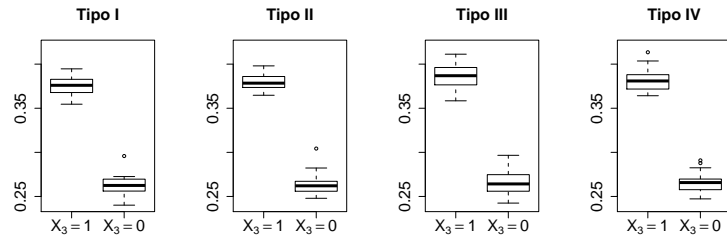
Observando os *box-plots* exibidos na Figura 2.45, vemos, em todos os modelos ajustados, que o comportamento das estimativas obtidas para a fração de cura não sofre alteração significativa em função do tipo de nódulo dos pacientes participantes do estudo E1684.



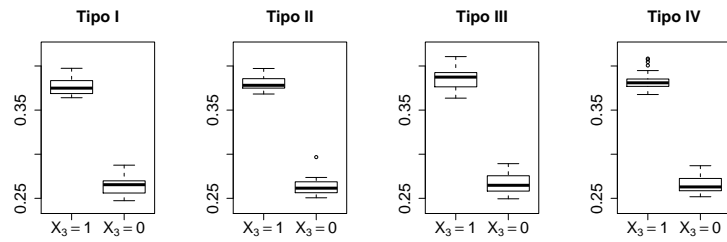
(a)



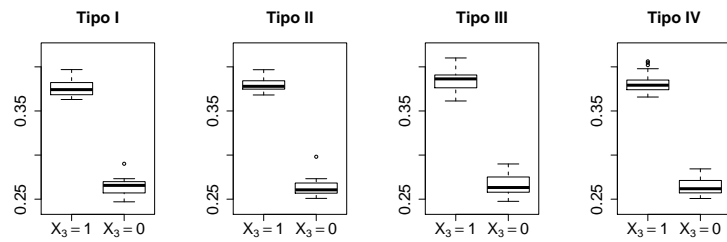
(b)



(c)



(d)



102

(e)

Figura 2.45: *Box-plots* das estimativas para a média *a posteriori* da fração de cura de todos os indivíduos observados que receberam doses de Interferon e possuem nódulo do Tipo I, II, III e IV, considerando os ajustes dos modelos MBFC1 (Gráfico a), 2 (Gráfico b) e 3 (Gráfico c), MTP (Gráfico d) e MBFC-Ma (Gráfico e).

Outro fator que suspeitamos ser significativo, ao analisarmos a Tabela 2.16, é a idade dos indivíduos ( $X_1$ ), visto que para todos os modelos os intervalos de credibilidade obtidos para o parâmetro  $\beta_1$  possuem o limite inferior muito próximo a zero, com limite superior positivo. Por essa razão, com o objetivo de analisar a relevância de tal covariável para o tempo de vida dos indivíduos, apresentamos nas Figuras 2.46 e 2.47 a seguir, *box-plots* das estimativas obtidas para a fração de cura dos indivíduos divididos por faixa etária, e também classificados por faixa etária e tratamento (de acordo com o valor observado para a covariável  $X_3$ ), considerando os cinco modelos ajustados. As faixas etárias consideradas são especificadas pelos intervalos: de 18 a 30.2 anos, de 30.2 a 42.4 anos, de 42.4 a 54.6 anos, de 54.6 a 66.8 anos e de 66.8 a 79 anos.

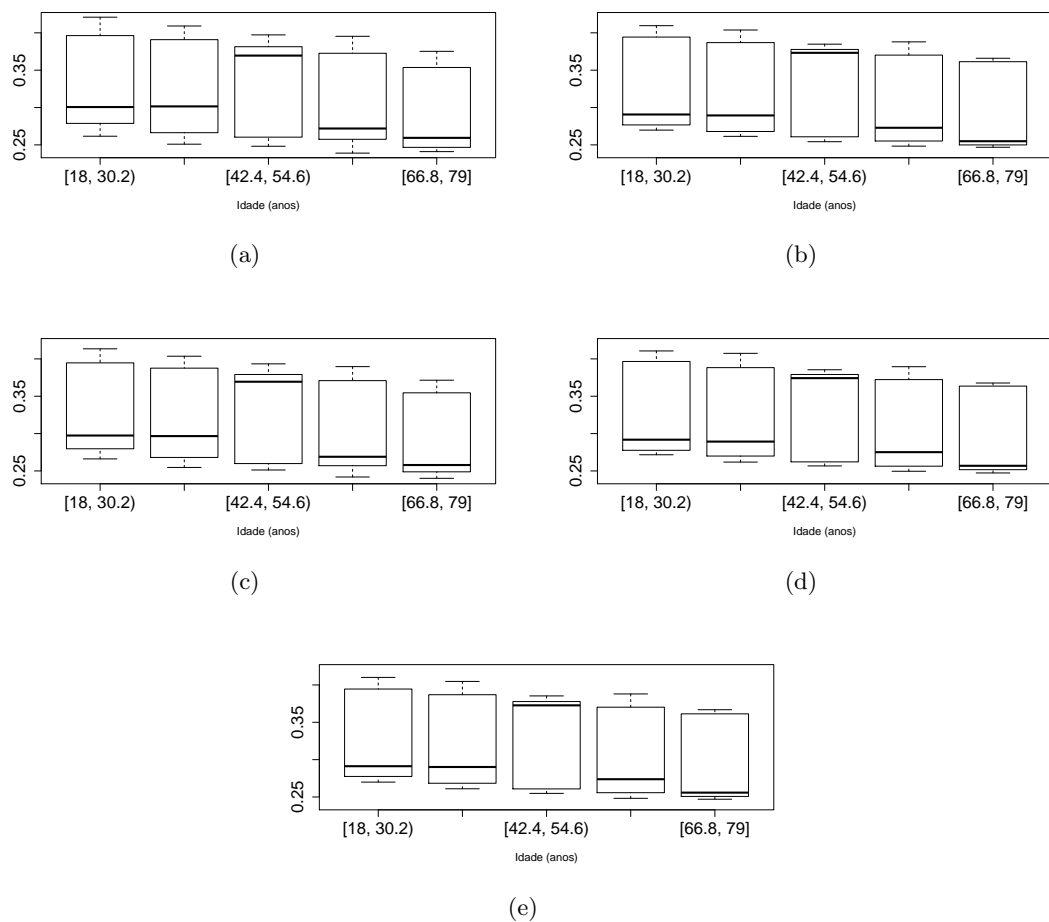


Figura 2.46: *Box-plots* das estimativas para a média *a posteriori* da fração de cura de todos os indivíduos observados para cada faixa etária considerada, sob o ajuste dos modelos MEFC1 (Gráfico a), 2 (Gráfico b) e 3 (Gráfico c), MTP (Gráfico d) e MEFC-Ma (Gráfico e).



Comparando os *box-plots* da Figura 2.46, nos quais temos as estimativas para a fração de cura dos indivíduos classificados segundo a faixa etária, vemos que há uma diminuição na estimativa de cura a medida que a idade avança. Tal comportamento sugere que a covariável  $X_1$ , para influenciar o tempo de vida dos pacientes observados, de modo que quanto maior é a idade do paciente, menor a probabilidade de cura.

Julgando que as covariáveis  $X_1$  e  $X_3$  são significativas para o tempo de vida dos indivíduos com melanoma cutâneo, na Figura 2.47 vemos como as estimativas da fração de cura se comportam quando observamos os indivíduos classificados por faixa etária e tratamento, notando que o efeito do tratamento por Interferon alpha-2b pode diminuir um pouco para os indivíduos com 42.4 anos ou mais. Isto pode ser observado para todos os modelos ajustados.

Para avaliar os ajustes fornecidos por MTP, MBFC1, MBFC2, MBFC3 e MBFC-Ma para os dados E1684, segundo alguns critérios de comparação de modelos, na Tabela 2.17 exibimos as medidas LPML, DIC e EBIC. Na análise apresentada na presente subseção o modelo eleito pelos critérios de avaliação de ajuste como o mais adequado é o MTP. Dentre aqueles que supõem erro de medida, o que melhor se ajusta aos dados é o MBFC-Ma, seguido do MBFC2, para todos os critérios de avaliação. É importante lembrarmos que o MTP, MBFC2 e MBFC-Ma apresentaram estimativas aproximadamente iguais para todos os parâmetros que são comuns a esses três modelos, entretanto há uma grande diferença entre os valores das medidas de comparação que foram obtidas sob cada um deles.

Tabela 2.17: Comparação dos Modelos ajustados aos dados do estudo E1684.

Modelo	Critério		
	LPML	DIC	EBIC
MTP	-73.211	140.124	173.218
MBFC1	-501.326	937.686	2155.457
MBFC2	-233.724	405.841	1535.035
MBFC3	-303.465	610.599	1787.216
MBFC-Ma	-161.585	274.050	1402.749

Na Figura 2.48 apresentamos os gráficos da divergência K-L para cada um dos modelos ajustados. Ao compararmos tais gráficos podemos concluir que o modelo ajustado com a presença de um menor número de pontos potencialmente influentes é o MTP, que assim como observado na análise da Subseção 2.4.1, é seguido pelo MBFC3. Por outro lado, o MBFC2 é novamente, o modelo que possui mais observações influentes. A diferença de qualidade de ajuste dos modelos segundo os critérios de avaliação utilizados, contraria a semelhança observada nas estimativas exibidas na Tabela 2.16.

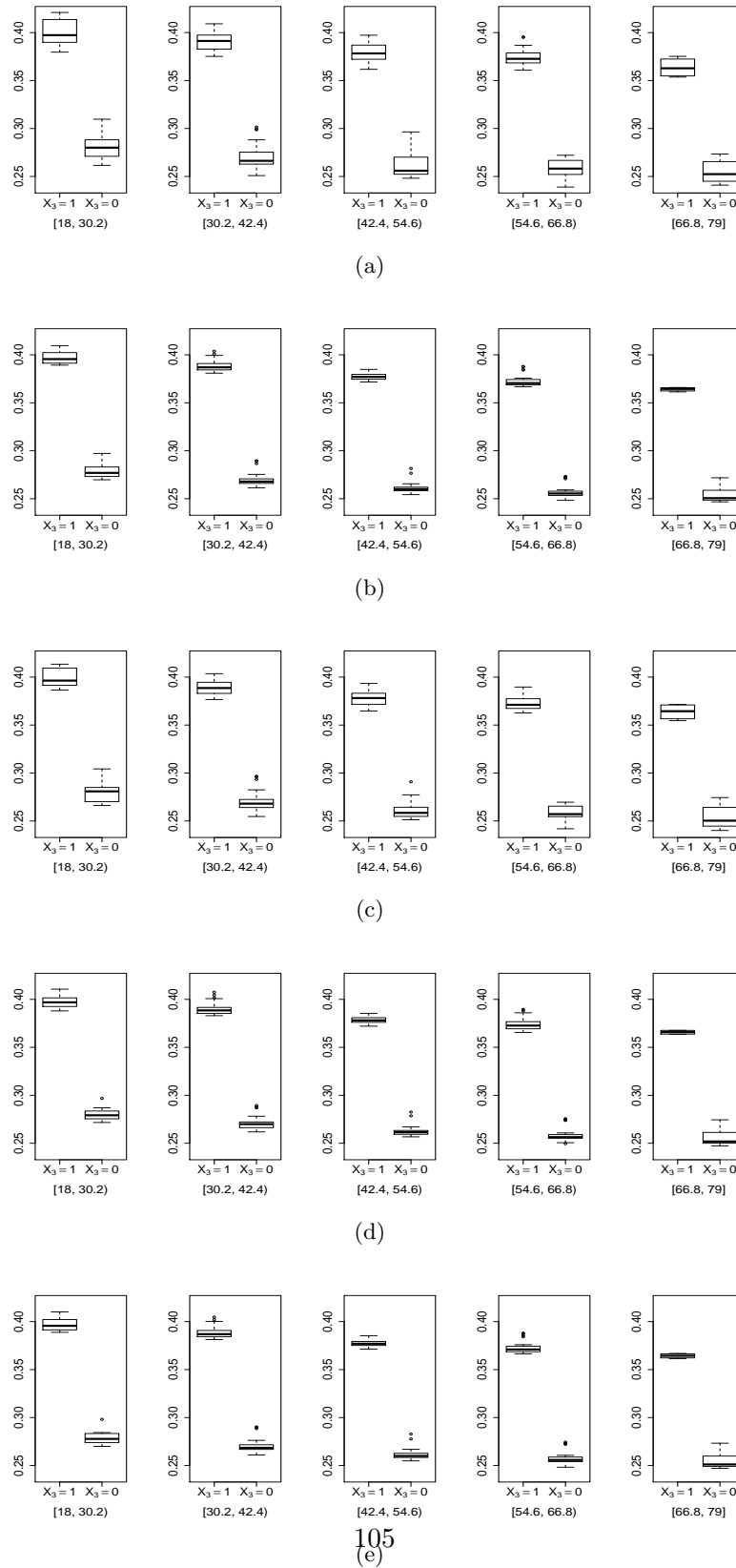


Figura 2.47: *Box-plots* das estimativas para a média *a posteriori* da fração de cura de todos os indivíduos observados para cada faixa etária considerada divididas entre os indivíduos que receberam Interferon ( $X_3 = 1$ ) e os que permaneceram apenas em observação ( $X_3 = 0$ ), sob o ajuste dos modelos MBFC1 (Gráfico a), 2 (Gráfico b) e 3 (Gráfico c), MTP (Gráfico d) e MEFC-Ma (Gráfico e).

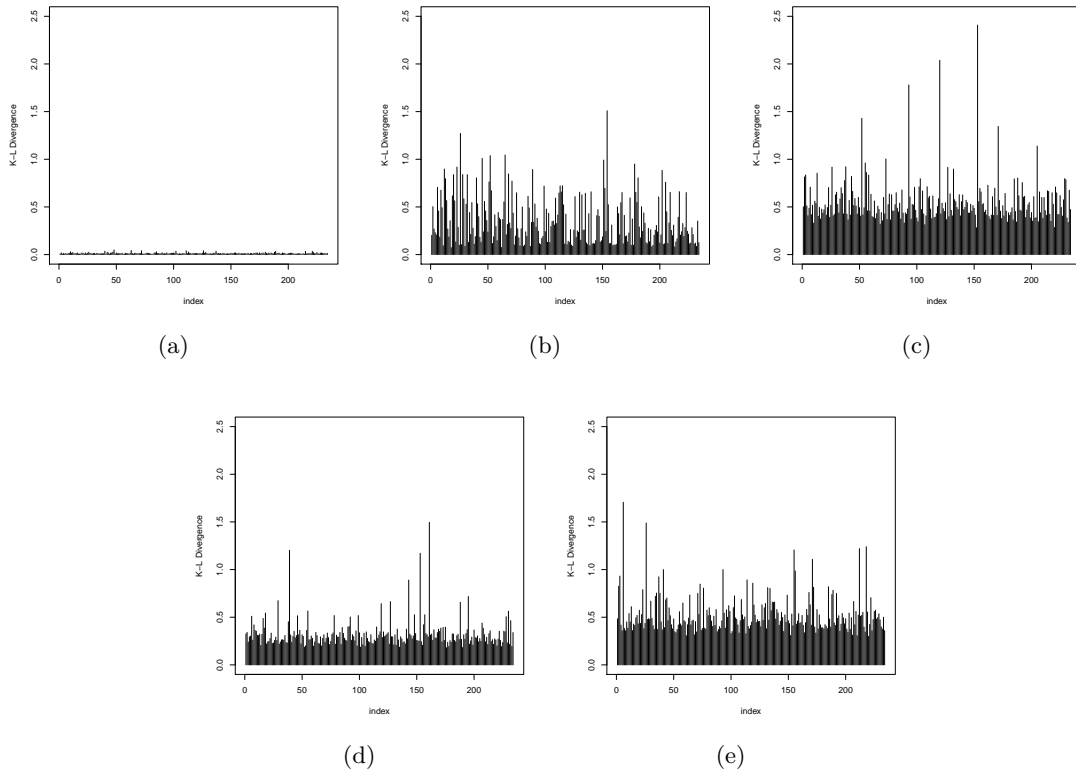


Figura 2.48: Divergência K-L. Os gráficos (a), (b), (c), (d) e (e) são relativos aos ajustes dos modelos MTP, MBFC1, MBFC2, MBFC3 e MBFC-Ma, respectivamente.

Para analisar a distribuição *a posteriori* da fração de cura de cada indivíduo, obtida em cada modelo ajustado, na Figura 2.49, apresentamos os gráficos de tais distribuições para os primeiros 30 indivíduos. Nestes gráficos, a exemplo do que indicam as estimativas apresentadas na Tabela 2.16, podemos observar que as distribuições *a posteriori* da fração de cura dos modelos MTP, MBFC2 e MBFC-Ma são muito próximas uma das outras e podemos notar uma certa diferença entre as curvas destas distribuições e as curvas das distribuições *a posteriori* que foram obtidas ajustando-se os modelos MBFC1 e MBFC3.

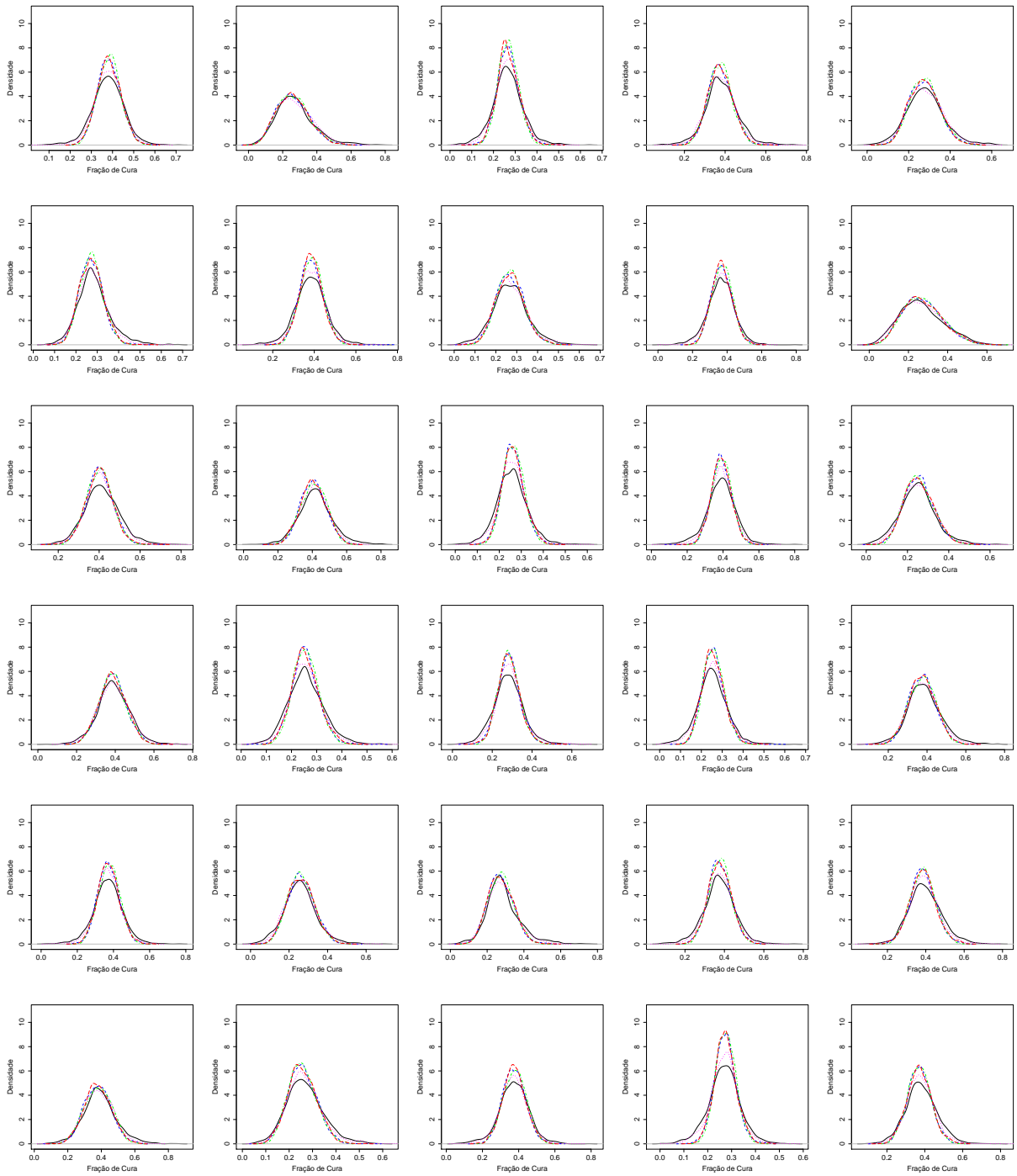


Figura 2.49: Distribuições *a posteriori* para a fração de cura. As linhas mista, sólida, traços pequenos, pontilhada e traços grandes são das distribuições obtidas ajustando-se os modelos MTP, MBFC1, MBFC2, MBFC3 e MBFC-Ma, respectivamente.

## Capítulo 3

# Modelo Bayesiano de Fração de Cura com Erro de Medida Normal Independente (MEFC-NI)

Neste capítulo, introduzimos um modelo estrutural de fração de cura com erro nas covariáveis no qual supomos que o erro de medida tem distribuição na família Normal independente (NI). A família NI de distribuições é obtida através de misturas de escala da distribuição Normal (Andrews e Mallows, 1974; Santos, 2015; Lange e Sinsheimer, 1993; Rogers e Tukey, 1972), entre outros. Esta é uma subclasse das distribuições elípticas bastante utilizadas na construção de modelos que são capazes de acomodar valores atípicos (*outliers*), pois a maioria das distribuições na família NI são distribuições com caudas mais pesadas que a distribuição Normal. Considerar esse tipo de distribuição para o erro de medida nos permite construir uma classe de modelos mais flexível a qual inclui o MBFC com erro normal definido no Capítulo 2 como caso particular.

Modelos de fração de cura com erro não Gaussianos foram considerados em Ma e Yin (2008) e Bertrand *et al.* (2017). Nestes trabalhos são propostas transformações nas covariáveis observadas  $\mathbf{W}$  de tal forma que a suposição de normalidade para os erros de medida sejam válidas. Uma modelagem funcional foi considerada para as covariáveis e a inferência sobre os parâmetros foi feita utilizando métodos clássicos.

Diferentemente da abordagem considerada por Ma e Yin (2008) e Bertrand *et al.* (2017) não nos utilizamos de transformações em  $\mathbf{W}$  para atingirmos normalidade nos erros. Propomos um modelo de sobrevivência com erro estrutural em que o comportamento do erro é modelado por distribuições com caudas pesadas e inferências são feitas assumindo tal estrutura. Especificamente, considere que o erro  $\epsilon_j$  associado a  $j$ -ésima covariável, é um vetor aleatório com distribuição na família NI e particularizamos o resultado assumindo as distribuições Normal e

$t$ -Student que são membros da família NI. De acordo com Lange e Sinsheimer (1993), um vetor aleatório  $\epsilon_j$  ( $n$ -dimensional) com distribuição na família NI ( $n$ -variada) com vetor de médias  $\mathbf{0}$  e matriz de covariância  $E(U_j^{-1})\Sigma_j$ , tem f.d.p

$$f(\epsilon_j) = \int_0^\infty \frac{u^{\frac{n}{2}}}{(2\pi)^{\frac{n}{2}}|\Omega|^{\frac{1}{2}}} \exp\left\{u\frac{v^2}{2}\right\} dH(u) \quad (3.1)$$

em que  $v^2 = \epsilon_j^\top \Omega^{-1} \epsilon_j$ ,  $U$  é chamado de fator misturador e  $H(u)$  é a distribuição de mistura e o erro  $\epsilon_j$  pode ser estocasticamente representado da seguinte forma:

$$\epsilon_j \stackrel{D}{=} U_j^{-\frac{1}{2}} \Sigma_j^{\frac{1}{2}} \mathbf{Z}_j, \quad j = 1, \dots, m, \quad (3.2)$$

em que  $U_j$  é uma variável aleatória com suporte positivo,  $\Sigma_j \in \mathbb{R}^n \times \mathbb{R}^n$  é uma matriz de covariâncias não singular e  $\mathbf{Z}_j$  é um vetor de variáveis aleatórias independentes normal  $n$ -variada com média  $\mathbf{0}$  e matriz de covariâncias  $\mathcal{I}_n$  (ver Apêndice A para mais detalhes). Tal representação será utilizada na construção do MBFC com erros NI distribuídos nas Seções 3.2.1 a 3.2.1.

Na Seção 3.3 apresentamos um estudo de simulação Monte Carlo para avaliarmos a capacidade de ajuste do modelo quando supomos que o erro de medida tem distribuição  $t$ -Student. Um dos grandes desafios no modelo  $t$ -Student é a estimação do número de graus de liberdade. Neste trabalho não focamos na estimação de tal parâmetro, porém investigamos o efeito de o assumirmos conhecido. Em todo o estudo de simulação realizado, consideramos uma distribuição *a priori* informativa para a variância das variáveis latentes, nos baseando nas conclusões obtidas no estudo de simulação apresentado na Seção 2.3. Comparamos o MBFC proposto assumindo erro  $t$ -Student e Normal com o modelo de tempo de promoção (Chen *et al.*, 1999) discutido no Capítulo 2.

A seguir, definimos o modelo estrutural de fração de cura com erro de medida na família normal independente, o qual é uma generalização do MBFC, apresentado no Capítulo 2. A construção da função de verossimilhança é análoga aquela obtida no Capítulo 2, excetuando-se a parte II da Equação (2.7), que insere a informação de que as covariáveis foram observadas com erro.

### 3.1 MBFC com Erro de Medida Normal Independente

Considere o modelo de tempo promoção definido no Capítulo 2, na Equação (2.2), cuja função de sobrevivência é dada pela equação a seguir:

$$S_p(t | \mathbf{X}) = \exp\{-\exp\{\mathbf{X}\beta\} F(t)\}, \quad (3.3)$$

em que  $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)^\top$ , denota a matriz de delineamento de experimento de ordem  $n \times p$ ,  $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})$  denota o vetor linha de ordem  $1 \times p$  das covariáveis observadas para o indivíduo  $i$  e  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$  é o vetor de coeficientes de regressão de ordem  $p \times 1$ .

Suponha que  $m$  covariáveis sejam medidas com erro de modo que a matriz  $\mathbf{X}$  seja particionada na submatriz  $\mathbf{X}_E$  de ordem  $n \times m$  dos valores reais e não observados das variáveis que experimentam erro de medição, e a submatriz  $\mathbf{X}_o$  de ordem  $n \times (p - m)$  correspondente as covariáveis completamente observadas. Dessa maneira, a matriz de covariáveis observadas de ordem  $n \times p$  é

$$\mathbf{W} = (\mathbf{X}_E + \mathbf{E}, \mathbf{X}_o),$$

em que  $\mathbf{E} = (\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n)^\top$  é uma matriz de erros de ordem  $n \times m$  tal que  $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{im})$ .

Denotemos por  $\mathbf{W}_{E_j}$  e  $\mathbf{X}_{E_j}$  os vetores formados pelas colunas de  $\mathbf{W}$  e  $\mathbf{X}_E$ , respectivamente, relacionados a  $j$ -ésima covariável observada com erro de medida, de tal forma que:

$$\mathbf{W}_{E_j} = \mathbf{X}_{E_j} + \boldsymbol{\epsilon}_j, \quad j = 1, \dots, m, \quad (3.4)$$

em que  $\boldsymbol{\epsilon}_j = (\epsilon_{j1}, \dots, \epsilon_{jn})^\top$  é o vetor de erros associado a covariável  $j$  para os  $n$  indivíduos da amostra. Suponhamos que os erros  $\epsilon_{j1}, \dots, \epsilon_{jn}$  são independentes com distribuição NI com média zero e variância  $E(U_{\epsilon_{ji}}^{-1})\sigma_\epsilon^2$ . Utilizando a representação estocástica descrita na Equação (3.2), podemos escrever

$$\epsilon_{ji} \stackrel{D}{=} U_{\epsilon_{ji}}^{-\frac{1}{2}} \sigma_\epsilon Z_{\epsilon_{ji}}. \quad (3.5)$$

Admitamos que  $U_{\epsilon_{j1}}, \dots, U_{\epsilon_{jn}}$  são variáveis aleatórias positivas i.i.d, e  $Z_{\epsilon_{j1}}, \dots, Z_{\epsilon_{jn}} \stackrel{i.i.d}{\sim} N(0, 1)$ . Portanto, como as variáveis aleatórias  $U_{\epsilon_{ji}}$ 's são i.i.d, temos que  $E(U_{\epsilon_{j1}}) = \dots = E(U_{\epsilon_{jn}})$  e, como consequência disso, os erros relacionados a  $j$ -ésima covariável são i.i.d com distribuição NI tendo média 0 e variância comum  $E(U_{\epsilon_{ji}}^{-1})\sigma_\epsilon^2$ . Além disso, assumimos que os erros observados para o indivíduo  $i$  nas  $m$  covariáveis medidas com erro são independentes, mas, não são identicamente distribuídos. Isto é obtido supondo que  $U_{\epsilon_{j1}}, \dots, U_{\epsilon_{mi}}$  são variáveis aleatórias independentes tais que  $E(U_{\epsilon_{ji}})$  são diferentes para todo  $j = 1, \dots, m$ . Dessa maneira, completamos a definição do modelo Bayesiano de fração de cura com erro na família normal independente (MBFC-NI).

Seguindo a mesma proposta que fizemos para o MBFC no Capítulo 2, consideremos, sem perda de generalidade, que apenas uma covariável seja medida com erro, isto é  $m = 1$ . Denotemos por  $\mathbf{X}_{\epsilon_i}$  e  $\epsilon_i$  o valor real desta covariável e do erro a ela associado para o indivíduo  $i = 1, \dots, n$ , respectivamente. Sob esses pressupostos, apresentamos a seguir a inferência Bayesiana no MBFC-NI.

### 3.2 Inferência Bayesiana no MBFC-NI

Nesta seção, a exemplo do que fizemos para o modelo estrutural de fração de cura com erro de medida nas covariáveis Normal, propomos três modelagens distintas para o MBFC-NI, tal como fizemos no Capítulo 2. Na primeira delas, denotada por MBFC-NI1, suponhamos que o parâmetro  $\sigma_\epsilon^2$  associado ao erro de medida é conhecido. Uma segunda modelagem é feita, quando supomos que  $\sigma_\epsilon^2$  é desconhecido e será estimado. Denotemos essa segunda versão por MBFC-NI2. Uma terceira modelagem é proposta, supondo que há uma relação entre as variâncias do erro e da variável latente, isto é, assumimos que  $k = \frac{\sigma_\epsilon^2}{\sigma_x^2}$ , em que  $k$  é uma constante desconhecida e  $\sigma_x^2$  denota a variância da variável latente  $X_{\epsilon_i}$ , para todo  $i = 1, \dots, n$ , definindo o modelo MBFC-NI3.

Considerando que os tempos de falha  $\mathbf{Y}$  possuem distribuição Weibull( $\alpha, \lambda$ ), a função de verossimilhança do MBFC-NI é dada por

$$\begin{aligned}
 L(\alpha, \lambda, \boldsymbol{\beta}, \sigma_\epsilon^2 | \mathbf{D}_{\text{obs}}) &\propto \left[ \prod_{i=1}^n \left( \exp \{ \mathbf{X}_{\epsilon_i} \boldsymbol{\beta}_\epsilon + \mathbf{X}_{oi} \boldsymbol{\beta}_o \} \exp \{ \lambda \} y_i^{\alpha-1} \exp \{ - \exp \{ \lambda \} y_i^\alpha \} \right)^{\delta_i} \right. \\
 &\quad \left. \times \exp \{ - \exp \{ \mathbf{X}_{\epsilon_i} \boldsymbol{\beta}_\epsilon + \mathbf{X}_{oi} \boldsymbol{\beta}_o \} [1 - \exp \{ - \exp \{ \lambda \} y_i^\alpha \}] \} \right] \\
 &\quad \times \prod_{i=1}^n \int_0^\infty \left( \frac{u_{\epsilon_i}}{2\pi\sigma_{\epsilon_i}^2} \right)^{\frac{1}{2}} \exp \left\{ \frac{u_{\epsilon_i}}{2\sigma_{\epsilon_i}^2} (w_{\epsilon_i} - x_{\epsilon_i})^2 \right\} dH(u_{\epsilon_i}), \quad (3.6)
 \end{aligned}$$

em que  $U_{\epsilon_i}$  é uma variável aleatória não negativa com medida de probabilidade  $H$ .

Distintos modelos de fração de cura com erro nas covariáveis podem ser definidos a partir da especificação de  $H$ . Se assumirmos que  $U_{\epsilon_i}$  tem distribuição degenerada em 1, retornamos ao MBFC com erro Normal discutido no Capítulo 2. Se  $U_{\epsilon_i} \stackrel{i.i.d}{\sim} \text{Gama} \left( \frac{\nu}{2}, \frac{\nu}{2} \right)$ , então,  $\epsilon_1, \dots, \epsilon_n$  serão independentes com distribuição comum  $t$ -Student com  $\nu$  graus de liberdade. A inferência neste modelo será discutida em mais detalhes nas próximas seções.

Para completarmos a especificação do modelo proposto, precisamos determinar as distribuições *a priori* dos parâmetros, das variáveis latentes, bem como dos seus hiperparâmetros e que indexam a distribuição de  $U_\epsilon$ . Em todos os modelos assumiremos que a variável latente  $\mathbf{X}_\epsilon | \mu_x, \sigma_x^2 \sim N_n(\boldsymbol{\mu}_x, \sigma_x^2 \mathcal{I}_n)$ . Suponhamos as seguintes especificações *a priori*:



$$\begin{aligned}
 X_{\epsilon_i} &| \mu_x, \sigma_x^2 \stackrel{i.i.d.}{\sim} N(\mu_x, \sigma_x^2), \\
 U_{\epsilon_i} &| \nu \stackrel{i.i.d.}{\sim} \pi(u), \\
 \boldsymbol{\beta} &\propto \mathbf{1}, \\
 \lambda &\sim N(m_1, v_1), \\
 \alpha &\sim G(a_1, b_1), \\
 \mu_x &\sim N(m_2, v_2), \\
 \sigma_x^2 &\sim GI(a_2, v_2), \\
 \nu &\sim \pi(\nu).
 \end{aligned} \tag{3.7}$$

Na próxima seção, obtemos as condicionais completas dos parâmetros do MBFC-NI considerando as três modelagens propostas: MBFC-NI1, MBFC-NI2 e MBFC-NI3. No modelo MBFC-NI1,  $\sigma_\epsilon^2$  é suposto conhecido, enquanto nos modelos MBFC-NI2 e MBFC-NI3 esse parâmetro é desconhecido e será estimado. Tal como fizemos para o MBFC3 (Capítulo 2), no MBFC-NI3 supomos  $\sigma_\epsilon^2 = k\sigma_x^2$ , sendo  $0 < k \leq 1$  uma constante desconhecida. O cálculo da distribuição *a posteriori* dos parâmetros e a demonstração de que esta é uma distribuição de probabilidade própria, pode ser consultada no Apêndice B desse texto.

### 3.2.1 MBFCNI1: MBFC-NI com $\sigma_\epsilon^2$ conhecido

Suponha que  $\sigma_\epsilon^2$  é uma constante conhecida e considere as especificações *a priori* dadas na Expressão (B.1). Sob tais suposições, as distribuições condicionais completas dos parâmetros  $\boldsymbol{\beta}, \alpha, \lambda, \mu_x$  e  $\sigma_x^2$  do MBFC-NI são idênticas àquelas obtidas no modelo MBFC1 nas Equações (3.8), (2.13), (2.12), (2.16) e (2.17), pois, tais d.c.c não dependem diretamente da f.d.p de  $\mathbf{W}_\epsilon$ . Assim teremos:

i) para cada regressor  $\beta_j, j = 1, \dots, p$

$$\pi(\beta_j | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}_{(-\beta_j)}) \propto \exp \left\{ \sum_{i=1}^n \delta_i (X_{ij} \beta_j) - \sum_{i=1}^n \exp \{ \mathbf{X}_{\epsilon_i} \boldsymbol{\beta}_\epsilon + \mathbf{X}_{oi} \boldsymbol{\beta}_o \} (1 - \exp \{ -\exp \{ \lambda \} y_i^\alpha \}) \right\}. \tag{3.8}$$

ii) para o parâmetro  $\lambda$  associado à distribuição dos tempos de falha,

$$\begin{aligned} \pi(\lambda | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}_{(-\lambda)}) &\propto \exp \left\{ \lambda \sum_{i=1}^n \delta_i - \exp \{ \lambda \} \sum_{i=1}^n \delta_i y_i^\alpha + \sum_{i=1}^n \exp \{ \mathbf{X}_{\epsilon_i} \boldsymbol{\beta}_\epsilon + \mathbf{X}_{o_i} \boldsymbol{\beta}_o - \exp \{ \lambda \} y_i^\alpha \} \right\} \\ &\times \exp \left\{ \frac{-(\lambda - m_1)^2}{2v_1} \right\}. \end{aligned}$$

iii) para o parâmetro  $\alpha$  associado à distribuição dos tempos de falha,

$$\begin{aligned} \pi(\alpha | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}_{(-\alpha)}) &\propto \exp \left\{ \sum_{i=1}^n [-\exp \{ \lambda \} \delta_i y_i^\alpha] + \exp \{ \mathbf{X}_{\epsilon_i} \boldsymbol{\beta}_\epsilon + \mathbf{X}_{o_i} \boldsymbol{\beta}_o - \exp \{ \lambda \} y_i^\alpha \} \right\} \\ &\times \left[ \prod_{i=1}^n y_i^{(\alpha-1)\delta_i} \right] \alpha^{a_1-1+\sum_{i=1}^n \delta_i} \exp \{ -b_1 \alpha \}. \end{aligned}$$

iv) Para a média  $\mu_x$  e variância  $\sigma_x^2$  das variáveis latentes, temos:

$$\mu_x | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}_{(-\mu_x)} \sim N \left( \frac{\mathbf{v}_2 \mathcal{I}_n \mathbf{x}_\epsilon + \sigma_x^2 m_2}{\sigma_x^2 + v_2}, \frac{\sigma_x^2 v_2}{\sigma_x^2 + v_2} \right),$$

em que  $\mathbf{v}_2 = v_2 \mathbf{1}$ , tal que  $\mathbf{1}$  é um vetor de dimensão  $n \times 1$  de uns, e

$$\sigma_x^2 | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}_{(-\sigma_x^2)} \sim GI \left( a_2 + \frac{n}{2}, \frac{1}{2} (\mathbf{x}_\epsilon - \boldsymbol{\mu}_x)^\top \mathcal{I}_n (\mathbf{x}_\epsilon - \boldsymbol{\mu}_x) + b_2 \right).$$

Considerando que  $X_{\epsilon_i}$  e  $U_{\epsilon_i}$  são, respectivamente a variável latente e a variável misturadora associadas ao  $i$ -ésimo indivíduo, a seguir obtemos suas d.c.c .

Para a variável latente  $X_{\epsilon_i}$ , a d.c.c é:

$$\begin{aligned} \pi(x_{\epsilon_i} | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}_{(-x_{\epsilon_i})}) &\propto \exp \{ \delta_i (x_{\epsilon_i} \boldsymbol{\beta}_\epsilon + \mathbf{x}_{o_i} \boldsymbol{\beta}_o) - \exp \{ x_{\epsilon_i} \boldsymbol{\beta}_\epsilon + \mathbf{x}_{o_i} \boldsymbol{\beta}_o \} \} \\ &\times \exp \left\{ \exp \{ x_{\epsilon_i} \boldsymbol{\beta}_\epsilon + \mathbf{x}_{o_i} \boldsymbol{\beta}_o \} \exp \left\{ -e^\lambda y_i^\alpha \right\} \right\} \\ &\times \left[ \int_0^\infty \frac{u_{\epsilon_i}^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}} \sigma_\epsilon} \exp \left\{ u_{\epsilon_i} \frac{v_i^2}{2} \right\} dH(u_{\epsilon_i}) \right] \pi(x_{\epsilon_i} | \mu_x, \sigma_x^2), \quad (3.9) \end{aligned}$$

em que  $v_i = \frac{(w_{\epsilon_i} - x_{\epsilon_i})}{\sigma_\epsilon}$  e  $w_{\epsilon_i}$  é o valor observado da covariável medida com erro para o indivíduo  $i$ . Observe, que no caso em que a variável misturadora  $U_{\epsilon_i}$  é degenerada em 1, a d.c.c de  $X_{\epsilon_i}$  sob o modelo MBFC-NI1 torna-se igual aquela obtida considerando o MBFC1, na Expressão 2.14.

Para construirmos o modelo com erro de medida com distribuição  $t$ -Student, assumi-

mos que as variáveis aleatórias  $U_{\epsilon_i}$ , nas Equações (3.2) e (3.4) são i.i.d e possuem distribuição Gama  $\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ . Nesse caso, dado  $X_{\epsilon_i}$  e  $U_{\epsilon_i}$ , temos que  $W_{\epsilon_i}$  possui distribuição  $t$ -Student com parâmetros de locação  $X_{\epsilon_i}$ , de escala  $\sigma_{\epsilon}^2 U_{\epsilon_i}^{-1}$  e graus de liberdade  $\nu$ . Denotemos  $W_{\epsilon_i} | X_{\epsilon_i}, U_{\epsilon_i}, \sigma_{\epsilon}^2, \nu \sim t(X_{\epsilon_i}, \sigma_{\epsilon}^2 U_{\epsilon_i}^{-1}, \nu)$ . Para completarmos a inferência no MBFC-NI1, obtemos a seguir as distribuições condicionais completas *a posteriori* de  $X_{\epsilon}$ ,  $U_{\epsilon_i}$  e  $\nu$ . As distribuições condicionais completas destes parâmetros podem ser obtidas de maneira mais simples se considerarmos uma modelagem hierárquica para  $W_{\epsilon_i}$ , inspirada pela representação estocástica dada na Expressão (3.4). Nesse caso, assumindo que  $W_{\epsilon_i}$  é uma variável aleatória com distribuição  $t$ -Student, então para o  $i$ -ésimo indivíduo temos

$$W_{\epsilon_i} | X_{\epsilon_i}, \sigma_{\epsilon}^2, U_{\epsilon_i} = X_{\epsilon_i} + U_{\epsilon_i}^{-\frac{1}{2}} \sigma_{\epsilon} Z_{\epsilon_i},$$

em que  $Z_{\epsilon_i} \stackrel{i.i.d}{\sim} \text{Normal}(0, 1)$  e  $U_{\epsilon_i} \stackrel{i.i.d}{\sim} \text{Gama}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ . Assim, o modelo MBFC-NI pode ser hierarquicamente representado como segue:

$$\begin{aligned} Y_i | W_{\epsilon_i}, X_{\epsilon_i}, \beta, \alpha, \lambda &\stackrel{indep.}{\sim} f_p(y_i | W_{\epsilon_i}, X_{\epsilon_i}, \beta, \alpha, \lambda), \\ W_{\epsilon_i} | X_{\epsilon_i}, \sigma_{\epsilon}^2, U_{\epsilon_i} &\stackrel{indep.}{\sim} N(X_{\epsilon_i}, U_{\epsilon_i}^{-1} \sigma_{\epsilon}^2), \\ X_{\epsilon_i} | \mu_x, \sigma_x^2 &\stackrel{i.i.d}{\sim} N(\mu_x, \sigma_x^2), \\ U_{\epsilon_i} | \nu &\stackrel{i.i.d}{\sim} G\left(\frac{\nu}{2}, \frac{\nu}{2}\right), \\ \mu_x &\sim N(m_2, v_2), \\ \sigma_x^2 &\sim GI(a_2, v_2), \\ \nu &\sim \pi(\nu). \end{aligned} \tag{3.10}$$

A representação hierárquica do modelo proposto em (3.10), simplifica bastante a implementação dos algoritmos MCMC que serão usados para obter amostras das distribuições *a posteriori* dos parâmetros, em todos os modelos definidos neste capítulo.

Para completarmos a especificação do MBFC-NI1, resta apenas determinarmos as distribuições condicionais completas de  $X_{\epsilon_i}$ , e  $U_{\epsilon_i}$ . A distribuição condicional completa para  $X_{\epsilon_i}$ ,  $i = 1, \dots, n$  é dada por:

$$\begin{aligned} \pi(x_{\epsilon_i} | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}_{(-x_{\epsilon_i})}, \mathbf{u}, \nu) &\propto \exp \left\{ \sum_{i=1}^n \left[ \delta_i (\mathbf{X}_{\epsilon_i} \boldsymbol{\beta} + \mathbf{X}_{o_i} \boldsymbol{\beta}_o) - \exp \{ \mathbf{X}_{\epsilon_i} \boldsymbol{\beta} + \mathbf{X}_{o_i} \boldsymbol{\beta}_o \} \right] \right\} \times \\ &\exp \left\{ \sum_{i=1}^n \exp \{ \mathbf{X}_{\epsilon_i} \boldsymbol{\beta} + \mathbf{X}_{o_i} \boldsymbol{\beta}_o \} \exp \{ - \exp \{ \lambda \} y_i^{\alpha} \} \right\} \times \\ &\phi(x_{\epsilon_i} | \mu_{x_i}, \sigma_{x_i}^2), \end{aligned} \tag{3.11}$$

em que  $\phi(x_{\epsilon_i} | \mu_{x_i}, \sigma_{x_i}^2)$  é a f.d.p da Normal com média  $\mu_{x_i} = \frac{\sigma_x^2 u_{\epsilon_i} w_i + \sigma_\epsilon^2 \mu_x}{u_{\epsilon_i} \sigma_x^2 + \sigma_\epsilon^2}$  e variância  $\sigma_{x_i}^2 = \frac{\sigma_x^2 \sigma_\epsilon^2}{u_{\epsilon_i} \sigma_x^2 + \sigma_\epsilon^2}$ . Para a variável misturadora  $U_{\epsilon_i}$  de cada indivíduo  $i$ , a d.c.c será:

$$\pi(u_{\epsilon_i} | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}, \nu) \propto u_{\epsilon_i}^{\frac{\nu+1}{2}-1} \exp \left\{ -u_{\epsilon_i} \left[ \frac{(w_{\epsilon_i} - x_{\epsilon_i})^2}{2\sigma_\epsilon^2} + \frac{\nu}{2} \right] \right\}, \quad (3.12)$$

a qual possui forma fechada sendo  $U_{\epsilon_i} | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}, \nu \stackrel{\text{indep.}}{\sim} G \left( \frac{\nu+1}{2}, \frac{(w_{\epsilon_i} - x_{\epsilon_i})^2}{2\sigma_\epsilon^2} + \frac{\nu}{2} \right)$ .

Para os graus de liberdade  $\nu$ , a condicional completa pode ser obtida da seguinte maneira:

$$\pi(\nu | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}, \mathbf{u}) \propto \pi(\nu) \prod_{i=1}^n \pi(u_{\epsilon_i}). \quad (3.13)$$

No estudo de simulação apresentado na Seção 3.3, consideramos  $\nu$  conhecido. É importante ressaltarmos o ganho computacional que temos ao utilizarmos uma modelagem hierárquica para o modelo proposto. No MCMC tal representação faz com que várias das distribuições condicionais completas dos parâmetros tenham forma fechada. Nesse caso, usamos um algoritmo *Gibbs Sampling* com passos de *Metropolis* Adaptativo (Migon *et al.*, 2015; Haario *et al.*, 2001, 2005), para amostrarmos das distribuições *a posteriori*.

### 3.2.2 MBFC-NI2: MBFC-NI com $\sigma_\epsilon^2$ desconhecido

Na segunda versão proposta para o MBFC-NI, a qual denotamos por MBFC-NI2, supomos que  $\sigma_\epsilon^2$  é desconhecido, ao contrário do que consideramos no MBFC-NI1. Para todos os parâmetros que são comuns aos modelos MBFC-NI1 e MBFC-NI2, as distribuições condicionais completas permanecem iguais as obtidas na subseção anterior, portanto, indicamos a seguir apenas a d.c.c de  $\sigma_\epsilon^2$ . Supondo  $\sigma_\epsilon^2$  desconhecido, inserimos uma maior flexibilidade ao modelo e o tornamos mais realístico em termos práticos, visto que no contexto de variáveis passíveis de erro de medida, a variância desse erro pode não ser precisamente conhecida como assumimos no MBFC-NI1.

Se considerarmos que *a priori*  $\sigma_\epsilon^2 \sim GI(a_3, b_3)$ , tal como fizemos no MBFC2, a d.c.c de  $\sigma_\epsilon^2$  sob o MBFC-NI2 é dada por:

$$\pi(\sigma_\epsilon^2 | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}_{(-\sigma_\epsilon^2)}, U_\epsilon, \nu) \propto (\sigma_\epsilon^2)^{-(a_3 + \frac{n}{2} - 1)} \exp \left\{ - \frac{[(\mathbf{w}_\epsilon - \mathbf{x}_\epsilon)^\top \boldsymbol{\Sigma}^{-1} (\mathbf{w}_\epsilon - \mathbf{x}_\epsilon) + b_2]}{\sigma_\epsilon^2} \right\},$$

em que  $\Sigma = \text{diag} \{U_{\epsilon_1}^{-1}, \dots, U_{\epsilon_n}^{-1}\}$ . Concluimos que:

$$\sigma_\epsilon^2 | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}_{(-\sigma_\epsilon^2)}, U_\epsilon, \nu \sim GI \left( a_3 + \frac{n}{2}, (\mathbf{w}_\epsilon - \mathbf{x}_\epsilon)^\top \Sigma^{-1} (\mathbf{w}_\epsilon - \mathbf{x}_\epsilon) + b_2 \right).$$

Esta forma fechada para d.c.c de  $\sigma_\epsilon^2$  foi obtida ao considerarmos a representação hierárquica do modelo proposto na Equação (3.10).

### 3.2.3 MBFC-NI3: MBFC-NI com $\sigma_\epsilon^2 = k\sigma_x^2$

De modo análogo ao que foi definido no MBFC3 (Seção 2.1.3), supomos que  $\sigma_\epsilon^2 = k\sigma_x^2$ , em que  $k$  é uma constante desconhecida tal que  $0 < k \leq 1$ . Por essa razão, a inferência a respeito de  $\sigma_\epsilon^2$  nesse caso, é feita através de  $k$  e  $\sigma_x^2$ . Admitimos, novamente, que *a priori*  $k \sim U(a, b)$  sendo  $0 < a < b \leq 1$ . Assim, obtemos que a distribuição condicional completa *a posteriori* para  $k$  é:

$$\pi(k | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}_{(-\sigma_\epsilon^2)}, U) \propto \frac{1}{k^{\frac{n}{2}}(b-a)} \exp \left\{ -\frac{1}{k\sigma_x^2} (\mathbf{w}_\epsilon - \mathbf{x}_\epsilon)^\top \Sigma^{-1} (\mathbf{w}_\epsilon - \mathbf{x}_\epsilon) \right\} \mathbf{1}_{(a,b)}(k), \quad (3.14)$$

em que  $\Sigma = \text{diag} \{U_{\epsilon_1}^{-1}, \dots, U_{\epsilon_n}^{-1}\}$ . Para a variância das variáveis latentes, temos que a d.c.c de  $\sigma_x^2$  é

$$\begin{aligned} \pi(\sigma_x^2 | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}_{(-\sigma_\epsilon^2)}, k, U_\epsilon) &\propto f_{\mathbf{w}_\epsilon}(\mathbf{w}_\epsilon | \mathbf{x}_\epsilon, \mathbf{x}_o, \sigma_x^2, k, U_\epsilon) \pi(\mathbf{x}_\epsilon | \mathbf{D}_{\text{obs}}, \mu_x, \sigma_x^2) \pi(\sigma_x^2 | a_2, b_2) \\ &\propto \left( \frac{1}{k\sigma_x^2} \right)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}} \exp \left\{ -\frac{1}{2k\sigma_x^2} (\mathbf{w}_\epsilon - \mathbf{x}_\epsilon)^\top \Sigma^{-1} (\mathbf{w}_\epsilon - \mathbf{x}_\epsilon) \right\} \\ &\quad \times \left( \frac{1}{\sigma_x^2} \right)^{\frac{n}{2}} |\mathcal{I}_n| \exp \left\{ -\frac{1}{2\sigma_x^2} (\mathbf{x}_\epsilon - \mu_x)^\top \mathcal{I}_n (\mathbf{x}_\epsilon - \mu_x) \right\} \left( \frac{1}{\sigma_x^2} \right)^{a_2+1} \\ &\quad \times \exp \left\{ -\frac{b_2}{\sigma_x^2} \right\} \\ &\propto \left( \frac{1}{\sigma_x^2} \right)^{a_2 + \frac{n}{2} + 1} \exp \left\{ -\frac{1}{\sigma_x^2} \left[ \frac{1}{2k} (\mathbf{w}_\epsilon - \mathbf{x}_\epsilon)^\top \Sigma^{-1} (\mathbf{w}_\epsilon - \mathbf{x}_\epsilon) \right. \right. \\ &\quad \left. \left. + \frac{1}{2} (\mathbf{x}_\epsilon - \mu_x)^\top \mathcal{I}_n (\mathbf{x}_\epsilon - \mu_x) + b_2 \right] \right\}, \end{aligned}$$

ou seja,

$$\sigma_x^2 | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}_{(-\sigma_\epsilon^2)}, k, U_\epsilon \sim GI \left( a_2 + \frac{n}{2}, \frac{1}{2k} (\mathbf{w}_\epsilon - \mathbf{x}_\epsilon)^\top \Sigma^{-1} (\mathbf{w}_\epsilon - \mathbf{x}_\epsilon) + \frac{1}{2} (\mathbf{x}_\epsilon - \mu_x)^\top \mathcal{I}_n (\mathbf{x}_\epsilon - \mu_x) + b_2 \right).$$

Sob essa estrutura, para a variável misturadora  $U_\epsilon$  de cada indivíduo  $i$ , a d.c.c *a posteriori*

torna-se

$$\pi(u_{\epsilon_i} | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}, \nu, k) \propto u_{\epsilon_i}^{\frac{\nu+1}{2}-1} \exp \left\{ -u_{\epsilon_i} \left[ \frac{(w_{\epsilon_i} - x_{\epsilon_i})^2}{2k\sigma_x^2} + \frac{\nu}{2} \right] \right\}, \quad (3.15)$$

portanto,  $U_{\epsilon_i} | \mathbf{D}_{\text{obs}}, \boldsymbol{\eta}, \nu, k \sim G \left( \frac{\nu+1}{2}, \frac{(w_{\epsilon_i} - x_{\epsilon_i})^2}{2k\sigma_x^2} + \frac{\nu}{2} \right)$ .

As três especificações propostas para o MBFC-NI tem aspectos que são comuns, o que torna a inferência Bayesiana, sobre os parâmetros em cada modelo, muito semelhante. O pressuposto de independência entre indivíduos e entre os parâmetros do modelo torna isso possível. Por esse motivo, o cálculo das condicionais completas dos parâmetros que são comuns as três modelagens é o mesmo.

Os métodos MCMC escolhidos para obter amostras das distribuições *a posteriori* dos parâmetros desses modelos, são os mesmos que utilizamos nas três modelagens propostas para o MBFC no Capítulo 2 desse texto.

### 3.3 Estudo de Simulação de Monte Carlo para Avaliar o MBFC-NI

Nesta seção, apresentamos alguns estudos de simulação de Monte Carlo para avaliar a capacidade de ajuste do modelo proposto no presente capítulo, supondo que o erro de medida tem distribuição *t-Student* com  $\nu$  graus de liberdade. Nos cenários analisados, consideramos  $\nu$  conhecido e no primeiro estudo de simulação apresentado avaliamos o efeito desta suposição. Em todo o texto que segue, nos referimos ao modelo estrutural de fração de cura com erro de medida Normal nas covariáveis como MBFC e ao modelo com erro *t-Student* por MBFC-NI. Apresentamos uma comparação entre tais modelos e o MTP de Chen *et al.* (1999). Para realizar os estudos de simulação expostos a seguir, consideramos três cenários distintos. Nos Cenários 1 e 3 supomos que uma das covariáveis foi observada com erro, enquanto que no Cenário 2 todas as covariáveis são completamente observadas. O objetivo de analisar os dados do Cenário 2, é avaliar se o modelo proposto é capaz de se ajustar bem a cenários nos quais as covariáveis não possuem erro de medida. No Cenário 1 os dados são gerados supondo erro de medida com distribuição *t-Student*, enquanto que no Cenário 3, supomos erro normal. Ao estudarmos esses dois cenários fazemos uma comparação entre o MBFC e o MBFC-NI com a finalidade de avaliar a robustez destes modelos.

Nos estudos de simulação realizados, consideramos que a variável misturadora  $U_{\epsilon}$ , nas Equações (3.2) e (3.4), tem distribuição  $G \left( \frac{\nu}{2}, \frac{\nu}{2} \right)$ , e portanto, o erro de medida para cada

indivíduo  $i$ , condicional em  $U_{\epsilon_i}$ ,  $\sigma_\epsilon^2$  e  $\nu$ , tem distribuição  $t(0, U_{\epsilon_i}^{-1}\sigma_\epsilon^2, \nu)$ .

A seguir, descreveremos detalhadamente os cenários que serão considerados nos estudos de simulação apresentados nesta seção. Em todos eles supomos amostras de tamanho 100 e no Monte Carlo fazemos 250 réplicas. O código em R (R Core Team, 2016) implementado para gerar os bancos de dados do Cenário 1 se encontra disponibilizado no Apêndice E.

**Cenário 1:** Neste cenário, os dados foram gerados de um modelo de fração de cura com erro de medida  $t$ -Student nas covariáveis, atendendo as seguintes especificações: os tempos de falha  $\mathbf{Y} = (y_1, \dots, y_n)$ , são gerados da distribuição Weibull( $\alpha, \lambda$ ), em que  $\alpha = 1$ ,  $\lambda = -1.2$  e consideramos 30% de censura; a matriz de delineamento de experimento é dada por  $\mathbf{W} = (\mathbf{X}_1, \mathbf{X}_2 + \boldsymbol{\epsilon})$ , em que  $X_1 \sim \text{Bernoulli}(0.4)$ ,  $X_2 \sim \text{Normal}(\mu_x, \sigma_x^2)$ , com  $\mu_x = 3.9$  e  $\sigma_x^2 = 10$ ; os erros de medida  $\epsilon_i \stackrel{i.i.d.}{\sim} NI(0, U_{\epsilon_i}^{-1}\sigma_\epsilon^2, \nu)$ , em que  $U_{\epsilon_i} \sim G\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ ,  $\sigma_\epsilon^2 = 1$  e  $\nu = 3$ ; e os coeficientes de regressão são  $\beta_1 = 1.8$  e  $\beta_2 = -0.15$ .

**Cenário 2:** O objetivo deste cenário é estudarmos o que acontece quando ajustamos o MBFC-NI no qual se supõe erro de medida, porém os dados foram medidos livres de erro. Nesse sentido, geramos os dados assumindo que os tempos de falha  $\mathbf{Y} = (y_1, \dots, y_n)$  possuem distribuição Weibull( $\alpha, \lambda$ ), com  $\alpha = 1$ ,  $\lambda = -1.2$  e considerando 30% de censura. A matriz de delineamento de experimento é dada por  $\mathbf{W} = \mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ , em que  $X_1 \sim \text{Bernoulli}(0.4)$  e  $X_2 \sim N(3.9, 10)$ ; novamente, o vetor de coeficientes de regressão é  $\boldsymbol{\beta} = (1.8, -0.15)^\top$ .

**Cenário 3:** Repetimos aqui o Cenário 1 definido na Seção 2.3 em que supomos erro de medida com distribuição Normal. Relembramos que os tempos de falha  $\mathbf{Y} = (y_1, \dots, y_n)$  foram gerados da distribuição Weibull( $\alpha, \lambda$ ), em que  $\alpha = 1$ ,  $\lambda = -1.2$ , estabelecendo uma censura de 30%; a matriz de delineamento de experimento é dada por  $\mathbf{W} = (\mathbf{X}_1, \mathbf{X}_2 + \boldsymbol{\epsilon})$ , em que  $X_1 \sim \text{Bernoulli}(0.4)$ ,  $X_2 \sim \text{Normal}(\mu_x, \sigma_x^2)$ , com  $\mu_x = 3.9$  e  $\sigma_x^2 = 10$ ; e o erro de medida para cada indivíduo é  $\epsilon_i \stackrel{i.i.d.}{\sim} N(0, \sigma_\epsilon^2)$ , com  $\sigma_\epsilon^2 = 1$ ; o vetor de coeficientes de regressão é  $\boldsymbol{\beta} = (1.8, -0.15)$ .

Em todos os estudos de simulação exibidos a seguir, consideramos 250 réplicas de Monte Carlo. No MCMC, utilizamos *lag* de tamanho 50 sem *burn-in*, uma vez que consideramos como valor inicial da cadeia o valor real dos parâmetros, da variável misturadora e variável latente, tal como consideramos nos estudos de simulação do Capítulo 2. Antes de realizarmos esse estudo, verificamos a convergência das cadeias, partindo de valores iniciais aleatórios (ver apêndice H). Em cada caso, fizemos um total de 100000 iterações, selecionando uma amostra da distribuição *a posteriori* de cada parâmetro de tamanho 2000, com a qual calculamos as estimativas *a posteriori* em todos os modelos analisados. Tal como na Seção 2.3, os programas implementados foram executados no software livre R (R Core Team, 2016).

### 3.3.1 Estudo 1: O Efeito de Supor $\nu$ Conhecido

No estudo de simulação apresentado nesta seção temos como objetivo avaliar as consequências de se assumir os graus de liberdade conhecido na distribuição do erro de medida. Para tanto, consideremos cinco valores distintos para fixarmos os graus de liberdade, supondo  $\nu = 2.5, 3, 5, 10$  e  $30$ . Na Figura 3.1 a seguir, temos uma comparação entre as funções densidade de probabilidade das distribuições  $t$ -Student obtidas ao fixarmos  $\nu$  em tais valores. Observe que modelos com cauda mais pesada são fornecidos a medida que diminuimos o valor de  $\nu$ . Portanto, aquele no qual assumimos  $\nu = 2.5$  é o que possui a cauda mais pesada entre as opções, sendo o mais adequado em situações com dados contendo valores atípicos.

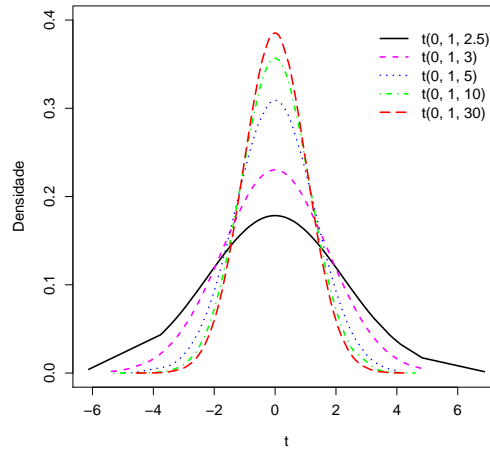


Figura 3.1: Comparação entre as f.d.p da distribuição  $t$ -Student considerando  $\nu = 2.5, 3, 5, 10$  e  $30$ .

Para o que segue, consideremos os dados do Cenário 1 descrito nesta seção. Tais dados serão analisados ajustando-se os modelos MBFC-NI1, MBFC-NI2 e MBFC-NI3. As estimativas médias obtidas no estudo de Monte Carlo para o valor esperado *a posteriori* de todos os parâmetros dos modelos são dadas na Tabela 3.1. Neste estudo de simulação, supomos *a priori* que  $\beta \propto 1$ ,  $\lambda \sim N(0, 100)$ ,  $\alpha \sim G(1, 1)$ ,  $X_{\epsilon_i} \sim N(\mu_x, \sigma_x^2)$ ,  $\mu_x \sim N(5, 10)$ ,  $\sigma_x^2 \sim GI(102, 1010)$  e  $U_{\epsilon_i} \sim G\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ , em que  $\nu$  é assumido conhecido. Tais escolhas de distribuições *a priori* foram feitas de forma análoga àquelas que fizemos no estudo de simulação da Seção 2.3. No modelo MBFC-NI1, além de fixarmos  $\nu$  como uma constante conhecida, também admitimos que o parâmetro de escala da distribuição do erro de medida  $\sigma_\epsilon^2$  é conhecido. Enquanto que nos modelos MBFC-NI2 e MBFC-NI3 exploramos a estimação de  $\sigma_\epsilon^2$ . No modelo MBFC-NI2, eliciamos  $\sigma_\epsilon^2 \sim GI(1, 1)$  e no MEFC-NI3 (no qual  $\sigma_\epsilon^2 = k\sigma_x^2$ ) supomos  $k \sim U(0, 0.5)$ . A eliciação de distribuição *a priori* para  $k$  e  $\sigma_x^2$  foram exploradas no estudo de simulação do MBFC com



erro Normal nas Subseções 2.3.1 e 2.3.2, e consideramos as conclusões obtidas em tais estudos para determinar aqui as distribuições *a priori* de tais parâmetros.

Tabela 3.1: Média, Viés e EQM das estimativas *a posteriori* obtidas ajustando-se os modelos MBFC-NI1, considerando  $\sigma_\epsilon^2 = 1$ , MBFC-NI2 e MBFC-NI3, supondo  $\nu = 2.5, 3, 5, 10$  e  $30$ , com os dados do Cenário 1( em que  $\nu = 3$ ).

Modelo	MBFC-NI1				MBFC-NI2			MBFC-NI3			
	$\eta_{(-\nu)}$	Real	Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM
$\nu = 2.5$	$\beta_1$	1.8	1.806	0.006	0.362	1.815	0.015	0.394	2.037	0.237	1.208
$\nu = 3$			1.786	-0.014	0.355	1.804	0.004	0.360	1.898	0.098	0.543
$\nu = 5$			1.758	-0.042	0.358	1.798	-0.002	0.513	1.787	-0.013	0.358
$\nu = 10$			1.758	-0.042	0.358	1.798	-0.002	0.513	1.761	-0.039	0.360
$\nu = 30$			1.758	-0.042	0.358	1.759	-0.041	0.361	1.758	-0.042	0.355
$\nu = 2.5$	$\beta_2$	-0.15	-0.146	0.004	0.008	-0.148	0.002	0.010	-0.235	-0.085	0.080
$\nu = 3$			-0.134	0.016	0.007	-0.142	0.007	0.008	-0.186	-0.036	0.023
$\nu = 5$			-0.113	0.037	0.007	-0.129	0.021	0.017	-0.135	0.015	0.008
$\nu = 10$			-0.113	0.037	0.007	-0.129	0.021	0.017	-0.116	0.034	0.007
$\nu = 30$			-0.112	0.038	0.007	-0.113	0.037	0.007	-0.113	0.037	0.007
$\nu = 2.5$	$\alpha$	1	1.106	0.106	0.054	1.110	0.110	0.010	1.185	0.185	0.124
$\nu = 3$			1.097	0.097	0.051	1.106	0.106	0.051	1.142	0.142	0.078
$\nu = 5$			1.083	0.083	0.046	1.096	0.096	0.057	1.098	0.098	0.052
$\nu = 10$			1.083	0.083	0.046	1.096	0.096	0.057	1.084	0.084	0.047
$\nu = 30$			1.082	0.082	0.046	1.082	0.082	0.046	1.082	0.082	0.046
$\nu = 2.5$	$\lambda$	-1.2	-1.318	-0.118	0.508	-1.317	-0.117	0.510	-1.149	0.051	0.597
$\nu = 3$			-1.356	-0.156	0.506	-1.334	-0.134	0.512	-1.213	-0.013	0.563
$\nu = 5$			-1.431	-0.231	0.518	-1.401	-0.201	0.511	-1.354	-0.154	0.513
$\nu = 10$			-1.431	-0.231	0.518	-1.401	-0.201	0.511	-1.421	-0.221	0.521
$\nu = 30$			-1.436	-0.236	0.515	-1.433	-0.233	0.522	-1.433	-0.233	0.514
$\nu = 2.5$	$\mu$	3.9	4.246	0.346	0.145	4.246	0.346	0.144	4.252	0.352	0.148
$\nu = 3$			4.246	0.346	0.108	4.245	0.345	0.149	4.250	0.350	0.147
$\nu = 5$			4.257	0.357	0.160	4.253	0.353	0.155	4.249	0.349	0.149
$\nu = 10$			4.256	0.356	0.160	4.253	0.353	0.155	4.257	0.357	0.160
$\nu = 30$			4.257	0.357	0.160	4.258	0.358	0.161	4.258	0.358	0.161
$\nu = 2.5$	$\sigma_x^2$	10	9.619	-0.381	0.193	9.657	-0.343	0.154	8.838	-1.162	1.375
$\nu = 3$			9.791	-0.209	0.145	9.730	-0.270	0.073	9.120	-0.880	0.811
$\nu = 5$			10.271	0.271	0.288	10.112	0.112	0.123	9.797	-0.203	0.125
$\nu = 10$			10.271	0.271	0.288	10.112	0.112	0.123	10.199	0.199	0.239
$\nu = 30$			10.286	0.286	0.300	10.276	0.276	0.276	10.283	0.283	0.297
$\nu = 2.5$	$\sigma_\epsilon^2$	1	-	-	-	0.994	-0.006	0.071	4.292	3.292	10.842
$\nu = 3$			-	-	-	1.413	0.413	1.309	4.442	3.442	11.854
$\nu = 5$			-	-	-	3.443	2.443	13.091	4.784	3.784	14.340
$\nu = 10$			-	-	-	8.963	7.963	386.611	4.981	3.981	15.895
$\nu = 30$			-	-	-	9.794	8.794	754.364	5.020	4.020	16.214
$\nu = 2.5$	$Var(\epsilon)$	3	5.000	-	-	4.970	-	-	21.460	-	-
$\nu = 3$			3.000	-	-	4.239	-	-	13.326	-	-
$\nu = 5$			1.667	-	-	5.739	-	-	7.975	-	-
$\nu = 10$			1.250	-	-	11.204	-	-	6.226	-	-
$\nu = 30$			1.071	-	-	10.489	-	-	5.376	-	-

### Efeitos nas Estimativas dos Parâmetros

As estimativas exibidas na Tabela 3.1 sugerem que o efeito de se assumir graus de liberdade conhecido é nulo ou insignificante para os parâmetros  $\beta_1, \alpha$  e  $\lambda$ , pois ao assumirmos diferentes valores para  $\nu$  os modelos fornecem estimativas muito similares, com viés e EQM bem próximos em todas as versões do MBFC-NI. Enquanto que para os parâmetros da variável latente e do erro de medida ( $\mu_x, \sigma_x^2$  e  $\sigma_\epsilon^2$ ) quando consideramos  $\nu$  fixado em um valor muito maior que o real, obtemos estimativas com maior viés em ambos os modelos. Percebemos ainda que o EQM obtido para a estimativa média de  $\sigma_\epsilon^2$  é muito maior quando assumimos que o erro de medida tem graus de liberdade muito maior que o verdadeiro. Por essa razão, focamos parte de nossa análise a seguir na estimação de tal parâmetro.

Os bancos de dados analisados no presente estudo de simulação foram gerados assumindo erro de medida  $t$ -Student com  $\nu = 3$  graus de liberdade e  $\sigma_\epsilon^2 = 1$ , desse modo a variância do erro é igual a 3. Na Tabela 3.1 podemos observar que os valores desta variância no modelo MBFC-NI1 é fixada de acordo com os valores de  $\nu$ . Nesse modelo, assumimos que  $\sigma_\epsilon^2 = 1$  para todo  $\nu$ . Quando fixamos  $\nu$  em um valor diferente do real, a variância do erro é subestimada ou superestimada a medida que diminuimos ou aumentamos, respectivamente, a densidade da cauda da distribuição do erro de medida. O MBFC-NI2 é o modelo que fornece melhores estimativas para tal variância. Podemos observar que quando assumimos  $\nu = 3$  ou supomos uma cauda mais pesada para a distribuição do erro, ao fixarmos  $\nu = 2.5$ , o modelo fornece estimativas mais próximas do valor real da variância desta distribuição. O fato que mais chama atenção nesses resultados é que as estimativas fornecidas para  $\sigma_\epsilon^2$  são infladas, exceto no caso em que fixamos  $\nu = 2.5$ , além disso, o viés e EQM aumentam a medida em que assumimos uma distribuição com cauda mais leve para a distribuição do erro de medida (mesmo quando fixamos  $\nu$  em seu valor real), o que nos parece uma forma de ajustar a variância que é subestimada quando assumimos distribuições com cauda mais leve que a real. Na Figura 3.2, podemos visualizar o comportamento das estimativas fornecidas para  $\sigma_\epsilon^2$  ajustando-se o MBFC-NI2 quando supomos  $\nu = 2.5, 3, 5, 10$  e  $30$ .

Observando a Figura 3.2 podemos ver que para  $\nu = 5, 10$  e  $30$  temos muitos valores discrepantes dentre as estimativas de  $\sigma_\epsilon^2$ , principalmente nos casos em que  $\nu$  é fixado em  $10$  e  $30$ . Os *box-plots* mostram que ao assumirmos cauda muito leve para a distribuição do erro de medida, quando a distribuição real possui cauda bastante pesada, os modelos ajustados tendem a estimar valores muito maiores do que o valor real do parâmetro de escala  $\sigma_\epsilon^2$ , parecendo ser uma compensação à estimação da variância do erro.

No estudo de simulação apresentado na Seção 2.3, concluímos que o modelo MBFC-NI3 com erro Normal é sensível a escolha da distribuição *a priori* de  $k$ . O mesmo parece ocorrer no MBFC-NI3 com erro  $t$ -Student, visto os valores que as estimativas obtidas para  $\sigma_\epsilon^2$ , supondo que *a priori*  $k \sim U(0, 0.5)$ , são semelhantes aquelas obtidas sob a mesma suposição no MBFC3

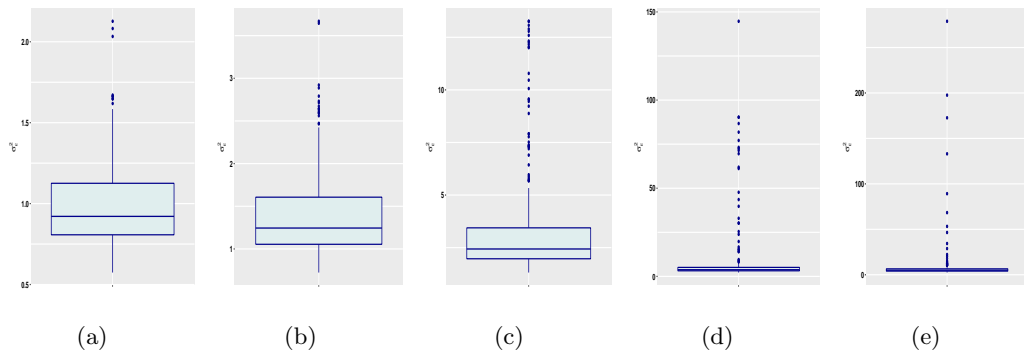


Figura 3.2: *Box-plots* das estimativas médias para  $\sigma_\epsilon^2$  ajustando-se o modelo MBFC-NI2. Os gráficos (a), (b), (c), (d) e (e) são relativos aos ajustes supondo  $\nu = 2.5, 3, 5, 10$  e  $30$ , respectivamente.

(ver Subseção 2.3.5). Ao ajustarmos o MBFC-NI3, no presente estudo de simulação, obtivemos estimativas para  $\sigma_\epsilon^2$  aproximadamente iguais em todos os casos analisados. Entretanto, se avaliarmos as estimativas fornecidas para a variância do erro de medida, observamos que considerando uma distribuição com cauda mais pesada para o erro, fixando  $\nu = 2.5$ , ou quando este parâmetro é fixado em seu valor real  $3$ , a variância do erro é superestimada. Por outro lado, quando assumimos  $\nu = 5, 10$  e  $30$  a variância é subestimada, ao contrário do que observamos ajustando o MBFC-NI2. Contudo, na Figura 3.3, em que exibimos *box-plots* das estimativas médias para o valor esperado *a posteriori* de  $\sigma_\epsilon^2$ , observamos um comportamento semelhante ao da Figura 3.2 para as estimativas via MEFC-NI2. Desse modo, notamos que ao assumirmos uma distribuição com cauda leve para modelarmos o erro de medida perdemos bastante a precisão na estimação de  $\sigma_\epsilon^2$  e o modelo ajustado tende a fornecer estimativas discrepantes para tal parâmetro.

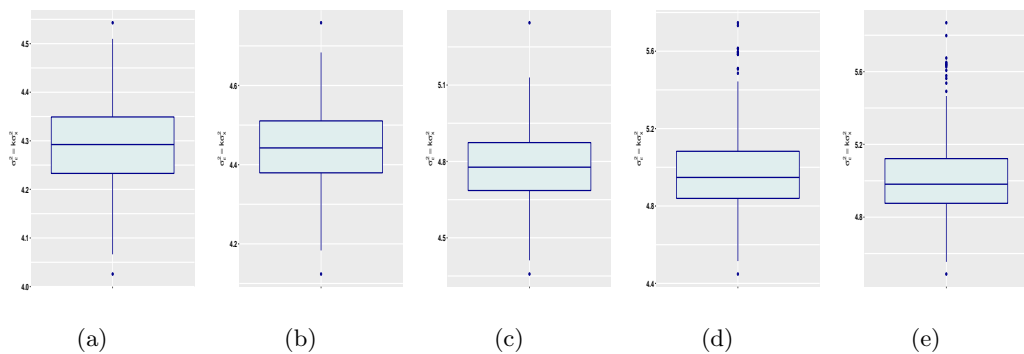


Figura 3.3: *Box-plots* das estimativas médias para  $\sigma_\epsilon^2$  ajustando-se o modelo MBFC-NI3. Os gráficos (a), (b), (c), (d) e (e) são relativos aos ajustes supondo  $\nu = 2.5, 3, 5, 10$  e  $30$ , respectivamente.

Em relação a estimação das variáveis latentes, na Tabela 3.1 notamos que, em geral, as

estimativas para  $\mu_x$  e  $\sigma_x^2$  são aproximadamente iguais com o MBFC-NI1 e o MBFC-NI2. O MBFC-NI3 apresenta comportamento diferente, fornecendo estimativas mais enviesadas para  $\mu_x$  quando assumimos  $\nu = 2.5$  e  $3$  do que quando fixamos tal parâmetro em  $5$ , o que não é esperado. Para a variância das variáveis latentes o melhor ajuste é fornecido quando fixamos  $\nu = 10$ , outro resultado bastante inesperado. Para visualizar o comportamento das estimativas para a variável latente de cada indivíduo, nas Figuras 3.4 a 3.6, exibimos diagramas de dispersão nos quais relacionamos as médias obtidas para o valor esperado *a posteriori* da variável latente de cada indivíduo com os valores reais de tal variável. Nestes gráficos, consideramos o ajuste dos modelos MBFC-NI1, MBFC-NI2 e MBFC-NI3, assumindo os diferentes valores para  $\nu$ . Nosso objetivo é avaliar se as estimativas fornecidas para a variável latente são afetadas pelas suposições a respeito dos graus de liberdade da distribuição do erro de medida.

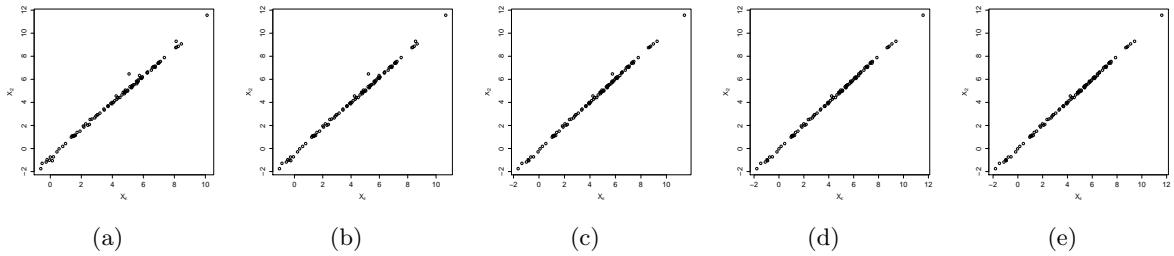


Figura 3.4: No eixo das abscissas temos as estimativas para  $\mathbf{X}_\epsilon$  ajustando-se o MBFC-NI1 considerando o Cenário 1, supondo  $\nu = 2.5, 3, 5, 10$  e  $30$  nos gráficos (a), (b), (c), (d) e (e), respectivamente.

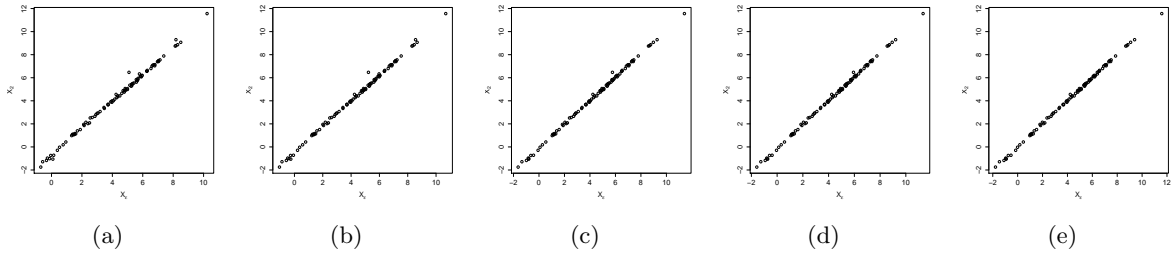


Figura 3.5: No eixo das abscissas temos as estimativas para  $\mathbf{X}_\epsilon$  ajustando-se o MBFC-NI2 considerando o Cenário 1, supondo  $\nu = 2.5, 3, 5, 10$  e  $30$  nos gráficos (a), (b), (c), (d) e (e), respectivamente.

Nas Figuras 3.4 a 3.6, podemos observar que o comportamento das estimativas para as variáveis latentes é bastante semelhante sob o ajuste dos modelos MBFC-NI1 e MBFC-NI2, tal como foi constatado na Tabela 3.1 para a média e variância de tais variáveis. Analogamente, observamos sob o ajuste do MBFC-NI3, uma maior dispersão nos casos em que assumimos  $\nu = 2.5$  e  $3$ , o que confirma as conclusões obtidas ao analisarmos a média e a variância de tais

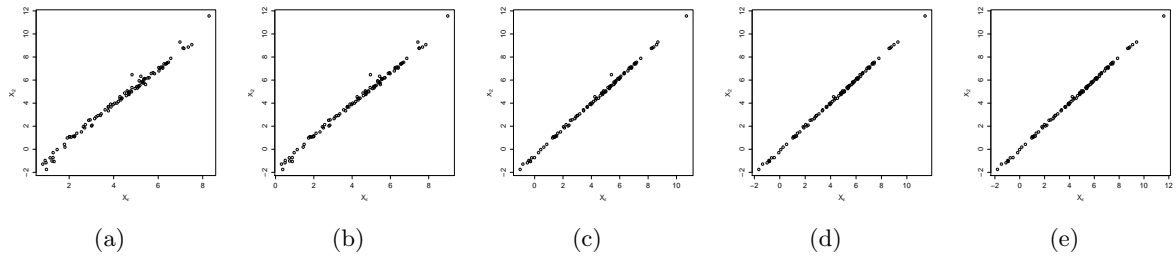


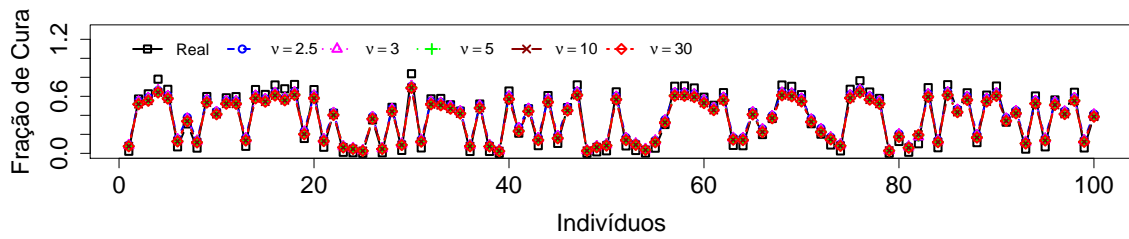
Figura 3.6: No eixo das abscissas temos as estimativas para  $X_\epsilon$  ajustando-se o MBFC-NI3 considerando o Cenário 1, supondo  $\nu = 2.5, 3, 5, 10$  e  $30$  nos gráficos (a), (b), (c), (d) e (e), respectivamente.

variáveis.

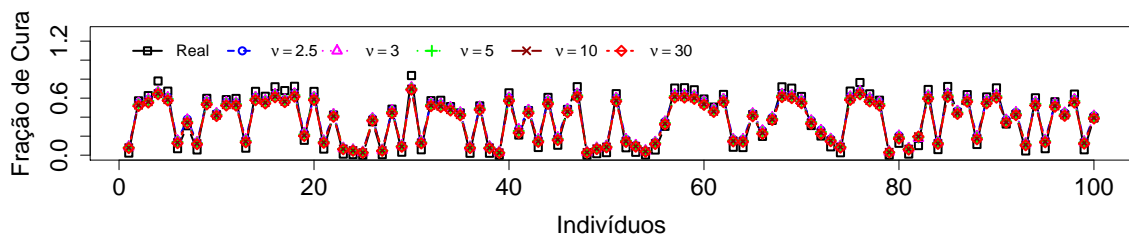
### Efeito nas Estimativas da Fração de Cura

Para avaliar os efeitos de supormos os graus de liberdade  $\nu$ , da distribuição do erro de medida, conhecido nas estimativas para a fração de cura, exibimos na Figura 3.7, gráficos nos quais comparamos o valor real da fração de cura de cada indivíduo com as estimativas médias para o seu valor esperado *a posteriori*. Na Figura 3.8, temos o viés de tais estimativas via modelos MBFC-NI1, 2 e 3, considerando  $\nu = 2.5, 3, 5, 10$  e  $30$ . Além disso, na Figura 3.9, podemos visualizar o histograma da fração de cura real superposto pelas curvas das distribuições empíricas das médias dos valores esperados *a posteriori* da fração de cura, considerando os três modelos, supondo os diferentes valores estabelecidos para  $\nu$  no presente estudo de simulação. Para finalizar nossa análise, na Tabela 3.2, apresentamos o percentual de cobertura dos intervalos de 95% de credibilidade (HPD 95%) para a fração de cura.

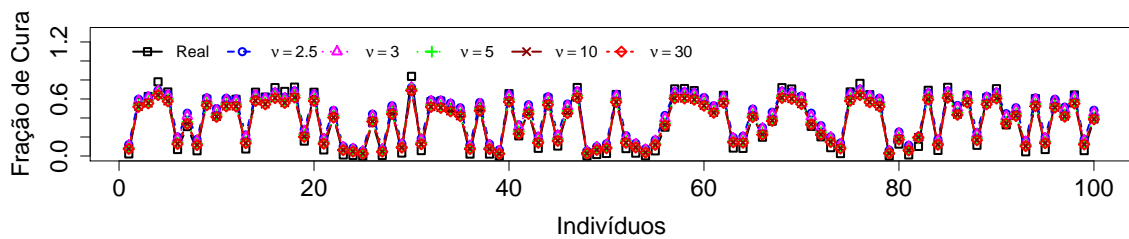
Observando a Figura 3.7 não é possível notar qualquer diferença entre os gráficos 3.7a e 3.7b, referentes a comparação entre as estimativas para a fração de cura sob os modelos MBFC-NI1 e MBFC-NI2 com os valores reais da fração de cura. O mesmo acontece ao compararmos os vieses dessas estimativas nas Figuras 3.8a e 3.8b. Isso nos mostra que quando fixamos o parâmetro de escala  $\sigma_\epsilon^2$  em seu valor verdadeiro (suposição do MBFC-NI1) temos estimativas aproximadamente iguais àquelas obtidas quando assumimos um modelo em que não inserimos informação a respeito de tal parâmetro, como foi considerado no MBFC-NI2. Neste segundo modelo, assumimos uma distribuição *a priori* não informativa  $GI(1, 1)$  para  $\sigma_\epsilon^2$ . Em relação as estimativas obtidas ajustando-se o MBFC-NI3 podemos notar um comportamento levemente diferente daquele sob o ajuste dos outros dois modelos, tal como observamos na Tabela 3.1 para os demais parâmetros dos modelos analisados. Podemos concluir que, de um modo geral, todos os modelos ajustados forneceram boas estimativas para a fração de cura de maneira bem similar. Os gráficos na Figura 3.8 são muito parecidos e esboçam viés próximo a zero para a fração de cura de todos os indivíduos. A Figura 3.9 a seguir, consolida estas conclusões mostrando que, para



(a)



(b)

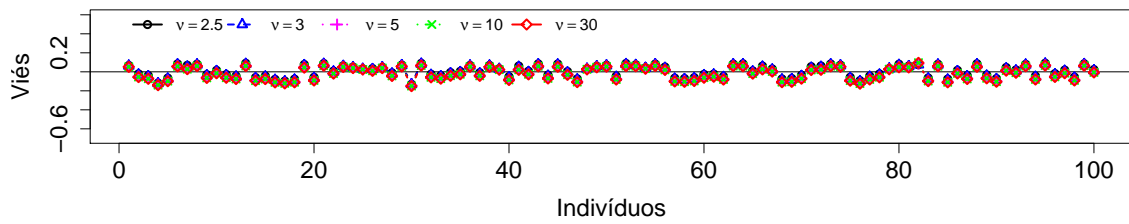


(c)

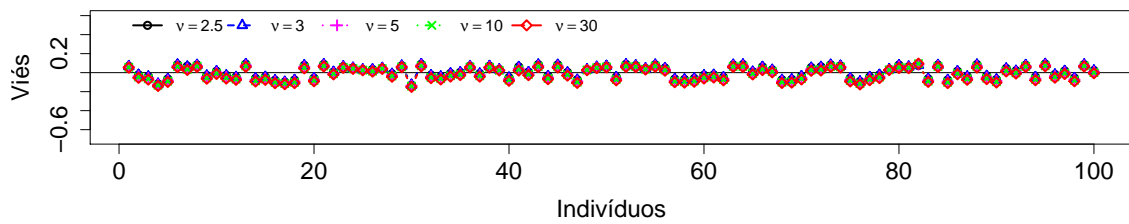
Figura 3.7: Estimativas médias do valor esperado *a posteriori* para a fração de cura de cada um dos 100 indivíduos, ajustando-se o MBFC-NI1 (a), o MBFC-NI2 (b) e o MBFC-NI3 (c) considerando o Cenário 1, assumindo  $\nu = 2.5, 3, 5, 10$  e 30

todos os valores estabelecidos para  $\nu$ , os modelos MBFC-NI 1 e 2 apresentam um comportamento aproximadamente igual nas estimativas fornecidas.

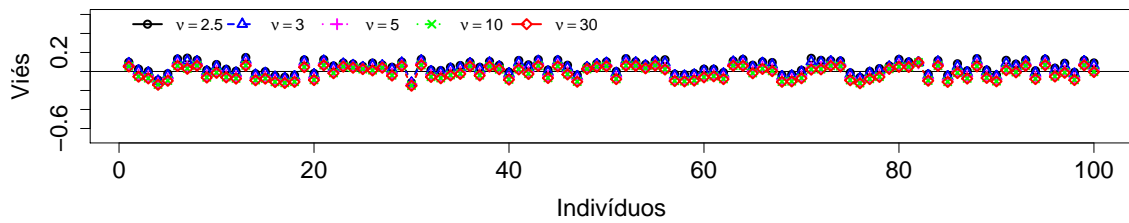
Na Figura 3.9c, podemos ver que para os indivíduos que possuem fração de cura entre 0.3 e 1 a curva que melhor se ajusta ao histograma é aquela das estimativas obtidas com o MBFC-NI3 assumindo-se  $\nu = 2.5$ . Além disso, sob os ajustes do MBFC-NI1 e e do MBFC-NI2 a conclusão para tais indivíduos é similar, isto é, as curvas que melhor aproximam o histograma são aquelas



(a)



(b)



(c)

Figura 3.8: Viés das estimativas do valor esperado *a posteriori* para a fração de cura de cada um dos 100 indivíduos, ajustando-se o MBFC-NI1 (a), o MBFC-NI2 (b) e o MBFC-NI3 (c) considerando o Cenário 1, assumindo  $\nu = 2.5, 3, 5, 10$  e 30

das estimativas obtidas quando assume-se  $\nu = 2.5$ , indicando que o modelo com cauda mais pesada fornece melhores estimativas para a fração de cura dos indivíduos no Cenário 1. Sob todos os modelos ajustados, o número de indivíduos que possuem fração de cura de até 10% é subestimado, já entre 10% e 30% estes valor é superestimado.

Ainda a respeito das estimativas obtidas para a fração de cura, neste estudo de simulação, podemos ver na Tabela 3.2 que os modelos ajustados apresentam um maior percentual de

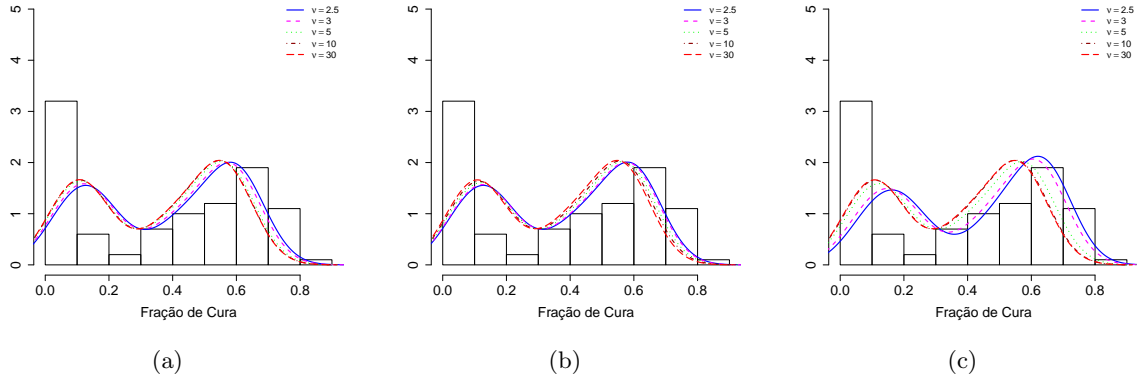


Figura 3.9: Histograma das frações de cura reais superposto pelas distribuições empíricas das médias dos valores esperados *a posteriori* da fração de cura, considerando o Cenário 1 e ajustando-se os modelos MBFC-NI1, MBFC-NI2 e MBFC-NI3, nos gráficos (a), (b) e (c), respectivamente, para os diferentes valores de  $\nu$ .

intervalos de credibilidade contendo o verdadeiro valor quando fixamos  $\nu = 2.5$  (isto é, quando assumimos um modelo com cauda mais pesada para o erro de medida). Este percentual é bem menor quando  $\nu$  é fixado em 30 (modelo com cauda mais leve), e isto ocorre nas três modelagens propostas para o MBFC-NI. De acordo com os percentuais observados na referida tabela, vemos que modelar o erro de medida supondo uma distribuição com cauda mais pesada proporcionou uma maior chance de cobertura do verdadeiro valor da fração de cura. Além disso, para os valores  $\nu \leq 5$  o MBFC-NI3 apresenta uma maior proporção de *HPD's* que contém a fração de cura real dos indivíduos, seguido pelo MBFC-NI2.

Tabela 3.2: Percentual de cobertura dos intervalos de 95% de credibilidade para a fração de cura, considerando o Cenário 1 e ajustando-se os modelos MBFC-NI1, MBFC-NI2 e MBFC-NI3 para os diferentes valores de  $\nu$ .

$\nu$	Modelo		
	MBFC-NI1	MBFC-NI2	MBFC-NI3
2.5	0.918	0.916	0.943
3	0.901	0.910	0.935
5	0.842	0.881	0.902
10	0.803	0.840	0.830
30	0.793	0.800	0.796

De um modo geral, tendo em vista que os dados do Cenário 1 foram gerados supondo erro com distribuição *t-Student* com  $\nu = 3$  graus de liberdade, o esperado é que os modelos MBFC-NI1, 2 e 3 nos quais assumimos  $\nu = 3$  ajustem-se melhor aos dados, entretanto isto não foi observado. Ajustando-se o MBFC-NI1 obtivemos estimativas menos enviesadas para os parâmetros  $\beta$ ,  $\lambda$  e  $\mu_x$  quando assumimos  $\nu = 2.5$ . Para  $\alpha$  e  $\sigma_x^2$ , estimativas menos enviesadas



foram fornecidas ao fixarmos  $\nu = 3$  (modelo real dos dados) neste modelo. Analisando as estimativas ao ajustarmos o MBFC-NI2, verificamos um comportamento análogo ao que foi percebido sob o ajuste do MBFC-NI1 para os parâmetros  $\beta_2$  e  $\lambda$ . Além destes, uma estimativa com menor viés foi obtida também para  $\sigma_\epsilon^2$ . Para  $\beta_1$  e  $\mu_x$ , o MBFC-NI2 no qual fixamos  $\nu = 3$  forneceu estimativas menos enviesadas e também com menor EQM para a variância  $\sigma_x^2$ . O pior ajuste do Modelo MEFC-NI3. Comparado-o aos demais modelos, este forneceu estimativas com viés e EQM elevados, principalmente para os parâmetros diretamente relacionados ao erro de medida:  $\sigma_\epsilon^2$ ,  $\sigma_x^2$  e  $\mu_x$ . Atribuímos este comportamento a eliciação *a priori* da distribuição da constante  $k$ , a qual está diretamente relacionada a estimação de  $\sigma_\epsilon^2$ . Vimos no Estudo 5, apresentado na Subseção 2.3.5, que o modelo MBFC3 (em que assumimos  $\sigma_\epsilon^2 = k\sigma_x^2$ ) é bastante sensível a escolha da distribuição *a priori* de  $k$ , sendo mais interessante utilizar uma distribuição bem informativa para tal parâmetro. A escolha da distribuição *a priori* Uniforme para  $k$  no presente estudo e nos que apresentamos a seguir, foi feita por questões computacionais, visto que experimentamos problemas numéricos no algoritmo MCMC ao usar distribuições *a priori* Gama pouco ou muito informativas para tal parâmetro. Este fato deverá ser objeto de estudo em trabalhos futuros. Um fato inesperado diante das estimativas obtidas para os parâmetros do modelo é que ao estimarmos a fração de cura sob o ajuste do MEFC-NI3, obtivemos melhores estimativas, em geral, que os demais modelos sempre que assumimos uma distribuição com cauda mais pesada para o erro de medida. Isto deve ocorrer devido a alta variabilidade estimada para o erro de medida.

### 3.3.2 Estudo 2: O Efeito de Pressupor Erro Quando Não Existe

Neste estudo de simulação analisamos os dados do Cenário 2 que não possuem erro de medida, comparando os modelos MBFC-NI1, MBFC-NI2, MBFC-NI3 e o MTP (Chen *et al.*, 1999) definido na Subseção 1.3.1. Como foi considerado no Estudo 1, apresentado na Subseção 3.3.1, no presente estudo também admitimos  $\nu$  conhecido na distribuição do erro de medida, mas desta vez fixamos tal parâmetro em apenas dois valores: 2.5 e 30. Fazemos isso com o objetivo de analisar o efeito de assumirmos cauda muito pesada ( $\nu = 2.5$ ) ou muito leve ( $\nu = 30$ ) para a distribuição do erro de medida, em um contexto no qual não existe qualquer covariável medida com erro. As distribuições *a priori* assumidas para os parâmetros do modelo foram iguais as utilizadas no Estudo 1 deste capítulo (ver Subseção 3.3.1), com exceção do parâmetro  $\alpha$  que optamos por eliciar uma distribuição pouco informativa  $G(0.5, 0.001)$ . Além de fixarmos  $\nu = 2.5$  e 30, no modelo MBFC-NI1 também fixamos  $\sigma_\epsilon^2 = 1$ .

Na Tabela 3.3 a seguir, exibimos as estimativas obtidas para os parâmetros dos modelos. Tais estimativas são referentes à média, viés e erro quadrático médio (EQM) do valor esperado

*a posteriori* de cada parâmetro.

Tabela 3.3: Média, Viés e EQM das estimativas *a posteriori* obtidas ajustando-se os modelos MTP, MBFC-NI1, considerando  $\sigma_\epsilon^2 = 1$ , MBFC-NI2 e MBFC-NI3, supondo  $\nu = 2.5$  e 30, com os dados do Cenário 2.

Modelo		MTP				MBFC-NI1			MBFC-NI2			MBFC-NI3		
$\nu$	$\eta_{(-\nu)}$	Real	Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM
2.5	$\beta_1$	1.798	-0.002	0.306		1.943	0.143	1.221	1.905	0.105	0.766	2.286	0.486	2.538
30						1.803	0.003	0.305	1.806	0.006	0.306	1.805	0.005	0.303
2.5	$\beta_2$	-0.15	-0.128	0.022	0.006	-0.181	-0.031	0.051	-0.167	-0.017	0.027	-0.302	-0.152	0.145
30						-0.127	0.023	0.006	-0.125	0.025	0.006	-0.126	0.024	0.006
2.5	$\alpha$	1	1.065	0.065	0.044	1.105	0.105	0.099	1.073	0.073	0.052	1.177	0.177	0.128
30						1.065	0.065	0.045	1.046	0.046	0.038	1.046	0.046	0.038
2.5	$\lambda$	-1.2	-1.316	-0.116	0.358	-1.230	-0.030	0.420	-1.260	-0.060	0.390	-1.087	0.113	0.504
30						-1.322	-0.122	0.362	-1.340	-0.140	0.360	-1.338	-0.138	0.360
2.5	$\mu_x$	3.9	-	-	-	3.913	0.013	0.051	3.914	0.014	0.000	3.897	-0.003	0.001
30						3.915	0.015	0.000	3.916	0.016	0.000	3.914	0.014	0.000
2.5	$\sigma_x^2$	10	-	-	-	9.569	-0.431	0.187	9.648	-0.352	0.126	8.790	-1.210	1.466
30						10.048	0.048	0.003	10.049	0.049	0.003	10.042	0.042	0.002
2.5	$\sigma_\epsilon^2$	-	-	-	-	-	-	-	0.864	-	-	4.266	-	-
30						-	-	-	3.425	-	-	4.903	-	-
2.5	$Var(\epsilon)$	-	-	-	-	5.000	-	-	4.320	-	-	21.330	-	-
30						1.071	-	-	3.670	-	-	5.253	-	-

### Efeito nas Estimativas dos Parâmetros

Ao observarmos as estimativas de  $\beta_1, \beta_2, \alpha$  e  $\lambda$ , que são comuns aos quatro modelos ajustados, na Tabela 3.3, vemos que o parâmetro  $\beta_1$  é melhor aproximado quando o modelo ajustado é o MTP. Entretanto, ajustando-se o MBFC-NI1 e o MBFC-NI2 supondo  $\nu = 30$  obtivemos uma estimativa para  $\beta_1$  aproximadamente igual a fornecida pelo MTP com uma diferença de EQM de 0.001. O mesmo comportamento é observado para os demais parâmetros indicando que a qualidade do ajuste do MTP é aproximadamente igual ao do MBFC-NI1 assumindo  $\nu = 30$  na distribuição do erro de medida. Para o parâmetro  $\beta_2$ , obtemos estimativa com menor viés e EQM ao ajustamos o MBFC-NI2 com  $\nu = 2.5$ . Por outro lado,  $\alpha$  é melhor aproximado quando ajustamos tal modelo supondo  $\nu = 30$ . A melhor aproximação para  $\lambda$  foi obtida via MBFC-NI1 no qual fixamos  $\nu = 2.5$ . Para todos estes parâmetros, o modelo MBFC-NI3 forneceu estimativas mais enviesadas e quase sempre com maior EQM. Diante dessas considerações notamos que assumir os modelos MBFC-NI1 e 2, que supõem erro de medida nas covariáveis, pode ser apropriado mesmo em situações em que não há covariáveis com erro. Considerando o modelo MBFC-NI3, vimos que assumindo  $k \sim U(0, 0.5)$  *a priori*, obtivemos estimativas muito enviesadas para os parâmetros comuns aos quatro modelos, no presente contexto, não se mostrando muito apropriado na situação analisada. Em relação aos parâmetros  $\mu_x$  e  $\sigma_x^2$ , notamos que nas três versões do MBFC-NI as estimativas são muito parecidas para todos estes parâmetros, com

pouquíssima diferença entre os vícios e EQMs. Vale destacar que as melhores aproximações para os parâmetros são observadas quando supomos  $\nu = 2.5$  em todos os MBFC-NI. Embora isso seja observado, calculando a variância estimada para o erro de medida  $Var(\epsilon)$  obtemos uma melhor aproximação desta quantidade quando ajustamos o MBFC-NI1 com cauda leve. Em todos os outros casos esta variância é inflada. Isto acontece porque fixando o parâmetro de escala  $\sigma_\epsilon^2$ , a variância em um modelo  $t$ -Student com cauda leve é bem menor do que no modelo  $t$ -Student com cauda mais pesada.

Nas Figuras 3.10 e 3.11, apresentamos diagramas de dispersão nos quais relacionamos as estimativas médias dos valores esperados *a posteriori* para a variável latente de cada indivíduo com o seu valor real, considerando cada modelo com erro nas covariáveis ajustado. Na Figura 3.12, exibimos *box-plots* das estimativas médias do valor esperado a posteriori de  $\sigma_\epsilon^2$ , ajustando-se os modelos MEFC-NI2 e 3, com o objetivo de visualizarmos o comportamento das estimativas para tal parâmetro.

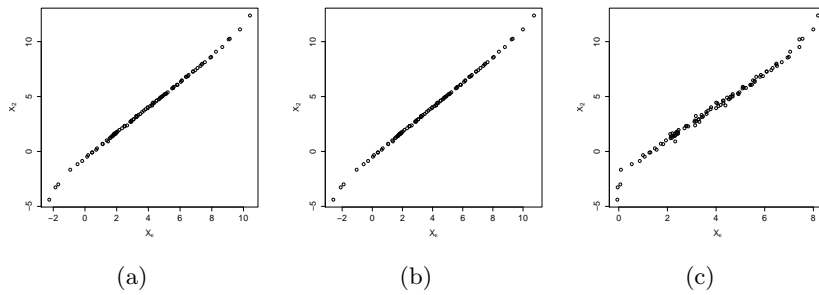


Figura 3.10: No eixo das abscissas temos as estimativas para  $X_\epsilon$  ajustando-se o MBFC-NI1, 2 e 3, considerando o Cenário 2, supondo  $\nu = 2.5$  nos gráficos (a), (b) e (c), respectivamente.

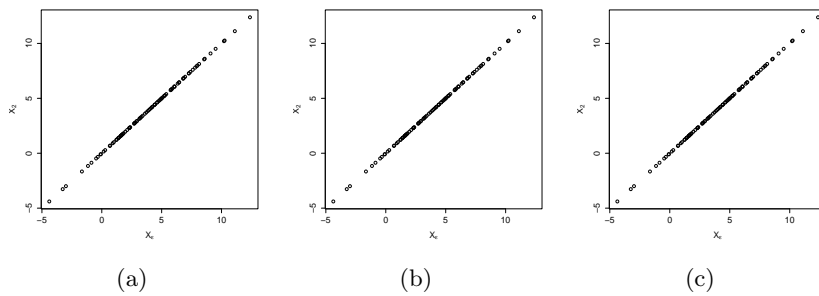


Figura 3.11: No eixo das abscissas temos as estimativas para  $X_\epsilon$  ajustando-se o MBFC-NI1, 2 e 3 considerando o Cenário 2, supondo  $\nu = 30$  nos gráficos (a), (b) e (c), respectivamente.

Comparando os valores estimados para  $X_\epsilon$  de cada indivíduo com o valor real  $X_2$ , na Figura

3.10 (assumindo  $\nu = 2.5$ ), vemos que há uma maior dispersão sob o ajuste do MBFC-NI3. Nos diagramas da Figura 3.11, observamos um comportamento muito similar para as estimativas obtidas sob os três ajustes, supondo  $\nu = 30$ . Visualmente não notamos qualquer diferença e as estimativas parecem aproximar muito bem os valores reais das variáveis latentes, sendo esta aproximação melhor que àquela na qual supomos  $\nu = 2.5$ . Deste modo, ao analisarmos as estimativas para os parâmetros da distribuição do erro de medida, incluindo as variáveis latentes, é difícil decidir se os modelos  $t$ -Student com cauda mais pesada de fato se ajustam melhor aos dados livres de erro de medida.

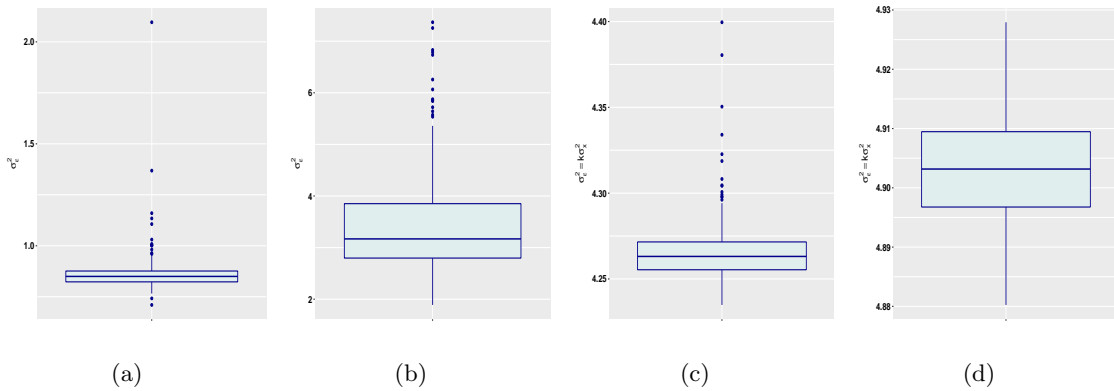
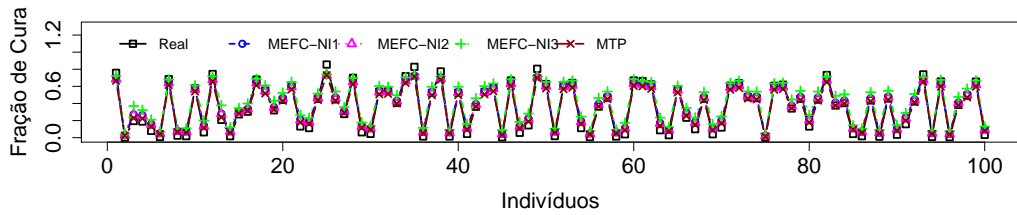


Figura 3.12: *Box-plots* das estimativas médias obtidas para  $\sigma_\epsilon^2$  ajustando-se os modelos MEFC-NI2 (Fig.a e b) e MEFC-NI3 (Fig.c e d), assumindo  $\nu = 2.5$  nas figuras (a) e (c), e  $\nu = 30$  nas figuras (b) e (d).

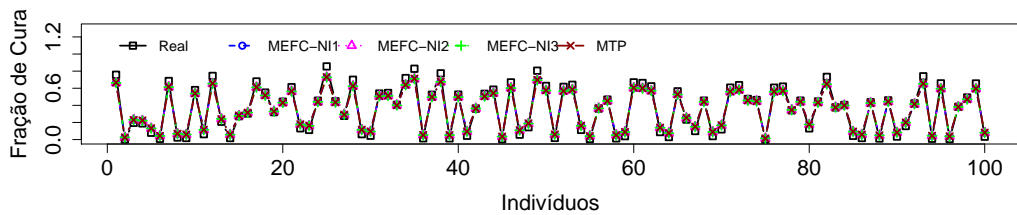
Na Figura 3.12 podemos observar como se comportam as estimativas médias obtidas para  $\sigma_\epsilon^2$  ajustando-se os modelos MBEFC-NI2 e MBFC-NI3. Com exceção do caso em que o modelo ajustado foi o MBFC-NI3 com  $\nu = 30$  (Fig.3.12d), vemos que a distribuição das estimativas médias é bem assimétrica, apresentando muitos valores atípicos maiores do que 75% dos valores obtidos. Nesse caso, a média utilizada para estimar o parâmetro de escala  $\sigma_\epsilon^2$  tende a ser superestimado por influência desses valores atípicos.

### Efeito nas Estimativas da Fração de Cura

Para analisar o efeito de se modelar dados livres de erro de medida nas estimativas da fração de cura, supondo erro nas covariáveis, exibimos alguns gráficos nas Figuras 3.13 a 3.15. Em tais gráficos, podemos visualizar um comparação entres as estimativas obtidas ajustando-se cada um dos quatro modelos analisados e avaliar o viés. Além disso, na Tabela 3.4, exibimos o percentual médio de cobertura dos intervalos de 95% de credibilidade para a fração de cura, obtidos ao ajustarmos cada um dos modelos aos 250 bancos do presente estudo de Monte Carlo.



(a)



(b)

Figura 3.13: Estimativas médias do valor esperado *a posteriori* para a fração de cura de cada um dos 100 indivíduos, ajustando-se os modelos MBFC-NI1, 2, 3 e MTP supondo  $\nu = 2.5$  (a) e  $\nu = 30$  (b).

Observando a Figura 3.13b, vemos que ao supormos  $\nu = 30$  nos modelos MBFC-NI não é possível notar diferença substancial entre as estimativas obtidas sob cada ajuste. Notamos ainda que os valores reais representados através de quadrados no gráfico são muito bem aproximados por todos os modelos analisados. Já na Figura 3.13a, na qual supomos  $\nu = 2.5$  nos modelos MBFC-NI, conseguimos distinguir bem as estimativas obtidas via MBFC-NI3 daquelas sob o MTP e os MBFC-NI1 e 2. Embora não consigamos distinguir esses três últimos modelos um do outro, eles parecem aproximar de maneira similar os valores reais da fração de cura, fornecendo estimativas mais próximas dos valores reais do que aquelas do MBFC-NI3. Tais impressões podem ser confirmadas ao observarmos o viés de cada estimativa em comparação com a linha horizontal passando pela origem na Figura 3.14.

Observemos na Figura 3.14a, na qual os resultados foram obtidos com  $\nu = 2.5$ , que não é possível distinguir o viés das estimativas ajustando-se os modelos MTP, MBFC-NI1 e 2, mas vemos uma certa diferença entre o viés observado ajustando tais modelos e aquele ajustando-se o MBFC-NI3. Nesse caso, o ajuste do MBFC-NI3 fornece estimativas mais enviesadas para uma maior quantidade de indivíduos em média. Já na Figura 3.14b,  $\nu = 30$  nos modelos MBFC-NI, não percebemos diferença entre os vícios obtidos aos ajustarmos os quatro modelos analisados. Com o objetivo de tentar observar alguma diferença nas estimativas obtidas por cada modelo

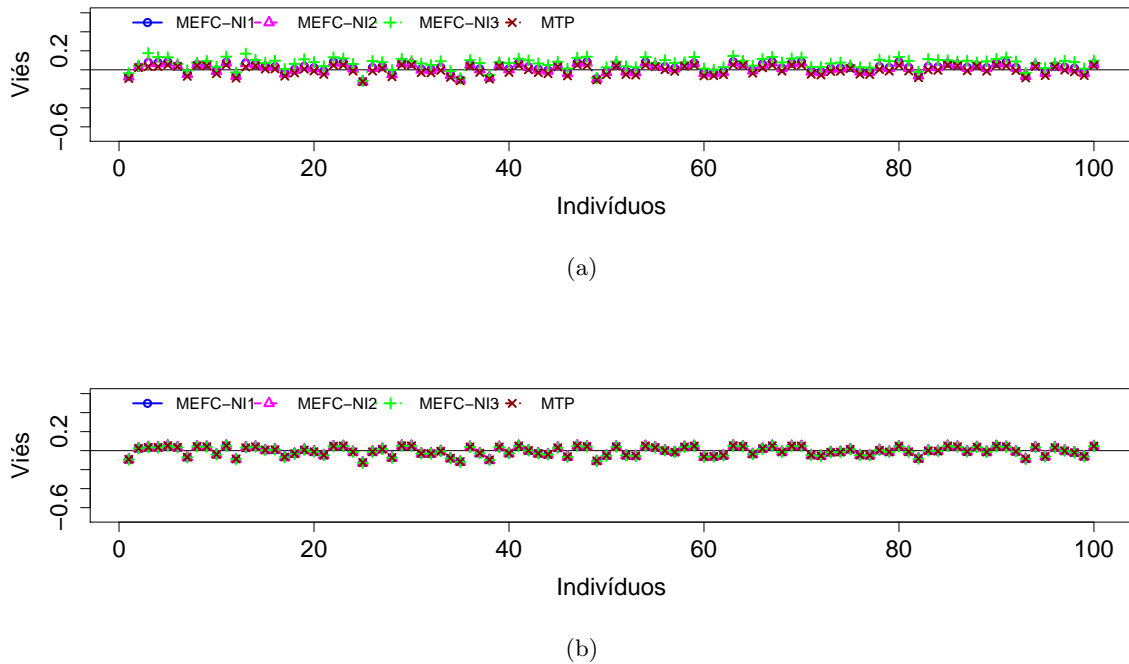


Figura 3.14: Viés médio das estimativas do valor esperado *a posteriori* para a fração de cura de cada um dos 100 indivíduos, ajustando-se os modelos MBFC-NI1, 2, 3 e MTP supondo  $\nu = 2.5$  (a) e  $\nu = 30$  (b).

ou evidenciar a semelhança entre os ajustes, vejamos na Figura 3.15 a seguir, o histograma das frações de cura reais superposto pelas distribuições empíricas das médias dos valores esperados *a posteriori* da fração de cura que foram obtidas ajustando-se os modelos MBFC-NI1, MBFC-NI2, MBFC-NI3 e MTP supondo  $\nu = 2.5$  (Fig.3.15a) e  $\nu = 30$  (Fig.3.15b).

Na Figura 3.15a, na qual assumimos  $\nu = 2.5$  nos modelos MBFC-NI, percebemos bastante diferença entre o histograma da fração de cura real e as curvas das distribuições empíricas das médias dos valores esperados *a posteriori* da fração de cura. Além disso, conseguimos distinguir cada curva obtida. Nota-se mais claramente que ajustando-se os modelos MBFC-NI1 e 2 obtemos distribuições empíricas aproximadamente iguais para o valor esperado da fração de cura. Podemos ver que o MTP (modelo real dos dados) só fornece uma melhor aproximação se comparada aos modelos MBFC-NI para os indivíduos com fração de cura de até 30%, embora para os indivíduos com até 10% de probabilidade de cura este valor seja subestimado, não só pelo MTP, mas por todos os modelos ajustados. Para os indivíduos com fração de cura acima de 60% o melhor ajuste observado é do MBFC-NI3. Olhando de um modo geral, este é o modelo cujo ajuste melhor aproximou o histograma da fração de cura através da distribuição empírica

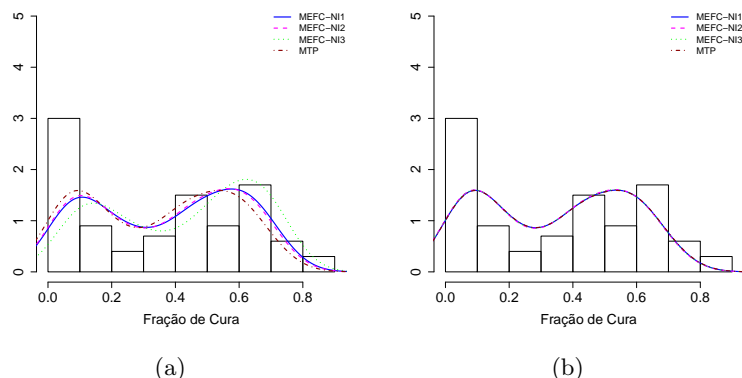


Figura 3.15: Histograma das frações de cura reais superposto pelas distribuições empíricas das médias dos valores esperados *a posteriori* da fração de cura, considerando o Cenário 2 e ajustando-se os modelos MBFC-NI1, MBFC-NI2, MBFC-NI3 e MTP supondo  $\nu = 2.5$  (a) e  $\nu = 30$  (b).

de suas estimativas. Em contrapartida, na Figura 3.15b, na qual assumimos  $\nu = 30$  nos modelos MBFC-NI ajustados, as distribuições empíricas são aproximadamente iguais àquela do MTP. As quatro curvas estão completamente sobrepostas. Nesse caso, a fração de cura de uma parte dos indivíduos é subestimada.

Na Tabela 3.4 a seguir, na qual exibimos os percentuais médios dos intervalos de credibilidade de 95% que contém o verdadeiro valor da fração de cura, obtidos ajustando-se os modelos MBFC-NI1, MBFC-NI2, MBFC-NI3 e MTP, assumindo  $\nu = 2.5$  ou 30, podemos ver que os MBFC-NIs fornecem um maior percentual de cobertura para o verdadeiro valor da fração de cura. Além disso, temos que quando assumimos que a distribuição do erro de medida tem cauda mais pesada (admitindo que  $\nu = 2.5$ ), os percentuais de cobertura são maiores ajustando-se qualquer um dos modelos MBFC-NI do que os percentuais observados quando supomos  $\nu = 30$ . Notemos também que assumindo  $\nu = 30$  nos MBFC-NIs, o percentual de cobertura dos intervalos fornecidos ajustando-se o MTP também é aproximadamente igual àqueles obtidos pelos modelos com erro nas covariáveis. Esse comportamento coincide com o observado na Figura 3.15.

Tabela 3.4: Percentual de cobertura dos intervalos de 95% de credibilidade para a fração de cura, considerando o Cenário 2 e ajustando-se os modelos MBFC-NI1, MBFC-NI2, MBFC-NI3 e MTP para os diferentes valores de  $\nu$ .

$\nu$	Modelo			
	MTP	MBFC-NI1	MBFC-NI2	MBFC-NI3
2.5	0.917	0.950	0.948	0.957
30	–	0.919	0.917	0.917

Através do presente estudo de simulação podemos concluir que a suposição de erro de medida nos modelos ajustados a dados livres de erro não afeta a qualidade das estimativas, visto que ao compararmos os resultados ajustando-se os modelos MBFC-NI1 a 3 com aqueles obtidos ajustando-se o MTP (o qual é o modelo de onde os dados foram gerados), as estimativas têm viés e EQM aproximadamente igual para os parâmetros. Além disso, a fração de cura foi melhor estimada em média através dos modelos MBFC-NI. Percebemos também que ao supormos um modelo com cauda mais pesada para a distribuição do erro de medida, as estimativas intervalares são significativamente melhores do que aquelas obtidas quando diminuimos a cauda de tal distribuição nos três modelos MBFC-NI. Vale ressaltar que, em todos os gráficos e estimativas pontuais analisadas, o comportamento dos quatro modelos pode ser considerado aproximadamente igual, porém, o modelo mais robusto ajustado que requer uma menor quantidade de informações *a priori* é o MBFC-NI2 no qual  $\nu = 2.5$ . Por essa razão, em situações práticas, tal modelo pode ser o mais interessante de se utilizar, dentre os analisados.

### 3.3.3 Estudo 3: Avaliando a Robustez do Modelo

Nesta subseção, apresentamos dois estudos de simulação para avaliar a robustez dos modelos propostos neste capítulo. Comparamos os modelos com erro de medida com distribuição Normal e *t-Student*. Para realizarmos tal comparação, consideramos primeiramente o Cenário 1, no qual o erro de medida possui distribuição *t-Student* com  $\nu$  graus de liberdade. Dessa maneira estudamos o que acontece quando o erro de medida tem distribuição com cauda pesada e ajustamos modelos que supõem erro Normal. As estimativas médias obtidas para o valor esperado *a posteriori* dos parâmetros dos modelos analisados em tal estudo estão expostas na Tabela 3.5. Um outro estudo que consideramos aqui consiste em compararmos o MBFC e o MBFC-NI, considerando dados com erro de medida Normal. Desta vez, analisamos o Cenário 3, com o objetivo de avaliar a qualidade do ajuste de modelos que supõem cauda pesada para o erro de medida, quando tal erro é proveniente de uma distribuição com cauda leve. Neste segundo estudo, supomos que os graus de liberdade em MBFC-NI1, 2 e 3 é  $\nu = 2.5$ , isto é, estabelecemos que o erro de medida possui distribuição com cauda consideravelmente mais pesada que a distribuição Normal. As estimativas obtidas para os parâmetros desse estudo de simulação são apresentadas na Tabela 3.7. As distribuições *a priori* consideradas para os parâmetros dos modelos analisados no presente estudo, são idênticas aquelas já consideradas nos Estudos 1 e 2 desta seção.



Tabela 3.5: Média, Viés e EQM das estimativas *a posteriori* obtidas ajustando-se os modelos MBFC1 e MBFC-NI1, considerando  $\sigma_\epsilon^2 = 1$ , MBFC2, MBFC-NI2, MBFC3 e MBFC-NI3, supondo  $\nu = 3$ , aos dados do Cenário 1.

Parâmetro	Modelo									
	MBFC1			MBFC2			MBFC3			
	Real	Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM
$\beta_1$	1.8	1.777	-0.023	0.352	1.797	-0.003	0.391	1.913	0.113	0.705
$\beta_2$	-0.15	-0.128	0.022	0.008	-0.136	0.014	0.009	-0.193	-0.043	0.040
$\alpha$	1	1.115	0.115	0.061	1.100	0.100	0.051	1.147	0.147	0.099
$\lambda$	-1.2	-1.365	-0.165	0.512	-1.358	-0.158	0.573	-1.209	-0.009	0.592
$\mu_x$	3.9	4.257	0.357	0.159	4.257	0.357	0.159	4.254	0.354	0.157
$\sigma_x^2$	10	9.952	-0.048	0.166	9.841	-0.159	0.097	9.119	-0.881	0.882
$\sigma_\epsilon^2$	1	-	-	-	1.509	0.509	0.822	4.431	3.431	11.797
$Var(\epsilon)$	1	-	-	-	1.509	-	-	4.431	-	-
Parâmetro	MBFC-NI1			MBFC-NI2			MBFC-NI3			
	Real	Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM
$\beta_1$	1.8	1.786	-0.014	0.358	1.804	0.004	0.376	1.898	0.098	0.543
$\beta_2$	-0.15	-0.134	0.016	0.007	-0.142	0.008	0.009	-0.187	-0.037	0.023
$\alpha$	1	1.097	0.097	0.051	1.106	0.106	0.056	1.142	0.142	0.078
$\lambda$	-1.2	-1.356	-0.156	0.506	-1.334	-0.134	0.512	-1.213	-0.013	0.563
$\mu$	3.9	4.246	0.346	0.145	4.245	0.345	0.145	4.251	0.351	0.147
$\sigma_x^2$	10	9.791	-0.209	0.108	9.730	-0.270	0.114	9.121	-0.879	0.811
$\sigma_\epsilon^2$	1	-	-	-	1.413	0.413	0.430	4.442	3.442	11.854
$Var(\epsilon)$	3	-	-	-	4.239	-	-	13.326	-	-

### Efeito nas Estimativas dos Parâmetros

Ao observarmos as estimativas expostas na Tabela 3.5 é interessante compararmos, principalmente, os modelos entre as modelagens denominadas 1, 2 e 3. Percebemos que os resultados vistos nessas perspectivas são mais comparáveis, dado que as suposições sobre o parâmetro de escala  $\sigma_\epsilon^2$  da distribuição do erro de medida são as mesmas em cada modelagem. Olhando dessa maneira, notamos que os modelos MBFC-NI parecem ter se ajustado melhor aos dados do que os modelos MBFC, visto que em geral o viés médio e EQM das estimativas fornecidas ao ajustarmos o MBFC-NI1 são menores do que aquelas ao ajustarmos o MBFC1. O mesmo é observado comparando os modelos MBFC-NI2 e MBFC-NI3 com os MBFC2 e MBFC3, respectivamente. O parâmetro que foge a essa conclusão é a variância  $\sigma_x^2$  das variáveis latentes, cujas estimativas obtidas nos modelos MBFC possuem menor viés e EQM do que aquelas ajustando-se os MBFC-NI. Além de tal parâmetro, vemos que a variância do erro de medida é bastante inflada pelos MBFC-NI, o que só acontece, dentre aqueles que supõem erro Normal, ao ajustarmos o MBFC3. Para entendermos um pouco melhor o que acontece com a variância do erro de medida, observemos: como se distribuem as estimativas obtidas para  $\sigma_\epsilon^2$  (na Figura 3.16) e uma comparação entre as estimativas médias obtidas para as variáveis latentes e seus reais valores (na Figura 3.17).

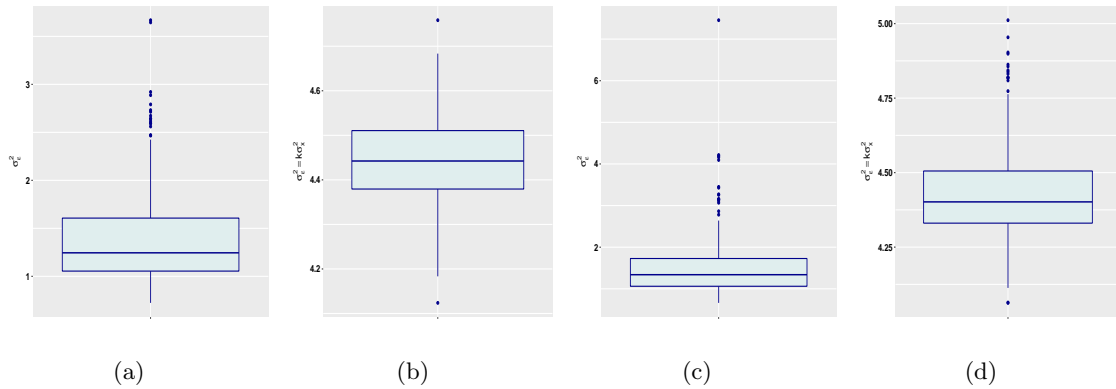


Figura 3.16: *Box-plots* das estimativas médias para  $\sigma_\epsilon^2$  ajustando-se os modelos MBFC-NI2(a), MBFC-NI3(b), MBFC2(c) e MBFC3(d), considerando os dados do Cenário 1.

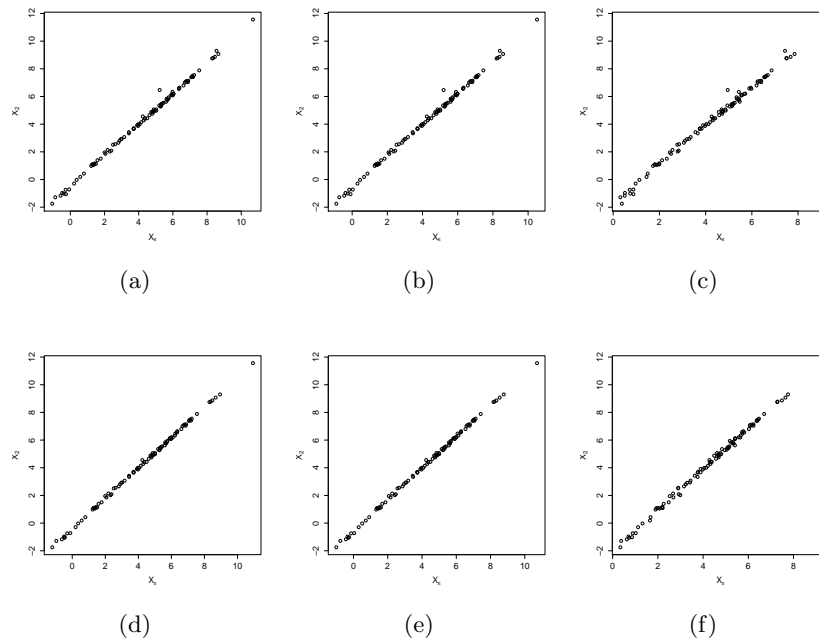


Figura 3.17: No eixo das abscissas temos as estimativas para  $\mathbf{X}_\epsilon$  ajustando-se os modelos MBFC-NI1, 2, 3, MBFC1, 2 e 3, considerando o Cenário 1, nos gráficos (a), (b), (c), (d), (e) e (f), respectivamente.

Ao analisarmos as estimativas para a variância do erro de medida, expostas na Tabela 3.5, vemos que ajustando-se o MBFC-NI2 obtivemos uma estimativa mais próxima ao valor real 3 de tal variância. O mesmo acontece em relação as estimativas para  $\sigma_\epsilon^2$ . Além disso, ao observarmos a Figura 3.16, percebemos que a variabilidade das estimativas para  $\sigma_\epsilon^2$  ajustando-se o modelo MBFC-NI2 é similar a variabilidade das estimativas obtidas para tal parâmetro ao ajustarmos

o MBFC2. Embora isso aconteça, ao observarmos os diagramas de dispersão na Figura 3.17, nos quais relacionamos as estimativas obtidas para as variáveis latentes com seus valores reais, notamos que os pontos da Figura 3.17e estão mais dispersos do que aqueles da Figura 3.17b. Isto indica que ao ajustarmos o MBFC2 obtemos em média, estimativas mais próximas dos valores reais para as variáveis latentes e dado o que observamos até aqui, esse resultado é diferente do esperado. O mesmo é constatado ao compararmos os MBFC1 e MBFC-NI1. No MBFC-NI1, fixamos  $\sigma_\epsilon^2 = 1$  e  $\nu = 3$ , em decorrência disso a variância do erro é fixada no valor real 3. O mesmo não ocorre no MBFC1, pois fixando  $\sigma_\epsilon^2 = 1$ , conseqüentemente a variância do erro é fixada em 1. Novamente observamos um pouco mais de dispersão na Figura 3.17a do que na Figura 3.17d, indicando que o MBFC1 fornece estimativas mais precisas para as variáveis latentes do que o MBFC-NI1. Ajustando-se os modelos MBFC3 e MBFC-NI3, obtivemos estimativas aproximadamente iguais para o parâmetro de escala do erro de medida, porém com viés e EQM bastante elevados. A variância das variáveis latentes foi inflada, sendo pior aproximada através do ajuste do MBFC-NI3 que estimou esse parâmetro em média como 13.326. Em relação às estimativas para as variáveis latentes, observamos um pouco mais de dispersão nas Figuras 3.17f e 3.17c, sendo a aproximação obtida ajustando-se o MBFC-NI3 a que parece menos precisa em média.

De acordo com os resultados exibidos na Tabela 3.5 e com o observado nas Figuras 3.16 e 3.17, supor uma distribuição Normal para o erro de medida quando este erro possui distribuição *t*-Student não causa grande impacto na estimação dos parâmetros. Obtemos estimativas em média mais precisas para as variáveis latentes quando ajustamos os modelos MEFC, embora os coeficientes de regressão e parâmetros da distribuição dos tempos de falha sejam melhores estimados quando ajustamos os modelos MBFC-NI com  $\nu = 3$ .

### **Efeito nas Estimativas da Fração de Cura**

Continuamos nossa análise, avaliando os efeitos de se ajustar os modelos MBFC a dados que foram gerados de um MBFC-NI. Com esta finalidade, nas Figuras 3.18 a 3.20, podemos observar uma comparação entre as estimativas médias obtidas para o valor esperado *a posteriori* da fração de cura dos 100 indivíduos analisados, nas 250 réplicas de Monte Carlo, ajustando-se os seis modelos propostos. Além disso, na Tabela 3.6 temos o percentual de cobertura dos intervalos de credibilidade de 95% para a fração de cura, ajustando-se os MBFC-NI supondo erro de medida Normal e *t*-Student.

Nas Figuras 3.18 e 3.19, vemos que as estimativas e viés médios obtidos pelos seis modelos são aproximadamente iguais, de modo que é difícil distinguir os pontos nos gráficos. Além disso, observamos que todas as estimativas para a fração de cura, sob todos os modelos ajustados, estão muito próximas dos valores reais, com viés médio entre -0.1 e 0.1. Podemos perceber isso

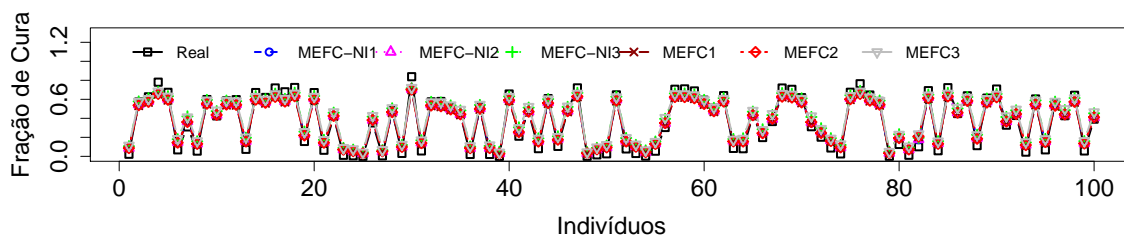


Figura 3.18: Estimativas médias do valor esperado *a posteriori* para a fração de cura de cada um dos 100 indivíduos, ajustando-se os modelos MBFC1, 2, 3, MBFC-NI1, 2 e 3.

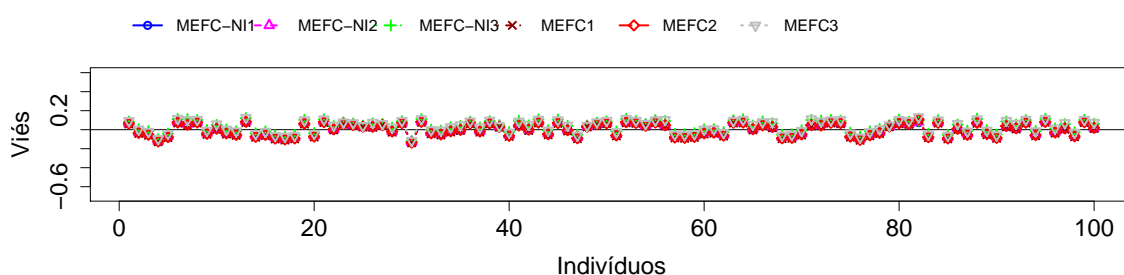


Figura 3.19: Viés médio das estimativas do valor esperado *a posteriori* para a fração de cura de cada um dos 100 indivíduos, ajustando-se os modelos MBFC1, 2, 3, MBFC-NI1, 2 e 3.

mais claramente na Figura 3.20, na qual apresentamos as distribuições empíricas das médias dos valores esperados *a posteriori* para a fração de cura sobrepostas ao histograma das frações de cura reais e é possível fazer distinção entre as curvas obtidas ajustando-se cada um dos modelos.

Observando a Figura 3.20, notamos que os modelos ajustados superestimam a fração de cura daqueles indivíduos que possuem probabilidade de cura de até 10%. Para tais indivíduos aqueles que apresentam um melhor ajuste, em média, são os MBFC1 e 2. As distribuições empíricas obtidas ajustando-se tais modelos estão praticamente sobrepostas, sendo difícil distingui-las para os valores de fração de cura entre 10% e 40%. Para os indivíduos com fração de cura acima de 70%, um melhor ajuste é observado com o MBFC-NI3. E de um modo geral, vemos que os modelos MBFC-NI1 e MBFC-NI2, que apresentam distribuições empíricas aproximadamente iguais para todos os valores de fração de cura, se ajustam melhor a distribuição da fração de cura real.

Na Tabela 3.6, observamos que em relação a cobertura dos intervalos de credibilidade obtidos ao ajustarmos os modelos MBFC-NI e MBFC, os MBFC-NI nos quais supomos graus de liberdade  $\nu = 3$  apresentam um percentual médio de cobertura maior. É interessante observar

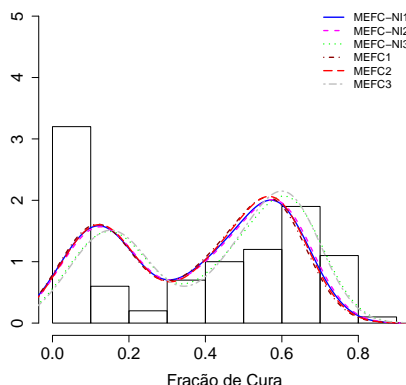


Figura 3.20: Histograma das frações de cura reais superposto pelas distribuições empíricas das médias dos valores esperados *a posteriori* da fração de cura, obtidas considerando os Cenário 1 e ajustando-se os modelos MBFC1, MBFC2, MEFC3, MBFC-NI1, MBFC-NI2 e MBFC-NI3.

que o modelo que resultou em um maior percentual médio de cobertura foi o MBFC-NI3, seguido por MBFC3, embora ajustando tais modelos tenhamos obtido estimativas mais enviesadas para todos os parâmetros e variáveis latentes do que quando ajustamos os MBFC-NI1, MBFC-NI2, MBFC1 e MBFC12.

Tabela 3.6: Percentual de cobertura dos intervalos de 95% de credibilidade para a fração de cura, considerando o Cenário 1 e ajustando-se os modelos MBFC1 a 3 e MBFC-NI1 a 3.

Modelos					
MBFC1	MBFC2	MBFC3	MBFC-NI1	MBFC-NI2	MBFC-NI3
0.885	0.901	0.927	0.918	0.916	0.943

Diante da análise realizada no presente estudo, vemos que é difícil eleger um modelo dentre os apresentados como sendo o que melhor se ajusta aos dados do Cenário 1. Entretanto, vimos que estimativas mais precisas para os parâmetros foram obtidas quando consideramos o MBFC-NI2, o qual é o modelo em que supomos mais incerteza sobre os parâmetros ao utilizarmos distribuições *a priori* pouco informativas. Além disso, consideramos  $\sigma_\epsilon^2$  completamente desconhecido, diferentemente do que supomos nos MBFC1 e MBFC-NI1 e de uma forma menos convicta nos MBFC-NI3 e MBFC3. Por serem sensíveis à escolha da distribuição *a priori* de  $k$ , os modelos MBFC-NI3 e MBFC3 também podem ser um pouco inapropriados em situações práticas, visto a necessidade de se ter informações sobre a relação entre as variâncias do erro de medida e variáveis latentes.

Na Tabela 3.7 a seguir, apresentamos as estimativas médias obtidas para os parâmetros

dos modelos MBFC e MBFC-NI, quando os ajustamos aos dados do Cenário 3, os quais foram gerados considerando que a covariável  $X_2$  foi observada com erro de medida Normal com média 0 e variância 1. Neste estudo utilizamos as mesmas distribuições *a priori* já mencionadas no primeiro estudo Monte Carlo desta subseção. Nos modelos em que supomos erro de medida *t-Student*, fixamos os graus de liberdade  $\nu = 2.5$ , isto é, consideramos uma cauda mais pesada para modelarmos a distribuição do erro de medida.

Tabela 3.7: Média, Viés e EQM das estimativas *a posteriori* obtidas ajustando-se os modelos MBFC1 e MBFC-NI1, considerando  $\sigma_\epsilon^2 = 1$ , MBFC2, MBFC-NI2, MBFC3 e MBFC-NI3, analisando os dados do Cenário 3.

Parâmetro	Modelo									
	MBFC1			MBFC2			MBFC3			
	Real	Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM
$\beta_1$	1.8	1.820	0.020	0.352	1.834	0.034	0.380	1.988	0.188	0.571
$\beta_2$	-0.15	-0.145	0.005	0.007	-0.152	-0.002	0.013	-0.220	-0.080	0.030
$\alpha$	1	1.078	0.078	0.052	1.085	0.085	0.055	1.139	0.139	0.080
$\lambda$	-1.2	-1.253	-0.053	0.410	-1.237	-0.037	0.381	-1.084	0.116	0.504
$\mu_x$	3.9	3.920	0.020	0.010	3.921	0.021	0.010	3.919	0.019	0.011
$\sigma_x^2$	10	10.028	0.028	0.032	9.990	-0.010	0.024	9.171	-0.829	0.708
$\sigma_\epsilon^2$	1	-	-	-	1.241	0.241	0.211	4.454	3.454	11.936
$Var(\epsilon)$	1	1	-	-	1.241	-	-	4.454	-	-
Parâmetro	MBFC-NI1			MBFC-NI2			MBFC-NI3			
	Real	Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM
	$\beta_1$	1.8	1.862	0.062	0.356	1.924	0.124	0.726	2.571	0.771
$\beta_2$	-0.15	-0.163	-0.013	0.011	-0.179	-0.029	0.031	-0.402	-0.252	0.663
$\alpha$	1	1.091	0.091	0.058	1.116	0.116	0.111	1.320	0.320	0.647
$\lambda$	-1.2	-1.208	-0.008	0.420	-1.206	-0.006	0.430	-1.038	0.162	0.502
$\mu$	3.9	3.919	0.019	0.010	3.918	0.018	0.011	3.893	-0.007	0.012
$\sigma_x^2$	10	9.808	-0.192	0.062	9.841	-0.159	0.045	8.975	-1.025	1.069
$\sigma_\epsilon^2$	1	-	-	-	1.039	0.039	0.043	4.362	3.362	11.311
$Var(\epsilon)$	1	5	-	-	5.195	-	-	21.810	-	-

### Efeito nas Estimativas dos Parâmetros

Ao observarmos a Tabela 3.7, não nos parece claro qual dos modelos melhor se ajusta aos dados, visto que  $\beta_1$  e  $\alpha$  são melhor aproximados quando ajustamos o modelo MBFC1 (modelo gerador dos dados),  $\beta_2$ ,  $\sigma_x^2$  e  $Var(\epsilon)$  são melhor aproximados quando ajustamos o modelo MBFC2 e os parâmetros  $\mu_x$ ,  $\lambda$  e  $\sigma_\epsilon^2$  são melhor aproximados ao ajustarmos o MBFC-NI2. Porém, notamos que a variância do erro de medida é inflada ao ajustarmos os modelos MBFC-NI e em contrapartida a variância das variáveis latentes são subestimadas. Os ajustes dos modelos MBFC3 e MBFC-NI3 fornecem estimativas muito enviesadas para todos os parâmetros com exceção da média  $\mu_x$  das variáveis latentes. Creditamos esse fato à sensibilidade que tais modelos possuem em relação a escolha da distribuição *a priori* de  $k$  (estudada no Estudo 5 da Seção 2.3). Vejamos nas Figuras 3.21 e 3.22 como se comportam as estimativas obtidas para o parâmetro

de escala  $\sigma_\epsilon^2$  e as estimativas médias das variáveis latentes, respectivamente.

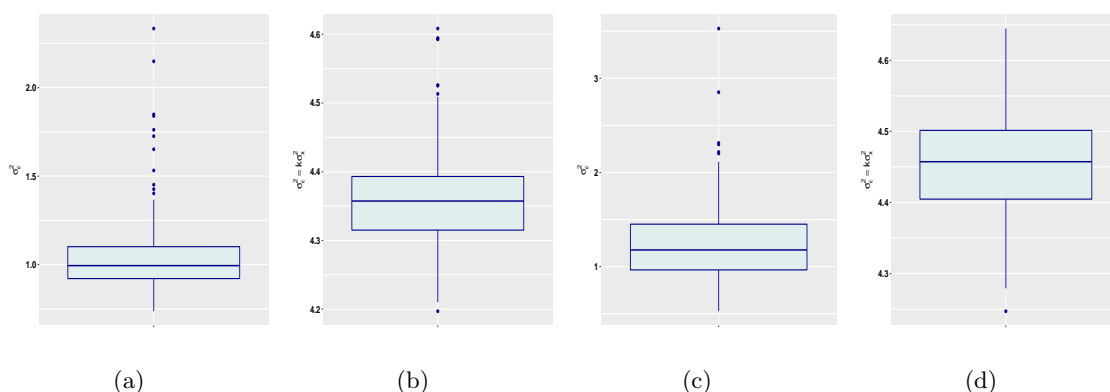


Figura 3.21: *Box-plots* das estimativas médias para  $\sigma_\epsilon^2$  ajustando-se os modelos MBFC-NI2(a), MBFC-NI3(b), MBFC2(c) e MBFC3(d), considerando os dados do Cenário 3.

Comparando os *box-plots* nas Figuras 3.21a e 3.21c, obtidos via MBFC-NI2 e MBFC2, respectivamente, percebemos que a distribuição das estimativas obtidas para  $\sigma_\epsilon^2$  ajustando-se o MBFC-NI2 é quase simétrica em torno da mediana, com pouca variabilidade entre o primeiro e terceiro quartil, porém, apresenta muitos pontos discrepantes. Por outro lado, a distribuição das estimativas obtidas sob o MBFC2, apresentam uma variabilidade maior entre o primeiro e o terceiro quartil, com a presença de menos *outliers* e com mediana um pouco maior que no caso anterior. Notamos que 75% das estimativas obtidas ao ajustarmos o MBFC2 são maiores do que o valor real de  $\sigma_\epsilon^2$ , enquanto que, sob o ajuste do MBFC-NI2 essa quantidade é um pouco menor do que 50%, indicando uma certa simetria em torno do valor real e pouca variabilidade. O comportamento das estimativas ao ajustarmos os MBFC-NI3 e MBFC3, as quais possuem mediana em torno de 4.3 e 4.4, respectivamente, é bem similar entre tais modelos. Os ajustes dos MBFC-NI3 e MBFC3 fornecem estimativas muito enviesadas para  $\sigma_\epsilon^2$  e essa pode ser a razão das estimativas obtidas para os demais parâmetros serem mais enviesadas do que aquelas obtidas ao ajustarmos os demais modelos. Vejamos como isso influencia na estimação das variáveis latentes.

Na Figura 3.22, vemos que as estimativas obtidas para  $X_2$  ao ajustarmos os modelos MBFC-NI1 (Fig.3.22a), MBFC-NI2 (Fig.3.22b), MBFC1 (Fig.3.22d) e MBFC2 (Fig.3.22e) parecem bem próximas dos valores reais desta covariável, porém percebemos que nos MBFC1 e MBFC2 o ajuste parece ser um pouco melhor, dada a não percepção de dispersão nos diagramas das Figuras 3.22a e 3.22b. Ajustando-se os modelos MBFC-NI3 (Fig.3.22c) e MBFC3 (Fig.3.22d) percebemos muito mais dispersão ao associarmos  $X_\epsilon$  a  $X_2$  do que o observado nos demais modelos, o que indica que as estimativas médias obtidas para  $X_2$ , em tais casos, são mais enviesadas.

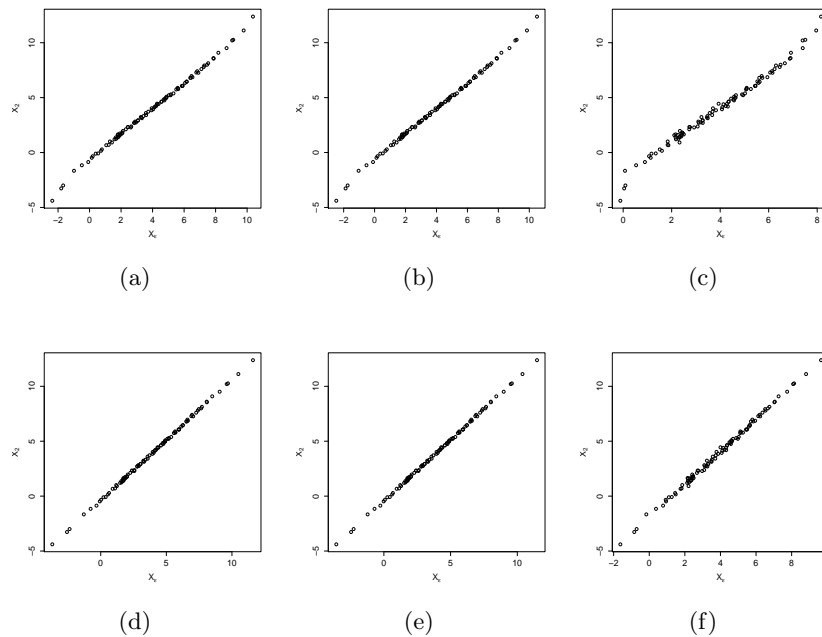


Figura 3.22: No eixo das abscissas temos as estimativas para  $X_\epsilon$  ajustando-se os modelos MBFC-NI1, 2, 3, MBFC1, 2 e 3, considerando o Cenário 3, nos gráficos (a), (b), (c), (d), (e) e (f), respectivamente.

### Efeito nas Estimativas da Fração de Cura

Para analisar o efeito de ajustarmos os modelos MBFC-NI assumindo uma distribuição  $t$ -Student com  $\nu = 2.5$  graus de liberdade, à dados com erro Normal, na estimação da fração de cura, exibimos nas Figuras 3.23 à 3.25, as estimativas médias para o valor esperado *a posteriori* da fração de cura ajustando-se os modelos MBFC-NIs e MBFCs. Além disso, na Tabela 3.8 são dados os percentuais médios de cobertura dos intervalos de 95% de credibilidade para a fração de cura dos 100 indivíduos nas 250 réplicas de Monte Carlo.

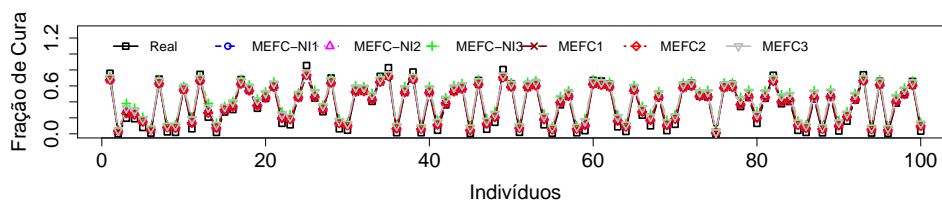


Figura 3.23: Estimativas médias do valor esperado *a posteriori* para a fração de cura de cada um dos 100 indivíduos, ajustando-se os modelos MBFC1, 2, 3, MBFC-NI1, 2 e 3, considerando o Cenário 3.

Ao compararmos as estimativas médias obtidas para o valor esperado *a posteriori* da



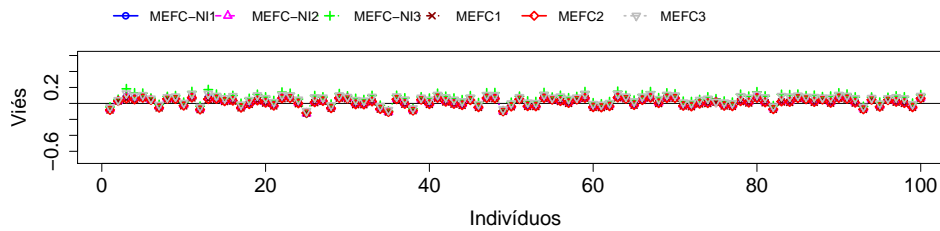


Figura 3.24: Viés médio das estimativas do valor esperado *a posteriori* para a fração de cura de cada um dos 100 indivíduos, ajustando-se os modelos MBFC1, 2, 3, MBFC-NI1, 2 e 3.

fração de cura nos seis modelos analisados, mais uma vez percebemos que tais estimativas não são significativamente diferentes de um modelo para o outro. Tanto as médias quanto os vieses obtidos são muito similares de modo que nas Figuras 3.23 e 3.24 é difícil distinguir os pontos relativos a cada modelo. Percebemos alguma diferença apenas para as estimativas de poucos indivíduos sob o ajuste do MBFC-NI3.

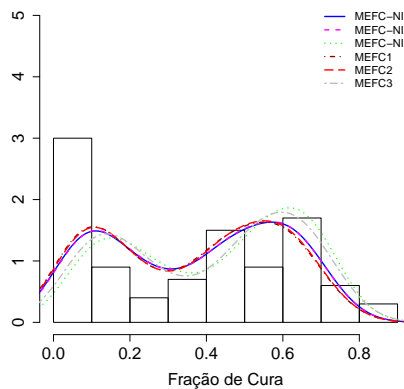


Figura 3.25: Histograma das frações de cura reais superposto pelas distribuições empíricas das médias dos valores esperados *a posteriori* da fração de cura, considerando o Cenário 3 e ajustando-se os modelos MBFC1, MBFC2, MBFC3, MBFC-NI1, MBFC-NI2 e MBFC-NI3.

Observando as curvas das distribuições empíricas na Figura 3.25, comparando entre modelos e em relação ao histograma das frações de curas reais, notamos que todos eles subestimam o número de indivíduos que possuem probabilidade de cura de até 10%. Percebe-se uma diferença considerável entre as curvas obtidas através de cada modelo ajustado e é notável que os modelos MBFC-NI3 e MBFC3 são aqueles que mais se afastam das quantidades reais. Para os indivíduos que possuem fração de cura acima de 10%, as curvas empíricas obtidas ajustando-se

os seis modelos estimam razoavelmente bem a forma como a fração de cura real está distribuída, com destaque para os modelos MBFCNI-3 e MBFC3, cujas curvas são mais precisas para valores de fração de cura acima de 60%. É interessante relacionarmos essa análise aos resultados apresentados na Tabela 3.8, pois vemos que os modelos MBFC-NI3 e MBFC3 fornecem os maiores percentuais médios de cobertura para as frações de cura reais.

Tabela 3.8: Percentual de cobertura dos intervalos de 95% de credibilidade para a fração de cura, considerando o Cenário 3 e ajustando-se os modelos MBFC1 a 3 e MBFC-NI1 a 3.

Modelos					
MBFC1	MBFC2	MBFC3	MBFC-NI1	MBFC-NI2	MBFC-NI3
0.925	0.932	0.955	0.944	0.944	0.947

De acordo com os resultados observados nesta subseção, podemos concluir que ajustar os modelos MBFC-NI, que supõem erro de medida Normal, a dados com erro de medida *t-Student* ou o oposto, não resulta em um aumento no viés das estimativas dos parâmetros ou fração de cura. Uma ressalva a essa conclusão deve ser feita acerca dos modelos MBFC-NI3 e MBFC3, que em geral forneceram estimativas mais enviesadas em ambas as situações. Mais uma vez queremos destacar que, baseando-se nas conclusões obtidas através do Estudo 5, apresentado na Seção 2.3, sabemos que o MBFC3 é sensível à escolha da distribuição *a priori* de  $k$ . Vimos que não é possível obtermos estimativas não enviesadas (ou com viés significativamente pequeno) quando utilizamos distribuições pouco informativas para este parâmetro. Ao ajustarmos o MBFC3 obtivemos melhores estimativas quando supomos uma distribuição *a priori* para  $k$  centrada em seu valor verdadeiro. No presente estudo de simulação  $k = 0.1$  e nesse caso, de acordo com o estudo Monte Carlo da Subseção 2.3.5, a melhor escolha *a priori*, dentre as analisadas é  $k \sim \text{Gama}(0.1, 1)$ . Porém, ao realizarmos a mesma escolha no MBFC-NI3, enfrentamos problemas numéricos no MCMC e, por essa razão, ajustamos tal modelo supondo  $k \sim U(0, 0.5)$  *a priori*, a qual é pouco informativa e tem média 0.25.

### Comentários Gerais

Nos estudos de simulação apresentados nesta seção, investigamos o desempenho dos modelos MBFC-NI supondo distribuição *t-Student* para o erro de medida. Em todos os casos, consideramos os graus de liberdade  $\nu$  como conhecido. Avaliamos o que acontece ao fixarmos  $\nu$  em determinados valores e percebemos que não há prejuízos substanciais à estimação dos parâmetros, visto que não observamos aumento considerável no viés das estimativas. Os resultados indicam ainda que ao supormos  $\nu$  fixo em valores menores, assumindo uma distribuição com cauda mais pesada para o erro de medida, obtivemos em geral estimativas mais precisas.

Uma outra preocupação que tivemos foi acerca do efeito de se ajustar os modelos MBFC-

NI a dados que não possuem erro de medida. Com o objetivo de avaliar tal efeito, no Estudo 2, analisamos dados desta natureza e comparamos as estimativas ajustando-se os modelos MBFC-NI às aquelas obtidas ao ajustarmos o modelo de tempo de promoção (MTP) proposto por Chen *et al.* (1999). Ao avaliarmos os resultados em tal estudo, constatamos que os modelos MBFC-NI também podem ser apropriados para analisar nesse tipo de situação, pois, embora tenhamos obtido estimativas um pouco mais enviesadas para alguns dos parâmetros, as estimativas para a fração de cura sob cada modelo foram aproximadamente iguais, sendo melhores (em média) aquelas estimativas intervalares obtidas via MBFC-NIs.

Notamos ainda, que os modelos MBFC-NI podem se ajustar bem a dados com erro de medida Normal ou *t-Student*, mas destacamos que o modelo MBFC-NI2, por comportar mais incerteza em relação aos dados e, além disso, fornecer em geral melhores resultados que os demais modelos analisados, tende a ser uma melhor alternativa em situações práticas. Destacamos também a necessidade de pensar mais a respeito do modelo MBFC-NI3, que pode oferecer melhor desempenho computacional, porém, precisa ser mais estudado no futuro em relação a eliciação de distribuições *a priori* para  $k$ .

### 3.4 Aplicação

Consideremos novamente os dados de 234 pacientes observados no estudo E1684, correspondente ao ensaio clínico pós cirúrgico de pacientes com melanoma cutâneo, realizado no Eastern Cooperative Oncology Group (ECOG). Este banco de dados foi analisado, entre outros autores da literatura de análise de sobrevivência, por Chen *et al.* (1999) e na Seção 2.4 deste trabalho, na qual ajustamos o modelo MTP proposto por Chen *et al.* (1999) e os modelos MBFC introduzidos no Capítulo 2. Relembramos que tal ensaio clínico consistiu na administração de um tratamento por Interferon alpha-2b em alguns pacientes, após terem sido submetidos a cirurgia para a extração do melanoma. Para avaliar o efeito do tratamento alguns dos pacientes permaneceram apenas em observação após a cirurgia. Para cada indivíduo participante do estudo foi registrado o tempo de falha em anos ( $Y$ ), o indicador de falha ( $\delta = 1$  caso o indivíduo tenha falhado e  $\delta = 0$  caso contrário), a idade em anos ( $X_1$ ), o sexo ( $X_2 = 0$  se masculino e 1 se feminino), tratamento ( $X_3 = 0$ , se o indivíduo ficou somente em observação e  $X_3 = 1$  se recebeu doses de Interferon alpha-2b), a profundidade de Breslow ( $X_4$ , medida em milímetros), e a performance física do paciente ( $X_5 = 1$  se o indivíduo praticar alguma atividade física e  $X_5 = 0$  caso seja sedentário). Embora na Seção 2.4 os critérios de seleção de modelo tenham apontado para o MTP, o qual não supõe erro nas covariáveis, os resumos da variância do erro obtidos em tal seção deram indícios que a covariável  $X_4$  pode ter sido medida com erro. Além do mais, sabemos que a avaliação resultante dos critérios de seleção de modelos não são con-

sistentes para comparar modelos Bayesianos com erro de medida aqueles que não supõem erro, visto que possuem níveis de complexidade diferentes. Os modelos com erro propostos são menos parcimoniosos que o MTP.

Por motivações práticas, suponhamos novamente que a covariável  $X_4$  foi observada com erro de medida, isto é, observou-se  $X_4 = X_\epsilon + \epsilon$ . Reanalisaremos os dados do E1684 na presente seção, ajustando os modelos MBFC-NI propostos neste capítulo, supondo que para cada indivíduo  $i = 1, \dots, 234$ , temos  $\epsilon_i \sim t(0, \sigma_\epsilon^2, \nu)$ . Em nossa análise assumimos a distribuição Weibull para os tempos de falha e exibimos alguns resumos das distribuições *a posteriori* dos parâmetros dos modelos ajustados na Tabela 3.9.

No algoritmo MCMC utilizamos um *lag* de tamanho 50 e *burn-in* de 500000, resultando em um total de 600000 iterações e uma amostra de tamanho 2000 da distribuição *a posteriori* dos parâmetros. Como distribuições *a priori*, eliciamos:  $\beta \propto 1$ ,  $\alpha \sim \text{Gama}(0.5, 0.001)$ ,  $\lambda \sim \text{Normal}(-0.63, 10)$ ,  $X_\epsilon \sim \text{Normal}(\mu_x, \sigma_x^2)$ , em que  $\mu_x \sim \text{Normal}(0, 10)$  e  $\sigma_x^2 \sim \text{GI}(100, 100)$ ,  $U \sim G\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ . No modelo MBFC-NI1 fixamos o parâmetro de escala do erro de medida  $\sigma_\epsilon^2 = 1$ , no MBFC-NI2 supomos *a priori* que  $\sigma_\epsilon^2 \sim \text{GI}(1, 1)$  e no MBFC-NI3 temos  $k \sim U(0, 0.5)$ . Nas três versões do MBFC-NI fixamos os graus de liberdade da distribuição do erro de medida em  $\nu = 2.5$ . Os hiperparâmetros da distribuição *a priori* do parâmetro de escala  $\lambda$  dos tempos de falha foram escolhidos com base nos resumos amostrais de tal variável. Para todos os outros parâmetros as distribuições *a priori* eliciadas são pouco informativas.

Ao analisarmos os dados do estudo E1684 na Seção 2.4 todos os modelos ajustados indicaram que apenas o tratamento, representado pela covariável  $X_3$ , é significativo para o tempo de vida dos pacientes. Observando a Tabela 3.9, notamos que todos os modelos aqui ajustados também apontam o tratamento como sendo significativo para o tempo de vida, mas, além do tratamento, outra covariável indicada pelos modelos MBFC-NI1 e MBFC-NI2, como significativa é a idade dos pacientes. Somente sob o ajuste do modelo MBFC-NI3 tal covariável não é apontada como significativa para o tempo de vida dos pacientes e como os limites inferiores dos intervalos de 95% de credibilidade obtidos ao ajustarmos os modelos MBFC-NI1 e 2 são muito próximos a zero, investigamos um pouco mais como tal covariável pode influenciar a fração de cura dos pacientes observados. Na análise que realizamos na Seção 2.4 os modelos ajustados também forneceram limites para os intervalos de credibilidade do preditor linear  $\beta_1$  bem próximos a zero.

Em relação ao erro de medida, o modelo MBFC-NI1 pressupõe  $\text{Var}(\epsilon) = 5$ , que é uma suposição mais forte de que a covariável  $X_4$  foi mensurada com erro do que aquela assumida no MBFC1 na Seção 2.4. Apesar dessa suposição, notamos que os resumos obtidos para todos os parâmetros não diferem substancialmente daqueles ao ajustarmos os modelos MBFC-NI2, 3, os quais não supõem variância conhecida para o erro de medida. Além disso, também não são tão diferentes daqueles obtidos com os modelos discutidos no Capítulo 2 na Seção 2.4, incluindo o

Tabela 3.9: Resumos das distribuições *a posteriori* dos parâmetros dos modelos MBFC-NI1, MBFC-NI2 e MBFC-NI3 para os dados do estudo E1684, assumindo-se distribuição Weibull para os tempos de falha.

$\eta$	Resumo	Modelo		
		MBFC-NI1	MBFC-NI2	MBFC-NI3
$\beta_1$	Média	0.079	0.078	0.083
	DP	0.039	0.037	0.048
	HPD 95%	(0.001, 0.151)	(0.002, 0.151)	(-0.015, 0.173)
$\beta_2$	Média	-0.013	0.010	-0.023
	DP	0.170	0.258	0.271
	HPD 95%	(-0.362, 0.303)	(-0.475, 0.547)	(-0.561, 0.504)
$\beta_3$	Média	-0.322	-0.331	-0.335
	DP	0.165	0.161	0.162
	HPD 95%	(-0.658, -0.010)	(-0.663, -0.038)	(-0.646, -0.004)
$\beta_4$	Média	-0.006	-0.015	-0.009
	DP	0.183	0.243	0.094
	HPD 95%	(-0.348, 0.357)	(-0.545, 0.422)	(-0.187, 0.179)
$\beta_5$	Média	-0.045	-0.045	-0.041
	DP	0.273	0.271	0.269
	HPD 95%	(-0.579, 0.503)	(-0.590, 0.447)	(-0.588, 0.491)
$\alpha$	Média	1.005	1.007	1.006
	DP	0.067	0.065	0.066
	HPD 95%	(0.874, 1.139)	(0.883, 1.128)	(0.880, 1.137)
$\lambda$	Média	-0.506	-0.503	-0.507
	DP	0.116	0.112	0.117
	HPD 95%	(-0.746, -0.290)	(-0.730, -0.293)	(-0.749, -0.293)
$\mu_x$	Média	0.396	0.394	0.400
	DP	0.078	0.056	0.062
	HPD 95%	(0.248, 0.550)	(0.282, 0.505)	(0.285, 0.521)
$\sigma_x^2$	Média	0.691	0.608	0.635
	DP	0.052	0.043	0.048
	HPD 95%	(0.584, 0.788)	(0.525, 0.697)	(0.550, 0.731)
$\sigma_\epsilon^2$	Média	-	0.083	0.311
	DP	-	0.019	0.023
	HPD 95%	-	(0.051, 0.121)	(0.264, 0.353)
$Var(\epsilon)$	Média	-	0.415	1.555

MTP que não supõe erro de medida nas covariáveis. Os resumos para  $\sigma_\epsilon^2$  ajustando os MBFC-NI2 e 3 (exibidos na Tabela 3.9) implicam em erro de medida na covariável  $X_4$ , visto que as médias *a posteriori* para a variância do erro de medida são, respectivamente, 0.415 e 1.555. É relevante lembrarmos que os estudos de simulação (apresentados na Seção 3.3) mostram que os modelos MBFC-NI com erro de medida *t-Student* tendem a inflacionar um pouco a variância do erro, entretanto a média e variância das variáveis latentes costumam ser bem aproximadas em média. Se comparamos os valores esperados *a posteriori* para  $\mu_x$  e  $\sigma_x^2$ , sob todos os ajustes com a média e variância amostral da variável  $X_4$ , vemos que os resumos para  $\mu_x$ , em torno de 0.39 sob todos os modelos, diferem consideravelmente da média amostral 0.913 desta covariável. Dessa

maneira, considerando que a amostra coletada é significativamente grande, a estimativa 0.913 deve ser bem próxima da verdadeira média de  $X_4$ , dando indícios de que tal covariável pode ter sido de fato observada com erro. Para visualizarmos as estimativas das variáveis latentes via MBFC-NIs, nas Figuras 3.26 a 3.28, podemos observar os intervalos de 95% de credibilidade para a variável  $X_\epsilon$  de cada indivíduo.

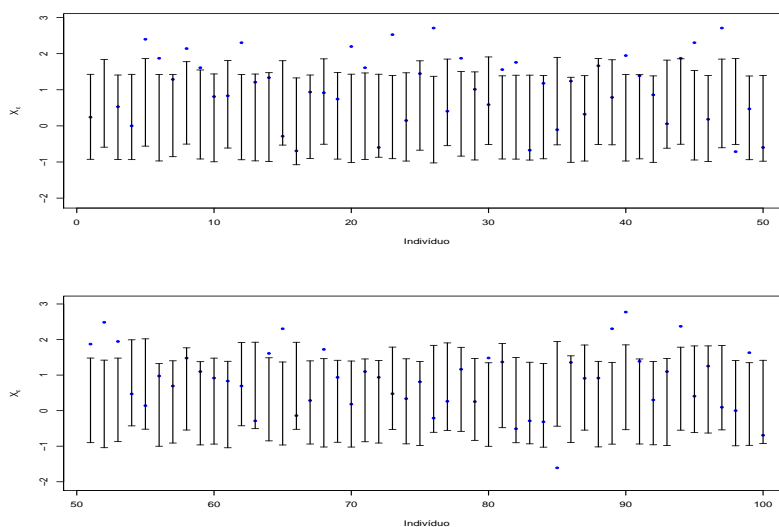


Figura 3.26: HPD's de 95% para a variável latente  $X_\epsilon$  dos primeiros 100 indivíduos, obtidos ajustando-se o MBFC-NI1. Os pontos destacados são os valores observados da covariável  $X_4$  para cada indivíduo.

Ao observarmos os HPD's de 95% exibidos nas Figuras 3.26 a 3.28 notamos que, para grande parte do indivíduos, cujos os intervalos de credibilidade estão expostos em tais figuras, alguns dos valores observados da covariável  $X_4$  não pertencem aos intervalos. A proporção de intervalos de credibilidade que contém o valor observado de  $X_4$  sob o ajuste dos modelos MBFC-NI1, MBFC-NI2 e MBFC-NI3 é, respectivamente, 72.6%, 37.6% e 54.7%. Nesse caso, percebemos que o MBFC-NI1 (modelo que supõe variância relativamente alta para o erro de medida) determina um percentual maior de intervalos de credibilidade para as variáveis latentes do que o MBFC-NI2 e 3, nos quais temos mais incerteza acerca de tal variância. Em todo caso, temos indícios de que há erro de medida na covariável  $X_4$ .

**Analisando o Efeito das Covariáveis na Fração de Cura**

Os resumos da distribuição *a posteriori* dos preditores lineares obtidos via MBFC-NIs, exibidos na Tabela 3.9, sugerem que a covariável  $X_4$  é significativa para o tempo de falha dos pacientes. Isto é, há indícios de que o tratamento por Interferon alpha-2b administrado em alguns pacientes após a cirurgia para a retirada do melanoma cutâneo é efetivo no tempo de

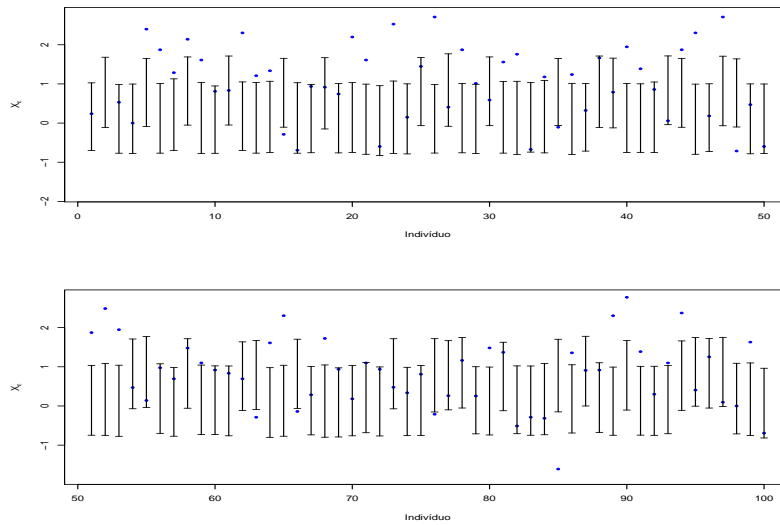


Figura 3.27: HPD's de 95% para a variável latente  $X_\epsilon$  dos primeiros 100 indivíduos, obtidos ajustando-se o MBFC-NI2. Os pontos destacados são os valores observados da covariável  $X_4$  para cada indivíduo.

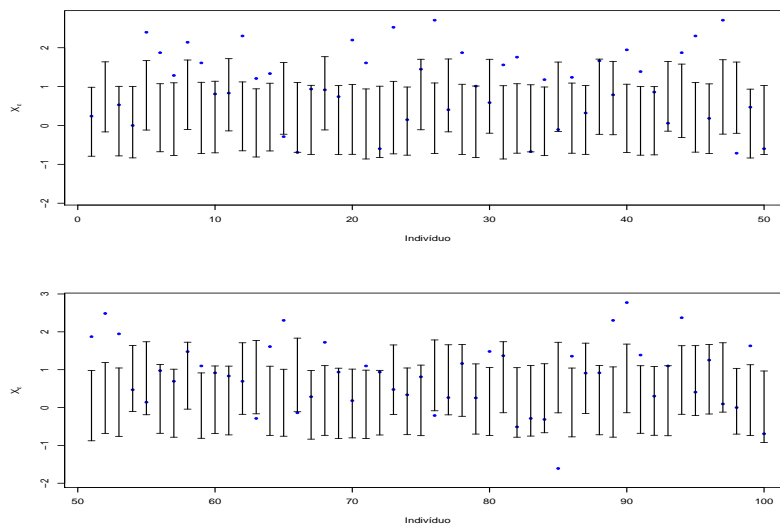


Figura 3.28: HPD's de 95% para a variável latente  $X_\epsilon$  dos primeiros 100 indivíduos, obtidos ajustando-se o MBFC-NI3. Os pontos destacados são os valores observados da covariável  $X_4$  para cada indivíduo.

vida. Além do tratamento, sob o ajuste dos modelos MBFC-NI1 e MBFC-NI2 percebemos que a covariável idade também pode ser significativa para a variável resposta. Para investigarmos melhor a relevância de tais covariáveis no prognóstico da doença, exibimos nas Figuras 3.29 a

3.31, alguns *box-plots* das estimativas para o valor esperado *a posteriori* da fração de cura de cada indivíduo, obtidas ao ajustarmos os três modelos MBFC-NI.

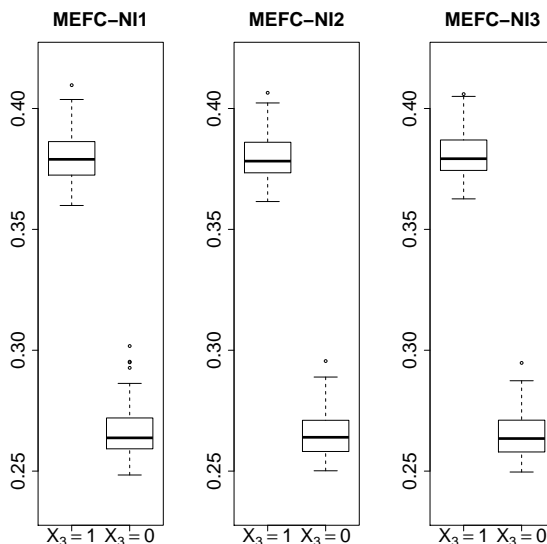


Figura 3.29: *Box-plots* das estimativas do valor esperado da distribuição *a posteriori* da fração de cura dos indivíduos observados que receberam doses de Interferon ( $X_3 = 1$ ) ou que ficaram apenas em observação ( $X_3 = 0$ ), considerando os ajustes dos modelos MBFC-NI1, 2 e 3, respectivamente.

Os *box-plots* apresentados na Figura 3.29 nos mostram que os indivíduos que receberam o tratamento apresentam em média uma probabilidade maior de serem curados do que aqueles que permaneceram apenas em observação após a cirurgia. O comportamento das estimativas sob cada ajuste foi semelhante. Em todos eles observamos que a probabilidade máxima de cura estimada entre os pacientes que estiveram somente em observação é em torno de 30%, que é um valor menor do que o mínimo estimado entre os pacientes que receberam o tratamento.

Para investigar a influência da idade sobre a fração de cura, dividimos os indivíduos em cinco faixa etárias considerando intervalos a cada 12.2 anos, sendo a menor e a maior idade 18 e 79 anos, respectivamente. Na Figura 3.30 notamos ao analisarmos os *box-plots* das estimativas para valor esperado da distribuição *a posteriori* da fração de cura, sob todos os modelos ajustados, que a medida que os pacientes possuem uma idade mais avançada há uma certa redução na probabilidade de cura estimada. Porém, vemos que a diferença entre as frações de cura do indivíduos com idades entre 18 e 30.2 anos e as frações de cura dos indivíduos com idades entre 66.8 e 79 anos não é tão grande. Essa deve ser o motivo de alguns dos modelos analisados não determinarem estimativas que apontam a idade como um fator significativo para o tempo de falha.



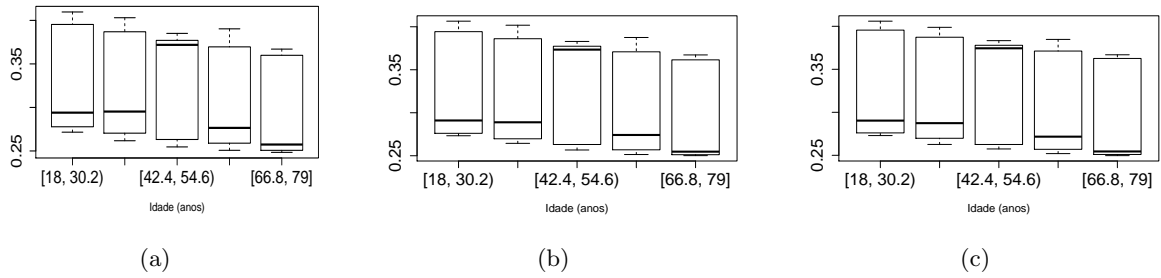


Figura 3.30: *Box-plots* das estimativas do valor esperado da distribuição *a posteriori* da fração de cura dos indivíduos observados, considerando as diferentes faixa etárias, sob o ajuste dos modelos MBFC-NI1 (a), 2 (b) e 3 (c).

Podemos observar também uma comparação entre as frações de cura estimadas dos indivíduos divididos entre os que receberam o tratamento por Interferon e aqueles que permaneceram apenas em observação, por faixa etária. Na Figura 3.31, apresentamos os *box-plots* das estimativas do valor esperado da distribuição *a posteriori* da fração de cura dos indivíduos especificados por tratamento e faixa etária. Percebemos que a idade dos indivíduos parece ser mais relevante entre aqueles que passaram pelo tratamento com a droga, pois observamos uma maior redução da probabilidade de cura para indivíduos com idade mais avançada. Por outro lado, entre os pacientes que permaneceram apenas em observação, o comportamento das estimativas é muito similar em todas as faixa etárias.

Tendo em vista os resultados obtidos, podemos concluir, tal como na Seção 2.4, que o tratamento pode ser indicado, pois temos indícios de que a probabilidade de sobrevivência dos pacientes que receberam doses da droga é maior do que entre aqueles que não utilizaram o medicamento. Além disso, a idade dos pacientes também pode ser significativa para a eficácia do tratamento. Vimos que para os indivíduos com idade mais avançada, o aumento na probabilidade de cura é um pouco menor que nos mais jovens.

A seguir, lançamos mão de algumas medidas de comparação de modelos, já consideradas na Seção 2.4, para investigar qual dos modelos é o mais apropriado para se ajustar aos dados do estudo E1684. As medidas de comparação aqui utilizadas são recorrentes na literatura de modelos Bayesianos e podem ser vistas com detalhes em Cancho *et al.* (2011); Gelman *et al.* (2013); Oliveira (2015); Spiegelhalter *et al.* (2002) ou mais resumidamente no Apêndice C desse texto. Na Tabela 3.10, exibimos o valor da *Conditional Predictive Ordinate* (CPO), do *Deviance Information Criterion* (DIC) e do *Expected Bayesian Information Criterion* (EBIC) para cada modelo. A medida LPML representa o resumo do logaritmo natural da CPO. Relembramos que a avaliação é feita da seguinte maneira: valores maiores de LPML indicam melhor ajuste, enquanto que as demais medidas julgam como mais adequado aquele modelo cujo valor é menor.

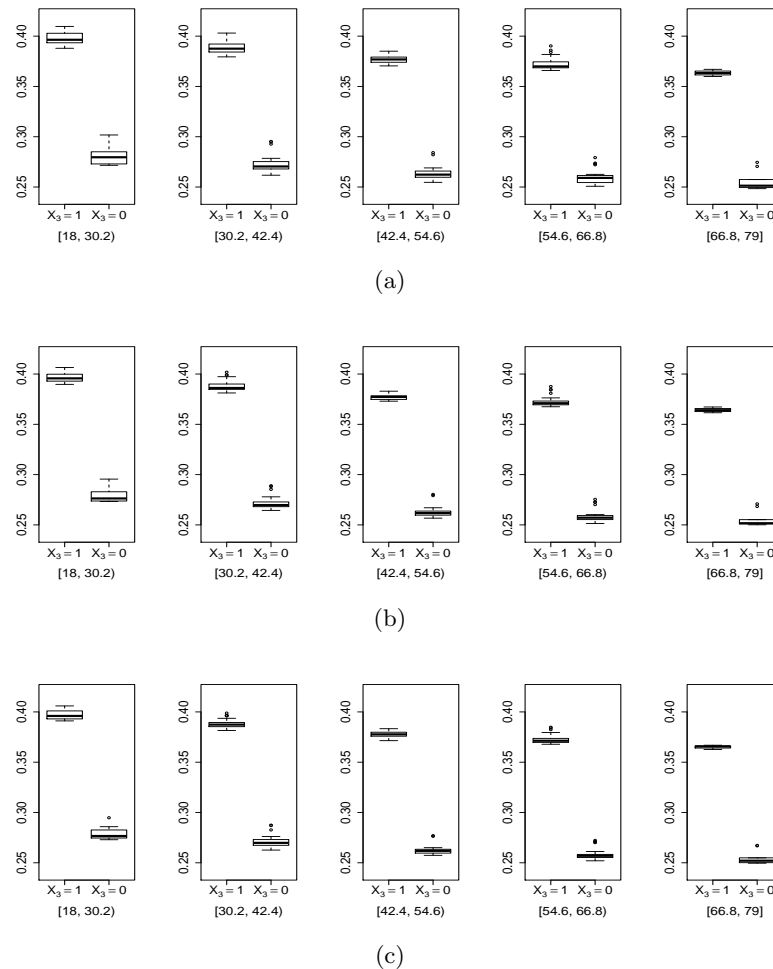


Figura 3.31: *Box-plots* das estimativas do valor esperado da distribuição *a posteriori* da fração de cura dos indivíduos observados para cada faixa etária considerada, divididas entre os indivíduos que receberam Interferon ( $X_3 = 1$ ) e aqueles que permaneceram apenas em observação ( $X_3 = 0$ ), sob o ajuste dos modelos MBFC-NI1 (a), 2 (b) e 3 (c).

Tabela 3.10: Comparação dos Modelos ajustados aos dados do estudo E1684.

Modelo	Critério		
	LPML	DIC	EBIC
MBFC-NI1	-374.548	820.254	3236.638
MBFC-NI2	-275.469	454.283	2767.185
MBFC-NI3	-312.071	677.031	3042.154

Considerando as medidas de comparação de modelos LPML, DIC e EBIC, dadas na Tabela 3.10, o modelo mais apropriado para se ajustar aos dados é o MBFC-NI2, seguido pelo MBFC-

NI3. Destacamos que o MBFC-NI1 é o modelo cuja suposição de erro de medida é mais forte entre os três considerados, pois tal modelo pressupõe uma variância relativamente grande para a distribuição do erro de medida, embora seja menos complexo que os outros dois, devido a suposição acerca do erro de medida. Vimos nos estudos de simulação apresentados na Seção 3.3 que o MBFC-NI2 costuma ser capaz de se ajustar bem a diversos tipos de dados, mesmo nos casos em que os dados são livres de erro de medição.

Se compararmos as medidas de seleção de modelos obtidas na presente análise com aquelas dadas na Tabela 2.15 para os modelos MBFC1, 2, 3 e MTP, vemos que tais medidas selecionam o MTP como o mais apropriado para ajustar-se aos dados, seguido por MBFC2 e depois por MBFC-NI2. sabemos que o modelo MTP é mais simples que o MBFC2, o qual por sua vez é um pouco menos complexo que o MBFC-NI2. Por outro lado, dentre os modelos que supõem erro de medida, as opções MBFC1 e MBFC-NI1 são mais simples, por suporem variância do erro conhecida, porém, nas análises apresentadas as variâncias escolhidas podem ter sido muito maiores do são na realidade. Talvez essa seja a razão de tais modelos serem considerados, pelos critérios de seleção utilizados, os menos apropriados.

Outra característica interessante a ser destacada, é a presença de pontos mais influentes em cada modelo com erro. Neste caso, na Figura 3.32, observemos os gráficos da divergência K-L dos MBFC-NIs.

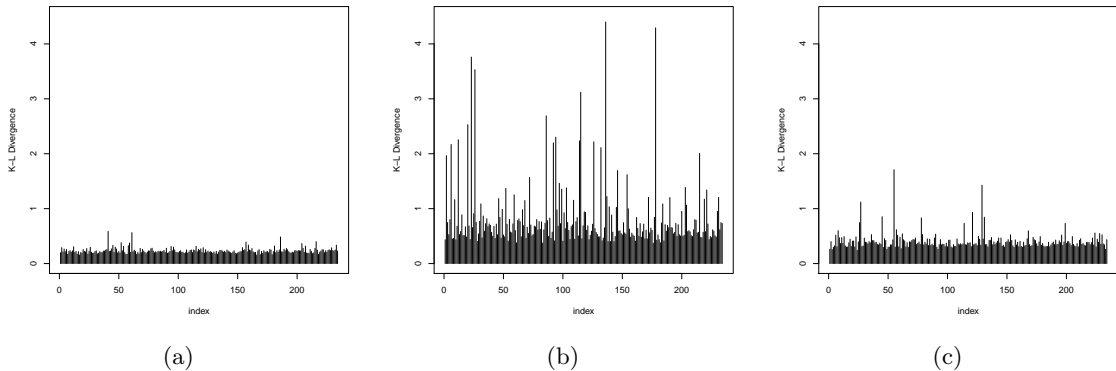


Figura 3.32: Divergência K-L nos modelos MBFC-NI1 (a), MBFC-NI2 (b) e MBFC-NI3 (c).

Uma característica já observada na Seção 2.4 (Figuras 2.35 e 2.48), ao ajustarmos os MBFCs, é a presença de muitos pontos influentes. Na Figura 3.32, percebemos que os MBFC-NIs também apresentam um grande número de observações influentes. Tal como vimos na Figuras 2.35 e 2.48 (dadas na Seção 2.4), em ralação aos MBFCs, o modelo com erro ajustado nesta seção, cujo número de observações influentes é menor, é aquele no qual supomos  $\sigma_\epsilon^2 = k\sigma_x^2$ . Comparando os modelos com erro Normal e com erro  $t$ -Student analisados, dois a dois, de acordo

com a suposição acerca do parâmetro de escala  $\sigma_{\epsilon}^2$  do erro de medida, percebemos que o modo como as medidas de seleção classificam tais modelos é similar nas duas análises apresentadas para os dados do estudo E1684, incluindo em termos de pontos influentes via divergência K-L.

## Capítulo 4

# Considerações Finais

Neste trabalho, introduzimos uma classe de modelos Bayesianos de fração de cura com erro de medida nas covariáveis com distribuição na família mistura de escala da Normal. Inicialmente, consideramos a distribuição *Normal* para o erro de medida, denominando o modelo obtido por MBFC. Estabelecemos três suposições distintas para o parâmetro de escala  $\sigma_\epsilon^2$  da distribuição do erro de medida e denotamos os modelos obtidos por MBFC1, MBFC2 e MBFC3. Supomos  $\sigma_\epsilon^2$  fixado em uma constante conhecida, completamente desconhecido e, por último, proporcional a variância das variáveis latentes supondo  $\sigma_\epsilon^2 = k\sigma_x^2$ , sendo  $k$  uma constante desconhecida. As considerações acerca de  $\sigma_\epsilon^2$ , nas quais o supomos desconhecido e a preocupação em estimar as variáveis latentes com o intuito de modelarmos a incerteza em relação a tais variáveis, são o grande diferencial deste trabalho.

Com o objetivo de avaliarmos o desempenho dos modelos propostos, realizamos alguns estudos de simulação. Primeiramente estudamos o efeito de supormos  $\sigma_\epsilon^2$  conhecido e percebemos que ao fixarmos tal parâmetro em um valor muito maior do que o real, obtemos estimativas demasiadamente imprecisas para todos os parâmetros. Investigamos também a eliciação de distribuições *a priori* para  $\sigma_\epsilon^2$  ao ajustarmos o MBFC2, percebendo não haver sensibilidade a esta escolha. Para analisar o MBFC3, propomos estudos de simulação nos quais avaliamos a eliciação da distribuição *a priori* da constante de proporcionalidade  $k$ . Notamos que o MBFC3 é bastante sensível a tal escolha, pois obtivemos estimativas mais enviesadas para todos os parâmetros a medida em que escolhemos distribuições *a priori* menos informativas para  $k$ . Outro fator investigado foi em relação a eliciação da distribuição *a priori* da variância  $\sigma_x^2$  das variáveis latentes, notando que quando inserimos mais informação sobre tal variância, reduzimos um pouco o viés e EQM médios das estimativas. O estudo de simulação apresentado na Seção 2.3 foi bastante extenso e pudemos estudar muitas características dos modelos propostos. Além de avaliarmos a sensibilidade em relação a escolha das distribuições *a priori* dos parâmetros  $\sigma_\epsilon^2, \sigma_x^2$  e  $k$ , com-

paramos os modelos introduzidos no Capítulo 2 com o MTP proposto por Chen *et al.* (2002d), ajustando-os a dados com e sem a presença de erro nas covariáveis. Em ambos os casos notamos que o MBFC2, em geral, é mais apropriado e concluímos que tal opção é mais interessante de ser aplicada em situações práticas, por ser mais flexível e, por apresentar um bom desempenho mesmo nos casos em que os dados são livres de erro de medida. Notamos também que se dispormos de informação *a priori* acerca do erro de medida, sendo possível eliciar uma distribuição *a priori* muito informativa para a constante  $k$ , o MBFC3 se mostra uma excelente opção, visto que possui um desempenho computacional melhor do que o MBFC2 em termos de velocidade. Tal característica do MBFC3 pode ser bastante desejável na prática. Na Seção 2.4, nos utilizamos de algumas medidas de seleção de modelos para avaliar se os MBFCs são apropriados para analisar o conjunto de dados E1684, referente a um ensaio clínico de pacientes com melanoma cutâneo, realizado no Eastern Cooperative Oncology Group e já analisado previamente por Chen *et al.* (1999). Consideramos as distribuições Weibull e Exponencial para modelarmos os tempos de falha. Ao supormos a distribuição Weibull, as estatísticas LPML, DIC e EBIC apontaram o MBFC2 como o modelo mais apropriado entre aqueles com erro de medida. Ao analisarmos os pontos influentes através dos gráficos da divergência de Kullback-Leibler, o MBFC3 foi o modelo que apresentou uma menor quantidade de observações influentes. Na análise em que assumimos distribuição Exponencial, consideramos uma versão Bayesiana e estrutural do modelo proposto por Ma e Yin (2008), denotada por MBFC-Ma, a qual definimos na Seção 2.2. Nesse caso, considerando apenas os modelos com a suposição de erro de medida nas covariáveis, os critérios de avaliação de ajuste apontaram o MBFC-Ma como o mais apropriado para o estudo E1684, seguido pelo MBFC2. Novamente o MBFC3 apareceu tendo a menor presença de observações influentes dentre os modelos com erro de medida.

No Capítulo 3, apresentamos uma extensão para o MBFC introduzido no Capítulo 2. Desta vez supomos erro de medida com distribuição na família *Normal Independente*, obtendo uma classe mais robusta de modelos com erro nas covariáveis. O modelo com erro Normal aparece como um caso particular da nova proposta. Denotamos o modelo estrutural com erro de medida nas covariáveis Normal Independente como MBFC-NI. Dentre as distribuições membros da família NI também temos a distribuição *t-Student*, a qual possui cauda mais pesada que a distribuição Normal e, por essa razão, é mais apropriada nos casos em que os dados apresentam observações atípicas. Definimos uma representação hierárquica para o MBFC-NI, na qual o erro de medida depende de uma variável misturadora  $U$ . A suposição acerca da variável  $U$  determina a distribuição do erro de medida. Tal representação hierárquica facilitou bastante a implementação dos algoritmos MCMC utilizados para amostrarmos das distribuições *a posteriori* dos parâmetros. Na Seção 3.3, apresentamos alguns estudos de simulação para avaliarmos o novo modelo proposto. Em todos os estudos de simulação assumimos  $U \sim G\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ , obtendo

distribuição *t-Student* com  $\nu$  graus de liberdade para o erro de medida. Consideramos os graus de liberdade  $\nu$  como conhecido, investigando o efeito de tal suposição. Percebemos que quando supomos um valor errado para  $\nu$  é melhor que tal valor seja menor que o real, ou seja, é mais apropriado optarmos por uma distribuição com cauda mais pesada do que a distribuição real. Ao assumirmos  $\nu = 2.5$ , por exemplo, as estimativas obtidas para os parâmetros são um pouco mais precisas. Além disso, também investigamos o que acontece ao utilizarmos os MBFC-NIs para analisar dados sem a presença de erro de medida, supondo  $\nu = 2.5$  e 30, comparando-o ao MTP (do qual os dados foram gerados). Para ambos os valores de  $\nu$ , ao ajustarmos os modelos MBFC-NI1 e 2, obtivemos estimativas aproximadamente iguais aquelas determinadas ao ajustarmos o MTP, dando indícios que tais modelos também são apropriados para ajustar dados desta natureza. Entretanto, o MBFC-NI2 forneceu uma estimativa relativamente grande para a variância do erro de medida, como se a covariável  $X_2$  tivesse sido observada com erro. É importante notar que ao utilizarmos os modelos MBFC-NI, supondo distribuição com cauda mais pesada para o erro de medida, teremos como consequência a obtenção de estimativas inflacionadas para a variância do erro, pois esta é uma suposição embutida na modelagem proposta. Contudo, vimos que isso não resulta na obtenção de estimativas mais imprecisas para os parâmetros, variáveis latentes e fração de cura. Isto é bem perceptível ao compararmos os modelos MBFC-NI com os MBFC e MTP, tanto ao ajustarmos os dados livres de erro de medida, quanto com erro *t-Student* (e também com erro Normal). Via MBFC-NI3, enfrentamos algumas dificuldades computacionais que não permitiram utilizarmos distribuições *a priori* Gama informativas para a constante  $k$ . Nesse caso, optamos pela distribuição Uniforme que, por ser pouco informativa, não pareceu tão apropriada para modelarmos  $k$  nos casos considerados. Uma alternativa para contornar tal dificuldade poderá ser objeto de estudos futuros.

Na Seção 3.4, consideramos o mesmo banco de dados previamente analisado no Capítulo 2, para ajustarmos os modelos MBFC-NI com erro *t-Student*, supondo  $\nu = 2.5$  graus de liberdade. Em nossa análise lançamos mão novamente das medidas de seleção LPML, DIC e EBIC para avaliarmos qual dos modelos ajustados é mais apropriado para os dados. O MBFC-NI2 foi o modelo selecionado pelas três estatísticas consideradas, sendo seguido pelo MBFC-NI3. Entretanto, ao compararmos os MBFC-NIs com os MBFCs (considerados na Seção 2.4), vemos que o MBFC2 apresenta um melhor desempenho que o MBFC-NI2. Além disso, notamos que, dentre os modelos com erro de medida analisados, o MBFC3 e o MBFC-NI3 apresentam mais estabilidade diante de pequenas modificações, pois os gráficos da divergência K-L os mostram como possuindo um menor número de observações influentes.

Uma limitação deste trabalho é a suposição dos graus de liberdade  $\nu$  conhecido, quando admitimos distribuição *t-Student* para o erro de medida. Em trabalhos futuros pode-se investigar o uso de algumas distribuições *a priori* já utilizadas na literatura para modelar os graus de

---

liberdade da distribuição *t-Student*, como por exemplo a distribuição proposta por Fonseca *et al.* (2008) e as distribuições Exponencial e Exponencial Truncada já consideradas por Rocha *et al.* (2016) e Villa *et al.* (2014). Além disso, a eliciação de distribuições *a priori* para o parâmetro  $k$  do MBFC-NI3 precisa ser investigada, com o objetivo de obtermos estimativas um pouco mais precisas para os parâmetros deste modelo.

Existem muitas outras alternativas de modelagem que poderiam flexibilizar ainda mais os modelos com erro de medida propostos neste artigo. Algumas suposições menos restritivas para modelar as variáveis latentes, os tempos de falha ou o erro de medida das covariáveis podem ser acrescentadas. Neste sentido a utilização de *clusters* para modelar algumas dessas variáveis pode ser um caminho promissor. Outra alternativa é considerar para os tempos de falha uma abordagem semiparamétrica, tal como em Demarqui *et al.* (2011), com o objetivo de adequar o uso de tais modelos a uma maior quantidade de situações práticas.



# Referências Bibliográficas

- Andrews, D. F. e Mallows, C. L. (1974). Scale mixtures of normal distributions, *Journal of the Royal Statistical Society. Series B* pp. 99–102.
- Apanasovich, T. V., Carroll, R. J. e Maity, A. (2009). Simex and standard error estimation in semiparametric measurement error models, *Electronic Journal of Statistics* **3**: 318.
- Berkson, J. e Gage, R. (1952). Survival curve for cancer patients following treatment, *Journal of the American Statistical Association* **47**(259): 501–515.
- Bertrand, A., Legrand, C., Carroll, R. J., De Meester, C. e Van Keilegom, I. (2016). Inference in a survival cure model with mismeasured covariates using a simulation-extrapolation approach, *Biometrika* **104**(1): 31–50.
- Bertrand, A., Legrand, C., Léonard, D. e Van Keilegom, I. (2017). Robustness of estimation methods in a survival cure model with mismeasured covariates, *Computational Statistics & Data Analysis* **113**: 3–18.
- Bolfarine, H. e Arellano-Valle, R. B. (1998). Weak nondifferential measurement error models, *Statistics & Probability Letters* **40**(3): 279–287.
- Bolfarine, H., Rodrigues, J. e Cordani, L. (1992). O modelo de regressão com erros nas variáveis, *Texto de Minicurso do X SINAPE, Sao Paulo, SP*.
- Cancho, V. G., Dey, D. K., Lachos, V. H. e Andrade, M. G. (2011). Bayesian nonlinear regression models with scale mixtures of skew-normal distributions: Estimation and case influence diagnostics, *Computational Statistics & Data Analysis* **55**(1): 588–602.
- Chen, M.-H. e Ibrahim, J. G. (2001). Maximum likelihood methods for cure rate models with missing covariates, *Biometrics* **57**(1): 43–52.

- 
- Chen, M.-H., Harrington, D. P. e Ibrahim, J. G. (2002a). Bayesian cure rate models for malignant melanoma: a case-study of eastern cooperative oncology group trial e1690, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **51**(2): 135–150.
- Chen, M.-H., Ibrahim, J. G. e Lipsitz, S. R. (2002b). Bayesian methods for missing covariates in cure rate models, *Lifetime Data Analysis* **8**(2): 117–146.
- Chen, M.-H., Ibrahim, J. G. e Sinha, D. (1999). A new bayesian model for survival data with a surviving fraction, *Journal of the American Statistical Association* **94**(447): 909–919.
- Chen, M.-H., Ibrahim, J. G. e Sinha, D. (2002c). Bayesian inference for multivariate survival data with a cure fraction, *Journal of Multivariate Analysis* **80**(1): 101–126.
- Chen, M.-H., Ibrahim, J. G. e Sinha, D. (2002d). Bayesian inference for multivariate survival data with a cure fraction, *Journal of Multivariate Analysis* **80**(1): 101–126.
- Chib, S. (1992). Bayes inference in the tobit censored regression model, *Journal of Econometrics* **51**(1-2): 79–99.
- Cowles, M. K. e Carlin, B. P. (1996). Markov chain monte carlo convergence diagnostics: a comparative review, *Journal of the American Statistical Association* **91**(434): 883–904.
- Cox, D. R. (1972). Regression models and life-tables, *Journal of the Royal Statistical Society. Series B* **34**(2): 527–541.
- Demarqui, F. N., Dey, D. K., Loschi, R. H. e Colosimo, E. A. (2011). Modeling survival data using the piecewise exponential model with random time grid, *Recent Advances in Biostatistics: False Discovery Rates, Survival Analysis, and Related Topics*, World Scientific, pp. 109–122.
- Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors, *Biometrics* pp. 1041–1046.
- Fonseca, T. C., Ferreira, M. A. e Migon, H. S. (2008). Objective bayesian analysis for the student-t regression model, *Biometrika* **95**(2): 325–333.
- Fuller, W. A. (2009). *Measurement error models*, Vol. 305, John Wiley & Sons, New York.
- Gelfand, A. E. (2000). Gibbs sampling, *Journal of the American statistical Association* **95**(452): 1300–1304.
- Gelfand, A. E., Hills, S. E., Racine-Poon, A. e Smith, A. F. (1990). Illustration of bayesian inference in normal data models using gibbs sampling, *Journal of the American Statistical Association* **85**(412): 972–985.

- 
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. e Rubin, D. B. (2013). *Bayesian Data Analysis*, CRC press, New York.
- Giolo, S. R. e Colosimo, E. A. (2006). Análise de sobrevivência aplicada, *Edgard Blucher*.
- Haario, H., Saksman, E. e Tamminen, J. (2001). An adaptive metropolis algorithm, *Bernoulli* **7**(2): 223–242.
- Haario, H., Saksman, E. e Tamminen, J. (2005). Componentwise adaptation for high dimensional mcmc, *Computational Statistics* **20**(2): 265–273.
- Ibrahim, J. G., Chen, M.-H. e Sinha, D. (2005). *Bayesian Survival Analysis*, John Wiley & Sons, Ltd, New York.
- Kirkwood, J. M., Strawderman, M. H., Ernstoff, M. S., Smith, T. J., Borden, E. C. e Blum, R. H. (1996). Interferon alfa-2b adjuvant therapy of high-risk resected cutaneous melanoma: the eastern cooperative oncology group trial est 1684., *Journal of clinical oncology* **14**(1): 7–17.
- Lange, K. e Sinsheimer, J. S. (1993). Normal/independent distributions and their applications in robust regression, *Journal of Computational and Graphical Statistics* **2**(2): 175–198.
- Lima, R. (2008). *Modelos de sobrevivência com fração de cura e erro de medida nas covariáveis*, Master's thesis, Programa de Pós Graduação em Matemática Aplicada e Estatística, Universidade Federal do Rio Grande do Norte, Natal.
- Lopes, C. M. C. e Bolfarine, H. (2012). Random effects in promotion time cure rate models, *Computational Statistics & Data Analysis* **56**(1): 75–87.
- Ma, Y. e Yin, G. (2008). Cure rate model with mismeasured covariates under transformation, *Journal of the American Statistical Association* **7**(4): 457–472.
- Migon, H. S., Gamerman, D. e Louzada, F. (2015). *Statistical Inference: An Integrated Approach*, Chapman & Hall/CRC press, New York.
- Mizoi, M. F. (2004). *Influência Local em modelos de sobrevivência com erro*, PhD thesis, Universidade de São Paulo, Instituto de Matemática e Estatística, São Paulo.
- Mizoi, M. F., Bolfarine, H. e Pedroso-De-Lima, A. C. (2007). Cure rate model with measurement error, *Communications in Statistics-Simulation and Computation* **36**(1): 185–196.
- Nakamura, T. (1990). Corrected score function for errors-in-variables models: Methodology and application to generalized linear models, *Biometrika* **77**(1): 127–137.

- 
- Nakamura, T. (1992). Proportional hazards model with covariates subject to measurement error, *Biometrics* **48**(3): 829–838.
- Oliveira, R. P. B. (2015). *Modelos robustos para degradação linear e tempo de falha*, PhD thesis, Universidade Federal de Minas Gerais, Programa de Pós Graduação em Estatística, Belo Horizonte.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
- Rocha, G. H., Loschi, R. H. e Arellano-Valle, R. B. (2016). Bayesian mismeasurement t-models for censored responses, *Statistics* **50**(4): 841–869.
- Rodrigues, J., Cancho, V. G., de Castro, M. e Louzada-Neto, F. (2009). On the unification of long-term survival models, *Statistics & Probability Letters* **79**(6): 753–759.
- Rodrigues, J., Cancho, V. G. e Castro, M. (2008). Teoria unificada de análise de sobrevivência, *Texto de Minicurso do XVIII SINAPE, São Pedro, SP*.
- Rogers, W. H. e Tukey, J. W. (1972). Understanding some long-tailed symmetrical distributions, *Statistica Neerlandica* **26**(3): 211–226.
- Santos, C. d. C. (2015). *Inferência e Interpretação de Parâmetros em Classes mais Flexíveis de Modelos Logísticos*, PhD thesis, Universidade Federal de Minas Gerais, Programa de Pós Graduação em Estatística, Belo Horizonte.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. e Van Der Linde, A. (2002). Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society: Series B* **64**(4): 583–639.
- Tanner, M. A. e Wong, W. H. (1987). The calculation of posterior distributions by data augmentation, *Journal of the American Statistical Association* **82**(398): 528–540.
- Tsodikov, A. (1998). A proportional hazards model taking account of long-term survivors, *Biometrics* pp. 1508–1516.
- Villa, C., Walker, S. G. *et al.* (2014). Objective prior for the number of degrees of freedom of at distribution, *Bayesian Analysis* **9**(1): 197–220.
- Wainstein, A. J. e Belfort, F. A. (2004). Conduta para o melanoma cutâneo, *Rev. Col. Bras. Cir* **31**(3): 204–214.
- Yakovlev, A. Y. e Tsodikov, A. D. (1996). *Stochastic models of tumor latency and their biostatistical applications*, Vol. 1, World Scientific, Singapore.

---

Zeng, D., Yin, G. e Ibrahim, J. G. (2006). Semiparametric transformation models for survival data with a cure fraction, *Journal of the American Statistical Association* **101**(474): 670–684.

## Apêndice A

# Família de Distribuições Normal Independente

Considere uma variável aleatória positiva  $U$  e um vetor  $\mathbf{Z}$  de variáveis aleatórias independentes normal  $n$ -variada com média  $\vec{\mathbf{0}}$ , um vetor de zeros em  $\mathbb{R}^n$ , e matriz de covariância não singular  $\mathbf{\Omega} \in \mathbb{R}^{n \times n}$ . Seja  $\boldsymbol{\mu} \in \mathbb{R}^n$ , dizemos que a variável aleatória

$$\mathbf{S} = \boldsymbol{\mu} + U^{-\frac{1}{2}} \mathbf{\Omega}^{\frac{1}{2}} \mathbf{Z} \quad (\text{A.1})$$

tem distribuição normal independente. Denotaremos  $\mathbf{S} \sim NI(\boldsymbol{\mu}, \mathbf{\Omega}, H(u))$ .

Suponha  $H(u)$  a função de distribuição acumulada de  $U$ . Desse modo, a função densidade de probabilidade de  $\mathbf{S}$  é dada por

$$\int_0^\infty \frac{u^{\frac{n}{2}}}{(2\pi)^{\frac{n}{2}} |\mathbf{\Omega}|^{\frac{1}{2}}} \exp\left\{u \frac{v^2}{2}\right\} dH(u) \quad (\text{A.2})$$

em que  $v^2 = (\mathbf{s} - \boldsymbol{\mu})^T \mathbf{\Omega}^{-1} (\mathbf{s} - \boldsymbol{\mu})$  e  $U$  é chamado de fator misturador e  $H(u)$  distribuição de mistura.

Algumas propriedades interessantes dessa família de distribuições podem ser vistas em Lange e Sinsheimer (1993). Destacamos aqui apenas as propriedades mais relevantes para o entendimento desse texto, que são:

- i) Se a esperança matemática  $E(\mathbf{S})$  existe, então,  $E(\mathbf{S}) = \boldsymbol{\mu}$ ;
- ii) Se a covariância  $cov(\mathbf{S})$  existe, então,  $cov(\mathbf{S}) = E(U^{-1}) \mathbf{\Omega}$ .

Além destas propriedades, vale salientar que dependendo da distribuição que atribuirmos a  $U$  obtemos uma distribuição diferente para  $\mathbf{S}$ . A saber,

### 1. Distribuição Normal

Suponha a distribuição de  $U$  degenerada em um, dessa forma,  $\mathbf{S} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Omega})$ .

### 2. Distribuição $t$ -Student

Suponha  $U \sim \text{Gama}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ , então, tem distribuição  $t$ -Student com parâmetro de locação  $\boldsymbol{\mu}$ , de escala  $\boldsymbol{\Omega}$  e  $\nu$  graus de liberdade, denotaremos por  $\mathbf{S} \sim t(\boldsymbol{\mu}, \boldsymbol{\Omega}, \nu)$ . A função densidade de  $\mathbf{S}$ , nessas condições, é dada por

$$f_{\mathbf{S}|\boldsymbol{\mu}, \boldsymbol{\Omega}, \nu} = \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{|\boldsymbol{\Omega}|^{\frac{1}{2}} \pi^{\frac{n}{2}} \Gamma\left(\frac{\nu}{2}\right) \nu^{\frac{n}{2}}} \left[1 + \frac{v^2}{\nu}\right]^{-\frac{(\nu+n)}{2}}. \quad (\text{A.3})$$

### 3. Distribuição *Slash*

Considere  $U \sim \text{Beta}(\nu, 1)$ , nesse caso,  $\mathbf{S}$  tem distribuição *Slash* multivariada com parâmetros de locação  $\boldsymbol{\mu}$ , de escala  $\boldsymbol{\Omega}$  e de forma  $\nu$ . Denotaremos  $\mathbf{S} \sim SL_n(\boldsymbol{\mu}, \boldsymbol{\Omega}, \nu)$ . A função densidade de  $\mathbf{S}$ , nesse caso, é dada por

$$f_{\mathbf{S}|\boldsymbol{\mu}, \boldsymbol{\Omega}, \nu} = \int_0^1 \nu u^{\nu-1} \phi_n(\mathbf{s} | \boldsymbol{\mu}, u^{-1}\boldsymbol{\Omega}) du, \quad (\text{A.4})$$

em que  $\phi_n(\mathbf{s} | \boldsymbol{\mu}, u^{-1}\boldsymbol{\Omega})$  denota a função densidade da distribuição normal multivariada com parâmetro de escala  $\boldsymbol{\mu}$  e matriz de covariância  $\boldsymbol{\Omega}$ .

### 4. Distribuição Normal Contaminada

Seja  $U$  uma variável aleatória discreta, tal que

$$\mathbb{1}_A(u) = \begin{cases} 1, & \text{se } u \in A; \\ 0, & \text{se } u \notin A, \end{cases}$$

em que  $\varsigma, \xi \in (0, 1)$ . Portanto, a função de probabilidade de  $U$  é dada por

$$f_{u|\varsigma, \xi} = \varsigma \mathbb{1}_{\{\xi\}}(u) + (1 - \xi) \mathbb{1}_{\{1\}}(u). \quad (\text{A.5})$$

Sob essas suposições, dizemos que  $\mathbf{S}$  tem distribuição Normal Contaminada com parâmetro de locação  $\boldsymbol{\mu}$ , de escala  $\boldsymbol{\Omega}$  e de forma  $\varsigma$  e  $\xi$ . Denotamos  $\mathbf{S} \sim NC(\boldsymbol{\mu}, \boldsymbol{\Omega}, \varsigma, \xi)$  e sua função densidade de probabilidade é dada por

$$f_{\mathbf{S}|\boldsymbol{\mu}, \boldsymbol{\Omega}, \varsigma, \xi} = \varsigma \phi_n(\mathbf{s} | \boldsymbol{\mu}, \xi^{-1}\boldsymbol{\Omega}) + (1 - \varsigma) \phi_n(\mathbf{s} | \boldsymbol{\mu}, \boldsymbol{\Omega}). \quad (\text{A.6})$$

Para mais detalhes e outros exemplos de distribuição dessa família, Lange e Sinsheimer (1993) pode ser consultado.

Na Figura A.1 a seguir, comparamos as curvas das densidades das distribuições da família normal independente que discutimos nesse texto. Para simplificar a compreensão através da visualização das curvas dessas distribuições, consideramos apenas o caso univariado. Escolhemos a seguinte configuração para os parâmetros:  $\mu = 0$ ,  $\Omega = 1$ ,  $\nu = 5$ ,  $\zeta = 0.7$  e  $\xi = 0.3$ .

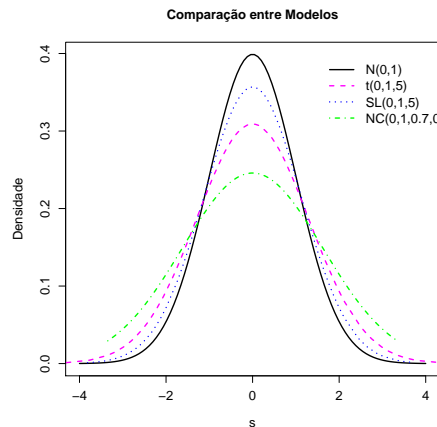


Figura A.1: Comparação entre as f.d.p. das distribuições normal (curva sólida), t-Student (curva tracejada), slash (curva pontilhada) e normal contaminada (curva mista – pontilhada e tracejada).

Podemos observar, comparando as curvas na Figura A.1, que as distribuições t-Student, Slash e normal contaminada, possuem cauda mais pesada que a distribuição normal, e se considerarmos os mesmos valores para os parâmetros que são comuns à ambas, quando  $\zeta = 0.7$  e  $\xi = 0.3$ , a distribuição que apresenta cauda mais pesada é distribuição normal contaminada, seguida pela t-Student e slash, nessa ordem.

Na Figura A.2 a seguir, comparamos as distribuições supracitadas dentro de suas próprias famílias, variando o valor dos parâmetros. Fazemos isso, com o objetivo de estudar o comportamento das caudas de cada uma delas.



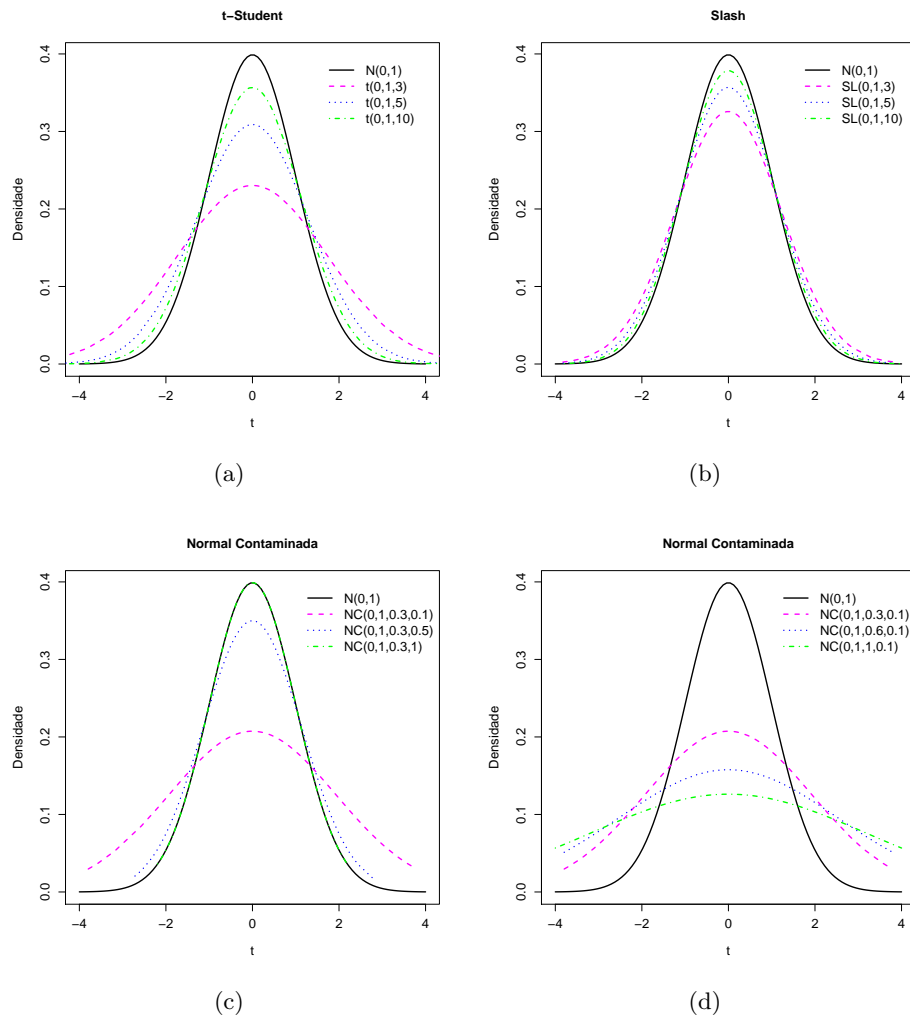


Figura A.2: Na Figura (a), temos a distribuição  $t$ -Student para diferentes valores de  $\nu$ . Na Figura(b), temos a distribuição slash para diferentes valores de  $\nu$  e nas Figuras (c) e (d) temos a distribuição normal contaminada para diferentes valores de  $\xi$  e  $\varsigma$ .

Ao observar os gráficos na Figura A.2a, podemos notar que para a distribuição  $t$ -Student, quanto maior é o valor de  $\nu$ , mais pesada é a cauda da distribuição. À medida que  $\nu$  aumenta, a distribuição  $t$ -Student se aproxima da distribuição normal. Um comportamento análogo pode ser observado para a distribuição Slash na Figura A.2b. Em relação à distribuição normal contaminada, as Figuras A.2c e A.2d sugerem que quando  $\xi \rightarrow 1$ , a distribuição normal contaminada tende à distribuição normal. Em contrapartida, quando  $\varsigma \rightarrow 1$ , a cauda da distribuição se torna mais densa.

## Apêndice B

# Distribuição *a posteriori* dos Parâmetros do MBFC-NI

A seguir, calculamos a distribuição *a posteriori* dos parâmetros do MBFC-NI e fornecemos uma demonstração de que tal distribuição é própria. Vale lembrarmos que o MBFC, introduzido no Capítulo 2, é um caso particular do MBFC-NI, no qual supomos a distribuição que *a priori* da v.a.  $U$  degenerada em 1.

Considere as seguintes distribuições *a priori*: para os parâmetros do MBFC-NI:

$$\begin{aligned} X_{\epsilon_i} | \mu_x, \sigma_x^2 &\stackrel{i.i.d}{\sim} \text{Normal}(\mu_x, \sigma_x^2), \\ U_{\epsilon_i} | \nu &\stackrel{i.i.d}{\sim} \pi(u | \nu), \\ \sigma_\epsilon^2 &\stackrel{i.i.d}{\sim} \pi(\sigma_\epsilon^2), \\ \beta &\propto \mathbf{1}, \\ \lambda &\sim \text{Normal}(m_1, v_1), \\ \alpha &\sim \text{Gama}(a_1, b_1), \\ \mu_x &\sim \text{Normal}(m_2, v_2), \\ \sigma_x^2 &\sim \text{GI}(a_2, v_2), \\ \nu &\sim \pi(\nu). \end{aligned} \tag{B.1}$$

Supondo  $\nu$  como uma constante conhecida, temos que a distribuição *a posteriori* dos

parâmetros do MBFC-NI é dada por:

$$\begin{aligned} \pi(\alpha, \lambda, \boldsymbol{\beta}, \sigma_\epsilon^2, \mathbf{X}_E, \mu_x, \sigma_x^2, U_\epsilon) &\propto L(\alpha, \lambda, \boldsymbol{\beta}, \sigma_\epsilon^2, \mathbf{X}_E | \mathbf{D}_{\text{obs}}) \pi(\alpha, \lambda, \boldsymbol{\beta}, \sigma_\epsilon^2, \mathbf{X}_E, \mu_x, \sigma_x^2, U_\epsilon) \\ &= L(\alpha, \lambda, \boldsymbol{\beta}, \sigma_\epsilon^2, \mathbf{X}_E | \mathbf{D}_{\text{obs}}) \pi(\alpha | a_1, b_1) \pi(\lambda | m_1, v_1) \\ &\quad \times \pi(\mathbf{X}_E | \mu_x, \sigma_x^2) \pi(u | \nu) \pi(\mu_x) \pi(\sigma_x^2) \pi(\sigma_\epsilon^2). \end{aligned} \quad (\text{B.2})$$

**Proposição B.0.1.** *Seja  $\mathbf{W}^* = (\delta_1 \mathbf{W}_1, \dots, \delta_n \mathbf{W}_n)^T$  uma matriz  $n \times p$ . Suponha as distribuições a priori dadas na Equação (B.1), sendo  $\pi(u)$  própria ou degenerada em 1. Além disso, assuma as três especificações, dadas na Tabela 2.1, para  $\sigma_\epsilon^2$ . Se  $\mathbf{W}^*$  é full rank, então a distribuição a posteriori dada na Equação (B.2) é própria.*

*Demonstração.* Utilizando o resultado dado na Equação (2.7), podemos reescrever a Expressão (B.2) da seguinte maneira:

$$\begin{aligned} \pi(\alpha, \lambda, \boldsymbol{\beta}, \sigma_\epsilon^2, \mathbf{X}_E, \mu_x, \sigma_x^2, U_\epsilon) &\propto \left[ \prod_{i=1}^n \left( \exp \{ \mathbf{X}_{\epsilon i} \boldsymbol{\beta}_\epsilon + \mathbf{X}_{oi} \boldsymbol{\beta}_o \} \exp \{ \lambda \} y_i^{\alpha-1} \exp \{ - \exp \{ \lambda \} y_i^\alpha \} \right)^{\delta_i} \right. \\ &\quad \left. \times \exp \{ - \exp \{ \mathbf{X}_{\epsilon i} \boldsymbol{\beta}_\epsilon + \mathbf{X}_{oi} \boldsymbol{\beta}_o \} [1 - \exp \{ - \exp \{ \lambda \} y_i^\alpha \}] \} \right] \\ &\quad \times \pi(\alpha | a_1, b_1) \pi(\lambda | m_1, v_1) \\ &\quad \times \prod_{i=1}^n \int_0^\infty \left( \frac{u_{\epsilon i}}{2\pi\sigma_{\epsilon i}^2} \right)^{\frac{1}{2}} \exp \left\{ - \frac{u_{\epsilon i}}{2\sigma_{\epsilon i}^2} (w_{\epsilon i} - x_{\epsilon i})^2 \right\} dH(u_{\epsilon i}) \\ &\quad \times \pi(\mathbf{X}_E | \mu_x, \sigma_x^2) \pi(u | \nu) \pi(\mu_x) \pi(\sigma_x^2) \pi(\sigma_\epsilon^2). \end{aligned} \quad (\text{B.3})$$

O Teorema 1 de Chen *et al.* (1999) garante que a Expressão

$$\begin{aligned} \pi(\alpha, \lambda, \boldsymbol{\beta}, \sigma_\epsilon^2, \mathbf{X}_E, \mu_x, \sigma_x^2, U_\epsilon) &\propto \left[ \prod_{i=1}^n \left( \exp \{ \mathbf{X}_{\epsilon i} \boldsymbol{\beta}_\epsilon + \mathbf{X}_{oi} \boldsymbol{\beta}_o \} \exp \{ \lambda \} y_i^{\alpha-1} \exp \{ - \exp \{ \lambda \} y_i^\alpha \} \right)^{\delta_i} \right. \\ &\quad \left. \times \exp \{ - \exp \{ \mathbf{X}_{\epsilon i} \boldsymbol{\beta}_\epsilon + \mathbf{X}_{oi} \boldsymbol{\beta}_o \} [1 - \exp \{ - \exp \{ \lambda \} y_i^\alpha \}] \} \right] \\ &\quad \times \pi(\alpha | a_1, b_1) \pi(\lambda | m_1, v_1), \end{aligned}$$

na Equação (B.3), que configura a distribuição *a posteriori* dos parâmetros do modelo de tempo de promoção, é própria. Portanto, como as distribuições *a priori* de  $\pi(\mathbf{X}_E | \mu_x, \sigma_x^2)$ ,  $\pi(\mu_x)$ ,  $\pi(\sigma_x^2)$  e  $\pi(\sigma_\epsilon^2)$  são próprias e, além disso,  $\pi(u | \nu)$  é própria ou degenerada em 1, segue diretamente do Teorema 1 de Chen *et al.* (1999), que a distribuição *a posteriori* do MBFC-NI (Equação B.3) também é própria.  $\square$

## Apêndice C

# Alguns Critérios de Seleção de Modelos

Na literatura existem muitas medidas de seleção de modelos usadas como critério de decisão para escolher qual dentre os modelos disponíveis é o mais apropriado para se ajustar a determinados tipos de dados. Embora muitos autores deixem claro que as medidas de comparação e ajuste são uma alternativa interessante para auxiliar na tarefa de selecionar o modelo mais apropriado, é importante lembrar que muitas características de um modelo que devem ser levadas em conta como por exemplo, a robustez e plausibilidade de utilização do modelo no determinado contexto não são captadas pelos critérios de seleção existentes. Entre os diversos critérios de seleção disponíveis, segundo Cancho *et al.* (2011) um dos mais utilizados em trabalhos aplicados que consideramos ao longo desse texto é derivado da estatística de previsão condicional ordenada (*CPO*). Outros critérios de seleção bastante utilizados para avaliar a qualidade de ajuste de modelos Bayesianos e que também nos utilizamos aqui são o *DIC* (*Deviance Information Criterion*) e o *EBIC* (*Expected Bayesian Information Criterion*). A seguir, apresentamos brevemente a definição de tais medidas. Mais detalhes podem ser encontrados em Cancho *et al.* (2011), Spiegelhalter *et al.* (2002) e Gelman *et al.* (2013).

### ***Conditional Predictive Ordinate (CPO)***

Defina  $\mathbf{D}$  os dados completos e denote por  $\mathbf{D}_{(-i)}$  os dados sem a  $i$ -ésima observação,  $i = 1, \dots, n$ . De acordo com Cancho *et al.* (2011) a  $CPO_i$  relacionada a  $i$ -ésima observação é dada por

$$CPO_i = \int_{\Theta} f(y_i|\boldsymbol{\theta})\pi(\boldsymbol{\theta} | \mathbf{D}_{(-i)})d\boldsymbol{\theta} = \left\{ \int_{\Theta} \frac{\pi(\boldsymbol{\theta} | \mathbf{D})}{f(y_i|\boldsymbol{\theta})} d\boldsymbol{\theta} \right\}^{-1}, i = 1, \dots, n,$$

em que  $\pi(\boldsymbol{\theta} | \mathbf{D}_{(-i)})$  denota a distribuição a posteriori de  $\boldsymbol{\theta}$  condicional nos dados sem a  $i$ -ésima

observação e  $\Theta$  é o espaço paramétrico.

Uma estimativa de Monte Carlo para a  $CPO_i$  pode ser obtida utilizando uma amostra da distribuição *a posteriori*  $\pi(\boldsymbol{\theta} | \mathbf{D})$  (Cancho *et al.*, 2011; Gelman *et al.*, 2013). Considerando  $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_Q$  uma amostra de tamanho  $Q$  de  $\pi(\boldsymbol{\theta} | \mathbf{D})$ , temos que a estimativa de Monte Carlo para a  $CPO_i$  é dada por

$$\widehat{CPO}_i = \left\{ \frac{1}{Q} \sum_{q=1}^Q \frac{1}{f(y_i | \boldsymbol{\theta}_q)} \right\}^{-1}. \quad (\text{C.1})$$

Como critério de seleção baseado na  $CPO_i$  podemos considerar a estatística de resumo:

$$LPML = \sum_{i=1}^n \widehat{CPO}_i,$$

em que um maior valor da  $LPML$  indica que melhor é o modelo ajustado.

#### ***Deviance Information Criterion (DIC)***

Alguns critérios de seleção de modelo são baseados no desvio médio *a posteriori* que pode ser aproximado da seguinte maneira:

$$\bar{D} = \sum_{q=1}^Q \frac{D(\boldsymbol{\theta}_q)}{Q}, \quad (\text{C.2})$$

em que  $D(\boldsymbol{\theta}) = -2 \sum_{i=1}^n \log(f(y_i | \boldsymbol{\theta}))$ .

O DIC é um dos critérios baseados na quantidade  $\bar{D}$  dada na Equação (C.2) e pode ser estimado por  $\widehat{DIC} = \bar{D} + \#(\theta)$ , em que  $\#(\theta)$  é o número de parâmetros do modelo. Quanto menor for o valor do  $DIC$  melhor é considerado o ajuste do modelo.

#### ***Expected Bayesian Information Criterion (EBIC)***

Outra medida de seleção de modelos que se baseia no desvio médio *a posteriori*  $\bar{D}$  dado na Equação (C.2) é a  $EBIC$ . Uma aproximação para a  $EBIC$  que pode ser obtida utilizando uma amostra da distribuição *a posteriori*  $\pi(\boldsymbol{\theta} | \mathbf{D})$  é dada por:

$$\widehat{EBIC} = \bar{D} + \#(\theta) \log(n).$$

Quanto menor for o valor do  $EBIC$  melhor é considerado o ajuste do modelo.

#### ***Divergência de Kullback-Leibler (K-L)***

A divergência de K-L é uma medida que quantifica a sensibilidade do modelo em relação

as observações influentes. Cancho *et al.* (2011) explica que modelos de regressão são sensíveis as suas suposições iniciais e por essa razão é importante que ao avaliá-los se realize uma análise de sensibilidade, pois, modelos com menos observações influentes são mais confiáveis, visto que terão mais estabilidade diante de pequenas modificações. Em modelos Bayesianos a divergência de K-L entre as distribuições *a posteriori*  $\pi(\boldsymbol{\theta} \mid \mathbf{D})$  e  $\pi(\boldsymbol{\theta} \mid \mathbf{D}_{(-i)})$  é dada por:

$$K(P, P_{(-i)}) = \int_{\Theta} \pi(\boldsymbol{\theta} \mid \mathbf{D}) \log \left( \frac{\pi(\boldsymbol{\theta} \mid \mathbf{D})}{\pi(\boldsymbol{\theta} \mid \mathbf{D}_{(-i)})} \right), \quad (\text{C.3})$$

em que  $K(P, P_{(-i)})$  denota a divergência de K-L entre  $P = \pi(\boldsymbol{\theta} \mid \mathbf{D})$  e  $P_{(-i)} = \pi(\boldsymbol{\theta} \mid \mathbf{D}_{(-i)})$ .  $K(P, P_{(-i)})$  quantifica o efeito de se retirar a  $i$ -ésima observação do modelo, na distribuição *a posteriori* dos parâmetros.

Uma aproximação para a divergência de K-L, considerando-se a observação  $i, i = 1, \dots, n$ , pode ser calculada em função da  $CPO_i$  da seguinte maneira:

$$K(\widehat{P}, \widehat{P}_{(-i)}) = -\log(\widehat{CPO}_i) + \frac{1}{Q} \sum_{q=1}^Q \log(f(y_i \mid \boldsymbol{\theta}_q)), \quad (\text{C.4})$$

em que  $\widehat{CPO}_i$  é dada na Equação (C.1). Consideramos que uma observação é muito influente se a quantidade  $p_i = \frac{1}{2} \left[ 1 + \sqrt{1 - \exp \{-2K(P, P_{(-i)})\}} \right]$  é muito maior do que  $\frac{1}{2}$ .

## Apêndice D

# Código para Gerar Dados com Erro de Medida Normal

O código disponibilizado a seguir foi utilizado para gerar os dados do Cenário 1 da Seção 2.3, no software estatístico R (R Core Team, 2016).

```
rm(list=ls(all=TRUE))
path <- "C:\\Users\\Anna Rafaela\\Dropbox\\Rafaela_Tese2\\EstudoExponencial\\Bancos\\"
fileName = "banco_dados"
amostra = 100
set.seed(7)
###covariáveis
##x1:=sexo dos individuos - 0 se mulher, 1 se homem
x1 = rbinom(amostra,1, 0.4)
##x2:=espessura do tumor
x2.real = rnorm(amostra,3.9,sqrt(10))
cov = cbind(x1,x2.real)
#parametros
beta.real = c(1.8, -0.15)
alpha.real = 1
lambda.real = -1.2
v.real = 1
mu.real = 3.9
sigma2.real = 10
theta = exp(as.vector(cov%*%beta.real))
N=rpois(amostra,theta)
```



```

### Fração de cura verdadeira
fracao.de.cura = exp(-theta)
n.replicas = 250 ## geramos 250 bancos de dados
#####
for(contador in 1:n.replicas){
  # Variável observada com erro
  erro = rnorm(amostra)
  x2 = x2.real + erro
  ##### Tempos #####
  Z=list()
  t = NULL
  c = NULL
  #tempos em que cada célula produz uma massa de tumor detectável
  for(i in 1:amostra){
    if(N[i]>0){
      Z[[i]]=rexp(N[i],exp(lambda.real))
      t[i] = round(min(Z[[i]]),2)
    }else{
      t[i] = 10000
    }
  }
}

#Para 20 % de censura,  $P(T>C)=0.8$ , logo, temos
muc = (exp(lambda.real)*0.8)/(0.2)
c = rexp(amostra,muc)
# auxiliar = t[t!=10000]
# t[t==10000] = max(auxiliar) +1
# y = apply(cbind(t,c), 1,min)
y = apply(cbind(t,c), 1,min)

#Indicador de falha ou censura
delta = NULL
for(i in 1:amostra){
  if(t[i]>c[i]){
    delta[i] = 0
  }else{

```

```
        delta[i] = 1
    }
}
dados = cbind(y, delta, x1, x2, x2.real, erro)
col.names = c("Tempos", "Indicador de Falha", "x1", "X2 com erro", "x2.real", "erro")
write.table(dados, (paste0(path, fileName, contador, ".txt", collapse = NULL)),
row.names=F, sep=" ", col.names=T)
}
```

## Apêndice E

# Código para Gerar Dados com Erro de Medida *t-Student*

O código disponibilizado a seguir foi utilizado para gerar os dados do Cenário 1 da Seção 3.3, no software estatístico R (R Core Team, 2016).

```
rm(list=ls(all=TRUE))
path <- "C:\\Users\\Anna Rafaela\\Dropbox\\Rafaela_Tese2\\T1\\1\\Bancos2\\"
fileName = "banco_dados"
amostra = 100
set.seed(7)
###covariáveis
##x1:=sexo dos individuos - 0 se mulher, 1 se homem
x1 = rbinom(amostra,1, 0.4)
##x2:=espessura do tumor
x2.real = rnorm(amostra,3.9,sqrt(10))
cov = cbind(x1,x2.real)
#parametros
beta.real = c(1.8, -0.15)
alpha.real = 1
lambda.real = -1.2
v.real = 1
mu.real = 3.9
sigma2.real = 10
nu.real = 3
###numeros de celulas cancerígenas de cada indivíduo
```

```

theta = exp(as.vector(cov%*%beta.real))
N=rpois(amostra,theta)
### Fração de cura verdadeira
fracao.de.cura = exp(-theta)
n.replicas = 250
u = rgamma(amostra, nu.real/2, nu.real/2)
erro = NULL
x2 = NULL

for(contador in 1:n.replicas){
  ##### Variável observada com erro
  for(i in 1:amostra){
    z = rnorm(1)
    erro[i] = u[i]^(-0.5)*sqrt(v.real)*z
    x2[i] = x2.real[i] + erro[i]
  }
  #####tempos de falha
  a = alpha.real
  b = exp(-lambda.real/a)
  Z=list()
  t = NULL
  c = NULL
  #tempos em que cada célula produz uma massa de tumor detectável
  for(i in 1:amostra){
    if(N[i]>0){
      Z[[i]]=rweibull(N[i],a, b)
      t[i] = min(Z[[i]])
    }else{
      t[i] = 10000
    }
  }
  m = (exp(lambda.real)*0.7)/(0.3)
  for(i in 1:amostra){
    c[i] = rweibull(1,1,m)
  }
  y = apply(cbind(t,c), 1,min)
}

```

```
#indicador de falha
delta = NULL

for(i in 1:amostra){
  if(min(t[i],c[i])==t[i]){
    delta[i] = 1
  }else{
    delta[i] = 0
  }
}

dados = cbind(y, delta, x1, x2, x2.real, erro, fracao.de.cura, u)
col.names = c("Tempos", "Indicador de Falha", "x1", "X2 com erro",
"x2.real", "erro", "fc", "U")
write.table(dados, (paste0(path, fileName, contador, ".txt", collapse = NULL)),
row.names=F, sep=" ", col.names=T)
}
```

## Apêndice F

# Programa Implementado para Simular o MBFC2

Neste apêndice disponibilizamos o código com o qual simulamos o MBFC2 no estudo de Monte Carlo da Subseção 2.3.2. Na construção do algoritmo, nos utilizamos de algumas funções que permitiram paralelizar a simulação com o objetivo de diminuir o tempo computacional. A linguagem de programação utilizada é R (R Core Team, 2016).

```
##Estudo de simulação de monte carlo
## Modelo com erro para os dados simulados sem intercepto
#### Variância do erro igual a um
rm(list=ls(all=TRUE))
require(coda)
library(parallel)
set.seed(7)
fileName = "banco_dados"
##### Trocar o caminho abaixo:
savePath = "C:\\Users\\Rafaela\\Documents\\Corrigidos\\1\\1\\"
imagePath <- "C:\\Users\\Rafaela\\Documents\\Corrigidos\\1\\1\\rdata.RData"
dbPath <- "C:\\Users\\Rafaela\\Documents\\Corrigidos\\Bancos\\"
firstDBPath <- paste0(dbPath, "banco_dados1.txt", collapse = NULL)
Npost = 2000
lag = 1
burn_in = 0
Nint = Npost*lag + burn_in
```

```
n.replicas = 250
amostra = 100
contador = n.replicas
censura = NULL
vecqq = NULL
someData <- rep(0, n.replicas*amostra*2);
obsW = array(someData, c(n.replicas, amostra, 2))
estimativas.betas = matrix(0,n.replicas,2)
estimativas.alpha = rep(0,n.replicas)
estimativas.lambda = rep(0,n.replicas)
estimativas.mu = rep(0,n.replicas)
estimativas.sigma = rep(0,n.replicas)
estimativas.sigma.erro = rep(0,n.replicas)
estimativas.cura = matrix(0,n.replicas,amostra)
estimativas.latente = matrix(0,n.replicas,amostra)
cont.int = rep(0,n.replicas)
#parametros
beta.real = c(1.8, -0.15)
alpha.real = 1
lambda.real = -1.2
v.real = 1 ##variância do erro de medida
mu.real = 3.9
sigma2.real = 10
##### Trocar o caminho abaixo:
Dados = read.table(file = firstDBPath, header=TRUE, sep="", dec=".")
x1 = Dados[,3]
x2.real = Dados[,5]
cov = cbind(x1,x2.real)
### Fração de cura verdadeira
fracao.de.cura = Dados[,7]
inicio=proc.time() # contando o tempo da simulação
# Calculate the number of cores#####
no_cores <-4 #detectCores() - 1
# Iniciando o cluster
cl <- makeCluster(no_cores)
clusterSetRNGStream(cl, 5)
```

```
# Exportando o Cluster
clusterEvalQ(cl, library(coda))
clusterEvalQ(cl, library(ggplot2))
##### VARIÁVEIS #####
clusterExport(cl, "dbPath")
clusterExport(cl, "fileName")
clusterExport(cl, "amostra")
clusterExport(cl, "Npost")
clusterExport(cl, "v.real")
clusterExport(cl, "beta.real")
clusterExport(cl, "alpha.real")
clusterExport(cl, "lambda.real")
clusterExport(cl, "v.real")
clusterExport(cl, "mu.real")
clusterExport(cl, "sigma2.real")
clusterExport(cl, "x2.real")
clusterExport(cl, "x1")
clusterExport(cl, "censura")
clusterExport(cl, "Nint")
clusterExport(cl, "burn_in")
clusterExport(cl, "lag")
clusterExport(cl, "estimativas.mu")
clusterExport(cl, "estimativas.betas")
clusterExport(cl, "estimativas.alpha")
clusterExport(cl, "estimativas.lambda")
clusterExport(cl, "estimativas.sigma")
clusterExport(cl, "estimativas.sigma.erro")
clusterExport(cl, "estimativas.cura")
clusterExport(cl, "estimativas.latente")
clusterExport(cl, "fracao.de.cura")
clusterExport(cl, "cont.int")
teste = function(contador) {
  require(mvtnorm)
  dados = read.table(file = (paste0(dbPath, fileName, contador, ".txt",
collapse = NULL)), header=TRUE, sep="", dec=".")
### Variáveis
```



```
Y = dados[,1]
delta = dados[,2]
x1 = dados[,3]
x2 = dados[,4]
W = cbind(x1, x2)
### log das condicionais completas
f_betas = function(betas, j, t, ind_falha, cov, lambda, alpha, x){
  cov[,2] = x
  aux1 = as.vector(cov%*%betas)
  aux2 = sum(ind_falha*cov[,j]*betas[j]) - sum(exp(aux1)*(1-exp(-exp(lambda)*
(t^alpha))))
  return(aux2)
}
f_alpha = function(alpha, t, ind_falha, cov, lambda, betas, forma, taxa, x){
  cov[,2] = x
  d = sum(ind_falha)
  aux = as.vector(cov%*%betas)
  aux3 = d*log(alpha) + (alpha-1)*sum(ind_falha*log(t))
  aux4 = sum(-exp(lambda)*ind_falha*(t^alpha)+exp(aux-exp(lambda)*(t^alpha)))
  aux5 = aux3+aux4+dgamma(alpha, forma, taxa, log=TRUE)
  return(aux5)
}
f_lambda = function(lambda, alpha, t, ind_falha, cov, betas, med, vari, x){
  cov[,2] = x
  d = sum(ind_falha)
  aux7 = as.vector(cov%*%betas)
  aux8 = -exp(lambda)*((sum(delta*(t^alpha))))
  aux9 = sum(exp(aux7-(exp(lambda)*(t^alpha))))
  aux10 = aux8+aux9+d*lambda+dnorm(lambda, mean=med, sd=sqrt(vari), log=TRUE)
  return(aux10)
}
f_x = function(x, i, cov, lambda, alpha, t, ind_falha, betas, v, med, vari){
  w = cov[i,2]
  cov[i,2] = x[i]
  aux11 = as.vector(cov[i,]%*%betas)
  aux12 = ind_falha[i]*x[i]*betas[2] - exp(aux11)*(1-exp(-exp(lambda)*
```

```
(t[i]^alpha))
  aux13 = aux12 + dnorm(w, x[i], sqrt(v), log=TRUE) + dnorm(x[i], med,
sqrt(vari), log=TRUE)
  return(aux13)
}
f_v = function(x, cov, forma, taxa, n){
  w = cov[,2]
  cov[,2] = x
  I = diag(n)
  A = as.numeric(t(as.matrix(w-x))%*%I%*(as.matrix(w-x)))
  aux14 = forma +(n/2)
  aux15 = (A/2)+taxa
  est = 1/rgamma(1, aux14, aux15)
  return(est)
}
### Hiperparametros
#alpha:
a = 0.5
b = 0.001
#lambda
media = 0
variancia = 100
#x:
mu = mu.real
sigma = sigma2.real
#mu
media_mu = 5
sigma_mu = 10
#sigma2_x
c = 102
d = 1010
#sigma2_erro
forma = 1
taxa = 1
###variaveis para guardar as amostras a posteriori
coefi = matrix(0,Npost,((dim(W)[2])))
```

```
lambdas = rep(0, Npost)
alphas = rep(0, Npost)
xlatente = matrix(0,Npost,amostra)
vs = rep(0, Npost)
mus = rep(0,Npost)
sigmas = rep(0,Npost)
f.curados = matrix(0,Npost,amostra)
###chutes iniciais
lambda0 = lambda.real
alpha0 = alpha.real
beta0 = beta.real
x0 = x2.real
v0 = v.real
###variaveis auxiliares
g1 = rep(0.2,(dim(W)[2])+4)
g = rep(0.5,length(Y))
k1 = 1; k2 = 1; k3 = 1; k4 = 1; k5 = 1; k6 = 1
k10 = rep(1,length(Y))
z = 0
mean_x = rep(0, amostra)
mean_x_ant = rep(0, amostra)
var_x = rep(0, amostra)
x.cand = x0
vetor = x0
nulo = rep(0,amostra)
razaox=rep(0,amostra)
fracao_de_cura = rep(0,amostra)
##### Gibbs com passos de Metropolis #####
for(k in 1:Nint){
  step = k-1
  #####beta1
  u1 = runif(1)
  b1.cand = rnorm(1, beta0[1], sqrt(g1[1]))
  vetorc = c( b1.cand, beta0[2])
  raza01 = f_betas(betas=vetorc, j=1, t=Y, ind_falha=delta, cov=W, lambda=lambda0,
alpha=alpha0, x=x0) - f_betas(betas=beta0, j=1, t=Y, ind_falha=delta, cov=W,
```

```
lambda=lambda0, alpha=alpha0, x=x0)
psi1 = min(1, exp(razao1))
  if(u1<psi1){
    beta0[1] = b1.cand
    k1 = k1+1
  }
  if(step == 0){ #Lembrar que contador step começa do zero
    mean_beta1 = beta0[1];
    mean_beta1_ant = beta0[1];
    var_beta1 = 0;
  }else{
    mean_beta1_ant = mean_beta1;
    mean_beta1 = (mean_beta1*step+beta0[1])/(step+1);
    var_beta1 = (step-1)/(step)*var_beta1+(mean_beta1_ant^2)+(beta0[1]^2)/(step)-
(step+1)/(step)*(mean_beta1^2);
  }
  #####beta2
  u2 = runif(1)
  b2.cand = rnorm(1, beta0[2], sqrt(g1[2]))
  vetorc = c(beta0[1], b2.cand)
  razao2 = f_betas(betas=vetorc, j=2, t=Y, ind_falha=delta, cov=W, lambda=lambda0,
alpha=alpha0, x=x0) - f_betas(betas=beta0, j=2, t=Y, ind_falha=delta, cov=W,
lambda=lambda0, alpha=alpha0, x=x0)
  psi2 = min(1, exp(razao2))
  if(u2<psi2){
    beta0[2] = b2.cand
    k2 = k2+1
  }
  if(step == 0){ #Lembrar que contador step começa do zero
    mean_beta2 = beta0[2];
    mean_beta2_ant = beta0[2];
    var_beta2 = 0;
  }else{
    mean_beta2_ant = mean_beta2;
    mean_beta2 = (mean_beta2*step+beta0[2])/(step+1);
    var_beta2 = (step-1)/(step)*var_beta2+(mean_beta2_ant^2)+(beta0[2]^2)/(step)-
```

```
(step+1)/(step)*(mean_beta2^2);
  }
  #####lambda
  u = runif(1)
  lambda.cand = rnorm(1, lambda0, sqrt(g1[3]))
  razao7 = f_lambda(lambda=lambda.cand, alpha=alpha0, t=Y, ind_falha=delta, cov=W,
betas=beta0, med=media, vari=variancia, x=x0) - f_lambda(lambda=lambda0, alpha=alpha0,
t=Y, ind_falha=delta, cov=W, betas=beta0, med=media, vari=variancia, x=x0)
  psi = min(1, exp(razao7))
  if(u<psi){
    lambda0 = lambda.cand
    k4 = k4+1
  }
  if(step == 0){ #Lembrar que contador step começa do zero
    mean_lambda = lambda0;
    mean_lambda_ant = lambda0;
    var_lambda = 0;
  }else{
    mean_lambda_ant = mean_lambda;
    mean_lambda = (mean_lambda*step+lambda0)/(step+1);
    var_lambda = (step-1)/(step)*var_lambda+(mean_lambda_ant^2)+(lambda0^2)/(step)-
(step+1)/(step)*(mean_lambda^2);
  }
  #####alpha
  u7 = runif(1)
  repeat{
    alpha.cand = rnorm(1, alpha0, sqrt(g1[4]))
    if (alpha.cand > 0) break
  }
  razao8 = f_alpha(alpha=alpha.cand, t=Y, ind_falha=delta, cov=W, lambda=lambda0,
betas=beta0, forma=a, taxa=b, x=x0) - f_alpha(alpha=alpha0, t=Y, ind_falha=delta,
cov=W, lambda=lambda0, betas=beta0, forma=a, taxa=b, x=x0)
  tau = min(1, exp(razao8))
  if(u7<tau){
    alpha0 = alpha.cand
    k5 = k5+1
```

```

}
if(step == 0){ #Lembrar que contador step começa do zero
  mean_alpha = alpha0;
  mean_alpha_ant = alpha0;
  var_alpha = 0;
}else{
  mean_alpha_ant = mean_alpha;
  mean_alpha = (mean_alpha*step+alpha0)/(step+1);
  var_alpha = (step-1)/(step)*var_alpha+(mean_alpha_ant^2)+(alpha0^2)/(step)-
(step+1)/(step)*(mean_alpha^2);
}
####variancia do erro
v0 = f_v (x=x0, cov=W, forma=forma, taxa=taxa, n=amostra)
####variaveis latentes
for(i in 1:amostra){
  ux = runif(1)
  x.cand[i] = rnorm(1, x0[i], sqrt(g[i]))
  razaox[i] = f_x(x=x.cand, i=i, cov=W, lambda=lambda0, alpha=alpha0, t=Y,
ind_falha=delta, betas=beta0, v=v0, med=mu, vari=sigma) - f_x(x=x0, i=i,
cov=W, lambda=lambda0, alpha=alpha0, t=Y, ind_falha=delta, betas=beta0,
v=v0, med=mu, vari=sigma)
  taux = min(1,exp(razaox[i]))
  if(ux<taux){
    x0[i] = x.cand[i]
  }
}
m = (amostra*sigma_mu*mean(x0) + media_mu*sigma)/(amostra*sigma_mu + sigma)
v = (sigma*sigma_mu)/(amostra*sigma_mu + sigma)
mu = rnorm(1, m, sqrt(v))
A = (amostra/2) + c
B = (1/2)*(sum((x0-mu)^2)) + d
sigma = 1/rgamma(1, shape=A, scale=1/B)
for(j in 1:amostra){
  if(step == 0){ # Lembrar que contador step começa do zero
    mean_x[j] = x0[j];
    mean_x_ant[j] = x0[j];

```

```

    var_x[j] = 0;
  }else{
    mean_x_ant[j] = mean_x[j];
    mean_x[j] = (mean_x[j]*step+x0[j])/(step+1);
    var_x[j] = (step-1)/(step)*var_x[j]+(mean_x_ant[j]^2)+(x0[j]^2)/(step)-
(step+1)/(step)*(mean_x[j]^2);
  }
}
##### Fração de Curados
for(i in 1:amostra){
  fracao_de_cura[i]=exp(-exp(W[i,1]*beta0[1]+x0[i]*beta0[2]))
}
if(k>100 & sum(vetor!= x0)==amostra){
  g1[1] = var_beta1
  g1[2] = var_beta2
  g1[3] = var_lambda
  g1[4] = var_alpha
  g = var_x
}
if (k > burn_in & k%%lag==0){
  z = z +1
  lambdas[z] = lambda0
  alphas[z] = alpha0
  vs[z] = v0
  coefi[z,] = beta0
  xlatente[z,] = x0
  mus[z] = mu
  sigmas[z] = sigma
  f.curados[z,] = fracao_de_cura
}
}
##### Visualizando as estimativas #####
### mu
Media_mu_estimado = mean(mus)
int_cred_mu = as.matrix(HPDinterval(as.mcmc(mus)))
dp_mu = sd(mus)

```

```
    estimativas.mu[contador] = Media_mu_estimado
### sigma2
Media_sigma2_estimado = mean(sigmas)
int_cred_sigma2 = as.matrix(HPDinterval(as.mcmc(sigmas)))
dp_sigma2 = sd(sigmas)
estimativas.sigma[contador] = Media_sigma2_estimado
### Betas
Media_beta_estimado = apply(coefi,2,mean)
int_cred_beta = as.matrix(HPDinterval(as.mcmc(coefi)))
dp_betas = apply(coefi,2,sd)
estimativas.betas[contador,] = Media_beta_estimado
### Alpha
Media_alpha_estimado = mean(alphas)
int_cred_alpha = as.matrix(HPDinterval(as.mcmc(alphas)))
dp_alpha = sd(alphas)
estimativas.alpha[contador] = Media_alpha_estimado
### Lambda
Media_lambda_estimado = mean(lambdas)
int_cred_lambda = as.matrix(HPDinterval(as.mcmc(lambdas)))
dp_lambda = sd(lambdas)
estimativas.lambda[contador] = Media_lambda_estimado
### variância do erro
Media_v_estimado = mean(vs)
int_cred_v = as.matrix(HPDinterval(as.mcmc(vs)))
dp_v = sd(vs)
estimativas.sigma.erro[contador] = Media_v_estimado
### variavel latente
Media_x_estimado = apply(xlatente,2,mean)
int_cred_x = as.matrix(HPDinterval(as.mcmc(xlatente)))
estimativas.latente[contador,] = Media_x_estimado
### fração de cura
Media_fracao_estimado = round(apply(f.curados,2,mean),5)
int_cred_fracao = as.matrix(HPDinterval(as.mcmc(f.curados)))
estimativas.cura[contador,] = Media_fracao_estimado
int = 0
for(i in 1:amostra){
```



```

    if(int_cred_fracao[i,1]<fracao.de.cura[i] & int_cred_fracao[i,2]>fracao.de.cura[i]){
      int = int+1
    }
  }
  pint = int/amostra
  #A ORDEM É: mu, sigma2x, beat1, beta2, alpha, lambda, sigma2e, pint, latentes,
  #fração de cura, W
  return (c(Media_mu_estimado, Media_sigma2_estimado, Media_beta_estimado,
Media_alpha_estimado, Media_lambda_estimado, Media_v_estimado, pint,
Media_x_estimado, Media_fracao_estimado, W))
}
vecqq = parLapply(cl, 1:n.replicas,
                  teste)
stopCluster(cl)
final = proc.time()
tempo = (final-inicio)/3600
####Guardar o resultado de cada réplica em sua variável
for (contador in 1 : n.replicas) {
  estimativas.mu[contador] = vecqq[[contador]][1]
  estimativas.sigma[contador] = vecqq[[contador]][2]
  estimativas.betas[contador,] = vecqq[[contador]][3:4]
  estimativas.alpha[contador] = vecqq[[contador]][5]
  estimativas.lambda[contador] = vecqq[[contador]][6]
  estimativas.sigma.erro[contador] = vecqq[[contador]][7]
  cont.int[contador] = vecqq[[contador]][8]
  npar = 8 #npar eh o numero de parametros
  estimativas.latente[contador,] = vecqq[[contador]][(npar+1):(amostra+npar)]
  estimativas.cura[contador,] = vecqq[[contador]][(amostra+npar+1):(2*amostra+npar)]
  obsW[contador, ,1] = vecqq[[contador]][(2*amostra+npar+1):(3*amostra+npar)]
  obsW[contador, ,2] = vecqq[[contador]][(3*amostra+npar+1):(4*amostra+npar)]
}
setwd(savePath)
#### aproximação para o viés
vies.mu = sum(estimativas.mu - mu.real)/n.replicas
vies.sigma = sum(estimativas.sigma-sigma2.real)/n.replicas
vies.beta1 = sum(estimativas.betas[,1]-beta.real[1])/n.replicas

```

```
vies.beta2 = sum(estimativas.betas[,2]-beta.real[2])/n.replicas
vies.alpha = sum(estimativas.alpha-alpha.real)/n.replicas
vies.lambda = sum(estimativas.lambda-lambda.real)/n.replicas
vies.sigma.erro = sum(estimativas.sigma.erro-v.real)/n.replicas
vies.cura = matrix(0,n.replicas,amostra)

for(vies in 1:n.replicas){
  vies.cura[vies,] = estimativas.cura[vies,] - fracao.de.cura
}

vies.curados = apply(vies.cura,2,sum)/n.replicas
vies.latentes = matrix(0,n.replicas,amostra)
for(vies in 1:n.replicas){
  vies.latentes[vies,] = estimativas.latente[vies,] - x2.real
}

vies.latente = apply(vies.latentes,2,sum)/n.replicas
write.table(round(vies.curados,3), (paste0(savePath, "vies_cura.txt",
collapse = NULL)), row.names=F, sep=" ", col.names=T)
write.table(round(vies.latentes,3), (paste0(savePath, "vies_latentes.txt",
collapse = NULL)), row.names=F, sep=" ", col.names=T)
###guardando os resumos
resumo.beta1 = as.vector(summary(estimativas.betas[,1]))
resumo_beta1 = rbind(resumo.beta1, NULL)
colnames(resumo_beta1) = c("Min.", "1st Qu", "Median", "Mean",
"3rd Qu", "Max.")
write.table(resumo_beta1, (paste0(savePath, "resumo_beta1.txt",
collapse = NULL)), row.names=F, sep=" ", col.names=T)
resumo.beta2 = as.vector(summary(estimativas.betas[,2]))
resumo_beta2 = rbind(resumo.beta2, NULL)
colnames(resumo_beta2) = c("Min.", "1st Qu", "Median", "Mean",
"3rd Qu", "Max.")
write.table(resumo_beta2, (paste0(savePath, "resumo_beta2.txt",
collapse = NULL)), row.names=F, sep=" ", col.names=T)
resumo.mu = as.vector(summary(estimativas.mu))
resumo_mu = rbind(resumo.mu, NULL)
colnames(resumo_mu) = c("Min.", "1st Qu", "Median", "Mean",
```

```
"3rd Qu", "Max.")
write.table(resumo_mu, (paste0(savePath, "resumo_mu.txt",
  collapse = NULL)), row.names=F, sep=" ", col.names=T)
resumo.sigma2 = as.vector(summary(estimativas.sigma))
resumo_sigma2 = rbind(resumo.sigma2, NULL)
colnames(resumo_sigma2) = c("Min.", "1st Qu", "Median", "Mean",
  "3rd Qu", "Max.")
write.table(resumo_sigma2, (paste0(savePath, "resumo_sigma2.txt",
  collapse = NULL)), row.names=F, sep=" ", col.names=T)
resumo.alpha = as.vector(summary(estimativas.alpha))
resumo_alpha = rbind(resumo.alpha, NULL)
colnames(resumo_alpha) = c("Min.", "1st Qu", "Median", "Mean",
  "3rd Qu", "Max.")
write.table(resumo_alpha, (paste0(savePath, "resumo_alpha.txt",
  collapse = NULL)), row.names=F, sep=" ", col.names=T)
resumo.lambda = as.vector(summary(estimativas.lambda))
resumo_lambda = rbind(resumo.lambda, NULL)
colnames(resumo_lambda) = c("Min.", "1st Qu", "Median", "Mean",
  "3rd Qu", "Max.")
write.table(resumo_lambda, (paste0(savePath, "resumo_lambda.txt",
  collapse = NULL)), row.names=F, sep=" ", col.names=T)
resumo.sigma.erro = as.vector(summary(estimativas.sigma.erro))
resumo_sigma.erro = rbind(resumo.sigma.erro, NULL)
colnames(resumo_sigma.erro) = c("Min.", "1st Qu", "Median", "Mean",
  "3rd Qu", "Max.")
write.table(resumo_sigma.erro, (paste0(savePath, "resumo_sigma_erro.txt",
  collapse = NULL)), row.names=F, sep=" ", col.names=T)
info.hpds = as.vector(cont.int)
write.table(info.hpds, (paste0(savePath, "pHPDs.txt", collapse = NULL)),
  row.names=F, sep=" ", col.names=F)
## Guardando os vicios
vicio = cbind(vies.mu, vies.sigma, vies.sigma.erro, vies.beta1, vies.beta2,
  vies.alpha, vies.lambda)
colnames(vicio) = c("mu", "sigma2 x", "sigma2 e", "beta1", "beta2", "alpha",
  "lambda")
write.table(vicio, (paste0(savePath, "resumo_vicio.txt", collapse = NULL)),
```

```
row.names=F, sep=" ", col.names=T)
#### aproximação para o EQM
eqm.mu = sum((estimativas.mu - mu.real)^2)/n.replicas
eqm.sigma = sum((estimativas.sigma-sigma2.real)^2)/n.replicas
eqm.beta1 = sum((estimativas.betas[,1]-beta.real[1])^2)/n.replicas
eqm.beta2 = sum((estimativas.betas[,2]-beta.real[2])^2)/n.replicas
eqm.alpha = sum((estimativas.alpha-alpha.real)^2)/n.replicas
eqm.lambda = sum((estimativas.lambda-lambda.real)^2)/n.replicas
eqm.sigma.erro = sum((estimativas.sigma.erro-v.real)^2)/n.replicas
eqm.cura = matrix(0,n.replicas,amostra)
for(eqm in 1:n.replicas){
  eqm.cura[eqm,] = (estimativas.cura[eqm,] - fracao.de.cura)^2
}
eqm.curados = apply(eqm.cura,2,sum)/n.replicas
eqm.x2 = matrix(0,n.replicas,amostra)
for(eqm in 1:n.replicas){
  eqm.x2[eqm,] = (estimativas.latente[eqm,] - x2.real)^2
}
eqm.latentes = apply(eqm.x2,2,sum)/n.replicas
## Guardando os EQMs
eqms = cbind(eqm.mu, eqm.sigma, eqm.sigma.erro, eqm.beta1, eqm.beta2,
  eqm.alpha, eqm.lambda)
colnames(eqms) = c("mu", "sigma2 x", "sigma2 e", "beta1", "beta2",
  "alpha", "lambda")
write.table(eqms, (paste0(savePath, "eqms.txt", collapse = NULL)),
  row.names=F, sep=" ", col.names=T)
write.table(round(eqm.curados,3), (paste0(savePath, "eqm_cura.txt",
  collapse = NULL)), row.names=F, sep=" ", col.names=T)
write.table(round(eqm.latentes,3), (paste0(savePath, "eqm_latentes.txt",
  collapse = NULL)), row.names=F, sep=" ", col.names=T)
library(ggplot2)
# ####fracao de cura
individuos=rep(1:50,each=n.replicas)
fcura=rep(fracao.de.cura[1:50],each=n.replicas)
fracaodecura1 = as.vector(estimativas.cura[,1:50])
data1 = data.frame(fracaodecura1,individuos,fcura)
```

```
x11(20,10)
ggplot(data1, main="") +
  geom_boxplot( aes(x = as.factor(individuos), y = fracaodecura1),
  color = "darkblue",fill = "azure2")+ ylab("Fração de Cura")+
  xlab("Indivíduo")+
  geom_point(aes(x = as.factor(individuos), y=fcura),shape=8,
  colour="darkblue", size=2)
savePlot(filename = "boxplot_cura1_erro",type = "eps")
savePlot(filename = "boxplot_cura1_erro",type = "pdf")
individuos1=rep(51:100,each=n.replicas)
fcura1=rep(fracao.de.cura[51:100],each=n.replicas)
fracaodecura11 = as.vector(estimativas.cura[,51:100])
data11 = data.frame(fracaodecura11,individuos1,fcura1)
x11(20,10)
ggplot(data11, main="") +
  geom_boxplot( aes(x = as.factor(individuos1), y = fracaodecura11),
  color = "darkblue", fill = "azure2")+ ylab("Fração de Cura")+
  xlab("Indivíduo")+
  geom_point(aes(x = as.factor(individuos1), y=fcura1),shape=8,
  colour="darkblue", size=2)
savePlot(filename = "boxplot_cura11_erro",type = "eps")
savePlot(filename = "boxplot_cura11_erro",type = "pdf")
save.image(imagePath)
```

## Apêndice G

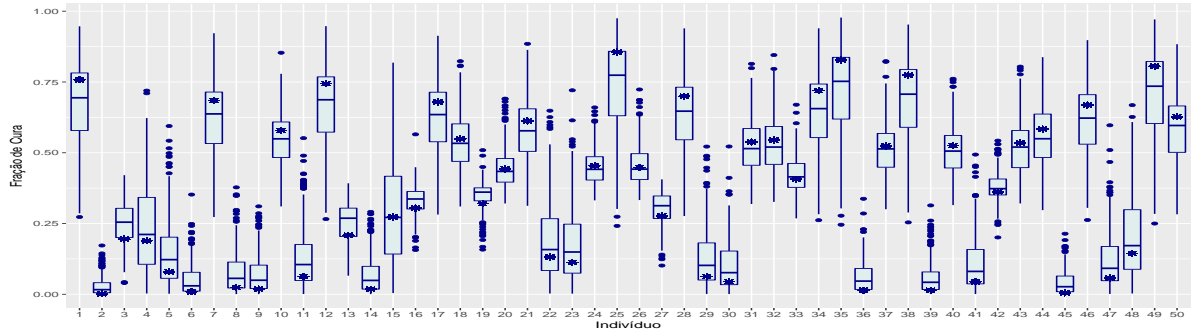
# ***Box-plots* das Estimativas da Fração de Cura Obtidas nos Estudos de Monte Carlo da Seção 2.3**

Nesta seção, apresentamos alguns *box-plots* das estimativas médias para o valor esperado *a posteriori* de cada indivíduo obtidas no estudos de simulação exibidos na Seção 2.3. Nos *box-plots* apresentados é possível visualizarmos melhor o comportamento das estimativas fornecidas para a fração de cura em tais estudos.

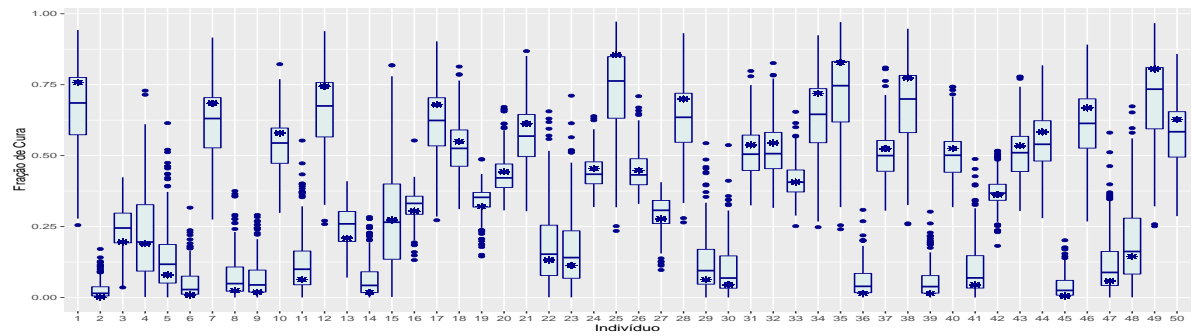
Nas Figuras G.1 a G.6, apresentamos *box-plots* das estimativas obtidas para a fração de cura, ajustando-se o MBFC1 fixando  $\sigma_\epsilon^2 = 1, 0.5$  e  $10$ . Nas Figuras G.1 e G.2 assumimos  $\sigma_x^2 \sim GI(102, 1010)$ , nas Figuras G.3 e G.4, assumimos  $\sigma_x^2 \sim GI(0.1, 0.1)$  e nas Figuras G.5 e G.6.

Nas Figuras G.7 a G.10 a seguir, apresentamos os *box-plots* das estimativas médias para o valor esperado *a posteriori* de cada indivíduo, que foram obtidas, ajustando-se os modelos MTP, MBFC1, MBFC2 e MBFC3, no Estudo 2. Na Figuras G.7 e G.8, no MBFC1 fixamos  $\sigma_\epsilon^2 = 1$ , enquanto que na Figura G.9 e G.10  $\sigma_\epsilon^2 = 5$ .

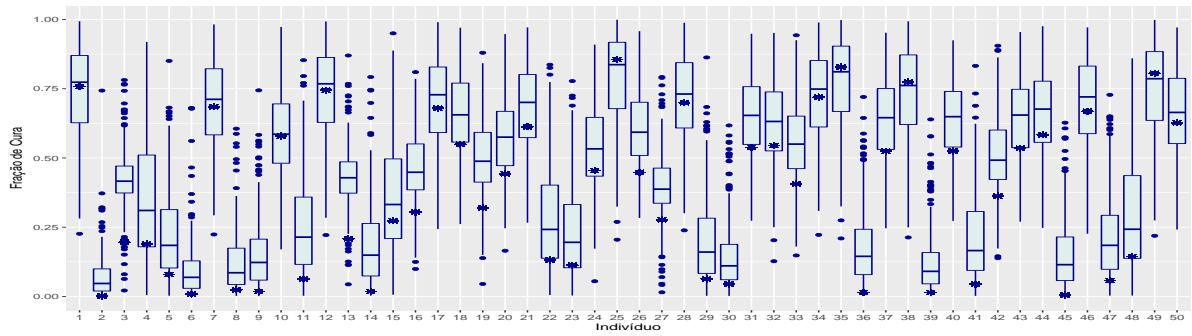
Nas Figuras G.11 e G.12, apresentamos os *box-plots* das estimativas médias para o valor esperado *a posteriori* de cada indivíduo, que foram obtidas no Estudo 3, ajustando-se os modelos MTP, MBFC1 em que fixamos  $\sigma_\epsilon^2 = 1$ , MBFC2 e MBFC3 e assumindo  $\sigma_x^2 \sim GI(102, 1010)$ .



(a)

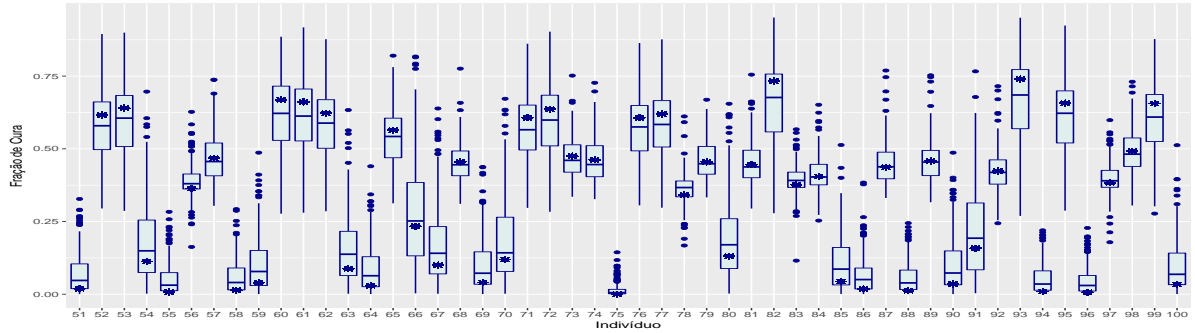


(b)

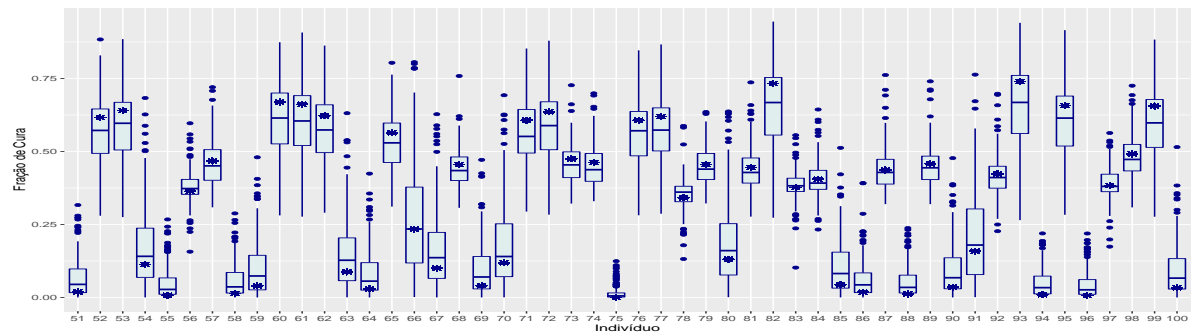


(c)

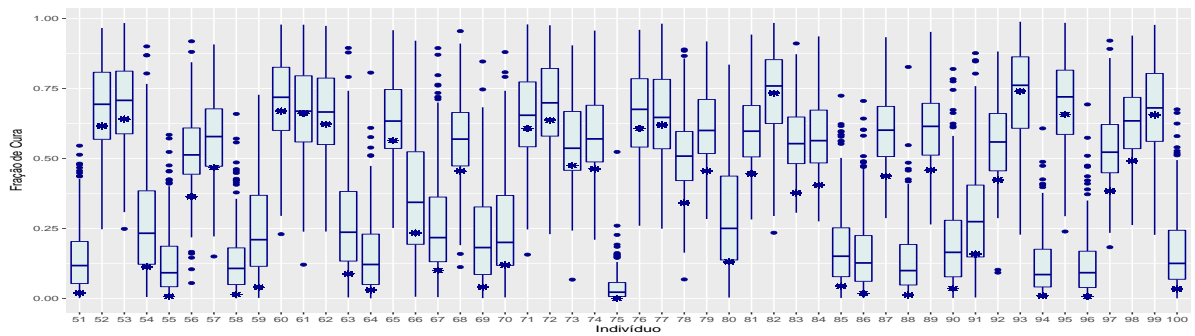
Figura G.1: *Box-plots* dos valores esperados *a posteriori* da fração de cura para cada um dos 50 primeiros indivíduos da amostra ajustando o MBFC1 e assumindo  $\sigma_x^2 \sim GI(102, 1010)$ , fixando  $\sigma_\epsilon^2 = 1$  (Fig.a), 0.5 (Fig.b) e 10 (Fig.c). O \* representa o valor real para a fração de cura de cada indivíduo.



(a)



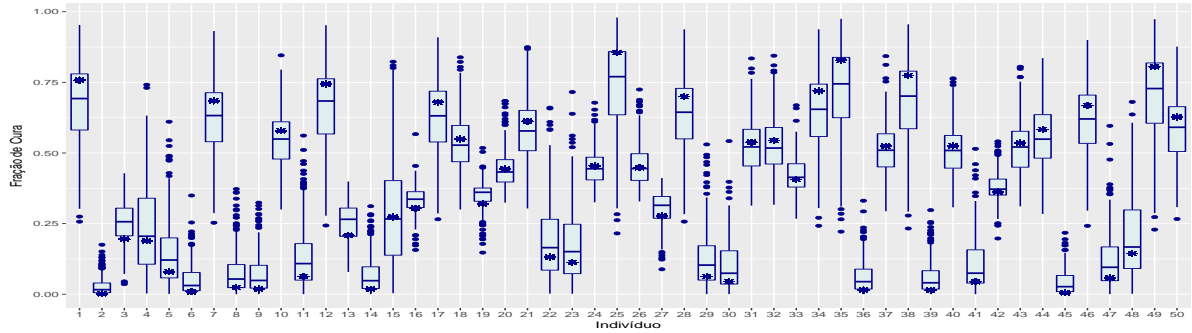
(b)



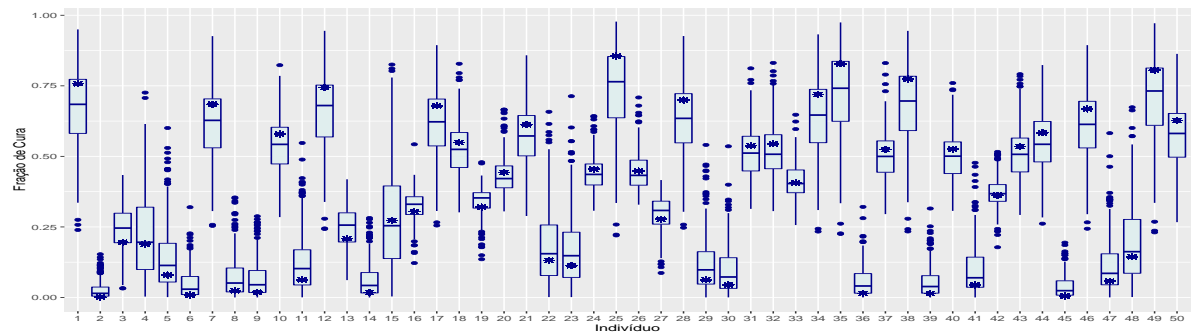
(c)

Figura G.2: *Box-plots* dos valores esperados *a posteriori* da fração de cura para cada um dos 50 últimos indivíduos da amostra ajustando o MBFC1 e assumindo  $\sigma_x^2 \sim GI(102, 1010)$ , fixando  $\sigma_\epsilon^2 = 1$  (Fig.a), 0.5 (Fig.b) e 10 (Fig.c). O \* representa o valor real para a fração de cura de cada indivíduo.

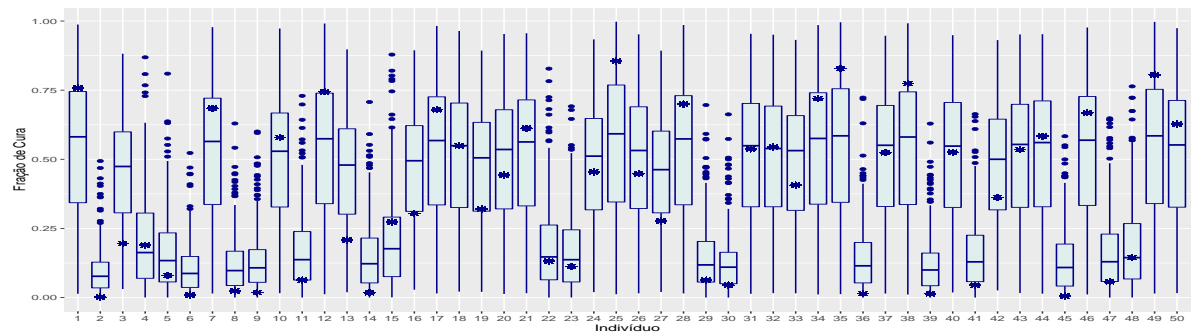




(a)

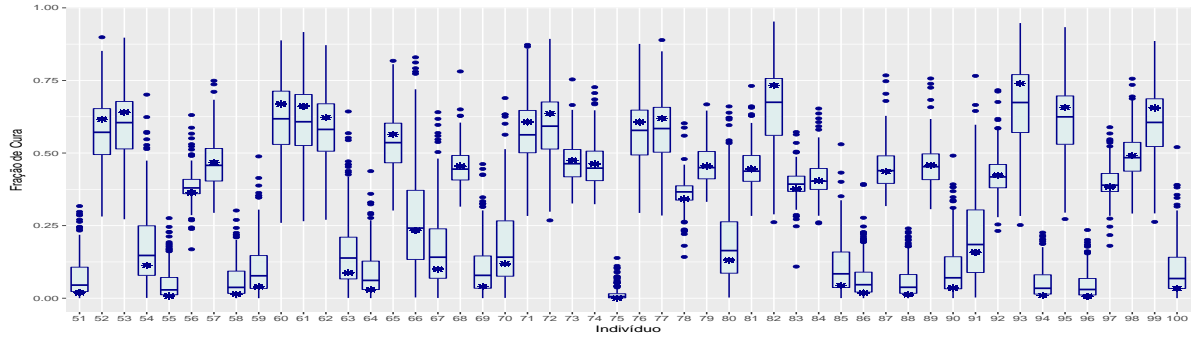


(b)

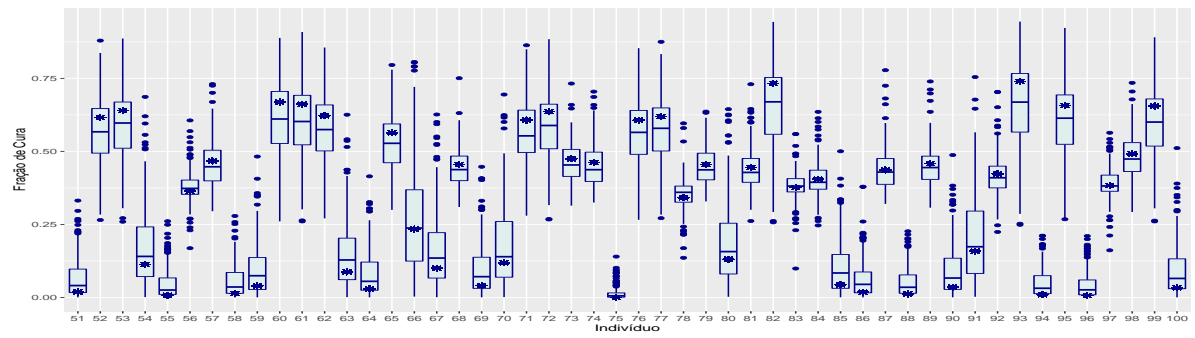


(c)

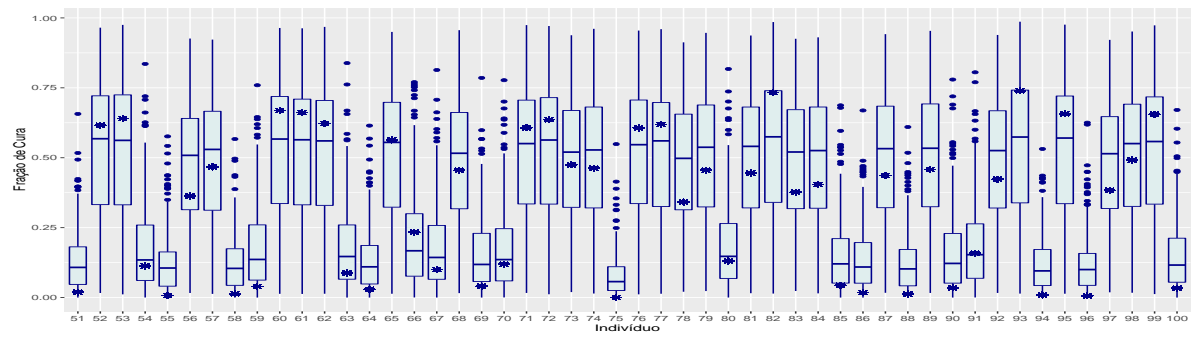
Figura G.3: *Box-plots* dos valores esperados *a posteriori* da fração de cura para cada um dos 50 primeiros indivíduos da amostra ajustando o MBFC1 e assumindo  $\sigma_x^2 \sim GI(0.1, 0.1)$ , fixando  $\sigma_c^2 = 1$  (Fig.a), 0.5 (Fig.b) e 10 (Fig.c). O \* representa o valor real para a fração de cura de cada indivíduo.



(a)

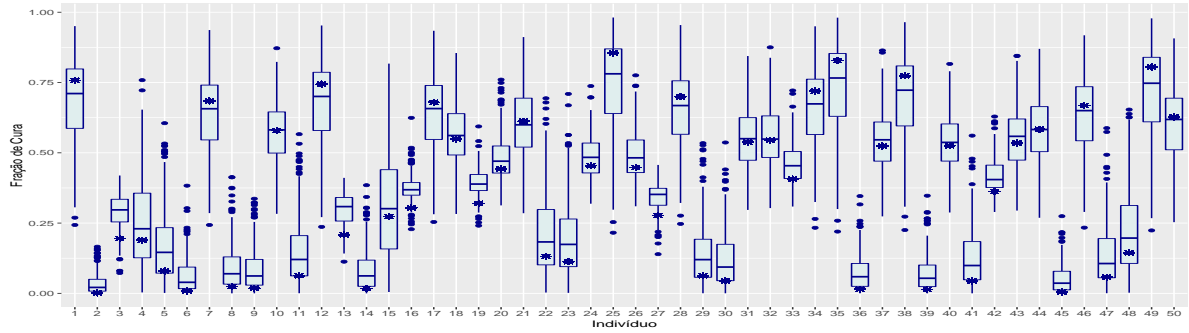


(b)

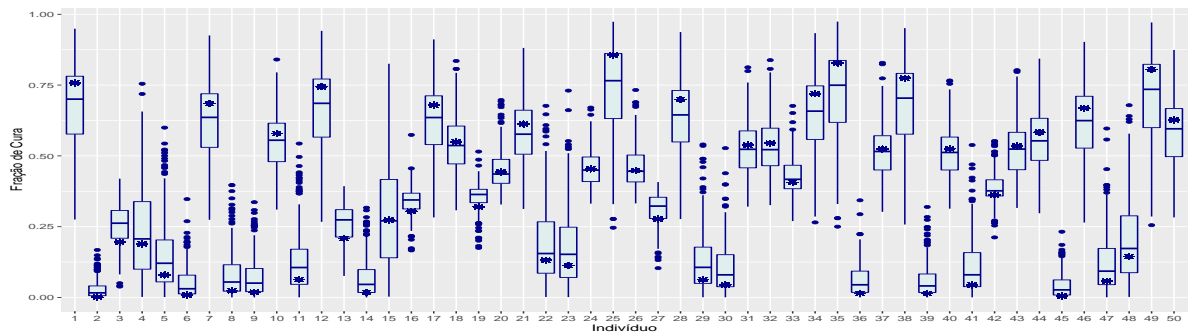


(c)

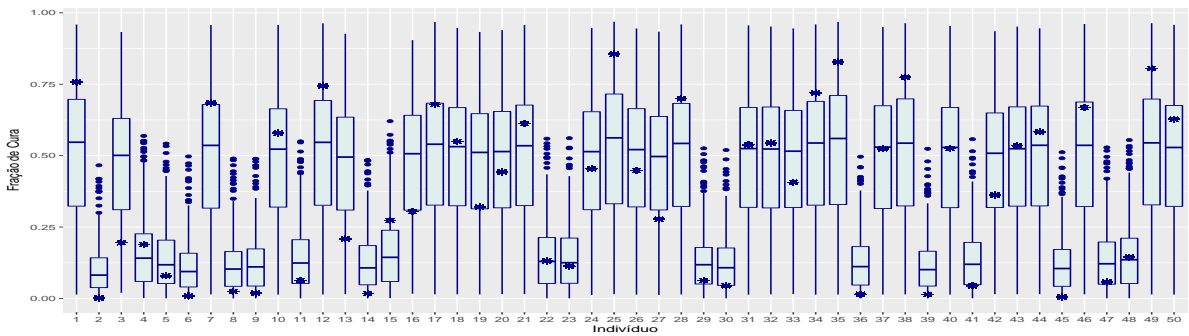
Figura G.4: *Box-plots* dos valores esperados *a posteriori* da fração de cura para cada um dos 50 últimos indivíduos da amostra ajustando o MBFC1 e assumindo  $\sigma_x^2 \sim GI(0.1, 0.1)$ , fixando  $\sigma_c^2 = 1$  (Fig.a), 0.5 (Fig.b) e 10 (Fig.c). O \* representa o valor real para a fração de cura de cada indivíduo.



(a)

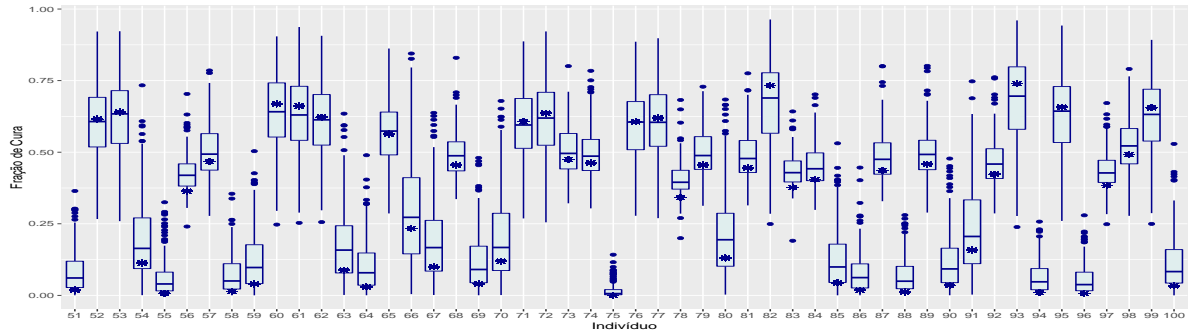


(b)

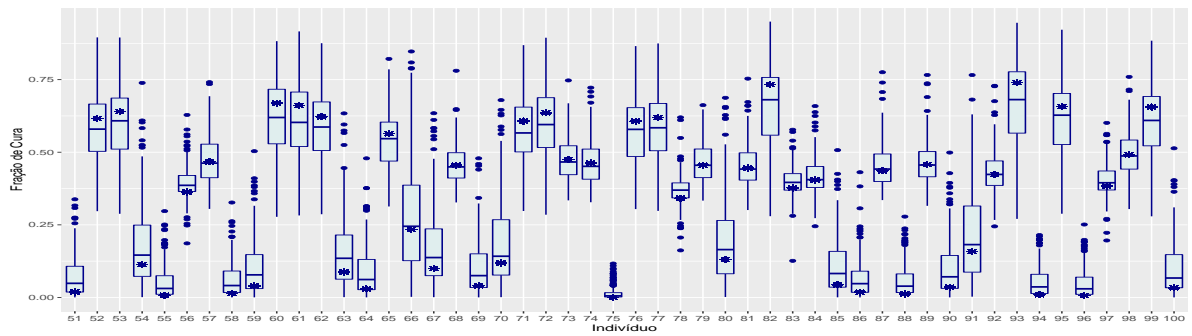


(c)

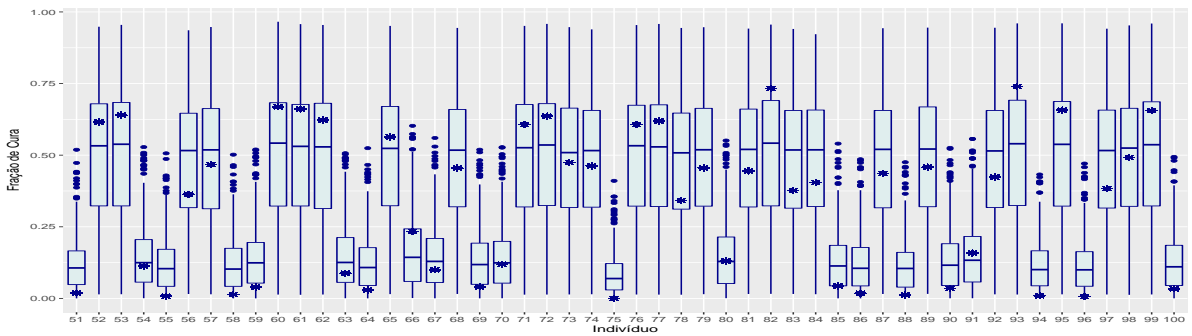
Figura G.5: *Box-plots* dos valores esperados *a posteriori* da fração de cura para cada um dos 50 primeiros indivíduos da amostra ajustando o MBFC1 e assumindo  $\sigma_x^2 \sim GI(100,100)$ , fixando  $\sigma_c^2 = 1$  (Fig.a), 0,5 (Fig.b) e 10 (Fig.c). O \* representa o valor real para a fração de cura de cada indivíduo.



(a)

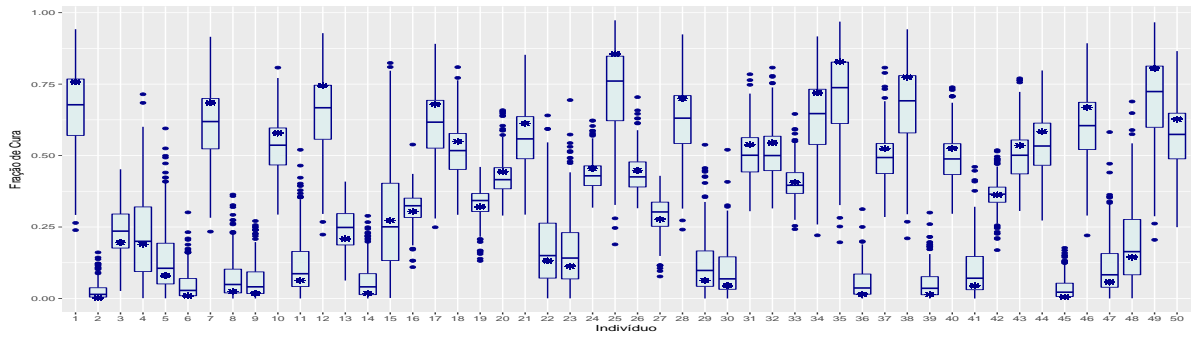


(b)

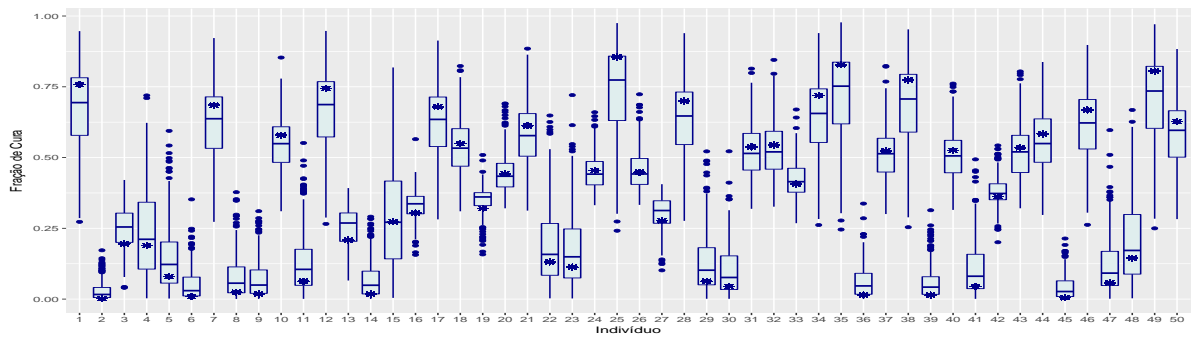


(c)

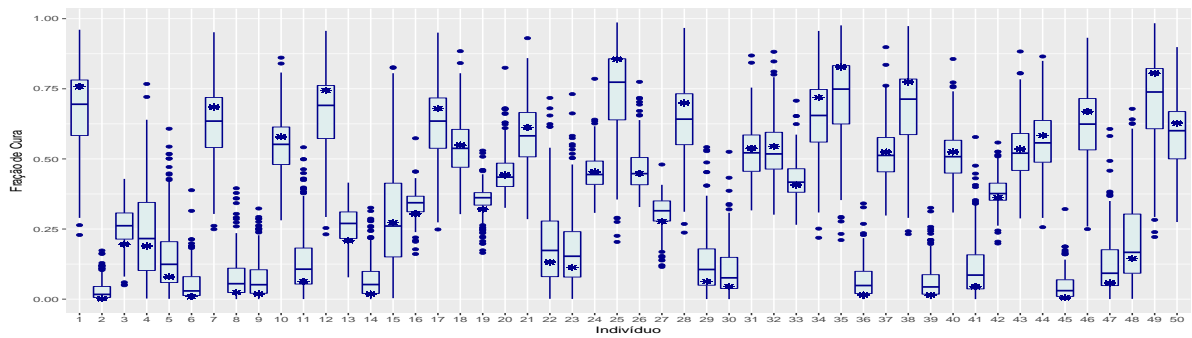
Figura G.6: *Box-plots* dos valores esperados *a posteriori* da fração de cura para cada um dos 50 últimos indivíduos da amostra ajustando o MBFC1 e assumindo  $\sigma_x^2 \sim GI(100,100)$ , fixando  $\sigma_c^2 = 1$  (Fig.a), 0,5 (Fig.b) e 10 (Fig.c). O \* representa o valor real para a fração de cura de cada indivíduo.



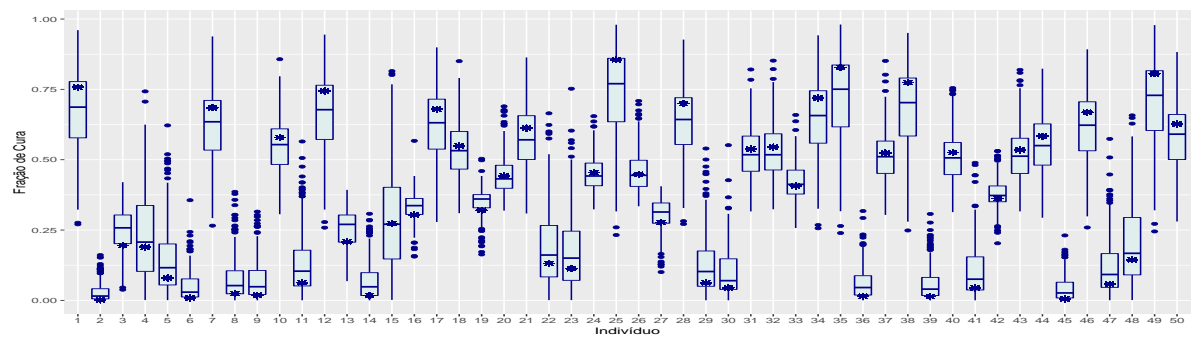
(a)



(b)

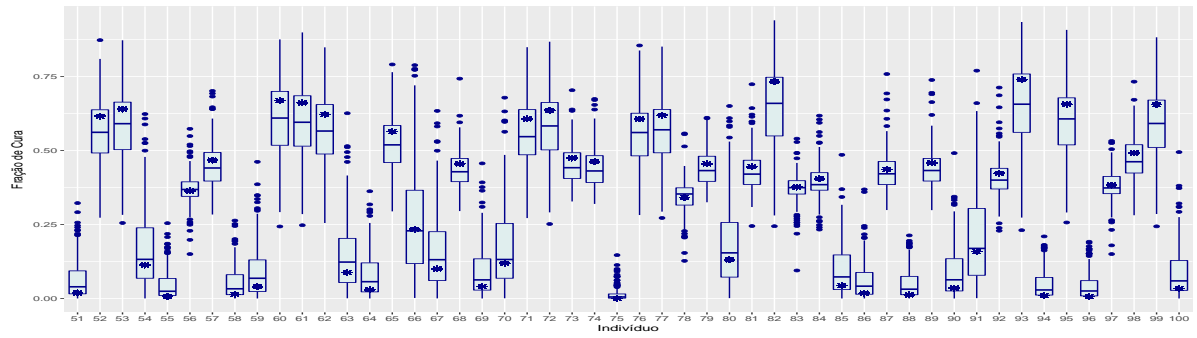


(c)

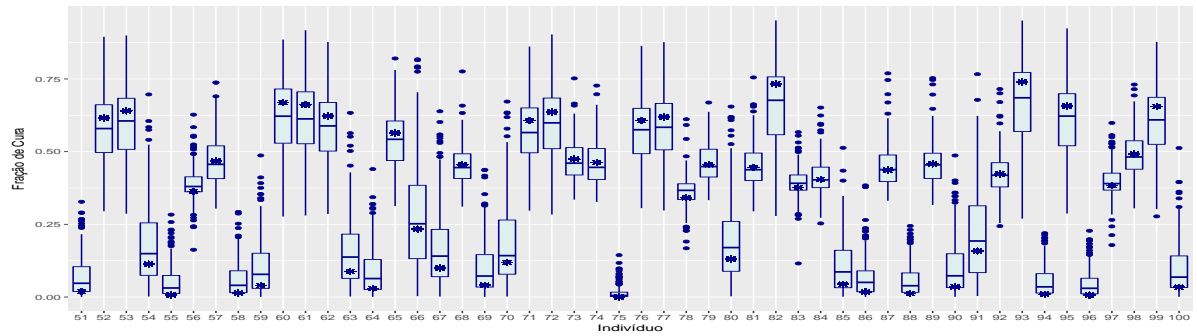


(d)

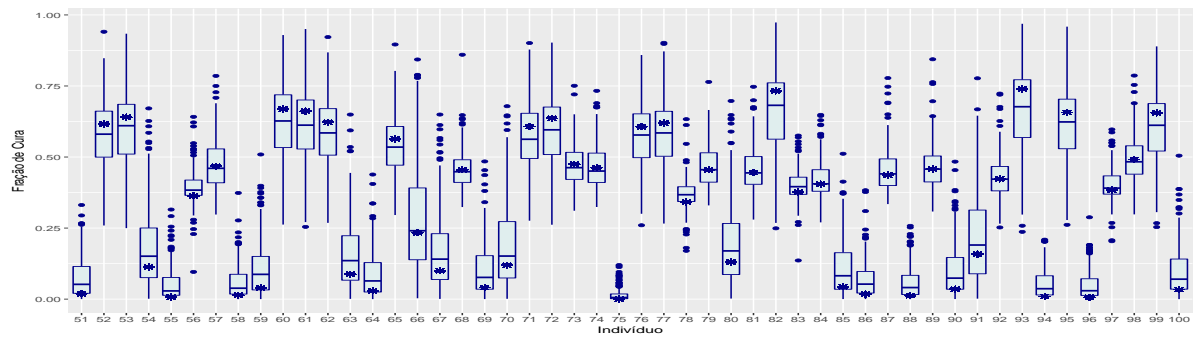
Figura G.7: *Box-plots* dos valores esperados *a posteriori* da fração de cura para cada um dos 50 primeiros indivíduos da amostra ajustando os modelos MTP (Fig.a), MBFC1 (fixando  $\sigma_\epsilon^2 = 1$ , Fig.b), MBFC2 (Fig.c) e MBFC3 (Fig.d), assumindo  $\sigma_x^2 \sim GI(102, 1010)$ . O \* representa o valor real para a fração de cura de cada indivíduo.



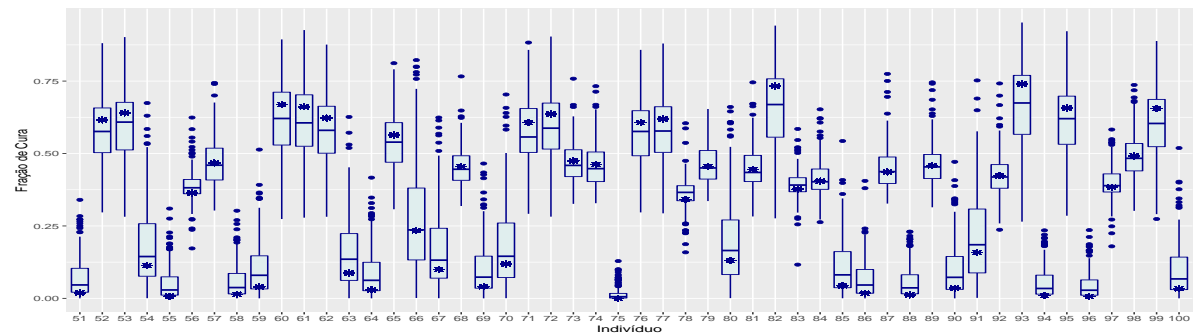
(a)



(b)

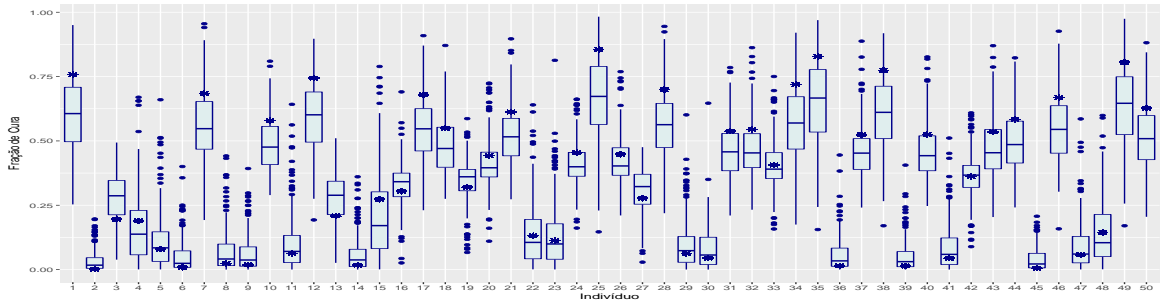


(c)

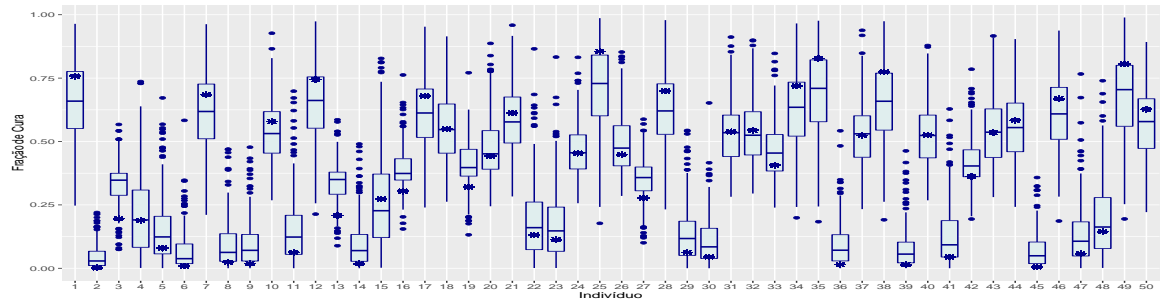


(d)

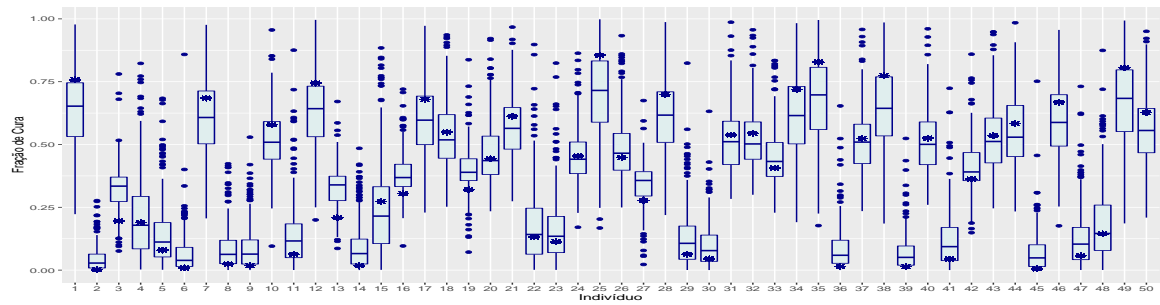
Figura G.8: *Box-plots* dos valores esperados *a posteriori* da fração de cura para cada um dos 50 últimos indivíduos da amostra ajustando os modelos MTP (Fig.a), MBFC1 (fixando  $\sigma_\epsilon^2 = 1$ , Fig.b), MBFC2 (Fig.c) e MBFC3 (Fig.d), assumindo  $\sigma_x^2 \sim GI(102, 1010)$ . O \* representa o valor real para a fração de cura de cada indivíduo.



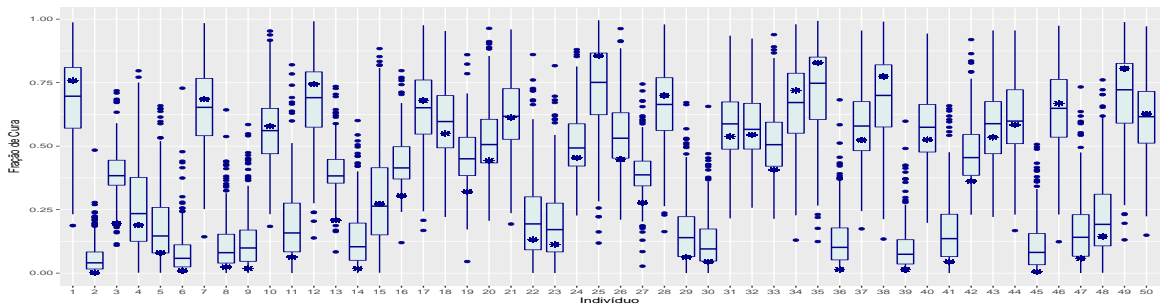
(a)



(b)

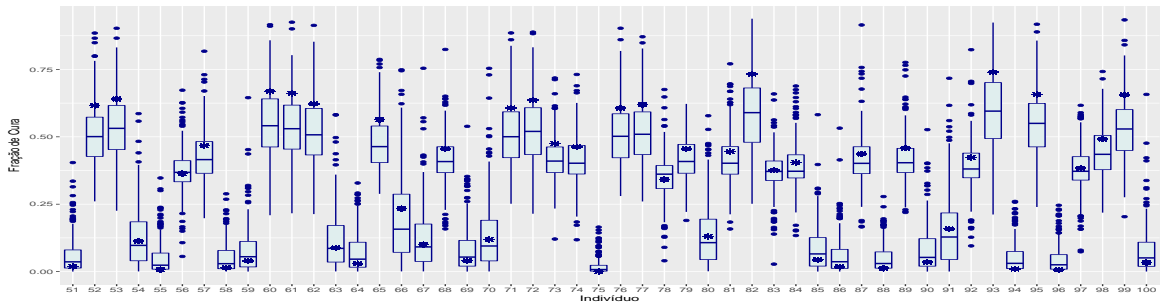


(c)

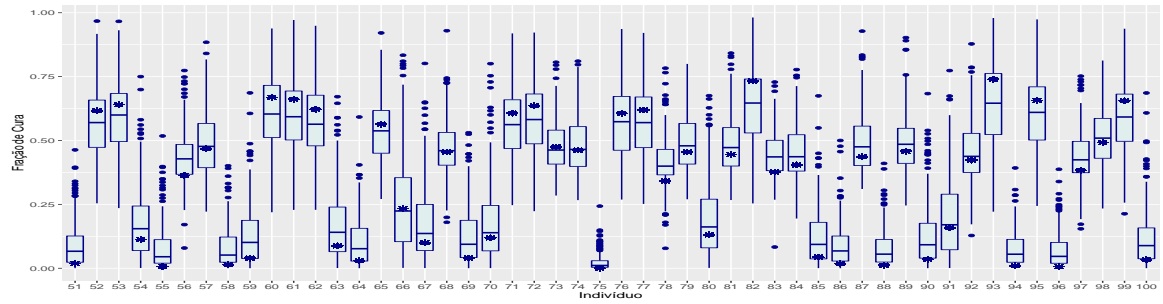


(d)

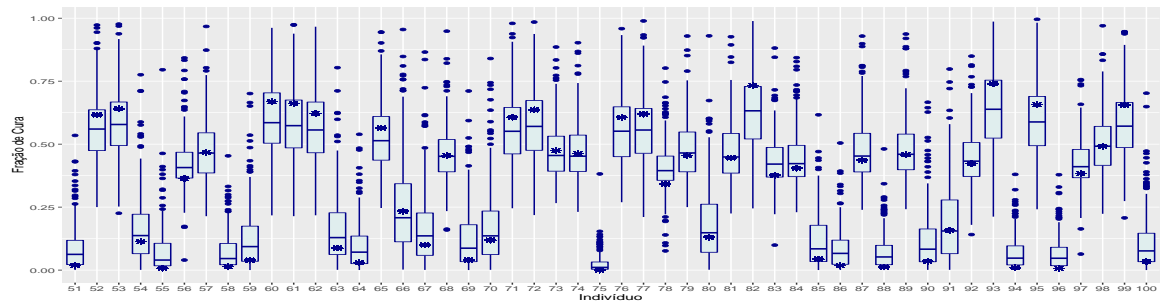
Figura G.9: *Box-plots* dos valores esperados *a posteriori* da fração de cura para cada um dos 50 primeiros indivíduos da amostra ajustando os modelos MTP (Fig.a), MBFC1 (fixando  $\sigma_\epsilon^2 = 5$ , Fig.b), MBFC2 (Fig.c) e MBFC3 (Fig.d), assumindo  $\sigma_x^2 \sim GI(102, 1010)$ . O \* representa o valor real para a fração de cura de cada indivíduo.



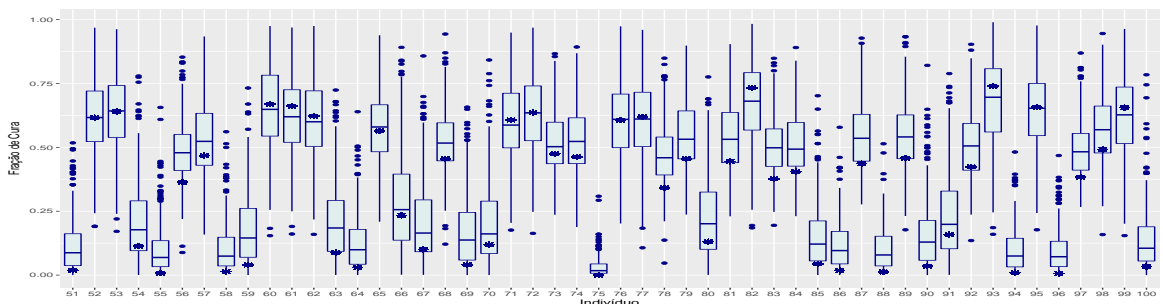
(a)



(b)



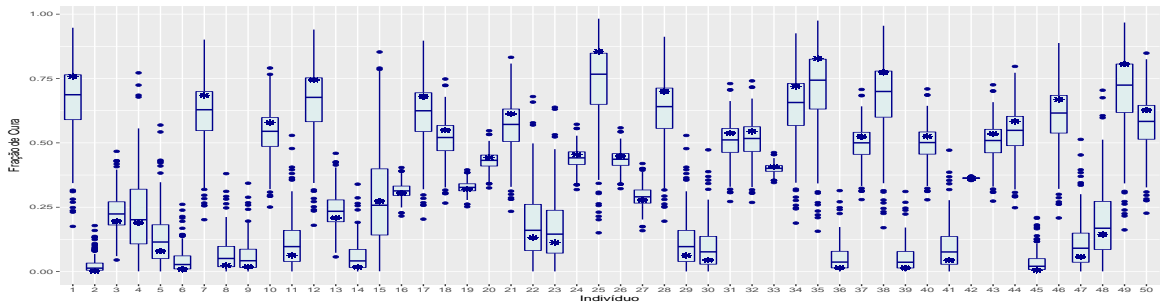
(c)



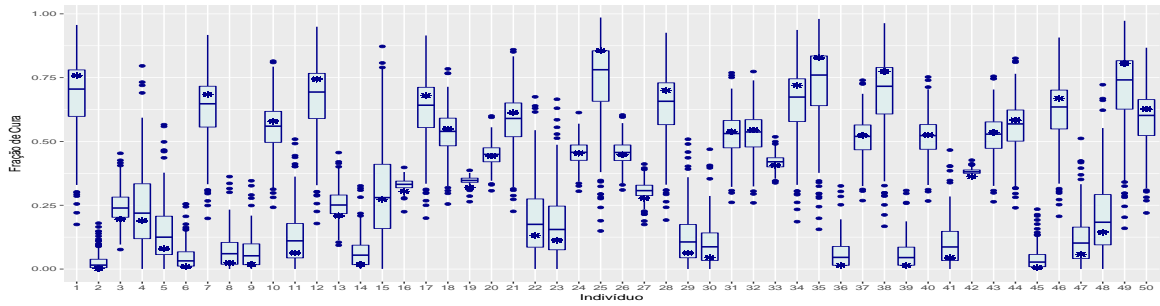
(d)

Figura G.10: *Box-plots* dos valores esperados *a posteriori* da fração de cura para cada um dos 50 últimos indivíduos da amostra ajustando os modelos MTP (Fig.a), MBFC1 (fixando  $\sigma_\epsilon^2 = 5$ , Fig.b), MBFC2 (Fig.c) e MBFC3 (Fig.d), assumindo  $\sigma_x^2 \sim GI(102, 1010)$ . O \* representa o valor real para a fração de cura de cada indivíduo.

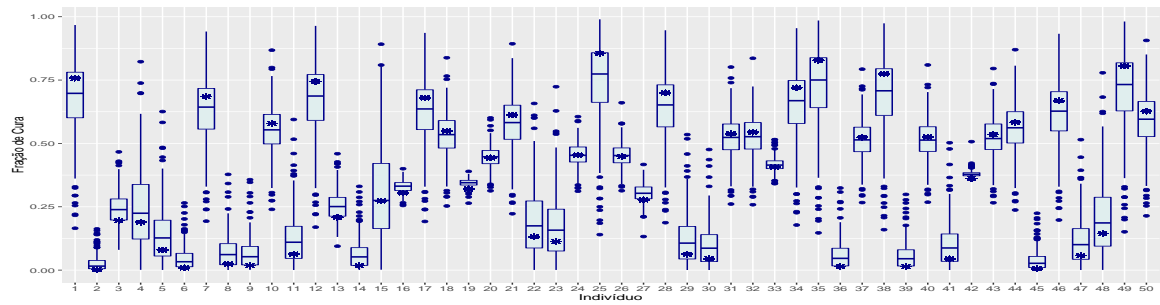




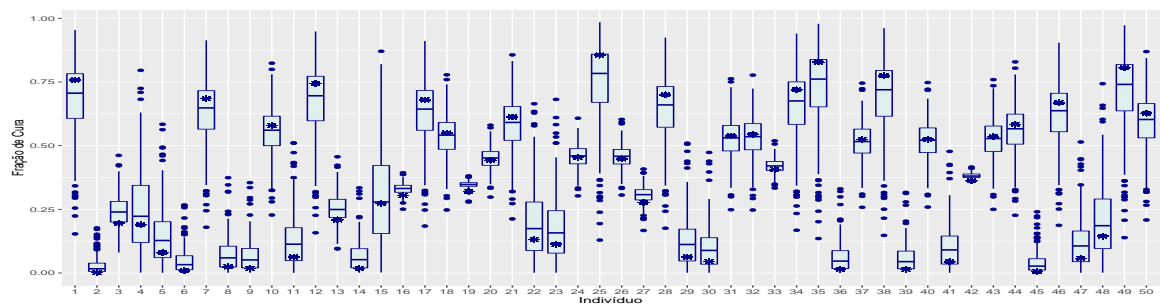
(a)



(b)

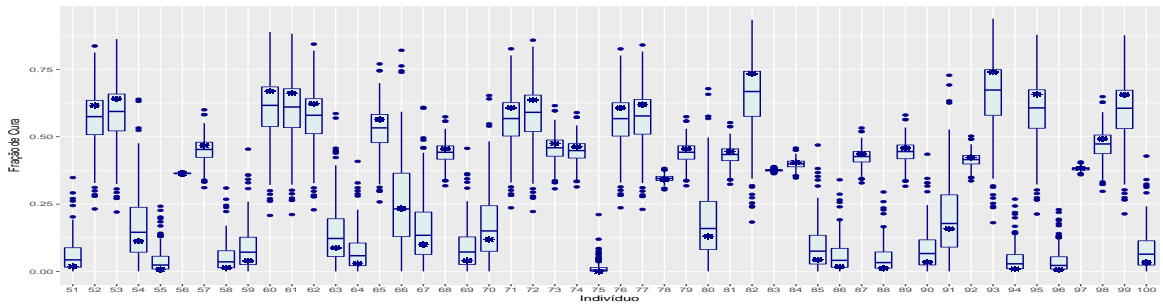


(c)

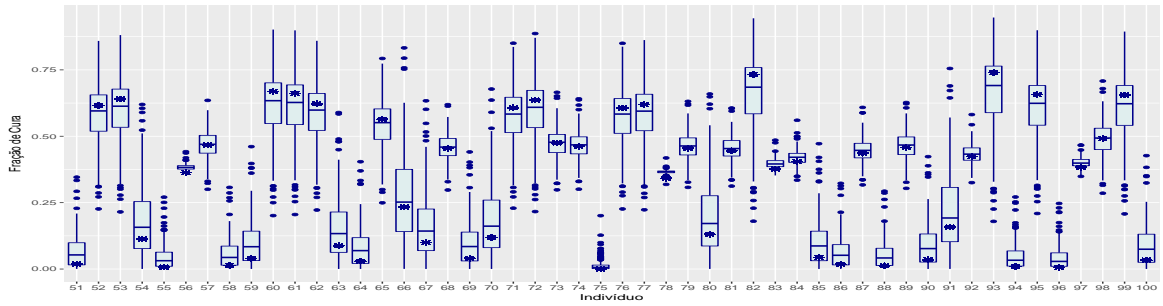


(d)

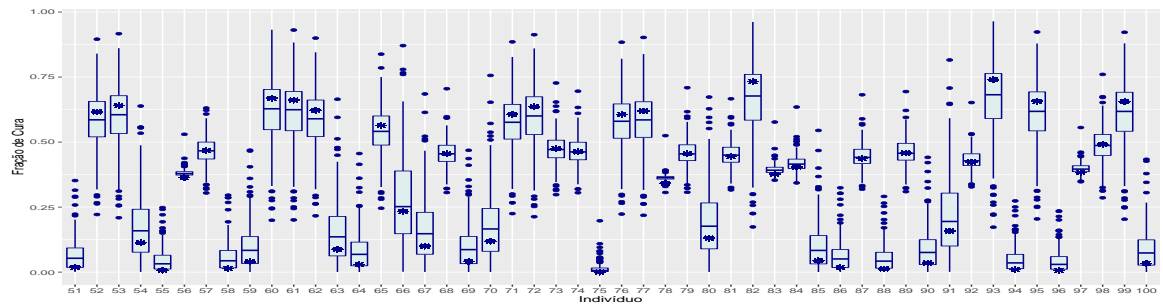
Figura G.11: *Box-plots* dos valores esperados *a posteriori* da fração de cura para cada um dos 50 primeiros indivíduos da amostra ajustando os modelos MTP (Fig.a), MBFC1 (fixando  $\sigma_\epsilon^2 = 1$ , Fig.b), MBFC2 (Fig.c) e MBFC3 (Fig.d), assumindo  $\sigma_x^2 \sim GI(102, 1010)$ . O \* representa o valor real para a fração de cura de cada indivíduo.



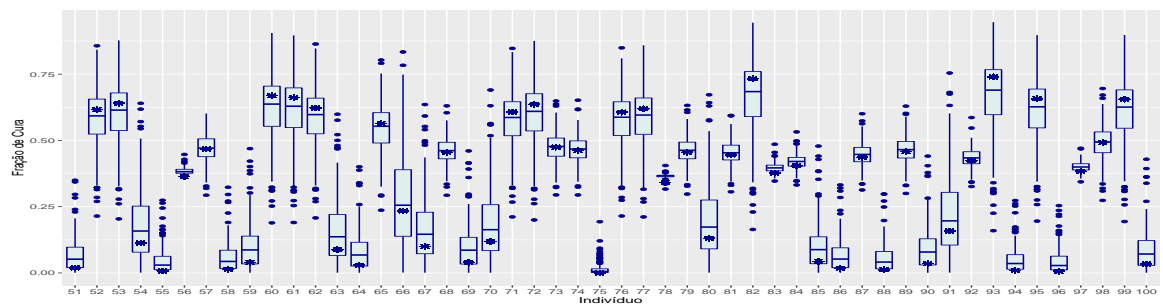
(a)



(b)



(c)



(d)

Figura G.12: *Box-plots* dos valores esperados *a posteriori* da fração de cura para cada um dos 50 últimos indivíduos da amostra ajustando os modelos MTP (Fig.a), MBFC1 (fixando  $\sigma_\epsilon^2 = 1$ , Fig.b), MBFC2 (Fig.c) e MBFC3 (Fig.d), assumindo  $\sigma_x^2 \sim GI(102, 1010)$ . O \* representa o valor real para a fração de cura de cada indivíduo.

## Apêndice H

# Gráficos Diagnósticos de Convergência

A seguir, temos alguns gráficos de diagnóstico de convergência. Na Figura H.1 são dados os gráficos da média ergódica e das autocorrelações de cada parâmetro do MBFC2. Além disso, na Figura H.2 temos gráficos da estatística  $Z$  de Geweke (Cowles e Carlin, 1996). Os dados analisados são da primeira réplica de Monte Carlo do Cenário 1 descrito na Seção 2.3. No MCMC utilizamos *lag* de tamanho 80 e *burn-in* de 100000, resultando em um total de 260000, cujas 2000 foram destinadas ao cálculo dos resumos das distribuições *a posteriori* dos parâmetros. Partimos de valores iniciais aleatórios, a saber:  $\lambda = 1, \alpha = 0.5, \sigma_\epsilon^2 = 0.5, \mu = 5, \sigma_x^2 = 10$ . Os coeficientes de regressão  $\beta_1$  e  $\beta_2$  foram gerados de uma distribuição  $U(-2, 2)$ . Da mesma maneira  $X_\epsilon$  foi gerado da  $U(0, 1)$ . O algoritmo foi implementado no software R (R Core Team, 2016) e as distribuições *a priori* eliciadas são aquelas do Estudo 2.

Analogamente, nas Figuras H.3 e H.4 são dados os gráficos da média ergódica, das autocorrelações e da estatística  $Z$  de Geweke de cada parâmetro do MBFC-NI2. Os dados analisados são referentes a réplica 16 do estudo de Monte Carlo do Cenário 1, descrito na Seção 3.3. No MCMC, consideramos a mesma configuração dada, acima, para o estudo de convergência do MBFC2. Além disso, utilizamos as mesmas distribuições *a priori*, com a inclusão de que  $U \sim \text{Gama}(\nu/2, \nu/2)$ , em que  $\nu = 2.5$ . Geramos os valores iniciais de  $U$  de uma distribuição  $U(0, 1)$ .

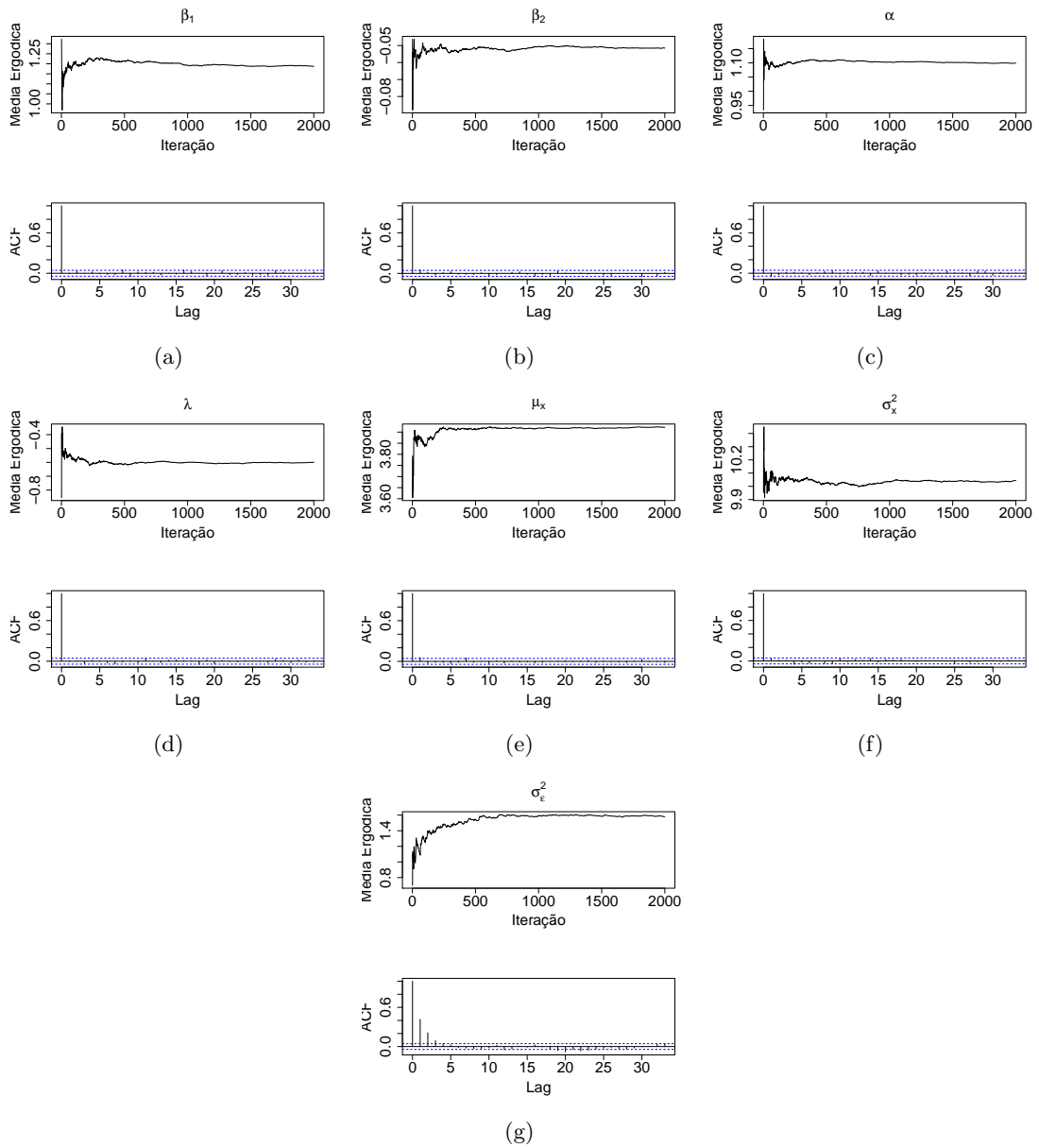


Figura H.1: Média Ergódica e autocorrelação dos parâmetros do MBFC2.

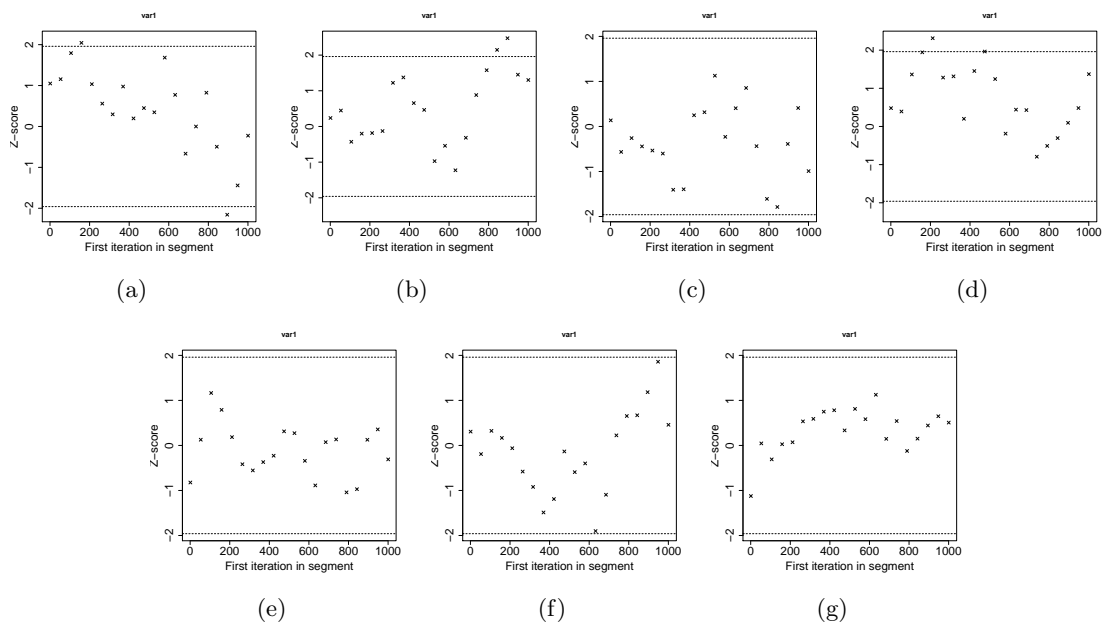


Figura H.2: Gráficos da estatística Z de Geweke para as cadeias dos parâmetros  $\beta_1$ (a),  $\beta_2$ (b),  $\alpha$ (c),  $\lambda$ (d),  $\mu_x$ (e),  $\sigma_x^2$ (f) e  $\sigma_\epsilon^2$ (g) do MBFC2

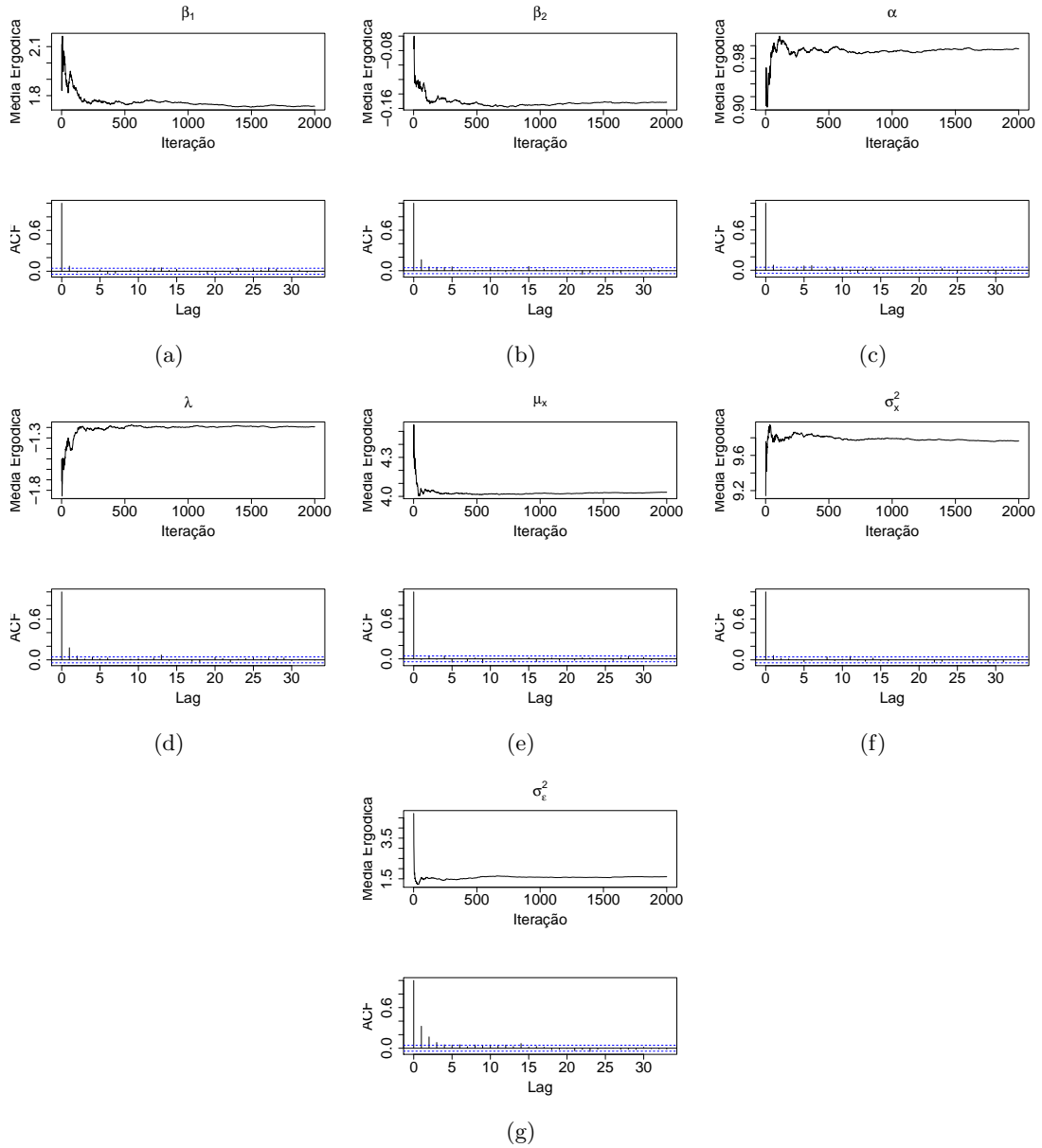


Figura H.3: Média Ergodica e autocorrelação dos parâmetros Gráficos da estatística Z de Geweke para as cadeias dos parâmetros do MBFC-NI2.

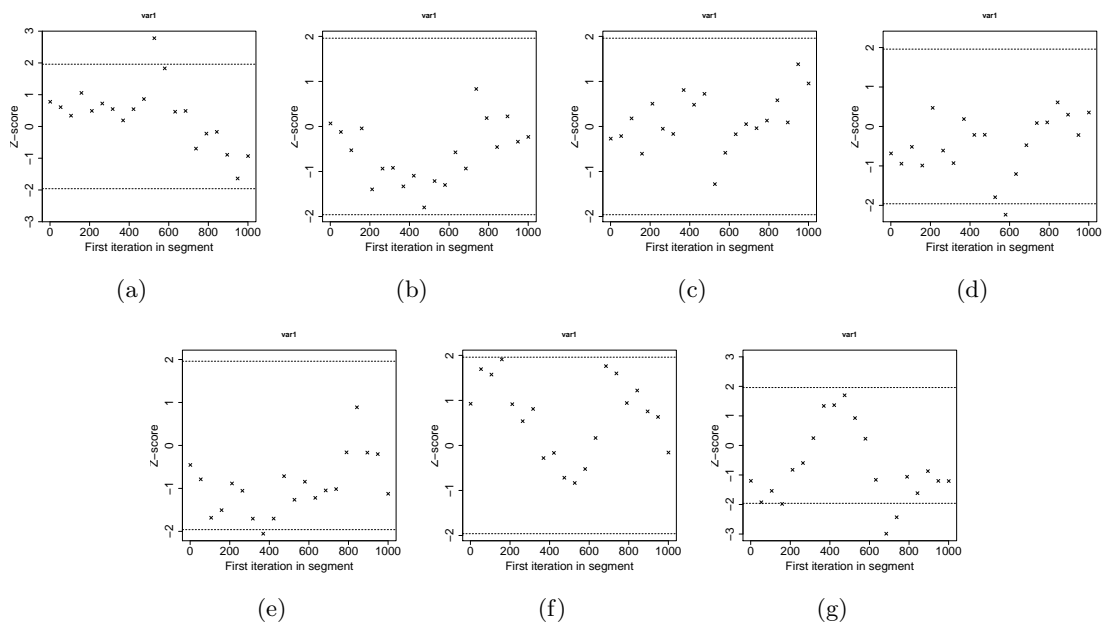


Figura H.4: Gráficos da estatística Z de Geweke para as cadeias dos parâmetros  $\beta_1$ (a),  $\beta_2$ (b),  $\alpha$ (c),  $\lambda$ (d),  $\mu_x$ (e),  $\sigma_x^2$ (f) e  $\sigma_\epsilon^2$ (g) do MBFC-NI2