

Universidade de Brasília

Análise de dados de transplante de medula óssea:  
Proposta do modelo de regressão Kumaraswamy-Weibull com fração de cura

Marcos Lima de Oliveira

Brasília  
2014



Marcos Lima de Oliveira  
Bacharel em Estatística

**Análise de dados de transplante de medula óssea:  
Proposta do modelo de regressão Kumaraswamy-Weibull com fração de cura**

Orientadora:  
Profa. Dra. **Juliana Betini Fachini Gomes**

**Brasília  
2014**



## DEDICATÓRIA

*À Deus, que é de onde vem toda  
minha força e sabedoria.*

*Aos meus amados pais,*

*Creusilene Lima Ferreira e José Carlos, por todos  
os momentos ao meu lado, me apoiando, me aconse-  
lhando e principalmente me suportando.*



## 1 AGRADECIMENTOS

A **Deus** por todos os dias eu ter a possibilidade de acordar, levantar e ver quão bonito é o dia que tenho pela frente, de todas as possibilidades e oportunidades que tive em minha vida. Por Ele sempre me dar paz, saúde, força, sabedoria e me ensinar a ter humildade, paciência, determinação e perseverança.

A minha mãe, **Creusilene Lima Ferreira** pelo amor, paciência, companheirismo, carinho e pelas orações.

Ao meu pai, **José Carlos Jesus de Oliveira** por toda ajuda e carinho.

A amiga, orientadora, doutora, professora e conselheira **Juliana Betini Fachini Gomes**, por toda carinho, preocupação, atenção e comprometimento.

Ao departamento de estatística da Universidade de Brasília e sua equipe de docentes, por toda a contribuição para o meu futuro profissional e pessoal.

Aos meus avós, **Neocelina Ângela Jesus de Oliveira** e **José de Oliveira** por me conceder alimentação, possibilitar que eu durma no chão da sala durante os últimos 3 anos e por suas torcidas e orações.

Ao meu irmão **Marcelo Ferreira Batista** por todo companheirismo, incentivo, motivação e contribuição para o meu crescimento pessoal.

A minha irmã **Mayra Lima** por toda a torcida.

As minhas primas **Jayne Lima** que iniciou todo esse processo ao meu lado e sempre me apoio e incentivou.

A toda minha família pela torcida, pelo carinho, pela as orações e pela paciência pela minhas ausências.

Pela minha companheira **Jessyca Gomes da Silva** por todo amor, carinho, companheirismo, dedicação, momentos e paciência pra me aguentar.

Aos meu amigos **Firmino da Silva** e **Lusineide Gomes** por todo apoio, ajuda e carinho.

Aos meus amigos-irmãos, Luiz Fernando, Matthaus, Daniel Rodrigues, Maria Emanuele, Ana Claudia e Rafael Brasil por estarem sempre por perto, por todos os nossos momentos e por estarem sempre de alguma forma me ensinando algo.

Aos meus amigos da UnB Thiago Pfaffiman, Tales Dias, Luiz Paulo Roriz,

Frederico Barros Diniz, Luciano Moura e Rayanny de Oliveira pelos semestres e matérias compartilhados e por todo o carinho e paciência.

Aos meus amigos do CIL Alex, João Paulo, Rafael e Kerrye por todas as conversas em inglês e momentos juntos.

Ao Roberto por compartilha seu conhecimento e por estar sempre à disposição.

A ESTAT Consultoria por contribuir para meu crescimento pessoal e profissional.





## SUMÁRIO

1 AGRADECIMENTOS . . . . .	5
RESUMO . . . . .	10
ABSTRACT . . . . .	12
2 INTRODUÇÃO . . . . .	14
3 REVISÃO DE LITERATURA . . . . .	17
3.1 Conceitos de Análise de Sobrevivência . . . . .	17
3.2 Estimador de Kaplan-Meier . . . . .	18
3.3 Comparação de curvas de sobrevivência . . . . .	19
3.4 Função do Tempo Total em Teste (Curva TTT) . . . . .	21
3.5 Distribuições . . . . .	22
3.6 Fração de Cura . . . . .	25
3.7 Modelo de Regressão . . . . .	26
3.7.1 Modelo de Locação e Escala . . . . .	26
3.7.2 Modelo de Regressão usando Reparametrização . . . . .	28
3.8 Análise de Resíduos . . . . .	29
3.8.1 Resíduos de Cox-Snell . . . . .	29
3.8.2 Resíduo de Martingal . . . . .	30
3.8.3 Resíduo de Deviance . . . . .	30
4 MATERIAL E MÉTODOS . . . . .	31
4.1 Material . . . . .	31
4.2 Métodos . . . . .	32
4.2.1 Modelo de Regressão Kumaraswamy Weibull com fração de cura . . . . .	32
4.2.2 Método de Máxima Verossimilhança . . . . .	32
5 RESULTADOS E DISCUSSÃO . . . . .	35
5.1 Análise descritivas dos dados . . . . .	35
5.2 Modelo de regressão Weibull . . . . .	43
5.3 Modelo Kumaraswamy-Weibull com fração de cura . . . . .	47
6 CONSIDERAÇÕES FINAIS . . . . .	51
REFERÊNCIAS . . . . .	53

ANEXOS ..... 55



## RESUMO

### **Análise de dados de transplante de medula óssea: Proposta do modelo de regressão Kumaraswamy-Weibull com fração de cura**

Neste trabalho serão apresentados estudos sobre os modelos de regressão Weibull e Kumaraswamy Weibull usando técnicas de análise de sobrevivência. A modelagem de fração de cura foi utilizada, pois observou que nos dados utilizados, quando o tempo é muito grande há indivíduos que não são mais suscetíveis à falha. Para a estimação dos parâmetros, foi utilizado o método de máxima verossimilhança sujeito a restrição nos espaço paramétrico. Para a avaliação do modelo, foram utilizados os dados de pessoas com leucemia mieloide crônica que receberam transplante de medula óssea.

Palavras-chave: Modelo de regressão; Análise de sobrevivência; Censura; Fração de cura, Weibull, Kumaraswamy, Kumaraswamy-Weibull;



## ABSTRACT

This work studies the Weibull regression models and Kumaraswamy Weibull regression model using survival analysis techniques. Fraction for healing was used, because we realized that the data used when time tends to infinity there are individuals who aren't longer susceptible to failure. For the estimation parameter, was used the maximum likelihood method subject to restriction at parameter space. For avaluation of model, was used the data of people with myeloid leukemia chronic that received bone marrow transplantation.

Keywords: Regression models; Survival analysis; censorship; Fraction of healing, Weibull, Kumaraswamy, Kumaraswamy-Weibull.





## 2 INTRODUÇÃO

A análise de sobrevivência ou também conhecida como análise de confiabilidade, tem o interesse de estudar dados relacionados ao tempo decorrido até a ocorrência de um determinado evento de interesse, o que é geralmente conhecido como tempo de falha. Na análise de sobrevivência há muitas aplicações, como por exemplo, na área da saúde com o estudo de um tratamento de câncer, no qual, o evento de interesse é o óbito do indivíduo devido ao câncer e o tempo de falha é o tempo desde o início do estudo até o óbito do indivíduo devido ao câncer. Na área financeira, um exemplo de tempo de falha é o tempo que uma pessoa leva para ficar inadimplente. Contudo, essa técnica pode ser aplicada em diversas áreas do conhecimento.

O que diferencia a análise de sobrevivência ou análise de confiabilidade da análise de regressão comum, seria a informação adicional chamada censura. A censura é a indicação de que por alguma razão o evento de interesse não ocorreu no período do estudo, como por exemplo, o paciente morrer por uma causa diferente do câncer, o tempo do estudo acabar e o evento de interesse ainda não ter ocorrido.

Geralmente nos estudos de sobrevivência quanto maior o tempo, a probabilidade de sobrevivência diminui, ou seja, quando o tempo tende a ser muito grande a probabilidade de sobrevivência tende a zero. Mas em alguns casos esse tipo de situação não acontece, o que pode indicar a presença de indivíduos curados. Indivíduos esses que não são suscetíveis ao evento de interesse, ou seja, indivíduos que nunca irão falhar. Nestes casos deve-se utilizar os modelos com fração de cura.

Considerando todos os atributos descritos e a presença de covariáveis vinculadas a cada indivíduo, o trabalho consiste em propor um modelo de regressão com fração de cura para analisar os dados de pessoas com leucemia mieloide crônica que foram submetidas a transplante de medula óssea.

Diante das diversas distribuições de probabilidade com inúmeras características, neste estudo foi utilizada a distribuição Kumaraswamy-Weibull e seu caso particular, a distribuição Weibull, para propor um modelo de regressão com fração de cura, pois essa família de distribuições tem uma grande flexibilidade, quando aplicado em dados reais em relação aos seus respectivos sub-modelos.

Para realizar as análises estatísticas desse estudo foi utilizado o software R versão 3.0.1 (R DEVELOPMENT CORE TEAM, 2012).



### 3 REVISÃO DE LITERATURA

#### 3.1 Conceitos de Análise de Sobrevivência

Na análise de sobrevivência um fator muito importante é o tempo de falha. O tempo de falha é o tempo até a ocorrência do fenômeno de interesse. Para medir esse tempo se tem três considerações, a primeira é o tempo inicial, ou seja, o tempo que marca o início de um estudo; em segundo lugar, a escala a ser medida e por último, considera-se o evento de interesse, também conhecido como falha.

Em estudos na área financeira, se costuma considerar como tempo de falha, desde o momento que um indivíduo entra no estudo até o momento que ele se torna inadimplente. Já na engenharia é comum considerar o tempo de falha, desde o momento que o equipamento entra em funcionamento até a falha do equipamento. E na área da saúde que é o caso a ser abordado nesse relatório, é mais comum considerar como tempo de falha, desde o momento que o indivíduo entra no estudo, por exemplo, receber transplante de medula óssea até óbito do indivíduo.

Existem informações que são incompletas, porém de extrema importância para uma boa análise dos dados, essas informações são chamadas de censuras. A censura é a perda de informação decorrente de não se ter observado ocorrência da falha ou a falha ter ocorrido, antes do início do estudo. Existem diversas razões e causas que definem observações censuradas.

Pode ser observado três tipos de censuras; a censura a direita, a censura a esquerda e a censura intervalar. A censura a esquerda é quando o evento de interesse acontece antes do início do estudo. A censura intervalar é quando o estudo é observado em intervalos de tempo, ou seja, o evento de interesse ocorre em um intervalo de tempo. A censura a direita é quando o tempo de ocorrência do evento está à direita do tempo registrado. Esta última censura pode ser dividida em três tipos de censuras, que são, as censuras do tipo I, do tipo II e a aleatória.

Quando o estudo acaba e o evento de interesse ainda não aconteceu para alguns indivíduos, chama-se esses indivíduos de censurados, para ser mais exato esses indivíduos sofreram a censura a direita do tipo I. A censura a direita do tipo II é quando é delimitado a quantidade de eventos de interesse a ser observado, como por exemplo, observar dez indivíduos

e o estudo vai acabar quando for observado cinco falhas, ou seja, todos os indivíduos que sobreviveram após as cinco falhas são classificados como censura à direita do tipo II. A censura a direita aleatória pode acontecer quando, o indivíduo sai do estudo por um evento que não seja o de interesse ou quando o estudo termina e o evento de interesse ainda não ocorreu, ou seja, a censura a direita do tipo I é um caso particular da censura a direita aleatória.

De acordo com o Colosimo e Giolo (2006), os dados de sobrevivência para o indivíduo  $i$  ( $i = 1, \dots, n$ ) sob estudo são representados, em geral, pelo par  $(t_i, \delta_i)$  sendo  $t_i$  o tempo de falha ou de censura e  $\delta_i$  a variável indicadora de falha ou censura, isto é:

$$\delta_i = \begin{cases} 1, & \text{se } t_i \text{ é um tempo de falha} \\ 0, & \text{se } t_i \text{ é um tempo de censura} \end{cases}$$

Desta forma, a variável aleatória resposta em análise de sobrevivência é representada por duas colunas no banco de dados. Na presença de covariáveis medidas no  $i$ -ésimo indivíduo, tais como,  $x_i = (\text{sexo}_i, \text{idade}_i, \text{tratamento}_i)$ , dentre outros, os dados ficam representados por  $(t_i, \delta_i, x_i)$  (Colosimo e Giolo, 2006).

Para a representação do tempo de sobrevivência será usado a função distribuição acumulada,  $F(t)$ ; função densidade de probabilidade,  $f(t)$ ; função de sobrevivência,  $S(t)$ ; função de risco,  $h(t)$  e a relação dessas distribuições.

### 3.2 Estimador de Kaplan-Meier

Na análise descritiva dos dados de análise de sobrevivência, não se pode usar técnicas comuns, pois nos dados há a existência de informações adicionais, informações essas que podem ser chamadas de censuras. Quando existe esse tipo de informação adicional nos dados, uma alternativa é usar o estimador de Kaplan-Meier, pois com esse estimador é possível calcular algumas medidas descritivas considerando as censuras. De acordo com Colosimo e Giolo (2006), a expressão geral do estimador de Kaplan-Meier pode ser apresentada após as seguintes considerações preliminares:

- (i)  $t_1 < t_2 \dots < t_k$ , os  $k$  tempos distintos e ordenados de falha
- (ii)  $d_j$  o número de falhas em  $t_j$ ,  $j = 1, \dots, k$  e

(iii)  $n_j$  o número de indivíduos sob risco e  $t_j$ , ou seja, os indivíduos que não falharam e não foram censurados até o instante imediatamente anterior a  $t_j$ .

Sendo assim, o estimador de Kaplan-Meier é, então, definido como:

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

As principais propriedades do estimador de Kaplan-Meier são basicamente as seguintes (Colosimo e Giolo, 2006) :

- é não viciado para amostras grandes,
- é fracamente consistente,
- converge assintoticamente para um processo gaussiano e
- é estimador de máxima verossimilhança de  $S(t)$ .

Quando não existe a presença de informações adicionais (censuras), a função de sobrevivência empírica é definida como:

$$\hat{S}(t) = \frac{n^\circ \text{ de observações que não falharam até o tempo } t}{n^\circ \text{ total de observações no estudo}} \quad (2)$$

O comportamento da função de sobrevivência quando o estudo envolve censuras do tipo I e II, também é o comportamento de função empírica, porém essa função não atinge zero, devido as últimas observações serem censuradas.

### 3.3 Comparação de curvas de sobrevivência

Um dos objetivos deste estudo é comparar o efeito das covariáveis no tempo de sobrevivência. Quando as covariáveis são categóricas, existem testes que podem ser aplicados, como o teste de logRank e o de Wilcoxon. Quando as variáveis são contínuas, elas podem ser categorizadas para que seja possível aplicar o teste.

O teste de logRank proposto por Mantel (1966) para dados censurados tem o pressuposto de risco proporcionais, ou seja, a razão das funções de risco dos grupos comparados são constantes. Essa suposição pode ser analisado quando é observado as curvas de sobrevivência dos grupos que estão sendo estudados.

Para esses testes, considere dois ou mais grupos de indivíduos. Assim, temos as seguintes hipóteses:

$H_0$  : Não existe diferença entre as curvas de sobrevivência

$H_1$  : Existe diferença entre as curvas de sobrevivência

Considere  $k$  falhas distintas e ordenadas no tempo,  $t_1 < t_2 < t_3 < \dots < t_k$ .

Considere também:

- $n_j$  o número total de indivíduos sob risco no tempo  $t_j$ ;
- $n_{1j}$  o número de indivíduos do primeiro grupo sob risco no tempo  $t_j$ ;
- $d_j$  o número total de falhas no tempo  $t_j$ ;
- $d_{1j}$  o número de falhas no primeiro grupo no tempo  $t_j$ .

A distribuição de  $d_{1j}$  é uma hipergeométrica com média  $E(d_{1j})$  e variância  $Var(d_{1j})$  onde:

$$E(d_{1j}) = d_j \frac{n_{1j}}{n_j} = w_{1j} \quad (3)$$

e

$$Var(d_{1j}) = \frac{n_{1j}d_j(n_j - d_j)(n_j - n_{1j})}{n_j^2(n_j - 1)} = V_{1j} \quad (4)$$

Assim,  $d_{1j} - w_{1j}$  tem média zero e variância  $V_{1j}$  e a estatística do teste de LogRank é definida por:

$$T = \frac{[\sum_{j=1}^k (d_{1j} - w_{1j})]^2}{\sum_{j=1}^k V_{1j}} \quad (5)$$

que tem distribuição aproximadamente qui-quadrado com 1 grau de liberdade.

O teste de Wilcoxon não necessita do pressuposto de riscos proporcionais. Esse teste coloca peso igual ao número de indivíduos sob risco. Assim, o peso fica maior na proporção inicial do eixo tempo. As hipóteses para esse teste são as mesmas que a do teste de LogRank. Porém a estatística do teste de Wilcoxon é representada por:

$$S = \frac{[\sum_{j=1}^k n_j(d_{1j} - w_{1j})]^2}{\sum_{j=1}^k n_j V_{1j}} \quad (6)$$

que tem distribuição aproximadamente qui-quadrado com 1 grau de liberdade.

### 3.4 Função do Tempo Total em Teste (Curva TTT)

Para avaliar possíveis distribuições de probabilidade que se ajustam aos dados e auxiliarão na procura do melhor modelo, será usado a curva TTT. A curva TTT também conhecido como *TTTplot*, é a metodologia do gráfico do tempo total em teste, que foi proposto por Aarset (1987) e serve para ajudar a escolher o modelo mais apropriado para os dados.

O *TTTplot* é definido por:

$$G(r/n) = \frac{[(\sum_{i=1}^r T_{i:n}) + (n-r)T_{r:n}]}{(\sum_{i=1}^n T_{i:n})} \quad (7)$$

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

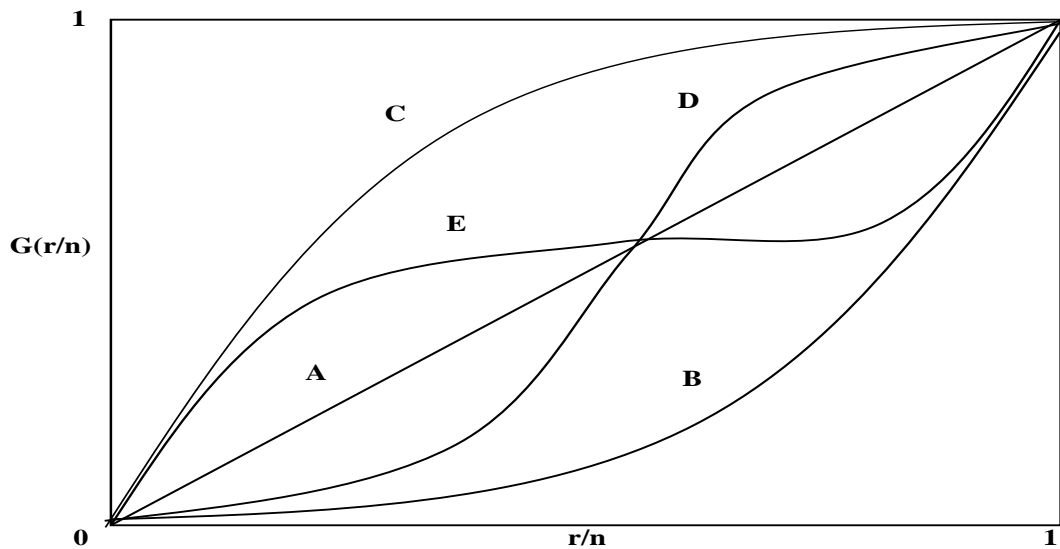


Figura 1 – Exemplo de curvas TTT



É possível visualizar algumas formas do *TTTplot*, na Figura 1. A partir destas formas é possível definir possíveis distribuições, com funções de risco, constante, crescente, decrescentes, unimodais, e algumas distribuições que contemplam uma função de risco na forma de banheira, ou, também conhecida como função de risco em forma de U.

As funções de riscos são definidas de acordo com as seguintes relações:

- Indicado pela a letra A, a reta diagonal supõe uma distribuição com função de risco constante;
- Indicado pela a letra B, a curva representa uma forma convexa, o que supõe uma distribuição com função de risco decrescente;
- Indicado pela a letra C, a curva representa uma forma côncava, o que supõe uma distribuição com forma de risco crescente;
- Indicado pela a letra D, a curva que se comporta primeiramente de forma convexa e depois de forma côncava, neste caso se supõe uma distribuição com função de risco na forma de U.
- Indicado pela letra E, a curva que se comporta primeiramente de forma côncava e depois convexa, neste caso se supõe uma distribuição com função de risco na forma unimodal.

### 3.5 Distribuições

Uma distribuição de probabilidade descreve a chance que uma variável aleatória pode assumir em um espaço amostral, ou seja, no conjunto de valores possíveis de assumir. A distribuição de probabilidade é uma função que tem como domínio os valores da variável aleatória e como o conjunto imagem o intervalo restrito entre 0 e 1.

Neste estudo será usado uma distribuição com atributos necessários para ajustar os dados de análise de sobrevivência com informações adicionais, ou seja, censuras e que consideram valores maiores ou iguais a zero. Para considerar estes fatores algumas distribuições de probabilidade serão consideradas, tais como: Weibull, Kumaraswamy, Kumaraswamy Generalizada e Kumaraswamy Weibull.

A distribuição Weibull (COLOSIMO e GIOLO, 2006) é uma das funções mais utilizadas na análise de sobrevivência, especialmente para dados na área da saúde e enge-

nharia. A distribuição Exponencial é um caso particular da Weibull (COLOSIMO e GIOLO, 2006, Santos, 2013). A função densidade de probabilidade, a função de sobrevivência e a função de risco da distribuição Weibull, são respectivamente representadas por:

$$g(t; \alpha, \gamma) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1} \exp \left[ - \left( \frac{t}{\alpha} \right)^\gamma \right], t \geq 0, \quad (8)$$

$$S(t; \alpha, \gamma) = \exp \left[ - \left( \frac{t}{\alpha} \right)^\gamma \right], t \geq 0, \quad (9)$$

e

$$h(t; \alpha, \gamma) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1}, t \geq 0, \quad (10)$$

em que  $\gamma$  e  $\alpha$  são positivos e representam, respectivamente os parâmetros de forma e de escala. Esta distribuição tem sua função de risco crescente quando  $\gamma > 1$ , a forma decrescente quando  $\gamma < 1$  e constante quando  $\gamma = 1$ , que resulta na distribuição Exponencial.

Em 1980 uma nova distribuição de probabilidade foi proposta por Kumaraswamy com muita utilidade na área hidrológica e com uma grande semelhança a distribuição Beta, o que pode se observado em Jones (2009). A função distribuição acumulada e função densidade de probabilidade Kumaraswamy são respectivamente:

$$W(t; \alpha, \beta) = 1 - (1 - t^\alpha)^\beta, 0 < t < 1, \quad (11)$$

$$w(t; \alpha, \beta) = \alpha \beta t^{\alpha-1} (1 - t^\alpha)^{\beta-1}, 0 < t < 1, \quad (12)$$

A distribuição Kumaraswamy tem dois parâmetros de forma  $\alpha$  e  $\beta$  que são positivos e permite uma maior flexibilidade da função densidade de probabilidade. Algumas formas da função densidade de probabilidade são:

- A forma unimodal quando  $\alpha > 1$  e  $\beta > 1$ ,
- A forma de U ou forma de banheira quando,  $\alpha < 1$  e  $\beta < 1$ ,
- A forma crescente quando  $\alpha > 1$  e  $\beta \leq 1$ ,
- A forma decrescente quando  $\alpha \leq 1$  e  $\beta > 1$  e

- A forma constante quando  $\alpha = 1$  e  $\beta = 1$

Segundo Jones (2009), a constante normalizadora em  $w(t; \alpha, \beta)$  tem forma muito simples e a distribuição não tem envolvimento de funções especiais nas funções quantílicas e na função distribuição acumulada apresentada acima. Porém em outras propriedades dessa distribuição se assemelham às da distribuição Beta.

Dada a grande utilidade da distribuição Kumaraswamy, Cordeiro e Castro (2010) sugeriram a distribuição Kumaraswamy generalizada. Ele propõem esta distribuição com base nos trabalhos de Eugene et al. (2002) e Jones (2009). Distribuição que tem função distribuição acumulada e função densidade de probabilidade, representadas por:

$$F(t; a, b) = 1 - [1 - G^a(t)]^b, \quad (13)$$

e

$$f(t; a, b) = abg(t)G^{a-1}(t)[1 - G^a(t)]^{b-1}. \quad (14)$$

em que  $a$  e  $b$  são parâmetros positivos, que estão relacionados com a assimetria e flexibilização dos pesos das caudas.  $G(t)$  é uma função distribuição acumulada arbitrária e  $g(t) = \frac{dG(t)}{dt}$  é a função densidade de probabilidade arbitrária.

A interpretação física dessa distribuição supõe um sistema composto de  $b$  componentes independentes sendo que cada componente é composto de  $a$  subcomponentes independentes. Supõe-se que o sistema falhe se algum dos  $b$  componentes falhem e que cada componente falhe se todos os subcomponentes falhem (Fachini, 2011).

Dada a utilidade da distribuição Weibull e a flexibilidade da Kumaraswamy Generalizada, neste trabalho será proposto o uso da distribuição Kumaraswamy Weibull. Esta distribuição é o resultado da mistura da distribuição Kumaraswamy Generalizada com a distribuição Weibull. A distribuição Kumaraswamy Weibull tem como, função distribuição acumulada, função densidade de probabilidade, função de sobrevivência e função de risco, respectivamente:

$$F(t; \alpha, \gamma, a, b) = 1 - \left\{ 1 - \left\{ 1 - \exp \left[ - \left( \frac{t}{\alpha} \right)^\gamma \right] \right\}^a \right\}^b, \quad t \geq 0, \quad (15)$$

$$f(t; \alpha, \gamma, a, b) = ab \frac{\gamma}{\alpha^\gamma} t^{\gamma-1} \exp \left[ - \left( \frac{t}{\alpha} \right)^\gamma \right] \left\{ 1 - \exp \left[ - \left( \frac{t}{\alpha} \right)^\gamma \right] \right\}^{a-1} \quad (16)$$

$$\left\{ 1 - \left\{ 1 - \exp \left[ - \left( \frac{t}{\alpha} \right)^\gamma \right] \right\}^a \right\}^{b-1} t \geq 0,$$

$$S(t; \alpha, \gamma, a, b) = \left\{ 1 - \left\{ 1 - \exp \left[ - \left( \frac{t}{\alpha} \right)^\gamma \right] \right\}^a \right\}^b t \geq 0, \quad (17)$$

e

$$h(t; \alpha, \gamma, a, b) = \frac{ab \frac{\gamma}{\alpha^\gamma} t^{\gamma-1} \exp \left[ - \left( \frac{t}{\alpha} \right)^\gamma \right] \left\{ 1 - \exp \left[ - \left( \frac{t}{\alpha} \right)^\gamma \right] \right\}^{a-1} \left\{ 1 - \left\{ 1 - \exp \left[ - \left( \frac{t}{\alpha} \right)^\gamma \right] \right\}^a \right\}^{b-1}}{1 - \left\{ 1 - \exp \left[ - \left( \frac{t}{\alpha} \right)^\gamma \right] \right\}^a} t \geq 0. \quad (18)$$

A flexibilidade da distribuição Kumaraswamy Weibull destaca-se ao compará-la aos seus respectivos sub-modelos em aplicações com dados reais (Fachini, 2011). Pode-se encontrar mais informações sobre esta distribuição em Cordeiro et al. (2010).

### 3.6 Fração de Cura

Dentro da análise de sobrevivência, considera-se que todos os indivíduos em estudos possivelmente falharão. Porém, em alguns casos mesmo com um tempo suficientemente longo, alguns indivíduos não falham e nem vão falhar. Isto indica a presença de fração de cura, ou seja, indivíduos que não apresentam a possibilidade de ocorrência do evento de interesse.

De acordo com Fachini(2011), há situações em que uma parte da amostra pode não ser suscetível ao evento de interesse e dessa forma nunca experimentar o evento de interesse (falha). Esses indivíduos são conhecidos como não suscetíveis ou curados.

A função de sobrevivência que considera os indivíduos curados é chamada de função impropria, ou, função de sobrevivência populacional. Para calcular esta função dividem-se as observações em duas partes, a primeira parte considera-se os indivíduos suscetíveis ao evento de interesse e a segunda parte considera-se os indivíduos que não são suscetíveis. Após a divisão se atribui as probabilidades  $\phi$  e  $(1 - \phi)$ , respectivamente. O modelo é representado por:

$$S_{pop}(t) = (1 - \phi) + \phi S(t), \quad (19)$$

sendo que  $\phi$  pertence ao intervalo  $[0, 1]$  e  $(1 - \phi)$  é o parâmetro que representa a fração de cura, ou seja,  $(1 - \phi)$  representa a porcentagem de indivíduos curados.

As propriedades da função são:

- $\lim_{t \rightarrow \infty} S_{pop}(t) = (1 - \phi)$  e
- $\lim_{t \rightarrow 0} S_{pop}(t) = 1$ .

Usando as relações da função densidade probabilidade, função de sobrevivência e função de risco, a função densidade populacional e a função de risco populacional são respectivamente representadas por:

$$f_{pop}(t) = \phi f(t), \quad (20)$$

e

$$h_{pop}(t) = \frac{\phi f(t)}{(1 - \phi) + \phi S(t)}. \quad (21)$$

### 3.7 Modelo de Regressão

Na prática, a grande maioria dos estudos envolvem a presença de covariáveis ou variáveis explicativas que representam heterogeneidade existente na população tais como, idade, sexo, tratamento que está sendo testado, entre outras, aos quais os indivíduos são submetidos.

Neste contexto, o objetivo da análise de sobrevivência está centrado na relação entre os tempos de sobrevivência e as variáveis explicativas de interesse, com o intuito de estudar essa relação, modelos de regressão podem ser formulados ao considerar uma reparametrização da distribuição de probabilidade dos tempos ou usando modelos de locação e escala.

#### 3.7.1 Modelo de Locação e Escala

O modelo de locação e escala leva em consideração uma variável aleatória representando o tempo até a ocorrência da falha. O vetor de covariáveis,  $\mathbf{x} = (x_1, x_2, \dots, x_p)$

formado por  $p$  variáveis regressoras. O modelo é caracterizado por  $Y = \log(T)$ , que tem uma distribuição com um parâmetro de locação representado por  $\mu(x)$  dependendo das variáveis regressoras e um segundo parâmetro denominado parâmetro de escala, que é representado por  $\sigma$ .

O modelo é:

$$Y = \mu(x) + w\sigma, \quad (22)$$

sendo que  $\sigma > 0$  e  $w$  é o erro aleatório que tem distribuição conhecida que é independente de  $\mathbf{x}$ .

A função de probabilidade tem a seguinte forma:

$$f(y|\mathbf{x}) = \frac{1}{\sigma} g\left(\frac{y - \mu(x)}{\sigma}\right), \quad -\infty < y < \infty, \quad (23)$$

em que  $g\left(\frac{y - \mu(x)}{\sigma}\right)$  é uma função densidade de probabilidade (f.d.p), totalmente especificada em  $(-\infty, \infty)$ .

A forma a ser assumida para  $\mu(x)$ , será:

$$\mu(x) = \mathbf{x}^T \boldsymbol{\beta} \quad (24)$$

Em que a  $\boldsymbol{\beta}^T = (\beta_0, \beta_1, \dots, \beta_p)$  é o vetor de parâmetros associado às covariáveis.

Sendo  $T$  uma variável aleatória com distribuição Weibull. E seja:

- $Y = \log(t)$ .
- $g_Y(y) = g_T(y)|J|$
- $|J| = \frac{dt}{dy} = \exp(y)$
- $g(t; \alpha, \gamma) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1} \exp\left[-\left(\frac{t}{\alpha}\right)^\gamma\right], t \geq 0,$

Assim,

$$g_y(y) = \frac{\gamma}{\alpha^\gamma} \exp(y)^{\gamma-1} \exp\left[-\left(\frac{\exp(y)}{\alpha}\right)^\gamma\right] \exp(y) \quad (25)$$

$$= \frac{\gamma}{\alpha^\gamma} \exp(y)^\gamma \exp \left[ - \left( \frac{\exp(y)}{\alpha} \right)^\gamma \right]$$

Substituindo a reparametrização  $\gamma = \frac{1}{\sigma}$  e  $\alpha = \exp(\mu(x))$  na equação (21), então será obtido a seguinte expressão:

$$\begin{aligned} g_y(y) &= \frac{1}{\sigma \exp(y)^\gamma} \exp(\mu(x))^{\frac{1}{\sigma}} \exp \left[ - \left( \frac{\exp(y)}{\exp(\mu(x))} \right)^{\frac{1}{\sigma}} \right] \\ &= \frac{1}{\sigma} \exp \left( \frac{y - \mu(x)}{\sigma} \right) \exp \left[ - \exp \left( \frac{y - \mu(x)}{\sigma} \right) \right] \\ &= \frac{1}{\sigma} \exp \left[ \left( \frac{y - \mu(x)}{\sigma} \right) - \exp \left( \frac{y - \mu(x)}{\sigma} \right) \right], -\infty < y < \infty. \end{aligned} \quad (26)$$

Note que a variável aleatória  $Y$  segue a distribuição do valor extremo com parâmetros  $\mu(x)$  e  $\sigma$  e é um modelo de locação e escala. Ao incluir as covariáveis no parâmetro de locação como mostra a equação (20), tem-se um modelo de regressão do valor extremo ou um modelo de regressão log Weibull.

### 3.7.2 Modelo de Regressão usando Reparametrização

Neste trabalho também será definido modelo de regressão ao utiliza reparametrização nos parâmetros das distribuições de probabilidade. Esse modelo leva em consideração uma variável aleatória representando o tempo desde o início do estudo até a ocorrência da falha.

Seja o vetor de covariáveis,  $\mathbf{x}^T = (x_1, x_2, \dots, x_p)$  formado por  $p$  variáveis regressoras, a reparametrização é caracterizada por relacionar uma distribuição qualquer com um modelo de regressão ao considerar que pelo menos um parâmetro depende do vetor de covariáveis. Essa relação pode ser representada genericamente por:

$$\alpha = g(\mathbf{x}^T \boldsymbol{\beta}) \quad (27)$$

em que  $g(\cdot)$  é uma função positiva e contínua, sendo  $\boldsymbol{\beta}$  é o vetor de parâmetros desconhecidos e associados as covariáveis.

Seja  $T$  uma variável aleatória com distribuição Kumaraswamy-Weibull definida na equação (12) por  $f(t; \alpha, \gamma, a, b)$ , em que  $t > 0$ ,  $a > 0$  e  $b > 0$  parâmetros adicionais,  $\alpha$

parâmetro de escala,  $\gamma$  parâmetro de forma. Neste estudo, propõem-se as reparametrizações  $\alpha = \exp\{\mu\}$  e  $\gamma = \frac{1}{\sigma}$ , em que  $-\infty < \mu < \infty$  e  $\sigma > 0$ . Ao considerar o vetor de variáveis regressoras  $\mathbf{x}^T = (x_1, x_2, \dots, x_n)$ , usualmente considera-se,  $\mu = \mathbf{x}^T \boldsymbol{\beta}$  em que  $\boldsymbol{\beta}^T = (\beta_0, \beta_1, \dots, \beta_p)$  é vetor de parâmetros desconhecido associados às covariáveis. Sendo assim, a reparametrização pode ser escrita por:  $\alpha = \exp\{(\beta_0, \beta_1 x_1, \dots, \beta_p x_p,)\}$  e  $\gamma = \frac{1}{\sigma}$ . O modelo de regressão Kumaraswamy Weibull tem como função distribuição acumulada, função densidade probabilidade e função de sobrevivência, respectivamente dada por:

$$F(t; \boldsymbol{\beta}, \sigma, a, b) = 1 - \left\{ 1 - \left\{ 1 - \exp \left[ - \left( \frac{t}{\exp\{\mathbf{x}^T \boldsymbol{\beta}\}} \right)^{\frac{1}{\sigma}} \right] \right\}^a \right\}^b, \quad (28)$$

$$f(t; \boldsymbol{\beta}, \sigma, a, b) = ab \frac{1}{\sigma} \frac{1}{\exp\{\frac{\mathbf{x}^T \boldsymbol{\beta}}{\sigma}\}} t^{\frac{1}{\sigma}-1} \exp \left[ - \left( \frac{t}{\exp\{\mathbf{x}^T \boldsymbol{\beta}\}} \right)^{\frac{1}{\sigma}} \right] \left\{ 1 - \exp \left[ - \left( \frac{t}{\exp\{\mathbf{x}^T \boldsymbol{\beta}\}} \right)^{\frac{1}{\sigma}} \right] \right\}^{a-1} \quad (29)$$

$$\left\{ 1 - \left\{ 1 - \exp \left[ - \left( \frac{t}{\exp\{\mathbf{x}^T \boldsymbol{\beta}\}} \right)^{\frac{1}{\sigma}} \right] \right\}^a \right\}^{b-1},$$

e

$$S(t; \boldsymbol{\beta}, \sigma, a, b) = \left\{ 1 - \left\{ 1 - \exp \left[ - \left( \frac{t}{\exp\{\mathbf{x}^T \boldsymbol{\beta}\}} \right)^{\frac{1}{\sigma}} \right] \right\}^a \right\}^b. \quad (30)$$

### 3.8 Análise de Resíduos

Com o objetivo de avaliar adequacidade do modelo ajustado em relação aos dados de pessoas com leucemia mieloide crônica que receberam transplante de medula óssea, será descrito nesta seção a análise dos resíduos de Cox-Snell, Martingal e Deviance.

#### 3.8.1 Resíduos de Cox-Snell

Para examinar o ajuste global dos modelos, usam-se os resíduos de Cox-Snell. Esses resíduos são quantidades determinadas por:

$$\hat{e}_i = \hat{\Lambda}(t_i | \mathbf{x}_i), \quad (31)$$



Em que  $\widehat{\Lambda}(\cdot)$  é a função de taxa de falha acumulada obtida do modelo ajustado. Segundo Lawless (2003) e Colosimo e Giolo (2006) esse resíduo vem de uma população homogênea e deve seguir uma distribuição exponencial padrão caso o modelo seja adequado. Para auxiliar na verificação do ajuste do modelo, será usado o gráfico das curvas de sobrevivência desses resíduos, obtidas pelo modelo exponencial padrão e pelo estimador de Kaplan-Meier, sendo que  $\widehat{S}(\widehat{e}_i)$  é a função de sobrevivência dos  $\widehat{e}_i$ s retirada do estimador de Kaplan-Meier.

### 3.8.2 Resíduo de Martingal

De acordo com o Colosimo e Giolo (2006) o resíduo de Martingal é como uma estimativa do número de falhas em excesso observada nos dados mas não predito pelo modelo. Esse resíduo é definido por:

$$\widehat{m}_i = \delta_i - \widehat{e}_i. \quad (32)$$

No qual  $\delta_i$  é a variável indicadora de falha e  $\widehat{e}_i$ , os resíduos de Cox-Snell. O resíduo de Martingal serve para verificar a melhor forma funcional de uma covariável quantitativa dos dados em estudo, ou seja, se ela será usada de forma linear ou será necessário algum tipo de transformação, como deixar a covariável na forma quadrática.

### 3.8.3 Resíduo de Deviance

O resíduo de deviance é definido por:

$$\widehat{d}_i = \text{sign}(\widehat{m}_i) [-2(\widehat{m}_i + \delta_i \log(\delta_i - \widehat{m}_i))]^{\frac{1}{2}}. \quad (33)$$

Esses resíduos são uma tentativa de tornar os resíduos de martingal mais simétrico em torno de zero, facilitam, em geral, a detecção de pontos atípicos (*outliers*). Se o modelo for apropriado, esses resíduos devem apresentar um comportamento aleatório em torno de zero. Gráficos dos resíduos martingal ou deviance, versus os tempos fornecem, assim, uma forma de verificar a adequação do modelo ajustado, bem como auxiliam na detecção de observações atípicas (Colosimo e Giolo, 2006).

## 4 MATERIAL E MÉTODOS

Nesta seção serão apresentados os métodos que serão usados no presente estudo, como o modelo de regressão e o processo que será utilizado para a estimação dos parâmetros.

### 4.1 Material

Os dados provêm de uma coorte de 96 pacientes submetidos a transplante de medula óssea para tratamento de leucemia mieloide crônica, no período de junho de 1986 a junho de 1998, no Centro de Transplante de Medula Óssea (CEMO) do Instituto Nacional do Câncer (INCA).

O acompanhamento dessa coorte possibilitou o estudo do efeito de fatores prognósticos para a ocorrência de doença do enxerto contra hospedeiro aguda e crônica, da sobrevivência livre de doença e da sobrevivência global. As covariáveis registradas para cada paciente estão descritas a seguir:

- Tempo observado até o óbito devido a leucemia mieloide crônica (em dias)
- Sexo (1=Masculino, 2=Feminino)
- Idade na data do transplante
- Status – Indicador de falha ou censura (0=censura, 1=óbito)
- Plaq – recuperação de plaquetas (0=não, 1=sim)
- Tempplaq – dias até a recuperação de plaquetas
- Deag – ocorrência de doença enxerto aguda (0=não, 1=sim)
- Tempdeag – dias até a ocorrência da doença enxerto aguda
- Decr – doença enxerto crônica (0=não, 1=sim)
- Tempdecr – tempo até a ocorrência da doença enxerto crônica
- Fase – fase da doença na data do transplante (1=crônica, 2=aguda, 3=crise blástica)

## 4.2 Métodos

### 4.2.1 Modelo de Regressão Kumaraswamy Weibull com fração de cura

Com base no modelo de fração de cura apresentado em (15) e (16). e com o modelos de regressão apresentado em (25) e (26). O modelo de regressão Kumaraswamy Weibull com fração de cura, tem função densidade de probabilidade populacional e função de sobrevivência populacional, respectivamente definidas como:

$$f_{pop} = \phi \left[ ab \frac{1}{\sigma} \frac{1}{\exp\{\frac{\mathbf{x}^T \boldsymbol{\beta}}{\sigma}\}} t^{\frac{1}{\sigma}-1} \exp \left[ - \left( \frac{t}{\exp\{\mathbf{x}^T \boldsymbol{\beta}\}} \right)^{\frac{1}{\sigma}} \right] \left\{ 1 - \exp \left[ - \left( \frac{t}{\exp\{\mathbf{x}^T \boldsymbol{\beta}\}} \right)^{\frac{1}{\sigma}} \right] \right\}^{a-1} \right] \quad (34)$$

$$\left\{ 1 - \left\{ 1 - \exp \left[ - \left( \frac{t}{\exp\{\mathbf{x}^T \boldsymbol{\beta}\}} \right)^{\frac{1}{\sigma}} \right] \right\}^a \right\}^{b-1},$$

e

$$S_{pop} = (1 - \phi) + \phi \left[ \left\{ 1 - \left\{ 1 - \exp \left[ - \left( \frac{t}{\exp\{\mathbf{x}^T \boldsymbol{\beta}\}} \right)^{\frac{1}{\sigma}} \right] \right\}^a \right\}^b \right]. \quad (35)$$

em que  $0 < \phi < 1$ ,  $\boldsymbol{\beta}$  é o parâmetro associado às variáveis regressores,  $\sigma > 0$  é o parâmetro de escala  $a$  e  $b$ , são os parâmetros que estão relacionados a assimetria e flexibilização dos pesos da caudas.

O modelo Kumaraswamy Weibull com fração de cura apresenta algumas distribuições como casos particulares. Quando  $\phi = 1, a = 1$  e  $b = 1$ , o resultado é a distribuição Weibull, quando  $\phi = 1, b = 1$ , o resultado é a distribuição Weibull exponenciada, quando  $\phi = 1, \sigma = 1$  e  $b = 1$ , o resultado é a distribuição Exponencial exponenciada, entre outros sub-modelos.

### 4.2.2 Método de Máxima Verossimilhança

Para este estudo será usado na estimação dos parâmetros o método de máxima verossimilhança. Este método consiste em procurar uma distribuição de maneira a maximizar a amostra observada, ou seja, encontrar a distribuição com maior probabilidade de ter

encontrado a amostra observada.

Para esta estimação usa-se o seguinte cálculo:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n f(t_i; \boldsymbol{\theta}), \quad (36)$$

em que  $n$ , é o número total de observações na amostra e  $\boldsymbol{\theta} = (\phi, a, b, \boldsymbol{\beta}, \sigma)$  é o vetor de parâmetros. O método de máxima verossimilhança procura maximizar este vetor de parâmetros, pois assim encontram-se os valores que mais se aproximam dos valores reais.

Dentro da análise de sobrevivência, a estimação pelo método de máxima verossimilhança, considera a variável indicadora de censura citada na secção 2.1. Sendo que quando a observação falha, ou seja, o evento de interesse ocorre, então é considerada a função densidade de probabilidade  $f(t_i; \boldsymbol{\theta})$  e quando a observação é censurada considera a função de sobrevivência  $S(t_i; \boldsymbol{\theta})$ . A expressão fica da seguinte forma:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n [f_{pop}(t_i; \boldsymbol{\theta})]^{\delta_i} [S_{pop}(t_i; \boldsymbol{\theta})]^{1-\delta_i} \quad (37)$$

Para obter o estimador de máxima verossimilhança, utiliza-se o logaritmo da função de verossimilhança, que é:

$$l(\boldsymbol{\theta}) = \log(L(\boldsymbol{\theta})) = \sum_{i=1}^n \delta_i \log[f(t_i; \boldsymbol{\theta})] + (1 + \delta_i) \log[S(t_i; \boldsymbol{\theta})]. \quad (38)$$

Depois dos passos anteriores, deriva-se parcialmente o logaritmo da função de verossimilhança  $l(\boldsymbol{\theta})$ , se resolve o sistema de equações:

$$U(\boldsymbol{\theta}) = \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0. \quad (39)$$

Neste trabalho será usado o método de máxima verossimilhança com restrição no espaço paramétrico. Este método estima os valores que maximizam a função de máxima verossimilhança. Ao considerar o modelo definido em (30) e (31) o vetor de parâmetros  $\boldsymbol{\theta}$  é definido por  $\boldsymbol{\theta} = (\phi, a, b, \boldsymbol{\beta}, \sigma)^T$  e neste modelo têm as seguintes restrições nos parâmetros:  $0 < \phi < 1$ ,  $a > 0$ ,  $b > 0$  e  $\sigma > 0$ . Dessa forma, neste trabalho será utilizada a função de verossimilhança com restrição representada por:

$$l(\boldsymbol{\theta}, v) = l(\boldsymbol{\theta}) + v \sum_{i=1}^k (u_i \boldsymbol{\theta} - c_i), \quad (40)$$

em que  $v$  é um parâmetro de ajuste que é sempre positivo e o vetor de parâmetro  $\boldsymbol{\theta}$  tem  $k$  restrições de inequações lineares  $u_i \boldsymbol{\theta} - c_i \leq 0$ , em que  $u_i, i = 1, 2, \dots, k$ , são  $k$  por 1 vetores e  $c_i$  são escalares que assumem valores 0 ou 1 (Santos, 2013). Para estimar  $\boldsymbol{\theta}$  será utilizado a função `constrOptim` do software R (R DEVELOPMENT CORE TEAM, 2012). Pode-se encontrar mais informações sobre este método em Lange (1999) e o sistema de ajuda do *software* R.

## 5 RESULTADOS E DISCUSSÃO

### 5.1 Análise descritivas dos dados

Com o interesse de conhecer um pouco mais os dados, optou-se por fazer uma análise exploratória. Sabendo que o banco de dados tem 48,96% de indivíduos que foram censurados. Primeiramente será utilizado o estimador de Kaplan-Meier, para o tempo de sobrevivência dos indivíduos com leucemia mieloide crônica que receberam transplante de medula óssea.

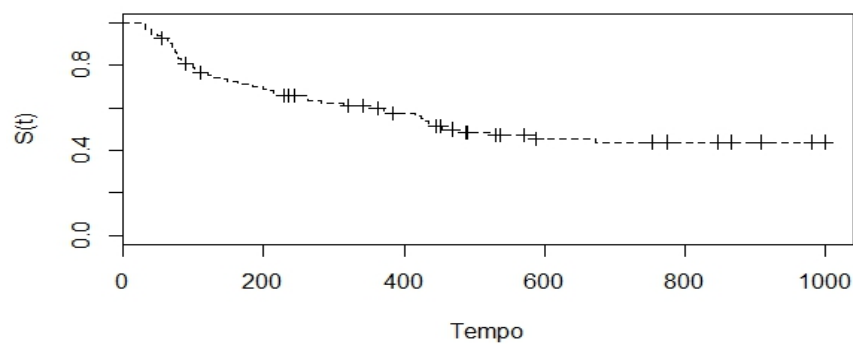


Figura 2 – Tempo de sobrevivência dos indivíduos com leucemia mieloide crônica que receberam transplante de medula óssea, estimada pelo método de Kaplan-Meier.

Pode-se observar pela Figura 2, que a função de sobrevivência do tempo, pode ser uma função imprópria, pois quando o tempo é assintótico, percebe-se que a função de sobrevivência não tende para zero. O que pode indicar uma existência de indivíduos curados nos dados observados.

Para verificar se a variável de sexo influencia no tempo de sobrevivência dos indivíduos com leucemia mieloide crônica que receberam transplante de medula óssea, foi feito um gráfico de Kaplan-Meier levando em consideração a covariável sexo.

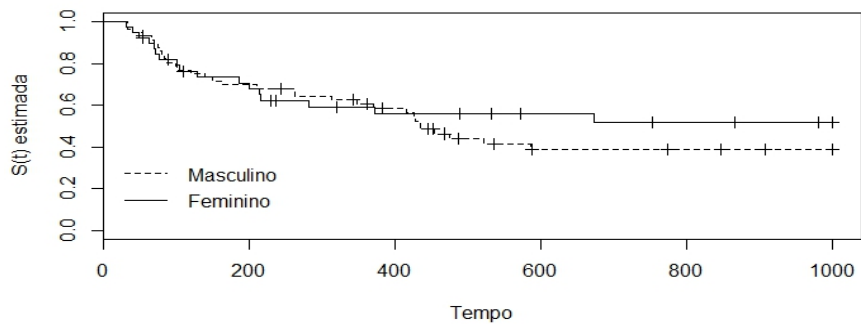


Figura 3 – Tempo de sobrevivência dos indivíduos com leucemia mieloide crônica que receberam transplante de medula óssea, nas categorias de sexo, estimado por Kaplan-Meier.

Pode-se perceber na Figura 3, que as curvas de sexo não se distanciam muito, chegando até mesmo a cruzar, ou seja, há indícios de que o sexo não influencia no tempo de sobrevivência dos indivíduos com leucemia mieloide crônica que receberam transplante de medula óssea, o que também pode ser observado na Figura 4.

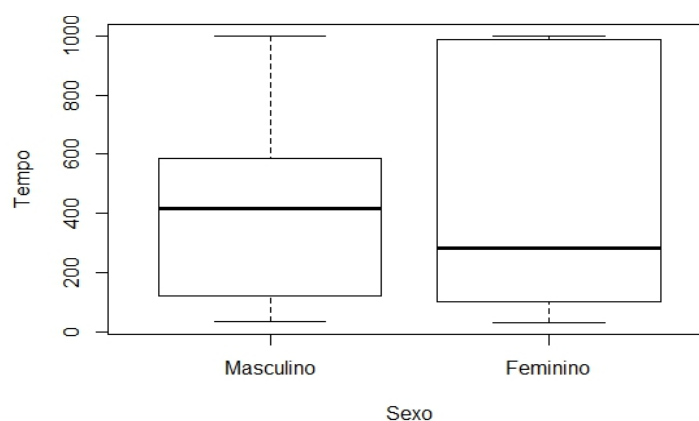


Figura 4 – Boxplot do Tempo de sobrevivência dos indivíduos com leucemia mieloide crônica que receberam transplante de medula óssea, considerando o sexo.

Para verificar se o grupo que teve ou não recuperação de plaquetas, influencia no tempo de sobrevivência dos indivíduos com leucemia mieloide crônica que receberam transplante de medula óssea, foi feito um gráfico de Kaplan-Meier levando em consideração a covariável plaq.

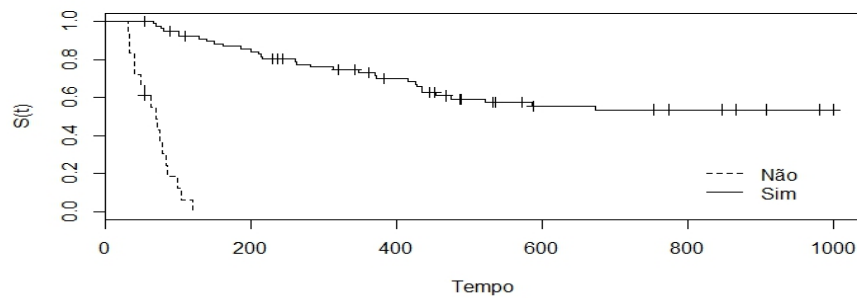


Figura 5 – Tempo de sobrevivência dos indivíduos com leucemia mieloide crônica que receberam transplante de medula óssea, estimado por Kaplan-Meier, considerando a existência de recuperação de plaquetas.

A Figura 5 mostra que o tempo de sobrevivência de quem teve recuperação de plaquetas é muito superior ao das pessoas que não tiveram essa recuperação. Como as curvas de Kaplan-Meier se diferenciam bastante, essa é uma possível covariável que influencia no tempo de sobrevivência das pessoas com leucemia mieloide crônica que receberam transplante de medula óssea. Essa influência também pode ser percebida analisando a Figura 6.

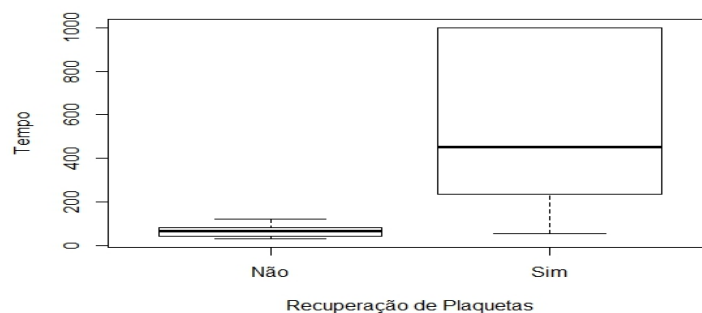


Figura 6 – Boxplot do Tempo de sobrevivência dos indivíduos com leucemia mieloide crônica que receberam transplante de medula óssea, considerando a recuperação de plaquetas.



Para verificar se o grupo que teve ou não a doença enxerto aguda, influencia no tempo de sobrevivência de quem tem leucemia mieloide crônica que receberam transplante de medula óssea, foi feito um gráfico de Kaplan-Meier levando em consideração a covariável deag.

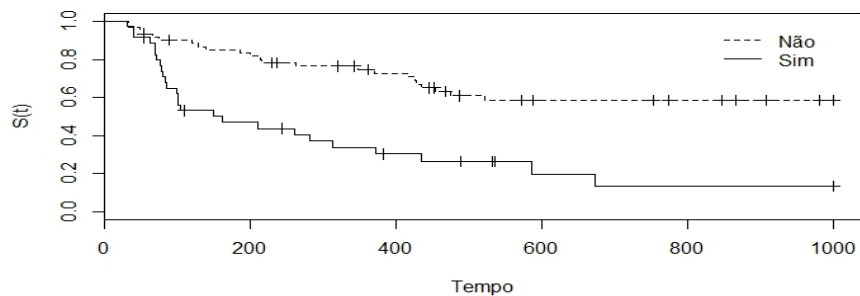


Figura 7 – Tempo de sobrevivência dos indivíduos com leucemia mieloide crônica que receberam transplante de medula óssea, estimado por Kaplan-Meier, considerando a existência de doença enxerto aguda.

A Figura 7 mostra que o tempo de sobrevivência de quem não teve doença enxerto aguda é superior ao das pessoas que tiveram essa doença. Como as curvas de Kaplan-Meier se diferenciam, essa é uma possível covariável que influencia no tempo de sobrevivência das pessoas com leucemia mieloide crônica. A influência pode ser percebida pela Figura 8 também.

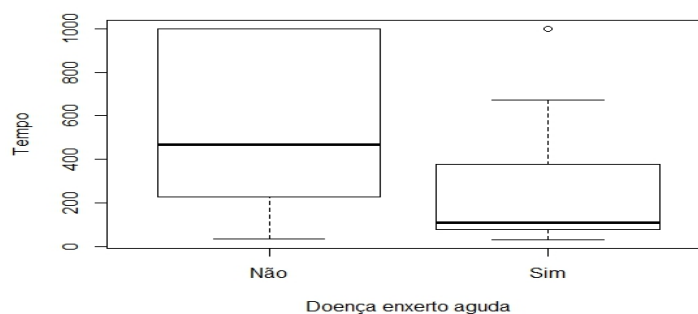


Figura 8 – Boxplot do Tempo de sobrevivência dos indivíduos com leucemia mieloide crônica que receberam transplante de medula óssea, considerando a ocorrência de doença enxerto aguda.

Para verificar se o grupo que teve ou não a doença enxerto crônica, influencia no tempo de sobrevivência dos indivíduos com leucemia mieloide crônica que receberam transplante de medula óssea, foi feito um gráfico de Kaplan-Meier levando em consideração a covariável *decr*.

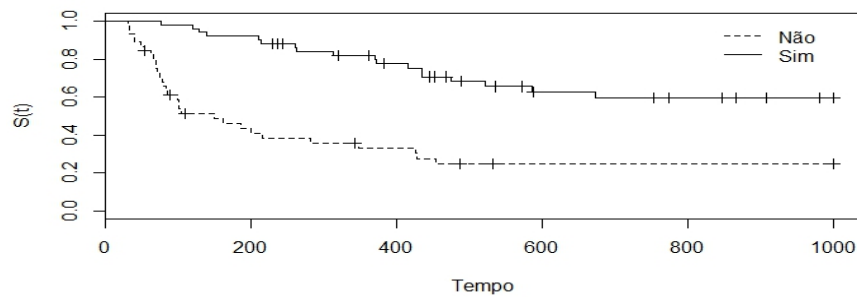


Figura 9 – Tempo de sobrevivência dos indivíduos com leucemia mieloide crônica que receberam transplante de medula óssea, estimado por Kaplan-Meier, considerando a existência de doença enxerto crônica.

A Figura 9 mostra que o tempo de sobrevivência de quem teve doença enxerto crônica é superior ao das pessoas que não tiveram essa doença. Como as curvas de Kaplan-Meier se diferenciam, essa é uma possível covariável que influencia no tempo de sobrevivência das pessoas com leucemia mieloide crônica. A influência pode ser percebida também na Figura 10.

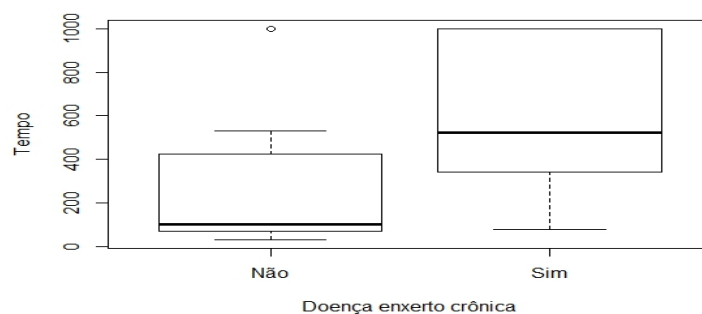


Figura 10 – Boxplot do Tempo de sobrevivência dos indivíduos com leucemia mieloide crônica que receberam transplante de medula óssea, considerando a ocorrência de doença enxerto crônica.

Para verificar o impacto da fase da doença no tempo de sobrevivência dos indivíduos com leucemia mieloide crônica que receberam transplante de medula óssea, foi feito um gráfico de Kaplan-Meier levando em consideração a covariável fase.

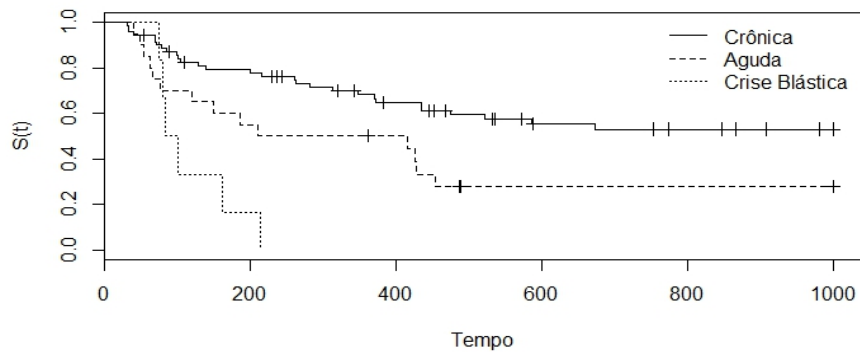


Figura 11 – Tempo de sobrevivência dos indivíduos com leucemia mieloide crônica que receberam transplante de medula óssea, estimado por Kaplan-Meier, considerando as fases.

A Figura 11 mostra que o tempo de sobrevivência de quem esteve na fase com crise blástica, diferencia bastante do tempo de sobrevivência dos que tiveram na fase crônica e aguda. Logo, há indícios que a covariável fase influencia no tempo de sobrevivência das pessoas com leucemia mieloide crônica. A influência também pode ser observada na Figura 12.

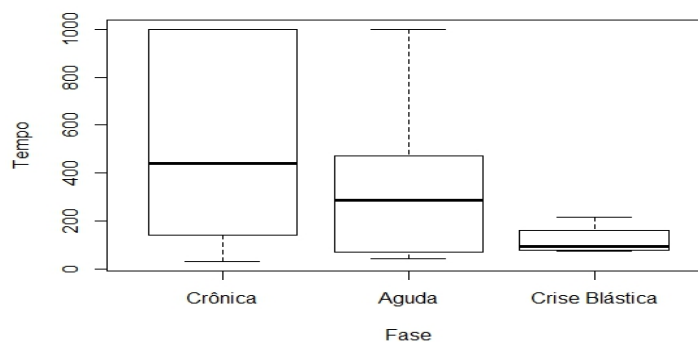


Figura 12 – Boxplot do Tempo de sobrevivência dos indivíduos com leucemia mieloide crônica que receberam transplante de medula óssea, considerando a fase da leucemia.

Como pode ser observado na Figura 2, a curva de Kaplan-Meier para os grupos de sexo não segue riscos proporcionais, ou seja, as curvas não são paralelas pois se cruzam em algum momento. Então decidiu-se utilizar o teste de Wilcoxon para verificar se o grupo de sexo tem influência no tempo de sobrevivência dos indivíduos com leucemia mieloide crônica.

O teste de Wilcoxon teve como hipótese:

$H_0$  : O sexo não influencia no tempo de sobrevivência

$H_1$  : O sexo influencia no tempo de sobrevivência

O Resultado do teste:

Tabela 1 – Resultado do teste de Wilcoxon para a variável sexo

Coefficiente Qui-Quadrado ( $\chi^2$ )	p-valor
0,2	0,673

Como o p-valor observado é maior que o nível de significância de 5%, não rejeita-se  $H_0$ , ou seja, há evidência para concluir que o grupo de sexo não influencia no tempo de sobrevivência dos indivíduos com leucemia mieloide crônica.

Como pode ser observado nas Figuras 4, 6, 8, 10, as curvas de sobrevivência estimada pelo método de Kaplan-Meier para as variáveis recuperação de plaquetas, ocorrência de doença enxerto aguda, ocorrência de doença enxerto crônica e fase da leucemia, segue riscos proporcionais, ou seja, as curvas de sobrevivência são paralelas pois não se cruzam. Então decidiu-se utilizar o teste de LogRank para verificar se essas variáveis tem influência no tempo de sobrevivência dos indivíduos com leucemia mieloide crônica que receberam transplante de medula óssea.

O Teste de LogRank teve como hipótese:

$H_0$  : Não existe diferença entre as curvas de sobrevivência

$H_1$  : Existe diferença entre as curvas de sobrevivência

O Resultado do teste:

Tabela 2 – Resultados do teste de LogRank para as variáveis recuperação de plaquetas, doença enxerto aguda, doença enxerto crônica e fases da leucemia

Variável	Estatística do Teste ( $\chi^2$ )	p-valor
Recuperação de Plaquetas	114	<0,0001
Doença enxerto aguda	18,8	0,0000145
Doença enxerto crônica	22,6	0,00000201
Fase	19,7	0,00000521

Como o p-valor observado é menor que o nível de significância de 5% em todos os casos, rejeita-se  $H_0$ , ou seja, há evidência para concluir que a recuperação de plaquetas, a doença enxerto aguda, a doença enxerto crônica e a fase da leucemia influenciam no tempo de sobrevivência dos indivíduos

Para avaliar as possíveis distribuições de probabilidade que ajustam aos dados dos indivíduos com leucemia mieloide crônica que receberam transplante de medula óssea, foi feita uma função taxa de falha empírica, o gráfico pode ser conhecido também como TTTplot.

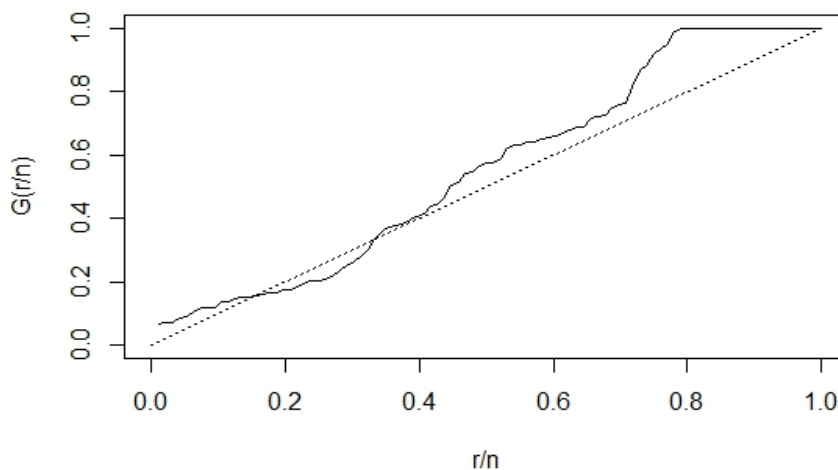


Figura 13 – Gráfico da função taxa de falha empírica (TTTplot) dos tempos de sobrevivência dos indivíduos com leucemia mieloide crônica que receberam o transplante de medula óssea.

Observando de uma forma geral a Figura 13, a curva TTT indica que possivelmente um bom ajuste aos dados será uma distribuição que assumi em sua função de risco um comportamento de banheira, porém considerando o início pode-se pensar também em uma função de risco unimodal. Sendo assim, possíveis distribuições a serem testadas são as distribuições da família Kumaraswamy e suas distribuições particulares.

## 5.2 Modelo de regressão Weibull

Feita a análise descritiva, deseja-se avaliar o melhor modelo que se adequa aos dados. Primeiramente, será proposto o modelo de regressão Weibull, para depois ser estudado mais profundamente o modelo de regressão Kumaraswamy-Weibull e verificar se realmente este, tem um efeito significativo para os dados de transplante de medula óssea.

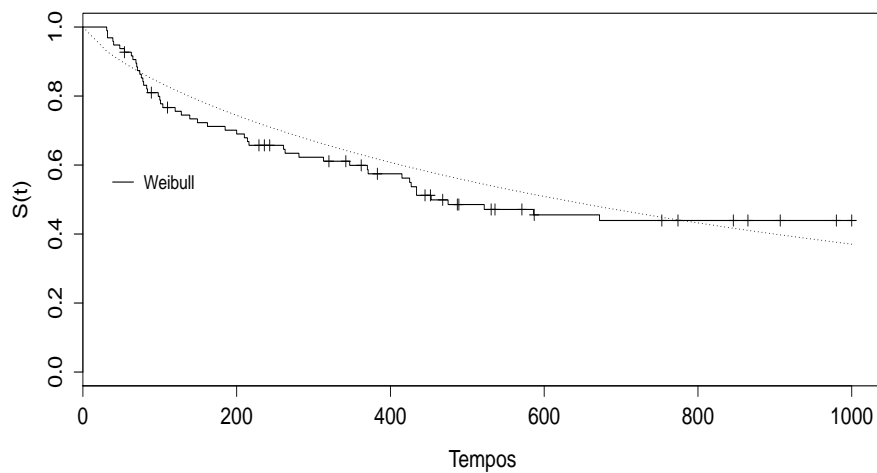


Figura 14 – Ajuste da curva de Kaplan-Meier e da curva de sobrevivência da distribuição Weibull estimados aos dados de transplante de medula óssea.

Na Figura 14, há indícios de que a distribuição Weibull é um bom ajuste para os dados de transplante de medula óssea. Assim o primeiro passo que será tomado, é a avaliação do modelo completo, ou seja, o modelo com todas as covariáveis.

Tabela 3 – Estimativas do modelo de regressão Weibull com todas covariáveis para os dados de transplante de medula óssea.

Parâmetros	Estimativas	Erro Padrão	Z	p-valor
$Log(\sigma)$	-0,9081	0,106949	-8,491	2,04E-17
$\beta_0$	3,84533	0,26645	14.431	3,290E-47
$\beta_1$ (Sexo masculino)	-0,0550	0,13019	-0,423	0,6720
$\beta_2$ (Teve recuperação de plaquetas)	0,61853	0,225134	2.747	0,0060
$\beta_3$ (Tempo de recuperação de plaquetas)	-0,0031	0,002817	-1.121	0,2620
$\beta_4$ (Teve doença enxerto aguda)	0,56666	0,162751	3.482	0,0005
$\beta_5$ (Tempo até a doença enxerto aguda)	0,00233	0,00052	4.479	7,480E-06
$\beta_6$ (Teve doença enxerto crônica)	0,97404	0,172237	5.655	1,560E-08
$\beta_7$ (Tempo até a doença enxerto crônica)	0,00246	0,000691	3.562	0,0004

A Tabela 3 mostra o modelo de regressão Weibull com todas as covariáveis. Pode ser observado que os parâmetros  $\beta_1$ ,  $\beta_3$  que acompanham as covariáveis sexo e tempo de recuperação de plaquetas, respectivamente, não são significativos. Sendo assim, essas covariáveis, serão retiradas do modelo para verificar se sem elas, ocorre um ajuste melhor.

A Tabela 4 mostra que todos os parâmetros são significativos para o modelo Weibull sem as covariáveis sexo e tempo de recuperação de plaquetas, considerando o nível de 5% de significância. Também foi feito o teste da razão de verossimilhança quando se tira as variáveis uma por uma, porém esse modelo continuou significativo e o mais apropriado. Conclui-se que as pessoas que tiveram recuperação de plaquetas, doença enxerto aguda e doença enxerto crônica tem uma probabilidade maior de sobreviver ao tempo  $t$ .

Tabela 4 – Estimativas do modelo de regressão Weibull sem as covariáveis sexo e tempo de recuperação de plaquetas para os dados de transplante de medula óssea.

Parâmetros	Estimativas	Erro Padrão	Z	p-valor
$Log(\sigma)$	-0,89973	0,105540	-8,52	1,53E-17
$\beta_0$	3,59627	0,143011	25,15	1,53E-139
$\beta_2$ (Plaquetas)	0,74584	0,192717	3,87	1,09E-04
$\beta_4$ (Doença enxerto aguda)	0,52992	0,161682	3,28	1,05E-03
$\beta_5$ (Tempo até a doença enxerto aguda)	0,00225	0,000514	4,37	1,22E-05
$\beta_6$ (Doença enxerto crônica)	0,97546	0,175972	5,655	2,97E-08
$\beta_7$ (Tempo até a doença enxerto crônica)	0,00230	0,000706	3,25	1,15E-03

Para analisar globalmente se o modelo Weibull sem as covariáveis sexo e tempo de recuperação de plaquetas, é um bom modelo para os dados de transplante de medula óssea, foi realizado uma análise de resíduos considerando os resíduos: Cox-Snell, Martingal e Deviance.

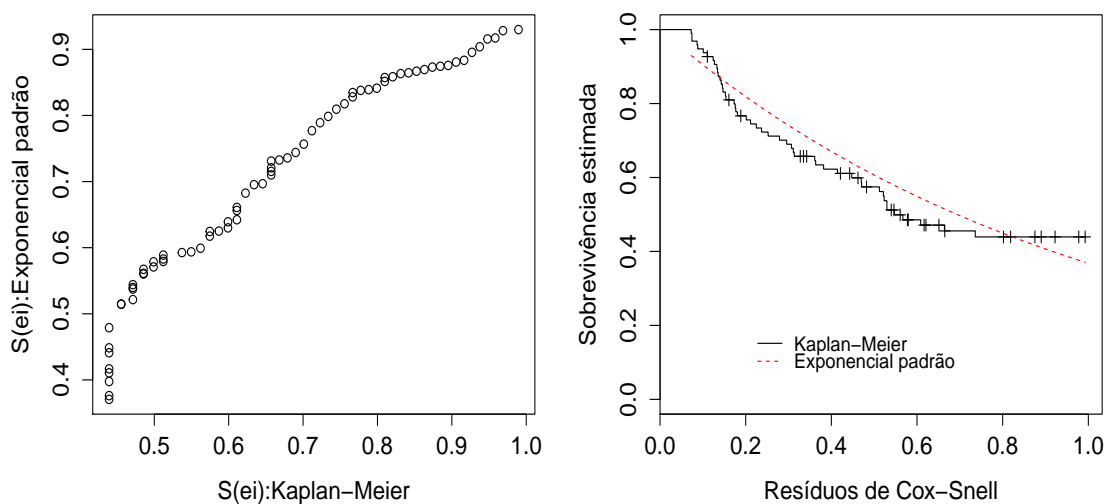


Figura 15 – Resíduos de Cox-Snell para a distribuição Weibull.



Na Figura 15 pode-se observar que, o gráfico de  $S(\hat{e}_i)$  tem um grande desvio de uma reta, ou seja, está mostrando um ajuste razoável, porém, o gráfico de resíduos de Cox-Snell mostra que o ajuste não está tão bom assim. Dessa forma, pelos gráficos da Figura 15 há indícios de que o modelo de regressão Weibull não possui um bom ajuste para os dados de indivíduos que receberam transplante de medula óssea.

O resíduo de Martingal e de Deviance será outra forma para verificar se os dados de transplante de medula óssea estão bem ajustados pelo modelo escolhido e também para verificar a existência de observações atípicas, ou seja, que sejam observações influentes.

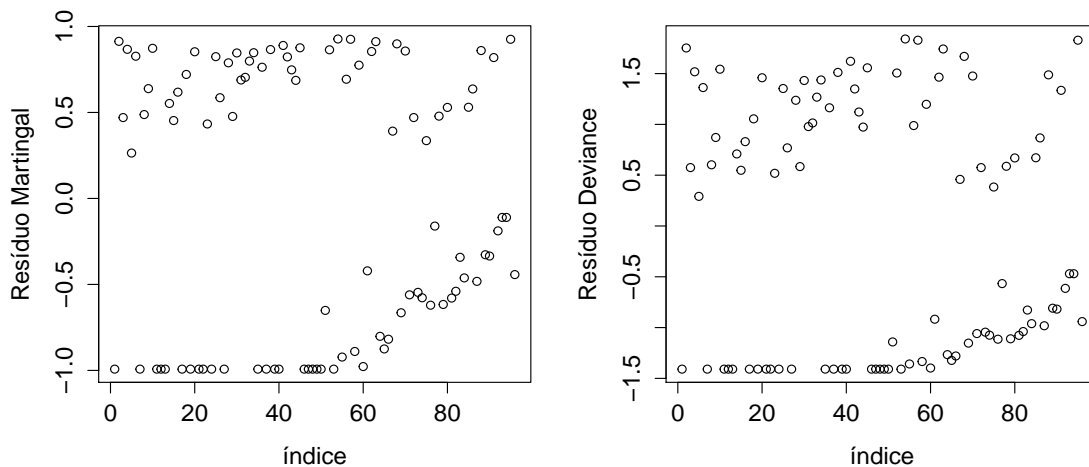


Figura 16 – Resíduos de martingal e de deviance para a distribuição Weibull.

A Figura 16 mostra que os resíduos parecem ter um comportamento aleatório, exceto pelas observações que no gráfico de resíduos de deviance formam uma linha no valor -1,5 que são observações censuradas e por isso essas observações não tem um comportamento tão aleatório.

Após a análise dos resíduos é possível perceber que o Modelo de Regressão Weibull não tem um ajuste adequado aos dados e que há indícios de que um modelo com fração de cura deve ser usado para analisar esses dados.

### 5.3 Modelo Kumaraswamy-Weibull com fração de cura

Após a observação de que o modelo de regressão Weibull não foi o mais adequado aos dados de transplante de medula óssea. Pretende-se investigar distribuições que modelem melhor esses dados. Sendo assim, foi proposto o modelo de regressão Kumaraswamy-Weibull e o modelo Kumaraswamy-Weibull com fração de cura. Nesses modelos, foi utilizado a função `constrOptim` do software R (R DEVELOPMENT CORE TEAM, 2012) para a estimação dos parâmetros e ainda não foi encontrado o ponto de máximo global, mas os resultados encontrados até o momento serão mostrados nas Tabelas 5 e 6 e Figura 18 para discutimos os problemas encontrados e também verificar que há um grande indicativo ao uso dos modelos Kumaraswamy Weibull e Kumaraswamy Weibull com fração de cura para os dados de indivíduos que receberam transplante de medula óssea.

Tabela 5 – Estimativas de máxima verossimilhança para o modelo Kumaraswamy-Weibull para os dados de transplante de medula óssea.

Parâmetro	Estimativas	Erro Padrão
$\alpha$	0,3474615	0,6387223
$\gamma$	0,3441323	0,12863048
$a$	178,12625	197,977586
$b$	0,0906400	0,06689261

As estimativas dos parâmetros do modelo Kumaraswamy-Weibull estão sendo apresentados na Tabela 5. Pode-se observar que o erro padrão para as estimativas dos parâmetros  $\alpha$  e  $a$  são grandes. Pela estimação usando função `constrOptim` do software R (R DEVELOPMENT CORE TEAM, 2012), foi possível verificar que os parâmetros estavam instáveis, ou seja, dependendo dos valores iniciais colocados na função `constrOptim`, os resultados mudavam. Apesar desse problema o valor da verossimilhança também mudava o que é um indicativo para eliminar o problema de indentificabilidade do modelo.

Tabela 6 – Estimativas de máxima verossimilhança para os modelos Kumaraswamy-Weibull com fração de cura

Parâmetro	Estimativas	Erro Padrão
$\alpha$	5302,4641	6856,435
$\gamma$	2,2616103	12,25521
$a$	0,5422988	2,944719
$b$	41,685130	67,74901
$\phi$	0,5628449	0,05570117

Na Tabela 6 estão às estimativas dos parâmetros do modelo Kumaraswamy-Weibull com fração de cura. Observa-se que em relação ao modelo Kumaraswamy-Weibull que o valor dos parâmetros variaram, ou seja, outro indicativo de que a estimação ainda está instável. Porém o trabalho continuará para achar o máximo global, e dar valores mais preciso das estimativas dos parâmetros. Percebe-se também que a fração de cura  $(1 - \phi)$ , ficou em volta do valor 0,437, o que já era esperado com base na função de sobrevivência estimado pelo o método de Kaplan-Meier.

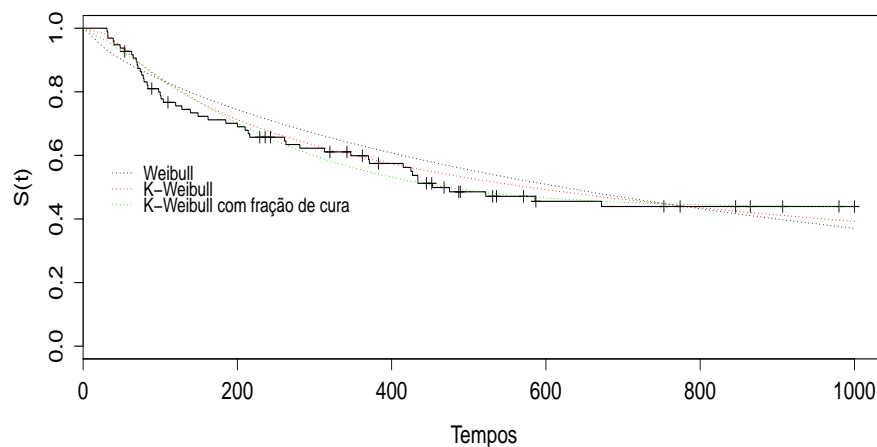


Figura 17 – Ajuste da curva de Kaplan-Meier e da curva de sobrevivência da distribuição Weibull, K-W e K-W com fração de cura estimados aos dados de transplante de medula óssea.

Apesar dos parâmetros ainda estarem instáveis e ainda com os resultados de um ponto de máximo local, já é possível perceber na Figura 18 um grande indicativo de que quando os dados de transplante de medula óssea são modelados pela distribuição Kumaraswamy-Weibull, há uma melhora em relação a modelagem com distribuição Weibull. E quando os dados são modelados com a distribuição Kumnnaraswmay-Weibull com fração de cura, há uma grande melhora em relação aos dois modelos anteriores.



## 6 CONSIDERAÇÕES FINAIS

A partir do estudo feito e de todos os resultados encontrados, observou-se que considerando o modelo Weibull o melhor ajuste foi desconsiderando as covariáveis sexo e tempo até a recuperação de plaquetas, porém o modelo Weibull não teve um bom ajuste aos dados de transplante de medula óssea como mostrou a análise dos resíduos.

Como esperado uma distribuição que apresenta em sua função de risco formas unimodal e de U ou também conhecida como forma de banheira, mostrou indicativos de modelar melhor os dados, essa distribuição chamada Kumaraswamy-Weibull.

Há também um grande indicativo de que realmente os dados de transplante de medula óssea tenha a presença de indivíduos curados, ou seja, que não são mais suscetíveis ao evento de interesse. Sendo assim, há um grande indicativo de que o melhor modelo para esses dados é o Kumaraswamy Weibull com fração de cura.

Porém ainda não foi possível finalizar a estimação dos parâmetros dos modelos Kumaraswamy Weibull e Kumaraswamy Weibull com fração de cura devido a instabilidade computacional e tempo hábil para a conclusão deste trabalho. Mas é importante ressaltar que haverá uma continuidade deste trabalho, pois o objetivo final é submeter um artigo.



## REFERÊNCIAS

- AASET, M.V. How to identify bathtub hazard rate. **IEEE Transactions on Reliability**, New York, v.36, p.106-108, 1987.
- BERKSON, J.; GAGE, R.P. Survival curve for cancer patients following treatment. **Journal of the American Statistical Association**, v.47, p. 501-515, 1952.
- COLOSIMO, E.A.; GIOLO, S.R. **Análise de Sobrevivência Aplicada**, São Paulo: Blucher, 2006. 392p.
- CORDEIRO G.M ; Castro M. A new Family of generalized distributions, **Journal of Statistical Computation and Simulation**, New York, v.0, p. 1-17, 2009.
- FACHINI, J.B, **Modelos de regressão com e sem fração de cura para dados bivariados em análise de sobrevivência**. 2011. 140p. Tese (Doutorado em Estatística e Experimentação Agrônômica), Escola Superior de Agricultura "Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2011.
- FARAGE, C.G.; **Análise de Diagnóstico para o Modelo de Regressão de Cox**. 2013. 67p. Monografia (Graduação em Estatística), Universidade de Brasília, 2013.
- GUPTA, R.D.; KUNDU, D. Exponentiated exponential family: An alternative to Gamma and Weibull distributions, **Biometrical Journal**, Berlin, v.43, p. 117-130, 2011.
- IBRAHIM, J. G.; CHEN, M. H.; SINHA, D. **Bayesian survival analysis**, New York: Springer, p.479, 2001.
- KALBFLEISCH, J.D.; PRENTICE, R.L. **The statistical analysis of failure time data**. 2nd ed. New York: John Wiley, 2002. 439p.
- KLEIN, J.P.; MOESCHBERGER, M. L. **Survival analysis: techniques for censored and truncated data**. New York: Springer Verlag, 1997. 536 p.
- LAI C.D., XIE M.;MURTHY, D.N.P. A modified Weibull Distribution. **IEEE Transactions on Reliability**, New York, v.52, p.33-37, 2003.
- LANGE, K. **Numerical analysis dor statisticians**. New York: Springer, 1999. 356 p.
- LAWLESS, J.F. **Statistical models and methods for lifetime data**, 2nd ed., New York: Wiley, 2003, 439 p.
- MALLER.; ZHOU, X. **Survival Analysis with long-Term Survivors**, John Wiley Sons, 1996.
- MUDHOLKAR, G.S.; SRVASTAVA, D.K.; Exponentiated Weibull Family for Analyzing Bathtub Failure-Rate Data, **IEEE Transactions on Reliability**, New York, v.42, p.229-302, 1993.
- MUDHOLKAR, G.S.; SRVASTAVA, D.K.; FREIMER, M. The exponentiated Weibull



Family a reanalysis of the bus-motor-failure data, **Technometrics**, v.37 p.436-445, 1995.

MUDHOLKAR, G.S.; SRVASTAVA, D.K.; KOLLIA, G.D. A generalization of the Weibull Distribution with Application to the Analysis of survival Data, Journal of the American Statistical Association, Boston, v.91 p.1575-1583, 1996.

MURTHY, D.N.P; XIE, M.; JIANG, R. Weibull models, wiley series in probability and statistics, New Jersey, Canada, 2004.

JONES, M.C. Kumaraswamy's distribution: A beta-type distribution with some tractability advantages. **Statistical Methodology**, v.6 p. 70-81, 2009.

LEAL, C.; **Modelo de regressão log-linear com fração de cura**. 2013. 46p. Monografia (Graduação em Estatística), Universidade de Brasília, 2013.

SANTOS, T.A.; **Modelo de regressão pertence à família Weibull com fração de cura**. 2013. 50p. Monografia (Graduação em Estatística), Universidade de Brasília, 2013.

R Core Team (2012). **R**: A language and environment for statistical computing, Vienna, Austria, Disponível em: <http://www.R-project.org/>

ANEXOS



## Anexo A: Programação

```

##Entrando com os dados dados<-read.table(file.choose(),sep="," ,dec="," ,
header=T)

##Olhando o Cabeçario dos dados head(dados)
##Olhando quais são as variáveis
names(dados)
id<-dados$id
sexo<-dados$sexo
idade<-dados$idade
status<-dados$status
tempo<-dados$tempo
plaq<-dados$plaq
tempplaq<-dados$tempplaq
deag<-dados$deag
tempdeag<-dados$tempdeag
decr<-dados$decr
tempdecr<-dados$tempdecr
fase<-dados$fase
censura<-dados$status
## Carregando Pacotes (require = library)
require(splines)
require (survival)
##Fazendo Análise Descritiva "Univariada" dos dados ##

## Comando para criar estatísticas básicas
require(fBasics)

# function basicStats(tempo)

#### Analisando Variáveis Quantitativas

```

```
##Analisando Tempo
plot(tempo)
hist(tempo)
boxplot(tempo)
summary(tempo sexo)
basicStats(tempo sexo)

##Analisando Idade
plot(idade, xlab=, ylab="Idade")
hist(idade, xlab="Idade", ylab="Frequência", main="Histograma de Idade")
boxplot(idade, xlab="Idade", ylab="Frequência")
summary(idade)
basicStats(idade)

##Analisando tempplaq
plot(tempplaq)
hist(tempplaq)
boxplot(tempplaq)
summary(tempplaq)
basicStats(tempplaq)

##Analisando tempdeag
plot(tempdeag)
hist(tempdeag)
boxplot(tempdeag)
summary(tempdeag)
basicStats(tempdeag)

##Analisando tempdecr
plot(tempdecr)
```

```
hist(tempdecr)
boxplot(tempdecr)
summary(tempdecr)
basicStats(tempdecr)

#####Analisando Variáveis qualitativas

##Analisando Status(Censuras)
a <- table(status)
pie(a)

##Analisando Plaq
b<-table(plaq)
pie(b)

##Analisando deag
c<-table(deag)
pie(c)

##Analisando decr
d<-table(decr,deag)
d

##Analisando fase
e<-table(fase)
pie(e)

##Analisando Sexo
f <- table(sexo)
pie(f)
```

```
##Fazendo Análise Descritiva "Bivariada"dos dados ##
```

```
##### Boxplot com o Tempo #####
```

```
##### Tempo #####
```

```
boxplot(tempo, xlab="Tempo", ylab="Frequência")
```

```
##### Tempo por sexo #####
```

```
boxplot(tempo sexo, xlab="Sexo", ylab="Tempo", na-
mes=c("Masculino", "Feminino"))
```

```
##### Tempo por recuperação de plaquetas #####
```

```
boxplot(tempo plaq, xlab="Recuperação de Plaquetas", ylab="Tempo", na-
mes=c("Não", "Sim"))
```

```
##### Tempo por Deag #####
```

```
boxplot(tempo deag, xlab="Doença enxerto aguda", ylab="Tempo", na-
mes=c("Não", "Sim"))
```

```
##### Tempo por Decr #####
```

```
boxplot(tempo decr, xlab="Doença enxerto crônica", ylab="Tempo", na-
mes=c("Não", "Sim"))
```

```
##### Tempo por fase #####
```

```
boxplot(tempo fase, xlab="Fase", ylab="Tempo", na-
mes=c("Crônica", "Aguda", "Crise Blástica"))
```

```
##### Lendo a Biblioteca Hmisc para fazer #as correlações
```

```
require(Hmisc)
```

```
##### Analisando as Quantitativas #####
```

```
plot(idade tempo,main="Gráfico de Tempo por Idade",
xlab="Tempo",ylab="Idade")
```

```
plot(tempplaq tempo,main="Gráfico de Tempo por Plaquetas",
xlab="Tempo",ylab="Tempo de Plaquetas")
```

```
plot(tempdeag tempo,main="Gráfico de Tempo por Tempo de ocorrência en-
xerto agúda", xlab="Tempo",ylab="Tempo de ocorrência enxerto agúda")
```

```
plot(tempdecr tempo,main="Gráfico de Tempo por Tempo de Ocorrência en-
xerto crônica", xlab="Tempo",ylab="Tempo de Ocorrência enxerto crônica")
```

```
##### Tempo por Decr #####
```

```
plot(tempdecr tempdeag, xlab="Doença enxerto aguda", ylab="Doença en-
xerto crônica")
```

```
##### Análise de correlação das variáveis #Quantitativas #####
```

```
rcorr(tempo, idade, type=c("pearson"))
```

```
rcorr(tempo, tempplaq, type=c("pearson"))
```

```
rcorr(tempo, tempdeag, type=c("pearson"))
```

```
rcorr(tempo, tempdecr, type=c("pearson"))
```

```
rcorr(idade, tempplaq, type=c("pearson"))
```

```
rcorr(idade, tempdeag, type=c("pearson"))
```

```
rcorr(idade, tempdecr, type=c("pearson"))
```

```
rcorr(tempplaq, tempdeag, type=c("pearson"))
```

```
rcorr(tempplaq, tempdecr, type=c("pearson"))
```

```
rcorr(tempdeag, tempdecr, type=c("pearson"))
```

```
cbind(tempdeag,deag,tempdecr,decr)
```

```
##### Fazendo Análise Descritiva "Kaplan-#Meier" dos dados #####
```



```

#####Análise sem grupos#####
kmsexo<-survfit(Surv(tempo,status) 1)
Kmsexo
summary(kmsexo)
plot(kmsexo, conf.int=F, lty=c(2,1) ,xlab="Tempo (em dias)", ylab="S(t)")

#####Análise para a o grupo de Sexo#####
kmsexo<-survfit(Surv(tempo,status) sexo)
kmsexo
summary(kmsexo)
plot(kmsexo, conf.int=F, lty=c(2,1) ,xlab="Tempo", ylab="S(t) estimada")
legend(1,0.4,lty=c(2,1), c("Masculino", "Feminino"),lwd=1,bty="n")

#####Análise para a o grupo de Plaq (Recuperação de Plaquetas)#####
kmplaq<-survfit(Surv(tempo,status) plaq)
kmplaq
summary(kmplaq)
plot(kmplaq, conf.int=F, lty=c(2,1) ,xlab="Tempo", ylab="S(t)")
legend(800,0.3,lty=c(2,1), c("Não", "Sim"),lwd=1,bty="n")

##### Análise para a o grupo de deag (Doença exerto agúda) #####
kmdeag<-survfit(Surv(tempo,status) deag)
kmdeag
summary(kmdeag)
plot(kmdeag, conf.int=F, lty=c(2,1) ,xlab="Tempo", ylab="S(t)")
legend(825,1,lty=c(2,1), c("Não", "Sim"),lwd=1,bty="n")

##### Análise para a o grupo de decr (Doença #exerto crônica) #####
kmdecr<-survfit(Surv(tempo,status) decr)
kmdecr

```

```

summary(kmdecr)
plot(kmdecr, conf.int=F, lty=c(2,1) ,xlab="Tempo", ylab="S(t)")
legend(825,1.02,lty=c(2,1), c("Não", "Sim"),lwd=1,bty="n")

##### Análise para a o grupo de fase (fase da ##doença) #####
kmfase<-survfit(Surv(tempo,status) fase)
kmfase
summary(kmfase)
plot(kmfase, conf.int=F, lty=c(1,2,3) ,xlab="Tempo", ylab="S(t)")
legend(750,1.04,lty=c(1,2,3), c("Crônica", "Aguda", "Crise
tica"),lwd=1,bty="n")

##### Coparando as cruvas de Sobrevivência #####

#H0: Não há influência no tempo - as #curvas são iguais
#H1: Há influência no tempo - pelo menos #uma curva é diferente

##### Aplicando o Teste de Parecido com #Wilcoxon para sexo#####
difere1 <-survdif(Surv(tempo,status) sexo, rho=1)

# rho=1 considera o Teste parecido com #Willcoxondifere1

##### Aplicando o Teste de LogRank para #Plaquetas#####
difere2 <-survdif(Surv(tempo,status) plaq, rho=0)

# rho=0 considera o Teste de LogRank
difere2
names(difere2)

RR<- (difere2obs[1]/diferere2exp[1])/(difere2obs[2]/diferere2exp[2])

```

RR

```
##### Aplicando o Teste de LogRank para #Doença enxerto aguda#####
difere3 <-survdif(Surv(tempo,status) deag, rho=0) # rho=0 considera o Teste
de LogRank
```

difere3

names(difere3)

RR<-(difere3obs[1]/difere3exp[1])/(difere3obs[2]/difere3exp[2])

RR

```
##### Aplicando o Teste de LogRank para Doença enxerto crônica#####
difere4 <-survdif(Surv(tempo,status) decr, rho=0)
```

# rho=0 considera o Teste de LogRank

difere4

names(difere4)

RR<-(difere4obs[1]/difere4exp[1])/(difere4obs[2]/difere4exp[2])

RR

```
##### Aplicando o Teste de LogRank para fase#####
```

difere5 <-survdif(Surv(tempo,status) fase, rho=0)

# rho=0 considera o Teste de LogRank

difere5

```
names(difere5)
```

```
RR<-(difere5obs[1]/difere5exp[1])/(difere5obs[2]/difere5exp[2])
```

```
RR
```

```
##### Como a Fase tem 3 categorias iremos #comparar elas 2 a 2 #####
```

```
##### Comparando fase 1 e fase 2 #####
```

```
tempo12<-subset(tempo, fase==1|fase==2)
```

```
cens12<-subset(status, fase==1|fase==2)
```

```
fase12<-subset(fase, fase==1|fase==2)
```

```
difere5.1<-survdiff(Surv(tempo12,cens12) fase12,rho=0)
```

```
difere5.1
```

```
RR<-(difere5.1obs[1]/difere5.1exp[1])/(difere5.1obs[2]/difere5.1exp[2])
```

```
RR
```

```
##### Comparando fase 1 e fase 3 #####
```

```
tempo13<-subset(tempo, fase==1|fase==3)
```

```
cens13<-subset(status, fase==1|fase==3)
```

```
fase13<-subset(fase, fase==1|fase==3)
```

```
difere5.2<-survdiff(Surv(tempo13,cens13) fase13,rho=0)
```

```
difere5.2
```

```
RR<-(difere5.2obs[1]/difere5.2exp[1])/(difere5.2obs[2]/difere5.2exp[2])
```

RR

```
##### Comparando fase 2 e fase 3 #####
```

```
tempo23<-subset(tempo, fase==2|fase==3)
```

```
cens23<-subset(status, fase==2|fase==3)
```

```
fase23<-subset(fase, fase==2|fase==3)
```

```
difere5.3<-survdif(Surv(tempo23,cens23) fase23,rho=0)
```

```
difere5.3
```

```
RR<-(difere5.3obs[1]/difere5.3exp[1])/(difere5.3obs[2]/difere5.3exp[2])
```

RR

```
##### Separando as variáveis para fazer o #TTTplot #####
```

```
require(AdequacyModel)
```

```
##Separando os sexos - e - Fazendo os #TTTplots
```

```
smax<-subset(tempo, sexo == 1)
```

```
sfem<-subset(tempo, sexo == 2)
```

```
par(mfrow=c(1,2))
```

```
TTT(smax,col="red",lwd=2.5,grid=T,lty=2)

TTT(sfem,col="red",lwd=2.5,grid=T,lty=2)

##Separando os Recuperação de plaquetas

pnao<-subset(tempo, plaq==0)

psim<-subset(tempo, plaq==1)

par(mfrow=c(1,2))

TTT(pnao,col="red",lwd=2.5,grid=T,lty=2)

TTT(psim,col="red",lwd=2.5,grid=T,lty=2)

##Separando os Deag - Doença enxerto ##aguda

deagnao<-subset(tempo, deag == 0)

deagsim<-subset(tempo, deag == 1)

par(mfrow=c(1,2))

TTT(deagnao,col="red",lwd=2.5,grid=T,lty=2)

TTT(deagsim,col="red",lwd=2.5,grid=T,lty=2)

##Separando os Decr - Doença enxerto crônica
```

```
decrnao<-subset(tempo, decr == 0)

decrsim<-subset(tempo, decr == 1)

par(mfrow=c(1,2))

TTT(decrnao,col="red",lwd=2.5,grid=T,lty=2)

TTT(decrsim,col="red",lwd=2.5,grid=T,lty=2)

##Separando os Fase

fasecr<-subset(tempo, fase == 1)

faseag<-subset(tempo, fase == 2)

fasecb<-subset(tempo, fase == 3)

par(mfrow=c(1,3))

TTT(fasecr,col="red",lwd=2.5,grid=T,lty=2)

TTT(faseag,col="red",lwd=2.5,grid=T,lty=2)

TTT(fasecb,col="red",lwd=2.5,grid=T,lty=2)

#TTTplot#

n<-length(tempo)
```

```

stemp<-sort(tempo)

G<-numeric()

for(i in 1:n) G[i]<-(sum(stemp[1:i])+(n-i)*stemp[i])/(sum(stemp))

A<-(1:n)/n

plot(A,G,type="l",xlim=c(0,1),ylim=c(0,1),xlab="r/n",ylab="G(r/n)")

points(c(0,1),c(0,1),type="l",lty=3)

##### #Distribuição Weibull -
survreg #####
##### Dist. Weibull #####

require(survival)

mwe<-survreg(Surv(tempo,censura) 1, dist="weibull")

mwe

alphaww<-exp(mwe$coefficients[1])

alphaww

gammaww<-1/mwe$scale

gammaww

```



```

##### Entrando com as covariáveis - Usando #o Modelo Weibull #####

#Modelo de Regressão - Weibull#

y<-log(tempo)

n<-length(y)

mregwei<-survreg(Surv(y,censura) 1, dist='extreme')

summary(mregwei)

verow<-mregwei$loglik[1]

verow

pw <- 3

AICw <- (-2*verow)+(2*pw)

AICw

AICcw<-AICw+((2*pw+(pw+1))/(n-pw-1))

AICcw

mregwei1<-survreg(Surv(y,censura) sexo+plaq+tempplaq+deag+tempdeag+decr+tempdecr
dist='extreme')

```

```
summary(mregwei1)
```

```
sigmamreg1<-1/mregwei1$scale
```

```
sigmamreg1
```

```
mregwei2<-survreg(Surv(y,censura)      plaq+tempplaq+deag+tempdeag
+decr+tempdecr, dist='extreme')
```

```
summary(mregwei2)
```

```
mregwei3<-survreg(Surv(y,censura)  plaq+deag+tempdeag+decr+tempdecr,
dist='extreme')
```

```
summary(mregwei3)
```

```
alphaww<-exp(mwe$coefficients[1])
```

```
alphaww
```

```
gammaww<-1/mwe$scale
```

```
gammaww
```

```
mregwei4<-survreg(Surv(y,censura)      plaq+deag+tempdeag+decr,
dist='extreme')
```

```
summary(mregwei4)
```

```
mregwei5<-survreg(Surv(y,censura) plaq+deag+decr, dist='extreme')

summary(mregwei5)

mregwei6<-survreg(Surv(y,censura) tempplaq+tempdeag+tempdecr,
dist='extreme') summary(mregwei6)

#### Análisiso as variáveis #individualmente ####

mregwei7<-survreg(Surv(y,censura) plaq, dist='extreme')

summary(mregwei7)

mregwei8<-survreg(Surv(y,censura) tempplaq, dist='extreme')

summary(mregwei8)

mregwei9<-survreg(Surv(y,censura) deag, dist='extreme')

summary(mregwei9)

mregwei10<-survreg(Surv(y,censura) tempdeag, dist='extreme')

summary(mregwei10)

mregwei11<-survreg(Surv(y,censura) decr, dist='extreme')

summary(mregwei11)
```

```

mregwei12<-survreg(Surv(y,censura) tempdecr, dist='extreme')

summary(mregwei12)

##### Resultado - Modelo sem ##Tempplaq #####

mregwei3<-survreg(Surv(y,censura)   plaq+deag+tempdeag+decr+tempdecr,
dist='extreme')
summary(mregwei3)
##### Estimativas de Sobrevivência ##usando kaplan-Meier, e os
##modelos#####

km<-survfit(Surv(tempo,censura) 1)

time <- km$time ### Tempo de KM ####

time

names(km)

skm<-km$surv ### Sobrevivência de Kaplan-Meier ###

skm

swe<-exp((time/alphaww)gammaww)

swe

##### Seleção de Modelos #####

```

```

#####par(mfrow=c(1,3)) ##### Separar os #gráficos

plot(km,conf.int=F,xlab="Tempos", ylab="S(t)")

lines(c(0,time),c(1,swe),lty=3,col=1)

legend(25,0.6,lty=1,col=1,c("Weibull"),bty="n",cex=0.8)

##### Análise de Sensibilidade #####

##### Á Fazer #####

### Análise de Resíduo ###

#-Weibull

#Cox-Snell

xxw<-((y-mregwei$coeff[1])/mregwei$scale)

Hw<-exp(xxw)

#função risco acumulado - resíduo de #CoxSnell

KMew<-survfit(Surv(Hw,censura) 1,conf.int=F)

te<-KMew$time

ste<-KMew$surv

```

```

sexpw<-exp(-te)

par(mfrow=c(1,2))

plot(ste,sexpw, xlab="S(ei):Kaplan-Meier", ylab="S(ei):Exponencial padrão")

plot(KMew,conf.int=F, xlab="Resíduos de Cox-Snell", ylab="Sobrevivência
estimada")

lines(te,sexpw,lty=2, col=2)
legend(0.2,0.2,lty=c(1,2), col=c(1,2),c("Kaplan-Meier", "Exponencial pa-
drão"), cex=0.8, bty="n")

#Resíduo martingal

martingalw<-censura-Hw
par(mfrow=c(1,1))
indice<-c(1:n)

plot(indice,martingalw,xlab="índice", ylab="Resíduo Martingal")

#Resíduo Deviance

devw<-((martingalw/abs(martingalw))*(-2*(martingalw+censura*log(censura-
martingalw))))^(1/2)

plot(indice,devw,xlab="índice", ylab="Resíduo Deviance")
##### Indentificando os pontos perto de -1,5 #####

#identify(indice,devw)

```

```
#1 7 11 12 13 17 19 21 22 24 27 35
```

```
#37 39 40 46 48 49 50 51 55 58 60 61
```

```
#64 65 66 69 71 73 74 76 79 77 83 82
```

```
#81 87 89 90 92 93 94 96
```

```
### Olhando se são observações #censuradas ###
```

```
head(devw)
```

```
devw
```

```
dados$devw<-devw
```

```
head(dados)
```

```
Rnale<-subset(dados,dados$status==0)
```

```
devwnale<-Rnale$devw
```

```
idnale<-Rnale$id
```

```
statusnale<-Rnale$status
```

```
cbind(idnale,devwnale,statusnale)
```

```
# Resultado: OK! São observações censuradas.
```

```
##### Distribuição Kumarasswamy-Weibull com fração de cura ,
```

Programação da Juliana #####

### Verossimilhança ###

```
verokw<-function(para) alfa<-para[1] gamma<-para[2] aa<-para[3] bb<-
para[4] ftt<-aabb(gamma/(alfa*gamma))(tempo^(gamma-1))*exp(-(tempo/alfa)*gamma) ((1-
exp(-(tempo/alfa)*gamma))^(aa-1))*((1-(1-exp(-(tempo/alfa)*gamma))^aa)^(bb-1))
stt<-((1-(1-exp(-(tempo/alfa)*gamma))^aa)^(bb))
ll<-censuralog(ftt)+(1-censura)log(stt)
sum(-ll)
```

```
gradkw<-function(para) alfa<-para[1] gamma<-para[2] aa<-para[3] bb<-
para[4]
```

```
derialf<-sum(censura (((aa bb (gamma/(alfa*gamma)) (tempo^(gamma -
1)) (exp(-(tempo/alfa)*gamma) ((tempo/alfa)^(gamma - 1) (gamma (tempo/alfa^2))))
- aa bb (gamma (alfa^(gamma - 1) gamma)/(alfa*gamma)^2) (tempo^(gamma -
1)) exp(-(tempo/alfa)*gamma)) ((1 - exp(-(tempo/alfa)*gamma))^(aa - 1)) - aa ×
bb (gamma/(alfa*gamma)) (tempo^(gamma - 1)) exp(-(tempo/alfa)*gamma) ((1
- exp(-(tempo/alfa)*gamma))^(aa - 1) - 1) ((aa - 1) (exp(-(tempo/alfa)*gamma)
((tempo/alfa)^(gamma - 1) (gamma (tempo/alfa^2)))))) ((1 - (1 - exp(-
(tempo/alfa)*gamma))^aa)^(bb - 1)) + aa bb (gamma/(alfa*gamma)) (tempo^(gamma - 1))
exp(-(tempo/alfa)*gamma) ((1 - exp(-(tempo/alfa)*gamma))^(aa - 1)) ((1 - (1 - exp(-
(tempo/alfa)*gamma))^aa)^(bb - 1) - 1) ((bb - 1) ((1 - exp(-(tempo/alfa)*gamma))^(aa - 1) (aa
(exp(-(tempo/alfa)*gamma) ((tempo/alfa)^(gamma - 1) (gamma (tempo/alfa^2)))))))/((aa
bb (gamma/(alfa*gamma)) (tempo^(gamma - 1)) exp(-(tempo/alfa)*gamma) ((1 - exp(-
(tempo/alfa)*gamma))^(aa - 1)) ((1 - (1 - exp(-(tempo/alfa)*gamma))^aa)^(bb - 1)))) +
(1 - censura) ((1 - (1 - exp(-(tempo/alfa)*gamma))^aa)^(bb - 1) ((bb) ((1 - exp(-
(tempo/alfa)*gamma))^(aa - 1) (aa (exp(-(tempo/alfa)*gamma) ((tempo/alfa)^(gamma - 1)
(gamma (tempo/alfa^2)))))))/((1 - (1 - exp(-(tempo/alfa)*gamma))^aa)^(bb))))
```



$$\begin{aligned}
\text{derigam} &<-\text{sum}(\text{censura} * (((((\text{aa} * \text{bb} * (1/(\text{alfa}\hat{\text{gamma}}) - \text{gamma} \\
&* (\text{alfa}\hat{\text{gamma}} * \log(\text{alfa}))/(\text{alfa}\hat{\text{gamma}})^2) * (\text{tempo}(\hat{\text{gamma}} - 1)) + \text{aa} * \text{bb} * \\
&(\text{gamma}/(\text{alfa}\hat{\text{gamma}})) * (\text{tempo}(\hat{\text{gamma}} - 1) * \log(\text{tempo}))) * \exp(-(\text{tempo}/\text{alfa})\hat{\text{gamma}}) \\
&- \text{aa} * \text{bb} * (\text{gamma}/(\text{alfa}\hat{\text{gamma}})) * (\text{tempo}(\hat{\text{gamma}} - 1)) * (\exp(-(\text{tempo}/\text{alfa})\hat{\text{gamma}}) \\
&* ((\text{tempo}/\text{alfa})\hat{\text{gamma}} * \log((\text{tempo}/\text{alfa})))))) * ((1 - \exp(-(\text{tempo}/\text{alfa})\hat{\text{gamma}}))^{\hat{\text{aa}} - 1}) \\
&+ \text{aa} * \text{bb} * (\text{gamma}/(\text{alfa}\hat{\text{gamma}})) * (\text{tempo}(\hat{\text{gamma}} - 1)) * \exp(-(\text{tempo}/\text{alfa})\hat{\text{gamma}}) * \\
&((1 - \exp(-(\text{tempo}/\text{alfa})\hat{\text{gamma}}))^{\hat{\text{aa}} - 1}) * ((\text{aa} - 1) * (\exp(-(\text{tempo}/\text{alfa})\hat{\text{gamma}}) * \\
&((\text{tempo}/\text{alfa})\hat{\text{gamma}} * \log((\text{tempo}/\text{alfa})))))) * ((1 - (1 - \exp(-(\text{tempo}/\text{alfa})\hat{\text{gamma}}))^{\hat{\text{aa}}})^{\hat{\text{bb}} - 1}) \\
&- \text{aa} * \text{bb} * (\text{gamma}/(\text{alfa}\hat{\text{gamma}})) * (\text{tempo}(\hat{\text{gamma}} - 1)) * \exp(-(\text{tempo}/\text{alfa})\hat{\text{gamma}}) * \\
&((1 - \exp(-(\text{tempo}/\text{alfa})\hat{\text{gamma}}))^{\hat{\text{aa}} - 1}) * ((1 - (1 - \exp(-(\text{tempo}/\text{alfa})\hat{\text{gamma}}))^{\hat{\text{aa}}})^{\hat{\text{bb}} - 1}) \\
&- 1) * ((\text{bb} - 1) * ((1 - \exp(-(\text{tempo}/\text{alfa})\hat{\text{gamma}}))^{\hat{\text{aa}} - 1}) * (\text{aa} * (\exp(-(\text{tempo}/\text{alfa})\hat{\text{gamma}}) \\
&* ((\text{tempo}/\text{alfa})\hat{\text{gamma}} * \log((\text{tempo}/\text{alfa})))))))/(\text{aa} * \text{bb} * (\text{gamma}/(\text{alfa}\hat{\text{gamma}})) * \\
&(\text{tempo}(\hat{\text{gamma}} - 1)) * \exp(-(\text{tempo}/\text{alfa})\hat{\text{gamma}}) * ((1 - \exp(-(\text{tempo}/\text{alfa})\hat{\text{gamma}}))^{\hat{\text{aa}} - 1}) \\
& * ((1 - (1 - \exp(-(\text{tempo}/\text{alfa})\hat{\text{gamma}}))^{\hat{\text{aa}}})^{\hat{\text{bb}} - 1}))) - (1 - \text{censura}) * ((1 - (1 - \exp(- \\
&(\text{tempo}/\text{alfa})\hat{\text{gamma}}))^{\hat{\text{aa}}})^{\hat{\text{bb}} - 1}) * ((\text{bb} * ((1 - \exp(-(\text{tempo}/\text{alfa})\hat{\text{gamma}}))^{\hat{\text{aa}} - 1}) * (\text{aa} * \\
&(\exp(-(\text{tempo}/\text{alfa})\hat{\text{gamma}}) * ((\text{tempo}/\text{alfa})\hat{\text{gamma}} * \log((\text{tempo}/\text{alfa})))))))/((1 - (1 - \exp(- \\
&(\text{tempo}/\text{alfa})\hat{\text{gamma}}))^{\hat{\text{aa}}})^{\hat{\text{bb}}}))
\end{aligned}$$

$$\begin{aligned}
\text{deriaa} &<-\text{sum}(\text{censura} (((\text{bb} (\text{gamma}/(\text{alfa}\hat{\text{gamma}})) (\text{tempo}(\hat{\text{gamma}} - 1)) \times \\
&\exp(-(\text{tempo}/\text{alfa})\hat{\text{gamma}}) ((1 - \exp(-(\text{tempo}/\text{alfa})\hat{\text{gamma}}))^{\hat{\text{aa}} - 1}) + \text{aa} \text{bb} \times \\
&(\text{gamma}/(\text{alfa}\hat{\text{gamma}})) (\text{tempo}(\hat{\text{gamma}} - 1)) \exp(-(\text{tempo}/\text{alfa})\hat{\text{gamma}}) ((1 - \exp(- \\
&(\text{tempo}/\text{alfa})\hat{\text{gamma}}))^{\hat{\text{aa}} - 1} * \log((1 - \exp(-(\text{tempo}/\text{alfa})\hat{\text{gamma}})))))) ((1 - (1 - \exp(- \\
&(\text{tempo}/\text{alfa})\hat{\text{gamma}}))^{\hat{\text{aa}}})^{\hat{\text{bb}} - 1}) - \text{aa} \text{bb} (\text{gamma}/(\text{alfa}\hat{\text{gamma}})) (\text{tempo}(\hat{\text{gamma}} - 1)) \\
&\exp(-(\text{tempo}/\text{alfa})\hat{\text{gamma}}) ((1 - \exp(-(\text{tempo}/\text{alfa})\hat{\text{gamma}}))^{\hat{\text{aa}} - 1}) ((1 - (1 - \exp(- \\
&(\text{tempo}/\text{alfa})\hat{\text{gamma}}))^{\hat{\text{aa}}})^{\hat{\text{bb}} - 1}) - 1) ((\text{bb} - 1) ((1 - \exp(-(\text{tempo}/\text{alfa})\hat{\text{gamma}}))^{\hat{\text{aa}}}) \times \\
&\log((1 - \exp(-(\text{tempo}/\text{alfa})\hat{\text{gamma}})))))))/(\text{aa} \text{bb} (\text{gamma}/(\text{alfa}\hat{\text{gamma}})) (\text{tempo}(\hat{\text{gamma}} - 1)) \\
&\exp(-(\text{tempo}/\text{alfa})\hat{\text{gamma}}) ((1 - \exp(-(\text{tempo}/\text{alfa})\hat{\text{gamma}}))^{\hat{\text{aa}} - 1}) ((1 - \\
&(1 - \exp(-(\text{tempo}/\text{alfa})\hat{\text{gamma}}))^{\hat{\text{aa}}})^{\hat{\text{bb}} - 1}))) - (1 - \text{censura}) ((1 - (1 - \exp(- \\
&(\text{tempo}/\text{alfa})\hat{\text{gamma}}))^{\hat{\text{aa}}})^{\hat{\text{bb}} - 1}) ((\text{bb} ((1 - \exp(-(\text{tempo}/\text{alfa})\hat{\text{gamma}}))^{\hat{\text{aa}}}) \log((1 - \exp(- \\
&(\text{tempo}/\text{alfa})\hat{\text{gamma}})))))))/((1 - (1 - \exp(-(\text{tempo}/\text{alfa})\hat{\text{gamma}}))^{\hat{\text{aa}}})^{\hat{\text{bb}}}))
\end{aligned}$$

$$\text{deribb} <-\text{sum}(\text{censura} ((\text{aa} (\text{gamma}/(\text{alfa}\hat{\text{gamma}})) (\text{tempo}(\hat{\text{gamma}} - 1)) \times$$

$$\begin{aligned} & \exp(-(\text{tempo}/\text{alfa})\hat{\gamma}) \left( (1 - \exp(-(\text{tempo}/\text{alfa})\hat{\gamma})) \hat{a}^a - 1 \right) \left( (1 - (1 - \exp(-(\text{tempo}/\text{alfa})\hat{\gamma})) \hat{a}) \hat{b}^b - 1 \right) + a a * b b * (\gamma/(\text{alfa}\hat{\gamma})) * (\text{tempo}\hat{\gamma} \\ & - 1) * \exp(-(\text{tempo}/\text{alfa})\hat{\gamma}) * ((1 - \exp(-(\text{tempo}/\text{alfa})\hat{\gamma})) \hat{a}^a - 1) * ((1 - (1 - \exp(-(\text{tempo}/\text{alfa})\hat{\gamma})) \hat{a}) \hat{b}^b - 1) * \log((1 - (1 - \exp(-(\text{tempo}/\text{alfa})\hat{\gamma})) \hat{a})) \\ & ) / (a a * b b * (\gamma/(\text{alfa}\hat{\gamma})) * (\text{tempo}\hat{\gamma} - 1) * \exp(-(\text{tempo}/\text{alfa})\hat{\gamma}) * ((1 - \exp(-(\text{tempo}/\text{alfa})\hat{\gamma})) \hat{a}^a - 1) * ((1 - (1 - \exp(-(\text{tempo}/\text{alfa})\hat{\gamma})) \hat{a}) \hat{b}^b - 1))) \\ & + (1 - \text{censura}) * ((1 - (1 - \exp(-(\text{tempo}/\text{alfa})\hat{\gamma})) \hat{a}) \hat{b}^b * \log((1 - (1 - \exp(-(\text{tempo}/\text{alfa})\hat{\gamma})) \hat{a})) / ((1 - (1 - \exp(-(\text{tempo}/\text{alfa})\hat{\gamma})) \hat{a}) \hat{b}^b))) \end{aligned}$$

```
c(-derialf,-derigam,-deriaa,-deribb)
```

```
require(numDeriv)
```

```
v1 <- constrOptim(c(0.33,0.34,170.27,0.09),verokw,gradkw,
ui=rbind(c(1,0,0,0),c(0,1,0,0),c(0,0,1,0),c(0,0,0,1)),ci=c(0,0,0,0),method
"BFGS",outer.iterations = 10000,hessian=T)
```

```
v1 v1$convergence
```

```
vero<-v1$value
```

```
paraR<-v1$par
```

```
hessiR<-v1$hessian
```

```
invR<-solve(hessiR)
```

```
varianciaR<-diag(invR)
```

```
eppR<-sqrt(varianciaR)
```

```
eppR
```

```
zR=paraR/eppR
```

```
pvalorR<-2*(1-pnorm(abs(zR)))
```

```
pvalorR
```

```
grad(vkwcura,paraR)
```

```
verow1<-v1$value*-1
```

```
verow1
```

```
pw1 <- 4
```

```
AICw1 <- (-2*verow1)+(2*pw1)
```

```
AICw1
```

```

AICcw1<-AICw1+((2*pw1+(pw1+1))/(n-pw1-1))
AICcw1
alphakw<-v1$par[1] gammakw<-v1$par[2]
aakw<-v1$par[3] bbkw<-v1$par[4]
##### Distribuição Kumarasswamy-Weibull com fração de cura
, Gradiente = gradgrr ##### verooo<-function(para) alfa<-para[1]
gamma<-para[2] aa<-para[3] bb<-para[4] phi<-para[5]
      ftt<-phi*(aa*bb*(gamma/(alfa*gamma))*(tempo^(gamma-1))*exp(-
(tempo/alfa)*gamma)*
      ((1-exp(-(tempo/alfa)*gamma))^(aa-1))*((1-(1-exp(-
(tempo/alfa)*gamma))^aa)^(bb-1)))
      stt<-(1-phi)+phi*(((1-(1-exp(-(tempo/alfa)*gamma))^aa)^(bb)))
      ll<-censura*log(ftt)+(1-censura)*log(stt)
      sum(-ll)

## Gradiente ###

gradgrr<-function(param) alfa<-param[1] gamma<-param[2] aa<-param[3]
bb<-param[4] phi<-param[5]
      derialfa <- sum(censura * (phi * (((aa * bb * (gamma/(alfa*gamma)) *
(tempo^(gamma - 1)) * (exp(-(tempo/alfa)*gamma) * ((tempo/alfa)^(gamma - 1) * (gamma
* (tempo/alfa^2)))))) - aa * bb * (gamma * (alfa^(gamma - 1) * gamma)/(alfa*gamma)^2) *
(tempo^(gamma - 1)) * exp(-(tempo/alfa)*gamma) * ((1 - exp(-(tempo/alfa)*gamma))^(aa -
1)) - aa * bb * (gamma/(alfa*gamma)) * (tempo^(gamma - 1)) * exp(-(tempo/alfa)*gamma)
* ((1 - exp(-(tempo/alfa)*gamma))^(aa - 1) - 1) * ((aa - 1) * (exp(-(tempo/alfa)*gamma)
* ((tempo/alfa)^(gamma - 1) * (gamma * (tempo/alfa^2)))))) * ((1 - (1 - exp(-
(tempo/alfa)*gamma))^aa)^(bb - 1)) + aa * bb * (gamma/(alfa*gamma)) * (tempo^(gamma -
1)) * exp(-(tempo/alfa)*gamma) * ((1 - exp(-(tempo/alfa)*gamma))^(aa - 1)) * ((1 - (1 - exp(-
(tempo/alfa)*gamma))^aa)^(bb - 1) - 1) * ((bb - 1) * ((1 - exp(-(tempo/alfa)*gamma))^(aa
- 1) * (aa * (exp(-(tempo/alfa)*gamma) * ((tempo/alfa)^(gamma - 1) * (gamma *
(tempo/alfa^2)))))))))))/(phi * (aa * bb * (gamma/(alfa*gamma)) * (tempo^(gamma - 1))

```



```
(gamma/(alfa*gamma)) * (tempo^(gamma - 1)) * exp(-(tempo/alfa)*gamma) * ((1 - exp(-
(tempo/alfa)*gamma))^(aa - 1)) * ((1 - (1 - exp(-(tempo/alfa)*gamma))^aa)^(bb - 1)))) - (1 -
censura) * (phi * ((1 - (1 - exp(-(tempo/alfa)*gamma))^aa)^(bb - 1)) * ((bb) * ((1 - exp(-
(tempo/alfa)*gamma))^aa * log((1 - exp(-(tempo/alfa)*gamma)))))))/((1 - phi) + phi * (((1 - (1
- exp(-(tempo/alfa)*gamma))^aa)^(bb))))))
```

```
deribb<- sum(censura * (phi * (aa * (gamma/(alfa*gamma)) * (tempo^(gamma
- 1)) * exp(-(tempo/alfa)*gamma) * ((1 - exp(-(tempo/alfa)*gamma))^(aa - 1)) * ((1 - (1 -
exp(-(tempo/alfa)*gamma))^aa)^(bb - 1)) + aa * bb * (gamma/(alfa*gamma)) * (tempo^(gamma
- 1)) * exp(-(tempo/alfa)*gamma) * ((1 - exp(-(tempo/alfa)*gamma))^(aa - 1)) * ((1 - (1 -
exp(-(tempo/alfa)*gamma))^aa)^(bb - 1)) * log((1 - (1 - exp(-(tempo/alfa)*gamma))^aa))))/(phi
* (aa * bb * (gamma/(alfa*gamma)) * (tempo^(gamma - 1)) * exp(-(tempo/alfa)*gamma)
* ((1 - exp(-(tempo/alfa)*gamma))^(aa - 1)) * ((1 - (1 - exp(-(tempo/alfa)*gamma))^aa)^(bb
- 1)))) + (1 - censura) * (phi * ((1 - (1 - exp(-(tempo/alfa)*gamma))^aa)^(bb) *
log((1 - (1 - exp(-(tempo/alfa)*gamma))^aa)))/((1 - phi) + phi * (((1 - (1 - exp(-
(tempo/alfa)*gamma))^aa)^(bb))))))
```

```
deriphi<- sum(censura * ((aa * bb * (gamma/(alfa*gamma)) * (tempo^(gamma
- 1)) * exp(-(tempo/alfa)*gamma) * ((1 - exp(-(tempo/alfa)*gamma))^(aa - 1)) * ((1 -
(1 - exp(-(tempo/alfa)*gamma))^aa)^(bb - 1)))/((phi * (aa * bb * (gamma/(alfa*gamma)) *
(tempo^(gamma - 1)) * exp(-(tempo/alfa)*gamma) * ((1 - exp(-(tempo/alfa)*gamma))^(aa
- 1)) * ((1 - (1 - exp(-(tempo/alfa)*gamma))^aa)^(bb - 1)))) + (1 - censura) * (((((1
- (1 - exp(-(tempo/alfa)*gamma))^aa)^(bb)) - 1)/((1 - phi) + phi * (((1 - (1 - exp(-
(tempo/alfa)*gamma))^aa)^(bb))))))
```

```
c(-derialfa,-derigamma,-deriaa,-deribb,-deriphi)
```

```
v4 <- constrOptim(c(5302.2568,gammakw,aakw,bbkw,0.5),verooo,gradgrr,
ui=rbind(c(1,0,0,0,0),c(0,1,0,0,0),c(0,0,1,0,0),c(0,0,0,1,0),c(0,0,0,0,1),c(0,0,0,0,-1)),
ci=c(0,0,0,0,0,-1),method = "BFGS",outer.iterations = 10000, hessian=T)
```

```

v4
v4$convergence
vero<-v4$value
paraR<-v4$par
hessiR<-v4$hessian
invR<-solve(hessiR)
varianciaR<-diag(invR)
eppR<-sqrt(varianciaR)
eppR
zR=paraR/eppR
pvalorR<-2*(1-pnorm(abs(zR)))
pvalorR
grad(vkwcura,paraR)
verow2<-v4$value*-1
verow2
pw2 <- 5
AICw2 <- (-2*verow2)+(2*pw2)
AICw2
AICcw2<-AICw2+((2*pw2+(pw2+1))/(n-pw2-1))
AICcw2
alphawkw<-v4$par[1]
gammawkw<-v4$par[2] aawkw<-v4$par[3]
bbwkw<-v4$par[4]
phiwkw<-v4$par[5]
##### Estimativas de Sobrevida usando kaplan-Meier, e os mode-
los #####
km<-survfit(Surv(tempo,censura) 1)
time <- km$time ### Tempo de KM ####
time
names(km)

```

```

skm<-km$surv ### Sobrevivência de Kaplan-Meier ###
skm
swe<-exp(-(time/alphaww)âgammaww) #### Sobrevivência do modelo Wei-
bull ####
swe
skw<-((1-(1-exp(-(time/alphakw)âgammakw))âakw)â(bkw))
skw
skwfc<-((1-phiwwk)+phiwwk*(((1-(1-exp(-(time/alphawwk)âgammawwk))âawwk)
â(bbwkw))))
skwfc
##### Seleção de Modelos #####
#####par(mfrow=c(1,3)) ##### Separar os gráficos
plot(km,conf.int=F,xlab="Tempos", ylab="S(t)")
lines(c(0,time),c(1,swe),lty=3,col=1)
legend(25,0.6,lty=1, col=1,c("Weibull"),bty="n",cex=0.8)
lines(c(0,time),c(1,skw),lty=3,col=2)
legend(25,0.55,lty=3, col=2,c("K-Weibull"),bty="n",cex=0.8)
lines(c(0,time),c(1,skwfc),lty=3,col=3)
legend(25,0.5,lty=3, col=3,c("K-Weibull com fração de
cura"),bty="n",cex=0.8)
##### Teste TRV Weibull x Kumaraswamy-Weibull #####

#Hipótese:
# H0: Modelo Simples - "Weibull"
# H1: Modelo Geral - "Kumaraswamy-Weibull"
lweib<-mwe$log[1] # logaritmo da verossimilhança do modelo weibull
lkw<-v1$value*-1 # logaritmo da verossimilhança do modelo Kumaraswamy-
Weibull

TRV<-2*(lkw-lweib)
TRV

```

```

1-pchisq(TRV,2)
#### Teste TRV Kumaraswamy-Weibull x K-W com fração de cura ####
#Hipótese:
# H0: Modelo Simples - "Weibull"

# H1: Modelo Geral - "Kumaraswamy-Weibull com fração de cura"

lweib<-mwe$log[1] # logaritmo da verossimilhança do modelo weibull
lkwfc<-v4$value*-1 # logaritmo da verossimilhança do modelo Kumaraswamy-
Weibull com #fração de cura
TRV<-2*(lkwfc-lweib)
TRV
1-pchisq(TRV,3)
#### Teste TRV Kumaraswamy-Weibull x K-W com fração de cura ####
#Hipótese:
# H0: Modelo Simples - "Kumaraswamy-Weibull"
# H1: Modelo Geral - "Kumaraswamy-Weibull com fração de cura"
lkw<-v1$value*-1 # logaritmo da verossimilhança do modelo Kumaraswamy-
Weibull
lkwfc<-v4$value*-1 # logaritmo da verossimilhança do modelo Kumaraswamy-
Weibull com
# fração de cura TRV<-2*(lkwfc-lkw)
TRV
1-pchisq(TRV,1)
##### Critério de Akaike #####
### Weibull ###
verow<-mregwei$loglik[1]
verow
pw <- 3
AICw <- (-2*verow)+(2*pw)

```



AICw

$AIC_{cw} <- AIC_w + ((2 * p_w + (p_w + 1)) / (n - p_w - 1))$

AICcw

### Kumaraswamy-Weibull ###

AICw1 #AIC normal

AICcw1 #AIC Corrigido

### Kumaraswamy-Weibull com fração de cura ###

AICw2 #AIC normal

AICcw2 #AIC Corrigido

`cbind(AICw,AICcw1,AICcw2)`