

ANÁLISE DE DADOS DE SOBREVIDA PARA PACIENTES COM CÂNCER NA MAMA: UMA APLICAÇÃO ASSUMINDO MODELOS DE FRAÇÃO DE CURA SOB O ENFOQUE BAYESIANO

Tatiana Reis ICUMA¹
Isabela Panzeri Carlotti BUZATTO²
Daniel Guimarães TIEZZI²
Jorge Alberto ACHCAR¹

- RESUMO: O câncer de mama é o segundo tipo de câncer mais comum no mundo (WORLD CANCER REPORT 2014 a, b). A evolução do tratamento do câncer da mama permite uma vida mais longa para as pacientes. Neste artigo, será considerado um estudo com um conjunto de dados relacionados a um estudo realizado no Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, Ribeirão Preto, Brasil, com pacientes em estágios II e III de câncer de mama HER2 atendidos na Divisão de Câncer de mama do HCFMRP- USP no período de 2008 a 2012, passando por quimioterapia neoadjuvante associado com o medicamento Herceptin[®]. O principal objetivo deste estudo é verificar se há diferenças significativas nos tempos de sobrevida das pacientes que receberam, pelo menos, quatro ciclos de Herceptin[®] e as que receberam menos ciclos da medicação durante a meodjuvância. Como objetivo secundário, será verificado se algumas covariáveis observadas (idade, estágio da doença, tipo de cirurgia, resposta patológica completa, positividade para receptor de estrogênio ou positividade para receptor) apresentam efeitos significativos sobre os tempos de sobrevivência das pacientes (SLD: sobrevida livre da doença e ST: sobrevida total, ambas em meses). Para isso, são discutidas as vantagens de um modelo estatístico com fração de curas em termos de ajuste aos dados quando comparado com modelos tradicionais baseados na distribuição Weibull sem fração de curas sob um enfoque clássico ou Bayesiano onde é feito um estudo comparativo. Sumários a posteriori de interesse são obtidos usando métodos MCMC (*Markov Chain Monte Carlo*) existentes.
- PALAVRAS-CHAVE: Dados censurados; covariáveis; modelos de fração de curas; dados de câncer de mama; análise Bayesiana.

1 Introdução

A cada ano estima-se que cerca de um milhão e meio de novos casos de câncer de mama são diagnosticados em todo o mundo. É a forma de câncer mais comum em mulheres, compreendendo 25% de todos os tipos de câncer (WORLD CANCER

¹ Universidade de São Paulo - USP, Faculdade de Medicina de Ribeirão Preto, Departamento de Medicina Social, CEP: 14049-900, Ribeirão Preto, SP, Brasil. E-mail: tati.icuma@usp.br; achcar@fmrp.usp.br

² Universidade de São Paulo - USP, Faculdade de Medicina de Ribeirão Preto, Departamento de Ginecologia e Obstetrícia, CEP: 14049-900, Ribeirão Preto, SP, Brasil. E-mail: isabelacarlotti@yahoo.com.br; dtiezzi@fmrp.usp.br

REPORT 2014 a, b; WORLD CANCER REPORT, 2008, KELSEY, 1993). Nos Estados Unidos, é a segunda maior causa de óbitos por câncer, sendo estimado que uma em cada oito mulheres desenvolva a doença em sua vida (DE SANTIS *et al.*, 2014). Sua incidência no Brasil é de aproximadamente 56,09 casos para 100 mil mulheres (INCA, 2012; TIEZZI, 2009).

Nos estudos em oncologia, muitas vezes o interesse do pesquisador é voltado a variáveis que representam o tempo até a ocorrência de um evento, como o tempo até uma recidiva da doença ou o tempo até o óbito.

As técnicas estatísticas de análise de sobrevivência são comumente utilizadas nestes estudos, com destaque ao modelo de riscos proporcionais de Cox (COX, 1972), aos modelos paramétricos baseados na distribuição de Weibull (ver, por exemplo, LAWLESS, 1982) e testes não paramétricos de comparações entre curvas de sobrevida, como os testes de Wilcoxon e do log-rank (ver, por exemplo, LEE e WENYUWANG, 2003). Na prática, em algumas situações não se verifica o pressuposto de riscos proporcionais, como assumido no modelo de Cox e assim outros modelos devem ser considerados na análise dos dados. Outra situação comum na análise de dados de sobrevivência: parte dos indivíduos não é suscetível ao evento de interesse. Nesses casos, modelos que incluem uma fração de cura são mais adequados à estrutura dos dados.

Neste artigo, considera-se uma aplicação relacionada a um estudo realizado no Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, Ribeirão Preto, referente a 54 pacientes do sexo feminino com câncer da mama localmente avançado (Estágio II e III) com super expressão do HER-2 (HER-2 positivo) atendidas no Ambulatório de Mastologia do HCFMRP-USP no período de 2008 a 2012, submetidas à quimioterapia neoadjuvante (Tabela 1, apêndice). O principal objetivo deste estudo é verificar se existem diferenças significativas nos os tempos de sobrevida das pacientes que receberam pelo menos quatro ciclos de Herceptin[®] uma droga antes da cirurgia ou menos de quatro ciclos. O trastuzumabe (nome comercial da droga Herceptin[®]) é um anticorpo monoclonal, derivado do DNA recombinante ligado com alta afinidade ao Receptor do Fator de Crescimento Epidermal Humano 2 (HER2). O Herceptin[®] é um pó estéril, amarelado, sem conservantes e liofilizado para administração intravenosa. Segundo o órgão europeu "European Medicines Agency" o Herceptin[®] é utilizado no tratamento de vários tipos de câncer: câncer da mama em fase precoce (quando o câncer se alastrou no interior da mama ou para as glândulas debaixo do braço mas não se alastrou a outras partes do corpo) após cirurgia, quimioterapia ou radioterapia. Pode também ser usado em uma fase prévia do tratamento, em associação com quimioterapia. No caso de tumores localmente avançados (incluindo os inflamatórios) ou com mais de 2 cm de largura, o Herceptin[®] é utilizado antes da cirurgia em associação com quimioterapia e, novamente após a cirurgia, em monoterapia; câncer da mama metastático (câncer que se alastrou a outras partes do corpo). É utilizado em monoterapia em doentes que não tenham respondido a tratamentos anteriores. O Herceptin[®] é também utilizado em associação com outros medicamentos contra o câncer. O Herceptin[®] pode igualmente ser utilizado para câncer gástrico (câncer do estômago) metastático, em associação com outros medicamentos. O Herceptin[®] só pode ser utilizado após ter sido demonstrado que o câncer apresenta "superexpressão" do HER2. (http://www.ema.europa.eu/docs/pt_PT/document_library/EPAR_Summary_for_the_public/human/000278/WC500049819.pdf).

Como objetivos secundários têm-se interesse em verificar se as covariáveis observadas (Idade, Estágio da doença, Tipo de cirurgia, Resposta patológica completa,

Positividade para receptor de Estrogênio ou Positividade para receptor de Progesterona têm efeitos significativos sobre os tempos de sobrevivência das pacientes (SLD: sobrevida livre da doença e ST: sobrevida total, ambas em meses). Além disso, são discutidas as vantagens de um modelo com fração de cura em termos de ajuste aos dados quando comparado com modelos tradicionais baseados na distribuição Weibull sem fração de curas sob um enfoque clássico ou Bayesiano. Outro ponto de interesse: são feitas algumas comparações entre o enfoque clássico e o enfoque Bayesiano usando técnicas de simulação de amostras para a distribuição a posteriori de interesse na obtenção das inferências de interesse.

2 Materiais e métodos

Cada paciente foi acompanhada desde a data de entrada no estudo (entre os anos de 2008 e 2012) até a data de encerramento (01/01/2014). Todas as pacientes receberam o medicamento Herceptin[®] sendo que 39 das pacientes em estudo receberam pelo menos 4 ciclos do medicamento antes da cirurgia, enquanto que 15 receberam menos do que 4 ciclos do medicamento pré-cirurgia. Os dados desse estudo apresentam duas variáveis respostas de interesse para cada paciente: o tempo de sobrevida livre da doença (SLD) (a paciente pode apresentar recidiva ou não) e o tempo de sobrevida total (ST) (óbito por câncer de mama ou sobrevida até o último tempo de seguimento), dados em meses. São as colunas “Recid” e “Óbito” do Anexo 1. Além disso, o Anexo 1 contém as informações de censuras associadas respectivamente aos tempos livres da doença e aos tempos de sobrevida total. No anexo 1 as dicotomizações das covariáveis, como por exemplo, idade e hercept foram feitas a partir de critérios estabelecidos por pesquisadores da área de oncologia.

A distribuição de Weibull (1951) tem sido muito utilizada na análise de dados médicos (ver, por exemplo, COX e OAKES, 1984; COLOSIMO e GIOLO, 2006) e industriais (ver, por exemplo; NELSON, 2004; LAWLESS, 1982; MEEKER e ESCOBAR, 1998) dada a sua grande flexibilidade de ajuste.

Inicialmente é assumida uma análise de dados baseada em uma distribuição de Weibull com dois parâmetros desprezando a possível presença de uma fração de cura na amostra. Sua função densidade de probabilidade é dada por:

$$f(t_i) = \frac{\alpha t_i^{\alpha-1} \exp \left\{ -\left(\frac{t_i}{\lambda}\right)^\alpha \right\}}{\lambda^\alpha}, \quad (1)$$

em que, $t_i > 0$ denota os tempos de sobrevida. Os parâmetros λ e α denotam respectivamente, os parâmetros de escala e de forma para a distribuição. Diferentes valores de α levam a diferentes formas para a distribuição o que a torna muito flexível na análise de dados para tempos de sobrevivência. O principal objetivo da análise de sobrevivência é estimar a função de sobrevivência, convencionalmente denotada por S e definida por:

$$S(t^*) = P(T > t^*), \quad (2)$$

em que t^* é um tempo qualquer fixado, T é o tempo de sobrevivência, e P denota probabilidade de um evento. Assumindo a distribuição de Weibull descrita em (1), tem-se que,

$$S(t^*) = \exp\left\{-\left(\frac{t^*}{\lambda}\right)^\alpha\right\}. \quad (3)$$

A função de risco $h(t)$ ou taxa instantânea de falha, da distribuição de Weibull (ver, por exemplo, Lawless, 1982) é dada de $h(t) = f(t)/S(t)$ por:

$$h(t) = \alpha \frac{t^{\alpha-1}}{\lambda^\alpha}. \quad (4)$$

Observar que se $\alpha = 1$, temos a distribuição exponencial, isto é, a distribuição exponencial é um caso particular da distribuição de Weibull. A função de risco $h(t)$ dada por (4) é estritamente crescente para $\alpha > 1$, estritamente decrescente para $\alpha < 1$ e constante para $\alpha = 1$. Assim, observa-se uma grande flexibilidade de ajuste aos dados. A média e a variância da distribuição de Weibull com densidade dada por (1) são dadas respectivamente por:

$$\begin{aligned} \mu = E(T) &= \lambda \Gamma\left(1 + \frac{1}{\alpha}\right), \\ \sigma^2 = Var(T) &= \lambda^2 \left\{ \Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma\left[\left(1 + \frac{1}{\alpha}\right)\right]^2 \right\}, \end{aligned} \quad (5)$$

em que $\Gamma(\cdot)$ denota uma função gama, $\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt$.

2.1 Modelo de Weibull com fração de cura

De acordo com Maller e Zhou (1996), em um modelo de fração de cura assume-se que certa fração p de indivíduos na população é curada ou nunca observou o evento de interesse, logo $(1-p)$ é a fração de indivíduos não curados, portanto a função de sobrevivência nesse caso pode ser escrita considerando uma mistura, ou seja,

$$S(t) = p + (1-p)S_0(t), \quad (6)$$

em que $p \in (0,1)$ é o parâmetro de mistura (proporção de imunes) e $S_0(t)$ é a função de sobrevivência basal para a população de indivíduos não curados (indivíduos suscetíveis).

Considerando uma amostra aleatória de tempos de sobrevida (t_i, δ_i) , $i = 1, \dots, n$, a contribuição do i -ésimo indivíduo para a função de verossimilhança é dada por:

$$L_i = [f(t_i)]^{\delta_i} [S(t_i)]^{1-\delta_i}, \quad (7)$$

em que δ_i é a variável indicadora de censura, ou seja, $\delta_i = 1$ quando o tempo de sobrevivência é completo e $\delta_i = 0$ quando o tempo é censurado para o i -ésimo indivíduo.

A partir da função de sobrevivência definida em (6), é possível obter a função densidade de probabilidade, utilizando o resultado $f(t_i) = \frac{d}{dt} S(t_i)$, dada por:

$$f(t_i) = (1 - p)f_0(t_i), \quad (8)$$

em que $f_0(t_i)$ é a função densidade de probabilidade para os indivíduos suscetíveis.

Substituindo a função de densidade (8) e a função de sobrevivência (6) na função de verossimilhança (7) obtêm-se a seguinte função de verossimilhança para o modelo de mistura com fração de cura (ou longa duração):

$$L_i = [(1 - p)f_0(t_i)]^{\delta_i} [p + (1 - p)S_0(t_i)]^{1 - \delta_i}. \quad (9)$$

Portanto a função log-verossimilhança considerando todas as observações é dada por:

$$l = r \log(1 - p) + \sum_{i=1}^n \delta_i \log f_0(t_i) + \sum_{i=1}^n (1 - \delta_i) \log [p + (1 - p)S_0(t_i)], \quad (10)$$

em que, $r = \sum_{i=1}^n \delta_i$ é o número de observações não censuradas.

Uma possibilidade para a análise dos dados é assumir uma distribuição de Weibull para indivíduos suscetíveis com função de densidade de probabilidade dada por:

$$f_0(t) = \alpha \theta t^{\alpha-1} \exp[-\theta t^\alpha], \quad (11)$$

e função de sobrevivência $S_0(t) = \exp[-\theta t^\alpha]$.

Supondo o modelo de misturas (6), o logaritmo da função de verossimilhança para p , α e θ é dado por:

$$l(p, \theta, \alpha) = r \ln(1 - p) + r \ln(\alpha) + r \ln(\theta) + (\alpha - 1)v - \theta A_1(\theta) + A_2(p, \theta, \alpha), \quad (12)$$

em que $A_1(\theta) = \sum_{i=1}^n \delta_i t_i^\alpha$, $A_2(p, \theta, \alpha) = \sum_{i=1}^n (1 - \delta_i) \ln [p + (1 - p)e^{-\theta t_i^\alpha}]$, $r = \sum_{i=1}^n \delta_i$ e $v = \sum_{i=1}^n \delta_i \ln(t_i)$

Na presença de um vetor de covariáveis $\mathbf{x} = (x_1, \dots, x_k)$ que afeta os parâmetros p e θ , mas não afeta o parâmetro de forma α , assume-se o seguinte modelo de regressão:

$$\theta_i = \beta_0 \exp(\beta_1 x_{1i} + \dots + \beta_k x_{ki}) \ln\left(\frac{p_i}{1 - p_i}\right) = \gamma_0 + \gamma_1 x_{1i} + \dots + \gamma_k x_{ki}. \quad (13)$$

2.2 Modelo de regressão de Weibull sem fração de cura

Assumir um modelo de regressão de Weibull para os tempos de sobrevida livre da doença ou tempos totais de sobrevida definido por:

$$\log(t_i) = \beta_0 + \beta_1 idade_i + \beta_2 hercep_i + \beta_3 estágio_i + \beta_4 cirur_i + \beta_5 pCR_i + \beta_6 estrog_i + \beta_7 proggest_i + \sigma^* \varepsilon_i, \quad (14)$$

em que t_i denotam os tempos de sobrevida livre da doença ou tempos totais, $i = 1, \dots, n$; $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6$ e β_7 são parâmetros de regressão.

O parâmetro σ^* está relacionado com o parâmetro de forma da distribuição de Weibull (1) pela relação $\sigma^* = 1/\alpha$. O termo ε_i em (14) é uma quantidade aleatória com

distribuição de valor extremo (ver NELSON, 2004 ou LAWLESS, 1982) também definida como distribuição de valor extremo de tipo I (mínimo) ou distribuição de Gumbel (ver, GUMBEL, 1954) com função densidade de probabilidade dada por:

$$f(\varepsilon) = \exp(\varepsilon - \exp(\varepsilon)), -\infty < \varepsilon < \infty. \quad (15)$$

Também observar que o parâmetro de escala λ definido em (1) está relacionado com as covariáveis a partir da relação,

$$\lambda_i = \exp(\beta_0 + \beta_1 idade_i + \beta_2 hercep_i + \beta_3 estágio_i + \beta_4 cirur_i + \beta_5 pCR_i + \beta_6 estrog_i + \beta_7 progest_i). \quad (16)$$

Isto é, o modelo de regressão definido por (14) define um modelo de regressão no parâmetro de escala (ver, por exemplo, COLOSIMO e GIOLO, 2006) assumindo mesmo parâmetro de forma.

Para o modelo de regressão (14), usualmente estima-se os parâmetros de regressão $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6$ e β_7 e o parâmetro σ^* usando métodos de máxima verossimilhança (ver por exemplo, MOOD *et al.*, 1974). Estimadores de máxima verossimilhança para os parâmetros $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7$ e σ^* são obtidos maximizando-se a função de verossimilhança, $L(\theta) = \prod f(\varepsilon_i)$, em que $f(\varepsilon_i) = \exp[\varepsilon_i - \exp(\varepsilon_i)]$, $i=1, \dots, n$, $\theta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \sigma^*)$ e,

$$\sigma^* \varepsilon_i = \log(t_i) - [\beta_0 + \beta_1 idade_i + \beta_2 hercep_i + \beta_3 estágio_i + \beta_4 cirur_i + \beta_5 pCR_i + \beta_6 estrog_i + \beta_7 progest_i]. \quad (17)$$

Na obtenção de inferências estatísticas para os modelos propostos são considerados métodos clássicos baseados em estimação de máxima verossimilhança para os parâmetros do modelo e resultados assintóticos na obtenção de intervalos de confiança e testes de hipóteses e métodos Bayesianos.

Os métodos Bayesianos são uma boa alternativa para os modelos de regressão na presença de fração de cura e dados censurados (ver, por exemplo, PAULINO *et al.*, 2003), pois os métodos clássicos baseados na distribuição normal assintótica para os estimadores de máxima verossimilhança podem não ser muito precisos com tamanhos amostrais não muito grandes e grande proporção de dados censurados. Além disso, consideram-se métodos de simulação MCMC (Monte Carlo em Cadeias de Markov) para a obtenção de sumários a posteriori de interesse (ver, por exemplo, GELFAND e SMITH, 1990; CASELA e GEORGE, 1992; CHIB e GREENBERG, 1995) com o uso do software OpenBugs (SPIEGELHALTER *et al.*, 2003).

3 Resultados e discussões

Na Figura 1 são apresentados os gráficos dos estimadores não paramétricos de Kaplan-Meier (1958) para as funções de sobrevivência dos tempos livres da doença e tempos de sobrevida total obtidos usando o *software* R (R CORE TEAM, 2015). Por esses gráficos observa-se que a função de sobrevivência decresce ao longo do período de seguimento, mas este decréscimo torna-se mais lento até tornar-se constante. A presença deste “platô” à direita das curvas de sobrevida sugere que em uma parte dos indivíduos amostrados não haverá recidiva da doença, enquanto uma parte dos mesmos indivíduos não deverá morrer devido ao câncer de mama (não necessariamente os mesmos

indivíduos). Este comportamento da curva de Kaplan-Meier sugere a presença de uma fração de cura, ou seja, uma proporção de indivíduos em que o evento de interesse não ocorrerá. Dessa forma conclui-se que o modelo de fração de cura é o modelo mais apropriado para a análise dos dados.

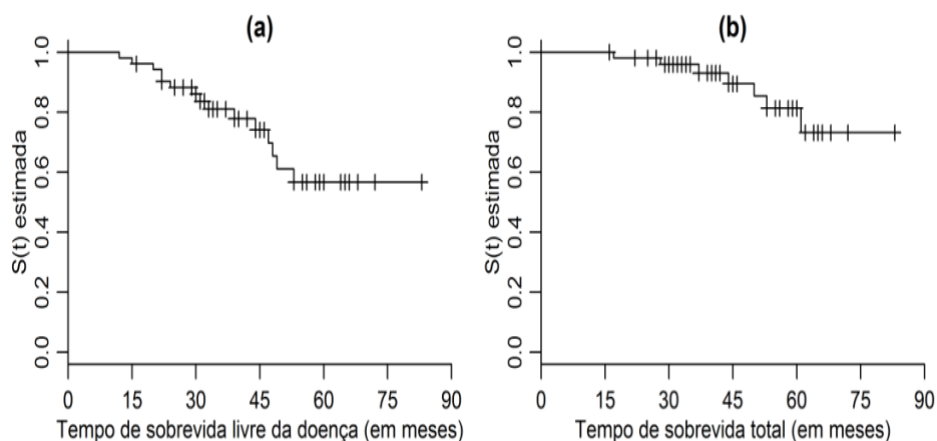


Figura 1 - Estimadores de Kaplan-Meier: (a) Tempos de sobrevida livre da doença, (b) Tempos de sobrevida total.

3.1 Análise clássica

Inicialmente apresentam-se os resultados de uma análise clássica assumindo a distribuição de Weibull na presença ou não de covariáveis e sem fração de cura.

3.1.1 Distribuição de Weibull sem presença de covariáveis

Assumindo a distribuição de Weibull (1), os estimadores para os parâmetros λ e α podem ser obtidos usando o método de máxima verossimilhança, maximizando a função de verossimilhança $L(\lambda, \alpha)$ na presença de dados censurados usando métodos numéricos.

Usando o *software* Minitab[®], versão 16, encontram-se os estimadores de máxima verossimilhança (EMV) e algumas estatísticas de interesse dadas na Tabela 1. Dos resultados da Tabela 1, observa-se que o tempo médio estimado de sobrevida livre da doença é 73,05 meses e o tempo mediano estimado de sobrevida livre da doença é 68,26 meses. Da mesma forma observa-se que o tempo médio estimado de sobrevida total é 96,61 meses e o tempo mediano estimado de sobrevida total é 94,34 meses.

Nos tempos de sobrevida livre da doença, 37 pacientes não apresentaram o evento de interesse, isto é, aproximadamente 71% da amostra e nos tempos totais de sobrevida, 45 pacientes não apresentaram o evento de interesse, isto é, aproximadamente 86% da amostra. Esses resultados sugerem um modelo com a fração de cura.

A partir de gráficos de probabilidade para a distribuição de Weibull obtidos pelo *software* Minitab[®], versão 16 (não apresentados por economia de espaço) observa-se bom

ajuste (pontos próximos da reta) da distribuição de Weibull aos dados (tempos livres da doença e tempos totais).

Tabela 1 - EMV para os parâmetros da distribuição de Weibull. Tempos de sobrevida livre da doença

Estimadores de Máxima Verossimilhança				
Parâmetro	Estimativa	Erro Padrão	Intervalo de Confiança (95%)	
			Limite Inferior	Limite Superior
α	1,95	0,4189	1,2795	2,9709
λ	82,38	14,0980	58,9076	115,2130
Características da Distribuição				
Média	73,05	12,7516	51,8838	102,850
Desvio Padrão	39,07	13,0767	20,2750	75,2910
Mediana	68,26	10,1620	50,9898	91,3920
Primeiro Quartil	43,48	5,8942	33,3372	56,7150
Terceiro Quartil	97,41	19,0841	66,3476	143,0080
Amplitude Interquartil	53,93	16,9049	29,1709	99,6860
Tempos de sobrevida totais				
Parâmetro	Estimativa	Erro Padrão	Intervalo de Confiança (95%)	
			Limite Inferior	Limite Superior
α	2,57	0,78450	1,4119	4,6741
λ	108,81	26,8791	67,0495	176,579
Características da Distribuição				
Média	96,61	23,2029	60,3397	154,69
Desvio Padrão	40,35	19,6078	15,5641	104,587
Mediana	94,34	20,1528	62,0696	143,395
Primeiro Quartil	67,00	10,4177	49,395	90,8666
Terceiro Quartil	123,56	34,4959	71,4906	213,561
Amplitude Interquartil	56,57	27,2469	22,0071	145,4

3.1.2 Modelo de regressão de Weibull na presença de covariáveis

Assumir o modelo de regressão de Weibull introduzido na seção 2 para os tempos livres da doença e tempos totais de sobrevida.

Os estimadores de máxima verossimilhança para os parâmetros do modelo são apresentados na Tabela 2 (uso do *software* Minitab[®], versão 16).

Dos resultados da Tabela 2, observa-se que todas as covariáveis não mostram efeitos significativos para os tempos livres da doença e para os tempos totais, pois todos os intervalos de confiança para os parâmetros de regressão correspondentes contém o valor zero. Além disso, nenhum valor-p é inferior do que 0,05 (nível de significância usual), o que comprova a não significância de todas as covariáveis.

É importante salientar que para os dados da Tabela 1, há uma grande proporção de dados censurados (71%), o que pode dificultar a descoberta de possíveis covariáveis significativas afetando os tempos de sobrevida livres da doença ou tempos totais.

Tabela 2 - EMV para os parâmetros de regressão de Weibull. Tempos de sobrevida livre da doença

Parâmetro	Estimativa	Erro Padrão	Z	P	Intervalo de Confiança (95%)	
					Limite Inferior	Limite Superior
β_0	4,60	1,6835	2,73	0,006	1,2992	7,8985
β_1	0,33	0,3030	1,09	0,277	-0,2651	0,9238
β_2	0,08	0,4250	0,19	0,846	-0,7508	0,9154
β_3	-0,20	0,5770	-0,35	0,725	-1,3341	0,9277
β_4	-0,18	0,3207	-0,55	0,581	-0,8056	0,4518
β_5	0,23	0,3063	0,74	0,462	-0,3752	0,8255
β_6	0,14	0,3774	0,36	0,715	-0,6021	0,8775
β_7	0,19	0,4567	0,41	0,679	-0,7065	1,0839
α	1,94	0,4233			1,2663	2,9766

Tempos totais de sobrevida

Parâmetro	Estimativa	Erro Padrão	Z	P	Intervalo de Confiança (95%)	
					Limite Inferior	Limite Superior
β_0	25,96	3511,26	0,01	0,994	-6855,98	6907,89
β_1	-0,16	0,32	-0,51	0,612	-0,79	0,47
β_2	-0,30	0,43	-0,71	0,478	-1,14	0,53
β_3	-5,03	1136,11	0,00	0,996	-2231,78	2221,71
β_4	-6,10	843,87	-0,01	0,994	-1660,05	1647,85
β_5	0,27	0,36	0,74	0,458	-0,44	0,98
β_6	-0,24	0,34	-0,69	0,489	-0,91	0,43
β_7	0,75	0,47	1,57	0,115	-0,18	1,68
α	2,82	0,90			1,51	5,27

(Z: estatística normal padronizada; P: valor-p)

Na Figura 2, tem-se os gráficos dos resíduos padronizados para o modelo de regressão de Weibull. Desses gráficos, observa-se um ajuste razoável do modelo de regressão de Weibull aos dados (tempos livres da doença e tempos totais).

Na análise de sobrevivência com dados médicos é comum essa dificuldade, pois estes dados geralmente possuem uma grande proporção de censuras e diversas covariáveis de interesse. Por isso a necessidade cada vez maior de modelos e técnicas estatísticas mais poderosas para analisar dados com estas características.

Outra possibilidade na procura de possíveis melhores inferências (maior precisão dos estimadores dos parâmetros do modelo, maior sensibilidade de detecção de possíveis efeitos das covariáveis na resposta de interesse) é reanalisar os dados sob o enfoque Bayesiano (ver, por exemplo, PAULINO *et al.*, 2003).

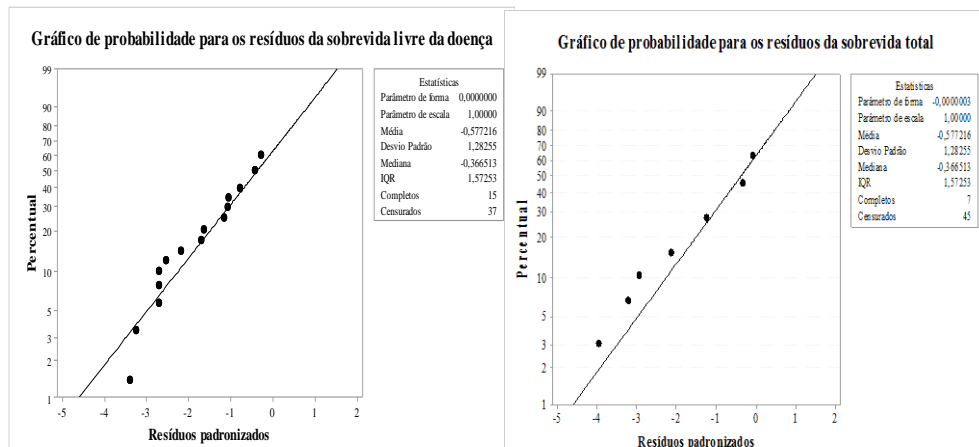


Figura 2 - Gráficos dos resíduos padronizados para a regressão de Weibull (tempos livres da doença e tempos totais).

3.2 Análise Bayesiana

3.2.1 Distribuição de Weibull sem fração de curas e sem a presença de covariáveis

Assumir uma distribuição de Weibull sem a presença de covariáveis considerando métodos MCMC (Monte Carlo em Cadeias de Markov) (ver, por exemplo, GELFAND e SMITH, 1990; CASELA e GEORGE, 1992; CHIB e GREENBERG, 1995) com o uso do *software* OpenBugs versão 3.2.2 (SPIEGELHALTER *et al.*, 2003). No *software* OpenBugs a densidade da distribuição de Weibull é dada em uma forma reparametrizada de (1), isto é:

$$f(t) = \alpha \theta t^{\alpha-1} * \exp\{-\theta t^\alpha\}, \text{ onde } \theta = 1/\lambda\alpha. \quad (18)$$

Para esta análise Bayesiana considera-se distribuições a priori gama $G(0,1; 0,1)$ não informativas para α e θ , onde $G(a, b)$ denota uma distribuição gama com média igual à a/b e variância igual à a/b^2 . Na simulação de amostras da distribuição a posteriori para α e θ , considera-se uma amostra de aquecimento “*burn-in sample*” de tamanho 1.000 para eliminar o efeito do valor inicial no processo iterativo; após essa amostra de aquecimento, foram geradas outras 600.000 amostras tomando amostras de 100 em 100, totalizando uma amostra final de tamanho 6.000 que será utilizada para obter as quantidades a posteriori de interesse.

Na Tabela 3, são apresentados os sumários a posteriori de interesse. Observar que nessa análise duas observações foram eliminadas por conter dados faltantes (ver Tabela 1). Utilizando distribuições a priori não informativas o tempo médio estimado de sobrevida livre da doença é de 81,71 meses (resultado próximo do obtido usando inferência clássica). O tempo médio estimado de sobrevida total é de 139,10 meses.

Tabela 3 - Sumários a posteriori de interesse. Tempos de sobrevida livre da doença.

	Média	Desvio Padrão	Intervalo de Credibilidade (95%)	
			Limite Inferior	Limite Superior
α	1,84	0,3717	1,12	2,64
λ	90,84	20,2915	65,79	144,10
Média	81,71	20,2226	58,27	139,62
θ	0,0007	0,0014	0,00001	0,00418
Tempos totais de sobrevida				
α	2,12	0,607	1,117	3,48
λ	153,20	86,34	84,56	344,1
Média	139,10	92,61	75,5	327,3
θ	0,000256	0,000688	0,000000136	0,001977

3.2.2 Modelo de Weibull com fração de cura sem a presença de covariáveis

Para uma análise Bayesiana dos tempos de sobrevida livre da doença ou tempos totais de sobrevida assume-se agora uma distribuição de Weibull sem a presença de covariáveis e na presença de fração de cura. São consideradas as seguintes distribuições a priori: $\alpha \sim \text{Gama}(1,1)$, $p \sim \text{Beta}(70,30)$ e $\theta \sim U(0,300)$, onde $U(a,b)$ denota uma distribuição uniforme no intervalo (a,b) e $\text{Beta}(a,b)$ denota uma distribuição beta com média igual à $a/(a+b)$ e variância igual $ab/[(a+b)^2(a+b+1)]$. Observar que os hiperparâmetros da distribuição beta dados por $a=70$ e $b=30$ foram escolhidos levando a uma priori informativa (uso de métodos Bayesianos empíricos, ver, por exemplo, Carlin e Louis, 2002) para p , com média igual à 0,70 (um valor próximo da proporção observada de dados censurados, interpretados como pacientes imunes ou curados).

Na simulação de amostras da distribuição a posteriori de interesse, foi considerada uma amostra de aquecimento “*burn-in sample*” de tamanho 1.000; a seguir foram geradas outras 600.000 amostras tomando amostras de 100 em 100 totalizando uma amostra final de tamanho 6.000 que será utilizada para obter as quantidades a posteriori de interesse. Na Tabela 4, são apresentados os sumários a posteriori de interesse.

Tabela 4 - Sumários a posteriori de interesse modelo com fração de cura sem covariáveis. Tempos de sobrevida livre da doença

	Média	Desvio Padrão	Intervalo de Credib (95%)	
			Limite Inferior	Limite Superior
α	2,52	0,5685	1,508	3,771
λ	42,94	6,5655	32,97	58,51
p	0,67	0,0409	0,5825	0,7411
Tempos totais de sobrevida				
α	2,52	0,850	1,095	4,359
λ	60,34	19,650	39,850	109,300
p	0,83	0,035	0,760	0,896

Sendo assim, as funções de sobrevivência sob uma abordagem Bayesiana considerando os modelos de Weibull sem e com fração de cura (ver estimadores Bayesianos nas Tabelas 3 e 4) são dadas, respectivamente (tempos livres da doença) por:

- Weibull sem fração de cura: $S(t) = \exp \left[-\left(\frac{t}{90,84}\right)^{1,84} \right]$
- Weibull com fração cura: $S(t) = 0,67 + (1 - 0,67)\exp \left[-\left(\frac{t}{42,94}\right)^{2,52} \right]$

Da mesma forma temos para os tempos totais de sobrevida,

- Weibull sem fração de cura: $S(t) = \exp \left[-\left(\frac{t}{153,2}\right)^{2,12} \right]$
- Weibull com fração cura: $S(t) = 0,83 + (1 - 0,83)\exp \left[-\left(\frac{t}{60,34}\right)^{2,51} \right]$

Na Figura 3 são apresentados os gráficos das funções de sobrevivência estimadas considerando os estimadores Kaplan-Meier, Weibull Bayesiano sem e com presença de fração de cura. Observa-se excelente ajuste do modelo de Weibull na presença de fração de cura, especialmente para tempos de sobrevida grandes. Isso não ocorre com o modelo de Weibull sem a presença de fração de cura.

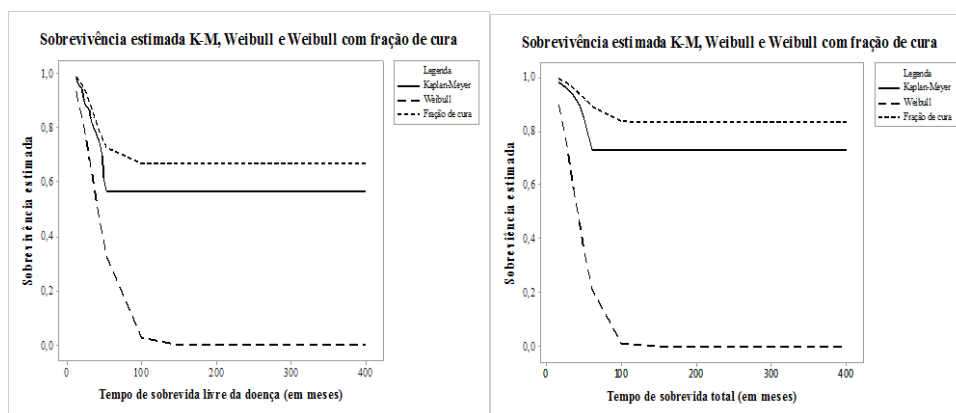


Figura 3 - Gráficos da função de sobrevivência estimada – Kaplan e Meier, Weibull Bayesiano sem e com fração de cura (tempos de sobrevida livre da doença e tempos totais de sobrevida).

3.2.3 Modelo de Weibull na presença de covariáveis e sem fração de cura para o tempo de sobrevida livre da doença

Considerar agora uma análise Bayesiana dos dados (SLD) assumindo uma distribuição de Weibull com densidade $f(t) = \alpha \theta t^{\alpha-1} \exp\{-\theta t^\alpha\}$, na presença de covariáveis e o modelo de regressão dado por:

$$\theta_i = \exp(\beta_0 + \beta_1 idade_i + \beta_2 hercep_i + \beta_3 estágio_i + \beta_4 cirur_i + \beta_5 pCR_i + \beta_6 estrog_i + \beta_7 proggest_i). \quad (19)$$

Assumindo distribuições a priori não-informativas normais $N(0,1)$ para todos os parâmetros de regressão β_r , $r = 0, 1, 2, \dots, 7$; uma priori Gama(1,1) para o parâmetro de forma α e usando o *software* OpenBugs, versão 3.2.2 (*burn-in sample* =1.000 e 6.000

amostras finais tomadas de 100 em 100), tem-se na Tabela 5 os sumários a posteriori de interesse.

Tabela 5 - Sumários a posteriori de interesse - Tempos de sobrevida livre da doença - Modelo de regressão sem fração de cura

	Média	Desvio Padrão	Intervalo de Credibilidade (95%)	
			Limite Inferior	Limite Superior
α	1,26	0,2569	0,7973	1,8050
β_0	-1,40	0,9062	-3,1800	0,3691
β_1	-0,62	0,4736	-1,5390	0,3062
β_2	-0,86	0,6658	-2,2810	0,3722
β_3	-1,01	0,4422	-1,8710	-0,1220
β_4	0,18	0,4889	-0,7601	1,1610
β_5	-0,61	0,4891	-1,5890	0,3473
β_6	-0,34	0,5584	-1,4510	0,7302
β_7	-0,49	0,6271	-1,7700	0,7180

Dos resultados da Tabela 5, a covariável estágio tem efeito significativo (intervalo de credibilidade para o parâmetro de regressão β_3 não inclui o valor zero) no tempo de sobrevida livre da doença. Observa-se que o modelo Bayesiano é mais sensível para detectar covariáveis significativas mesmo assumindo distribuições a priori não informativas para os parâmetros do modelo, pois o modelo de regressão sob o enfoque clássico mostrou não significância para todas as covariáveis (ver Tabela 3).

3.2.4 Modelo de Weibull com fração de cura e com covariáveis afetando o parâmetro de escala da distribuição Weibull para o tempo de sobrevida livre da doença

Considerar agora uma análise Bayesiana dos dados assumindo uma distribuição de Weibull na presença de covariáveis, fração de cura e o modelo de regressão afetando o parâmetro de escala da distribuição Weibull θ dado por (19).

Assumindo distribuições a priori não-informativas normais $N(0,1)$ para todos os parâmetros de regressão β_r , $r = 0, 1, 2, \dots, 7$; $\theta \sim \text{Gama}(1,1)$, $p \sim \text{Beta}(70,30)$ e usando o *software* OpenBugs, versão 3.2.2, (burn-in sample =1.000 e 6.000 amostras finais tomadas de 100 em 100), tem-se na Tabela 6 os sumários a posteriori de interesse.

Dos resultados da Tabela 6, também se observa que a covariável estágio tem um efeito significativo (intervalo de credibilidade para o parâmetro de regressão β_3 correspondente não inclui o valor zero).

Considerar uma análise Bayesiana dos tempos de sobrevida livre da doença assumindo uma distribuição de Weibull na presença de covariáveis, fração de cura e os modelos de regressão afetando o parâmetro de escala da distribuição Weibull dado por (19) e o modelo para a fração de cura p , dado por:

$$\begin{aligned} \text{logito}(p_i) = & \gamma_0 + \gamma_1 \text{idade}_i + \gamma_2 \text{hercep}_i + \gamma_3 \text{estágio}_i + \gamma_4 \text{cirur}_i \\ & + \alpha_5 pCR_i + \gamma_6 \text{estrog}_i + \gamma_7 \text{progest}_i \end{aligned} \quad (20)$$

Tabela 6 - Sumários a posteriori de interesse – Tempo de sobrevida livre da doença (Modelo de regressão na presença de fração de curas afetando o parâmetro de escala)

	Média	Desvio Padrão	Intervalo de Credibilidade (95%)	
			Limite Inferior	Limite Superior
β_0	0,93	0,8956	-0,8246	2,6940
β_1	-0,08	0,3585	-0,7525	0,6769
β_2	0,40	0,5216	-0,5865	1,5180
β_3	0,79	0,3697	0,0622	1,5090
β_4	-0,32	0,4163	-1,1710	0,5006
β_5	0,54	0,3811	-0,1921	1,3150
β_6	0,48	0,4443	-0,3829	1,3940
β_7	0,23	0,4823	-0,6975	1,2290
θ	1,94	0,5117	1,0960	3,0770
ρ	0,66	0,0422	0,5743	0,7381

3.2.5 Modelo de Weibull com fração de cura com covariáveis afetando o parâmetro de escala da distribuição Weibull e a probabilidade de cura para o tempo de sobrevida livre da doença

Assumindo distribuições a priori não-informativas normais $N(0,1)$ para todos os parâmetros de regressão $\beta_r, \gamma_r; r = 0, 1, 2, \dots, 7$ e $\theta \sim \text{Gama}(1,1)$ e usando o software OpenBugs (*burn-in sample* = 1.000 e 6.000 amostras finais tomadas de 100 em 100), temos na Tabela 7, os sumários a posteriori de interesse.

Tabela 7 - Sumários a posteriori de interesse – Tempo de sobrevida livre da doença - Modelos de regressão afetando parâmetro de escala da distribuição Weibull e a fração de cura

	Média	Desvio Padrão	Intervalo de Credibilidade (95%)	
			Limite Inferior	Limite Superior
γ_0	0,05	0,9623	-1,8350	1,8970
γ_1	0,22	0,9528	-1,7280	2,0480
γ_2	-0,12	0,9283	-1,9680	1,6770
γ_3	-0,61	0,6212	-1,9360	0,4957
γ_4	-0,06	0,9323	-1,9330	1,7260
γ_5	0,00	0,9300	-1,9020	1,7820
γ_6	-0,15	0,9509	-1,9850	1,7190
γ_7	0,04	0,9526	-1,8600	1,8500
β_0	1,11	0,9333	-0,7087	2,9540
β_1	0,32	0,4848	-0,5951	1,3140
β_2	0,52	0,5931	-0,5760	1,7580
β_3	0,77	0,3861	0,0097	1,5350
β_4	-0,24	0,4839	-1,1760	0,7276
β_5	0,52	0,4599	-0,3784	1,4320
β_6	0,39	0,5120	-0,6171	1,3970
β_7	0,32	0,5483	-0,7373	1,4100
θ	1,54	0,4563	0,8375	2,6090

Dos resultados da Tabela 7, observa-se que a covariável estágio tem efeito significativo (intervalo de credibilidade do parâmetro de regressão β_3 não inclui o valor zero).

3.2.6 Modelo de Weibull na presença de covariáveis sem fração de cura para o tempo de sobrevida total (ST)

Considerar uma análise Bayesiana dos tempos de sobrevida total assumindo uma distribuição de Weibull (ver seção 2) na presença de covariáveis e o modelo de regressão dado por (19).

Assumindo distribuições a priori não-informativas normais $N(0,1)$ para todos os parâmetros de regressão β_r , $r = 0, 1, 2, \dots, 7$; uma priori Gama(1,1) para o parâmetro de forma α e usando o software OpenBugs, versão 3.2.2 (*burn-in sample* = 1.000 e 6.000 amostras finais tomadas de 100 em 100), tem-se na Tabela 8 os sumários a posteriori de interesse.

Dos resultados da Tabela 8, observa-se que nenhuma covariável mostrou efeito significativo, pois os intervalos de credibilidade 95% para todos os parâmetros de regressão incluem o valor zero. O modelo de regressão Weibull (19) não detectou nenhuma covariável que afete o tempo de sobrevida total das pacientes.

Tabela 8 - Sumários a posteriori de interesse - Tempo de sobrevida total - Modelo de regressão

	Média	Desvio Padrão	Intervalo de Credibilidade (95%)	
			Limite Inferior	Limite Superior
α	1,53	0,4528	0,7313	2,5030
β_0	-5,88	2,3360	-10,3100	-1,2930
β_1	-0,07	0,6505	-1,3240	1,2430
β_2	-0,43	0,7706	-2,0050	1,0360
β_3	-0,70	0,7007	-1,9590	0,7388
β_4	1,00	0,6971	-0,3052	2,4290
β_5	-0,57	0,6555	-1,8560	0,6831
β_6	0,02	0,6662	-1,2950	1,3150
β_7	-0,72	0,7373	-2,2180	0,6894

3.2.7 Modelo Weibull com fração de cura e com covariáveis afetando o parâmetro de escala da distribuição Weibull para o tempo de sobrevida total (ST)

Considerar agora uma análise Bayesiana dos tempos de sobrevida total assumindo uma distribuição de Weibull na presença de covariáveis, fração de curas e o modelo de regressão afetando o parâmetro de escala da distribuição Weibull (19).

Assumindo distribuições a priori não-informativas normais $N(0,1)$ para todos os parâmetros de regressão β_r , $r = 0, 1, 2, \dots, 7$; $\gamma \sim \text{Gama}(1,1)$, $p \sim \text{Beta}(86,14)$ e usando o software OpenBugs, versão 3.2.2 (*burn-in sample*=1.000 e 6.000 amostras finais tomadas de 100 em 100), são apresentados na Tabela 9 os sumários a posteriori de interesse.

Dos resultados da Tabela 9, observa-se que todas as covariáveis não mostram efeitos significativos, pois os intervalos de credibilidade 95% para todos os parâmetros de regressão incluem o valor zero.

Tabela 9 - Sumários a posteriori de interesse - tempo de sobrevida total - Modelo de regressão na presença de fração de cura afetando o parâmetro de escala

	Média	Desvio Padrão	Intervalo de Credibilidade (95%)	
			Limite Inferior	Limite Superior
β_0	0,48	0,9666	-1,3980	2,3230
β_1	0,05	0,5617	-1,0700	1,1780
β_2	0,51	0,6053	-0,6691	1,7740
β_3	0,97	0,4926	-0,0120	1,9210
β_4	-0,33	0,9468	-2,0480	1,7200
β_5	0,10	0,5777	-0,9722	1,3220
β_6	0,62	0,5320	-0,4528	1,7130
β_7	0,19	0,6149	-0,9521	1,5190
θ	2,25	0,9291	0,8718	4,5130
p	0,84	0,0330	0,7677	0,8965

3.2.8 Modelo Weibull com fração de cura e com covariáveis afetando o parâmetro de escala da distribuição Weibull e a probabilidade de cura para o tempo de sobrevida total (ST)

Considerar agora uma análise Bayesiana dos tempos de sobrevida total assumindo uma distribuição de Weibull na presença de covariáveis, fração de curas e os modelos de regressão afetando parâmetro de escala da distribuição Weibull e a fração de cura, dados por (19 e 20).

Assumindo distribuições a priori não-informativas normais $N(0,1)$ para todos os parâmetros de regressão $\gamma_r, \beta_r; r = 0, 1, 2, \dots, 7$ e $\theta \sim \text{Gama}(1,1)$ e usando o software OpenBugs (*burn-in sample* = 1.000 e 6.000 amostras finais tomadas de 100 em 100), são apresentados na Tabela 10, os sumários a posteriori de interesse.

Dos resultados da Tabela 10, observa-se que a covariável estágio tem efeito significativo (intervalo de credibilidade para o parâmetro de regressão β_3 não inclui o valor zero).

Tabela 10 - Sumários a posteriori de interesse - tempo de sobrevida total - modelos de regressão afetando parâmetro de escala da distribuição Weibull e a fração de cura

	Média	Desvio Padrão	Intervalo de Credibilidade (95%)	
			Limite Inferior	Limite Superior
γ_0	0,23	0,9730	-1,6980	2,1140
γ_1	-0,03	0,9333	-1,8910	1,7760
γ_2	0,03	0,8987	-1,6890	1,7340
γ_3	-0,17	0,7217	-1,7820	1,0820
γ_4	-0,50	0,9942	-2,4110	1,4790
γ_5	0,32	0,9498	-1,6030	2,0760
γ_6	-0,36	0,9344	-2,1460	1,5000
γ_7	0,19	0,9811	-1,7850	2,0760
β_0	0,87	0,9749	-1,0440	2,7060
β_1	0,11	0,6219	-1,0720	1,3650
β_2	0,54	0,6835	-0,7076	1,9840
β_3	1,13	0,4875	0,1646	2,0770
β_4	-0,65	0,8564	-2,1670	1,2760
β_5	0,37	0,6786	-0,9687	1,7370
β_6	0,33	0,6548	-0,9529	1,6490
β_7	0,50	0,6867	-0,8443	1,9080
θ	1,80	0,7722	0,7507	3,7380

4 Resultados e discussão

O uso de modelos de fração de cura pode ser de grande interesse na análise de dados de sobrevida para pacientes com câncer de mama, dado que novas terapias levam a tempos de sobrevida maiores ou mesmo a cura de muitas pacientes, sendo assim em uma grande parcela da amostra não ocorre o evento de interesse. Dessa forma, modelos tradicionais sem a presença de fração de cura sob o enfoque clássico ou Bayesiano podem não ser apropriados. Esses resultados são de grande interesse médico.

O uso de métodos Bayesianos tem crescido de forma substancial na análise de dados médicos de sobrevivência na presença de censuras e covariáveis, especialmente usando métodos MCMC de simulação de amostras da distribuição a posteriori de interesse para determinar os sumários a posteriori de interesse com o uso do *software* OpenBugs, na obtenção de inferências muito precisas que independem do tamanho amostral. É importante salientar que os métodos clássicos usualmente baseados na distribuição normal assintótica para os estimadores de máxima verossimilhança podem não ser muito precisos com tamanhos amostrais não muito grandes e grande proporção de dados censurados, especialmente com modelos na presença de fração de cura.

A análise dos dois tempos de sobrevida disponíveis deste estudo não evidenciou diferença significativa nos tempos de sobrevida das pacientes que receberam pelo menos quatro ciclos de Herceptin[®] antes da cirurgia ou menos de quatro ciclos.

Os modelos Weibull sob enfoque clássico na presença de covariáveis obtiveram desempenho inferior em termos de sensibilidade na determinação de fatores significativos nos tempos de sobrevida das pacientes com câncer na mama em relação aos modelos Bayesianos apresentados neste artigo. Os modelos Bayesianos se ajustam melhor aos

dados desse estudo, possibilitando a detecção de fatores relevantes que afetam os tempos de sobrevida das pacientes, tanto até a recidiva da doença quanto até o óbito.

Os modelos Bayesianos com fração de cura demonstraram serem bastante sensíveis para detectar covariáveis significativas e podem ser muito úteis na pesquisa médica de novas terapias para pacientes com câncer na mama, pois as probabilidades de cura podem ser estimadas para cada paciente (valores fixados das covariáveis), possibilitando ao médico, uma classificação de pacientes com maiores ou menores chances de cura.

Em ambos os tempos de sobrevida (SLD e ST) a covariável estágio se mostrou significativa um efeito esperado pelos pesquisadores médicos e não detectado pelo modelo clássico; gráficos com as curvas de Kaplan-Meier estimadas encontram-se na Figura 4.

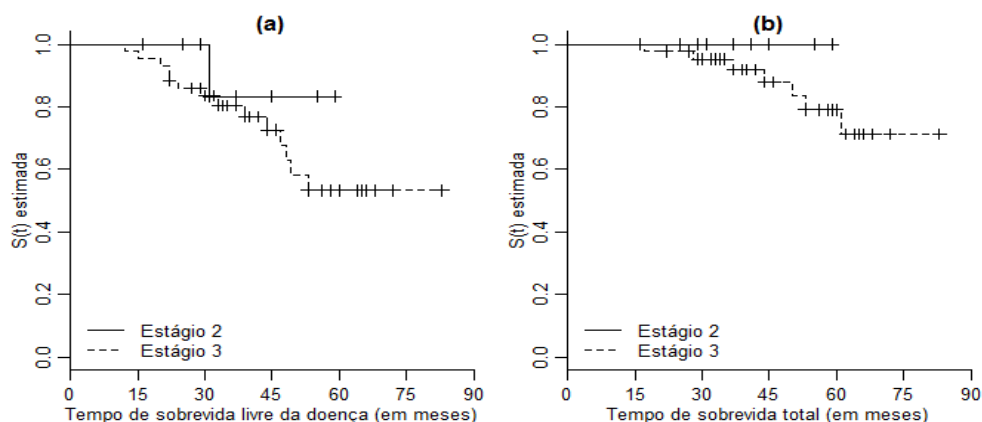


Figura 4. Estimadores de Kaplan-Meier da covariável estágio: (a) Tempo de sobrevida livre da doença, (b) Tempo de sobrevida total.

Nos tempos de sobrevida livre da doença, estágio foi significativa nos modelos Weibull (Bayesiano) sem a presença de fração de cura, no modelo Weibull com fração de cura afetando o parâmetro de escala e no modelo Weibull Bayesiano com fração de cura afetando o parâmetro de escala e a probabilidade de cura. No modelo Weibull (Bayesiano) sem a presença de fração de cura, o tempo mediano de sobrevida livre da doença de pacientes do estágio 2 é 2,77 vezes maior do que o tempo mediano de pacientes do estágio 3, os outros modelos estimaram este tempo mediano como 2,20 vezes e 2,15 vezes respectivamente. Dos resultados da análise do tempo de sobrevida total, o modelo Weibull Bayesiano com fração de cura afetando o parâmetro de escala e a probabilidade de cura traz evidências de que a covariável estágio afeta o tempo de sobrevida total das pacientes, estimando que o tempo mediano de sobrevida de pacientes do estágio 2 é 3,09 vezes maior do que pacientes do estágio 3. Todas as outras covariáveis do estudo não apresentaram evidências de influencia nos tempos de sobrevida das pacientes.

É interessante observar que outras distribuições paramétricas para dados de sobrevivência poderiam ser usados na análise desses dados, mas a distribuição Weibull apresentou um bom ajuste para os dados da Tabela 1. Nessa direção podemos mencionar várias distribuições exponenciais generalizadas (ver, por exemplo, MUDHOLKAR e SRIVASTAVA, 1993; GUPTA e KUNDU, 1999, 2007; RAQAB e AHSANULLAH,

2001; RAQAB, 2002; SARHAN, 2007; CARRASCO *et al.*, 2008; ACHCAR e BOLETA, 2009).

Outra possibilidade que poderia ser uma sugestão de estudo seria a análise dos tempos de sobrevida livre da doença e tempos de sobrevida total usando distribuições de sobrevivência bivariadas (ver por exemplo, ARNOLD e STRAUSS, 1988; MARSHALL e OLKIN, 1967; SARKAR, 1987; BLOCK e BASU, 1974).

Para finalizar a discussão dos resultados obtidos é importante reenfatar que o estudo realizado nesse artigo para dados de câncer da mama, poderia ser de grande interesse para pesquisadores médicos que trabalham com dados de câncer, pois o uso de modelos de fração de cura sob o enfoque Bayesiano pode ser uma alternativa promissora dada a eficácia de novas terapias recentes que tem prolongado a sobrevida dos pacientes e muitas vezes levando à cura. Esses modelos são amplamente pesquisados e propostos pela comunidade estatística, mas muitas vezes de pouco conhecimento pela comunidade médica que muitas vezes fica restrita ao uso de técnicas não-paramétricas de comparação de curvas de sobrevida e/ou o uso do modelo de riscos proporcionais de Cox.

Agradecimentos

O presente trabalho foi realizado com apoio do CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico – Brasil. Os autores agradecem às excelentes sugestões apresentadas pelos revisores que levaram à grande melhoria na apresentação do manuscrito.

ICUMA, T. R.; BUZZATTO, I. P. C.; TIEZZI, D. G.; ACHCAR, J. A. Analysis of survival data for breast cancer patients: a case study assuming cure fraction models under the Bayesian approach. *Rev. Bras. Biom.*, Lavras, v.34, n.4, p.669-692, 2016.

- **ABSTRACT:** *Breast cancer is the second most common cancer in the world (WORLD CANCER REPORT 2014 a, b). The evolution of the treatment of breast cancer allows a longer life for the patients. In this article, it will be considered a study with a data set related to a study conducted at the Hospital das Clinicas, Faculty of Medicine, University of São Paulo, Ribeirão Preto, Brazil, with patients in stages II and III of HER2 breast cancer treated at the HCFMRP USP breast cancer division from 2008 to 2012, undergoing neoadjuvant chemotherapy combined with Herceptin drug. The aim of this study is to verify that there are significant differences in survival times of patients who received at least four cycles of Herceptin. As a secondary objective, it will be checked if some observed covariates (age, stage of disease, type of surgery, pathologic complete response, positive for estrogen receptor or positive for receptor have significant effects on the patients survival times (SLD: free survival disease and ST: overall survival, both in months) for this, we discuss the advantages of a statistical cure fraction model in terms of fit to the data when compared with traditional models based on the Weibull distribution without cure fraction under an classical or a Bayesian approach where a comparative study is done. Posterior summaries of interest are obtained using existing MCMC methods (MCMC).*
- **KEYWORDS:** *Censored data; covariates; fraction models cures; breast cancer data; Bayesian analysis.*

Referências

- ACHCAR, J. A.; BOLETA, J. Distribuição exponencial generalizada: uso de métodos Bayesianos. *Revista Brasileira de Biometria*, v.27, n.4, p.644-658, 2009.
- ARNOLD, B.C.; STRAUSS, D. Bivariate distributions with exponential conditionals, *Journal of the American Statistical Association*, v. 83, pp. 522-527, 1988.
- BLOCK, H. W.; BASU, A. P. A continuous bivariate exponential extension, *Journal of the American Statistical Association*, p.1031-1037, 1974.
- CARLIN, B. P.; LOUIS, T. A. *Bayes and Empirical Bayes Methods for Data Analysis*. London: Chapman Hall, 2002.
- CARRASCO, J. M.; ORTEGA, E. M. M.; CORDEIRO, G. M. A generalized modified Weibull distribution for lifetime modeling. *Computational Statistics and Data Analysis*, v.53, p.450-462, 2008.
- CASELLA G.; GEORGE, E. I. Explaining the Gibbs sampler. *The American Statistician*, v.46, p.167-174, 1992.
- CHIB, S.; GREENBERG, E. Understanding the Metropolis-Hastings algorithm. *The American Statistician*, v.49, p.327-335, 1995.
- COLOSIMO, E. A.; GIOLO, S. R. *Análise de Sobrevivência Aplicada*. São Paulo: Edgard Blucher Ltda., 2006, 205 p.
- COX, D. R. Regression models and life tables. *Journal of the Royal Statistical Society, B*, v.34, p.187-220, 1972.
- COX, D. R.; OAKES, D. *Analysis of Survival Data*. London: Chapman & Hall, , 1984, 198p.
- DeSANTIS, C; M. A, J.; BRYAN, L. J. A. Breast cancer statistics. *Cancer Journal Clinics*, v.64, n.1, p.52-62, 2014.
- GELFAND, A. E.; SMITH, A. F. M. Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, v.85, p.398-409, 1990.
- GUMBEL, E. J. Statistical theory of extreme values and some practical applications, *Applied Mathematics*, series 33, U.S. Department of Commerce, National Bureau of Standards, 1954.
- GUPTA, R. D.; KUNDU, D. Generalized exponential distributions. *Australian and New Zealand Journal of Statistics*, v.41, p.173-188, 1999.
- GUPTA, R. D.; KUNDU, D. Generalized exponential distribution: existing results and some recent developments. *Journal of Statistical Planning and Inference*, doi:10.1016/j.jspi.2007.03.030, 2007.
- INCA - Instituto Nacional de Câncer. *Incidência de Câncer no Brasil*, 2012. Disponível em: <<http://www.inca.gov.br/estimativa/2014/>>. Último acesso: 01 Ago. 2015.
- KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, v.53 , n.282, p.457-481, 1958.

- KELSEY, J. L. Breast cancer epidemiology: summary and future directions. *Epidemiological Review*, v.15, n.1, p.256-63, 1993.
- LAWLESS, J. F. *Statistical Models and Methods for Lifetime Data*. New York: John Wiley, 1982, 580 p.
- LEE, E. T, WENYUWANG, J. *Statistical methods for survival data analysis*. 3.ed. New Jersey: John Wiley & Sons, 2003, 535 p.
- MALLER, R. A.; ZHOU, X. *Survival analysis with long-term survivors*. Chichester: John Wiley & Sons, 1996, 753 p.
- MARSHALL, A. W.; OLKIN, I. A generalized bivariate exponential distribution, *Journal of Applied Probability*, p.291-302, 1967.
- MEEKER, W. Q.; ESCOBAR, L. A. *Statistical methods for reliability data*. New York: Wiley & Sons, 1998, 712 p.
- MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. *Introduction to the Theory of Statistics*. New York: McGraw-Hill, 1974, 577 p.
- MUDHOLKAR, G. S.; SRIVASTAVA, D. K. Exponentiated Weibull family for analyzing bathtub failure-rate data, *IEEE Transactions on Reliability*, v.42 , n.2, p. 299–302, 1993.
- NELSON, W. *Applied life data analysis*. Oxford: Wiley - Blackwell, 2004, 662 p.
- PAULINO, C. D.; TURKMAN, M. A. A.; MURTEIRA, B. *Estatística Bayesiana*. Lisboa: Fundação Calouste Gulbenkian, 2003, 446 p.
- R CORE TEAM. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing, 2015. Disponível em: <<http://www.r-project.org>>. Acesso em: 10 dez. 2015.
- RAQAB, M. Z. Inferences for generalized exponential distribution based on record statistics. *Journal of Statistical Planning and Inference*, v.104, p.339-350, 2002.
- RAQAB, M. Z.; AHSANULLAH, M. Estimation of the location and scale parameters of generalized exponential distribution based on order statistics. *Journal of Statistical Computation and Simulation*, v.69, p.109-124, 2001.
- SARHAN, A. M. Analysis of incomplete, censored data in competing risks model with generalized exponential distributions. *IEEE Transactions on Reliability*, v.56, p.132-138, 2007.
- SARKAR, S. K. A continuous bivariate exponential distribution. *Journal of the American Statistical Association*, n.82, p.667-675, 1987.
- SPIEGELHALTER, D. J.; THOMAS, A.; BEST, N. G.; GILKS, W. R. *WinBUGS User Manual (version 1.4)*. Cambridge: MRC Biostatistics Unit, 2003.
- TIEZZI, D. G. Epidemiologia do câncer de mama. *Revista Brasileira de Ginecologia e Obstetrícia*, v.31, n.5, p.13-215, 2009.
- WEIBULL, W. A Statistical distribution function of wide applicability. *Journal of Applied Mechanics*, p.292–297, 1951.
- WORLD CANCER REPORT - a. World Health Organization, Chapter 5.2, 2014.

WORLD CANCER REPORT - b. International Agency for Research on Cancer, World Health Organization, 2014

WORLD CANCER REPORT. *Cancer statistics often exclude non-melanoma skin cancers such as basal-cell carcinoma, which are common but rarely fatal.* International Agency for Research on Cancer, 2008.

Recebido em 03.02.2016

Aprovado após revisão em 03.06.2016

Anexo 1 - Dados de 54 pacientes com câncer de mama.

Ident	Idade	Hercep	Est	Cirur	pCR	Estr	Proge	Recid	SLD	Óbito	ST
2	1	1	3	1	1	0	0	0	60	0	60
3	1	2	3	0	0	0	0	0	42	0	42
4	0	2	3	0	1	0	0	0	30	0	30
7	1	2	3	0	0	1	1	0	30	0	30
8	0	1	2	0	1	1	1	0	45	0	45
9	1	1	3	1	1	1	1	0	83	0	83
10	1	2	3	1	0	1	1	0	56	0	56
11	1	1	3	1	1	1	1	0	53	0	53
13	1	2	3	1	1	0	0	0	32	0	32
14	0	1	3	0	0	1	1	1	30	0	58
16	0	1	3	1	0	0	0	0	72	0	72
18	1	1	3	1	1	1	1	0	30	0	30
19	1	1	3	0	0	0	0	0	65	0	65
20	1	1	3	0	1	0	0	0	68	0	68
21	1	1	3	1	1	0	0	1	20	0	60
22	1	1	3	0	0	1	0	0	64	0	64
23	1	2	3	0	0	1	1	0	40	0	40
25	1	1	3	1	0	0	1	0	33	0	33
26	1	2	3	1	1	0	0	1	39	1	44
27	1	2	3	1	0	1	1	0	37	0	37
29	1	1	2	1	1	0	0	0	37	0	37
30	1	1	3	1	0	1	0	0	27	0	27
31	1	2	3	0	0	0	0	1	24	0	34
32	0	1	3	1	0	0	0	1	22	0	59
33	1	1	3	1	0	0	0	0	34	0	34
35	1	1	3	1	1	0	0	0	34	0	34
37	1	2	3	0	1	1	0	0	58	0	58
39	1	2	2	1	1	1	1	0	29	0	29
40	1	1	3	0	1	1	0	0	46	0	46
41	0	1	3	1	0	0	0	0	32	0	32
42	1	1	3	1	0	0	0	0	66	0	66
43	1	2	3	1	1	0	0	0	44	0	44
48	0	2	3	1	1	1	1	0	22	0	22
49	0	1	3	1	0	1	0	1	53	1	61
50	1	1	2	0	1	0	0	0	55	0	55
51	1	1	3	1	1	0	0	0	65	0	65
52	0	1	3	1	0	1	1	1	44	1	50
53	1	1	3	1	0	1	0	1	15	1	53
54	1	1	3	1	0	0	0	1	33	1	37
55	1	1	3	1	0	0	0	1	12	1	17
57	1	1	2	0	0	1	1	0	59	0	59
58	1	1	3	1	0	1	0	0	39	0	39
60	1	1	3	1	0	1	1	0	29	0	29
61	1	2	2	1	1	1	1	0	31	0	31
63	1	1	3	1	0	0	0	0	35	0	35
65	1	1	3	1	1	0	0	1	22	1	28
66	1	1	2	*	*	0	0	*	8	1	8
68	1	1	3	1	0	0	0	*	22	0	22
70	0	1	3	0	0	0	0	1	47	0	62
71	0	1	2	1	0	0	0	0	25	0	25
72	0	1	3	1	1	0	0	1	48	0	58
73	1	1	3	1	0	1	1	1	49	0	83
74	1	1	2	0	1	0	0	1	31	0	41
75	1	2	2	0	1	1	1	0	16	0	16

Foram utilizadas as seguintes codificações nas variáveis do anexo 1: Ident: Identificação da Paciente, Idade: Idade da Paciente (0: ≤ 40 anos; 1: > 40 anos), Hercep: Uso do medicamento Herceptin[®] (1: ≥ 4 ciclos; 2: < 4 ciclos), Est: Estágio da doença (2 ou 3), Cirur: Tipo de cirurgia realizada na Paciente (1: radical; 0: conservadora), pCR: Resposta Patológica Completa (1: Sim; 0: Não), Estr: Receptor de Estrogênio (1: positivo; 0: negativo), Proge: Receptor de Progesterona (1: positivo; 0: negativo), Recid: Recidiva (1: Sim; 0: Não), SLD: Sobrevida livre da doença (em meses), Óbito: Óbito da paciente, por câncer de mama (1: Sim; 0: Não), ST: Sobrevida total (em meses)