Revista Eletrônica Paulista de Matemática

ISSN 2316-9664
Volume 14, fev. 2019
Edição Ermac

André Jun Shiroma
UNESP-Universidade Estadual Paulista "Júlio de Mesquita Filho"
junshiromax@hotmail.com

Fernando Antônio Moala
UNESP-Universidade Estadual Paulista "Júlio de Mesquita Filho"
f.moala@unesp.br

Sérgio Minoru Oikawa
UNESP-Universidade Estadual Paulista "Júlio de Mesquita Filho"
sm.oikawa@unesp.br

Distribuição exponencial logarítmica com presença de censura do tipo 1 e tipo 2

Logarithmic exponential distribution in presence of type I and II censorship

Resumo

Neste artigo foi realizado uma análise estatística baseada na estimação de máxima verossimilhança dos parâmetros da distribuição Exponencial Logarítmica aplicada a dados de tempos de falha ou sobrevivência, considerando os dois tipos principais de dados censurados à direita como censura tipo 1 e tipo 2. Esta distribuição apresenta uma função de taxa de risco decrescente e pode ser utilizada para ajustar dados nas áreas da biologia e engenharia. Os estimadores de máxima verossimilhança obtidos sob dados completos e censurados tipo 1 e tipo 2, foram avaliados através de simulações utilizando medidas de vício, erro quadrático médio e probabilidade de cobertura. Três problemas reais considerando dados completos e censurados tipo 1 e tipo 2 foram analisados para fins ilustrativos. Concluiu-se que o método de máxima verossimilhança proporcionou boas estimativas para tamanhos amostrais relativamente grandes ($n > 30$) mesmo na presença de censura.

Palavras-chave: Análise de confiabilidade. Estimador de máxima verossimilhança. Matriz de informação de Fisher.

Abstract

In this paper, we performed a statistical analysis based on the maximum likelihood estimation of the Logarithmic Exponential distribution parameters applied to failure or survival time data, considering the two main types of censored data like type 1 and type 2 censorship. This distribution has a decreasing risk rate function and can be used to fit data in the areas of biology and engineering. The maximum likelihood estimators obtained under complete and censored data type 1 and type 2 were evaluated through simulations using measures of bias, mean square error and coverage probability. Three real problems considering complete and censored data type 1 and type 2 were analyzed for illustrative purposes. It was concluded that the maximum likelihood method provided good estimates with relatively large sample sizes (n > 30) even with the presence of censorship.

Keywords: Survival analysis. Maximum likelihood estimator. Fisher information matrix.

Artigo recebido em set. 2018 e aceito em fev. 2019.

1 Introdução

A análise de confiabilidade é um método estatístico importante para avaliar a confiabilidade de componentes e sistemas eletrônicos. Com o advento de novas tecnologias e processos de fabricação, os itens eletrônicos operam por muito mais tempo antes de falhar, portanto, o custo para a realização de uma análise de confiabilidade onde os itens são postos em testes até todos falharem torna-se oneroso. Neste caso, os testes de confiabilidade geralmente são finalizados após um período específico de tempo (censura Tipo 1) ou após ter sido observado um número específico de falhas (censura Tipo 2). Há inúmeras referências para o planejamento de testes de confiabilidade disponíveis na literatura estatística e de engenharia, dentre elas destacam-se Lawless (2011), e algumas mais específicas em engenharia abordadas por O'Connor e Kleyner (2012), Karbasian; Rostami Mehr; Agharajabi (2012) e Jalali Naini et al (2009).

A distribuição Exponencial Logarítmica, proposta por Tahmasbi e Rezaei (2008), é uma distribuição de confiabilidade com taxa de falha decrescente, com interessantes aplicações nas áreas de ciências biológicas e engenharia. Esta distribuição pode ser usada para estudar sobrevivência de organismos, dispositivos, materiais, etc. Neste caso, pode ser uma boa alternativa para estudos de tempo de vida devido às formas simples das suas funções de sobrevivência e risco, entre algumas distribuições com taxa de falha decrescente como a Gama.

Tahmasbi e Rezaei (2008) apresentam uma discussão detalhada sobre sua aplicação e propriedades probabilísticas, além da estimação de máxima verossimilhança para dados completos. Moala e Garcia (2013) desenvolveram uma análise Bayesiana para a distribuição Exponencial Logarítmica sob dados completos, isto é, dados não-censurados.

Neste artigo o principal objetivo é realizar uma estimação de máxima verossimilhança para obtenção dos estimadores dos parâmetros da distribuição Exponencial Logarítmica na presença de dados censurados à direita. Essa análise ocorre por meio da obtenção dos estimadores dos parâmetros da distribuição e da matriz de informação de Fisher. Avaliar os desempenhos dos estimadores, através de simulações tais como, as medidas de vício, erro quadrático médio e probabilidade de cobertura. Três problemas reais considerando dados completos e censurados tipo 1 e tipo 2 foram analisados para fins ilustrativos.

O artigo está organizado da seguinte forma: na Seção 2, a distribuição Exponencial Logarítmica é apresentada e algumas de suas propriedades são revisadas; na Seção 3, realizamos uma estimação da máxima verossimilhança dos parâmetros e a matriz de informação observada de Fisher para dados censurados do tipo 1 e tipo 2; a seção 4 ilustra e discute os resultados do desempenho da simulação; três exemplos abordando aplicações práticas sob dados completos e censurados do tipo 1 e tipo 2 são introduzidos na Seção 5 e finalmente na seção 6, apresentamos as conclusões do estudo proposto neste artigo.

As principais diferenças desse artigo em relação ao resumo enviado para o ERMAC foram os acréscimos de simulações usando censura do tipo 1 e utilização de três problemas reais.

2 Distribuição exponencial logarítmica: definição e propriedades

A distribuição Exponencial Logarítmica é uma composição da distribuição Exponencial com a distribuição Logarítmica.

Se X é uma variável aleatória (v.a.) com distribuição Exponencial Logarítmica, então sua função densidade de probabilidade é definida por:

$$f(x;p,\beta) = \left(-\frac{1}{\ln p}\right)\left(\frac{\beta(1-p)\mathrm{e}^{-\beta x}}{1-(1-p)\mathrm{e}^{-\beta x}}\right) \quad (1)$$

onde $0 < p < 1$ e $\beta > 0$ são os parâmetros de forma e escala, respectivamente.
A função de sobrevivência, definida como a probabilidade de não ocorrer a falha até certo tempo t, é uma das principais funções utilizadas nos estudos de sobrevivência e definida por $S(x;p,\beta) = P\{X > x\}$. Para a distribuição Exponencial Logarítmica, a função de sobrevivência é dada por:

$$S(x;p,\beta) = \frac{\ln(1-(1-p)\mathrm{e}^{-\beta x})}{\ln p} \quad (2)$$

Outra função bastante utilizada na análise de sobrevivência é a função de taxa de falha (ou risco). Ela pode assumir diferentes formas de curva como risco crescente, decrescente, constante, banheira e unimodal. As funções de taxa de falha se diferenciam mais entre si do que as funções de sobrevivência, tornando-as mais informativas para o ajustamento dos dados. A função de taxa de falha da distribuição Exponencial Logarítmica tem a seguinte forma:

$$h(x;p,\beta) = \frac{-\beta(1-p)\mathrm{e}^{-\beta x}}{(1-(1-p)\mathrm{e}^{-\beta x})\ln(1-(1-p)\mathrm{e}^{-\beta x})} \quad (3)$$

As (Figuras 1 e 2) apresentam as formas das funções densidade e de taxa de falha para diferentes valores dos parâmetros p e β.

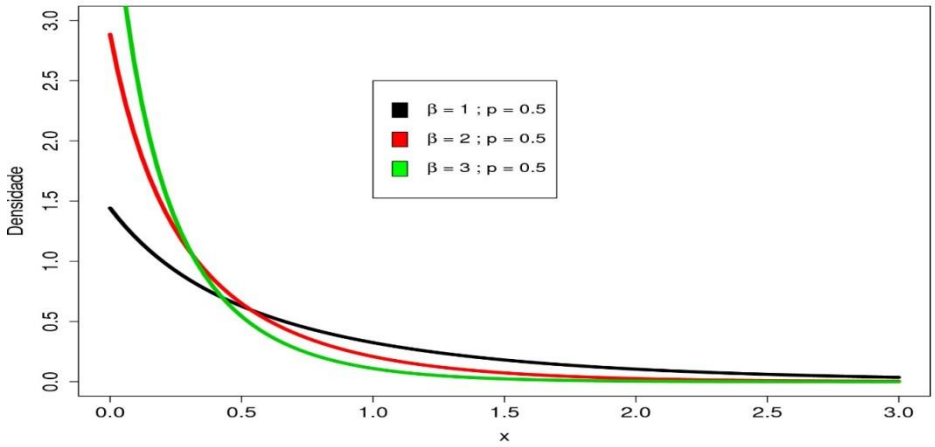

Figura 1 – Função densidade de probabilidade com diferentes parâmetros

Tahmasbi e Rezaei (2008) obtêm a função geratriz de momentos da distribuição Exponencial Logarítmica como:

$$M_x(t) = -\frac{\beta(1-p)}{\ln p(\beta-t)}\, hipergeom_{2,1}\left(\left[1,\frac{\beta-t}{\beta}\right],\left[\frac{2\beta-t}{\beta}\right],1-p\right) \quad (4)$$

em que hipergeom é a função hipergeométrica estendida de Barnes, definida em Barnes (1908) definida como:

$$F_{p,q}(n,d,\lambda) = \sum_{k=0}^{\infty} \frac{\lambda^k \prod_{i=1}^{p} \Gamma(n_i + k)\Gamma^{-1}(n_i)}{\Gamma(k+1) \prod_{i=1}^{q} \Gamma(d_i + k)\Gamma^{-1}(d_i)} \quad (5)$$

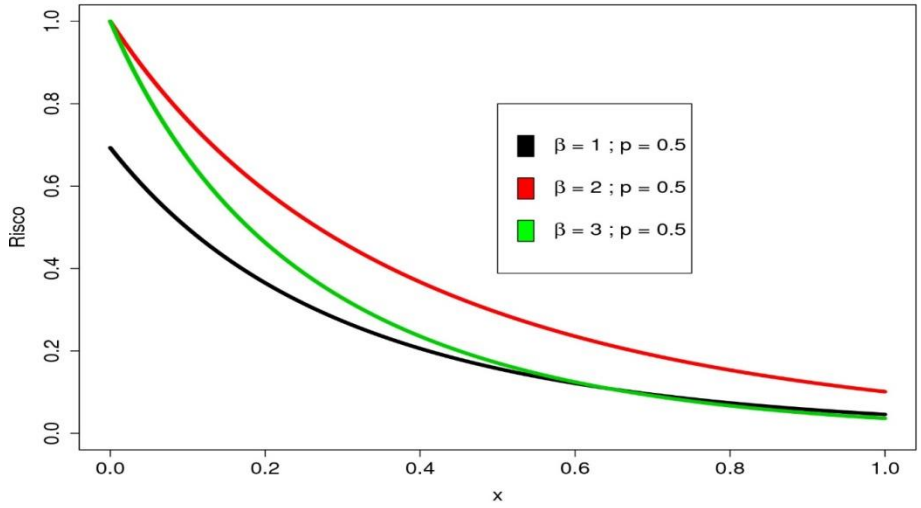

Figura 2 – Função de taxa de falha com diferentes parâmetros

Por meio da função geratriz $M_x(t)$ é possível determinar os momentos da v.a. X considerando t = 0, sendo o primeiro momento a esperança, E(X), e o segundo a variância, Var(X). Dessa forma a esperança e a variância da distribuição Exponencial Logarítmica são expressas, respectivamente, por:

$$E(X) = -\frac{polilog(2,1-p)}{\beta \ln p} \quad (6)$$

$$var(X) = \frac{2polilog(3,1-p)}{\beta^2 \ln p} - \frac{polilog^2(2,1-p)}{\beta^2 \ln^2 p} \quad (7)$$

em que polilog é a função poli-logaritmo, definida em Lewin (1981) e definida como:

$$polilog(a,z) = \sum_{k=1}^{\infty} \frac{z^k}{k^a} \quad (8)$$

A geração de amostras aleatórias da distribuição Exponencial Logarítmica se baseia no método da transformação inversa em que devemos igualar U tendo distribuição Uniforme, U(0,1), a uma distribuição acumulada, $F(x)$, dada por:

$$F(x;p,\beta) = 1 - \frac{\ln\left(1-(1-p)\mathrm{e}^{-\beta x}\right)}{\ln p} \tag{9}$$

Assim, usando a transformação inversa geramos os dados para a distribuição Exponencial Logarítmica. Dada uma variável aleatória sob a distribuição Uniforme no intervalo (0,1), então a variável:

$$X = \frac{1}{\beta}\ln\left(\frac{1-p}{1-p^{1-U}}\right) \tag{10}$$

tem distribuição Exponencial Logarítmica com parâmetros p e β.

3 Estimação de máxima verossimilhança

3.1 Dados completos

Suponha uma amostra aleatória (completa) aleatória $X_1, X_2, \ldots, X_n$ da distribuição Exponencial Logarítmica dada em (1), em que os parâmetros p e β são desconhecidos, então a função de verossimilhança é definida por:

$$L(p,\beta) = \prod_{i=1}^{n} f(x_i,p,\beta) \tag{11}$$

ou seja:

$$L(p,\beta) = \left(-\frac{1}{\ln p}\right)^n \frac{\beta^n (1-p)^n \mathrm{e}^{-\beta\sum_{i=1}^n x_i}}{\prod_{i=1}^n [1-(1-p)\mathrm{e}^{-\beta x_i}]} \tag{12}$$

Os EMV dos parâmetros p e β são os valores, baseado na amostra observada, que maximizam o logaritmo da função de verossimilhança dada por:

$$\log L(p,\beta) = -n\ln(-\ln p) + n\ln(1-p) + n\ln(\beta) - \beta\sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\ln\left(1-(1-p)\mathrm{e}^{-\beta x_i}\right) \tag{13}$$

Assim, os estimadores de máxima verossimilhança (EMV) para os parâmetros p e β são obtidos como as soluções das derivadas de $log\, L(p,\beta)$ descritas por:

$$\frac{\partial}{\partial p} logL(p,\beta) = -\frac{n}{p\log(p)} - \frac{n}{1-p} - \sum_{i=1}^{n}\frac{\mathrm{e}^{-\beta x_i}}{1-(1-p)\mathrm{e}^{-\beta x_i}} \tag{14}$$

$$\frac{\partial}{\partial \beta} logL(p,\beta) = \frac{n}{\beta} - \sum_{i=1}^{n} x_i - (1-p)\sum_{i=1}^{n}\frac{x_i \mathrm{e}^{-\beta x_i}}{1-(1-p)\mathrm{e}^{-\beta x_i}} \tag{15}$$

Tahmasbi e Rezaei (2008) também mostraram em detalhes que, para a distribuição Exponencial Logarítmica, a matriz de informação de Fisher é dada por:

$$I(p,\beta) = \begin{bmatrix} I_{1,1} & I_{1,2} \\ I_{2,1} & I_{2,2} \end{bmatrix}$$

$$I_{1,1} = \frac{-n(\ln p+1)}{(p\ln p)^2} + \frac{n}{(1-p)^2} - \sum_{i=1}^{n} \frac{e^{-2\beta x_i}}{(1-(1-p)e^{-\beta x_i}} \quad (16)$$

$$I_{1,2} = I_{2,1} = -\sum_{i=1}^{n} \frac{x_i e^{-\beta x_i}}{(1-(1-p)e^{-\beta x_i}} \quad (17)$$

$$I_{2,2} = \frac{n}{\beta^2} - (1-p)\sum_{i=1}^{n} \frac{x_i^2 e^{-\beta x_i}}{(1-(1-p)e^{-\beta x_i}} \quad (18)$$

Para amostras grandes, os intervalos de confiança aproximados para os parâmetros podem ser construídos através da distribuição assintoticamente normal dos estimadores de máxima verossimilhança, isto é:

$$(\hat{p},\hat{\beta}) \sim N_2\left((p,\beta), I^{-1}(\hat{p},\hat{\beta})\right) \text{ , for } n \to \infty .$$

em que $I(p,\beta)$ é a matriz de informação de Fisher observada.

3.2 Censura tipo 1

Um esquema de censura do tipo 1 pode ser descrito como se segue.
Seja uma amostra aleatória $X_1, X_2, \ldots, X_n$ de tamanho n de tempos de vida sob distribuição com densidade $f(x \mid p,\beta)$. Na presença de dados censurados ou não-censurados define-se a variável indicadora δ_i, como:

$$\delta_i = \begin{cases} 1 \,;\, T_i \le L_i \\ 0 \,;\, T_i > L_i \end{cases}$$

em que $x_i = \min(T_i, C_i)$ é o tempo observado, T_i é o tempo de sobrevivência e L_i é o tempo de censura assumido fixo para o i-ésimo indivíduo. A variável δ_i indica se X_i é censurado ou não. Neste caso, a função de verossimilhança para o parâmetro é dada (ver, por exemplo, Lawless (2011)) por:

$$L(p,\beta) = \prod_{i=1}^{n} f(x_i \mid p,\beta)^{\delta_i} S(x_i \mid p,\beta)^{1-\delta_i} \quad (19)$$

Sendo $r = \sum_{i=1}^{n} \delta_i$ o número de tempos de vida observados então os estimadores de máxima verossimilhança de (p,β) maximizam a função de verossimilhança obtida da equação (19) como:

$$L(p,\beta) = \prod_{i=1}^{n}\left[\left(-\frac{1}{\ln p}\right)\left(\frac{\beta(1-p)\mathrm{e}^{-\beta x_i}}{1-(1-p)\mathrm{e}^{-\beta x_i}}\right)\right]^{\delta_i}\prod_{i=1}^{n}\left[\frac{\ln\left(1-(1-p)\mathrm{e}^{-\beta x_i}\right)}{\ln p}\right]^{1-\delta_i} \quad (20)$$

Desse modo, o logaritmo da função de máxima verossimilhança (20) é dado por:

$$log\, L(p,\beta) = -r\ln(-\ln p) + r\ln(1-p) + r\ln\beta - \beta\sum_{i=1}^{n}\delta_i x_i - \sum_{i=1}^{n}\delta_i \ln\left(1-(1-p)\mathrm{e}^{-\beta x_i}\right) - (n-r)\ln(\ln p) + \sum_{i=1}^{n}(1-\delta_i)\ln(\ln\left(1-(1-p)\mathrm{e} \quad (21)$$

Derivando (21) com relação a p e β, chegamos às equações de verossimilhança, que são expressas por:

$$\frac{\partial}{\partial p} logL(p,\beta) = \frac{n}{p\log(p)} - \frac{r}{1-p} + \sum_{i=1}^{n}\frac{\delta_i \mathrm{e}^{-\beta x_i}}{1-(1-p)e^{-\beta x_i}} + \sum_{i=1}^{n}\frac{(1-\delta_i)\mathrm{e}^{-\beta x_i}}{\log(1-(1-p)\mathrm{e}^{-\beta x_i})\,(1-(1-p)\mathrm{e}^{-\beta x_i})} \quad (22)$$

$$\frac{\partial}{\partial \beta}\, logL(p,\beta) = \frac{r}{\beta} - \sum_{i=1}^{n}\delta_i x_i - (1-p)\sum_{i=1}^{n}\frac{\delta_i x_i \mathrm{e}^{-\beta x_i}}{1-(1-p)\mathrm{e}^{-\beta x_i}} + (1-p)\sum_{i=1}^{n}\frac{(1-\delta_i)x_i \mathrm{e}^{-\beta x_i}}{\log(1-(1-p)\mathrm{e}^{-\beta x_i})\,(1-(1-p)\mathrm{e}^{-\beta x_i})} \quad (23)$$

Os estimadores de máxima verossimilhança de p e β podem ser obtidos resolvendo simultaneamente as equações:

$$\frac{\partial}{\partial p}\log L(\mathrm{p},\beta)=0 \quad \text{e} \quad \frac{\partial}{\partial \beta}\log L(\mathrm{p},\beta)=0\,. \quad (24)$$

Uma vez que não se pode resolver o sistema de equações (22 e 23) analiticamente para obter os EMV $\hat{p}$ e $\hat{\beta}$, algum método numérico deve ser aplicado.

Teorema: A matriz de informação de Fisher observada para os parâmetros (p,β) da distribuição Exponencial Logarítmica dada em (1) sob esquema de censura tipo 1 é dada por:

$$I(p,\beta) = \begin{bmatrix} I_{11} & I_{12} \\ I_{12} & I_{22} \end{bmatrix}$$

em que:

$$
\begin{aligned}
I_{11} = & -\frac{r}{(1-p)^2} + \frac{r}{p^2\ln(p)} + \frac{r}{p^2\ln(p)^2} - \left(\sum_{i=1}^{n}\left(-\frac{\delta_i\left(\mathrm{e}^{-\beta x_i}\right)^2}{\left(1-(1-p)\,\mathrm{e}^{-\beta x_i}\right)^2}\right)\right) \\
& + \frac{1}{\ln\left(1-(1-p)\,\mathrm{e}^{-\beta x_i}\right)}\left(\left(1-\delta_i\right)\left(-\frac{\left(\mathrm{e}^{-\beta x_i}\right)^2}{\left(1-(1-p)\,\mathrm{e}^{-\beta x_i}\right)^2\ln(p)}\right.\right. \\
& - \frac{2\,\mathrm{e}^{-\beta x_i}}{\left(1-(1-p)\,\mathrm{e}^{-\beta x_i}\right)\ln(p)^2 p} + \frac{2\ln\left(1-(1-p)\,\mathrm{e}^{-\beta x_i}\right)}{\ln(p)^3 p^2} \\
& \left.\left. + \frac{\ln\left(1-(1-p)\,\mathrm{e}^{-\beta x_i}\right)}{\ln(p)^2 p^2}\right)\ln(p)\right) \\
& - \frac{\left(1-\delta_i\right)\left(\dfrac{\mathrm{e}^{-\beta x_i}}{\left(1-(1-p)\,\mathrm{e}^{-\beta x_i}\right)\ln(p)} - \dfrac{\ln\left(1-(1-p)\,\mathrm{e}^{-\beta x_i}\right)}{\ln(p)^2 p}\right)\ln(p)\,\mathrm{e}^{-\beta x_i}}{\ln\left(1-(1-p)\,\mathrm{e}^{-\beta x_i}\right)^2\left(1-(1-p)\,\mathrm{e}^{-\beta x_i}\right)} \\
& + \frac{\left(1-\delta_i\right)\left(\dfrac{\mathrm{e}^{-\beta x_i}}{\left(1-(1-p)\,\mathrm{e}^{-\beta x_i}\right)\ln(p)} - \dfrac{\ln\left(1-(1-p)\,\mathrm{e}^{-\beta x_i}\right)}{\ln(p)^2 p}\right)}{\ln\left(1-(1-p)\,\mathrm{e}^{-\beta x_i}\right)p}
\end{aligned}
\tag{25}
$$

$$
\begin{aligned}
I_{12} = I_{21} = & -\left(\sum_{i=1}^{n}\left(-\frac{\delta_i x_i\,\mathrm{e}^{-\beta x_i}}{1-(1-p)\,\mathrm{e}^{-\beta x_i}} - \frac{\delta_i\left(\mathrm{e}^{-\beta x_i}\right)^2(1-p)\,x_i}{\left(1-(1-p)\,\mathrm{e}^{-\beta x_i}\right)^2}\right)\right) \\
& + \frac{1}{\ln\left(1-(1-p)\,\mathrm{e}^{-\beta x_i}\right)}\left(\left(1-\delta_i\right)\left(-\frac{x_i\,\mathrm{e}^{-\beta x_i}}{\left(1-(1-p)\,\mathrm{e}^{-\beta x_i}\right)\ln(p)}\right.\right. \\
& \left.\left. - \frac{\left(\mathrm{e}^{-\beta x_i}\right)^2(1-p)\,x_i}{\left(1-(1-p)\,\mathrm{e}^{-\beta x_i}\right)^2\ln(p)} - \frac{(1-p)\,x_i\,\mathrm{e}^{-\beta x_i}}{\left(1-(1-p)\,\mathrm{e}^{-\beta x_i}\right)\ln(p)^2 p}\right)\ln(p)\right) \\
& - \frac{1}{\ln\left(1-(1-p)\,\mathrm{e}^{-\beta x_i}\right)^2\left(1-(1-p)\,\mathrm{e}^{-\beta x_i}\right)}\left(\left(1\right.\right. \\
& \left.\left. - \delta_i\right)\left(\frac{\mathrm{e}^{-\beta x_i}}{\left(1-(1-p)\,\mathrm{e}^{-\beta x_i}\right)\ln(p)} - \frac{\ln\left(1-(1-p)\,\mathrm{e}^{-\beta x_i}\right)}{\ln(p)^2 p}\right)\ln(p)\,(1-p)\,x_i\,\mathrm{e}^{-\beta x_i}\right)
\end{aligned}
\tag{26}
$$

$$
\begin{aligned}
\mathrm{I}_{22} = & -\frac{r}{\beta^2} - \left(\sum_{i=1}^{n} \left(-\frac{\delta_i (1-p)\, x_i^2\, \mathrm{e}^{-\beta x_i}}{1-(1-p)\,\mathrm{e}^{-\beta x_i}} - \frac{\delta_i (1-p)^2 x_i^2 \left(\mathrm{e}^{-\beta x_i}\right)^2}{\left(1-(1-p)\,\mathrm{e}^{-\beta x_i}\right)^2} \right) \right) \\
& - \frac{\left(1-\delta_i\right)(1-p)\, x_i^2\, \mathrm{e}^{-\beta x_i}}{\left(1-(1-p)\,\mathrm{e}^{-\beta x_i}\right) \ln\left(1-(1-p)\,\mathrm{e}^{-\beta x_i}\right)} \\
& - \frac{\left(1-\delta_i\right)(1-p)^2 x_i^2 \left(\mathrm{e}^{-\beta x_i}\right)^2}{\left(1-(1-p)\,\mathrm{e}^{-\beta x_i}\right)^2 \ln\left(1-(1-p)\,\mathrm{e}^{-\beta x_i}\right)} \\
& - \frac{\left(1-\delta_i\right)(1-p)^2 x_i^2 \left(\mathrm{e}^{-\beta x_i}\right)^2}{\left(1-(1-p)\,\mathrm{e}^{-\beta x_i}\right)^2 \ln\left(1-(1-p)\,\mathrm{e}^{-\beta x_i}\right)^2}
\end{aligned}
\tag{27}
$$

prova: Os elementos da matriz de Fisher observada são obtidos através das derivadas de segunda ordem do logaritmo da função de verossimilhança em relação a cada um dos parâmetros como:

$$I_{11} = \frac{\partial^2}{\partial p^2} \log L(\mathrm{p}, \beta) \tag{28}$$

$$I_{12} = I_{21} = \frac{\partial^2}{\partial p^2} \log L(\mathrm{p}, \beta) \tag{29}$$

$$I_{22} = \frac{\partial^2}{\partial p^2} \log L(\mathrm{p}, \beta) \tag{30}$$

3.3 Censura tipo 2

Num esquema de censuras tipo 2, as unidades são colocadas em teste e o experimento termina após a ocorrência da r-ésima falha (r fixo, $0 \leq r \leq n$).
A função de verossimilhança baseada nos r tempos ordenados $x_{(1)}, x_{(2)}, \ldots, x_{(r)}$ de falhas é dada por:

$$L(p,\beta) = \frac{n!}{(n-r)!} f(x_{(1)} \mid p,\beta) \ldots f(x_{(r)} \mid p,\beta)[S(x_{(r)} \mid p,\beta)]^{n-r} \tag{31}$$

em que $x_{(r)}$ é o tempo de vida correspondente à ocorrência da r-ésima falha.
Deste modo, para os tempos de falha dos componentes ajustados a uma distribuição Exponencial Logarítmica dada em (1), a função de verossimilhança correspondente ao esquema de censura tipo 2 é dada por:

$$L(p,\beta)=\frac{n!}{(n-r)!}\prod_{i=1}^{r}\left[\left(-\frac{1}{\ln p}\right)\left(\frac{\beta(1-p)e^{-\beta x_i}}{1-(1-p)e^{-\beta x_i}}\right)\right]\left[\frac{\ln\left(1-(1-p)e^{-\beta x_i}\right)}{\ln p}\right]^{n-r} \quad (32)$$

Assim, o logaritmo da função de máxima verossimilhança (32) é dado por:

$$\log L(p,\beta) = -r\ln(-\ln p) + r\ln(1-p) + r\ln(\beta) - \beta\sum_{i=1}^{r} x_i - \sum_{i=1}^{r}\ln\left(1-(1-p)e^{-\beta x_i}\right) + (n-r)\sum_{i=1}^{n-r}\ln(\ln\left(1-(1-p)e^{-\beta x_i}\right)) - (n-r)\ln(\ln(p)) \quad (33)$$

Derivando (33) com relação a p e β, obtém-se as equações de verossimilhança, que são expressas por:

$$\frac{\partial}{\partial p}logL(p,\beta) = -\frac{r}{1-p} - \frac{r}{p\ln(p)} - \left(\sum_{i=1}^{n}\frac{e^{-\beta x_i}}{1-(1-p)e^{-\beta x_i}}\right) + (n-r)\left(\sum_{i=1}^{n-r}\frac{e^{-\beta x_i}}{\left(1-(1-p)e^{-\beta x_i}\right)\ln\left(1-(1-p)e^{-\beta x_i}\right)}\right) - \frac{n-r}{p\ln(p)}$$

$$\frac{\partial}{\partial \beta}logL(p,\beta) = \frac{r}{\beta} - \left(\sum_{i=1}^{n}x_i\right) - \left(\sum_{i=1}^{n}\frac{(1-p)x_i e^{-\beta x_i}}{1-(1-p)e^{-\beta x_i}}\right) + (n-r)\left(\sum_{i=1}^{n-r}\frac{(1-p)x_i e^{-\beta x_i}}{\left(1-(1-p)e^{-\beta x_i}\right)\ln\left(1-(1-p)e^{-\beta x_i}\right)}\right) \quad (34)$$

Os Estimadores de Máxima Verossimilhança de (p,β) da distribuição Exponencial Logarítmica dada em (1), sob esquema de censura tipo 2, são obtidos resolvendo simultaneamente as equações:

$$\frac{\partial}{\partial p}\log L(p,\beta)=0 \quad \text{e} \quad \frac{\partial}{\partial \beta}\log L(p,\beta)=0\,. \quad (35)$$

Teorema: A matriz de informação de Fisher observada para os parâmetros (p,β) da distribuição EL dada em (1) sob esquema de censura tipo 2 é dada por:

$$I(p,\beta) = \begin{bmatrix} I_{11} & I_{12} \\ I_{12} & I_{22} \end{bmatrix}$$

em que:

$$I_{11} = -\frac{r}{(1-p)^2} + \frac{r}{p^2\ln(p)} + \frac{r}{p^2\ln(p)^2} - \left(\sum_{i=1}^{n}\left(-\frac{\left(e^{-\beta x_i}\right)^2}{\left(1-(1-p)e^{-\beta x_i}\right)^2}\right)\right) + (n-r)\left(\sum_{i=1}^{n-r}\left(-\frac{\left(e^{-\beta x_i}\right)^2}{\left(1-(1-p)e^{-\beta x_i}\right)^2\ln\left(1-(1-p)e^{-\beta x_i}\right)} - \frac{\left(e^{-\beta x_i}\right)^2}{\left(1-(1-p)e^{-\beta x_i}\right)^2\ln\left(1-(1-p)e^{-\beta x_i}\right)^2}\right)\right) + \frac{n-r}{p^2\ln(p)} + \frac{n-r}{p^2\ln(p)^2} \quad (36)$$

$$I_{12} = I_{21} = -\left(\sum_{i=1}^{n}\left(-\frac{x_i e^{-\beta x_i}}{1-(1-p)e^{-\beta x_i}} - \frac{\left(e^{-\beta x_i}\right)^2(1-p)x_i}{\left(1-(1-p)e^{-\beta x_i}\right)^2}\right)\right) + (n-r)\left(\sum_{i=1}^{n-r}\left(-\frac{x_i e^{-\beta x_i}}{\left(1-(1-p)e^{-\beta x_i}\right)\ln\left(1-(1-p)e^{-\beta x_i}\right)} - \frac{\left(e^{-\beta x_i}\right)^2(1-p)x_i}{\left(1-(1-p)e^{-\beta x_i}\right)^2\ln\left(1-(1-p)e^{-\beta x_i}\right)} - \frac{\left(e^{-\beta x_i}\right)^2(1-p)x_i}{\left(1-(1-p)e^{-\beta x_i}\right)^2\ln\left(1-(1-p)e^{-\beta x_i}\right)^2}\right)\right) \quad (37)$$

$$I_{22} = -\frac{r}{\beta^2} - \left(\sum_{i=1}^{n}\left(-\frac{(1-p)x_i^2 e^{-\beta x_i}}{1-(1-p)e^{-\beta x_i}} - \frac{(1-p)^2 x_i^2\left(e^{-\beta x_i}\right)^2}{\left(1-(1-p)e^{-\beta x_i}\right)^2}\right)\right) + (n-r)\left(\sum_{i=1}^{n-r}\left(-\frac{(1-p)x_i^2 e^{-\beta x_i}}{\left(1-(1-p)e^{-\beta x_i}\right)\ln\left(1-(1-p)e^{-\beta x_i}\right)} - \frac{(1-p)^2 x_i^2\left(e^{-\beta x_i}\right)^2}{\left(1-(1-p)e^{-\beta x_i}\right)^2\ln\left(1-(1-p)e^{-\beta x_i}\right)} - \frac{(1-p)^2 x_i^2\left(e^{-\beta x_i}\right)^2}{\left(1-(1-p)e^{-\beta x_i}\right)^2\ln\left(1-(1-p)e^{-\beta x_i}\right)^2}\right)\right) \quad (38)$$

prova: Os elementos da matriz de Fisher observada são obtidos através das derivadas de segunda ordem do logaritmo da função de verossimilhança em relação a cada parâmetros como:

$$I_{11} = \frac{\partial^2}{\partial p^2} \log L(\mathrm{p}, \beta) \tag{39}$$

$$I_{12} = I_{21} = \frac{\partial^2}{\partial p^2} \log L(\mathrm{p}, \beta) \tag{40}$$

$$I_{22} = \frac{\partial^2}{\partial p^2} \log L(\mathrm{p}, \beta) \tag{41}$$

4 Simulação

Nesta seção, foi desenvolvido um estudo de simulação via métodos de Monte Carlo. O principal objetivo desta simulação foi estudar a eficiência da abordagem proposta. Nesta simulação, foi investigada a influência do tamanho da amostra e o percentual de censura na precisão das estimativas. O desempenho dos estimadores de máxima verossimilhança dos parâmetros p e β foram avaliados em termos da média e os desvios padrões dos estimadores, erro quadrático médio e probabilidade de cobertura de 95%. Para isto, foram considerados diferentes tamanhos amostrais e taxas de censura.

Para isso 500 simulações foram realizadas, gerando amostras com (n = 30, 50, 70), utilizando os parâmetros $p = 0{,}5$ e $\beta = 2$, utilizando a censura do tipo 1. E no caso da censura do tipo 2, o mesmo número de simulações utilizando os mesmos parâmetros, porém com tamanho amostral diferente (n = 100, 110 e 120). Por conta da impossibilidade de se calcular a raiz quadrada da inversa da matriz de informação de Fisher, com n < 100 e 10% de censura.

Cens	n	$\hat{p}$	$\hat{\beta}$	Dp $\hat{p}$	Dp $\hat{\beta}$	EQM $\hat{p}$	EQM $\hat{\beta}$	Prob.Cob.95% $\hat{p}$	Prob.Cob.95% $\hat{\beta}$
0%	100	0,584	2,051	0,373	0,407	0,081	0,131	92,6%	96,8%
	110	0,585	2,050	0,358	0,389	0,082	0,126	92,4%	97,0%
	120	0,572	2,036	0,334	0,370	0,072	0,109	92,6%	96,8%
5%	100	0,568	2,038	0,378	0,422	0,086	0,160	90,0%	93,6%
	110	0,571	2,041	0,365	0,405	0,085	0,150	89,8%	94,4%
	120	0,560	2,034	0,339	0,384	0,076	0,141	89,4%	93,6%
10%	100	0,577	2,039	0,420	0,455	0,097	0,187	88,4%	93,2%
	110	0,575	2,038	0,397	0,434	0,093	0,179	87,8%	94,0%
	120	0,571	2,042	0,378	0,415	0,086	0,166	87,6%	93,6%

Quadro 1 – Resultados da Simulação/Censura do tipo 1

Cens	n	$\hat{p}$	$\hat{\beta}$	Dp $\hat{p}$	Dp $\hat{\beta}$	EQM $\hat{p}$	EQM $\hat{\beta}$	Prob.Cob.95% $\hat{p}$	Prob.Cob.95% $\hat{\beta}$
0%	30	0,621	2,118	0,667	0,737	0,136	0,363	87,0%	97,4%
	50	0,597	2,070	0,518	0,570	0,114	0,236	89,2%	97,6%
	70	0,597	2,063	0,449	0,485	0,102	0,173	90,6%	97,0%
10%	30	0,630	2,124	0,671	0,734	0,150	0,468	85,8%	95,4%
	50	0,609	2,070	0,528	0,568	0,131	0,312	84,2%	94,2%
	70	0,604	2,070	0,451	0,484	0,114	0,236	88,4%	94,6%
20%	30	0,638	2,149	0,673	0,735	0,157	0,608	84,4%	91,0%
	50	0,608	2,072	0,516	0,562	0,136	0,383	82,4%	89,8%
	70	0,598	2,053	0,443	0,479	0,119	0,289	85,6%	91,2%
30%	30	0,654	2,163	0,701	0,739	0,174	0,791	81,4%	85,8%
	50	0,621	2,091	0,519	0,559	0,149	0,521	81,6%	86,0%
	70	0,601	2,053	0,436	0,473	0,131	0,395	83,8%	83,2%

Quadro 2 – Resultados da Simulação/Censura do tipo 2

5 Aplicações

5.1 Primeiro exemplo (dados não-censurados)

Este exemplo foi proposto por Lawless (2011) e trata do tempo de falha de um tipo de isolador elétrico a um esforço de tensão constante. Com o propósito de verificar o melhor ajuste aos dados, foi comparada a distribuição proposta (Exponencial Logarítmica) com a distribuição Weibull. Os dados são os seguintes: 0,96; 4,15; 0,19; 0,78; 8,01; 31,75; 7,35; 6,50; 8,27; 33,91; 32,52; 3,16; 4,85; 2,78; 4,67; 1,31; 12,06; 36,71; 72,89.
A estimação dos parâmetros das distribuições Weibull e Exponencial Logarítmica, através do método da máxima verossimilhança, resultaram nos seguintes valores:

Distribuição Weibull				
	EMV	Desvio	Variância	*IC (95%)*
$\hat{\alpha}$	0,008	0,017	$2{,}811.10^{-4}$	(-0,025 ; 0,041)
$\hat{\beta}$	0,283	0,056	0,003	(0,173 ; 0,393)
Distribuição Exponencial Logarítmica				
	EMV	Desvio	Variância	*IC (95%)*
$\hat{p}$	0,115	0,163	0,027	(-0,2057 ; 0,434)
$\hat{\beta}$	0,040	0,020	$4{,}137.10^{-5}$	$(5{,}396.10^{-5}$; 0,080)

Quadro 3 - Resultados do EMV

A partir da Figura 3, pode-se observar que as funções de densidade de probabilidade das distribuições Weibull e Exponencial Logarítmica, acompanham de forma satisfatória a função de densidade empírica.

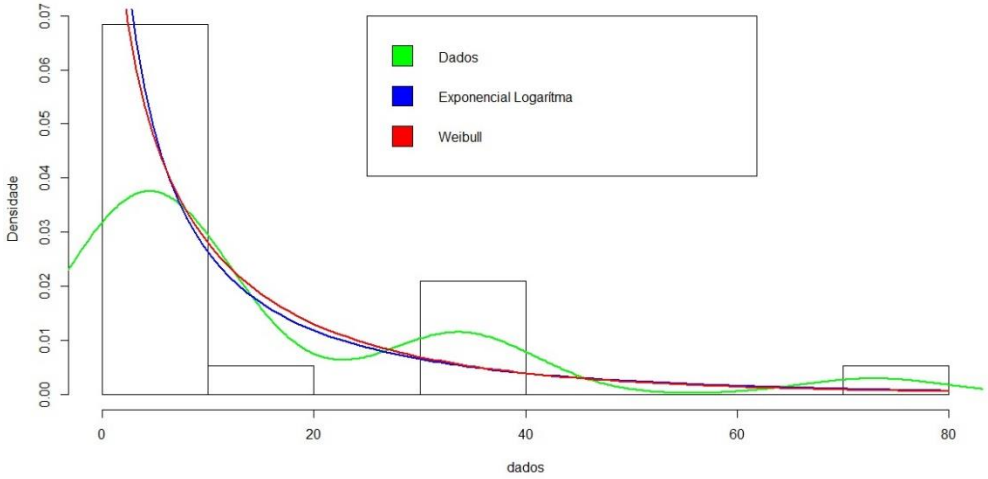

Figura 3 – Função Densidade de probabilidade

O gráfico TTT-plot serve para detectar se a função apresenta risco crescente, decrescente, constante, banheira ou unimodal. A partir da Figura 4, pode-se observar que o comportamento da função de risco dos dados apresenta forma decrescente:

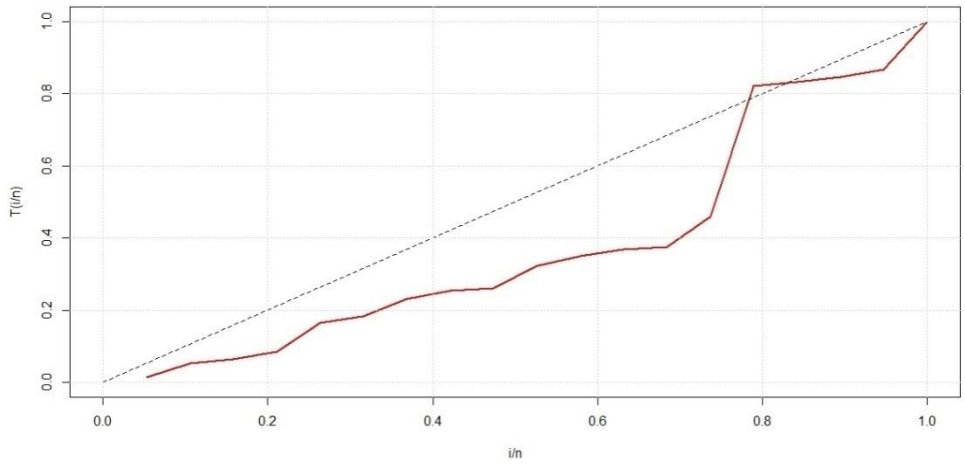

Figura 4 – TTT-plot

As curvas de sobrevivência empírica e das distribuições utilizando o método de Kaplan-Meier podem ser visualizadas na Figura 5:

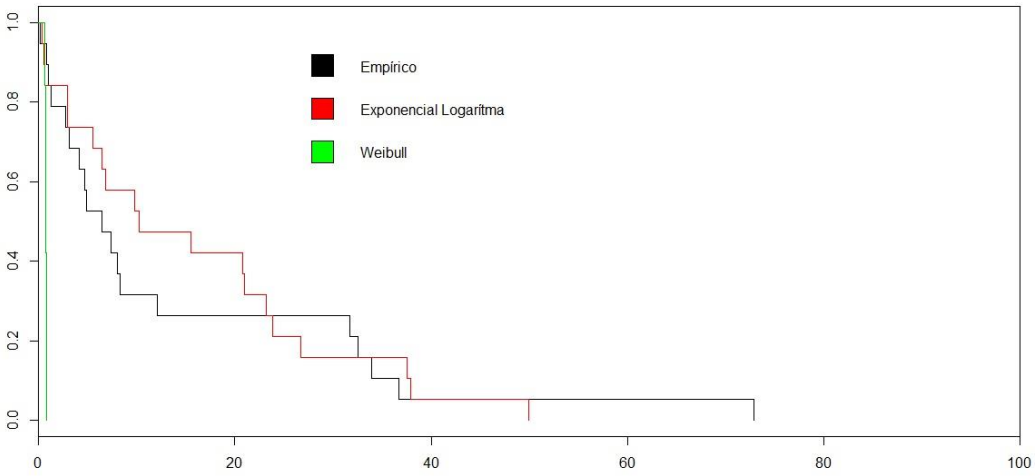

Figura 5 – Função de Sobrevivência utilizando Kaplan-Meier

A partir das Figuras 3 e 5, pode-se perceber que a distribuição Exponencial Logarítmica acompanha melhor a curva de sobrevivência empírica do que a distribuição Weibull.
Para comprovar que o uso da distribuição Exponencial Logarítmica é o mais adequado em relação à distribuição Weibull, para este conjunto de dados, o teste de Kolmogorov-Smirnov foi utilizado obtendo os seguintes resultados:

	Exponencial Logarítmica	Weibull
p-valor	0,796	$7{,}101.10^{-8}$

Quadro 4 - Kolmogorov-Smirnov

Então, com 5% de significância não se rejeita a hipótese H_0: Os dados seguem a distribuição Exponencial Logarítmica, ou seja, não se pode rejeitar a hipótese de que os dados seguem a distribuição Exponencial Logarítmica.
Os resultados apresentados anteriormente evidenciam que o uso da distribuição Exponencial Logarítmica é mais adequado do que a distribuição Weibull, para este conjunto de dados.

5.2 Segundo exemplo (presença de dados censurados tipo 2)

Mann e Fertig (1973) fornecem dados sobre os tempos de falha dos componentes da aeronave sujeitos a um teste de vida. Os dados são obtidos de $n = 13$ itens do teste selecionados aleatoriamente e o teste de vida termina na décima falha observada dos itens sob estudo. Os $r = 10$ tempos de vida observados foram (em horas):

0,22; 0,50; 0,88; 1,00; 1,32; 1,33; 1,34; 1,76; 2,50; 3,00.

Esses dados foram discutidos por Lawless (2011) que ajustou uma distribuição Weibull a esses dados.

Utilizando o gráfico TTT-plot, pode-se observar que o comportamento da função de risco dos dados apresenta forma crescente:

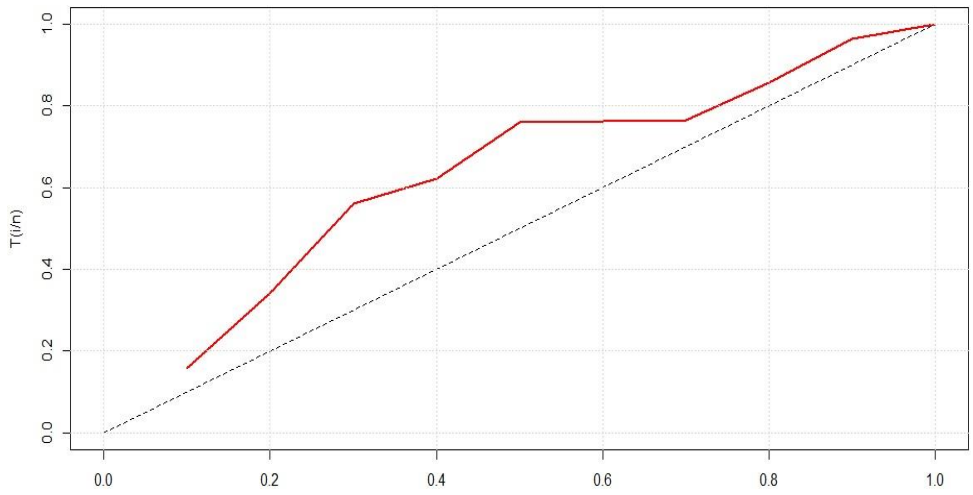

Figura 6 – TTT-plot

A estimação dos parâmetros das distribuições Weibull e Exponencial Logarítmica, através do método da máxima verossimilhança, resultaram nos seguintes valores:

Distribuição Weibull				
	EMV	Desvio	Variância	*IC (95%)*
$\hat{\alpha}$	1,763	0,442	0,195	(0,897;2,628)
$\hat{\beta}$	1,556	0,294	0,086	(0,980;2,132)
Distribuição Exponencial Logarítmica				
	EMV	Desvio	Variância	*IC (95%)*
$\hat{p}$	0,998	1,002	1,005	(-0,967;2,963)
$\hat{\beta}$	0,738	0,326	0,106	(0,098;1,377)

Quadro 5 - Resultados do EMV

As curvas de sobrevivência empírica e das distribuições utilizando o método de Kaplan-Meier são as seguintes:

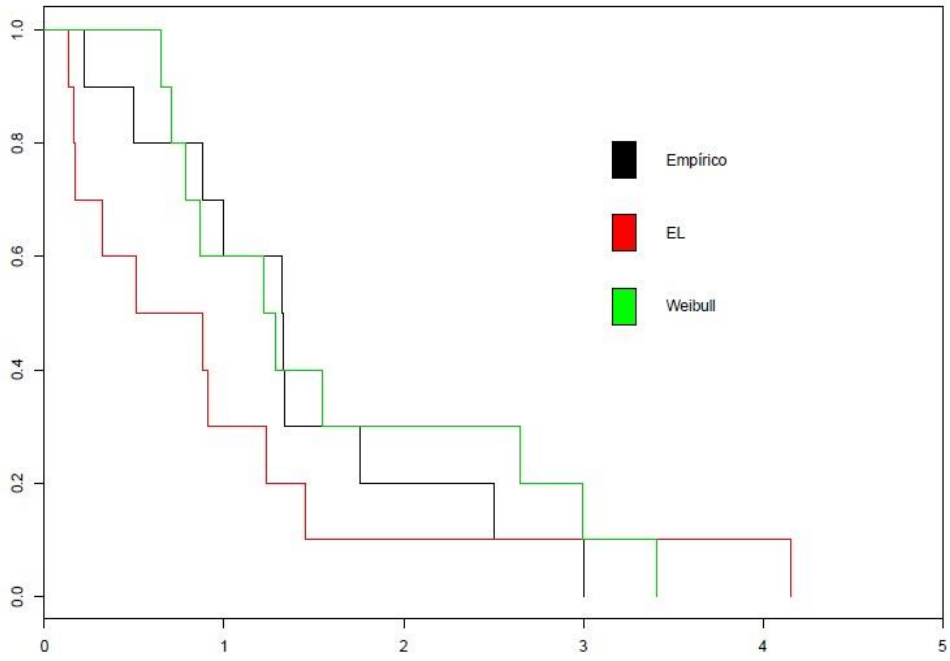

Figura 7 – Função de Sobrevivência utilizando Kaplan-Meier

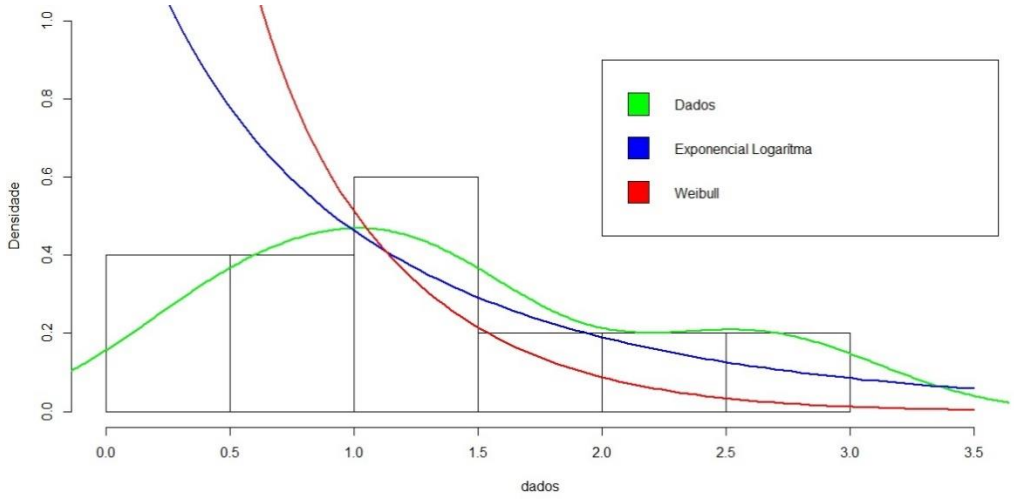

Figura 8 – Função Densidade de Probabilidade

A partir das Figuras 7 e 8, pode-se perceber que a distribuição Weibull, acompanha melhor a curva de sobrevivência empírica, ao invés da distribuição Exponencial Logarítmica.
Para comprovar que o uso da distribuição Weibull é o mais adequado em relação à distribuição Exponencial Logarítmica, para este conjunto de dados, o teste de Kolmogorov-Smirnov foi utilizado obtendo os seguintes resultados:

	Exponencial Logarítmica	Weibull
p-valor	0,355	0,913

Quadro 6 - Kolmogorov-Smirnov

Então, com 5% de significância não se rejeita a hipótese H_0: Os dados seguem a distribuição Exponencial Logarítmica e Weibull, ou seja, não se pode rejeitar a hipótese de que os dados seguem a distribuição Exponencial Logarítmica e Weibull.
Os resultados apresentados nesta Seção evidenciam que o uso da distribuição Weibull é mais adequado ao invés da distribuição Exponencial Logarítmica, para este conjunto de dados.

5.3 Terceiro exemplo (presença de dados censurados)

Os dados são tempos de remissão, em semanas, de um grupo de 30 pacientes com leucemia em um determinado tipo de terapia, as observações com asterisco (*) são tempos de censura.
Os dados foram retirados do Lawless (2011), e são descritos por: 1, 1, 2, 4, 4, 6, 6, 6, 7, 8, 9, 9, 10, 12, 13, 14, 18, 19, 24, 26, 29, 31*, 42, 45*, 50*, 57, 60, 71*, 85*, 91.
Utilizando o gráfico TTT-plot, pode-se observar que o comportamento da função de risco dos dados apresenta forma decrescente:

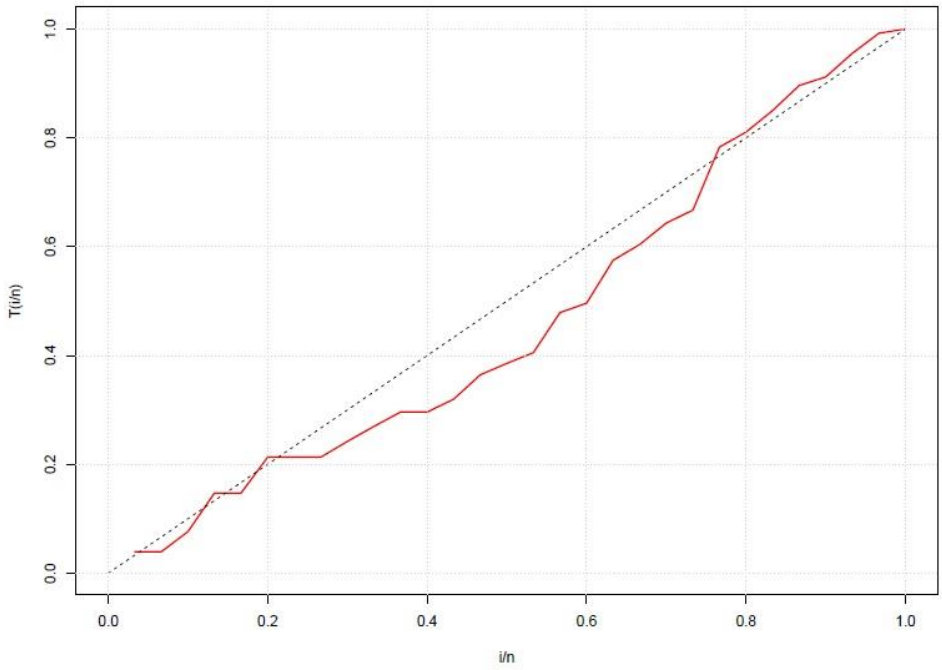

Figura 9 – TTT-plot

A estimação dos parâmetros das distribuições Weibull e Exponencial Logarítmica, através do método da máxima verossimilhança, resultaram nos seguintes valores:

Distribuição Weibull				
	EMV	Desvio	Variância	*IC (95%)*
$\hat{\alpha}$	0,753	0,050	0,002	(0,655;0,851)
$\hat{\beta}$	14,958	1,831	3,354	(11,368;18,547)
Distribuição Exponencial Logarítmica				
	EMV	Desvio	Variância	*IC (95%)*
$\hat{p}$	0,144	0,164	0,027	(-0,178;0,466)
$\hat{\beta}$	0,018	0,008	$5{,}942.10^{5}$	(0,003;0,033)

Quadro 7: Resultados do EMV

A curva de sobrevivência empírica e das distribuições utilizando o método de Kaplan-Meier são as seguintes:

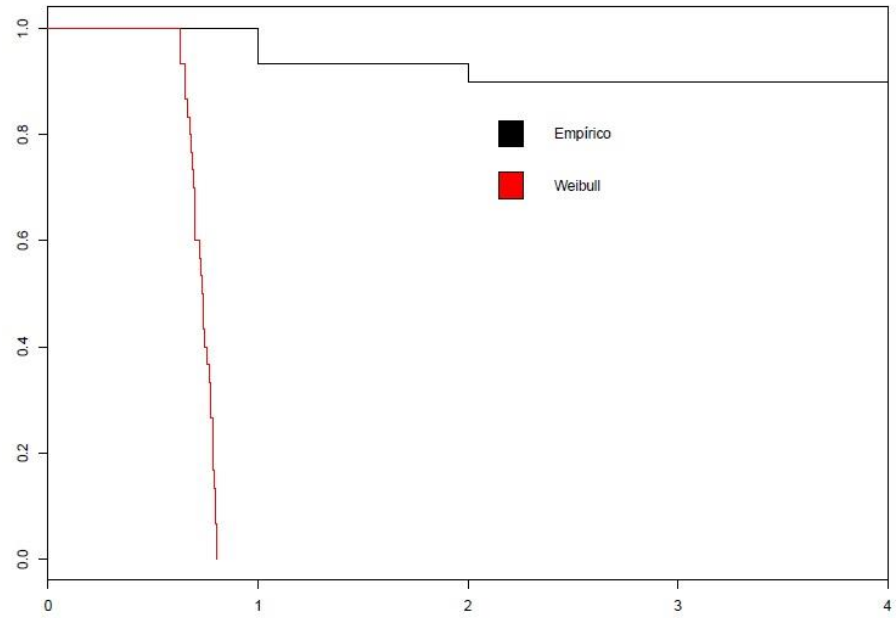

Figura 10 – Função de Sobrevivência utilizando Kaplan-Meier

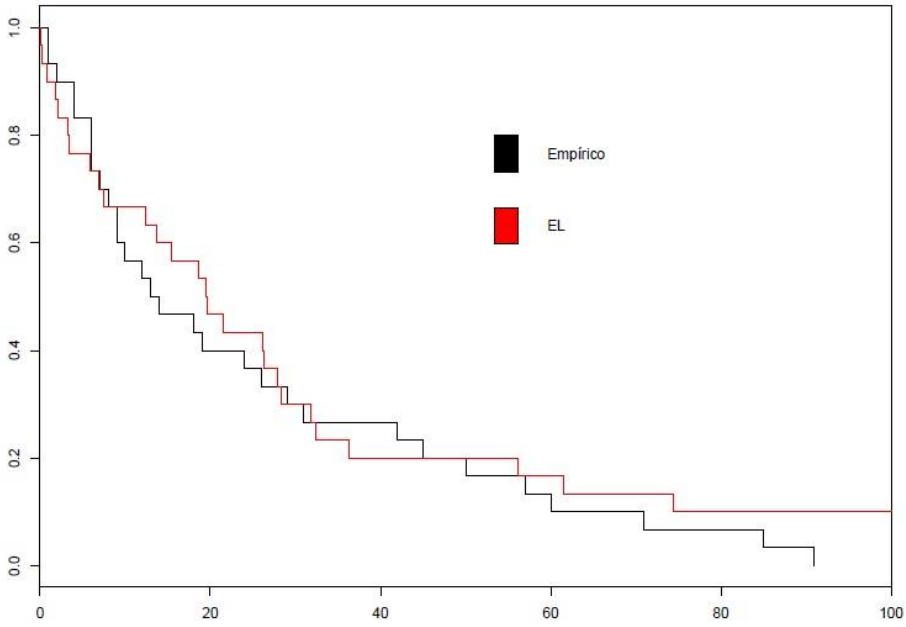

Figura 11 – Função de Sobrevivência utilizando Kaplan-Meier

A partir das Figuras 10 e 11, pode-se perceber que a distribuição Exponencial Logarítmica acompanha melhor a curva de sobrevivência dos dados, ao invés da distribuição Weibull.
Para comprovar que o uso da distribuição Exponencial Logarítmica é adequada para este conjunto de dados, o teste de Kolmogorov-Smirnov foi utilizado obtendo o seguinte resultado:

	Exponencial Logarítmica	Weibull
p-valor	0,944	$2{,}2.10^{-6}$

Quadro 8 - Kolmogorov-Smirnov

Então, com 5% de significância não se rejeita a hipótese H_0: Os dados seguem a distribuição Exponencial Logarítmica, ou seja, não se rejeita a hipótese de que os dados seguem a distribuição Exponencial Logarítmica.

6 Conclusão

A partir dos resultados obtidos pode-se concluir que o método de máxima verossimilhança (Clássico), fornece boas estimativas quando os tamanhos amostrais são relativamente grandes ($n > 30$), mesmo com a presença de censura. A partir dos resultados pode-se notar que quanto maior for o tamanho amostral, para uma porcentagem de censura fixada, menor será o EQM. Pode-se notar também que quanto maior a porcentagem de censura, para um tamanho amostral fixado, maior será o EQM. Porém, os intervalos de confiança dos parâmetros $\hat{p}$ e $\hat{\beta}$ estão incluindo valores negativos e maiores que 1, sendo que $0 < p < 1$ e $\beta > 1$. Uma alternativa seria aplicar a inferência Bayesiana.

Os estudos mostraram que a distribuição Exponencial Logarítmica pode ser uma opção para dados de sobrevivência que apresentam função de taxa de risco decrescente.

7 Referências

BARNES, E. W. A new development of the theory of the hypergeometric functions. **Proceedings of the London Mathematical Society**, v. s2-6, n. 1, 141-177, 1908.

JALALI NAINI, S. G. et al. Condition based maintenance for two-component systems with reliability and cost considerations. **International Journal of Industrial Engineering & Production Research**. v. 20, n. 3, p. 107-116, 2009.

KARBASIAN, M.; ROSTAMI MEHR, M.; AGHARAJABI, M. Improvement of the reliability of automatic manufacture systems by using FTA technique. **International Journal of Industrial Engineering & Production Research**, v. 23, n. 2, p. 85-89, 2012.

LAWLESS, J. F. **Statistical models and methods for lifetime data**. New York: Wiley & Sons, 2011.

MANN, N. R.; FERTIG, K. W. Tables for obtaining Weibull confidence bounds and tolerance bound, based on best linear invariant estimates of the extreme-value distribution. **Technometrics**, v. 15, n. 1, p. 87-101, 1973.

MOALA, F. A.; GARCIA, L. M. A Bayesian analysis for the parameters of the exponential-logarithmic distribution. **Quality Engineering**, v. 25, n. 3, p. 282-291, 2013.

O'CONNOR, P. P.; KLEYNER, A. **Practical reliability engineering**. 5th ed. New York: Wiley & Sons, 2012.

TAHMASBI, R.; REZAEI, S. A two-parameter lifetime distribution with decreasing failure rate. **Computational Statistics and Data Analysis**, v. 52, n. 8, p. 3889-3901, 2008.