



Universidade Estadual da Paraíba  
Centro de Ciências e Tecnologia  
Departamento de Estatística

Priscilla dos Santos Cabral

**A Distribuição Weibull Inversa  
Generalizada na Modelagem de dados de  
Sobrevivência: Uma aplicação a dados de  
recorrentes químicos.**

Campina Grande  
Dezembro de 2012

Priscilla dos Santos Cabral

# **A Distribuição Weibull Inversa Generalizada na Modelagem de dados de Sobrevivência: Uma aplicação a dados de recorrentes químicos.**

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Orientador:

Tiago Almeida de Oliveira

Coorientador:

Marcelino Alves Rosa de Pascoa

Campina Grande

Dezembro de 2012

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL – UEPB

C117d Cabral, Priscilla dos Santos.  
A distribuição weibull inversa generalizada na modelagem de dados de sobrevivência [manuscrito] : uma aplicação a dados de recorrentes químicos / Priscilla dos Santos Cabral. – 2012.  
45f. : il. color.

Trabalho de Conclusão de Curso (Graduação em Estatística)  
– Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2012.  
“Orientação: Prof. Dr. Tiago Almeida de Oliveira, Departamento de Estatística”.

1. Estatística. 2. Análise de Sobrevivência. 3. Dependentes Químicos. I. Título.

21. ed. CDD 519.5

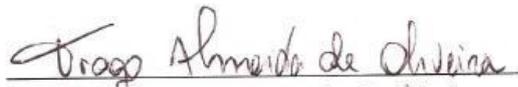
Priscilla dos Santos Cabral

# A Distribuição Weibull Inversa Generalizada na Modelagem de dados de Sobrevivência: Uma aplicação a dados de recorrentes químicos.

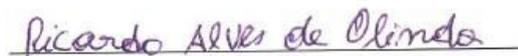
Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Aprovado em: 14 / 12 / 2012 .

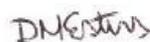
## Banca Examinadora:



Prof. Dr. Tiago Almeida de Oliveira  
Orientador



Prof. Dr. Ricardo Alves de Olinda  
Universidade Estadual da Paraíba



Profa. Dr. Divamilda Maia Esteves  
Universidade Estadual da Paraíba

# Dedicatória

*Dedico esta minha vitória a Deus e principalmente a meus pais que me incentivaram a luta pela vida fazendo com que me tornasse uma pessoa de princípios e caráter. Tenho muito a agradecer aos ensinamentos e amor que me dedicaram para que pudesse continuar esta caminhada que é a minha conclusão de curso. Dedico também a meu orientador. Aqui fica uma palavra de amizade e admiração, desta sua aluna que muito lhe estima, e que pede ao nosso maravilhoso Deus que Ele esteja sempre em sua companhia e de sua família, para que nada possa lhe faltar em nenhum momento.*

# Agradecimentos

Agradeço ...

Primeiramente a **Deus**, por sua bondade em ter dado a vida, a coragem, a paciência para transpor obstáculos por caminhos com portas largas e também com portas estreitas para que houvesse crescimento e a inteligência necessária para que eu pudesse realizar mais um objetivo.

Aos meus pais (**in memoriam**) Rosileide Bezerra e Gilmar dos Santos que não tive a oportunidade de conhecer. E aos meus pais “padrinhos” que me adotaram, **Maria de Fátima dos Santos Cabral e Gildo Soares Cabral**, pela educação que me proporciona, por fazer parte da minha vida, pelo apoio de sempre e por nunca deixar desistir dos meus objetivos.

A **meus irmãos** Evaldo, Gitana, Jacqueline, Bruno, Gustavo e a minha irmã torta Zezé, por ter me ajudado na minha trajetória e em especial a **Wagner** por sempre me guiar meu caminho nas horas difíceis e boas.

Aos **meus sobrinhos** Arthur, Brendow, João Victor, Luís Henrique e Wallison pelas risadas, loucuras feitas e pelas as raivas que me propuseram.

Ao **meu orientador Tiago Almeida de Oliveira** pela sua imensa dedicação, paciência para conosco, por ter acreditado no meu potencial, e a sua esposa **Ana Patricia Bastos Peixoto** por nunca me deixar abater diante dos desafios, pelo incentivo, competência e disponibilidade oferecida, que além de ser minha professora, para que a realização deste trabalho da ocorresse da melhor forma possível. E aos demais professores, em especial a **Ricardo Alves de Olinda, Juarez Fernandes de Oliveria** por fortalecerem a minha experiência acadêmica e pessoal.

Ao **meu namorado** Rodrigo Gustavo Amorim Barros, companheiro amigo, pelo amor, paciência, e compreensão e incentivador do futuro profissional.

Aos **amigos da Universidade**, em especial a **Alessandra, Erasilson** sempre dedicados, companheiros de aulas, imprescindíveis à conclusão desta etapa de minha vida, e aos demais amigos Allana, Sidcleide, Fernanda Matias, Michele, Fernanda Souza, por se

fazerem sempre presentes com belos risos, críticas, confidências, choros e principalmente por nossa amizade, que apesar dos anos e distância após o curso, sempre estarão presentes em minhas melhores lembrança. Na lembrança ficaram as alegrias e angústias desses cinco anos.

A **FIEP**, por ter me permitido realizar o estágio. A **Cláudia Lopes** e **Rosemere Barbosa** pelo apoio, convívio, compreensão, aprendizagem, e por sua sabedoria.

Aos **professores** do curso de Estatística da UEPB, que dedicam seus ensinamentos e engrandecimentos para a nossa formação profissional e a todos os funcionários do departamento.

Aos que direta ou indiretamente contribuíram para a conclusão desta etapa de grande importância na minha vida, e outros, que por qualquer motivo deixei de nomear.

Muito obrigada a todos por essa honrosa jornada.

# Resumo

Neste trabalho foi feita uma breve apresentação da distribuição Weibull Inversa Generalizada (WIG) introduzida por Gusmão et al. (2011). A distribuição WIG possui três parâmetros sendo uma nova proposta para a análise de sobrevivência. Após a apresentação de alguns casos particulares da WIG, estimaram-se seus parâmetros pelo método da máxima verossimilhança, a partir de uma amostra de dados de sobrevivência, contendo dados censurados relativos a tempo de recorrência de dependentes químicos na cidade de Caratinga no Estado de MG. A WIG se ajustou bem aos dados e sua função de risco estimada foi caracterizada por uma distribuição unimodal

**Palavras-chaves:** Análise de sobrevivência, Recorrentes químicos, Distribuições generalizadas

# Abstract

In this work, a brief presentation of the generalized inverse Weibull distribution (WIG) introduced by Gusmao et al. (2009). The WIG distribution has three parameters being a new proposal for the survival analysis. After the presentation of some particular cases of WIG, its parameters were estimated by maximum likelihood method, from a sample of survival data containing censored data with respect to time to recurrence of drug addicts in the city of Caratinga, in the State of Minas Gerais. The WIG good fit to the data and their estimated risk function was characterized by a unimodal distribution

**Key-words:** Survival analysis, Chemical recurrent, Generalized distributions.

# Sumário

## Lista de Figuras

## Lista de Tabelas

<b>1</b>	<b>Introdução</b>	p. 11
<b>2</b>	<b>Fundamentação Teórica</b>	p. 12
2.1	Análise de sobrevivência . . . . .	p. 12
2.1.1	Tempo . . . . .	p. 12
2.1.2	Censura . . . . .	p. 15
2.2	Testes não paramétricos . . . . .	p. 18
2.2.1	Estimador de Kaplan-Meier . . . . .	p. 18
2.3	Modelos probabilísticos . . . . .	p. 19
2.3.1	Distribuição exponencial . . . . .	p. 20
2.3.2	Distribuição Weibull . . . . .	p. 21
2.3.3	Distribuição Weibull Inversa Generalizada . . . . .	p. 23
2.4	Estimação dos parâmetros . . . . .	p. 24
2.4.1	Máxima verossimilhança . . . . .	p. 25
2.4.2	Intervalo de confiança . . . . .	p. 26
2.4.3	Teste de hipótese . . . . .	p. 27
2.5	Distribuição Log Weibull Inversa Generalizada . . . . .	p. 28
2.6	Modelo probabilístico . . . . .	p. 29
2.6.1	Método Gráfico . . . . .	p. 29

<b>3</b>	<b>Aplicação</b>	p. 30
<b>4</b>	<b>Conclusão</b>	p. 36
<b>5</b>	<b>Propostas Futuras</b>	p. 37
<b>6</b>	<b>Referências Bibliográficas</b>	p. 38
	<b>Apêndice A - Primeiro apêndice (SAS)</b>	p. 40
	<b>Apêndice B - Segundo apêndice (R)</b>	p. 41

# Lista de Figuras

1	Gráficos ilustrativos de alguma curva TTT. . . . .	p. 16
2	Representação gráfica de censura, em que • representa falha e o censura.	p. 17
3	Análise de Sobrevivência aos dados de recorrentes químicos. . . . .	p. 33
4	Tempo Total em Teste (Curva TTT) aos dados de recorrentes químicos.	p. 34
5	Risco Ajustado as variáveis. . . . .	p. 34
6	Ajuste da Kaplan-Meier e a WIG aos dados de recorrentes químicos. . .	p. 35

# Lista de Tabelas

1	Representação dos dados de dependentes químicos da cidade de Caratinga - MG . . . . .	p. 32
2	Estatística Descritiva das variáveis dos dados de recorrentes químicos .	p. 32
3	Estimativas dos Parâmetros e erros padrão da Weibull Inversa aplicada a dados de recorrentes químicos . . . . .	p. 33
4	Estimativas dos Parâmetros e erros padrão da Weibull Inversa Generalizada aplicada a dados de recorrentes químicos . . . . .	p. 33

# 1 Introdução

A análise de sobrevivência é um conjunto de processos estatísticos, utilizados na análise de dados, para qual a variável de interesse é o tempo que decorre até que um acontecimento de interesse se verifique. Estes acontecimentos são chamados de falhas. A análise de sobrevivência é uma das técnicas estatísticas que mais cresceram nas últimas duas décadas. A razão deste crescimento é o desenvolvimento e aprimoramento de técnicas estatísticas combinado com computadores cada vez mais ágeis (COLOSIMO; GIOLO, 2006). O termo análise de sobrevivência refere-se a situações médicas envolvendo dados censurados.

A principal característica do conjunto dados de sobrevivência é a presença de censura, isto refere-se ao acompanhamento do paciente ser interrompido, por diversos motivos como por exemplo o de residir mas não na mesma cidade, ou que o paciente faleceu de causa diferente do tratamento. O objetivo da análise de sobrevivência é estimar, comparar, analisar e interpretar as variáveis com o tempo de sobrevivência e as funções de risco.

Para que se possa conhecer o comportamento de determinadas doenças e/ou fenômenos é necessário que se acompanhe os indivíduos sobre risco, para isto é necessário uma amostra (conjunto de dados), neste contexto a análise de sobrevivência é uma das formas de se obter algumas dessas informações estimando-se os parâmetros de uma distribuição para o conjunto de dados coletados.

O objetivo deste trabalho foi o de ajustar a distribuição Weibull Inversa Generalizada à dados de dependentes químicos da cidade de Caratinga no Estado de Minas Gerais e encontrar a função de Risco Estimada.

## 2 Fundamentação Teórica

Neste capítulo serão abordados algumas das principais distribuições que são casos particulares da Weibull Inversa Generalizada, bem como todo o procedimento inferencial para ajuste e comparação de distribuições, sendo que nem todos os tópicos aqui elucidados farão parte da aplicação.

### 2.1 Análise de sobrevivência

A análise de sobrevivência é o conjunto de técnicas e modelos estatísticos usados na análise do comportamento de variáveis positivas, tais como: tempo decorrido entre o início do tratamento até a morte do paciente, período de remissão de uma doença, tempo até o desenvolvimento de uma doença ou simplesmente tempo até a morte. Segundo Pascoa (2008), em análise de sobrevivência, a variável resposta é, geralmente, o tempo até a ocorrência de um evento de interesse. Esses eventos na maioria dos casos, indesejáveis e usualmente chamados de falhas. Os conjuntos de dados de sobrevivência são caracterizados pelos tempos de falhas, cuja característica importante é a presença de censura, que representa a observação parcial da resposta.

#### 2.1.1 Tempo

Seja  $T$  uma variável aleatória, não-negativa, absolutamente contínua, cujo tempo de sobrevivência  $T \geq 0$  é expresso por meio de várias funções matematicamente equivalentes, tais que, se uma delas é especificada, as outras podem ser derivadas. Segundo Colosimo (2006) estas funções são usadas para descrever diferentes aspectos apresentado pelo conjunto de dados e utilizada para caracterizar o comportamento de dados de tempo de sobrevivência, em que  $t$  representa o tempo de falha especificada em análise de sobrevivência, cuja distribuição pode ser caracterizada por qualquer umas das seguintes funções: A função densidade de probabilidade,  $f(t)$ , a função de sobrevivência,  $S(t)$ , função de risco,  $h(t)$ .

A função de densidade de probabilidade é caracterizada pelo evento de interesse ao observar um indivíduo no intervalo de tempo  $[t, t+\Delta t]$  por unidade de tempo que é definida como limite da probabilidade, (LOUZADA, 2012). Expressa por,

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t} \quad (2.1)$$

em que,  $f(t) \geq 0$  para todo  $t$ , e tem a área abaixo da curva igual a 1.

Enquanto a função de sobrevivência, denotada por  $S(t)$ , é definida como a probabilidade de um indivíduo sobreviver até um certo tempo  $t$ , sem o evento. Sendo uma das principais funções probabilísticas usadas para descrever dados de tempo de sobrevivência, definida por

$$S(t) = P(T > t) = 1 - F(t) = \int_0^t f(u) du \quad , \quad (2.2)$$

sendo que  $S(t) = 1$  quando  $t = 0$  e  $S(t) = 0$  quando  $t \rightarrow \infty$  e  $F(t) = \int_0^t f(u)$  representa a função de distribuição acumulada.

A função de risco, ou taxa de falha descreve a forma com que a taxa de falha muda com o tempo, ou seja, demonstra o risco do indivíduo falhar no tempo. É definida como o risco instantâneo de um indivíduo sofrer o evento entre o tempo  $t$  e  $t + \Delta t$ , dado que ele sobreviveu ao tempo  $t$ , uma definição formal é apresentada por Louzada (2012),

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t \mid T \geq t)}{\Delta t}. \quad (2.3)$$

A função de risco pode ser definida, em termos da função de distribuição  $F(t)$  e da função de densidade de probabilidade  $f(t)$ , da seguinte forma:

$$h(t) = \frac{f(t)}{1 - F(t)} \quad (2.4)$$

A função de risco fornece a taxa instantânea de falha, por unidade de tempo, isto é, pode-se caracterizar classes especiais de distribuições de tempo de sobrevivência de acordo com o comportamento em relação ao tempo. Conhecida como força de mortalidade ou taxa de mortalidade condicional. A função de risco pode ser constante, crescente, decrescente. A distribuição exponencial apresenta função de risco constante e a distribuição de weibull apresenta função de risco crescente, decrescente ou ainda constante, as quais serão apresentadas no decorrer deste trabalho.

Estas funções são utilizadas na prática com o objetivo de descrever os aspectos apre-

sentados pelo conjunto de dados. As funções densidade de probabilidade definidas anteriormente são matematicamente equivalentes, ou seja, a obtenção das funções descritas são especificada. A função densidade de probabilidade é definida como a derivada da função densidade de probabilidade acumulada dada em (2.1)

$$f(t) = \frac{\partial F(t)}{\partial t}. \quad (2.5)$$

Como  $F(t) = 1 - S(t)$  pode-se escrever

$$f(t) = \frac{\partial[1 - S(t)]}{\partial t} = -S'(t), \quad (2.6)$$

substituindo (2.6) em (2.4) obte-se

$$h(t) = -\frac{S'(t)}{S(t)} = -\frac{\partial[\log S(t)]}{\partial t}. \quad (2.7)$$

Dessa forma tem-se

$$\log S(t) = -\int_0^t h(u) du \quad (2.8)$$

ou seja,

$$S(t) = \exp\left(-\int_0^t h(u) du\right). \quad (2.9)$$

Uma outra função importante é a função risco acumulada, definida como

$$H(t) = \int_0^t h(u) du \quad (2.10)$$

Substituindo-se (2.10) em (2.9) tem-se que

$$S(t) = \exp[-H(t)]. \quad (2.11)$$

Como,  $\lim_{t \rightarrow \infty} S(t) = 0$  então

$$\lim_{t \rightarrow \infty} H(t) = \infty.$$

Além disso, de (2.4) e seleção de variáveis

$$f(t) = h(t)S(t) \quad (2.12)$$

Substituindo-se (2.11) em (2.12) tem-se

$$f(t) = h(t) \exp\left(-\int_0^t h(u)du\right). \quad (2.13)$$

A expressão (2.13) é muito importante quando desenvolve-se os procedimentos de estimação somente sobre a função de risco.

Uma forma empírica de determinar o comportamento da função de risco é expresso por meio da construção do gráfico do tempo total em teste (Curva TTT), proposto por Aarset (1987). A curva de TTT é obtida construindo-se um gráfico de

$$G\left(\frac{r}{n}\right) = \frac{\sum_{i=1}^r T_{i:n} + (n-r)T_{r:n}}{\sum_{i=1}^n T_{i:n}} \quad \text{por } \frac{r}{n}, \quad (2.14)$$

dado que  $r = 1, \dots, n$   $T_{i:n}^i = 1, \dots, n$  é estatística de ordem da amostra e  $n$  o tamanho da amostra (MUDHOLKAR, SRIVASTAVA e KOLLIA, 1996).

Na Figura 1 tem-se a representação da curva TTT em que tem-se uma reta diagonal se o risco for constante (reta A), uma curva convexa se a função de risco for decrescente (curva B) e côncava se o risco for crescente (curva C), e uma curvatura primeiramente convexa e depois côncava (curva D) se o risco for em forma de “U”, e no caso reverso (curva E) é unimodal.

## 2.1.2 Censura

Dados censurados são dados coletados ao longo de um tempo pré-determinado (por exemplo, tempo inicial até a ocorrência de um determinado evento). A ocorrência de dados censurados se deve ao fato que em um estudo nem todos os indivíduos, por diversos motivos, o evento de interesse ocorre. A informação obtida naqueles elementos ou indivíduos em que o evento não ocorre é apenas parcial sendo chamada assim de “dado censurado”, ou censura. Por exemplo, em ensaios clínicos ao longo de um tempo, em que o resultado de interesse é o tempo de sobrevivência de pacientes que continuam vivos ao término do ensaio, correspondente ao tempo decorrido entre o início e o fim do estudo. Se o indivíduo não manifestou o evento de interesse, então ao fim do ensaio é um dado censurado.

Segundo Strapasson (2007), para análise de sobrevivência é necessário que as observações sejam representadas por um vetor  $(t_i, \delta_i, x_i)$  em que,  $t_i$  é o tempo observado de falha ou censura e  $\delta_i$  uma variável indicadora de censura, em que  $\delta_i = 1$ , o tempo obser-

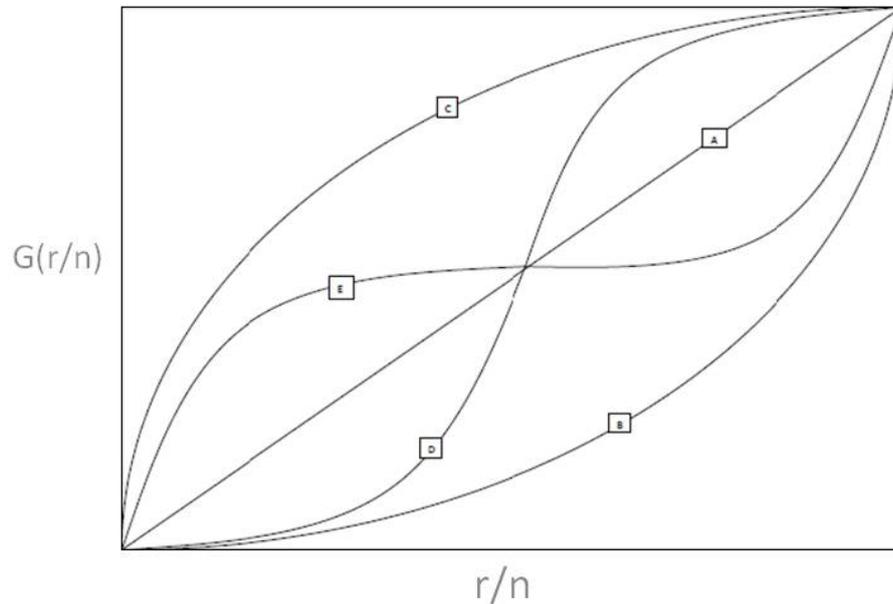


Figura 1: Gráficos ilustrativos de alguma curva TTT.

vado corresponde a uma falha ou  $\delta_i = 1$ , corresponde a uma censura. Para cada indivíduo observado tem-se uma covariável  $x_i$ , em que  $i, i=1, \dots, n$  são observações representadas pelo um par  $(t_i, \delta_i)$ .

$$\delta_i = \begin{cases} 1, & \text{quando } T \leq C, \\ 0, & \text{quando } T > C. \end{cases}$$

Pode-se ainda ocorrer outros dois tipos de censuras à esquerda e a censura intervalar. Segundo Strapasson (2007), censura à esquerda ocorre quando o evento de interesse já aconteceu, quando o indivíduo foi observado: ou seja, o tempo de vida é menor que o observado.

Censura intervalar é quando não se sabe o tempo exato de ocorrência do evento de interesse, sabe-se que ele ocorreu dentro de um intervalo especificado, por exemplo, ocorre quando não se conhece o exato momento da morte, mas sabe-se que ocorreu no intervalo de tempo.

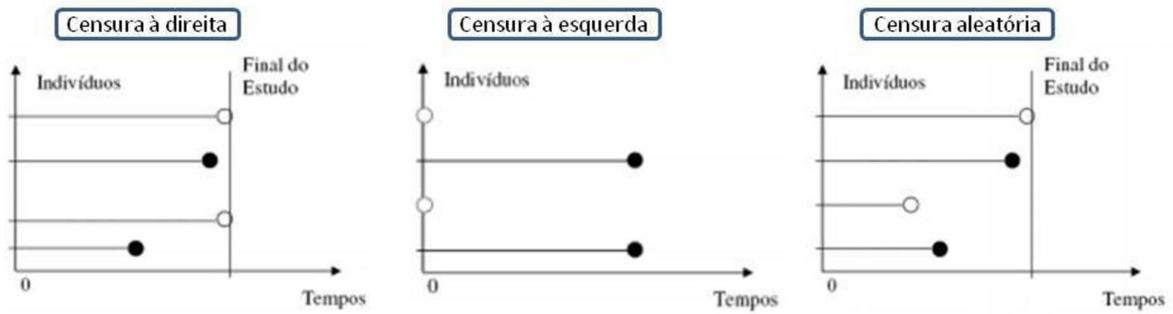


Figura 2: Representação gráfica de censura, em que ● representa falha e ○ censura.

As censuras podem ocorrer de várias formas com diferentes mecanismos. Correspondente a situações de censura à direita tem-se:

i) Censura tipo I

São observações acompanhadas de estudos que terminaram após um período pré-estabelecido de tempo, ou seja, só pode-se saber o tempo de vida de um indivíduo se a morte ocorrer antes do instante pré-definido. Neste caso, o número de mortes observadas é aleatório.

ii) Censura tipo II

São observações acompanhadas de estudos que após um determinado número de pacientes experimentar o evento de interesse terminará em número pré-estabelecido de indivíduos. Neste caso, o tempo de duração do estudo é uma variável aleatória.

iii) Censura tipo III aleatória

Ocorre quando o indivíduo é retirado no decorrer do estudo sem ter ocorrido falha. Neste caso, os indivíduos podem entrar no estudo em diferentes tempos e as censuras não podem ocorrer pelas seguintes razões: paciente morrer de causa diferente da estudada, paciente recusa-se a continuar no tratamento, paciente muda de cidade (então seu tempo de sobrevivência é maior ou igual ao valor esperado).

Segundo Colosimo e Giolo (2006) o mecanismo de censura aleatória é aquele em que os tempos de censura são variáveis aleatórias mutuamente independentes e ainda independentes dos tempos de vida. A censura do tipo I é um caso particular da aleatória, cuja variável aleatória  $t$ , tem uma probabilidade maior do que zero, ou seja,  $t$  é uma variável aleatória mista com um componente contínuo e outro discreto. Dados censurados são representados por sinal “+”.

## 2.2 Testes não paramétricos

A função de sobrevivência pode ser estimada considerando-se modelos paramétricos e técnicas não-paramétricas. As técnicas não-paramétricas podem indicar evidências de que o modelo paramétrico não está adequado. Segundo Colosimo e Giolo (2006) existem técnicas não-paramétricas para estimar parâmetros em análise de sobrevivência, obtendo a opção de ajustar os dados utilizando-se os modelos paramétricos probabilístico para tempo de falha.

Os procedimentos não paramétricos são usados para estimação das funções de densidade de probabilidade,  $f(t)$ , da função de sobrevivência,  $S(t)$ , e da função de risco,  $h(t)$ . A função de densidade de probabilidade,  $\hat{f}(t)$ , pode ser estimada a partir dos dados amostrais se não existirem observações censuradas, a função de sobrevivência,  $\hat{S}(t)$ , é estimada a partir dos dados, como a proporção de pacientes que sobreviveram após um certo período de tempo,  $t$ , e a função de risco,  $\hat{h}(t)$ , é estimada a partir dos dados amostrais quando não existirem observações censuradas.

Os estimadores de probabilidade de sobrevida,  $S(t)$ , utilizados nos teste não-paramétricos se resumem em três que são: o teste de Kaplan-Meier, a tabela de vida ou actuarial, que é uma das mais antigas técnicas estatística para estimar o tempo de falha, sendo utilizada apenas em grandes amostras. E o estimador de Nelson-Aalen que apresenta propriedades similares a Kaplan-Meier.

### 2.2.1 Estimador de Kaplan-Meier

A análise descritiva consiste essencialmente em encontrar medidas de tendência central e variabilidade. Com a presença de censuras inválidas, este tipo de tratamento aos dados de sobrevivência, o principal componente da análise descritiva envolvendo dados de tempo de vida é a função de sobrevivência. O objetivo de uma análise estatística envolvendo dados de sobrevivência está relacionado com a identificação de fatores de prognóstico para uma certa doença ou à comparação de tratamentos em estudos clínicos. Quando a amostra não contiver observações censuradas, há ocorrências de falha em um certo número no intervalo, para a função de sobrevivência é utilizado o estimador não-paramétrico de Kaplan-Meier, dada por César (2005) e Colosimo e Giolo (2006).

Portanto, como a função de densidade de probabilidade e a função de risco a partir dos dados amostrais, estas não permitem a presença de observações censuradas as quais são comuns os dados de sobrevivência e confiabilidade. As estimativas podem ser obtidas

a partir de métodos não-paramétricos, que não supõem nenhuma distribuição conhecida, como a utilização da função de Kaplan-Meier no qual permite a presença de observações censuradas.

O estimador de Kaplan-Meier permite realizar testes de hipóteses que não requerem pressupostos sobre a forma da distribuição subjacentes aos dados, é usado para analisar dados medidos apenas numa escala ordinal, podendo ocorrer para dados categorizados que são medidos em escala nominal. É adequada para amostras provenientes de diversas populações, sendo usado se todas as observações falharam, ou seja, não existiram censuras. As observações censuradas informam que o tempo até a falha é maior do que aquele que foi registrado. O estimador não-paramétrico de Kaplan-Meier considera a ocorrência de falhas distintas em intervalos de tempo, onde os tempos de sobrevivência são ordenados, isto é,  $t_1 \leq t_2 \leq t_3 \leq \dots \leq t_k$ , podendo ocorrer mais de uma falha no mesmo tempo, expressado por Colosimo e Giolo (2006) por,

- i)  $t_1 \leq t_2 \leq t_3 \leq \dots \leq t_k$  tempos distintos e ordenados de falha,
- ii)  $d_j$ : número de falhas até o tempo  $t_j$ ,  $j= 1, 2, \dots, k$  e
- iii)  $n_j$ : número de itens sob risco, ou seja, os indivíduos não falharam e não censurados até  $t_j$

Segundo César (2005) e Colosimo e Giolo (2006) o estimador  $\hat{R}(t)$  de Kaplan-Meier, é definido por,

$$\hat{R}(t) = \left( \frac{n_1 - d_1}{n_1} \right) \cdot \left( \frac{n_2 - d_2}{n_2} \right) \cdot \dots \cdot \left( \frac{n_{t_o} - d_{t_o}}{n_{t_o}} \right) = \prod_{j, t_j < t} \frac{n_i - d_i}{n_i}$$

em que  $t_o$  é o maior tempo de falha menor que  $t$ .

As principais propriedades do estimador são: ele é não viciado para amostras grandes, é fracamente consistente, converge assintoticamente para um processo gaussiano e é estimador de máxima verossimilhança de  $S(t)$ .

## 2.3 Modelos probabilísticos

Embora existam vários modelos probabilísticos, alguns ocupam maior destaque por sua comprovada adequação a várias situações reais, ou seja, por modelar os tempos de sobrevivência. Os principais modelos probabilísticos utilizados na análise de sobrevivência

são o Exponencial, o Weibull e o Log-Normal, pois as variáveis tratam do tempo até a falha sendo positivos, por outro lado, a Gaussiana (normal) e a binomial são adequadas para variáveis clínicas e industriais.

A distribuição Exponencial é das mais simples e importantes distribuições de probabilidade utilizadas para modelagem de dados que representam o tempo até a ocorrência do evento em interesse, apresentando a função de risco constante. A distribuição Weibull é a generalização da distribuição exponencial, sendo bastante utilizada no ajuste de dados de confiabilidade em diversas áreas do conhecimento, apresenta função de risco crescente, decrescente ou ainda constante. A distribuição Log-Normal é usada para ajustar dados referentes à confiabilidade, como a distribuição Weibull, sendo que a Weibull e Log-Normal são caracterizados por dois parâmetros e a Exponencial por apenas um.

### 2.3.1 Distribuição exponencial

A distribuição exponencial é um dos modelos mais simples e uma importantes distribuição de sobrevivência, ela descreve o tempo da função de risco (taxa de falha) constante. A função de densidade de probabilidade para uma variável aleatória tempo de falha,  $T$ , tem distribuição exponencial com parâmetro  $\alpha > 0$ , definida por:

$$f(t) = \frac{1}{\alpha} \exp \left\{ - \left( \frac{t}{\alpha} \right) \right\}, \quad \text{para } t \geq 0$$

em que  $\alpha > 0$  é o tempo médio de vida. O parâmetro  $\alpha$  tem a mesma unidade do tempo de falha  $t$ . Isto é, se  $t$  é medido em horas,  $\alpha$  também será fornecido em horas.

A função de sobrevivência  $S(t)$  e de taxa de falha  $\lambda(t)$  são definidos, respectivamente, por (COLOSIMO e GIOLO, 2006):

$$S(t) = \exp \left\{ - \left( \frac{t}{\alpha} \right) \right\}$$

e

$$\lambda(t) = \frac{1}{\alpha} \quad \text{para } t \geq 0 \quad (2.15)$$

Como a distribuição exponencial apresenta uma taxa de falha constante ao longo do tempo e a sua função de risco não depende do tempo, então, esta propriedade é chamada falta de memória. Outras características importantes da distribuição exponencial são a média, a variância e os percentis, onde a média é o  $\alpha$ , a variância  $\alpha^2$  e o percentil  $100_p\%$  corresponde ao tempo em que os indivíduos falharam.

Para o caso da exponencial, o percentil  $100_p\%$ ,  $t_p$ , é dado respectivamente:

$$t_p = -\alpha \log(1 - p)$$

então, o valor de  $\alpha$ , corresponde ao percentil da mediana, seguindo aproximadamente  $t_{0,63}$ , ou seja, o 63-ésimo percentil, facilmente obtida por  $t_{0,5} = -\alpha \log(1 - 0,5)$ .

Uma extensão da distribuição exponencial é obtida quando,  $t$  é substituído por  $(t - t_o)$ , em que  $t_o < t$  é o parâmetro conhecido como o tempo de garantia, em que a falha nunca ocorre. Esta extensão é conhecida como distribuição exponencial com dois parâmetros. A distribuição exponencial é usada para situação onde a morte de um indivíduo, ou a falha de um componente ocorra aleatoriamente no tempo, ou seja, a distribuição do tempo de sobrevivência adicional não é afetada pela informação do mesmo que sobreviveu a algum tempo.

### 2.3.2 Distribuição Weibull

A distribuição Weibull foi proposta originalmente por Walodi Weibull (1939) discutida em (1951, 1954), por este mesmo autor. Esta distribuição é usada frequentemente para estudos biomédicos e industriais, pois apresenta uma grande variedade de formas devido à sua simplicidade, todas com propriedades básicas: função de taxa de falha é monótona, isto é, crescente, decrescente ou constante. O autor resalta que essa distribuição é tão importante para análise paramétrica de dados de sobrevivência quanto a distribuição normal é para modelos lineares.

De acordo com Colosimo e Giolo (2006), a sua densidade é da forma,

$$f(t) = \frac{\gamma}{\alpha\gamma} t^{\gamma-1} \exp \left\{ - \left( \frac{t}{\alpha} \right)^\gamma \right\}, \quad t \geq 0 \quad (2.16)$$

em que  $\gamma > 0$  e  $\alpha > 0$  são parâmetros de forma e escala. Pode-se observar, que para  $\gamma < 1$ , tem-se função de taxa de falha decrescente, enquanto  $\gamma > 1$  as funções de taxa de falha são crescente, e  $\gamma = 1$  a função de taxa de falha é constante. O parâmetro  $\alpha$  tem mesma unidade de medida de  $t$ ,  $\gamma$  não tem unidade.

As funções de risco e de sobrevivência são, respectivamente,

$$S(t) = \exp \left\{ - \left( \frac{t}{\alpha} \right)^\gamma \right\} \quad (2.17)$$

e

$$\lambda(t) = \frac{\gamma}{\alpha\gamma} t^{\gamma-1} \quad (2.18)$$

para  $t \geq 0$ ,  $\alpha$  e  $\gamma > 0$ .

Quando  $\gamma = 1$ , obte-se a distribuição exponencial como caso particular da distribuição Weibull, sendo algumas formas das funções de densidade de sobrevivência e de taxa de falha (risco) da variável  $T$ .

O modelo Weibull é um caso particular do modelo de Cox para dados de sobrevivência intervalar. Segundo Strapasson (2007) pode-se ainda mostrar que se  $T \sim \text{Weibull}(\alpha, \gamma)$ , então,  $Y = \log(T) \sim \text{Gumbel}$ , ou seja,  $Y$  tem distribuição do valor extremo, com função de densidade de probabilidade, função de sobrevivência e função da taxa dado respectivamente, por

$$f(y) = \frac{1}{\sigma} \exp \left\{ \left( \frac{y - \mu}{\sigma} \right) - \exp \left\{ \frac{y - \mu}{\sigma} \right\} \right\}, \quad (2.19)$$

$$S(y) = \exp \left\{ - \exp \left\{ \frac{y - \mu}{\sigma} \right\} \right\} \quad (2.20)$$

e

$$\lambda(y) = \frac{1}{\sigma} \exp \left\{ \frac{y - \mu}{\sigma} \right\} \quad (2.21)$$

em que  $Y$  e  $\mu \in \mathbb{R}$  e  $\sigma > 0$ . Se  $\mu = 0$  e  $\sigma = 1$ , tem-se a distribuição do valor extremo padrão. Os parâmetros  $\mu$  e  $\sigma$  são denominados parâmetros de locação e escala, respectivamente, e relacionam-se com os parâmetros da distribuição de Weibull e do valor extremo apresentando as seguintes relações de igualdade,  $\gamma = \frac{1}{\sigma}$  e  $\alpha = \exp\{\mu\}$ .

A constante de Euler é conhecido pela média e variância, ou seja,  $\mu - v\sigma$  e  $(\frac{\pi^2}{6})\sigma^2$  com  $v = 0,5772\dots$ . O percentil  $100_p\%$ ,  $t_p$ , expresso por (COLOSIMO, 2006) é,

$$t_p = \mu + \sigma \log[-\log(1 - p)] \quad (2.22)$$

Na análise de dados de sobrevivência, é muita vezes conveniente trabalhar com o logaritmo dos tempo de vida dos indivíduos.

### 2.3.3 Distribuição Weibull Inversa Generalizada

Segundo Gusmão et al. (2011), a distribuição Weibull Inversa Generalizada estuda as formas de densidade. Em 1982 Keller e Kamath estudaram as formas de densidade e as funções da taxa de falha para a Weibull Inversa, em que a variável aleatória  $Y$  e sua função de distribuição acumulada, sendo também uma limitação de distribuição da estatística de ordem maior.

A distribuição Weibull Inversa Generalizada (GIW) foi discutida por Drapella(1993), Mudholkar e Kollia(1996) e Jiang et al.(1999) entre outros. A GIW é uma função de densidade cumulativa definida por  $G(t)$  para  $\gamma > 0$ , então diz-se que  $F(t) = G(t)^\gamma = \exp \left[ \gamma - \left( \frac{\alpha}{t} \right)^\beta \right]$ . Assim tem-se, a função de densidade GIW com três parâmetros  $\alpha > 0$ ,  $\beta > 0$  e  $\gamma > 0$  expressa por

$$f(t) = \gamma\beta\alpha^\beta t^{-(\beta+1)} \exp \left[ -\gamma \left( \frac{\alpha}{t} \right)^\beta \right], \quad t > 0 \quad (2.23)$$

A distribuição Weibull Inversa é um caso especial de (2.23) quando  $\gamma = 1$ . Em que  $T$  é uma variável aleatória com densidade (2.23), então  $T \sim \text{GIW}(\alpha, \beta, \gamma)$ . A sobrevivência correspondente a função de risco são expressada por (GUSMÃO et al, 2011)

$$S(t) = 1 - F(t) = 1 - \exp \left[ -\gamma \left( \frac{\alpha}{t} \right)^\beta \right]$$

e

$$h(t) = \gamma\beta\alpha^\beta t^{-(\beta+1)} \exp \left[ -\gamma \left( \frac{\alpha}{t} \right)^\beta \right] \left\{ 1 - \exp \left[ -\gamma \left( \frac{\alpha}{t} \right)^\beta \right] \right\}^{-1},$$

em que, pode-se simular a distribuição GIW utilizando-se a equação não linear

$$t = \alpha \left[ -\frac{\log(\mu)}{\gamma} \right]^{-\frac{1}{\beta}} \quad (2.24)$$

tendo  $\mu$  como distribuição uniforme (0, 1).

A função da taxa de falha é caracterizada pela primeira derivada  $h(t) = \frac{dh(t)}{dt}$  e é expressa por (GUSMÃO et al, 2011)

$$h'(t) = h(t)t^{-(\beta+1)} \left\{ \gamma\beta\alpha^\beta \left\{ 1 - \exp \left[ \gamma - \left( \frac{\alpha}{t} \right)^\beta \right] \right\}^{-1} - (\beta + 1)t^\beta \right\}$$

sendo que  $h(t)$  é unimodal em que  $t$  satisfaz a equação não-linear.

$$\gamma \frac{\alpha}{t^*} \left\{ 1 - \exp \left[ -\gamma \left( \frac{\alpha}{t^*} \right)^\beta \right] \right\}^{-1} = 1 + \beta^{-1}$$

respectivamente,

$$\lim_{t \rightarrow 0} h(t) = \lim_{t \rightarrow \infty} h(t) = 0, 3.$$

A distribuição Weibull Inversa Generalizada foi introduzida para distribuição de vida útil às estimativas, podendo-se aumentar ou diminuir a taxa de risco unimodais.

Segundo Gusmão et al. (2011), a característica mais importante de uma distribuição pode ser estudada por meio de momentos, ou seja, precisa-se enfatizar a importância dos momentos para qualquer análise estatística, sendo ela de tendência, dispersão, assimetria ou curtose. Se  $T$  é uma variável aleatória com função de densidade GIW (2.23), o  $k$ -ésimo momento definida por.

$$E(T^k) = \gamma^{\frac{k}{\beta}} \alpha^k \Gamma(1 - k\beta^{-1})$$

A função geradora de momentos  $M(z)$  de  $T$  para  $|z| < 1$  é

$$M(z) = E(e^{zT}) = \sum_{k=0}^{\infty} \left\{ \frac{z^k}{k!} \gamma^{\frac{k}{\beta}} \alpha^k \Gamma(1 - k\beta^{-1}) \right\}. \quad (2.25)$$

A distribuição Weibull Inversa é expressa pela entropia de Shannon, uma variável  $T$  com densidade  $f(t)$  medida por  $E \{-\log[f(t)]\}$  dada por (GUSMÃO et al., 2011)

$$E\{-\log[f(t)]\} = -\log(\alpha\beta\gamma) + (\beta + 1)\{\log(\alpha) + \beta^{-1}[\log(\gamma) + 0,577216]\} + 1$$

em que, 0,577216 é o valor aproximado para constante de Euler.

## 2.4 Estimação dos parâmetros

Segundo Colosimo e Giolo (2006), os parâmetros são características dos modelos de probabilidade para estudos de tempo de vida, existindo-se alguns métodos de estimação. O método de máxima verossimilhança é uma opção apropriada para dados censurados, incorporando-se as censuras relativamente simples por possuir propriedades para grandes amostras.

### 2.4.1 Máxima verossimilhança

O método de máxima verossimilhança apresenta os procedimentos de estimação para os parâmetros dos modelos de sobrevivência. A distribuição do tempo de falha é a Weibull, para cada combinação diferente  $\gamma$  e  $\alpha$ , tendo diferentes distribuições de Weibull. O estimador de máxima verossimilhança escolhe o par de  $\gamma$  e  $\alpha$  que melhor explique a amostra observada.

Segundo Colosimo e Giolo (2006), considera-se uma amostra de observações aleatórias  $t_1, \dots, t_n$  de uma variável aleatória  $T$  com tempos de sobrevivência e de confiabilidade de uma certa população de interesse com  $n$  observações independentes de  $t_i$ , em que  $t_i$ ,  $i = 1, \dots, n$ , indica o tempo de falha ou censura, onde todas são não-censuradas. Com um vetor de parâmetros  $\boldsymbol{\theta} = (\alpha, \beta, \gamma)$ , tem-se a função de verossimilhança para um parâmetro genérico  $\theta$  da população.

O método de máxima verossimilhança é baseado geralmente para modelo em inferência paramétrica e sua teoria assintótica, onde a função de verossimilhança para o vetor de parâmetros  $\boldsymbol{\theta}$  é expressa por (GUSMÃO et al., 2011)

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

cujo logaritmo é

$$\begin{aligned} l(\boldsymbol{\theta}) = & r[\log(\gamma) + \log(\beta) + \log(\alpha)] - (\beta + 1) \sum_{i \in F} \log(ti) - \gamma \alpha^\beta \sum_{i \in F} \log(ti)^{-\beta} \quad (2.26) \\ & + \sum_{i \in C} \log \left\{ 1 - \exp \left[ -\gamma \left( \frac{\alpha}{ti} \right)^\beta \right] \right\} \end{aligned}$$

A função de verossimilhança  $l(\boldsymbol{\theta})$  mostra que a contribuição de cada observação denota a não-censurada é a sua função de densidade, em que a contribuição censurada não é a sua função de densidade, onde  $r$  é o número de falhas em F e C a não censuradas e censurados indicaram os conjuntos de censura e observações censuradas. Os estimadores de máxima verossimilhança são os valores de  $\boldsymbol{\theta}$  que maximizam  $l(\boldsymbol{\theta})$ , onde temos a dependência de  $f$  em  $\boldsymbol{\theta}$ , em que  $l$  é função de  $\boldsymbol{\theta}$ .

As funções de pontuação para os parâmetros  $\alpha, \beta, \gamma$  são expressadas por (GUSMÃO et al., 2011)

$$U_\alpha(\boldsymbol{\theta}) = \frac{r\beta}{\alpha} - \gamma\beta\alpha^{\beta-1} \sum_{i \in F} ti^{-\beta} + \gamma\beta\alpha^{\beta-1} \sum_{i \in C} ti^{-\beta} \left( \frac{1 - ui}{ui} \right),$$

$$U_\beta(\boldsymbol{\theta}) = \frac{r}{\beta} + r \log(\alpha) - \sum_{i \in F} \log(ti) - \gamma \alpha^\beta \sum_{i \in F} ti^{-\beta} \log\left(\frac{\alpha}{ti}\right) + \gamma \alpha^\beta \sum_{i \in C} ti^{-\beta} \log\left(\frac{\alpha}{ti}\right) \left(\frac{1-ui}{ui}\right)$$

e

$$U_\gamma(\boldsymbol{\theta}) = \frac{r}{\gamma} - \alpha^\beta \sum_{i \in F} ti^{-\beta} + \alpha^\beta \sum_{i \in C} ti^{-\beta} \left(\frac{1-ui}{ui}\right),$$

em que  $ui = 1 - \exp\left[-\gamma\left(\frac{\alpha}{ti}\right)^\beta\right]$  é a  $i$ -ésima observação transformada.

A estimativa do logaritmo da função de verossimilhança é obtida por meio das probabilidades de equações não-lineares,  $U_\alpha(\boldsymbol{\theta})=0$ ,  $U_\beta(\boldsymbol{\theta})=0$  e  $U_\gamma(\boldsymbol{\theta})=0$  funções escores usando-se o algoritmo de Newton-Raphson.

Segundo Strapasson (2007), as propriedades assintóticas dos estimadores de máxima verossimilhança parcial são necessárias para construção de intervalos de confiança e testes de hipóteses sobre os parâmetros do modelo sob condições de regularidade com média  $\boldsymbol{\theta}$ , matriz de variância e covariância dada pelo inverso da matriz de Fisher ( $I(\boldsymbol{\theta})^{-1}$ ).

## 2.4.2 Intervalo de confiança

Para a construção de intervalo de confiança é necessário que tenhamos grandes amostras e certas condições para estabelecer regularidade às propriedades da distribuição assintótica do estimador de máxima verossimilhança, com o vetor  $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_1, \dots, \hat{\boldsymbol{\theta}}_k)$ , sendo  $k$  a dimensão de  $\boldsymbol{\theta}$ , ou seja, a inversa da esperança da matriz de variâncias-covariâncias é aproximadamente negativo devido a matriz de derivadas parciais do logaritmo de  $l(\boldsymbol{\theta})$ .

Segundo Colosimo e Giolo (2006), para a construção de intervalos de confiança é necessário uma estimativa do erro padrão de  $\hat{\boldsymbol{\theta}}$ , isto é, para  $[Var(\hat{\boldsymbol{\theta}})]^{\frac{1}{2}}$ . No caso especial em que  $\boldsymbol{\theta}$  é um escalar, um intervalo aproximado de  $(1 - \alpha)100\%$  de confiança para  $\boldsymbol{\theta}$  é dado por

$$\hat{\boldsymbol{\theta}} \pm z_{\frac{\alpha}{2}} \sqrt{Var(\hat{\boldsymbol{\theta}})}$$

em que  $\boldsymbol{\theta}$ , é o valor da estimativa de máxima verossimilhança do  $j$ -ésimo parâmetro e  $z_{\frac{\alpha}{2}}$  o percentil superior  $\frac{\alpha}{2}$  de uma distribuição normal padrão. No caso do modelo exponencial, temos um intervalo de confiança para o parâmetro  $\alpha$ , definido por

$$\hat{\alpha} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{\alpha}^2}{r}},$$

pois  $E\left[\frac{r}{\alpha^2} - \frac{\alpha \sum_i^n ti}{\alpha^3}\right] = -\frac{r}{\alpha^2}$ . A partir da matriz de variâncias e covariâncias  $Var(\hat{\boldsymbol{\theta}})$ ,

pode-se obter a estimativa do erro padrão para qualquer vetor de parâmetros dentro de um intervalo de confiança.

Supondo-se que  $\boldsymbol{\theta}=(\gamma, \alpha)$ . Como no modelo Weibull, o interesse é estimar a função dos parâmetros  $\boldsymbol{\phi} = g(\gamma, \alpha)$ . O estimador de máxima verossimilhança para  $\boldsymbol{\phi}$  e  $\widehat{\boldsymbol{\phi}}=g(\widehat{\gamma}, \widehat{\alpha})$ . Se além de estimar  $\boldsymbol{\phi}$ , existir interesse em construir um intervalo de confiança, é necessário obter uma estimativa para o erro padrão de  $\widehat{\boldsymbol{\phi}}$ . Isto é feito usando-se o método delta,

$$Var(\widehat{\boldsymbol{\phi}}) \doteq Var(\widehat{\alpha}) \left( \frac{\partial \boldsymbol{\phi}}{\partial \alpha} \right)^2 + 2Cov(\widehat{\alpha}, \widehat{\gamma}) \left( \frac{\partial \boldsymbol{\phi}}{\partial \alpha} \right) \left( \frac{\partial \boldsymbol{\phi}}{\partial \gamma} \right) + Var(\widehat{\gamma}) \left( \frac{\partial \boldsymbol{\phi}}{\partial \gamma} \right)^2$$

### 2.4.3 Teste de hipótese

Teste de hipótese é utilizado para modelos relacionados com um vetor  $\boldsymbol{\theta}=(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_n)'$  de parâmetro ou um subconjunto dele. E a comparação entre ajustes de modelos de sobrevivência aos dados, quando os mesmos são hierárquicos, é facilmente avaliada por meio de testes formais de ajuste, como o de Wald, o Razão de Verossimilhança e o Escore.

Segundo Colosimo e Giolo (2006), o teste Wald é baseado na distribuição assintótica de  $\widehat{\boldsymbol{\theta}}$  e é uma generalização do teste t de Student (Wald, 1943). É conhecida por testar um único parâmetro  $\boldsymbol{\theta}_j$ , tendo aproximadamente uma distribuição qui-quadrado com  $p$  graus de liberdade ( $\chi_p^2$ ).

O teste da razão de máxima verossimilhança é a comparação entre o modelo paramétrico e o sub-modelo o modelo Weibull. Segundo Strapasson (2007), pelo teste da razão verossimilhança, Gomes (2005) faz uma discriminação entre o modelo proposto por Freitas, Borges e Ho (2003) que é baseado no modelo Weibull e o modelo de risco proporcionais, proposto por ela.

As estatísticas da razão de máxima verossimilhança, TRV e do teste escore,  $S$  é obtida pelas hipóteses de interesse definidas por

$$\begin{cases} H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0, & \text{o modelo Weibull esta adequado aos dados} \\ H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0, & \text{o modelo Weibull não esta adequado aos dados} \end{cases}$$

Para testar  $H_0$  pode-se utilizar a estatística de razão de verossimilhança, definida por Colosimo e Giolo (2006),

$$TRV = -2 \log \left[ \frac{L(\widehat{\boldsymbol{\theta}}_0)}{L(\widehat{\boldsymbol{\theta}})} \right] = 2[\log L(\widehat{\boldsymbol{\theta}}) - \log L(\widehat{\boldsymbol{\theta}}_0)] \quad (2.27)$$

em que, sob  $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ , dado  $\widehat{\boldsymbol{\theta}}$  é o modelo geral e  $\widehat{\boldsymbol{\theta}}_0$  modelo de interesse, seguem aproximadamente uma distribuição qui-quadrado com  $p$  graus de liberdade. Como critério de decisão, a hipótese  $H_0$  é rejeitada, a um nível de  $100\alpha\%$  de significância se  $S > \chi_{p,1-\alpha}^2$ .

Na presença de covariáveis é interessante avaliar a significância das mesmas para o ajuste do modelo. A partir das estatísticas da razão de verossimilhança e escore, em que temos  $\{\boldsymbol{\theta} \mid TRV(\boldsymbol{\theta}) < \chi_{p,1-\alpha}^2\}$  é intervalo de  $(1 - \alpha)$  100% de confiança para  $\boldsymbol{\theta}$ .

E por meio da estatística escore em que pode-se construir intervalo de confiança, expressado por (COLOSIMO, 2006).

$$S = U'(\boldsymbol{\theta}_0)[\mathfrak{F}(\boldsymbol{\theta}_0)]^{-1}U(\boldsymbol{\theta}_0)$$

em que  $U(\boldsymbol{\theta}_0)$  é a função escore  $U(\boldsymbol{\theta}) = \frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$  avaliada em  $\boldsymbol{\theta}_0$ , e  $\mathfrak{F}(\boldsymbol{\theta}_0)$  a matriz de variância-covariância observada de  $\widehat{\boldsymbol{\theta}}$  também avaliada por  $\boldsymbol{\theta}_0$ . Segundo Colosimo e Giolo (2006), as três estatísticas de teste podem ser adaptadas para o caso em que se tenha interesse somente em um subconjunto de  $\boldsymbol{\theta}$ .

Para Elizabeth (2007) e Cordeiro (1992), as estatísticas da razão de verossimilhança (TRV) e do teste de escore ( $S$ ) são assintoticamente equivalentes, sob a hipótese nula  $H_0$ , à distribuição qui-quadrado. O problema de escolha entre elas surge quando a estimação segundo ambas as hipóteses apresentar o mesmo grau de dificuldade. As estatísticas (TRV) e ( $S$ ) são invariantes em relação à parametrização da distribuição dos dados.

## 2.5 Distribuição Log Weibull Inversa Generalizada

Segundo Gusmão et al. (2011), a distribuição Log-Weibull Inversa Generalizada (LGIW) tem uma função de densidade, cuja variável aleatória  $Y = \log(T)$  tem parametrização  $\sigma = \frac{1}{\beta}$  e  $\mu = \log(\alpha)$  expressa por:

$$f(y; \gamma, \sigma, \mu) = \frac{\gamma}{\sigma} \exp \left\{ - \left( \frac{y - \mu}{\sigma} \right) - \gamma \left[ - \left( \frac{y - \mu}{\sigma} \right) \right] \right\}, \quad -\infty < y < \infty \quad (2.28)$$

em que  $\gamma > 0$ ,  $\sigma > 0$  e  $-\infty < \mu < \infty$ . Tem uma grande flexibilidade de densidade para o novo parâmetro  $\gamma$ .

A função de sobrevivência correspondente (2.28) é

$$S(y) = 1 - \exp \left\{ -\gamma \exp \left[ - \left( \frac{y - \mu}{\sigma} \right) \right] \right\} \quad (2.29)$$

A variável aleatória padronizada  $Z = \frac{Y-\mu}{\sigma}$  com densidade

$$\pi(z; \gamma) = \gamma \exp[-z - \gamma \exp(-z)], \quad -\infty < z < \infty \quad e \quad \gamma > 0, \quad (2.30)$$

em que, a distribuição segue um distribuição valor extremo inverso, segue como um caso especial quando  $\gamma = 1$ .

## 2.6 Modelo probabilístico

Segundo Colosimo e Giolo (2006), a proposta empírica consiste em ajustar os modelos probabilístico (Exponencial, Weibull), com base na comparação entre valores estimados e observados, decidir qual deles “melhor” explica os dados amostrais. A forma mais simples e eficiente de selecionar o “melhor” modelo a ser usado é dado por meio de técnicas gráficas.

### 2.6.1 Método Gráfico

O método gráfico consiste em dois métodos: o primeiro é a comparação da função de sobrevivência proposto pelo estimador de Kaplan-Meier, ou seja a curva de sobrevivência se aproxima do estimador de Kaplan-Meier e o segundo consiste na linearização da função de sobrevivência tendo a construção de gráficos aproximadamente lineares. A função de sobrevivência ou a função da taxa de falha acumulada são úteis para discriminar modelos por meio do gráfico. A ideia é comparar estas funções com estimador de Kaplan-Meier e selecionar o modelo cuja curva melhor se aproxima da curva de Kaplan-Meier.

## 3 Aplicação

Nesta seção do trabalho, são apresentados um conjunto de dados relacionados com dependência química, para ilustrar o uso da distribuição de WIG. Todos os resultados foram obtidos nos *softwares* SAS versão 9.1 (SAS institute inc, 2006) e R versão 2.15.2 (R development core team, 2012). O conjunto de dados reais refere-se a 141 residentes dependentes químicos, no período de 2000 a 2005. Fornecido por Pascoa (2008) por meio da Associação Mãe Admirável, situada na cidade de Caratinga-MG.

Esse conjunto de dados foi utilizado por Pascoa (2008), em credibilidade para a razão de risco. Os dados cujo tamanho é  $n=141$ , representam a resistência na comunidade até a desistência do tratamento, considerando-se que cada residente mantém-se na comunidade por um período de 270 dias, sem obter nenhum contato com drogas. Quem atingir essa meta é considerado, neste trabalho, como um dado censurado. A variável resposta foi o tempo de permanência na comunidade até a desistência do tratamento.

A dependência química é uma doença progressiva, incurável e fatal, por ser, uma doença crônica do cérebro. Uma das principais características é o episódio da recaída, em que o indivíduo entrega-se ao consumo da substância em questão, seguida por uma nova tentativa de alcançar ou restringir este consumo (BRANDON, VIDRINE e LITVIN, 2007, KOOB, LE MOAL, 1997). A recaída não exprime falha no tratamento, ao contrário: é um segmento do processo de reabilitação (MARLATT, 2001). Ao entrar em recaída, o paciente que apresenta débito no conjunto de procedimento disfuncional apresenta princípios antes do tratamento, ou seja, a recaída começa antes de se recuperar.

Segundo Pascoa (2008) os fatores de risco associados a recaída em dependência química são muito importantes, pois aumentam as consequências iniciais de interferência clínica. Contudo, estudos de risco concedem a identificação de fatores de defesa durante a resistência da fragilidade à desejo de recaída.

As variáveis envolvidas neste caso são:

- i)**  $t_i$ : tempo de permanência na comunidade até a desistência do tratamento (dias);
- ii)**  $cens_i$ : o indicador de censura (0= censura, 1= falha);
- iii)**  $x_{i1}$ : problemática (0= álcool ou outro tipo de droga, 1= álcool e outro tipo de droga);
- iv)**  $x_{i2}$ : estado civil do residente (0= solteiro, 1= casado ou separado);
- v)**  $x_{i3}$ : filhos (0= não tem filhos, 1= tem filhos);
- vi)**  $x_{i4}$ : estado civil dos pais (0= solteiro, 1= casado ou separado);
- vii)**  $x_{i5}$ : fonte de renda própria (0=não tem fonte de renda, 1=tem fonte de renda);
- viii)**  $x_{i6}$ : escolaridade (0=nenhuma ou ensino fundamental incompleto, 1= ensino fundamental completo ou ensino superior completo ou não);
- ix)**  $x_{i7}$ : idade ( $0 < 30 \text{ anos}$ ,  $1 \geq 30 \text{ anos}$ ).
- x)**  $x_{i8}$ : diversidade de drogas (0= uma, 1= mais de uma);
- xi)**  $x_{i9}$ : caso de droga / álcool na família (0= não tem, 1= tem);
- xii)**  $x_{i10}$ : tratamento anterior (0= não tem, 1= tem);
- xii)**  $x_{i11}$ : problemas com justiça (0= não tem, 1= tem);
- xiv)**  $x_{i12}$ : tempos de uso (em anos) ( $0 < 15 \text{ anos}$ ,  $1 \geq 15 \text{ anos}$ );

em que  $i$  representa o  $i$ -ésimo indivíduo.

O estudo foi realizado com o objetivo principal de ajustar a distribuição WIG ao conjunto de dados de dependentes químicos fornecido por Pascoa (2008). Na Tabela 1 tem-se um layout dos dados.

Tabela 1: Representação dos dados de dependentes químicos da cidade de Caratinga - MG

Observações	Covariáveis													
	Tempo	Censura	$x_{i1}$	$x_{i2}$	$x_{i3}$	$x_{i4}$	$x_{i5}$	$x_{i6}$	$x_{i7}$	$x_{i8}$	$x_{i9}$	$x_{i10}$	$x_{i11}$	$x_{i12}$
1	7	1	0	0	0	0	1	1	1	0	1	0	0	1
2	25	1	0	1	1	0	0	1	1	0	1	1	1	0
3	48	1	0	1	1	0	0	1	1	0	1	1	0	1
4	55	1	0	0	1	0	0	1	1	0	1	0	0	1
5	76	1	1	0	0	1	0	0	0	1	1	0	0	0
⋮														
100	270	0	0	1	1	0	1	0	1	0	1	0	1	1
101	270	0	1	0	0	0	1	1	0	1	0	1	1	1
102	270	0	0	0	0	0	0	0	1	0	1	0	0	0
103	270	0	0	1	1	0	0	1	1	0	1	0	0	1
104	270	0	0	0	0	0	0	1	1	0	1	0	1	1
⋮														
141	270	0	0	0	0	0	0	0	1	0	0	1	0	1

A análise descritiva consiste essencialmente em encontrar medidas de tendência central e variabilidade, de forma que se possa caracterizar o banco de dados, para uma melhor compreensão do mesmo. A Tabela 2 refere-se a caracterização do banco de dados.

Tabela 2: Estatística Descritiva das variáveis dos dados de recorrentes químicos

Covariáveis	Média	Mediana	Moda	Desvio Padrão	Variância	C.Variação	Assimetria	Curtose
Tempo	132,21	92,00	270,00	106,38	11317	80,45	0,26	-1,64
Censura	0,70	1	1	0,45	0,21	65,36	-0,89	-1,21
xi1	0,18	0	0	0,38	0,15	211,06	1,64	0,71
xi2	0,41	0	0	0,49	0,24	120,05	0,36	-1,89
xi3	0,57	1	1	0,49	0,24	86,37	-0,30	-1,93
xi4	0,16	0	0	0,37	0,13	227,31	1,84	1,41
xi5	0,24	0	0	0,42	0,18	178,03	1,22	-0,51
xi6	0,29	0	0	0,45	0,21	154,07	0,89	-1,21
xi7	0,63	1	1	0,48	0,23	75,54	-0,58	-1,68
xi8	0,29	0	0	0,45	0,21	154,07	0,89	-1,21
xi9	0,67	1	1	0,47	0,22	69,83	-0,74	-1,45
xi10	0,25	0	0	0,43	0,19	171,39	1,13	-0,72
xi11	0,35	0	0	0,48	0,23	135,38	0,61	-1,64
xi12	0,58	1	1	0,49	0,24	83,89	-0,36	-1,89

Pelos resultados da Tabela 2, percebe-se que o coeficiente de variação é alto para todas as variáveis, destacando-se a variável censura (com valor obtido de 65,36) sendo este o menor coeficiente de variação em comparação aos demais. Algumas variáveis têm uma assimetria negativa, sendo elas as covariáveis censura,  $x_{i3}$ ,  $x_{i7}$ ,  $x_{i9}$  e  $x_{i12}$ . Em relação ao coeficiente de curtose, verifica-se que a variável  $x_{i1}$  é platicúrtica e as demais são leptocúrticas. A quantidade de indivíduos censurados foi de 99 e os que falharam foram 42. Para o ajuste da distribuição WIG ao banco de dados, considerou-se apenas as variáveis tempo e censura.

De início ajustou-se a distribuição Weibull inversa, por esta se tratar de um caso particular da WIG. Utilizando-se a função nlmixed SAS 9.1 (SAS inc., 2002), iniciou-se o algoritmo iterativo com dois valores iniciais arbitrários ( $\alpha = 0,1$  e  $\beta = 0,1$ ). A Tabela 3 apresenta as estimativas dos parâmetros com respectivos erros padrão e estatística  $t$  para os parâmetros  $\alpha$  e  $\beta$ .

Tabela 3: Estimativas dos Parâmetros e erros padrão da Weibull Inversa aplicada a dados de recorrentes químicos

Parâmetros	Estimativa	Erro Padrão	Teste t	Valor p
$\alpha$	48,2783	7,8285	6,17	< 0,001
$\beta$	0,5631	0,04235	13,30	< 0,001

Pela estatística  $t$  no ajuste na Weibull Inversa os parâmetros  $\alpha$  e  $\beta$  foram significativos (*valor p* < 0,01). Desde modo sugeriu-se estas estimativas de  $\alpha$  e  $\beta$  como valores iniciais a WIG, entretanto, a mesma não convergiu, partiu-se então para novos valores iniciais por tentativa e erro até que a convergência fosse atingida. Na Tabela 4, tem-se as estimativas dos parâmetros, erros padrão e estatística  $t$  para a WIG.

Tabela 4: Estimativas dos Parâmetros e erros padrão da Weibull Inversa Generalizada aplicada a dados de recorrentes químicos

Parâmetro	Estimativas	Erro Padrão	Estatística t	Valor P
$\alpha$	200,14	0,000072	2775103	< 0,001
$\beta$	0,5631	0,04235	13,30	< 0,001
$\gamma$	0,4490	0,05709	7,86	< 0,001

Por meio da Figura 3 o modelo WIG é apropriado para os dados de recorrentes químicos, uma vez que a curva de  $-\log$  da função de sobrevivência estimada em relação ao tempo de sobrevivência se aproxima de uma reta que passa pela origem, sendo este um indício de que a distribuição está bem ajustada. Nota-se que o gráfico do  $\log$  de  $-\log$  da função de sobrevivência versus o logaritmo do tempo é praticamente uma reta (Figura 3).

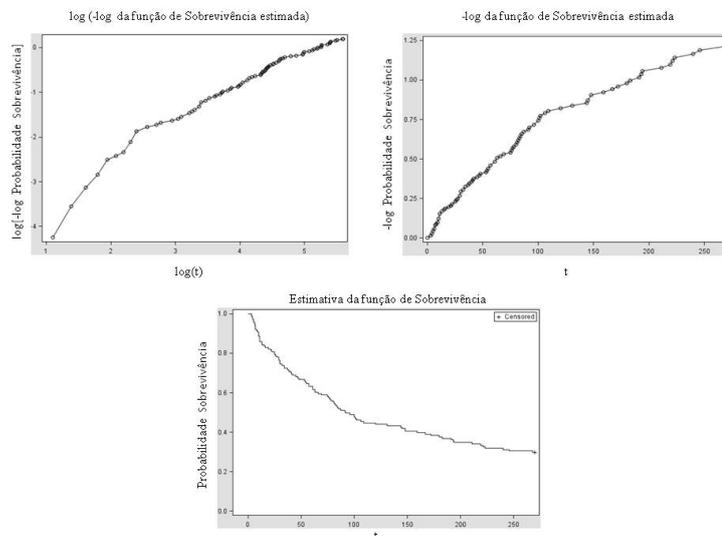


Figura 3: Análise de Sobrevivência aos dados de recorrentes químicos.

A curva TTT ajustada aos dados de recorrentes químicos teve um comportamento que a caracteriza como risco unimodal (Figura 4), então a função de risco ajustada aos dados ficou representada como na Figura 5.

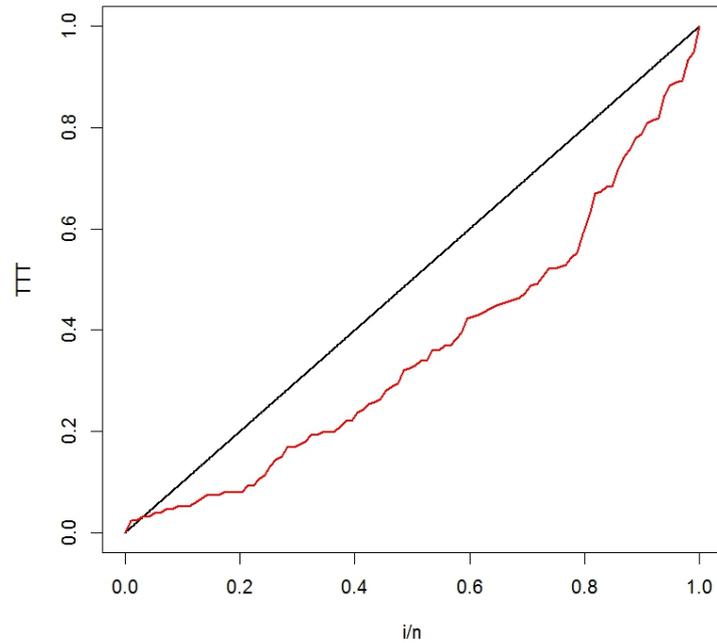


Figura 4: Tempo Total em Teste (Curva TTT) aos dados de recorrentes químicos.

Como a função de risco foi unimodal, a taxa de falha não é constante ao longo do tempo.

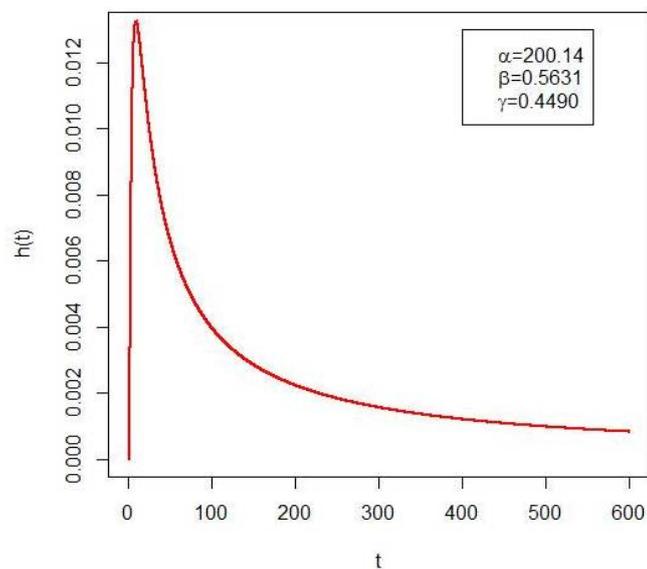


Figura 5: Risco Ajustado as variáveis.

Como discutido em Lee (1992), o gráfico da função de risco estimada observa-se que a taxa de risco é mais elevada nos primeiros dias de diagnóstico. A partir do final do centésimo dia até o fim dos duzentos e setenta dias do ano, a taxa de risco permanece relativamente constante, oscilando entre 0,02 e 0,04. A taxa de risco é geralmente mais alta após o décimo dia. Isso pode indicar que um paciente que sobreviveu aos duzentos e setenta dias tem uma chance melhor do que um paciente que acaba de ser diagnosticado. O perfil das vidas médias residuais também apoia esta interpretação.

O ajuste do modelo aos dados foi checado por meio da curva de sobrevivência estimada, em que, se a mesma ficar cruzando com o Kaplan-Meier, é um indício que o modelo se ajustou bem aos dados. Como pode-se observar na Figura 6 a sobrevivência estimada ficou justaposta com a curva de Kaplan-Meier indicando-se um bom ajuste do modelo aos dados.

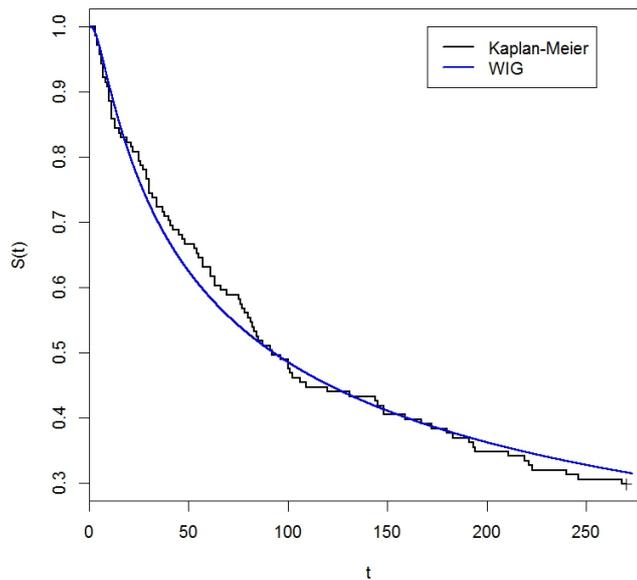


Figura 6: Ajuste da Kaplan-Meier e a WIG aos dados de recorrentes químicos.

## 4 Conclusão

A distribuição WIG proposta por (Gusmão et al., 2011) ajustou-se bem aos dados de recorrentes químicos da cidade de Caratinga - MG , com sua função de risco caracterizada por uma distribuição unimodal, em que, observa-se que a taxa de risco foi mais elevada após o décimo dia de diagnóstico, e a partir do centésimo dia até o fim dos duzentos e setenta dias, a taxa de risco permaneceu relativamente constante. Sendo a WIG uma distribuição bastante flexível para o risco estimado, ela torna-se um excelente modelo para ser comparado com as distribuições mais tradicionais da área de análise de sobrevivência. É possível considerar que a principal vantagem da WIG é a flexibilidade da sua função de taxa de falha, de forma a deixar o modelo mais maleável e com a possibilidade de ajustá-la a variados conjuntos de dados em análise de sobrevivência.

## 5 Propostas Futuras

- Pretende-se ajustar a distribuição Log Weibull Inversa Generalizada aos dados de recorrentes químicos da cidade de Caratinga no Estado de Minas Gerais;
- Ajustar ambas distribuições em um modelo de regressão, utilizando-se as covariáveis associadas ao tempo de falha dos dependentes químicos;
- Realizar comparação de modelos casos particulares da WIG e da LWIG por meio de critérios de informação de modelos (AIC e BIC);
- Fazer um estudo dos resíduos do Modelo ajustado.

## 6 Referências Bibliográficas

AASET, M.V. . “How to indentify bathtub harzard rate”. Transactions on Reliability, New York, v. 36, 106-108 p., 1987.

BRANDON, T.H.; VIDRINE, J.I.; LITVIN, E.B. “Relapse and Relapse Prevention.” Annual Review of Psychology, York, v.3, 257-84p., 2007.

CÉSAR, K.A. . “Análise estatística de sobrevivência: Um estudo com pacientes com câncer de mama”. UCB, graduação do curso de matemática pela Universidade Católica de Brasília, 100-103 p., 2005.Disponível em: [www.ucb.br/sites/100/103/tcc/12005/KellyAraujoCesar.pdf](http://www.ucb.br/sites/100/103/tcc/12005/KellyAraujoCesar.pdf) Acesso em: 18-03-2012 às 16:07.

COLOSIMO, E.A.; GIOLO, S.R. , **Análise de Sobrevivência Aplicada**. Projeto Fisher - ABE. São Paulo: Edgard Blucher Ltda., 1ª edição 354 p., 2006.

CORDEIRO, G.M.. “Introdução à teoria de verossimilhança.”In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, v.10., Resumos Rio de Janeiro: UFRJ, 174p., 1992.

COX, D.R. “Regression models and Life Tables”. Journal os the Royal Statistical Society B,34, 187-220p., 1972. Disponível em: [www.ressources-actuarielles.net](http://www.ressources-actuarielles.net) Acesso em: 01/12/2012 ás 17:07.

COX, D.R.; HINKLEY, D.V.. “Theoretical Statistics.”Chapman and Hall, London, 551p., 1974.

DRAPELLA, A. “Complementary Weibull distribution: unknown or just forgotten.”Qual Reliab Eng Int 9: 383-385p., 1993.

FREITAS, M.A.; BORGES, W.; HO, L.L. “A statistical model for shelf life estimation using sensory evaluations scores.”Communications in Statistics - Theory and Methods, New York, v.32, n.8, 1559-1589p., 2003.

GOMES, R.C.D., “Estimando o tempo de vida de produto em prateleira utilizando modelo

de riscos proporcionais em dados oriundos de avaliações sensoriais.” Universidade Federal de Minas Gerais, Belo Horizonte, 178p., 2005. Tese de Mestrado

GUSMÃO, F.R.S.; ORTEGA, E.M.M.; CORDEIRO, G.M. . **The generalized inverse Weibull distribution**. Stat Papers, v. 52, 591 - 619 p., 2011.

JIANG, R.; ZUO, M.J.; LI, H.X. “Weibull and Weibull inverse mixture models allowing negative weights.” Reliab Eng Syst Saf 66:227-234p., 1999.

KOOB, G.F.; LE MOAL, M. “Drug abuse: hedonic homeostatic dysregulation.” Science, New York, v.278, 52-58p., 1997.

LEE, E. T., **Statistical Methods for Survival Data Analysis**. 2<sup>nd</sup>. New York: John Wiley & Sons, 1992.

LOUZADA, F.; DINIZ, C., **Modelagem Estatística para Risco de Crédito**. São Paulo: COPYRIGHT, 2<sup>a</sup> edição 172 p., 2012.

MARLATT, G.A. “Integrating contingency management with relapse prevention skills: comment on Silverman et al. 2001.” Experimental and Clinical Psychopharmacology, Washington, v.9, 33-34p., 2001.

MUDHOLKAR, G.S.; SRIVASTAVA, D.K.; KOLLIA, G.D.; **A generalization of the Weibull**. Statistical Association, New York, v.91, 1575-1583 p., 1996.

PASCOA, M.A.R de. **Intervalos de credibilidade para razão de risco do modelo de Cox, considerando estimativas pontuais *bootstrap***. Minas Gerais, UFLA 84 p., 2008. Tese de Mestrado.

STRAPASSON, E. . “Comparação de modelos com censura intervalos em análise de sobrevivência”. Piracicaba, ESALQ 135 p., 2007. Tese de Doutorado.

WALD, A. **Tests of Statistical Hypotheses concerning Several Parameters when the Number of Observations is Large**. Trans. Amer. Math. Soc., 54, 426-482p., 1943.

WEIBULL, W. “A Statistical Theory of the Strength of Materials.” Ingeniors Vetenskaps Akademien Handlingar, n.151: The Phenomenon of Rupture in Solid, 293-297p., 1939.

WEIBULL, W. “A Statistical Distribution of Wide Applicability.” Journal of Applied Mechanics, 18, 293-297, 1951.

WEIBULL, W. “A Statistical Representation of Fatigue Failure in Solids.” Royal Institute of Technology, 1954.

## APÊNDICE A – Primeiro apêndice (SAS)

```

/*ods html;*/ proc print data=dq; run; /*ods html close;*/ proc univariate data=dq;
run; proc lifetest graphics plots=(s,ls,lls);/ time t*censur(0); run;

/*Weibull inversa*/

proc nlmixed cov data=dq;

parms alfa=0.1 beta=1; bounds alfa>0; bounds beta>0;

logp=censur*(log(beta)+beta*log(alfa)-(beta+1)*log(t)-((alfa/t)**beta))+
(1-censur)*(log(1-exp(-((alfa/t)**beta))));

model t general(logp); run;

/*Weibull inversa generalizada*/

proc nlmixed cov data=dq;

parms alfa=200 beta=1 gama=0.01; bounds alfa>0; bounds beta>0; bounds gama>0;

logp=censur*(log(gama)+log(beta)+beta*log(alfa)-
(beta+1)*log(t)-(gama*((alfa/t)**beta)))+(1-censur)*(log(1-exp(-(gama*(alfa/t)**beta))));

model t general(logp); run;

```

## APÊNDICE B – Segundo apêndice (R)

```

/*Sobrevivencia ajustada como o kaplan-meier*/

dados ← read.table('dados_aq.txt', head = T) dados require(survival) tempos=dados[,13]
cens=dados[,14] ekm=survfit(Surv(tempos,cens) 1) time=seq(0,273,0.01)

/*WIG*/

alfa=200.14 beta=0.5631 gama=0.4490

s = 1 - exp(-(gama * (alfa/time)^beta))

plot(ekm,conf.int=F,ylim=c(0.3,1), xlab="t",ylab="S(t)") lines(ekm,col='black', lty=1,
lwd=2, bty="n") lines(time,s,col='blue', lty=1, lwd=2)

legend(170,1,c('Kaplan-Meier','WIG'),bty='o',cex=1, col=c('black','blue'), lty=c(1,1),
lwd=c(2,2))

/*Risco Ajustado*/

plot(c(0,600), c(0,1), type="n",ylim=c(0,0.013),
xlab="t", ylab=expression(paste("h(t)"))) t_i=seq(0,600,.01)

alfa=200.14 beta=0.5631 gama=0.4490

h = gama * beta * (alfa^beta) * (t - (beta + 1)) * exp(-gama * ((alfa/t)^beta)) * ((1 -
exp(-gama * ((alfa/t)^beta)))^-1)

/*title("a)*/

lines(t,h,col='red', lty=1, lwd=2)

legend(400,0.013, expression(paste(alpha,"=200.14"), paste(beta,"=0.5631"),
paste(gama,"=0.4490")),bty="o",cex=1)

/*Gráfico TTT-Plot*/

i=scan() 0 0.01010101 0.02020202 0.03030303 0.04040404 0.05050505 0.06060606 0.07070707

```

```

0.080808081 0.090909091 0.101010101 0.111111111 0.121212121 0.131313131 0.141414141
0.151515152 0.161616162 0.171717172 0.181818182 0.191919192 0.202020202 0.212121212
0.222222222 0.232323232 0.242424242 0.252525253 0.262626263 0.272727273 0.282828283
0.292929293 0.303030303 0.313131313 0.323232323 0.333333333 0.343434343 0.353535354
0.363636364 0.373737374 0.383838384 0.393939394 0.404040404 0.414141414 0.424242424
0.434343434 0.444444444 0.454545455 0.464646465 0.474747475 0.484848485 0.494949495
0.505050505 0.515151515 0.525252525 0.535353535 0.545454545 0.555555556 0.565656566
0.575757576 0.585858586 0.595959596 0.606060606 0.616161616 0.626262626 0.636363636
0.646464646 0.656565657 0.666666667 0.676767677 0.686868687 0.696969697 0.707070707
0.717171717 0.727272727 0.737373737 0.747474747 0.757575758 0.767676768 0.777777778
0.787878788 0.797979798 0.808080808 0.818181818 0.828282828 0.838383838 0.848484848
0.858585859 0.868686869 0.878787879 0.888888889 0.898989899 0.909090909 0.919191919
0.929292929 0.939393939 0.949494949 0.95959596 0.96969697 0.97979798
0.98989899 1

```

```

/* ttt=scan()

```

```

0 0.022792176 0.022792176 0.030281804 0.030281804 0.037663667 0.037663667 0.044937766
0.044937766 0.0521041 0.0521041 0.0521041 0.059108788 0.066059594 0.0729565 0.0729565
0.0729565 0.079691794 0.079691794 0.079691794 0.079691794 0.092731289 0.092731289
0.105555256 0.111913357 0.130826014 0.143326688 0.149523142 0.167950859 0.167950859
0.173985667 0.179966593 0.19182068 0.19182068 0.197639959 0.197639959 0.197639959
0.208955224 0.220162724 0.220162724 0.236650682 0.242092785 0.252869228 0.258203567
0.263484024 0.279163748 0.289509133 0.294627943 0.319952584 0.32496363 0.329920793
0.339727356 0.339727356 0.358909424 0.358909424 0.368284929 0.368284929 0.382024894
0.395603211 0.422436554 0.426854895 0.431219354 0.435529932 0.444043321 0.448246134
0.452395064 0.456490113 0.460531279 0.464518562 0.472385366 0.487903443 0.49172908
0.5068161 0.521687591 0.521687591 0.525297699 0.528853925 0.542863301 0.553208686
0.59054906 0.627296729 0.670025325 0.673258257 0.682795409 0.682795409 0.716579557
0.740718789 0.755536397 0.778813514 0.787380786 0.809795786 0.815291772 0.817985883
0.862869767 0.883560537 0.888625465 0.893582628 0.934802522 0.949027426 1

```

```

/*

```

```

y=seq(0,1,0.001) x=seq(0,1,0.001)

```

```

plot(i,ttt, type="n", xlab="i/n", ylab="TTT") /*title("(a)") lines(x,y,col='black',
lty=1, lwd=2) lines(i,ttt,col='red', lty=1, lwd=2) lines(i,ttt,col='red', lty=1, lwd=2)

```