



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CENTRO DE CIÊNCIAS**  
**DEPARTAMENTO DE ESTATÍSTICA E MATEMÁTICA APLICADA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM E MÉTODOS**  
**QUANTITATIVOS**

**DANIEL JACINTO DOS SANTOS FILHO**

**DISTRIBUIÇÃO LOGÍSTICA EXPONENCIAL INVERSA LEHMANN**  
**TIPO II: MODELO DE REGRESSÃO E APLICAÇÃO A DADOS**  
**CENSURADOS À DIREITA**

**FORTALEZA**

**2023**

DANIEL JACINTO DOS SANTOS FILHO

DISTRIBUIÇÃO LOGÍSTICA EXPONENCIAL INVERSA LEHMANN TIPO II:  
MODELO DE REGRESSÃO E APLICAÇÃO A DADOS CENSURADOS À DIREITA

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem e Métodos Quantitativos, do Centro de Ciências, da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Modelagem e Métodos Quantitativos. Área de Concentração: Modelagem e Análise Quantitativa.

Orientador: Prof. Dr. Gualberto Segundo Agamez Montalvo.

FORTALEZA

2023

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Sistema de Bibliotecas  
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

S234d Santos Filho, Daniel Jacinto dos.

Distribuição logística exponencial inversa Lehmann tipo II: modelo de regressão e aplicação a dados censurados à direita / Daniel Jacinto dos Santos Filho. – 2023.  
38 f. : il.

Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Ciências, Programa de Pós-Graduação em Modelagem e Métodos Quantitativos, Fortaleza, 2023.

Orientação: Prof. Dr. Gualberto Segundo Agamez Montalvo.

1. Análise de sobrevivência. 2. Alternativa de Lehmann. 3. Dados censurados. 4. Logística exponencial inversa. 5. Modelo de regressão log-linear. I. Título.

CDD 510

---

DANIEL JACINTO DOS SANTOS FILHO

DISTRIBUIÇÃO LOGÍSTICA EXPONENCIAL INVERSA LEHMANN TIPO II:  
MODELO DE REGRESSÃO E APLICAÇÃO A DADOS CENSURADOS À DIREITA

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem e Métodos Quantitativos, do Centro de Ciências, da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Modelagem e Métodos Quantitativos. Área de Concentração: Modelagem e Análise Quantitativa.

Aprovada em: 29/09/2023

BANCA EXAMINADORA

---

Prof. Dr. Gualberto Segundo Agamez Montalvo (Orientador)  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Rafael Bráz Azevedo Farias  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Francisco Moisés Cândido de Medeiros  
Universidade Federal do Rio Grande do Norte (UFRN)

Aos meus pais, pela capacidade de acreditar  
e de investir em mim.

## AGRADECIMENTOS

À Deus pelo dom da vida, por guiar os meus passos e ter me sustentado durante todo esse período. À Ele seja dado toda honra e toda glória!

Aos meus pais por todo apoio, carinho, amor e torcida pelas minhas conquistas. Amo vocês, Daniel (Pai), Fabiana (Mãe) e Débora (Irmã).

A todos os professores do Programa de Pós Graduação em Modelagem e Métodos Quantitativos (PPGMMQ), pelo conhecimento que transmitem, pelo tempo que disponibilizam para com os alunos e pelos os incentivos. Em especial agradeço ao prof. Dr. Gualberto Segundo Agamez Montalvo que inicialmente me acolheu em um momento de dificuldade nessa trajetória de estudos/pesquisa do mestrado, sendo capaz de me incentivar e me apoiar nessa caminhada. Parabéns pelo excelente trabalho que realiza, professor Gualberto! Sua ajuda e seu apoio foram capazes de me fazer crescer e amadurecer quanto ser humano e quanto profissional.

Aos colegas de classe pelas discussões e conversas nas horas de estudos.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

De modo geral agradeço a todos que de forma direta e indireta contribuíram para a elaboração deste trabalho.

“O temor do Senhor é o princípio do conhecimento; os loucos desprezam a sabedoria e a instrução. Filho meu, ouve a instrução de teu pai, e não deixes o ensinamento de tua mãe, porque serão como diadema gracioso em tua cabeça, e colares ao teu pescoço” (BÍBLIA, 1969, Prov. 1, 7-9, p. 864).

## RESUMO

Neste trabalho, é introduzida uma nova distribuição contínua de probabilidade com suporte em  $\mathbb{R}^+$ , denominada de Logística Exponencial Inversa Lehmann tipo II, obtida a partir da alternativa de Lehmann tipo II. O principal propósito desse trabalho é aplicar a nova distribuição no contexto da análise de sobrevivência, e obter uma estrutura de regressão log-linear para os dados do tipo censurados à direita. Discutimos as principais propriedades da nova distribuição como quantis, assimetria, curtose, momentos, estatísticas de ordem e entropia de Rényi. A nova distribuição possui uma função taxa de risco que pode assumir diferentes comportamentos segundo os parâmetros associados. Realizou-se estudos de simulação para avaliar o desempenho das estimativas obtidas via o método da máxima verossimilhança. Para ilustrar a aplicabilidade da nova distribuição, foram utilizados conjunto de dados reais.

**Palavras-chave:** análise de sobrevivência; alternativa de Lehmann; dados censurados; logística exponencial inversa; modelo de regressão log-linear.

## ABSTRACT

In this academic work, a new continuous probability distribution with support in  $\mathbb{R}^+$  is introduced, called Inverse Exponential Logistic Lehmann type II, obtained from Lehmann type II alternative. The main purpose of this work is to apply the new distribution obtained in the context of survival analysis, and to obtain a log-linear regression structure for right-censored data. We discuss the main properties of the new distribution such as quantiles, asymmetry, kurtosis, moments, order statistics and Rényi entropy. The new distribution has a risk rate function that assumes different behaviors according to the associated parameters. Simulation studies were carried out to evaluate the performance of the estimates obtained via the maximum likelihood method. To illustrate the applicability of the new distribution, real data sets were used.

**Keywords:** survival analysis; Lehmann's alternative; censored data; inverse exponential logistics; log-linear regression model.

## LISTA DE FIGURAS

Figura 1 – Função densidade (à esquerda) e função taxa de risco (à direita), para a distribuição Logística Exponencial Inversa Lehmann tipo II, considerando alguns valores para os parâmetros . . . . .	14
Figura 2 – Comportamento da assimetria (a-b) e da curtose (c-d) para determinados valores dos parâmetros da distribuição Logística Exponencial Inversa Lehmann tipo II . . . . .	15
Figura 3 – Ajuste das distribuições e gráfico de quantis Kaplan-Meier versus Logística Exponencial Inversa de Lehmann tipo 2 . . . . .	21
Figura 4 – Ajuste das distribuições e gráfico de quantis Kaplan-Meier versus Logística Exponencial Inversa de Lehmann tipo 2 . . . . .	23
Figura 5 – Ajuste Kaplan-Meier e para distribuição LIEL2 para os dados tempos de falhas em relés . . . . .	29
Figura 6 – Resíduos deviance e gráfico de envelope Normal para os resíduos deviance do modelo LLIEL2 para os tempos de falhas em relés . . . . .	30
Figura 7 – Distância de Cook generalizada para o modelo LLIEL2 considerando os dados tempos de falhas em relés . . . . .	31

## LISTA DE TABELAS

Tabela 1 – Resultados do estudo de simulação para diferentes tamanhos amostrais e taxas de censuras considerando a distribuição LEIL2 com parâmetros $\alpha = 2, \gamma = 3, \lambda = 5$ . . . . .	20
Tabela 2 – Critérios AIC e BIC para as diferentes distribuições, considerando o conjunto de dados tempo de remissão em 929 indivíduos que apresentam câncer de cólon . . . . .	21
Tabela 3 – Estimativas e Erro padrão dos estimadores de máxima verossimilhança dos parâmetros da distribuição LEIL2, considerando o conjunto de dados tempo de remissão . . . . .	22
Tabela 4 – Critérios AIC e BIC para diferentes distribuições, considerando o conjunto de dados tempo de vida de pacientes com câncer de mama . . . . .	22
Tabela 5 – Estimativas e Erro padrão dos estimadores de máxima verossimilhança dos parâmetros da distribuição LEIL2, considerando o conjunto de dados tempo de vida de pacientes com câncer de mama . . . . .	23
Tabela 6 – Resultados do estudo de simulação para diferentes tamanhos amostrais e taxas de censuras considerando um modelo de regressão simples com $\beta_0 = 3, \beta_1 = 5, \sigma = 1$ e $\nu \sim LLEI2(\alpha, \gamma, \lambda)$ com parâmetros $\alpha = 2, \gamma = 1$ e $\lambda = 3$ . . . . .	28
Tabela 7 – Estimativas e Erro padrão dos estimadores de máxima verossimilhança dos parâmetros da distribuição LEIL2, considerando o conjunto de dados tempos de falhas em relés . . . . .	30
Tabela 8 – Critérios AIC e BIC para os diferentes modelos de regressão, considerando o conjunto de dados de tempos de falhas de relé . . . . .	30
Tabela 9 – Estimativas dos parâmetros do modelo LLIEL2 quando excluídas as observações 5, 23, e 24 . . . . .	31
Tabela 10 – Variação percentual das estimativas dos parâmetros do modelo LLIEL2 quando excluídas as observações 5, 23, e 24 . . . . .	31

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> . . . . .	11
<b>2</b>	<b>DISTRIBUIÇÃO LOGÍSTICA EXPONENCIAL INVERSA LEHMANN TIPO II</b> . . . . .	13
<b>2.1</b>	<b>Definição</b> . . . . .	13
<b>2.2</b>	<b>Quantis, assimetria e curtose</b> . . . . .	13
<b>2.3</b>	<b>Momentos</b> . . . . .	14
<b>2.4</b>	<b>Estatísticas de ordem</b> . . . . .	16
<b>2.5</b>	<b>Entropia de Rényi</b> . . . . .	16
<b>2.6</b>	<b>Estimação via máxima verossimilhança na presença de dados censurados à direita</b> . . . . .	17
<b>2.7</b>	<b>Estudo de Simulação</b> . . . . .	19
<b>2.8</b>	<b>Aplicação</b> . . . . .	20
<b>2.8.1</b>	<i>Tempo de remissão em indivíduos com câncer de cólon</i> . . . . .	20
<b>2.8.2</b>	<i>Tempo de vida de pacientes com câncer de mama</i> . . . . .	22
<b>3</b>	<b>MODELO DE REGRESSÃO.</b> . . . . .	24
<b>3.1</b>	<b>Modelo de Regressão Log-Logística Exponencial Inversa Leh- mann tipo II para dados censurados à direita</b> . . . . .	24
<b>3.2</b>	<b>Adequação do modelo</b> . . . . .	26
<b>3.3</b>	<b>Estudo de Simulação</b> . . . . .	27
<b>3.4</b>	<b>Aplicação</b> . . . . .	29
<b>3.4.1</b>	<i>Tempos de falhas em relés</i> . . . . .	29
<b>4</b>	<b>CONSIDERAÇÕES FINAIS</b> . . . . .	32
	<b>REFERÊNCIAS</b> . . . . .	33
	<b>APÊNDICE A– ELEMENTOS DA MATRIZ DE INFORMAÇÃO DE FISHER OBSERVADA DA DISTRIBUIÇÃO LOG- LOGÍSTICA EXPONENCIAL INVERSA LEHMANN TIPO II...</b>	36

## 1 INTRODUÇÃO

Nos estudos de análise de sobrevivência, a distribuição de probabilidade exponencial é um dos modelos contínuos mais utilizados, uma vez que possui fácil tratamento analítico e também por conta da importante propriedade da falta de memória. Contudo, tal distribuição possui a limitação de sua função taxa de risco ser constante, o que reduz sua utilidade em muitas outras situações práticas. Dado isso, diversas outras extensões da distribuição exponencial foram propostas a fim de obter outras formas da função taxa de risco.

É possível citar alguns artigos recentes que trabalham nessa linha como Mansoor *et al.* (2019) que propõem a distribuição Marshall-Olkin Exponencial Logística, Okorie *et al.* (2017) que propõem a distribuição exponencial Log-Logística Ajustada, George e Thobias (2019) realizam um estudo da distribuição Marshall-Olkin Exponencial Kumaraswamy, Nassar *et al.* (2019) introduzem a família de distribuição Marshall-Olkin Exponencial alfa potência, enquanto que Basheer (2019) aplica tal desenvolvimento para a distribuição Exponencial Inversa; Fallah e Kazemi (2020) realizam um estudo inferencial da distribuição exponencial ponderada generalizada, Chaudhary e Kumar (2020) trabalham com a distribuição Logística Exponencial Inversa, Sobhi e Mashail (2020) estudam a distribuição logística exponencial inversa potência no contexto de dados de seguros, Eghwerido *et al.* (2022) realizam uma extensão da distribuição exponencial generalizada alfa potência e, Ikechukwu e Eghwerido (2022) propõem a distribuição exponencial deslocada transmutada.

Neste trabalho será proposto a alternativa de Lehmann tipo II para distribuição Logística Exponencial Inversa (LEI) (Chaudhary; Kumar, 2020) a fim de obter uma distribuição mais flexível no contexto de análise de dados de sobrevivência na presença de dados censurados à direita. Seja  $X$  a variável aleatória Logística Exponencial Inversa,  $X \sim \text{LEI}(\gamma, \lambda)$ , cuja função de distribuição, densidade e taxa de risco são dadas, respectivamente, por

$$F_X(x; \gamma, \lambda) = \frac{1}{1 + (\exp(\lambda/x) - 1)^\gamma}, \quad x > 0 \text{ e } \gamma, \lambda > 0,$$

$$f_X(x; \gamma, \lambda) = \frac{\gamma \lambda \exp(\lambda/x) (\exp(\lambda/x) - 1)^{\gamma-1}}{x^2 [1 + (\exp(\lambda/x) - 1)^\gamma]^2}, \quad x > 0 \text{ e } \gamma, \lambda > 0,$$

$$h_X(x; \gamma, \lambda) = \frac{\gamma \lambda \exp(\lambda/x)}{x^2 [1 + (\exp(\lambda/x) - 1)^\gamma] [\exp(\lambda/x) - 1]} \quad x > 0 \text{ e } \gamma, \lambda > 0.$$

Para obter a alternativa de Lehmann tipo II, será considerado a transformação usando a função de distribuição acumulada de uma variável aleatória absolutamente contínua  $G(y) = 1 - [1 - F(y)]^\alpha$ ,  $\alpha > 0$ , em que  $F(\cdot)$  é a função de distribuição de uma variável aleatória absolutamente contínua. As funções densidade e taxa de risco associadas à  $G(y)$  são dadas respectivamente por

$$g(y; \alpha) = \alpha[1 - F(y)]^{\alpha-1} f(y) \text{ e } h(y; \alpha) = \frac{\alpha f(y)}{1 - F(y)}.$$

Para outros desenvolvimentos que aplicam as alternativas Lehmann, os leitores podem consultar trabalhos como: Chaubey e Zhang (2015) que propõem uma alternativa de Lehmann para a família de distribuição Chen, Tomazella *et al.* (2020) que fazem uma aplicação de uma alternativa de Lehmann para a distribuição Weibull inversa, Awodutire *et al.* (2020) em que desenvolvem um estudo para a distribuição half-Logística, Ogunde *et al.* (2020) que propõem uma alternativa de Lehmann para a distribuição Gumbel tipo-II estendida, enquanto em Ogunde *et al.* (2021) os autores fazem um desenvolvimento de uma alternativa de Lehmann para a distribuição Frèchet-Poisson.

O estudo do tempo de vida em análise de sobrevivência é bastante influenciado pela presença de covariáveis. Nesse contexto, os modelos de regressão são amplamente utilizados. Alguns modelos de regressão para dados censurados foram propostos na literatura recentemente. Tem-se, por exemplo, o desenvolvimento de Korkmaz *et al.* (2019) para a família de distribuição Marshall-Olkin Weibull, Cordeiro *et al.* (2019) propõem um modelo de regressão para uma classe da Weibull Marshall-Olkin, Yousof *et al.* (2019) realizam um estudo para a distribuição Topp Leone gerada da distribuição Burr XII e Afify *et al.* (2018) propõem um modelo de regressão para a distribuição Weibull Burr XII. Neste trabalho foi desenvolvido de forma introdutória o modelo de regressão log-Logística Exponencial Inversa Lehmann tipo II.

Com relação a estrutura desta dissertação, temos que no Capítulo 2 é introduzida a distribuição Logística Exponencial Inversa Lehmann tipo II, são apresentadas algumas propriedades e os procedimentos para obtenção de estimação via método de máxima verossimilhança na presença de dados censurados, e no Capítulo 3 é apresentado um modelo de regressão log-linear na presença de dados censurados à direita, e obtida as medidas para avaliar a adequação do modelo. Para as aplicações foram utilizados conjunto de dados reais em que a fim de comparar a distribuição proposta com outros modelos já desenvolvidos em outros trabalhos.

## 2 DISTRIBUIÇÃO LOGÍSTICA EXPONENCIAL INVERSA LEH-MANN TIPO II

Neste capítulo são abordadas algumas propriedades da distribuição Logística Exponencial Inversa Lehmann tipo II como quantis, assimetria, curtose e momentos. Além disso, são obtidas as expressões para cálculo da função escore e da matriz de informação de Fisher. Os resultados das aplicações podem ser conferidos na última seção do capítulo.

### 2.1 Definição

Seja  $T$  uma variável aleatória com distribuição Logística Exponencial Inversa Lehmann tipo II, denotada por  $T \sim \text{LEIL2}(\alpha, \gamma, \lambda)$ , sua função de distribuição acumulada é dada por

$$F_T(t; \alpha, \gamma, \lambda) = 1 - \left[ \frac{(\exp(\lambda/t) - 1)^\gamma}{1 + (\exp(\lambda/t) - 1)^\gamma} \right]^\alpha, \quad t > 0 \text{ e } \alpha, \gamma, \lambda > 0, \quad (2.1)$$

e sua função densidade é dada por

$$f_T(t; \alpha, \gamma, \lambda) = \frac{\alpha\gamma\lambda \exp(\lambda/t)(\exp(\lambda/t) - 1)^{\alpha\gamma-1}}{t^2 [1 + (\exp(\lambda/t) - 1)^\gamma]^{\alpha+1}} \quad t > 0 \text{ e } \alpha, \gamma, \lambda > 0. \quad (2.2)$$

Para  $\alpha = 1$  tem-se como caso particular a distribuição Logística Exponencial Inversa (Chaudhary; Kumar, 2020), e quando  $\alpha = \gamma = 1$  obtém-se a distribuição Exponencial Inversa como caso particular (Keller *et al.*, 1982).

A função taxa de risco é representada por

$$h_T(t; \alpha, \gamma, \lambda) = \frac{\alpha\gamma\lambda}{t^2} \frac{\exp(\lambda/t)}{[1 + (\exp(\lambda/t) - 1)^\gamma][\exp(\lambda/t) - 1]} \quad t > 0 \text{ e } \alpha, \gamma, \lambda > 0. \quad (2.3)$$

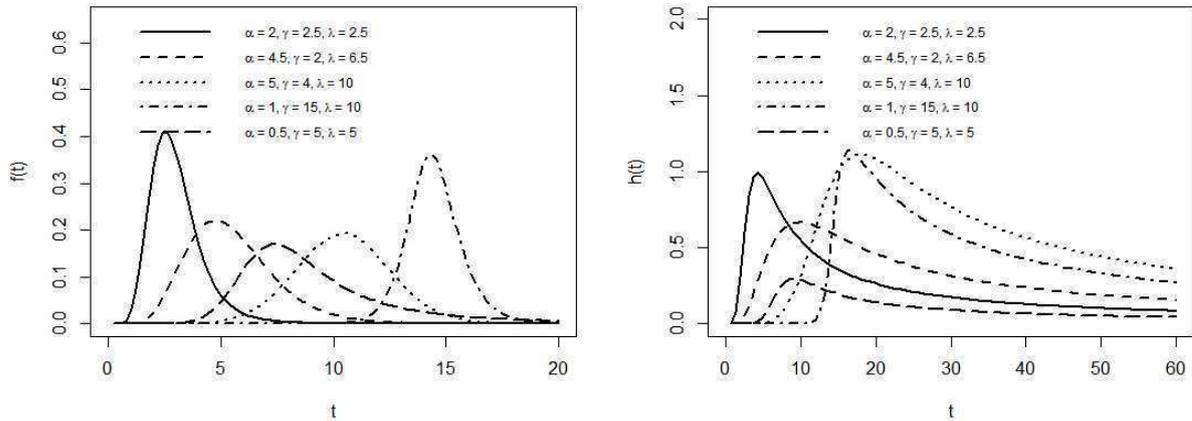
Na Figura 1 tem-se as funções densidades e de risco para a distribuição LEIL2. Note que a densidade assume diferentes formas (de assimetria e de curtose), o que é bastante útil nas aplicações práticas.

### 2.2 Quantis, assimetria e curtose

A função quantílica denotada por  $Q(u)$ ,  $u \in (0, 1)$ , da distribuição LEIL2 pode ser obtida através da inversa da função (2.1)

$$Q(u) = \frac{\lambda}{\log \left[ 1 + \left( \frac{1}{1-(1-u)^{1/\alpha}} - 1 \right)^{1/\gamma} \right]} \quad \alpha, \gamma, \lambda > 0,$$

Figura 1 – Função densidade (à esquerda) e função taxa de risco (à direita), para a distribuição Logística Exponencial Inversa Lehmann tipo II, considerando alguns valores para os parâmetros



Fonte: elaborado pelo autor.

caso haja interesse particular na mediana (Med), temos que

$$\text{Med} = \frac{\lambda}{\log \left[ 1 + \left( \frac{1}{1-2^{-1/\alpha}} - 1 \right)^{1/\gamma} \right]} \quad \alpha, \gamma, \lambda > 0.$$

Para o cálculo da assimetria podemos utilizar o coeficiente de assimetria quantílico de Bowley (Kenney; Keeping, 1962) dado por

$$S = \frac{Q(0, 25) - 2Q(0, 5) + Q(0, 75)}{Q(0, 75) - Q(0, 25)},$$

e para a curtose utilizar o coeficiente de curtose quantílico proposto por Moors (Moors, 1988)

$$K = \frac{Q(0, 875) - Q(0, 625) + Q(0, 375) - Q(0, 125)}{Q(0, 75) - Q(0, 25)}.$$

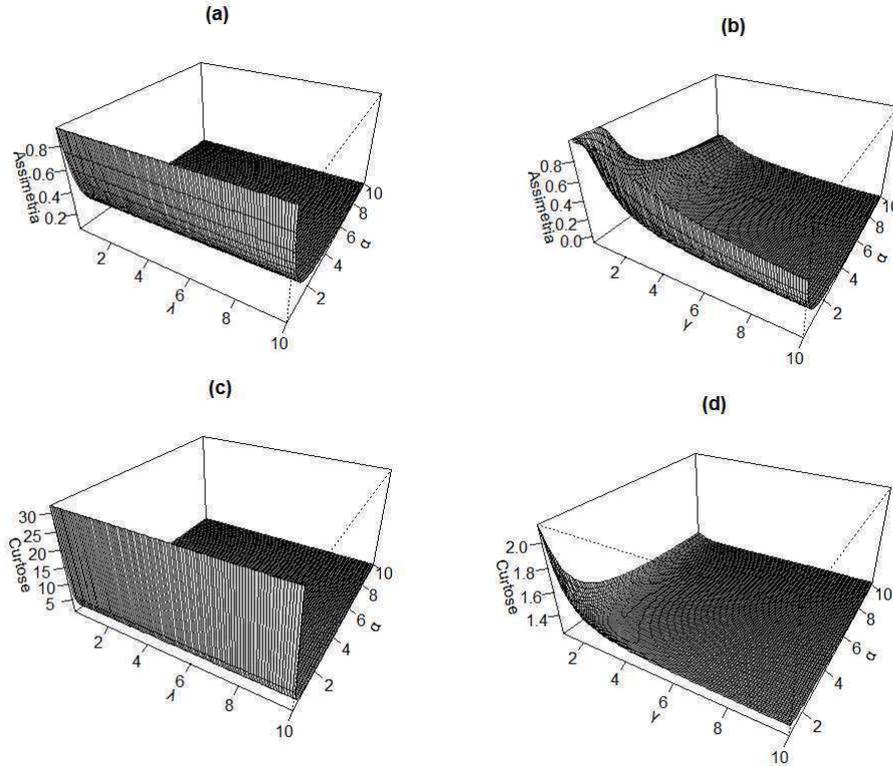
Na Figura 2 observa-se o comportamento da assimetria e da curtose considerando alguns valores dos parâmetros. Veja que para os menores valores do parâmetro  $\alpha$  tanto a assimetria como a curtose tendem a ter os maiores resultados.

### 2.3 Momentos

Seja  $T$  uma variável aleatória seguindo uma distribuição LEIL2( $\alpha, \gamma, \lambda$ ), seu  $r$ -ésimo momento é dado por

$$\begin{aligned} \mathbb{E}(T^r) &= \int_0^\infty t^r \frac{\alpha\gamma\lambda \exp(\lambda/t) (\exp(\lambda/t) - 1)^{\alpha\gamma-1}}{t^2 [1 + (\exp(\lambda/t) - 1)^\gamma]^{\alpha+1}} dt \\ &= \alpha\gamma\lambda \int_0^\infty t^{r-2} \frac{\exp(\lambda/t) (\exp(\lambda/t) - 1)^{\alpha\gamma-1}}{[1 + (\exp(\lambda/t) - 1)^\gamma]^{\alpha+1}} dt. \end{aligned}$$

Figura 2 – Comportamento da assimetria (a-b) e da curtose (c-d) para determinados valores dos parâmetros da distribuição Logística Exponencial Inversa Lehmann tipo II



Fonte: elaborado pelo autor.

Reescrevendo a fração a seguir em formato de uma série de potência binomial, resulta que

$$\frac{1}{[1 + (\exp(\lambda/t) - 1)^\gamma]^{\alpha+1}} = \sum_{k=0}^{\infty} \binom{k + \alpha}{k} (\exp(\lambda/t) - 1)^{\gamma k}.$$

Ao multiplicar pelo termo  $(\exp(\lambda/t) - 1)^{\alpha\gamma-1}$  em ambos os lados tem-se

$$\begin{aligned} \frac{(\exp(\lambda/t) - 1)^{\alpha\gamma-1}}{[1 + (\exp(\lambda/t) - 1)^\gamma]^{\alpha+1}} &= \sum_{k=0}^{\infty} \binom{k + \alpha}{k} (\exp(\lambda/t) - 1)^{\gamma(k+\alpha)-1} \\ &= \sum_{k=0}^{\infty} \binom{k + \alpha}{k} (-1)^{\gamma(k+\alpha)-1} (1 - \exp(\lambda/t))^{\gamma(k+\alpha)-1}. \end{aligned}$$

Representando  $(1 - \exp(\lambda/t))^{\gamma(k+\alpha)-1}$  em série binomial,  $\sum_{j=0}^{\infty} (-1)^j \binom{\gamma(k+\alpha)-1}{j} \exp(j\lambda/t)$ , e multiplicando pelo termo  $\exp(\lambda/t)$ , resulta

$$\frac{\exp(\lambda/t)(\exp(\lambda/t) - 1)^{\alpha\gamma-1}}{[1 + (\exp(\lambda/t) - 1)^\gamma]^{\alpha+1}} = \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \binom{k + \alpha}{k} \binom{\gamma(k + \alpha) - 1}{j} (-1)^{\gamma(k+\alpha)+j-1} \exp((j+1)\lambda/t).$$

O  $r$ -ésimo momento então é dado por

$$\begin{aligned}\mathbb{E}(T^r) &= \alpha\gamma\lambda \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \binom{k+\alpha}{k} \binom{\gamma(k+\alpha)-1}{j} (-1)^{\gamma(k+\alpha)+j-1} \int_0^{\infty} t^{r-2} \exp((j+1)\lambda/t) dt \\ &= \alpha\gamma\lambda \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \binom{k+\alpha}{k} \binom{\gamma(k+\alpha)-1}{j} (-1)^{\gamma(k+\alpha)+j-1} \frac{\Gamma(1-r)}{[-(j+1)\lambda]^{1-r}}.\end{aligned}$$

## 2.4 Estatísticas de ordem

Seja  $T_1, T_2, \dots, T_n$  uma amostra aleatória de  $T \sim \text{LEIL2}(\alpha, \gamma, \lambda)$ ,  $\alpha, \gamma, \lambda > 0$ . Considere suas respectivas estatísticas de ordem  $T_{(1)}, T_{(2)}, \dots, T_{(n)}$ . A função densidade da  $l$ -ésima ( $0 < l \leq n$ ) estatística de ordem é dada por

$$\begin{aligned}g_{T_{(l)}}(t) &= \frac{n!}{(l-1)!(n-l)!} g(t) [G(t)]^{l-1} [1-G(t)]^{n-l} \\ &= \frac{n!}{(l-1)!(n-l)!} \frac{\alpha\gamma\lambda \exp(\lambda/t) (\exp(\lambda/t) - 1)^{\alpha\gamma[1+(n-l)]-1}}{t^2 [1 + (\exp(\lambda/t) - 1)\gamma]^{(\alpha+1)[1+(n-l)]}} \\ &\quad \times \left[ 1 - \frac{(\exp(\lambda/t) - 1)^{\alpha\gamma}}{[1 + (\exp(\lambda/t) - 1)\gamma]^{\alpha}} \right]^{l-1}.\end{aligned}$$

Utilizando a expansão binomial, é possível reescrever a função densidade da  $l$ -ésima estatística de ordem como

$$\begin{aligned}g_{T_{(l)}}(t) &= \frac{n!}{(l-1)!(n-l)!} \frac{\alpha\gamma\lambda \exp(\lambda/t) (\exp(\lambda/t) - 1)^{\alpha\gamma[1+(n-l)]-1}}{t^2 [1 + (\exp(\lambda/t) - 1)\gamma]^{(\alpha+1)[1+(n-l)]}} \\ &\quad \times \sum_{k=0}^{l-1} (-1)^k \binom{l-1}{k} \frac{(\exp(\lambda/t) - 1)^{\alpha\gamma k}}{[1 + (\exp(\lambda/t) - 1)\gamma]^{\alpha k}}.\end{aligned}$$

Através de uma organização algébrica, a função densidade da  $l$ -ésima estatística fica

$$g_{T_{(l)}}(t) = \frac{n!}{(l-1)!(n-l)!} \sum_{k=0}^{l-1} \frac{(-1)^k}{[1 + (\exp(\lambda/t) - 1)\gamma]^{n-l}} \binom{l-1}{k} g(t; \alpha^*, \gamma, \lambda), \quad (2.4)$$

em que  $g(t; \alpha^*, \gamma, \lambda)$  é a função densidade de  $T \sim \text{LEIL2}(\alpha^*, \gamma, \lambda)$ , sendo  $\alpha^* = 1 + k + n - l$ .

## 2.5 Entropia de Rényi

Seja  $X$  variável aleatória com função densidade  $g(x)$ , define-se a entropia de Rényi (RÉNYI, 1961) de ordem  $\rho$  como

$$H_R = \frac{1}{1-\rho} \log \left[ \int_{-\infty}^{\infty} [g(x)]^{\rho} dx \right], \quad \rho > 0, \quad \rho \neq 1.$$

Para o caso de  $\rho \uparrow 1$ , tem-se como o caso particular a entropia de Shannon (Shannon, 1948). Considerando a distribuição  $T \sim \text{LEIL2}(\alpha, \gamma, \lambda)$  a entropia de Rényi é dada por

$$\begin{aligned} H_R &= \frac{1}{1-\rho} \log \left[ \int_0^\infty \frac{(\alpha\gamma\lambda)^\rho \exp(\rho\lambda/t) (\exp(\lambda/t) - 1)^{\alpha\gamma\rho-\rho}}{t^{2\rho} [1 + (\exp(\lambda/t) - 1)^\gamma]^{\rho(\alpha+1)}} dt \right] \\ &= \frac{1}{1-\rho} \log \left[ (\alpha\gamma\lambda)^\rho \int_0^\infty (1/t)^{2\rho} \frac{\exp(\rho\lambda/t) (\exp(\lambda/t) - 1)^{\alpha\gamma\rho-\rho}}{[1 + (\exp(\lambda/t) - 1)^\gamma]^{\rho(\alpha+1)}} dt \right]. \end{aligned} \quad (2.5)$$

Reescrevendo a fração a seguir em formato de uma série de potência binomial, resulta que

$$\frac{1}{[1 + (\exp(\lambda/t) - 1)^\gamma]^{\rho(\alpha+1)}} = \sum_{k=0}^{\infty} \binom{k + \rho(\alpha + 1) - 1}{k} (\exp(\lambda/t) - 1)^{\gamma k}$$

Multiplicando pelo termo  $(\exp(\lambda/t) - 1)^{\rho(\alpha\gamma-1)}$  a série fica

$$\begin{aligned} \frac{(\exp(\lambda/t) - 1)^{\rho(\alpha\gamma-1)}}{[1 + (\exp(\lambda/t) - 1)^\gamma]^{\rho(\alpha+1)}} &= \sum_{k=0}^{\infty} \binom{k + \rho(\alpha + 1) - 1}{k} (\exp(\lambda/t) - 1)^{\gamma(k+\alpha\rho)-\rho} \\ &= \sum_{k=0}^{\infty} \binom{k + \rho(\alpha + 1) - 1}{k} (-1)^{\gamma(k+\alpha\rho)-\rho} (1 - \exp(\lambda/t))^{\gamma(k+\alpha\rho)-\rho} \end{aligned}$$

Representando  $(1 - \exp(\lambda/t))^{\gamma(k+\alpha\rho)-\rho}$  em série binomial,  $\sum_{j=0}^{\infty} (-1)^j \binom{\gamma(k+\alpha\rho)-\rho}{j} \exp(j\lambda/t)$ , e multiplicando pelo termo  $\exp(\lambda/t)$  obtém-se

$$\sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \binom{k + \rho(\alpha + 1) - 1}{k} \binom{\gamma(k + \alpha\rho) - \rho}{j} (-1)^{\gamma(k+\alpha\rho)+j-\rho} \exp((j+1)\lambda/t)$$

Assim, a entropia de Rényi é dada por

$$\begin{aligned} H_R &= \frac{1}{1-\rho} \log \left[ (\alpha\gamma\lambda)^\rho \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \binom{k + \rho(\alpha + 1) - 1}{k} \binom{\gamma(k + \alpha\rho) - \rho}{j} (-1)^{\gamma(k+\alpha\rho)+j-\rho} \right. \\ &\quad \times \left. \int_0^\infty (1/t)^{2\rho} \exp((j+1)\lambda/t) dt \right] \\ &= \frac{1}{1-\rho} \log \left[ (\alpha\gamma\lambda)^\rho \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \binom{k + \rho(\alpha + 1) - 1}{k} \binom{\gamma(k + \alpha\rho) - \rho}{j} (-1)^{\gamma(k+\alpha\rho)+j-\rho} \right. \\ &\quad \times \left. \frac{\Gamma(2\rho - 1)}{[-(j+1)\lambda]^{2\rho-1}} \right]. \end{aligned} \quad (2.6)$$

## 2.6 Estimação via máxima verossimilhança na presença de dados censurados à direita

Seja  $Y_i$  uma variável aleatória que representa o tempo de falha da  $i$ -ésima unidade, e  $C_i$  a variável aleatória que representa o tempo de censura da respectiva

unidade. As variáveis  $Y_i$  e  $C_i$  são assumidas independentes,  $Y_i$  possui distribuição LEIL2 parametrizada por  $\boldsymbol{\theta} = (\alpha, \gamma, \lambda)^\top$ . Observando os pares de dados  $(T_i, \delta_i)$ , em que  $T_i = \min(Y_i, C_i)$  com  $\delta_i = 1$ , quando  $Y_i \leq C_i$  e  $\delta_i = 0$  caso contrário.

Dado uma amostra de tamanho  $n$  dos pares  $(T_i, \delta_i)$ , o logaritmo da função de verossimilhança é dada por

$$\begin{aligned} \ell(\boldsymbol{\theta}) &= \sum_{i=1}^n \delta_i \log[f_T(t_i)] + \sum_{i=1}^n (1 - \delta_i) \log[S_T(t_i)] \\ &= \log(\alpha\gamma\lambda) \sum_{i=1}^n \delta_i - 2 \sum_{i=1}^n \delta_i \log(t_i) + \lambda \sum_{i=1}^n \frac{\delta_i}{t_i} + \sum_{i=1}^n (\alpha\gamma - \delta_i) \log(\exp(\lambda/t_i) - 1) \\ &\quad - \sum_{i=1}^n (\alpha + \delta_i) \log(1 + (\exp(\lambda/t_i) - 1)^\gamma), \end{aligned} \quad (2.7)$$

em que  $f_T(\cdot)$  é dado por (2.2) e  $S_T(\cdot) = 1 - F_T(\cdot)$  com  $F_T(\cdot)$  dado por (2.1).

A função escore para os parâmetros fica

$$U_\alpha(\boldsymbol{\theta}) = \frac{1}{\alpha} \sum_{i=1}^n \delta_i + \gamma \sum_{i=1}^n \log(\eta_i) - \sum_{i=1}^n \log(1 + \eta_i^\gamma), \quad (2.8)$$

$$U_\gamma(\boldsymbol{\theta}) = \frac{1}{\gamma} \sum_{i=1}^n \delta_i + \alpha \sum_{i=1}^n \log(\eta_i) - \sum_{i=1}^n (\alpha + \delta_i) \frac{\log(\eta_i) \eta_i^\gamma}{1 + \eta_i^\gamma}, \quad (2.9)$$

$$U_\lambda(\boldsymbol{\theta}) = \frac{1}{\lambda} \sum_{i=1}^n \delta_i + \sum_{i=1}^n \frac{\delta_i}{t_i} + \sum_{i=1}^n \frac{(\alpha\gamma - \delta_i)(1 + \eta_i)}{t_i \eta_i} - \gamma \sum_{i=1}^n \frac{(\alpha + \delta_i) \eta_i^{\gamma-1} (1 + \eta_i)}{t_i (1 + \eta_i^\gamma)}, \quad (2.10)$$

em que  $\eta_i = \exp(\lambda/t_i) - 1$ . Os estimadores de máxima verossimilhança são obtidos através da solução do sistema de equações  $U_\alpha(\boldsymbol{\theta}) = 0$ ,  $U_\gamma(\boldsymbol{\theta}) = 0$  e  $U_\lambda(\boldsymbol{\theta}) = 0$ . Como esse sistema não possui solução analítica fechada, utilizou-se rotinas computacionais em R (R core team, 2023) através da utilização da função *optim*, que por sua vez utiliza o método de otimização de Nelder e Mead (1965) para encontrar uma solução numérica.

Assintoticamente, e sob condição dos parâmetros serem pontos interiores do espaço paramétrico tem-se que  $\hat{\boldsymbol{\theta}}$  possui distribuição Normal trivariada dada por

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim \mathcal{N}_3(\mathbf{0}, \mathbf{I}^{-1}(\boldsymbol{\theta})) \text{ quando } n \rightarrow \infty,$$

em que  $\mathbf{I}(\boldsymbol{\theta})$  é matriz de informação de Fisher esperada. Tal comportamento ainda pode ser esperado se a matriz de informação  $\mathbf{I}(\boldsymbol{\theta})$  for substituída pela matriz de informação

observada avaliada em  $\hat{\boldsymbol{\theta}}$ ,  $\mathbf{J}(\hat{\boldsymbol{\theta}})$ , em que  $\mathbf{J}(\hat{\boldsymbol{\theta}}) = -\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$  é representada por

$$\mathbf{J}(\hat{\boldsymbol{\theta}}) = \begin{pmatrix} \frac{d^2 \ell(\boldsymbol{\theta})}{d\alpha^2} & \frac{d^2 \ell(\boldsymbol{\theta})}{d\alpha d\gamma} & \frac{d^2 \ell(\boldsymbol{\theta})}{d\alpha d\lambda} \\ \cdot & \frac{d^2 \ell(\boldsymbol{\theta})}{d\gamma^2} & \frac{d^2 \ell(\boldsymbol{\theta})}{d\gamma d\lambda} \\ \cdot & \cdot & \frac{d^2 \ell(\boldsymbol{\theta})}{d\lambda^2} \end{pmatrix},$$

sendo

$$\frac{d^2 \ell(\boldsymbol{\theta})}{d\alpha^2} = -\frac{1}{\alpha^2} \sum_{i=1}^n \delta_i,$$

$$\frac{d^2 \ell(\boldsymbol{\theta})}{d\alpha d\gamma} = \sum_{i=1}^n \log(\eta_i) - \sum_{i=1}^n \frac{\log(\eta_i) \eta_i^\gamma}{1 + \eta_i^\gamma},$$

$$\frac{d^2 \ell(\boldsymbol{\theta})}{d\alpha d\lambda} = \sum_{i=1}^n \frac{\gamma(1 + \eta_i)}{t_i \eta_i} - \sum_{i=1}^n \frac{\gamma \eta_i^{\gamma-1} (1 + \eta_i)}{t_i (1 + \eta_i^\gamma)},$$

$$\frac{d^2 \ell(\boldsymbol{\theta})}{d\gamma^2} = -\frac{1}{\gamma^2} \sum_{i=1}^n \delta_i - \sum_{i=1}^n \frac{(\alpha + \delta_i) \log^2(\eta_i) \eta_i^\gamma}{(1 + \eta_i^\gamma)^2},$$

$$\frac{d^2 \ell(\boldsymbol{\theta})}{d\gamma d\lambda} = \sum_{i=1}^n \frac{\alpha(1 + \eta_i)}{t_i \eta_i} - \sum_{i=1}^n \frac{(\alpha + \delta_i) \eta_i^{\gamma-1} (1 + \eta_i) (\gamma \log(\eta_i) + \eta_i^\gamma + 1)}{t_i (1 + \eta_i^\gamma)^2},$$

$$\frac{d^2 \ell(\boldsymbol{\theta})}{d\lambda^2} = -\frac{1}{\lambda^2} \sum_{i=1}^n \delta_i - \sum_{i=1}^n \frac{(\alpha \gamma - \delta_i) (1 + \eta_i)}{t_i^2 \eta_i^2} - \sum_{i=1}^n \frac{(\alpha + \delta_i) \gamma \eta_i^{\gamma-2} (1 + \eta_i) (\gamma (1 + \eta_i) - \eta_i^\gamma - 1)}{t_i^2 (1 + \eta_i^\gamma)^2},$$

com  $\eta_i = \exp(\lambda/t_i) - 1$ .

## 2.7 Estudo de Simulação

Com o propósito de avaliar as estimativas de máxima verossimilhança para grandes e pequenos tamanhos amostrais, realizou-se um estudo de simulação Monte Carlo. Para obter as amostras da distribuição LEIL2 utilizou-se o método da transformação inversa a partir de

$$t = \lambda \left\{ \log \left[ \left( \frac{1}{1 - (1 - \mathcal{U})^{1/\alpha}} - 1 \right)^{1/\gamma} + 1 \right] \right\}^{-1},$$

em que  $\mathcal{U}$  representa uma variável aleatória uniforme padrão. Assim, tomou-se  $N = 5000$  réplicas de Monte Carlo da distribuição LEIL2( $\alpha, \gamma, \lambda$ ) para os tamanhos amostrais  $n = 50, 100, 200, 500, 1000, 2000$ , além disso avaliou-se para as seguintes taxas de censuras 5%, 10% e 20%. A obtenção das variáveis de censura,  $\delta_i$ , se deu a partir da variável aleatória de Bernoulli( $p$ ), em que  $p$  são as taxas de censuras. Para cada tamanho amostral e taxa de censura foram observados o viés e o erro quadrático médio. Foram fixados os

seguintes valores para os parâmetros  $\alpha = 2, \gamma = 3, \lambda = 5$  e utilizou-se 513776 como semente da geração de dados. As estimativas de máxima verossimilhança foram computadas a partir de dos recursos computacionais em R (R core team, 2023) através da função *optim*. Na Tabela 1 são apresentados os resultados obtidos. É possível observar que os valores do

Tabela 1 – Resultados do estudo de simulação para diferentes tamanhos amostrais e taxas de censuras considerando a distribuição LEIL2 com parâmetros  $\alpha = 2, \gamma = 3, \lambda = 5$

Censura	$n$	Viés			Erro Padrão		
		$\hat{\alpha}$	$\hat{\gamma}$	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\gamma}$	$\hat{\lambda}$
5%	50	4,834	-0,056	2,558	10,937	1,289	5,522
	100	2,694	-0,086	1,758	6,705	0,966	4,355
	200	1,561	-0,082	1,050	4,868	0,711	3,284
	500	0,496	-0,044	0,426	2,572	0,453	1,830
	1000	0,108	-0,025	0,134	1,183	0,258	0,796
	2000	-0,064	-0,003	0,029	0,306	0,164	0,274
10%	50	4,495	-0,039	2,766	9,792	1,507	5,641
	100	2,522	-0,121	1,778	6,515	0,937	4,447
	200	1,473	-0,115	1,107	4,699	0,702	3,248
	500	0,370	-0,047	0,410	2,420	0,432	1,829
	1000	-0,070	-0,008	0,091	0,680	0,251	0,562
	2000	-0,129	-0,011	0,047	0,329	0,167	0,288
20%	50	3,767	-0,004	2,664	9,415	1,484	5,661
	100	2,534	-0,130	1,962	6,437	0,986	4,642
	200	1,203	-0,114	1,262	4,246	0,762	3,602
	500	0,170	-0,049	0,467	2,495	0,465	2,109
	1000	-0,287	-0,009	0,150	1,222	0,297	1,059
	2000	-0,337	-0,001	0,046	0,590	0,192	0,451

Fonte: elaborado pelo autor.

viés e a medida do erro padrão calculado para as estimativas dos parâmetros diminuíram com o aumento dos tamanhos amostrais para cada taxa de censura.

## 2.8 Aplicação

### 2.8.1 Tempo de remissão em indivíduos com câncer de cólon

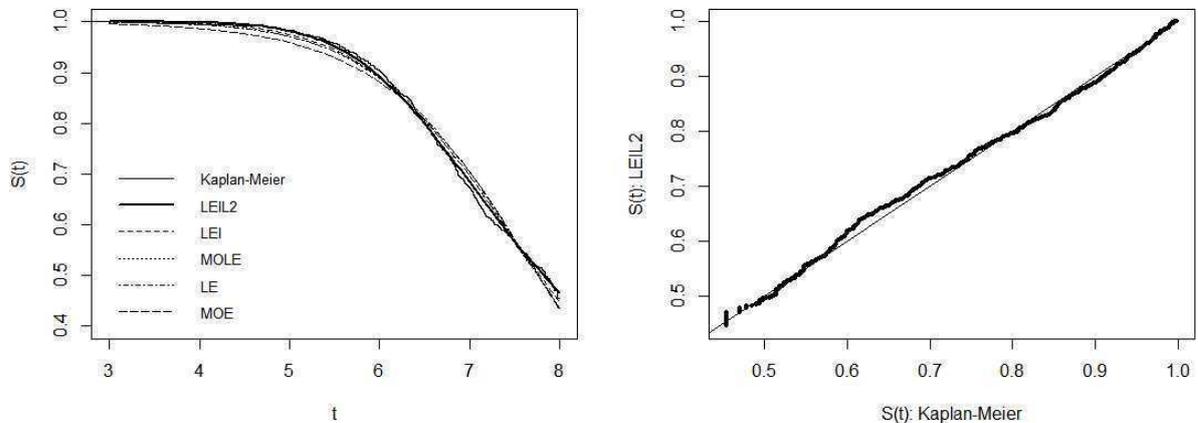
Nesta seção apresentamos os resultados do ajuste do modelo LEIL2 ao conjunto de dados tempo de remissão em 929 indivíduos que apresentam câncer de cólon. Tal conjunto de dados pode ser encontrado na biblioteca *survival* (Therneau, 2023).

A partir dos resultados do ajuste do modelo LEIL2, comparou-se com outros desenvolvimentos encontrados na literatura como, Logística Exponencial Inversa (LEI) (Chaudhary; Kumar, 2020) e Marshall-Olkin Exponencial Logística (MOLE) (Mansoor *et al.*, 2019) e seus casos particulares, Marshal-Olkin Exponencial (MOE) e Logística Exponencial (LE).

Para comparar os diferentes modelos, foram utilizados algumas medidas de critério de informação como o critério de informação de Akaike dado por  $AIC = -2\ell(\hat{\theta}) + 2p$  e o critério de informação Bayesiano dado por  $BIC = -2\ell(\hat{\theta}) + p \times \log(n)$  em que  $p$  e  $n$  são, respectivamente, o número de parâmetros e o tamanho amostral.

Na Figura 3 é apresentado o ajuste para o conjunto de dados considerando a distribuição LEIL2 e outras distribuições, além do gráfico de quantis de Kaplan-Meier versus LEIL2. Com base nisso, pode-se observar que a distribuição LEIL2 apresenta um ajuste mais adequado frente as outras distribuições candidatas. Tal indicativo também é observado na Tabela 2, nota-se que a distribuição proposta apresenta os menores valores para os critérios de informações adotados. Contudo, observando os resultados do critério de informação BIC, as distribuições LEIL2 e LEI apresentam valores muito próximos. Realizando uma análise comparativa dessas distribuições através do teste da razão de verossimilhanças para as hipóteses  $\mathcal{H}_0 : \alpha = 1$  versus  $\mathcal{H}_1 : \alpha \neq 1$ , temos que modelo LEIL2 é significativo (p-valor: 0,003).

Figura 3 – Ajuste das distribuições e gráfico de quantis Kaplan-Meier versus Logística Exponencial Inversa de Lehmann tipo 2



Fonte: elaborado pelo autor.

Tabela 2 – Critérios AIC e BIC para as diferentes distribuições, considerando o conjunto de dados tempo de remissão em 929 indivíduos que apresentam câncer de cólon

Distribuição	AIC	BIC
<b>LEIL2</b>	<b>2253,598</b>	<b>2268,100</b>
MOLE	2267,523	2282,025
LE	2268,521	2278,190
MOE	2289,171	2298,839
LEI	2260,016	2269,684

Fonte: elaborado pelo autor.

Na Tabela 3 são apresentados as estimativas e erro padrão para os os parâmetros da distribuição LEIL2.

Tabela 3 – Estimativas e Erro padrão dos estimadores de máxima verossimilhança dos parâmetros da distribuição LEIL2, considerando o conjunto de dados tempo de remissão

Parâmetros	Estimativa	Erro Padrão
$\alpha$	0,388	0,007
$\gamma$	7,319	0,466
$\lambda$	4,614	0,024

Fonte: elaborado pelo autor.

### 2.8.2 Tempo de vida de pacientes com câncer de mama

Nesta segunda aplicação, a distribuição LEIL2 foi ajustada ao conjunto de dados de tempos de vida de 272 indivíduos que apresentam câncer de mama. A descrição e o conjunto de dados completo podem ser encontrados em Ramanan (2016).

A distribuição LEIL2 foi comparada com as mesmas distribuições consideradas anteriormente, Logística Exponencial Inversa (LEI) (Chaudhary; Kumar, 2020) e Marshall-Olkin Exponencial Logística (MOLE) (Mansoor *et al.*, 2019) e seus casos particulares, Marshal-Olkin Exponencial (MOE) e Logística Exponencial (LE). Foram utilizados o critério de informação de Akaike e o critério de informação Bayesiano. Como pode ser observado na Tabela 4, existem indicativos a favor da distribuição LEIL2, pois esta apresenta os menores valores para os critérios adotados. Uma vez que o critério de informação BIC novamente apresenta valores próximos para as distribuições LEIL2 e LEI, realizou-se um teste da razão de verossimilhanças para as hipóteses  $\mathcal{H}_0 : \alpha = 1$  versus  $\mathcal{H}_1 : \alpha \neq 1$ , temos que modelo LEIL2 é significativo (p-valor: 0,005).

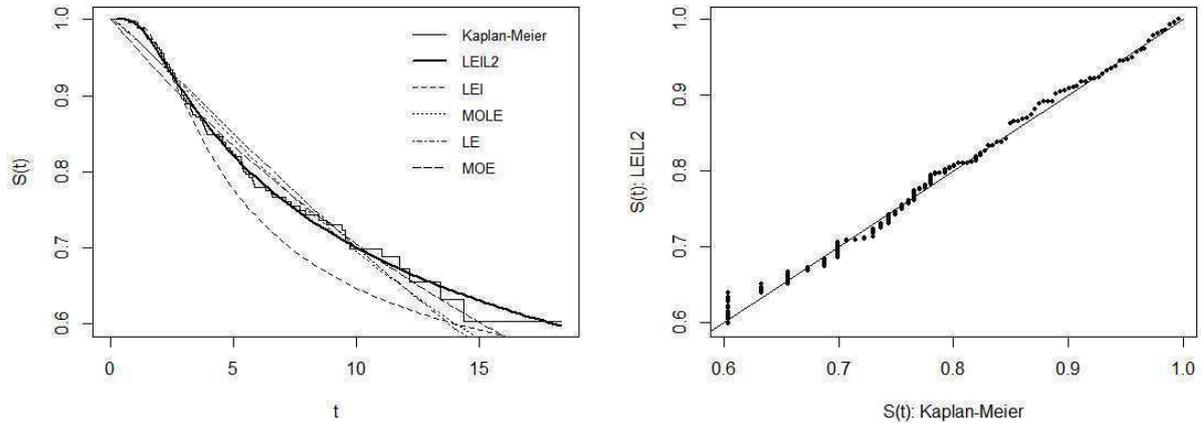
Tabela 4 – Critérios AIC e BIC para diferentes distribuições, considerando o conjunto de dados tempo de vida de pacientes com câncer de mama

Distribuição	AIC	BIC
<b>LEIL2</b>	<b>655,98</b>	<b>666,79</b>
MOLE	671,04	681,86
LE	672,79	680,01
MOE	673,90	681,11
LEI	661,54	668,75

Fonte: elaborado pelo autor.

Na Figura 4 tem-se os ajustes para as distribuições consideradas para a análise, percebe-se como a distribuição LEIL2 se ajusta bem aos dados. Na Tabela 5 são apresentadas as estimativas dos parâmetros e o erro padrão dos parâmetros da distribuição LEIL2.

Figura 4 – Ajuste das distribuições e gráfico de quantis Kaplan-Meier versus Logística Exponencial Inversa de Lehmann tipo 2



Fonte: elaborado pelo autor.

Tabela 5 – Estimativas e Erro padrão dos estimadores de máxima verossimilhança dos parâmetros da distribuição LEIL2, considerando o conjunto de dados empo de vida de pacientes com câncer de mama

Parâmetros	Estimativa	Erro Padrão
$\alpha$	0,452	0,085
$\gamma$	0,764	0,196
$\lambda$	5,821	22,923

Fonte: elaborado pelo autor.

### 3 MODELO DE REGRESSÃO

Neste capítulo é tratado o modelo de regressão log-linear para dados censurados à direita, sendo obtida as expressões algébricas para a função escore e matriz de informação de Fisher. Um estudo de aplicação e adequação do modelo pode ser visto na última seção do capítulo.

#### 3.1 Modelo de Regressão Log-Logística Exponencial Inversa Lehmann tipo II para dados censurados à direita

Seja  $V$  uma variável aleatória com distribuição LEIL2 cuja densidade é dada em (2.2). Define-se uma nova variável aleatória dada por  $Y = \log(V)$  que denominaremos de Log-Logística Exponencial Inversa de Lehmann tipo 2 (LLEI2) cuja função densidade é dada por

$$f_Y(y; \alpha, \gamma, \lambda) = \frac{\alpha\gamma\lambda}{[\exp(y)]^2} \frac{\exp(y + \lambda/\exp(y))(\exp(\lambda/\exp(y)) - 1)^{\alpha\gamma-1}}{[1 + (\exp(\lambda/\exp(y)) - 1)^\gamma]^{\alpha+1}} \quad y > 0 \text{ e } \alpha, \gamma, \lambda > 0. \quad (3.1)$$

Conforme à estrutura apresentada em Korkmaz *et al.* (2019), Cordeiro *et al.* (2019), Yousof *et al.* (2019) e Afify *et al.* (2018), considere um modelo linear simples em que  $y_i$  representa a variável resposta do  $i$ -ésimo indivíduo, e  $x_{ij}$  a  $j$ -ésima variável explicativa do  $i$ -ésimo indivíduo, e associado a ela o  $j$ -ésimo parâmetro da regressão  $\beta_j$  com,

$$y_i = \sum_{j=0}^k x_{ij}\beta_j + \sigma\nu,$$

o parâmetro  $\sigma > 0$  é classificado como parâmetro de escala, e  $\nu$  é uma variável aleatória com densidade dada por (3.1) e  $x_{i0} = 1$  para  $i = 1, \dots, n$ . A função densidade e de sobrevivência de  $Y_i$  são dadas, respectivamente, por

$$f_{Y_i}(y_i; \alpha, \gamma, \lambda, \boldsymbol{\beta}, \sigma) = \frac{\alpha\gamma\lambda}{\sigma[\exp(z_i)]^2} \frac{\exp(z_i + \lambda/\exp(z_i))(\exp(\lambda/\exp(z_i)) - 1)^{\alpha\gamma-1}}{[1 + (\exp(\lambda/\exp(z_i)) - 1)^\gamma]^{\alpha+1}}, \quad \alpha, \gamma, \lambda, \sigma > 0 \quad (3.2)$$

e

$$S_{Y_i}(y_i; \alpha, \gamma, \lambda, \boldsymbol{\beta}, \sigma) = \left[ \frac{(\exp(\lambda/\exp(z_i)) - 1)^\gamma}{1 + (\exp(\lambda/\exp(z_i)) - 1)^\gamma} \right]^\alpha \quad \alpha, \gamma, \lambda, \sigma > 0, \quad (3.3)$$

com  $z_i = \frac{y_i - \sum_{j=0}^k x_{ij}\beta_j}{\sigma}$ , e  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)^\top$  com  $\beta_j \in \mathbb{R}$ ,  $j = 1, 2, \dots, k$ . Tomando uma amostra de tamanho  $n$  dos vetores  $(T_i, \delta_i)$  em que  $T_i = \min\{Y_i, C_i\}$ , com  $\delta_i = 1$ , quando

$Y_i \leq C_i$  e  $\delta_i = 0$  caso contrário, sendo  $C_i$  a variável aleatória que representa o tempo de censura, independente de  $Y_i$ , e seja o vetor de parâmetros  $\boldsymbol{\theta} = (\alpha, \gamma, \lambda, \boldsymbol{\beta}, \sigma)^\top$ , temos que o logaritmo da função de verossimilhança de  $\boldsymbol{\theta}$  é dado por

$$\begin{aligned} \ell(\boldsymbol{\theta}) &= \sum_{i=1}^n \delta_i \log[f_T(t_i)] + \sum_{i=1}^n (1 - \delta_i) \log[S_T(t_i)] \\ &= \log\left(\frac{\alpha\gamma\lambda}{\sigma}\right) \sum_{i=1}^n \delta_i - \sum_{i=1}^n \delta_i z_i + \sum_{i=1}^n \delta_i \frac{\lambda}{\exp(z_i)} + \sum_{i=1}^n (\alpha\gamma - \delta_i) \times \\ &\quad \times \log(\exp(\lambda/\exp(z_i)) - 1) - \sum_{i=1}^n (\alpha + \delta_i) \log(1 + (\exp(\lambda/\exp(z_i)) - 1)^\gamma), \end{aligned}$$

com  $f_T(t_i)$  e  $S_T(t_i)$  dados respectivamente por (3.2) e (3.3).

A função escore para os parâmetros fica

$$U_\alpha(\boldsymbol{\theta}) = \frac{1}{\alpha} \sum_{i=1}^n \delta_i + \gamma \sum_{i=1}^n \log(\eta_i) - \sum_{i=1}^n \log(1 + \eta_i^\gamma), \quad (3.4)$$

$$U_\gamma(\boldsymbol{\theta}) = \frac{1}{\gamma} \sum_{i=1}^n \delta_i + \alpha \sum_{i=1}^n \log(\eta_i) - \sum_{i=1}^n (\alpha + \delta_i) \frac{\log(\eta_i) \eta_i^\gamma}{1 + \eta_i^\gamma}, \quad (3.5)$$

$$U_\lambda(\boldsymbol{\theta}) = \sum_{i=1}^n \delta_i \frac{\lambda + \exp(z_i)}{\lambda \exp(z_i)} + \sum_{i=1}^n \frac{(1 + \eta_i)}{\exp(z_i)} \left( \frac{\alpha\gamma - \delta_i}{\eta_i} - \frac{\gamma \eta_i^{\gamma-1} (\alpha + \delta_i)}{1 + \eta_i^\gamma} \right), \quad (3.6)$$

$$\begin{aligned} U_{\beta_j}(\boldsymbol{\theta}) &= \sum_{i=1}^n \frac{\delta_i x_{ij}}{\sigma} (\lambda \exp(-z_i) + 1) + \sum_{i=1}^n \frac{x_{ij} \lambda (1 + \eta_i) \exp(-z_i)}{\sigma} \times \\ &\quad \times \left( \frac{\alpha\gamma - \delta_i}{\eta_i} - \frac{\gamma \eta_i^{\gamma-1} (\alpha + \delta_i)}{1 + \eta_i^\gamma} \right), \end{aligned} \quad (3.7)$$

$$\begin{aligned} U_\sigma(\boldsymbol{\theta}) &= \sum_{i=1}^n \frac{\delta_i}{\sigma} [z_i (1 + \lambda \exp(-z_i)) - 1] + \sum_{i=1}^n \frac{z_i \lambda (1 + \eta_i) \exp(-z_i)}{\sigma} \times \\ &\quad \times \left( \frac{\alpha\gamma - \delta_i}{\eta_i} - \frac{\gamma \eta_i^{\gamma-1} (\alpha + \delta_i)}{1 + \eta_i^\gamma} \right), \end{aligned} \quad (3.8)$$

em que  $\eta_i = \exp(\lambda/\exp(z_i)) - 1$ , sendo  $z_i = \frac{y_i - \sum_{j=0}^k x_{ij} \beta_j}{\sigma}$ . Os estimadores de máxima verossimilhança são obtidas através da solução do sistema de equações  $U_\alpha(\boldsymbol{\theta}) = 0$ ,  $U_\gamma(\boldsymbol{\theta}) = 0$ ,  $U_\lambda(\boldsymbol{\theta}) = 0$ ,  $U_{\beta_i}(\boldsymbol{\theta}) = 0$  e  $U_\sigma(\boldsymbol{\theta}) = 0$ . Como esse sistema não possui solução analítica fechada, utilizou-se rotinas computacionais em R (R core team, 2023) através da utilização da função *optim*, que utiliza o método de otimização de Nelder e Mead (1965) para encontrar uma solução numérica.. Assintoticamente, e sob condição dos parâmetros serem pontos interiores do espaço paramétrico tem-se que  $\hat{\boldsymbol{\theta}}$  possui distribuição Normal multivariada ( $p = 4 + k + 1$ ), sendo  $k + 1$  o número de parâmetros da regressão linear

simples; então temos

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}^{-1}(\boldsymbol{\theta})) \text{ quando } n \rightarrow \infty,$$

em que  $\mathbf{I}(\boldsymbol{\theta})$  é matriz de informação de Fisher esperada. Esse mesmo resultado pode ser observado se a matriz de informação  $\mathbf{I}(\boldsymbol{\theta})$  for substituída pela matriz de informação observada avaliada em  $\hat{\boldsymbol{\theta}}$ ,  $\mathbf{J}(\hat{\boldsymbol{\theta}})$ , sendo

$$\mathbf{J}(\hat{\boldsymbol{\theta}}) = \begin{pmatrix} \frac{d^2 \ell(\boldsymbol{\theta})}{d\alpha^2} & \frac{d^2 \ell(\boldsymbol{\theta})}{d\alpha d\gamma} & \frac{d^2 \ell(\boldsymbol{\theta})}{d\alpha d\lambda} & \frac{d^2 \ell(\boldsymbol{\theta})}{d\alpha d\beta_0} & \cdots & \frac{d^2 \ell(\boldsymbol{\theta})}{d\alpha d\beta_k} & \frac{d^2 \ell(\boldsymbol{\theta})}{d\alpha d\sigma} \\ \cdot & \frac{d^2 \ell(\boldsymbol{\theta})}{d\gamma^2} & \frac{d^2 \ell(\boldsymbol{\theta})}{d\gamma d\lambda} & \frac{d^2 \ell(\boldsymbol{\theta})}{d\gamma d\beta_0} & \cdots & \frac{d^2 \ell(\boldsymbol{\theta})}{d\gamma d\beta_k} & \frac{d^2 \ell(\boldsymbol{\theta})}{d\gamma d\sigma} \\ \cdot & \cdot & \frac{d^2 \ell(\boldsymbol{\theta})}{d\lambda^2} & \frac{d^2 \ell(\boldsymbol{\theta})}{d\lambda d\beta_0} & \cdots & \frac{d^2 \ell(\boldsymbol{\theta})}{d\lambda d\beta_k} & \frac{d^2 \ell(\boldsymbol{\theta})}{d\lambda d\sigma} \\ \cdot & \cdot & \cdot & \frac{d^2 \ell(\boldsymbol{\theta})}{d\beta_0^2} & \cdots & \frac{d^2 \ell(\boldsymbol{\theta})}{d\beta_0 d\beta_k} & \frac{d^2 \ell(\boldsymbol{\theta})}{d\beta_0 d\sigma} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \frac{d^2 \ell(\boldsymbol{\theta})}{d\beta_k^2} & \frac{d^2 \ell(\boldsymbol{\theta})}{d\beta_k d\sigma} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \frac{d^2 \ell(\boldsymbol{\theta})}{d\sigma^2} \end{pmatrix},$$

os elementos desta matriz estão descritos no Apêndice A.

### 3.2 Adequação do modelo

A análise de sensibilidade global é uma abordagem que consiste na avaliação das estimativas de máxima verossimilhança quando se descarta a  $i$ -ésima observação do conjunto de dados. Tomando as estimativas de máxima verossimilhança desconsiderando a  $i$ -ésima observação  $\hat{\boldsymbol{\theta}}_{(i)} = (\hat{\alpha}_{(i)}, \hat{\gamma}_{(i)}, \hat{\lambda}_{(i)}, \hat{\boldsymbol{\beta}}_{(i)}^\top, \hat{\sigma}_{(i)})^\top$ , a ideia básica é verificar a diferença entre  $\hat{\boldsymbol{\theta}}_{(i)}$  e  $\hat{\boldsymbol{\theta}}$ , uma vez que se a  $i$ -ésima exercer uma forte influência nas estimativas, maior será a diferença entre  $\hat{\boldsymbol{\theta}}_{(i)}$  e  $\hat{\boldsymbol{\theta}}$ . Cook (1986) propôs uma medida padronizada para avaliar a influência global das observações, dada por

$$\text{GD}_i(\boldsymbol{\theta}) = (\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}})^\top \mathbf{J}(\hat{\boldsymbol{\theta}})^{-1} (\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}}).$$

Diversas propostas de resíduos são encontradas na literatura com a finalidade de verificar os desvios das suposições do erro do modelo, veja por exemplo Klein e Moeschberger (2003) e Lawless (2011). Os resíduos de deviance são muito utilizados no contexto da análise de sobrevivência, uma vez que consideram as observações censuradas. Por meio gráfico que considera os resíduos de deviance versus os tempos observados é possível verificar adequabilidade do modelo e detectar as observações atípicas. Por definição, os

resíduos de deviance são dados por

$$r_{D_i} = \text{sinal}(r_{M_i}) \sqrt{-2(r_{M_i} + \delta_i \log(\delta_i - r_{M_i}))},$$

sendo  $r_{M_i}$  os resíduos de martingale - veja Colosimo e Giolo (2006).

### 3.3 Estudo de Simulação

Para obter as amostras do modelo de regressão cuja densidade é dada por (3.2), utilizou-se o método da transformação inversa, sendo gerado por

$$y = \sigma \left[ \log \left( \lambda \left\{ \log \left[ \left( \frac{1}{1 - (1 - \mathcal{U})^{1/\alpha}} - 1 \right)^{1/\gamma} + 1 \right] \right\}^{-1} \right) \right] + \sum_{j=0}^k x_{ij} \beta_j,$$

em que  $\mathcal{U}$  representa uma variável aleatória Uniforme(0, 1).

Simulou-se um modelo de regressão linear simples  $y_i = \beta_0 + \beta_1 x_{i1} + \sigma \nu$ , aqui fixou-se  $\beta_0 = 3$ ,  $\beta_1 = 5$  e  $\sigma = 1$  e  $\nu \sim \text{LLEI2}(\alpha, \gamma, \lambda)$  com  $\alpha = 2, \gamma = 1$  e  $\lambda = 3$ . A variável explicativa  $x_1$  foi obtida a partir da variável aleatória Uniforme(1, 3).

Foram consideradas  $N = 5000$  réplicas de Monte Carlo utilizando a semente 513776, considerando os tamanhos amostrais  $n = 50, 100, 200, 500, 1000, 2000$ , foram utilizadas as taxas de censuras 5%, 10% e 20%.

A obtenção das variáveis de censura,  $\delta_i$ , se deu a partir da variável aleatória de Bernoulli( $p$ ), em que  $p$  são as taxas de censuras. Para cada tamanho amostral e taxa de censura avaliou-se o viés e o erro quadrático médio.

Na Tabela 6 estão apresentados os resultados da simulação realizada, é possível notar que os valores do viés e erro padrão diminuíram com o aumento dos tamanhos amostrais para cada taxa de censura, indicando a propriedade de consistência do estimadores obtidos por meio de métodos numéricos.

Tabela 6 – Resultados do estudo de simulação para diferentes tamanhos amostrais e taxas de censuras considerando um modelo de regressão simples com  $\beta_0 = 3$ ,  $\beta_1 = 5$ ,  $\sigma = 1$  e  $\nu \sim \text{LLEI2}(\alpha, \gamma, \lambda)$  com parâmetros  $\alpha = 2$ ,  $\gamma = 1$  e  $\lambda = 3$

Censura	n	Viés						Erro Padrão					
		$\alpha$	$\gamma$	$\lambda$	$\beta_0$	$\beta_1$	$\sigma$	$\alpha$	$\gamma$	$\lambda$	$\beta_0$	$\beta_1$	$\sigma$
5%	50	0,589	0,136	0,429	-0,012	0,009	0,080	1,908	0,975	1,805	1,126	0,207	0,595
	100	0,299	0,080	0,508	0,052	-0,043	0,052	1,293	0,590	1,481	0,708	0,143	0,324
	200	0,275	0,063	0,453	-0,032	-0,011	0,054	1,076	0,445	1,321	0,507	0,099	0,204
	500	0,141	0,014	0,236	-0,006	0,002	0,031	0,798	0,243	0,888	0,332	0,056	0,120
	1000	0,110	0,000	0,207	0,008	-0,004	0,028	0,668	0,182	0,771	0,249	0,039	0,089
	2000	0,056	0,001	0,153	0,004	-0,004	0,018	0,597	0,146	0,656	0,214	0,029	0,071
10%	50	0,372	0,147	0,659	0,283	-0,137	0,094	2,292	0,940	2,073	1,150	0,208	0,671
	100	0,354	0,053	0,336	-0,020	0,038	0,073	1,296	0,549	1,472	0,698	0,134	0,301
	200	0,170	0,007	0,509	0,010	-0,000	0,056	0,981	0,360	1,374	0,523	0,091	0,200
	500	0,142	0,002	0,232	-0,028	0,022	0,036	0,770	0,240	0,945	0,337	0,062	0,124
	1000	-0,033	0,007	0,277	-0,031	-0,000	0,025	0,623	0,184	0,804	0,277	0,041	0,090
	2000	-0,004	-0,012	0,254	-0,012	-0,000	0,021	0,516	0,137	0,702	0,215	0,029	0,068
20%	50	0,138	0,014	0,522	0,054	0,042	0,048	1,446	0,744	2,340	1,168	0,192	0,552
	100	0,282	0,050	0,557	0,091	-0,046	0,060	1,287	0,549	1,479	0,697	0,123	0,337
	200	0,041	0,006	0,480	0,091	-0,027	0,061	0,879	0,464	1,300	0,623	0,100	0,236
	500	0,010	-0,032	0,414	-0,004	0,015	0,050	0,713	0,248	1,020	0,421	0,062	0,125
	1000	-0,051	-0,051	0,381	-0,001	0,017	0,043	0,587	0,181	0,820	0,382	0,042	0,089
	2000	-0,064	-0,054	0,370	0,029	-0,001	0,038	0,522	0,140	0,721	0,303	0,029	0,070

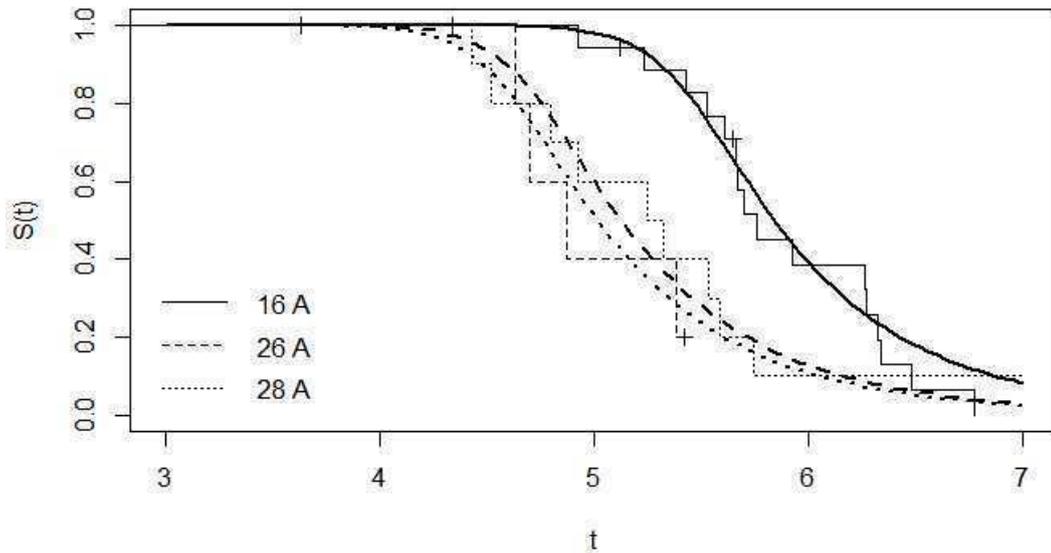
Fonte: elaborado pelo autor.

### 3.4 Aplicação

#### 3.4.1 Tempos de falhas em relés

Neste conjunto de dados ajustou-se um modelo de regressão linear simples  $y_i = \beta_0 + \beta_1 x_i$  em que  $y_i$  representa os tempos de falhas (número de ciclo em milhares) dos relés e  $x_i$  a variável explicativa que mede a intensidade de corrente elétrica (em Amperes - A) 16A, 26A e 28A. Os dados podem ser encontrados inicialmente em Nelson (2009). Em Cordeiro *et al.* (2019) tais dados são analisados utilizando o modelo de regressão Marshall-Olkin Weibull flexível, e em Cordeiro *et al.* (2017) é feito um ajuste para esses dados utilizando a família de distribuições Odd Log-Logística Generalizada. Na Figura 5 temos o ajuste do modelo LLIEL2 ao conjunto de dados e na Tabela 7 as estimativas dos parâmetros com os respectivos erros padrão.

Figura 5 – Ajuste Kaplan-Meier e para distribuição LIEL2 para os dados tempos de falhas em relés



Fonte: elaborado pelo autor.

Na Tabela 8 são apresentados os valores para os critérios de informação AIC e BIC para os modelos LLIEL2 e seus casos particulares LLEI e LEI. Os resultados mostram que o modelo LLEIL2 é preferível quanto aos demais. Além disso, através do teste da razão de verossimilhanças comparou-se os modelos LLIEL2 e LEI por meio das hipóteses  $\mathcal{H}_0 : \alpha = \gamma = 1$  e  $\mathcal{H}_1 : \alpha \neq 1$  ou  $\gamma \neq 1$ , e os modelos LLIEL2 e LEI por meio das hipóteses  $\mathcal{H}_0 : \alpha = 1$  e  $\mathcal{H}_1 : \alpha \neq 1$ . Em ambos os casos, o modelo LLIEL2 obteve resultado significativo (valor-p: 0,006 e 0,029, respectivamente).

Tabela 7 – Estimativas e Erro padrão dos estimadores de máxima verossimilhança dos parâmetros da distribuição LEIL2, considerando o conjunto de dados tempos de falhas em relés

Parâmetros	Estimativa	Erro Padrão
$\alpha$	0,21	0,007
$\gamma$	3,46	3,004
$\lambda$	2,09	0,018
$\beta_0$	8,14	1,077
$\beta_1$	-1,26	0,092
$\sigma$	0,64	0,083

Fonte: elaborado pelo autor.

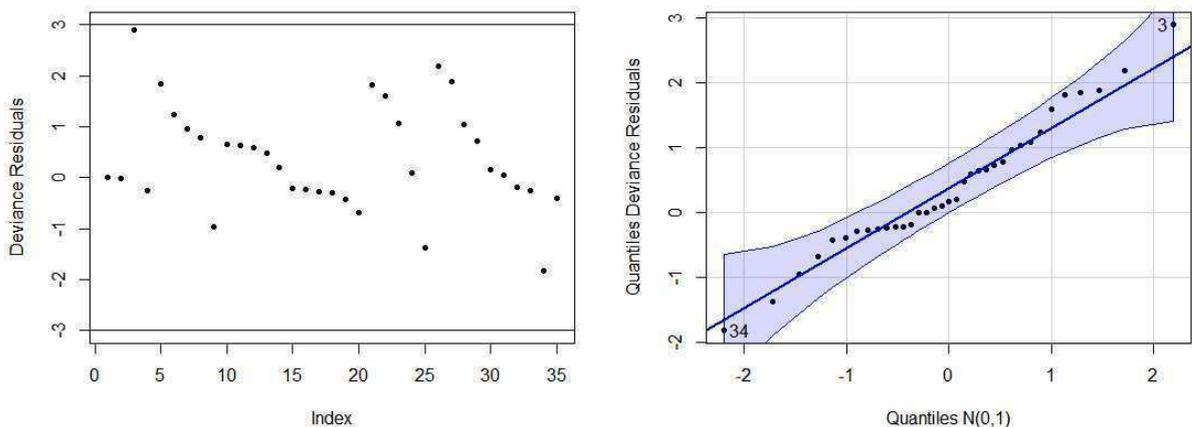
Tabela 8 – Critérios AIC e BIC para os diferentes modelos de regressão, considerando o conjunto de dados de tempos de falhas de relé

Distribuição	AIC	BIC
<b>LLEIL2</b>	<b>68,63</b>	<b>77,96</b>
LLEI	71,39	79,17
LEI	74,73	80,95

Fonte: elaborado pelo autor.

Com relação à análise de resíduos deviance, observa-se na Figura 6 que não foi identificado pontos discrepantes ao conjunto de dados. O gráfico de envelope considerando a distribuição Normal corrobora com a hipótese de que modelo LLIEL2 apresenta ajuste adequado aos dados.

Figura 6 – Resíduos deviance e gráfico de envelope Normal para os resíduos deviance do modelo LLIEL2 para os tempos de falhas em relés

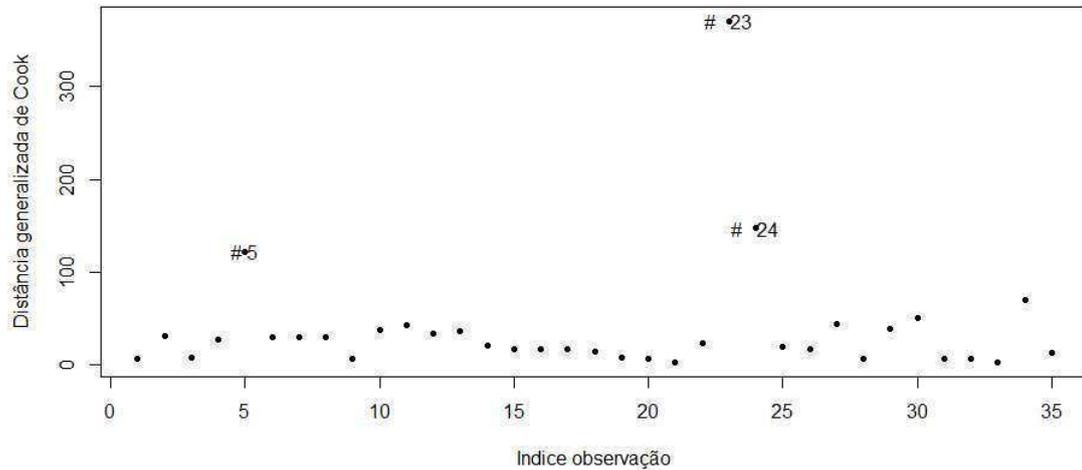


Fonte: elaborado pelo autor.

Por meio Figura 7, é possível identificar as observações 5,23 e 24 como sendo influentes nas estimativas dos parâmetros do modelo. Na Tabela 9 estão as estimativas dos parâmetros do modelo quando se exclui tais observações e na Tabela 10 a respectiva variação percentual das estimativas. Nota-se que as observações 23 e 24 exercem uma maior

influencia nas estimativas dos parâmetros da distribuição e no parâmetro de dispersão do modelo de regressão, uma vez que possuem as maiores variações percentuais.

Figura 7 – Distância de Cook generalizada para o modelo LLIEL2 considerando os dados tempos de falhas em relés



Fonte: elaborado pelo autor.

Tabela 9 – Estimativas dos parâmetros do modelo LLIEL2 quando excluídas as observações 5, 23, e 24

Observação	$\hat{\alpha}$	$\hat{\gamma}$	$\hat{\lambda}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\sigma}$
5	0,23	9,85	1,68	8,06	-1,52	1,81
23	1,75	12,29	1,37	2,86	-0,88	8,49
24	0,52	7,51	6,39	4,29	-1,64	2,67
5, 23	0,34	6,45	2,94	7,36	-1,53	1,68
5, 24	0,36	0,94	0,36	8,90	-1,21	0,26
23, 24	0,61	14,07	0,67	9,82	-1,45	5,84
5, 23, 24	0,36	3,38	1,26	9,37	-1,57	0,88

Fonte: elaborado pelo autor.

Tabela 10 – Variação percentual das estimativas dos parâmetros do modelo LLIEL2 quando excluídas as observações 5, 23, e 24

Observação	$\hat{\alpha}$	$\hat{\gamma}$	$\hat{\lambda}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\sigma}$
5	9,52%	184,68%	-19,61%	-0,98%	20,63%	182,81%
23	<b>733,33%</b>	<b>255,20%</b>	-34,44%	-64,86%	-30,15%	<b>1226,65%</b>
24	147,61%	117,05%	<b>205,74%</b>	-47,29%	30,15%	<b>317,18%</b>
5, 23	61,90%	86,41%	40,66%	9,58%	21,42%	162,50%
5, 24	71,42%	-72,83%	-82,77%	9,33%	-3,96%	-59,37%
23, 24	190,47%	<b>306,64%</b>	-67,94%	20,63%	15,07%	<b>812,50%</b>
5, 23, 24	71,42%	-2,31%	-39,71%	15,11%	24,60%	37,50%

Fonte: elaborado pelo autor.

## 4 CONSIDERAÇÕES FINAIS

Nesta dissertação foi proposta uma nova distribuição de probabilidade contínua com suporte em  $\mathbb{R}^+$  a partir da alternativa de Lehmann tipo II, denominada de Logística Exponencial Inversa de Lehmann tipo II. A nova distribuição apresenta como casos particulares as distribuições Logística Exponencial Inversa e Exponencial Inversa. Derivou-se as principais propriedades estatísticas da nova distribuição.

Além disso, obteve-se o modelo de regressão LLEIL2 para dados censurados à direita bem como as expressões para a função score e para a matriz de informação observada, uma vez que os parâmetros não possuem expressão algébrica fechada. Realizou-se também para o modelo de regressão um estudo de adequação a saber a análise de resíduos e a análise de sensibilidade global.

Através dos estudos de simulações realizados na presença de censura à direita foi verificado que as estimativas de máxima verossimilhança obteve resultados satisfatórios, uma vez que as medidas de viés e o erro padrão diminuíram com o aumento do tamanho amostral para diferentes taxas de censuras.

Nos estudos de aplicação a dados reais, além da distribuição LEIL2 e do sub-modelo LEI comparou-se com os resultados da distribuição proposta por Mansoor *et al.* (2019). Notou-se que a distribuição proposta nesta dissertação mostrou como uma boa candidata frente aos outros desenvolvimentos, o que pode ser indiciado pelos resultados dos critérios de informação e do teste de hipóteses.

Na aplicação do modelo de regressão proposto, também foi observado que o modelo LLEIL2 obteve bons resultados tanto em termos de critérios de informação como em teste de hipóteses quando comparado com seus sub-modelos. Através do estudo de adequação foi possível identificar as observações que foram mais influentes nas estimativas dos parâmetros do modelo.

Para os trabalhos futuros, ainda no contexto de dados censurados à direita, propõe-se a extensão da distribuição Logística Exponencial Inversa utilizando as alternativas de Lehmann tipo I e de Marshall-Olkin, bem como suas respectivas extensões em modelos de regressão. Outra sugestão para os trabalhos futuros é a disponibilização dos resultados obtidos nesta dissertação através de pacote no software R.

## REFERÊNCIAS

- AFIFY, A. Z. *et al.* The four-parameter burr xii distribution: properties, regression model, and applications. **Communications in Statistics-Theory and Methods**, v. 47, n. 11, p. 2605–2624, 2018. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/03610926.2016.1231821>. Acesso em: 20 abr. 2023.
- AWODUTIRE, P. O.; NDUKA, E. C.; IJOMAH, M. A. Lehmann type ii generalized half logistic distribution: properties and application. **Mathematical Theory and Modeling**, v. 10, n. 2, p. 103–115, 2020. Disponível em: <https://www.iiste.org/Journals/index.php/MTM/article/view/51765>. Acesso em: 20 abr. 2023.
- BASHEER, A. M. Marshall–olkin alpha power inverse exponential distribution: properties and applications. **Annals of Data Science**, p. 1–13, 2019. DOI: 10.1007/s40745-019-00229-0. Disponível em: <https://link.springer.com/article/10.1007/s40745-019-00229-0>. Acesso em: 20 abr. 2023.
- BÍBLIA. **Antigo e Novo Testamento**. Brasília: Tradução de João Ferreira de Almeida. Edição revista e atualizada no Brasil, 1969.
- CHAUBEY, Y. P.; ZHANG, R. An extension of chen’s family of survival distributions with bathtub shape or increasing hazard rate function. **Communications in Statistics-Theory and Methods**, v. 44, n. 19, p. 4049–4064, 2015. DOI: 10.1080/03610926.2014.997357. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/03610926.2014.997357>. Acesso em: 20 abr. 2023.
- CHAUDHARY, A. K.; KUMAR, V. Logistic inverse exponential distribution with properties and applications. **International Journal of Mathematics Trends and Technology (IJMTT)**, v. 66, n. 10, p. 151–162, 2020. DOI: 10.14445/22315373/IJMTT-V66I10P518. Disponível em: <https://ijmttjournal.org/archive/ijmtt-v66i10p518>. Acesso em: 20 abr. 2023.
- COLOSIMO, E. A.; GIOLO, S. R. **Análise de Sobrevivência Aplicada**. São Paulo: Editora Blucher, 2006.
- COOK, R. D. Assessment of local influence. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 48, n. 2, p. 133–155, 1986.
- CORDEIRO, G. M. *et al.* The generalized odd log-logistic family of distributions: properties, regression models and applications. **Journal of Statistical Computation and Simulation**, v. 87, n. 5, p. 908–932, 2017. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/00949655.2016.1238088>. Acesso em: 22 abr. 2023.
- CORDEIRO, G. M. *et al.* The marshall-olkin extended flexible weibull regression model for censored lifetime data. **Model Assisted Statistics and Applications**, v.14, n.1, p. 1–17, 2019. Disponível em: <https://content.iospress.com/articles/model-assisted-statistics-and-applications/mas180455>. Acesso em: 22 abr. 2023.

EGHWERIDO, J. T.; NZEI, L. C.; ZELIBE, S. C. The alpha power extended generalized exponential distribution. **Journal of Statistics and Management Systems**, v. 25, n. 1, p. 187–210, 2022. DOI: 10.1080/09720510.2021.1872692. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/09720510.2021.1872692>. Acesso em: 22 abr. 2023.

FALLAH, A.; KAZEMI, R. Statistical inference for the generalized weighted exponential distribution. **Communications in Statistics-Simulation and Computation**, v. 51, n. 8, p. 1–18, 2020. DOI: 10.1080/03610918.2020.1740267. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/03610918.2020.1740267>. Acesso em: 22 abr. 2023.

GEORGE, R.; THOBIAS, S. Kumaraswamy marshall-olkin exponential distribution. **Communications in Statistics-Theory and methods**, v. 48, n. 8, p. 1920–1937, 2019. DOI: 10.1080/03610926.2018.1440594. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/03610926.2018.1440594>. Acesso em: 22 abr. 2023.

IKECHUKWU, A. F.; EGHWERIDO, J. T. Transmuted shifted exponential distribution and applications. **Journal of Statistics and Management Systems**, p. 1–22, 2022. DOI: 10.1080/09720510.2021.1958517. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/09720510.2021.1958517>. Acesso em: 22 abr. 2023.

KELLER, A.; KAMATH, A.; PERERA, U. Reliability analysis of cnc machine tools. **Reliability Engineering**, v. 3, n. 6, p. 449–473, 1982.

KENNEY, J. F.; KEEPING, E. S. **Mathematics of Statistics**. NJ: VanNostrand, Princeton, v. 3, n. 1, 1962.

KLEIN, J. P.; MOESCHBERGER, M. L. **Survival Analysis: techniques for censored and truncated data**. New York, 2003. v. 1230.

KORKMAZ, M. Ç. *et al.* The weibull marshall-olkin family: regression model and application to censored data. **Communications in Statistics-Theory and Methods**, v. 48, n. 16, p. 4171–4194, 2019. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/03610926.2018.1490430>. Acesso em: 22 abr. 2023.

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

MANSOOR, M. *et al.* The marshall-olkin logistic-exponential distribution. **Communications in Statistics-Theory and Methods**, v. 48, n. 2, p. 220–234, 2019. DOI: 10.1080/03610926.2017.1414254. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/03610926.2017.1414254>. Acesso em: 22 abr. 2023.

MOORS, J. A quantile alternative for kurtosis. **Journal of the Royal Statistical Society: Series D (The Statistician)**, v. 37, n. 1, p. 25–32, 1988.

NASSAR, M. *et al.* The marshall–olkin alpha power family of distributions with applications. **Journal of Computational and Applied Mathematics**, v. 351, p. 41–53, 2019. DOI: 10.1016/j.cam.2018.10.052. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0377042718306605>. Acesso em: 22 abr. 2023.

NELDER, J. A.; MEAD, R. A simplex method for function minimization. **The Computer Journal**, The British Computer Society, v. 7, n. 4, p. 308–313, 1965.

NELSON, W. B. **Accelerated testing: statistical models, test plans, and data analysis**. Hoboken, New Jersey: John Wiley & Sons, 2009.

OGUNDE, A. *et al.* Extended gumbel type-2 distribution: properties and applications. **Journal of Applied Mathematics**, v. 2020, 2020. DOI: 10.1155/2020/2798327. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1155/2020/2798327>. Acesso em: 23 abr. 2023

OGUNDE, A. A. *et al.* Lehmann type ii frechet poisson distribution: Properties, inference and applications as a life time distribution. **International Journal of Statistics and Probability**, v. 10, n. 3, p. 1–8, 2021. DOI: 10.5539/ijsp.v10n3p8. Disponível em: <https://ccsenet.org/journal/index.php/ijsp/article/view/0/44974>. Acesso em: 23 abr. 2023.

OKORIE, I. *et al.* The adjusted log-logistic generalized exponential distribution with application to lifetime data. **International Journal of Statistics and Probability**, v. 6, n. 4, 2017. DOI: 10.5539/ijsp.v6n4p1. Disponível em: <https://ccsenet.org/journal/index.php/ijsp/article/view/68740>. Acesso em: 23 abr. 2023.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2023. Disponível em: <http://www.R-project.org/>. Acesso em: 20 abr. 2023.

RAMANAN, D. **Data World**. [S. l.], 2016. Disponível em: <https://data.world/deviramanan2016/nki-breast-cancer-data>.

RÉNYI, A. On measures of entropy and information. *In: PROCEEDINGS of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: Laboratory of the University of California, 1961. v. 4, p. 547–562.

SHANNON, C. E. A mathematical theory of communication. **The Bell System Technical Journal**, v. 27, n. 3, p. 379–423, 1948.

SOBHI, A.; MASHAIL, M. The inverse-power logistic-exponential distribution: Properties, estimation methods, and application to insurance data. **Mathematics**, v. 8, n. 11, p. 2060, 2020. DOI: 10.3390/math8112060. Disponível em: <https://www.mdpi.com/2227-7390/8/11/2060>. Acesso em: 20 abr. 2023.

THERNEAU, T. M. **A Package for Survival Analysis in R**. [S. l.], 2023. R package version 3.5-3. Disponível em: <https://CRAN.R-project.org/package=survival>. Acesso em: 20 abr. 2023.

TOMAZELLA, V. L. *et al.* The lehmann type ii inverse weibull distribution in the presence of censored data. **Communications in Statistics-Simulation and Computation**, p. 1–17, 2020. DOI: 10.1080/03610918.2020.1823000. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/03610918.2020.1823000>. Acesso em: 21 abr. 2023.

YOUSOF, H. M. *et al.* A new lifetime model with regression models, characterizations and applications. **Communications in Statistics-Simulation and Computation**, v. 48, n. 1, p. 264–286, 2019. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/03610918.2017.1377241>. Acesso em: 22 abr. 2023.

APÊNDICE A – ELEMENTOS DA MATRIZ DE INFORMAÇÃO DE  
FISHER OBSERVADA DA DISTRIBUIÇÃO LOG-LOGISTICA  
EXPONENCIAL INVERSA LEHMANN TIPO II

$$\begin{aligned}
\frac{d^2\ell(\boldsymbol{\theta})}{d\alpha^2} &= -\frac{1}{\alpha^2} \sum_{i=1}^n \delta_i, \\
\frac{d^2\ell(\boldsymbol{\theta})}{d\alpha d\gamma} &= \sum_{i=1}^n \log(\eta_i) - \sum_{i=1}^n \frac{\log(\eta_i)\eta_i^\gamma}{1 + \eta_i^\gamma}, \\
\frac{d^2\ell(\boldsymbol{\theta})}{d\alpha d\lambda} &= \sum_{i=1}^n \frac{\gamma(1 + \eta_i)}{\exp(z_i)\eta_i} - \sum_{i=1}^n \frac{\gamma\eta_i^{\gamma-1}(1 + \eta_i)}{\exp(z_i)(1 + \eta_i^\gamma)}, \\
\frac{d^2\ell(\boldsymbol{\theta})}{d\alpha d\beta_j} &= \sum_{i=1}^n \frac{x_{ij}\lambda(1 + \eta_i)\exp(-z_i)}{\sigma} \left( \frac{\gamma}{\eta_i} + \frac{\gamma\eta_i^{\gamma-1}}{1 + \eta_i^\gamma} \right), \\
\frac{d^2\ell(\boldsymbol{\theta})}{d\alpha d\sigma} &= \sum_{i=1}^n \frac{z_i\lambda(1 + \eta_i)\exp(-z_i)}{\sigma} \left( \frac{\gamma}{\eta_i} + \frac{\gamma\eta_i^{\gamma-1}}{1 + \eta_i^\gamma} \right), \\
\frac{d^2\ell(\boldsymbol{\theta})}{d\gamma^2} &= -\frac{1}{\gamma^2} \sum_{i=1}^n \delta_i - \sum_{i=1}^n \frac{(\alpha + \delta_i)\log^2(\eta_i)\eta_i^\gamma}{(1 + \eta_i^\gamma)^2}, \\
\frac{d^2\ell(\boldsymbol{\theta})}{d\gamma d\lambda} &= \sum_{i=1}^n \frac{\alpha(1 + \eta_i)}{\exp(-z_i)\eta_i} - \sum_{i=1}^n \frac{(\alpha + \delta_i)\eta_i^{\gamma-1}(1 + \eta_i)(\gamma\log(\eta_i) + \eta_i^\gamma + 1)}{\exp(-z_i)(1 + \eta_i^\gamma)^2}, \\
\frac{d^2\ell(\boldsymbol{\theta})}{d\gamma d\beta_j} &= \sum_{i=1}^n \frac{x_{ij}\lambda(1 + \eta_i)\exp(-z_i)}{\sigma} \left( \frac{\alpha}{\eta_i} + \frac{(\alpha + \delta_i)\eta_i^{\gamma-1}(\gamma\log(\eta_i) + \eta_i^\gamma + 1)}{(1 + \eta_i^\gamma)^2} \right), \\
\frac{d^2\ell(\boldsymbol{\theta})}{d\gamma d\sigma} &= \sum_{i=1}^n \frac{z_i\lambda(1 + \eta_i)\exp(-z_i)}{\sigma} \left( \frac{\alpha}{\eta_i} + \frac{(\alpha + \delta_i)\eta_i^{\gamma-1}(\gamma\log(\eta_i) + \eta_i^\gamma + 1)}{(1 + \eta_i^\gamma)^2} \right), \\
\frac{d^2\ell(\boldsymbol{\theta})}{d\lambda^2} &= -\frac{1}{\lambda^2} \sum_{i=1}^n \delta_i - \sum_{i=1}^n \frac{(\alpha\gamma - \delta_i)(1 + \eta_i)}{t_i^2\eta_i^2} - \sum_{i=1}^n \left[ \frac{(\alpha + \delta_i)\gamma\eta_i^{\gamma-2}(1 + \eta_i)}{t_i^2(1 + \eta_i^\gamma)^2} \times \right. \\
&\quad \times \left. \frac{\gamma(1 + \eta_i) - \eta_i^\gamma - 1}{t_i^2(1 + \eta_i^\gamma)^2} \right], \\
\frac{d^2\ell(\boldsymbol{\theta})}{d\lambda d\beta_j} &= \sum_{i=1}^n \frac{\delta_i x_{ij} \exp(-z_i)}{\sigma} + \sum_{i=1}^n \left[ \frac{(\alpha\gamma - \delta_i)x_{ij} \exp(-2z_i)(1 + \eta_i)}{\sigma\eta_i^2} \times \right. \\
&\quad \times \left. \frac{(1 + \eta_i)\exp(z_i) - \lambda - \exp(z_i)}{\sigma\eta_i^2} \right] \\
&\quad - \sum_{i=1}^n \frac{(\alpha + \delta_i)x_{ij}\gamma\exp(-2z_i)\eta_i^{\gamma-2}(1 + \eta_i)\{(1 + \eta_i)(\exp(z_i) + \eta_i^\gamma + \gamma\lambda + \exp(z_i))\}}{\sigma(1 + \eta_i)^2}
\end{aligned}$$

$$\begin{aligned}
\frac{d^2\ell(\boldsymbol{\theta})}{d\lambda d\sigma} &= \sum_{i=1}^n \frac{\delta_i z_i \exp(-z_i)}{\sigma} + \sum_{i=1}^n \left[ \frac{(\alpha\gamma - \delta_i) z_i \exp(-2z_i) (1 + \eta_i)}{\sigma \eta_i^2} \times \right. \\
&\times \left. \frac{(1 + \eta_i) \exp(z_i) - \lambda - \exp(z_i)}{\sigma \eta_i^2} \right] \\
&- \sum_{i=1}^n \frac{(\alpha + \delta_i) z_i \gamma \exp(-2z_i) \eta_i^{\gamma-2} (1 + \eta_i) \{ (1 + \eta_i) (\exp(z_i) + \eta_i^\gamma + \gamma\lambda + \exp(z_i)) \}}{\sigma (1 + \eta_i)^2} \\
&- \frac{\eta_i^\gamma (\lambda + \exp(z_i)) - \lambda - \exp(z_i)}{\sigma (1 + \eta_i)^2}, \\
\frac{d^2\ell(\boldsymbol{\theta})}{d\beta_j^2} &= \sum_{i=1}^n \frac{\delta_i x_{ij}^2 \lambda \exp(-z_i)}{\sigma^2} + \sum_{i=1}^n \left[ \frac{(\alpha\gamma - \delta_i) x_{ij}^2 \lambda (1 + \eta_i) \exp(-2z_i)}{\sigma^2 \eta_i^2} \times \right. \\
&\times \left. \frac{(1 + \eta_i) \exp(z_i) - \exp(z_i) - \lambda}{\sigma^2 \eta_i^2} \right] \\
&- \sum_{i=1}^n \frac{(\alpha - \delta_i) x_{ij}^2 \gamma \lambda \eta_i^{\gamma-2} (1 + \eta_i) \exp(-2z_i) \{ (1 + \eta_i) (\exp(z_i) \eta_i^\gamma + \exp(z_i) + \gamma\lambda) \}}{\sigma^2 (1 + \eta_i)^2} \\
&- \frac{\eta_i^\gamma (\exp(z_i) + \lambda) - \exp(z_i) - \lambda}{\sigma^2 (1 + \eta_i)^2}, \\
\frac{d^2\ell(\boldsymbol{\theta})}{d\beta_j d\sigma} &= - \sum_{i=1}^n \frac{\delta_i x_{ij}}{\sigma^2} + \sum_{i=1}^n \frac{\delta_i x_{ij} \lambda \exp(-z_i) (z_i - 1)}{\sigma^2} \\
&- \sum_{i=1}^n \frac{(\alpha\gamma - \delta_i) x_{ij} \lambda (1 + \eta_i) \exp(-2z_i) \{ (1 + \eta_i) \exp(z_i) (\sigma - y_i + x_{ij} \beta_j) \}}{\sigma^3 \eta_i^2} \\
&+ \frac{\exp(z_i) (y_i - x_{ij} \beta_j - \sigma) + (y_i - x_{ij} \beta_j) \lambda}{\sigma^3 \eta_i^2} \\
&+ \sum_{i=1}^n \frac{(\alpha + \delta_i) x_{ij} \gamma \lambda \exp(-2z_i) \eta_i^{\gamma-2} (1 + \eta_i) \{ (1 + \eta_i) (\exp(z_i) \eta_i^\gamma (\sigma - y_i + x_{ij} \beta_j) \}}{\sigma^3 (1 + \eta_i)^2} \\
&+ \frac{\exp(z_i) (\sigma - y_i + x_{ij} \beta_j) - (y_i - x_{ij} \beta_j) \lambda \gamma}{\sigma^3 (1 + \eta_i)^2} \\
&+ \frac{\eta_i^\gamma ((y_i - x_{ij} \beta_j - \sigma) \exp(z_i) + (y_i - x_{ij} \beta_j) \lambda)}{\sigma^3 (1 + \eta_i)^2} \\
&+ \frac{\exp(z_i) (y_i - x_{ij} \beta_j - \sigma) + (y_i - x_{ij} \beta_j) \lambda}{\sigma^3 (1 + \eta_i)^2},
\end{aligned}$$

$$\begin{aligned}
\frac{d^2\ell(\boldsymbol{\theta})}{d\sigma^2} &= \sum_{i=1}^n \frac{\delta_i \{1 - z_i(2 + \lambda \exp(z_i)(2 - z_i))\}}{\sigma^2} \\
&- \sum_{i=1}^n \frac{(\alpha\gamma - \delta_i)(y_i - x_{ij}\beta_j)\lambda(1 + \eta_i)\exp(-2z_i)\{(1 + \eta_i)\exp(z_i)(2\sigma - y_i + x_{ij}\beta_j)\}}{\sigma^4\eta_i^2} \\
&+ \frac{\exp(z_i)(y_i - x_{ij}\beta_j - 2\sigma) + (y_i - x_{ij}\beta_j)\lambda}{\sigma^4\eta_i^2} \\
&+ \sum_{i=1}^n \frac{(\alpha + \delta_i)(y_i - x_{ij}\beta_j)\gamma\lambda\exp(-2z_i)\eta_i^{\gamma-2}(1 + \eta_i)}{\sigma^4(1 + \eta_i)^2} \times \\
&\times \left\{ \frac{(1 + \eta_i)(\exp(z_i)\eta_i\gamma(2\sigma - y_i + x_{ij}\beta_j))}{\sigma^4(1 + \eta_i)^2} \right. \\
&+ \frac{\exp(z_i)(2\sigma - y_i + x_{ij}\beta_j) - (y_i - x_{ij}\beta_j)\gamma\lambda}{\sigma^4(1 + \eta_i)^2} \\
&+ \frac{\eta_i^\gamma(\exp(z_i)(y_i - x_{ij}\beta_j - 2\sigma) + (y_i - x_{ij}\beta_j)\lambda)}{\sigma^4(1 + \eta_i)^2} \\
&\left. + \frac{\exp(z_i)(y_i - x_{ij}\beta_j - 2\sigma) + (y_i - x_{ij}\beta_j)\lambda}{\sigma^4(1 + \eta_i)^2} \right\}
\end{aligned}$$

com  $\eta_i = \exp(\lambda/\exp(z_i)) - 1$ , sendo  $z_i = \frac{y_i - \sum_{j=0}^k x_{ij}\beta_j}{\sigma}$ .