



**UEPB**

**UNIVERSIDADE ESTADUAL DA PARAÍBA  
CAMPUS I  
CENTRO DE CIÊNCIAS E TECNOLOGIAS  
DEPARTAMENTO DE ESTATÍSTICA  
CURSO DE GRADUAÇÃO EM ESTATÍSTICA**

**JOSEFA ANUCIADA PEREIRA DA SILVA**

**FAMILIA GAMA GENERALIZADA APLICADA A DADOS DE SOBREVIVÊNCIA**

**CAMPINA GRANDE  
2021**

JOSEFA ANUCIADA PEREIRA DA SILVA

**FAMILIA GAMA GENERALIZADA APLICADA A DADOS DE SOBREVIVÊNCIA**

Trabalho de Conclusão de Curso apresentado ao curso de bacharelado em Estatística da Universidade Estadual da Paraíba, em cumprimento as exigências legais para à obtenção do título de Bacharel em Estatística.

**Orientador:** Prof. Dr. Tiago Almeida de Oliveira.

**Coorientador:** Prof. Dr. Marcelino Alves Rosa de Pascoa.

**CAMPINA GRANDE  
2021**

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

S586f Silva, Josefa Anuciada Pereira da.  
Família gama generalizada aplicada a dados de sobrevivência [manuscrito] / Josefa Anuciada Pereira da Silva. - 2021.  
34 p.  
  
Digitado.  
Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2021.  
"Orientação : Prof. Dr. Tiago Almeida de Oliveira, Coordenação do Curso de Estatística - CCT."  
1. Gama generalizada. 2. Análise de sobrevivência. 3. Máxima verossimilhança. 4. Distribuição de Weibull. I. Título  
21. ed. CDD 519.5

JOSEFA ANUCIADA PEREIRA DA SILVA

FAMILIA GAMA GENERALIZADA APLICADA A DADOS DE SOBREVIVÊNCIA.

Trabalho de Conclusão de Curso apresentado ao curso de bacharelado em Estatística da Universidade Estadual da Paraíba, em cumprimento as exigências legais para à obtenção do título de Bacharel em Estatística.

Aprovada em: 30 /11/2021.

**BANCA EXAMINADORA**



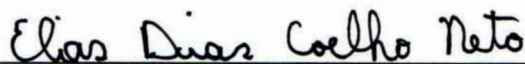
---

Prof. Dr. Tiago Almeida De Oliveira(Orientador)  
Universidade Estadual da Paraíba (UEPB)



---

Prof. Dr. Marcelino Alves Rosa De Pascoa(Coorientador)  
Universidade Federal De Mato Grosso (UFMT)



---

Prof. Dr. Elias Dias Coelho Neto  
Universidade Estadual da Paraíba (UEPB)

Ao meu DEUS, pela Força e Fé em mim  
depositadas por todos esses anos,  
DEDICO.

## **AGRADECIMENTOS**

Primeiramente, quero agradecer a Deus por estar comigo em cada decisão que eu tome me dando forças para conseguir concretizar meus objetivos.

A minha mãe, por ter me educado e me ensinado a ser a mulher que sou, a buscar coragem para realizar meus propósitos.

Ao meu pai, por todas as vezes que me acompanhava ao sair e ao chegar em casa, dias de frio, dias de calor, deixando seus afazeres para me ajudar.

A minha irmã, por todo apoio psicológico, todo amor e afeto.

Ao meu irmão, por não me deixar passar aperto financeiro.

Ao meu marido, por ter chegado na minha vida, por me mostrar a importância de se ter amigos, de me incentivar a concluir uma graduação e buscar a independência financeira por meio da realização profissional.

Claro que eu não poderia esquecer de uma pessoa muito especial na minha vida, a minha grande amiga Elizandra (e a sua família também), que esteve comigo em todos os momentos de tristeza e alegria da minha vida acadêmica, me acolheu, me ajudou. Isso me faz deixar um recadinho para você que está lendo, tenha um amigo porque a vida acadêmica é muito difícil, não fique sozinho.

Ao professor Tiago, por ter me orientado, por estar aqui deixando sua parcela de contribuição para a construção do meu futuro.

Ao professor Marcelino, por ter contribuído com todo o seu conhecimento sobre o tema, me motivando a concluir este trabalho.

## RESUMO

No presente trabalho é feita uma apresentação da gama generalizada exponenciada (GGE) introduzida por Cordeiro *et al.* (2009). A distribuição GGE possui quatro parâmetros e é geralmente utilizada em análise de sobrevivência. Após a apresentação de alguns casos particulares da GGE, estimaram-se seus parâmetros pelo método da máxima verossimilhança. Os ajustes foram comparados com o modelo da distribuição gama generalizada (GG) (STACY, 1962), e também comparado com o da distribuição Weibull. A utilidade do modelo proposto é ilustrada com a aplicação a dois conjuntos de dados de sobrevivência, um relativo ao valor médio do *Nursing Activities Score* (NAS) e o outro contendo dados censurados relativos à saúde (AIDS). A comparação dos ajustes mostrou que a distribuição gama generalizada exponenciada, apresentou melhor ajuste para os dados em estudo, segundo o teste da razão de verossimilhança e as estatísticas AIC, BIC e CAIC, se mostrando mais flexível quando comparada com a distribuição GG e a Weibull.

**Palavras-chave:** Gama Generalizada. Gama Generalizada Exponenciada. Weibull. Análise de Sobrevivência. Máxima Verossimilhança.

## ABSTRACT

In the present work, a presentation of the generalized exponentiated gamma (GGE) introduced by Cordeiro et al. (2009). The GGE distribution has four parameters and is generally used in survival analysis. After the presentation of some particular cases of the GGE, its parameters were estimated by the maximum likelihood method. The fits were compared with the generalized gamma distribution (GG) model (STACY, 1962), and also compared with the Weibull distribution. The usefulness of the proposed model is illustrated by applying two sets of survival data, one relating to the mean value of the Nursing Activities Score (NAS) and the other containing censored data relating to health (AIDS). The comparison of fits showed that the exponentiated generalized gamma distribution presented a better fit for the data under study, according to the likelihood ratio test and the AIC, BIC and CAIC statistics, being more flexible when compared to the GG and Weibull distributions.

**Keywords:** Generalized Gamma. Exponentiated Generalized Gamma. Weibull. Survival Analysis. Maximum Likelihood.



## LISTA DE ILUSTRAÇÕES

- Figura 1 – Gráfico da função densidade da distribuição GG para diferentes 18 parâmetros.
- Figura 2 – Gráfico da função de risco e da função de sobrevivência da 19 distribuição GG.
- Figura 3 – Gráfico da densidade da GGE com parâmetros. 21
- Figura 4 – Gráfico da densidade da GGE com valores. 21
- Figura 5 – Gráfico da função de sobrevivência da GGE. 22
- Figura 6 – Gráfico da função de taxa de falha assumindo as formas 23 crescente, decrescente e constante.
- Figura 7 – Gráfico da função de taxa de falha assumindo as formas U e 23 unimodal.
- Figura 8 – Gráfico da função de taxa de falha da GGE na forma crescente 28 ajustada aos dados de NAS.
- Figura 9 – (a)Densidade estimadas dos modelos para os dados NAS e 29 (b)estimativas da função de sobrevivência klapan-Méier, GGE, GG e weibull.
- Figura 10 – Gráfico da função de taxa de falha da GGE na forma unimodal 30 ajustada aos dados de saúde.
- Figura 11 – (a)Densidades estimadas dos modelos para os dados de AIDS e 31 (b)ajustes das curvas de sobrevivência da GGE, GG e weibull comparadas ao klapan-Méier.

## LISTA DE TABELAS

Tabela 1 – Ajuste final dos modelos comparados, para os dados de NAS	27
Tabela 2 – Teste de razão de verossimilhança, para os dados de NAS	28
Tabela 3 – Ajuste final dos modelos comparados	29
Tabela 4 – Teste de razão de verossimilhança, para os dados de AIDS	30

## LISTA DE ABREVIATURAS E SIGLAS

AIDS	Síndrome da imunodeficiência adquirida
GG	Gama Generalizada
GGE	Gama generalizada exponenciada
NAS	<i>Nursing Activities Score</i>

## LISTA DE SÍMBOLOS

$\Gamma$	Gama
$\alpha$	Alpha
$\beta$	Beta
$\Delta$	Delta
$\theta$	Teta
$\lambda$	Lambda
$\psi$	Psi

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>12</b>
<b>2</b>	<b>REVISÃO BIBLIOGRAFICA</b> .....	<b>13</b>
<b>2.1</b>	<b>Conceitos básicos em análise de sobrevivência</b> .....	<b>13</b>
<b>2.1.1</b>	<b>Representação dos dados em análise de sobrevivência</b> .....	<b>13</b>
<b>2.1.2</b>	<b>Distribuição do tempo de sobrevivência</b> .....	<b>14</b>
<b>2.1.3</b>	<b>O estimador klapan-Meier</b> .....	<b>15</b>
<b>2.1.4</b>	<b>A função de verossimilhança em análise de sobrevivência</b> .....	<b>15</b>
<b>2.2</b>	<b>Distribuição Generalizada</b> .....	<b>16</b>
<b>2.2.1</b>	<b>Classe de distribuição exponenciada</b> .....	<b>16</b>
<b>2.2.2</b>	<b>Distribuição gama generalizada</b> .....	<b>17</b>
<b>2.2.3</b>	<b>A distribuição gama generalizada exponenciada</b> .....	<b>20</b>
2.2.3.1	As formas da função de taxa de falha.....	22
2.2.3.2	Casos particulares da GGE .....	23
<b>2.3</b>	<b>Critérios de comparação entre modelos e sub-modelos</b> .....	<b>26</b>
<b>3</b>	<b>APLICAÇÕES</b> .....	<b>27</b>
<b>3.1</b>	<b>Análise dos dados de <i>Nursing Activities score</i> (não censura)</b> .....	<b>27</b>
<b>3.2</b>	<b>Dados de AIDS (com censura)</b> .....	<b>39</b>
<b>4</b>	<b>CONSIDERAÇÕES FINAIS</b> .....	<b>32</b>
	<b>REFERÊNCIAS</b> .....	<b>33</b>

## 1 INTRODUÇÃO

As distribuições gama generalizada, exponencial, Weibull, log-logística e log-normal são tidas como as distribuições de probabilidade mais clássicas no ramo de análise de sobrevivência da área de estatística. As distribuições de probabilidade são de extrema importância em análise de sobrevivência, pois é a partir delas que modelamos os dados de sobrevivência coletados e podemos fazer inferências sobre os casos em estudo. Essas distribuições são denominadas modelos probabilísticos ou paramétricos (CORDEIRO e CASTRO, 2003).

Os dados em análise de sobrevivência referem-se ao tempo até a ocorrência de um evento de interesse. Este tempo é denominado tempo de falha e apresenta algumas características especiais. A primeira é que a variável resposta, geralmente, apresenta distribuição assimétrica positiva, não sendo, portanto, adequado assumir que tenha distribuição normal. A segunda característica é uma espécie de dado incompleto onde a falha não ocorreu até o fim do estudo ou simplesmente porque o acompanhamento do paciente foi perdido ou interrompido, o que acaba se tornando uma das principais características dos dados em análise de sobrevivência pois é muito improvável que em estudos nessa área não ocorra algum tipo de censura, que é a observação parcial da resposta (COLOSIMO e GIOLO, 2006).

Para inferir como determinada doença ou fenômeno age sobre um paciente ao longo do tempo é essencial obter informações de uma determinada amostra (conjunto de dados), no contexto de análise de sobrevivência uma das formas de obter algumas dessas informações é estimando os parâmetros de uma distribuição para o conjunto de dados coletados e é nesse sentido que este trabalho se apresenta fazendo uma breve apresentação e aplicação da distribuição gama generalizada exponenciada proposta por Cordeiro *et al.* (2009) em contrapartida com as distribuições gama generalizada (STACY, 1972) e Weibull.

Nesse contexto, o objetivo deste trabalho foi estimar os parâmetros da distribuição gama generalizada exponenciada, gama generalizada e Weibull pelo método de máxima verossimilhança a partir de uma amostra de dados de sobrevivência, relativa ao valor médio do *Nursing Activities Score* (NAS) e AIDS. Comparando o ajuste das distribuições segundo o teste da razão de verossimilhança e as estatísticas AIC, BIC e CAIC.

## 2 REVISÃO BIBLIOGRAFICA

Nesta seção é feita uma revisão sucinta sobre análise de sobrevivência assim como os métodos inferenciais para análise de dados de sobrevivência, formando o embasamento teórico necessário para o entendimento do trabalho relacionado ao objetivo da pesquisa.

### 2.1 Conceitos básicos em análise de sobrevivência

A análise de sobrevivência é caracterizada pelo fato que a variável resposta é composta de dois fatores, o tempo até a ocorrência de um evento de interesse e frequentemente pelas censuras (Colosimo e Giolo, 2006). O evento em estudo é denominado falha e o tempo até a ocorrência da falha é chamado de tempo de falha.

A censura é o registro parcial do tempo de falha, devido à perda ou retirada de um elemento do estudo. A censura é dita ser do tipo I quando ocorre devido ao término do estudo após um período de tempo pré-determinado e, do tipo II quando ocorre devido ao término do estudo após um número de falhas fixado previamente e censura aleatória, sendo a mais comum em situações práticas, quando um elemento deixa o estudo sem que o evento de interesse tenha ocorrido.

As censuras previamente citadas são conhecidas como censura à direita, pois a falha ocorre sempre à direita do tempo registrado. Existem ainda outros mecanismos de censura, como por exemplo, as censuras à esquerda, em que o tempo registrado é maior que o tempo de falha; e censura intervalar, na qual não se sabe o tempo exato de falha, e a única informação disponível é que o tempo de falha ocorreu em certo intervalo de tempo. Neste trabalho, entretanto, será adotado apenas o mecanismo de censura aleatória à direita que será denominada simplesmente por censura.

#### 2.1.1 Representação dos dados em análise de sobrevivência

Considere  $T$  uma variável aleatória que representa o tempo de falha de um paciente e  $C$ , outra variável aleatória independente de  $T$ , representando o tempo de censura associado a este paciente. Logo o que se observa para este paciente é, portanto,

$$t = \min(T, C)$$

e

$$\delta = \begin{cases} 1 & \text{se } T \leq C \\ 0 & \text{se } T > C \end{cases} .$$

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 que indica a falha ou censura, ou seja,

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

Desta maneira, a variável aleatória resposta em análise de sobrevivência é representada por duas colunas no banco de dados.

A distribuição da variável aleatória  $T$  pode ser especificada por meio da função de densidade de probabilidade, função de sobrevivência ou função de taxa de falha.

### 2.1.2 Distribuição do tempo de sobrevivência

Seja  $T$  uma variável aleatória, não negativa, que representa o tempo de vida de um indivíduo proveniente de uma dada população homogênea. A distribuição de  $T$  pode ser caracterizada por meio da função de densidade, função de sobrevivência ou função de risco.

A função densidade de probabilidade  $f(t)$  é definida por:

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t)}{\Delta t} = \frac{\partial F(t)}{\partial t} ,$$

em que  $F(t) = P[T \leq t]$  é a função distribuição acumulada (fda) de  $T$ . No contexto da análise de sobrevivência, a função  $f(t)$  pode ser interpretada como a probabilidade de um indivíduo sofrer o acontecimento de interesse no intervalo de tempo  $\Delta t$  e possui duas propriedades:

$$f(t) \geq 0 \text{ e } \int_0^{\infty} f(t) dt = 1$$

A função de sobrevivência,  $S(t)$ , é definida como sendo a probabilidade de ocorrência do acontecimento de interesse após o instante  $t$ . A função de sobrevivência define-se então da seguinte forma:

$$S(t) = P(T \geq 0) = \int_t^{\infty} f(x) dx ,$$

em que:  $\lim_{t \rightarrow 0} S(t) = 1$  e  $\lim_{t \rightarrow \infty} S(t) = 0$ .



Em termos clínicos, a função de sobrevivência representa a probabilidade de um indivíduo sobreviver além do instante  $t$ .

A função risco,  $h(t)$ , representa o risco instantâneo no instante  $t$  condicional à sobrevivência até o tempo  $t$ . Define-se da seguinte forma:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}$$

### 2.1.3 O estimador Kaplan-Meier

As Em 1958, Kaplan e Meier propuseram um estimador não-paramétrico para a função de sobrevivência, quando se tem a presença de uma amostra censurada. Este estimador é designado por estimador de Kaplan-Meier ou estimador produto-limite.

O estimador considera na sua construção um número de intervalos de tempo igual ao número de falhas distintas, sendo os limites dos intervalos os tempos de falha da amostra.

Dessa forma, o estimador de Kaplan-Meier, pode ser definido como:

$$S(t) = \prod_{j:t_{(j)} \leq t} \left( \frac{n_j - d_j}{n_j} \right) = \prod_{j:t_{(j)} \leq t} \left( 1 - \frac{d_j}{n_j} \right),$$

em que  $t_{(1)}, t_{(2)}, \dots, t_{(k)}$  representam os  $k$  tempos de falha distintos e ordenados,  $d_j$  é o número de falhas em  $t_{(j)}$ ,  $j = 1, \dots, k$ , e  $n_j$  é o número de indivíduos sob risco em  $t_{(j)}$ , ou seja, os indivíduos que não falharam e não foram censurados até o instante imediatamente anterior a  $t_{(j)}$ .

### 2.1.4 A função de verossimilhança em análise de sobrevivência

Através do método de máxima verossimilhança é possível incorporar as censuras, presentes em muitos dados de tempo de vida.

Considere uma amostra de variáveis aleatórias independentes  $T_1, T_2, \dots, T_n$  de tempos de sobrevivência.

Supondo que os dados consistem de  $n$  pares observados  $(t_1, \delta_1), (t_2, \delta_2), \dots, (t_n, \delta_n)$  em que  $t_i$  é o tempo de falha ou censura,  $\delta_i$  é o indicador de censura cuja distribuição envolve um vetor de parâmetros  $\theta$ . Nesse caso, a função de verossimilhança considerando censura não informativa é dada por:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n [f(t_i, \boldsymbol{\theta})]^{\delta_i} [S(t_i, \boldsymbol{\theta})]^{1-\delta_i},$$

em que  $f(t_i)$  e  $S(t_i)$  são as funções densidade probabilidade e de sobrevivência para cada variável aleatória  $T_i$ , respectivamente. Observe que, a contribuição de cada observação não censurada é a sua função de densidade e que cada observação censurada contribui com a função de sobrevivência.

## 2.2 Distribuição Generalizada

Durante os últimos anos diversos pesquisadores têm se esforçado para conseguirem generalizações de distribuições já conhecidas, sempre no intuito de tornarem os modelos mais flexíveis de forma a apresentarem um ajuste melhor a uma ampla gama de dados e também outras distribuições como casos particulares. Nesse contexto existem as classes de distribuições generalizadas que são métodos para obter generalizações de distribuições, dentre essas classes temos como exemplos: a classe das distribuições Marshall e Olkin, a classe de distribuições estendidas, a classe das distribuições exponenciadas, a classe de distribuições beta, a classe Kumaraswamy entre outras.

A distribuição gama generalizada exponenciada é uma família pertencente à classe exponenciada. Deste modo, daremos maior atenção a essa classe de distribuições modificadas.

### 2.2.1 Classe de distribuição exponenciada

A classe de distribuições exponenciadas é obtida elevando a função de distribuição acumulada de uma distribuição já conhecida ao expoente  $\lambda$  ( $\lambda > 0$ ), ou seja:

$$F(t) = [G(t)]^\lambda.$$

Em que  $g(u)$  é a distribuição conhecida previamente e  $G(t) = \int_0^t g(u)du$  é a acumulada dessa distribuição.

Deste modo a nova função de densidade é dada por  $f(t) = \lambda g(t)[G(t)]^{\lambda-1}$ . A função de sobrevivência e risco são dadas respectivamente por:

$$S(t) = 1 - F(t) \text{ e } h(t) = \frac{\lambda g(t)[G(t)]^{\lambda-1}}{1 - [G(t)]^\lambda}.$$

Nesta forma,  $\lambda$  é visto como um novo parâmetro e da mais flexibilidade ao modelo, além de que geralmente a nova distribuição possuirá pelo menos a antiga como caso particular,  $\lambda = 1$ .

### 2.2.2 Distribuição gama generalizada

A distribuição gama generalizada (GG) foi introduzida por Stacy (1962) e despertou o interesse de diversos pesquisadores pelo fato de representar uma família paramétrica que possui como casos particulares outras distribuições amplamente utilizadas na análise de dados de tempo de vida. Esta propriedade da distribuição GG, faz com que a mesma seja de grande utilidade, como por exemplo, na discriminação de modelos probabilísticos (ORTEGA; CORDEIRO; PASCOA, 2011).

Nos últimos anos diversos trabalhos envolvendo a distribuição GG foram propostos, dentre os quais destacam-se, por exemplo, Nadarajah e Gupta (2007) que usaram a distribuição com aplicações em dados de seca. Ali *et al.* (2008) derivou a distribuição exata do produto  $X_1 X_2$  em que  $X_1$  e  $X_2$  são variáveis aleatórias independentes GG, provenientes dos resultados de dados de seca. Ortega *et al.* (2008) comparou três tipos de resíduos baseado no componente de deviance no modelo de regressão GG usando observações censuradas. Nadarajah (2008) apresentou um estudo sobre o uso da distribuição GG em engenharia elétrica e eletrônica. Cox (2008) discutiu e comparou a família F-generalizada com o modelo GG. Malhotra *et al.* (2009) modelou o sinal *wireless* por meio da distribuição GG e analisou o desempenho do sinal. Cordeiro *et al.* (2009) propôs a distribuição gama generalizada exponenciada.

Seja  $T$  uma variável aleatória com distribuição GG proposta por Stacy (1962), cuja função de densidade, denotada por  $f(t)$ , é dada por:

$$f(t) = \frac{\tau}{\alpha \Gamma(k)} \left(\frac{t}{\alpha}\right)^{\tau k - 1} \exp\left[-\left(\frac{t}{\alpha}\right)^\tau\right], t > 0, \quad (1)$$

em que  $\alpha > 0$  é o parâmetro de escala,  $\tau > 0$  e  $k > 0$  são os parâmetros de forma e  $\Gamma(k)$  é a função gama, definida por:

$$\Gamma(k) = \int_0^{\infty} t^{k-1} \exp(-t) dt.$$

A Figura 1 mostra o gráfico da função densidade probabilidade da distribuição GG para diferentes valores de parâmetros. Se  $T$  é uma variável aleatória positiva com distribuição GG com parâmetros  $\alpha$ ,  $\tau$  e  $k$ , então, denota-se que  $T \sim GG(\alpha, \tau, k)$ .

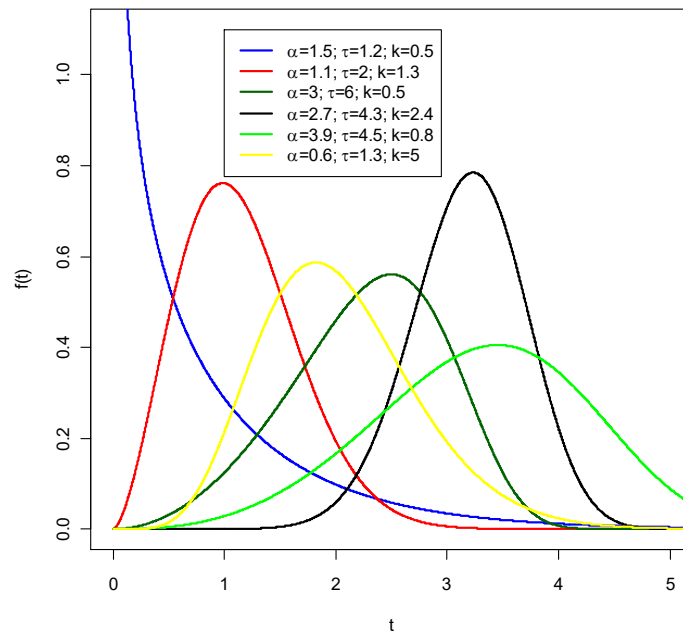


Figura 1: Gráfico da função densidade da distribuição GG para diferentes valores de parâmetros.

Vários trabalhos já estudaram as propriedades da distribuição GG, dentre os quais destacam-se Stacy e Mihram (1965), Prentice (1974), Farewell e Prentice (1977) e Lawless (1980, 2003). A média e a variância da distribuição GG são dadas por:

$$E(T) = \frac{\alpha \Gamma\left(\frac{\tau k + 1}{\tau}\right)}{\Gamma(k)} \quad e \quad V(T) = \frac{\alpha^2}{\Gamma(k)} \left\{ \Gamma\left(\frac{\tau k + 2}{\tau}\right) - \frac{\left[\Gamma\left(\frac{\tau k + 1}{\tau}\right)\right]^2}{\Gamma(k)} \right\}.$$

A distribuição GG é uma família paramétrica que possui alguns casos particulares. Quando  $\tau = k = 1$  na equação (1) temos a distribuição exponencial, quando  $\tau = 1$  temos a distribuição gama. Para  $k = 1$  temos a distribuição Weibull, para  $\alpha = 1/\lambda$ ,  $\tau = 2$  e  $k = 1$  temos a distribuição Rayleigh, outros submodelos podem ser vistos, por exemplo, em Lawless (2003). As funções de distribuição acumulada

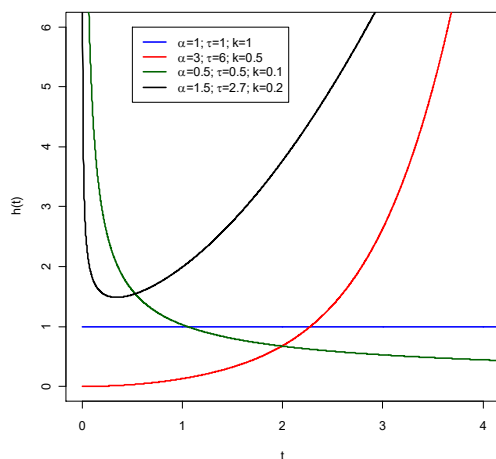
$F(t)$ , de sobrevivência  $S(t)$  e de risco  $h(t)$  são expressas, respectivamente, por;

$$F(t) = P[T \leq t] = \frac{\gamma(k, (t/\alpha)^\tau)}{\Gamma(k)} = \frac{1}{\Gamma(k)} \int_0^{\left(\frac{t}{\alpha}\right)^\tau} w^{k-1} \exp(-w) dw = \left\{ \gamma_1 \left[ k, \left( \frac{t}{\alpha} \right)^\tau \right] \right\},$$

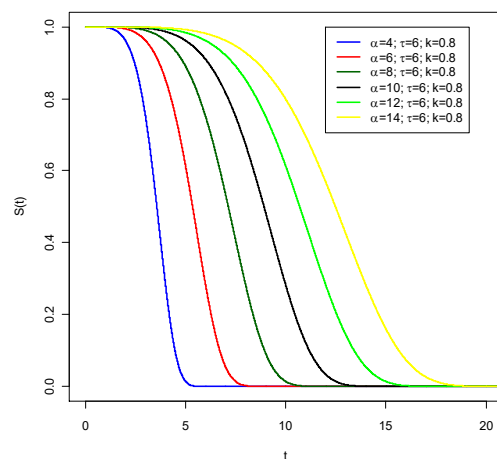
$$S(t) = 1 - F(t) = 1 - \left\{ \gamma_1 \left[ k, \left( \frac{t}{\alpha} \right)^\tau \right] \right\} e$$

$$h(t) = \frac{f(t)}{S(t)} = \frac{t^{\tau k-1} \exp \left[ - \left( \frac{t}{\alpha} \right)^\tau \right]}{\int_0^\infty x^{\tau k-1} \exp \left[ - \left( \frac{x}{\alpha} \right)^\tau \right] dx},$$

em que  $\gamma(k; x) = \int_0^\infty w^{k-1} e^{-w} dw$  é a função gama incompleta e  $\gamma_1(k; x)$  é a razão da função gama incompleta, definida por  $\gamma_1(k; x) = \gamma(k; x)/\Gamma(k)$ , que é facilmente implementada em vários pacotes estatísticos (R, SAS, Ox). A Figura 2, mostra em (a) o gráfico da função risco da distribuição GG para diferentes valores de parâmetros. Em (b) o gráfico da função de sobrevivência da distribuição GG variando o parâmetro  $\alpha$ .



(a)



(b)

Figura 2: (a) gráfico da função risco da distribuição GG para diferentes valores de parâmetros. (b) gráfico da função de sobrevivência da distribuição GG variando o parâmetro  $\alpha$ .

Algumas outras propriedades da distribuição GG podem ser encontradas em Lawless (2003).

### 2.2.3 A distribuição gama generalizada exponenciada

Como já foi dito, a distribuição gama generalizada exponenciada é uma distribuição pertencente à classe de distribuições exponenciadas, desta forma a GGE (CORDEIRO *et al.*, 2009) pode ser obtida elevando-se a acumulada da GG a potencia  $\lambda$ , ou seja,  $F(t) = [G(t)]^\lambda$ . Sendo assim a expressão da função densidade de probabilidade da GGE com quatro parâmetros  $\alpha$ ,  $\beta$ ,  $k$  e  $\lambda$ , pode ser escrita como

$$f(t) = \frac{\partial F(t)}{\partial t} = \frac{\lambda\beta}{\alpha\Gamma(k)} \left(\frac{t}{\alpha}\right)^{\beta k - 1} \exp\left\{-\left(\frac{t}{\alpha}\right)^\beta\right\} \left\{\gamma_1\left[k, \left(\frac{t}{\alpha}\right)^\beta\right]\right\}^{\lambda - 1}, \quad t > 0. \quad (1)$$

Para verificar que  $f(t)$  é uma função de densidade de probabilidade, basta provar que

$$f(t) \geq 0 \text{ e } \int_0^\infty f(t) dt = 1.$$

A desigualdade  $f(t) \geq 0$  decorre diretamente do fato de  $G(t) = \int_0^t g(u) du \geq 0$ , pois  $g(t) \geq 0$ . Para provar que  $\int_0^\infty f(t) dt = 1$  basta fazer a substituição

$$u = \gamma_1\left[k, \left(\frac{t}{\alpha}\right)^\beta\right], \text{ desta forma teremos } du = \frac{\beta}{\alpha\Gamma(k)} \left(\frac{t}{\alpha}\right)^{\beta k - 1} \exp\left\{-\left(\frac{t}{\alpha}\right)^\beta\right\} \text{ sendo assim a}$$

integral fica da seguinte forma:

$$\int_0^\infty f(t) dt = \int_0^\infty \lambda u^{\lambda - 1} du = \left[u^\lambda\right]_0^\infty = 1,$$

Note que a última igualdade é válida, pois  $u = G(t)$  com  $G(0) = 0$  e  $\lim_{t \rightarrow \infty} G(t) = 1$ .

A Figura 3 mostra o gráfico da função densidade variando apenas o novo parâmetro e fixando os valores de  $\alpha$ ,  $\beta$  e  $k$ , a Figura 4 mostra o gráfico da função de densidade da GGE para alguns valores de  $\alpha$ ,  $\beta$ ,  $k$  e  $\lambda$ .

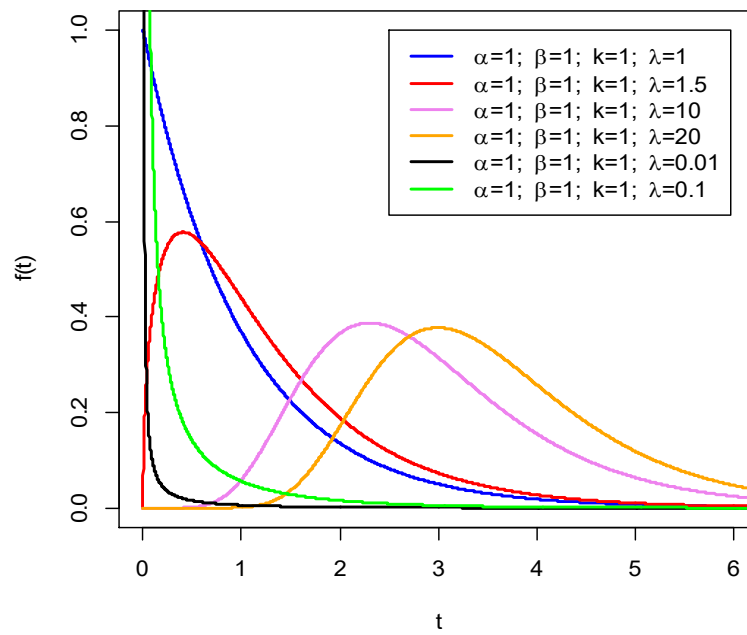


Figura 3 - Gráfico da densidade da GGE com os parâmetros  $\alpha = \beta = k = 1$  e  $\lambda$  variando.

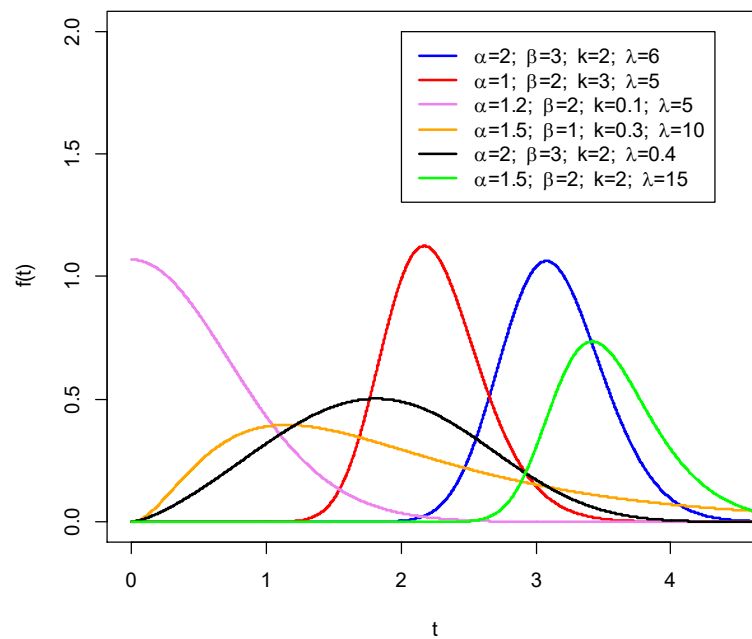


Figura 4 - Gráfico da função densidade da GGE para alguns valores de  $\alpha, \beta, k$  e  $\lambda$ .

A função de distribuição acumulada, a função sobrevivência e a função taxa de falha são, respectivamente, expressas por:

$$F(t) = \int_0^t f(t) = \left\{ \frac{1}{\Gamma(k)} \int_0^{\left(\frac{t}{\alpha}\right)^\beta} w^{k-1} e^{-w} \right\}^\lambda = \left\{ \gamma_1 \left[ k, \left(\frac{t}{\alpha}\right)^\beta \right] \right\}^\lambda,$$

$$S(t) = 1 - F(t) = 1 - \left\{ \gamma_1 \left[ k, \left( \frac{t}{\alpha} \right)^\beta \right] \right\}^\lambda \quad (2)$$

$$h(t) = \frac{f(t)}{S(t)} = \frac{\lambda \beta}{\alpha \Gamma(k)} \left( \frac{t}{\alpha} \right)^{\beta k - 1} \exp \left\{ - \left( \frac{t}{\alpha} \right)^\beta \right\} \left\{ \gamma_1 \left[ k, \left( \frac{t}{\alpha} \right)^\beta \right] \right\}^{\lambda - 1} \left\{ 1 - \left\{ \gamma_1 \left[ k, \left( \frac{t}{\alpha} \right)^\beta \right] \right\}^\lambda \right\}^{-1}$$

A Figura 5 mostra o gráfico da função de sobrevivência para alguns valores de  $\alpha$ ,  $\beta$ ,  $k$  e  $\lambda$ .

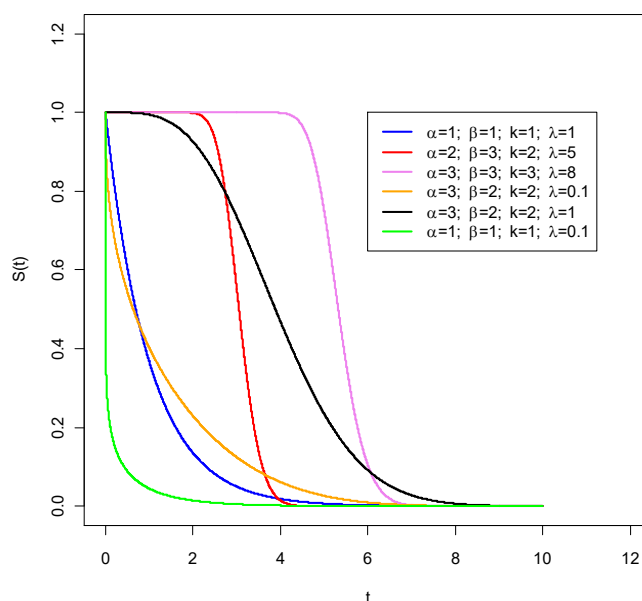


Figura 5 - Gráfico da função de sobrevivência da GGE para alguns valores de  $\alpha$ ,  $\beta$ ,  $k$  e  $\lambda$ .

### 2.2.3.1. As formas da função de taxa de falha

Algo que está diretamente interligado a flexibilidade de um modelo em análise de sobrevivência são as formas que a função de taxa de falha apresenta, nesse sentido uma importante característica da GGE é que sua função de taxa de falha acomoda as formas constante, crescente, decrescente, forma de U e unimodal. Basicamente estas formas dependem dos parâmetros  $\beta$ ,  $k$  e  $\lambda$ , um estudo mais detalhado sobre as formas da função de taxa de falha pode ser encontrado em Cordeiro *et al.* (2009).

As figuras 6 e 7 mostram as diferentes formas da função de taxa de falha da GGE para diferentes valores dos parâmetros  $\alpha$ ,  $\beta$ ,  $k$  e  $\lambda$ .



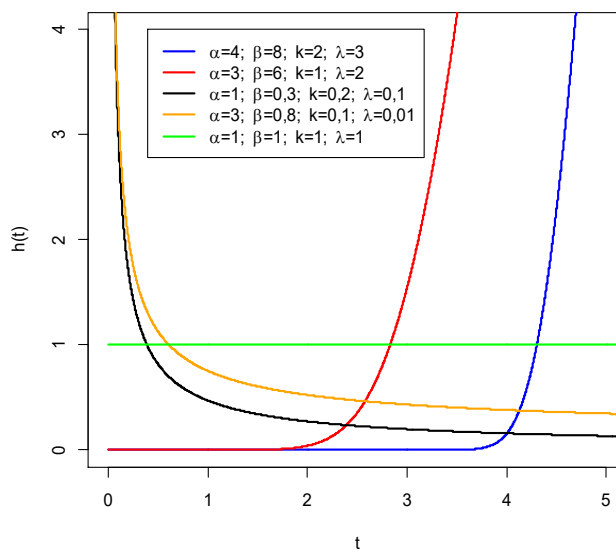


Figura 6 - Gráfico da função taxa de falha assumindo as formas crescente, decrescente e constante.

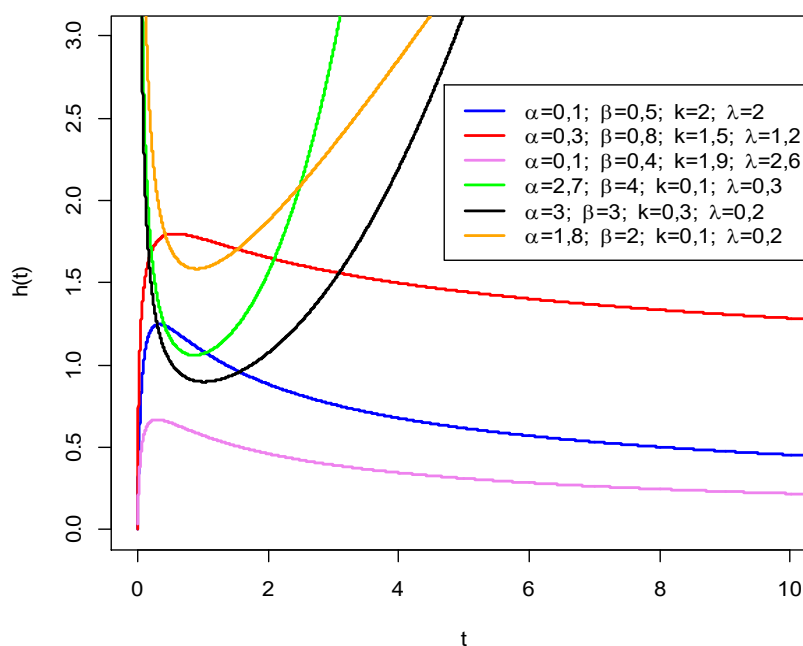


Figura 7 - Gráfico da função de taxa de falha assumindo as formas U e unimodal.

#### 2.2.3.2. Casos particulares da GGE

Outra característica interessante da GGE é que ela possui várias distribuições como casos particulares, entre elas citarei as mais clássicas em análise de sobrevivência:

- Para  $\beta = 1, k = 1$  e  $\lambda = 1$  na equação (1) a distribuição GGE se reduz a uma distribuição exponencial, com função de densidade da forma:

$$f(t) = \frac{1}{\alpha} \exp\left\{-\left(\frac{t}{\alpha}\right)\right\}, t > 0.$$

- Para  $\lambda = 1$  e  $k = 1$  na equação (1) a distribuição GGE se reduz a uma distribuição Weibull, com função de densidade da forma:

$$f(t) = \frac{\beta}{\alpha} \left(\frac{t}{\alpha}\right)^{\beta-1} \exp\left\{-\left(\frac{t}{\alpha}\right)^\beta\right\}, t > 0.$$

- Para  $\beta = 1$  e  $\lambda = 1$  na equação (1) a distribuição GGE se reduz a uma distribuição Gama, com função de densidade da forma:

$$f(t) = \frac{t^{k-1}}{\Gamma(k)\alpha^k} \exp\left\{-\left(\frac{t}{\alpha}\right)\right\}, t > 0.$$

- Para  $\lambda = 1$  na equação (1) a distribuição GGE se reduz a uma distribuição Gama Generalizada, com função de densidade da forma:

$$g(t) = \frac{\beta}{\Gamma(k)\alpha^{\beta k}} t^{\beta k-1} \exp\left\{-\left(\frac{t}{\alpha}\right)^\beta\right\}, t > 0.$$

Considerando  $t_1, t_2, \dots, t_n$  uma amostra aleatória de uma população com distribuição GGE e com vetor de parâmetros  $\theta = (\alpha, \beta, k, \lambda)^T$ , a estimação dos parâmetros pode ser feita pelo método da máxima verossimilhança. Substituindo a equação (1) e (2) na equação (3) o logaritmo da função verossimilhança da GGE é dada por:

$$l(\theta) = r \left\{ \log(\lambda) + \log(\beta) - \log(\alpha) - \log[\Gamma(k)] \right\} + (\beta k - 1) \sum_{i=1}^r \log\left(\frac{t_i}{\alpha}\right) - \sum_{i=1}^r \left(\frac{t_i}{\alpha}\right) +$$

$$+(\lambda - 1) \sum_{i=1}^r \log \left\{ \gamma_1 \left[ k, \left(\frac{t_i}{\alpha}\right)^\beta \right] \right\} + \sum_{i=r+1}^n \log \left\{ 1 - \left\{ \gamma_1 \left[ k, \left(\frac{t_i}{\alpha}\right)^\beta \right] \right\}^\lambda \right\}$$

em que  $r$  é o número de falhas e  $n$  o tamanho da amostra.

As derivadas parciais de primeira ordem do logaritmo da função verossimilhança em relação aos parâmetros do vetor  $\theta = (\alpha, \beta, k, \lambda)^T$  são expressas por:

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \alpha} = -\frac{r\beta k}{\alpha} + \frac{\beta}{\alpha} \sum_{i=1}^n u_i - \frac{\beta(\lambda-1)}{\alpha\Gamma(k)} \sum_{i=1}^r \frac{u_i^k \exp(-u_i)}{\gamma_1(k, u_i)} + \frac{\lambda\beta}{\alpha\Gamma(k)} \sum_{r+1}^n \frac{[\gamma_1(k, u_i)]^{\lambda-1} u_i^k \exp(-u_i)}{1 - [\gamma_1(k, u_i)]^\lambda},$$

$$\begin{aligned} \frac{\partial l(\boldsymbol{\theta})}{\partial \beta} &= \frac{r}{\beta} + \frac{r}{\beta} \sum_{i=1}^r \log(u_i) - \frac{1}{\beta} \sum_{i=1}^r u_i \log(u_i) + \frac{(\lambda-1)}{\beta\Gamma(k)} \sum_{i=1}^r \frac{u_i^k \exp(-u_i) \log(u_i)}{\gamma_1(k, u_i)} + \\ &- \frac{\lambda}{\beta\Gamma(k)} \sum_{r+1}^n \frac{[\gamma_1(k, u_i)]^{\lambda-1} u_i^k \exp(-u_i) \log u_i}{1 - [\gamma_1(k, u_i)]^\lambda}, \end{aligned}$$

$$\begin{aligned} \frac{\partial l(\boldsymbol{\theta})}{\partial k} &= -r\lambda\psi(k) + \sum_{i=1}^r \log(u_i) + \frac{(\lambda-1)}{\Gamma(k)} \sum_{i=1}^n \frac{[\dot{\gamma}(k, u_i)]_k}{\gamma_1(k, u_i)} + \lambda\psi(k) \sum_{i=r+1}^n \frac{[\gamma_1(k, u_i)]^\lambda}{1 - [\gamma_1(k, u_i)]^\lambda} + \\ &+ \frac{1}{\Gamma(k)} \sum_{i=r+1}^n \frac{[\dot{\gamma}(k, u_i)]_k [\gamma_1(k, u_i)]^{\lambda-1}}{1 - [\gamma_1(k, u_i)]^\lambda}, \end{aligned}$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \lambda} = \frac{r}{\lambda} + \sum_{i=1}^r \log[\gamma_1(k, u_i)] - \sum_{i=r+1}^n \frac{[\gamma_1(k, u_i)]^\lambda \log[\gamma_1(k, u_i)]}{1 - [\gamma_1(k, u_i)]^\lambda},$$

em que  $u_i = \left(\frac{t_i}{\alpha}\right)^\beta$ ,  $\psi(k)$  é a função digama definida como  $\psi(k) = \frac{\partial(\log[\Gamma(k)])}{\partial k} = \frac{\Gamma'(k)}{\Gamma(k)}$

e  $[\dot{\gamma}(k, u_i)]_k = \int_0^{u_i} w^{k-1} e^{-w} \log(w) dw$ . Sendo assim os estimadores de máxima

verossimilhança (EMV) dos parâmetros de  $\boldsymbol{\theta}$  são obtidos resolvendo o sistema

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \alpha} = \frac{\partial l(\boldsymbol{\theta})}{\partial \beta} = \frac{\partial l(\boldsymbol{\theta})}{\partial k} = \frac{\partial l(\boldsymbol{\theta})}{\partial \lambda} = 0.$$

Em muitas aplicações, há uma informação qualitativa sobre o formato da função de taxa de falha, que pode ajudar na seleção de um modelo específico. Neste contexto, um dispositivo chamado de tempo total em teste (TTT) plot (Aarset, 1987) é útil. O gráfico TTT plot é obtido plotando  $G(r/n) = [(\Pr i=1 T_{i:n}) + (n-r)T_{r:n}]/(P_n i=1 T_{i:n})$ , em que  $r = 1, \dots, n$  e  $T_{i:n}$ ,  $i = 1, \dots, n$ , são as estatísticas de ordem da amostra, contra  $r/n$  (Mudholkar et al., 1996).

### 2.3 Critérios de comparação entre modelos e sub-modelos

Quando queremos comparar dois modelos probabilísticos, como por exemplo, o modelo GGE e seu sub-modelo GG, é interessante revisar os principais critérios de seleção mais utilizados.

Dentre os principais critérios de seleção de modelos utilizados em programas computacionais estão o critério de informação de Akaike – AIC (AKAIKE, 1974) e o critério de informação de Schwarz – BIC (SCHWARZ, 1978), que são em função do valor do logaritmo da função verossimilhança do modelo, dependem também do tamanho da amostra e do número de parâmetros. Liang e Zou (2007) propuseram um AIC melhorado (CAIC).

Considerando uma amostra  $T = (T_1, \dots, T_n)$  de tamanho  $n$  e o vetor de parâmetros  $\theta$ , sendo assim podemos calcular os critérios AIC, BIC e CAIC com as expressões:

$$AIC = -2l(\theta) + 2p,$$

$$BIC = -2l(\theta) + p \log(n),$$

$$CAIC = AIC + \frac{2(p+2)(p+3)}{n-p-3},$$

onde  $l(\theta)$  é o logaritmo da função de verossimilhança,  $p$  é a quantidade de parâmetros e  $n$  o tamanho da amostra.

Desta forma um ajuste mais apropriado dos dados é representado pelo modelo que apresentar os menores valores para os critérios AIC, BIC e CAIC.

### 3 APLICAÇÕES

Nesta seção do trabalho, são apresentados dois conjuntos de dados para ilustrar o uso da distribuição de probabilidade gama generalizada (GGE). Todos os resultados foram obtidos nos *softwares* R (Core Team, 2021) e SAS versão 9.1 (SAS, 2004).

#### 3.1 Análise dos dados de *Nursing Activities Score* (não censura)

Este estudo, prospectivo, de abordagem quantitativa, do tipo exploratório descritivo, foi desenvolvido com o objetivo de identificar o tempo médio de assistência de enfermagem necessário para assistir ao paciente adulto, classificado como alta dependência de enfermagem (ADE). Participaram da pesquisa todos os pacientes classificados como ADE, internados na Unidade de Clínica Médica do Hospital Universitário da USP por, no mínimo, 24 horas, durante o período de coleta de dados.

Para a identificação dos pacientes de ADE foi utilizado o Instrumento de Classificação de Pacientes (ICP) de Fugulinet *al.* (2008). O tempo médio de cuidado necessário para assistir o paciente de ADE foi calculado por meio da aplicação de uma equação matemática disponível na literatura, a partir da identificação do valor médio do *Nursing Activities Score* (NAS), mais detalhes, Tsukamoto (2010).

Os dados foram coletados das evoluções e prescrições de enfermagem e das condições de cada paciente, registradas nos impressos que compõem seu prontuário. Foram identificados  $n=97$  pacientes classificados como ADE, no período de 06 de maio a 21 de julho e 15 a 31 de agosto de 2009. Neste caso a variável em estudo é denominada de NAS.

Na Tabela 1 são apresentadas as estimativas de máxima verossimilhança de cada parâmetro com os respectivos erros padrões (entre parênteses) e os valores das estatísticas AIC, BIC e CAIC, para as distribuições comparadas.

Tabela 1: Ajuste final dos modelos comparados, para os dados de NAS.

Modelo	$\alpha$	$\tau$	$K$	$\lambda$	AIC	CAIC	BIC
GGE	0,1907 (0,5573)	0,7113 (0,2335)	41,89 (40,61)	12,53 (13,50)	622,1	632,1	632,3
GG	0,0872 (0,3701)	0,7577 (0,3570)	124,95 (118,45)	1 (-)	627,9	628,1	635,6
Weibull	53,9774 (0,7748)	7,5264 (0,5172)	1 (-)	1 (-)	658,1	658,2	663,3

Observa-se, na Tabela 1, que a distribuição GGE, apresentou menor AIC, BIC e CAIC, quando comparada com a distribuição GG e Weibull para o ajuste do conjunto de dados de NAS.

A comparação da distribuição GG com a distribuição Weibull, utilizando a estatística  $RV$  pode ser observada na Tabela 2, os números na tabela, especialmente o valor  $p$ , sugere que o modelo GG produz um melhor ajuste a esses dados do que a distribuição Weibull. Assim como a comparação da GGE com a GG em que é possível ver pela Tabela 2 que a GGE produz um melhor ajuste que a GG.

Tabela 2: Teste da razão de verossimilhança, para os dados de NAS.

Modelo	Hipótese	Estatística $w$	Valor $p$
GGE vs GG	$H_0: \lambda = 1$ vs $H_1: \lambda \neq 1$	8,9	0,0113
GG vs Weibull	$H_0: k=1$ vs $H_1: k \neq 1$	64,4	<0,001

Na Figura 8 tem-se que a curva empírica de risco é crescente o que faz parte das possibilidades de risco das distribuições estudadas.

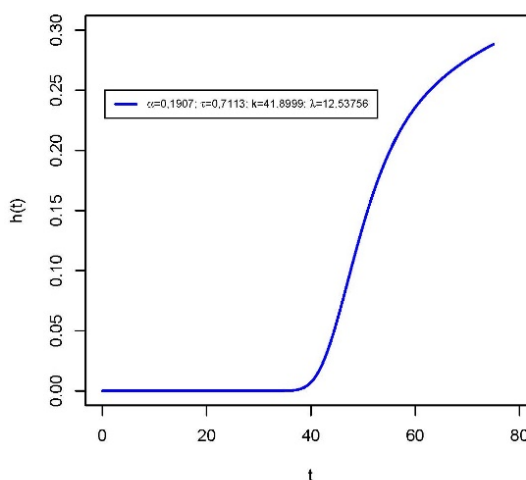


Figura 8: – Gráfico da função de taxa de falha da GGE na forma crescente ajustada aos dados de NAS.

A Figura 9, mostra em (a) o ajuste das densidades dos modelos GGE, GG e Weibull com o histograma empírico dos dados observados. Em (b) apresenta a comparação das estimativas segundo os modelos GGE, GG e Weibull com a estimativa não paramétrica de Kaplan-Meier. Observa-se pela figura que a distribuição GGE tem um ajuste satisfatório para os dados de NAS.

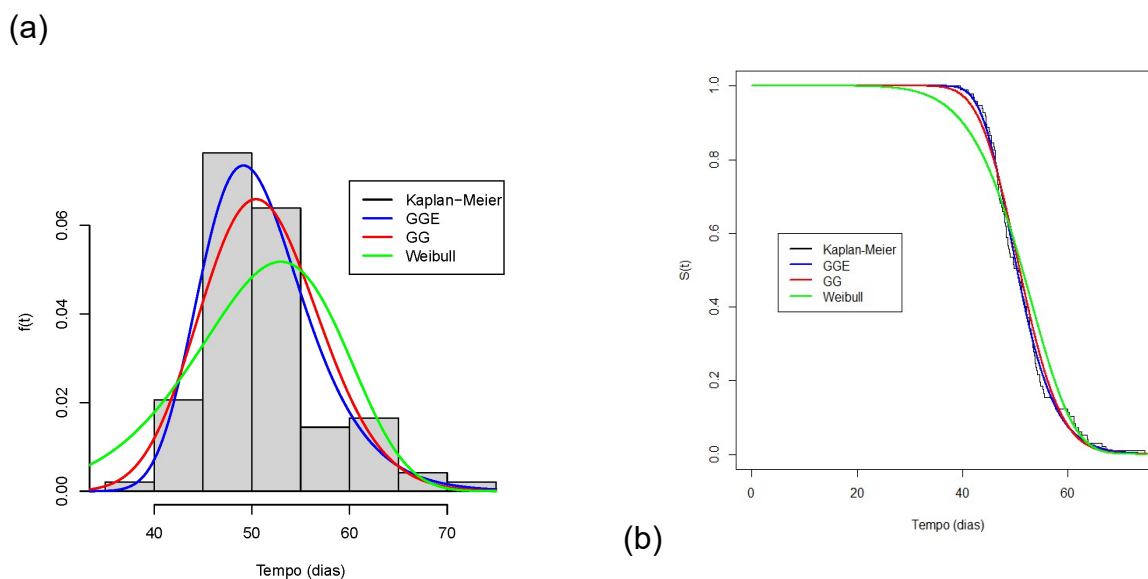


Figura 9: (a) Densidades estimadas dos modelos para os dados de NAS. (b) Estimativas da função de sobrevivência segundo Kaplan Meier, e segundo os modelos GGE, GG e Weibull, para os dados de NAS.

### 3.2 Dados de AIDS (com censuras)

Neste estudo foram coletados dados provenientes de pacientes portadores de HIV atendidos entre 1986 e 2000 no Instituto de Pesquisa Clínica Evandro Chagas (Ipec-Fiocruz/RJ). Obteve-se uma amostra de 194 indivíduos que foram diagnosticados como portadores de AIDS (critério CDC 1993) durante o período de acompanhamento. Foi definido como o tempo de falha o tempo de sobrevivência desde o diagnóstico da AIDS até óbito (em dias), sendo o evento de interesse o óbito do paciente. Foram observadas 91 falhas e 103 censuras.

Na Tabela 3 são apresentadas as estimativas de máxima verossimilhança (e os correspondentes erros-padrão que estão entre parênteses) de cada parâmetro e os valores dos critérios AIC, BIC e CAIC para os modelos comparados.

Tabela 3 – Ajustes finais dos modelos comparados.

Modelo	A	B	K	$\Lambda$	AIC	BIC	CAIC
GGE	49,9392 (95,4707)	0,2492 (0,1512)	0,7940 (1,9541)	9,9647 (42,6270)	1544,1	1544,1	1557,1
GG	2012,06 (3101,40)	0,7172 (0,4588)	1,1082 (0,9949)	1 (-)	1546,3	1546,4	1556,1
Weibull	572,72 (1,04)	60,81 (0,0881)	(-)	(-)	1545,6	1552,1	1545,7

Analisando a Tabela 3 verifica-se que a GGE ajustou os dados de AIDS de maneira satisfatória se comparada a GG devido aos menores valores de AIC e BIC. Porém é importante salientar que de acordo com a Tabela 4 não houve diferença no ajuste da GGE e da GG, havendo diferença entre a GG vs a Weibull.

Tabela 4: Teste da razão de verossimilhança, para os dados de AIDS.

Modelo	Hipótese	Estatística $w$	Valor $p$
GGE vs GG	$H_0: \lambda = 1$ vs $H_1: \lambda \neq 1$	4,4	0,1108
GG vs Weibull	$H_0: k=1$ vs $H_1: k \neq 1$	64,4	< 0,0001

A Figura 10 mostra o gráfico do ajuste da curva de risco da GGE para esse conjunto de dados o que evidencia comportamento de risco unimodal.

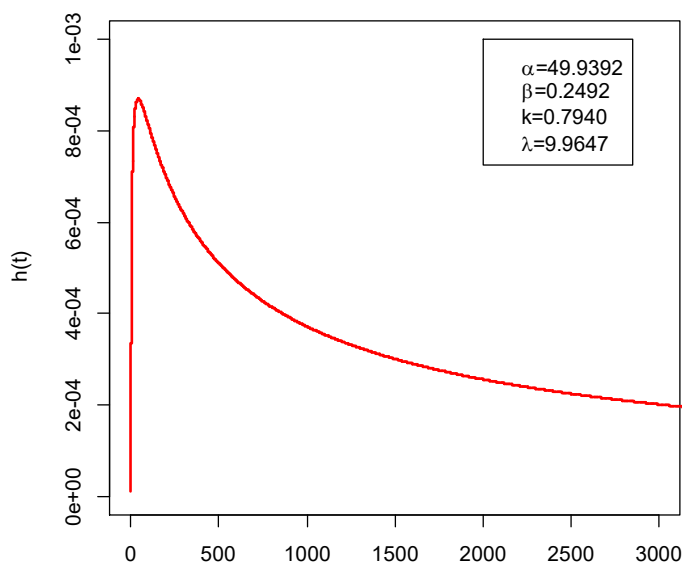


Figura 10 – Gráfico da função de taxa de falha da GGE na forma unimodal ajustada aos dados de saúde.

A Figura 11 ilustra o gráfico da função de sobrevivência da distribuição gama generalizada exponenciada comparada a distribuição gama generalizada e a curva de sobrevivência empírica estimada pelo método de Kaplan-Meier.

Pelas figuras observa-se uma maior flexibilidade da distribuição gama generalizada exponenciada.



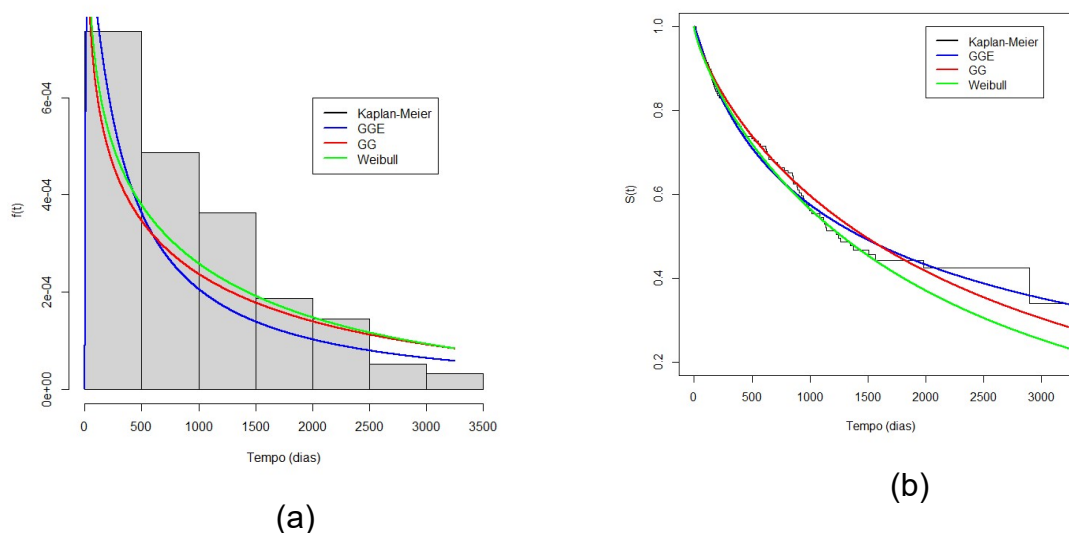


Figura 11 – (a) Densidades estimadas dos modelos para os dados de AIDS  
 (b) Ajustes das curvas de sobrevivência da GGE, GG e Weibull comparadas ao Kaplan-Meier para os dados de AIDS.

Segundo COUTO(2010) onde os resultados obtidos foram parecidos, pois de acordo com seus estudos concluiu também que, a principal vantagem do modelo GGE é a sua flexibilidade em acomodar diferentes formas da função de taxa de falha assim como possuir diversas distribuições muito utilizadas em análise de sobrevivência. Além disso, foi proposto por COUTO(2010) um modelo de regressão baseado na distribuição log-gama generalizada exponenciada para dados considerando censura à direita e com mecanismo aleatório mostrando a flexibilidade em relação aos seus novos parâmetros, aparti do método da máxima verossimilhança.

#### 4 CONSIDERAÇÕES FINAIS

Pelo teste da razão de verossimilhança e as estatísticas AIC, BIC e CAIC verificou-se que a distribuição gama generalizada exponenciada, proposta por Cordeiro *et al.* (2009) ajustou melhor os dois conjuntos de dados do que seus submodelos, a distribuição gama generalizada (STACY, 1962) e Weibull. Sendo assim é possível considerar que a principal vantagem da GGE é a flexibilidade da sua função de taxa de falha, de forma a deixar o modelo mais maleável e com a possibilidade de ajustar variados conjuntos de dados em análise de sobrevivência.

## REFERÊNCIAS

- AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, Boston, v.6, p.716-723, 1974.
- ALI, M. M.; WOO, J.; NADARAJAH, S. Generalized gamma variables with drought application. *Journal of the Korean Statistical Society*, Seul, v.37, p.37-45, 2008.
- COLOSIMO, E. A.; GIOLO, S. R. **Análise de sobrevivência aplicada**. São Paulo: Edgard Blucher, 2006. 370p.
- CORDEIRO, G. M.; SILVA, G. O.; ORTEGA, E. M. M. The exponentiated generalized gamma distribution with application to lifetime data, 2009. In press.
- CORDEIRO, G. M.; CASTRO, M. A new family of generalized distributions. *Journal of Statistical Computation and Simulation*, London, p.1-17, 2010.
- COUTO, Epaminondas de Vasconcellos. **Modelo de regressão log-gama generalizado exponenciado com dados censurados**. 2010. Tese de Doutorado. Universidade de São Paulo.
- FAREWELL, V. T.; PRENTICE, R. L. A study of distributional shape in life testing. *Technometrics*, Wisconsin, v.19, p.69-75, 1977.
- FUGULIN, R. M. T.; GAIDZINSKI, R. R.; KURCGANT, P. (2005) Sistema de classificação de pacientes: identificação do perfil assistencial dos pacientes das unidades de internação do HU-USP. *Revista Latino-Americana de Enfermagem*, 13, 72-78.
- KAPLAN, E.L.; MEIER, P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, Boston, v.53, p.457-481, 1958.
- LAWLESS, J. F. Inference in the generalized gamma and log-gamma distributions. *Technometrics*, Wisconsin, v.22, n.3, p.409-419, 1980.
- LAWLESS, J. F. **Statistical models and methods for lifetime data**. 2nd ed. New York: Wiley, 2003. 630p.
- LIANG, H.; ZOU, G. Improved AIC selection strategy for survival analysis. *Computational Statistics and Data Analysis*, New York, v.52, p.2538-2548, 2007.
- MALHOTRA, J.; SHARMA, A. K.; KALER, R. S. On the performance analysis of wireless receiver using generalized-gamma fading model. *Annals of Telecommunications*, Paris, v.64, p.147-153, 2009.
- MUDHOLKAR, G. S., SRIVASTAVA, D. K. AND KOLLIA, G. D. A generalization of the Weibull distribution with application to the analysis of survival data. *Journal of American Statistical Association*, 91, 1575-1583, 1996.

- NADARAJAH, S. On the use of generalized gamma distribution. *International Journal of Electronic*, New York, v.95, p.1029-1032, 2008.
- NADARAJAH, S.; GUPTA, A. K. A generalized gamma distribution with application to drought data. *Mathematics and Computers in Simulation*, Elsevier, v. 74, p.1-7, 2007.
- ORTEGA, E. M. M.; PAULA, G. A.; BOLFARINE, H. Deviance residuals in generalized log-gamma regression models with censored observations. *Journal of Statistical Computation and Simulation*, New York, v.78, p.747-764, 2008.
- PAL, M.; ALI, M. M.; WOO, J. Exponentiated Weibull distribution. *Statistica*, n.2, 2006.
- PESCIM, R. R.; DEMÉTRIO, C. G. B.; CORDEIRO, G. M.; ORTEGA, E. M. M.; URBANO, M. R. The Beta Generalized Half-Normal Distribuiton. *Computacional Statistics and Data Analysis*, New York, v.54, n.4, p.945-957, 2010.
- PRENTICE, R. L. A log-gamma model and its maximum likelihood estimation. *Biometrika*, London, v.61, n.3, p.539-544, 1974.
- R DEVELOPMENT CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL Disponivelem: <http://www.R-project.org>.
- SAS INSTITUTE. SAS/STAT User's Guide: Version 9. Cary: SAS Institute. 2004. 5121p.
- SCHWARZ, G. Estimating the dimension of a model. *Annals of Statistics*, Philadelphia, v.6, p.461-464, 1978
- SILVA, G. O.; ORTEGA, E. M. M.; CORDEIRO, G. M. The Beta Modified Weibull Distribution. *Lifetime Data Analysis*, v.16, p.409-430, 2010.
- STACY, E. W. A generalization of the gamma distribution. *Annals of Mathematical Statistics*, v.33, p.1187-1192, 1962.
- STACY, E. W.; MIHRAM, G. A. Parameter estimation for a generalized gamma distribution. *Technometrics*, New York, v.7, n.3, p.349-358, 1965.
- TSUKAMOTO, R. (2010). **Tempo Médio de Cuidado ao Paciente de Alta Dependência de Enfermagem Segundo o Nursing Activities Score (NAS)**. Dissertação, Escola de Enfermagem da Universidade de São Paulo, Brasil.