



Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística

**UM ESTUDO DA DISTRIBUIÇÃO BETA WEIBULL
MODIFICADA: MODELANDO A FUNÇÃO DE TAXA DE FALHA NÃO
MONÓTONA**

**LARISSA FERREIRA SOARES ALMEIDA
ALAN CAIRO FERREIRA ROSA**

Brasília
2011

LARISSA FERREIRA SOARES ALMEIDA

ALAN CAIRO FERREIRA ROSA

**UM ESTUDO DA DISTRIBUIÇÃO BETA WEIBULL MODIFICADA:
MODELANDO A FUNÇÃO DE TAXA DE FALHA NÃO MONÓTONA**

Monografia apresentada junto ao
Curso de Estatística da Universidade de Brasília,
na área de concentração de Ciências Exatas,
como requisito parcial à obtenção do título de
Bacharel.

Orientador: Prof. Démerson André Polli

Brasília

2011

DEDICATÓRIA

À minha mãe, Maria Elizabeth Malaquias Ferreira,
pelo amor, confiança e dedicação infinita.
Aos meus familiares, que são essenciais à minha vida.
Aos amigos que caminham comigo e alegam a jornada.
Ao meu grande companheiro, Alan, pela paciência e compromisso.

Larissa F. S. Almeida

À minha mãe, Lucivane Rosa Pereira, à minha irmã, Laís Roberta
Rosa Patrício, e ao meu padrasto, Carlos Roberto,
tudo o que eu produzir com algum valor.
À minha amiga, Larissa, pela paciência e compromisso.

Alan C. F. Rosa

RESUMO

Em alguns modelos de sobrevivência, a suposição de que a função de taxa de falhas tem um crescimento ou decrescimento monótono pode não ser adequado. A proposta deste trabalho é estudar os resultados obtidos por Silva (2008) para as distribuições Beta Weibull Modificada e Weibull Modificada na modelagem de dados com função de taxa de falha não monótona – em especial aquelas em formato de banheira (*bathtub shape*). O uso do modelo Weibull Modificado apresentou uma melhora considerável no ajuste aos conjuntos de dados analisados de sobrevivência quando comparado a modelos usuais da teoria de sobrevivência.

ABSTRACT

In some survival analysis models, the assumption that the hazard function is monotonic may not be appropriate. The Beta Weibull and Weibull Modified distributions – studied by Silva (2008) - are capable of adjust survival data, especially those in which the hazard function is bathtub shaped. The Weibull Modified model showed better fit for the analysed data sets if compared to the results obtained with the usual models.

SUMÁRIO

1 INTRODUÇÃO	6
2 ANÁLISE DE SOBREVIVÊNCIA	7
Função de sobrevivência e função de taxa de falha.....	10
Estimador de Kaplan-Meier	13
3 MODELO BETA MODIFICADO	15
Aplicação do modelo em Análise de Sobrevivência	15
Distribuição Weibull.....	16
Distribuição Weibull Modificada.....	17
Distribuição Beta Weibull Modificada	17
Casos especiais da Beta Weibull Modificada	18
4 MÉTODO DE MÁXIMA VEROSSIMILHANÇA	20
Estimação de Máxima Verossimilhança.....	20
5 FORMA DA FUNÇÃO DE TAXA DE FALHA	22
6 SELEÇÃO DO MODELO	24
Critério de Informação de Akaike (AIC).....	24
Teste de Kolmogorov-Smirnov	24
7 ANÁLISE DOS CONJUNTOS DE DADOS	26
Curva TTT.....	26
Comparação entre Estimador de Kaplan-Meier, Modelo Weibull e Modelo Weibull Modificado...27	
Estimativas de Máxima Verossimilhança e AIC	30
Teste de Kolmogorov-Smirnov.....	30
8 CONSIDERAÇÕES FINAIS	31
Referências Bibliográficas	32
APÊNDICE A – Algoritmo para implementação do Modelo Weibull Modificado	35

1 INTRODUÇÃO

Em alguns modelos de sobrevivência a suposição de que a função de taxa de falha tem um crescimento ou decrescimento monótono pode não ser adequado, pelo fato do risco ter a chamada forma de banheira (*bathhtub shape*). Exemplos de cenários nos quais a função de risco apresenta tal formato são dados de tratamento cirúrgico para tumores, no qual ocorre a recidiva – o risco de óbito no pós-operatório imediato é alto, sofre um rápido decréscimo e, na recidiva, o risco de óbito sobe rapidamente – e confiabilidade de equipamentos eletrônicos – os equipamentos tendem a falhar nos primeiros dias de uso, e caso não falhem neste período somente terão um risco considerável após algum tempo de uso. Nos dois cenários apresentados a função taxa da falha apresenta o chamado formato de banheira.

Uma alternativa para modelar o risco de falha com comportamento similar ao exposto acima são as chamadas distribuições Beta Generalizadas que tem recebido considerável atenção nos últimos anos, em particular depois dos recentes trabalhos de Eugene et al. (2002) e Jones (2004). A proposta deste trabalho é estudar os resultados obtidos por Silva (2008) para as distribuições Beta Weibull Modificada e Weibull Modificada na modelagem de dados com função taxa de risco não monótona – em especial aquelas com formato de banheira (*bathhtub shape*). Esta distribuição apresenta função de sobrevivência e função de taxa de falha analiticamente explícitas, o que favorece o seu uso para estudar o comportamento do tempo de sobrevivência.

Serão aplicadas técnicas paramétricas e não paramétricas para verificar o ajuste dos conjuntos de dados à situação proposta. Em seguida, verificar-se-á a adequação dos modelos supracitados. O resultado esperado é que este modelo ajuste melhor o conjunto de dados cuja função de taxa de falha tenha forma de banheira.

Em suma, são objetivos deste trabalho estudar as distribuições Beta Generalizadas em Análise de Sobrevivência e ajustar modelos de sobrevivência para dados cuja função de taxa de falha seja não monótona, usando tais distribuições.

2 ANÁLISE DE SOBREVIVÊNCIA

A Análise de Sobrevivência engloba um conjunto de métodos e modelos destinados à análise estatística de dados de sobrevivência – cuja característica é ter como variável resposta o tempo até a ocorrência de um evento de interesse, o qual pode ser a recidiva de uma doença, a falência de uma empresa ou a quebra de um aparelho eletrônico. Nesses casos, nem sempre as técnicas de análise tradicionais são válidas devido à presença de censura, ou observação parcial da resposta, cujo conceito será definido mais a frente.

Cabe ressaltar que a Análise de Sobrevivência é uma das áreas da Estatística que mais cresceu nos últimos anos e que esse crescimento está intrinsecamente ligado ao aprimoramento de técnicas estatísticas combinadas ao uso de computadores cada vez mais velozes e com poderosa capacidade para processamento de dados.

O Modelo de Regressão de Cox (Cox, 1972) foi uma grande contribuição na área, pois abriu uma nova fase na modelagem de dados clínicos por sua versatilidade. A principal razão dessa popularidade é a presença do componente não paramétrico, o que torna o modelo bastante flexível. Na década de 80, o artigo de Cox em que o modelo foi descrito constava como o segundo mais citado na literatura estatística, ultrapassado apenas pelo artigo de Kaplan e Meier (1958).

A variável resposta nos estudos de sobrevivência é, geralmente, o tempo até a ocorrência de um evento de interesse. Esse período é denominado tempo de falha, podendo ser o tempo até a morte de um paciente, cura de uma doença ou até mesmo sua recidiva. O tempo inicial, a escala de medida e o evento de interesse (falha) são os elementos que compõem o tempo de falha. Por isso, o marco inicial do estudo deve ser claramente especificado, assim como a definição de falha. Em estudos aleatorizados, a data de aleatorização é a escolha natural para o começo do acompanhamento – data do diagnóstico ou início do tratamento, por exemplo.

A principal característica que diferencia a Análise de Sobrevivência das outras áreas da Estatística é que os dados de sobrevivência, geralmente, apresentam

censura. As censuras acontecem quando a informação do tempo de sobrevivência para algumas observações é incompleta. Isso implica que toda a informação referente à variável resposta se resume ao conhecimento de que o tempo de falha é diferente do tempo de censura.

A presença de censura nos conjuntos de dados amostrais requer técnicas estatísticas especializadas para acomodar a informação contida nessas observações. Com raras exceções, os mecanismos de censura na maior parte dos estudos observacionais são desconhecidos. Então, é necessário fazer suposições sobre eles quando são utilizadas técnicas estatísticas usuais para análise dos dados.

Há três mecanismos conhecidos de censura: a censura do tipo I que ocorre quando o estudo é terminado após um período pré-estabelecido de tempo – os elementos cujo evento de interesse não foi observado são considerados censurados; a do tipo II é aquela em que o estudo é concluído assim que o evento de interesse ocorrer em um número determinado de observações; e, por fim, a censura aleatória acontece quando há perda de observações no decorrer do estudo sem que o evento de interesse tenha ocorrido. É comum em estudos clínicos a ocorrência simultânea das censuras do tipo I e aleatória.

Além disto, as censuras podem ser classificadas como à direita, à esquerda ou intervalar. Partindo da suposição do modelo que todos os elementos amostrais sofrerão a falha, se a falha ocorre após o elemento amostral ser censurado (a falha fica à direita do tempo observado) se diz que houve uma censura à direita; por outro lado, se o elemento amostral sofreu a falha antes da observação (a falha fica à esquerda do tempo observado) se diz que houve uma censura à esquerda; por último, se a falha ocorre entre dois instantes de observação (em um intervalo) se diz que houve uma censura intervalar.

Um tipo mais geral de censura acontece quando o tempo de sobrevivência de um indivíduo, T_i , ocorre entre dois valores, isto é, $T_i \in [L_i, U_i]$, sendo que $L_i \leq T_i \leq U_i$. Para indivíduos cujo evento de interesse ocorreu nesse intervalo, sabe-se que o tempo de sobrevivência é, no mínimo, L_i e, no máximo, U_i . Note que tempos de

falha exatos, bem como tempos censurados à direita e à esquerda, são casos especiais de dados de sobrevivência intervalar.

Uma representação simples do mecanismo de censura à direita - presente nos conjuntos de dados analisados no capítulo 7 deste relatório - é feita utilizando duas variáveis aleatórias. Suponha que o tempo de falha de uma observação seja representado pela variável aleatória T_i e o tempo de censura associado a essa observação seja C_i . Então, os tempos observados consistem em $t_i = \min(T_i, C_i)$ e o indicador de falha é dado por

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

Observe que se todo $C_i = C$, uma constante fixada para o estudo, tem-se a censura do tipo I.

A censura pontual ocorre quando as observações são feitas entre os tempos T_0 e T_1 e o tempo de sobrevivência é conhecido. Isso ilustra situações em que o monitoramento dos dados é frequente ou a ocorrência do evento é bem documentada.

Na Figura 1, cada linha representa o período de risco para cada observação. Aquelas terminadas em asterisco (*) indicam que o evento de interesse ocorreu e as terminadas em um ponto aberto (o) indicam a ocorrência de um evento diferente daquele esperado.

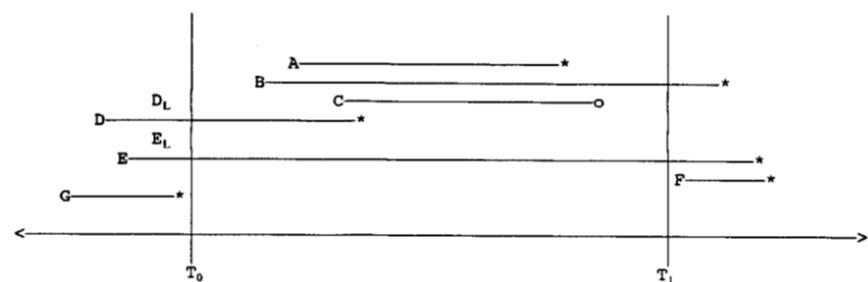


Figura 1: Exemplos de censura pontual

Supondo que as observações se refiram a pacientes acompanhados em um estudo clínico, para o indivíduo A o tempo de falha é conhecido; logo, ele não é censurado. Já para o indivíduo B, o período de risco começa durante o tempo de observação e o evento ocorre após T_1 . O tipo de censura nesse caso é à direita. Para o indivíduo C, a observação também é censurada à direita, pois um evento diferente daquele de interesse ocorre, e assim por diante.

É importante notar que mesmo as observações censuradas devem ser usadas na análise estatística, pois, mesmo incompletas, fornecem informações sobre o tempo de sobrevivência e a sua omissão pode resultar em conclusões viciadas (em geral subestimando o tempo médio de sobrevivência).

2.1 Função de sobrevivência e função de taxa de falha

O tempo de sobrevivência de um indivíduo é denotado pela variável aleatória T . A distribuição de probabilidade destes tempos de sobrevivência pode ser representada por duas funções que são particularmente úteis e bastante usadas para ilustrar diferentes aspectos dos dados em aplicações de sobrevivência: a função de sobrevivência e a função de taxa de falha (ou risco).

Suponha que a variável aleatória T tenha uma distribuição de probabilidade com função densidade de probabilidade $f(t)$. A função de distribuição de T é então dada por

$$F(t) = P(T \leq t) = \int_0^t f(u) du \quad (2.1)$$

e representa a probabilidade de que o tempo de sobrevivência seja menor ou igual a t . A função de sobrevivência denotada por $S(t)$ é definida, então, como a probabilidade do tempo de sobrevivência ser maior que t . Em termos probabilísticos,

$$S(t) = P(T > t). \quad (2.2)$$

ou, escrevendo em termos da função de distribuição,

$$S(t) = 1 - F(t). \quad (2.3)$$

Para descrever a função de sobrevivência é comumente utilizada uma representação gráfica de $S(t)$ que é chamada curva de sobrevivência. Ela pode ser usada para comparar distribuições de sobrevivência entre dois ou mais grupos. É importante salientar que se tratando de distribuições de sobrevivência assimétricas, a média não deve ser usada para descrever a tendência central da distribuição, sendo que o tempo mediano de sobrevivência (o tempo no qual a probabilidade de sobreviver é 50%) é o usualmente aceito.

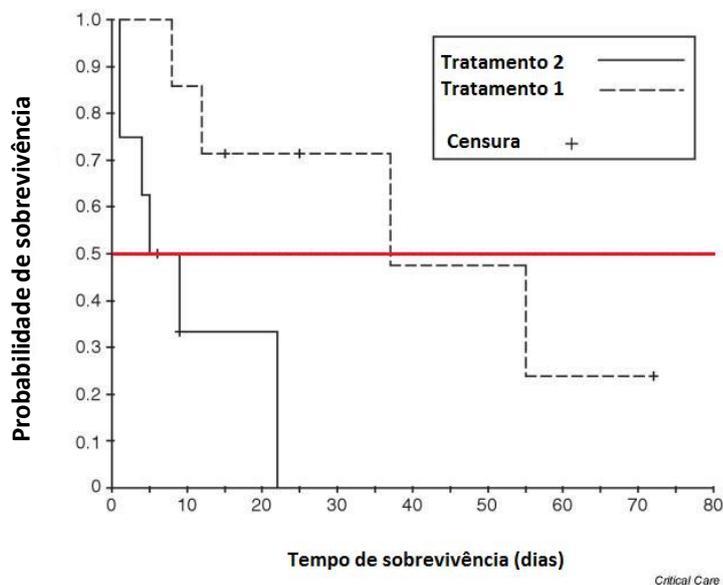


Figura 2: Exemplo de curvas de sobrevivência

Fonte: BEWICK, V.; CHEEK, L.; BALL, J. (2004). *Statistics Review 12: survival analysis*. Crit Care.

As curvas de sobrevivência ilustradas na Figura 2 indicam que durante todo o acompanhamento o tempo de vida é maior para as observações referentes ao Tratamento 1 do que para aquelas referentes ao Tratamento 2, bem como a probabilidade de sobrevivência ao longo do tempo. Podem ser estimados também os tempos medianos de sobrevivência, que são os tempos nos quais 50% das observações falham. Observe a linha vermelha destacada no gráfico.

A função de risco ou *hazard rate*, como também é chamada a função de taxa da falha, é a definida por

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} . \quad (2.4)$$

Assumindo Δt bem pequeno, $h(t)$ representa a taxa de falha instantânea no tempo t condicional à sobrevivência até o tempo t . As taxas de falha são números positivos, mas sem limite superior. A função de taxa de falha é muito útil, pois descreve como a taxa instantânea de falha muda com o tempo.

A função de risco também pode ser descrita como a razão entre a função densidade de probabilidade e a própria função de sobrevivência e como o resultado da derivação do logaritmo neperiano da função de sobrevivência. Dessa forma, uma importante relação entre a função de sobrevivência e a função de risco, quando se trata de variáveis contínuas, é apresentada a seguir:

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} (\log S(t)). \quad (2.5)$$

Neste trabalho será dado enfoque em funções taxa de falha não monótonas, em especial aquelas tratadas na literatura como *bathhtub shaped* ou em formato de banheira. Elas caracterizam processos nos quais há três padrões para a função ao longo do tempo: o período em que $h(t)$ decresce à medida que t cresce, a região de risco padrão ($h(t)$ não se altera significativamente com t) e o período em que $h(t)$ cresce à medida que t cresce. O exemplo clássico de taxa de falha em formato de banheira é a ocorrência de óbito por causas naturais (um recém-nascido tem alto risco mas decresce logo no início do tempo de vida, à medida que envelhece este risco fica praticamente constante e volta a crescer na terceira idade). O mecanismo de formação deste tipo de curva encontra-se explicitado na Figura 3, onde está definida pela linha azul no gráfico.

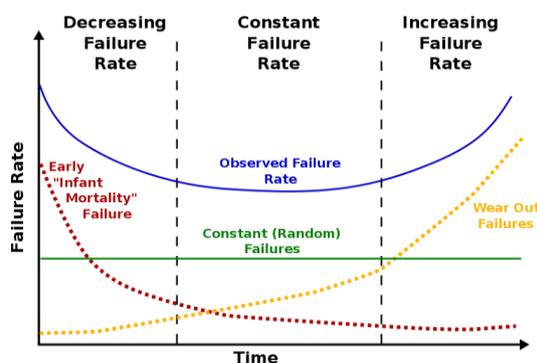


Figura 3: Estrutura da função taxa da falha em forma de banheira

Fonte: BATHTUB Curve. In: *Wikipédia: a enciclopédia livre*.

Outra função útil em análise de dados de sobrevivência é a função de taxa de falha acumulada. Esta função, como o próprio nome sugere, fornece a taxa de falha acumulada do indivíduo e é definida por

$$\Lambda(t) = \int_0^t h(u)du = -\log S(t). \quad (2.6)$$

Essa função não tem uma interpretação direta, mas pode ser útil na avaliação da função de taxa de falha. Isto acontece essencialmente na estimação não paramétrica em que $\Lambda(t)$ apresenta um estimador com propriedades ótimas e $h(t)$ é difícil de ser estimada.

2.2 Estimador de Kaplan-Meier

Em 1958, Kaplan e Meier propuseram um estimador não paramétrico para a função de sobrevivência para uma amostra que continha dados censurados. Este estimador ficou conhecido como estimador de Kaplan-Meier ou estimador limite-produto. Segundo Colosimo e Giolo (2006), este estimador é o mais utilizado em estudos clínicos e vem ganhando cada vez mais espaço em estudos de confiabilidade.

O estimador não paramétrico de Kaplan-Meier para a função de sobrevivência é uma adaptação da função de sobrevivência empírica que é definida como

$$\hat{S}(t) = \frac{\text{n}^\circ \text{ de observações que não falharam até o tempo } t}{\text{n}^\circ \text{ total de observações no estudo}} \quad (2.7)$$

em que $\hat{S}(t)$ é uma função escada com degraus nos tempos observados de falha. Como o que determina o tamanho do degrau é o número de falhas que ocorrem no tempo t , para uma amostra de tamanho n em que não ocorrem censuras ou empates, o tamanho dos degraus será $1/n$.

Na sua construção, o estimador de Kaplan-Meier considera tantos intervalos de tempo quantos forem o número de falhas distintas e os limites dos intervalos de tempo são os tempos de falhas da amostra.

O estimador de Kaplan-Meier é definido como:

$$\hat{S}(t) = \prod_{j:t_j < t} \left(\frac{n_j - d_j}{n_j} \right) = \prod_{j:t_j < t} \left(1 - \frac{d_j}{n_j} \right) \quad (2.8)$$

em que $t_1 < t_2 < \dots < t_k$, os k tempos distintos e ordenados de falha, d_j o número de falhas em t_j , $j = 1, \dots, k$ e n_j o número de indivíduos sob risco em t_j , ou seja, os indivíduos que não falharam e não foram censurados até o instante imediatamente anterior a t_j .

Em suas publicações, autores como Breslow e Crowley (1974) e Kaplan e Meier (1958) mostram a consistência e a normalidade assintótica de $\hat{S}(t)$ sob certas condições, bem como provam que este é o estimador de máxima verossimilhança de $S(t)$.

3 MODELO BETA GENERALIZADO

Em análise de sobrevivência, há conjuntos de dados em que a função de taxa de falha é não monótona – e um caso particular de interesse é o formato de banheira (*bathhtub*). Com isso, os modelos tradicionais de sobrevivência não apresentam um bom ajuste. Uma alternativa que recentemente tem sido explorada são as chamadas distribuições Beta Generalizadas, que consistem em uma extensão da densidade de probabilidade original por meio de uma composição com uma densidade da distribuição Beta.

Partindo da densidade de uma distribuição Beta dada por

$$f(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} \quad (3.1)$$

, na qual $B(a,b) = \int_0^1 \omega^{a-1}(1-\omega)^{b-1}d\omega$, a distribuição Beta Generalizada tem a função de probabilidade acumulada dada por

$$G(x) = \int_0^{F(x)} \frac{1}{B(a,b)} \omega^{a-1}(1-\omega)^{b-1}d\omega, \quad \omega \in [0,1] \quad (3.2)$$

em que $F(x)$ é a probabilidade acumulada original da variável aleatória X , $a > 0$ e $b > 0$ são os parâmetros adicionais ao modelo, introduzidos pela transformação da distribuição Beta. A densidade de probabilidade da distribuição Beta Generalizada é dada por

$$g(x) = \frac{1}{B(a,b)} F(x)^{a-1}(1-F(x))^{b-1} f(x). \quad (3.3)$$

3.1 Aplicação do modelo em Análise de Sobrevivência

A função de probabilidade acumulada de uma distribuição Beta Generalizada pode ser escrita como

$$G(x) = \frac{B_{F(x)}(a,b)}{B(a,b)} \quad (3.4)$$

em que $B_{F(x)}(a, b)$ representa a função Beta Incompleta, ou seja, a função Beta integrada no limite entre 0 e $F(x)$.

A respectiva função de sobrevivência é

$$S(x) = 1 - G(x) = \frac{B(a,b) - B_{F(x)}(a,b)}{B(a,b)} \quad (3.5)$$

e a função de taxa de falha correspondente é

$$h(x) = \frac{g(x)}{s(x)} = \frac{F(x)^{a-1}(1-F(x))^{b-1}f(x)}{B(a,b) - B_{F(x)}(a,b)}. \quad (3.6)$$

Ghitany (2004) mostra que a função de taxa de falha é não monótona quando $0 < a < 1$.

3.2 Distribuição Weibull

Nesta seção será apresentada brevemente a distribuição Weibull. Estas explanações são essenciais para dar suporte ao desenvolvimento teórico da distribuição Beta Weibull Modificada.

A distribuição Weibull é amplamente utilizada em análise de sobrevivência por apresentar várias formas com uma mesma característica: função de taxa de falha é monótona.

A função de distribuição Weibull com três parâmetros é dada pela equação

$$F(x) = 1 - \exp\left[-\left(\frac{x-\tau}{\delta}\right)^\gamma\right], \quad x \geq \tau. \quad (3.7)$$

Os parâmetros $\tau \geq 0$, e $\gamma, \delta > 0$ são, respectivamente, locação, forma e escala da distribuição. Fixando $\tau = 0$, se obtém a distribuição Weibull usual, que é dada por

$$F(x) = 1 - \exp\left[-\left(\frac{x}{\delta}\right)^\nu\right]. \quad (3.8)$$

3.3 Distribuição Weibull Modificada

Lai et al. (2003) propõe uma generalização da distribuição Weibull usual. Tal distribuição, que tem como parâmetros $\lambda, \nu > 0$ e $\beta \geq 0$, apresenta a função de distribuição acumulada dada por

$$F(x) = 1 - \exp\{-\lambda x^\beta \exp(\nu x)\} \quad (3.9)$$

e a função densidade dada por

$$f(x) = \lambda x^{\beta-1}(\beta + \nu x)\exp\{-\lambda x^\beta \exp(\nu x)\}, \quad x > 0. \quad (3.10)$$

Os parâmetros λ e β são, respectivamente, parâmetros de escala e forma da distribuição Weibull Modificada (WM). Segundo Silva (2008), o parâmetro ν é uma espécie de aceleração do tempo de sobrevivência e funciona como um fator de fragilidade na sobrevivência do indivíduo quando o tempo aumenta. Quando $0 < \beta < 1$, a função de taxa de falha apresenta forma de banheira.

Quando $\nu = 0$ na equação (3.10), obtém-se a função de densidade da distribuição Weibull. Fixando $\nu = 0$ e $\beta = 1$ a distribuição WM torna-se a Exponencial, a distribuição Rayleigh é obtida com $\nu = 0$ e $\beta = 2$.

3.4 Distribuição Beta Weibull Modificada

A distribuição Beta Weibull Modificada (BWM) proposta por Silva (2008) será utilizada neste trabalho para modelar dados cuja função de taxa de falha é não monótona. Entretanto, a BWM não é a única distribuição Beta Generalizada que pode ser usada para modelar as taxas de falha em forma de banheira (*bathtub*).

A modificação Beta não define apenas uma distribuição de probabilidades, mas sim uma família de distribuições. Apesar de existirem alternativas para a

modelagem da taxa de falha em forma de banheira (bathtub), a vantagem de se usar a distribuição BWM é que a partir dela se obtém um conjunto de novas distribuições da mesma família (ver detalhes na próxima seção).

Substituindo a função de distribuição $F(x)$ da equação (3.2) pela função de distribuição Weibull Modificada, se obtém a função de distribuição Beta Weibull Modificada, que é dada por

$$G(x) = \frac{1}{B(a,b)} \int_0^{1-\exp\{-\lambda x^\beta \exp(vx)\}} \omega^{a-1} (1-\omega)^{b-1} d\omega. \quad (3.11)$$

A função de densidade, que é obtida utilizando-se a equação (3.3), é

$$g(x) = \frac{\lambda x^{\beta-1} (\beta + vx) \exp(vx)}{B(a,b)} [1 - \exp\{-\lambda x^\beta \exp(vx)\}]^{a-1} \exp\{-b\lambda x^\beta \exp(vx)\}, x > 0. \quad (3.12)$$

A função de taxa de falha da BWM é dada por

$$h(x) = \frac{\lambda x^{\beta-1} (\beta + vx) \exp(vx)}{B(a,b) B_{1-\exp\{-\lambda x^\beta \exp(vx)\}}(a,b)} [1 - \exp\{-\lambda x^\beta \exp(vx)\}]^{a-1} \exp\{-b\lambda x^\beta \exp(vx)\}, x > 0 \quad (3.13)$$

em que $B_{1-\exp\{-\lambda x^\beta \exp(vx)\}}(a,b) = \int_0^{1-\exp\{-\lambda x^\beta \exp(vx)\}} \omega^{a-1} (1-\omega)^{b-1} d\omega$.

Observe que a distribuição BWM possui cinco parâmetros, isto torna a função de taxa de falha mais flexível com relação à sua forma.

3.5 Casos especiais da Beta Weibull Modificada

Distribuição Beta Weibull

Fixando o parâmetro $v = 0$ da BWM, tem-se a distribuição Beta Weibull (BW), em que a função de densidade é

$$g(x) = \frac{\lambda x^{\beta-1} \beta}{B(a,b)} [1 - \exp\{-\lambda x^\beta\}]^{a-1} \exp\{-b\lambda x^\beta\}, x > 0. \quad (3.14)$$

A função de taxa de falha da BW é dada por

$$h(x) = \frac{\lambda x^{\beta-1} \beta}{B(a,b) B_{1-\exp\{-\lambda x^\beta\}}(a,b)} [1 - \exp\{-\lambda x^\beta\}]^{a-1} \exp\{-b\lambda x^\beta\}, x > 0. \quad (3.15)$$

Distribuição Beta Exponencial

Quando $\nu = 0$ e $\beta = 1$ da BWM, obtém-se a distribuição Beta Exponencial (BE) com densidade dada pela equação

$$g(x) = \frac{\lambda}{B(a,b)} [1 - \exp\{-\lambda x\}]^{a-1} \exp\{-b\lambda x\}, x > 0. \quad (3.16)$$

A função de taxa de falha é dada pela equação

$$h(x) = \frac{\lambda}{B(a,b) B_{1-\exp\{-\lambda x\}}(a,b)} [1 - \exp\{-\lambda x\}]^{a-1} \exp\{-b\lambda x\}, x > 0. \quad (3.17)$$

Esses não são os únicos casos especiais da distribuição BWM, a própria distribuição Weibull é obtida com $a = b = 1$ e $\nu = 0$. Para detalhes, ver Silva (2008).

4 MÉTODO DE MÁXIMA VEROSSIMILHANÇA

De maneira geral, dado um conjunto de dados e um modelo estatístico, o método de máxima verossimilhança estima os valores dos diferentes parâmetros do modelo estatístico de maneira a maximizar a probabilidade de a amostra observada ocorrer. A função de verossimilhança para um parâmetro genérico θ é expressa por

$$L(\theta) = \prod_{i=1}^n f(t_i; \theta) \quad (4.1)$$

supondo que $f(t)$ seja a função de densidade da população. Para compor essa função, as observações não censuradas contribuem com a sua função de densidade e as censuradas contribuem com a sua função de sobrevivência, em se tratando de casos de censura à direita. Então, assumindo que sua distribuição não depende dos parâmetros de interesse, tem-se a função de verossimilhança com censura,

$$L(\theta) = \prod_{i=1}^n f(t_i; \theta)^{\delta_i} S(t_i; \theta)^{1-\delta_i} . \quad (4.2)$$

4.1 Estimação de Máxima Verossimilhança

Seja uma amostra aleatória $(t_1, \delta_1), \dots, (t_n, \delta_n)$, em que $t_i = \min(T_i, C_i)$, δ_i é o indicador de falha, T_i segue distribuição BWM $(a, b, \alpha, \gamma, \lambda)$ e T_i e C_i são variáveis aleatórias independentes que representam o tempo de falha e o tempo de censura, respectivamente. Considerando a reparametrização $\lambda = \lambda_1^{-\beta}$ e a presença de censura à direita, Silva (2008) define

$$l(\theta) = -r \log[B(a, b)] + \sum_{i \in F} \{[\log(\beta + vt_i)] + \log \xi_i - \log t_i\} + (a - 1) \sum_{i \in F} \log\{1 - \exp(-\xi_i)\} - b \sum_{i \in F} \xi_i + \sum_{i \in C} \log\{1 - R_{[1 - (\exp(-\xi_i))]}(a, b)\} \quad (4.3)$$

em que $\xi_i = \lambda_1^{-\beta} t_i^\beta \exp(vt_i)$, r é o número de observações não-censuradas, F é o conjunto de observações não-censuradas, C é o conjunto de observações censuradas e $R_{[1 - (\exp(-\xi_i))]}(a, b) = B_{[1 - (\exp(-\xi_i))]}(a, b)/B(a, b)$, que denota a função

razão Beta incompleta. O indicador de censura δ_i é considerado nos somatórios – quando $\delta_i = 1$ a observação pertence ao conjunto F ou a observação pertence ao conjunto C caso contrário. O uso de logaritmo é justificado pelo fato de a maximização do logaritmo de uma função é, em geral, mais simples e produz os mesmos resultados da maximização da função original.

Para encontrar as estimativas de máxima verossimilhança (EMV) da distribuição BWM há a necessidade de trabalhar com métodos numéricos, isto ocorre porque a expressão não apresenta forma analítica fechada para os EMV dos parâmetros da distribuição BWM. Então, maximizando o logaritmo da função de verossimilhança e empregando um algoritmo de otimização não linear, como Newton-Raphson ou um algoritmo quase-Newton (sendo o BFGS é um dos mais populares dentre a classe dos métodos quase-Newton), pode-se obter os estimadores de máxima verossimilhança.

As estimativas apresentadas no capítulo 7 foram calculados usando o *software* R por meio do comando *constrOptim* (ver Apêndice A).

5 FORMA DA FUNÇÃO DE TAXA DE FALHA

A função de taxa de falha definida anteriormente pode assumir diversas formas – no caso deste trabalho, o foco são aquelas que apresentam forma de banheira. Portanto, fez-se necessário definir um método para identificação de modelos mais adequados para essas formas. A curva TTT (Tempo Total em Teste) possibilita a análise gráfica do comportamento do tempo de sobrevivência. Proposta por Barlow & Campo (1975) e estudada por Bergman (1979), é uma ferramenta muito útil em situações em que há informações qualitativas sobre a curva de risco.

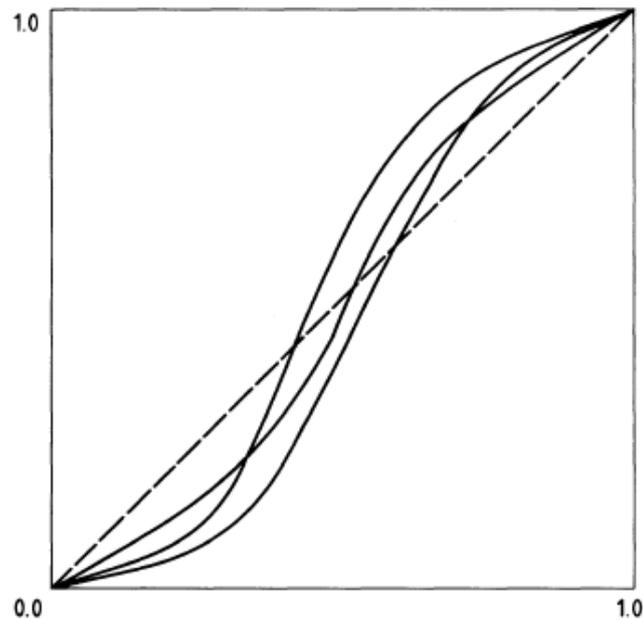


Figura 4: Curvas TTT de distribuição exponencial (diagonal pontilhada) e de algumas distribuições com função de taxa de falha em formato bathtub

Fonte: AARSET, M. V. (1985). *The Null Distribution for a Test of Constant versus "Bathtub" Failure Rate*.

Bergman (1979) sugere que se faça um procedimento para identificação de distribuições com função de taxa de falha em formato *bathtub* pela comparação com a distribuição exponencial.

Este gráfico é construído a partir das quantidades

$$G(r/n) = [(\sum_{i=1}^n T_{i:n}) + (n - r)T_{r:n}] / \sum_{i=1}^n T_{i:n} \quad \text{e} \quad r/n \quad (5.1)$$

em que $r = 1, \dots, n$ e $T_{i:n}, i = 1, \dots, n$ são estatísticas de ordem da amostra.

Aarset (1987) demonstra a relação entre as possíveis formas assumidas pela curva TTT e a forma da função de taxa de falha. Caso a curva seja convexa ou côncava, a função tem decrescimento ou crescimento monótono, respectivamente; caso resulte em uma reta diagonal, trata-se de uma função de risco constante. Por fim, é possível determinar que se trata de uma função de taxa de falha em forma de U caso comece como uma curva convexa e depois torne-se côncava.

6 SELEÇÃO DO MODELO

6.1 Critério de Informação de Akaike (AIC)

O critério de informação proposto por Akaike (1974) é uma medida relativa da qualidade do ajuste de um modelo estatístico paramétrico estimado pelo método de máxima verossimilhança. O AIC é definido por

$$AIC = -2(\log EMV) + 2(\text{número de parâmetros}). \quad (6.2)$$

Esta é uma ferramenta para seleção de modelos, não é um teste de hipóteses – portanto, não faz sentido falar de significância ou p-valor. Dado vários modelos concorrentes, pode-se classifica-los de acordo com o seu AIC, sendo que aqueles que apresentarem o menor valor para este critério são considerados melhores do que os outros. Ou seja, essa medida não define se um modelo é apropriado, apenas permite concluir se este é melhor do que os demais testados.

6.2 Teste de Kolmogorov-Smirnov

Para verificar se a distribuição de probabilidade utilizada ajusta-se a série de dados disponível, pode-se aplicar alguns testes não paramétricos, tais como o do X^2 (Qui-quadrado) e o de Kolmogorov-Smirnov.

Assis *et al.* (1996) define que no teste de aderência por Kolmogorov-Smirnov considera-se $F(x)$ a função de distribuição acumulada assumida para os dados e $H(x)$ a função de distribuição acumulada empírica, com D_{obs} igual ao módulo do desvio máximo observado:

$$D_{obs} = \text{máx}|F(x) - H(x)|. \quad (6.3)$$

Comparando-se D_{obs} com o desvio máximo tabelado, conclui-se que existe concordância entre as funções de distribuição se D_{obs} for menor do que o valor tabelado.

No contexto deste relatório, considere $F(x)$ como sendo a estimativa obtida pelo estimador de Kaplan-Meier e $H(x)$ como sendo a estimativa obtida pela distribuição em teste.

7 ANÁLISE DOS CONJUNTOS DE DADOS

O conjunto de dados analisado nas seções subsequentes está disponível no sítio <http://my.clevelandclinic.org/professionals/software/hazard/downloads.aspx>, cujos arquivos e informações disponibilizados são mantidos e desenvolvidos por The Cleveland Clinic Foundation. A referida clínica é em um centro de medicina sem fins lucrativos, que oferece tratamento clínico e hospitalar e é tida como referência no desenvolvimento de pesquisas e na formação de profissionais na área de saúde. Sua sede está localizada na cidade de Cleveland, que fica no estado de Ohio nos Estados Unidos.

Neste Capítulo, será adotada a seguinte legenda referente aos bancos de dados analisados:

- KUL DEATH: intervalo, em meses, para a ocorrência de óbito após cirurgia realizada para o tratamento de isquemias cardíacas em 5880 pacientes atendidos entre 1971 e 1987 na Catholic University of Leuven;
- VVS DEATH: intervalo, em meses, para a ocorrência de óbito após primeira cirurgia para substituição de válvula cardíaca em 1533 pacientes.

A porcentagem de observações censuradas é de 77,95% e de 90,73%, respectivamente.

7.1 Curva TTT

Para ilustrar com clareza a diferença entre as possíveis formas assumidas pela curva TTT, as Figuras 5 e 6 dão indicações do comportamento da função de taxa de falha.

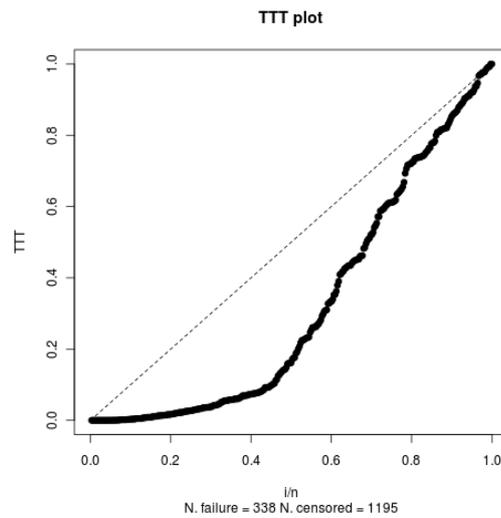


Figura 5: Curva TTT para a variável VVS DEATH

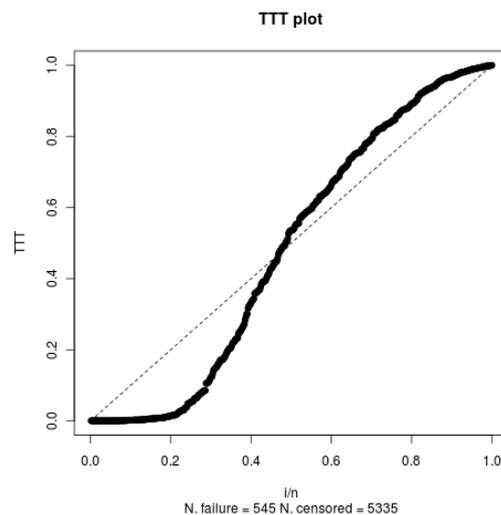


Figura 6: Curva TTT para a variável KUL DEATH

Conforme explicado no capítulo 5, a variável KUL DEATH apresenta função de taxa de falha em forma de banheira. Já a variável VVS DEATH apresenta a referida função monotonamente decrescente.

7.2 Comparação entre Estimador de Kaplan-Meier, Modelo Weibull e Modelo Weibull Modificado

Uma ideia amplamente utilizada para seleção de modelos é comparar as estimativas para a função de sobrevivência e optar pelo modelo paramétrico cujos pontos no gráfico estejam próximos da curva obtida pela função de sobrevivência empírica dada pelo estimador não paramétrico de Kaplan-Meier, assumido como

padrão de comparação por ser o estimador de máxima verossimilhança de $S(t)$. Observa-se abaixo as curvas de sobrevivência estimadas pelo modelo Weibull (linha pontilhada em vermelho) e pelo estimador de Kaplan-Meier e seus intervalos de confiança para variáveis VVS DEATH e KUL DEATH, respectivamente. Confirmou-se que o modelo paramétrico supracitado não é adequado ao segundo conjunto de dados.

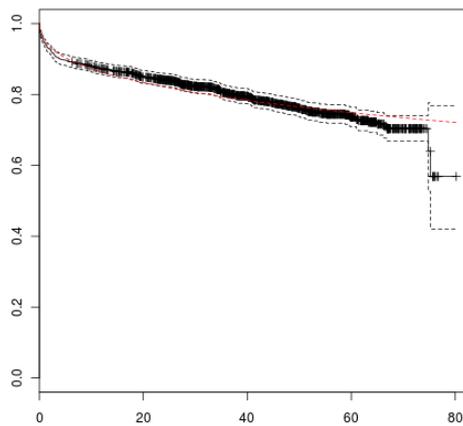


Figura 7: Funções de sobrevivência estimadas a partir do modelo Weibull e do estimador de Kaplan-Meier para a variável VVS DEATH

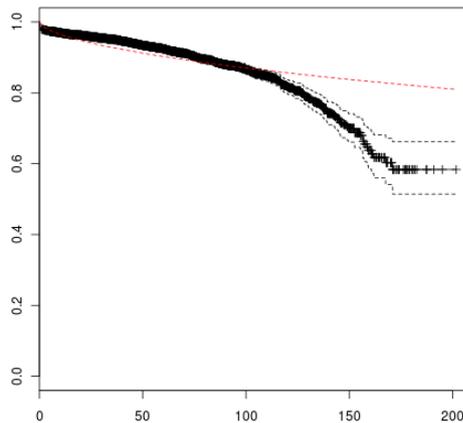


Figura 8: Funções de sobrevivência estimadas a partir do modelo Weibull e do estimador de Kaplan-Meier para a variável KUL DEATH

É interessante notar que o modelo Weibull superestimou a função de sobrevivência nos dois casos.

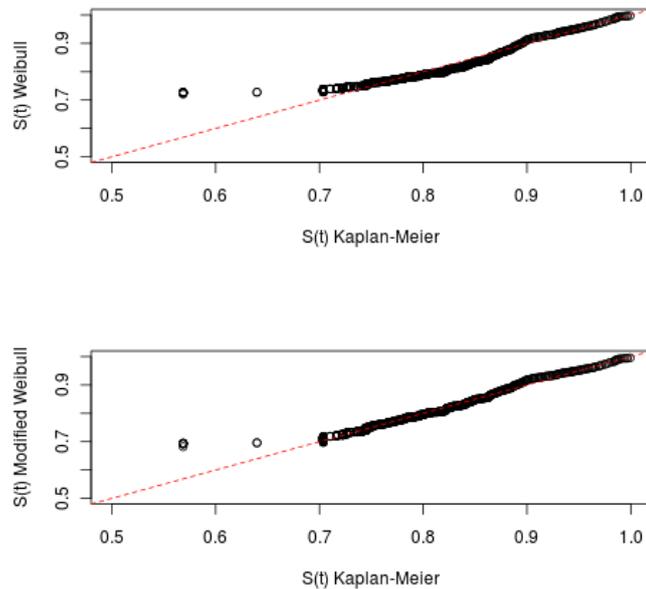


Figura 9: Gráficos das sobrevivências estimadas por Kaplan-Meier *versus* as sobrevivências estimadas pelos modelos Weibull e Weibull Modificado para a variável VVS DEATH

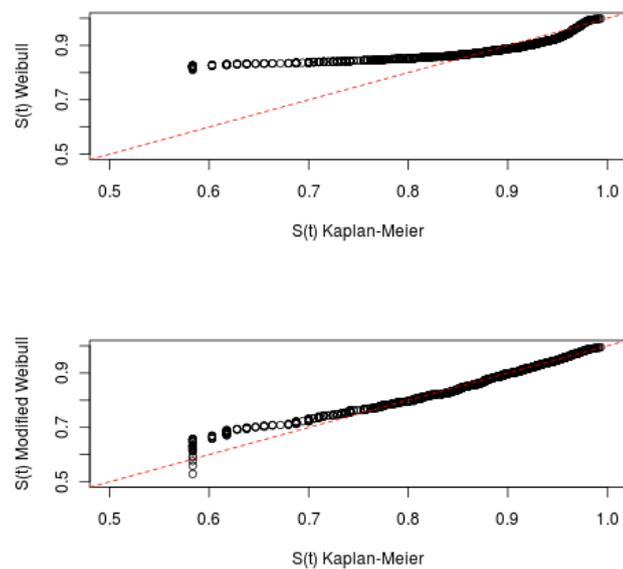


Figura 10: Gráficos das sobrevivências estimadas por Kaplan-Meier *versus* as sobrevivências estimadas pelos modelos Weibull e Weibull Modificado para a variável KUL DEATH

A Figura 10 ilustra com clareza que o modelo Weibull Modificado apresentou um ajuste melhor aos dados da variável KUL DEATH. Isso era esperado, pois, como já foi mencionado, a distribuição Weibull apresenta função de taxa de falha estritamente crescente, decrescente ou constante. No caso da variável VVS DEATH, a melhora na acomodação dos dados pelo modelo modificado foi discreta.

7.3 Estimativas de Máxima Verossimilhança e AIC

Tabela 1: EMV para sub-modelos da BWM e os valores do critério AIC

Variável	Modelo	Parâmetro					AIC
		a	b	λ	β	u	
VVS DEATH	Weibull	-	-	-	0,8598	7,025	3765,8
	Weibull Modificado	1	1	0,0502	0,347	0,059	3750,4
KUL DEATH	Weibull	-	-	-	0,5199	7,925	7875,4
	Weibull Modificado	1	1	0,0128	0,2787	0,012	7577,8

7.4 Teste de Kolmogorov-Smirnov

Tabela 2: Teste de Kolmogorov-Smirnov

Variável	Modelo	Teste de Kolmogorov-Smirnov	
		Estatística	p-Valor
VVS DEATH	Weibull vs. Kaplan-Meier	0,15	<0,0001
	Weibull Modificado vs. Kaplan-Meier	0,068	0,01394
KUL DEATH	Weibull vs. Kaplan-Meier	0,1758	<0,0001
	Weibull Modificado vs. Kaplan-Meier	0,0389	0,01873

Os resultados acima indicam que o modelo Weibull Modificado apresenta o menor valor para o AIC entre os modelos ajustados e, portanto, pode ser escolhido como mais indicado para análise dos dados. Além disso, o teste de aderência de Kolmogorov-Smirnov reforça essa escolha, pois mostra que não há evidências para rejeitar a hipótese de aderência entre o modelo Weibull Modificado e o estimador de Kaplan-Meier. Para esse teste, rejeita-se H_0 ao nível de significância de 1%, assim definido pois as amostras são grandes nos dois casos.

8 CONSIDERAÇÕES FINAIS

Neste trabalho, foram estudadas distribuições modificadas que, por sua flexibilidade, acomodam dados que frequentemente não apresentam bom ajuste quando submetidos a técnicas da teoria de sobrevivência clássica. O algoritmo para implementação do modelo BWM apresentou dificuldade para convergir. Em um primeiro momento, tentou-se obter chutes para os parâmetros a partir de submodelos - isso porque o modelo é muito sensível aos valores iniciais dos parâmetros - e, ainda assim, não foi possível fazer com que houvesse convergência. Uma provável justificativa seria a de que o algoritmo estaciona em pontos de máximos locais, não atingindo o máximo global. Com isso, foi necessário fixar $a = b = 1$, reduzindo o modelo ao submodelo Weibull Modificado, que é capaz de modelar dados cuja função de taxa de falha apresenta forma de banheiro quando $0 < \beta < 1$.

O modelo WM mostrou-se mais adequado para acomodar os dados de sobrevivência estudados. Esse resultado já era esperado, pois este apresenta maior número de parâmetros do que o modelo Weibull tradicional.

Referências bibliográficas

1. AARSET, M. V. (1987). *How to identify bathtub hazard rate*. IEEE Transactions Reliability, 36, 106-108.
2. AARSET, M. V. (1985). *The Null Distribution for a Test of Constant versus "Bathtub" Failure Rate*. Scand J Statist, 12, 55-61.
3. AKAIKE, H. (1974). *A New Look at the Statistical Model Identification*. IEEE Transactions on Automatic Control, 19, 716 – 723.
4. ASSIS, F. N.; ARRUDA, H. V.; PEREIRA, A. R. (1996). *Aplicações de estatística à climatologia*. Pelotas-RS: Universitária, 161p.
5. BARLOW, R . E. & CAMPO ,R . (1975). *Total time on test processes and applications to failure data analysis*. Reliability and fault tree analysis (ed. Barlow, Fussel & Singpurwalla).
6. BARRETO-SOUZA, W.; MORAIS, A. L.; CORDEIRO, G. M.(2008). *The Weibull-Geometric Distribution*. Journal of Statistical Computation and Simulation, p. 1-14.
7. BATHTUB Curve. In: Wikipédia: a enciclopédia livre. Disponível em: <http://en.wikipedia.org/wiki/Bathtub_curve>. Acesso em 17 de Janeiro de 2011.
8. BEWICK, V.; CHEEK, L.; BALL, J. (2004). *Statistics Review 12: survival analysis*. Crit Care., v. 8, p. 389–394. Disponível em <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1065034/>>. Acessado em 17 de Janeiro de 2011.
9. BERGMAN, B. (1979). *On age replacement and the total time on test concept*. Scand. J. Statist, 6 , 161-168.
10. BRESLOW, N. & CROWLEY, J. (1974). *A large Study of the Life Table and Product Limit Estimates Under Random Censorship*. Annals of Statistics, 2, 437- 453.
11. CARRASCO, J. M. F.; ORTEGA, E. M. M.; CORDEIRO, G. M (2009). *The Generalized Modified Weibull Distribution*, p. 1-12.
12. CARRASCO, J. M. F., ORTEGA, E. M. M., CORDEIRO, G. M. (2008). *A generalized modified Weibull distribution for lifetime modeling*. Computational Statistics and Data Analysis, 53, 450-462.

13. COLOSIMO, E. A., & GIOLO, S. R. (2006). *Análise de Sobrevida Aplicada*. São Paulo: Edgard Blücher.
14. COX, D.R. (1972). *Regression Models and life Tables (with discussions)*. Journal Royal Statistical Society, B, 34, 187-220
15. EUGENE, N.; LEE, C.; FAMOYE, F. (2002). *Beta-Normal Distribution and its Applications*. Communication in Statistics – Theory and Methods, Nova York, v.31, p.497-512.
16. FAMOYE, F.; LEE, C.; OLUMALADE, O. (2005) *The Beta-Weibull Distribution*. Journal of Statistics – Theory and Applications, Holanda, v.4, p. 121-136.
17. FLETCHER, R. (1987). *Practical Methods of Optimization*, Second Edition, Chichester, UK: John Wiley & Sons.
18. GHITANY, M.E. (2004). *The monotonicity of the reliability measures of the beta distribution*. Appl. Math. Lett. 17,1277–1283 .
19. KWAN-MOON, L.; ELASHOFF, R.M.; AFIFI, A.A. (1997). *Censored Issues in Survival Analysis*. Annual Reviews of Public Health, v. 18, p. 83-104.
20. LAI, C. D., XIE, M. and MURTHY, D. N. P. (2003). *A modified Weibull distribution*. Transactions on Reliability, v. 52, n. 1, p. 33-37.
21. KAPLAN, E. L., MEIER, P. (1958). *Nonparametric estimation from incomplete observations*. Journal of the American Statistical Association, 53, P. 457-481
22. JONES, M. C. (2004) *Families of Distributions Arising from Distributions of Order Statistics*. TEST, 13, P. 1, 43.
23. NADARAJAH, S. (2009). *Bathtub-shaped Failure Rate Functions*. Qual Quant, v. 43, n. 5, p. 855-863.
24. NADARAJAH, S.; CORDEIRO, G. M.; ORTEGA, E. M. M. (2009). *General Results for the Beta Modified Weibull Distribution*, p. 1-23.
25. NADARAJAH, S.; KOTZ, S. (2006) *Some Beta Distributions*. Bulletin of the Brazilian Mathematics Society, v. 37, n. 1, p. 103-125.

26. R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
27. SILVA, G. O. (2008) *Modelos de regressão quando a função de taxa de falha não é monótona e o modelo probabilístico beta Weibull modificado*. Tese (Doutorado em Agronomia) – Universidade de São Paulo, Escola Superior de Agricultura Luiz de Queiroz, São Paulo.

APÊNCIDE A – Algoritmo para implementação do Modelo Weibull Modificado

```

library(survival)

ibeta = function(x, a, b) { pbeta(x, a, b) * beta(a, b) }

# Ajusta o modelo Modified Weibull
# -----
-
ll.mw = function(par, T, dt) {
  alpha = par[1]
  gamma = par[2]
  lambda = par[3]

  failure = sum(dt) * log(alpha) +
            sum(dt * (log(gamma + lambda * T) + (gamma - 1) * log(T)
+ lambda * T)) -
            sum(dt * alpha * T^(gamma) * exp(lambda * T))
  censored = sum((1 - dt) * alpha * T^(gamma) * exp(lambda * T))
  failure - censored
}

S.mw = function(par, T) {
  alpha = par[1]
  gamma = par[2]
  lambda = par[3]
  exp(-alpha * T^(gamma) * exp(lambda * T))
}

# Parametrizacao da Weibull no modelo do Gauss:
#
#  $f(x) = \alpha * x^{(\gamma - 1)} * (\gamma) * \exp(-\alpha * x^{(\gamma)})$ 
#
# Parametrizacao da Weibull no rweibull - shape a, scale b
#
#  $f(x) = a * x^{(a - 1)} * (1/b)^a * \exp(-x^a * (1/b)^a)$ 
#
# entao,
#
#  $\gamma = a$ 
#  $\alpha = (1/b)^a \rightarrow 1/\alpha^{(1/a)} = b$ 
#
# Parametrizacao do survreg()
#
#  $f(x) = (1/s) * x^{((1/s) - 1)} * (1/\exp(1))^{(1/s)} * \exp(-x^{(1/s)} * (1/b)^{(1/s)})$ 
#
#
#
#  $scale = 1/a \rightarrow a = 1/scale$ 
#  $intercept = \log(b) \rightarrow b = \exp(intercept)$ 
#
# survreg's scale = 1/(rweibull shape)
# survreg's intercept = log(rweibull scale)
#
# Finalmente: reparametrizando o do Gauss para o do survreg
#
#  $\gamma = 1/scale$ 

```

```

# alpha = exp(-gamma * intercept)

# KUL DEATH
# =====
fit = survreg(Surv(INT_DEAD, DEAD) ~ 1, data = d.kul, dist = "weibull")
summary(fit)

# Ajusta o modelo MW para obter os chutes iniciais de (alpha, gamma,
lambda)
# -----
ipar = with(fit, c(exp(-as.numeric(coefficients)/scale), 1/scale,
0.010000))

tm = proc.time()
fit.op = constrOptim(ipar, ll.mw, grad = NULL, control = list(fnscale =
-1),
  T = with(d.kul, INT_DEAD), dt = with(d.kul, DEAD),
  ui = matrix(c(1, 0, 0, 0,
                0, 1, -1, 0,
                0, 0, 0, 1), nrow = 4), ci = c(0, 0, -1, 0))
fit.op
proc.time() - tm

fit.km = survfit(Surv(INT_DEAD, DEAD) ~ 1, data = d.kul)
fit.wb = survreg(Surv(INT_DEAD, DEAD) ~ 1, data = d.kul, dist =
"weibull")
st = with(fit.km, surv)
time = with(fit.km, time)

prob.wb = pweibull(time, shape = 1/fit.wb$scale, scale =
exp(coef(fit.wb)), lower.tail = FALSE)
prob.mw = S.mw(with(fit.op, par), time)

mean((prob.wb - st)^2)
mean((prob.mw - st)^2)

png("kul-mw-death.png")
opar = par(mfrow = c(2, 1))
plot(st, prob.wb, xlab="S(t) Kaplan-Meier", ylab="S(t) Weibull",
      xlim=c(0.5, 1), ylim=c(0.5, 1))
abline(a = 0, b = 1, lty = 2, col = 2)
plot(st, prob.mw, xlab="S(t) Kaplan-Meier", ylab="S(t) Modified
Weibull",
      xlim=c(0.5, 1), ylim=c(0.5, 1))
abline(a = 0, b = 1, lty = 2, col = 2)
par(opar)
dev.off()

ks.test(st, prob.wb)
ks.test(st, prob.mw)

aic.wb = 2*2 - 2*fit.wb$loglik[1]
aic.mw = 2*3 - 2*fit.op$value

# VVS DEAD
# =====
fit = survreg(Surv(INT_DEAD, DEAD) ~ 1, data = d.vvs, dist = "weibull")
summary(fit)

```

```

# Ajusta o modelo MW para obter os chutes iniciais de (alpha, gamma,
lambda)
# -----
ipar = with(fit, c(exp(-as.numeric(coefficients)/scale), 1/scale,
0.010000))
ipar

tm = proc.time()
fit.op = constrOptim(ipar, ll.mw, grad = NULL, control = list(fnscale =
-1),
  T = with(d.vvs, INT_DEAD), dt = with(d.vvs, DEAD),
  ui = matrix(c(1, 0, 0, 0,
                0, 1, -1, 0,
                0, 0, 0, 1), nrow = 4), ci = c(0, 0, -1, 0))
fit.op
proc.time() - tm

fit.km = survfit(Surv(INT_DEAD, DEAD) ~ 1, data = d.vvs)
fit.wb = survreg(Surv(INT_DEAD, DEAD) ~ 1, data = d.kul, dist =
"weibull")

st = with(fit.km, surv)
time = with(fit.km, time)

prob.wb = pweibull(time, shape = 1/fit.wb$scale, scale =
exp(coef(fit.wb)), lower.tail = FALSE)
prob.mw = S.mw(with(fit.op, par), time)

mean((prob.wb - st)^2)
mean((prob.mw - st)^2)

png("vvs-mw-death.png")
opar = par(mfrow = c(2, 1))
plot(st, prob.wb, xlab="S(t) Kaplan-Meier", ylab="S(t) Weibull",
      xlim=c(0.5, 1), ylim=c(0.5, 1))
abline(a = 0, b = 1, lty = 2, col = 2)
plot(st, prob.mw, xlab="S(t) Kaplan-Meier", ylab="S(t) Modified
Weibull",
      xlim=c(0.5, 1), ylim=c(0.5, 1))
abline(a = 0, b = 1, lty = 2, col = 2)
par(opar)
dev.off()

ks.test(st, prob.wb)
ks.test(st, prob.mw)

aic.wb = 2*2 - 2*fit.wb$loglik[1]
aic.mw = 2*3 - 2*fit.op$value

```